Acta Cryst. (1996). A52, 480–489

Direct Phasing in Protein Electron Crystallography – Phase Extension and the Prospects for *Ab Initio* Determinations

DOUGLAS L. DORSET

Electron Diffraction Department, Hauptman–Woodward Medical Research Institute, Inc., 73 High Street, Buffalo, NY 14203–1196, USA. E-mail: dorset@hwi.buffalo.edu

(Received 19 August 1995; accepted 31 January 1996)

Abstract

Zonal diffraction amplitudes and crystallographic phases, derived from an averaged electron micrograph of two-dimensionally crystalline *E. coli* Omp F outer membrane porin (plane group $p31m$, $a = 72$ Å), embedded in glucose, were used as a model data set to test the feasibility of direct phase extension and *ab initio* direct phase determination. If 17 phase terms derived from *e.g.* a 10 Å (diffraction) resolution image are expanded to 6 Å by the Sayre–Hughes equation, the unknown phases are found with reasonable accuracy (mean error 43° for 25 reflections). This, however, is not the most optimal starting point. As a function of initial image resolution, the accuracy of the phase extension to 6 Å is approximately a parabolic function. That is, an optimal basis resolution, found at 11 Å (*i.e.* 14 defined reflections), produces a least mean error of 18° for 28 new reflections. In addition, *ab initio* phase determination is possible *via* a multisolution technique, using a test for density flatness as a figure of merit. The success of the determination, again, is sensitive to the size of the starting basis set generated from the permuted unknown reflections. If an annealing step is used to improve the basis set, the test for flatness will identify which reflections should be changed in phase. However, this figure of merit is not absolutely reliable for finding the exact value of the unknown phases.

1. Introduction

Electron crystallographic techniques have been extremely important for the elucidation of integral biomembrane protein structures. This is because such globular macromolecules are often most conveniently organized into two-dimensional microcrystalline arrays (*e.g.* in a phospholipid bilayer matrix). Given the enhanced scattering cross section of matter for electrons (compared to X-rays or neutrons), electron diffraction is the only feasible technique for obtaining unit-cell dimensions and symmetry information from single crystals, as well as well resolved intensities of individual reflections. Since electron crystallography is also an optical method, electron micrographs have been a major source of crystallographic phases for such structure analyses, based on the averaging of crystalline lattices by Fourier filtration or cross-correlation techniques (Amos, Henderson & Unwin, 1982; Henderson, Baldwin, Downing, Lepault & Zemlin, 1986). [Patterson search techniques, *e.g.* molecular replacement, have been much less frequently exploited (Rossman & Henderson, 1982; Earnest, Walian, Gehring & Jap, 1992).] Although initial work on protein microcrystals was carried out mainly on negatively stained preparations, higher-resolution information is nowadays obtained from unstained preparations, where the aqueous environment is either frozen into a vitreous ice or replaced by a non-volatile hydrogen-bonding substance, such as a saccharide.

Although the analysis of low-contrast unstained micrographs is important for the visualization of the native protein structures, there are challenges to this method for obtaining crystallographic phases that become more critical as the desired resolution increases (Henderson *et al.*, 1986). For example, radiation damage, induced by the inelastic interaction of the incident beam with the sample, becomes more problematic as more and more pixels are needed to resolve a structural detail, even though the crystallographic repeat allows a somewhat noisy image, collected by so-called 'low-dose techniques', to be averaged. The actual phase-contrast transfer function of the electron-microscope objective lens is also difficult to define exactly for a given experimental image, especially at higher spatial frequencies, when it is recorded at low beam doses, even though approximate methods have been found to retrieve it (Unwin & Henderson, 1975; Li, 1991). Finally, reconstitution of a membrane protein in a lipid matrix often induces a certain amount of paracrystalline disorder to the specimen, *i.e.* the space lattice is locally curved. Thus, the electron diffraction pattern from a sample may be found to extend out to 3 Å resolution or better, while the optical transform of the recorded micrograph will be observed to a lower resolution limit, *e.g.* from 10 to 6 Å (Henderson *et al.*, 1986). Is there some way, therefore, to use the lower-resolution image information, most conveniently obtained from the electron microscope, to

extend directly to the resolution limit of the electron diffraction pattern? Are there even suitable methods for determining crystallographic phases directly for thin macromolecular crystals without recourse to electron micrographs?

In response to the first question, Gilmore, Shankland & Fryer (1993) pioneered the application of direct methods to phase extension for such protein structures. In this work, they improved the resolution of 15 Å image-derived phases from bacteriorhodopsin (Henderson *et al.*, 1986) to better than 3 Å electron diffraction resolution by employing maximum-entropy and likelihood methods. More recently, this laboratory (Dorset, Kopp, Fryer & Tivol, 1995) has investigated the utility of more simplistic approaches to such phase extension, successfully applying the Sayre (1952) equation to data sets from bacteriorhodopsin and halorhodopsin (Havelka, Henderson, Heymann & Oesterhelt, 1993), respectively expanding image-derived phases in 10 and 15 Å resolution basis sets. The prospect for *ab initio* determinations has also been explored (Dorset, 1995) successfully with the latter structure, for which the projection is centrosymmetric. The phase determination employed a multisolution approach, also *via* the Sayre (1952) equation. The optimal structure solution was identified by a figure of merit used in early X-ray studies of amphiphile mesophases (Luzzati, Tardieu & Taupin, 1972), later proposed for macromolecules (Luzzati, Mariani & Delacroix, 1988).

It might be argued from the preliminary work that, since the two structures examined both contain a large amount of α-helix (projecting, moreover, down the helical axes), a certain amount of 'pseudo-atomicity' [*i.e.* the Fourier transform of a Gaussian 'glob' (Harker, 1953)] might presuppose a favorable outcome for extensions with the Sayre equation. In this paper, a protein with predominately β-sheet structure is examined, *i.e.* the Omp F porin from the outer membrane of *E. coli* (Sass *et al.*, 1989), to test the generality of procedures developed so far.

