research papers

SAD phasing by combination of direct methods with the *SOLVE/RESOLVE* procedure

**J. W. Wang,[a] J. R. Chen,[a]
Y. X. Gu,[a] C. D. Zheng,[a] F. Jiang,[a]
H. F. Fan,[a]* T. C. Terwilliger[b]
and Q. Hao[a,c]**

[a]Institute of Physics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China, [b]Los Alamos National Laboratory, Los Alamos, NM 87545, USA, and [c]MacCHESS, 273 Wilson Synchrotron Laboratory, Cornell University, Ithaca, NY 14853-8001, USA

Correspondence e-mail: fan@mail.iphy.ac.cn

In the initial stage of SAD phasing, the essential point is to break the intrinsic phase ambiguity. The presence of two kinds of phase information enables the discrimination of phase doublets from SAD data prior to density modification. One is from the heavy atoms (anomalous scatterers), while the other is from the direct-methods phase relationships. The former can be expressed by the Sim distribution, while the latter can be expressed by the Cochran distribution. Typically, only the Sim distribution has been used to yield initial phases for subsequent density modification. However, it has been demonstrated that using direct-methods phases based on the product of the Sim and Cochran distributions can lead to improved initial phases. In this paper, the direct-methods phasing procedure *OASIS* has been improved and combined with the *SOLVE/RESOLVE* procedure. Experimental SAD data from three known proteins with expected Bijvoet ratios $\langle|\Delta F|\rangle/\langle F \rangle$ in the range 1.4–7.0% were used as test cases. In all cases, the phases obtained using the program *RESOLVE* beginning with initial phases based on experimental phases plus Sim and direct-methods information were more accurate than those based on experimental plus Sim phase information alone.

1. Introduction

Single-wavelength anomalous diffraction (SAD) phasing has become increasingly popular in protein crystallography (Dauter *et al.*, 2002; Olczak *et al.*, 2003; Ramagopal *et al.*, 2003). Apart from model building and refinement, there are two main steps in solving protein structures from SAD data: (i) obtaining the initial phases and (ii) improving the electron-density map calculated with the initial phases. Density-modification methods (Abrahams, 1997; Bricogne, 1984, 1988; Cowtan & Main, 1993, 1996; Giacovazzo & Siliqi, 1997; Goldstein & Zhang, 1998; Lunin, 1993; Perrakis *et al.*, 1997; Podjarny *et al.*, 1987; Prince *et al.*, 1988; Refaat *et al.*, 1996; Roberts & Brünger, 1995; Rossmann & Arnold, 2001; Terwilliger, 1999, 2000, 2001, 2002, 2003; Vellieux *et al.*, 1995; Wang, 1985; Wilson & Agard, 1993; Xiang *et al.*, 1993; Zhang & Main, 1990; Zhang, 1993; Zhang *et al.*, 1997) are powerful tools for handling SAD data. However, they focus mainly on the second step. It is reasonable to expect that with better initial phases density-modification methods may give improved output electron-density maps. This paper targets the first step. Test calculations with experimental SAD data from three known proteins show that by incorporating direct methods much better initial phases can be obtained, leading to improved quality of the output maps from density-modification methods.

2. The phase ambiguity intrinsic to SAD

When anomalous scattering is present, the structure factor of a reflection with reciprocal vector **h** can be expressed as

$$F(\mathbf{h}) = \sum_{j=1}^{N} (f_j^o + f_j' + if_j'') \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j), \qquad (1)$$

where f_j^o is the normal atomic scattering factor, f_j' is the real part and f_j'' is the imaginary part of the correction to the atomic scattering factors. (1) can be rewritten as

$$F^+ = F^o + F' + F''. \qquad (2)$$

From (1) and (2), we have

$$F^*(-\mathbf{h}) \equiv F^{-*} = F^o + F' - F'', \qquad (3)$$

where $F^*(-\mathbf{h})$ is the conjugate complex of $F(-\mathbf{h})$. Subtracting (3) from (2) it follows that

$$F^+ - F^{-*} = 2F''. \qquad (4)$$

(4) defines a phase triangle. In an experiment, we can measure the magnitudes of F^+ and F^{-*}, i.e. $|F^+|$ and $|F^-|$, from which the heavy-atom (anomalous scatterer) sites can be located. F'' can be calculated using the expression

$$F''(\mathbf{h}) = i \sum_{j=1}^{N} f_j'' \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j). \qquad (5)$$

There will now be two possible ways to draw the triangle of (4), as shown in Fig. 1. This leads in an ideal case to two possible phases, one correct and the other incorrect, for a given reflection. The averaged magnitude given by (6) (see Fig. 1) is normally used in the calculation of electron-density maps,
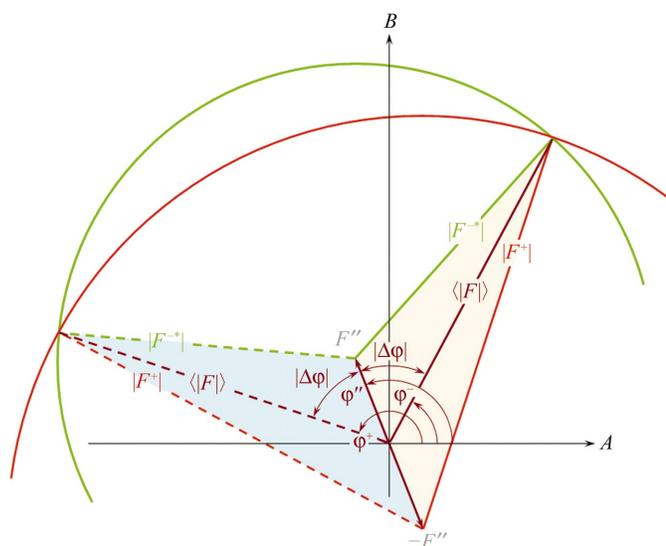
$$\langle |F| \rangle \simeq (|F^+| + |F^-|)/2. \qquad (6)$$

The two possible phases associated with $\langle F \rangle$ can be expressed as

$$\varphi = \varphi'' \pm |\Delta\varphi|, \qquad (7)$$

where φ'' is the phase of $F''(\mathbf{h})$ (see equation 5) and $|\Delta\varphi|$ is the absolute difference between φ'' and the phase associated with $\langle F \rangle$. The value of $|\Delta\varphi|$ can be calculated as (see Blundell & Johnson, 1976)

$$|\Delta\varphi| \simeq \cos^{-1}[(|F^+| - |F^-|)/2|F''|]. \qquad (8)$$

Hence, the phase ambiguity in SAD can also be considered to be the sign ambiguity of $\Delta\varphi$.

3. Phase information available in the initial stage of SAD phasing

3.1. Bimodal phase probability distribution from SAD

Taking into account the lack-of-closure error (Blow & Crick, 1959), instead of the phase doublet expressed in (7) there will be a bimodal phase-probability distribution. Given the magnitude $\langle |F| \rangle$ (6) and the structure factor contributed from the imaginary-part scattering of the heavy-atom substructure (5), the relative probability of a possible phase of $\langle F \rangle$, i.e. $\varphi_{\mathbf{h}}$, can be estimated from the agreement between the observed anomalous difference (Δ_{ANO}^o) and the anomalous difference calculated using this phase $[\Delta_{ANO}^c(\varphi_{\mathbf{h}})]$

$$p_{SAD}(\varphi_{\mathbf{h}}) \propto \exp -\{[\Delta_{ANO}^o - \Delta_{ANO}^c(\varphi_{\mathbf{h}})]^2/(E_{ANO}^2 + 2\sigma_{ANO}^2)\}, \qquad (9)$$

where σ_{ANO}^2 is the variance in the measurement of Δ_{ANO}^o and E_{ANO}^2 is the square of the lack-of-closure error (Blow & Crick, 1959).

