Acta Cryst. (1982). A**38**, 13–20

Phasing Electron Diffraction Amplitudes with the Molecular Replacement Method

BY MICHAEL G. ROSSMANN* AND RICHARD HENDERSON

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

(*Received* 7 *May* 1981; *accepted* 29 *June* 1981)

Abstract

The protein from purple membrane, bacteriorhodopsin, together with lipid, forms two-dimensional membranous arrays of symmetry $p3$. The protein can also be induced to reform into a planar orthorhombic lattice having at least two slightly different cell dimensions. The $p3$ and $p22_12_1$ (small cell) amplitudes have been combined with the known molecular orientation and envelope to show that it is possible to determine phases to the presently known resolution of 6 Å, with the molecular replacement method. This method was then used to extend phases beyond the present 6 Å resolution limit imposed by low-dose imaging, and to obtain a projection showing the structure of purple membrane protein at 3·3 Å resolution. Data from the $p3$ structure and both orthorhombic forms were used simultaneously for this extension. The procedure suggests a way of extending the resolution of structures based on electron diffraction data from two-dimensional arrays when the transform can be sampled as finely as desired.

Introduction

Electron diffraction data from molecules organized into two-dimensional arrays can be phased by analysis of electron-microscope images (*e.g.* Unwin & Henderson, 1975; DeRosier & Klug, 1968; Erickson & Klug, 1971). Resolution will be limited to around 30 Å by the properties of the stain in stained biological specimens. For unstained specimens, the information from a single coherent image is currently limited to about 7 Å by virtue of radiation damage, intrinsic lack of coherent long-range order and other instrumental factors such as mechanical and magnetic perturbances. Nevertheless, the diffraction data can often extend to much higher resolution as in the case of the purple-membrane protein arrays from *Halobacterium halobium* (Henderson, 1975; Henderson & Unwin, 1975).

Attempts have been made to extend phasing by a variety of methods. The isomorphous replacement method is not as useful for electron diffraction as for X-ray diffraction owing to the slower increase in potential scattering factors with atomic number compared to electron scattering factors (Dumont, Wiggins & Hayward, 1981). Attempts at combining phase information from different incoherently related portions of the electron micrograph, thereby increasing the number of unit cells in the averaging process, possibly improves phase, but in one case did not extend resolution (Crowther & Sleytr, 1977). Hayward & Stroud (1981), however, recently report for the projected structure of purple membrane a method of combining the very weak data at high resolution that is below the noise level in each of a large number of incoherently related image areas. The method gives estimates of the accuracy of the phases of the combined images which are good out to 3·7 Å resolution. Possibly this method could be extended to three-dimensional data, but in general the requirements for imaging (coherence and flatness of the specimen) are much more stringent than for electron diffraction and so will always be more limited in resolution than is the diffraction pattern. Finally, model building at atomic resolution based on only 7 Å structural data may not be sufficiently accurate to provide good enough phase information at atomic resolution (*e.g.* Engelman, Henderson, McLachlan & Wallace, 1980). This paper therefore proposes and tests the method of molecular replacement as a viable alternative for the phase determination of electron diffraction amplitudes beyond the resolution limits of images.

The molecular replacement technique depends on the availability of samples of the molecular transform at reciprocal-lattice coordinates closer than the inverse size of the molecule. This can be achieved in practice by either having more than one molecule per asymmetric unit of the unit cell or by having the same molecule in more than one crystalline array. Furthermore, in the case of arrays constrained to two-dimensional periodicity by, for instance, a lipid membrane, the transform is continuous along reciprocal-lattice lines perpendicular to the array. Hence the molecular replacement method should be useful to extend the resolution along each line.

* Present address: Department of Biological Sciences, Purdue University, Lilly Hall of Life Sciences, West Lafayette, Indiana 47907, USA.

0567-7394/82/010013-08$01.00

Two space groups have been reported for the purple membrane of *Halobacterium halobium*. The native membrane has the space group $p3$ with $a = b = 61{\cdot}8$ Å (Henderson, 1975) and a reconstituted form has the space group $p22_12_1$ with $a = 57{\cdot}6$, $b = 73{\cdot}5$ Å (Michel, Oesterhelt & Henderson, 1980). We have used the non-centric $p3$ ($hk0$) amplitudes in conjunction with the centric $p22_12_1$ ($hk0$) amplitudes, both at zero tilt, to interpolate the molecular transform in projection. The procedure was used (i) to demonstrate that, given only these amplitudes, the molecular envelope, and the non-crystallographic relationship between the two space groups, it would be possible to reproduce the image phases, and (ii) to extend the phases to 3·3 Å resolution starting with the image phases for each space group to around 6 Å resolution.

