Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Jolly SAD

**Zbigniew Dauter,[a]* Miroslawa Dauter[b] and Eleanor Dodson[c]**

[a]Synchrotron Radiation Research Section, Macromolecular Crystallography Laboratory, NCI, Brookhaven National Laboratory, Building 725A-X9, Upton, NY 11973, USA, [b]SAIC Intramural Research Support Program, Brookhaven National Laboratory, Building 725A-X9, Upton, NY 11973, USA, and [c]Department of Chemistry, University of York, Heslington, York YO10 5DD, England

Correspondence e-mail: dauter@bnl.gov

Received 14 September 2001
Accepted 21 January 2002

Examples of phasing macromolecular crystal structures based on single-wavelength anomalous dispersion (SAD) show that this approach is more powerful and may have more general application in structural biology than was anticipated. Better data-collection facilities and cryogenic techniques, coupled with powerful programs for data processing, phasing, density modification and automatic model building, means that the SAD approach may gain wide popularity owing to its simplicity, less stringent wavelength requirements and faster data collection and phasing than the multi-wavelength (MAD) approach. It can be performed at any wavelength where anomalous scattering can be observed, in many cases using laboratory X-ray sources.

## 1. Introduction

Formally, an experimental phase for any reflection can be uniquely estimated from three measurements of the amplitudes, provided that the coordinates of a model describing the differences can be determined. The method of multiple isomorphous replacement (MIR) uses related crystals where additional atoms, usually heavy metals, are 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 was augmented by exploiting the anomalous dispersion differences between $F(h, k, l)$ and $F(-h, -k, -l)$ resulting from the resonant scattering of the heavy atoms within a derivative (single isomorphous replacement with anomalous scattering; SIRAS). Such differences were not affected by non-isomorphism, but were usually much weaker than the isomorphous differences and therefore harder to determine accurately.

When more than one derivative is available (multiple isomorphous replacement with anomalous scattering; MIRAS), the technique can also provide the correct enantiomer of the alternative heavy-atom model solutions. The discussion of the underlying theory is given in many classic texts, *e.g.* Blow & Rossmann (1961), North (1965), Mathews (1966), Dodson & Vijayan (1971), Blundell & Johnson (1976), Wang (1985), Fourme *et al.* (1996) and Drenth (1999).

The multi-wavelength anomalous dispersion (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, usually three or four 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. Ideally, all measurements should be made from a single crystal, which must therefore be robust and able to survive lengthy irradiation. This almost always requires cryogenic techniques. The anomalous scatterer for MAD may be inherently contained in the metalloprotein (*e.g.* Zn, Cu or Fe), introduced by soaking (classic heavy atoms, *e.g.* Hg, Pt, Au compounds or Br ions introduced into the ordered solvent shell) or by genetic or chemical modification, as for selenomethionine in proteins or bromouracil in DNA (Boggon & Shapiro, 2000). If conditions are favorable, the phasing power is excellent and MAD is presently the method of choice for solving new macromolecular crystal structures.

In the present era of high-throughput structure determination, new ideas are being put forward to make more efficient use of synchrotron time. It has been proposed (González *et al.*, 1999) that good phase estimates may be obtained by collecting more accurate data at fewer wavelengths. In some cases, data collected at one wavelength were 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 as advocated by B. C. Wang in his classic work (Wang, 1985). In this paper, we demonstrate that this approach of single-wavelength anomalous dispersion (SAD) coupled with increasingly powerful phasing and density-modification algorithms (de La Fortelle & Bricogne, 1997; Hauptman, 1996; Langs *et al.*, 1999; Cowtan, 1999; Terwilliger, 2000) can solve the phase problem for macromolecular structures and demonstrate that the methodology is generally useful.

## 2. Background of phase determination

This is covered in several reviews (*e.g.* Blundell & Johnson, 1976; Drenth, 1999) 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 which are able to jump from a lower to higher energy shell, 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

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

The anomalous effect increases both with the atomic number of the scatterer and as the X-ray energy (inversely proportional to the wavelength) moves to the point where it corresponds to the resonant energy for the excitation of electrons from particular orbital levels. As the X-ray energy moves below such an absorption edge, the absorption and anomalous scattering effects diminish abruptly. The absorption edges are classified according to the electron energy shells as $K$, $L_{\rm I}$, $L_{\rm II}$, $L_{\rm III}$ or $M$ edges. Moreover, for some atom types such as the lanthanides the $f''$ spectrum near $L$ edges shows additional features, often increasing significantly before dropping down abruptly beyond the edge. These white lines in $f''$ can be significantly higher than the predicted level. In most macromolecules, there are few anomalous scatterers and their anomalous dispersion generates only small differences in intensity, so the diffraction data have to be measured very accurately to allow its utilization for phasing.

When the atomic form factor is real, *i.e.* the $f''$ contribution is zero, it is clear from the structure-factor equation that $F(h, k, l)$ and $F(-h, -k, -l)$ will have the same magnitude and that $\varphi(h, k, l) = -\varphi(-h, -k, -l)$. This is known as Friedel's law. However, when the form factor contains an imaginary contribution $if''$, the reflections $F(h, k, l)$ and $F(-h, -k, -l)$ will have different intensities and their phases are no longer complementary. In the MAD technique, several data sets are measured at different wavelengths $\lambda_i$, with different values for the dispersive difference $f'$ and the anomalous difference $f''$. This results in differences between both the reflection amplitudes at the different wavelengths and between the amplitudes of the Friedel-related reflections measured at the same wavelength (Bijvoet differences). Once the positions of the anomalous scatterers are known, the protein phases for the reflections can be derived from these amplitude differences in an analogous way to the well established MIRAS approach.

The procedure has two independent stages. Firstly, the positions of 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. In principle, the positions of anomalous scatterers can be found from the Bijvoet differences for single-wavelength data (Rossmann, 1961; Mukherjee *et al.*, 1989).

Once a model of anomalous scatterers is obtained, the partial structure needs to be refined to improve its ability to predict the observed differences. Simultaneously, it is used to deduce protein phases from these differences and from the calculated anomalous model phases (Blundell & Johnson, 1976; Drenth, 1999). Formally, a phase can be determined from three observations and the appropriate models for MAD, MIR or SIRAS experiments, but the errors in both measurements and models mean that it is essential to use a probabilistic approach to obtain appropriate weights for the phase information. If there is only one partial structure model (as is the case for SAD, MAD or SIRAS experiments), the correct enantiomer can usually only be chosen by assessing which

hand for the anomalous scatterers generates the better electron-density map.

## 3. Single-wavelength phasing

It is formally not possible to evaluate the protein phases exactly if only two experimental measurements are available. This is the case when the data are restricted to one wavelength (SAD), with only the Bijvoet 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 $\varphi_A$, respectively, are error-free, there will be a twofold ambiguity in the estimation of the protein phase (Ramachandran & Raman, 1956).

Fig. 1 shows that for the SAD case with all anomalous scatterers of the same kind, the two possible phase values of the protein structure factor $\varphi_T$ are symmetrically oriented around $(\varphi_A - 90^\circ)$. The phase error for either solution is $(\varphi_T - \varphi_A + 90^\circ)$ and the figure of merit is $\cos(\varphi_T - \varphi_A + 90^\circ)$.

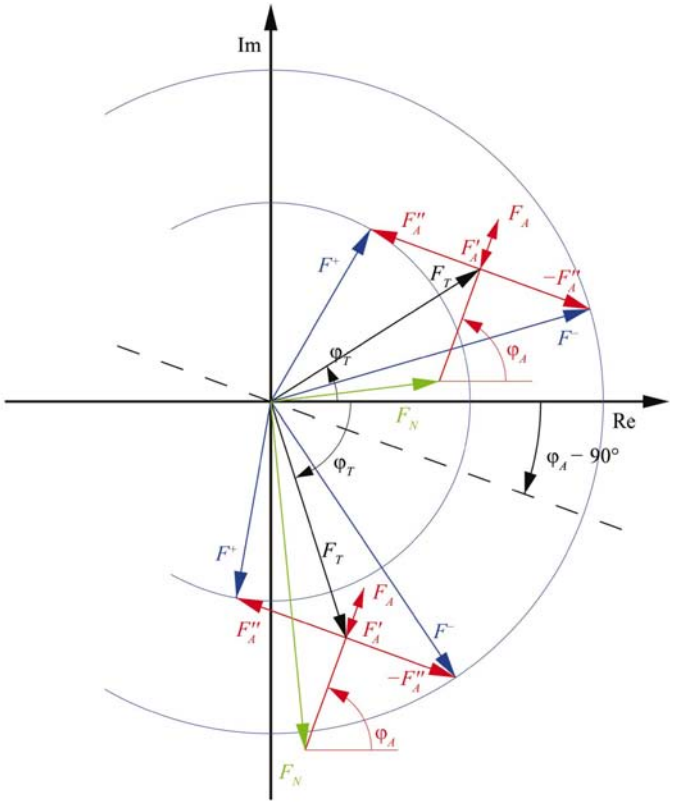


**Figure 1**
The Argand diagram showing various contributions to the scattering factors. The measured amplitudes of both Friedel mates ($F^+$ and $F^-$) are shown in blue, the total normal scattering ($F_T$) in black, the contribution of normally scattering atoms ($F_N$) in green and that of the anomalous scatterers ($F_A$, $F'_A$ and $F''_A$) in red. If the anomalous substructure has been solved, the red vectors are known in length and phase. Two solutions are then possible, giving the same Bijvoet difference $|F^+| - |F^-|$, with the total phase, $\varphi_T$, symmetrically placed on both sides of $\varphi_A - 90^\circ$. These two solutions have a different contribution of the normal scatterers, $F_N$.