(9) can be rewritten in an approximate form emphasizing the bimodal distribution of phase probabilities,

$$P_{SAD}(\varphi_{\mathbf{h}}) = \frac{1}{2\sigma_{\mathbf{h}}(2\pi)^{1/2}} \exp\{-[\varphi_{\mathbf{h}} - (\varphi_{\mathbf{h}}'' + |\Delta\varphi_{\mathbf{h}}|)]^2/2\sigma_{\mathbf{h}}^2\} \qquad (10)$$

$$+ \frac{1}{2\sigma_{\mathbf{h}}(2\pi)^{1/2}} \exp\{-[\varphi_{\mathbf{h}} - (\varphi_{\mathbf{h}}'' - |\Delta\varphi_{\mathbf{h}}|)]^2/2\sigma_{\mathbf{h}}^2\},$$

where

$$\sigma_{\mathbf{h}}^2 = \frac{D^2}{4|F''|^2 \sin^2 \Delta\varphi_{\mathbf{h}}}. \qquad (11)$$

Figure 1
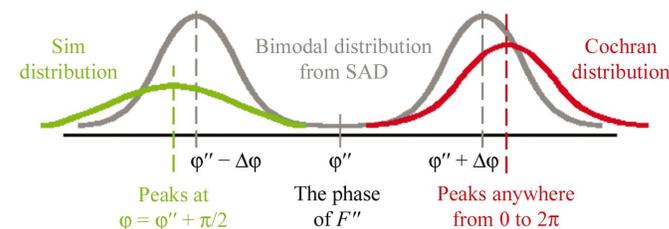The phase doublet of SAD. The green circle is centred at the end of the vector F'' with radius $|F^-|$. The red circle is centred at the end of the vector $-F''$ with radius $|F^+|$. F'' is the structure factor contributed from the imaginary-part scattering of the heavy-atom substructure. φ'' is the phase of F''. F^+ and F^- are structure factors of a Bijvoet pair. $\langle |F| \rangle \simeq (|F^+| + |F^{-*}|)/2$. The two possible phase triangles corresponding to (4) (filled in orange and cyan) lead to two possible phases, $\varphi^+ = \varphi'' + |\Delta\varphi|$ or $\varphi^- = \varphi'' - |\Delta\varphi|$ for the 'averaged' structure factor $\langle F \rangle$.

Figure 2
Schematic plot of phase-distribution curves available for initial phasing of SAD data.

In (10), $\sigma_{\mathbf{h}}^2$ is the standard deviation of the lack-of-closure error expressed as a phase difference and D is the lack-of-closure error defined by Blow & Crick (1959). As can be seen, (10) has two identical maxima symmetrically positioned at both sides of φ'' on the abscissa (see the grey curve in Fig. 2).

3.2. The Sim distribution

When the heavy-atom (anomalous scatterer) substructure is known, we have the Sim distribution (Sim, 1959),

$$P_{\text{Sim}}(\varphi_{\mathbf{h}}) = [2\pi I_0(x)]^{-1} \exp[x \cos(\varphi_{\mathbf{h}} - \varphi'_{\mathbf{h}})], \qquad (12)$$

where x is proportional to the product of the scattering contribution from the whole structure and that from the heavy-atom substructure; φ' is the phase contributed from the real-part scattering of the heavy-atom substructure. When there is only one type of anomalous scatterer, (12) becomes

$$P_{\text{Sim}}(\varphi_{\mathbf{h}}) = [2\pi I_0(x)]^{-1} \exp(-x \sin \Delta\varphi_{\mathbf{h}}), \qquad (13)$$

which is shown as the green curve in Fig. 2. (13) will still approximately hold when there is more than one type of anomalous scatterer in the structure. The Sim distribution provides discrimination between the two peaks of (10). However, the Sim distribution always peaks at $\Delta\varphi_{\mathbf{h}} \simeq -\pi/2$. Consequently, of the two identical peaks in (10), the Sim distribution will always favour that at $\varphi = \varphi'' - |\Delta\varphi|$, even though the opposite one ($\varphi = \varphi'' + |\Delta\varphi|$) may be closer to the true phase.

3.3. The Cochran distribution

Apart from the phase information based on the experimental bimodal distribution and the Sim distribution, there is additional information based on the direct-methods theory of phase relationship, which can be expressed by the Cochran distribution (Cochran, 1955),

$$P_{\text{Cochran}}(\varphi_{\mathbf{h}}) = [2\pi I_0(\kappa)]^{-1} \exp\left[\sum_{\mathbf{h}'} \kappa \cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{h}'} - \varphi_{\mathbf{h}-\mathbf{h}'})\right], \qquad (14)$$

where

$$\kappa = 2\sigma_3 \sigma_2^{-3/2} |E_{\mathbf{h}} E_{\mathbf{h}'} E_{\mathbf{h}-\mathbf{h}'}|. \qquad (15)$$

Unlike the Sim distribution, the Cochran distribution may peak anywhere from $\varphi = 0$ to $\varphi = 2\pi$. This may provide a useful discrimination of the two identical maxima produced by (10).

4. Phasing strategies in the initial stage of SAD phasing

4.1. Use of heavy-atom information

Ramachandran & Raman (1956) first proposed a method of breaking the phase ambiguity in SAD. The method makes a choice between the two possible phases by taking the one which is closer to the heavy-atom phase. This is equivalent to setting all $\Delta\varphi$ to be negative (see Figs. 1 and 2). Hendrickson & Teeter (1981) used a similar but improved procedure to solve the originally unknown protein crambin. Their method

also excludes possibilities where $\Delta\varphi$ is positive. The method of Wang (1985) directly uses phase doublets (see equation 7) as input for density modification. Recently, Pannu & Read (2004) presented the use of heavy-atom information with a new error treatment.

4.2. Use of direct methods

The use of direct methods to break the phase ambiguity in SAD was originally proposed by Fan (1965). Later, Karle (1966) reported a similar method. Hazell (1970), Sikka (1973) and Heinerman et al. (1978) tried to use direct methods in different ways. Extensive studies on direct-methods phasing of SAD data have been carried out since the 1980s. Hauptman (1982) and Giacovazzo (1983) integrated direct methods with SAD data to derive joint probability distributions for three-phase structure invariants. Fan and coworkers (Fan, Han, Qian & Yao, 1984; Fan et al., 1984; Fan & Gu, 1985) proposed a direct-methods approach based on the concept of Fan (1965), which differs from other direct-methods approaches in that (i) the probability distribution is given to individual phases rather than three-phase structure invariants, (ii) the method makes use of the known heavy-atom substructure so that the phase problem in the range of 0 to 2π is reduced to the sign problem of making a choice between $+|\Delta\varphi|$ and $-|\Delta\varphi|$ and (iii) the lack-of-closure error (Blow & Crick, 1959) in protein crystallography is introduced into the formulation of phase relationships in direct methods. The procedure has been successfully tested with experimental SAD data of a number of known proteins (Fan et al., 1990; Bing-Dong et al., 1995; Xiao-Feng et al., 1996). It has also been applied for the original solution of unknown proteins (Harvey et al., 1998; Huang et al., 2004; Chen et al., 2004). The program OASIS (Hao et al., 2000) based on the principle of Fan & Gu (1985) and the practical implementation of Fan et al. (1990) is available in the CCP4 suite (Collaborative Computational Project, Number 4, 1994) for phasing SAD and SIR (single isomorphous replacement) protein data.