Molecular replacement

The power of the molecular replacement method to improve the accuracy of phase determination has been well demonstrated (*e.g.* Buehner, Ford, Moras, Olsen & Rossmann, 1974; Fletterick & Steitz, 1976; Champness, Bloomer, Bricogne, Butler & Klug, 1976; Harrison, Olson, Schutt, Winkler & Bricogne, 1978; Abad-Zapatero *et al.*, 1980). The procedure (Buehner *et al.*, 1974; Bricogne, 1976; Fletterick & Steitz, 1976; Johnson, 1978) consists of averaging the electron density of molecules or molecular subunits related by non-crystallographic symmetry within the same or in different crystals. The improved molecular density is used to compute structure factors for the known molecular arrangement within each cell. New electron densities can then be computed with the observed amplitudes and the presumed improved phases, and the process repeated until it converges.

The method has not been used frequently to extend phasing beyond that of the initial approximate phase determination. If this were possible then, by induction, it should be possible to determine phases *ab initio* given only the nature of the non-crystallographic symmetry and the definition of the molecular envelope. That indeed was the original aim of molecular replacement (Rossmann & Blow, 1962; Rossmann, 1972). This hope has been tested only partially (Argos, Ford & Rossmann, 1975; Johnson, Akimoto, Suck, Rayment & Rossmann, 1976; Nordman, 1980). Apart from the viability of the method itself, another obvious difficulty is the precise determination of the molecular envelope (Main & Rossmann, 1966), but this problem has been found tractable in the phase improvement of the hemagglutinin structure of influenza virus (Wilson, Skehel & Wiley, 1981).

Real-space averaging is equivalent to interpolation of the complex molecular transform sampled at discrete reciprocal-lattice points (Main & Rossmann, 1966; Main, 1967; Crowther, 1967; Bricogne, 1974). Indeed, it is easily shown that the interpolated structure factor $\mathbf{F}_{h'}$ at a non-integral reciprocal-lattice position h' generated by rotating an integral lattice point h with the nth non-crystallographic operator is given by

$$\mathbf{F}_{h'} = \frac{U}{V} \sum_{h=-\infty}^{+\infty} \sum_{n=1}^{N} \mathbf{F}_p \, \mathbf{G}_{hpn}, \qquad (1)$$

where $\mathbf{F}_p$ are the structure factors at integral reciprocal-lattice points, $\mathbf{G}_{hpn}$ is a function relating the points h and p after operating with the nth non-crystallographic symmetry operator, U is the volume of one molecule, and V is the volume of the unit cell (Rossmann & Blow, 1962).

One cycle of real-space molecular replacement is equivalent in reciprocal space to substituting the approximate phases into the right-hand side of (1) and obtaining an improved set of phases on the left-hand side. Clearly, only the terms $\mathbf{F}_p$ in the immediate environment of h' will affect the determination of $\mathbf{F}_{h'}$. Extension of phasing to higher resolution thus involves the evaluation of $\mathbf{F}_{h'}$ at a resolution only just higher than the known phases where only about half the terms in the sum (1) are known. Hence the initial determination of $\mathbf{F}_{h'}$ is approximate and will need to be refined at that resolution. In real space this implies extension of resolution in steps less than the inverse of the molecular diameter using the averaged density and then refinement at that resolution before further extension. That this procedure is in general valid can be recognized from Shannon's sampling theorem (Shannon, 1949; Brillouin, 1956). Its precise relationship to the isomorphous replacement method is considered in the Appendix. For centric reflections the method was used elegantly in the sign determination of the ($h0l$) reflections of horse oxyhemoglobin by studying the nature of the continuous molecular transform as mapped with shrinkage stages (Perutz, 1953).

The purple-membrane structure

Phases for the trigonal form (Henderson & Unwin, 1975; Unwin & Henderson, 1975; Henderson, 1975) had shown the protein molecule to have seven, roughly parallel, α-helices running more or less perpendicular to the plane of the membrane (*i.e.* parallel to the z axis). The membrane itself was about 45 Å thick. A similar result had been obtained by phasing the centric ($hk0$) projection data for the orthorhombic reconstituted arrays. Not only was the packing quite different in the latter but the molecules were alternatively up and down by virtue of the twofold screw axis in the plane of the membrane, thus removing the polar nature of a natural membrane. Combination of these two results confirmed the previously chosen molecular boundary, and

also demonstrated that the orientation of the molecules relative to the z axis was essentially the same. Any tilt greater than 3° would alter the appearance of the helices, and the molecular width in projection by more than 2 Å. Hence the molecular replacement method could be applied by combining the two structures in projection. That is, the molecular projections were assumed to be the same (Fig. 1).

During the progress of this work a slightly different orthorhombic $p22_12_1$ form was discovered (Leifer & Henderson, unpublished results). This had cell dimensions of $a = 59{\cdot}2$, $b = 76{\cdot}6$ Å with structure amplitudes similar at low resolution, but rather different at high resolution ($R = 13{\cdot}8\%$ overall in structure amplitudes). An electron density map of this structure was computed initially at 6 Å resolution on the basis of the known phases for the smaller orthorhombic cell. This density was also used for comparison with the trigonal cell as described below. The two orthorhombic forms will be referred to as the 'smaller' and 'larger' cells.