Analogously, for the SIR case the two possible values of the protein phase are symmetrically oriented about the heavy-atom phase $\varphi_H$. The protein phase is only determined uniquely when the protein and anomalous phases differ by 90° and both solutions coincide. This corresponds to 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 (following Hendrickson, 1979),

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

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

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

When $\varphi_T \neq \varphi_A \pm 90^\circ$ there is an ambiguity, since

$$\sin(\varphi_T - \varphi_A) = \sin(180^\circ - \varphi_T + \varphi_A)$$

and, following Ramachandran & Raman (1956),

$$\varphi_T = \varphi_A + 90^\circ + \theta$$

or

$$\varphi_T = \varphi_A + 90^\circ - \theta,$$

where $\theta = \cos^{-1}(\Delta F^\pm / 2F''_A)$.

However, if the anomalous scatterers make up a substantial part of the structure, $\langle F_A \rangle$ is comparable in magnitude to $\langle F_T \rangle$ and $\varphi_T$ will be correlated with $\varphi_A$. In the limiting case, when the whole structure consists of identical anomalous scatterers, $\varphi_T = \varphi_A$, $|F^+| = |F^-|$ and no Bijvoet differences are observed. The probability of phase distribution resulting from anomalous scattering (Hendrickson, 1979) can be expressed as

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

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

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 $\varphi_A$. This is the basis for the proposal of Ramachandran & Raman (1956) that the value of $\varphi_T$ closer to $\varphi_A$ should be selected for initial phasing. Sim (1959) derived the statistical probability of the protein phase estimated from the known anomalous partial structure,

$$P_{\rm par}(\varphi_T) = N \exp[2(|F_T||F_A|/F_U^2)\cos(\varphi_T - \varphi_A)],$$

where $F_U^2$ is the contribution of the normally scattering (unknown) atoms.

The above relations are illustrated for few test data sets in Fig. 2, where the differences of phases calculated for the whole model and from the anomalous scatterers only are shown as a function of experimentally measured Bijvoet differences. The sinusoidal dependence between $\Delta F^\pm$ and $(\varphi_T - \varphi_A)$ shows that when the partial structure of anomalous scatterers constitutes a larger fraction of the total scattering, as for $Li_2SO_4$ ($Z_S^2 / \sum Z_i^2$ = 38%) or ferredoxin (21%), the protein

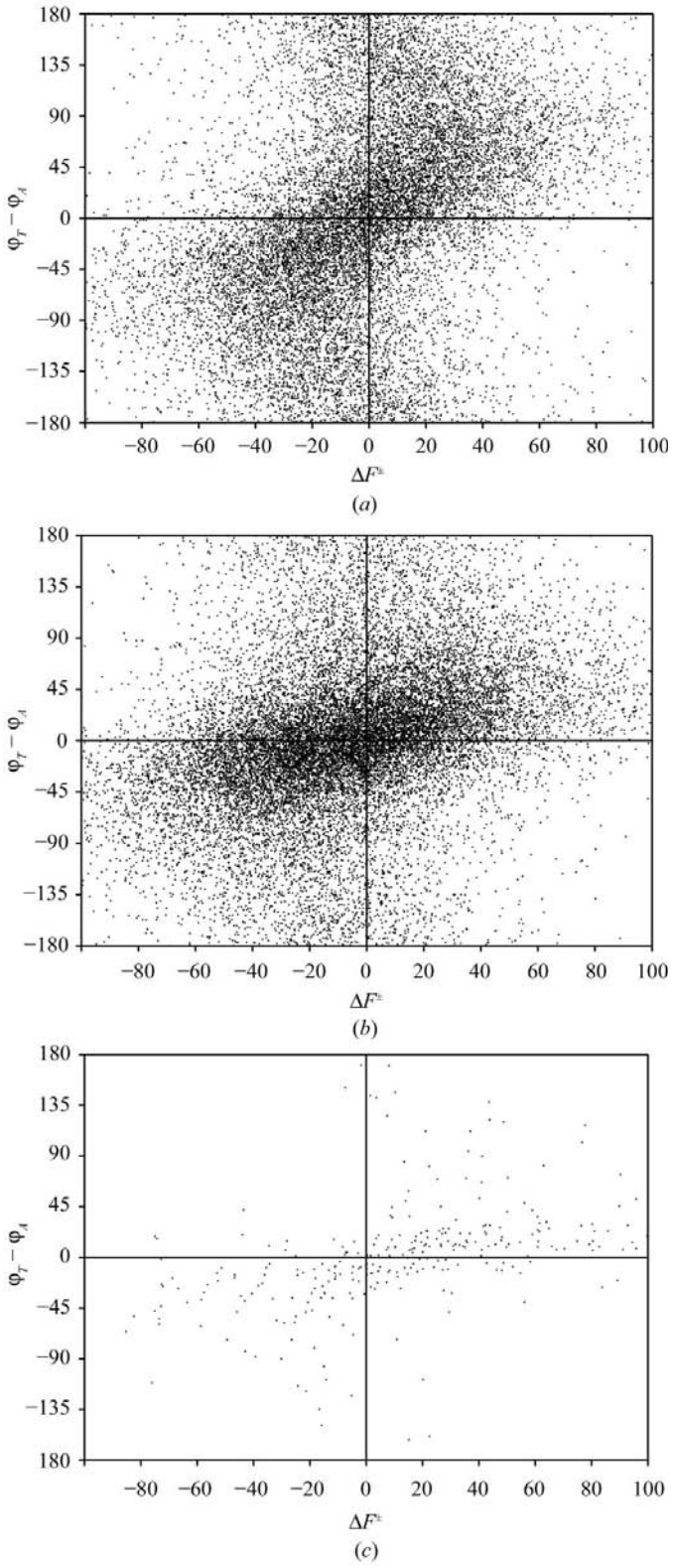


**Figure 2**
Dependence of the difference between the total protein phase and the phase of anomalous scatterers (both calculated from the model) on the measured Bijvoet difference for (*a*) lysozyme, (*b*) *Clostridium acidurici* ferredoxin with two $Fe_4S_4$ clusters, (*c*) lithium sulfate. The distribution follows the sinusoidal relationship $\Delta F^{\pm} = 2F''_A \sin(\varphi_T - \varphi_A)$ and shows that if the anomalous scatterers constitute a large part of the structure, the total protein phase $\varphi_T$ tends to be closer to the anomalous phase $\varphi_A$ of the two possible values $\varphi_T = \varphi_A + 90^\circ \pm \cos^{-1}(\Delta F^{\pm}/2F''_A)$.

phases ($\varphi_T$) tend to be closer to the anomalous scatterers phase ($\varphi_A$) than when they amount to only a small fraction, as for lysozyme (9%).

Several approaches for breaking the phase ambiguity have been used for solving crystal structures by SAD. Earlier methods included the following.

(i) Resolved anomalous phasing, as used originally by Hendrickson & Teeter (1981) for the solution of of crambin. One of the alternative phases is selected according to the combination of probabilities resulting from the partial structure, $P_{\rm par}$, and from anomalous scattering, $P_{\rm anom}$, except when there is a strong unimodal distribution of $P_{\rm anom}$ for the largest Bijvoet differences.

(ii) Iterative single-wavelength anomalous scattering (ISAS) approach, introduced by Wang (1985). This method utilizes the 'noise-filtering' procedure in direct space by iteratively smoothing the electron density within the solvent region in the macromolecular crystal, which enhances the meaningful features within the protein region until they become interpretable.

(iii) Direct-methods applications as proposed by Hauptman (1982, 1996) or by Fan *et al.* (1990), *e.g.* implemented in the program *OASIS* (Hao *et al.*, 2000). In this approach, the initial phases estimated from the anomalous differences are refined on the basis of probabilistic relations between large normalized structure factors.

The modern approach uses carefully weighted probabilistic methods to determine the initial phases and their reliability. Maximum likelihood-based phasing is programmed in *SHARP* (de La Fortelle & Bricogne, 1997), taking into account various effects such as intensity measurement uncertainty. *MLPHARE* (Otwinowski, 1991) integrates the phase probabilities around the phase circle. These phase distributions are improved by density-modification procedures such as those programmed in *DM* (Cowtan, 1999), *SOLOMON* (Abrahams & Leslie, 1996) and *RESOLVE* (Terwilliger, 2000). Providing the phase errors are properly estimated, this is an extremely powerful tool. There is no difference in the formal error estimate for the phase probabilities derived from the two enantiomorphs, but the maps differ in quality and only one will show interpretable structural features.

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

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

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

This formula is analogous to that given by Hendrickson & Teeter (1981),

$$\langle \Delta F^{\pm} \rangle / \langle F \rangle = 2^{1/2} (N_a^{1/2} f''_A) / (N_P^{1/2} Z_{\rm eff}).$$

Here, $f_{\rm eff}(\theta)$ is the effective scattering by the average atom in the whole structure at the diffraction angle $\theta$, instead of the value at zero angle, equal to the average atomic number $Z_{\rm eff}$. The anomalous scattering signal $f''$ does not depend on the

**Table 1**
X-ray data statistics.

Values in parentheses are for the highest resolution shell.

