

## Is it jolly SAD?

**Eleanor Dodson**

Department of Chemistry, University of York,  
Heslington, York YO10 5DD, England

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

Examples of phasing macromolecular crystal structures based on single-wavelength anomalous dispersion (SAD) have demonstrated that this approach may have general applications in structural biology. With better data-collection facilities and cryogenic techniques, combined with powerful data-processing, phasing and density-modification programs, the SAD approach may prove simpler than phasing from multi-wavelength (MAD) measurements. It can be performed at any wavelength where anomalous scattering can be observed, in many cases using laboratory X-ray sources. However, there is still a need for accurate data, successful phase improvement and a certain amount of luck. This paper extends the discussion of Jolly SAD in Dauter *et al.* [Dauter, Z., Dauter, M. & Dodson, E. (2002), *Acta Cryst.* **D58**, 494–506].

Received 4 June 2003

Accepted 22 September 2003

### 1. Introduction

Structure determination is now recognized as one of the most effective methods for acquiring biological insight and there is pressure on crystallographers to provide accurate information as quickly and painlessly as possible. The diffraction experiment only gives the intensities arising from the atomic distribution of the molecule within a lattice, so in order to produce an interpretable image it is necessary to determine the associated phases. The most unbiased approach to finding these is by modifying the observations in some predictable way. Formally, an experimental phase for any reflection can be uniquely estimated from three measurements of associated amplitudes, provided that the vector describing the differences can be calculated. This is easy if the coordinates of those atoms generating the differences, described as the substructure, are known. Once phases and error estimates have been obtained, there are powerful ways to refine them using known properties of the protein electron density, *e.g.* flattening the density in the solvent region, modifying the density within the protein or averaging the density for different copies of a molecule.

#### 1.1. Historical background

The method of multiple isomorphous replacement (MIR) uses related crystals where the substructure consists of additional atoms, usually heavy metals, soaked into the crystal from appropriate salt solutions. There is always a problem with isomorphism with respect to the native crystal; the salts often cause other rearrangements within the lattice apart from introducing the heavy atoms.

This approach is augmented by exploiting the anomalous dispersion differences between  $F(h, k, l)$  and  $F(-h, -k, -l)$  resulting from the resonant scattering of the substructure within the derivative(s) (the acronyms MIRAS or SIRAS stand for multiple or single isomorphous replacement with

anomalous scattering). Such differences are not affected by non-isomorphism, but are usually much weaker than the isomorphous differences and therefore harder to measure accurately.

### 1.2. Current practice

Once a model of the substructure is obtained, it must be refined in order to improve its ability to predict the observed differences. Simultaneously, it is used to deduce protein phases from these differences and from the calculated heavy-atom model structure factors (Blundell & Johnson, 1976; Drenth, 1999). Although formally a phase can be determined from three observations and the appropriate model, the errors in both measurements and models mean that it is essential to use a probabilistic approach to assign an appropriate weight to the phase. In addition, the methods used for finding the positions of the model atoms cannot distinguish the hand of the solution and the phasing geometry is equally well described by the model or by its mirror image. If there is only one partial structure (as for SAD, MAD or SIRAS experiments), the correct enantiomer can only be chosen by assessing which hand generates the better electron-density map. The underlying theory is reviewed in this volume and described in many classic texts, *e.g.* Blow & Rossmann (1961), North (1965), Mathews (1966), Dodson & Vijayan (1971), Fourme *et al.* (1996), Blundell & Johnson (1976), Drenth (1999).

### 1.3. Using the anomalous signal alone

The multi-wavelength anomalous diffraction (MAD) method uses only the wavelength-dependence of the atomic structure factor of the anomalously scattering atoms for solving the phase problem (Phillips & Hodgson, 1980; Karle, 1980; Hendrickson, 1991, 1999). In this approach, several data sets are collected at various wavelengths around the absorption edge of the anomalous scatterer present in the crystal and the differences in the  $f'$  and  $f''$  contributions are utilized for phase calculation. Such MAD experiments are possible only at synchrotron X-ray sources, where the X-ray wavelength can be tuned to the desired values. The anomalous scatterer used for MAD phasing may be inherently contained in the metalloprotein (*e.g.* Zn, Cu, Fe), introduced by soaking (classic heavy atoms, *e.g.* Hg, Pt, Au compounds or halide ions in the solvent shell) or by metabolic or chemical modification, such as those used to incorporate selenomethionine in proteins or bromouracil in DNA (Boggon & Shapiro, 2000). If conditions are favourable, the phasing power is excellent.

However, it has been proposed (González *et al.*, 1999) and demonstrated that sufficiently good phase estimates may be obtained by collecting more accurate data at fewer wavelengths. In some cases, data collected at one wavelength have been sufficient to determine the phases of both test and novel structures, as demonstrated by the solution of the structure of crambin (Hendrickson & Teeter, 1981) and advocated by Wang (1985). Dauter *et al.* (2002) give many examples where this approach of single-wavelength anomalous dispersion (SAD) coupled with increasingly powerful phasing and

density-modification algorithms (La Fortelle & Bricogne, 1997; Hauptman, 1996; Langs *et al.*, 1999; Cowtan, 1999; Terwilliger, 2000) can solve the phase problem for macromolecular structures.