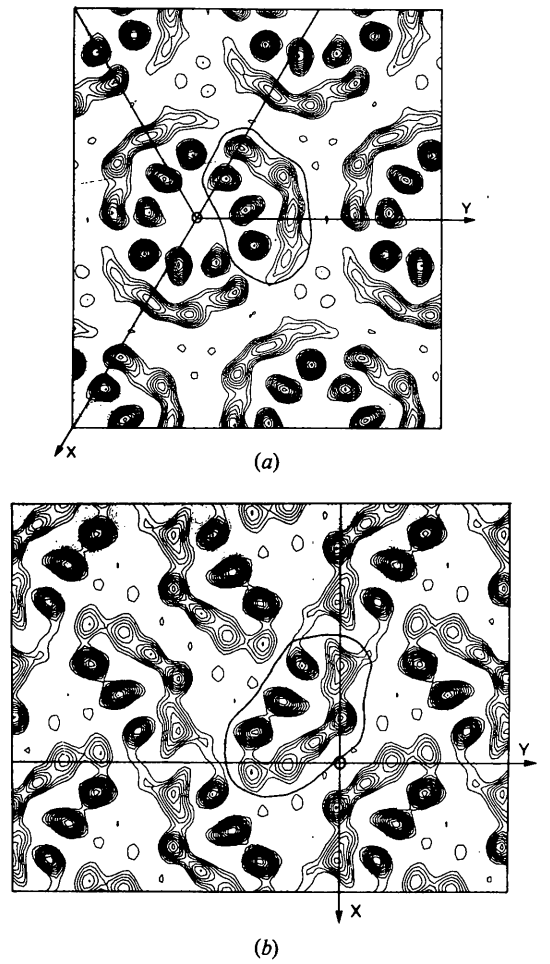


Fig. 1. Comparison of the (*a*) $p3$ and (*b*) $p22_12_1$ (smaller cell) purple-membrane structures projected down the z axis at 6 Å resolution. The maps are drawn on the same scale. The envelope used for defining a single molecule is shown in each case relative to the chosen axes which were used in the definition of the translations. The orthogonal system of axes referred to in the text has the same Y axis as marked (*i.e.* horizontal) with X vertically downwards.

The first task was to determine accurately the relative orientation and position of the molecules. This was achieved by maximizing the criterion C

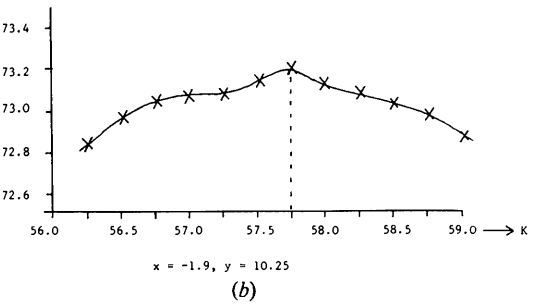
$$C = \sum_U \rho_1 \rho_2 \qquad (2)$$

where the sum is taken over all grid points (X_1, Y_1) in cell 1 within the molecular envelope U, and where ρ_1 and ρ_2 are the densities in structure 1 and structure 2 at equivalent molecular positions, respectively.

Equivalent points $(X_1, Y_1$ and $X_2, Y_2)$ in the two cells were defined by

$$X_2 = \cos\theta X_1 + \sin\theta Y_1 + d_x$$
$$Y_2 = -\sin\theta X_1 + \cos\theta Y_1 + d_y \qquad (3)$$

where θ is the amount of rotation of one structure relative to the other and d_x, d_y define the translational elements relative to the chosen origin and given molecular envelopes (Fig. 1). The density ρ_2 was determined by linear interpolation of the surrounding four grid points in structure 2. The coordinates (X,Y) and (d_x, d_y) were expressed in Å on orthogonal axial systems. Moreover, for convenience, coordinates expressed in grid points (P,Q) were used as input–output measurements within the program which evaluated C. Thus the translation parameters d_x, d_y could be defined



	9.00	9.25	9.50	9.75	10.00	10.25	10.50	10.75	11.00	11.25	11.50
-3.50	58.80	60.05	60.40	59.77	58.32	55.95	52.60	48.75	44.50	39.80	34.43
-3.25	61.31	63.13	63.88	63.74	62.85	60.69	57.71	54.08	49.96	45.35	39.78
-3.00	63.01	65.39	66.69	67.01	66.53	64.85	62.23	58.86	54.93	50.48	44.93
-2.75	64.08	66.95	68.01	69.40	69.41	68.26	66.01	63.05	59.39	55.12	49.61
-2.50	64.28	67.55	69.85	70.82	71.47	70.71	69.03	66.43	63.16	59.10	53.94
-2.25	63.61	67.44	69.98	71.58	72.58	72.32	71.46	68.92	66.17	62.42	57.44
-2.00	62.21	66.44	68.28	71.46	72.88	73.12	72.26	70.54	68.14	64.89	60.22
-1.75	60.14	64.57	67.04	70.32	72.32	73.07	72.63	71.32	69.38	66.49	62.15
-1.50	57.27	61.96	65.69	68.72	70.76	71.99	72.14	71.23	69.82	67.16	63.39
-1.25	53.72	58.68	62.46	65.46	68.24	69.96	70.45	70.12	69.24	67.00	63.65
-1.00	49.77	54.81	58.57	61.89	65.01	67.11	68.00	68.11	67.65	66.01	63.10

