Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Multiple isomorphous replacement on merohedral twins: structure determination of deacetoxycephalosporin C synthase

Anke C. Terwisscha van Scheltinga,[a]* Karin Valegård,[b] S. Ramaswamy,[a]† Janos Hajdu[b] and Inger Andersson[a]

[a]Department of Molecular Biology, Swedish University of Agricultural Sciences, Box 590, S-751 24 Uppsala, Sweden, and [b]Department of Biochemistry, Uppsala University, Box 576, S-751 23 Uppsala, Sweden

† Present address: Department of Biochemistry, University of Iowa, Iowa City, IA 52242, USA.

Correspondence e-mail: anke@xray.bmc.uu.se

Received 27 December 2000
Accepted 28 August 2001

Merohedral twinning is a packing anomaly that seriously impairs the determination of macromolecular crystal structures. Crystals of deacetoxycephalosporin C synthase (DAOCS), an enzyme involved in the expansion of the penicillin nucleus to form the core structure of the cephalosporin antibiotics, were found to be merohedrally twinned by many diagnostic criteria. Here, the structure determination of DAOCS from twinned crystals based on a combination of isomorphous replacement and the use of a multiple-wavelength diffraction data set is described.

1. Introduction

Twinning, caused by crystal packing or growth irregularities, is a frequently encountered obstacle in the determination of a macromolecular crystal structure. Two basic types of twinning may be identified (Koch, 1995): (i) epitaxic twinning, in which the lattices from the individual repeating units (distinct domains) match in one or two dimensions, and (ii) merohedral twinning, in which the distinct domains in the crystal superimpose completely in three dimensions. The first type, which is often referred to as 'non-merohedral' twinning, results in a crystal that is composed of two or more interpenetrating lattices, each leading to separate diffraction spots. In this case detection is trivial, even if a remedy may be far from easy to achieve. In merohedrally twinned crystals, however, the separate domain lattices have identical dimensions. Therefore, the diffraction spots from the domains superimpose exactly, so that each measured diffraction intensity contains contributions from two (or more) twin reflections. Because of this exact overlap, the diffraction pattern from a merohedrally twinned crystal appears quite 'normal' and the anomaly is therefore easy to overlook. The phenomenon of merohedral twinning was realised at an early stage (Friedel, 1926; Buerger, 1960) and seems to be fairly common in crystals of small inorganic and organic compounds (Catti & Ferraris, 1976; Araki, 1991; Yeates, 1997; Kahlenberg, 1999). In macromolecular crystallography, on the other hand, very few structures from merohedrally twinned crystals were reported until relatively recently. It is quite conceivable that many cases of merohedral twinning in crystals of macromolecules go undetected and that the phenomenon is under-reported (Yeates & Fam, 1999).

Most protein structures from merohedrally twinned crystals were determined using phases from molecular replacement (MR; Rees & Lipscomb, 1980; Reynolds *et al.*, 1985; Redinbo & Yeates, 1993; Gomis-Rüth *et al.*, 1995; Luecke *et al.*, 1998; Breyer *et al.*, 1999; Chandra *et al.*, 1999). Although structure determination by multiple isomorphous replacement (MIR) methods has been shown to be possible in theory (Yeates &

Rees, 1987), only a few cases have been reported, mostly from crystals with twin fractions less than 0.10 (Fisher *et al.*, 1980; Cheah *et al.*, 1994; Igarashi *et al.*, 1997). A recent study reports structure determination by multiwavelength anomalous diffraction (MAD) from crystals with a twin fraction of 0.35 (Yang *et al.*, 2000).

Here, we describe the structure determination of deacetoxycephalosporin C synthase (DAOCS; Valegård *et al.*, 1998) and show that it is possible to extract useful phases by MIR methods from twinned crystals even with high twin fractions. Aspects of merohedral twinning in macromolecular crystals are presented in the context of this particular case. More detailed descriptions of twinning can be found in a number of reviews (Fisher & Sweet, 1980; Rees, 1982; Yeates, 1997; Yeates & Fam, 1999).

2. Twinning by merohedry

The geometrical relation between twin components is called the twin law. In a crystal that is twinned by merohedry, the asymmetric units of the twin components have identical dimensions, requiring the twin law to be part of the lattice symmetry (Koch, 1995). This is possible if the twin operation, which is the symmetry operation associated with the twin law, is the crystallographic symmetry of a higher symmetry space group, *i.e.* the point-group symmetry of the crystal is lower than that of the lattice. For example, the twin operator for *R*3 crystals generates the twofold axis of the higher symmetry space group *R*32. Applying this operation results in the overlap of (h, k, l) and (k, h, $-l$) reflections, which is equivalent to the two ways *R*3 data can be indexed. More generally, in a space group where the reflections can be indexed in more than one way, the reindexing matrix describes a possible twin law. Additionally, the lattice symmetry can be higher than the space-group symmetry in special cases, *e.g.* if two non-identical axes are similar in length (Andersson & Brändén, 1984; Knight *et al.*, 1990; Yang *et al.*, 2000). These cases are referred to as pseudo-merohedral twins.

For three-dimensional protein crystals, the most common type of merohedral twinning generates only two twin components (hemihedral twinning). The fraction of the minor component is by convention called the twin fraction, α.

Assuming that the twin domains are larger than the coherence length of the X-ray beam, the observed intensities can be described by a summation of the intensities of the twin components: each observed intensity consists of contributions from two reflections related by the twin law. For the two intensities I_1 and I_2 related by a twin law, the observed intensities can be described by

$$I_{\rm obs,1} = (1-\alpha)I_1 + \alpha I_2, \quad (1a)$$

$$I_{\rm obs,2} = \alpha I_1 + (1-\alpha)I_2. \quad (1b)$$

Once the twin operator and α have been determined, the contributions of each twin can in principle be separated, *i.e.* the data can be detwinned. A problem inherent in this correction is that it amplifies the experimental errors by a factor of approximately $1/(1-2\alpha)$ (Grainger, 1969; Redinbo & Yeates, 1993); the error increases as α increases and becomes infinite when α is 0.50 (perfect twinning). Consequently, perfect twins cannot be detwinned without additional information (Redinbo & Yeates, 1993).

3. Detection of merohedral twinning in DAOCS crystals

DAOCS belongs to the extended family of mononuclear ferrous enzymes (for reviews, see Feig & Lippard, 1994; Que & Ho, 1996; Schofield & Zhang, 1999). The structure of the