## 2. Background of phase determination

Phase determination is covered elsewhere in this volume and only a brief outline is presented here. X-rays are diffracted by atoms positioned within a crystal lattice. Most diffraction arises from the electrons surrounding the atomic nucleus and since this electron cloud has a radius comparable to the X-ray wavelength, the contribution falls off at higher diffraction angles, *i.e.* at higher resolution. This is represented by the atomic form factor. Such a signal from the whole atom is isotropic and can be treated as a real number,  $f^0(\theta)$ .

If X-rays can excite those electrons that are able to jump from lower to higher energy shells, an auxiliary resonant anomalous signal is observed and the atomic form factor can be expressed as a complex number  $f' + if''$ . Generally,  $f''$  is proportional to the atomic absorption of the X-rays and to their fluorescence and  $f'$  follows the derivative of this function, according to the Kramer–Kronig transformation (James, 1958). In contrast to the normal atomic scattering factor  $f^0$ , the anomalous dispersion corrections  $f'$  and  $f''$  depend only on the wavelength  $\lambda$  of the X-rays used for the diffraction experiment and do not diminish with the diffraction angle. The full atomic form factor is

$$f(\theta, \lambda) = f^0(\theta) + f'(\lambda) + if''(\lambda).$$

In macromolecules, most of the atoms have negligible  $f'(\lambda)$  and  $f''(\lambda)$  and there are only a few anomalous scatterers, so that the total anomalous dispersion generates only small differences in intensity. The diffraction data must be measured very accurately to allow these differences to be utilized for phasing.

When all atomic form factors are real with zero  $f''$  contribution Friedel's law holds, so that  $F(h, k, l)$  and  $F(-h, -k, -l)$  have the same magnitude and  $\varphi(h, k, l) = -\varphi(-h, -k, -l)$ . However, when the form factor contains an imaginary contribution  $if''$ , the reflections  $F(h, k, l)$  and  $F(-h, -k, -l)$  have different intensities and their phases are no longer complementary. In the MAD technique, where several data sets are measured at different wavelengths  $\lambda_i$  with different values for the dispersive difference  $f'$  and the anomalous difference  $f''$ , two associated but different measurements of the amplitudes are obtained for each wavelength. Once the positions of the anomalous scatterers are known and the magnitudes of  $f'$  and  $f''$  for this wavelength have been estimated, the protein phases for the reflections can formally be derived in an analogous way to the MIRAS approach.

Once again, the procedure has two independent stages. Firstly, the positions of the anomalous scatterers have to be deduced from Patterson or direct-methods searches using coefficients derived from either dispersive or anomalous differences or from a combination of both; secondly, the position and precise values of  $f'$  and  $f''$  for the partial structure

needs to be refined in order to maximize its ability to predict the observed differences.

### 3. Single-wavelength phasing

It is not formally possible to evaluate a protein phase exactly if there are only two experimental measurements, *e.g.* when the data are restricted to one wavelength (SAD) with only a single anomalous difference available or in the SIR case when only the native and one derivative data set is measured. Even assuming that the measured protein amplitudes,  $F^+$  and  $F^-$ , and the calculated amplitude and phase contributions of the anomalous partial structure,  $F_a$  and  $\varphi_A$ , are error-free, there is a twofold ambiguity in the estimation of the protein phase (Ramachandran & Raman, 1956). Fig. 1 shows that for the SAD case, where all the anomalous scatterers are of the same kind, the two possible phase values of the protein structure factor,  $\varphi_T$ , are symmetrically oriented around  $(\varphi_A - 90^\circ)$ . There is a phase error for either solution of  $(\varphi_T - \varphi_A + 90^\circ)$ , with an associated figure of merit of  $\cos(\varphi_T - \varphi_A + 90^\circ)$ . Note that all centric reflections where  $(\varphi_T - \varphi_A)$  must be either 0 or  $180^\circ$  have figures of merit of zero. (Analogously, for the SIR case the two possible values of the protein phase are symmetrically oriented about the heavy-atom phase,  $\varphi_H$ .) Thus, a unique protein phase could only be determined if the protein and anomalous scatterer phases differ by  $90^\circ$ , when the two solutions would coincide. (These reflections also have the maximum possible Bijvoet difference.)

The relation between the Bijvoet difference,  $\Delta F^\pm$ , the phase of the protein,  $\varphi_T$ , and that of the anomalous substructure,  $\varphi_A$ , can be deduced from Fig. 1,

$$F^{+2} - F^{-2} = 4F_i F_a'' \sin(\varphi_T - \varphi_A).$$

If the contribution of the anomalous scattering to the total diffracting power of the crystal is small,  $F_a \ll F_i$ , then  $(|F^+| + |F^-|)/2 \simeq F_i$  and

$$\Delta F^\pm = |F^+| - |F^-| \simeq 2F_a'' \sin(\varphi_T - \varphi_A).$$

Defining  $\theta = \cos^{-1}(\Delta F^\pm/2F_a'')$  and since  $\sin(\varphi_T - \varphi_A) = \sin(180^\circ - \varphi_T + \varphi_A)$ ,

$$\varphi_T - \varphi_A = 90^\circ + \theta \text{ or } 90^\circ - \theta.$$

Except when  $\theta = 90^\circ$ , the ambiguity follows.

The probability of phase distribution resulting from anomalous scattering (Hendrickson, 1979) can be expressed,

$$P_{\text{anom}}(\varphi) = N \exp\{-[\Delta F^\pm + 2F_a'' \sin(\varphi_T - \varphi_A)]^2/2E^2\},$$

where  $N$  is the normalizing factor and  $E$  the standard error estimation.

