Acta Cryst. (1999). A**55**, 448–456

Direct methods in protein electron crystallography – beef liver catalase in its fully hydrated form at room temperature

Douglas L. Dorset[a]* and Christopher J. Gilmore[b]

[a]*Electron Diffraction Department, Hauptman–Woodward Medical Research Institute, Inc., 73 High Street, Buffalo, New York 14203-1196, USA, and* [b]*Chemistry Department, Glasgow University, Glasgow G12 8QQ, Scotland. E-mail: dorset@hwi.buffalo.edu*

(*Received 4 June 1998; accepted 14 August 1998*)

Abstract

The crystal structure of beef liver catalase was determined *ab initio* in projection to 9 Å resolution using electron diffraction data at room temperature from hydrated specimens maintained in an environmental chamber in the electron microscope. A conservative combination of symbolic addition with maximum entropy and likelihood led to a model with a Patterson correlation coefficient $C = 0.89$ to the observed data. This independent solution could then be compared favorably to a previous 23 Å analysis of electron micrographs from frozen hydrated preparations. Prediction of the higher-resolution structure by extension of the lower-resolution image-based phase basis set also gave a good match to the direct-methods solution, particularly for the most intense reflections.

1. Introduction

Electron crystallography is an important technique for the characterization of proteins that are otherwise difficult to crystallize with sample sizes sufficient for collection of X-ray intensity data. Much attention has been devoted to the integral membrane proteins, largely because two-dimensional crystalline sheets of the macromolecule embedded in a lipid bilayer are most conveniently prepared (Jap *et al.*, 1992). Structural results from pioneering high-resolution three-dimensional electron crystallographic studies of integral membrane proteins, such as bacteriorhodopsin (Kimura *et al.*, 1997) and the bacterial porins (Jap *et al.*, 1991), have been verified recently by independent X-ray crystal structure analyses of materials crystallized in the presence of detergent (Pebay-Peyroula *et al.*, 1997; Cowan *et al.*, 1992). Although there are weaknesses in the electron crystallographic technique, primarily due to the limited tilt of conventional electron-microscope goniometer stages (Amos *et al.*, 1982), it is now clear that careful structure analyses, based on electron diffraction amplitudes and the phases obtained from the Fourier transform of experimental electron micrographs, will lead to a viable result.

Direct phasing methods have been applied recently to electron diffraction data from two-dimensional protein crystals, partly to tackle the problem of finding the low-resolution envelope of the macromolecule, in order to differentiate it from a solvent space. Previous experience from X-ray crystallography had shown that, within a low-resolution domain (*e.g.* within 6 Å), conventional direct methods are useful for extension of a basis set to unphased reflections (Podjarny *et al.*, 1981). This has also been demonstrated in electron crystallography. For example, the phases provided from an image transform can be extended, with reasonable accuracy, to the higher-resolution limit of the measured electron diffraction data by a variety of techniques including maximum entropy and likelihood (Gilmore *et al.*, 1993) and the Sayre equation (Dorset *et al.*, 1995).

True *ab initio* phase determination *via* multisolution techniques has also been successful. Convolutional phasing methods or maximum entropy and likelihood have been exploited and compared, and shown to give equivalent results (Dorset, 1995*a*, 1996; Gilmore *et al.*, 1996). Following an early suggestion by Harker (1953), properties of the scattering density in the macromolecule, such as the pseudo-atomic nature of globular subunits, have also been exploited and, when applicable, have been shown to be quite useful for finding *e.g.* projected helix sites (Dorset, 1997*a*,*b*, 1998; Dorset & Jap, 1998).

There is less experience in the determination of soluble protein structures by electron crystallographic methods. A significant difficulty can arise from the crystallization of proteins in space groups in which more than one protein layer, n, is found in the projected unit-cell direction, where it is all too possible to find thin crystals with $(mn + 1)$ layers. If m is small, there can be extreme variations in the zonal intensities measured in the electron diffraction patterns, not to mention the violation of systematic absences expected for a particular space group. One of the earliest examples of a soluble protein investigated in the electron microscope, initially with negatively stained preparations, was beef liver catalase (Valentine, 1964; Vainshtein *et al.*, 1976).

Preparations of catalase embedded in glucose (Unwin, 1975), as well as hydrated specimens quickly frozen in a cryogen (Taylor & Glaeser, 1974), have been extensively studied. There has also been an effort to characterize hydrated specimens at room temperature, utilizing a differentially pumped environmental chamber to prevent the protein from dehydrating in the microscope vacuum (Matricardi *et al.*, 1972). While electron diffraction data have been recorded to 2.8 Å resolution, very little else has been performed with these intensities except to demonstrate that they should conform adequately to a kinematical scattering model to permit a structure determination to be carried out (Dorset & Parsons, 1975*a*). In this paper, the first attempt to analyze these intensity data is reported as a test of *ab initio* phasing procedures.

2. Materials and methods

2.1. *Catalase crystals*

Beef liver catalase was recrystallized from a commercial preparation (Boehringer Mannheim) using a procedure described by Sumner & Dounce (1937), except that no ammonium sulfate was used to bring down the crystals. After solubilization and recrystallization at an acidic pH, plate-like crystals of the protein were observed that were stable in distilled water. A combination of methods (Dorset & Parsons, 1975*b*,*c*) to measure the crystal thickness revealed that a skewed Gaussian distribution, centered in the 100–160 nm interval, was contained in a typical preparation. It was possible, for example, to use a Beer's law relationship of electron beam attenuation after a diffraction experiment on a given crystal (Dorset & Parsons, 1975*b*) to estimate the thickness of an individual crystalline object. Laterally, the crystals measured, typically, 10×20 μm.