K = 57.50

(*a*)

Fig. 2. Exploration of the criterion C (equation 2) to determine the exact relationship between the $p3$ and the $p22_12_1$ (small cell) structures in the vicinity of the expected peak. (*a*) Exploration of P_1 and Q_1 at $\theta = 57{\cdot}5°$ with $P_2 = -14{\cdot}8$ and $Q_2 = -11{\cdot}0$ grid divisions. (*b*) Exploration of θ at $P_1 = -1{\cdot}9$, $Q_1 = 10{\cdot}25$ grid divisions.

Table 1. *Search results for determining the best relative relationships of the trigonal and orthorhombic purple membranes*

	Structure 1	Structure 2	θ (°)	P_1	Q_1	P_2	Q_2	d_x	d_y
(*a*)	*p*3	$p22_12_1$ (small)	57·8	−2·0	10·1	−14·8	−11·0	−23·4	−21·7
	*p*3	$p22_12_1$ (large)	58·7	−1·8	9·9	−14·8	−11·0	−23·3	−21·2
(*b*)	$p22_12_1$ (small)	*p*3	−57·8	3·8	1·3	−5·4	30·8	−5·5	30·6
	$p22_12_1$ (large)	*p*3	−58·7	3·6	0·7	−5·4	30·8	−5·9	31·0

Notes: (1) P_1,Q_1.P_2,Q_2 are measured on non-orthogonal crystallographic axes in 1/60th of the appropriate cell edge, and d_x, d_y in Å on an orthogonal axial system.
(2) Although the results for (*b*) can be derived from (*a*), independent searches were performed and found to be in excellent agreement.

by exploring grid points P_1, Q_1 relative to a given arbitrary point P_2, Q_2 about which rotation θ was performed.

Each cell edge was divided into 60 parts, or roughly 1 Å intervals. Bricogne (1974) shows that sampling should be at about 1/6th of the resolution although in practice 1/4th has been found sufficient. Hence the current choice was certainly ample for 6 Å resolution, sufficient for 4 Å resolution, but questionable for the final 3·3 Å resolution.

Fig. 2 shows the results for the exploration of *p*3 (structure 1) *versus* $p22_12_1$ – small cell (structure 2) in the vicinity of the expected peak. Table 1 lists the results for all comparisons and these were used in the subsequent molecular replacement experiments.

The iterative procedure

The averaged electron density was determined at grid points in each structure within the molecular boundary. The area between molecules was set equal to $\rho_{\text{mean}} - 1{\cdot}5\sigma(\rho)$, where ρ_{mean} was the mean of the averaged density and $\sigma(\rho)$ is the r.m.s. deviation from the mean of the averaged density. In practice the mean averaged density was slightly positive with a skew distribution having fewer large values and many more small values. Thus the setting of the background took into account the anticipated low mean density for the lipid between the molecules.

The electron density distribution for the entire unit cell was then transformed back to give a set of calculated structure factors. Scale factors were found in resolution ranges (n = 1, 2, *etc.*), where limits were given by $(30/n)$ Å. Linear interpolation between these factors was used for each reflection according to its interplanar spacing. R factors were computed for each range. Figures of merit were derived from phase probability curves given by

$$P(\alpha) \propto \exp[A_{\text{MR}} \cos\alpha + B_{\text{MR}} \sin\alpha],$$

where

$$A_{\text{MR}} = \left[\frac{F_o F_c}{E^2_{\text{MR}}}\right] \cos\alpha_c,$$

$$B_{\text{MR}} = \left[\frac{F_o F_c}{E^2_{\text{MR}}}\right] \sin\alpha_c,$$

and

$$E_{\text{MR}} = \left[\sum_n \frac{(F_o - F_c)^2}{n}\right]^{1/2},$$

where F_c and α_c are the calculated amplitude and phase, and n is the number of reflections. These expressions can be derived as shown by Rossmann & Blow (1963). When the image phases were also to be considered, as in the extension of phases beyond 6 Å resolution, then the above phase probability curves were multiplied (*i.e.* components were added) by the curve

$$P(\alpha) \propto \exp[A_{\text{IM}} \cos\alpha + B_{\text{IM}} \sin\alpha],$$

where

$$A_{\text{IM}} = \frac{F_o^2}{E^2_{\text{IM}}} \cos\alpha_{\text{IM}}, \quad B_{\text{IM}} = \frac{F_o^2}{E^2_{\text{IM}}} \sin\alpha_{\text{IM}}.$$

Here α_{IM} is the image-determined phase and E^2_{IM} is essentially a weighting factor relative to the molecular replacement determination.