2. Materials and methods

2.1. *Experimental data*

Experimental image data (amplitudes and phases in the Fourier transform) were obtained from Dr H. J. Sass, who supervised the earlier analysis of this structure from low-dose high-resolution (3.2 Å) electron micrographs of the reconstituted protein, embedded in glucose. The original data were obtained at 100 kV and 5 K on the Suleika electron microscope with a liquid-helium-cooled superconducting lens at the Fritz Haber Institut in Berlin. There were two forms of the protein observed in this original work (Sass *et al.*, 1989; Sass, 1990), one in the single membrane layer with projected $p3$ symmetry and another in an aligned stack of two membranes with plane-group symmetry $p31m$. In both cases, the hexagonal unit-cell constant was $a = 72$ Å.

The data provided by Dr Sass were among the sets used for determination of the projected structure by image analysis with the computer program package *IMAGIC* (Van Heel & Keegstra, 1981). In order to assess the practical resolution limits (Glaeser & Downing, 1992) that would be suitable for direct phase determination/extension, mean structure-factor amplitudes, found from the image transform and averaged in resolution shells, were plotted as a function of median reciprocal spacing for each shell (Fig. 1). Although the diffraction resolution extends formally to 3.2 Å, most of the information content is expressed within a 6 Å limit, where most of the diffracted energy is found.

An image was generated at this lower 6 Å diffraction limit assuming $p1$ plane-group symmetry. It was then re-scanned and analyzed using the image-averaging program package *CRISP* (Hovmöller, 1992), which permits any possible projection symmetry to be assessed after a suitable shift is made to an allowable plane-group origin. Under these conditions, while the lowest phase residual (19.3°) was found for plane group $p3$, a nearly equivalent value (26.1°) was found for plane group $p31m$, while a significantly poorer residual (43.7°) was detected, for example, when plane-group symmetry $p3m1$ was imposed. The potential map for the $p31m$ structure (Fig. 2) closely resembled the average image obtained for the double membrane stack (Sass *et al.*, 1989). Thus, the derived phases and amplitudes from the structure in plane group $p31m$ were used as a model set of 42 unique reflections for the phase determinations described below. From the structure-factor amplitudes,

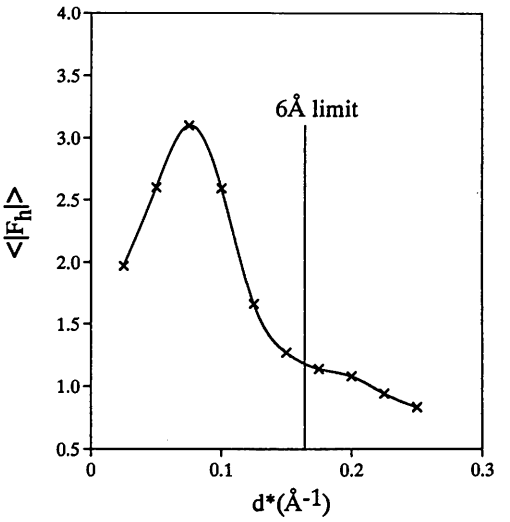


Fig. 1. Average structure-factor amplitudes $|\overline{F_h}|$ for increasing overlapping shells of diffraction resolution centered at d^*. Although the nominal resolution of the diffraction data (image transform) is found to be near 3.2 Å, most of the structural detail can be observed at 6 Å, coincident with the most intense part of the transform.

normalized values were calculated in the way usually carried out for protein data (Blundell & Johnson, 1976), *i.e.* $|E_{\mathbf{h}}| = |F_{\mathbf{h}}|/[\overline{F_{\mathbf{h}}^2}]^{1/2}$. An approximate value for $E_{000} = N^{1/2}$ was found from the molecular weight of the porin monomer (Rosenbusch, 1974), *i.e.* 36500 daltons, multiplied by 3, assuming that the average heavy atom is carbon. As shown earlier (Dorset, Kopp, Fryer & Tivol, 1995), this value need not be extremely accurate and is used only to stabilize the basis-phase-set values used in the Sayre equation. [If it is omitted, the phase values in the basis set are rapidly changed by the convolution since the condition of positivity must be imposed (Sayre, 1952).] Among the factors that would contribute to the inaccuracy of the E_{000} estimate are the limited resolution of the data set and the presence of only zonal data (hence an inaccurate scaling of the normalized structure factors), as well as the non-inclusion of solvent (or lipid) molecular scattering contributions.

2.2. *Phase determination*

In initial tests, it was assumed that an electron micrograph, obtained at a specified diffraction resolution, could be averaged to yield a set of useful crystallographic phases that could be used as a basis for resolution enhancement. The phases were then extended by means of the Sayre–Hughes equation (Hughes, 1953; Sayre, 1980): $E_{\mathbf{h}} = N^{1/2}\langle E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}\rangle_{\mathbf{k}}$. A mean phase error $\langle|\Delta\varphi|\rangle = \langle|\varphi_{\text{Sayre}} - \varphi_{\text{image}}|\rangle$ was then calculated, to evaluate the fit of the overall phase set as well as the newly found terms from the Sayre expansion to the original phases derived from the electron micrograph. (Thus, φ_{image} denotes the phase of any reflection obtained from the Fourier transform of the averaged highest-resolution micrograph after translation to an allowed unit-cell origin.)

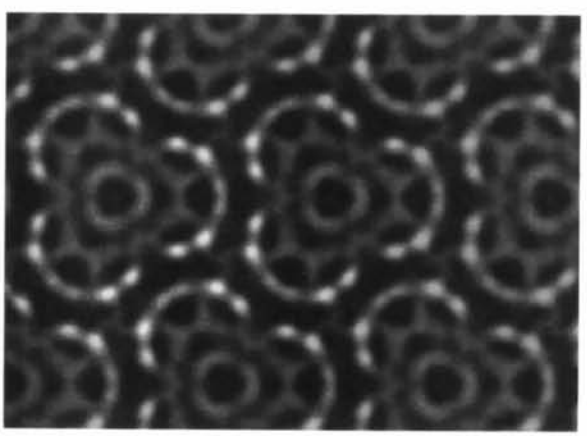


Fig. 2. Average structure of Omp F porin at 6 Å diffraction resolution when $p31m$ symmetry is imposed. (Here, positive density is represented by white areas.)