2.2. *Electron diffraction*

As described earlier (Dorset & Parsons, 1975*a*), electron diffraction experiments were carried out at 200 kV using a JEOL JEM-200A electron microscope fitted with a differentially pumped environmental chamber. The first demonstration of its efficacy for obtaining useful diffraction results from this dehydration-sensitive protein was presented by Matricardi *et al.* (1972). Electron diffraction patterns were recorded on a sensitive screenless X-ray film. After calibration against a gold powder standard, the rectangular lattice constants were observed to be $a = 69.7$, $b = 177$ Å, in agreement with measurements made at other laboratories (Wrigley, 1968). Systematic absences along the reciprocal-lattice rows were in accord with a *pgg* plane-group assignment, also in agreement with the identified space group, $P2_12_12_1$ (Unwin, 1975). It is clear, by inspection of the lowest-angle diffraction region, that the intensity distributions of the patterns from hydrated specimens and glucose-embedded preparations are somewhat different. [This was tested by our studies of glucose-embedded preparations but can be verified independently by comparing results from other laboratories, *e.g.* Unwin's (1975) work on saccharide-impregnated crystals with Taylor's (1978) study of frozen hydrated crystals.] Qualitatively, our electron diffraction intensities from the hydrated specimens at room temperature compare well with those in patterns from frozen hydrated preparations (Schröder & Burmester, 1993), except that the former were screened to minimize the appearance of space-group-forbidden reflections.

Films were scanned on flat-bed microdensitometers (Joyce-Loebl Mk III C or CS) and an approximation of the integrated intensity was made for the peak traces. No Lorentz-type correction was applied. Using diffraction patterns extending to high resolution, as well as those emphasizing the low-angle region, intensities from a number of patterns were averaged to accumulate a representative composite intensity set. Although, compared to molecular organics, the crystals are very flat, compensation for Ewald-sphere curvature was not considered to be important for the 9 Å resolution limit of the structure determination, aided in part by low electron wavelength. A plot of intensity averaged over shells of $\sin\theta/\lambda$ reveals a scattering characteristic typical of many proteins (Pauling & Corey, 1951), *i.e.* a minimum near $d^* = 6$ Å^{-1}. As discussed earlier (Podjarny *et al.*, 1981; Dorset *et al.*, 1995), phase determination within the first intensity envelope is accessible to direct methods. In all, 116 unique $hk0$ reflections were considered.

Some error in the intensity average over a number of diffraction patterns is expected. For example, the $R_{\rm merge}$ value for a composite intensity set to the one initially analyzed [one of the diffraction patterns used for the calculation of a Patterson map in the earlier study of Dorset & Parsons (1975*a*)] is 0.27, which is a rather large value that could be problematic for data collection from small organics. The value of $R_{\rm sym}$, on the other hand, was generally more acceptable ($<$0.20), since an *mm* distribution of intensity was imposed as a criterion in the search for suitable diffraction patterns for use in the average. The error in measured intensities is caused by a number of experimental difficulties. Primarily, it was all too easy to tilt slightly the rod used to hold the specimen in the environmental chamber, so that good zonal patterns were sometimes difficult to find. The Beer's law estimate of crystal thickness by beam attenuation was used to screen the patterns further. If the crystal was found to be very thin, then a value that would correspond to an odd number of layers (where two are in the projected unit cell) would also be rejected. A plot of determined crystal thicknesses indicated that the projected unit-cell c dimension should be near 200 Å, in agreement with the value 205 Å obtained from powder X-ray patterns (Unwin, 1975). Although the problem

affected many of the diffraction patterns recorded in original work (less than 10% were found suitable for consideration as sources of intensity data), the locations of intense reflections in the low-angle diffraction pattern again agreed very well with positions found in the Fourier transform of an image taken of a frozen hydrated specimen (Taylor, 1978) as well as other published electron diffraction patterns (Schröder & Burmester, 1993). This point should be emphasized. Efforts to improve the accuracy of electron diffraction intensities from frozen hydrated samples have been made recently (Schröder & Burmester, 1993; Burmester *et al.*, 1995), including the use of energy filtration.

2.3. *Data normalization and direct methods*

In the direct phase determination, the results from previous image analyses were ignored initially so that the phase determination could truly test the application of these methods on an 'unknown' structure, without readily available phase information from images at hand. When, however, a comparison of phase results was necessary, two papers reporting the results from the frozen hydrated samples were consulted. In one (Akey & Edelstein, 1983), a subset of phases had been communicated to the authors by Dr K. Taylor. In another paper (Taylor, 1978), the phase information could be obtained directly after digitizing a published electron micrograph *via* a fast charge-coupled-device (CCD) camera for analysis with the computer program package *CRISP* (Hovmöller, 1992), providing values for a few weaker reflections but otherwise agreeing completely with the previously published table. Phase values derived from this image deviated from the distribution expected for a projected *pgg* symmetry by only 12.1°.