The improved maps were then calculated with the 'best' phases and with amplitudes weighted by the figure of merit. These maps then acted as a start for the next cycle of molecular replacement. Phase extension proceeded in steps of $(1/30)$ Å^{-1}. About three cycles were required after each extension to achieve convergence, measured by no further reduction in R factors.

The *ab initio* phase determination was started by setting every grid point within the molecular envelope to an arbitrary positive constant and the space between molecules to zero.

Data

The *p*3 electron diffraction intensities were based on the averaged film records of three large fused membranes (Henderson, unpublished results) and five small un-

fused membranes. The intensities were corrected for the small amount of twinning present in patterns from fused membranes by comparison with data from unfused membranes (Henderson, unpublished results). Phases were those of the original structure determination (Unwin & Henderson, 1975; Henderson & Unwin, 1975). The $p22_12_1$ (small cell) and $p22_12_1$ (large cell) data were based on two and one films, respectively. Signs were taken as those derived by Michel *et al.* (1980).

In comparing the different cells it is essential to know cell dimensions on the same relative scale. The ratio of a/b for the orthorhombic cells is equally important, but there is no intrinsic problem in measuring this on any given photograph. It was found that this ratio had a variation of less than 1% for patterns obtained from different grids on the same day. This reproducibility permitted us to measure the relationship of the $p3$ and $p22_12_1$ large and small cells. Nevertheless, the phase determination is dependent on the accuracy of the cell dimension measurements.

Results

Twelve cycles of *ab initio* phase extension and refinement produced a set of phases (α_{MR}) for the trigonal and orthorhombic (small cell) structures which were in moderate agreement with the known phases (α_{IM}) derived from imaging. The distribution of the differences $\alpha_{MR} - \alpha_{IM}$ for the $p3$ case is shown in Table 2. The mean difference was 55°. For the orthorhombic structure 49 out of 74 signs had been determined correctly. Furthermore, the majority of signs which were in error occurred at higher resolution or were on weaker spots, that is those signs which might be most affected by higher resolution amplitudes not included here [see (1)]. The convergence of the process is shown in Table 2 for the last five cycles. A density map for the orthorhombic structure using *ab initio* determined signs is shown in Fig. 3. The map shows all the essential structural features of Fig. 1(*a*). The scale factors (Table 2) show low-order estimates of F_c as being too large, possibly indicating that there is structural information in the regions between molecules which has been obscured by the flattening process. A different constant value for the density between the molecules was found to lead to scale factors which were more (or less) dependent on resolution. Any slight differences between the molecular structure in the two crystal forms, in disagreement with our assumption that they are identical, would also be expected to give the sort of scale-factor variation with resolution that is observed.

Success with the *ab initio* structure determination encouraged us to attempt phasing of these projections to the limit of the resolution of the electron diffraction data. To aid us in this we used simultaneously the amplitudes from the $p3$, the $p22_12_1$ (small cell) and the

Table 2. *Some statistics of the ab initio phase determination during refinement after initial phases had been extended to* 6 Å *resolution*

Resolution	Number of reflections	*R* factors, Cycle number 8	9	10	11	12	Final scale factors on F_c
Space group $p3$							
30·0	2	12·8	13·1	12·3	12·3	11·8	0·50
15·0	5	23·0	24·0	23·2	20·5	20·2	0·67
10·0	11	25·1	22·1	21·7	21·5	19·9	0·95
7·5	13	46·1	25·3	17·3	14·8	13·9	1·18
6·0	18	55·6	51·1	50·3	46·0	44·9	1·38
Overall	49	38·2	29·5	26·4	24·3	23·2	
E^2_{MR}		410	303	273	253	246	
$\langle m \rangle$		0·63	0·70	0·73	0·75	0·76	
Space group $p22_12_1$ (small)							
30·0	3	34·2	38·1	33·2	28·9	30·4	0·55
15·0	11	27·6	26·1	28·3	28·0	26·4	0·70
10·0	18	22·6	19·1	18·9	17·0	17·0	1·00
7·5	27	38·0	31·1	32·7	31·4	30·7	1·16
6·0	33	50·2	37·2	31·6	31·3	32·5	1·26
Overall	92	33·8	28·0	27·7	26·5	26·2	
E^2_{MR}		350	263	245	232	237	
$\langle m \rangle$		0·63	0·69	0·72	0·73	0·73	

Agreement between α_{MR} and α_{IM} in final $p3$ phase determination (mean deviation = 55°)

$(\alpha_{MR} - \alpha_{IM})$	−180	−165	−135	−105	−75	−45	−15	+15	+45	+75	+105	+135	+165	+180
Frequency	2	0	2	2	1	6	12	3	3	0	1	2	1	

(14 reflections were phased where no α_{IM} was available)