4.3. The improved OASIS procedure

The direct-methods procedure used in this paper is essentially the same as that in OASIS (Hao et al., 2000), but with the following improvements: (i) figures of merit for individual reflections derived from direct methods are fitted to a uniform distribution in the range zero to unity, (ii) the lack-of-closure error D in (11) for individual reflections is derived using an algorithm similar to that used in SOLVE (Terwilliger & Berendzen, 1999) and (iii) in addition to the best phases and figures of merit, the Hendrickson–Lattman coefficients (Hendrickson & Lattman, 1970) are also output and can be used by the subsequent density modification. These improvements will be included as options in the new version of OASIS to be released in due course, presumably first on the web site http://cryst.iphy.ac.cn. Details will be described in forthcoming papers. The main points of the phasing procedure are summarized as follows.

(i) Normalized structure-factor magnitudes $|E_\mathbf{h}|$ are derived from intensity data through Wilson statistics (Wilson, 1949). The heavy-atom substructure is determined and refined using the program *SOLVE*. φ'' and $|\Delta\varphi|$ values are then calculated by (5) and (8), respectively, making use of the known heavy-atom substructure. The entire set of observed reflections are used in the direct-methods phasing procedure with the cutoff κ ($= 2\sigma_3\sigma_2^{-3/2}|E_\mathbf{h}E_{\mathbf{h}'}E_{\mathbf{h}-\mathbf{h}'}|$) set to 0.02. This usually yields more than ten million three-phase structure invariants for a protein of moderate size. About 2 h of CPU time with a Pentium IV PC are needed for this part of the calculation.

(ii) The values of $m_\mathbf{h}$ and $\Delta\varphi_{\mathbf{h}\,\text{best}}$ for individual reflections are calculated from (16) and (17), respectively (see Fan & Gu, 1985), given the starting value $P_+ = 1/2$.

$$m_\mathbf{h} = \exp(-\sigma_\mathbf{h}^2/2)\{[2(P_+ - \tfrac{1}{2})^2 + \tfrac{1}{2}](1 - \cos 2\Delta\varphi_\mathbf{h}) + \cos 2\Delta\varphi_\mathbf{h}\}^{\frac{1}{2}},$$ (16)

$$\tan(\Delta\varphi_{\mathbf{h}\,\text{best}}) = 2(P_+ - \tfrac{1}{2})\sin|\Delta\varphi_\mathbf{h}|/\cos\Delta\varphi_\mathbf{h}.$$ (17)

(iii) New values of P_+, $m_\mathbf{h}$ and $\Delta\varphi_{\mathbf{h}\,\text{best}}$ are recalculated using (18), (16) and (17), while values of $\varphi_{\mathbf{h}\,\text{best}}$ are calculated using (19) (see Fan & Gu, 1985),

$$P_+(\Delta\varphi_\mathbf{h}) = \tfrac{1}{2} + \tfrac{1}{2}\tanh\bigg\{\sin|\Delta\varphi_\mathbf{h}|\bigg[\sum_{\mathbf{h}'} m_\mathbf{h}m_{\mathbf{h}-\mathbf{h}'}\kappa$$
$$\times \sin(\Phi_3' + \Delta\varphi_{\mathbf{h}'\,\text{best}} + \Delta\varphi_{\mathbf{h}-\mathbf{h}'\,\text{best}}) + x\sin\delta_\mathbf{h}\bigg]\bigg\}$$ (18)

$$\varphi_{\mathbf{h}\,\text{best}} = \varphi''_\mathbf{h} + \Delta\varphi_{\mathbf{h}\,\text{best}}.$$ (19)

In (18), κ is from the Cochran distribution (see equation 14) and x is from the Sim distribution (see equation 12); $\Phi_3' = \varphi''_\mathbf{h} + \varphi''_{\mathbf{h}'} + \varphi''_{\mathbf{h}-\mathbf{h}'}$, where φ'' is the phase contributed from the imaginary-part scattering of the heavy-atom substructure (see equation 5); $\delta_\mathbf{h} = \varphi'_\mathbf{h} - \varphi''_\mathbf{h}$, where $\varphi'_\mathbf{h}$ is the phase contributed from the real-part scattering of the heavy-atom substructure. For centrosymmetric reflections or unpaired non-centrosymmetric reflections that are missing the Bijvoet partner, the value of $|\Delta\varphi_\mathbf{h}|$ is set to $\pi/2$. These reflections thus have a starting $m_\mathbf{h}$ equal to zero, having no effect on other reflections in the first cycle run of (18). However, other reflections can derive the sign of $\Delta\varphi_\mathbf{h}$ and a new $m_\mathbf{h}$ for such reflections through (18) and (16), respectively. From the second cycle onward, all reflections can affect the phase of others.

(iv) The P_+-modified bimodal distribution is calculated according to (20), from which the Hendrickson–Lattman coefficients A, B, C and D can be derived.

Table 1
Summary of the test data sets.

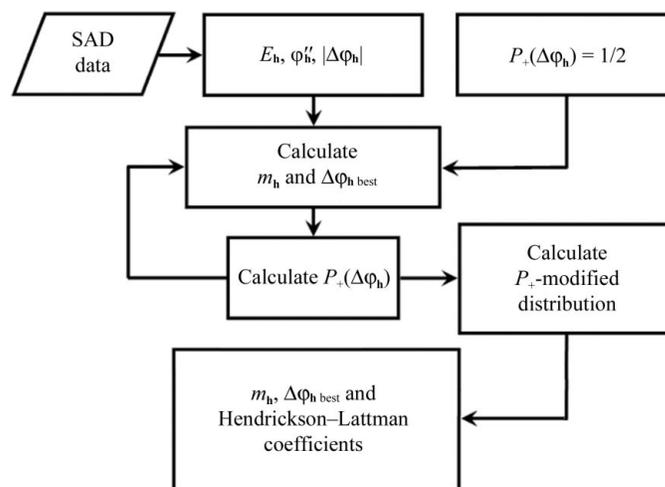| Protein	Histone methyl transferase SET 7/9	Rusticyanin	Azurin
Space group	$P2_12_12_1$	$P2_1$	$P4_122$
Unit-cell parameters (Å, °)	$a = 66.90$, $b = 83.44$, $c = 116.70$	$a = 32.43$, $b = 60.68$, $c = 38.01$, $\beta = 107.82$	$a = 52.65$, $b = 52.65$, $c = 100.63$
Residues in AU	586	154	129
Resolution (Å)	19.9–2.8	36.2–2.1	34.9–1.9
Data completeness (%)	98.5	94.0	60.3
Unique reflections	16352	7763	7097
Multiplicity	3.8	10.2	10.0
Anomalous scatterer	Se (12)	Cu (1)	Cu (1)
Symmetry of the heavy-atom substructure	Non-centric	Centric	Non-centric
Wavelength (Å)	0.9797	1.376	0.970
f'	−7.5	−5.397	0.004
f''	6.5	3.879	2.168
Bijvoet ratio (%)			
Measured	6.69	4.25	6.16
Expected	7.03	2.36	1.44

Figure 3
Flow chart of the direct-methods SAD phasing

$$P(\varphi_\mathbf{h}) = \frac{P_+(\Delta\varphi_\mathbf{h})}{2\sigma_\mathbf{h}(2\pi)^{1/2}}\exp\{-[\varphi_\mathbf{h} - (\varphi''_\mathbf{h} + |\Delta\varphi_\mathbf{h}|)]^2/2\sigma_\mathbf{h}^2\}$$
$$+ \frac{P_-(\Delta\varphi_\mathbf{h})}{2\sigma_\mathbf{h}(2\pi)^{1/2}}\exp\{-[\varphi_\mathbf{h} - (\varphi''_\mathbf{h} - |\Delta\varphi_\mathbf{h}|)]^2/2\sigma_\mathbf{h}^2\}.$$ (20)

$\sigma_\mathbf{h}$ in (20) is given by (11).

