Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 4 April 2000
Accepted 28 June 2000

Low-resolution direct phase determination in protein electron crystallography – breaking globular constraints

Douglas L. Dorset

Structural Biology Department, Hauptman-Woodward Medical Research Institute, 73 High Street, Buffalo, New York 14203-1196, USA. Correspondence e-mail: dorset@hwi.buffalo.edu

Although the assumption of an overall globular scattering entity can be useful for determining crystallographic phases for a protein at low resolution, there is a point where this pseudoatomic model must be abandoned for further phase refinement. Using 6 Å resolution electron diffraction data from aquaporin (AQP-CHIP) as an example, phases of the 16 most intense reflections from a previous direct solution (Dorset & Jap (1998). *Acta Cryst.* D**54**, 615–621) were modified with a Hadamard error-correcting code to produce potential maps very similar to the ones obtained using phases from the Fourier transform of averaged electron micrographs. The choice of the optimal phase set was made *via* the cross correlation of experimental with anticipated density histograms using the autocorrelation function of the latter histogram as the desired endpoint.

1. Introduction

In recent years, there has been considerable interest in solving the crystallographic phase problem for macromolecules at low resolution. One reason for this interest has been the need to define the molecular-envelope boundary accurately for interpretation of high-resolution structural maps (Podjarny & Urzhumtsev, 1997). One important direct phasing approach explores a logic tree wherein likely solution pathways can be identified by a suitable figure of merit such as maximum likelihood (Gilmore, Shankland & Bricogne, 1993; Gilmore *et al.*, 1996, 1999). In other approaches to the phasing problem, either in image or diffraction space, globular approximates to the density have been exploited (Anderson & Hovmöller, 1996; Lunin *et al.*, 1990, 1995, 1998; Dorset, 1997*a*), based on Harker's (1953) argument that the phenomenological scattering factor of identified globular subunits should normalize the diffracted intensities more accurately than the sum of the component atomic scattering factors.

In this laboratory (Dorset, 1997*a*,*b*, 1998, 2000*a*,*b*,*c*), direct methods, also constrained by an assumed globular macromolecular substructure, have been exploited in reciprocal space, with observed diffraction amplitudes out to 6 Å resolution. The resolution limit is near a point where a plot of averaged intensity from macromolecular crystals in shells of reciprocal space *versus* increasing resolution reaches a minimum (Pauling & Corey, 1951; Rossmann *et al.*, 1967). It is known, first of all, that traditional direct methods are, on average, valid within this diffraction intensity envelope (Podjarny *et al.*, 1981; Gilmore, Shankland & Fryer, 1993; Dorset *et al.*, 1995). Secondly, if the subunit globular cross section is similar to that of an α-helix, then the pseudoatomic scattering factor can be well treated by the scattering factor of a single atom after the dimensions of the diffraction problem (*e.g.* unit-cell constants) are reduced tenfold (Dorset, 1997*a*). This also assumes that the globular cross section is approximately Gaussian so that its nearly Gaussian scattering factor can be modeled reasonably well by the Lorentzian shape of a typical atomic scattering factor after the dimensional rescaling.

The success of this rather simplistic diffraction model for the analysis of low-resolution protein phasing problems has been surprisingly good. In obvious applications, where the view down α-helix axes immediately suggests the pseudoatomicity of the phasing model, traditional direct methods have, in two centrosymmetric projections, respectively, of halorhodopsin and orthorhombic bacteriorhodopsin, produced map density profiles that are virtually identical in appearance to the potential maps calculated from electron diffraction amplitudes and crystallographic phases derived from the Fourier transform of averaged electron micrographs (Dorset, 1997*a*, 1998). For two non-centrosymmetric projections of bacteriorhodopsin, reasonably accurate phases can be found by the Sayre equation, thereby locating the seven α-helix sites in the asymmetric unit (Dorset, 1997*b*, 2000*b*,*c*). Extension of these methods to X-ray data sets from two small proteins (crambin and monoclinic rubredoxin) has shown again that the centrosymmetric projections can be determined rather easily and even that some progress can be made to determine directly the three-dimensional electron-density profile (Dorset, 2000*a*).

There is no reason, however, to suppose that the pseudo-atomic model should be rigorously valid for such low-resolution phase determinations, even when projections of protein secondary structure would suggest that this initial assumption might be useful. For example, in the direct phase determination of a red blood cell membrane protein, aquaporin channel-forming integral membrane protein (AQP-CHIP), projected helix centers could be located within ±1.9 Å, even though it could not be claimed that the final map (Fig. 1*b*) otherwise closely resembled the one (Fig. 1*a*) calculated from the electron-micrograph-derived phases (Dorset & Jap, 1998).

In this paper, a possible method for breaking out of the initial globular symmetry imposed on the initial direct phase determination will be described for this protein.