However, since the anomalous scatterers are part of the structure,  $\varphi_T$  will be correlated with  $\varphi_A$  and of the two possibilities resulting from the sine ambiguity, there is a slightly higher probability that the protein phase,  $\varphi_T$ , has the value closer to  $\varphi_A$ . Sim (1959) derived the statistical probability of the protein phase estimated from the known partial structure as

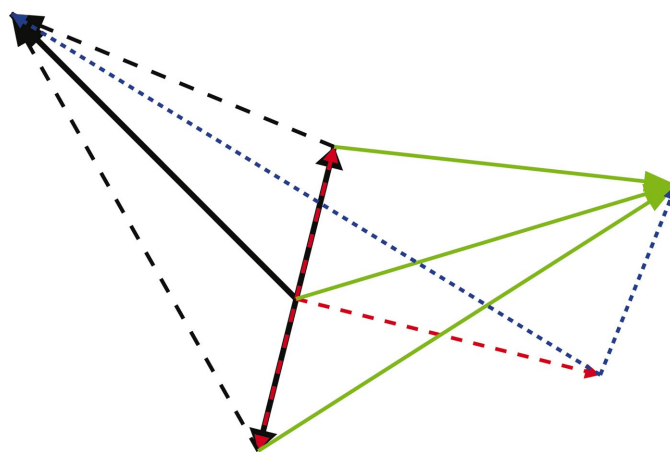
$$P_{\text{par}}(\varphi_T) = N \exp[2(|F_i||F_a|/F_u^2) \cos(\varphi_T - \varphi_A)],$$

where  $F_u^2$  is the contribution of the normally scattering (unknown) atoms. The total phase probability is obtained from a combination of the Sim-weighted estimate and that derived from the SAD equation. Modern programs such as *SHARP* (de La Fortelle & Bricogne, 1997), *SOLVE* (Terwilliger & Berendzen, 1999), *BP3* (Pannu *et al.*, 2003) and *MLPHARE* (Otwinowski, 1991) endeavour to provide realistic starting probabilistic estimates of the initial phases and figures of merit.

### 4. Phase-improvement techniques

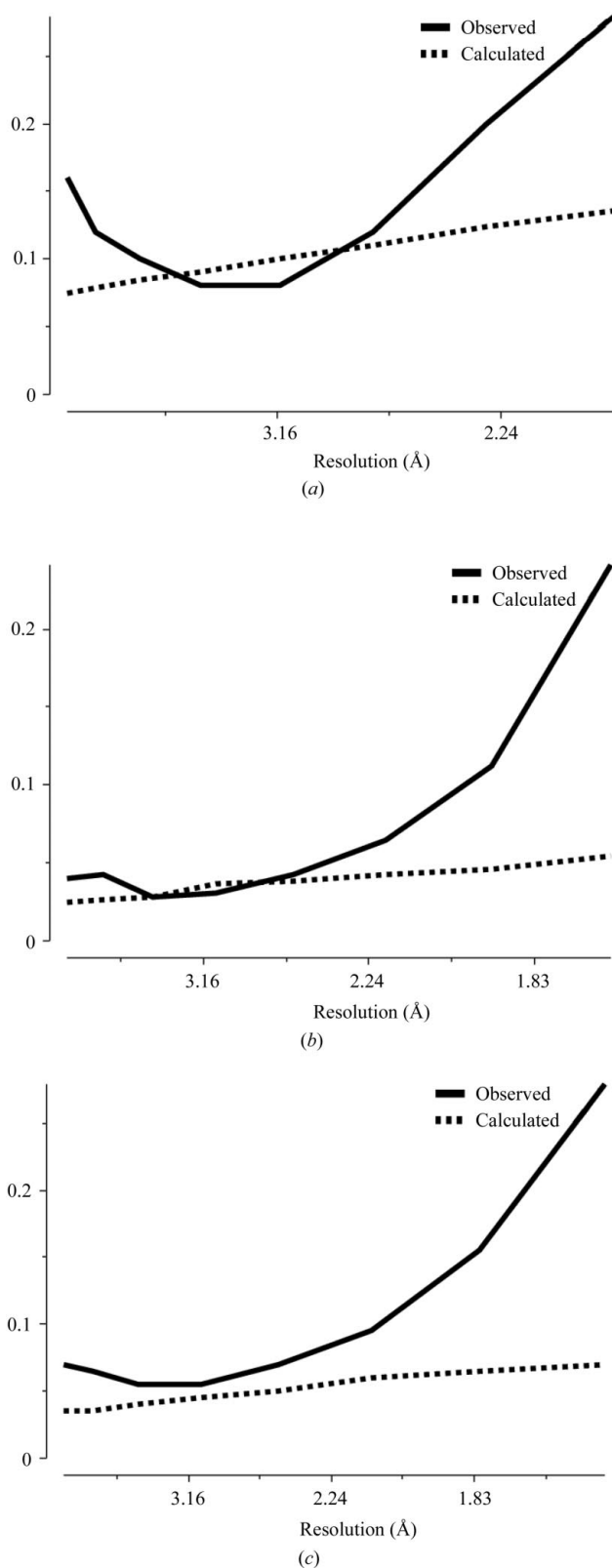
The problem of resolving the SAD phase ambiguity for reflections has been tackled by various methods: resolved anomalous phasing, used originally by Hendrickson & Teeter (1981) for the solution of crambin, the iterative single-wavelength anomalous scattering (ISAS) approach, introduced by Wang (1985), and direct-methods applications as proposed by Hauptman (1982, 1996) or by Fan *et al.* (1990).

