

MIR phasing using merohedrally twinned crystals

Anke C. Terwisscha van
Scheltinga,^{a*} Karin Valegård,^b
Janos Hajdu^b and Inger
Andersson^a

^aDepartment of Molecular Biology, Swedish
University of Agricultural Sciences, Box 590,
S-751 24 Uppsala, Sweden, and ^bDepartment of
Biochemistry, Uppsala University, Box 576,
S-751 23 Uppsala, Sweden

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

Merohedral twinning is a crystal-growth disorder that seriously hinders the determination of macromolecular crystal structures by isomorphous replacement. The strategies used in the structures solved so far are discussed. Several methods can be used to determine the extent of twinning, the twin fraction and to detwin the data. Accurate determination of the twin fraction by analysing heavy-atom refinement statistics is possible, but only influences the resulting phases slightly. It seems more crucial to restrict the variation in twin fractions between data sets, either by making the twin fractions of some data sets artificially higher or by screening crystals to obtain data with a low twin fraction.

Received 15 April 2003

Accepted 14 September 2003

1. Introduction

Solving a structure by isomorphous replacement can be hampered by non-isomorphism. In such a case, the *R* factors between the data sets are considerably higher than would be expected from the contribution of the heavy-atom substitution. Most of the time, this is caused by a heavy-atom-induced change in the protein conformation. However, in certain cases non-isomorphism can also occur between two native data sets. One possible explanation for this phenomenon may be merohedral twinning, a crystal-growth disorder. If it is not possible to eliminate the disorder by improving the crystallization conditions, it will be necessary to find a way to use the data from twinned crystals.

Twinned crystals are intergrown in such a way that one or more of the lattice directions of the twin components are parallel. For twinning by merohedry, the lattices of the distinct domains in the crystal superimpose completely in three dimensions, which is only possible when the lattice has a higher symmetry than the crystal itself. The relation between the lattices is described by the twin operation. Since the twin operation is part of the lattice symmetry, it must be a crystallographic symmetry operation of a higher symmetry space group. In the special case of pseudo-merohedral twinning, the crystal symmetry is higher than the space-group symmetry because of special unit-cell dimensions (*e.g.* if two non-identical axes are similar in length; Andersson & Brändén, 1984; Neidhart *et al.*, 1987; Knight *et al.*, 1990; Yang *et al.*, 2000).

For three-dimensional protein crystals, the most common type of merohedral twinning generates only two twin components (hemihedral twinning) and the twin operation that relates them is a twofold rotation axis. By convention, the fraction of the minor component is called the twin fraction, α .

Since the lattices of the twin domains are identical, the reciprocal lattices overlap, 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 'normal' and the anomaly is therefore easy to overlook; it is impossible to recognize this type of twinning directly from a diffraction image.

Assuming that the twin domains are larger than the coherence length of the X-ray beam, they will scatter independently. In this case, 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 operation,

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

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

Once the twin operation and α are known, the contributions of each twin can be separated, *i.e.* the data can be detwinned. A problem inherent in this correction is that it amplifies the experimental errors: the error increases as α increases. According to the error analysis given by Fisher & Sweet (1980), the standard deviation of a detwinned intensity becomes that of the measured intensity multiplied by 1.1 for $\alpha = 0.1$, 1.8 for $\alpha = 0.3$ and 5.5 for $\alpha = 0.45$, and becomes infinite when α is 0.50 (perfect twinning). Consequently, perfect twins cannot be detwinned without additional information (Redinbo & Yeates, 1993).

Until the late 1990s, merohedral twinning had only been reported for a few protein crystals. Recently, several reviews increased alertness for the disorder (Yeates, 1997; Chandra *et al.*, 1999; Dumas *et al.*, 1999; Yeates & Fam, 1999) and many cases have since been described. On the other hand, the use of twinned crystals for MIR structure solution has not increased accordingly: only six structures have been reported and two of these appeared as early as 1980 (Fisher & Sweet, 1980; Rees & Lipscomb, 1980; Cheah *et al.*, 1994; Igarashi, Moriyama, Mikami *et al.*, 1997; Valegård *et al.*, 1998; Declercq & Evrard, 2001).