| Crystal | Space group | Unit-cell parameters | | | | Wavelength (Å) | Resolution (Å) | Measured reflections | Unique reflections | Multiplicity | $R_{merge}$† (%) | $R_{anom}$‡ (%) | $I/\sigma(I)$ | Completeness (%) | Anomalous completeness (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $a$ (Å) | $b$ (Å) | $c$ (Å) | $\beta$ (°) | | | | | | | | | | |
| $Li_2SO_4.xH_2O$ | $P2_1$ | 5.45 | 4.87 | 8.17 | 107.3 | 0.98 | 0.82 | 3282 | 425 | 7.7 | 3.5 (3.7) | 2.0 (3.2) | 38.2 (21.3) | 89.9 (86.0) | 89.6 (85.7) |
| Lysozyme | $P4_32_12$ | 71.81 | 71.81 | 36.80 | | 1.54 | 1.53 | 403234 | 17964 | 22.4 | 4.3 (17.9) | 1.4 (5.7) | 73.0 (10.3) | 100.0 (99.9) | 99.9 (99.9) |
| DNA | $P2_12_12_1$ | 17.81 | 31.42 | 44.00 | | 1.54 | 1.5 | 80781 | 4272 | 18.7 | 1.7 (6.0) | 1.4 (2.5) | 66.9 (23.1) | 100.0 (100.0) | 99.9 (100.0) |
| 2Zn insulin | $R3$ | 80.92 | 80.92 | 33.50 | | 0.93 | 1.0 | 150461 | 46849 | 3.2 | 4.2 (32.5) | 3.0 (26.2) | 28.2 (2.3) | 99.3 (98.9) | 70.4 (69.2) |
| Fe rubredoxin | $R3$ | 64.04 | 64.04 | 32.51 | | 0.92 | 1.1 | 64115 | 20052 | 3.2 | 4.7 (26.2) | 3.1 (17.2) | 23.6 (3.2) | 99.9 (89.6) | 76.8 (68.9) |
| Zn rubredoxin | $R3$ | 64.10 | 64.10 | 33.05 | | 0.93 | 1.2 | 45983 | 15659 | 2.9 | 5.9 (26.0) | 4.0 (20.7) | 19.0 (3.6) | 98.9 (99.0) | 77.0 (75.7) |
| CauFd | $P4_32_12$ | 33.95 | 33.95 | 74.82 | | 0.88 | 0.94 | 109090 | 28087 | 3.9 | 7.3 (48.3) | 5.0 (21.6) | 15.1 (2.1) | 98.7 (93.7) | 83.5 (81.1) |
| Ca subtilisin | $P2_12_12_1$ | 75.06 | 47.37 | 60.94 | | 1.54 | 1.75 | 214936 | 22601 | 9.5 | 2.9 (4.5) | 1.5 (2.3) | 68.3 (59.5) | 99.5 (99.5) | 99.0 (96.2) |
| Lu subtilisin | $P2_12_12_1$ | 75.00 | 47.07 | 61.22 | | 1.34 | 1.75 | 123792 | 20897 | 5.9 | 4.4 (21.5) | 6.5 (15.8) | 19.7 (3.7) | 92.0 (60.7) | 92.8 (65.2) |
| Glucose isomerase | $I222$ | 92.76 | 97.99 | 102.69 | | 1.34 | 1.5 | 581270 | 74887 | 7.8 | 2.6 (3.7) | 1.3 (2.6) | 41.5 (22.4) | 99.6 (97.4) | 99.2 (96.8) |
| dUTPase | $R3$ | 86.61 | 86.61 | 62.27 | | 0.99 | 2.0 | 27381 | 11737 | 2.3 | 5.9 (24.1) | 4.7 (11.4) | 11.3 (2.8) | 94.1 (89.6) | 85.1 (79.2) |
| Decaplanin | $P6_122$ | 60.5 | 60.5 | 133.1 | | 1.54 | 1.6 | 367077 | 19777 | 18.6 | 5.4 (15.3) | 1.6 (5.6) | 43.1 (10.3) | 100.0 (100.0) | 100.0 (100.0) |
| PSCP | $P6_2$ | 97.2 | 97.2 | 83.4 | | 0.92 | 1.8 | 228827 | 41341 | 5.5 | 4.1 (32.7) | 4.1 (19.3) | 23.4 (3.2) | 98.9 (96.7) | 94.9 (98.2) |
| APT | $P2_1$ | 39.59 | 127.89 | 39.66 | 102.8 | 0.92 | 1.8 | 129461 | 35250 | 3.7 | 2.8 (10.0) | 4.0 (9.4) | 22.1 (6.3) | 99.9 (98.9) | 99.6 (98.6) |
| MRD | $C2$ | 44.66 | 61.19 | 43.94 | 106.5 | 0.98 | 1.3 | 119002 | 50803 | 2.3 | 3.8 (26.4) | 6.0 (23.1) | 25.8 (2.7) | 99.8 (98.8) | 92.5 (88.9) |

† $R_{merge} = \sum_h \sum_i |I_i - \langle I \rangle| / \sum_h |I_i|$. ‡ $R_{anom} = \sum_h (|I^+| - |I^-|) / \sum_h [(|I^+| + |I^-|)/2]$.

resolution, but $f_{eff} = (1/N)\sum f_i$ reduces with resolution and thus the percentage of anomalous signal could be expected to increase at high resolution. In addition, this effect may be enhanced if the temperature factors of the anomalous scatterers are lower than the average value for all atoms of the macromolecule. However, the weak intensities at high 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 seriously overestimated $\langle \Delta F^{\pm} \rangle$. If $\varepsilon$ represents the measurement error, $\Delta F^{\pm}_{obs} = \Delta F^{\pm}_{true} \pm \varepsilon$ and $\langle \Delta F^{\pm}_{obs} \rangle = [(\Delta F^{\pm}_{true})^2 + \varepsilon^2 \pm 2\varepsilon \Delta F^{\pm}_{true}]$.

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 seperate. Also, the list of potential outliers should reveal significant and consistent differences between some of the Bijvoet-related intensities.

## 5. Test examples on previously known structures

We present here the results of phasing of various diffraction data collected from crystals of known structures and solved again by the SAD approach. The statistics of diffraction data and phasing for each data set are given in Tables 1 and 2, respectively. The amount of anomalous signal in each data set is illustrated in Fig. 4, 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. All models were refined with *REFMAC* (Murshudov *et al.*, 1997) and *wARP* (Perrakis *et al.*, 1999) or, if data extended to atomic resolution (beyond 1.2 Å), with *SHELXL* (Sheldrick & Schneider, 1997).

### 5.1. A small structure with very weak anomalous signal: lithium sulfate monohydrate

This salt crystallizes in space group $P2_1$, with a unit-cell volume of only 207 Å$^3$. Diffraction data from a crystal of $Li_2SO_4.xH_2O$ were collected at the NSLS synchrotron beamline X9B with the ADSC Quantum-4 CCD detector. The wavelength used was 0.98 Å, where sulfur has very small values of $f' = 0.183$ and $f'' = 0.234$ electrons as calculated by *CROSSEC* (Cromer, 1983). The expected value of $\langle \Delta F^{\pm} \rangle / \langle F \rangle$ is about 3% and the observed ratio is 2.85%. Two exposure passes were recorded and 3282 intensity measurements were merged to give 99 centric and 326 acentric reflections (for which Friedel mates were kept separate) to a resolution of 0.80 Å, corresponding to a complete copper-radiation sphere of reflections. $R_{merge}$ was 3.5% and $R_{anom}$ was 2.0%.

The anomalous difference Patterson synthesis (Fig. 3*a*) clearly showed the position of sulfur. Starting from the single atom of sulfur, a few cycles of refinement and difference Fourier synthesis revealed the whole structure, including both water H atoms. In fact, the $F_{obs},\alpha(S)_{calc}$ map with phases calculated solely from sulfur showed the complete image of the structure superimposed on its centrosymmetric representation. The subsequent refinement was performed against $F^2$ with Friedel mates treated as separate reflections. It converged with a final value of $R1 = 2.37\%$ and the Flack parameter $x = -0.15$ ($\pm 0.12$). The inverted enantiomer gave an $R1$ of 2.73% and $x = 1.14$ ($\pm 0.14$). The Flack absolute structure parameter $x$ should refine to 0.0 for correct chirality and to 1.0 for inverted chirality (Flack, 1983).

**Table 2**
Details of SAD phasing.

| Crystal | Wavelength (Å) | Resolution (Å) | Protein size (kDa) | Anomalous scatterers | $f''$ (electron units) | $\langle\Delta F^{\pm}\rangle/\langle F\rangle$ estimated (%) | FOM† *SHARP* | FOM† *DM* | CC‡ | $\Delta\varphi$§ *SHARP* (°) | $\Delta\varphi$§ *DM* (°) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $Li_2xSO_4.H_2O$ | 0.98 | 0.82 | 0.07 | 1 S | 0.23 | 3.0 | — | — | — | — | — |
| Lysozyme | 1.54 | 1.55 | 13.3 | 10 S + 7 Cl | 0.56 + 0.70 | 1.5 | 0.40 | 0.89 | 0.79 | 58.3 | 41.9 |
| DNA | 1.54 | 1.5 | 3.4 | 10 P | 0.43 | 2.0 | 0.52 | 0.60 | 0.84 | 47.9 | 37.9 |
| 2Zn insulin | 0.93 | 1.0 | 10.8 | 2 Zn | 2.3 | 2.3 | 0.22 | 0.81 | 0.73 | 71.3 | 65.9 |
| Fe rubredoxin | 0.92 | 1.1 | 5.6 | 1 Fe | 1.35 | 1.4 | 0.32 | 0.81 | 0.77 | 65.9 | 63.6 |
| Zn rubredoxin | 0.93 | 1.2 | 5.6 | 1 Zn | 2.3 | 2.3 | 0.47 | 0.83 | 0.86 | 58.9 | 59.8 |
| CauFd | 0.88 | 0.94 | 5.8 | 8 Fe | 1.25 | 3.7 | 0.26 | 0.54 | 0.70 | 49.0 | 67.4 |
| Ca subtilisin | 1.54 | 1.75 | 25.0 | 3 Ca | 1.28 | 1.1 | 0.34 | 0.77 | 0.70 | 66.4 | 54.7 |
| Lu subtilisin | 1.34 | 1.75 | 25.0 | 4 Lu | ∼10.0 | | 0.46 | 0.75 | 0.51 | 63.4 | 49.5 |
| Glucose isomerase | 1.34 | 1.5 | 40.9 | 1 Mn | 2.23 | 0.9 | 0.15 | 0.78 | 0.78 | 65.7 | 41.9 |
| dUTPase | 0.99 | 2.0 | 13.6 | 1 Hg | 9.5 | 6.6 | 0.38 | 0.53 | 0.54 | 75.6 | 54.2 |
| Decaplanin | 1.54 | 1.6 | 6.1 | 8 Cl | 0.70 | 2.0 | 0.39 | 0.89 | 0.84 | 59.7 | 33.2 |
| PSCP | 0.92 | 1.8 | 36.4 | 9 Br | ∼5.0 | ? | 0.21 | 0.74 | 0.62 | 77.3 | 58.1 |
| APT | 0.92 | 1.8 | 56.0 | 22 Br | ∼5.0 | ? | 0.40 | 0.85 | 0.76 | 64.8 | 48.7 |
| MRD | 0.98 | 1.3 | 13.3 | 3 Se | ∼5.0 | 5.2 | 0.57 | 0.69 | 0.82 | 54.7 | 46.2 |

† Overall figure of merit after *SHARP* or *DM*. ‡ Correlation coefficient between ($F_{obs}$, $\varphi_{calc}$) Fourier map and the map calculated after *DM*. § Average difference between phases calculated from the refined model and those obtained from *SHARP* or *DM*.

