Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Two-wavelength MAD phasing: in search of the optimal choice of wavelengths

A. González,[a]* J.-D. Pédelacq,[b] M. Solà,[c] F. X. Gomis-Rüth,[c] M. Coll,[c] J.-P. Samama[b] and S. Benini[a]

[a]European Molecular Biology Laboratory (EMBL), Hamburg Outstation, Notkestrasse 85, 22603 Hamburg, Germany, [b]Groupe de Cristallographie Biologique, Institut de Pharmacologie et de Biologie Structurale du CNRS, 205 Route de Narbonne, 31077 Toulouse, France, and [c]Centre d'Investigació i Desenvolupament, CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain

Correspondence e-mail: ana@embl-hamburg.de

Received 19 January 1999
Accepted 12 May 1999

The multiwavelength anomalous dispersion (MAD) method is increasingly being used to determine protein crystal structures. In theory, data collection at two wavelengths is sufficient for the determination of MAD phases, but three or even more wavelengths are used most often. In this paper, the results of the phasing procedure using only two wavelengths for proteins containing different types of anomalous scatterers are analyzed. In these cases, it is shown that this approach leads to interpretable maps, similar in quality to those obtained with data collected at three wavelengths, provided that the wavelengths are chosen so as to give a large contrast in the real part of the anomalous scattering factor **f**. The consequences for a rational MAD data-collection strategy are discussed.

1. Introduction

Within a few years, the multiwavelength anomalous dispersion (MAD) method has moved from being a theoretical approach with limited application and one which many protein crystallographers only contemplated attempting as the last resort, to an almost routine approach to solving the phase problem in protein crystallography (Fourme *et al.*, 1996; Hendrickson & Ogata, 1997; Smith, 1997). During the 1990s, the number of new crystal structures determined by MAD increased more rapidly than the total number of structures solved (Smith *et al.*, 1996).

The MAD method exploits the structure-factor differences arising from the variation of the complex anomalous scattering factor **f** at wavelengths around the absorption edges of heavy atoms within the protein crystal. These differences are typically much smaller than isomorphous differences. Nevertheless, MAD maps tend to be of comparable or higher quality than those obtained from classical multiple isomorphous replacement techniques (Blow & Crick, 1959), because the phases are unaffected by lack of isomorphism and because scaling of the data is straightforward. On the other hand, a typical MAD experiment can take up to six or eight times the amount of beam time required for a single data-set collection, depending on the space group and crystal orientation (Dauter, 1997).

It is possible to shorten a MAD experiment considerably by collecting data at fewer wavelengths. As shown by a Harker diagram (Helliwell, 1992) and assuming no errors in the diffraction data, any two distinct wavelengths which exploit the rapid variation of the imaginary and the real part of the anomalous scattering factor f'' and f' (Helliwell, 1997) would suffice to solve the phase problem. Although data errors do

swamp part of the anomalous signal present in the data, Peterson *et al.* (1996), Friedman *et al.* (1995) and Wimberly *et al.* (1997) have presented examples of successful two-wavelength MAD experiments. If a two-wavelength experiment is in general sufficient for a definitive evaluation of unknown parameters, there should be methods provided to predict, from the anomalous scattering-factor values, which combination of two wavelengths will give the most accurate phases in any given MAD experiment and how much a third wavelength would add to the quality of the electron-density maps.

The question of which pair of wavelengths gives the most useful information may also be an important aspect of any MAD experiment, since MAD data collections may sometimes not be completed because of crystal damage or because of technical constraints. It may, therefore, be essential to collect the data in the right order for the benefit of the whole experiment.

In a two-wavelength MAD experiment, it is usually not possible to optimize both the Bijvoet or anomalous differences at a wavelength λ_1 [given by $2f''(\lambda_1)$] and the dispersive differences between wavelengths λ_1 and λ_2, $[|f'(\lambda_1) - f'(\lambda_2)|]$. Narayan & Ramaseshan (1981) predicted that for two wavelengths and assuming the same error in the determination of the anomalous and dispersive contributions, the error in the MAD phases would be inversely proportional to the expression

$$[f'(\lambda_1) - f'(\lambda_2)]^2[f''(\lambda_1)^2 + f''(\lambda_2)^2]. \quad (1)$$

Under the assumption of equal errors for both anomalous and dispersive differences, the higher magnitude of the anomalous differences should, in principle, provide the largest contribution to the phasing power (see Fig. 1), making it desirable to maximize the Bijvoet differences in the MAD data set. Here, we show the results of analyzing several combinations of two-wavelength MAD data so as to optimize dispersive and anomalous differences separately and present our conclusions regarding data-collection strategies for MAD experiments.

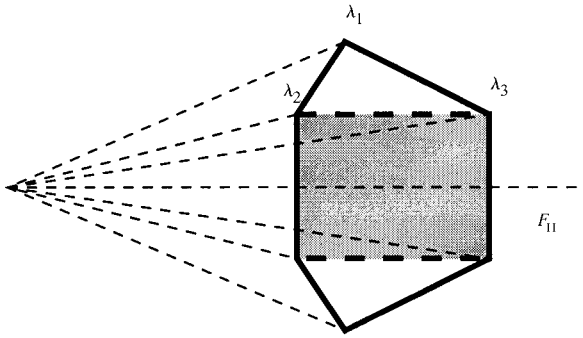

Figure 1
Phasing power of a MAD data set. The horizontal dotted line represents the heavy-atom structure factor assuming no anomalous scattering effects. The other dotted lines represent the heavy-atom structure factor at wavelengths λ_1, λ_2 and λ_3, taking into account the contributions from f' and f'' (the lines in the bottom half of the figure correspond to the Friedel pair). In the absence of errors, the expected phasing power from this MAD data set is correlated to the area enclosed by the hexagon defined by the structure-factor vectors at the three wavelengths (thick line). The more extreme the values of f' and f'', the larger the separation between the vectors and the higher the phasing power. The grey-filled quadrangle represents the phasing power in a two-wavelength experiment using λ_2 and λ_3. In principle, for an optimal two-wavelength experiment one should select the pair of wavelengths which define the largest area, but errors in the measured intensities can affect the accurate measurement of the heavy-atom contribution and decrease the effective separation of the phase circle centres.

2. Methods

The data sets used to determine the quality of two-wavelength MAD phases were obtained from standard MAD experiments at three or four wavelengths. They provided phases and electron-density maps of such quality that the polypeptide main chain and many side chains could be identified and allowed modelling of the protein structures. These full data maps and the associated phasing statistics were used as 'yardsticks' against which the two-wavelength maps and statistics were compared. In the two-wavelength work, each pair of different wavelengths was treated as a totally independent MAD experiment and only the reflections collected at these wavelengths were used for scaling, heavy-atom parameter refinement, phasing and map calculation.

