Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The use of wavelet transforms in low-resolution phase extension

Julie Wilson[a]* and Peter Main[b]

[a]Department of Chemistry, University of York, Heslington, York YO10 5DD, England, and [b]Department of Physics, University of York, Heslington, York YO10 5DD, England

Correspondence e-mail: julie@ysbl.york.ac.uk

Received 11 August 1999
Accepted 28 February 2000

A method to extend low-resolution phases has been developed using histogram matching not only of the electron density itself but also of histograms obtained from the different levels of detail provided by the wavelet transform of the electron density. It is shown that the method can extend phases from 10 Å to around 6–7 Å on a wide range of trial structures differing in size, space group and solvent content. This level of phase extension can improve the electron-density map from little more than a molecular envelope to one in which secondary structure can often be identified.

1. Introduction

In order to calculate an electron-density map which can be interpreted in terms of an atomic model, both the phases and the amplitudes of the structure factors are required. However, only the diffraction amplitudes can be measured experimentally, giving rise to the so-called 'phase problem' in crystallography. Methods which involve the collection of more than one X-ray data set, such as multiple isomorphous replacement and anomalous scattering, or the use of a model from a closely related structure in the molecular-replacement method can often provide the required phases, although this is not always the case.

The use of heavy atoms often means that the phases obtained from further experiments are to a lower resolution than is available for the original diffraction amplitudes. Density-modification and image-processing techniques have been developed which can extend the phases to the resolution of the native data (Zhang & Main, 1990*b*; Cowtan & Main, 1993; Vellieux, 1995; Abrahams, 1995). These include electron-density histogram matching over the protein region, solvent flattening or 'flipping', Sayre's equation and, where appropriate, the use of averaging both within a crystal, through non-crystallographic symmetry, and between multiple crystal forms. These methods require the available phases to be of sufficiently high resolution to treat the protein and solvent regions separately; in particular, the 'atomicity' constraints of Sayre's equation restrict their use. Typically, phases to about 3–4 Å are required. Problems in obtaining suitable heavy-atom derivatives or in measuring small anomalous differences mean there any many cases where no such phases are available and the protein crystal structure must be solved *ab initio*, *i.e.* from the diffraction amplitudes alone.

'Direct methods' which exploit relationships between structure factors *via* a probabilistic approach are now used routinely to solve the structures of crystals of small molecules (Main *et al.*, 1980; Yao, 1981; Sheldrick, 1985). However, the probabilities are dependent on the number of atoms in the

structure and the reliability of the phases obtained decreases as the number of atoms increases. This has limited the success of these methods to structures of less than 200 atoms. More recently, the combination of these reciprocal-space constraints with real-space constraints has increased their efficiency enormously (Miller *et al.*, 1994; Sheldrick & Gould, 1995). In this approach, automatic electron-density map interpretation in the form of 'peak picking' is used alternately with reciprocal-space refinement to improve the phases and has led to the solution of structures of around 1000 atoms. Although the complexity of structures which can now be solved in this way has increased so dramatically, the necessary assumption of atomic resolution currently restricts the methods to the solution of structures for which data to a resolution of at least 1.2 Å are available.

Significant progress has also been made at very low resolution and methods have been developed which can determine the molecular envelope of macromolecular structures, providing low-resolution phases in the most general case. The problem is then how to extend these phases to a resolution at which the existing density-modification and image-processing techniques can be applied. Increasing the resolution can be seen as adding the right amount of detail to the correct place in the electron-density map and it has been found that wavelet analysis, which effectively divides the electron density into different levels of detail, can give precise control over this.

2. Low-resolution phasing

Methods to determine the molecular envelope rely on the fact that there is little fluctuation within the protein region at low resolution and the electron density can be considered smooth and uniform within the envelope. Furthermore, an estimate for the volume of protein to be determined can be calculated when the molecular weight of the protein is known. Assuming the protein to be roughly spherical allows it to be modelled by a sphere of appropriate volume. This simple approach was taken by Kraut (1958) and has recently been used again with some success by a number of authors (see, for example, Harris, 1995; Andersson, 1999). The sphere is systematically moved around the unit cell and structure factors are calculated for each position and compared with the observed amplitudes. In an extension of this idea, a small number of large spheres are used in the 'few atoms model' method (Lunin *et al.*, 1995; Podjarny & Urzhumtsev, 1997) in which the spheres are positioned randomly throughout the cell. A very large number of these simple models are generated and a set of structure factors calculated for each. The phase sets of those models considered to have suitable amplitude correlations are then merged in a clustering procedure. In contrast, a large number of point scatterers, again consistent with the expected solvent content, can be used instead to model the electron density. Subbiah (1991) used minimization of the difference between the observed and calculated amplitudes in order to adjust such an initial random model until the point scatterers converged in a cluster. Often, the solvent is identified in this manner rather than the protein region; a method to determine which has been found has also been suggested by Subbiah (1993). A Monte Carlo approach to the determination of the molecular envelope from randomly generated electron density was developed by Lunin *et al.* (1990), where electron-density histograms are used to determine plausible phase sets from which cluster analysis leads to a final electron-density map showing the molecular envelope. A similar method has also been implemented by one of the present authors (Main, unpublished results), in which electron-density histogram matching, a technique which is used routinely in density modification at high resolution (Zhang & Main, 1990*a*), is used to systematically alter the random starting density. The success of this technique at both high and low resolution has motivated the current research using histogram matching of the wavelet coefficients. Until the protein and solvent regions can be properly identified, the histogram matching must be performed over the entire unit cell, unlike its implementation at high resolution where only the protein region is considered. The wavelet transform of the electron density allows histograms corresponding to different levels of detail to be matched simultaneously, thus extending the power of the method.