For the direct phase analysis, it was assumed that the protein density distribution could be simulated by an aggregation of pseudo-atomic 'globs' (Harker, 1953) to sufficient accuracy. As discussed in previous papers utilizing this approximation (Dorset, 1997*a*,*b*, 1998), the unit-cell constants were reduced by a factor of 10, so that the glob transform could be modeled by the carbon electron scattering factor. Thus, it is assumed that the glob profile is nearly Gaussian. Intensity data to a 9 Å limit (or 0.9 Å, after rescaling) were used to calculate a Wilson plot, which indicated a negative overall displacement parameter $B = -4.0$ Å^2, used as a parameter to rescale the scattering-factor model for calculation of $|E_h|$ values. (The negative displacement parameter is, in itself, meaningless; it is merely a device to optimize the shape of a phenomenological scattering factor.) Previously, in the analysis of proteins with a high amount of α-helix, and projecting down the helix axes, there was a specific reason for the tenfold dimensional rescaling (Dorset, 1997*a*,*b*, 1998; Dorset & Jap, 1998), *i.e.* comparison of center-to-center distances for two touching helices to the length of a carbon–carbon single bond. In this application, the appropriateness of this rescaling assumed, without prior proof, that there were features of the catalase substructure that would conform to this scattering-factor model. However, the approximate tenfold rescaling of macromolecular phasing problems has been mentioned by other authors (Podjarny & Urzhumtsev, 1997).

Since this is an unknown structure, the method of finding phase information from the recorded diffraction intensity data was the most conservative approach possible. In an overview, two direct phasing approaches were used separately to find possible phase values for a subset of reflections (generally those with the largest normalized structure-factor magnitudes). The results from the identified 'best' solutions, established from a criterion of map density flatness (see below), were then compared. A potential map was then calculated from the reflections common to both methods. From this map, pseudo-atom peaks were used for a structure-factor calculation, giving a phase set on a common basis for the two approaches. These phases could then be compared to the values obtained from the Fourier transform of electron micrographs of frozen hydrated samples.

In the first direct methods approach, after generation of Σ_2 three-phase invariants ('triples'), $\varphi_{\mathbf{h}} = \varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}$ symbolic addition (Karle & Karle, 1966) was used to assign phase values to the 35 triples with largest values of $A = (2/N^{1/2})|E_{\mathbf{h}}E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}|$, where N is the number of 'globs' in the unit cell. [Symbolic addition applied to electron diffraction data sets is demonstrated in a recent monograph (Dorset, 1995*b*).] After origin definition *via* the assignments $\varphi_{550} = 0$ and $\varphi_{630} = 0$, two algebraic terms (φ_{690}, φ_{170}) were included to assign phase values to unknown reflections. Also, a value for another phase, $\varphi_{060} = \pi$, was indicated from a highly probable Σ_1 phase invariant, but it was not included in the basis set. Potential maps generated from these trial phase sets were evaluated using the Luzzati figure of merit (FOM) $\langle \Delta\rho^4 \rangle$ (Luzzati *et al.*, 1972). In this evaluation, $\Delta\rho = \rho - \overline{\rho}$, where the average map density, $\overline{\rho} = 0$, is established by setting $F_{000} = 0.0$ when computing potential maps from the trial phases sets. As found in a previous study (Dorset, 1996), this is not an absolutely rigorous FOM for discriminating optimal phase sets from multiple solutions, although it is nearly correct for low-resolution phase determinations. It had been a useful FOM in one *ab initio* phase determination of a membrane protein, when the globular pseudo-atomic approximation was not utilized (Dorset, 1995*a*).

In a second, separate, phase determination, the maximum-entropy (ME) method, coupled with likelihood evaluation, was employed. The method is based on the Bricogne (1984) formalism, as implemented by Bricogne & Gilmore (1990) and reviewed in detail by Gilmore (1996). This is a technique that can also begin with origin-defining reflections (here values $\varphi_{140} = 0$ and

$\varphi_{170} = 0$ were chosen). It then would proceed by testing possible phase permutations for unassigned reflections as 'branches' of a 'tree', which are then 'pruned' by an appropriate FOM, here the maximum likelihood. The data were normalized using the *MITHRIL* computer program (Gilmore, 1984) with a fixed displacement parameter $B = 10.0$ Å^2 to give a set of unitary structure factors $|U_{\mathbf{h}}|^{\text{obs}}$. The origin was defined, as for the symbolic addition procedure described in the previous paragraph, to generate the root node of a maximum entropy phasing tree. Twelve reflections were given permuted phase values in a full factorial design, thus generating $2^{12} = 4096$ nodes on the second level of the tree. Each one of these nodes or phase choices was subjected to constrained entropy maximization, in which the amplitudes and phases of the origin and permuted reflections (the basis set) were used as the constraints, using the *MICE* computer program (Gilmore & Bricogne, 1997). The corresponding likelihood estimates were analyzed (Shankland *et al.*, 1993; Gilmore *et al.*, 1997), and those eight nodes with the highest scores were kept, and the corresponding maps examined. These centroid maps (Bricogne & Gilmore, 1990) were computed as Sim-filtered (Sim, 1959) U maps, where the basis-set reflections contributed with full weight, and the extrapolated reflections, predicted by the ME regularization procedure, were given coefficients, computed *via*

$$|U_{\mathbf{k}}|^{\text{obs}} \tanh(X_{\mathbf{k}}) \exp(i\varphi_{\mathbf{k}}^{\text{ME}}),$$

where

$$X_{\mathbf{k}} = (N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\text{obs}}|U_{\mathbf{k}}^{\text{ME}}|$$

and $\varphi_{\mathbf{k}}^{\text{ME}}$ is the phase angle predicted from the ME optimization with a corresponding amplitude $|U_{\mathbf{k}}^{\text{ME}}|$; N is the number of atoms ('globs') in the unit cell and $\varepsilon_{\mathbf{k}}$ is the statistical weight of reflection **k**. Although a limited number of trial phase solutions were culled *via* the maximum-likelihood criterion, the Luzzati FOM was also chosen to select one of these, again acknowledging that the effective imposition of density flatness may not be a rigorous criterion. On the other hand, it was hoped that the comparison of two independent phasing methods would be somewhat self-correcting.