2. Materials and methods

2.1. Protein data and initial phase determination

While *hk*0 electron diffraction patterns from untilted reconstituted two-dimensional crystals of AQP-CHIP extend to about 3.0 Å, when the samples are embedded in vitreous ice (Jap & Li, 1995) only the 106 unique data to 6 Å were considered in the initial phase determination (for reasons given in the *Introduction*). The potential map using phases from averaged low-dose electron micrographs of the [001] projection with electron diffraction amplitudes is shown in Fig. 1(*a*). The three-dimensional space group is $P42_12$, so that the projected centrosymmetric plane group is *p*4*mg*. Unit-cell constants are $a = b = 96.4\,(2)$ Å.

Details of the original direct phase determination are given in a previous paper (Dorset & Jap, 1998). The phase of one reflection had been determined from a highly probable Σ_1 triple (10,0,0) and the origin was set by defining $\varphi_{580} = 0$. Symbolic addition (Karle & Karle, 1966) *via* algebraic phase unknowns with the two defined phases for the centrosymmetric projection determined new phase terms from the ten most probable Σ_2-invariant sums. With the criterion of map density flatness (Luzzati *et al.*, 1986) used to reduce the number of solutions to be evaluated, the best solution was then identified by a Patterson correlation coefficient (Drenth, 1994). With identified peak positions used as pseudo-atomic sites, the model, refined by Fourier methods, gave an overall mean phase error of 29° for the 25 most intense reflections. This can be compared to a mean 14.4° error when helix sites from the image-derived potential map were used to calculate structure factors *via* the pseudo-atomic scattering factors, and really reflects some problems with the original phase determination.

2.2. Phase refinement *via* an error-correcting code

It is reasonable to expect that the most significant deviations from the pseudo-atomic model imposed on the initial phase determination should be found in the most intense reflections in the data set, since their weighting of incorrect phases would account for the differences noted between the maps in Figs. 1(*a*) and (*b*). If the phase set from the direct determination can be regarded as a distorted signal, then, using techniques borrowed from communications theory (Brillouin, 1962; Peterson & Wilson, 1972; Anderson, 1974; MacWilliams & Sloane, 1977; Hill, 1986), a better approximation of the correct signal might be found using error-correcting codes. This is particularly advantageous for a centrosymmetric projection since the aggregate of $0,\pi$ phase choices for a set of reflections can be considered as a train of 1 or −1 pulses to be corrected by the code. The application of error-correcting codes to the crystallographic phasing problem was pioneered by Bricogne (1993). It was further explored by Gilmore *et al.* (1999) who, in their paper, have discussed the theoretical background and its possible applications to representative structural problems. Generally, a matrix of polarity corrections is applied to the pulse train (phase set). For example, a variant of the symmetric 8×8 Hadamard matrix

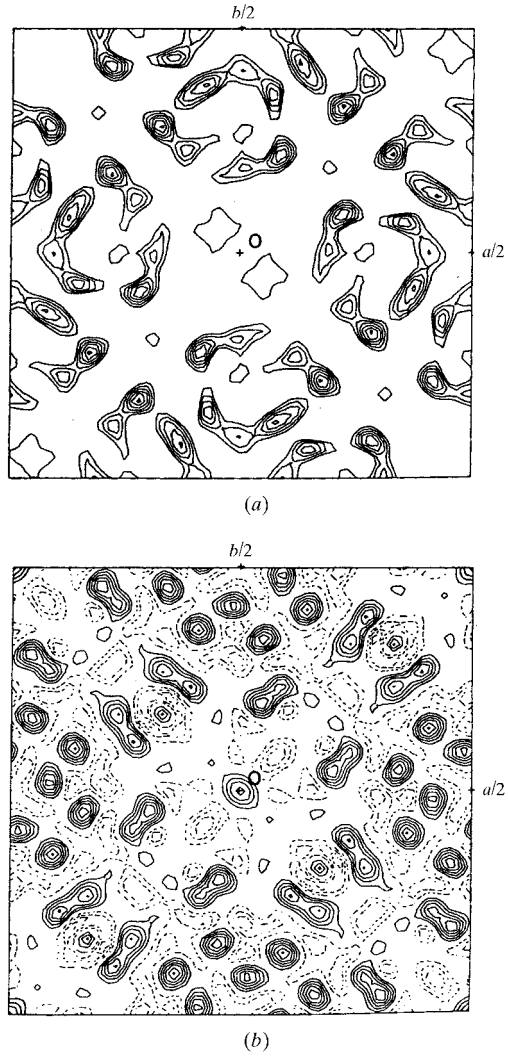

Figure 1
Potential maps for AQP-CHIP. (*a*) Crystallographic phases obtained from the Fourier transform of averaged electron-microscope images (Jap & Li, 1995). (*b*) Crystallographic phases from direct methods assuming pseudo-atomic substructure (Dorset & Jap, 1998).