3. The wavelet histograms

The procedure for histogram matching is described in detail in Zhang (1993) and will not be repeated here. Its success relies on the fact that electron-density histograms can be predicted. At high resolution, the protein region can be considered separately and it has been shown that the histograms for different proteins are extremely similar and depend only on the resolution (Lunin & Skovoroda, 1991; Zhang & Main, 1990*a*). At very low resolution, the electron density from the entire unit cell must be used and molecular packing as well as the percentage of solvent play a part in determining the histogram. However, it has been found that considering the asymmetric unit only reduces the dependence on packing

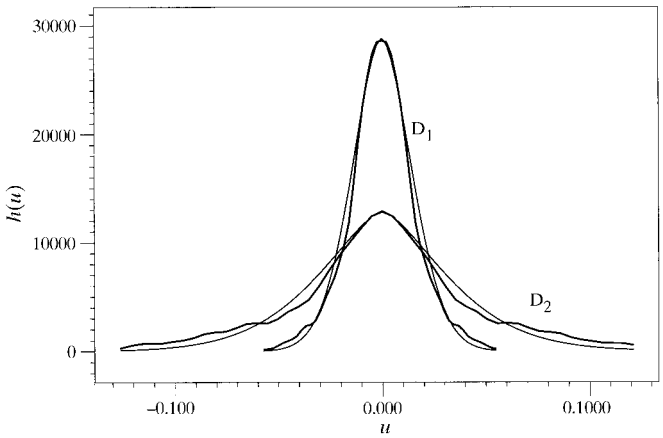

Figure 1
Detail histograms from a two-level transform; D_1 is calculated from the coefficients corresponding to the smallest details and D_2 from the next level of detail. The thick lines show the actual histograms calculated from a typical protein after smoothing with a Gaussian kernel and the thin lines show the histograms predicted from Mallat's mathematical model.

arrangements sufficiently so that only resolution and solvent content need to be addressed (Main, 1998). In both cases, a library of histograms can be compiled.

We have described the basic mathematics of the discrete wavelet transform (Main & Wilson, 2000). Intuitively, each level of the transform can be thought of as a pair of filters: a smoothing filter, S, which provides a kind of average between neighbouring points, and a related filter, D, which stores the differences between this smoothed version and the original data (which we call the details). The information can be stored in the coefficients of certain functions which depend on the particular filter used. A single level of a two-dimensional transform would require smoothing and detail filters in both directions and we would need to store the coefficients of functions corresponding to a smoothing in both directions (SS), the details in both directions (DD) and a smoothing in one direction but detail in the other (SD). Initially, three-dimensional wavelet transforms were considered with a single level, giving four different types of coefficient, so that four different histograms can be accumulated and denoted SSS, SSD, SDD and DDD. The SSS histogram looks like a low-resolution electron-density histogram as might be expected, since regardless of the wavelet functions used, what we have here are coefficients of a 'smoother electron density', *i.e.* the electron density with the finer details removed. More surprisingly, the detail histograms look very similar to those obtained by Mallat (1989) for two-dimensional images. An effective separation of the electron density is required in order to obtain different histograms which can be matched simultaneously, but the three-dimensional transform resulted in the important information being compressed into a small number of coefficients. In particular, the coefficients in the DDD histogram were all very close to zero. This is not so surprising as these coefficients encode the changes in the electron density, so that the DDD coefficients can only be significant when there are notable changes in all three of the x, y and z directions. In fact, the SDD histogram also contained very little information and this effect was even more marked if more than one level of the transform was performed. This demonstrates the power of the technique for compression, one of the important uses of wavelet analysis, but reduces the histogram matching to little more than electron-density histogram matching.

Consequently, one-dimensional wavelet transforms are used in the current method, leading to three possible phase sets for the resolution increase which can then be merged. Unless the symmetry is such that the transforms are equivalent, for example the x and y transforms in space group $P3_121$, the three transforms give independent results. Several ways of combining the results based on how well the results from the three transforms agree have been tried, but very little improvement has been gained over a straightforward averaging. The procedure is outlined in the next section. In this case a two-level transform is used, giving coefficients corresponding to the smoothed electron density and two different levels of detail. Therefore three histograms are available, denoted S, D_1 and D_2, for each of x, y and z. The S histograms for one-dimensional transforms again look like low-resolution electron-density histograms and therefore could also be predicted. However, we have found that these coefficients change least as the resolution increases, with the corresponding wavelet functions forming a base to which new details can be added, and the S histograms are not predicted but the coefficients from the previous resolution level are used in the transform, enabling these coefficients to keep their position. Indeed, we have found that this is more successful than using the actual coefficients from an S histogram calculated at the increased resolution, where this positional information is lost. Both the D_1 and D_2 histograms contain significant information and typical examples are shown in Fig. 1. These are independent of the protein in question and in fact look remarkably similar to the detail histograms obtained by Mallat (1989) for photographs. Mallat provided a mathematical model to describe these histograms:

$$h(u) = K \exp[-(|u|/\alpha)^{\beta}].$$

Since

$$\int_{-\infty}^{\infty} h(u)\,\mathrm{d}u = N,$$

where N is the number of coefficients, we can obtain the normalizing constant K from

$$K = N\beta/2\alpha\Gamma(1/\beta),$$

where $\Gamma(x)$ is the familiar gamma function defined by

$$\Gamma(x) = \int_{0}^{\infty} x^{t-1} \exp(-x)\,\mathrm{d}x.$$

The parameters α and β can be determined from the first and second moments of the histogram:

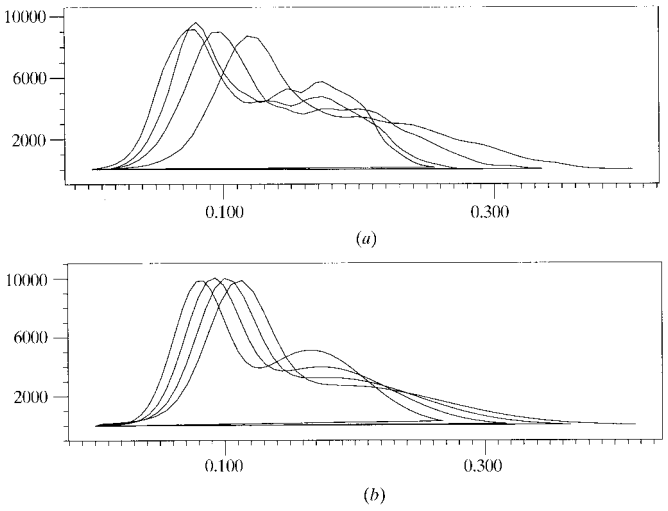

Figure 2
Electron-density histograms at 9, 8, 7 and 6 Å (*a*) calculated from a typical protein after smoothing with a Gaussian kernel and (*b*) calculated from a two-Gaussian model.

$$m_1 = \int_{-\infty}^{\infty} |u| h(u)\,\mathrm{d}u$$

$$m_2 = \int_{-\infty}^{\infty} u^2 h(u)\,\mathrm{d}u,$$

giving

$$\alpha = m_2\Gamma(2/\beta)/m_1\Gamma(3/\beta)$$

and

$$\Gamma^2(2/\beta)/\Gamma(1/\beta)\Gamma(3/\beta).$$

Fig. 1 shows histograms which have been calculated using parameters obtained in this way. As the parameters depend only on the resolution and the percentage of solvent, we have been able to predict values for α and β as functions of resolution for different solvent contents. These histograms have been shown to work as efficiently as histograms calculated using the actual values of the coefficients.

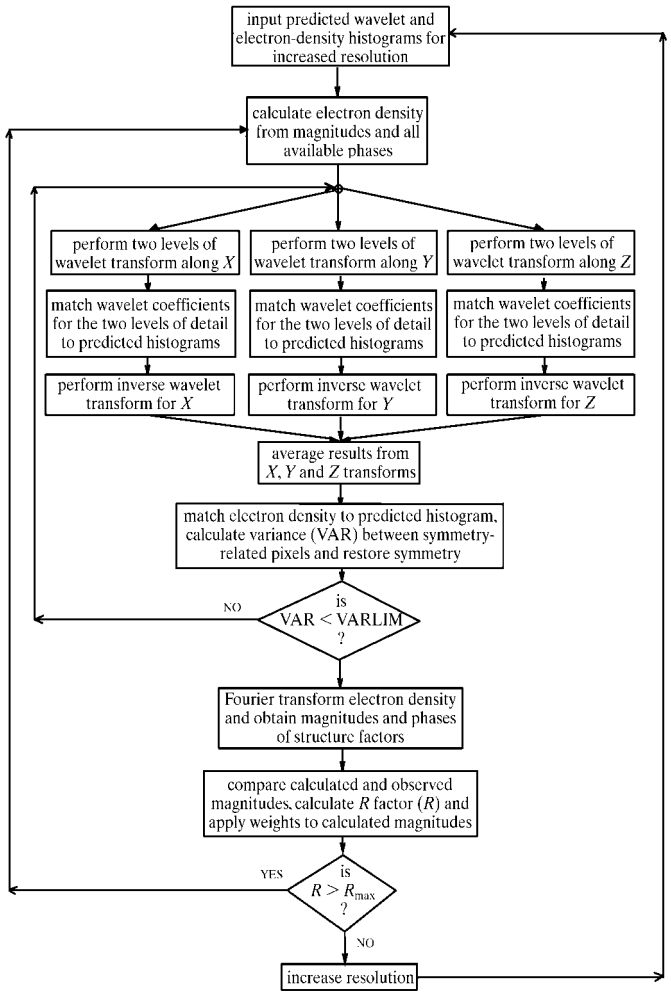

Figure 3
Flowchart showing the various stages in the phase-extension procedure.

4. The electron-density histograms

Although electron-density histogram matching is used both at high resolution, where only the protein region is used, and at very low resolution over the whole unit cell, histogram libraries relating to the resolution considered here have not previously been compiled and the wide range of resolution to be taken into account for various solvent contents requires a library of considerable size. As an alternative to such a library, Main (1990) provided a formula for calculating electron-density histograms at high resolution, in which the high-density values are modelled by a histogram of Gaussian peaks and the low-density values by a histogram of a randomly distributed background. Similarly, Lunin & Skovoroda (1991) used the two-component histogram model

$$v(t) = (F_{000}/V_{\mathrm{cell}})v^0(t) + q^0(t),$$

where the distributions $v^0(t)$ and $q^0(t)$ are calculated empirically to correspond with the histograms of known protein