3. Results

A potential map at 9 Å resolution, based on phase assignments for 26 unique reflections by symbolic addition, is shown in Fig. 1(*a*). It is a phase solution that corresponded to the lowest value of the Luzzati FOM. In this, the value $\varphi_{060} = \pi$ agrees with the Σ_1 estimate cited above. On the other hand, the values $\varphi_{140} = \pi$ and $\varphi_{170} = \pi$ are the reverse of the origin-defining reflections used for the maximum-entropy and likelihood determination. There are similarities in the density distributions in all $2^2 = 4$ solutions.

Next, the maximum-entropy results were evaluated. The maps from the top eight nodes, ranked by likelihood scoring methods, were broadly similar and were also clearly wrong since they exhibited strong density clustered around the dyad axis, as shown in Fig. 1(*b*), based on 52 phased structure-factor magnitudes. However, the Babinet representation of the centroid maps, in which each phase was shifted by π, exhibited a much more uniform and homogeneous positive density, as shown in Fig. 1(*c*). In this overlap reflection set from the maximum-entropy solution, after the Babinet phase shift, the shifted values of the two origin-defining reflections, 140 and 170, are now in accord with the identified symbolic addition solution. Also, the matching of the Σ_1 estimates, as well as the one symbolic addition origin-defining reflection, $\varphi_{550} = 0$, found in the overlap set, are in accord for the two determinations. For the comparison of the determinations by the two methods, five of 15 reflections differ in their phase assignments, but the differences are expressed for medium-intensity reflections and not for the strongest ones.

The Babinet-shift behavior is interesting, and only seems to occur in a maximum-entropy phasing environment where there are low-contrast diffraction data from biological specimens, and may be a consequence of the antiphase property of macromolecule and solvent contributions to the structure factor at low resolution (Podjarny & Urzhumtsev, 1997), particularly for a crystal with a large solvent content. It was seen in the study of Omp F porin (Gilmore *et al.*, 1996), although it was not as serious as found for the current example. Choosing to invoke Babinet's principle here was, in part, an imposition of the flatness criterion for the positive density distribution; it might, therefore, be expected that the preferred maps from the ME calculations would exhibit maximum entropy instead of likelihood, but an analysis of node scores, based on the latter criterion or a weighted sum of entropy and likelihood, was no more successful. Nonetheless, once the Babinet map had been calculated, it provided a density representation more in accord with the symbolic addition phase solution.

In order to compare the phase results from symbolic addition with those from maximum entropy and likelihood, reflections from the corresponding 'best' solutions were chosen for calculation of potential maps. In the case of maximum-entropy solutions, the subset with overlaps to the symbolic addition phases was used (Fig. 1*d*), again corresponding to the lowest Luzzati FOM. Peaks were taken from the maps as pseudo-atoms for a structure-factor calculation $F_{\mathbf{h}} = \sum f_{\text{C}} \exp(2\pi i \mathbf{h} \cdot \mathbf{r})$. Again, after rescaling unit-cell parameters, it was assumed that the carbon scattering factor, f_{C}, adjusted for the phenomenological displacement parameter, would model the glob transform. The solutions were then screened by a Patterson correlation coefficient (Drenth, 1994) defined as $C = \sum m_o m_c \big/ \sum |m_o||m_c|$,