For *ab initio* phase determinations, a basis resolution, equivalent to the limit used for two of the modeled image-derived phase sets, was defined. Reflections with suitably high $|E_{\mathbf{h}}|$ values within this low-resolution set were then assigned symbolic values. Permutation of algebraic unknowns was required for the plane group since all $hk0$ reflections are seminvariants (Rogers, 1980), allowing no origin-defining values to be specified *a priori*. If the reflections (*i.e.* the axial $h00$ values) had centrosymmetric phase values, the symbolic value was then permuted through $0,\pi$. When a general, non-centrosymmetric, $hk0$ term was chosen, it was cycled through $\pi/4 + n(\pi/2)$ around the four quadrants of phase space. From the initial phase sets found after this generation of multiple trials, expansions were then made with the Sayre–Hughes equation to fill in all reflections to the resolution limit of the defined basis set. [Unlike the previous experience with halorhodopsin (Dorset, 1995), all phases were defined after just one convolution cycle.] Possible solutions, generated from the permuted starting phase terms, were sought in the multiple list after calculating all possible potential maps and testing the flatness of the density distribution. This test was made using the figure of merit suggested by Luzzati, Tardieu & Taupin (1972), *i.e.* $q = \langle\Delta\rho^4\rangle$, where $\Delta\rho = \rho - \rho_{\text{avg}}$. When the F_{000} term is set to 0.0 for calculation of the potential map, $\rho_{\text{avg}} = 0.0$. The sampling of pixels in potential maps generated from the phase sets was well within the Nyquist limit (see Gaskill, 1978). Phases of the most intense reflections in this built-up lower-resolution set could then be annealed. A sequence of $|E_{\mathbf{h}}|$ values was so treated, starting from highest to lowest magnitudes. For centrosymmetric phases, the consequence of adding π to the generated value was evaluated by observing the flatness of the associated potential map; for non-centrosymmetric phases, the value a given by the Sayre equation was cycled, again through n $(\pi/2)$ (where $n = 1, 2, 3, 4$) for the same test. After derivation of a stable low-resolution set, it was then expanded *via* the Sayre equation as before for image-derived phases.

In none of this work was the suitability of density modification considered for further phase refinement. While this may be a desirable goal, the sole object of this study was merely to evaluate the accuracy of direct methods themselves for predicting new phases.

3. Results

3.1. *Expansions from a 10 Å basis*

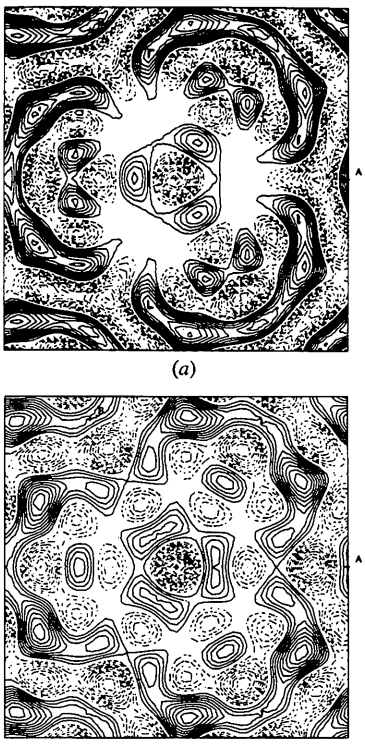
A potential map, based just on image-derived phases at 10 Å resolution, is shown in Fig. 3(*a*). The ideal potential map at 6 Å resolution, based on all 42 unique reflections, is shown in Fig. 4(*a*). Much of the structural information is contained in the phases of the 17 most intense reflections, as shown by Fig. 4(*b*) but, since

most of the diffracted intensity is found at the lower resolution, many details are also visible in Fig. 3(*a*).

Starting with 17 unique phases to 10 Å resolution, the Sayre–Hughes expansion yielded a reasonably accurate prediction of new terms, as outlined in Table 1, which was in accord with the results found earlier for bacteriorhodopsin or halorhodopsin to the 6 Å limit (Dorset, Kopp, Fryer & Tivol, 1995). If, on the other hand, an incorrect *p*3 symmetry was assumed for the projection, the extension would not produce accurate phase estimates. Certainly, this serves as an illustration of the well known fact that choice of the correct symmetry plays a significant role in the accuracy of the structure determination, also at low resolution. The comparison of phase sets is given in Table 2. The resultant potential map also closely resembles the one calculated from the image phase set (Fig. 4*c*). The map calculated, assuming random phases beyond 10 Å, is shown in Fig. 4(*d*). Thus, although many structural details are already present in the 10 Å resolution map, they can be degraded somewhat by a poor phase extension – the random assignment being the worst case.

3.2. *Expansions from other basis sets*

The choice of a 10 Å starting point was, of course, arbitrary. (Experimentally, this would be determined by the resolution of the electron micrograph.) Other resolutions were tested as possible basis sets, as if these were the (accurate) image transform limit of the respective electron micrographs. The mean phase accuracy for all 42 unique reflections improved as the resolution of the basis set was increased, as expected. On the other hand, the accuracy of just the newly defined phases did not conform similarly to the chosen starting basis resolution. As shown in Table 3, an optimal starting resolution of 11 Å was found where the newly assigned reflections had the highest accuracy. A comparison of ideal and expanded phases is given in Table 4 for this resolution. On either side of this starting point, the mean phase error for newly assigned reflections was found to increase, so that there is an approximate parabolic dependence of this accuracy on basis resolution. A dramatic change is found when a slight decrease of resolution from 12 to 12.5 Å is considered. [A potential map calculated just from image phases at the lower resolution is shown in Fig. 3(*b*).] This distinction amounts to the interaction of the intense 330 reflection with the other starting phases. Such a



(*a*) (*b*)

Fig. 3. Potential maps for two starting basis set resolutions. (*a*) 10 Å, (*b*) 12.5 Å

Table 1. *Mean phase errors (°) for Omp F porin*

	Image phase extension to 6 Å			*Ab initio* phase determination	
	*p*31*m*		*p*3	*p*31*m*	
	10 Å basis	12.5 Å basis	10 Å basis	10 Å basis	12.5 Å basis
All data	26	44	49	60	57
New phases	43	57	82		

