

Analysis and characterization of data from twinned crystals

N. Chandra,^{a,b*} K. Ravi Acharya^b
and P. C. E. Moody^a^aDepartment of Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, England, and ^bDepartment of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, England

Correspondence e-mail: nc42@le.ac.uk

Received 11 March 1999
Accepted 21 July 1999

It is difficult but not impossible to determine a macromolecular structure using X-ray data obtained from twinned crystals, providing it is noticed and corrected. For perfectly twinned crystals, the structure can probably only be solved by molecular replacement. It is possible to detect and characterize twinning from an analysis of the intensity statistics and crystal packing density. Tables of likely twinning operators and some examples are discussed here.

1. Introduction

Recent advances in the theory and practice of molecular crystallography have increased the speed and scope of structure determination enormously, but suitable crystals are still an essential pre-requisite. Twinning is a possible complication; its symptoms can be recognized at the data-processing stage and if suitably treated, the solution of the structure is possible. It is frequently observed in crystals of small molecules and is not regarded as an insoluble problem (Dunitz, 1964; Wei, 1969; Gao *et al.*, 1994).

Twinning can be described as a crystal-growth anomaly whereby the orientations of individual crystalline domains within the crystal specimen differ in such a way that their diffraction lattices overlap, either completely or partially (Redinbo & Yeates, 1993; Giacovazzo, 1992; Yeates & Fam, 1999). There is some twinning operator which can be applied to align the crystal axes of each domain (Fig. 1). For instance, a trigonal crystal could contain three blocks, *B*₁, *B*₂ and *B*₃, where the *B*₁ axes (*a*, *b*, *c*) are aligned with the *B*₂ axes (*b*, *a*, $-c$) and the *B*₃ axes ($-a$, $-b$, *c*). The real-space twinning operators [*T*_{real}] to convert *B*₂ or *B*₃ to *B*₁ would be

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \text{ and } \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix},$$

respectively. The reciprocal-space twinning operators [*T*_{recip}] are the inverse of [*T*_{real}]; so that [*h*₁*k*₁*l*₁] will overlap [*h*₂*k*₂*l*₂] [*T*_{recip}]. (In both these cases [*T*_{recip}] = [*T*_{real}], but this is not necessarily so.)

The different types of twinning have been described and classified (Donnay & Donnay, 1974; Giacovazzo, 1992). There are two families: (i) quasi-twin-lattice symmetry (QTLS) and (ii) twin-lattice symmetry (TLS) or merohedral twinning (Catti & Ferraris, 1976; Giacovazzo, 1992). The QTLS twins have two or more lattices which do not completely overlap; they give rise to multiple diffraction spots and can often be recognized under a microscope with a polarizing attachment. If only two lattice axes can be aligned, these twins are called

non-merohedral or epitaxial. This is easily recognizable by inspection of the observed three-dimensional diffraction pattern, which reveals clearly distinct interpenetrating reciprocal lattices (for details, see Liang *et al.*, 1996; Lietzke *et al.*, 1996). Usually a single lattice can be identified and the unique data integrated.

On the other hand, TLS twins are generally indistinguishable even under a powerful optical microscope and the reciprocal lattices of each domain completely overlap, giving rise to single diffraction spots. TLS twins are further divided into class I, where the twin fragments do not have the same diffracting volume (referred to as partial merohedral twinning) and the apparent lattice symmetry is the same as the true Laue symmetry, and class II, where the twin fragments have identical volumes (perfect merohedral twinning) and the apparent twin lattice symmetry appears greater than the true Laue symmetry. The independent twin operation imposes an extra relationship in the lattice symmetry which is not part of the Laue symmetry of the single crystal (Yeates, 1997). (This of course is only possible if the Laue group is capable of supporting a higher order of symmetry, and hence twinning usually occurs in $P4$, $P3$, $P6$ or cubic systems.)

The twinning can either involve two domains (hemihedral), four domains (tetartohedral) or, for some centric forms, eight domains (ogdohedral).

To deal with twinned crystals it is essential to (i) identify the nature of the twin operator and (ii) estimate the volume ratio (the twinning fraction) of the domains in the crystals. Likely merohedral twin operators for different space groups are tabulated in Table 1. All these operators can be considered as a rotation about some axis defined as the twin axis. Extra complexities in the analysis are introduced when this axis is parallel to some axis of non-crystallographic symmetry (see, for example, Lea & Stuart, 1995).

We describe here some general strategies for detecting and characterizing the merohedral twinning problem by analyzing the data on the basis of diffraction pattern, intensity statistics and packing density. Some troubleshooting ideas are also discussed.

2. Data collection in the case of twinned crystals

For QTLS, the crystal lattices are not superposable in all three dimensions, so care must be taken to make sure that the data from each lattice can be integrated separately and overlaps