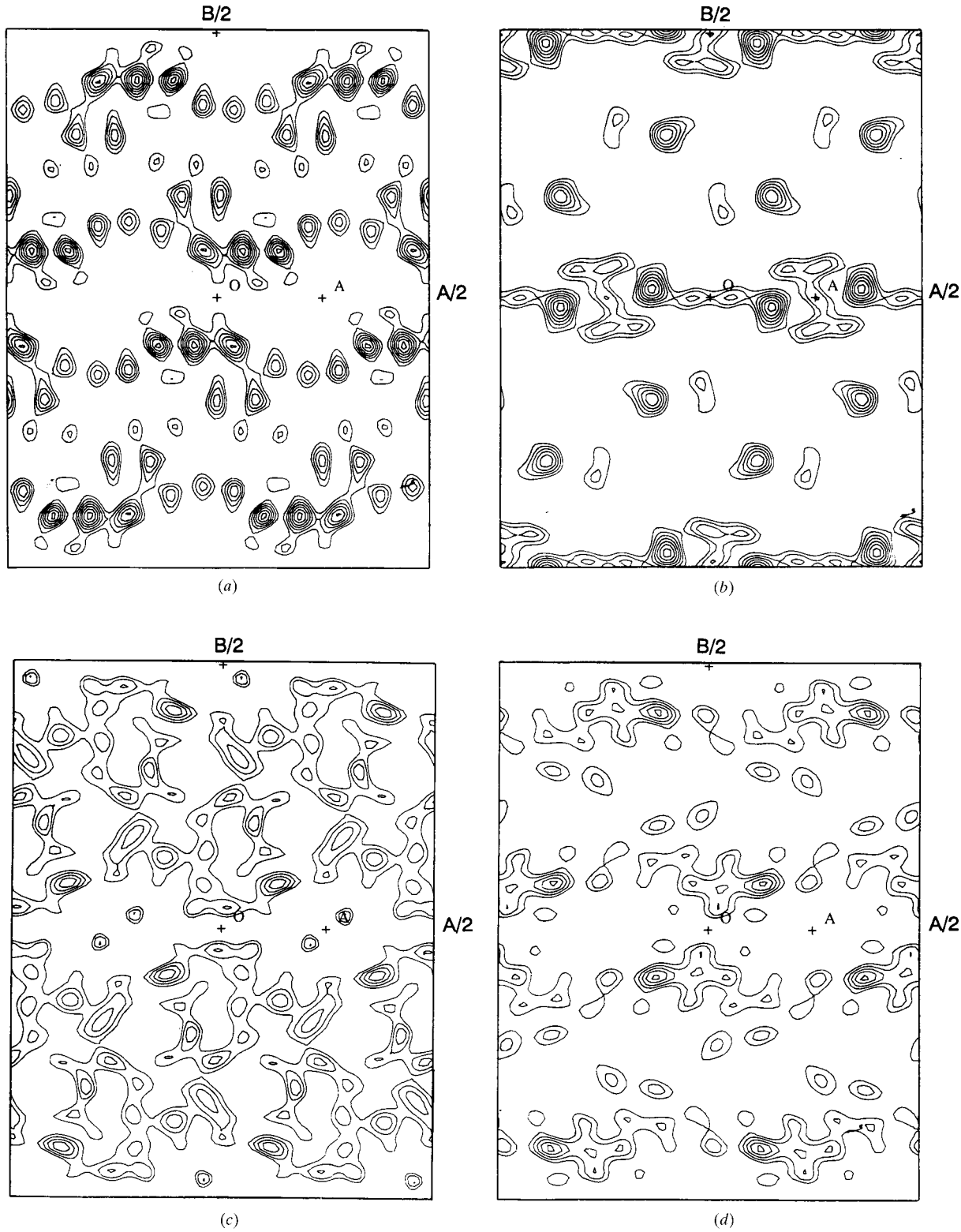

Fig. 1. Potential maps (approximately 9 Å resolution) after direct phasing: (*a*) symbolic addition, lowest $\langle \Delta\rho^4 \rangle$; (*b*) maximum entropy and likelihood, all assigned phases, lowest $\langle \Delta\rho^4 \rangle$; (*c*) Babinet phases of previous solution (*b*); (*d*) subset of Babinet set in (*c*), where reflection indices overlap with those used to generate (*a*).

where $m_o = |F_o|^2 - \langle |F_o|^2 \rangle$, *etc*. (subscripts c and o denote calculated and observed values, respectively). This has been suggested to be a suitable FOM for evaluating low-resolution phase determinations (Podjarny & Urzhumtsev, 1997).

When peaks in the map generated from the complete maximum-entropy solution were used as pseudo-atoms in a structure-factor calculation (subset of 17 most intense reflections), the Patterson correlation was, in fact, negative, *i.e.* $C = -0.012$. This solution, obviously, is not a likely one. Peaks from a map generated with the symbolic addition set (Fig. 1*a*) gave a reasonable value, $C = 0.74$, but the map generated from the subset of maximum-entropy phases, using reflections also appearing in the symbolic addition sets (Fig. 1*d*), gave the best value, $C = 0.89$. (The value is $C = 0.85$ when the sample is expanded to 24 most intense reflections.) Attempts to improve this figure by Fourier refinement were not successful; the correlation coefficient did not change very much. Thus, to a first approximation, this phase solution, of which the values for strongest reflections are given in Table 1, would be thought to be the best possible result from direct methods for this protein. Note that the values of origin-defining reflections for maximum entropy, but Babinet-shifted, and one of the symbolic addition origin-defining reflections, are retained with their original values, as is the suggested Σ_1 phase. On the other hand, the symbolic addition origin definer, φ_{630}, which did not appear on the overlap-reflection list for combining the separate solutions, is now shifted by π.

Table 1. *Phases, φ(electron diffraction), of 24 most intense reflections for hydrated catalase after structure-factor calculation*

The phases φ(electron diffraction) are compareed to another phase set, φ(Taylor, refined), from the extended image-based set, obtained after refinement *via* structure-factor calculation.

*hk*0	$\lvert F_o \rvert$	φ(electron diffraction)	φ(Taylor, refined)
0 2 0	6.87	0	0
0 4 0	8.46	π	π
0 6 0	6.42	π	π
0 10 0	3.03	π	0
0 16 0	3.73	π	0
0 18 0	4.72	π	π
1 1 0	4.37	0	0
1 4 0	5.25	π	π
1 7 0	4.59	π	π
1 11 0	3.34	π	π
1 14 0	3.12	π	π
2 4 0	3.39	0	0
2 6 0	3.10	π	π
3 2 0	3.12	0	π
3 5 0	3.13	0	π
3 7 0	3.23	π	π
3 14 0	3.17	0	π
4 7 0	3.01	π	π
5 4 0	3.24	π	π
5 5 0	3.66	0	π
6 0 0	3.08	π	π
6 3 0	3.14	π	π
6 4 0	3.13	π	π
7 3 0	3.18	π	π

4. Discussion

While the magnitude of the Patterson correlation coefficient for the phase solution in Table 1 compares well with results from the recent analysis of a membrane protein (Dorset & Jap, 1998), for which helix positions were accurately located, how can this solution otherwise be justified? As shown by Unwin (1975), as well as Akey & Edelstein (1983), the potential maps given for orthorhombic catalase after analysis of images of preparations embedded in negative or positive stains, glucose or other media are quite variable in appearance so that it is difficult to guess what is a correct structure. [Correspondingly, there are significant differences at 23 Å resolution for image-derived crystallographic phases (Akey & Edelstein, 1983) from preparations embedded in various media.] Radiation damage can also affect the density distribution in the average image (Unwin, 1975; Akey & Edelstein, 1983).

A 9 Å resolution density map of glucose-embedded catalase has been published by Unwin & Henderson (1975). Although there are some density features that may also fit our solution for hydrated catalase, there are also significant differences. Extraction of peak positions from the glucose-embedded structure for structure-factor calculation reveals that the Patterson correlation ($C = 0.35$) is poorer than the solution found in this analysis, consistent with the observed differences in the electron diffraction patterns.