However, the most powerful approach to improving the phase distributions uses density-modification procedures such as those programmed in *SOLOMON* (Abrahams & Leslie, 1996), *DM* (Cowtan, 1999) and *RESOLVE* (Terwilliger, 2000). The methods all modify the initial density, use this to generate a new set of phases which are combined with the experimental ones and then repeat the cycle. Providing the phase errors are properly estimated and the solvent boundary correctly outlined, this is extremely effective. Since the method depends on the recognition and enhancement of interpretable features



**Figure 1**

Part of the Argand diagram showing various contributions to the scattering factors. The measured amplitudes of both Friedel mates and their mean ( $F_i^+$ ,  $F_i^-$  and  $F_i$ ) are shown in black and green, those of the anomalous scatterers ( $F_a$  and  $F_a''$ ) in red and the resultant contribution of the normally scattering atoms ( $F_p$ ) in blue. (The likely contribution of  $F_a$  has been grossly exaggerated to clarify the figure.) The magnitudes of  $F_i^+$  and  $F_i^-$  are known and once the anomalous substructure has been positioned, the red vectors can be calculated. Two solutions for  $F_i$  are then possible, with their phase,  $\varphi_T$ , symmetrically placed on either side of  $\varphi_A - 90^\circ$ . The contribution of the normal scatterers,  $F_p$  will be different in the two cases.



**Figure 2**

The calculated and observed  $\langle \Delta F^{\pm} \rangle / \langle F \rangle$  ratio as a function of resolution for (a) Cbm27(i), (b) ProtE and (c) Cel5a. Cbm27(i) and ProtE were successfully phased; although the high-resolution differences were overestimated, they provided some phase information which was then improved by density modification. The very weak signal from the S atoms in Cel5a was not sufficient to trigger useful phasing.

in the electron density and the maps based on phases derived from the two enantiomorphs differ in quality, this procedure should also select the correct enantiomorph.

After this procedure, automated model building, cycled with maximum-likelihood weighted refinement of the partial model to further improve the phasing, can lead to a near-complete model in a very short time. The method is available in software packages such as *ARP/wARP* (Perrakis *et al.*, 1999) and *RESOLVE*.

### 5. Finding the positions of the substructure atoms

These have to be deduced from Patterson or direct-methods searches using coefficients derived from isomorphous, dispersive or anomalous differences or from a combination of both. The methodology has been described in detail in many places, *e.g.* Weeks *et al.* (2003) and references therein. To summarize, for isomorphous differences

$$|F_{PH}| - |F_P| \simeq 2|F_H| \cos(\varphi_T - \varphi_H)$$

and for anomalous differences

$$|F^+| - |F^-| \simeq 2F'_A \sin(\varphi_T - \varphi_A).$$

Thus, in principle, the positions of anomalous scatterers can be found from the Bijvoet differences for single-wavelength data. There are several powerful automated search procedures such as those programmed in *SnB* (Miller *et al.*, 1994), *SOLVE* (Terwilliger & Berendzen, 1999), *SHELXD* (Schneider & Sheldrick, 2002) or *ACORN* (Foadi *et al.*, 2000) which are usually successful, even with incomplete difference data extending to a resolution sufficient to separate the sites, perhaps to 3.5 Å. However, all depend primarily on the large differences and are likely to fail if there are even a few overestimated outliers. Collecting multiple observations seems to be almost essential for detecting and removing such rogue reflections. Fortunately, Patterson search methods work well when there are only a few substructure sites, whereas direct-methods procedures work best when there are many sites scattered throughout the unit cell, as is often the case for Se substitution. If the differences are reliable, it is possible to find many sites.

### 6. Estimation of the amount of anomalous signal in diffraction data

The mean ratio of the Bijvoet difference to the total protein amplitude is

$$\langle \Delta F^{\pm} \rangle / \langle F \rangle = 2^{1/2} (N_A^{1/2} f''_A) / [N_P^{1/2} f_{\text{eff}}(\theta)],$$

where  $f_{\text{eff}} = (1/N) \sum f_i$  is the effective scattering of an average atom at diffraction angle  $\theta$ . The anomalous scattering signal  $f''$  does not depend on the resolution, but  $f_{\text{eff}}$  reduces with resolution and thus the percentage of anomalous signal could be expected to increase at high resolution, especially if the temperature factors of the anomalous scatterers are lower than the average value for all atoms of the macromolecule. However, weak intensities, which are more likely at high

**Table 1**

Statistics of X-ray data.

Values in parentheses are for the highest resolution shell.

