Acta Cryst. (1975). A31, 499

An Application of the Molecular Replacement Technique in Direct Space to a Known Protein Structure

BY PATRICK ARGOS,* GEOFFREY C. FORD AND MICHAEL G. ROSSMANN

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, *U.S.A.*

(*Received* 16 *January* 1975; *accepted* 12 *February* 1975)

The molecular replacement technique in real space (successive cycles of electron density calculations, averaging among non-crystallographically equivalent subunits, and structure factor calculations based upon the improved averaged density) was applied to glyceraldehyde-3-phosphate dehydrogenase. It was shown possible (i) to extend the phases from a known 6.0 Å single isomorphous replacement set to at least 4.9 Å resolution; (ii) to determine a reasonable structure at low resolution given the molecular envelope and non-crystallographic symmetry; and (iii) to use the resulting low-resolution phases in the determination of heavy-atom sites. The application of such procedures to the determination of virus structures and heavy-atom derivative sites is discussed.

Introduction

Attention is progressively turning to the structure determination of subunit proteins (Matthews & Bernhard, 1973), enzyme systems (*e.g.* DeRosier & Oliver, 1971) and small viruses (Harrison, 1971; Lentz & Strandberg, 1974; Johnson, Rossmann, Smiley & Wagner, 1974). Such complex biological systems generally incorporate non-crystallographic symmetry and thus exhibit redundancy of information within the crystallographic asymmetric unit. Methods for determining the position and orientation of such systems are well established and have often been applied recently (*e.g.* Czerwinski & Mathews, 1974). The use of the redundant structural information for phase determination has not yet been satisfactorily accomplished in three dimensions for direct phase determination.

Bricogne (1974) has written a review of the molecular replacement technique outlining its theoretical development and cataloging the few limited applications that have so far been published. The method has been generally developed in reciprocal space where the computational problems have been severe. However, molecular replacement has been used in real space to improve isomorphous-replacement phases in the structural determination of low-resolution human hemoglobin (Muirhead, Cox, Mazzarella & Perutz, 1967) and more recently in the high-resolution glyceraldehyde-3-phosphate dehydrogenase (GAPDH) structure (Buehner, Ford, Moras, Olsen & Rossmann, 1974*a*). The method in real space was examined by Bricogne and shown theoretically to be equivalent to the reciprocal-space approach, as had been suggested by Rossmann (1972).

The present paper shows that low-resolution phases for the tetrameric GAPDH molecule can be determined by application of the molecular replacement technique. These phases were tested objectively by their successful use in the determination of the heavy atom sites for the GAPDH K_2HgI_4 derivative. As it is difficult to find sites in multiple subunit proteins or viruses, this provides a useful procedure linking the molecular and isomorphous-replacement techniques. That is, at low resolution the molecular replacement technique allows heavy-atom position determination in such difficult cases. The isomorphous-replacement technique can then extend the phase determination to high resolution where the molecular replacement method becomes computationally cumbersome. This paper also shows that starting with low-resolution isomorphous-replacement phases, the molecular replacement technique can extend phases to higher resolution, although the latter was computationally expensive.

Crystallographic data for lobster GAPDH

The holo-enzyme of lobster GAPDH crystallizes in the orthorhombic space group $P2_12_12_1$ with $a=149{\cdot}0$, $b=139{\cdot}1$, $c=80{\cdot}7$ Å, containing one tetramer (M.W. 143000 Dalton) per asymmetric unit (Watson & Banaszak, 1964). The orientations of the three mutually perpendicular molecular twofold axes were determined by Rossmann, Ford, Watson & Banaszak (1972) and the position of the molecule by Buehner *et al.* (1974*a*), who also defined the P,Q,R nomenclature for these axes. The structure was solved at 3·0 Å resolution utilizing single isomorphous-replacement phases with subsequent electron density averaging over the four equivalent subunits. The known structure (Buehner, Ford, Moras, Olsen & Rossmann, 1974*b*) permitted the definition of a reasonable molecular envelope which was used for all of the work reported here. At a later stage a second heavy-atom derivative was introduced (Moras *et al.*, 1975) and it is the results of this work which are taken as the 'standard of truth' here (see below).

* Present Address: Department of Physics, Southern Illinois University, Edwardsville, Illinois 62025.

Molecular replacement procedures

The method consisted of the following steps applied iteratively:

(1) The electron density, based upon the observed structure amplitudes and a given set of phases, is computed on a grid relative to the crystal axes.

(2) The electron density is re-oriented into sections of constant Q with the molecular center as origin by use of linear interpolation with a 'skew plane' program.

(3) The electron density (now referred to the molecular axes) is averaged over the points PQR, $\bar{P}\bar{Q}R$, $\bar{P}Q\bar{R}$ and $P\bar{Q}\bar{R}$. The density outside the assumed molecular envelope is set equal to the average electron density of the molecule, leaving a tetramer with 222 symmetry.

(4) The averaged density is transformed back into crystal space by sectioning in planes of constant z, and then placed into the four symmetry-related positions of the orthorhombic unit cell, with background set as before.

(5) Structure factors are recalculated with respect to the reconstructed cell. The resulting phases are then used in step one, after passing GO and collecting \$200 for computing costs (this work was done on Purdue University's CDC 6500 computer).

The above density calculations were made on grids of around 2 Å spacing both in crystal and molecular space. Such a procedure was outlined by Rossmann (1972), Buehner *et al.* (1974*a*) and Bricogne (1974).

The molecular-replacement phase refinement results could be compared at every stage with double isomorphous-replacement results (Moras *et al.*, 1975). The latter were taken as a 'standard of truth', although this is by no means absolute. When single isomorphous-replacement phases were used as a starting set, it was necessarily closely related to the double isomorphous 'standard of truth' in that the second derivative causes only small perturbations on the original set. The mean figure of merit for the double isomorphous-replacement results to 4·9 Å resolution was 0·8 (Moras *et al.* 1975), which implies a standard error in phase of 36° (Dickerson, Kendrew & Strandberg, 1961). That such an estimate of error is reasonable is made evident in the work on rubredoxin. The mean figure of merit to 2·0 Å resolution was also roughly 0·8 (Herriott, Sieker & Jensen, 1970) ,where the agreement between the 'best' isomorphous phases and calculated (*hk*0) structure factors was 39° (Watenpaugh, Sieker, Herriott & Jensen, 1973). Thus the agreement between molecular replacement phases and the double isomorphous-replacement phases will be bounded by the standard error in phase (representing the best agreement) and by 90° (representing no correlation at all).