At this point, results from the earlier investigation of the frozen hydrated structure (Taylor, 1978) were consulted. (They had not been consulted during the course of the direct phase analysis.) A comparison of phases to 23 Å resolution from the image analysis of the frozen hydrated specimen to the phases obtained from a structure-factor calculation from the glob model is given in Table 2. After shift of the image-derived phase set to another permissible origin, in accord with the one used for direct phase determination by maximum entropy and likelihood, there are only six errors in a list of 18 reflections, none of which is associated with the strongest reflections. The map generated from the phases listed by Akey & Edelstein (1983) for the frozen hydrated structure is shown in Fig. 2(*a*). The map from the glob structure-factor phases, including the complete list of reflections indicated in the *CRISP* phase determination from Taylor's (1978) image (agreeing with the original published list but adding a few more reflections), is shown in Fig. 2(*b*). These density distributions are virtually indistinguishable.

The similarity of the derived structure to the one found from image analysis of frozen hydrated specimens

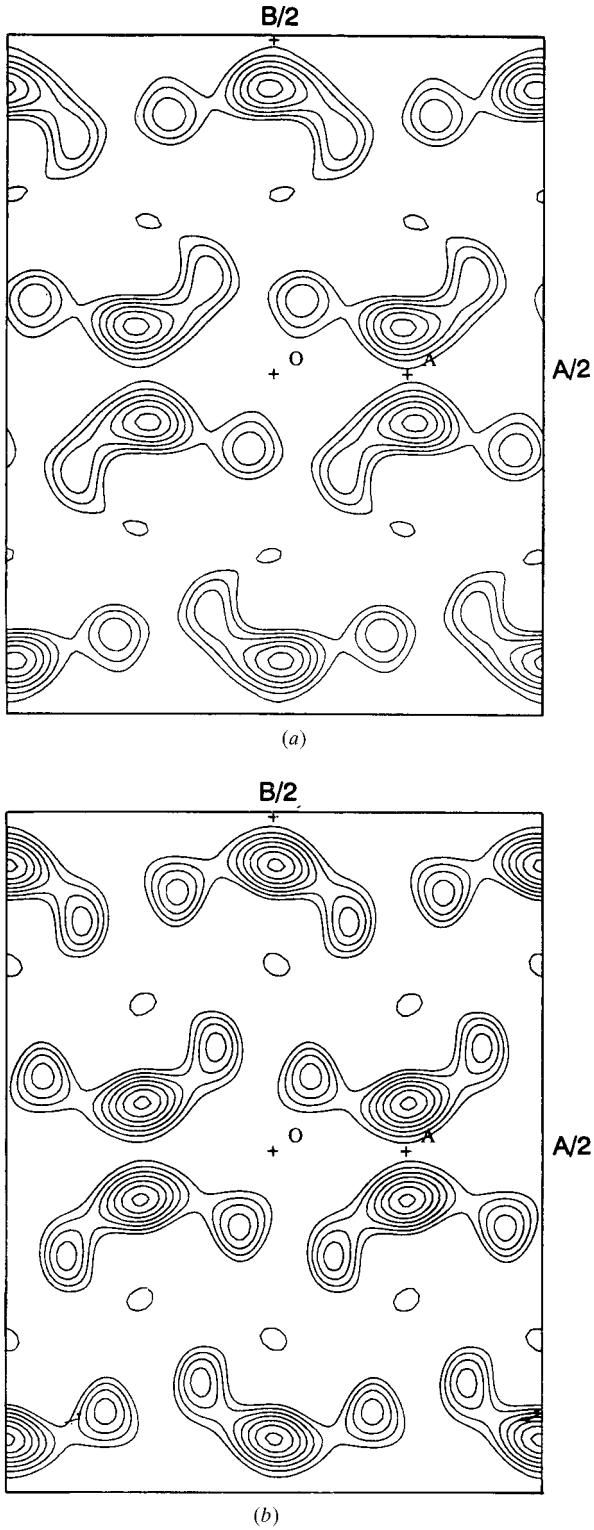

Fig. 2. Potential maps from 23 Å resolution phases. (*a*) Image-derived phases from the list given by Akey & Edelstein (1983) based on Taylor's (1978) work. The original origin definition is used. (*b*) Phases from the direct phasing model after an origin shift.

Table 2. *Comparison of image-derived phases, φ(Taylor), from frozen hydrated catalase to structure-factor phases, φ(electron diffraction), calculated from the direct phasing model in Fig. 1(d)*

*hk*0	$\|F_o\|$	φ(electron diffraction)†	φ(Taylor)‡
0 2 0	6.87	0 0	0
0 4 0	8.46	π π	π
0 6 0	6.42	π π	π
0 8 0	1.95	π π	(0)
1 1 0	4.37	0 0	0
1 2 0	2.56	0 0	0
1 3 0	2.95	0 0	π
1 4 0	5.25	π π	π
1 5 0	0.50	0 0	0
1 6 0	0.80	π 0	0
1 7 0	4.59	π π	0
2 0 0	2.81	0 0	0
2 1 0	2.65	0 0	0
2 2 0	1.65	0 0	π
2 3 0	1.88	π 0	0
2 4 0	3.39	0 0	π
2 5 0	0.42	π π	π
2 6 0	3.10	π π	π
2 7 0	1.64	π π	(π)
2 8 0	0.53	π π	(π)
3 1 0	2.75	0 0	(0)
3 2 0	3.12	0 π	π
3 3 0	2.71	π π	(π)
3 4 0	1.92	π π	(π)