Fig. 3 shows a flow chart of the phasing procedure.

5. Tests and results

5.1. Data

SAD data from three known proteins (summarized in Table 1) were used in the test. The proteins are histone methyltransferase SET 7/9 (Wilson *et al.*, 2002), rusticyanin (Harvey *et al.*, 1998) and azurin (Dodd *et al.*, 1995). The expected Bijvoet ratios in Table 1 were calculated according to Hendrickson & Teeter (1981). For the protein histone methyltransferase SET 7/9, the experimental Bijvoet ratio is
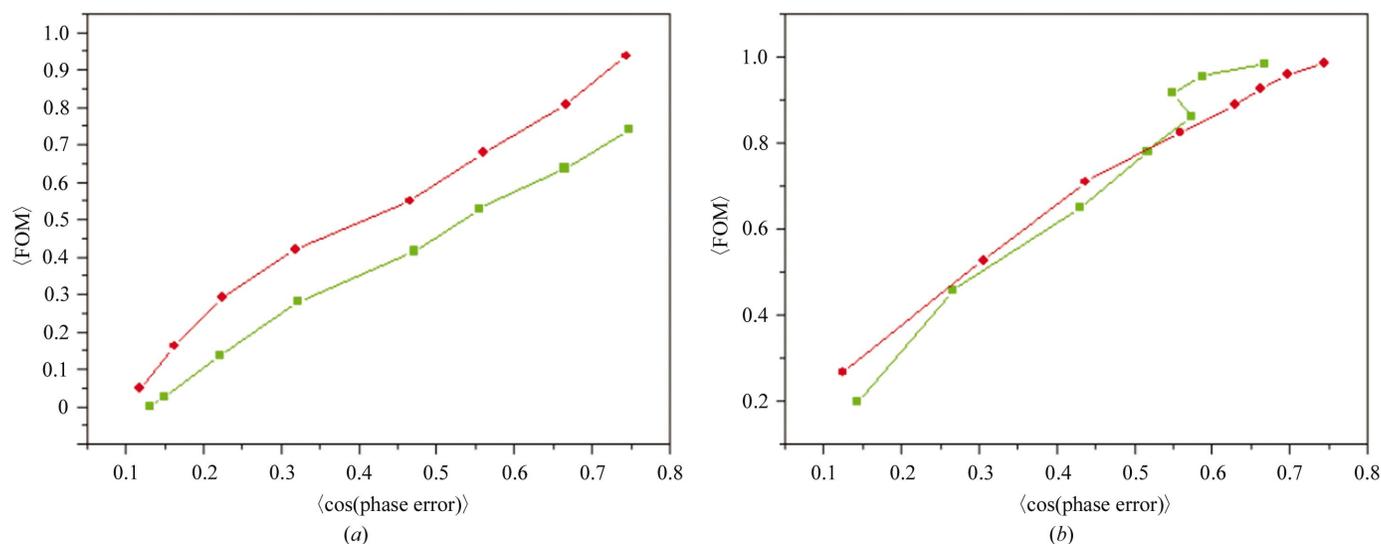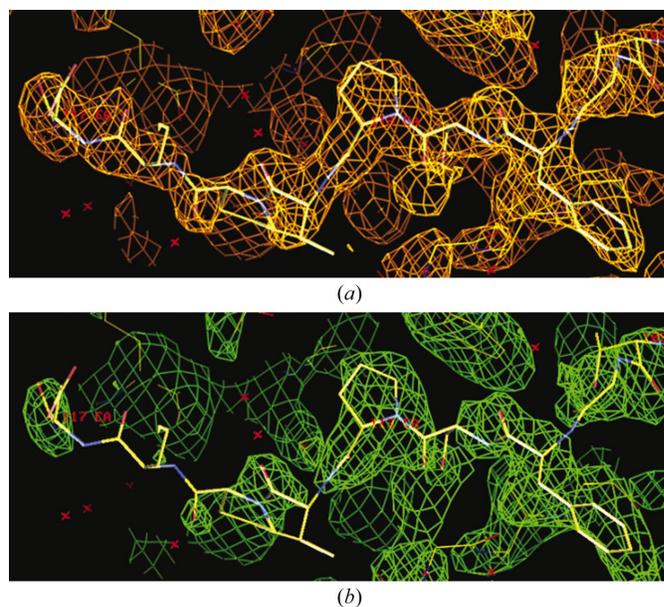
Figure 4
Plots of averaged FOM *versus* averaged cos(phase error). (*a*) Before density modification; (*b*) after density modification by *RESOLVE* (Terwilliger, 2000). Each point on the plots represents an average of 1000 reflections.

somewhat less than that expected. This may be a consequence of the fact that only ten of the 12 Se atoms can be found from the measured Bijvoet differences. For the proteins rusticyanin and azurin, the abnormally large experimental Bijvoet ratio may arise from large errors in the measured Bijvoet differences ΔF.

5.2. Comparison between results based on fitted and unfitted figures of merit

Estimates of the figure of merit of each phase are important in obtaining an optimal electron-density map. The distribution of the figures of merit (FOMs, values of $m_\mathbf{h}$ in equation 16) is difficult to predict. There are many factors affecting the distribution: the crystallographic and non-crystallographic symmetry, the heavy-atom substructure and ratio of its contribution to the whole structure, the lack-of-closure error, the implementation of direct methods *etc.* However, it is reasonable to expect that the values of $m_\mathbf{h}$ should be more or less smoothly distributed in the range between zero and unity. If there is evidence of an unreasonable distribution of $m_\mathbf{h}$, fitting them to a uniform distribution in the range zero to unity may lead to improved maps. For this purpose, the whole set of reflections are sorted in ascending order of their figure of merit. A new figure of merit $m_\mathbf{h} = n_\mathbf{h}/N$ is then assigned to each reflection, where $n_\mathbf{h}$ is the sequence number of the reflection in the sorted array and N is the total number of reflections. Test calculations showed that in most cases the fitted $m_\mathbf{h}$ led to improved maps being obtained from the subsequent density modification. The test results for the protein rusticyanin are given here. Initial phases and FOMs were calculated according to the flow chart shown in Fig. 3. Phase errors were then calculated based on the refined model. Averaged FOMs are plotted against the averaged 'cosine of phase error' in green in Fig. 4(*a*). While there is some correspondence between FOMs and their 'cosine of phase error',

values of FOM are systematically underestimated. Fitting of FOMs resulted in the red curve in Fig. 4(*a*). Fig. 4(*b*) shows the results after density modification by the program *RESOLVE* (Terwilliger, 2000). The green curve corresponds to the result based on the original direct-methods FOM, while the red curve corresponds that based on the fitted FOM. The latter appears to be more reasonable. The electron-density maps output by *RESOLVE* are shown in Fig. 5. It can be seen that the map based on fitted FOMs is better than that based on the original FOMs.