2.1. MAD experiments

2.1.1. Cytochrome *c*553. The cytochrome *c*553 from the ureolytic soil bacterium *Bacillus pasteurii* is one of the very few *c*-type cytochromes isolated from Gram-positive bacteria. This cytochrome is a small protein (71 residues) with an iron-containing haem group and has a functional role in the bacterial respiratory chain. The crystals belong to the space group $P2_12_12_1$ (Benini *et al.*, 1997) and data were collected to 2.1 Å resolution[1] on the EMBL Hamburg Outstation beamline BW7A. The MAD experiment was carried out on the Fe *K* edge (Benini *et al.*, in preparation). The wavelengths used were above the *K* edge (λ_1 = 1.735 Å), at the inflexion point (λ_2 = 1.743 Å) and a remote wavelength (λ_3 = 1.0 Å).

2.1.2. Receiver domain of the transcription factor PhoB. PhoB is a transcription factor which activates more than 30 genes of the bacterial phosphate-assimilation pathways. The receiver domain of PhoB is a 125-residue protein with two molecules in the asymmetric unit. The MAD experiment was performed on a co-crystallized gold derivative (250 residues per gold ion[2]). Crystals belong to the space group $P2_12_12_1$. MAD data were collected at the EMBL Hamburg Outstation beamline X31 above and at the inflexion point of the $L_{\rm III}$ edge (λ_1 = 1.035 and λ_2 = 1.040 Å, respectively) and above the $L_{\rm I}$ edge at λ_3 = 0.830 Å (Solà *et al.*, 1999).

2.1.3. Panton–Valentine leucocidin F component (LukF-PV). LukF-PV monomer is a 301-residue protein (34 kDa) which forms, with the LukS-PV protein (32 kDa), the Panton–Valentine leucocidin. This toxin, secreted by

[1] At one wavelength data were collected to 1.7 Å resolution, but only the data to 2.1 Å resolution were used for phasing.
[2] A second site with low occupancy was found after phasing, but was not used.

Staphyloccoccus aureus, acts upon polymorphonuclear neutrophils, monocytes, and human and rabbit macrophages, and forms transmembrane pores. An iridium derivative, crystallized in the space group $P2_12_12_1$, with three major and four minor sites was used for the MAD experiment. Data were collected at the EMBL Hamburg outstation beamline X31 at four wavelengths: at the top of the white line above the L_{III} edge (λ_1 = 1.1048 Å), on the inflexion point of the same edge (λ_2 = 1.1053 Å) and at remote wavelengths above the L_I edge (λ_3 = 0.827 Å) and below the L_{III} edge (λ_4 = 1.127 Å) (Pédelacq *et al.*, 1999).

2.1.4. C-terminal domain of initiation elongation factor 3 (IF3-C). IF3-C is a 94-residue protein which crystallizes in space group *C*2. A selenomethionyl three-wavelength MAD data set was collected around the *K* edge at Brookhaven (Biou *et al.*, 1995). This MAD data set is distributed as a tutorial for the program *SHARP* (de La Fortelle *et al.*, 1997*a*). Selenium is a very common atom for MAD experiments (Biou, 1997). In this example, the anomalous signal from two selenium sites is close to the average value expected for proteins, which contain, on average, one methionine per 59 residues (Hendrickson *et al.*, 1990).

The above experiments cover three different types of MAD experiments.

(i) At *K* edges. The electrons involved in these transitions are those from the inner core, and f' becomes close to zero relatively near the high-energy side of the edge (Fig. 2). The dispersive differences can be nearly as important as the anomalous differences, and in many cases they can be optimally measured at most dedicated beamlines (Deacon *et al.*, 1995).

(ii) At *L* edges with a small or no white line. Dispersive differences are relatively small because of the structure of the edge and because of the contribution of the *K* electrons to the anomalous effect at energies larger than the absorption edge. The anomalous differences in these MAD data sets will be larger than the dispersive differences.

(iii) At *L* edges with a white line. Anomalous differences can become rather large – over twice as large as dispersive differences.

The wavelengths were selected according to the following criteria.

(i) λ_1, at the peak of the *K* or L_{III} absorption edge. This wavelength was chosen to maximize the value of f''.

(ii) λ_2, at the inflexion point of the absorption edge. f' should reach its minimum value at this wavelength, although small deviations from this point, arising from either errors in its calculation or instability of the wavelength of the beam, will cause the f' value to increase.

(iii) λ_3, a remote wavelength on the high-energy side of the edge, chosen so that the value of f' is as large (close to zero) as possible. When the wavelength is chosen above the L_I edge, the largest magnitude of f'' is also attained, but only in the absence of a strong white line on the L_{III} edge.

(iv) In one case, a remote wavelength (λ_4) about 200 eV to the low-energy side of the edge was selected. The f' value does not increase as fast as it does on the high-energy side (see Fig. 2) and the f'' contribution is usually close to its minimum value. This limits the phasing power derived from data collected at this wavelength.

2.2. Data collection, analysis and phasing

All data sets were collected on cryo-cooled crystals at 100 K. The *c*553, PhoB and LukF data were processed using the program *DENZO* from the *HKL* package (Otwinowski, 1993) and scaled and merged with *SCALA* (Collaborative Computational Project, Number 4, 1994). The scaling statistics, shown in Table 2, indicate that the diffraction data were of reasonable quality. In the case of the LukF-PV, the lower overall completeness was caused by the presence of solvent rings in the images. In the unaffected resolution bins at low and medium resolution, the completeness was over 95%.

Since only one crystal was used for each MAD experiment, no information loss owing to lack of isomorphism was expected and phases were calculated to the highest resolution. The MAD phasing statistics, represented by the figure of merit