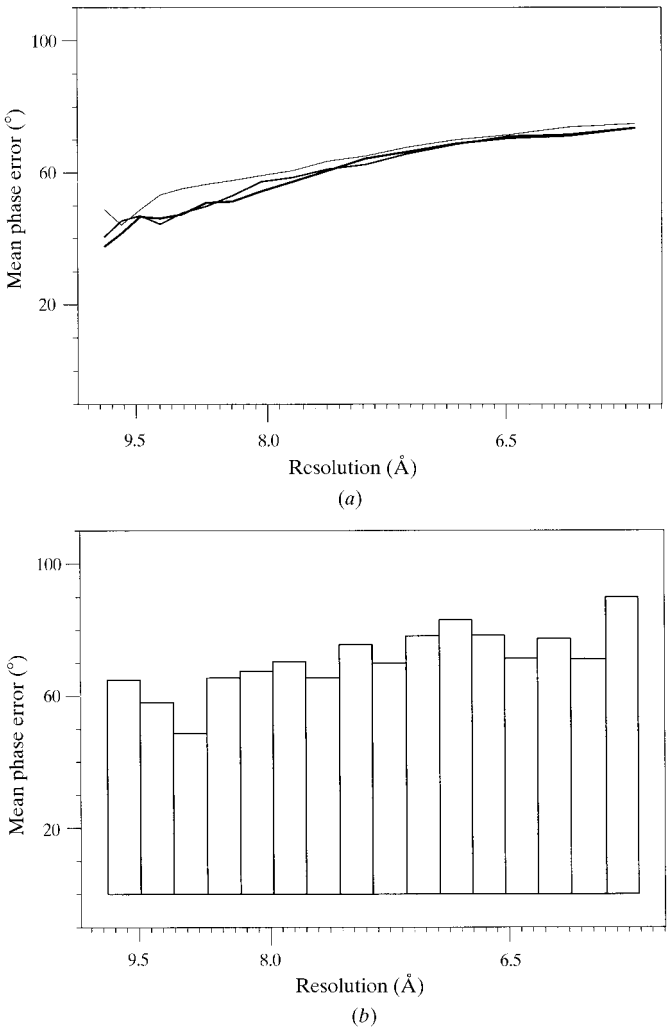

Figure 4
(*a*) Cumulative mean phase errors for three proteins: 1an8 (thin line), 1myg (medium line) and 2aai (thick line). The phase errors given are for all new phases between 10 Å and the resolution shown. (*b*) The mean phase errors for each resolution cycle for 1an8, *i.e.* the phase errors given are only for those phases within the resolution bin indicated by the bars.

Table 1
Cumulative mean phase errors (MPE) for new phases.

The resolution levels shown are representative and do not indicate the increments used.

PDB code	Solvent (%)	Symmetry	MPE on phases (°)				Starting phases†	New phases‡
			10–9.0 Å	10–8.0 Å	10–7.0 Å	10–6.0 Å		
1an8	75	$P4_32_12$	60.2	60.6	68.8	72.4	356	1062
1cbf	69	$P3_121$	48.9	58.0	68.7	77.8	273	901
1myg	65	$I21$	46.7	57.5	66.6	73.8	608	2019
1wsy	65	$C2$	54.0	65.0	73.0	79.5	890	3079
3gly	55	$P2_12_12_1$	52.4	65.9	71.4	73.9	406	1211
2aai	54	$P2_12_12_1$	48.9	60.3	70.1	76.8	440	1394
1olb	52	$P2_12_12_1$	58.9	63.4	73.8	76.8	412	1271
1aac	51	$P212_12_1$	54.5	63.7	70.9	76.4	504	1619
1ajg	50	$P3_121$	53.5	66.6	72.8	78.2	311	1069
1alu	49	$P3_121$	62.4	62.9	70.0	72.2	134	410
1an9	47	$P2_12_12_1$	48.5	64.8	72.9	77.8	477	1546
1ako	47	$P3_121$	57.9	58.2	70.3	79.8	187	598
1aqb	46	$P2_12_12_1$	48.9	62.7	68.3	77.4	135	409
1al3	45	$P2_12_12$	66.6	69.8	69.2	75.6	179	537
1am7	43	$P2_12_12_1$	65.1	64.2	72.2	77.2	326	1024

† Number of independent starting reflections at 10 Å. ‡ Number of independent new reflections, 10–6 Å.

Table 2
Use of starting phases with errors.

The mean phase errors (MPE) shown are in relation to those phases calculated from the refined structure. Similarly, the correlation coefficient given is for the correlation with the electron-density map with both magnitudes and phases calculated from this structure. Case 1: starting phases to 10.0 Å were calculated from the refined structure and calculated magnitudes were used throughout. Case 2: starting phases to 10.0 Å were calculated from a good molecular-replacement model and observed magnitudes were used where available. Case 3: starting phases to 10.0 Å were calculated from a poor molecular-replacement model and observed magnitudes were used where available.

Case	MPE on phases (°)					Total MPE (°)	Correlation coefficient
	To 10.0 Å	10.0–9.0 Å	10.0–8.0 Å	10.0–7.0 Å	10.0–6.0 Å		
1	0.0	62.3	60.8	69.4	74.4	58.5	0.73
2	22.3	62.5	64.3	74.2	77.6	65.8	0.59
3	47.2	67.7	73.7	79.5	82.9	75.3	0.49

structures. However, the resolution we are considering here is much lower and as it is too low for protein and solvent to be distinguished accurately, we must consider the electron density from the entire cell. The solvent has a large effect on the histograms and we obtain bimodal distributions such as those shown in Fig. 2(*a*). We have found that these histograms can be described as a sum of two Gaussian functions, one which corresponds roughly to the solvent and one to the protein region (Fig. 2*b*),

$$h(u) = \exp\{-[(u-\mu_1)/\sigma_1]^2\} + K\exp\{-[(u-\mu_2)/\sigma_2]^2\}.$$

This model requires five parameters to be determined: a mean and variance for each Gaussian and a scale between the two functions. These have been calculated empirically as functions of resolution which depend only on the solvent content. At resolutions beyond 10 Å, we have found no dependence on molecular packing and the electron-density histograms calculated in this way have been used successfully. Moreover, it has been found that this description allows systematic changes to the histograms which improve the results. For example, if the variance of the first Gaussian σ_1 is set at a lower level than expected, the phase errors are reduced. Since σ_1 corresponds to differences in the solvent density, this is equivalent to solvent flattening without the need to define the solvent region. Similarly, increasing the variance of the second Gaussian σ_2 can be seen as sharpening the electron density in the protein region. Using density histograms at slightly higher resolution than required can also improve results and this is easily achieved as the parameters describing the histograms are all functions of the resolution.