The general strategy for these structure determinations is to detwin the data in order to extract the heavy-atom signal. Problems may arise when the twin fractions and therefore the errors upon detwinning are very high. Other problems occur when the differences between twin fractions of data sets are large, which will seriously affect difference Fourier and Patterson maps, even when the data have been corrected for twinning. Below, the strategies used to solve protein structures by isomorphous replacement from merohedrally twinned data are described. We will focus on the factors that are important to detect twinning and to arrive at a successful structure solution.

2. Carboxypeptidase A

Crystals of a carboxypeptidase A complex grew in space group $P3_2$ and showed a high degree of twinning, with twin fractions

of between 0.30 and 0.40 (Rees & Lipscomb, 1980). Since the twin fractions were quite similar, the R factors between data sets were low and a mercury-derivative data set could be obtained by merging twinned data from ten crystals. A platinum-derivative data set was collected from a single crystal with a twin fraction of 0.43. To overcome the problems caused by the high degree of twinning for all data sets, the data were processed in $P3_221$. This was possible (to 5.5 Å resolution) since a non-crystallographic twofold axis was nearly parallel to the twin operation, which in turn was equivalent to the twofold axis of the higher symmetry space group.

Yeates and Rees developed an algorithm to calculate structure-factor amplitudes and phases for a protein structure using MIR data from perfect twins, where $I_{\text{obs},1} = I_{\text{obs},2}$ and the data cannot be detwinned (Yeates & Rees, 1987). In such a case, the phasing problem becomes four-dimensional and four independent isomorphous derivatives are needed to give a unique solution for each phase.

The method was tested using the carboxypeptidase A complex. Structure factors were calculated in the resolution range 20–4.5 Å. Five different derivative data sets were calculated using structures with an Hg atom introduced at random positions. The intensities of twin-related reflections were averaged to simulate data from a perfectly twinned crystal. As expected, four or more derivatives gave a unique most probable phase solution. Fewer derivatives resulted in an ambiguous phase solution, but by averaging the degenerate solutions it was possible to estimate phases reasonably well.

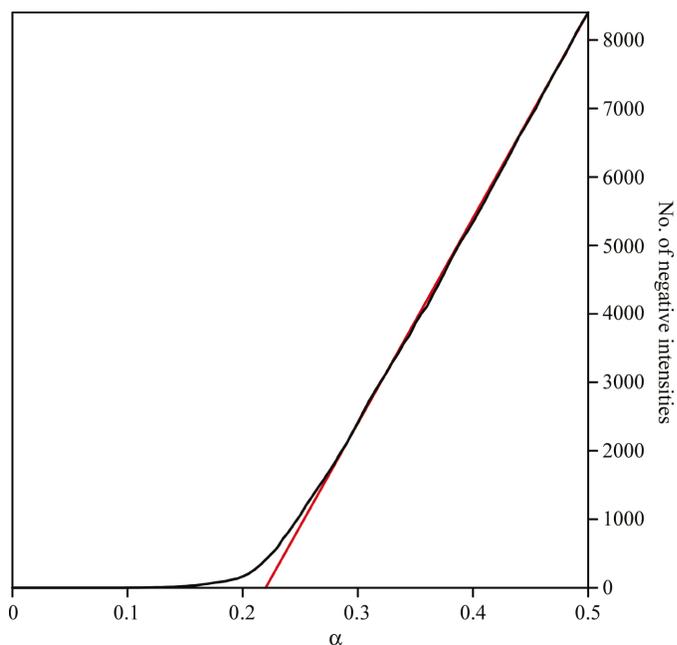


Figure 1
Britton plot (Britton, 1972; Fisher & Sweet, 1980) for a DAOCS crystal, showing the number of negative intensities as a function of the assumed twin fraction. An estimation of the twin fraction that is higher than its real value will result in several detwinned intensities with a negative value. The monitoring of the number of negative intensities as a function of the hypothetical α allows the actual twin fraction to be extrapolated (red line).