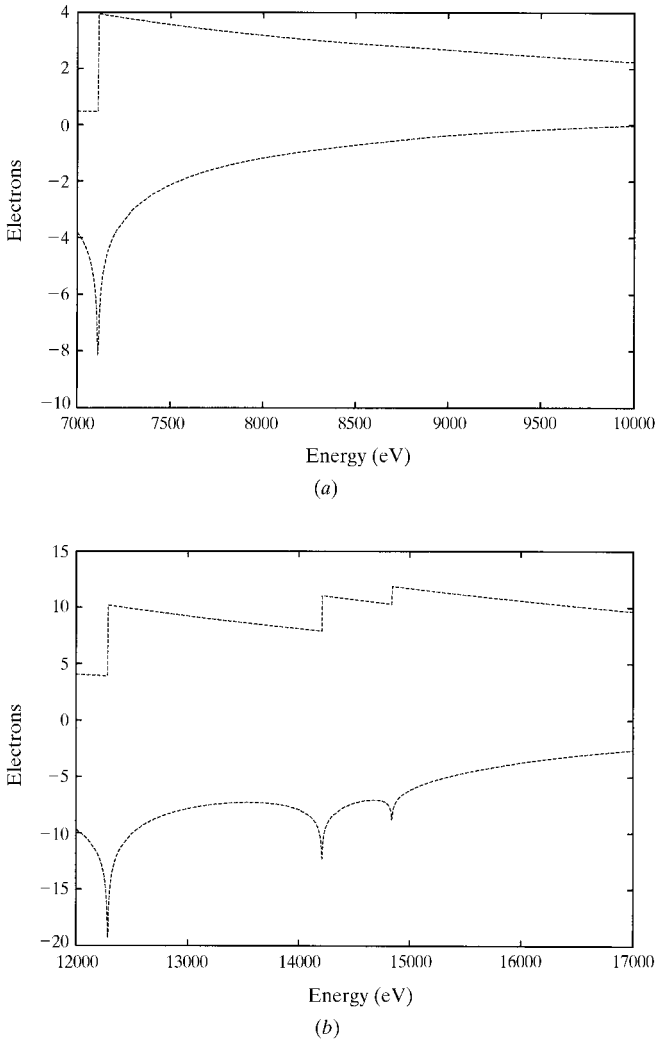

Figure 2
Theoretical plots of f'' and f' over (*a*) a *K* edge and (*b*) an *L* edge (Merritt, 1998).

Table 1
Correlation coefficient of the two-wavelength electron-density map with respect to the three-wavelength map, as a function of the fractional anomalous and dispersive differences.

The anomalous and dispersive mean fractional differences (%) were calculated using refined scattering factors from expressions (3) and (4). The values in parentheses correspond to the ones given by *SCALEIT*.

*c*553 (Fe *K* edge).

	$\lambda_1\lambda_3$	$\lambda_2\lambda_3$	$\lambda_1\lambda_2$
Correlation	0.87	0.95	0.58
Anomalous	5.0 (5.9)	3.0 (3.9)	5.0 (5.8)
Dispersive	3.4 (4.5)	4.9 (5.9)	1.5 (2.4)

PhoB (Au *L* edge).

	$\lambda_1\lambda_3$	$\lambda_2\lambda_3$	$\lambda_1\lambda_2$
Correlation	0.78	0.94	0.72
Anomalous	5.5 (7.9)	5.5 (7.9)	4.8 (7.2)
Dispersive	3.0 (5.3)	4.1 (5.7)	1.1 (4.4)

LukF-PV (Ir *L* edge).

	$\lambda_1\lambda_3$	$\lambda_2\lambda_3$	$\lambda_1\lambda_2$	$\lambda_1\lambda_4$	$\lambda_2\lambda_4$	$\lambda_3\lambda_4$
Correlation	0.80	0.81	0.71	0.6	0.62	0.59
Anomalous	11.8 (5.9)	8.1 (5.6)	11.8 (6.1)	11.8 (5.9)	8.1 (5.6)	7.4 (5.0)
Dispersive	5.1 (4.2)	8.5 (5.6)	3.7 (3.2)	2.6 (3.3)	5.9 (4.1)	2.6 (3.4)

IFC-3 (Se *K* edge).

	$\lambda_1\lambda_3$	$\lambda_2\lambda_3$	$\lambda_1\lambda_2$
Correlation	0.87	0.90	0.76
Anomalous	7.4	4.4	7.2
Dispersive	4.0	4.6	1.0

(f.o.m.) and Cullis *R* factor,[3] are given in Tables 3, 4 and 5. Unless stated otherwise, these statistics were obtained using *MLPHARE* (Collaborative Computational Project, Number 4, 1994). In addition, scaling and phasing were also performed with the programs *SHARP* (de La Fortelle *et al.*, 1997*a*), *SOLVE* (Terwilliger & Berendzen, 1997) and *MADSYS* (Hendrickson, 1991), except for the IF3-C case where only *SHARP* was used. Similar qualitative results were obtained in all cases, which suggests that the conclusions are independent of the chosen algorithm or data-analysis method. The electron-density maps were calculated with the *CCP*4 program *FFT* (Collaborative Computational Project, Number 4, 1994; Read & Schierbeek, 1988). In order to quantify the differences between the 'three-wavelength' maps (calculated with phases obtained from all wavelengths) and the 'two-wavelength' maps, the correlation coefficient between the two maps given by

$$\mathrm{CC} = \sum(f_1-\langle f_1\rangle)(f_2-\langle f_2\rangle)/\sum[(f_1-\langle f_1\rangle)^2(f_2-\langle f_2\rangle)^2]^{1/2} \quad (2)$$

was used. f_1 and f_2 are the density values of each map for each grid point. The higher this value is, the more similar the two maps are. Correlation values higher than 0.9 suggest that adding a third wavelength does not substantially increase the quality of the map. No density-modification methods were applied to the MAD phases, in order to analyze the performance of the MAD method by itself without contribution from any other information or assumptions. Map figures were displayed with *TURBO-FRODO* (Roussel & Cambillau, 1989) at the 1σ level on a Silicon Graphics Workstation.

2.3. Estimation of anomalous scattering factors and anomalous signal

A fluorescence spectrum was measured from the crystal in the near-edge region. The f'' values at the wavelengths in this region were calculated by scaling the fluorescence spectrum data to the theoretical absorption edge and the f' values were derived from the Kramers–Krönig transformation (Evans, 1994). For the remote wavelengths, the f'' and f' values were estimated from theoretical curves (Brennan & Cowan, 1992). The initial values were subsequently refined during phasing with *SHARP* (de La Fortelle *et al.*, 1997*a*).

From the knowledge of the f' and f'' values, the anomalous signal in the data set can be estimated as the ratio of the expected anomalous (1) or dispersive differences (2) to the total scattering (Smith, 1997),

$$\langle|\mathbf{F}_\lambda^+| - |\mathbf{F}_\lambda^-|\rangle/\langle|\mathbf{F}_T|\rangle \simeq (N/2)^{1/2}2f_\lambda''/\langle|\mathbf{F}_T|\rangle \quad (3)$$

and

$$\langle|\mathbf{F}_{\lambda a}| - |\mathbf{F}_{\lambda b}|\rangle/\langle|\mathbf{F}_T|\rangle \simeq (N/2)^{1/2}(|f'_{\lambda a} - f'_{\lambda b}|)/\langle|\mathbf{F}_T|\rangle, \quad (4)$$

where N is the number of anomalous scatterers in the sample, $\mathbf{F}_T$ is the total structure factor $\langle|\mathbf{F}_T|\rangle = (\sum f_i^2)^{1/2}$ (f_i is the complex atomic scattering factor for each atom in the unit cell) and $\mathbf{F}^+$ and $\mathbf{F}^-$ are the structure factors of the positive and negative Bijvoet pairs.