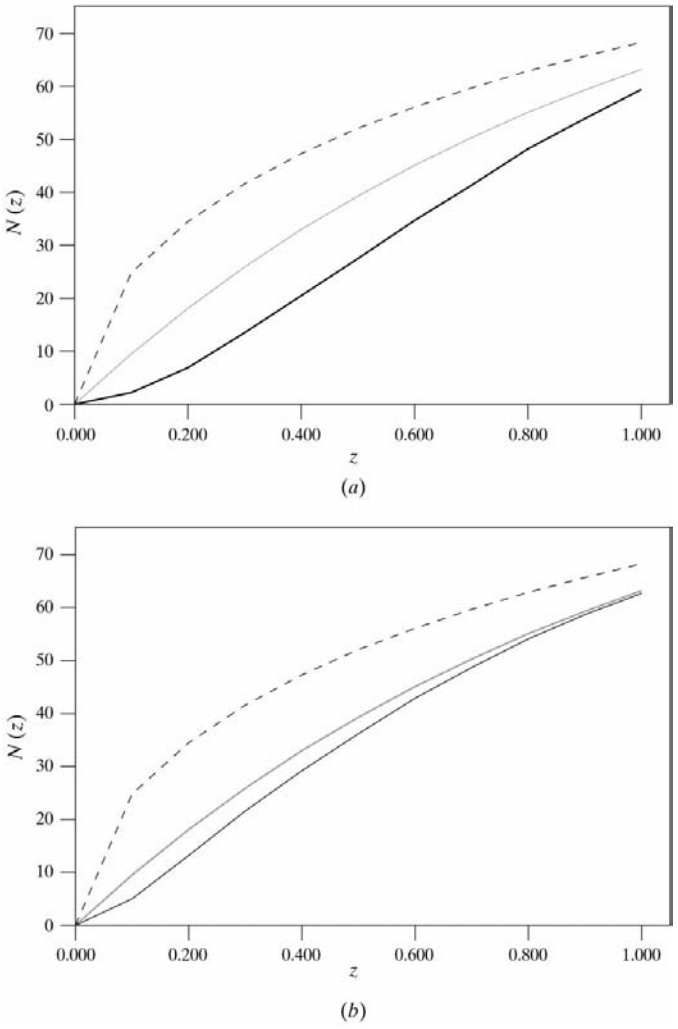

Figure 1
Cumulative intensity-distribution curves, produced with *TRUNCATE* (Collaborative Computational Project, Number 4, 1994). In the graph $N(z)$ is plotted as a function of z, in which $z = I/\langle I\rangle$ and $N(z)$ is the fraction of the reflections whose intensities are less than or equal to z (Howells *et al.*, 1950). The dashed line shows the theoretical distribution for centric reflections and the grey line that for non-centric reflections. The black line shows the observed distribution for non-centric reflections (DAOCS crystals belong to space group *R*3, hence there are no observed centric reflections). Since for twinned crystals observed intensities consist of two (or more) components, there will be fewer observations with extremely low intensity, which results in a sigmoid cumulative intensity-distribution curve. (*a*) A highly twinned crystal (α = 0.45). (*b*) A crystal displaying a low twin fraction (SeMet, α = 0.07).

Table 1
Data-collection statistics for DAOCS crystals, including estimates of twin fractions determined by Britton plots or by heavy-atom refinement.

Values in parentheses refer to the highest resolution bin.

Data set	Resolution	Completeness	R_{merge}	Twin fraction (Britton plot)	Twin fraction (MIR)
SeMet, λ = 0.9787 Å	19–2.8 (2.85–2.80)	0.990 (0.959)	0.055 (0.059)	0.07	n.a.
SeMet, λ = 0.9788 Å	19–2.8 (2.85–2.80)	0.982 (0.923)	0.055 (0.062)	0.07	0.057
SeMet, λ = 0.689 Å	19–2.7 (2.8–2.7)	0.985 (0.988)	0.044 (0.063)	0.07	0.057
Native-1	20–1.9 (1.93–1.90)	0.855 (0.490)	0.038 (0.130)	0.11	n.a.
Native-2	22–1.3 (1.33–1.30)	0.980 (0.985)	0.040 (0.148)	0.14	0.141
Pt-1	19–2.7 (2.76–2.70)	0.958 (0.870)	0.042 (0.097)	0.33	n.a.
Pt-2	25–2.6 (2.69–2.60)	0.999 (0.998)	0.046 (0.099)	0.19	0.177
Xe-1	19–2.5 (2.59–2.50)	0.939 (0.963)	0.056 (0.121)	0.45	n.a.
Xe-2	25–2.1 (2.18–2.10)	0.938 (0.631)	0.052 (0.093)	0.28	0.271

ring-expanding DAOCS from *Streptomyces clavuligerus* is the first from the family of 2-oxoacid-dependent ferrous enzymes to be determined (Valegård *et al.*, 1998).

The DAOCS apo enzyme crystallizes in space group *R*3, with unit-cell parameters $a = b = 106.4$, $c = 71.2$ Å (in the hexagonal setting) and one monomer in the asymmetric unit (Valegård *et al.*, 1998; Lloyd *et al.*, 1999). Statistics for native and derivative data sets used in the structure determination are summarized in Table 1 (see also Valegård *et al.*, 1998). The crystals have a normal morphology, look uniform under polarized light and diffract X-rays beyond 1.3 Å resolution with good scaling statistics. However, a close inspection of the intensity data from these crystals reveals all the possible symptoms of merohedral twinning.

(i) Sigmoidal probability distribution for intensities: since each observed intensity consists of two (or more) components, there will be fewer observations with extremely high or low intensity. This is visible in a plot of the cumulative intensity distribution according to Howells *et al.* (1950) (Fig. 1). Whereas for a non-twinned crystal the plot shows an exponential curve, the characteristic small number of low-intensity observations for twinned crystals will give the curve a sigmoidal shape. Other similar and useful indicators of a non-standard probability distribution include the Wilson ratio $\langle|F|\rangle^2/\langle F^2\rangle$, which for non-centrosymmetric has the value of 0.885 for non-twinned data and 0.785 for twinned data (Stanley, 1972) and the $\langle I^2\rangle/\langle I\rangle^2$ ratio, which is 2 for normal crystals and 1.5 for twinned crystals (Yeates, 1988). Since these tests are based on the intensity distribution, they have the major advantage that they do not rely on the data to be processed in the true space group. They are implemented in many programs such as *DATAMAN* (Kleywegt & Jones, 1996), *EDEN* (A. and H. Szöke, manuscript in preparation), *DETWIN* (Taylor & Leslie, 1998), *ECALC* and *TRUNCATE* (Collaborative Computational Project, Number 4, 1994). An advantage of the cumulative intensity distribution plot over the other methods is that it gives a straightforward visual indicator of merohedral twinning.

(ii) Additional symmetry in the self-rotation function: the twin operation for DAOCS crystals (space group *R*3) creates a twofold axis perpendicular to the crystallographic threefold axis, running along the diagonal between the *a* and *b* axes (Fig. 2). The height of the peaks in the self-rotation function differ between data sets and is roughly proportional to the twin fraction. For data with a twin fraction lower than 0.15, it is difficult to detect the twofold axis. Since the twin operator corresponds to the twofold axis in the higher symmetry space group *R*32, the twin-related reflections are reflections that are equivalent in *R*32. Consequently, reducing data from a highly twinned crystal in space group *R*32 only slightly increases the merging *R* factor: *e.g.* for a crystal with $\alpha = 0.36$, it was 0.037 for *R*3 and 0.080 for *R*32.

(iii) Non-isomorphism: although the merged data from one single crystal show excellent internal statistics (merging *R* around 0.03), scaling two data sets together gives a much higher *R* factor, *e.g.* the scaling together of two data sets with $\alpha = 0.01$ and 0.36 gave an *R* factor of 0.295.

(iv) Heavy-atom sites: in a structure determination by MIR, identification of suitable heavy-atom derivatives may be seriously hindered. The intensity changes between the native and derivative data sets will include a contribution from the difference in twin fractions of the two data sets. In an extreme case, a non-substituted native crystal may appear to be a potential derivative owing to intensity differences caused by twinning (Rees, 1982). In the case of DAOCS, some data sets showed peaks in the Patterson maps at special positions. The corresponding heavy-atom site could not be refined and recollecting the data from different crystals did not reproduce the sites.