5. Outline of the method

In addition to the *ab initio* methods for producing very low resolution phases, experimental techniques are available which will determine the molecular envelope. Chemical-contrast variation or multiple-wavelength anomalous solvent-contrast variation can be used to define the solvent region (Carter *et al.*, 1990; Fourme *et al.*, 1995) and electron-microscopy images can be used to provide a low-resolution model (Ban *et al.*, 1998). However, the methods are all limited in their phasing power and additional information is required for further progress. The current research aims to extend such low-resolution phases using wavelet analysis as part of the procedure shown in the flowchart in Fig. 3. In order to test the method, magnitudes and starting phases to 10 Å were calculated from known model structures with solvent accounted for using Babinet's principle. The wavelet transform is performed on the electron density which has been calculated using all available phases, *i.e.* the original 10 Å starting phases as well as the phases obtained in previous cycles. The wavelet coefficients obtained are then matched to the histograms which have been predicted for an increase in resolution as described above. The relationship between the wavelet coefficients arising from space-group symmetry is extremely complex and it is not practical to implement symmetry at this stage. Thus, when the inverse transform is performed to retrieve the higher resolution electron density, the symmetry has been destroyed. However, the differences between symmetry-related pixels can be used to indicate the correctness of the map and the symmetry is reimposed as the electron density is matched to its predicted

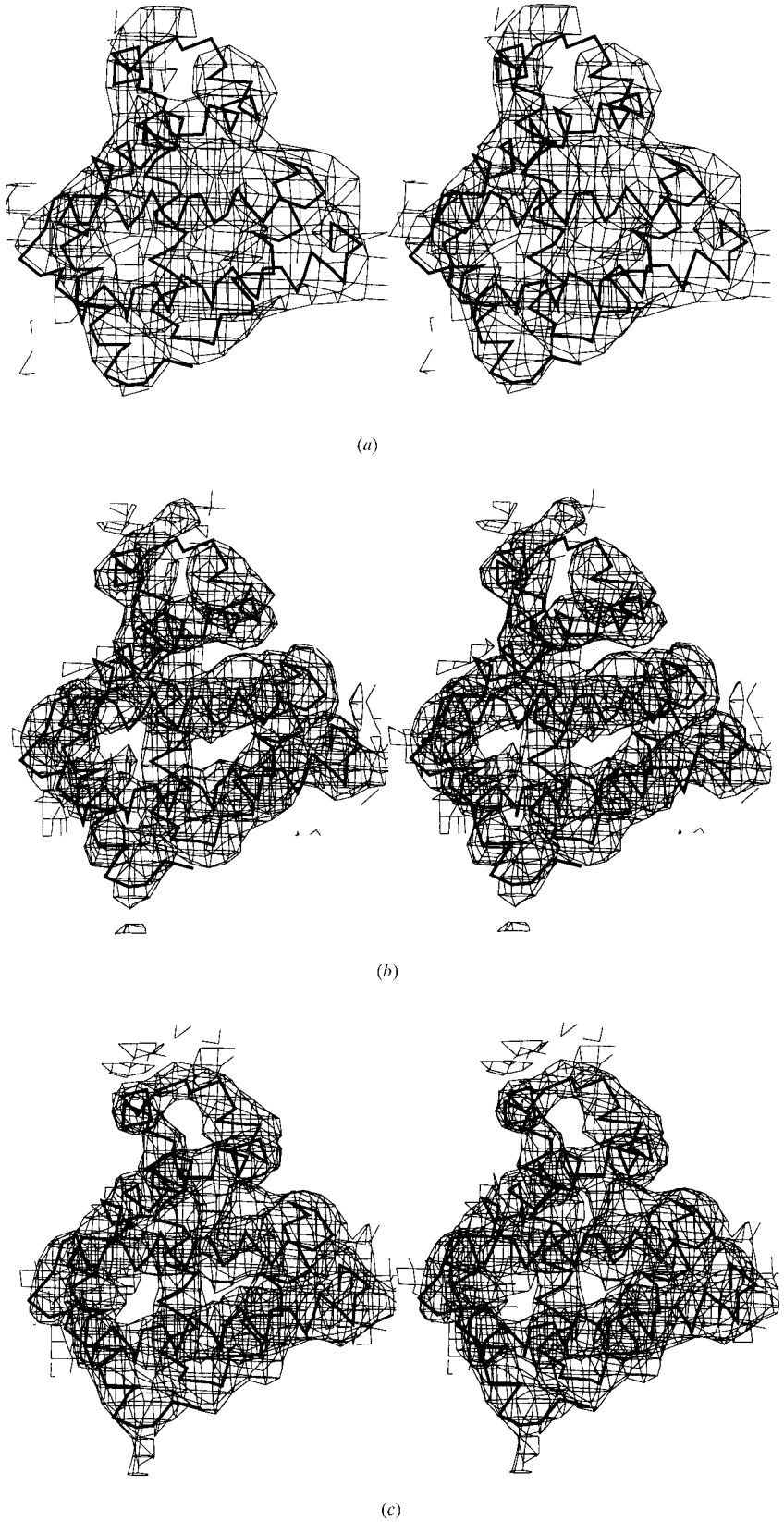

Figure 5
Stereoviews showing electron-density maps for myoglobin (PDB code 1myg) with the C^{α} trace shown in thick lines. (*a*) 10 Å starting map calculated from the final model. (*b*) 6 Å map with phases between 10 and 6 Å given by our method. (*c*) 6 Å map calculated from the final model. All electron-density maps are calculated at 1.25σ, where σ is the r.m.s. deviation from the mean density of the map.