† First column, phases from structure factors, using globs in Fig. 1(*d*); second column, after one cycle of Fourier refinement. ‡ Phases in parentheses are from the *CRISP* analysis of an electron-microscope image of a frozen hydrated crystal published by Taylor (1978). Other phases are from the *CRISP* analysis and/or a list given by Akey & Edelstein (1983), which are in total agreement.

could be demonstrated in another way. The Fourier transform of the 23 Å resolution image was used as a basis set for extension to the 9 Å resolution electron diffraction amplitudes by the Sayre–Hughes equation (Sayre, 1980), $E_{\mathbf{h}} = N^{1/2}\langle E_{\mathbf{k}} E_{\mathbf{h-k}}\rangle$. After these new phases were generated, the match with values for the 24 intense reflections used to calculate the Patterson correlation coefficients above was satisfied by 15 data, with the differences expressed in medium-intensity reflections but not for the strong ones. The accuracy of the phase extension is similar to the result found for electron diffraction data from other proteins from a lower-resolution basis (Gilmore *et al.*, 1993; Dorset *et al.*, 1995; Dorset, 1996) or from X-ray data from a transfer RNA, starting with an incomplete MIR phase set (Podjarny *et al.*, 1981).

If these extended phases were used to calculate a potential map, and peak positions from this map were chosen as pseudo-atoms to calculate structure factors, the phase agreement was found for 18 out of 24 reflections (Table 1). The expanded image-derived model was then compared to the phases from the structure-factor calculation from peaks in Fig. 1(*d*). Note that some of

Table 3. *Final listing of amplitudes and phases for refined structural model*

Phases from Fourier transform of Fig. 1(*d*).

*hk*0	$\|F\|$	φ	*hk*0	$\|F\|$	φ	*hk*0	$\|F\|$	φ
0 2 0	6.87	0	2 12 0	0.85	0	4 14 0	2.05	π†
0 4 0	8.46	π	2 13 0	0.45	0†	4 15 0	1.19	0†
0 6 0	6.42	π	2 14 0	2.46	0	4 16 0	1.02	π†
0 8 0	1.95	π†	2 15 0	2.11	0†	5 1 0	1.49	π
0 10 0	3.03	π†	2 17 0	0.75	0	5 2 0	0.87	0†
0 12 0	0.72	0	2 18 0	1.09	0†	5 3 0	1.33	π†
0 14 0	2.12	0†	2 19 0	0.46	π	5 4 0	3.24	π
0 16 0	3.73	π†	3 1 0	2.75	0†	5 5 0	3.66	0†
0 18 0	4.72	π	3 2 0	3.12	0†	5 6 0	1.34	π†
1 1 0	4.37	0	3 3 0	2.71	π	5 7 0	0.68	π†
1 2 0	2.56	0	3 4 0	1.92	π	5 8 0	1.39	0
1 3 0	2.95	0	3 5 0	3.13	0†	5 9 0	2.63	0
1 4 0	5.25	π	3 6 0	2.70	π†	5 10 0	1.40	0
1 5 0	0.50	0	3 7 0	3.23	π	5 11 0	2.40	0†
1 6 0	0.80	π	3 8 0	2.60	π†	5 12 0	1.04	π†
1 7 0	4.59	π	3 9 0	2.16	0	5 13 0	0.97	π†
1 9 0	2.05	π	3 10 0	0.75	0	5 14 0	1.10	π†
1 10 0	1.58	0	3 11 0	0.30	π†	5 15 0	1.02	π†
1 11 0	3.34	π	3 12 0	1.01	π	6 0 0	3.08	π
1 12 0	0.62	0	3 13 0	1.19	0	6 1 0	0.24	0†
1 13 0	1.07	π	3 14 0	3.17	0†	6 2 0	0.53	π
1 14 0	3.12	π	3 15 0	1.10	0†	6 3 0	3.14	π
1 15 0	0.59	0	3 16 0	0.78	0†	6 4 0	3.13	π
1 16 0	0.73	0†	3 17 0	0.50	π†	6 5 0	2.44	0
1 17 0	1.12	0	3 18 0	0.53	π	6 6 0	2.80	0
1 18 0	1.08	0	4 0 0	2.86	0	6 7 0	2.63	π†
1 19 0	0.63	π	4 1 0	2.69	π†	6 8 0	2.77	0
2 0 0	2.81	0	4 2 0	2.14	π	6 9 0	2.90	π†
2 1 0	2.65	0	4 3 0	0.88	π	6 10 0	2.22	0†
2 2 0	1.68	0†	4 4 0	2.57	π	6 11 0	1.14	0†
2 3 0	1.88	π	4 5 0	2.67	0	6 12 0	1.25	π
2 4 0	3.39	0	4 6 0	2.63	π	7 1 0	2.30	π†
2 5 0	0.42	π†	4 7 0	3.01	π	7 2 0	2.26	π†
2 6 0	3.10	π	4 8 0	2.24	π†	7 3 0	3.18	π
2 7 0	1.64	π	4 9 0	2.71	π	7 4 0	1.99	0
2 8 0	0.53	π†	4 10 0	2.41	0†	7 5 0	0.75	0†
2 9 0	0.74	π	4 11 0	1.33	0†	7 6 0	2.30	0†
2 10 0	0.98	π	4 12 0	1.92	0	7 7 0	0.40	0
2 11 0	0.47	0†	4 13 0	2.31	0			

† Differences found in refinement with Sayre expansion of image phases.

the phases from the Taylor (1978) model changed value after this refinement. The Patterson correlation coefficient was then $C = 0.79$. A complete list of phases is given in Table 3 with associated structure-factor amplitudes. Comparing those from the complete Fourier transform of Fig. 1(*d*) to the refined 9 Å extension of Taylor's 23 Å resolution image-based phase set, the mean phase difference is 74.5° for 116 reflections.