Table 2. *Phases for Omp F porin by direct methods (°) (10 Å basis)*

*hk*0	Ideal	10 to 6 Å	*Ab initio*	*hk*0	Ideal	10 to 6 Å	*Ab initio*
100*	180	180	180	320*	118	118	89
200*	180	180	180	420*	305	305	306
300*	180	180	180	520*	36	36	2
400*	0	0	0	620	259	240	308
500*	180	180	180	720	98	84	69
600*	0	0	0	820	348	283	264
700	0	180	180	920	173	194	188
800	0	0	0	330*	261	261	179
900	180	180	180	430*	159	159	138
10,0,0	0	0	0	530	33	8	−10
110*	324	324	161	630	70	81	91
210*	43	43	44	730	297	284	262
310*	214	214	135	830	245	102	21
410*	237	237	275	440	42	78	104
510*	58	58	225	540	290	274	226
610	104	128	235	640	195	74	327
710	346	48	196	740	129	141	76
810	184	113	84	840	272	335	31
910	34	47	200	550	16	−91	177
10,1,0	217	226	83	650	30	17	18
220*	230	230	247	660	264	246	267

* Reflections within 10 Å limit.

difference in derived phase accuracy, dependent on one added phase value, was found earlier when the Sayre equation was used to phase zonal data from a polymer (Dorset, Kopp, Fryer & Tivol, 1995).

The success of the phase extension can also be evaluated with the resultant potential maps – comparing Sayre expansions to the results of adding random phases beyond the starting point (Fig. 5). Of course, since most map details are contained in the 17 accurately phased reflections with the highest intensity, the consequence of random phases on the quality of the image will be less severe as the starting resolution is increased. This is because more and more of the mean phase error will correspond to the weaker reflections.

Table 3. *Mean phase errors for Sayre expansion from various starting resolutions*

Basis resolution (Å)	No. of initial reflections	$\langle\|\Delta\varphi\|\rangle$ (°) All data	New phases	17 most intense reflections
12.5	11	44	57	33
12.0	12	29	41	17
11.0	14	29	18	6
10.0	17	26	43	3
9.0	21	26	48	4
8.0	27	23	58	0

3.3. *Ab initio phase determinations*

The possibility of carrying out *ab initio* phase determination was also evaluated. For these tests, starting resolutions were selected at 10 and 12.5 Å, respectively. As usual (Dorset, 1995), the initial solution was expanded in shells to the diffraction limit.

If, for an *ab initio* phase determination at 10 Å, the values for 200, 300, 500 and 210 were chosen (considering both $|E_{\mathbf{h}}|$ magnitude and index connectivity *via* the convolution operation to be desirable parameters), then, for 16 unique phase sets, an identified 'best' solution was found for the respective starting phase values, $\pi, \pi, \pi, \pi/4$, corresponding to the flattest distribution of density (Fig. 6*a*). The Luzzati figure of merit q was calculated for this map (Table 5). Although it does not closely resemble the map calculated at 10 Å from image-derived phases (Fig. 6*g*), phase annealing could be started with this initial set. A shift of the 400 phase caused the map to become less flat (Fig. 6*b*) so the change was rejected. A shift in the 220 phase also found no improvement in q (Fig. 6*c*). For the 320 reflection, the optimal value of q was found for a phase value of

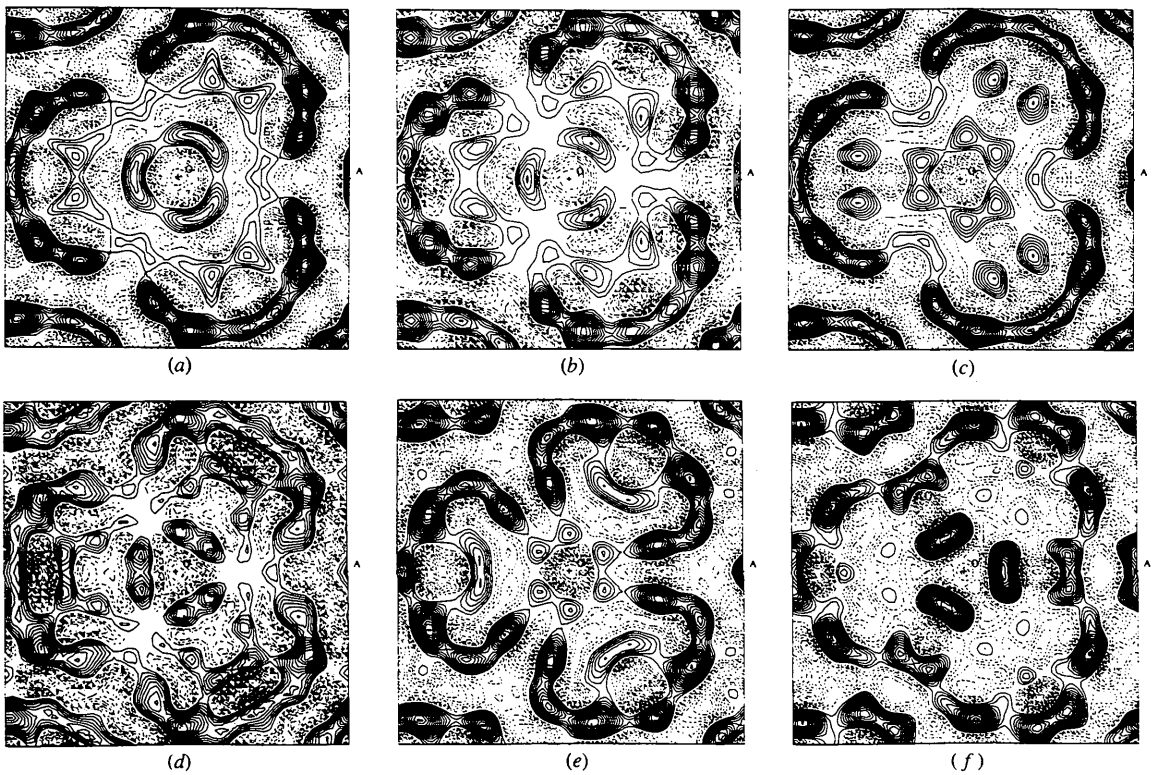


Fig. 4. Potential maps for Omp F porin at 6 Å diffraction resolution. (*a*) Phases from the Fourier transform of the image imposing symmetry *p*31*m*; 'ideal' phase solution. (*b*) Ideal (image) phases of the 17 most intense reflections. (*c*) Image-derived 10 Å basis set expanded to 6 Å by the Sayre equation. (*d*) 10 Å basis set; random phases at higher resolution. (*e*) Phase expansion from *ab initio* basis set in Table 5 (Fig. 6*f*). (*f*) Phase expansion of an *ab initio* basis set at 12.5 Å that, however, cannot be identified *a priori* as an optimal solution.