histogram. The procedure is repeated until the wavelet coefficients can be matched to within predefined symmetry errors. The electron density is then Fourier transformed to obtain structure factors and the calculated and 'observed' magnitudes are compared. At this point, the flowchart indicates that weights should be applied to the calculated magnitudes before the whole process is repeated until convergence is achieved and the resolution increased. We have tried several weighting schemes, but have so far found that none has been able to improve the results obtained by simply replacing the calculated magnitudes by the observed magnitudes. Although some weighting schemes give improved results in the early cycles, after several resolution increments the results are no better and of course any weighting scheme which very gradually steps towards the correct magnitudes is also much slower. One weighting scheme in particular that we have investigated uses the information we already have about the molecular envelope at this resolution to provide phase-probability distributions based on structure invariants. Although we saw an improvement in the 10–9 Å range, this was lost as the resolution was increased further; this particular weighting scheme is being investigated further. We believe that if a suitable weighting scheme can be found, the results could be improved significantly.

6. Results

The method has been tried on a large number of model structures varying in size, solvent content and space group. Stucture-factor amplitudes were calculated from the coordinates of a variety of proteins (ID codes are given in Table 1) selected from the Protein Data Bank (Berman *et al.*, 2000). There is a build-up of phase errors as the calculation proceeds, but starting from a good 10 Å map we are currently able to produce new phases to about 6–7 Å with reasonable phase errors on all the structures tested. Table 1 shows the results for some of these. The phase errors shown in the table are cumulative phase errors, as are those in Fig. 4(*a*) which gives a graphical representation of the results for three of the proteins. However, the histogram in Fig. 4(*b*) is not cumulative and shows the phase errors for individual resolution bins. Although a mean phase error of around 70° for the new phases may seem high, the elec-

tron-density maps can still be seen to contain significant new information with this level of error. In most cases, the 10 Å map is little more than a mask roughly covering the molecule as in Fig. 5(*a*), whereas secondary structure can often be identified in the maps obtained to 6 Å with this level of phase error. Fig. 5(*b*) shows the corresponding 6 Å map calculated from phases which were extended from 10 Å and have a mean error of 74°. This can be compared with Fig. 5(*c*), which shows the real 6 Å map as calculated from the final model. Helices can easily be identified in either of the 6 Å maps and it is likely that further information, in the form of chemical knowledge, can be introduced at this stage and the method extended. Although the secondary structure of most proteins is not as readily identified as the helices of myoglobin, the new phases do in general add extra detail to the electron-density maps in the form of gaps between secondary-structure elements and in adding electron density missing from the 10 Å map, as can be seen in Fig. 6.

In order to make the situation more realistic, random noise corresponding to 30–40° mean phase error (MPE) on the starting phases was added. Although the first cycles showed higher phase errors, as the calculation proceeded the difference between the phase errors obtained in this case and those from 'perfect' starting phases became less and less. At the end of the calculation, the phases obtained were no worse.

The method has also proved successful when the 10 Å starting phases were obtained from a molecular-replacement solution. A model constructed from the coordinates of human deoxyhaemoglobin (Fermi *et al.*, 1984) was positioned using the program *AMoRe* (Navaza, 1990) in order to solve the structure of trout carbonmonoxyhaemoglobin I (Tame *et al.*, 1996). This model was sufficiently close to the true structure to allow phases between 8 and 4 Å to be used for molecular replacement and consequently the structure could be solved easily. However, in our tests we only used phases to 10 Å from this 'correctly positioned' starting model, giving a set of starting phases with a mean phase error of 22°. Experimental magnitudes were used unless they had not been measured, in which case magnitudes calculated from the starting model were used. Table 2 shows that the mean phase errors are only slightly worse than when the 'correct' 10 Å phases and