Figure 5
Electron-density maps of the protein rusticyanin contoured at 1σ after density modification by *RESOLVE* (Terwilliger, 2000): (*a*) with direct-method phases and fitted FOMs as input; (*b*) with direct-method phases and original FOMs as input.

Table 2
Phasing results for four typical reflections from histone methyltransferase SET 7/9 (Wilson *et al.*, 2002).

FOM_{P_+} denotes the direct-method FOM. FOM_{Sim} denotes the FOM from multiplying the Sim distribution by the experimental bimodal distribution. φ_{P_+} is the P_+-modified phase. φ_{Sim} is the Sim-modified phase. φ_{model} denotes phases calculated against the final structure model.

hkl	F_{obs}	φ'' (°)	$\Delta\varphi$ (°)	P_+	$\|P_+ - \tfrac{1}{2}\|$	FOM_{P_+}	φ_{P_+} (°)	φ_{model} (°)	φ_{Sim} (°)	FOM_{Sim}
4 8 3	852	209	132	0.94	0.44	0.84	345	326	34	0.49
2 9 23	794	3	80	0.18	0.32	0.38	288	263	330	0.04
8 15 2	1004	58	95	0.69	0.19	0.23	161	165	300	0.06
10 4 11	862	15	139	0.59	0.09	0.67	187	141	205	0.57

5.3. Comparison of Sim-modified phases and P_+-modified phases

Until recently, initial SAD phases were generally derived from the product of the Sim distribution and the experimental bimodal phase distribution from SAD data. Hereafter, we

Table 3
Cumulative phase errors for the entire set of SAD data from histone methyltransferase SET 7/9 (Wilson *et al.*, 2002).

Reflections were sorted in descending order of F_{obs}.

No. reflections	Error in the Sim-modified phases (°)	Error in the P_+-modified phases (°)
1500	57.1	45.8
3000	57.1	49.1
4500	56.5	50.0
6000	57.0	51.2
7500	57.8	52.9
9000	58.7	54.1
10500	59.4	55.6
12000	60.8	56.9
13500	61.9	58.4
15000	63.4	60.2
16352	65.2	62.3

regard these phases as the Sim-modified phases. In our direct method, the initial SAD phases are derived from (20). Here-
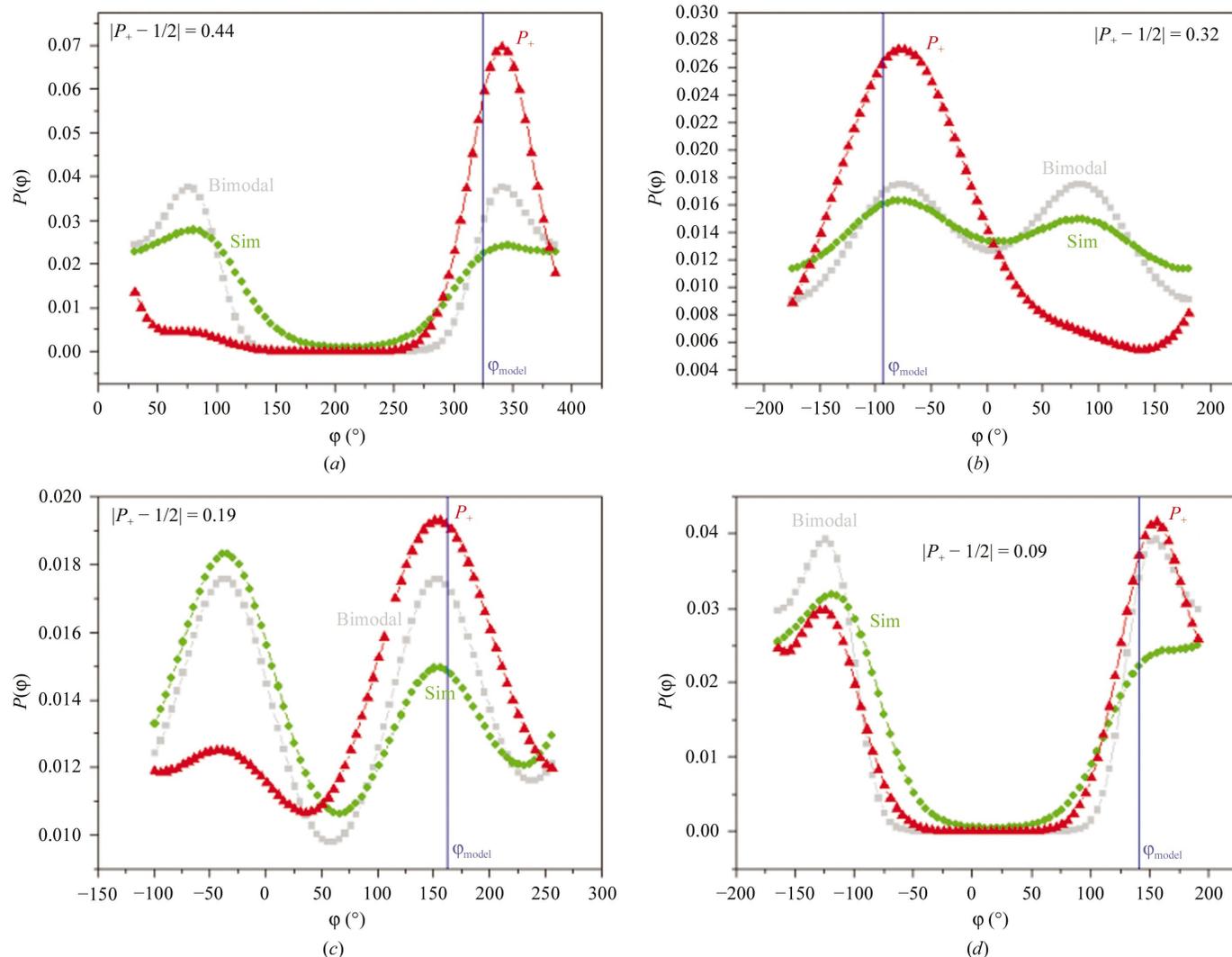

Figure 6
Phase-probability distribution curves for four reflections of the protein histone methyltransferase *SET* 7/9. (*a*) (4 8 3), (*b*) (2 9 23), (*c*) (8 15 2), (*d*) (10 4 11). Gray, experimental bimodal distribution calculated by *SOLVE* (Terwilliger, 1999); green, Sim-modified distribution calculated by *SOLVE*; red, P_+-modified distribution calculated according to (20).

