Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Equations for determining tetartohedral twin fractions

Todd O. Yeates[a,b]* and Feng Yu[c]

[a]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA, [b]UCLA–DOE Institute for Genomics and Proteomics, Los Angeles, CA, USA, and [c]Shanghai Institute of Applied Physics, Chinese Academy of Sciences, People's Republic of China

Correspondence e-mail: yeates@mbi.ucla.edu

Received 11 July 2008
Accepted 9 September 2008

In most cases of merohedral twinning, two different twin-domain orientations are present. A rarer type of merohedral twinning exists in which there are four different twin-domain orientations. The former case is referred to as hemihedral twinning, while the latter more complex type is referred to as tetartohedral twinning. In tetartohedral twinning, each observed reflection is the weighted sum of four twin-related but otherwise independent reflection intensities. The weights that determine how the true crystallographic intensities combine to give the observed intensities are described by four twin fractions representing the fractional volumes of the four different domain orientations within the specimen. Here, equations are developed to determine values for the four tetartohedral twin fractions based on a statistical comparison of quadruplets of twin-related reflections. Knowledge of the twin fractions is important in working backwards to obtain values for the true crystallographic intensities. Use of the equations is demonstrated with synthetic intensity data simulated according to given values of the twin fractions.

1. Introduction

Molecular crystals can suffer from a variety of disorders. A growth disorder referred to as merohedral twinning is observed fairly often in crystals of macromolecules such as proteins and nucleic acids (reviewed in Yeates, 1997; Dauter *et al.*, 2005; Parsons, 2003). Merohedral twinning typically occurs in space groups that have lower point symmetries than the lattices that support them. For example, a crystal in space group $P4$ falls on a tetragonal lattice which has 422 symmetry. In such a case, aberrant crystal growth can produce a specimen that is comprised of two different domains related to each other by a rotation that belongs to the lattice symmetry but not to the crystal space-group symmetry. In the case of $P4$, this rotation (or twin operation) could be any one of the twofold-symmetry axes in the ab plane. Because the differently oriented twin domains have indistinguishable lattices, the diffraction patterns they produce are precisely superimposed on each other. The resulting diffraction pattern therefore appears unremarkable, yet each of the recorded reflection intensities is actually a sum of contributions from two twin-related but crystallographically independent reflections (Buerger, 1960).

The fractional volume of a crystal specimen comprised by the differently oriented twin domains is referred to as the twin fraction. Most cases of merohedral twinning involve just two possible twin-domain orientations and are referred to as hemihedral twinning. In hemihedral twinning only one twin fraction, α, needs to be specified; a twin fraction of 0.30 means

© 2008 International Union of Crystallography
Printed in Singapore – all rights reserved

that the specimen is comprised of a 70:30 mixture of the two twin domains. If the twin fraction is sufficiently close to 1/2, a situation referred to as 'perfect twinning' (as opposed to 'partial twinning'), the diffraction data set acquires erroneously high symmetry. In such cases, the presence of twinning can usually be detected by its effect on the intensity statistics of the data set (Rees, 1980; Stanley, 1972). However, it is not possible to recover the true individual intensities from the observed intensities in cases of perfect or near-perfect twinning. This is so because an equal mixture of twin-related reflections produces only a single unique observation, namely the average of the two underlying crystallographic intensities. Nonetheless, it is sometimes possible to determine macromolecular structures from perfectly twinned specimens. If reasonable phases can be obtained despite the absence of accurate structure-factor amplitudes, then a starting model can be obtained and the effects of twinning on the observed intensities can be taken into account later during atomic refinement.

When the hemihedral twin fraction is not near 1/2, then it is possible in principle to recover the true intensities from the observed intensities (Grainger, 1969). Because of the unequal weighting of the separate twin components, the twin-related observations are not equal to each other and two independent equations can be written in terms of the two unknown true intensities. Having an accurate value for the twin fraction is necessary in order to obtain accurate values for the crystallographic intensities, which may be vital at early stages of structure determination. In addition, accurate values for the twin fractions are required during the later stages (*e.g.* during atomic refinement) in order to obtain good agreement between calculated quantities and observed intensities. Multiple methods have been developed for estimating the hemihedral twin fraction α. Twinning affects the diffraction intensity statistics (Stanley, 1972; Rees, 1980, 1982). In particular, it causes twin-related observations to be more similar to each other than would be expected for two independent reflection intensities and the magnitude of the effect depends on the degree of twinning (*i.e.* the twin fraction). Good estimates of the twin fraction can therefore be obtained from statistical analyses of pairs of twin-related reflections (Britton, 1972; Yeates, 1988; Fisher & Sweet, 1980; Rees, 1982).