Crystal	Cbm27(i)	ProtE(i)	ProtE(ii)	Cbm27(ii)	Lipase	Cel5A
Space group	<i>P</i> 4 <sub>1</sub> 2 <sub>1</sub> 2	<i>I</i> 4 <sub>1</sub> 22	<i>I</i> 4 <sub>1</sub> 22	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 4 <sub>3</sub> 2 <sub>1</sub> 2	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit-cell parameters (Å)						
<i>a</i> (Å)	69.9	70.2	70.2	31.7	92.2	55.3
<i>b</i> (Å)	69.9	70.2	70.2	48.9	92.2	69.7
<i>c</i> (Å)	229.5	71.9	71.9	96.9	299.4	76.9
Wavelength (Å)	0.98	1.54/0.93	0.93	0.98	0.98	1.54
Resolution (Å)	2.00	1.7/1.3	1.3	1.78	2.10/2.78	1.68
Measured reflections	422308	162960	110389	97781	—	184654
Unique reflections	38522	11210	22383	15281	75988/33124	33982
Multiplicity	5.2	13	5.0	7.2	—	5.3
<i>R</i> <sub>merge</sub> <sup>†</sup> (%)	7.2 (34.7)	4.5 (34.8)	4.2 (47.0)	6.6 (26.9)	—	3.7 (12.5)
<i>I</i> / <i>σ</i> ( <i>I</i> )	10.5 (2.7)	23.6 (3.2)	—	28.2 (5.3)	6.4 (2.2)	40.7 (13.8)
Completeness (%)	100.0 (100.0)	99.9 (99.8)	99.9 (89.6)	96.0 (76.0)	99.9 (—)	98.3 (95.9)
Completeness (anomalous) (%)	99.1 (99.4)	70.4 (69.2)	76.8 (68.9)	96.0 (71.0)	—	95.7 (91.2)

resolution, are measured with lower accuracy, spoiling the practical advantage of these effects. The true  $\Delta F^{\pm}$  values are often of the same order as the measurement errors, leading to a seriously overestimated  $\langle \Delta F^{\pm} \rangle$ . If  $\varepsilon$  represents the measurement error,  $\Delta F_{\text{obs}}^{\pm} = \Delta F_{\text{true}}^{\pm} \pm \varepsilon$ , so  $\langle \Delta F_{\text{obs}}^{\pm} \rangle = (\Delta F_{\text{true}}^{\pm 2} + \varepsilon^2)^{1/2}$ .

The pattern of the average ratios of anomalous difference to total amplitude,  $\langle \Delta F^{\pm} \rangle / \langle F \rangle$ , for some of the test structures is shown as a function of resolution in Fig. 2.

In practice, the significance of the anomalous signal contained in the measured set of intensities can be roughly estimated at the data-merging stage. If Friedel mates are treated as equivalent, the true differences between the intensities of the Friedel-related reflections will lead to increased merging *R* factors and distorted normal probability plots compared with the results obtained when the Friedel mates are kept separate. Also, the list of potential outliers

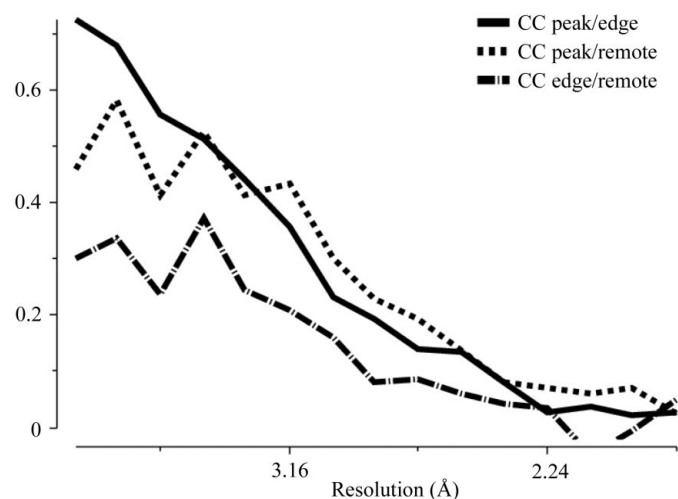
should reveal significant and consistent differences between some of the Bijvoet-related intensities.

An elegant method for assessing data quality, suggested by Schneider & Sheldrick (2002), is to verify that the anomalous signal from different sets of measurements is correlated. The data sets can either come from different wavelength measurements or the data can be arbitrarily partitioned. It is of course necessary to have some level of multiplicity; *i.e.* anomalous pairs must have been measured more than once. An illustration of this is given in Fig. 3. They found in practice that once the correlation falls below 0.25 the differences are too unreliable to be useful in placing the substructure or in estimating phases.

## 7. Test examples to assess success or failure

There is often a wide gap between theory and practice and this is especially so with SAD phasing. There are many successful applications discussed in Dauter *et al.* (2002). In most of these cases, the data were of extremely high quality. The following examples have been chosen to examine the power of the technique with more 'normal' data sets and to allow us to pinpoint the reasons for success or failure as a prerequisite for designing better protocols. In fact, like molecular replacement, applications seem often to be either trivial or impossible! The method certainly works well if data quality is excellent, if there are reasonable experimental phases extending to 2 Å, if the solvent content is greater than 50% or if the diffraction data extend to 1.5 Å or beyond.

The statistics of diffraction data and phasing for each example data set is given in Tables 1 and 2. The amount of anomalous signal in several of the data sets is illustrated in Fig. 2, where the average ratio of anomalous difference to total amplitude,  $\langle \Delta F^{\pm} \rangle / \langle F \rangle$ , is given as a function of resolution. The results reported here are for models phased using *MLPHARE* and with the density-modification steps performed with *DM*. All models were refined with *REFMAC* (Murshudov *et al.*, 1997) and the phase comparisons and map

**Figure 3**

The correlation between the anomalous signal for three data sets collected for Cmb27(i). If there was no error, the correlation would be 100%. When it falls below 25% there will be little useful phasing and below 40% it is not useful to position the substructure. In this case, only data to 3.5 Å was used to find the Se sites.