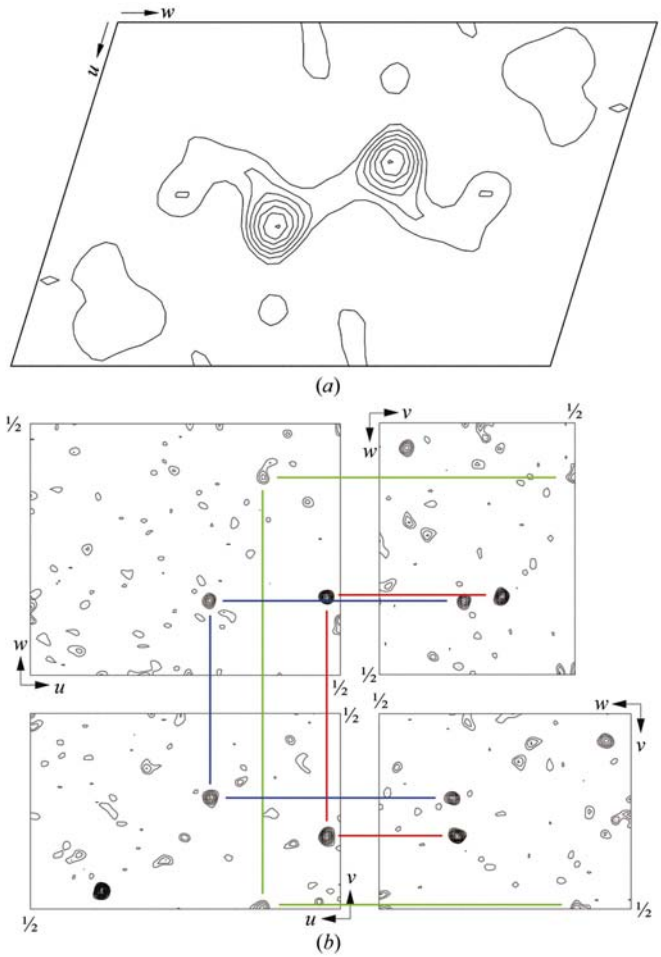


**Figure 3**
(*a*) Harker section $v = \frac{1}{2}$ of the anomalous difference Patterson for $Li_2SO_4.xH_2O$; (*b*) interpretation of three calcium sites in the Patterson of Ca subtilisin.

This example shows that even very small accurately measured anomalous scattering contribution can be quite meaningful. In this case, the anomalous signal was used to locate the S atom and to distinguish between two alternative enantiomers after refinement. The single atom of sulfur provides 43% of the total scattering of this crystal.

### 5.2. Anomalous signal of sulfurs and chlorines: lysozyme

The results of single-wavelength phasing of the tetragonal HEW lysozyme have been published previously (Dauter *et al.*, 1999). X-ray data to 1.55 Å resolution were collected from a native crystal grown from 'standard' conditions containing 1 *M* of NaCl, using a wavelength of 1.54 Å. *SHELXD* (Sheldrick, 1998) using the observed anomalous differences as input data found 17 sites corresponding to all ten S atoms in the protein and to seven chloride ions located within the ordered solvent shell around the lysozyme molecule. Phasing using the program *SHARP* followed by density modification using the program *DM* produced a clear electron-density map of the protein with a correlation coefficient with the $F_{obs}$ map of 0.79.

These data were collected in a standard way, with the crystal in arbitrary orientation, without employing inverse-beam or mirror strategies. However, the four data-collection passes with different exposure times led to a high redundancy of measurements, which contributed to the accuracy of the estimated intensities and anomalous differences. The expected amount of anomalous signal was 1.5% and indeed the calculated $\langle\Delta F^{\pm}\rangle/\langle F\rangle$ approached this value in the lower resolution shells (Fig. 4*b*).

The presence of chloride ions positioned around the surface of the lysozyme molecule held only by hydrophobic van der Waals or polar hydrogen-bond interactions suggested the possibility of using the anomalous signal of heavier halides, bromides or iodides, soaked into protein crystals for phasing (Dauter & Dauter, 2001).

### 5.3. DNA oligomer phased on phosphorus

Diffraction data from the crystal of $d(CGCGCG)_2$ DNA hexamer duplex in the Z-form were collected using a wavelength of 1.54 Å to 1.5 Å resolution in three passes with different exposures (Dauter & Adamiak, 2001). The orthorhombic crystal in space group $P2_12_12_1$ is densely packed, with about 25% solvent content. It contains 12 nucleosides connected by ten phosphates, one molecule of spermine and one magnesium ion in the asymmetric unit. Only P atoms

display a significant anomalous signal at this wavelength, with an $f''$ value of 0.43 electron units.

The diffraction data were collected with high redundancy and quality. The expected amount of anomalous signal is about 2%, which is a higher value than that for the sulfur signal in crambin or lysozyme (about 1.5%). Such signal can be expected for all DNA or RNA structures since, in contrast to the variable sulfur content in proteins, there is always one phosphate per nucleotide. The $\langle\Delta F^{\pm}\rangle/\langle F\rangle$ ratio for measured data is shown in Fig. 4(*b*).

The direct-methods programs *SHELXS* (Sheldrick, 1986), *SHELXD* or *SnB* (Weeks & Miller, 1999) run against $\Delta F^{\pm}$ data all gave clear solutions for the positions of anomalous scatterers. Phasing by *SHARP* and *DM* resulted in a very clear electron-density map unambiguously showing the complete molecule. More details of the phasing of d(CGCGCG)$_2$ against data processed with different redundancies are given elsewhere (Dauter & Adamiak, 2001).

### 5.4. Small protein with weak anomalous signal: 2Zn insulin

The results obtained on the 2Zn insulin with 102 amino acids in the asymmetric unit are somewhat unexpected and strongly underline the role of density modification in obtaining a good quality electron-density map from rather poor initial phases, especially at high resolution. 1.0 Å diffraction data from a rhombohedral 2Zn insulin crystal have been collected in order to carry out atomic resolution refinement. The conditions were optimized for recording high-resolution reflections, not the anomalous dispersion signal, so that although the completeness of unique reflections was 99%, only 72% (31 475 out of 43 476) had both Friedel mates measured. There are no centric reflections in this space group (*R*3). The wavelength of synchrotron radiation used, 0.927 Å, was distant from the absorption edge of zinc (1.28 Å) and $f''$ is 2.26 electrons. The expected $\langle\Delta F^{\pm}\rangle/\langle F\rangle$ ratio for two Zn atoms per about 2400 atoms in the protein is 1.4%. The $\langle\Delta F^{\pm}\rangle/\langle F\rangle$ ratio calculated from the data as a function of resolution is shown in Fig. 4(*a*). It is much higher than the expected value; a result of the inaccurate estimates of Bijvoet differences.

The positions of the two Zn atoms were easily obtained from the anomalous difference Patterson. They lie on the threefold axis, about 16 Å away from each other. This constellation of two atoms in *R*3 is centrosymmetric, with the center of symmetry placed on the threefold axis at the midpoint between two zinc sites. *SHARP* luckily chose the proper handedness by refining the temperature factors of the two Zn atoms to slightly different values, $B$(Zn1) = 2.607 and $B$(Zn2) = 2.619 Å$^2$, consistent with the correct enantiomer. The initial phasing gave an overall figure of merit (FOM) of