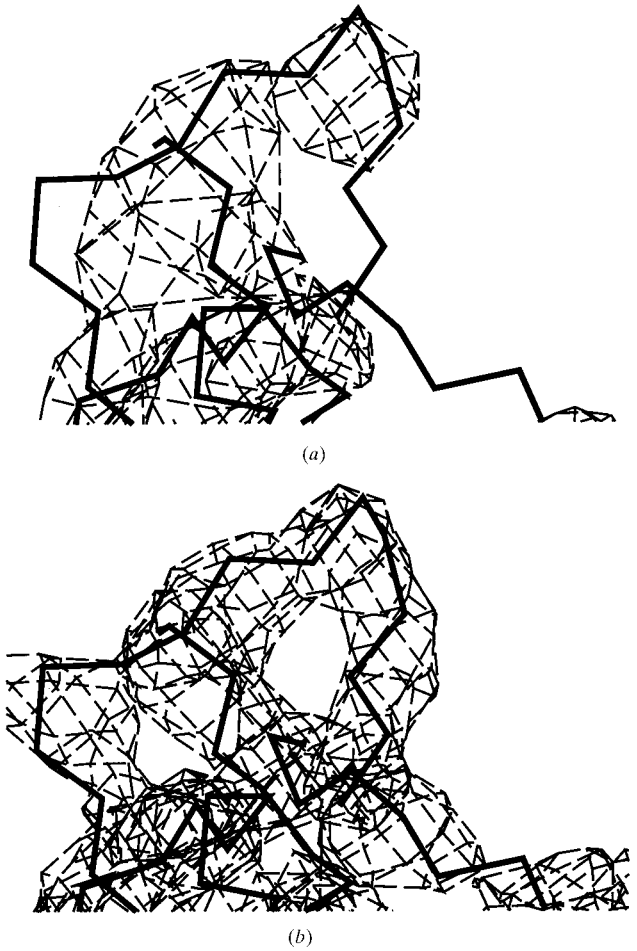

Figure 6
Electron-density maps showing a loop region in oppA (PDB code 1olb), with the C^{α} trace in thick lines. (*a*) shows the 10 Å starting map and (*b*) the map after phase extension to 6 Å.

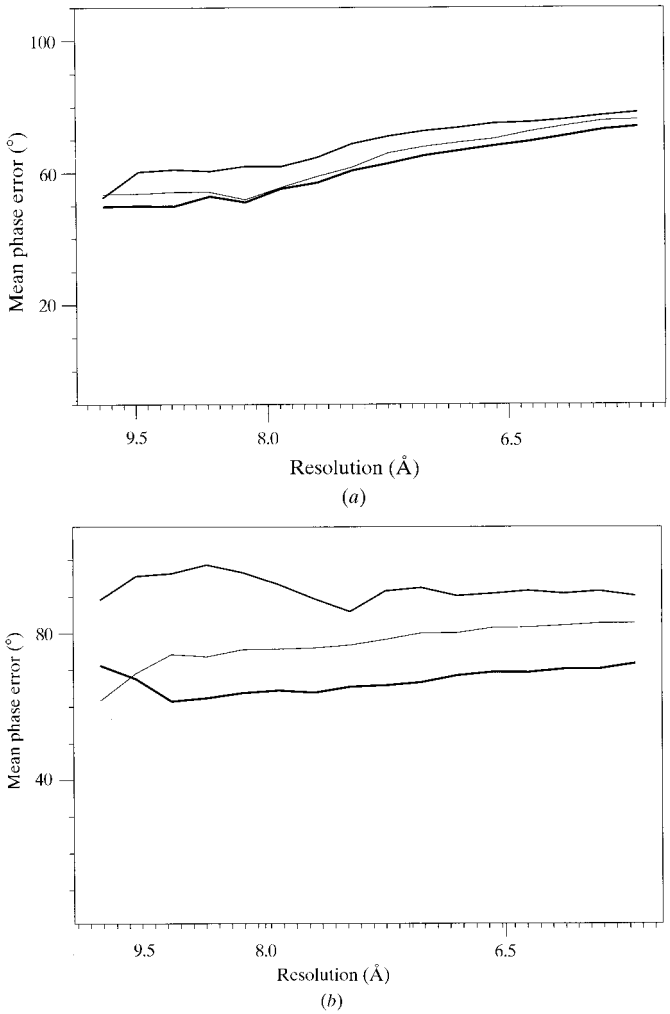

Figure 7
Cumulative phase errors for the proteins (*a*) 1wsy (56% solvent) and (*b*) 1an8 (75% solvent) when 'correct' electron-density histograms are used (medium line), when model electron-density histograms are used (thin line) and when model electron-density histograms and wavelet-coefficient histograms are used (thick line).

calculated magnitudes are used, showing that experimentally measured magnitudes can be used successfully. The correlation with the electron-density map calculated from the refined model may also be lower because experimental magnitudes are being compared with calculated magnitudes.

As only phases to 10 Å are needed, it is hoped that the method can be used in cases when the molecular-replacement model is much poorer and the usual procedure does not work. Further tests were carried out with a starting model which had been significantly altered using the graphics program *QUANTA* (Biosym-MSI, San Diego, CA). Some parts of the model were deleted entirely and others, such as helices, were moved in relation to each other to give a set of 10 Å phases with a mean phase error of 47°. Again, the measured magnitudes were used where available and the results are also shown in Table 2.

7. Discussion

We have tried to assess the relative importance of the electron-density and wavelet-coefficient histogram matching. As electron-density histogram matching was found to be limited in the resolution that could be attained in the *ab initio* calculations (Main, 1998), we had not previously tried to use this alone in the phase-extension procedure here. Furthermore, when we originally implemented the wavelet-histogram matching, we were still using electron-density histograms which were calculated from the known protein structure in question and so we were using information which we would not normally have. It therefore made no sense to try extending the resolution using these histograms alone, although for the sake of comparison we have performed exactly this and the associated phase errors are shown by the medium line in Fig. 7. The thick line in this figure shows the phase errors when wavelet-histogram matching is used together with electron-density histogram matching with histograms calculated from our two-Gaussian model. It can be seen that even where the electron-density histogram matching works quite well (Fig. 7*a*), this consistently reduces the phase errors by about 10° though no phase information beyond 10 Å is used. However, we were surprised to find that in several test cases the electron-density histogram matching alone works almost as well when the two-Gaussian model is used (the thin line in Fig. 7*a*). While Fig. 7(*a*) is typical for a number of proteins, the situation is different for those with particulary high solvent content, for example the proteins 1an8 (75% solvent) and 1cbf (69% solvent). Fig. 7(*b*) shows the results for 1an8 and it can be seen that electron-density histogram matching with the 'correct' electron-density histograms (*i.e.* those calculated from the actual values of the electron density) is now ineffective. Although using the two-Gaussian model does improve the results, the phase errors are still much higher than those seen when wavelet-histogram matching is also used.