$$H = \begin{vmatrix} M & M \\ M & \bar{M} \end{vmatrix},$$

where

$$M = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & \bar{1} & 1 & \bar{1} \\ 1 & 1 & \bar{1} & \bar{1} \\ 1 & \bar{1} & \bar{1} & 1 \end{vmatrix},$$

is a generator matrix and has useful properties when combined with its negative array to construct a perfect code as an 8×16 matrix (MacWilliams & Sloane, 1977; Hill, 1986)

$$T = \begin{vmatrix} H \\ \bar{H} \end{vmatrix}.$$

The resultant Hadamard [8,4,4] code, tabulated in the paper by Gilmore *et al.* (1999), has the property that one corrected set will contain a maximum of two errors for one of the solutions (or there may be none). [With the parenthetic notation (n,k,d) for any code, n is the number of characters (*i.e.* phase terms accessed) in a code word, k is the number of rows in a generator matrix (Gilmore *et al.*, 1999) and d is a minimum Hamming distance (MacWilliams & Sloane, 1977; Hill, 1986) accounting for the smallest character difference between code words. From the latter quantity, the error detection and correcting power of any code may be derived.] In terms of phase permutations, then $2^8 = 256$ for 8 reflections (characters in a code word) reduces to just 16 overall phase adjustments. If the code is 'punctured', *i.e.* the last column is omitted (MacWilliams & Sloane, 1977), then seven phases can be corrected in the Hadamard [7,4,3] code, with a solution containing a maximum of only one error, again with 16 permutations instead of $2^7 = 128$. As discussed by Gilmore *et al.* (1999), there are other codes (see also MacWilliams & Sloane, 1977) sufficient for handling larger phase sets (*e.g.* Nordström–Robinson, Golay), significantly reducing the number of required phase sets to be sampled and also guaranteeing one solution that has maximally one phase error. (They also discuss how these codes can be applied to non-centrosymmetric sets.)

2.3. Identification of the best phase solution

Ordinarily, error-correcting codes ensure that a noisy transmission of a binary signal will not lead to a mistaken interpretation of the message. For this, the signal and a parity check are embedded in the same code word (*e.g.* MacWilliams & Sloane, 1977). In the crystallographic application, the entire code sequence becomes a phase-correction device. Thus, the identity of the best phase solution, chosen from a group of 16 code words for a cluster of 7 or 8 phases, must still be established since the assembly of code words merely represents a reduced number of phase permutations, many of which will still contain many errors. In the present work, this identification has been made by exploiting the properties of cross-correlation functions (Gaskill, 1978), comparing experimental density histograms, assembled from maps generated for a particular phase set, to the histogram expected for a low-resolution map from a typical protein (Zhang & Main, 1990; Lunin, 1993). This concept is also adapted from communications theory or signal analysis (Mason & Zimmermann, 1960; Raemer, 1969; Balmer, 1991). It is expected that the density histograms $v(t)$ for potential maps (electron crystallographic determinations) should not be qualitatively different from those of electron-density maps (X-ray crystallographic determinations) as recent studies indicate (Dorset, 2000*a*,*b*,*c*).

Given an experimental map density histogram $v_1(t)$ and its expected value $v_2(t)$, then the cross-correlation function can be defined as

$$\psi_{12}(\tau) = \int_{-\infty}^{+\infty} v_1(t+\tau)v_2(t)\,\mathrm{d}t. \tag{1}$$

If the expected value can be described *a priori*, then the endpoint of a search for a correct match of $v_1(t)$ to $v_2(t)$ should be revealed by the properties of the autocorrelation function:

$$\psi(\tau) = \int_{-\infty}^{+\infty} v(t+\tau)v(t)\,\mathrm{d}t. \tag{2}$$

As is well known, the autocorrelation function is an even function, maximally peaked at the origin (Mason & Zimmermann, 1960; Gaskill, 1978; Frank, 1980). On the other hand, the cross-correlation function is not necessarily an even function, but it should more and more resemble the autocorrelation function as $v_1(t)$ and $v_2(t)$ become more similar. The property of skewness in the cross correlation

$$S = \sum_i |\psi_{12}(\tau_i) - \psi_{12}(-\tau_i)|/\psi_{12}(0) \tag{3}$$

measuring across some pre-determined origin position, therefore, is a possible way of distinguishing the dissimilarity of the two correlated functions (Dorset, 2000*a*,*b*) after they have been aligned to one another by an overall shift γ. The value of γ can be found directly by plotting the cross correlation so that the (skewed) function can be centered at its maximum peak height, where the greatest similarity should be observed for two compared functions (Frank, 1980). (An example of cross-correlation functions for density histograms with origins that are mutually offset from one another by different amounts is shown in Fig. 2(*f*). The values γ required to align peak positions differ, even though the overall cross-correlation functional shapes remain unchanged.) The shift value can then be applied to the original density histograms $v_1(t)$ and $v_2(t)$ to place them on a common basis so that a similarity function