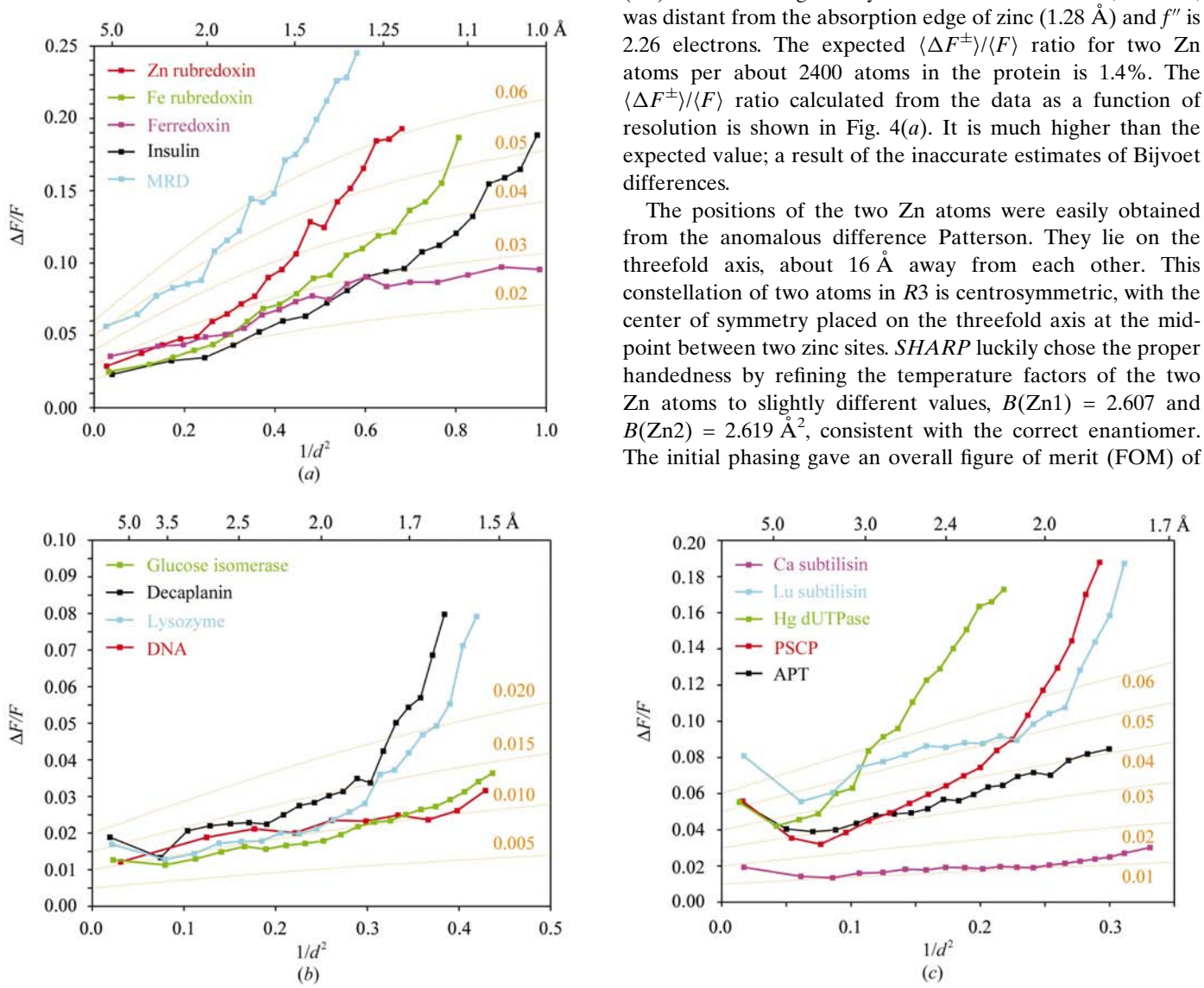


**Figure 4**
The $\langle\Delta F^{\pm}\rangle/\langle F\rangle$ ratio as a function of resolution for all data quoted in §5 and §6. (*a*) Atomic resolution data sets, extending beyond 1.3 Å; (*b*) data extending beyond 1.7 Å resolution; (*c*) data sets not exceeding 1.7 Å. The theoretically predicted curves taking into account the atomic scattering factors diminishing with resolution are drawn in brown.

0.23. The subsequent density modification by *DM* (in OMIT PERTURB mode) decisively broke the centrosymmetry and resulted in a phase set with an FOM = 0.81 and an interpretable map (Fig. 5*a*).

### 5.5. Fe and Zn rhombohedral rubredoxins

Two data sets on *R*3 *Clostridium pasteurianum* rubredoxin were collected for refinement using synchrotron radiation of 0.92 Å wavelength (Dauter *et al.*, 1996); the first to 1.1 Å resolution from a crystal containing one $FeS_4$ cluster and 55 amino-acid residues and the second from the Zn-substituted variant to 1.2 Å resolution. In both data sets only 77% of all reflections had both Friedel mates measured. At this wavelength the anomalous $f''$ contribution of Fe is 1.35 and that of Zn is 2.23 electrons. The theoretical amount of anomalous signal $\langle\Delta F^{\pm}\rangle/\langle F\rangle$ is 1.41% for Fe-substituted and 2.32% for Zn-substituted proteins, whereas the data show much higher values, 5.4 and 6.9% on average, suggesting that the Bijvoet differences contain significant errors.

Anomalous difference Pattersons revealed clearly the $M-M$ peaks at the 6.6$\sigma$ (Fe) and 14.2$\sigma$ (Zn) levels. Phasing based on these metal sites by *SHARP* and *DM* produced easily interpretable electron-density maps for both proteins.

### 5.6. Ferredoxin with two $Fe_4S_4$ clusters

Data to a resolution of 0.94 Å from a crystal of *C. acidurici* ferredoxin (CauFd) were collected with a wavelength of 0.88 Å and used to refine the structure at atomic resolution (Dauter *et al.*, 1997). However, inspection of this data set has shown a considerable amount of anomalous signal (Fig. 4*a*). The $f''$ value of Fe at this wavelength is 1.25 electrons, so that eight Fe atoms in the 55-residue protein gives the expected $\langle\Delta F^{\pm}\rangle/\langle F\rangle$ of 3.8%.

All direct-methods attempts to find the positions of Fe atoms based on Bijvoet differences have failed. However, the iron substructure was easily solved by *SHELXD* run against the native non-anomalous data, with a 95% success rate of phasing trials. *SHARP* phasing based on anomalous differences followed by *DM* led to a clearly interpretable map.

Eight Fe atoms constitute 21% of the total scattering of the 396 non-H atoms in this molecule and, according to Sim phase-probability distribution (Sim, 1959), the protein phases are on average close to the anomalous substructure phases (Fig. 2*b*). It is possible to partially interpret the electron-density map calculated with protein amplitudes and substructure phases and build the complete protein in this way after a few iterations. This can be achieved completely automatically by the application of *wARP* in the 'warp_solve' mode. CauFd therefore behaves like a typical 'small structure' and the anomalous diffraction signal is not necessary to solve it. A final analysis of the phase errors against the final model indicated two interesting features in this case: the *SHARP* phases were excellent but the figures of merit were much too low and the *DM* phases were distorted away from the Fe

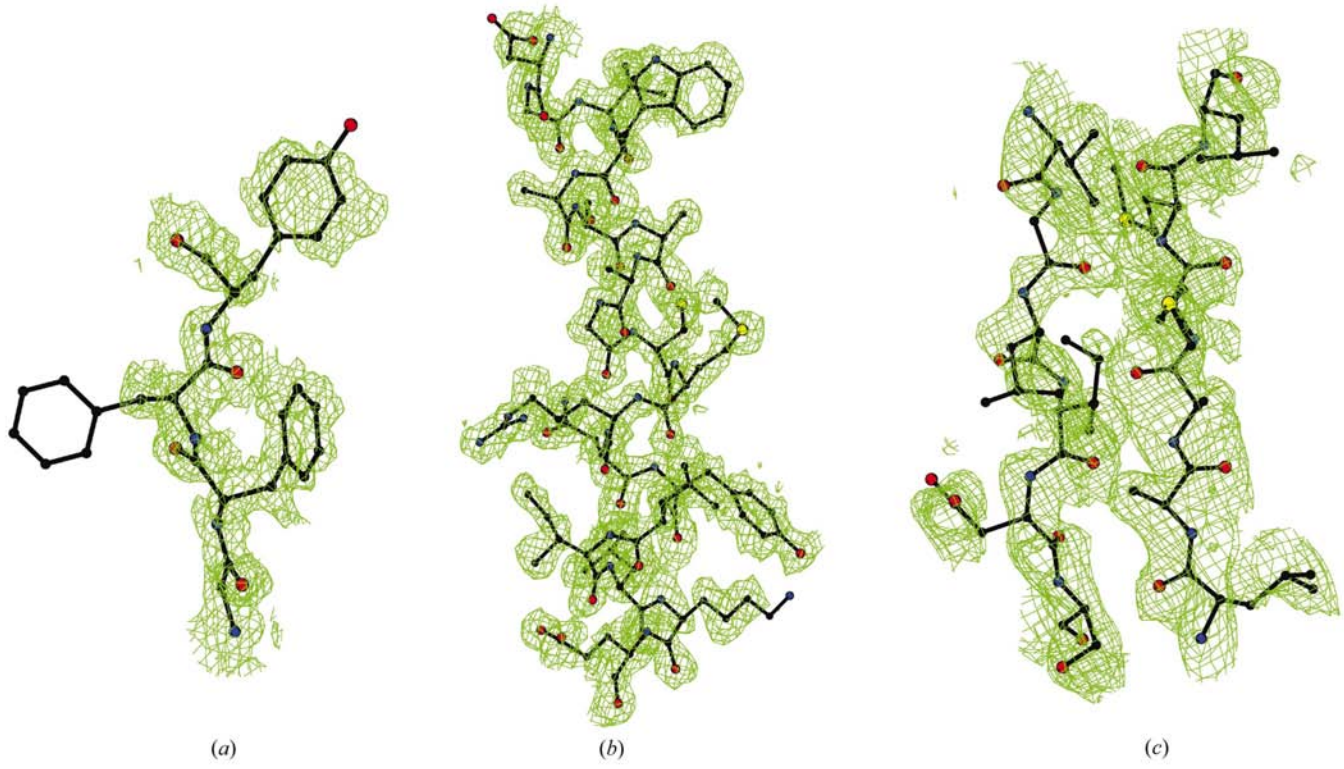


**Figure 5**
Electron-density maps contoured at the 1$\sigma$ level calculated with phases from *DM* for examples from §5: (*a*) 2Zn insulin, where the phenylalanine side chain sticking out into the solvent region has been wiped out by density modification; (*b*) a helix region of glucose isomerase, (*c*) dUTPase, a portion of a $\beta$-sheet.

**Table 3**
Anomalous scatterer sites found by *SHELXS*.

Peaks in the *E* map and *B* factors of the corresponding atoms in the refined model are given.