Table 4. *Phases for Omp F porin by direct methods (°) (11 Å basis)*

*hk*0	Ideal	11 to 6 Å	*hk*0	Ideal	11 to 6 Å
100*	180	180	320*	118	118
200*	180	180	420*	305	305
300*	180	180	520	36	−8
400*	0	0	620	259	252
500*	180	180	720	98	108
600	0	0	820	348	327
700	0	0	920	173	149
800	0	0	330*	261	261
900	180	180	430	159	192
10,0,0	0	0	530	33	24
110*	324	324	630	70	71
210*	43	43	730	297	286
310*	214	214	830	245	99
410*	237	237	440	42	88
510*	58	58	540	290	250
610	104	143	640	195	98
710	346	302	740	129	120
810	184	164	840	272	301
910	34	36	550	16	22
10,1,0	217	240	650	30	−1
220*	230	230	660	264	179

* Basis reflections.

Table 5. *Annealing of 10 Å starting phase set*

Initial value: $q = 0.138$.

*hk*0	$\|F\|$	$\|E\|$	Initial phase (°)	Change (q)	Decision
100	7.8	0.12	180		
200*	74.4	1.13	180		
300*	107.8	1.64	180		
400	107.6	1.02	0	180 (0.142)	Reject
500*	229.5	2.18	180		
600	11.9	0.11	0		
110	33.9	0.54	156		
210*	57.8	0.88	44		
310	28.4	0.27	135		
410	56.9	0.54	276		
510	45.8	0.44	225		
220	93.7	0.89	248	338 (0.142)	Reject
				68 (0.140)	Reject
				158 (0.138)	Reject
320	98.9	0.94	3	93 (0.126)	Accept
				183 (0.171)	Reject
				273 (0.133)	Reject
420	74.2	0.71	212	302 (0.121)	Accept
				32 (0.122)	Reject (?)
				122 (0.142)	Reject
330	149.3	1.42	90	180 (0.132)	Accept
				270 (0.143)	Reject
				0 (0.160)	Reject

* Generating reflections.

93° (Fig. 6*d*). When the 420 reflection was annealed, two maps were found with nearly the same q value, the marginally lower value was accepted, corresponding to a phase of 302° (Fig. 6*e*). Although it has a large $|E_{\mathbf{h}}|$ value, the 330 reflection was fairly insensitive to phase shifts, although a somewhat improved value of $q = \langle \Delta\rho^4 \rangle$ was found for a phase value of π, which, unlike the previous choices, is incorrect. (The insensitivity of this reflection in the annealing step is somewhat puzzling, given its decisive role in producing a more accurate phase set *via* the Sayre equation as shown above. It is obvious that the interactive role of the 330 reflection with other reflections is more important, therefore, than its own isolated value.) Based on these observations, a new low-resolution set was constructed to include revised values for the 320, 420 and 330 reflections (Fig. 6*f*) and the resultant map could then be compared to the one calculated from ideal image phases (Fig. 6*g*). The annealed phases were then extended to 6 Å resolution by the Sayre–Hughes equation. Although the phase accuracy (Table 1) was not so great as when the image-derived phases at the same resolution were extended, the resultant map still contains features of the one derived from images to the highest resolution (Fig. 4*e*). Moreover, the predominance of the β-barrel structure in the trimer can still be ascertained from the final map. A list of phases can be compared to the one found by expansion of image-derived phases (Table 2).

An attempt was then made to improve the 10 Å set by clustering the changes to the phases for the three reflections identified, by a minimization of q, as has been described earlier for halorhodopsin (Dorset, 1995). The initial change, starting with the 320 reflection, is nearly correct – in fact, this phase seems to be the most decisive for changing the appearance of the projected potential distribution in Fig. 6. When the additional change to φ_{420} was considered next, the best value for q (0.112) was found when the phase was 32°, not the more correct value, 302°, identified earlier. Obviously, a more accurate structure was not found by this path. Therefore, false minima can appear in such searches. Finally, in evaluating criteria similar to those used to prepare Fourier maps to be improved by density modification (Leslie, 1987), negative regions were arbitrarily set to zero and the figure of merit was applied only to the positive pixels. This attempt also led to a poorer selection of phase values.

The lowering of the initial resolution to 12.5 Å had a negative influence on the *ab initio* determination. In one trial, the 300, 400, 500 and 320 reflections were used to generate 16 unique solutions by phase permutation. Of these, a test for flatness only permitted rejection of 8 of the possible solutions. Even though the correct solution could not be chosen from this group, it could be narrowed to two choices if the (chemical) knowledge of β-sheet secondary structure was used to interpret the map – *i.e.* all maps containing large 'blobs' of density were rejected. Arbitrary selection of the actual starting phase set to expand to the 12.5 Å limit by the Sayre–Hughes equation resulted in a mean phase error of only 40°. Further extension to 6 Å (again *via* two cycles of the convolution) produced phases with an overall

accuracy of 57°, yielding a potential map where features of the β-barrel could be inferred (Fig. 4*f*) but with significant distortion of this structure. When the alternative list of symbolic 'generator' reflections considered in the previous section were used instead to expand to all reflections within 12.5 Å, similar difficulties were experienced, since the flatness criterion q, again, did not single out the best starting phase set. It cannot be claimed, therefore, that this lower-resolution start leads to a satisfactory outcome.