(v) Poor phase information from a selenomethionine derivative, even though the incorporation of the Se atoms into DAOCS had been clearly shown by mass spectroscopy (Lloyd *et al.*, 1999) and by a fluorescence spectrum collected on SeMet crystals.

4. Determination of the twin law and twin fractions

A very likely twin operator is a symmetry operator of a higher symmetry space group. The space group of DAOCS crystals is *R*3. As the only higher symmetry space group is *R*32, a likely twin operator is described by the additional twofold axis. This twin operator has been confirmed by the self-rotation functions for all DAOCS crystals with a twin fraction higher than 0.15 (Fig. 2).

When the twin law is known, the twin fraction α can be estimated by analysing statistics of detwinned data as a function of hypothetical values of α to find the conditions for which the detwinned data follow theoretical statistics. A number of methods have been described (Murray-Rust, 1973; Fisher &

Sweet, 1980; Rees, 1980; Yeates, 1988, 1997; Dumas *et al.*, 1999; Kahlenberg, 1999). These methods are implemented in the program *DETWIN* (Collaborative Computational Project, Number 4, 1994; Taylor & Leslie, 1998) and in a web-based twinning server (Yeates & Fam, 1999).

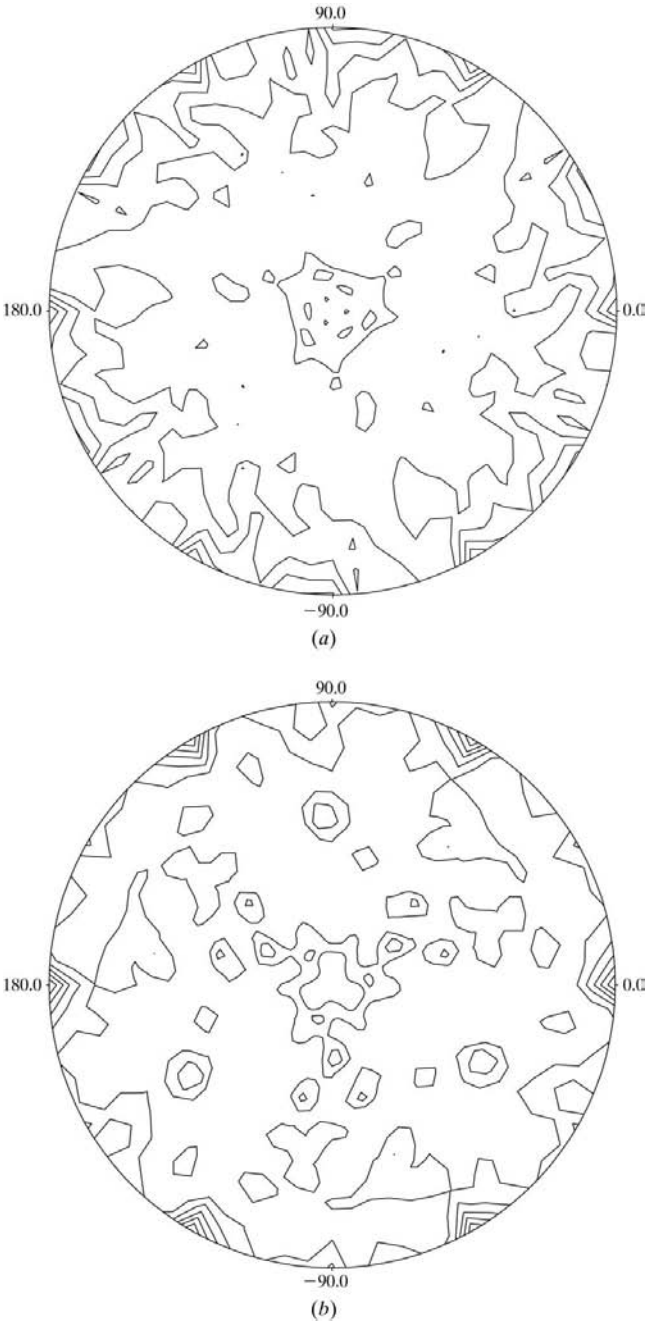

Figure 2
$\kappa = 180^\circ$ section for self-rotation functions (data from 7 to 5 Å) using the program *POLARRFN* (Collaborative Computational Project, Number 4, 1994) showing a twofold axis in addition to the *R*3 symmetry. As there is only one molecule present in the asymmetric unit, this twofold axis is the twin operator. The height of the peaks roughly corresponds to the twin fraction, so that the twofold axis is only clearly observed when the twin fraction is larger than ∼0.15. The twofold axis is perpendicular to the crystallographic threefold axis and is equivalent to the twofold axis in the higher symmetry space group *R*32. (*a*) Native-1, $\alpha = 0.11$. (*b*) Pt-2 derivative, $\alpha = 0.18$.

Apart from being utilized as a diagnostic of merohedral twinning, analysis of the cumulative intensity distribution (Stanley, 1955, 1972) gives a rough indication of the twin fraction: a higher α results in a more sigmoid curve (Fig. 1). The advantage of using this procedure is that no prior knowledge of the twin operation is required. A drawback is that the method is sensitive to the range and number of intensities used. When applied to DAOCS data, this method did not yield accurate enough data for use in phase determination.

Instead, the method first proposed by Britton (1972) and further elaborated by Fisher & Sweet (1980) was used. (1*a*) and (1*b*) can be rewritten as

$$I_1 = [I_{\rm obs,1} - \alpha(I_{\rm obs,1} + I_{\rm obs,2})]/(1 - 2\alpha), \qquad (2a)$$
$$I_2 = [I_{\rm obs,2} - \alpha(I_{\rm obs,1} + I_{\rm obs,2})]/(1 - 2\alpha). \qquad (2b)$$

The application of a twin fraction higher than its true value will result in several detwinned intensities with a negative value. From monitoring the number of negative intensities as a function of the hypothetical α, the actual twin fraction can be extrapolated from the intersection of a fitted straight line with the *x* axis (Fig. 3). If necessary, this curve can be sharpened by omitting low-intensity data (Igarashi *et al.*, 1997). With this method, an initial estimation of the twin fraction was made for DAOCS crystals. The DAOCS crystals display a high degree of twinning, typically between 0.2 and 0.5, with only approximately one out of 10 crystals showing a twin fraction less than 0.25.

5. Trials to determine the DAOCS structure using MR

Since DAOCS shows weak sequence homology with isopenicillin N synthase (IPNS; Roach *et al.*, 1995), extensive trials were performed to determine the structure by molecular replacement with the known IPNS structure as a search model, using data from a crystal with an estimated twin fraction of 0.11. Initially, these were attempted with data that were not corrected for merohedral twinning, producing no conclusive results. Subsequent trials with detwinned data and using phased translation searches (phases from the SeMet dispersive data, see below) did not produce a clear solution either. This indicates that this failure is more likely owing to the low level of homology between the structures than to the presence of merohedral twinning.

6. Structure determination by MIR methods

Although it has been shown that protein structures can be solved by MIR methods from twinned data (Yeates & Rees, 1987), this has never been reported in practice and has not been implemented in any of the current phasing programs. If instead the data are detwinned, this will generate an error which will increase with the twin fraction. Particularly for highly twinned crystals, there is a danger that the heavy-atom information will be lost.