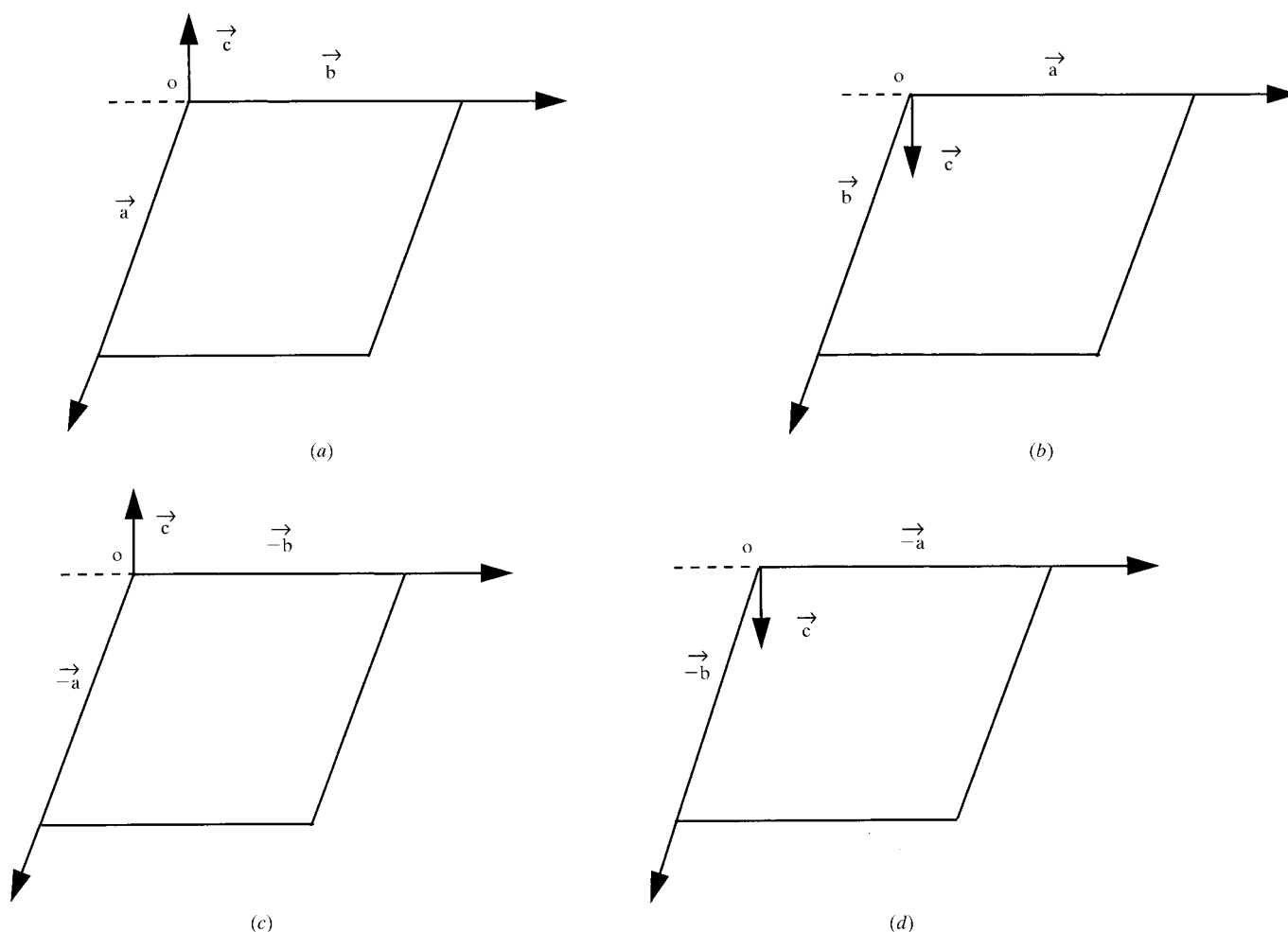


Figure 1

The alignment of crystal axes by twinning operator. All the four blocks (a)–(d) align axes related by crystal symmetry in the case of trigonal $P3_1$ (see Table 1).

Table 1

Lookup tables.

Possible merohedral twin operators for tetragonal, trigonal, hexagonal and cubic space groups. [Source: likely twinning operators (CCP4: General) under CCP4 program suite.]

It is a requirement for merohedral twinning that two independent lattices overlap. Since for tetragonal, trigonal, hexagonal or cubic systems the symmetry requires that two cell axes are equal, twinning often occurs in these systems. For these space groups axes can be aligned in the following ways: (a, b, c) or $(-a, -b, c)$ or $(b, a, -c)$ or $(-b, -a, -c)$, with corresponding reciprocal axes: (a^*, b^*, c^*) or $(-a^*, -b^*, c^*)$ or $(b^*, a^*, -c^*)$ or $(-b^*, -a^*, -c^*)$. The corresponding indexing systems are: (h, k, l) or $(-h, -k, l)$ or $(k, h, -l)$ or $(-k, -h, -l)$. All $P4_i$ and related 4_i space groups: (hkl) is equivalent to $(\bar{h}\bar{k}l)$, so we only need to check real axes (a, b, c) and $(b, a, -c)$ and reciprocal axes (b^*, a^*, c^*) and $(b^*, a^*, -c^*)$, i.e. check if reindexing (hkl) to $(k\bar{h}l)$ gives a better match to previous data sets. Twinning possible with this operator: apparent Laue symmetry for perfect twin would be $P422$.

Space-group number	Space group	Point group	Possible twin operator
75	$P4$	PG4	$k, h, -l$
76	$P4_1$	PG4	$k, h, -l$
77	$P4_2$	PG4	$k, h, -l$
78	$P4_3$	PG4	$k, h, -l$
79	$I4$	PG4	$k, h, -l$
80	$I4_1$	PG4	$k, h, -l$

For all $P4_22$ and related 4_22 space groups, (hkl) is equivalent to $(\bar{h}\bar{k}l)$, $(k\bar{h}l)$ and $(\bar{k}\bar{h}l)$, so all axial pairs are already equivalent as a result of the crystal symmetry. No twinning possible, but a perfect twin for the Laue group $P4$ might appear to have this symmetry.

Space-group number	Space group	Point group	Twin operators
89	$P422$	PG422	None
90	$P4_22$	PG422	None
91	$P4_122$	PG422	None
92	$P4_22$	PG422	None
93	$P4_322$	PG422	None
94	$P4_22$	PG422	None
95	$P4_322$	PG422	None
96	$P4_322$	PG422	None
97	$I422$	PG422	None
98	$I4_122$	PG422	None

All $P3_i$ and $R3$: (hkl) neither equivalent to $(\bar{h}\bar{k}l)$ nor $(k\bar{h}l)$ nor $(\bar{k}\bar{h}l)$, so we need to check all four possibilities. These are the only cases where tetartohedral twinning can occur: real axes (a, b, c) $(-a, -b, c)$, $(b, a, -c)$ and $(-b, -a, c)$, reciprocal axes (a^*, b^*, c^*) , $(-a^*, -b^*, c^*)$, $(b^*, a^*, -c^*)$ and $(-b^*, -a^*, c^*)$, i.e. for $P3$, consider reindexing (hkl) to $(\bar{h}\bar{k}l)$ or $(k\bar{h}l)$ or $(\bar{k}\bar{h}l)$. For $R3$, the indices must satisfy the relationship $-h + k + l = 3n$, so it is only possible to reindex as $(k\bar{h}l)$. For trigonal space groups, symmetry equivalents do not seem as 'natural' as in other systems. Replacing the four basic sets with other symmetry equivalents gives a bewildering range of apparent possibilities, but all are equivalent to one of the above. Twofold twinning possible with this operator: apparent Laue symmetry or twofold perfect twin could be $P312$ (operator $k, h, -l$), $P321$ (operator $-k, -h, -l$) or $P6$ (operator $-h, -k, l$). Fourfold twinning with these operators could generate apparent Laue symmetry $P622$.

Space-group number	Space group	Point group	Possible twin operator
143	$P3$	PG3	$-h, -k, l; k, h, -l; -k, -h, -l$
144	$P3_1$	PG3	$-h, -k, l; k, h, -l; -k, -h, -l$
145	$P3_2$	PG3	$-h, -k, l; k, h, -l; -k, -h, -l$
146	$R3$	PG3	$k, h, -l$

All $P3_12$: (hkl) already equivalent to $(\bar{k}\bar{h}l)$, so we only need to check real axes (a, b, c) and $(b, a, -c)$ and reciprocal axes (a^*, b^*, c^*) and $(b^*, a^*, -c^*)$, i.e. reindex (hkl) to $(k\bar{h}l)$ [or its equivalent operator $(-h, -k, l)$]. Twinning possible with this operator: apparent symmetry for twofold perfect twin would be $P622$ (operator $-h, -k, l$).

Space-group number	Space group	Point group	Possible twin operator
149	$P312$	PG312	$-h, -k, l$ or $k, h, -l$
151	$P3_12$	PG312	$-h, -k, l$ or $k, h, -l$
153	$P3_212$	PG312	$-h, -k, l$ or $k, h, -l$

All $P3_21$: (hkl) already equivalent to $(k\bar{h}l)$, so we only need to check real axes (a, b, c) and $(-a, -b, -c)$ and reciprocal axes (a^*, b^*, c^*) and $(-a^*, -b^*, -c^*)$, i.e. reindex (hkl) to $(\bar{h}\bar{k}l)$ [or its equivalent operator $(-k, -h, -l)$]. Twinning possible with this operator: apparent symmetry for twofold perfect twin would be $P622$ (operator $-h, -k, l$).