**Table 2**  
Details of SAD phasing.

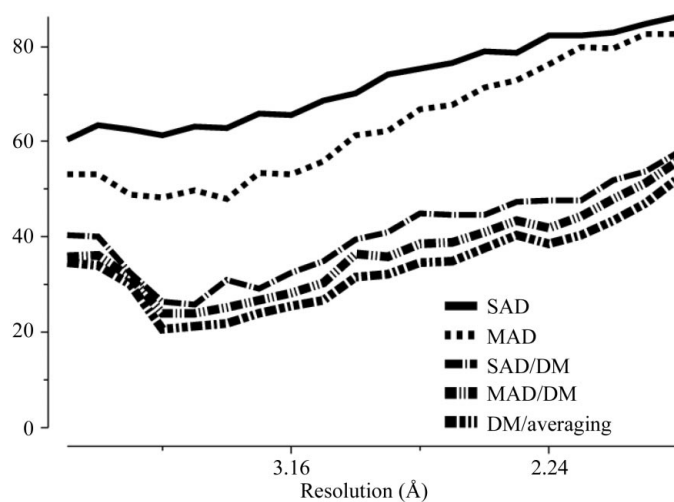
FOM, overall figure of merit after *MLPHARE* or *DM*. CC, correlation coefficient between ( $F_{\text{obs}}, \varphi_{\text{calc}}$ ) Fourier map and the map calculated after *DM*.  $\Delta\varphi$ , average difference between phases calculated from the refined model and those obtained from *MLPHARE* or *DM*.

Crystal	Cbm27(i)	ProtE(i)	ProtE(ii)	Cbm27(ii)	Lipase	Cel5A
Wavelength (Å)	0.98	1.54	0.93	0.98	0.98	1.54
Resolution (Å)	2.00	1.7	1.3	1.78	2.10 (2.78)	1.68
Protein size (kDa)	80 (2 mols)	9	9	40	45	34
Solvent fraction	0.55	0.52	0.52	0.34	0.55	0.42
Substructure	2 Se	9 S	9 S	1 Se	2 U, 4 Au	8 S, 1 Br
$f''$ (electron units)	4	0.5	0.2	4	—	0.5
$\langle \Delta F^{\pm} \rangle / \langle F \rangle$ est. (%)	~4	~1.5	—	~4	3.0	~1
FOM <i>MLPHARE</i>	0.1	—	—	0.3	0.4 (0.34)	0.03
FOM <i>DM</i>	0.80	0.55	—	0.65	0.75/0.35 (0.75/0.55)	0.7
CC <i>DM</i>	0.86	0.52	0.51 ( <i>ACORN</i> , 0.58)	0.37	0.79 (0.5)	0.25
$\Delta\varphi$ <i>MLPHARE</i> (°)	75.4	75.1	—	65.0 (2.8 Å)	68 (72)	78
$\Delta\varphi$ <i>DM</i> (°)	43.0	61.9	—	70.0 (2.8 Å)	71 (50)	78

correlations are all performed against these models. All this software is available within the *CCP4* suite (Collaborative Computational Project, Number 4, 1994). The failures, Cbm27(ii) and Cel5A, were also phased using the more sophisticated procedures coded in *SHARP* and *SOLOMON*, but without success.

#### 7.1. Success – Cbm27(i) crystal form A: anomalous signal of Se

These well diffracting crystals have two copies of the molecule in the asymmetric unit, with 55% solvent. Se-containing protein was prepared and a MAD phasing experiment was carried out at the ESRF (Boraston *et al.*, 2003). Two sites were positioned from the Patterson, with two more somewhat disordered ones positioned later from difference Fourier and the structure phased in a straightforward manner using *MLPHARE* and *DM* with and without



**Figure 4**  
The phase differences for Cbm27(i) between those calculated from the final model and those derived from different phasing procedures. The initial phase error after SAD phasing was almost  $10^\circ$  greater than that for the MAD phasing, but the density-modification errors were very similar. Averaging the density for the two molecules in the asymmetric unit gave a further improvement.

averaging. The phasing protocol was repeated using data from only one wavelength, which also gave excellent maps of comparable quality. The comparisons of the relative phase errors after the different procedures are illustrated in Fig. 4.

#### 7.2. Partial success – ProtE (E-fragment of human fibrinogen): anomalous signal of S

Excellent crystals of ProtE, a 90-residue fragment of the human fibrinogen were available, with one molecule in the asymmetric unit and 52% solvent (Brzozowski, 2003). The fragment contained nine S atoms, one methionine and four disulfide bridges and therefore seemed an ideal case for SAD phasing from the S signal alone. Highly redundant data were collected in-house to 1.7 Å. In fact, the structure was solved by molecular replacement before the experimental phasing was completed, but the phasing exercise was also successful. The sites for the four disulfides and the Met S atom were difficult to find and in somewhat special positions. We carried out many unsuccessful searches with different resolution limits and exclusion criteria. The correct set was found using carefully screened anomalous differences to 2.5 Å as input to *SHELXD*. All observations less than  $3\sigma$  were excluded and any differences greater than four times the mean value for the resolution range were omitted. (This is equivalent to only using  $E$  values of less than 4 for the direct-methods step in *SHELXD*.) It is hard to judge success or failure at this stage, except by trying the phasing from many solutions. The ‘anomalous Cullis  $R$  factor’, the ratio of the anomalous lack of closure to observed anomalous difference plotted by the *MLPHARE* program, is a useful criteria. This gives  $[\Delta F_{\text{obs}}^{\pm} - \Delta F_{\text{calc}}^{\pm}] / \Delta F_{\text{obs}}^{\pm}$  as a function of resolution. In our experience, for correct solutions it should be less than 0.65 for at least the low-resolution bins. Anisotropic refinement of the correct sites to 1.7 Å using *MLPHARE* indicated how to split the disulfides. The SAD FOM fell off rapidly after 3 Å, but provided sufficient information for the density modification to improve the phases dramatically. The overall map correlation leapt from 0.3 for the SAD phases to 0.55. This is illustrated in Fig. 5. An interesting extension of the structure-solution method was provided by *ACORN*. Synchrotron data were