$$E(\gamma)) = \left\{ \frac{\sum_j [v_1(r_j) - v_2(r_j+\gamma)]^2}{\left[\sum_j v_1^2(r_j) \sum_j v_2^2(r_j)\right]^{1/2}} \right\}^{1/2} \tag{4}$$

for the density histogram can be computed for the appropriate shift value γ (Frank, 1980).

In the above, the integrals of the correlation functions are approximated by summations using histograms sampled (sifted) by a train of δ functions (*i.e.* a 'comb' function). The summations, moreover, are not entirely accurate representations of the above integrals (DeFelice, 1981; Balmer, 1991).

One significant source of error can arise if individuals in pairs of cross-correlated density histograms are not sampled by comb functions with the same interval between δ functions. Analysis of a cross-correlation function is permitted if the more coarsely sampled histogram is spaced appropriately with respect to the finely sampled histogram.

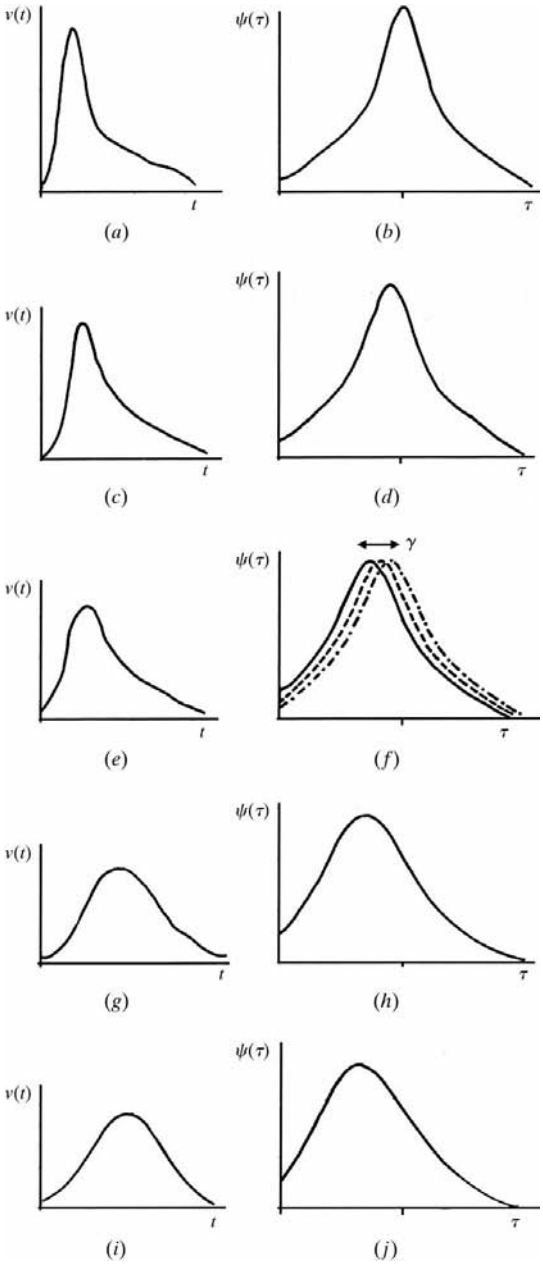

Figure 2
Density histograms (Lunin, 1993) and their cross-correlation functions. (*a*) 0° mean phase error. (*b*) Cross correlation of (*a*) with itself (autocorrelation function). (*c*) Average of histograms generated for 0 and 15° mean phase errors. (*d*) Cross correlation of (*a*) with (*c*). (*e*) 15° mean phase error. (*f*) Cross correlation of (*a*) with (*e*). Here different relative origin shifts between (*a*) and (*e*) are imposed before cross correlation to demonstrate that the otherwise invariant functional shape also shifts position. (*g*) 36° mean phase error. (*h*) Cross correlation of (*a*) with (*g*). (*i*) 90° mean phase error. (*j*) Cross correlation of (*a*) with (*i*).

Lunin (1993) depicted idealized density histograms for proteins at various mean phase errors. Representative histograms from this earlier work are shown in Fig. 2 along with their respective cross correlations with the error-free histogram. Since the widths of the autocorrelation functions of the individual density histograms increase at half height with increasing phase error, the cross correlations of the error-free histogram to those experimental distributions for maps calculated from erroneous phase sets also increase in width as the experimental phase error increases. (Here the reference origin for the cross-correlation function is determined by an optimal shift value, again located at its peak value.) The skewness parameter also increases for cross correlations with histograms from erroneous maps as the mean error increases. Finally, there is an increase in value of the similarity function, E, with the phase error when the density histograms are shifted with respect to one another by the γ value indicated in the cross correlation.