Space-group number	Space group	Point group	Possible twin operator
150	$P321$	PG321	$-h, -k, l$ or $-k, -h, -l$
152	$P3_121$	PG321	$-h, -k, l$ or $-k, -h, -l$
154	$P3_221$	PG321	$-h, -k, l$ or $-k, -h, -l$

All $P6_i$: (hkl) already equivalent to $(\bar{h}\bar{k}l)$, so we only need to check real axes (a, b, c) and $(b, a, -c)$ and reciprocal axes (a^*, b^*, c^*) and $(b^*, a^*, -c^*)$, i.e. reindex (hkl) to $(k\bar{h}l)$. Twinning possible with this operator: apparent symmetry for twofold perfect twin would be $P622$ (operator $k, k, -l$).

Space-group number	Space group	Point group	Possible twin operator
168	$P6$	PG6	$k, h, -l$
169	$P6_1$	PG6	$k, h, -l$
170	$P6_5$	PG6	$k, h, -l$
171	$P6_2$	PG6	$k, h, -l$
172	$P6_4$	PG6	$k, h, -l$
173	$P6_3$	PG6	$k, h, -l$

All $P6_22$: (hkl) already equivalent to $(\bar{h}\bar{k}l)$ and $(k\bar{h}l)$ and $(\bar{k}\bar{h}l)$, so no twinning is possible. However, a perfect twin for the Laue group $P312$, $P321$ or $P6$ might appear to have this symmetry.

Space-group number	Space group	Point group	Possible twin operator
177	$P622$	PG622	$k, h, -l$
178	$P6_122$	PG622	None
179	$P6_522$	PG622	None
180	$P6_222$	PG622	None
181	$P6_422$	PG622	None
182	$P6_322$	PG622	None

All $P2_13$ and related 2_13 space groups: (hkl) already equivalent to $(\bar{h}\bar{k}l)$, so we only need to check real axes (a, b, c) and $(b, a, -c)$ and reciprocal axes (a^*, b^*, c^*) and $(b^*, a^*, -c^*)$, i.e. reindex (hkl) to $(k\bar{h}l)$. Twinning possible with this operator: apparent symmetry for twofold perfect twin would be $P43$ (operator $k, h, -l$).

Space-group number	Space group	Point group	Possible twin operator
195	$P23$	PG23	$k, h, -l$
196	$F23$	PG23	$k, h, -l$
197	$I23$	PG23	$k, h, -l$
198	$P2_13$	PG23	$k, h, -l$
199	$I2_13$	PG23	$k, h, -l$

All $P4_32$ and related 4_32 space groups: (hkl) already equivalent to $(\bar{h}\bar{k}l)$ and $(k\bar{h}l)$ and $(\bar{k}\bar{h}l)$, so we do not need to check.

Space-group number	Space group	Point group	Possible twin operator
207	$P432$	PG432	None
208	$P4_232$	PG432	None
209	$F432$	PG432	None
210	$F4_32$	PG432	None
211	$I432$	PG432	None
212	$P4_332$	PG432	None
213	$P4_132$	PG432	None
214	$I4_132$	PG432	None

excluded, the data can perhaps then be treated for non-merohedral twinning effects (Liang *et al.*, 1996; Lietzke *et al.*, 1996). No examples of this type are discussed here.

For TLS, the twinning fraction may vary from crystal to crystal (for details, see Valegard *et al.*, 1998). In order to estimate the twin volume fraction accurately, it is essential that a complete data set is collected from a single crystal (for details, see Ito *et al.*, 1995). When the merohedral twinning is near-perfect, it is possible to mistake the space group; if only the supposed unique data were collected there would be no chance of deconvolution.

3. Diagnostic signals of twinning

It is important to detect twinning before embarking on the structure solution, and if it is possible to find an untwinned crystal this is by far the best approach!

3.1. Packing density

If the supposed asymmetric unit volume of the crystal is too small to hold the molecule, it is likely that there is perfect merohedral twinning and the space group has been wrongly assigned.

3.2. Intensity statistics

The intensity statistics from an untwinned crystal are quite different from those of twinned crystals. Wilson showed that for a single crystal the mean and higher moments of centric and acentric intensities and amplitudes follow a predictable pattern (Wilson, 1949). These may be tabulated either as $\langle I^k \rangle / \langle I \rangle^k$, $\langle E^k \rangle / \langle E \rangle^k$ or as functions of Z (defined as $I / \langle I \rangle$), but they are all related to each other. For measurements I_{iH} from a twinned crystal, each 'intensity' is in fact the sum of two or more I_H ,

$$I_{iH_1} = [(1-x)I_{H_1} + xI_{H_2}], \quad (1)$$

$$I_{iH_2} = [xI_{H_1} + (1-x)I_{H_2}], \quad (2)$$

$$I_{H_1} = [(1-x)I_{iH_1} - xI_{iH_2}] / (1-2x), \quad (3)$$

$$I_{H_2} = [(1-x)I_{iH_2} - xI_{iH_1}] / (1-2x). \quad (4)$$

3.3. Useful distributions to inspect

(i) The $N(Z)$ plot (given in *TRUNCATE* output) shows the number of observed weak reflections plotted against the expected value. For twinned data, there are many fewer weak reflections than expected and hence the acentric distribution is always sigmoidal. This follows from the fact that each observed I_{iH} is a sum of two or more I_H , and it is unlikely that all I_H contributions will be weak (it is wise to exclude the centric terms, since if the space group has been wrongly assigned these may be wrongly flagged).

(ii) The k th moments of I or E also provide useful indicators. The expected values are given in various references (Stanley, 1972; Redinbo & Yeates, 1993; Breyer *et al.*, 1999; Yeates, 1997) and the observed values can be extracted from several commonly used programs, for example *TRUNCATE*

and *ECALC* (Collaborative Computational Project, Number 4, 1994). The ratio of $\langle I^2 \rangle / \langle I \rangle^2$ for acentric reflections against resolution gives expected values of 2.0 for cases without twinning and 1.5 for perfect hemihedral twinning, respectively. Similarly, the acentric Wilson ratio, $\langle |E|^2 \rangle / \langle |E|^2 \rangle$, where E is the normalized structure factor, is expected to give values of 0.785 for twinned and 0.885 for untwinned data.

(iii) Once the twinning operator is known and the twinned partner intensity can be selected, it is often possible to 'detwin' the data assuming different values of the twinning ratio x . If the indices (hkl) of I_{iH_1} and I_{iH_2} are related so that $T[h_1k_1l_1]$ overlaps $T[h_2k_2l_2]$ [T_{recip}], then (1) and (2) are valid and, providing x is not equal to 0.5, the true I_{H_1} and I_{H_2} can be determined from (3) and (4). For the correct value of x the detwinned data should satisfy certain criteria. The intensity statistics of the detwinned data should be more 'normal'. The number of negative I_H should be small (this is the basis of the Britton plot; Britton, 1972). The correlation between I_{H_1} and I_{H_2} should be minimum. This may not actually fall to zero, since if there is non-crystallographic symmetry with an axis parallel to the twin axis, there may well be real correlation between the two observations.