The calculated theoretical values were compared with the mean fractional dispersive and anomalous differences evaluated during scaling of the data (see Table 1). Discrepancies between theoretical and calculated values arise from inaccurate measurements of scattered intensities, which affect the scaling procedure, and from the occupancy and temperature factor of the anomalous scatterers, which are not taken into account in the expressions (3) and (4).

3. Results

Tables 3, 4, 5 and 6 show that the figures of merit and Cullis *R* factor were, as expected, somewhat better when more than two wavelengths were used, since the phasing power of a MAD data set always increases slightly with the number of wavelengths, as each value of the anomalous scattering factor $\mathbf{f}(\lambda)$ gives a different origin and radius for the phase circle (Fig. 1). However, the statistics obtained for some two-wavelength combinations (see below) indicated good phase quality

[3] The figure of merit is defined as the inverse of the spread of the probability of the best estimate of the phase $\Phi_{\rm fom} = \sum P(\Phi)\exp(2\pi i\Phi)/\sum P(\Phi)$. $R_{\rm Cullis}$ is the ratio between the average dispersive and anomalous differences and their respective lack-of-closure error $\mathrm{err} = |\mathbf{F}_T| - |\mathbf{F}_N + \mathbf{F}_A|$, where $|\mathbf{F}_T|$ is the measured amplitude value, $|\mathbf{F}_N|$ the contribution from the reference and $|\mathbf{F}_A|$ the contribution from the anomalous scatterer. For acentric reflections $R^a_{\rm disp} = \sum_a \mathrm{err}_{\rm disp}/\sum_a \Delta_{\rm disp}$ and for centric reflections $R^c_{\rm disp} = \sum_c \mathrm{err}_{\rm disp}/\sum_c \Delta_{\rm disp}$. Similarly, $R_{\rm ano} = \sum \mathrm{err}_{\rm ano}/\sum \Delta_{\rm ano}$.

Table 2
Statistics on MAD data sets after internal scaling.

The values in parentheses correspond to the last resolution bin.

Cytochrome *c*553.

Wavelength (Å)	Resolution†	I/σ	R_{merge}‡	R_{anom}§	Completeness	Multiplicity
λ_1, 1.735	2.1	7.4 (3.2)	0.068 (0.153)	0.079	98.4	3.1
λ_2, 1.743	2.1	7.4 (3.2)	0.068 (0.151)	0.040	98.3	3.1
λ_3, 1.000	1.7	8.6 (5.4)	0.040 (0.101)	0.027	99.5	3.5

PhoB.

Wavelength (Å)	Resolution†	I/σ	R_{merge}‡	R_{anom}§	Completeness	Multiplicity
λ_1, 1.035	2.5	10.6 (3.6)	0.065 (0.204)	0.060	99.1	3.4
λ_2, 1.040	2.5	12.3 (4.3)	0.057 (0.175)	0.050	99.1	3.4
λ_3, 0.830	2.5	9.0 (3.3)	0.072 (0.213)	0.067	99.6	3.1

LukF-PV.

Wavelength (Å)	Resolution†	I/σ	R_{merge}‡	R_{anom}§	Completeness	Multiplicity
λ_1, 1.1048	2.0	7.4 (5.8)	0.064 (0.111)	0.056	86.7	3.7
λ_2, 1.1053	2.0	6.6 (5.8)	0.071 (0.123)	0.048	85.1	4.4
λ_3, 0.8269	2.0	8.2 (6.1)	0.063 (0.121)	0.050	83.3	3.6
λ_4, 1.1270	2.0	8.0 (7.2)	0.060 (0.101)	0.037	85.0	4.6

† $\lambda/2\sin\theta$ at the edge of the detector. ‡ $\sum|I-\langle I\rangle|/\sum\langle I\rangle$. § $\sum|\langle I^+\rangle-\langle I^-\rangle|/\sum(\langle I^+\rangle+\langle I^-\rangle)$.

and the maps calculated from those two-wavelength data sets were clearly interpretable.

3.1. Maximizing dispersive differences

The combined use of λ_2 and λ_3 optimizes the dispersive differences, as confirmed by the measured mean fractional values shown in Table 1. All electron-density maps computed using these data sets were of good quality and their correlation with the three-wavelength electron-density maps was very high. Close inspection of these maps (Fig. 5) revealed some breaks in the electron density which were not present in the three-wavelength maps (Fig. 3), but they did not affect overall interpretability.

The good quality of the maps obtained by maximizing the dispersive differences is not surprising when the MAD data were collected around the gold *L* edge (PhoB). In this case, the data set collected at the remote wavelength λ_3 (above the L_I edge) also provides the highest anomalous differences. Therefore, this wavelength combination gives the maximum achievable separation between the centres of the phase circles (Fig. 1) and the phasing power is expected to be higher than for other wavelength combinations. This is not the case for the LukF-PV, IF3-C and *c*553 data sets. While the first two still contain large Bijvoet differences, in the case of *c*553 (Table 1) the calculated anomalous signal at the remote wavelengths after refining f'' is only about 3%. This difference is a limit below which MAD experiments are considered difficult (Thompson, 1997). The high quality of the *c*553 electron-density map calculated with these data sets suggests that measuring the maximum Bijvoet differences is not critical, as long as the dispersive differences are sufficiently large.

3.2. Maximizing anomalous differences

As shown in Table 1, the λ_1 and λ_3 combination decreases the dispersive signal significantly, but optimizes the anomalous signal in the test cases investigated in this study. However, in the cases of *c*553, IF3-C and PhoB, both the phasing statistics for centric and acentric reflections (Tables 3, 6 and 4, respectively) and the maps obtained with this combination (Fig. 4) are worse than the map resulting from optimizing the dispersive differences (Fig. 5). Larger and more numerous breaks in the electron density were found and some side chains were less identifiable.

This wavelength combination still provided good maps when using the data collected at the iridium *L* edge. In this case, the electron-density map appears similar to the one obtained by maximizing the dispersive differences. Both maps are highly correlated to the three-wavelength map, as shown in Table 1. It seems that the large anomalous differences present in the data collected at the white line provided an important contribution to the phasing, but