All MIR structures before DAOCS were based on data with twin fractions less than 0.10. For DAOCS, the twin fraction varies considerably between data sets and for most crystals it is higher than 0.3. This variation in twin fraction poses another problem: the magnitude of the errors caused by detwinning depends on α. Therefore, features found in the Patterson map might be caused by a difference in twin fraction. To minimize this risk, we collected data sets with identical twin fractions by using MAD data from a selenomethionine derivative. As even in one crystal the twin fraction can vary considerably within different parts, we exposed the same part of the crystal when collecting the three data sets. Owing to technical reasons, it was not possible to collect data to the diffraction limit of the crystal.

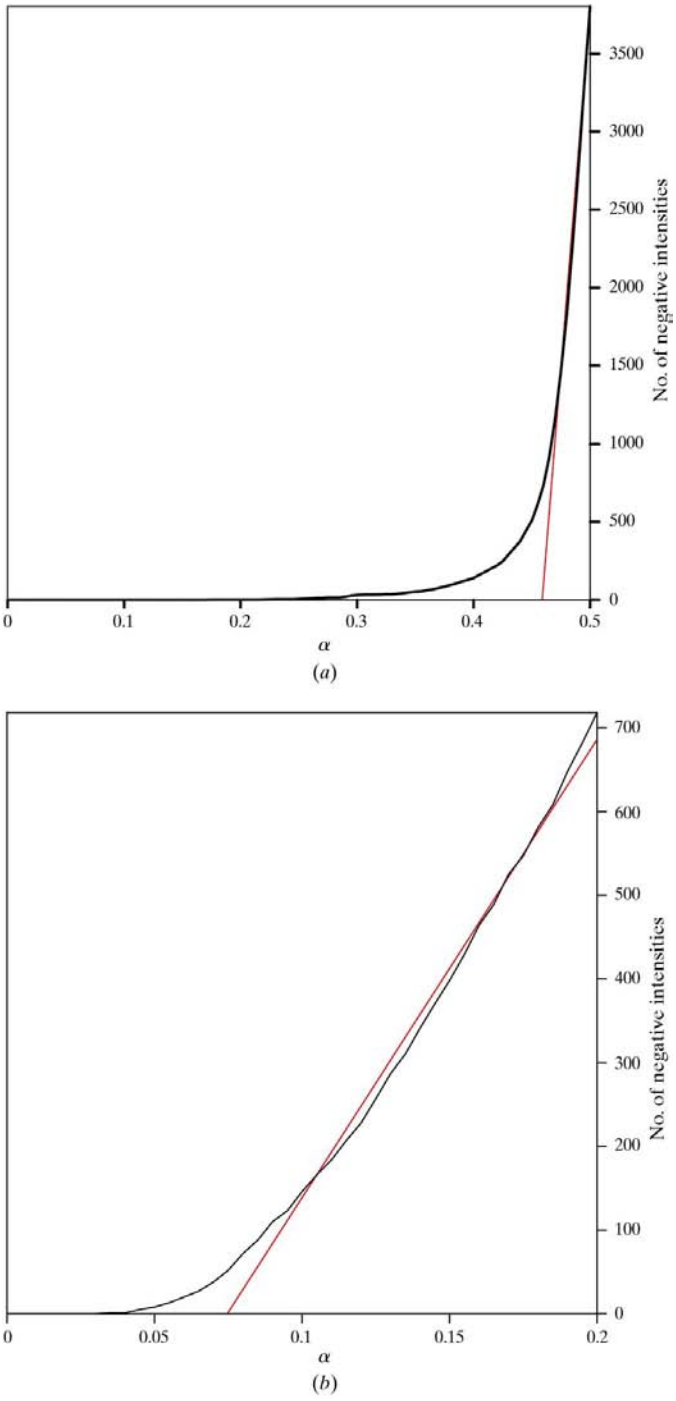

Figure 3
Britton plot (Britton, 1972; Fisher & Sweet, 1980), showing the number of negative intensities as a function of the assumed twin fraction. An estimation of the twin fraction that is higher than its real value will result in several detwinned intensities with a negative value. The monitoring of the number of negative intensities as a function of hypothetical α allows the actual twin fraction to be extrapolated (red line). (*a*) A crystal with high twin fraction (Xe-2, α = 0.45). (*b*) A crystal with low twin fraction (SeMet, α = 0.07).

Table 2
Refinement statistics.

Refinement step	No. of protein atoms	No. of solvent molecules	*R* factor	R_{free}	Twin fraction (*SHELXL*)
Unrefined model	1695	0	n.a.	n.a.	n.a.
REFMAC	2152	146	0.205	0.248	n.a.
SHELXL (detwinned reflections)	2165	160	0.202	0.239	n.a.
SHELXL (twinned intensities)	2187	160	0.167	0.184	0.175
Anisotropic temperature factors	2187	160	0.133	0.156	0.163
Riding H atoms	2187	160	0.129	0.150	0.161

We estimated the twin fraction for the selenomethionine data sets using Britton plots (Britton, 1972; Fisher & Sweet, 1980) and used this value to detwin the data. From the dispersive differences between data collected at the inflection point and the remote wavelength (Table 1) a Patterson map was calculated, which showed the positions of two heavy-atom scatterers (Fig. 4*a*). Subsequent refinement using the program *SHARP* (de La Fortelle & Bricogne, 1997) yielded three more sites. These five sites were used to obtain the twin fraction for both data sets more accurately. We evaluated the twin fractions by monitoring the statistics for heavy-atom refinement as a function of imposed twin fraction. The statistics of refinement showed a clear and consistent optimum for α = 0.057 (Fig. 5), which matches the value obtained from the Britton plot (α = 0.07).

With the twin fraction for the SeMet data exactly known, we were able to refine the twin fraction of the native data set (Native-1) in a similar fashion, using the SeMet data from the inflection point as a derivative. Again, the twin fraction obtained this way was consistent with the value obtained from the Britton plot (Table 2). Further potential derivative data sets were subsequently detwinned using Britton plots. Difference Fourier maps, using the phases from the SeMet data, showed that two additional derivatives (a platinum and a xenon derivative, Pt-1 and Xe-1) contained heavy-atom scatterers specifically bound. These derivatives probably went unnoticed before because of their high twin fractions (0.33 and 0.45, respectively). For both data sets no consistent optimum could be found by refining their heavy-atom parameters, which might be because of these high twin fractions. Despite this, the resulting electron-density map was much improved (Fig. 6*a*); the figure of merit increased from 0.34 to 0.42.

Since detwinning the Xe-1 and Pt-1 derivatives with their high twin fractions introduced large errors in the data, we

performed an extensive search for heavy-atom derivatives from crystals with less high twin fractions. We recollected both the platinum and the xenon derivative from crystals with twin fractions lower than 0.3 (Pt-2 and Xe-2; Table 1). For these data sets, optima for the twin fractions could be found in the *SHARP* refinement (Table 1). The Pt-2 and Xe-2 sites were refined against Native-1 using the SeMet phases. These MIR phases were improved by solvent flipping using *SOLOMON* (Abrahams & Leslie, 1996). Although the Pt-2 and Xe-2 data sets cannot be considered to have low twin fractions (0.19 and 0.27, respectively), their use instead of the very highly twinned Pt-1 and Xe-1 data improved the quality of the electron-density map considerably (Fig. 6*b*) and raised the figure of merit to 0.47.