3.4. Using twinned data

Twinned data has been used to solve many structures. The techniques fall into two groups, depending on the methods to be used. If the structure solution requires the use of amplitudes, as is the case when using heavy-atom derivative or

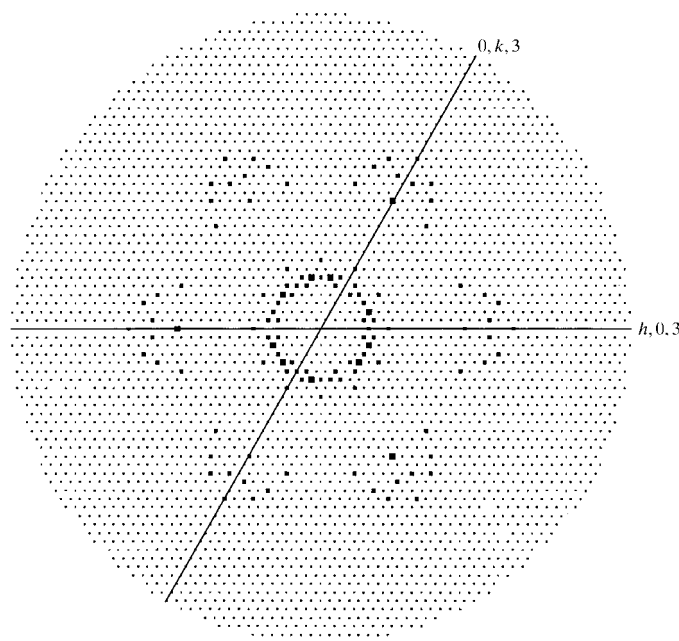


Figure 2

A *HKLVIEW* (Collaborative Computational Project, Number 4, 1994) of the pseudo-precession photograph of the diffraction pattern (for $l = 3$) of the crystals of α -LA. The diffraction pattern indicates an apparent (hexagonal) symmetry. The observed diffraction appears to be normal, but each of the observed intensities contains equal contributions from the two domains. This is an indication of the symptoms of merohedral twinning.

MAD data to give experimental phases, it is essential to detwin the intensities. This can be performed with programs such as *DETWIN* (Leslie, 1998), providing both the 3×3 twinning matrix $[T]$ and the twinning ratio x are known and the value of x is not equal to 0.5 (for details, see Valegard *et al.*, 1998).

Possible twinning operators can usually be deduced from the space-group symmetry and the apparent symmetry of the diffraction pattern. The second technique is not to bother to detwin the data but to use the summed intensities for structure solutions. Any technique which exploits intensities (I s) rather than structure factors (F s) is suitable.

3.4.1. Molecular replacement. MR should be able to find solutions for each crystal block. However, the usual test of whether these solutions overlap cannot be performed since each is partially occupied, but the results are often clear. Such a molecular-replacement search would in fact also indicate the real-space twinning operator.

3.4.2. Refinement against I s. Once the solution is obtained, *SHELXL* can be used with the option to refine the twinning factor x as well as the coordinates. *SHELXL* is a versatile program which can handle refinement with twinned data of any kind, be it non-merohedral, merohedral or racemic twinning.

3.5. Example: native bovine α -lactalbumin

Both native recombinant and native bovine α -lactalbumin (α -LA; Sigma) crystallize as hemihedral twins (N. Chandra, K. Brew & K. R. Acharya, unpublished results). The unit-cell parameters are $a = 93.5$, $b = 93.5$, $c = 67.0$ Å; $\alpha = 90.0$, $\beta = 90$, $\gamma = 120^\circ$. The pseudo-precession picture (Fig. 2) seems to show that the Laue group is either $P6/m$ or $P6/mmm$, as the

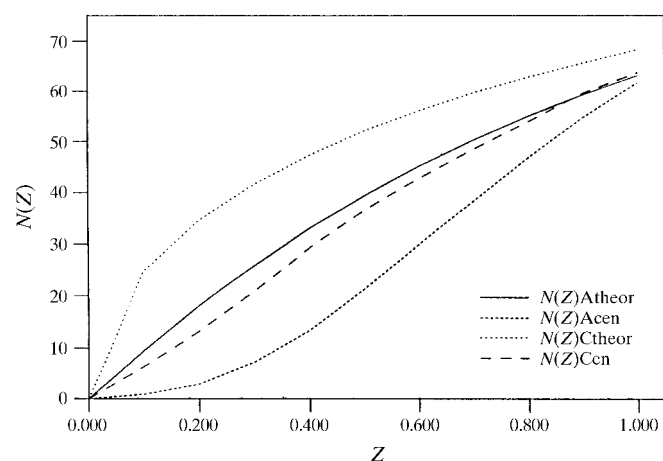


Figure 3
A graphical representation of $N(Z)$ curves against Z for acentric data from α -LA. The term $N(Z)$ represents the cumulative distribution function, where Z represents the intensity relative to the mean intensity (i.e. $Z = I/\langle I \rangle$, where I is the intensity), gives the fraction of reflections having an intensity less than Z . As indicated, a comparison with the theoretical distribution shows that a sigmoid habitus follows the distribution for perfect merohedral twinning ($x = 1/2$) (for details, see Gomis-Rüth *et al.*, 1995).

diffraction pattern nicely displays an apparent hexagonal symmetry.

A data set to 2.5 Å resolution was collected from a single crystal on a Siemens area detector using Cu $K\alpha$ radiation and the data were reduced in Laue symmetry groups $\bar{3}$, $\bar{3}1m$, $\bar{3}m1$, $6/m$ and $6/mmm$, corresponding to the space groups $P3$, $P312$, $P321$, $P6$ and $P622$, respectively. The corresponding R_{sym} values for these space groups are 6.9, 7.9, 8.5, 8.4 and 8.9% (99.9% complete in each space group), respectively.

The problem of twinning was addressed on the basis of the intensity statistics (Fig. 3), an intensity distribution which did not follow the Wilson statistics (Fig. 4), and the crystal packing density. Calculation of the solvent content for 12 molecules per unit cell gives a reasonable value of 57% [the V_m (Matthews, 1968) value is $2.99 \text{ \AA}^3 \text{ Da}^{-1}$]. Therefore, none of the above space groups could be eliminated. The ratio of $\langle I^2 \rangle / \langle I \rangle^2$ for α -LA is found to be 1.47 (Fig. 4) indicating the crystals might be hemihedrally twinned. Similarly, the Wilson ratio $\langle |E|^2 \rangle / \langle |E|^2 \rangle$ calculated for acentric reflections gave a value of 0.886, consistent with the data from α -LA crystals being perfectly twinned.

We estimated the twinning fraction x from the parameter H (Yeates, 1997), where

$$H = |I_{h_1} - I_{h_2}| / (I_{h_1} + I_{h_2}). \quad (5)$$

H is a function of x [from 0 to $(1 - 2x)$] and the true crystallographic intensities. The value of x is determined using the fact that

$$\langle H \rangle = 0.5 - x, \quad (6)$$

$$\langle H^2 \rangle = (0.5 - x)^2 / 3. \quad (7)$$

The mean value of x directly estimated from H was found to be 0.49, compared with the average value of 0.50 obtained based on intensity statistics described by Britton (1972), Rees (1982) and Fisher & Sweet (1980).