Table 3. *Statistics on phase determination to* 3·3 Å *resolution*

	$p3$			$p22_12_1$ (small cell)			$p22_12_1$ (large cell)		
Resolution	n	R%	Scale	n	R%	Scale	n	R%	Scale
30·0	2	11·0	0·34	1	18·9	0·38	3	37·2	0·44
15·0	5	16·3	0·45	10	15·9	0·52	11	13·3	0·62
10·0	11	22·4	0·70	18	23·5	0·72	20	24·2	0·78
7·5	13	24·0	0·85	27	32·8	1·06	27	35·8	1·14
6·0	18	41·4	1·41	33	36·0	1·19	36	45·6	1·27
5·0	21	35·2	1·40	38	60·8	1·15	39	43·9	0·98
4·3	23	31·6	1·39	44	43·8	1·35	44	43·2	1·19
3·8	30	25·8	1·43	54	50·0	1·35	49	42·0	1·27
3·3	31	54·1	1·73	13	51·1	1·42	37	48·5	1·69
Overall									
E^2_{MR}		206			105			122	
$\langle m \rangle$		0·61			0·59			0·57	

$p22_12_1$ (large cell) structures. In this case the determination of the best phase and figure of merit was influenced by the image phases where these were known by using $E_{IM} = 5{\cdot}0$. (The mean value of F_o was 24·4 and therefore overwhelming weight was given to the image phases.) The final R factors and other statistics are shown in Table 3. Combination of the two centric data sets was closely similar to the use of two shrinkage states of hemoglobin (Perutz, 1953) although in our case the non-centric $p3$ data were also included. The final averaged map is shown in Fig. 4 for the $p22_12_1$ (small cell) structure. As might be anticipated when viewing seven helices end-on through 45 Å thickness, no significant additional structural information can be seen. However, there is a certain amount of flattening in the center of the three helices 5, 6, 7 (Engelman *et al.*, 1980) as well as some additional detail in the unresolved helices 1, 2, 3, 4. Some features in the lipid regions might perhaps represent the positioning of hydrocarbon chains or ordered lipid.

Conclusions

It has been shown that, under even the minimal conditions of one centric and one non-centric structure (see Appendix), it is possible to determine phases *ab initio*. That is, phase extension is practical, given a molecular envelope and a sufficient sampling of the molecular transform at intervals less than the inverse of the molecular diameter. Hence it should be possible to extend the phase information along a well-sampled reciprocal-lattice line perpendicular to a two-dimensional array.

In practice there may be difficulties when the amplitude of the transform falls below the observational level over an extended region. In the case of the purple membrane, there are large observational gaps (where accuracy of measurement is low) at around 10 Å resolution perpendicular to the membrane. It may be possible to bypass this informational gap by using the phase extension to 3·3 Å resolution in the plane of

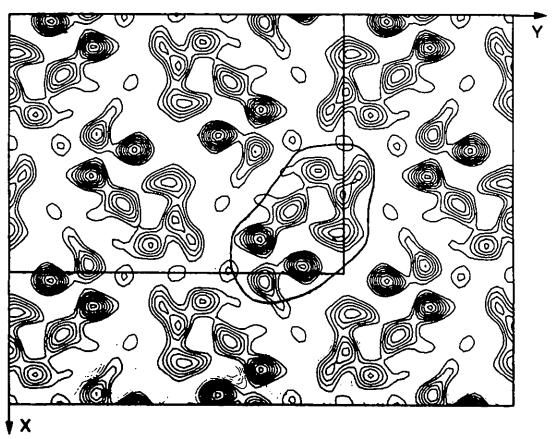


Fig. 3. The orthorhombic (small cell) purple-membrane structure determined by the *ab initio* comparison of structure amplitudes of the $p3$ and $p22_12_1$ cells. This map should be compared with that derived from image phases shown in Fig. 1(*b*).

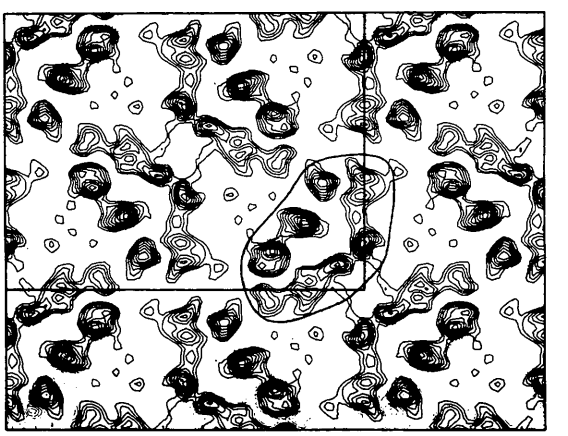


Fig. 4. The final averaged map for the $p22_12_1$ (small cell) structure. Fourier components are included to a resolution of 3·3 Å. Phases are heavily weighted towards the observed structure out to about 6 Å, and determined entirely by molecular replacement between 6 and 3·3 Å.

the membrane as shown here, and then extending the resolution in a shell at higher resolution by using two or more lattices. The intermediate lattice lines would bridge the gap from one line to another at high resolution (around 4·0 Å) where the molecular transform amplitude is moderately large. This technique might provide sufficient detail for phase improvement with a molecular model (Engelman *et al.*, 1980).