A new native data set collected to higher resolution (Native-2, Table 1), allowed the phases to be extended to 1.3 Å using *ARP*/*REFMAC*. An electron-density map on a grid of approximately 0.4 Å was used to construct a free-atom model as described for the *wARP* procedure (Perrakis *et al.*, 1997). This model was improved by 150 cycles alternated with refinement with a maximum-likelihood target as implemented in *REFMAC* (Murshudov *et al.*, 1997) and automated model update with *ARP* (Lamzin & Wilson, 1993). This reduced the *R* factor to 0.185 for the final free-atom model. Using the

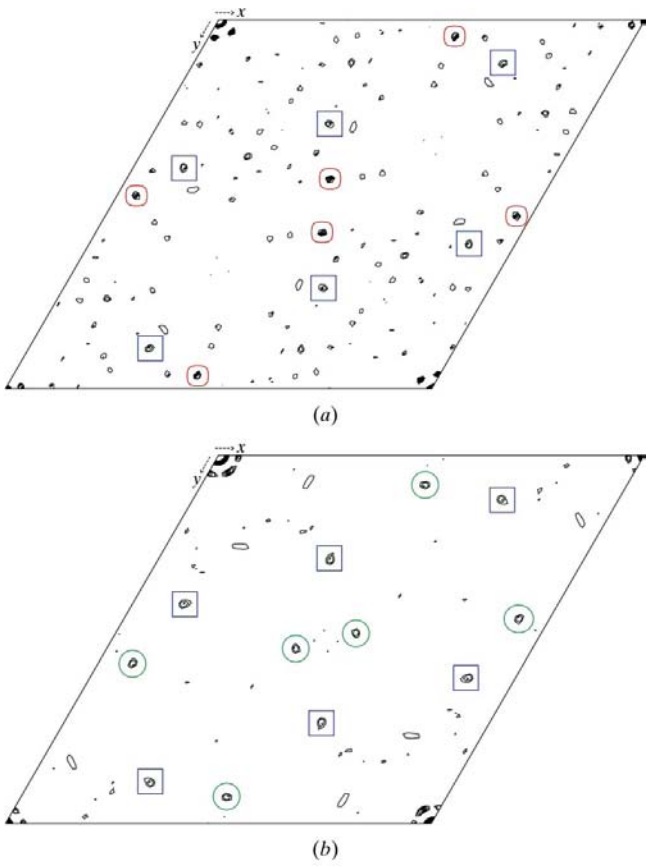

Figure 4
Harker section (w = 0) for the SeMet data. (*a*) Dispersive Patterson function after detwinning using the inflection point and remote data of the MAD data set. The map clearly shows two sets of peaks, marked in blue and red, which are related to two Se sites. Mass spectroscopy showed the presence of six Se atoms per DAOCS molecule. Heavy-atom refinement in *SHARP* suggested five selenium sites; subsequent structure refinement showed the sixth methionine side chain to be disordered. (*b*) Anomalous Patterson function using data which were detwinned with the *DE-TWIN* program. This Patterson function also shows two sets of peaks related to Se sites, marked in blue and green, whereas the anomalous Patterson function without detwinning only reveals one set (data not shown). The blue sets in (*a*) and (*b*) refer to the same Se site.

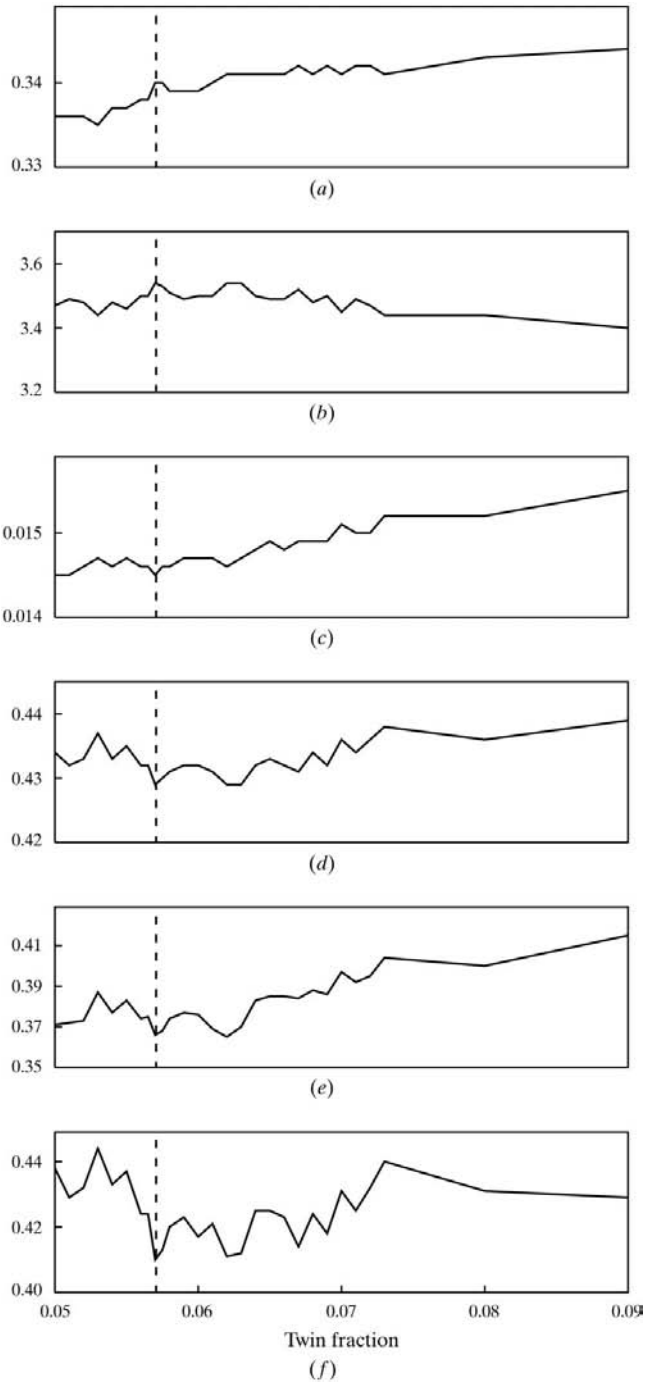

Figure 5
Variation of the statistics for refinement of the selenium sites as a function of the twin fraction used for detwinning the data. (*a*) Figure of merit. (*b*) Phasing power. (*c*) $R_{\rm Cullis}$. (*d*) $R_{\rm Kraut}$. (*e*) Lack of isomorphism (lack of closure). (*f*) Lack of isomorphism (model parameters). The statistics show a weak but consistent optimum for α = 0.057.

procedure described in Perrakis *et al.* (1999), we were able to automatically trace ten main-chain fragments consisting of four to 57 residues each and comprising a total of 262 residues out of the 311 expected. These fragments where automatically 'docked' into their corresponding place in the known amino-acid sequence. 163 side chains were then automatically constructed and the fragments were assembled into a single molecule.

7. The influence of twinning on anomalous diffraction

A space group general detwinning program (*DE-TWIN*) was developed which includes an option to take Friedel mates into account and to detwin these. The program was tested on the MAD data from the selenomethionine derivative. The detwinning proved to be sensitive enough to identify two selenium sites from the anomalous Patterson map (Fig. 4*b*). The program is available on request from SR.