The self-rotation function also showed more peaks than were expected (Fig. 5). The molecular-replacement translation search fixed the correct space group as $P3$. The crystals are twinned along the a , b or $-a$, $-b$ axes, all of which are

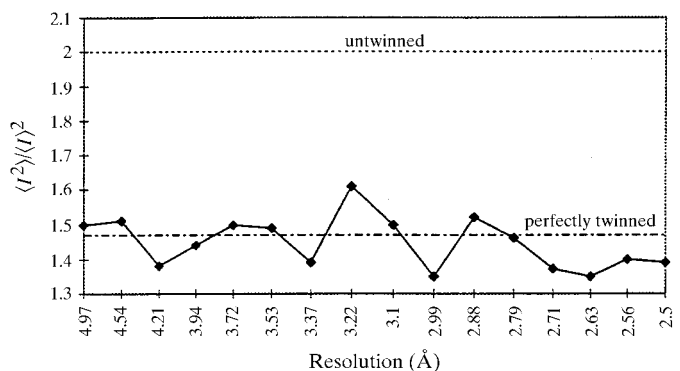


Figure 4
A test for perfect merohedral (hemihedral) twinning for acentric data from α -LA. The ratios are computed in thin resolution bins of 300–400 reflections.

equivalent by the symmetry of the space group. That is, the real-space twin matrix in this case is

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Detwinning of the observed twinned structure factors was carried out according to the method of Redinbo & Yeates (1993); following this, the structure was refined to R_{free} and R values of 19.8 and 18.8%, respectively. The *X-PLOR* data-input files were modified and used as a means of a convenient way to detwin the twinned data.

```
REMARKS DETWINNING OF PERFECTLY TWINNED (TWIN VOL. FRACTION X=1/2) DATA.
REMARKS THE INPUT FILES SHOULD BE MODIFIED TO DECONVOLUTE TWINNED DATA.
REMARKS CORRECT TWIN MATRIX SHOULD BE SPECIFIED IN XPLOR CONVENTION.
REMARKS XPLOR3.1, 3.843, 3.851, AND CNS VERSIONS CAN BE USED.
REMARKS FOLLOWING THE OVERALL LOGIC TELLS HOW DATA CAN BE DETWINNED.
REMARKS SEE FOR DETAILS (REDINBO & YEATES, 1993; YEATES, 1997).
REMARKS AS AN EXAMPLE, THE INPUT FILE GIVEN IS FOR ALPHA-LA.
!
!
!parameter @param19x.pro end      !{SPECIFY PARAMETER FILE}
parameter
@TOPPAR:parhcsdx.pro end
structure @alpha_LA_gen.psf end   !{SPECIFY STRUCTURE FACTOR FILE}
coordinate @alpha_LA_gen.pdb     !{SPECIFY COORDINATE FILE}

xrefin

a=93.5 b=93.5 c=67.0 alpha=90. beta=90. gamma=120.

symmetry=(x,y,z)                  !{SPACE GROUP P3}
symmetry=(-Y,X-Y,Z)
symmetry=(-X+Y,-X,Z)

SCATter { chemical C* }
2.31000 20.8439 1.02000 10.2075 1.58860 .568700 .865000 51.6512 .215600
SCATter { chemical N* }
12.2126 .005700 3.13220 9.89330 2.01250 28.9975 1.16630 .582600 -11.529
SCATter { chemical O* }
3.04850 13.2771 2.28680 5.70110 1.54630 .323900 .867000 32.9089 .250800
SCATter { chemical S* }
6.90530 1.46790 5.20340 22.2151 1.43790 .253600 1.58630 56.1720 .866900
SCATter { chemical P* }
6.43450 1.90670 4.17910 27.1570 1.78000 0.52600 1.49080 68.1645 1.11490
SCATter { chemical CU* }
11.8168 3.37484 7.11181 0.24408 5.78135 1.14520 0.07240 19.8970 1.14431

nreflections=30000
reflection @alph LA.fob end      !{SPECIFY TWINNED DATA}
resolution 40.0 2.5             !{SPECIFY RESOLUTION RANGE}

selection={known}
method=fft
fft
memory=2000000
end

update                          !{FCALC FOR MOL 1}
do amplitude (fpart=fcalc)      !{FPART=FCALC FOR MOL 1}
end

coord rotate
matrix (1 0 0)                  !{SPECIFY TWIN MATRIX}
(0 1 0)                          !{THE MATRIX SHOULD BE DERIVED BY
(0 0 -1)                          !{ANALYSIS OF THE X-RAY DATA.
!THE GIVEN MATRIX IS FOR ALPHA-LA}

end

xrefin
update                          !{FCALC FOR MOL 2}

do amplitude (fobs=fobs^2)      !{FOBS=IOBS}
do amplitude (fcalc=fcalc^2)    !{FCALC=I(2)CALC}
do amplitude (fpart=fpart^2)    !{FPART=I(1)CALC}

do amplitude (sigma=(fcalc+fpart))
do scale (fobs=sigma)          !{SCALE THE DATA}
do amplitude (sigma=(fpart-fcalc))
!do amplitude (sigma=(fcalc-fpart))

!IF FOB > |SIGMA|,                !{OVERALL LOGIC BASED ON
!FPART= (FOB+SIGMA)/2            !THE THEORY OF
!FCALC= (FOB-SIGMA)/2          !REDINBO & YEATES.
!IF |SIGMA| > FOB,              !SEE FOR DETAILS
!IF SIGMA > 0,                  !REDINBO & YEATES, 1993}
!FPART = FOB
!FCALC = 0
!IF SIGMA < 0,
!FPART = 0
!FCALC = FOB

!{THE FOLLOWING EQUATIONS ARE USED FOR DETWINNING THE DATA BASED ON
!THE LINES OF LOGIC AS MENTIONED ABOVE}

do amplitude (weight=heavy(fobs-abs(sigma)))
do amplitude (fpart=weight*(fobs+sigma)/2)
do amplitude (fpart=fpart+((1-weight)*fobs*heavy(sigma)))
do amplitude (fcalc=weight*(fobs-sigma)/2)
do amplitude (fcalc=fcalc+((1-weight)*fobs*heavy(-1.*sigma)))

do amplitude (fpart= sqrt(fpart)) !{FPART=F(1)UN}
do amplitude (fcalc= sqrt(fcalc)) !{FCALC=F(2)UN}
do amplitude (fobs= sqrt(fobs))  !{FOBS=FOBS}
do amplitude (sigma=fcalc)       !{STORE FCALC IN SIGMA}
do amplitude (weight=fpart)      !{STORE FPART IN WEIGHT}

write reflc fobs fcalc output=fof2u_alpha_LA.fob end
print rfactor                    !{AS A CHECK FOR R VALUES
!DURING REFINEMENT OR
!MODEL BUILDING}

do amplitude (fcalc=fpart)
write reflc fobs fcalc output=foflu_alpha_LA.fob end
print rfactor                    !{SAME LOGIC FOR MOL 1}

do amplitude (fobs=weight)
write reflc fobs output=alpha_LA11_untw.fob end
print rfactor                    !{DETWINNED DATA FOR MOL 1
!USED FOR REFINEMENT
!OR MODEL BUILDING}

do amplitude (fobs=sigma)
!write reflc fobs output=alpha_LA12_untw.fob end
print rfactor                    !{DO FOR MOL 2}

end
stop
```

corrected using (3) and (4) (Britton, 1972; Murray-Rust, 1973; Rees, 1982; Fisher & Sweet, 1980; Redinbo & Yeates, 1993; Yeates, 1997).