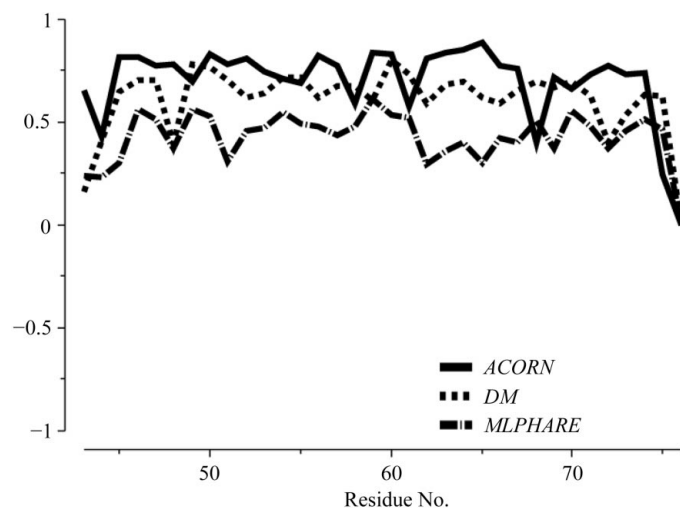
available to 1.3 Å and this was sufficient for *ACORN* to generate an excellent phase set starting from the S positions alone.

### 7.3. Total failure – Cbm27(ii) crystal form B: anomalous signal of Se

A second crystal form of Cbm27 with one molecule in the asymmetric unit and 34% solvent was obtained which diffracted better than form A. Again, Se-containing protein was prepared and a SAD phasing experiment carried out at the ESRF. The major sites were positioned from the Patterson and the structure phased as before. However, this time, with such low solvent content, the density modification did not give any significant phase improvement, the overall phase error stuck at more than 60° and it would have been difficult to interpret the resultant maps. The structure was easily solved by molecular replacement using one of the Cbm27 form A molecules as a model.

### 7.4. Partial failure – a lipase solved by isomorphous phasing

This example, provided by Jan Dohnalek (private communication), is included only to illustrate that whilst density modification is a powerful phase-improvement technique, it is much more effective if the starting phase set is of reasonable quality and extends to higher resolution. This structure has a solvent content of 55% and was solved by isomorphous replacement. Native data was available to 2.1 Å and U- and Au-derivative data were collected, at first only to 2.8 Å. The sites were easily obtained from a Patterson search, but the substitution was incomplete and the average figure of merit was 0.31. The map was not sufficiently clear to allow the tracing of an accurate solvent boundary and density-modification procedures failed to either improve the experimental phases or to extend the phase set to higher resolution. Once a second set of derivatives were prepared with more concen-



**Figure 5**  
The map correlations for ProtE for the main-chain residues of fragment B. The best agreement is for the 1.3 Å *ACORN*-generated set. The density modification after SAD phasing to 1.7 Å also gave greatly improved agreement with the final model.

trated solutions, resulting in higher substitution, and derivative data were collected to 2.1 Å, the figure of merit increased by about 0.15 in all resolution ranges. The same density-modification procedure now rapidly improved these MIRAS phases. The overall map correlation to the final model increased from 0.45 to 0.7, with very few breaks in the chain density. It demonstrated that the initial phase set must be good enough to allow the solvent boundary to be recognized.

### 7.5. Phasing failure – Cel5A complex

As a final test of the power of SAD phasing from S atoms, we investigated a complex of Cel5A (Varrot, 2000). This had been solved very easily by molecular replacement to investigate the substrate binding, but since in-house anomalous data to 1.68 Å was available and the structure contained eight S atoms and a bromine, we used it to test whether SAD phasing from these sites with such an unoptimized data set would have been possible. However, this proved to be a complete failure: the FOM from the SAD phasing was only 0.1 at 3 Å and fell to near zero at 1.6 Å. The density modification gave no improvement, demonstrating that to exploit such a weak signal, considerable care and time must be given at the data-collection stage.

## 8. Conclusions

The introduction of more accurate automatic detectors, as well as the use of crystal cryoprotection techniques, has now made it possible to collect diffraction intensities very accurately. Precise control of the wavelength of synchrotron radiation allows both anomalous and dispersive differences to be varied. At the same time, phasing software has improved dramatically. These developments have contributed to the popularity of the MAD method of phasing.