The use of wavelets to employ information at different levels of detail inspired the incorporation of multi-resolution histogram matching in the density-modification program *DM* (Cowtan *et al.*, 2000). Cowtan has so far been unable to demonstrate whether the multi-resolution histograms add new information or if they are merely an indirect means of downweighting higher resolution terms (Cowtan, 1999). We have not been able to improve the results obtained by using multi-resolution electron-density histogram matching without the use of the wavelet transform.

To improve the method further, we propose to average over a number of different trials in a Monte Carlo approach. This will require differences to be introduced without the individual runs deteriorating significantly. It has been found that sharpening the electron density by different amounts when matching this histogram can provide such results, although the reduction in phase error (about 2–6° from those shown in Table 1) is slight. This can also be achieved by running the program with different wavelet bases and averaging the results. A discussion on wavelet transforms is given in Main & Wilson (2000) and we only note here that a family of wavelet functions is defined by a set of filter coefficients which give the relationship between the wavelets at one scale and those at a scale twice as fine. The number of non-zero filter coefficients gives the support of the wavelets and as we are dealing with a relatively small number of grid points, those of particularly compact support are most suitable. A series of wavelet functions have been constructed by Daubechies (1992) with 2, 4, 6, ... non-zero filter coefficients and to date we have only used those of order 2, 4 and 6, as all have similar detail histograms. Individually, these have all given similar results, but when results from all three are averaged at the end of each resolution cycle we have obtained a slight improvement. It is hoped that the use of different wavelets, of which there are many available, will increase the improvement. Those considered have detail histograms with a greater variance than those of Daubechies which are used at present and the histogram prediction will need to be revised. It is hoped that the differences between a wider variety of wavelet bases will produce a more effective Monte Carlo procedure.

This work was supported by the BBSRC Structural Biology and Design Application Initiative.

References

Abrahams, J. P. (1995). *Acta Cryst.* D**52**, 30–42.
Andersson, K. M. (1999). *J. Appl. Cryst.* **32**, 530–535.
Ban, N., Freeborn, B., Nissen, P., Penczek, P., Grassucci, R. A., Sweet, R., Frank, J., Moore, P. B. & Steitz, T. A. (1998). *Cell*, **93**(7), 1105–1115.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H, Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Carter, C. W., Crumley, K. V., Coleman, D. E. & Hage, F. (1990). *Acta Cryst.* A**46**, 57–68.
Cowtan, K. D. (1999). Personal communication.
Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.
Cowtan, K. D., Zhang, K. Y. J. & Main, P. (2000). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, ch. 25.2.2. Dordrecht: Kluwer Academic Publishers. In the press.
Daubechies, I. (1992). *Ten Lectures on Wavelets.* Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.

Fermi, G., Perutz, M., Shaanan, B. & Fourme, R. (1984). *J. Mol. Biol.* **175**, 159–174.
Fourme, R., Shepard, W., Kahn, R., l'Hermite, G. & de la Sierra, I. L. (1995). *J. Synchrotron Rad.* **2**, 36–48.
Harris, G. W. (1995). *Acta Cryst.* D**51**, 695–702.
Kraut, J. (1958). *Biochim. Biophys. Acta*, **30**, 265–270.
Lunin, V. Y., Lunina, N. L., Petrova, T. E., Vernoslova, E. A., Urzhumtsev, A. G. & Podjarny, A. D. (1995). *Acta Cryst.* D**51**, 896–903.
Lunin, V. Y. & Skovoroda, T. P. (1991). *Acta Cryst.* A**47**, 45–52.
Lunin, V. Y., Urzhumtsev, A. G. & Skovoroda, T. P. (1990). *Acta Cryst.* A**46**, 540–544.
Main, P. (1990). *Acta Cryst.* A**46**, 507–509.
Main, P. (1998). Unpublished results.
Main, P., Fiske, S. J., Hull, S. E., Lessinger, L., Germaine, G., Declercq, J. P. & Woolfson, M. M. (1980). *MULTAN*80: *A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data*, Universities of York, England and Louvain, Belgium.
Main, P. & Wilson, J. (2000). *Acta Cryst.* D**56**, 618–624.
Mallat, S. (1989). *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 674–693.
Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
Navaza, J. (1990). *Acta Cryst.* A**46**, 619–620.
Podjarny, A. D. & Urzhumtsev, A. G. (1997). *Methods Enzymol.* **276**, 641–648.
Sheldrick, G. M. (1985). *SHELXS*86. *Program for the Solution of Crystal Structures*, University of Göttingen, Germany.
Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.
Subbiah, S. (1991). *Science*, **252**, 128–132.
Subbiah, S. (1993). *Acta Cryst.* D**49**, 108–119.
Tame, J. R. H., Wilson, J. C. & Weber, E. (1996). *J. Mol. Biol.* **259**, 749–760.
Vellieux, F. (1995). *J. Appl. Cryst.* **28**, 347–351.
Yao, J.-X. (1981). *Acta Cryst.* A**37**, 642–664.
Zhang, K. (1993). PhD thesis, University of York, England.
Zhang, K. & Main, P. (1990*a*). *Acta Cryst.* A**46**, 41–46.
Zhang, K. & Main, P. (1990*b*). *Acta Cryst.* A**46**, 377–381.