A high-resolution (1.509 Å) data set (R_{sym} is 0.052 and the completeness of the data is 95.9%) was collected using synchrotron radiation at Daresbury Station 7.2 from crystals of xenobiotic acetyltransferase (XAT). The crystals are in space group $R3$, with unit-cell parameters $a = 123.64$, $b = 123.64$, $c = 63.08$ Å, $\alpha = 90$, $\beta = 90$, $\gamma = 120^\circ$, $Z = 1$ molecule per asymmetric unit, $V_m = 3.96$ Å³ Da⁻¹ (Matthews, 1968), and exhibit partial merohedral twinning with an average twin fraction volume x , calculated from (3), (4) and (5), of 0.1982.

4. Partial twinning: XAT

When the twin volume fraction x is less than 0.5, the twinning is called partial merohedral twinning; the diffraction pattern does not reveal a higher apparent symmetry, but the observed intensities still contain contributions from both the domains. If there is a good estimate of x and it is below a value of about 0.45, the twinning can reliably be

The only possible twin operator for $R3$ generates a rotation about the diagonal of the a and b axes, *i.e.* the real-space twin operator matrix is

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

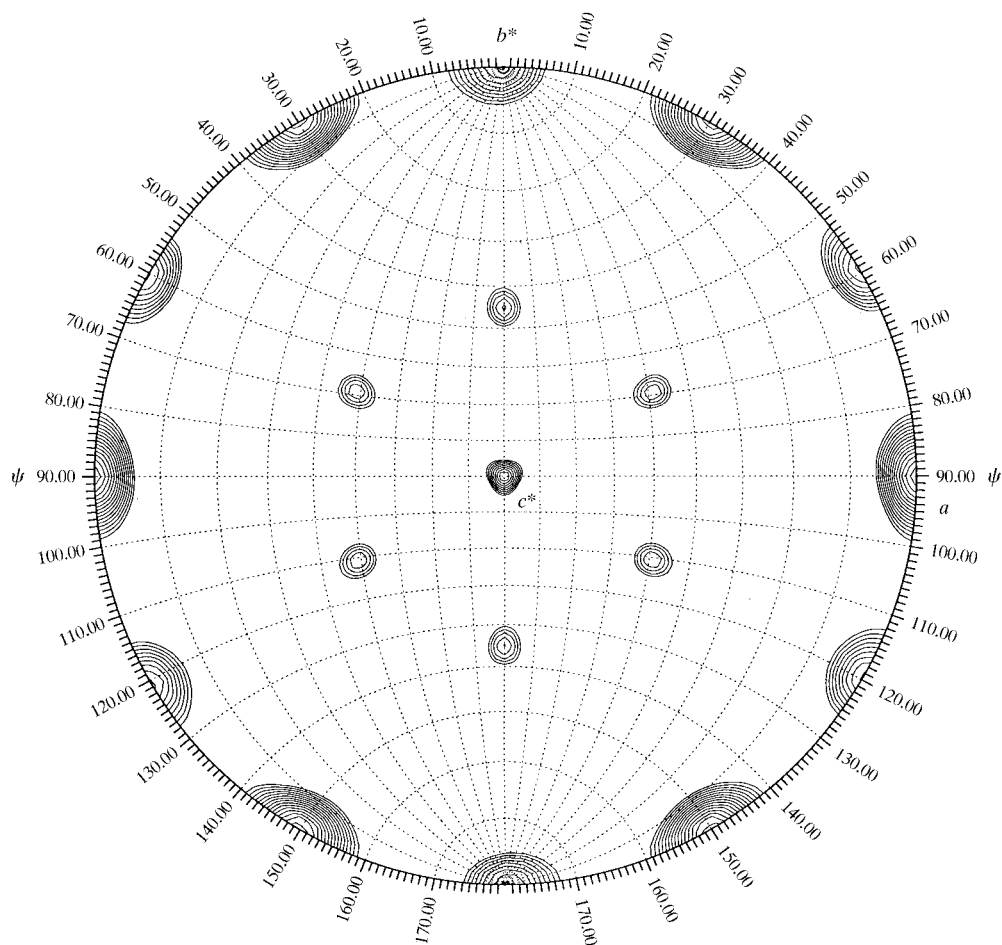


Figure 5

Representation of the Patterson self-rotation function of the set of structure factors in space group $P3$ for α -LA (using a spherical integration radius of 21 Å and data between 8 and 4.5 Å resolution) was produced using *GLRF* (Tong & Rossmann, 1990). The diagram indicates the presence of a strong twofold along the crystallographic c axis ($\psi = 90^\circ$, $\varphi = 90^\circ$, $\kappa = 180^\circ$), as is usually expected for a crystallographic sixfold axis. Additionally, twofold peaks occur every 30° in the ab plane. The total number of peaks is 12 and this would suggest an erroneously high space group of the $P6/mmm$ Laue family with 12 copies of the asymmetric unit in the unit cell (for a nice discussion on the interpretation of peaks in the Patterson self-rotation function in the case of human lactoferrin, see Breyer *et al.*, 1999).

When using the measured data, the MR solution of XAT did not refine well, with the R_{free} and R values converging at 39.4 and 36.1%, respectively. There was only one MR solution with a high correlation coefficient (after rigid-body refinement) of 71.2% with an R factor of 35.9% and, despite the high solvent content (70%), there was no sign of a second molecule. Furthermore, the self-rotation function showed only one significant peak.

To investigate whether the data could be partially twinned, we re-examined the data-processing statistics more carefully. The *TRUNCATE* program in the *CCP4* package (Collaborative Computational Project, Number 4, 1994) produces a table for acentric and centric reflections of the second, third and fourth moments of I . The expected value of the second moment for untwinned acentric data is 2.0 and any gross deviation from this is a probable indication of partial twinning.

The partially twinned data of XAT were detwinned (Fig. 6) using the *DETWIN* program (Leslie, 1998), which is now

released as a part of the *CCP4* package. This requires as input the reciprocal-space twinning operator, a range of values of the twin volume fraction x and the input data as intensities, and outputs a list of supposedly detwinned intensities. It tabulates the correlation coefficient between I_{H_1} and I_{H_2} after detwinning, which should have its minimum value for the best estimate of x .