Two criteria were used to determine the convergence of the molecular replacement (MR) results to those of the double isomorphous (DIR) data. The first was based upon the amplitudes (F):

$$R_F = \frac{\sum |(|F_{\mathrm{OBS}}| - k|F_{\mathrm{MR}}|)|}{\sum |F_{\mathrm{OBS}}|} ,$$

and is thus available for all problems.

The other was based upon the phases (α):

$$\Delta_\alpha = \frac{\sum m|(\alpha_{\mathrm{DIR}} - \alpha_{\mathrm{MR}})|}{\sum m}$$

where m is the figure of merit. This criterion was of value here in order to observe the progress of phase determination and refinement but will not in general be available for an unknown structure. The referee has pointed out that the residual Δ_α is not useful or appropriate in those regions of reciprocal space in which the starting phases are correlated with the 'true' phases. In the limiting case in which the 'true' phases and the starting phases for the refinement are identical, one has the absurdity that as the refinement phases move towards really correct values, the residual rises. This can also happen in the more realistic case in which the phases from the second heavy-atom derivative exert only a small effect upon the single isomorphous phase.

A residual using $m=1$ was found to give results closely correlated to Δ_α. As the use of a figure of merit weighting scheme seemed more reasonable it was adopted as a better measure. An estimate of the limit for the convergence of Δ_α can be made from the known distribution of the figures of merit of the double isomorphous-replacement phases according to

$$\frac{\sum m \cos^{-1} m}{\sum m}$$

which was found to be 29° for all phases to a resolution of 4·9 Å. As resolution increases this limit must also increase.

Mention must be made of the effect of the background structure. It is usual for there to be a significant amount of structure in the liquid regions between the molecules. Each new map contains evidence for this structure which would normally be neglected in the reconstitution of the crystal unit cell with the averaged molecule. To avoid losing this information (relating to almost half of the contents of the cell) the averaged cut out molecule was fitted back into the hole from which the unaveraged density was removed. The structure factors, F_B and F_M, were computed for the background density and for the reconstructed cell containing only the averaged molecules, respectively. These were then scaled together by minimizing the expression, E, for k_B, k_M and κ where

$$E = \sum [|F_{\mathrm{OBS}}| - |\{k_B F_B + k_M(F_M + \kappa)\}|]^2 .$$

In practice it was found that k_B was negligibly small with respect to k_M and κ tended to zero.

Extension of phase information from 6.0 Å to 4.9 Å resolution

The first test of the molecular replacement technique was to extend the phases of 6·0 Å resolution single isomorphous set to 4·9 Å resolution. In choosing only single isomorphous phases an attempt has been made

Table 1. *Mean weighted difference, Δ_α, of phases (°) and R value R_F, during phase extension from 6·0 to 4·9 Å resolution*

Resolution of data	Number of reflections	Single isomorphous phases	Cycle 0 (6 Å)	Cycle 1 (6 Å)	Cycle 2 (6 Å)	Cycle 3 (extension to 5·6 Å)	Cycle 4 (extension to 5·45 Å)	Cycle 5 (extension to 5·3 Å)	Cycle 6 (extension to 5·1 Å)	Cycle 7 (extension to 4·9 Å)	Cycle 8 (4·9 Å)
25·0—6·0	4284	38·1	46·5								44·4
6·0—5·9	224		76·7	70·6	67·5	62·0	59·3	55·8	54·7	53·7	53·1
5·9—5·8	206		83·8	74·0	74·3	69·5	64·5	62·7	61·8	60·8	60·4
5·8—5·7	244		84·7	73·1	75·5	69·6	62·7	59·1	58·0	57·0	57·0
5·7—5·6	253		83·6	82·0	78·4	73·4	68·8	64·0	63·0	62·3	61·5
5·6—5·5	256		90·2	84·4	85·2	77·4	72·4	67·1	66·3	66·4	66·5
5·5—5·4	260		81·0	76·5	78·9	81·8	76·2	73·3	71·1	69·6	68·4
5·4—5·3	293		82·4	82·1	81·2	79·8	81·1	80·0	78·3	76·9	5·74
5·3—5·2	286		82·8	86·7	86·7	84·1	84·1	81·4	79·6	78·9	78·0
5·2—5·1	300		89·6	88·7	89·9	82·4	79·7	80·3	78·4	77·4	77·6
5·1—5·0	310		90·0	90·0	90·4	86·9	85·5	83·1	80·9	77·7	76·3
5·0—4·9	335		89·4							85·2	83·6
4·9—4·82	347		89·1							81·6	81·2
4·82–4·74	348		91·2							84·2	84·6
4·74–4·67	365		92·4							85·7	86·4
4·67–4·60	396		90·1							85·6	85·1
R_F			0·425	0·398	0·303	0·248	0·293	0·248	0·335	0·369	0·357
Resolution limits for computing R_F (Å)			25–6	25–6	25–6	25–6	25–5·6	25–5·6	25–5·1	25–4·9	25–4·9

Notes:

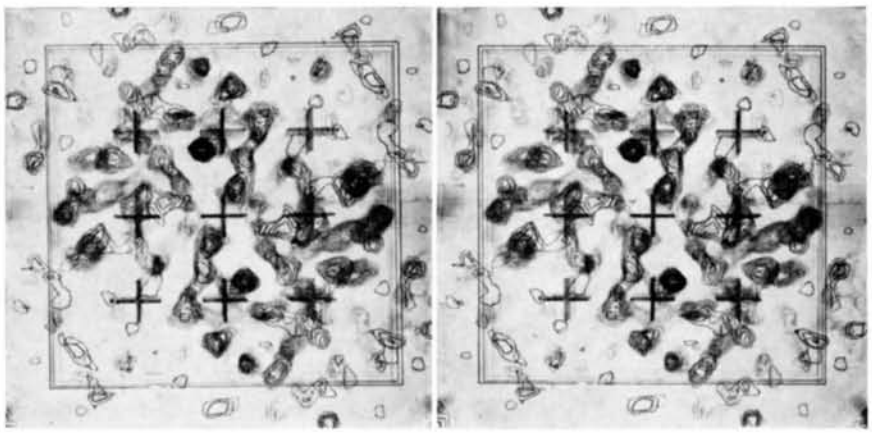
(i) All Δ_α values are calculated with respect to the double isomorphous-replacement phases. Centric reflections have been omitted.
(ii) Cycle 0 represents the unaveraged SIR electron density map at 6·0 Å resolution that was cut out and replaced into the crystal cell.
(iii) Each subsequent molecular replacement cycle uses the phases from the previous cycle only to the limit of extension shown.
(iv) In cycle 4 the calculated electron density map was weighted by figures of merit computed according to Sim (1959, 1960). This, however, did not substantially improve the technique.

Table 2. *Mean weighted differences, Δ_α (°), during ab initio phase determination*

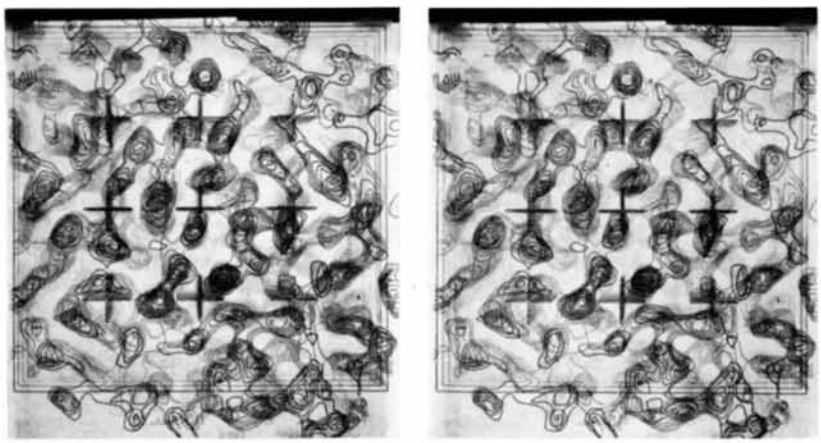
Resolution of data	Number of reflections	Cycle 0	Cycle 1 (extension to 13 Å)	Cycle 2 (extension to 11·5 Å)	Cycle 3 (extension to 9·7 Å)	Cycle 4 (9·7 Å)	Cycle 5 (9·7 Å)	Cycle 6 (9·7 Å)	Cycle 7 (extension to 8·4 Å)	Cycle 8 (extension to 8·4 Å)	Cycle 9 (6·3 Å)	Cycle 10 (6·3 Å)
21·4–18·3	21	69·2	59·9	75·0	80·2	84·0	83·6	82·8	80·4	75·2	70·5	66·5
18·3–16·0	47	59·6	62·1	71·1	67·2	65·6	65·8	65·5	66·1	64·8	61·6	59·3
16·0–14·2	74	77·6	69·6	71·2	67·9	67·1	67·2	66·6	67·3	65·4	63·4	60·3
14·2–12·7	92	79·5	72·9	71·8	72·4	71·7	70·2	68·9	69·1	68·3	67·6	66·2
12·7–11·5	106	78·3	64·3	68·1	64·9	63·2	62·1	61·3	59·2	55·7	54·4	53·8
11·5–10·6	139	80·0	85·5	77·0	73·4	69·4	65·6	63·6	61·8	60·7	59·7	58·9
10·6—9·7	180	89·8	76·3	74·4	69·0	66·1	64·3	62·9	60·6	59·0	57·9	57·3
9·7—9·0	191	91·6	82·9	80·2	75·5	73·2	72·0	70·3	68·3	67·3	65·8	65·6
9·0—8·4	243	88·4	91·0	81·4	81·6	80·9	79·0	79·1	73·8	70·7	68·6	66·8
8·4—7·9	273	87·8	83·8	87·8	87·4	86·4	87·8	88·3	82·5	79·4	77·7	77·3
7·9—7·4	314	89·0	82·1	87·0	91·3	87·7	86·5	86·5	91·4	90·0	88·6	87·5
7·4—7·0	355	89·3	84·2	81·3	81·1	80·7	81·2	82·1	85·7	86·6	86·2	85·4
7·0—6·6	398	96·7	87·4	88·6	85·6	84·1	83·7	82·4	84·6	82·5	81·1	80·3
6·6—6·3	445	87·5	91·6	88·2	89·2	88·2	86·5	85·2	83·4	85·3	85·8	86·3
6·3—6·0	502	91·2	90·5	90·0	90·5	92·9	93·8	93·2	89·8	87·1	87·8	88·7
R_F			0·424		0·452	0·456			0·448	0·517	0·444	0·416
Resolution limits for computing R_F			21·4–13·0 Å		21·4–11·0 Å	21·4–9·7 Å			21·4–8·4 Å	21·4–6·3 Å	21·4–6·3 Å	21·4–6·3 Å

Notes:

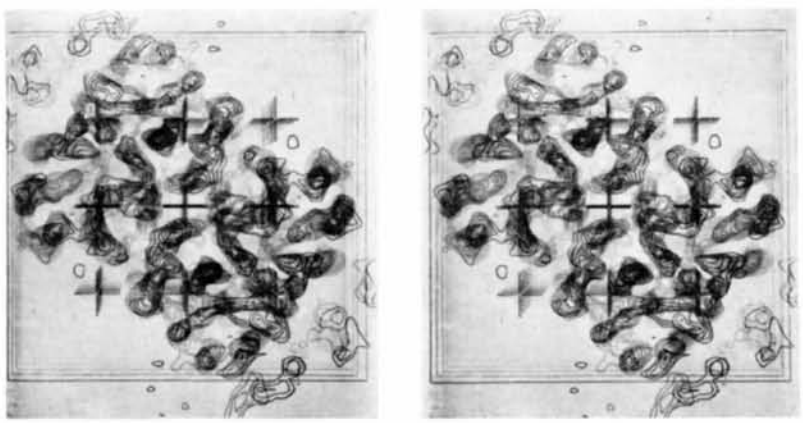
(i) All Δ_α values are calculated with respect to the double isomorphous-replacement phases. Centric phases have been omitted.
(ii) Cycle 0 represents the phases based upon the molecular envelope filled with uniform density.
(iii) Each subsequent cycle uses the phases from the previous cycle only to the limit of extension shown.
(iv) The background was set to a uniform value equal to the mean within the averaged molecule on each cycle. This value was normally close to zero.



(*a*)



(*b*)



(*c*)

Fig. 1. Stereo views of electron density maps including the coenzyme binding area computed during different steps of phase extension from 6·0 Å to 4·9 Å resolution: (*a*) the 'standard of truth' represented by an averaged multiple isomorphous-replacement map computed at 5·0 Å resolution; (*b*) the 6·0 Å resolution unaveraged single isomorphous-replacement map; and (*c*) the final averaged map after eight cycles of molecular replacement and extension to 4·9 Å resolution. Map (*a*) shows seven sections perpendicular to Q ($8 \leq Q \leq 20$ Å) separated by 2·0 Å, while maps (*b*) and (*c*) show five sections perpendicular to Q ($7{\cdot}5 \leq Q \leq 19{\cdot}5$ Å) separated by 3·0 Å. The origin for Q is at the center and crosses are separated by 20 Å. Contours are at arbitrary equal intervals. P is horizontal and R is vertical in all maps.

[*To face p.* 501

to minimize the amount of starting information. The K_2HgI_4 heavy-atom derivative data (Buehner *et al.*, 1974*a*) was used for this purpose. The process (Table 1) was one of a few iterations at the current resolution and then extension of the resolution by approximately one reciprocal-lattice point, followed by further iterations at the extended resolution. Justification that extension should proceed about one reciprocal-lattice point at a time is best seen from the reciprocal-lattice approach (Rossmann & Blow, 1963) and depends upon the ratio of the cut out volume, U, with respect to the unit cell volume, V.

As the initial agreement between all double isomorphous-replacement phases (the 'standard of truth') and the starting single isomorphous phase was $\Delta_\alpha = 31°$ within the 6 Å resolution sphere, the agreement was already as good as might be expected on the basis of the figures of merit. Thus it was not surprising that there occurred no substantial apparent improvement of Δ_α, although the final molecular replacement phases were possibly better than the initial single isomorphous set (Table 1) to 6 Å resolution. However, Table 1 does show a significant improvement of the phases on the original apparently-random set in the 6·0–4·9 Å resolution range after successive cycles and gradual extension of resolution. The major improvement of phases occurs in the extension steps, with only gradual but consistent improvement thereafter. Thus convergence to the probable limit of $\Delta_\alpha = 29°$ would require many cycles and becomes computationally expensive. An essentially similar situation is true for the centric reflections, except that these lock in almost immediately after extension of resolution with little change in subsequent cycles. In contrast, the structure amplitudes converge rapidly towards the observed values (Table 1), thus indicating that R_F is not a good measure of the degree of convergence to the 'real' structure. It can also be observed that there is greater economy in extending the resolution on every cycle (thus paying perhaps a further $50 for the extra amount of computing in order to come out of jail) rather than remaining at one resolution for two or three cycles before extension.

As an alternative measure of the quality of the phases determined by the molecular replacement technique, in the absence of an absolute 'standard of truth', electron density maps were stacked and compared visually. In Fig. 1 are shown sections from (*a*) a map at 5·0 Å resolution based upon essentially three isomorphous heavy atom derivatives and subsequent electron density averaging, (*b*) the unaveraged electron density from the starting 6·0 Å map (cycle 0 in Table 1) and (*c*) the averaged map at 4·9 Å resolution after eight cycles of molecular replacement. It can be clearly seen that the final map (*c*) is a far better approximation to the multiple-isomorphous map (*a*) than the starting SIR map (*b*). Indeed, it is probably a better approach to the real situation as judged from our 3·0 Å map (Buehner *et al.*, 1974*a*) than map (*a*) and certainly map (*b*). It is useful to observe that at the end of refinement reached in cycle eight there is very little difference in the appearance of an unaveraged map and an averaged map (Fig. 2). This is no doubt the cause of the slow convergence as averaging cannot then produce any substantial improvement.

Owing to the slow rate of convergence, the decreasing step size required for successive extensions of resolution, and the necessity of decreasing the electron density grid size to below 2 Å for better resolution, further phase determination became an overly expensive computer operation. As the structure was known already, and as the phase determination to 4·9 Å resolution had clearly been successful, it was not considered