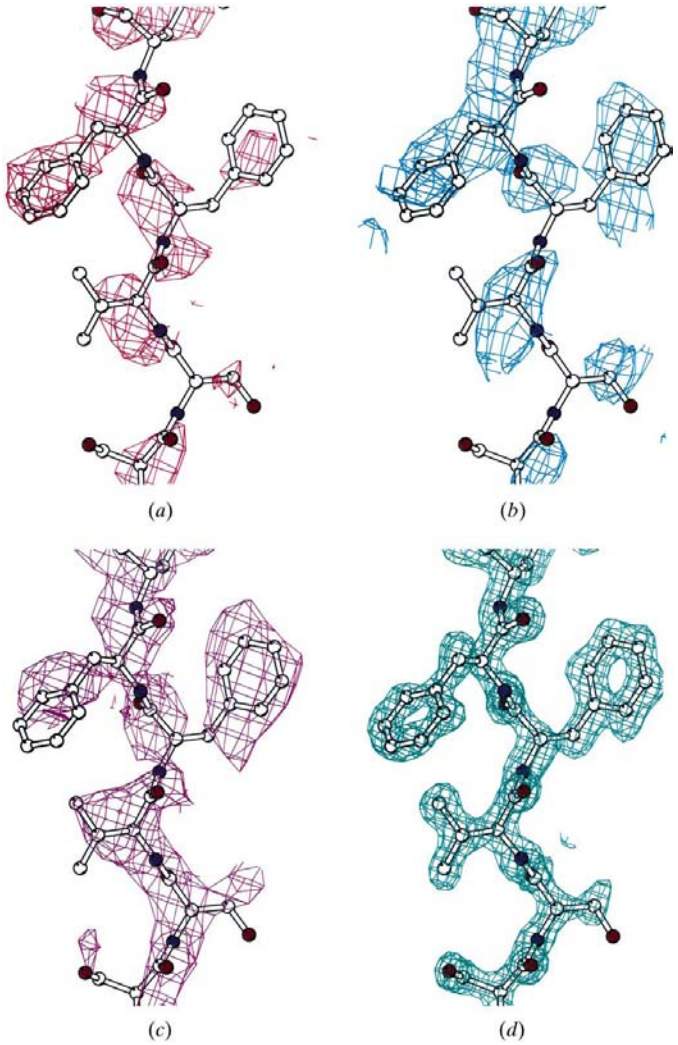

Figure 6
Examples of electron density during the various stages of structure determination and refinement, drawn with a modified version of *MOLSCRIPT* (Esnouf, 1997). (*a*) MIR phases after solvent flipping using SeMet data (twin fraction 0.07; figure of merit 0.34). (*b*) MIR phases after solvent flipping using SeMet, Native-1, Xe-1 and Pt-1 data (twin fractions 0.07, 0.11, 0.45 and 0.33, respectively; figure of merit 0.42). (*c*) Similar to (*b*), but using Xe-2 and Pt-2 (twin fractions 0.27 and 0.19, respectively) instead of Xe-1 and Pt-1 (figure of merit 0.47). (*d*) After phase extension using data to 1.3 Å from Native-2 and starting with the phases from (*c*).

8. Refinement

For the refinement of the structure, there is a choice to use either twinned (composite) or detwinned data. Since detwinning introduces errors, the magnitude of which is dependent on the size of α, refinement against detwinned data may be unstable (Jameson, 1982). An additional disadvantage with detwinning is that reflections whose twin-related partners are missing cannot be detwinned and will therefore not be used. Refinement against twinned data does not suffer from these particular disadvantages and introduces only one extra parameter to refine, the twin fraction α. Refinement using twinned data has been implemented in *SHELXL* (Herbst-Irmer & Sheldrick, 1998) and subsequently also in *CNS* (Brunger *et al.*, 1998).

To evaluate refinement of a protein structure against either twinned or detwinned data, refinement of DAOCS was started with the two alternatives in parallel. The model was separately refined against twinned data using *SHELXL* (Sheldrick & Schneider, 1997) and against detwinned data using *REFMAC* (Murshudov *et al.*, 1997). To reduce bias on the free *R* factor, especially when refining against twinned data, the test set was chosen such that pairs of twin-related reflections were either both in the working set or both in the test set.

Refinement against twinned data was unstable in the initial stages, which resulted in deteriorating quality of the maps and fluctuations in the value for the twin fraction: whereas the Britton plot, which is based only on the data and not on the model, gave a twin fraction of only 0.14, refinement with *SHELXL* allowed it to increase to 0.264. Refinement against twinned data was therefore abandoned, continuing only refinement against detwinned data using *REFMAC*. In this way, an additional 15 residues were modelled using *O* (Jones *et al.*, 1990). Once refinement using *REFMAC* had converged (progress of refinement is shown in Table 2), refinement with *SHELXL* was resumed. The model was first refined against detwinned structure factors after convergence against twinned intensities, refining the twin fraction. As refinement progressed, multiple conformations were modelled where appropriate and options for refinement of anisotropic *B* factors and riding H atoms were used (Table 2). The final *R* factor and $R_{\rm free}$ are 0.129 and 0.150, respectively; the refined twin fraction is 0.161.

9. Molecular interactions and merohedral twinning in DAOCS

The main chain of DAOCS folds into a distorted jelly-roll motif with flanking helices (Fig. 7*a*), with the active site buried within the central core of β-strands (Valegård *et al.*, 1998; Andersson *et al.*, 2001). In the crystal, residues 308–311 (the carboxy-terminal arm) of the apo enzyme participate in an

intermolecular contact. The arm penetrates the active site of the neighbouring molecule in a cyclic fashion, forming a crystallographic trimer (Fig. 7*b*). Dynamic laser light-scattering experiments on the apo enzyme (Lloyd *et al.*, 1999) indicate that the trimeric structure is present also in solution at concentrations used in the crystallization experiments.

The DAOCS apo enzyme crystallizes in space group *R*3. Since in this space group the origin is not fixed along the *z* axis, the exact position of twin molecules cannot be determined, so that interactions between molecules from different twin components cannot be analysed. However, some considerations regarding the size of the twin domains and the interactions between them can still be made. The building blocks of the twin domains can be individual molecules, trimers, stacks of trimers or layers of trimers. One possibility, the single molecule as a building block, is ruled out by the presence of strong interactions in the crystallographic trimer (Lloyd *et al.*, 1999). As the trimer has a fairly symmetrical shape, a 180° rotation around the twin axis leaves the trimer occupying approximately the same space (Figs. 7*b* and 7*c*) and one could imagine that the flipping of trimers, conceivably in stacks, is a possible event. When considering the crystal packing, it is observed that the trimers form tightly packed sheets in the *xy* plane of the crystal lattice, with few interactions along the *z* axis (Fig. 8). Therefore, flipping of whole layers or multiple stacks of layers, affecting only the interactions in the *z* direction, is the most likely possibility of formation of twin domains.