3. B-phycoerythrin

Fisher and Sweet solved the structure of B-phycoerythrin from twinned crystals (Fisher *et al.*, 1980). To assess twinning they used Britton plots (Britton, 1972), in which the number of detwinned reflections with negative intensities is given as a function of the assumed twin fraction (*e.g.* Fig. 1). The twin fractions can be estimated from extrapolation of the intersection of a fitted straight line with the x axis and can be used to detwin the corresponding data sets. Besides these plots, several other methods were tested to evaluate crystal twinning. With one crystal detwinned, the twin fractions of other data sets were optimized by minimizing the R factors between these and the detwinned data set. This method proved to be difficult owing to the shallow minima that were obtained. Better results could be obtained from the detwinned data set by pairwise comparison of twin-related reflections with the equivalent detwinned reflections. The twin fraction α was calculated for each pair of measured and detwinned reflections and the values obtained were fitted to a Gaussian function. Another approach was to assess twin fractions by minimizing the lack of closure during heavy-atom refinement. All four methods gave very similar twin fractions.

Finally, it was tested how the structure solution would be influenced when the twinning disorder was ignored. Heavy-atom refinement using twinned data resulted in a phase difference of only 22° . This small difference could be because of the low twin fractions of the native and derivative data sets, which were between 0.01 and 0.08.

4. Hydroxylamine oxidoreductase

For hydroxylamine oxidoreductase crystals, twinning was first suspected from the high R factors between data sets and a peak in the self-rotation function that varied in height between 60 and 98% of the origin peak (Igarashi, Moriyama, Mikami *et al.*, 1997). Moreover, difference Patterson maps for the two heavy-atom derivatives suggested two sets of heavy atoms for each derivative. The Patterson maps showed no cross-vectors between these sets, which were related by the twin operation.

Initially, twin fractions were estimated from Britton plots. To obtain native detwinned data, 16 data sets with twin fractions less than 0.33 were selected and weighted averages of equivalent detwinned reflections from different data sets were calculated. Subsequently, twin fractions were refined to minimize the residual in these equations.

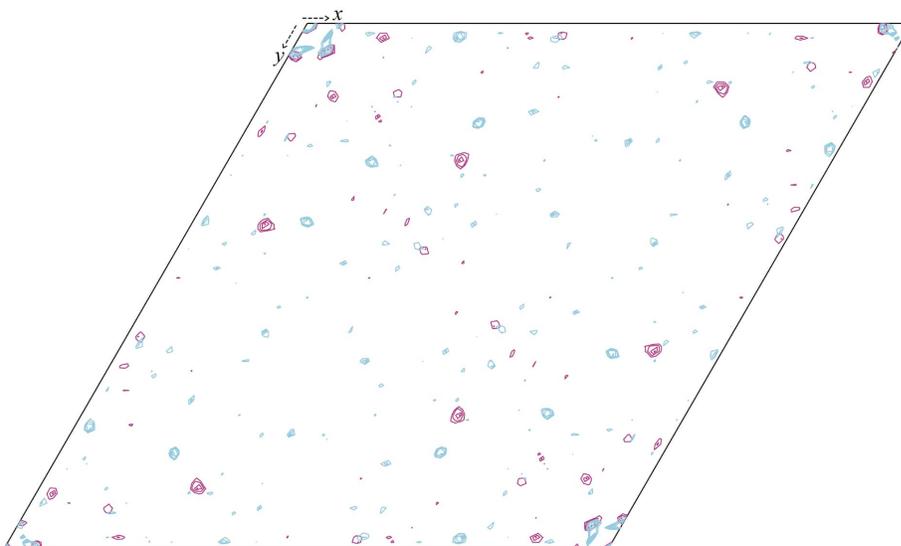


Figure 2

Harker section ($w = 0$) for the Xe derivative. The twin fraction of the derivative data is 0.45 and that of the native data is 0.14. The difference Patterson function calculated using twinned data is displayed in pink and shows a set of peaks which do not correspond to a heavy-atom site. The difference Patterson function for which detwinned data is used is coloured blue. The peaks that were present when twinned data was used have disappeared; instead, a set of peaks related to the true xenon-binding site has emerged.