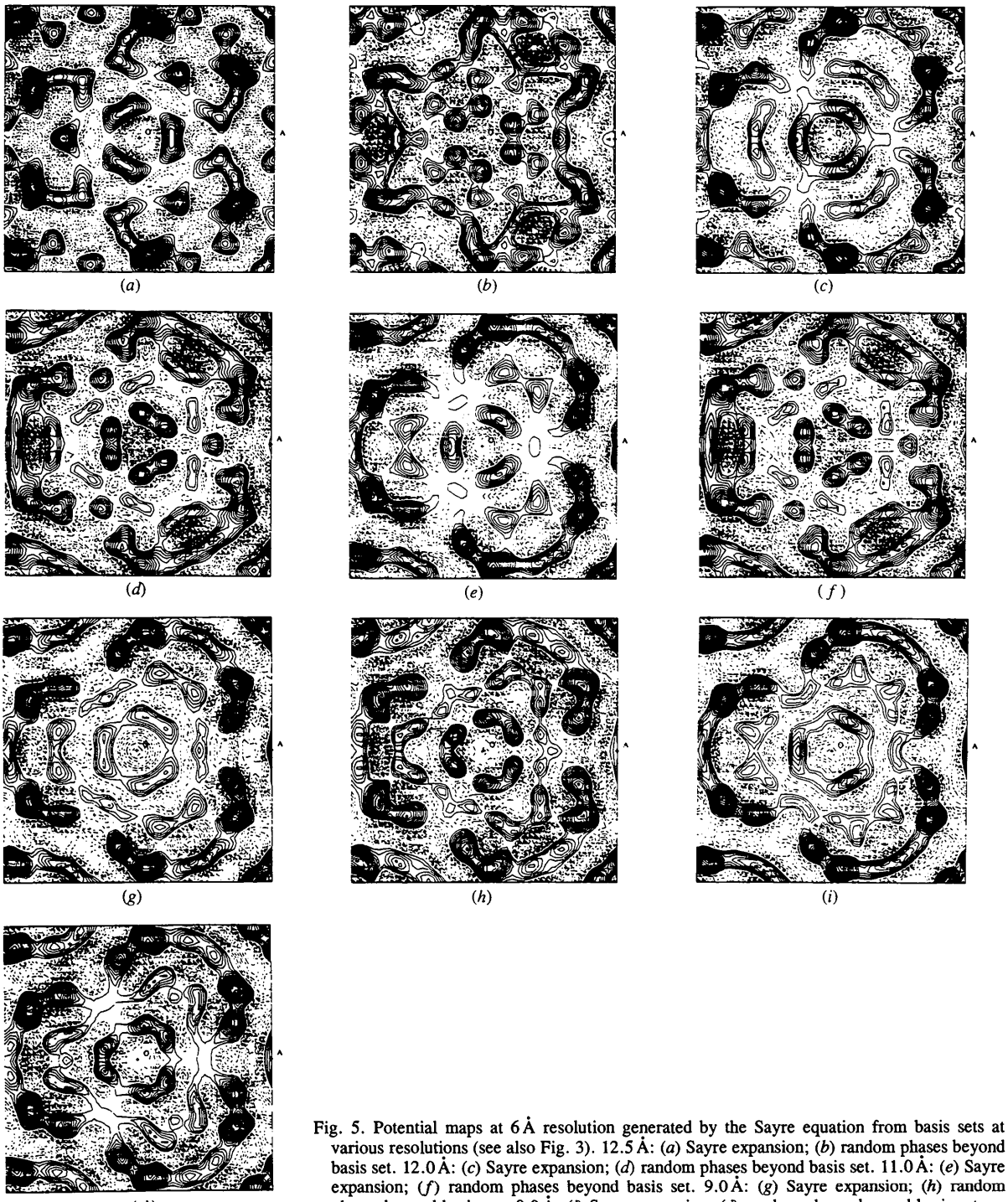


Fig. 5. Potential maps at 6 Å resolution generated by the Sayre equation from basis sets at various resolutions (see also Fig. 3). 12.5 Å: (*a*) Sayre expansion; (*b*) random phases beyond basis set. 12.0 Å: (*c*) Sayre expansion; (*d*) random phases beyond basis set. 11.0 Å: (*e*) Sayre expansion; (*f*) random phases beyond basis set. 9.0 Å: (*g*) Sayre expansion; (*h*) random phases beyond basis set. 8.0 Å: (*i*) Sayre expansion; (*j*) random phases beyond basis set.

4. Discussion

The idea of using direct phasing techniques to predict the values of unknown terms in protein structure analyses has now been applied for three separate macromolecular zonal data sets, including two examples with predominately α-helical geometry and one with extensive β-sheet substructure. While the resultant phase accuracy may not be as good as the best analysis of electron micrographs obtained to high resolution [with all due care that all possible perturbations mentioned above are minimized – see Henderson *et al.* (1986)], it is still good enough to determine many of the characteristic structural details within the molecular envelope. However, the accuracy of the phase extension seems to be related also to an optimal basis resolution. There does not seem to be a preliminary requirement for any particular structural motif, including a putative 'pseudo-atomicity' of a helix projection, for such phase extensions by the Sayre convolution to be successful. Furthermore, the reciprocal-space domain out to *e.g.* 6 Å resolution is particularly useful for such phase extensions. Somewhat beyond this resolution, a 'nodal point' is encountered, where reversals in the overall phase envelope occur (Dorset, Kopp, Fryer & Tivol, 1995). Fan, Hao & Woolfson (1991) have discussed how low-angle data from proteins should be particularly useful for phase determination because the intensities in this region are the most prominent part of the diffraction pattern. While, unfortunately, the lowest-resolution intensities are often not measured in X-ray crystallography, they are a prominent feature of an electron crystallographic data set and also represent the region where phases are most conveniently obtained from