3. Results

3.1. Density histograms

If, as in earlier work (Dorset, 2000*a*,*b*,*c*), an approximate value of F_{000} was used to generate potential maps where negative density regions were minimized, the resultant density histogram generated for the correct (image-based) crystallographic phase set did not at all resemble the skewed distribution shown in Fig. 2(*a*) but was very much like the nearly Gaussian histogram anticipated for the mean value of 90° (Fig. 2*i*), *i.e.* a random phase set. No great difference was seen for histograms generated from trial phase solutions. Therefore, a phase model could not be identified.

The discrepancy between anticipated and experimental density histograms was the result of including pixel values within solvent regions, consistent with the need for coupling solvent flattening with histogram matching (Zhang & Main, 1990). The 'generic' quality of the histogram depicted for example in Fig. 2(*a*) is, therefore, only valid for pixel values within the macromolecular envelopes (Lunin, 1993). New maps were then generated using $F_{000} = 0$ so that the mean density value of the map would also be 0.0. Only pixel values within the positive areas of the map were used to generate the test histograms. Since actual histograms should also contain a small sample of pixels within negative density regions (Zhang & Main, 1990), it is anticipated that there should be some error in the generated density distributions. Nevertheless (Fig. 3), there is a skewed density distribution within the positive part of the experimental histogram that somewhat resembles the anticipated envelope.

3.2. Most intense reflection set

In order to analyze the most intense set of eight reflections, the condition $\varphi_{580} = 0$ (origin definition) was imposed and the remaining seven phases were modified by a punctured Hadamard [7,4,3] code. (The full Hadamard code might have been used equally well, adding another reflection. However, in

this study, two blocks of eight phases each were considered. The first set already includes an origin-defining phase, hence the choice of the punctured code.) The symbolic identities, $\varphi_{150} = \varphi_{340}$ and $\varphi_{350} = \varphi_{570}$, found in the original direct phase analysis (Dorset & Jap, 1998) were also imposed. One set of eight possible solutions could be rejected by violation of these algebraic phase relationships alone. After generating potential maps for the remaining eight phase sets, their density histograms were cross correlated to the positive part of the histogram anticipated (Lunin, 1993) for a solution with a mean phase error between 0 and 15° (Figs. 2*c*, 3*a*), suitably scaled to match nearly the peak magnitudes. From the skewness parameter *S*, defined above, it is apparent that the best solution corresponded to the case where no phase modifications were made by the error-correcting code. This, in fact, verified the earlier observation (Dorset & Jap, 1998) that the eight most intense reflections were correctly assigned phase values by direct methods. In other words, despite the difference in the histograms shown in Fig. 3, the 'generic' estimate in Fig. 3(*a*) is good enough to find the correct phase solution for the most intense reflections.

3.3. Second most intense reflection set

The eight most intense reflections were then assigned invariant phases and the values for the next eight most intense group generated by direct methods were corrected by the Hadamard [8,4,4] code. Again, 16 phase sets (corrected by code words) were evaluated.

When potential maps were generated for the 16 phase combinations and the experimental density histograms (positive pixel values only for maps generated where $F_{000} = 0$) were cross correlated to the Lunin (1993) test histogram used above (Figs. 2*c*, 3*a*), no decision about the best phase set could be made if the skewness property of this function, *S*, or the similarity function, *E*, were consulted as FOM's. A density histogram (Fig. 3*b*) determined from maps calculated from the correct set of phases [Fourier transform of the electron micrographs determined by Jap & Li (1995)] was then cross correlated with the experimental histograms. In this case, the best solution, containing only one phase error (Table 1) was clearly identified, both by the skewness parameter *S* as well as by the similarity term *E*. The corresponding potential map (Fig. 4*b*) was then much more similar in appearance to Fig. 1(*a*) than it was to Fig. 1(*b*).

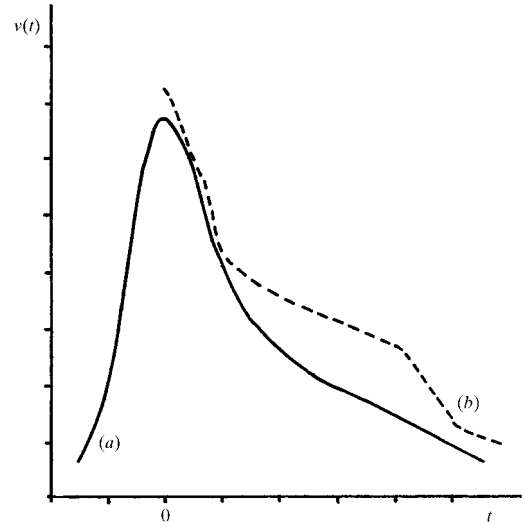

Figure 3
Comparison of predicted (Lunin, 1993) and actual density histograms. (*a*) Predicted histogram for average of 0 and 15° mean phase error. (*b*) Density histogram for map calculated from image-based crystallographic phases.