However, it has been demonstrated in many examples that it is now feasible to obtain interpretable electron-density maps from intensities containing the anomalous signal within a single data set recorded using only one X-ray wavelength (SAD). The accuracy of measurements required for successful SAD phasing seems to be comparable with that in routine MAD experiments, but in contrast to the MAD method, the wavelength need not be so finely tuned and indeed it is possible to use Cu K $\alpha$  radiation for many problems. At this wavelength, S and Br both have a detectable signal ( $f_A'' \approx 0.5$ ) and most metals have a very significant  $f_A''$ .

This flexibility in wavelength used makes the experiment much less demanding; small fluctuations in the wavelength are not disastrous since the data do not need to be recorded at the peak of the anomalous scattering. Cross-correlation of the anomalous signal to assess the resolution limit for useful differences, routinely used in MAD data sets, can be still be used by partitioning the observations randomly. This feature is already incorporated in the CCP4 program *SCALA* (Evans, 1997).

However, the quality of a phase estimate from only two measurements is inevitably limited, with a near-bimodal

probability distribution, no matter how accurate the data are, so success depends crucially on the density-modification step, which moves the probability distribution towards a unimodal form and gives phase estimates for centric reflections where the SAD experiment provides virtually no phasing information at all. These are most successful when there is limited phase information across the whole range of resolution, so in some ways the SAD (or MAD) experiment is ideal when there is no perfect isomorphism. However, if the initial phasing is poor, the solvent content is too low or the sites are in special positions (e.g. all sites in a polar space group at the same height along the rotation axis), this can fail.

It is also worth remembering that all crystallography becomes easier the higher the data resolution and in favorable cases it might prove a more effective strategy to collect one set of atomic resolution data and restrict the high-redundancy data sets required for experimental phasing techniques to a more limited resolution.

All workers in this field owe an immense debt of gratitude to Z. Dauter who demonstrated convincingly the value of SAD phasing in routine experiments. Members of the York Structural Biology Laboratory, in particular Marek Brzozowski, Jan Dohnalek, Didier Nurizzo and Annabelle Varrot, have provided both data and valuable discussions. The Wellcome Trust provided ED with funding.

## References

- Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* **D52**, 30–42.
- Blow, D. M. & Rossmann, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.
- Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. London: Academic Press.
- Boggon, T. J. & Shapiro, L. (2000). *Structure*, **8**, R143–R149.
- Boraston, A. B., Revett, T. J., Boraston, C. M., Nurizzo, D. & Davies, G. J. (2003). In the press.
- Brzozowski, M. (2003). In preparation.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cowtan, K. (1999). *Acta Cryst.* **D55**, 1555–1567.
- Dauter, Z., Dauter, M. & Dodson, E. (2002). *Acta Cryst.* **D58**, 494–506.
- Dodson, E. J. & Vijayan, M. (1971). *Acta Cryst.* **B27**, 2402–2411.
- Drenth, J. (1999). *Principles of Protein X-ray Crystallography*, 2nd ed. Heidelberg: Springer.
- Evans, P. R. (1997). *CCP4 Newslett. Protein Crystallogr.* **33**, 22–24.
- Fan, H.-F., Hao, Q., Gu, Y.-X., Qian, J.-Z., Zheng, C.-D. & Ke, H. (1990). *Acta Cryst.* **A46**, 935–939.
- Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S. & Yao, J.-X. (2000). *Acta Cryst.* **D56**, 1137–1147.
- Fourme, R., Shepard, W. & Kahn, R. (1996). *Prog. Biophys. Mol. Biol.* **64**, 167–199.
- González, A., Pedelacq, J. D., Sola, M., Gomis-Rüth, F.-X., Coll, M., Samama, J. P. & Benini, S. (1999). *Acta Cryst.* **D55**, 1449–1458.
- Hauptman, H. A. (1982). *Acta Cryst.* **A38**, 289–294.
- Hauptman, H. A. (1996). *Acta Cryst.* **A52**, 490–496.
- Hendrickson, W. A. (1979). *Acta Cryst.* **A35**, 245–247.
- Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
- Hendrickson, W. A. (1999). *J. Synchrotron. Rad.* **6**, 845–851.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
- James, R. W. (1958). *The Optical Principles of the Diffraction by X-rays*. London: Bell & Sons.
- Karle, J. (1980). *Int. J. Quant. Chem.* **7**, 357–367.
- La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
- Langs, D. A., Blessing, R. H. & Guo, D. Y. (1999). *Acta Cryst.* **A55**, 755–760.
- Mathews, B. W. (1966). *Acta Cryst.* **20**, 230–239.
- Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- North, A. C. T. (1965). *Acta Cryst.* **18**, 212–216.
- Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by Wolf, P. R., Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
- Pannu, N. S., McCoy, A. J. & Read, R. J. (2003). *Acta Cryst.* **D59**, 1801–1808.
- Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
- Phillips, J. C. & Hodgson, K. O. (1980). *Acta Cryst.* **A36**, 856–864.
- Ramachandran, G. N. & Raman, S. (1956). *Curr. Sci.* **25**, 348–351.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772–1779.
- Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.
- Terwilliger, T. C. (2000). *Acta Cryst.* **D56**, 965–972.
- Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* **D55**, 849–861.
- Varrot, A. (2000). PhD thesis. University of York, England.
- Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
- Weeks, C. M., Adams, P. D., Berendzen, J., Brunger, A. T., Dodson, E. J., Grosse-Kunstleve, R. W., Schneider, T. R., Sheldrick, G. M., Terwilliger, T. C., Turkenburg, M. & Usón, I. (2003). In the press.