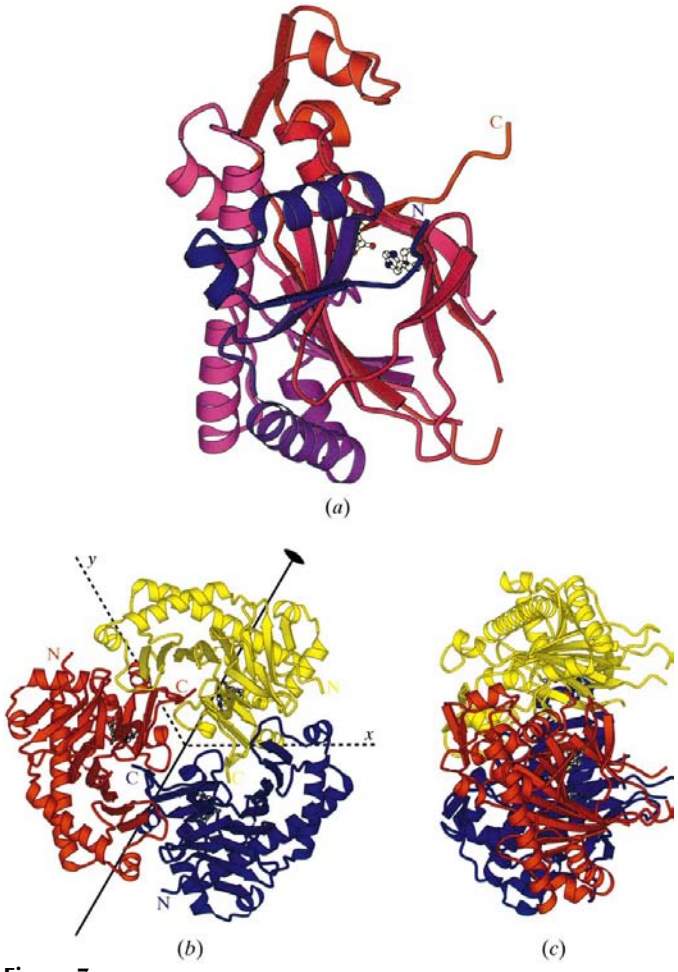

Figure 7
(*a*) The structure of DAOCS, color-ramped from blue at the N-terminus to red at the C-terminus. (*b*) The crystallographic trimer, viewed along the *z* axis. The *x* and *y* axes of the (hexagonal) crystal are shown, as well as the twofold axis along the diagonal that corresponds to the twin operator. (*c*) Edge-on view of the crystallographic trimer.

10. Discussion

We have described the determination of useful MIR phases from data from merohedrally twinned crystals. The presence of twinning by merohedry can most easily be detected by looking at intensity statistics or by searching for additional symmetry. The cumulative intensity distribution is a much more sensitive indicator for twinning than the self-rotation function; for instance, the initial native data set of DAOCS

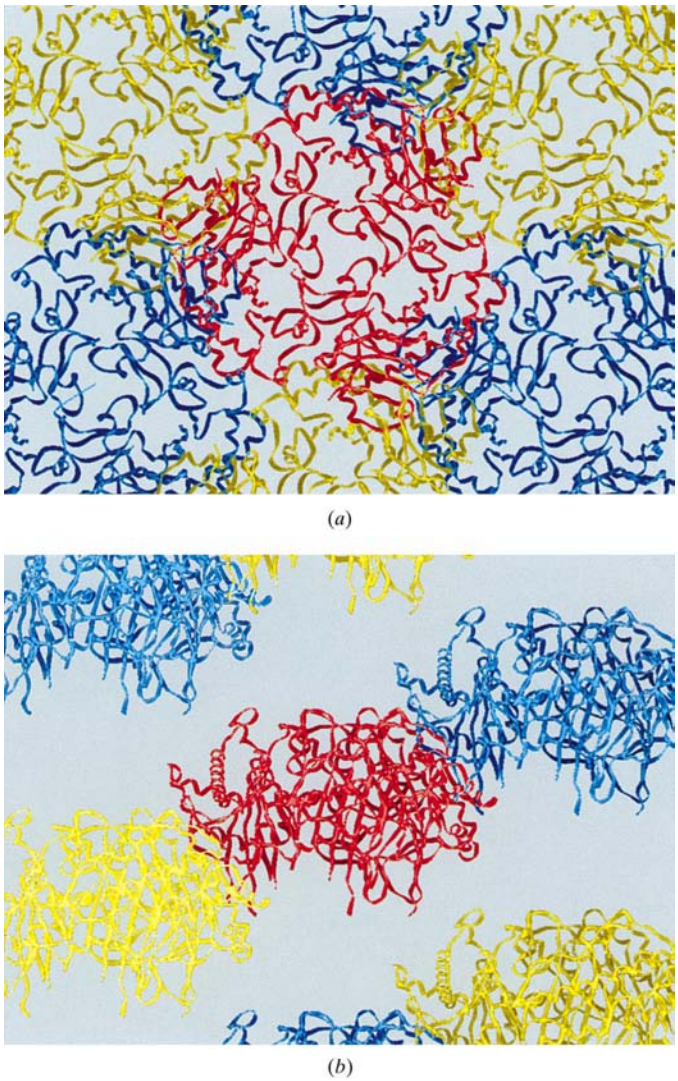

Figure 8
Crystal packing in *R*3 crystals of DAOCS. Each coloured unit represents a crystallographic trimer. (*a*) Looking in the *xy* plane. (*b*) View of the *xz* plane.

Native-1, which has a twin fraction of 0.11, does not show a clear twofold axis (Fig. 2*a*), whereas the cumulative intensity distribution already gives a clear indication of twinning for data with a twin fraction of 0.07 (Fig. 1*b*) or less. Using the cumulative intensity distribution, we have since discovered a number of merohedral twins by analysing 'problem crystals' in our laboratory.

The corrections based on Britton plots were sufficient to provide crude initial phases for a first model. Even the very highly twinned derivatives Pt-1 (α = 0.33) and Xe-1 (α = 0.45) had reasonable phasing powers (1.3 and 0.8, respectively) and helped to raise the figure of merit from 0.34 (SeMet phases only) to 0.42. The subsequent search for crystals with lower twin fractions, leading to Pt-2 (α = 0.19) and Xe-2 (α = 0.27), further increased the figure of merit to 0.47 and improved the electron density considerably (Fig. 6*b*). It should be emphasized that although it is possible to use data from very highly twinned crystals, using crystals with a lower twin fraction is obviously highly recommendable. From our data it appears that twin fractions lower than about 0.3 significantly facilitate the structure solution.

The use of the selenomethionine derivative had two main advantages. Firstly, the heavy atom was firmly integrated in the protein and its presence could be confirmed by independent means, *e.g.* mass spectrometry. Therefore, the poor phase information from this derivative could not be caused by the lack of (specific) binding of the heavy atom. Secondly, the possibility of using the data in dispersive mode, assigning a data set at one wavelength as 'native' and a second data set at a different wavelength as 'derivative', eliminated any errors arising from differences in twin ratios between the data sets. Fig. 5 shows that the method is sensitive enough to allow refinement of the twin fraction using the statistics obtained from heavy-atom refinement. However, as the values obtained from Britton plots are very close to the refined values, this additional refinement mainly played a confirmative role. It is worth noting that although the use of the data in anomalous mode was not practically possible in this case, it has been shown to be possible in a recent paper (Yang *et al.*, 2000).

For initial refinement with an incomplete model, it appears to be preferable not to refine the twin fraction. Using the partial model obtained from autotracing after *ARP*/*REFMAC*, refinement against twinned intensities with *SHELXL* increased the twin fraction to almost double the value suggested by the Britton plot. These observations are unlike those made by Jameson from refinement of inorganic structures (Jameson, 1982). It seems that the value of the twin-fraction refinement compensates for errors that are unrelated to twinning, *e.g.* the use of an incomplete model. The erroneous estimation of α appears to result in model bias; uninterpreted electron density deteriorates, even at 1.3 Å resolution. On the other hand, it is also conceivable that the model bias reflects the difference in the methods of refinement (least squares *versus* maximum likelihood), rather than differences in errors of the data sets.