We are grateful for the $p22_12_1$ (large cell) zero-tilt data collected by Mr Dana Leifer, and for many useful discussions with Drs R. A. Crowther and J. Baldwin. One of us (MGR) was supported by a NATO grant (No. 039.80) while on sabbatical leave.

APPENDIX

BY MICHAEL G. ROSSMANN

Equivalence between molecular and isomorphous replacement

Electron diffraction of two-dimensional arrays gives rise to a continuous transform along reciprocal-lattice lines perpendicular to the array. It will be assumed that phases have been determined only at lower resolution along each line. It will now be shown that phases at higher resolution can then be determined by a construction similar to that used for phase determination of isomorphous reflections (Harker, 1956).

Let it be required that structure factors are to be determined at intervals of $1/D$ along each line, where D is greater than the thickness of the array or membrane. Let us suppose that structure factors have been determined at $p = 0, \pm1, \pm2, \ldots, \pm(n-1)$. Hence, the next structure factor which must be determined will be $\mathbf{F}_{p=n}$. Let the transform amplitudes have been sampled (by use of tilted specimens) at the non-integral lattice points h.

It can be easily shown, either from (1) or from first principles, that

$$\mathbf{F}_h = \frac{U}{V} \sum_{p=-\infty}^{+\infty} \mathbf{F}_p \, \mathbf{G}_{hp}, \qquad (A1)$$

where $\mathbf{G}_{hp} = \{\sin \pi(h-p)/\pi(h-p)\}$ for a one-dimensional transform, U is the molecular diameter, and V is the length of the unit cell.

If $n-1 < h < n$, (A1) can be written as

$$\mathbf{F}_h = \frac{U}{V}\left[\sum_{p=-(n-1)}^{+(n-1)} \mathbf{F}_p \, \mathbf{G}_{hp} + \mathbf{F}_{p=n}\,\mathbf{G}_{hn} + \sum_{p=-\infty}^{n} \mathbf{F}_p \, \mathbf{G}_{hp} + \sum_{p=n+1}^{+\infty} \mathbf{F}_p \, \mathbf{G}_{hp}\right]. \qquad (A2)$$

The first term can be completely evaluated as all phases in this range are known. Hence, set

$$\mathbf{A}_h = \frac{U}{V} \sum_{p=-(n-1)}^{+n-1} \mathbf{F}_p \, \mathbf{G}_{hp}.$$

The third term can be neglected as all the $\mathbf{G}_{hp}$ coefficients will be extremely small. The fourth term will, hopefully, also be small. The biggest term in the summation of the fourth terms will be for $p = n+1$ so that $1 < |(h-p)| < 2$. This will ensure that $\mathbf{G}_{h,n+1}/\mathbf{G}_{hn} < 1/3$. Extensive further discussion of the error introduced by neglecting the last two terms in (A2) is not considered essential for the rest of this discussion. Hence,

$$\mathbf{F}_h \simeq \mathbf{A}_h + \frac{U}{V}\,\mathbf{F}_{p=n}\,\mathbf{G}_{hn}$$

or

$$\mathbf{F}_{p=n} \simeq \frac{V}{U\mathbf{G}_{hn}}\,\mathbf{F}_h - \frac{V}{U\mathbf{G}_{hn}}\,\mathbf{A}_h. \qquad (A3)$$

The problem now is to determine the amplitude and phase of $\mathbf{F}_p$ given the known magnitude $\mathbf{F}'_h = (V|\mathbf{F}_h|)/U\mathbf{G}_{hn}$ and the known vector $\mathbf{A}'_h = -V\mathbf{A}_h/\bar{U}\mathbf{G}_{hn}$ of phase φ_h.

Equation (A3) can be represented on an Argand diagram (Fig. 5) as a vector $\mathbf{A}'_h$ and a circle of radius $\mathbf{F}'_h$ drawn at the end of the vector. Hence, the structure factor $\mathbf{F}_{p=n}$ must have its vector end somewhere on this circle. If this construction is repeated three times, then there will be a unique intersection of the circles [other than the approximation due to neglecting the high-resolution terms of (A2)].

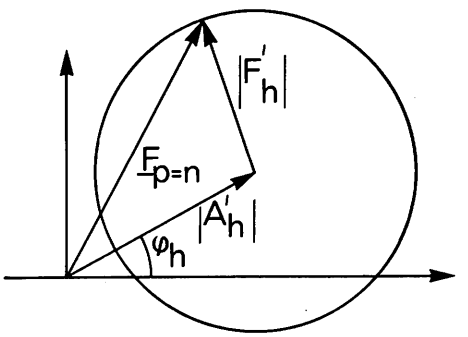


Fig. 5. Determination of the amplitude and phase of $\mathbf{F}_{p=n}$ given at least three observed amplitudes $\mathbf{F}_h$ between the known structure factor at $p = n-1$ and the unknown structure factor at $p = n$. The Argand diagram shows the circle of radius $|\mathbf{F}'_h|$ drawn at the end of the known vector $\mathbf{A}'_h$. Hence from (A3) the unknown vector $\mathbf{F}_{p=n}$ must lie somewhere on the circumference of the circle. The intersection of three circles, corresponding to three different positions h at which the transform has been sampled, determines the unknown vector $\mathbf{F}_{p=n}$.