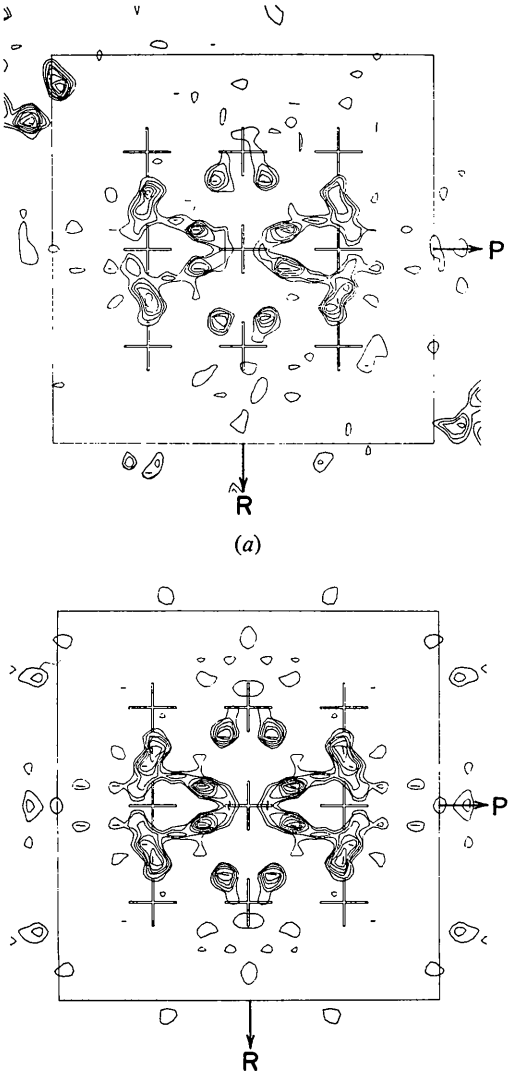


Fig. 2. Section $Q = 0$ for (*a*) the unaveraged and (*b*) the averaged electron density map after the eighth molecular replacement cycle when the phases had been extended to 4·9 Å resolution. The small differences in these maps indicate that the averaging procedure can only produce very slow convergence. Crosses are at 20 Å intervals. Contours are at arbitrary equal intervals.

worth further refinement and extension. There is, however, ample reason to think that molecular replacement can extend phases to whatever resolution is necessary given the observed data, some starting phases, a reasonable molecular envelope and the reality of the non-crystallographic equivalence.

Ab initio phase determination to 6.3 Å resolution

The second test of the molecular replacement technique was to determine phases *ab initio* using only a knowledge of the molecular envelope and the non-crystallographic symmetry. As a first approximation to the structure, the envelope was filled with uniform density. Successive steps in the phase determination are shown in Table 2. These follow essentially the same course as the phase extension results shown in Table 1, although the values of Δ_α are somewhat larger. This may be caused possibly by salt effects or difficulty in phase determination. Progress in the refinement can be seen particularly during the last three cycles, where a substantial improvement of R_F is paralleled by a slow but steady convergence in Δ_α. This again suggests that R_F is not very sensitive to the quality of the structure determination, although its refinement is a crude measure of convergence. Table 3 records the progress of the refinement of the centric reflections. Their behavior does not differ substantially from that of the non-centric phases in that sign determination has not yet converged beyond 7·9 Å resolution at the tenth cycle. The large increase in resolution accepted on the eighth cycle was clearly of little value beyond the one reciprocal-lattice point extension from 8·4 Å to 7·9 Å resolution even after the final three cycles of refinement.

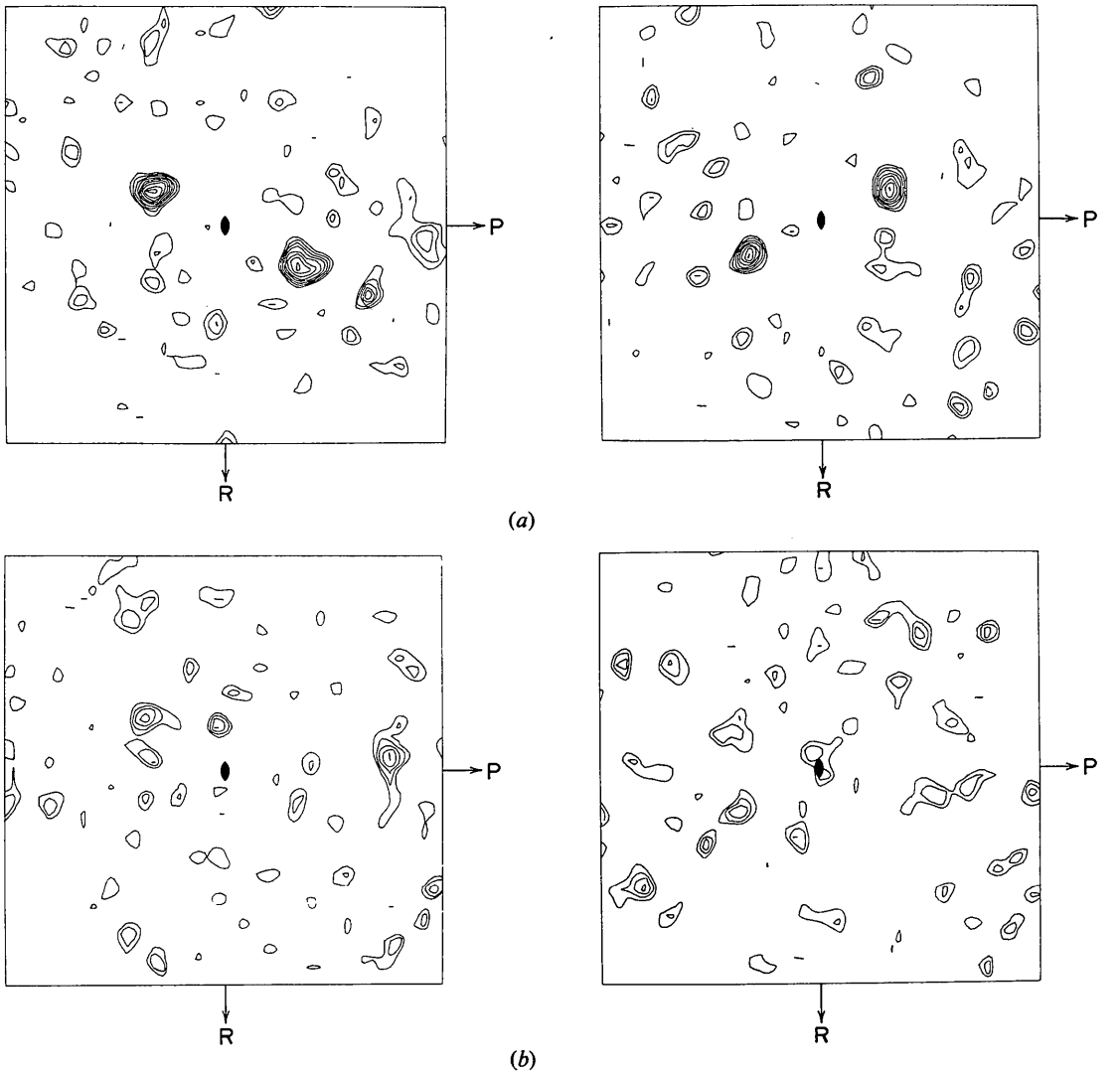


Fig. 3. Sections $Q = \pm 15$ Å for K_2HgI_4 difference Fourier maps at 6·3 Å resolution using (*a*) phases from cycle 10 and (*b*) phases from cycle 0. Sections $Q = -15$ Å (left) and $+15$ Å (right) contain the four major *A* sites. These are evident in the map based upon the improved phases from cycle 10. Contours are at arbitrary equal intervals. Each section represents a 50 Å square.