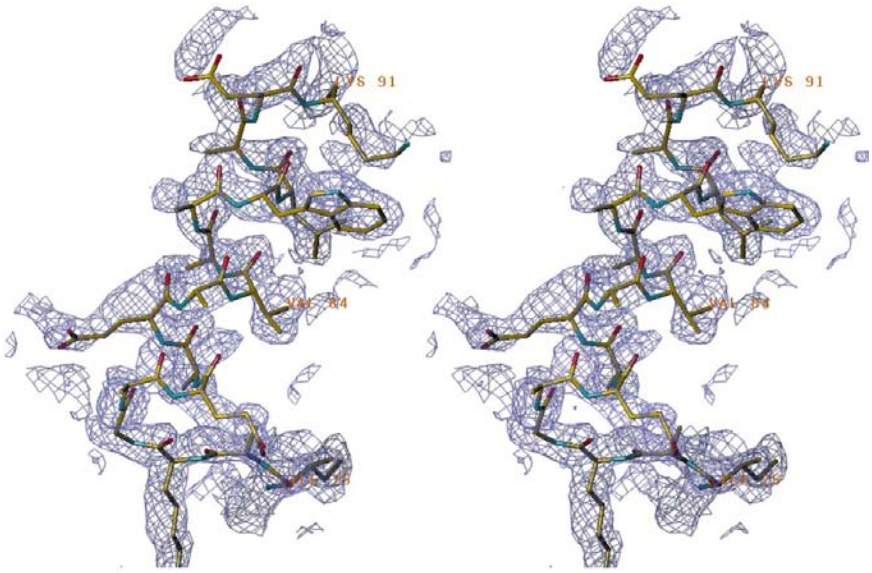

Figure 3
Detail of the 2.1 Å MAD map of cytochrome *c*553 calculated with phases from the three-wavelength data set, showing an α-helix and the refined model.

Table 3
MAD phasing statistics for cytochrome *c*553 at the iron *K* edge.

	Figure of merit			$R_{\rm Cullis}$		
	Acentric	Centric	All	Anomalous	Centric	All
$\lambda_1\lambda_2\lambda_3$	0.51	0.63	0.53	0.84	0.45	0.66
$\lambda_1\lambda_3$ (maximum f'' and remote)	0.25	0.32	0.26	0.91	0.81	0.95
$\lambda_2\lambda_3$ (minimum f' and remote)	0.33	0.60	0.36	0.98	0.52	0.78
$\lambda_1\lambda_2$ (maximum f'' and minimum f')	0.17	0.10	0.16	0.92	0.96	1.0

Table 4
MAD phasing statistics for PhoB at the gold *L* edge.

	Figure of merit			$R_{\rm Cullis}$		
	Acentric	Centric	All	Anomalous	Centric	All
$\lambda_1\lambda_2\lambda_3$	0.33	0.26	0.32	0.90	0.90	0.65
$\lambda_1\lambda_3$ (above L_{111} and remote)	0.21	0.02	0.18	0.92	1.00	1.00
$\lambda_2\lambda_3$ (minimum f' and remote)	0.23	0.18	0.22	0.92	0.94	0.97
$\lambda_1\lambda_2$ (maximum f'' and minimum f')	0.20	0.02	0.17	0.93	1.00	1.00

not as much as would be predicted from the relative magnitude of anomalous and dispersive differences.

The importance of a large dispersive difference is even more obvious when the data measured at the wavelengths λ_1 and λ_2 were used. In this case, the anomalous differences were maximized, but the dispersive signal was minimal; the quality of both the phasing and the maps decreases significantly and the breaks in the main density of the maps (Fig. 6) render interpretation more difficult. At the same time, it is worth noting that when the Bijvoet differences at a wavelength are too small, as in the example of the data collected below the iridium $L_{\rm III}$ absorption edge, the data will not contribute significantly to the phasing, as shown in Table 5.

4. Discussion

Our analysis suggests that measuring large dispersive differences provides more precise phase information than that obtained from large anomalous differences, at least in the absence of a strong white line. While a good dispersive difference can compensate for a poor Bijvoet difference, the inverse is in general not true. The quality of the maps seems to be more directly related to the magnitude of the dispersive signal. In the case of the LukF-PV, the electron-density maps were of similar quality when the Bijvoet differences were optimized. However, it is worth noting that the size of the calculated fractional anomalous difference is 75% larger than the maximum dispersive difference (Table 1).

Several reasons may be pointed out to support the importance of the dispersive contribution. The first is that a good dispersive signal allows a fairly accurate estimate of the errors in the heavy-atom model; these are calculated from the amplitude differences of the centric reflections for which the structure-factor vectors are almost parallel at different wavelengths. In this case, the lack-of-closure error can be calculated from the expression

$$\mathrm{err}_{\rm disp} = |\mathbf{F}_T| - |\mathbf{F}_N| - |\mathbf{F}_A|, \qquad (5)$$

with $|\mathbf{F}_T|$ being the measured amplitude values, $|\mathbf{F}_N|$ the contribution from the reference and $|\mathbf{F}_A|$ the contribution from the anomalous scatterer. $\mathrm{err}_{\rm disp}$ is derived solely from the dispersive differences, as Bijvoet differences are not measured in centric reflections. From (5), it is easy to estimate the error in the heavy-atom calculated structure factor err_A^2 as

$$\mathrm{err}_A^2 = \mathrm{err}_{\rm disp}^2 - \mathrm{err}_T^2, \qquad (6)$$

where err_T is the statistical error on the measured amplitudes, which is estimated from counting statistics. This is equivalent to the situation in isomorphous replacement (Blow & Crick, 1959), where better refinement of the heavy-atom parameters results in more reliable phases for the acentric reflections. On the other hand, since Bijvoet differences cannot be derived from centric reflections, large values will not contribute as much to the accurate estimation and refinement of the heavy-atom parameters. This may explain why a larger dispersive difference results in a better anomalous $R_{\rm Cullis}$, even if the anomalous differences are smaller.

The importance of the accurate refinement of the anomalous scatterer parameters was illustrated by refining the

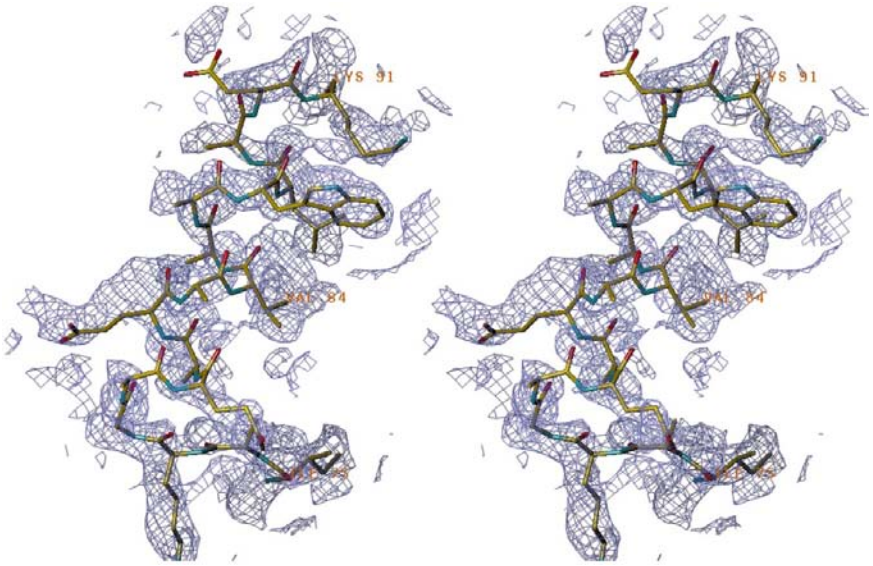

Figure 4
As Fig. 3, calculated with phases from λ_1 and λ_3 (above absorption edge and remote).

heavy-atom positions and occupancies against a MAD data set with large dispersive differences and using the resultant refined parameters for phasing the two-wavelength data set containing the smallest dispersive difference (λ_1 and λ_2). The results for the cytochrome *c*553 data are shown in Fig. 7.