Table 1
Results of direct phasing of AQP-CHIP, corrected by Hadamard code.

*hk*0	|*F*|	φ (best single solution)	φ (average of solutions)
150	99.63	0	0
340	72.18	0	0
350	70.23	0	0
570	77.99	0	0
580	77.93	0	0
670	68.43	π	π
680	62.39	π	π
770	64.38	0	0
060	55.63	π	π
160	52.73	π	π
250	45.48	0	0
390	57.94	π	π
440	47.09	π	π
4,10,0	46.23	0	π†
660	43.63	0†	π
990	51.55	0	0

† Incorrect phase value.

In an alternative search for a reference density histogram, an average was made over all of the experimental histograms determined for the 16 phase sets. This reference was used in the search for a phase solution by the cross-correlation function. From the standpoint of the skewness parameter, a correct solution could be found in a cluster of four solutions within the smallest *S* values. In terms of the similarity parameter, it could be located within three solutions giving the smallest *E* value. An analogy exists to the case of globular model solutions generated in image space and then averaged according to the similarity of density histograms (Lunin *et al.*, 1990). The average of phase sets (most represented phase value for any reflection) in both tests with the average histogram gave the same result, containing only one phase error (but differing from the one generated above), as shown in Table 1. The potential map generated from these 16 phased reflections (Fig. 4*c*) again is similar to the image-based structure (Fig. 1*a*).

4. Discussion

The assumption of pseudo-atomic globular arrays in protein structures has imposed a useful constraint for extracting approximate low-resolution phase information by conventional direct methods. This assumption has also been useful for the Fourier refinement of such structures, since the globular array can be modeled conveniently by a pseudo-atom for structure-factor calculations, after dimensional re-scaling

(Dorset, 1997*a*). In order to refine these initial phase assignments effectively, however, some means must be found to break out of the overall globular scattering-factor constraint. As demonstrated by Bricogne (1993) and Gilmore *et al.* (1999), error-correcting codes appear to be an efficient device for testing a reasonable number of phase permutations with the least amount of error. The Hadamard error-correcting codes used in this example are obviously not the optimal one for reasons previously discussed by these authors, who prefer to use codes (*e.g.* Nordström–Robinson, Golay) with longer words. Nevertheless, the use of seven- or eight-character code sequences has led to a useful result where the directly derived structure is much closer to the one found in the original electron microscopical study (Jap & Li, 1995). It also has demonstrated that much relevant structural information is contained in a relatively small subset of intense reflections, for which the derived phases must be highly accurate – in this case, much more so than provided by conventional direct methods – to provide a faithful reproduction of the potential map. This is illustrated conveniently by the strong similarity between the maps in Fig. 4(*a*) and Fig. 1(*a*), the former generated from just 16 intense reflections and the latter from a complete phased data set. Further, the successful phase refinement *via* Hadamard codes demonstrates where the deviations from the initial globular model become significant. For this example, the deviations occur at the second level of intense reflections whereas the most intense ones conform well to the original pseudoatom model. Hence, it was informative to use shorter code words in this feasibility study.

Nevertheless, there are 16 code words to be evaluated in each level of phase correction. Any phasing procedure involving permutations also must permit the *a priori* identification of an optimal solution *via* some figure of merit. While the minimization of the required number of choices by an error-correcting code is highly useful, there are no conservative structural constraints, on the order of bonding parameters in small-molecule crystallography, that can be used to identify a 'correct' low-resolution protein structure just from its appearance. In earlier studies from this laboratory (Dorset, 1997*a*,*b*, 1998, 2000*a*,*b*,*c*) and elsewhere (Podjarny & Urzhumtsev, 1997; Lunin *et al.*, 1998), a number of FOM's have been suggested to facilitate this structure identification and all of them have been shown to fail to find a unique phase solution in some cases.