Meanwhile, by collecting multiple data sets from a single (needle-shaped) crystal, it was established that the twin fraction gradually increased from 0.00 at one end of the crystal to 0.98 at the other. Knowing this, use of a narrow X-ray beam allowed the collection of data with a low twin fraction by exposing only one of the ends of the crystal to X-rays. Two heavy-atom derivatives, an $\text{Hg}(\text{OAc})_2$ derivative (two crystals, 3 Å) and a K_2PtCl_4 derivative (one crystal, 4 Å), were used and in both cases information from the anomalous signal was included. The resulting phases were of good quality, with a figure of merit of 0.616 (Igarashi, Moriyama, Fujiwara *et al.*, 1997).

5. Peroxiredoxin 5

The structure of peroxiredoxin 5 was solved by SIR (Declercq & Evrard, 2001). The crystals were monoclinic with a β angle close to 120° and could be indexed (but not merged) as C -centred orthorhombic. A native data set, a methylmercurichloride derivative and a potassium iodide derivative data set were collected. Analysis of the mercury derivative gave several sets of possible heavy-atom sites, which could not be cross-refined but gave a good isomorphous signal on their own. No calculated phases could reproduce the iodine sites suggested by the iodine Patterson map.

Further analysis revealed the data to be twinned, with the twin operation in the ac plane perpendicular to the a axis, resulting in superimposition of (h, k, l) with $(h, -k, h + l)$. Data were detwinned using Yeates' H distribution (Yeates, 1997), as implemented in the CCP4 program *DETWIN* (Taylor & Leslie, 1998). The native, mercury and iodine data had twin

fractions of 0.24, 0.42 and 0.46, respectively. The detwinned iodine data could not be used, since its high twin fraction caused the error upon detwinning to be too large. For the mercury derivative, eight heavy-atom sites could be identified and refined. Although the resulting phases only had a figure of merit of 0.22, they were good enough to discern the protein molecules in the electron-density map. Subsequent solvent flattening and eightfold averaging resulted in an interpretable map.

6. Deacetoxycephalosporin C synthase

Deacetoxycephalosporin C synthase (DAOCS) crystallized in space group *R3* with one monomer in the asymmetric unit (Valegård *et al.*, 1998; Lloyd *et al.*, 1999). The crystals had a normal morphology, looked uniform under polarized light and diffracted X-rays to beyond 1.3 Å resolution with good scaling statistics. However, a close inspection of the intensity data from these crystals revealed all possible symptoms of merohedral twinning.

The non-isomorphism was clearly detected by the high *R* factors obtained from the scaling of different data sets. While the merging *R* factor for a single data set is usually around 0.03, the scaling together of two native data sets with $\alpha = 0.01$ and 0.36, respectively, gave an *R* factor of 0.295 (Terwisscha van Scheltinga *et al.*, 2001).

In our search for heavy-atom derivatives, Patterson peaks showed up at special positions. An example is the Harker section for a putative Xe derivative shown in Fig. 2. Refinement of the site corresponding to these peaks failed and after detwinning the Patterson peaks had disappeared. These clues all pointed to some sort of non-isomorphism. To investigate this further, a selenomethionine derivative was produced. The sequence contains six methionines and although incorporation was shown by mass spectroscopy (Lloyd *et al.*, 1999), the MAD data collected from crystals of the SeMet protein were not readily interpretable.

The cumulative intensity distribution (calculated by *TRUNCATE*; Collaborative Computational Project, Number 4, 1994) showed a marked deviation from the theoretical curve. Its sigmoidal shape (Fig. 3) suggested that the crystals were twinned by merohedry. This was confirmed by analysing the self-rotation function, which showed peaks at $\kappa = 180^\circ$ consistent with the presence of an additional crystallographic twofold axis and suggesting a shift to space group *R32* (Fig. 4). Since forcing this space group upon the DAOCS data would require a solvent content of just 3%, this suggested the true space group to be *R3*, with a twin operation equivalent to the twofold axis in *R32*. The twin operation results in overlap of the (h, k, l) and $(k, h, -l)$ reflections and is equivalent to the two ways *R3* data can be indexed. Once the twin operation was determined, twin fractions were estimated from Britton plots (Britton, 1972; Fisher & Sweet, 1980). DAOCS crystals display a high degree of twinning, typically between 0.2 and 0.5, with only approximately one in ten crystals showing a twin fraction less than 0.25.