A second factor to consider is that MAD data are almost always collected in a way (same crystal in the same orientation for all wavelengths) which minimizes systematic errors for the dispersive differences, but not necessarily for the anomalous differences, even in a well oriented crystal (Evans, 1997). Moreover, the anomalous difference measurements suffer to a larger extent from a lack of data redundancy. For example, by collecting the minimum angular slice required for a complete data set in the space group $P2_12_12_1$, the redundancy for each dispersive difference measurement will be larger than 3, but will be approximately half this for anomalous differences. This will lead, in practice, to less accuracy in the measurement of a Bijvoet difference compared with that of a dispersive difference of the same magnitude.

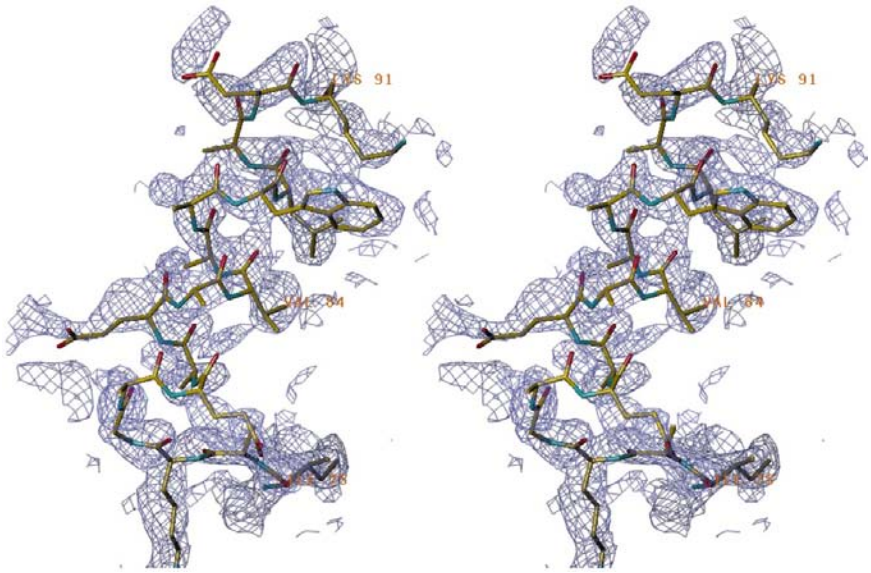

Figure 5
As Fig. 3, calculated with phases from λ_2 and λ_3 (inflexion point and remote). This map is the one which most closely resembles the one calculated with data at all three wavelengths.

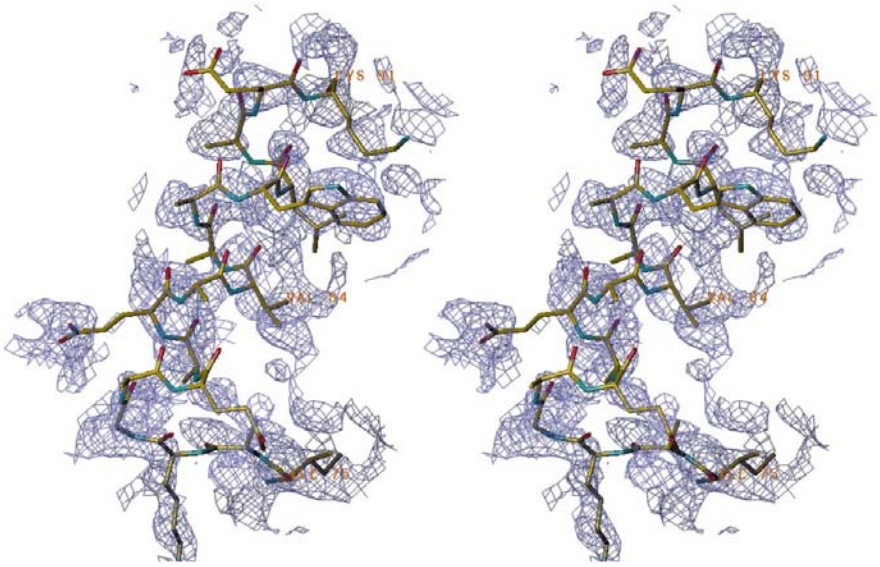

Figure 6
As Fig. 3, calculated with phases from λ_1 and λ_2 (inflexion point and above the edge).

5. Conclusions

5.1. Two-wavelength MAD experiments?

This study shows not only that two-wavelength MAD phasing can produce electron-density maps which can be fully interpretable, but also that the maps are not substantially worse than those obtained with three or more wavelengths. This conclusion is valid provided that the wavelengths are chosen in such a way that the dispersive differences between the two data sets are around 3–4% or better. One consequence of this is that MAD experiments can be considerably shortened. On third-generation undulator sources, where exposure time is not so much of an issue but where crystal decay might well be, a two-wavelength medium-resolution (2–2.5 Å) MAD experiment would be a fast and safe way to secure a set of phases from which the structure could be determined. With longer lasting crystals it would be advisable to pursue data collection at the same wavelength in order to increase redundancy, which will improve the phasing statistics. In any case, increasing the resolution of the data beyond 2 Å whenever possible would be a better option than collecting a data set at a third wavelength, since it has been proved for a substantial number of cases that, provided high-resolution data are available, good results can be obtained using the *wARP* program for automated electron-density map interpretation (Perrakis *et al.*, 1999).

When the available dispersive signal is small, because of low occupancy of the anomalous scatterer or in the case of large molecules, we should assess whether a two-wavelength experiment is still a valid approach or not. Obtaining more phase information from the anomalous differences would require an experimental setup which further decreases systematic errors. This can be achieved by orienting the crystal perfectly, so that Bijvoet pairs are recorded in the same image, or by collecting highly redundant data sets. The question of whether improved accuracy of the data can supply as much information as a third wavelength in borderline cases is

Table 5
MAD phasing statistics for LukF-PV at the iridium *L* edge.