(*a*) *SHELXS* results for Ca subtilisin.

| *SHELXS* peaks | | Refined atoms | |
|---|---|---|---|
| Rank | Height | Atom | *B* (Å$^2$) |
| 1 | 420 | Ca1 | 5.2 |
| 2 | 268 | Ca3 | 8.5 |
| 3 | 142 | Ca2 | 18.5 |
| 4 | 123 | SD222 | 4.8 |
| 5 | 119 | SD119 | 5.6 |
| 6 | 117 | Cl1 | 10.8 |
| 7 | 100 | Cl3 | 20.6 |
| 8 | 99 | SD175 | 7.5 |
| 9 | 91 | Cl2 | 15.8 |
| 10 | 88 | Cl4 | 25.2 |
| 11 | 68 | First wrong peak | |

(*b*) *SHELXS* results for glucose isomerase.

| *SHELXS* peaks | | Refined atoms | |
|---|---|---|---|
| Rank | Height | Atom | *B* (Å$^2$) |
| 1 | 453 | Mn | 10.2 |
| 2 | 119 | Mg† | 8.9 |
| 3 | 93 | SD307 | 8.3 |
| 4 | 84 | SD84 | 10.6 |
| 5 | 83 | SD370 | 8.9 |
| 6 | 79 | SD88 | 7.7 |
| 7 | 77 | SD380 | 11.4 |
| 8 | 77 | SG306 | 6.9 |
| 9 | 70 | SD158 | 11.1 |
| 10 | 59 | SD223 | 8.0 |
| 11 | 52 | First wrong peak | |

† This site is partially occupied by Mn.

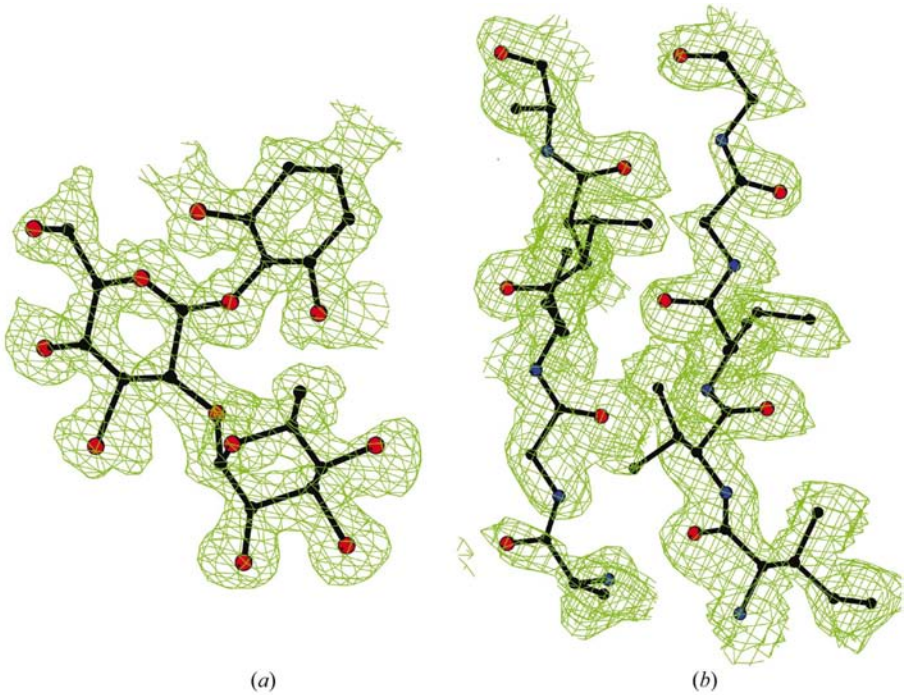


**Figure 6**
Electron-density maps (1$\sigma$ level) calculated with phases from *DM* for examples from §6: (*a*) decaplanin, a di-sugar fragment; (*b*) a $\beta$-sheet region in APT.

phases by the assumptions in the default histogram matching applied to the electron density, which effectively 'flattened' the Fe contribution.

### 5.7. Subtilisin phased on three calcium ions

Subtilisins are known to have calcium-binding sites important for protein stability. Diffraction data from an orthorhombic crystal of subtilisin from *Bacillus lentus* (Betzel *et al.*, 1988) grown in the presence of 1 *M* $CaCl_2$, were collected using 1.54 Å synchrotron radiation. A total of 270° rotation was covered in a single pass to the rather modest resolution limit of 1.75 Å and the exposure time was adjusted to avoid overloaded detector pixels. The data merged with $R_{\rm merge}$ = 2.9% and $R_{\rm anom}$ = 1.5%. The expected value of $\langle\Delta F^{\pm}\rangle/\langle F\rangle$ for three $Ca^{2+}$ ions, three methionine sulfurs and four chlorides per 270 residues is 1.35%, close to the observed value (Fig. 4*c*). The Harker sections of the anomalous difference Patterson synthesis (Fig. 3*b*) clearly revealed three calcium sites and some additional peaks resulting from sulfur sites.

*SHELXS* run against Bijvoet differences gave a clear solution, where ten highest peaks corresponded to three calcium ions (peaks 1, 2 and 3; $f''_{\rm Ca}$ = 1.28), three methionine sulfurs (peaks 4, 5 and 8; $f''_{\rm S}$ = 0.56) and four chloride ions (peaks 6, 7, 9 and 10; $f''_{\rm Cl}$ = 0.70). Table 3(*a*) lists the *SHELXS E*-map peaks and *B* factors of all anomalous scatterers in the refined structure. There is a clear correlation between *SHELXS* rank and the atomic anomalous scattering and *B* factor. In fact, the high temperature factors of the chloride ions may take into account their partial occupancies, so that sulfurs show up as higher peaks than chlorides. Analogous behavior of sulfurs and chlorides has been observed in lysozyme (Dauter *et al.*, 1999)

Only three calcium sites were used for phasing with *SHARP*, which gave an FOM of 0.338 and of 0.769 after density modification by *DM*. The resulting electron-density map was easily interpretable.

### 5.8. Subtilisin phased on $Lu^{3+}$ ions

Another crystal of subtilisin grown from 1 *M* NaCl was soaked in a 1 *M* solution of $LuCl_3$ for a short time and MAD data to 1.75 Å were collected at four wavelengths in the vicinity of the lutetium $L_{\rm III}$ absorption edge (1.34 Å), at the peak, inflection, high-energy and low-energy remote points of the fluorescence spectrum, each with a total rotation of 180° (Table 1). *SHELXD* run against peak, inflection or high-energy remote data clearly identified four Lu sites which had various fractional occupancies. Only one of the Lu

**Table 4**
Phasing of Lu subtilisin against various SAD and MAD data.

| Data | $f''_{XPREP}$ | % *SHELXD* | Lu No. | $\Delta\varphi_{SHARP}$ (°) | $\Delta\varphi_{DM}$ (°) | $CC_{SHARP}$ | $CC_{DM}$ |
|---|---|---|---|---|---|---|---|
| $\lambda_{peak}$ | 10.7 | 98.9 | 1 | 71.2 | 61.6 | 0.39 | 0.53 |
| | | | 4 | 63.4 | 49.5 | 0.51 | 0.71 |
| $\lambda_{infl}$ | 6.9 | 99.3 | 1 | 72.1 | 63.3 | 0.37 | 0.49 |
| | | | 4 | 64.4 | 50.1 | 0.48 | 0.68 |
| $\lambda_{remote}$ | 4.2 | 100.0 | 1 | 76.1 | 73.2 | 0.30 | 0.34 |
| | | | 4 | 69.7 | 58.2 | 0.42 | 0.59 |
| $\lambda_{low}$ | 1.9 | 63.7 | 1 | 81.8 | 85.7 | 0.19 | 0.09 |
| | | | 4 | 78.0 | 80.1 | 0.26 | 0.21 |
| 4λ | All | 98.0† | 1 | 65.1 | 55.4 | 0.46 | 0.61 |
| | | | 4 | 55.8 | 45.2 | 0.63 | 0.77 |

† *SHELXD* run against $F_A$ values prepared by *XPREP* (Bruker–Nonius Inc.).

sites (the third strongest) corresponded to a calcium sites in the Ca-substituted subtilisin.

The strongest Lu site was sufficient for successful phasing of the protein by MAD based on all four wavelength data and also for phasing the single data sets at peak, inflection and remote wavelengths (Table 4). According to expectation, the low-energy remote data did not contain enough anomalous signal to provide meaningful phasing power in the SAD mode. Also, the phasing from the high-energy remote data using only one Lu site did not generate an interpretable map.

### 5.9. Glucose isomerase: a 44 kDa protein phased on a single $Mn^{2+}$ ion

Glucose isomerase from *Streptomyces rubiginosus* (Carrell *et al.*, 1989) was supplied by Hampton Research. The *I*222 crystals were grown from 11% MPD, 0.1 *M* $MgCl_2$, 0.05 *M* Tris buffer pH 7.0 with a protein concentration of 15 mg $ml^{-1}$. The glucose isomerase monomer has 388 amino acids and contains two $Mn^{2+}$ ions as a natural cofactor.

Diffraction data were collected at a synchrotron beamline to a resolution of 1.5 Å using a wavelength of 1.34 Å in two exposure passes. The crystal was rotated through 360°, which ensured high multiplicity. The data set was very complete and accurate (Table 1). For subsequent comparisons with reliable calculated phases, the available enzyme model was refined by *REFMAC* and *wARP* against these data to an *R* factor of 18.3%. The anomalous difference Fourier synthesis using these phases showed clear peaks for one metal (at 81$\sigma$), a second metal (24$\sigma$) and nine sulfurs (18–12$\sigma$, eight Met SD and one Cys SG). Only the highly disordered N-terminal methionine sulfur was not visible. The strong metal site proved to be a fully occupied $Mn^{2+}$ ion and the second metal site is highly substituted by $Mg^{2+}$, which is present in the crystallization medium. At this wavelength the $f''$ values are 2.23 electrons for Mn and 0.43 electrons for S; Mg does not show any significant anomalous scattering effect. The theoretically expected amount of anomalous signal is about 0.9%. The $\langle\Delta F^{\pm}\rangle/\langle F\rangle$ ratio in the diffraction data shown in Fig. 4(*b*) approximates this value almost to the highest resolution limit.