When using detwinned data in MIR phasing there is one important point to consider: detwinning of the data results in errors, the magnitude of which will depend on the twin fraction. A difference in twin fractions between data sets may result in false Patterson peaks which will interfere with the MIR analysis. We therefore started our analysis with the SeMet data, which were all collected from the same volume of the same crystal with one identical twin fraction. At the time, we considered it too difficult to detwin the anomalous data and focused on the dispersive signal. Two selenium sites could be identified directly in the Patterson map. After refinement of these peaks with *SHARP* (de La Fortelle & Bricogne, 1997), three additional sites were identified in the difference Fourier map.

To obtain a more accurate estimation of the twin fraction, we detwinned the SeMet data by varying the twin fraction in the range 0–0.15. For each of these detwinned data sets, the positions of the selenium sites were refined. The optimum values for all refinement statistics were weak but consistent for a twin fraction of 0.057, which was quite close to the value suggested from the Britton plot (0.07). In a similar fashion, the twin fraction for the native data set was determined to be 0.108: the selenium sites were refined against native data and the statistics were monitored as a function of imposed twin fraction.

All putative derivative data sets were detwinned using their twin fractions estimated from Britton plots. Analysis of the calculated difference Patterson and difference Fourier maps identified an Xe and a Pt derivative with twin fractions of 0.45 and 0.33, respectively. As mentioned before, detwinning intensities with a twin fraction of 0.45 will increase their

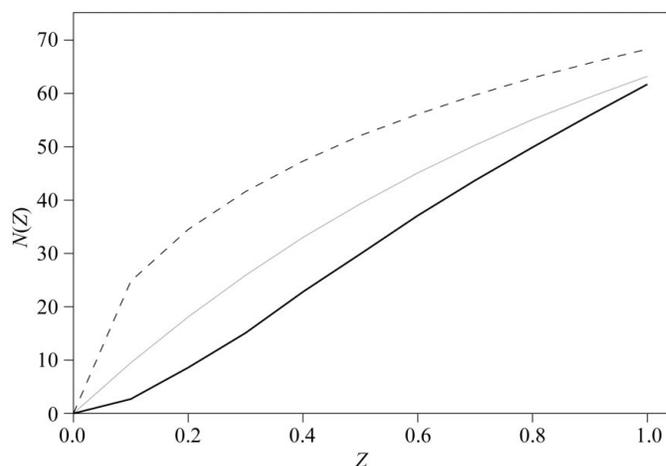


Figure 3

Cumulative intensity distribution curves for a DAOCS crystal with a twin fraction of 0.32, 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. 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.

standard deviation by a factor of 5.5. Since for the twinned Xe data the average ratio between detwinned intensities and their standard deviations is 22 and for data between 2.7 and 2.5 Å it is 10, the detwinned data still contain information. Indeed, the spurious peaks that were present in the twinned difference Patterson maps were absent (Fig. 2) and for both derivatives peaks consistent with one heavy-atom binding site were found. Although the twin fractions of both derivatives were too high to be refined by analysing MIR statistics, the use of these data resulted in an improved electron-density map. The addition of the Pt site increased the figure of merit from 0.34 to 0.39 and subsequent addition of the Xe site resulted in a figure of merit of 0.42.

Since the error introduced by detwinning increases with increasing twin fraction, we searched for crystals with lower twin fractions. After considerable screening, crystals with lower twin fraction were found for both derivatives. The twin fractions of the Xe- and Pt-derivative crystals were estimated from the corresponding Britton plots and were determined more accurately from heavy-atom statistics to be 0.271 and 0.177 for the Xe and Pt derivatives, respectively.