	Figure of merit			R_{Cullis}		
	Acentric	Centric	All	Anomalous	Centric	All
$\lambda_1\lambda_2\lambda_3\lambda_4$	0.43	0.42	0.43	0.77	0.77	0.85
$\lambda_1\lambda_2\lambda_3$	0.41	0.35	0.40	0.78	0.79	0.87
$\lambda_1\lambda_3$ (white line and high-energy remote)	0.24	0.21	0.23	0.84	0.97	0.99
$\lambda_2\lambda_3$ (minimum f' and high-energy remote)	0.23	0.24	0.23	0.93	0.86	0.93
$\lambda_1\lambda_2$ (white line and minimum f')	0.24	0.05	0.22	0.84	0.99	1.00
$\lambda_1\lambda_4$ (white line and low-energy remote)	0.15	0.04	0.14	0.91	1.00	1.00
$\lambda_2\lambda_4$ (minimum f' and low-energy remote)	0.13	0.05	0.13	0.95	1.00	0.99
$\lambda_3\lambda_4$ (high- and low-energy remote)	0.11	0.04	0.10	0.96	1.00	0.99

very important and should be addressed in a separate experiment.

An extension of the two-wavelength MAD approach would be to measure data at a single wavelength where f'' is large and to use density modification to solve the phase ambiguity. Single anomalous dispersion (SAD) phasing has been used, for example, by de La Fortelle *et al.* (1997*b*) and Benini *et al.* (in preparation). Although this approach did not appear to be successful for the examples presented in this paper, it is a possibility to consider when a good anomalous signal combined with high-resolution data are available.

5.2. Strategy for MAD data collection

When the MAD experiment involves a heavy-atom derivative without significant white-line features at the absorption edge, which is the usual case for atoms with atomic number greater than 78, the optimal setting would be as follows.

(i) At the L_{III} edge inflexion point.

(ii) A remote wavelength with a high f'' value. Although sucessful two-wavelength experiments can be carried out selecting this wavelength above the L_{III} edge (see, for example, Wimberly *et al.*, 1997), it is possible to obtain both better dispersive and anomalous signals by choosing a wavelength corresponding to an energy 200–300 eV above the L_I edge. This choice is better than measuring data closer to the edge, because a 30% increase of the dispersive signal can be obtained while the anomalous signal hardly decreases.

If the L_{III} absorption edge occurs at an energy greater than 15 keV (approximately 0.83 Å), it may be necessary to choose a remote wavelength below the edge, since the L_I edge will not be accessible at most protein crystallography beamlines. For example, in the case of uranium derivatives, where the L_{III} absorption edge is around 17.166 keV (λ = 0.722 Å), data collected at 9.184 keV (λ = 1.35 Å) will contribute −4 electrons for f' and a slightly larger f'' value than that available above the absorption edge.

Table 6
MAD phasing statistics (from the program *SHARP*) for IF3-C at the selenium *K* edge.

	Figure of merit			R_{Cullis}		
	Acentric	Centric	All	Anomalous	Centric	All
$\lambda_1\lambda_2\lambda_3$	0.65	0.73	0.66	0.37	0.29	0.35
$\lambda_1\lambda_3$ (maximum f'' and remote)	0.56	0.59	0.56	0.64	0.46	0.45
$\lambda_2\lambda_3$ (minimum f' and remote)	0.57	0.64	0.58	0.79	0.42	0.41
$\lambda_1\lambda_2$ (maximum f'' and minimum f')	0.47	0.33	0.46	0.64	0.49	0.52

In the case of synchrotron beamlines operating at high energies, both a low-energy and a high-energy remote wavelength giving a good contrast in f' can be attained. In this case, one should take into account other factors, such as beamline radiation spectrum, detector parameters and crystal size (Popov & Lamzin, 1998), to select the optimal wavelength for the experiment.

On those occasions where information about the anomalous scatterer sites is known prior to the experiment, if the mean dispersive difference evaluated from expression (4) is less than 3%, it would be advisable to sacrifice some anomalous signal and to collect the first wavelength at even higher energy, assuming that the beamline configuration allows it. In this particular case, a third wavelength just above the absorption edge might provide a slightly higher f'', but in most cases two wavelengths should yield most of the phasing power available from the derivative.

There is a wider choice of strategies when the heavy-atom derivative present is a transition metal with an atomic number between 62[4] and 78. The presence of empty states in the 5*d* atomic band of these elements allows high-probability 2*p*–5*d* transitions, which cause strong white-line features (Brown *et al.*, 1977). Although the optimization of the dispersive signal should give good results in this case, first collecting the data set which provides the largest amount of phasing information by itself (maximum f'' at the white-line peak on the L_{III} or even at the L_{II} edge) is the safest way to proceed. The second data set could be measured at a remote high-energy wavelength above the L_I edge or, in the case of lanthanides, at the descending inflexion point of the white line (Hendrickson & Ogata, 1997), in order to improve the dispersive differences. Data collection at a third wavelength at the L_{III} edge to obtain the minimum f' should be optional.

For a MAD experiment on a *K* edge, we suggest the collection of one data set at the edge and another one at a remote high energy where f' is close to zero (at least 2–3 keV

[4] For atoms lighter than samarium, the edge transition takes place at too low an energy for a conventional MAD experiment to be considered feasible.

above the absorption edge) as suggested by Fig. 2. Such a remote wavelength would be a good compromise between obtaining a large dispersive difference and a reasonable Bijvoet difference. However, if data collection in the region where f' becomes positive is possible and the expected anomalous signal is no worse than that present in the iron MAD experiment discussed above, the experiment could benefit from the additional one or two electrons contributing to the dispersive differences. When the edge occurs at high energy, as it is the case for strontium, a low-energy remote should be considered instead.

If a strong white line is present, as is often the case with selenium, the same considerations as for the L edges would apply and data collection at the f'' maximum is preferred. However, at beamlines with a bandpass larger than around 1×10^{-4}, the spectral resolution may not be sufficient to obtain a significant increase of f'' at the white line (Thompson, 1997). In this case, data collection at the inflexion point would be preferred. The depth of the f' resonance is also affected by a wide bandpass, but collection at this wavelength still provides the smallest f' value. In the case of L edges, white lines tend to be wider than the ones at the K edges (Krause & Oliver, 1979) and are less likely to be significantly affected by a larger bandpass.

6. Summary

In the three different cases presented in this work, where medium-resolution data of reasonable quality were available, we have shown that a two-wavelength MAD experiment would be a valid strategy to solve protein structures.

In theory, the anomalous differences should contribute more than the dispersive differences to the solution of the phase problem, but in practice the latter appear to be more important. Two reasons can be put forward. Firstly, the refinement of the anomalous scatterer positions and occupancies is highly dependent on the dispersive differences measured from centric reflections. Secondly, a MAD data collection is usually performed in a way which partially cancels out systematic errors in the structure factors at different wavelengths, leading to dispersive differences which are less affected by errors in the data. Although it would be possible to minimize errors in the anomalous differences by using local scaling procedures (see Hendrickson & Teeter, 1981; Friedman *et al.*, 1995; Evans, 1997) or by orienting the crystal so that Bijvoet mates appear on the same diffraction image, the redundancy of anomalous pairs will always remain small. We therefore consider that the best strategy for a successful MAD experiment is to choose the wavelengths which will provide a large difference in the real part of scattering factor f' between data sets, even when this implies a small decrease of the magnitude of the imaginary part f''.