If a generic density histogram can be defined *a priori* to sufficient accuracy for a protein, then there is every reason to expect that the cross-correlation function of this expected histogram with experimental ones generated from various phase combinations should be useful for finding a solution, based on ideal match to the autocorrelation function for the correct phase solution. In this study, it was shown that the search with the ideal histogram for the projected protein structure, generated from the map calculated with electron-micrograph-derived phases, was a useful criterion for this search. However, a more 'generic' histogram, *e.g.* one proposed by Lunin (1993) (Fig. 2*a*), was not necessarily useful for the evaluation of the second block of intense reflections that account for actual structural deviations from the globular model. In previous work with 'generic' histograms (Dorset, 2000*a*,*b*,*c*), the cross correlation was made generally with histogram distributions from alternative maps based on total phase sets generated by the Sayre (1952) equation, after permuting an algebraic unknown. The current study demonstrates that the histogram changes for density profiles from phase permutations in a cluster of medium-density reflections may, in fact, be very subtle, especially when the dominant intensities are assigned correct phases. Thus, an attempt to identify the best solution with an imperfect approximate to the exact histogram model (*e.g.* Fig. 3*a*) can fail.

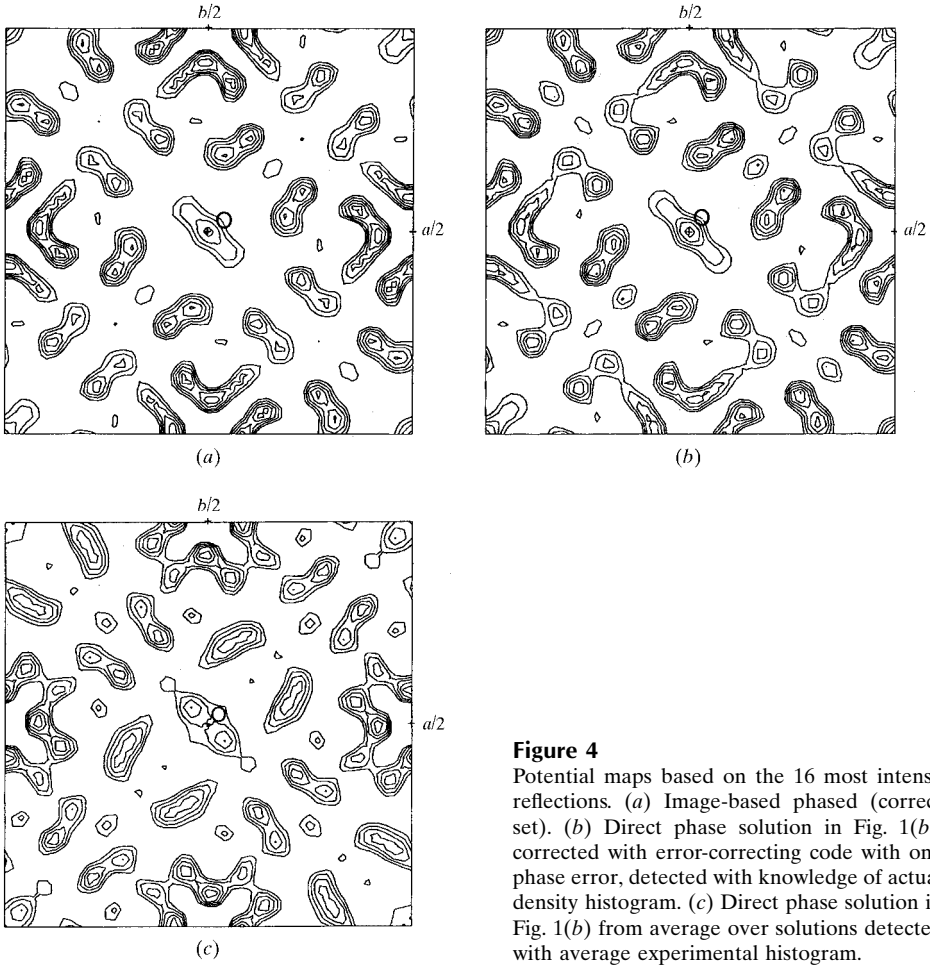

Figure 4
Potential maps based on the 16 most intense reflections. (*a*) Image-based phased (correct set). (*b*) Direct phase solution in Fig. 1(*b*), corrected with error-correcting code with one phase error, detected with knowledge of actual density histogram. (*c*) Direct phase solution in Fig. 1(*b*) from average over solutions detected with average experimental histogram.

An alternative procedure of finding a starting approximate histogram model by averaging the histograms generated for all phase choices may have some merit in further restricting the number of structural solutions to be tested. By analogy to the real-space globular model generations that are screened and sorted by a density histogram, the cluster of solutions might then be averaged to give an acceptable phase model. Previously it was assumed without proof (Dorset, 2000*a*,*b*,*c*) that the density histogram might be a reliable test of structural solutions even though other laboratories have shown that such searches can also go astray (Lunin *et al.*, 1990) in individual cases. The alternative approach with an approximate histogram model, generated by averaging over all possibilities from the phasing experiment, may, nevertheless, prove more useful in future applications if the most likely phase solutions are then averaged.