after, we regard these phases as P_+-modified phases. Test calculations for comparing these two kinds of phases were performed using the data of histone methyltransferase SET7/9 (Wilson *et al.*, 2002). Phasing results of four typical reflections are listed in Table 2. They are all strong reflections with $|P_+ - \fr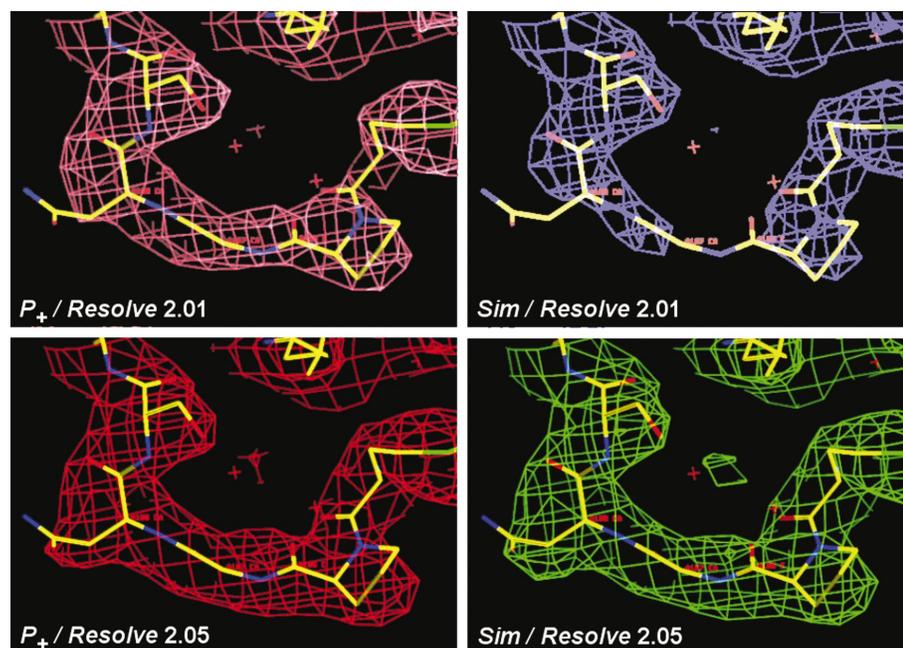ac{1}{2}|$ approximately equal to 0.4, 0.3, 0.2 and 0.1, respectively. It can be seen that the P_+-modified phases are much more accurate than Sim-modified phases. This can be seen more intuitively from the phase-probability distribution curves for the four reflections, as shown in Fig. 6. We see that the Sim-modified distribution gives no clear phase indication for any of the four reflections. Furthermore, three of the four predictions from the Sim-modified distribution are incorrect. On the other hand, the P_+-modified distribution gives a correct and strong indication of the true phase for the four reflections even when the value of $|P_+ - \frac{1}{2}|$ is as low as 0.1. Cumulative phase errors of Sim-modified phases and of P_+-modified phases are compared in Table 3 for the whole set of diffraction data. In the next section, we will see how the difference between these two kinds of phases affects the quality of resultant maps after density modification by *RESOLVE*.

Figure 7
Electron-density maps of histone methyltransferase SET 7/9 contoured at 1σ after density modification by *RESOLVE* (Terwilliger, 2000). All maps show the same portion of the unit cell. P_+/*Resolve* 2.01 means the result of *RESOLVE* v.2.01 with input of P_+-modified phases. *Sim*/*Resolve* 2.05 means the result of *RESOLVE* v.2.05 with input of Sim-modified phases.
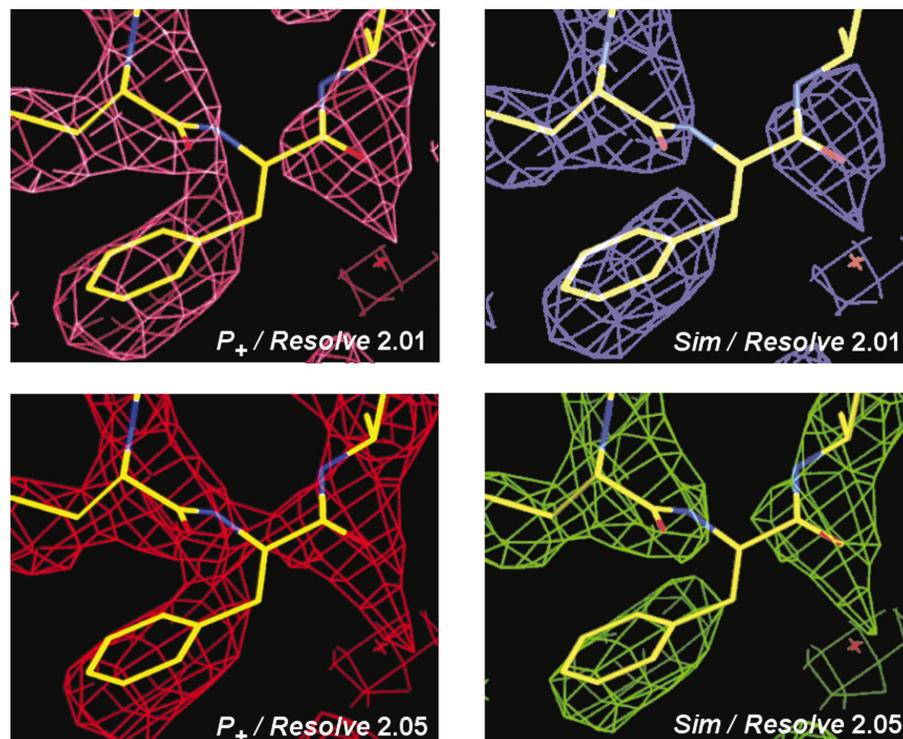
5.4. Comparison of output maps based on different initial phases and from different density-modification procedures

5.4.1. Histone methyltransferase SET 7/9. The 'peak' wavelength SAD data were used in the test. Owing to the good quality of the data and the relatively high ratio of anomalous contribution (see Table 1), both P_+-modified phases and Sim-modified phases led to a traceable electron-density map after density modification by the program *RESOLVE*. Figs. 7 and 8 show electron-density maps of two different portions of the unit cell. We see in Fig. 7 that when the program *RESOLVE* v.2.01 was used, P_+-modified phases led to a better map than did Sim-modified phases. However, when the program *RESOLVE* v.2.05 was used, the two resultant maps are both of high quality and appear to be nearly identical. *RESOLVE* v.2.05 differs from *RESOLVE* v.2.01 in the calculation of the probability that each point in the map is in the protein region, as well as in a doubling of the default weight on the map-probability phase information (T. C. Terwilliger, unpublished work). Again,

Figure 8
Electron-density maps of histone methyltransferase SET 7/9. See the caption of Fig. 7 for details.

in Fig. 8 we also see that maps from *RESOLVE* v.2.05 are better than those from *RESOLVE* v.2.01. However, here maps based on P_+-modified phases are better than those based on Sim-modified phases even after density modification by *RESOLVE* v.2.05.

5.4.2. Rusticyanin. The main differences between rusticyanin and histone methyltransferase SET 7/9 in view of SAD phasing are that (i) the expected Bijvoet ratio is much lower for rusticyanin and (ii) the anomalous scatterers in rusticyanin are centrosymmetric (see Table 1). The latter could cause problems in SAD phasing if the starting phases are Sim-modified phases. The Sim distribution in this case would shift the protein phases towards 0 or π, leading to enantiomorph ambiguity in the initial electron-density map. If, on the other hand, both possible SAD phases of each reflection are used as input, as in the method of Wang (1985), the initial electron-density map will be a superposition of the structure with its negative enantiomorph. The cancellation of positive and negative densities leads to some parts of the protein portion having an electron density that is lower than the solvent level. All these would present obstacles to finding a proper mask with density modification (solvent flattening). However, the situation is different when direct methods are used. Since $\Delta F\, (= |F^+| - |F^-|)$ is involved in the calculation of $|\Delta\varphi_\mathbf{h}|$ (see equation 8), the whole set of $|\Delta\varphi_\mathbf{h}|$ is consistent with the absolute configuration of the protein regardless of the symmetry of the heavy-atom substructure. Consequently, (18) is enantiomorph sensitive; it tends to give a set of phases corresponding to the absolute configuration of the protein, provided the calculation is accompanied by the correct configuration (centrosymmetric or non-centrosymmetric) of the heavy-atom substructure. As seen in Fig. 9, in the case of rusticyanin the resultant map from *RESOLVE* v.2.05 based on P_+-modified phases can easily be traced. However, tracing the map based on Sim-modified phases is rather difficult if not impossible.