The use of derivatives with twin fractions lower than 0.3 improved the electron density considerably and increased the figure of merit from 0.42 to 0.47. The optimization of twin fractions using heavy-atom refinement statistics, however, hardly had any effect on the quality of the electron density,

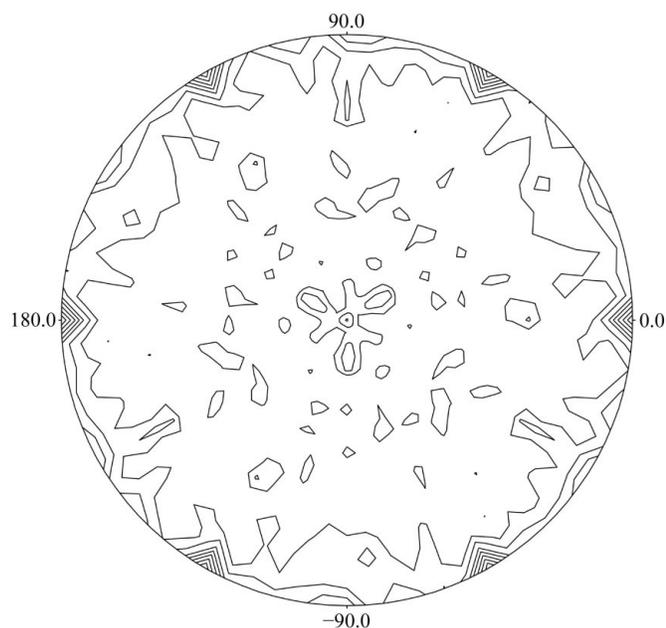


Figure 4

Section $\kappa = 180^\circ$ for the self-rotation function of a DAOCS crystal with twin fraction 0.30 (data from 7 to 5 Å) using the program *POLARRFN* (Collaborative Computational Project, Number 4, 1994) show a twofold axis in addition to the $R3$ symmetry. The twofold axis is perpendicular to the crystallographic threefold axis and is equivalent to the twofold axis in the higher symmetry space group $R32$. Since the size of the unit cell did not allow 18 molecules, non-crystallographic symmetry or the higher symmetry space group could be ruled out, identifying the axis as the twin operator.

confirming that the estimation from Britton plots was accurate enough to solve the structure of DAOCS.

7. Discussion

From the cases described here, several methods proved to be suitable for the detection of merohedral twinning. In most cases, the cumulative intensity distribution provides a sensitive means, even when the twin fraction is low. Another indication of merohedral twinning may be found in the self-rotation function, where twinning will result in additional peaks. However, care should be taken not to confuse these with peaks arising from non-crystallographic symmetry. Because the height of the peaks is roughly proportional to the twin fraction, the self-rotation function is less sensitive for low twin fractions: the peaks are not visible for DAOCS data when the twin fraction is less than 0.15 (Terwisscha van Scheltinga *et al.*, 2001).

Corrections based on Yeates' H distribution (Yeates, 1997) or Britton plots (Britton, 1972) are sufficient to give useful data for MIR phasing. Other methods described by Fisher & Sweet (1980) give similar twin fractions but rely on comparison to a non-twinned data set and are therefore less useful. For very high twin fractions and low signal-to-noise ratios, the error upon detwinning can be too large to give useful information. In this case, one could try the method described by Yeates & Rees (1987), which requires four derivatives to obtain unambiguous phases from perfect twins. In the special case where the twin operation is very close to a non-crystallographic symmetry operation, the data can be merged using the higher symmetry and at least low-resolution phases can be obtained (Rees & Lipscomb, 1980).

Twin fractions can be refined using phasing statistics. To be able to only refine one data set at a time, a non-twinned or detwinned data set is needed or, as in the case of DAOCS SeMet data, two data sets with the same twin fraction. If the twin fraction is lower than approximately 0.3, this value can be refined by looking for a consistent optimum in the heavy-atom statistics. However, this does not change the phases or statistics dramatically and is therefore optional.

High twin fractions constitute a considerable stumbling block for MIR structure solution and variable twin fractions even more so. Rees showed that the mean-square intensity change arising from heavy-atom binding can be calculated from merohedrally twinned data, provided that the differences between native and derivative data are relatively small (Rees, 1982). If, however, there is a considerable difference in the twin fractions between data sets, this may seriously affect difference Fourier and Patterson maps. For a native and derivative data set of the carboxypeptidase A complex with twin fractions of 0.10 and 0.42, respectively, a difference Patterson map of detwinned data was not interpretable. Creating perfectly twinned data for the native data set, however, resulted in a map where heavy-atom sites could be located (Rees, 1982).