The author is grateful to Chris Gilmore for introducing him to the concept of error-correcting codes and to Harvey Fishman for fruitful discussions about correlation functions. Bing Jap is thanked for providing the original data set. Research was funded by a grant from the National Institute for General Medical Sciences (GM-46733).

References

Anderson, I. (1974). *A First Course in Combinatorial Mathematics*, pp. 77–95. Oxford: Clarendon Press.
Anderson, K. M. & Hovmöller, S. (1996). *Acta Cryst.* D**52**, 1174–1180.
Balmer, L. (1991). *Signals and Systems. An Introduction*, 2nd ed., pp. 445–455. Englewood Cliffs, NJ: Prentice-Hall.
Bricogne, G. (1993). *Acta Cryst.* D**49**, 37–60.
Brillouin, L. (1962). *Science and Information Theory*, 2nd ed., pp. 62–70. New York: Academic Press.
DeFelice, L. J. (1981). *Introduction to Membrane Noise*, pp. 139–140. New York: Plenum.
Dorset, D. L. (1997*a*). *Proc. Natl Acad. Sci. USA*, **94**, 1791–1794.
Dorset, D. L. (1997*b*). *Acta Cryst.* A**53**, 445–455.
Dorset, D. L. (1998). *Acta Cryst.* A**54**, 290–295.
Dorset, D. L. (2000*a*). *Proc. Natl Acad. Sci. USA*, **97**, 3982–3986.
Dorset, D. L. (2000*b*). *Z. Kristallogr.* **215**, 265–271.
Dorset, D. L. (2000*c*). *Micron.* In the press.
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.
Drenth, J. (1994). *Principles of Protein X-ray Crystallography*, p. 229. New York: Springer Verlag.
Frank. J. (1980). *Computer Processing of Electron Microscope Images*, edited by P. W. Hawkes, pp. 187–222. Berlin: Springer Verlag.
Gaskill, J. D. (1978). *Linear Systems, Fourier Transforms, and Optics*, pp. 172–176. New York: Wiley.
Gilmore, C. J., Dong, W. & Bricogne, G. (1999). *Acta Cryst.* A**55**, 70–83.
Gilmore, C. J., Nicholson, W. V. & Dorset, D. L. (1996). *Acta Cryst.* A**52**, 937–946.
Gilmore, C. J., Shankland, K. & Bricogne, G. (1993). *Proc. R. Soc. London Ser. A*, **442**, 97–101.
Gilmore, C. J., Shankland, K. & Fryer, J. R. (1993). *Ultramicroscopy*, **49**, 132–146.
Harker, D. (1953). *Acta Cryst.* **6**, 731–736.
Hill, R. (1986). *A First Course in Coding Theory.* Oxford: Clarendon.
Jap, B. K. & Li, H. L. (1995). *J. Mol. Biol.* **251**, 413–420.
Karle, J. & Karle, I. L. (1966). *Acta Cryst.* **21**, 849–859.
Lunin, V. Yu. (1993). *Acta Cryst.* D**49**, 90–99.
Lunin, V. Yu., Lunina, N. L., Petrova, T. E., Urzhumtsev, A. G. & Podjarny, A. D. (1998). *Acta Cryst.* D**54**, 726–734.
Lunin, V. Yu., Lunina, N. L., Petrova, T. E., Vernoslova, E. A., Urzhumtsev, A. G. & Podjarny, A. D. (1995). *Acta Cryst.* D**51**, 896–903.
Lunin, V. Yu., Urzhumtsev, A. G. & Skovoroda, T. P. (1990). *Acta Cryst.* A**46**, 540–544.
Luzzati, V., Mariani, P. & Delacroix, H. (1986). *Makromol. Chem. Macromol. Symp.* **15**, 1–17.
MacWilliams, F. J. & Sloane, N. J. A. (1977). *The Theory of Error-Correcting Codes.* Amsterdam: North-Holland.
Mason, S. J. & Zimmermann, H. J. (1960). *Electronic Circuits, Signals and Systems*, pp. 206–222. New York: Wiley.
Pauling, L. & Corey, R. B. (1951). *Proc. Natl Acad. Sci. USA*, **37**, 282–285.
Peterson, W. W. & Wilson, E. J. Jr (1972). *Error Correcting Codes*, 2nd ed. Cambridge, MA: MIT Press.
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.
Raemer, H. R. (1969). *Statistical Communications Theory and Applications*, pp. 65–66. Englewood Cliffs, NJ: Prentice-Hall.
Rossmann, M. G., Jeffery, B. A., Main, P. & Warren, S. (1967). *Proc. Natl Acad. Sci. USA*, **57**, 515–524.
Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.
Zhang, K. Y. J. & Main, P. (1990). *Acta Cryst.* A**46**, 41–46.