5.4.3. Azurin. Of the three test samples, azurin has the lowest expected Bijvoet ratio and the lowest data completeness (see Table 1). For this sample, *RESOLVE* v.2.05 did not produce better maps than *RESOLVE* v.2.01. Once again, P_+-modified phases but not Sim-modified phases led to a traceable electron-density map after density modification by *RESOLVE*. Partial electron-density maps based on different initial phases are compared in Fig. 10.

6. Discussion

One of the most important factors affecting the result of SAD phasing is the Bijvoet ratio, $\langle|\Delta F|\rangle/\langle F\rangle$. This is why sulfur-SAD data are regarded as challenging data in protein diffraction phasing. Many proteins with low Bijvoet ratios have been solved *ab initio* or tested with SAD data and some of these are listed in a recent paper by Ramagopal *et al.* (2003). Those proteins with sulfur as the only or the main anomalous scatterer are listed in Table 4. The Bijvoet ratio in Table 4 ranges from 2 to 0.68%, with an average of 1.31%. As a comparison, the lowest Bijvoet ratio of our three test samples is 1.44% (azurin), which is close to the average value. Furthermore,
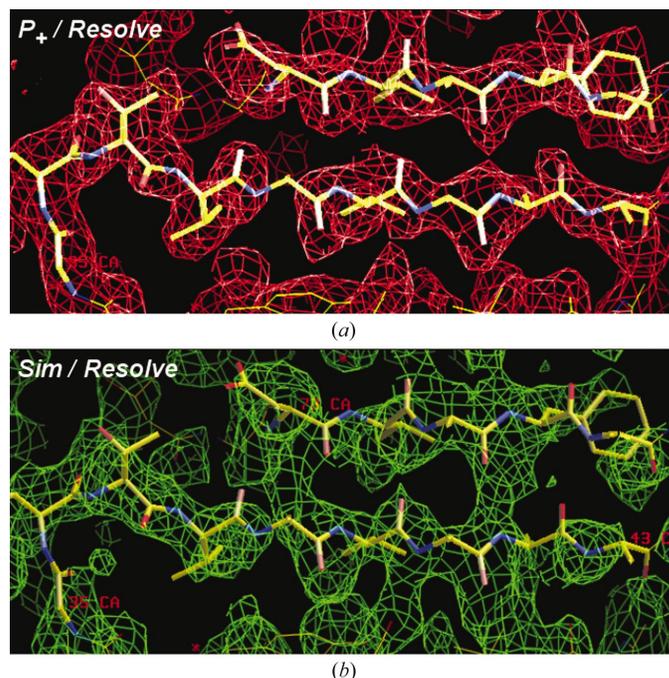

Figure 9
Electron-density maps of rusticyanin contoured at 1σ after density modification by *RESOLVE* (Terwilliger, 2000). (*a*) Result based on P_+-modified phases, (*b*) result based on Sim-modified phases.

Figure 10
Electron-density maps of the protein azurin contoured at 1σ after density modification by *RESOLVE* (Terwilliger, 2000). (*a*) Result based on P_+-modified phases, (*b*) result based on Sim-modified phases.
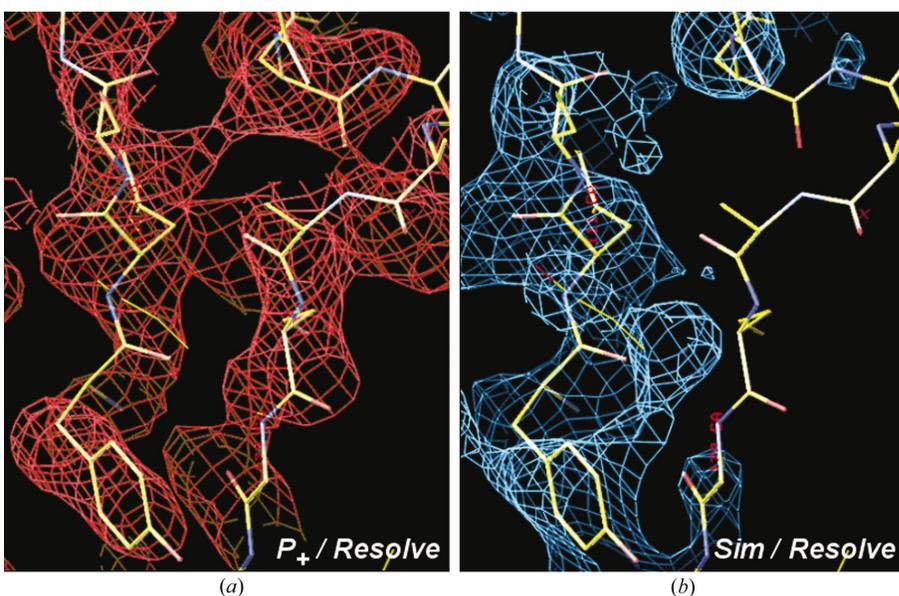
Table 4
Proteins solved by sulfur-SAD phasing.

Protein	Atoms in ASU	Scatterers in ASU	λ (Å)	f'' (e)	$\langle|\Delta F|\rangle/\langle F\rangle$ (%)	Reference
Crambin	344	6S	1.54	0.56	1.45	Hendrickson & Teeter (1981)
Lysozyme	1001	10S + 7Cl	1.54	0.56 + 0.70	1.55	Dauter et al. (1999)
Obelin	1553	8S + 1Cl	1.74	0.70 + 0.86	1.12	Liu et al. (2000)
Lysozyme	1001	12S + 7Cl	1.54	0.56 + 0.70	1.55	de Graaff et al. (2001)
Thaumatin	1551	17S	1.54	0.56	1.70	Yang & Pflugrath (2001)
Thiostrepton	105	5S	1.54	0.56	2.00	Bond et al. (2001)
Apocrustacyanin C1	2911	18S	1.77	0.72	1.00	Gordon et al. (2001)
IGF2R fragment	1006	11S	1.77	0.72	1.60	Brown et al. (2002)
EAS	3627	16S	1.54	0.56	0.86	Lemke et al. (2002)
CAP-Gly domain	813	3S + 1Cl	1.74	0.70 + 0.86	1.11	Li et al. (2002)
Tryparedoxin form I	2376	13S	1.77	0.72	1.2	Micossi et al. (2002)
Tryparedoxin form II	2360	13S	1.77	0.72	1.2	Micossi et al. (2002)
LBTI	501	14S	1.54	0.56	1.96	Debreczeni et al. (2003)
Glucose isomerase	3050	9S + 1Cl	1.54	0.56	0.68	Ramagopal et al. (2003); Wang et al. (2004)
Xylanase	2300	5S	1.74	0.70	0.69	Ramagopal et al. (2003); Wang et al. (2004)

Wang *et al.* (2004) showed that the two proteins with the lowest Bijvoet ratios in Table 4, glucose isomerase and xylanase (0.68 and 0.69%, respectively) could have been solved in a straightforward way using the improved *OASIS* procedure combined with the program *DM* (Cowtan, 1994) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) running with default parameters.