The detwinned data of XAT using the twinning ratio $x = 0.1941$ was used in MR and gave the same solution as that obtained from twinned data. The detwinned data were used in refinement of the model to see whether there was an improvement in the values of R_{free} and R . We found that the values of R_{free} and R were reduced by 4.3 and 4.5%, respectively. There was a remarkable improvement in the electron density of the side chains of many of the lysines and arginines and also some of the residues which it was not possible to fit into the electron-density map previously (before detwinning).

Further refinement of the XAT structure was subsequently carried out with *SHELXL* (Sheldrick & Schneider, 1997) using the *TWIN* command option available in the program. We found that the twin volume fraction x was refined from 0.194 to 0.191, showing that the other methods used are remarkably sensitive.

5. Correct estimation of x

The exact value of the twin volume fraction x can be computed if the calculated structure-factor amplitudes for the given model are available. Calculating the correlation coefficient [$\text{Corr}(x)$] using the expressions of Gomis-Rüth *et al.* (1995) gives

$$\text{Corr}(x) = \sum_H [(\Delta_{\text{calc}})(\Delta_{\text{obs}})] / \left[\sum_H (\Delta_{\text{calc}})^2 \sum_H (\Delta_{\text{obs}})^2 \right]^{1/2}, \quad (8)$$

where

$$\Delta_{\text{cal}} = \left| I_{iH_1}^{\text{calc-twin}}(x) - \langle I_{iH_1}^{\text{calc-twin}}(x) \rangle \right|, \quad (9)$$

$$\Delta_{\text{obs}} = \left| I_{iH_1}^{\text{obs}}(x) - \langle I_{iH_1}^{\text{obs}}(x) \rangle \right|, \quad (10)$$

$\langle I \rangle$ is the mean intensity and

$$I_{iH_1}^{\text{calc-twin}}(x) = (1-x)I_{iH_1}^{\text{calc}} + xI_{iH_2}^{\text{calc}}. \quad (11)$$

The correlation coefficient $\text{Corr}(x)$ attains a unique maximum as a function of x , and the optimum value of x may be accurately determined by setting the derivative in (8): $d\text{Corr}(x)/dx = 0$. Therefore, the optimal value of x is calculated from

$$\alpha = \{[\langle C_{iH_1}^{\text{calc}} C_{iH_1}^{\text{calc}} \rangle - (C_{iH_1}^{\text{calc}} S_{iH_{12}}^{\text{calc}}) \langle C_{iH_1}^{\text{obs}} C_{iH_1}^{\text{calc}} \rangle]\} \\ \times [C_{iH_1}^{\text{calc}} (C_{iH_1}^{\text{calc}} - C_{iH_2}^{\text{calc}})], \quad (12)$$

$$x = \alpha / (\langle C_{iH_1}^{\text{obs}} S_{iH_{12}}^{\text{calc}} \rangle), \quad (13)$$

where

$$S_{iH_{12}}^{\text{calc}} = C_{iH_1}^{\text{calc}} + C_{iH_2}^{\text{calc}} \quad (14)$$

and $C_{iH_i}^{\text{calc}}$ represents the calculated structure factors.

However, the accurate value of x , which is quite different from 0.5 and is independent of resolution, can be evaluated from (12) and (13). Then, the intensities of the twin-related reflections are detwinned using (3) and (4). In the case of

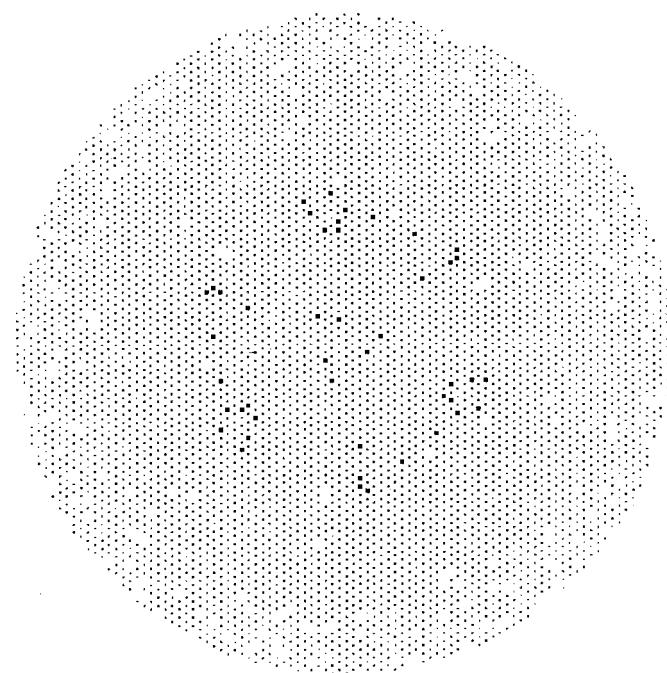


Figure 6
Differences observed between partial merohedral and perfect merohedral twinning. A *HKLVIEW* (Collaborative Computational Project, Number 4, 1994) of the pseudo-precession photograph of the diffraction pattern of the crystals of XAT. The diffraction pattern indicates no apparent symmetry. The observed diffraction appears to be normal with a threefold symmetry. Because the observed intensities of the twin-related reflections are not equal, the diffraction pattern does not acquire an additional symmetry unlike in the case of perfect twinning. Therefore, it is very difficult to recognize partial merohedral twinning from the observed diffraction pattern.

XAT, we did not try to obtain the accurate value of the twin volume fraction x using (13), since the *SHELXL* program is being used and takes care of the refinement of the twin factor. The structural results and other details of the XAT structure will be published elsewhere (N. Chandra, J. Snidwongse, W. V. Shaw, I. A. Murray & P. C. E. Moody, manuscript in preparation).

6. Troubleshooting

It is always better to avoid twinning altogether so that many difficulties can be eliminated. Growing the crystals under different crystallization conditions may be one way of overcoming the problem. A new crystal form in an entirely different space group may be obtained. For example, for phosphopantetheine adenylyltransferase (PPAT) from *Escherichia coli*, the data collected from PPAT crystals were reduced in *P3*, *P3₁*, *P6* and *P6₃* space groups (unit-cell parameters: $a = 65.15$, $b = 65.15$, $c = 119.06$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$) and exhibit an apparent Laue symmetry of *6/m*. The $N(z)$ plots indicated that twinning was likely and thus that the true Laue symmetry is probably 3 (Fig. 7). However, by changing the crystallization conditions, the crystals of the same protein were regrown in an entirely different space group *I23* free from twinning. The structure of this protein has been solved using MAD data (Izard & Geerlof, 1999).

Cephalosporin synthase protein structure was solved by MIR using several data sets collected from different crystals which exhibited different values of twin volume fraction (Valegard *et al.*, 1998). This demonstrates the accuracy with which the twinning fraction can be determined and used to deconvolute the data.

7. Summary

We have discussed the general strategy regarding the identification, analysis, characterization and correction of the data collected from twinned crystals based on the X-ray diffraction pattern, intensity statistics, packing density and refinement statistics. Some suggestions have been given for overcoming merohedral twinning.