An interesting artifact observed in the initial stages of the phase determination was that the major features of electron density were along the surface of the molecular envelope as shown by Gibbs (1898, 1899). With further iterations this disappeared and molecular features started to appear. Judgement of the quality of the molecular representation is at best subjective; thus an alternative method of estimating the quality of the phase determination was used. The phases were applied to structure factor differences to observe the heavy-atom positions of the K_2HgI_4 and PCMS derivatives (Buehner *et al.*, 1974*a*).

In Figs. 3(*a*) and 4(*a*) are shown sections $Q = \pm 15$ Å and $Q = \pm 10{\cdot}5$ Å for the K_2HgI_4 difference Fourier maps with phases taken from cycle 10 (Table 2). These contain the major *A* and minor *B* heavy-atom sites, respectively. These maps show the height of the *A* site peaks to be at least twice that of background while the *B* site peaks are positive but not significantly above background. Further improvement was obtained by 222 averaging which has the same beneficial effect on the difference map as on the electron density and is similar to the suggestion made by Rossmann (1967) in reciprocal space. The averaged difference maps are shown in Figs. 5(*a*) and 6(*a*) for the $Q = -15$ Å and $Q = -10{\cdot}5$ Å sections. The *A* site [Fig. 5(*a*)] has a height of four times the maximum background peak while the minor *B* site [Fig. 6(*a*)] is 2·5 times the maximum background peak. Clearly the molecular replacement phases determined *ab initio* can be used to find heavy atoms easily even when the compound contains four major and four minor sites per crystallographic asymmetric unit. As a control, the same series of maps shown in Figs. 3(*a*)–6(*a*) were calculated for the phases

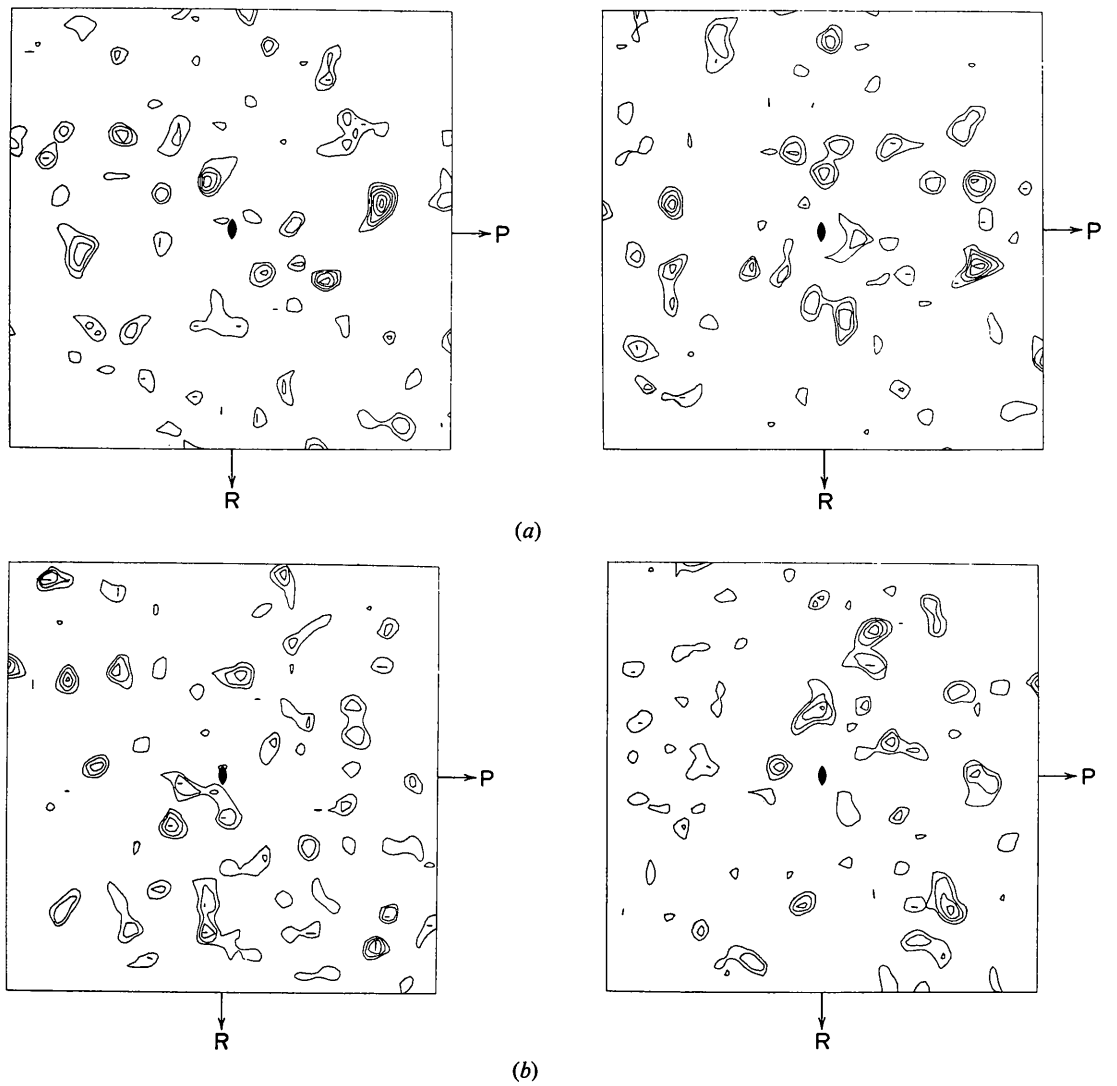


Fig. 4. Sections $Q = \pm 10{\cdot}5$ Å for the K_2HgI_4 difference Fourier maps at 6·3 Å resolution using (*a*) phases from cycle 10 and (*b*) phases from cycle 0. Sections $Q = -10{\cdot}5$ Å (left) and $+10{\cdot}5$ Å (right) contain the four minor *B* sites. Contours are at arbitrary equal intervals. Each section represents a 50 Å square.