In the later stages of refinement, the use of twinned intensities gave a stable estimate of α and we continued by using *SHELXL*. Even then, refinement of the twin fraction was dependent on the model. When using the final DAOCS model, adding refinement of anisotropic temperature factors decreased α from 0.175 to 0.163 and adding riding H atoms lowered it further to 0.161. This is still higher than the value obtained from the Britton plot. A possible reason for the discrepancy is that the final model does not contain the entire protein, as three loops (32 out of 311 residues) are disordered.

Twinning by merohedry represents an often unexpected obstacle to structure determination by X-ray crystallography which requires special attention. Because of the complete overlap of reflections it probably goes undetected in many cases, leading to difficulties in determining the initial phases or to problems with refinement. Here, we have shown that it is possible to use data from twinned crystals, even highly twinned crystals, to determine a protein structure by MIR.

We thank Gerard Kleywegt for providing a program to produce Britton plots and detwin data, and Andy Thompson for discussion. This work was supported by an EU-BIOTECH grant and by the Swedish Research Councils NFR and SJFR.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Andersson, I. & Brändén, C.-I. (1984). *J. Mol. Biol.* **172**, 363–366.
Andersson, I., Terwisscha van Scheltinga, A. C., Ranghino, G. & Valegård, K. (2001). *Handbook of Metalloproteins*, edited by A. Messerschmidt, R. Huber, T. L. Poulos & K. Weighart, pp. 643–651. Chichester: John Wiley & Sons.
Araki, T. (1991). *Z. Kristallogr.* **194**, 161–191.
Breyer, W. A., Kingston, R. L., Anderson, B. F. & Baker, E. N. (1999). *Acta Cryst.* D**55**, 129–138.
Britton, D. (1972). *Acta Cryst.* A**28**, 296–297.
Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Buerger, M. J. (1960). *Crystal Structure Analysis*. New York: John Wiley & Sons.
Catti, M. & Ferraris, G. (1976). *Acta Cryst.* A**32**, 163–165.
Chandra, N., Acharya, K. R. & Moody, P. C. E. (1999). *Acta Cryst.* D**55**, 1750–1758.
Cheah, E., Carr, P. D., Suffolk, P. M., Vasudevan, S. G., Dixon, N. E. & Ollis, D. L. (1994). *Structure*, **2**, 981–990.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dumas, P., Ennifar, E. & Walter, P. (1999). *Acta Cryst.* D**55**, 1179–1187.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.
Feig, A. L. & Lippard, S. J. (1994). *Chem. Rev.* **94**, 759–805.
Fisher, R. G. & Sweet, R. M. (1980). *Acta Cryst.* A**36**, 755–760.
Fisher, R. G., Woods, N. E., Fuchs, H. E. & Sweet, R. M. (1980). *J. Biol. Chem.* **255**, 5082–2089.
Friedel, G. (1926). *Leçons de Cristallographie*. Nancy/Paris/Strasbourg: Berger-Leurault.
Gomis-Rüth, F. X., Fita, I., Kiefersauer, R., Huber, R., Avilés, F. X. & Navaza, J. (1995). *Acta Cryst.* D**51**, 819–823.
Grainger, C. T. (1969). *Acta Cryst.* A**25**, 427–434.
Herbst-Irmer, R. & Sheldrick, G. M. (1998). *Acta Cryst.* B**54**, 443–449.
Howells, E. R., Phillips, D. C. & Rogers, D. (1950). *Acta Cryst.* **3**, 210–214.

Igarashi, N., Moriyama, H., Mikami, T. & Tanaka, N. (1997). *J. Appl. Cryst.* **30**, 362–367.
Jameson, G. B. (1982). *Acta Cryst.* A**38**, 817–820.
Jones, T. A., Bergdoll, M. & Kjeldgaard, M. (1990). *Crystallographic and Modelling Methods (in Molecular Design)*, edited by C. Bugg & S. Ealick, pp. 189–190. New York: Springer Verlag.
Kahlenberg, V. (1999). *Acta Cryst.* B**55**, 745–751.
Kleywegt, G. J. & Jones, T. A. (1996). *Acta Cryst.* D**52**, 826–828.
Knight, S., Andersson, I. & Brändén, C.-I. (1990). *J. Mol. Biol.* **215**, 113–160.
Koch, E. (1995). *International Tables for Crystallography*, Vol. C, edited by A. J. C. Wilson, pp. 10–14. Dordrecht: Kluwer Academic Publishers.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Lloyd, M. D., Lee, H.-J., Zhang, Z.-H., Baldwin, J. E., Schofield, C. J., Charnock, J. M., Garner, C. D., Hara, T., Terwisscha van Scheltinga, A. C., Valegård, K., Hajdu, J., Andersson, I., Danielsson, Å. & Bhikhabhai, R. (1999). *J. Mol. Biol.* **287**, 943–960.
Luecke, H., Richter, H. T. & Lanyi, J. K. (1998). *Science*, **280**, 1934–1937.
Murray-Rust, P. (1973). *Acta Cryst.* B**29**, 2559–2566.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Perrakis, A., Morris, R. & Lamzin, V.S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* D**53**, 448–455.
Que, L. Jr & Ho, R. Y. N. (1996). *Chem. Rev.* **96**, 2607–2624.
Redinbo, M. R. & Yeates, T. O. (1993). *Acta Cryst.* D**49**, 375–380.
Rees, D. C. (1980). *Acta Cryst.* A**36**, 578–581.
Rees, D. C. (1982). *Acta Cryst.* A**38**, 201–207.
Rees, D. C. & Lipscomb, W. N. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 277–280.
Reynolds, R. A., Remington, S. J., Weaver, L. H., Fisher, R. G., Anderson, W. F., Ammon, H. L. & Matthews, B. W. (1985). *Acta Cryst.* B**41**, 139–147.
Roach, P. L., Clifton, I. J., Fülop, V., Harlos, K., Barton, G., Hajdu, J., Andersson, I., Schofield, C. J. & Baldwin, J. E. (1995). *Nature (London)*, **375**, 700–704.
Schofield, C. J. & Zhang, Z. (1999). *Curr. Opin. Struct. Biol.* **9**, 722–731.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Stanley, E. (1955). *Acta Cryst.* **8**, 351–352.
Stanley, E. (1972). *J. Appl. Cryst.* **5**, 191–194.
Taylor, H. O. & Leslie, A. G. W. (1998). *CCP4 Newsl.* **35**, 9.
Valegård, K., Terwisscha van Scheltinga, A. C., Lloyd, M. D., Hara, T., Lee, H. J., Subramanian, R., Perrakis, A., Thompson, A., Baldwin, J. E., Schofield, C. J., Hajdu, J. & Andersson, I. (1998). *Nature (London)*, **394**, 805–809.
Yang, F., Dauter, Z. & Wlodawer, A. (2000). *Acta Cryst.* D**56**, 959–964.
Yeates, T. O. (1988). *Acta Cryst.* A**44**, 142–144.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.
Yeates, T. O. & Fam, B. C. (1999). *Structure*, **7**, R25–R29.
Yeates, T. O. & Rees, D. C. (1987). *Acta Cryst.* A**43**, 30–36.