Once an approximate phase has been determined, it can be substituted into (1) and an iterative least-squares procedure can refine the phase to give best agreement to the amplitude of all the sampled values of the transform at the non-integral values h. This process can then be repeated to extend phasing to the next position $p = n + 1$ outwards, thus also improving the previous phase determination.

It follows that the molecular transform amplitudes must be known at three or more positions within a radius of $1/D$ in order to interpolate from one reciprocal-lattice point to the next during phase extension. Should one set of points be known to be centric, then the vector $\mathbf{F}'_h$ is constrained to be real and only two intermediate points will be required. Indeed, the latter occurred in the case of *ab initio* phase determination described in this paper, showing that the conditions of the current work were the minimal for success.

It is easily shown (Main & Rossmann, 1966; Main, 1967; Bricogne, 1974) that (1) or (A1) is exactly equivalent to molecular averaging. Hence, the analysis given above in reciprocal space also applies in real space. The construction also shows that the amount of data which must be gathered to solve a structure with the molecular replacement method corresponds closely to that required by the isomorphous replacement technique.

References

ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., JOHNSON, J. E., LESLIE, A. G. W., RAYMENT, I., ROSSMANN, M. G., SUCK, D. & TSUKIHARA, T. (1980). *Nature (London)*, **286**, 33–39.

ARGOS, P., FORD, G. C. & ROSSMANN, M. G. (1975). *Acta Cryst.* A**31**, 499–506.

BRICOGNE, G. (1974). *Acta Cryst.* A**30**, 395–405.

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.

BRILLOUIN, L. (1956). *Science and Information Theory*, p. 93. New York: Academic Press.

BUEHNER, M., FORD, G. C., MORAS, D., OLSEN, K. W. & ROSSMANN, M. G. (1974). *J. Mol. Biol.* **82**, 563–585.

CHAMPNESS, J. N., BLOOMER, A. C., BRICOGNE, G., BUTLER, P. J. G. & KLUG, A. (1976). *Nature (London)*, **259**, 20–24.

CROWTHER, R. A. (1967). *Acta Cryst.* **22**, 758–764.

CROWTHER, R. A. & SLEYTR, U. B. (1977). *J. Ultrastruct. Res.* **58**, 41–49.

DEROSIER, D. J. & KLUG, A. (1968). *Nature (London)*, **217**, 130–134.

DUMONT, M. E., WIGGINS, J. W. & HAYWARD, S. B. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 2947–2951.

ENGELMAN, D. M., HENDERSON, R., MCLACHLAN, A. D. & WALLACE, B. A. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 2023–2027.

ERICKSON, H. P. & KLUG, A. (1971). *Philos. Trans. R. Soc. London Ser. B*, **261**, 105–118.

FLETTERICK, R. J. & STEITZ, T. A. (1976). *Acta Cryst.* A**32**, 125–132.

HARKER, D. (1956). *Acta Cryst.* **9**, 1–9.

HARRISON, S. C., OLSON, A. J., SCHUTT, C. E., WINKLER, F. K. & BRICOGNE, G. (1978). *Nature (London)*, **276**, 368–373.

HAYWARD, S. B. & STROUD, R. M. (1981). In the press.

HENDERSON, R. (1975). *J. Mol. Biol.* **93**, 123–138.

HENDERSON, R. & UNWIN, P. N. T. (1975). *Nature (London)*, **257**, 28–32.

JOHNSON, J. E. (1978). *Acta Cryst.* B**34**, 576–577.

JOHNSON, J. E., AKIMOTO, T., SUCK, D., RAYMENT, I. & ROSSMANN, M. G. (1976). *Virology*, **75**, 394–400.

MAIN, P. (1967). *Acta Cryst.* **23**, 50–54.

MAIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **21**, 67–72.

MICHEL, H., OESTERHELT, D. & HENDERSON, R. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 338–342.

NORDMAN, C. E. (1980). *Acta Cryst.* A**36**, 747–754.

PERUTZ, M. F. (1953). *Proc. R. Soc. London Ser. B*, **141**, 69–71.

ROSSMANN, M. G. (1972). *The Molecular Replacement Method*. New York: Gordon & Breach.

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

ROSSMANN, M. G. & BLOW, D. M. (1963). *Acta Cryst.* **16**, 39–45.

SHANNON, C. E. (1949). *Rochester Inst. Radio Eng. NY*, **37**, 10–41.

UNWIN, P. N. T. & HENDERSON, R. (1975). *J. Mol. Biol.* **94**, 425–440.

WILSON, I. A., SKEHEL, J. J. & WILEY, D. C. (1981). *Nature (London)*, **289**, 366–373.