In agreement with these findings, high twin fractions give few problems for MAD structure solution: the two MAD

structures determined from merohedral twins were both solved without realising the twinning disorder, even though in both cases the twin fractions of the data used were very high. The crystal structure of the capsid-stabilizing protein of bacteriophage λ was solved using data with a twin fraction of 0.36 (Yang *et al.*, 2000). A double mutant of interleukin-1 β was solved using SeMet data with a twin fraction of 0.40 (Rudolph *et al.*, 2003). Since the different data sets in MAD data collection are collected from one crystal using the same orientation, they will have one identical twin fraction.

When the MIR approach fails owing to a large variation in twin fraction, there are several options. It is possible to impose an artificially high twin fraction on the data with the lower twin fraction, as described by Rees (1982). On the other hand, crystals can be screened to obtain data sets with low twin fractions. The Xe derivative discussed above had a twin fraction of 0.45. When the data was recollected with a twin fraction of 0.27, its phasing power increased from 0.8 to 2.8 (Terwisscha van Scheltinga *et al.*, 2001). Therefore, if there are crystals available, it will be worth while to recollect a derivative with a low twin fraction.

In conclusion, it is clearly possible to solve a macromolecular structure from merohedrally twinned crystals using MIR. Determination of the twin fraction by conventional programs is accurate enough. The twin fraction may be improved by analysing heavy-atom refinement statistics, but this hardly improves the resulting phases. Most care is needed to restrict the variation in twin fractions between data sets. This is possible either by making the twin fractions of some data sets artificially higher or by screening crystals to obtain data with a lower twin fraction.

This work was supported by an EU-BIOTECH grant and by the Swedish Research Council.

References

- Andersson, I. & Brändén, C.-I. (1984). *J. Mol. Biol.* **172**, 363–366.
- Britton, D. (1972). *Acta Cryst.* **A28**, 296–297.
- Chandra, N., Acharya, K. R. & Moody, P. C. E. (1999). *Acta Cryst.* **D55**, 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.* **D50**, 760–763.
- Declercq, J.-P. & Evrard, C. (2001). *Acta Cryst.* **D57**, 1829–1835.
- Dumas, P., Ennifar, E. & Walter, P. (1999). *Acta Cryst.* **D55**, 1179–1187.
- Fisher, R. G. & Sweet, R. M. (1980). *Acta Cryst.* **A36**, 755–760.
- Fisher, R. G., Woods, N. E., Fuchs, H. E. & Sweet, R. M. (1980). *J. Biol. Chem.* **255**, 5082–2089.
- Howells, E. R., Phillips, D. C. & Rogers, D. (1950). *Acta Cryst.* **3**, 210–214.
- Igarashi, N., Moriyama, H., Fujiwara, T., Fukumori, Y. & Tanaka, N. (1997). *Nature Struct. Biol.* **4**, 276–284.
- Igarashi, N., Moriyama, H., Mikami, T. & Tanaka, N. (1997). *J. Appl. Cryst.* **30**, 362–367.
- Knight, S., Andersson, I. & Brändén, C.-I. (1990). *J. Mol. Biol.* **215**, 113–160.
- La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
- 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.
- Neidhart, D. J., Distefano, M. D., Tanizawa, K., Soda, K., Walsh, C. T. & Petsko, G. A. (1987). *J. Biol. Chem.* **262**, 15323–15326.
- Redinbo, M. R. & Yeates, T. O. (1993). *Acta Cryst.* **D49**, 375–380.
- Rees, D. C. (1982). *Acta Cryst.* **A38**, 201–207.
- Rees, D. C. & Lipscomb, W. N. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 277–280.
- Rudolph, M. G., Kelker, M. S., Schneider, T. R., Yeates, T. O., Oseroff, V., Heidary, D. K., Jennings, P. A. & Wilson, I. A. (2003). *Acta Cryst.* **D59**, 290–298.
- Taylor, H. O. & Leslie, A. G. W. (1998). *CCP4 Newsl.* **35**, 9.
- Terwisscha van Scheltinga, A. C., Valegård, K., Ramaswamy, S., Hajdu, J. & Andersson, I. (2001). *Acta Cryst.* **D57**, 1776–1785.
- 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.* **D56**, 959–964.
- 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.* **A43**, 30–36.