Density modification is a powerful tool in treating SAD data; direct methods, combined with standard SAD phasing methods and Sim weighting, are powerful in deriving initial SAD phases. Combination of these two can improve the quality of SAD phases, especially when the anomalous signals are weak (*e.g.* azurin) or the heavy-atom substructure is centrosymmetric (*e.g.* rusticyanin).

References

Abrahams, J. P. (1997). *Acta Cryst.* D**53**, 371–376.
Bing-Dong, S., Shen-Ping, L., Yuan-Xin, G., Hai-Fu, F., Ke, H., Jia-Xing, Y. & Woolfson, M. M. (1995). *Acta Cryst.* D**51**, 342–346.
Sikka, S. K. (1973). *Acta Cryst.* A**29**, 211–212.
Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*, p. 177. London: Academic Press.
Bond, C. S., Shaw, M. P., Alphey, M. S. & Hunter, W. N. (2001). *Acta Cryst.* D**57**, 755–758.
Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–445.
Bricogne, G. (1988). *Acta Cryst.* A**44**, 517–545.
Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). *EMBO J.* **21**, 1054–1062.
Chen, J. R., Gu, Y. X., Zheng, C. D., Jiang, F., Jiang, T., Liang, D. C. & Fan, H. F. (2004). In preparation.
Cochran, W. (1955). *Acta Cryst.* **8**, 473–478.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.
Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.

Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Dauter, Z., Dauter, M. & Dodson, E. (2002). *Acta Cryst.* D**58**, 494–506.
Debreczeni, J. E., Bunkoczi, G., Girmann, B. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 393–395.
Dodd, F., Hasnain, S. S., Abraham, Z. H., Eady, R. R. & Smith, B. E. (1995). *Acta Cryst.* D**51**, 1052–1064.
Fan, H. F. (1965). *Chin. Phys.* pp. 1429–1435.
Fan, H. F. & Gu, Y. X. (1985). *Acta Cryst.* A**41**, 280–284.
Fan, H. F., Han, F. S. & Qian, J. Z. (1984). *Acta Cryst.* A**40**, 495–498.
Fan, H. F., Han, F. S., Qian, J. Z. & Yao, J. X. (1984). *Acta Cryst.* A**40**, 489–495.
Fan, H. F., Hao, Q., Gu, Y. X., Qian, J. Z., Zheng, C. D. & Ke, H. (1990). *Acta Cryst.* A**46**, 935–939.
Giacovazzo, C. (1983). *Acta Cryst.* A**39**, 585–592.
Giacovazzo, C. & Siliqi, D. (1997). *Acta Cryst.* A**53**, 789–798.
Goldstein, A. & Zhang, K. Y. J. (1998). *Acta Cryst.* D**54**, 1230–1244.
Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst.* D**57**, 1230–1237.
Graaff, R. A. G. de, Hilge, M., van der Plaas, J. L. & Abrahams, J. P. (2001). *Acta Cryst.* D**57**, 1857–1862.
Hao, Q., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2000). *J. Appl. Cryst.* **33**, 980–981.
Harvey, I., Hao, Q., Duke, E. M. H., Ingledew, W. J. & Hasnain, S. S. (1998). *Acta Cryst.* D**54**, 629–635.
Hauptman, H. A. (1982). *Acta Cryst.* A**38**, 632–641.
Hazell, A. C. (1970). *Nature (London)*, **227**, 269.
Heinerman, J. J. L., Krabbendam, H., Kroon, J. & Spek, A. L. (1978). *Acta Cryst.* A**34**, 447–450.
Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* B**26**, 136–143.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Huang, Q. Q., Liu, Q. & Hao, Q. (2004). In preparation.
Karle, J. (1966). *Acta Cryst.* **21**, 273–276.
Lemke, C. T., Smith, G. D. & Howell, P. L. (2002). *Acta Cryst.* D**58**, 2096–2101.
Li, S., Finley, J., Liu, Z.-J., Qiu, S.-H., Chen, H., Luan, C.-H., Carson, M., Tsao, J., Johnson, D., Lin, G., Zhao, J., Thomas, W., Nagy, L. A., Sha, B., DeLucas, L. J., Wang, B.-C. & Luo, M. (2002). *J. Biol. Chem.* **277**, 48596–48601.
Liu, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
Lunin, V. Y. (1993). *Acta Cryst.* D**49**, 90–99.
Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). *Acta Cryst.* D**58**, 21–28.

Olczak, A., Cianci, M., Hao, Q., Rizkallah, P. J., Raftery, J. & Helliwell, J. R. (2003). *Acta Cryst.* A**59**, 327–334.

Pannu, N. S. & Read, R. J. (2004). *Acta Cryst.* D**60**, 22–27.

Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* D**53**, 448–455.

Podjarny, A. D., Bhat, T. N. & Zwick, M. (1987). *Annu. Rev. Biophys. Chem.* **16**, 351–373.

Prince, E., Sjolin, L. & Alenljung, R. (1988). *Acta Cryst.* A**44**, 216–222.

Ramachandran, J. N. & Raman, S. (1956). *Curr. Sci.* **25**, 348–51.

Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.

Refaat, L. S., Tate, C. & Woolfson, M. M. (1996). *Acta Cryst.* D**52**, 252–256.

Roberts, A. L. U. & Brünger, A. T. (1995). *Acta Cryst.* D**51**, 990–1002.

Rossmann, M. G. & Arnold, E. (2001). *International Tables for Crystallography*, Vol. *B*, edited by U. Shmueli, pp. 230–258. Dordrecht: Kluwer Academic Publishers.

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

Terwilliger, T. C. (1999). *Acta Cryst.* D**55**, 1863–1871.

Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.

Terwilliger, T. C. (2001). *Acta Cryst.* D**57**, 1755–1762.

Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 2082–2086.

Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 1688–1701.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Vellieux, F. M. D. A. P., Hunt, J. F., Roy, S. & Read, R. J. (1995). *J. Appl. Cryst.* **28**, 347–351.

Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.

Wang, J. W., Gu, Y. X., Zheng, C. D., Jiang, F. & Fan, H. F. (2004). *S-SAD phasing by OASIS-2004. Sixth Conference of the Asian Crystallogr. Assoc.*, 27–30 June, Hong Kong, People's Republic of China. Abstract Wang_A191.

Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

Wilson, C. & Agard, D. A. (1993). *Acta Cryst.* A**49**, 97–104.

Wilson, J. R., Jing, C., Walker, P. A., Martin, S. R., Howell, S. A., Blackburn, G. M., Gamblin, S. J. & Xiao, B. (2002). *Cell*, **111**, 105–115.

Xiang, S., Carter, C. W. Jr, Bricogne, G. & Gilmore, C. J. (1993). *Acta Cryst.* D**49**, 193–212.

Yang, C. & Pflugrath, J. W. (2001). *Acta Cryst.* D**57**, 1480–1490.

Zhang, K. Y. J. (1993). *Acta Cryst.* D**49**, 213–222.

Zhang, K. Y. J., Cowtan, K. D. & Main, P. (1997). *Methods Enzymol.* **277**, 53–64.

Zhang, K. Y. J. & Main, P. (1990). *Acta Cryst.* A**46**, 41–46.

Xiao-Feng, Z., Hai-fu, F., Hao, Q., Dodd, F. E. & Hasnain, S. S. (1996). *Acta Cryst.* D**52**, 937–941.