Table 3. *Refinement of the centric reflections during the ab initio phase determination using the double isomorphous-replacement determination as a measure of right and wrong*

Resolution of data	Cycle 0		Cycle 10	
	Right	Wrong	Right	Wrong
21·4–18·3	10	7	9	8
18·3–16·0	19	7	21	5
16·0–14·2	22	10	25	7
14·2–12·7	24	13	28	9
12·7–11·5	27	13	31	9
11·5–10·6	27	23	31	19
10·6–9·7	27	26	37	16
9·7—9·0	29	23	34	18
9·0—8·4	31	29	38	22
8·4—7·9	25	35	40	20
7·9—7·4	35	33	30	38
7·4—7·0	26	44	38	32
7·0—6·6	40	37	47	30
6·6—6·3	41	38	50	29
6·3—6·0	47	36	47	36

of cycle 0 and are shown in Figs. 3(*b*)–6(*b*). It is clear that although the major site *A* does eventually show a split positive peak, it is not above background and was indeed only the fourth-highest peak on the map. There is no evidence at all for the *B* site. Similar results were obtained for the PCMS difference maps. There can thus be no doubt, on the basis of the objective criterion of heavy-atom determination as well as the subjective criterion of observing protein features within the electron density map, that the method of molecular replacement has succeeded.

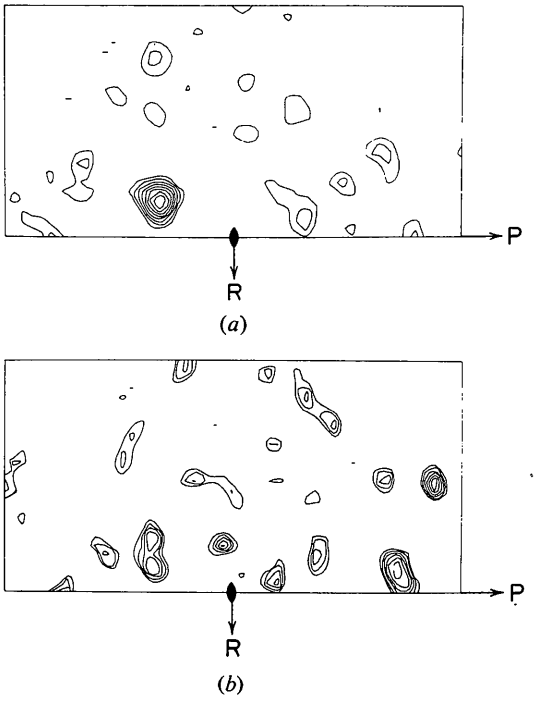


Fig. 5. Section $Q = -15$ Å through the 222 averaged K_2HgI_4 difference Fourier maps using (*a*) phases from cycle 10 and (*b*) phases from cycle 0. The major *A* site is clearly evident in the sections calculated with the improved phases taken from cycle 10. Contours are at arbitrary equal intervals. Each section represents a 50 × 25 Å rectangle.

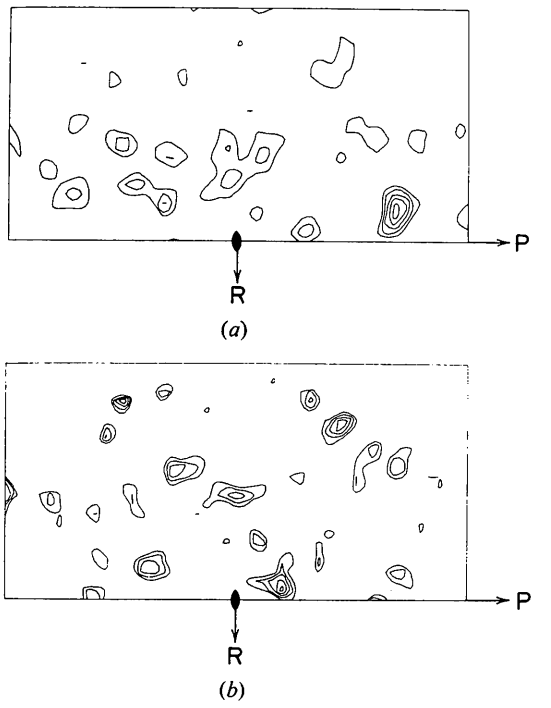


Fig. 6. Section $Q = -10{\cdot}5$ Å through the 222 averaged K_2HgI_4 difference Fourier maps using (*a*) phases from cycle 10 and (*b*) phases from cycle 0. Even the minor *B* site is now clearly evident in the maps based upon improved phases from cycle 10. Contours are at arbitrary equal intervals. Each section represents a 50 × 25 Å rectangle.

General discussion

The third test of the molecular replacement method should be phase determination without a prior knowledge of the molecular envelope. To use this technique as a direct method it would be necessary to assume a molecular shape based upon a reasonable geometric assumption. This might then be refined as the molecular details become apparent. The use of a sphere with the GAPDH data was, however, not immediately successful. In many cases, in particular for spherical viruses (*e.g.* in the structure determination of southern uean mosaic virus; Johnson *et al.*, 1974; or Tomato bushy stunt virus; Harrison, 1971) assuming a spherical envelope should be equivalent to knowing the GAPDH envelope to high resolution.