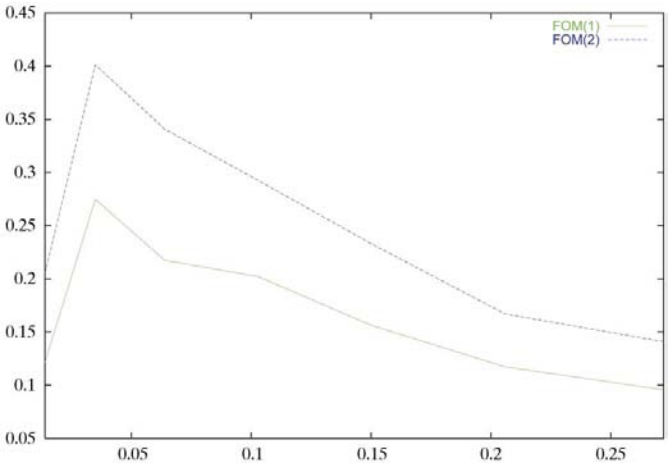

Figure 7
Plot of the mean figure of merit against resolution for the *c*553 cytochrome. FOM(1) represents the figure of merit after phasing refinement of the iron contribution to structure factor against the two-wavelength MAD data set at λ_1 and λ_2. FOM(2) depicts the values obtained when phasing against the same two wavelengths, but after refining the iron contribution against a three-wavelength data set. The maps calculated with the resultant phases had correlations of 0.58 and 0.69, respectively, with the three-wavelength map.

The IF3-C MAD data were used with permission from Valérie Biou and Venki Ramakrishnan. The authors wish to thank Paul Tucker (EMBL) for helpful comments on the manuscript. The experiments at the EMBL Hamburg outstation were supported by the European Union through the Access to Large Installations project. The PhoB project was supported by grants PB95-224 (Ministerio de Educación y Cultura) and 19975GR-275 (Generalitat de Catalunya). SB thanks the EMBL for a predoctoral fellowship.

References

Benini, S., Ciurli, S., Rypniewski, W. & Wilson, K. (1997). *Proteins Struct. Funct. Genet.* **28**, 580–585.
Biou, V. (1997). *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 97–102. Warrington: Daresbury Laboratory.
Biou, V., Shu, F. & Ramakrishnan, V. (1995). *EMBO J.* **14**, 4056–4064.
Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
Brennan, S. & Cowan, P. L. (1992). *Rev. Sci. Instrum.* **63**, 850–853.
Brown, M., Peierls, R. E. & Sterm, E. A. (1977). *Phys. Rev. B*, **15**, 738–749.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dauter, Z. (1997). *Methods Enzymol.* **276**, 326–344.
Deacon, A., Habash, J., Harrop, S. J., Helliwell, J. R., Hunter, W. N., Leonard, G. A., Peterson, M., Hadener, A., Kalb, A. J., Allinson, N. M., Castelli, C., Moon, K., McSweeney, S., González, A., Thompson, A. W., Ealick, S., Szebenyi, D. M. & Walter, R. (1995). *Rev. Sci. Instrum.* **66**(2), 1287–1292.
Evans, G. (1994). PhD thesis, University of Warwick, England.
Evans, P. R. (1997). *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 97–102. Warrington: Daresbury Laboratory.
Friedman, A. M., Fischman, T. O. & Steitz, T. A. (1995). *Science*, **268**, 1721–1727.
Fourme, R., Shepard, W. & Kahn, R. (1996). *Prog. Biophys. Mol. Biol.* **64**, 167–199.

Helliwell, J. R. (1992). *Macromolecular Crystallography with Synchrotron Radiation*, p. 345. Cambridge University Press.
Helliwell, J. R. (1997). *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 1–7. Warrington: Daresbury Laboratory.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Hendrickson, W. A. & Teeter, M. (1981). *Nature (London)*, **290**, 107–113.
Krause, M. O. & Oliver, J. H. (1979). *J. Phys. Chem. Ref. Data*, **8**(2), 329–338.
La Fortelle, E. de, Irwin, J. J. & Bricogne, G. (1997*a*). In *Crystallographic Computing 7*, edited by P. Bourne & K. Watenpaugh. Oxford: IUCr/Oxford University Press.
La Fortelle, E. de, Irwin, J. J. & Bricogne, G. (1997*b*). *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 25–39. Warrington: Daresbury Laboratory.
Merritt, E. A. (1998). *Anomalous Scattering Coefficients*, http://brie.bmsc.washington.edu:80/scatter/AS_form.html.
Narayan, R. & Ramaseshan, S. (1981). *Acta Cryst.* A**37**, 636–641.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 80–86. Warrington: Daresbury Laboratory.
Pédelacq, J.-D., Maveyraud, L., Prévost, G., Lamine, B.-M., González, A., Courcelle, E., Shepard, W., Monteil, H., Samama J.-P. & Mourey, L. (1999). *Structure*, **7**(3), 277–287.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). In the press.
Peterson, M. R., Harrop, S. J., McSweeney, S. M., Leonard, G. A., Thompson, A. W., Hunter, W. N. & Helliwell, J. R. (1996). *J. Synchrotron Rad.* **3**, 24–34.
Popov, A. & Lamzin, V. (1998). *Sixth International Conference on Biophysics and Synchrotron Radiation*, *Poster abstracts*, p. 50. Argonne, IL, USA.
Read, R. J. & Schierbeek, A. J. (1988). *J. Appl. Cryst.* **21**, 490–495.
Roussel, A. & Cambillau, C. (1989). *Silicon Graphics Geometry Partners Directory*, pp. 77–79. Silicon Graphics, Mountain View, California, USA.
Smith, J. L., Thompson, A. & Ogata, C. M. (1996). *Synchrotron Radiat. News*, **9**(6), 12–14.
Smith, J. L. (1997). *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 25–39. Warrington: Daresbury Laboratory.
Solà, M., Gomis-Rüth, F. X., Serrano, L., González, A. & Coll, M. (1999). *J. Mol. Biol.* **285**, 675–687.
Terwilliger, T. C. & Berendzen, J. (1997). *Acta Cryst.* D**53**, 571–579.
Thompson, A. (1997). *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 8–24. Warrington: Daresbury Laboratory.
Wimberly, B. T., White, S. W. & Ramakrishnan, V. (1997). *Structure*, **5**(9), 1187–1197.