The anomalous difference Patterson revealed one prominent peak corresponding to the stronger metal site. The direct-methods program *SHELXS* using Bijvoet differences gave a clear solution with ten anomalous scatterers correctly positioned (Table 3*b*).

The single strongest metal site positioned from the anomalous difference Patterson was input to *SHARP* and the phases with very low FOM of 0.15 submitted to density modification by *DM*, resulting in a phase set with FOM = 0.78 and an excellent electron-density map (Fig. 5*b*).

According to the estimation of Wang (1985) the anomalous signal of about 0.6%, corresponding to one disulfide group in a 12.5 kDa protein, can lead to structure solution. The glucose isomerase data are close to this limit and indeed suggest that a signal smaller than 1% could be successfully used for phasing, provided the differences are measured sufficiently accurately.

### 5.10. d-UTPase: a single-site Hg derivative

The structure of *Escherichia coli* dUTPase in space group *R*3 was originally solved by the MIRAS approach (Cedergren-Zeppezauer *et al.*, 1992; Dauter *et al.*, 1998) using a single-site mercury and a second, somewhat less substituted, platinum derivative. The data collected from the Hg-derivatized crystal extended to 2.0 Å and contained a significant amount of anomalous signal, but were not measured with high redundancy and only 85% (9924 of 11 687) reflections had both Friedel mates.

The Hg atom identified from the anomalous difference Patterson was input to *SHARP* which, after density modification by *DM*, led to an interpretable electron-density map (Fig. 5*c*).

This structure was recently successfully phased against newly collected atomic resolution Hg-derivative data by the *wARP* approach (González *et al.*, 2001).

## 6. Examples of novel structures solved by SAD

Several new structures were recently solved using SAD data collected at the X9B beamline of NSLS, Brookhaven National Laboratory and the ADSC Quantum-4 CCD detector. All data were processed with *HKL*2000 (Otwinowski & Minor, 1997) and their statistics are given in Table 1.

### 6.1. Oligopeptide antibiotic phased on chlorines

Decaplanin, a cyclic nonapeptide antibiotic composed from unusual amino acid and glycosyl residues, contains one covalently bound Cl atom. Small polypeptides are notoriously difficult to derivatize. Their conformational flexibility means that it is also not easy to apply molecular replacement, even if the similar structure is known. Direct methods can be applied only if crystals diffract to atomic resolution.

The $P6_122$ crystal form of decaplanin was recently solved (Lehmann, 2000) based on the anomalous signal within the data set collected to 1.6 Å resolution with a wavelength of 1.54 Å, where chlorine has an $f''$ of 0.70 electrons. The data were of high quality, characterized by an $R_{merge}$ of 5.6%, a redundancy of 10 for individual Friedel mates and an $I/\sigma(I)$ of 54 overall and 14 in the highest resolution shell.

It was not known *a priori* how many antibiotic molecules could be expected in the asymmetric unit of the crystal, since such compounds can crystallize with a wide range of packing densities. *SHELXD* gave good solutions (in 20% of phase sets) with eight independent anomalous scatterer sites and eight molecules could be accommodated in the asymmetric unit of the crystal. *SHARP* gave an FOM of 0.39 and a subsequent *DM* run assuming 40% solvent in the crystal (corresponding to eight independent molecules) gave an FOM of 0.63. Surprisingly, the map showed only four independent molecules, each with one bound Cl atom and an additional four chloride ions each hydrogen bonded by three main-chain NH groups of the antibiotic. The *DM* run repeated with the solvent content appropriate for four molecules, 70%, produced phases with an FOM of 0.89 and an excellent map (Fig. 6*a*). The expected $\langle \Delta F^{\pm} \rangle / \langle F \rangle$ ratio for eight independent Cl atoms is 2.0% and the amount of the anomalous signal in the data is close to that value (Fig. 4*b*). Four molecules of decaplanin contain 448 atoms and this structure is larger than rubredoxin or ferredoxin.

### 6.2. Serine-carboxyl proteinase: cryosoaked bromides

The structure of the 44 kDa *Pseudomonas* serine-carboxyl proteinase (PSCP) has recently been solved (Dauter *et al.*, 2001) using single-wavelength data collected from the crystal soaked in a cryosolution containing 1 *M* NaBr. The PSCP crystals resisted all derivatization attempts with classic heavy-atom reagents. The three-wavelength MAD data to 1.8 Å were collected around the bromine absorption edge, but only the peak-wavelength data were used for structure solution.

Nine sites from the *SHELXD* solution of the anomalous scatterers were selected for subsequent protein phasing. The 1.8 Å *SHARP* phases were extended by *DM* to the full resolution of the native data (1.4 Å) with an FOM of 0.74. The resulting map, with a correlation coefficient (CC) to an $F_{\rm obs}$ map of 0.62, was interpretable and was subjected to automatic model building by *wARP*. The protein model thus obtained was almost complete and required only a small number of side chains to be built by hand.

Since the MAD data were available, it was possible to compare the results of single-wavelength and three-wavelength phasing, which was performed *a posteriori* in an analogous way using *SHARP* and *DM*. The resulting three-wavelength *DM* map was superior, with a CC to the $F_{\rm obs}$ map of 0.76, but the *wARP* procedure led to a comparable result. However, this procedure required significantly more time both for data collection and for phasing (*SHARP* took 75 h for three-wavelength data and 15 h for single-wavelength data).

### 6.3. Acyl-protein thioesterase: phasing on cryosoaked bromide ions

The structure of human acyl-protein thioesterase (APT) was also solved recently (Devedjiev *et al.*, 2000) using single-wavelength data collected from a crystal soaked for 20 s in cryosolution containing 1.0 *M* NaBr. No separate native data were available. The bromide sites were found and promising preliminary phasing results were obtained shortly after diffraction images from the peak wavelength had been processed, so no further data collection with additional wavelengths was attempted.

There are two molecules of the protein, 56 kDa in total, in the asymmetric unit of the monoclinic $P2_1$ cell. The 1.8 Å data was measured at the Br fluorescence peak wavelength in 360 images of 0.5° rotation each. The measured intensities were strong, with an $I/\sigma(I)$ of 30 overall and of 8 in the highest resolution shell. The $R_{\rm merge}$ was 2.8% with Friedel mates not merged (average redundancy 1.9) and 4.8% with Friedel equivalents merged (average redundancy 3.8). The bromide sites were obtained easily both from *SnB*, with 309 solutions with $R_{\rm min}$ in the range 0.29–0.35 against 691 unsuccessful phase sets with $R_{\rm min}$ in the range 0.50–0.75, and from *SHELXD*, which gave 1000 successful solutions with the CC figure-of-merit above 0.30 per 1000 attempted phase sets.

However, as usual with a number of halide sites having different occupancies, there was no clear demarcation in the peak heights to suggest which sites are meaningful. The eight strongest sites were used for the first round of *SHARP* phasing and the set was expanded to 22 sites after inspecting residual peaks in two more rounds of *SHARP*. 18 of these sites could be grouped into two identical constellations, clearly confirming the presence of two independent molecules in the asymmetric unit of the crystal related by a noncrystallographic twofold axis. The density modification by *DM* increased the FOM from 0.40 to 0.85 and gave a map of high quality (Fig. 6*b*).

### 6.4. Mitomycin C resistance protein (MRD) phased on selenomethionines

The diffraction data were collected on a variant of MRD containing three SeMet residues in a 15 kDa protein. The wavelength was set to the peak of the selenium fluorescence spectrum at 0.979 Å. The data extended to a resolution of 1.3 Å and scaled with an $R_{\rm merge}$ of 3.8% and an $R_{\rm anom}$ of 6.8%, suggesting a strong anomalous signal, despite each Friedel mate being measured only twice on average. For $f''_{\rm Se} = 4.5$, the estimated value of $\langle \Delta F^{\pm} \rangle / \langle F \rangle$ is 5.2%.

*SHELXD* run against Bijvoet differences gave the correct solution for three selenium sites in 100% of the trials. *SHARP* produced a phase set with an average FOM of 0.57. After *DM* (FOM = 0.69 at 1.3 Å) the map was very clear and *wARP* automatically built 110 out of 130 residues with most of the side chains. The details of the structure solution and refinement will be given elsewhere (Martin, Devedjiev, Dauter, He, Sherman, Derewenda & Derewenda, in preparation).

## 7. Discussion

Traditionally in macromolecular crystallography the anomalous signal of heavy atoms has been used as an auxiliary source of phasing in the MIR method of solving new crystal structures, with the isomorphous differences in intensities between the native and derivative data serving as a main

source of phasing power. There were difficulties in accurately measuring the small anomalous differences using photographic film. However, the potential power of the anomalous scattering signal was recognized in early days of macromolecular crystallography (Ramachandran & Raman, 1956; Green *et al.*, 1954; Rossmann, 1961). The use of anomalous scattering of sulfur for phasing crambin (Hendrickson & Teeter, 1981) and the simulations by Wang (1985) showed the great phasing potential of the anomalous signal even with single-wavelength data.

The introduction of more accurate automatic detectors, such as multiwire chambers, imaging plates and CCDs, as well as the use of crystal-freezing techniques, has now made it possible to collect diffraction intensities very accurately. These developments contributed to the popularity of the MAD method of phasing, which is currently the method of choice for solving novel crystal structures of macromolecules. The MAD method completely alleviates all problems of non-isomorphism of derivatives characteristic of the classic MIR approach. However, it requires precise control of the synchrotron-radiation wavelength for the proper estimation of both the anomalous and dispersive differences.