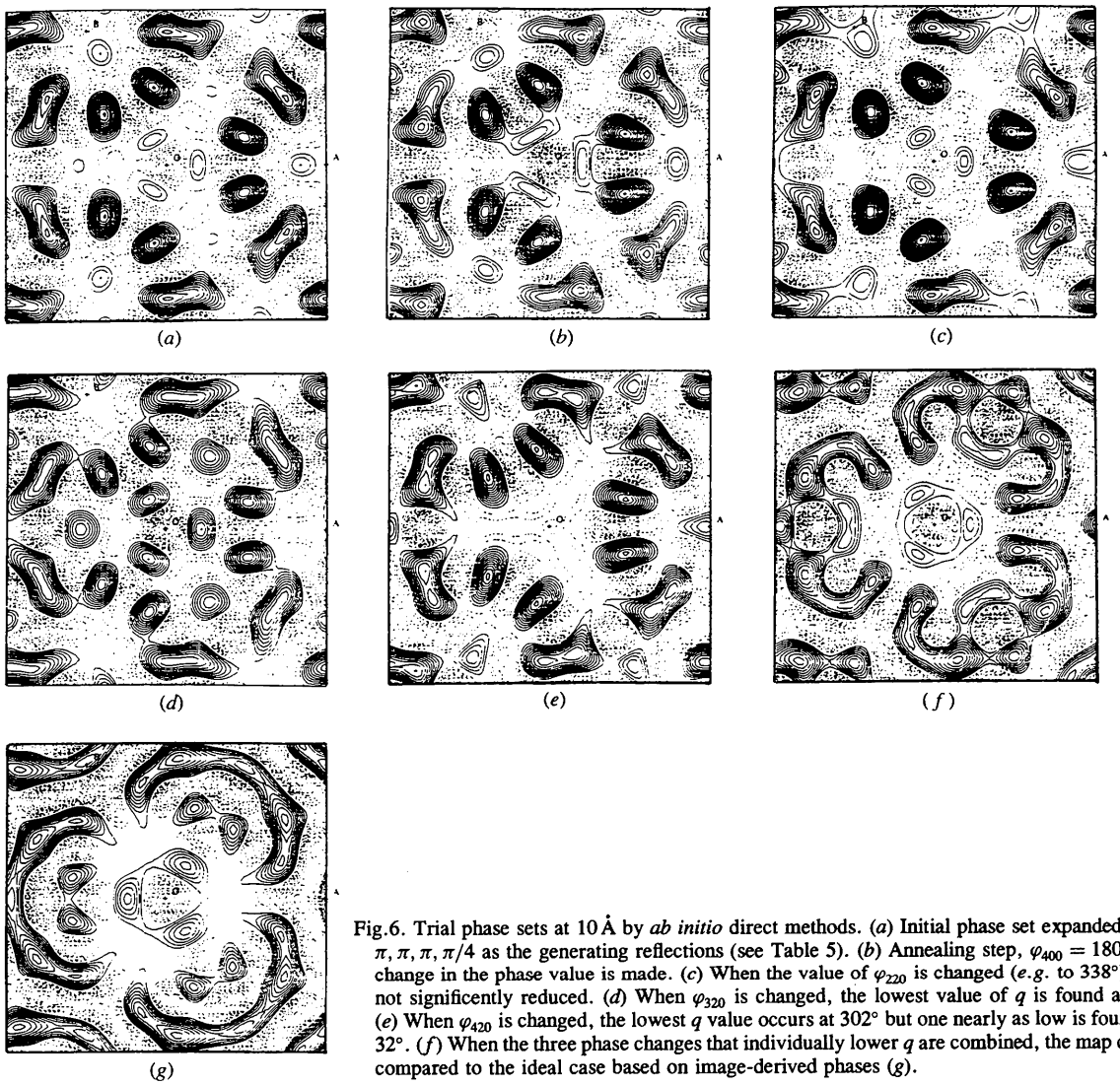


Fig.6. Trial phase sets at 10 Å by *ab initio* direct methods. (*a*) Initial phase set expanded from $\pi, \pi, \pi, \pi/4$ as the generating reflections (see Table 5). (*b*) Annealing step, $\varphi_{400} = 180°$. No change in the phase value is made. (*c*) When the value of φ_{220} is changed (*e.g.* to 338°), q is not significently reduced. (*d*) When φ_{320} is changed, the lowest value of q is found at 93°. (*e*) When φ_{420} is changed, the lowest q value occurs at 302° but one nearly as low is found for 32°. (*f*) When the three phase changes that individually lower q are combined, the map can be compared to the ideal case based on image-derived phases (*g*).

electron micrographs. The potential advantage of electron crystallography over X-ray crystallography is therefore clear, since the amplitudes and phases of the most intense reflections are most easily found by electron diffraction and image analysis.

The results of two *ab initio* phase determinations for projected protein structures (*e.g.* halorhodopsin and now the bacterial porin) show some promise, even though the results are not as useful as those from the phase extension. In these analyses, the need for a suitable figure of merit (FOM) that is different from the ones commonly used in crystallography is hardly surprising, since there is no reason to expect that the Cochran (1952) condition, important for atomic resolution structure analyses, should have any particular relevance to the application of direct methods to macromolecules (or anything else) in this low-resolution domain. Indeed, it is more reasonable to expect that the exact opposite of 'peakiness', *i.e.* density flatness or smoothness, should be the deciding factor in seeking a correct solution. While multisolution techniques also appear to be useful for phase determination, particularly when they are coupled with the Sayre equation, there are, obviously, implied caveats for their use. For example, there is a requirement that an initial phase set be built up to a sufficient resolution limit so that a test of density flatness will be decisive for choosing a useful starting point for annealing and then expansion. When the permuted generator reflections are expanded by the convolution to fill the limits of the defined low-resolution basis reflections, the starting phases must, therefore, produce a significant approximation to the molecular envelope. Seeking an absolute minimum of the density flatness criterion, q, in such analyses, however, can be dangerous. As a relative indicator for finding possible phase solutions, it seems to be useful, just as it is for finding likely phases that should be changed on annealing. (However, annealing reflections with weak amplitudes also is not useful.) False minima of q occur, therefore, and also shallow ones, as experienced before when this figure of merit was used to direct the phase refinement of phospholipid bilayers (Dorset, 1991). Thus, while this concept seems to provide a useful starting point, there may be other FOM's that will prove to be more useful for identification of correct phase changes. For example, one proposed recently by Mishnev & Woolfson (1994) could be tested in this context.

Despite these difficulties, the results of this work compare quite favorably to earlier refinement of multiple isomorphous replacement (MIR) X-ray phases for macromolecules at similar resolution by direct methods. For example, Reeke & Lipscomb (1969) reported mean phase errors in the range 37 to 79° for carboxypeptidase at 6 Å, for data refined by the tangent formula. A low-resolution expansion of *t*-RNA MIR phases by matricial direct methods resulted in an average error of 73°, or 44° for the 36% of the reflections with the best figure of merit (Podjarny, Schevitz & Sigler, 1981). In the electron crystallographic approach, it is, obviously, always advantageous to start with as much phase information as possible. If just the 17 most intense reflections for the Omp F porin are monitored, it can be shown that the mean phase error is only 17° when 12 Å image phases are extended by 34° when 12.5Å image phases are used and 42° when the *ab initio* approach is taken.