5. Conclusions

This direct analysis of a globular protein crystal structure from electron diffraction data indicates that the solution of an 'unknown' structure is indeed possible if a rather conservative approach is used for accepting structure solutions, after comparing results from two independent methods for phase determination. It also requires that the pseudo-atomic globular transform model is appropriate for the molecule. As seen in earlier analyses, the Luzzati figure of merit has approximate validity and is appropriate for screening choices among multiple solutions. The Patterson correlation coefficient is also a useful indicator for choosing one solution over another. After the assignment of phase values, there was a reasonable agreement with the phase model found from image analysis of frozen hydrated specimens.

Research was supported in part by a grant from the National Institute for General Medical Sciences (GM-46733). DLD thanks Dr D. F. Parsons for suggesting this problem a quarter of a century ago.

References

Akey, C. W. & Edelstein, S. J. (1983). *J. Mol. Biol.* **163**, 575–612.

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

Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–445.

Bricogne, G. & Gilmore, C. J. (1990). *Acta Cryst.* A**46**, 284–297.

Burmester, C., Holmes, K. C. & Schröder, R. R. (1995). *Proceedings, Microscopy and Microanalysis 1995*, edited by G. W. Bailey, M. H. Ellisman, R. A. Henniger and N. J. Zaluzec, pp. 856–857. New York: Jones and Begill.

Cowan, S. W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R. A., Jansonius, J. N. & Rosenbusch, J. P. (1992). *Nature (London)*, **358**, 727–733.

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

Dorset, D. L. (1995*b*). *Structural Electron Crystallography*, pp. 109–119. New York: Plenum.

Dorset, D. L. (1996). *Acta Cryst.* A**52**, 480–489.

Dorset, D. L. (1997*a*). *Acta Cryst.* A**53**, 445–455.

Dorset, D. L. (1997*b*). *Proc. Natl Acad. Sci. USA*, **94**, 1791–1794.

Dorset, D. L. (1998). *Acta Cryst.* A**54**, 290–295.

Dorset, D. L. & Jap, B. K. (1998). *Acta Cryst.* D**54**, 615–621.

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

Dorset, D. L. & Parsons, D. F. (1975*a*). *Acta Cryst.* A**31**, 210–215.

Dorset, D. L. & Parsons, D. F. (1975*b*). *J. Appl. Cryst.* **8**, 12–14.

Dorset, D. L. & Parsons, D. F. (1975*c*). *J. Appl. Phys.* **46**, 938–940.

Drenth, J. (1994). *Principles of Protein X-ray Crystallography*, p. 224. New York: Springer.

Gilmore, C. J. (1984). *J. Appl. Cryst.* **17**, 42–46.

Gilmore, C. J. (1996). *Acta Cryst.* A**52**, 561–589.

Gilmore, C. J. & Bricogne, G. (1997). *Methods Enzymol.* **277**, 65–78.

Gilmore, C. J., Marks, L. D., Grozea, D., Collazo, C., Landree, E. & Twesten, R. (1997). *Surf. Sci.* **381**, 77–91.

Gilmore, C. J., Nicholson, W. V. & Dorset, D. L. (1996). *Acta Cryst.* A**52**, 937–946.

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

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

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

Jap, B. K., Walian, P. J. & Gehring, K. (1991). *Nature (London)*, **350**, 167–170.

Jap, B. K., Zulauf, M., Scheybani, T., Hefti, A., Baumeister, W., Aebi, U. & Engel, A. (1992). *Ultramicroscopy*, **46**, 45–84.

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

Kimura, Y., Vassylyev, D. G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., Murata, K., Hirai, T. & Fujiyoshi, Y. (1997). *Nature (London)*, **389**, 206–211.

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

Matricardi, V. R., Moretz, R. C. & Parsons, D. F. (1972). *Science*, **177**, 268–270.

Pauling, L. & Corey, R. B. (1951). *Proc. Natl Acad. Sci. USA*, **37**, 282–285.

Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P. & Landau, E. M. (1997). *Science*, **277**, 1676–1681.

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

Podjarny, A. D. & Urzhumtsev, A. G. (1997). *Methods Enzymol.* **276**, 641–658.

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.

Schröder, R. R. & Burmester, C. (1993). *Proceedings 51st Annual Meeting of the Microscopy Society of America*, edited by G. W. Bailey & C. L. Rieder, pp. 666–667. San Francisco: San Francisco Press.

Shankland, K., Gilmore, C. J., Bricogne, G. & Hashizume, H. (1993). *Acta Cryst.* A**49**, 493–501.

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

Sumner, J. B. & Dounce, A. L. (1937). *J. Biol. Chem.* **121**, 417–424.

Taylor, K. A. (1978) *J. Microsc. (Oxford)*, **112**, 115–125.

Taylor, K. A. & Glaeser, R. M. (1974). *Science*, **186**, 1036–1037.

Unwin, P. N. T. (1975). *J. Mol. Biol.* **98**, 235–242.

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

Vainshtein, B. K., Sherman, M. B. & Barynin, V. V. (1976). *Sov. Phys. Crystallogr.* **21**, 287–291.

Valentine, R. C. (1964). *Nature (London)*, **204**, 1262–1264.

Wrigley, N. G. (1968). *J. Ultrastruct. Res.* **24**, 454–464.