Several examples quoted above suggest that with the current availability of the state-of-the-art technical hardware (synchrotron beamlines, detectors) and computational software (data-processing programs, phasing and density-modification algorithms) it is 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. In contrast to the MAD method, the wavelength need not correspond to the maximum anomalous effect near the absorption edge of the corresponding anomalous scatterer.

The accuracy of measurements required for successful phasing by SAD seems to be comparable to that acquired in routine MAD experiments. In favorable cases, satisfactory results can be obtained with data of relatively poor quality with low completeness and redundancy, as in the case of 2Zn insulin or dUTPase. It should, however, be stressed that excellent data quality, achieved mainly by recording high multiplicity of measurements, leads to more accurate phases and as a result more interpretable electron-density maps. Inspection of Tables 1 and 2 confirms that the most accurate phases were obtained for data collected with the highest redundancy of measurements, *e.g.* lysozyme, DNA, Ca subtilisin, glucose isomerase and decaplanin.

The resolution of diffraction data is not so important for SAD phasing as the data quality. In the examples discussed above as well as those available in the literature [*e.g.* crambin (Hendrickson & Teeter, 1981; Wang, 1985), neurophysin (Chen *et al.*, 1991), IF3 (Biou *et al.*, 1995), rusticyanin (Harvey *et al.*, 1998), psoriasin (Brodersen *et al.*, 2000) and obelin (Liu *et al.*, 2000)], the resolution limits of the data have varied from 1.05 to 3.0 Å. The clear indication of the importance of data accuracy, which can be best achieved by multiple measurements of equivalent reflections, is shown by phasing a DNA oligonucleotide against a series of data sets of varying redundancy (Dauter & Adamiak, 2001).

The SAD phasing procedure generally consists of three stages: (i) finding anomalous scatterers, (ii) evaluation of initial phases and (iii) phase improvement by density modification. The first stage can utilize the anomalous difference Patterson vector search, direct methods or a combination of both. The evaluation of experimental phases can be achieved in various ways, for example by maximum-likelihood estimation. The third stage involves phase-modification procedures such as solvent flattening and histogram matching.

In all the examples presented above, a similar protocol and programs were used for identification of anomalous sites, phasing and density modification to ensure compatibility of results. However, other available programs were also tried and gave positive results, such as *SOLVE* (Terwilliger & Berendzen, 1999) and *ACORN* (Foadi *et al.*, 2000) for searching for anomalous sites and phasing, *MLPHARE* (Otwinowski, 1991) for initial phasing or *SOLOMON* (Abrahams & Leslie, 1996) for density modification.

In some cases, the full three-wavelength data were collected according to the principle of the MAD technique, but only one data set was eventually used for successful structure solution by SAD (§5.2). In these cases, the common MAD anomalous scatterers such as Se or Br were used and diffraction data were collected in a standard way, with crystals in arbitrary orientation and without using the inverse-beam approach. The redundancy was not high enough to ensure the full completeness of anomalous data. In addition, the intensities were not collected to absolutely the highest potentially achievable resolution; instead, attention was directed to appropriate estimation of the lowest resolution strongest reflections, which play the most important role in the Patterson search, direct methods and phasing and are important for the correct appearance of electron-density maps. In these cases the anomalous signal amounted to about 4–5% of the total scattering and all structure-solution steps proceeded smoothly; easily interpretable electron-density maps were obtained when protein models were constructed in an automatic way by *wARP*.

A few examples quoted above are based on the data collected without optimizing the anomalous signal in mind. During data collection of 2Zn insulin or rubredoxins, no effort was directed towards enhancing the anomalous signal contained in the reflection intensities. The data-collection protocol was instead optimized to achieve the highest possible resolution and completeness within the asymmetric unit of the reciprocal lattice. Nevertheless, even the rather weak anomalous signal present in these incomplete data led to successful phasing of the crystal structure. In these examples, the weakness of the anomalous signal has been compensated by the very high resolution of the diffraction data.

## 8. Conclusions

The theoretical potential of the anomalous scattering signal as a sole source of phasing information has been known since the early days of macromolecular crystallography (Ramachandran

& Raman, 1956). The practical utilization of SAD by Hendrickson & Teeter (1981) has shown that it can indeed lead to successful phasing, provided the data can be measured accurately. Wang (1985) estimated that the anomalous signal at a level lower than 1% of the total scattering may be sufficient. To measure diffraction intensities at such levels of accuracy is a challenge, but the progress in data-collection techniques observed in recent years (cryofreezing, two-dimensional detectors and data-reduction programs) makes such tasks feasible. On the other hand, the progress and automation in phasing and density-modification algorithms makes the SAD structure solution easier and faster.

A limited number of novel structures have been solved by the SAD approach since crambin. The examples presented in this paper suggest that the strength of SAD phasing is greater than has been anticipated. In many cases of MAD phasing a successful result can be obtained from a single data set (Rice *et al.*, 2000), which requires much less time and effort for data collection and handling and minimizes the crystal radiation decay. It can be expected that the SAD approach will gain wider applicability, especially in the current era of high-throughput structural projects.

The anomalous diffraction data discussed here are available on request from ZD.

## References


Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Betzel, C., Dauter, Z., Dauter, M., Ingelman, M., Papendorf, G., Wilson, K. S. & Branner, S. (1988). *J. Mol. Biol.* **204**, 803–804.
Biou, V., Shu, F. & Ramakrishnan, V. (1995). *EMBO J.* **14**, 4056–4064.
Blow, D. M. & Rossmann, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.
Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. New York: Academic Press.
Boggon, T. J. & Shapiro, L. (2000). *Structure*, **8**, R143–R149.
Brodersen, D. E., de La Fortelle, E., Vonrhein, C., Bricogne, G., Nyborg, J. & Kjeldgaard, M. (2000). *Acta Cryst.* D**56**, 431–441.
Carrell, H. L., Glusker, J. P., Burger, V., Manfre, F., Tritsch, D. & Biellmann, J. F. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 4440–4444.
Cedergren-Zeppezauer, E. S., Larsson, G., Nyman, P. O., Dauter, Z. & Wilson, K. S. (1992). *Nature (London)*, **355**, 740–743.
Chen, L., Rose, J. P., Breslow, E., Yang, D., Chang, W. R., Furey, W. F., Sax, M. & Wang, B. C. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 4240–4244.
Cowtan, K. (1999). *Acta Cryst.* D**55**, 1555–1567.
Cromer, D. T. (1983). *J. Appl. Cryst.* **16**, 437–438.
Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* D**57**, 990–995.
Dauter, Z. & Dauter, M. (2001). *Structure*, **9**, R21–R26.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Dauter, Z., Li, M. & Wlodawer, A. (2001). *Acta Cryst.* D**57**, 239–249.
Dauter, Z., Wilson, K. S., Larsson, G., Nyman, P. O. & Cedergren-Zeppezauer, E. S. (1998). *Acta Cryst.* D**54**, 735–749.
Dauter, Z., Wilson, K. S., Sieker, L. C., Meyer, J. & Moulis, J. M. (1997). *Biochemistry*, **36**, 16065–16073.
Dauter, Z., Wilson, K. S., Sieker, L. C., Moulis, J. M. & Meyer, J. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 8836–8840.
Devedjiev, Y., Dauter, Z., Kuznetsov, S. R., Jones, T. L. Z. & Derewenda, Z. S. (2000). *Structure*, **8**, 1137–1465.
Dodson, E. J. & Vijayan, M. (1971). *Acta Cryst.* B**27**, 2402–2411.
Drenth, J. (1999). *Principles of Protein X-ray Crystallography*, 2nd ed. Heidelberg: Springer.
Fan, H. F., Hao, Q., Gu, Y. X., Qian, J. Z., Zheng, C. D. & Ke, H. (1990). *Acta Cryst.* A**46**, 935–939.
Flack, H. D. (1983). *Acta Cryst.* A**39**, 876–881.
Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *Acta Cryst.* D**56**, 1137–1147.
Fourme, R., Shepard, W. & Kahn, R. (1996). *Prog. Biophys. Mol. Biol.* **64**, 167–199.
González, A., Larsson, G., Persson, R. & Cedergren-Zeppezauer, E. (2001). *Acta Cryst.*, D**57**, 767–774.
González, A., Pedelacq, J. D., Sola, M., Gomis-Rüth, F. X., Coll, M., Samama, J. P. & Benini, S. (1999). *Acta Cryst.* D**55**, 1449–1458.
Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 287–307.
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., Ingledew, J. & Hasnain, S. S. (1998). *Acta Cryst.* D**54**, 629–635.
Hauptman, H. A. (1982). *Acta Cryst.* A**38**, 289–294.
Hauptman, H. A. (1996). *Acta Cryst.* A**52**, 490–496.
Hendrickson, W. A. (1979). *Acta Cryst.* A**35**, 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.* A**55**, 755–760.
Lehmann, C. (2000). Thesis, University of Göttingen, Germany.
Liu, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
Mathews, B. W. (1966). *Acta Cryst.* **20**, 230–239.
Mukherjee, A. K., Helliwell, J. R. & Main, P. (1989). *Acta Cryst.* A**45**, 715–718.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
North, A. C. T. (1965). *Acta Cryst.* **18**, 212–216.
Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Scattering and Anomalous Replacement*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Phillips, J. C. & Hodgson, K. O. (1980). *Acta Cryst.* A**36**, 856–864.
Ramachandran, G. N. & Raman, S. (1956). *Curr. Sci.* **25**, 348–351.
Rice, L. M., Earnest, T. N. & Brunger, A. T. (2000). *Acta Cryst.* D**56**, 1413–1420.
Rossmann, M. G. (1961). *Acta Cryst.* **14**, 383–388.
Sheldrick, G. M. (1986). *SHELXS*86. *Program for Crystal Structure Solution*. University of Göttingen, Germany.
Sheldrick, G. M. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 401–411. Dordrecht: Kluwer Academic Publishers.
Sheldrick, G. M. & Schneider, T. (1997). *Methods Enzymol.* **277**, 319–343.
Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.
Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.