We wish to acknowledge the significant contribution of Eleanor J. Dodson to this work and to thank Todd O. Yeates, Matthew R. Redinbo and S. Ramaswamy for helpful discussion and suggestions on the merohedral twinning problem, George Sheldrick and Geoffrey B. Jameson on the refinement strategy using twinned data, Andrew G. W. Leslie for allowing us to use the *DETWIN* program as one of the beta testers, Michael I. Wilson for helping to write a program for detwinning and Tina Izard for providing twinned data. NC thanks the CCP4 organizers for the opportunity to speak at the CCP4 meeting. The α -LA project is supported by a Leverhulme Trust grant (F351/U) to Ravi Acharya and the XAT project is supported by a BBSRC grant to Peter C. E. Moody.

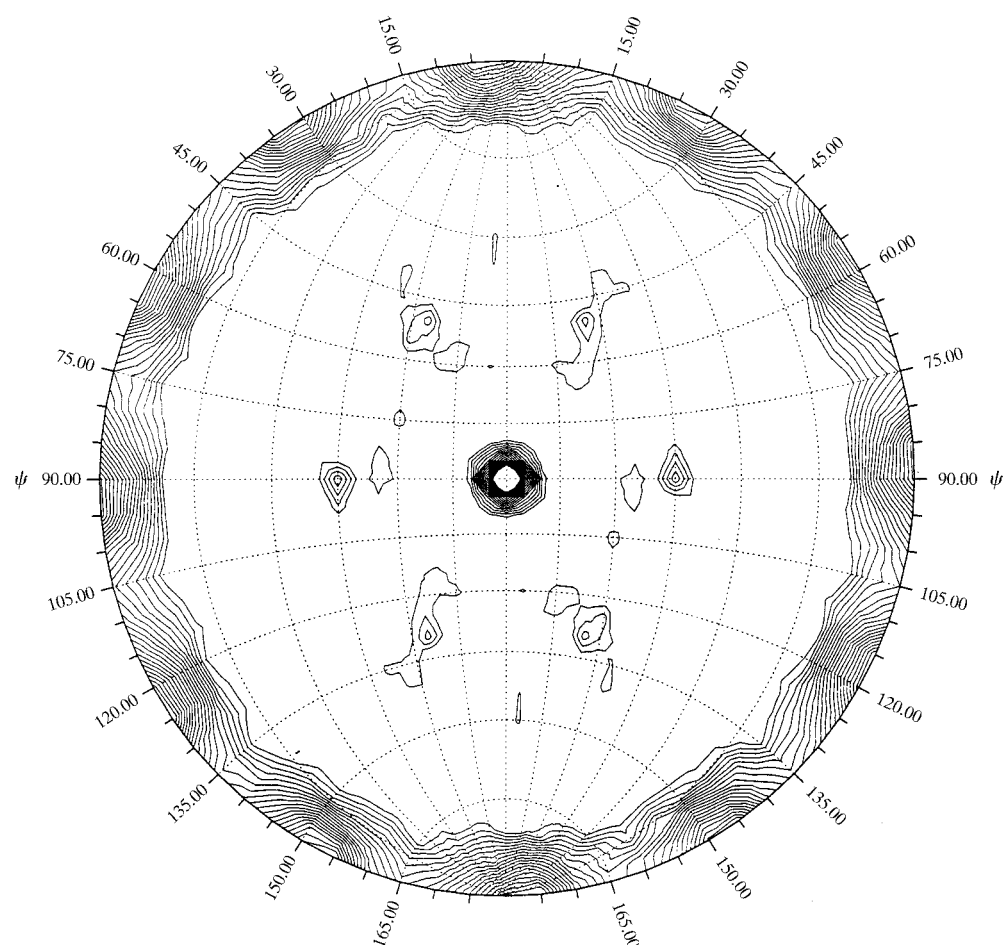


Figure 7

Representation of the Patterson self-rotation function of the set of structure factors in space group $P6_3$ for PPAT (using a spherical integration radius of 30 Å and diffraction data between 10 and 3.0 Å resolution) was produced using *GLRF* (Tong & Rossmann, 1990). The diagram shows the presence of a strong twofold along the crystallographic c axis ($\psi = 90$, $\varphi = 90$, $\kappa = 180^\circ$) as is usually expected for a crystallographic sixfold axis. Additionally, there are twofold peaks occurring every 30° in the ab plane. The total number of peaks is 12 and this suggests an erroneously high space group of the $P6/mmm$ Laue family.

References

Breyer, W. A., Kingston, R. L., Anderson, B. F. & Baker, E. N. (1999). *Acta Cryst.* **D55**, 129–138.
 Britton, D. (1972). *Acta Cryst.* **A28**, 296–297.
 Catti, M. & Ferraris, G. (1976). *Acta Cryst.* **A32**, 163–165.
 Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
 Donnay, G. & Donnay, J. D. H. (1974). *Can. Miner.* **12**, 422–425.

Dunitz, D. (1964). *Acta Cryst.* **17**, 1299–1304.
 Fisher, R. G. & Sweet, R. M. (1980). *Acta Cryst.* **A36**, 755–760.
 Gao, Q., Weber, H.-P. & Craven, B. M. (1994). *Acta Cryst.* **B50**, 695–703.
 Giacovazzo, C. (1992). Editor. *Fundamentals of Crystallography*, pp. 61–140. IUCr/Oxford University Press.
 Gomis-Rüth, F. X., Fita, I., Kiefer-sauer, R., Huber, R., Avilés, F. X. & Navaza, J. (1995). *Acta Cryst.* **D51**, 819–823.
 Ito, N., Komiyama, N. H. & Fermi, G. (1995). *J. Mol. Biol.* **250**, 648–658.
 Izard, T. & Geerlof, A. (1999). *EMBO J.* **18**, 2021–2030.
 Lea, S. & Stuart, D. (1995). *Acta Cryst.* **D51**, 160–167.
 Leslie, A. G. W. (1998). *DETWIN. A Program for Detwinning of Merohedrally Twinned Data*.
 Liang, J., Ealick, S., Nielsen, S., Schreiber, S. L. & Clardy, J. (1996). *Acta Cryst.* **D52**, 207–210.
 Lietzke, S. E., Carperos, V. E. & Kundrot, C. E. (1996). *Acta Cryst.* **D52**, 687–692.
 Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
 Murray-Rust, P. (1973). *Acta Cryst.* **B29**, 2559–2566.
 Redinbo, M. R. & Yeates, T. O. (1993). *Acta Cryst.* **D49**, 375–380.
 Rees, D. C. (1982). *Acta Cryst.* **A38**, 201–207.
 Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
 Stanley, E. (1972). *J. Appl. Cryst.* **5**, 191–194.
 Tong, L. & Rossmann, M. G. (1990). *Acta Cryst.* **A46**, 783–792.
 Valegard, K., Scheltinga, A. C. T. V., Lloyd, M. D., Hara, T., Ramaswamy, S., Perrakis, A., Thompson, A., Lee, H., Baldwin, J. E., Schofield, C. J., Hajdu, J. & Andersson, I. (1998). *Nature (London)*, **394**, 805–809.
 Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.
 Yeates, T. O. & Fam, B. C. (1999). *Structure*, **7**, R25–R29.
 Wei, C. H. (1969). *Inorg. Chem.* **8**, 2384–2397.
 Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.