The proven ability to find heavy atoms given a reasonable envelope and data of sufficient quality will be particularly useful in the determination of virus structures. In these cases the computational problems of molecular replacement can be managed at low resolution but is probably beyond reasonable computer limits at high resolution. Furthermore, it is a difficult

problem to determine the heavy-atom positions for the highly symmetrical icosahedral particles (Argos & Rossmann, unpublished results), which becomes almost trivial by the use of molecular replacement. However, the greater redundancy of information may induce more rapid convergence, reduce the computational problem and make molecular replacement a realizable direct method.

The method of averaging the heavy-atom difference map could well be extended. Instead of extensively applying the molecular replacement technique to obtain reasonable phases for the computation of the difference map, several cycles of molecular replacement might be applied directly onto the differences. In this case, the knowledge of the molecular envelope would be less critical as long as a reasonable number of heavy atoms lie within the designated geometric shape. Furthermore, the occurrence of multiple substitutions per protein subunit could be a positive advantage! Recent results have shown this technique to be successful in determining heavy-atom positions of GAPDH when starting with a spherical envelope. The validity of the molecular replacement technique when applied to heavy-atom differences lies in the approximation that

$$f_H \exp(i\varphi_H) \simeq |(|F_{PH}| - |F_P|)| \exp(i\alpha_P)$$

where $f_H \exp(i\varphi_H)$ is the structure factor of the heavy atoms alone, $|F_{PH}|$ and $|F_P|$ are the structure amplitudes of the heavy-atom derivative and native molecule or particle, and α_P is the phase of the native structure.

An important requirement in the further development of the molecular replacement method will be to increase the rate of convergence. A number of possibilities have been tried in this work but without significant improvement. These included:

(i) the use of squared electron density and truncation of top and bottom peaks for the structure factor computation (Hoppe & Gassman, 1968; Barrett & Zwick, 1971);

(ii) parabolic rather than linear interpolations during the skew operations to avoid degradation of density; and

(iii) a reciprocal-space averaging procedure which consists of rotating the molecule about each of the non-crystallographic symmetry elements. Structure factors were then calculated in terms of the resulting electron density and summed vectorially. The resulting phases were used in the next cycle of replacement.

It seems possible that faster convergence might be achieved by use of reciprocal-space equalities.

In spite of problems associated with the definition of the envelope and rate of convergence, it should be clear that molecular replacement is likely to take an important role in the solution of subunit protein and virus structures.

G. Bricogne has independently used methods similar to those presented here for the improvement of phasas of the 2.7 Å double isomorphous-replacement map of *B. stearothermophilus* GAPDH and in the determination of 5.0 Å resolution phases of Tobacco Mosaic Virus protein crystals (private communication).

We are grateful to Dr Dino Moras for providing us with the definition of the GAPDH molecular shape, and for helpful suggestions from the Parker Brothers. We would like to thank Mrs Sharon Wilder, Mrs Chris Kraemer and Mr William Boyle for technical assistance. This work was supported by the National Science Foundation (grant #GB 29596x) and the National Institutes of Health (grant #AI 11219).

References

BARRETT, A. N. & ZWICK, M. (1971). *Acta Cryst.* A**27**, 6–11.
BRICOGNE, G. C. (1974). *Acta Cryst.* A**30**, 395–405.
BUEHNER, M., FORD, G. C., MORAS, D., OLSEN, K. W., ROSSMANN, M. G. (1974*a*). *J. Mol. Biol.* **82**, 563–585.
BUEHNER, M., FORD, G. C., MORAS, D., OLSEN, K. W., ROSSMANN, M. G. (1974*b*). *J. Mol. Biol.* **90**, 25–49.
CZERWINSKI, E. W. & MATHEWS, F. S. (1974). *J. Mol. Biol.* **86**, 49–57.
DEROSIER, D. J. & OLIVER, R. M. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 199–203.
DICKERSON, R. E., KENDREW, J. C. & STRANDBERG, B. (1961). *Acta Cryst.* **14**, 1188–1195.
GIBBS, J. W. (1898). *Nature, Lond.* **59**, 200.
GIBBS, J. W. (1899). *Nature, Lond.* **59**, 606.
HARRISON, S. C. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 495–501.
HERRIOTT, J. R., SIEKER, L. C. & JENSEN, L. H. (1970). *J. Mol. Biol.* **50**, 391–406.
HOPPE, W. & GASSMAN, J. (1968). *Acta Cryst.* B**24**, 97–107.
JOHNSON, J. E., ROSSMANN, M. G., SMILEY, I. E. & WAGNER, M. A. (1974). *J. Ultrastruct. Res.* **46**, 441–451.
LENTZ, P. J. JR & STRANDBERG, B. (1974). *Acta Cryst.* A**30**, 552–559.
MATTHEWS, B. W. & BERNHARD, S. A. (1973). *Ann. Rev. Biophys. Bioeng.* **2**, 257–312.
MUIRHEAD, H. ,COX, J. M., MAZZARELLA, C. & PERUTZ, M. F. (1967). *J. Mol. Biol.* **28**, 117–156.
MORAS, D., OLSEN, K. W., SABESAN, M. N., BUEHNER, M., FORD, G. C. & ROSSMANN, M. G. (1975). *J. Biol. Chem.* In the press.
ROSSMANN, M. G. (1967). *Acta Cryst.* **23**, 173.
ROSSMANN, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon and Breach.
ROSSMANN, M. G. & BLOW, D. M. (1963). *Acta Cryst.* **16**, 39–45.
ROSSMANN, M. G., FORD, G. C., WATSON, H. C. & BANASZAK, L. J. (1972). *J. Mol. Biol.* **64**, 237–249.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
SIM, G. A. (1960). *Acta Cryst.* **13**, 511–512.
WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1973). *Acta Cryst.* B**29**, 943–956.
WATSON, H. C. & BANASZAK, L. J. (1964). *Nature, Lond.* **204**, 918–920.