From such work, the challenges to *ab initio* phase determinations in protein crystallography are made somewhat more clear than before. It is reasonable, first of all, to consider just the low-angle region as being reasonably accessible to phase analysis. Nevertheless, even though multiple-solution techniques *via* phase convolutions might provide a reasonable start for an *ab initio* analysis, is it possible to find better figures of merit for identifying the best solution and also improving the initial estimate? Similar success at structure determination has recently been realized when atomic resolution data were being analyzed by direct methods (Weeks *et al.*, 1995), using, of course, quite different criteria for identifying the correct phase solution. (There, the current resolution limit seems to be somewhere around 1.2 Å.) As pointed out previously (Podjarny & Yonath, 1977), the greatest challenge to phase determination in macromolecular crystallography currently seems to be the intermediate resolution region near 5 Å, where rather simple assumptions, useful for the other extremes of resolution, will then break down, so that one is faced with defining a criterion for structure identification somewhere between flatness and peakiness.

5. Conclusions

(i) Given a sufficiently accurate lower-resolution phase set from a protein, *e.g.* from the Fourier transform of an electron micrograph, it is possible to extend to higher resolution by rather simple convolutional techniques, such as the Sayre equation. The success of such an expansion does not seem to depend upon a specific secondary structure for the protein since similar success is observed from quite dissimilar examples with, respectively, high amounts of α-helix or β-sheet.

(ii) Surprisingly, phase extensions from a starting basis set are found to have an optimal resolution at low angle for assigning accurate values to higher-resolution reflections. On either side of this optimal basis resolution, the mean phase error increases.

(iii) *Ab initio* phase determinations are, in principle, possible for noncentrosymmetric projections, just as they are for centrosymmetric projections. However, even though the criteria of density flatness or smoothness are reasonably used as figures of merit below 6 Å,

it is clear that more robust FOM's than the $\langle \Delta\rho^4 \rangle$ value, tested in this and a previous study (Dorset, 1995), must be found.

Research was funded by a grant from the National Institute of General Medical Sciences (GM46733). Dr H. J. Sass is thanked for providing the data set used for this analysis.

References

Amos, L. A., Henderson, R. & Unwin, P. N. T. (1982). *Prog. Biophys. Mol. Biol.* **39**, 183–231.

Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*, p. 142. New York: Academic Press.

Cochran, W. (1952). *Acta Cryst.* **5**, 65–67.

Dorset, D. L. (1991). *Biophys. J.* **60**, 1356–1365.

Dorset, D. L. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 10074–10078.

Dorset, D. L., Kopp, S., Fryer, J. R. & Tivol, W. F. (1995). *Ultramicroscopy*, **57**, 59–89.

Earnest, T. N., Walian, P. J., Gehring, K. & Jap, B. K. (1992). *Trans. Am. Crystallogr. Assoc.* **28**, 159–164.

Fan, H. F., Hao, Q. & Woolfson, M. M. (1991). *Z. Kristallogr.* **197**, 197–208.

Gaskill, J. D. (1978). *Linear Systems, Fourier Transforms, and Optics*, pp. 266–285. New York: Wiley.

Gilmore, C. J., Shankland, K. & Fryer, J. R. (1993). *Ultramicroscopy*, **49**, 132–146.

Glaeser, R. M. & Downing, K. H. (1992). *Ultramicroscopy*, **47**, 256–265.

Harker, D. (1953). *Acta Cryst.* **6**, 731–736.

Havelka, W. A., Henderson, R., Heymann, J. A. W. & Oesterhelt, D. (1993). *J. Mol. Biol.* **234**, 837–846.

Henderson, R., Baldwin, J. M., Downing, K. H., Lepault, J. & Zemlin, F. (1986). *Ultramicroscopy*, **19**, 147–178.

Hovmöller, S. (1992). *Ultramicroscopy*, **41**, 121–135.

Hughes, E. W. (1953). *Acta Cryst.* **6**, 871.

Leslie, A. G. W. (1987). *Acta Cryst.* A**43**, 134–136.

Li, F. H. (1991). *Electron Crystallography of Organic Molecules*, edited by J. R. Fryer & D. L. Dorset, pp. 153–167. Dordrecht, Kluwer.

Luzzati, V., Mariani, P. & Delacroix, H. (1988). *Makromol. Chem. Makromol. Symp.* **15**, 1–17.

Luzzati, V., Tardieu, A. & Taupin, D. (1972). *J. Mol. Biol.* **64**, 269–286.

Mishnev, A. F. & Woolfson, M. M. (1994). *Acta Cryst.* D**50**, 842–846.

Podjarny, A. D., Schevitz, R. W. & Sigler, P. B. (1981). *Acta Cryst.* A**37**, 662–668.

Podjarny, A. D. & Yonath, A. (1977). *Acta Cryst.* A**33**, 655–661.

Reeke, G. N. & Lipscomb, W. N. (1969). *Acta Cryst.* B**25**, 2614–2623.

Rogers, D. (1980). *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. Ladd & R. A. Palmer, pp. 23–92. New York: Plenum.

Rosenbusch, J. P. (1974). *J. Biol. Chem.* **249**, 8019–8029.

Rossmann, M. G. & Henderson, R. (1982). *Acta Cryst.* A**38**, 13–20.

Sass, H. J. (1990). PhD thesis, Free University of Berlin, Germany.

Sass, H. J., Büldt, G., Beckmann, E., Zemlin, F., Van Heel, M., Zeitler, E., Rosenbusch, J. P., Dorset, D. L. & Massalski, A. (1989). *J. Mol. Biol.* **209**, 171–175.

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

Sayre, D. (1980). *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. Ladd & R. A. Palmer, pp. 271–286. New York: Plenum.

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

Van Heel, M. & Keegstra, W. (1981). *Ultramicroscopy*, **7**, 113–130.

Weeks, C. M., Hauptman, H. A., Smith, G. D., Blessing, R. H., Teeter, M. M. & Miller, R. (1995). *Acta Cryst.* D**51**, 33–38.