In recent years, at least four examples have been reported of macromolecular crystals that exhibit a higher form of merohedral twinning referred to as tetartohedral twinning (Barends *et al.*, 2005; Gayathri *et al.*, 2007; Rosendal *et al.*, 2004; Anand *et al.*, 2007). In tetartohedral twinning there are four different orientations for the twin domains and each observed intensity is therefore the weighted sum of four independent crystallographic intensities. For macromolecular crystals, the space groups $P3$, $P3_1$ and $P3_2$ are the only ones that allow true tetartohedral twinning – the underlying hexagonal lattice has 622 rotational symmetry (order 12), whereas the rotational symmetry of these space group is only of order 3 – but other space groups can allow pseudo-tetartohedral twinning for fortuitous unit-cell geometries (Anand *et al.*, 2007). From another perspective, a crystal that supports tetartohedral twinning is one that could be hemihedrally twinned in multiple distinct ways. For example, a crystal in space group $P3$ can be hemihedrally twinned in ways that cause the diffraction data to approach either $P321$, $P312$ or $P6$. Tetartohedral twinning causes a crystal to have the appearance of being twinned in these multiple ways simultaneously. We are not aware of previous methods for estimating the twin fractions for tetartohedral twinning. Here, we provide equations for estimating tetartohedral twin fractions from intensity data by a statistical comparison of quadruplets of twin-related reflections. Tests on synthetic data are provided for illustration. Application of the equations to the structure determination of a real protein crystal twinned by tetartohedry will be described in a subsequent paper (Yu *et al.*, in preparation).

2. Results and discussion

2.1. Determination of tetartohedral twin fractions

As noted above, partial tetartohedral twinning gives rise to partial symmetry about three distinct types of twin operations

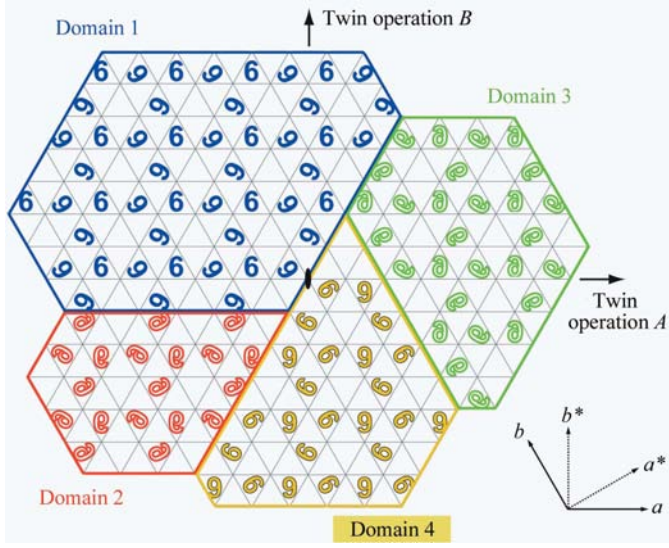

Figure 1
An illustration of a hypothetical tetartohedrally twinned crystal specimen. The true space group is $P3$. The figure '6' illustrates a single molecule. The four differently oriented twin domains are shown in different colors; in a real crystal these domains might be interspersed throughout the specimen. The red and green molecules are drawn as hollow outlines to emphasize their upside-down orientation in the layer. Three distinct twin operations are present. They constitute symmetry elements of the hexagonal lattice, but not of the $P3$ space-group symmetry. The two twin operations that lie perpendicular to the threefold axis of symmetry are shown as arrows. The third twin operation is parallel to the threefold axis and is indicated by the black symbol in the center. Each of the twin operations exchanges the four twin domains in pairs. For example, operation A exchanges twin domain 1 with 2 and 3 with 4. Operation B exchanges domain 1 with 3 and 2 with 4. Operation C exchanges domain 1 with 4 and 2 with 3. The diffraction patterns from the four domains overlap, so that each observed intensity is the sum of four distinct crystallographic intensities weighted according to the twin fractions (or fractional volumes) of the four domains. With the choice of unit-cell directions shown, the three twin operators A, B and C would correspond to exchanging reflection hkl with $(h, -h-k, -l)$, $(-h, h+k, -l)$ and $(-h, -k, l)$, respectively.

(Fig. 1). Each twin operation interrelates all of the domains so that the degree to which the diffraction data set obeys any of the twin operations depends on all four of the twin fractions simultaneously. As a result, the tetartohedral twin fractions cannot be adequately described in terms of degrees of partial twinning about the separate twin operations. Instead, the twin fractions must be understood according to their combined effects on quadruplets of twin-related reflections. The observed intensities of four twin-related reflections depend on four true intensities as follows:

$$J_1 = \alpha_1 I_1 + \alpha_2 I_2 + \alpha_3 I_3 + \alpha_4 I_4 \quad (1a)$$
$$J_2 = \alpha_2 I_1 + \alpha_1 I_2 + \alpha_4 I_3 + \alpha_3 I_4 \quad (1b)$$
$$J_3 = \alpha_3 I_1 + \alpha_4 I_2 + \alpha_1 I_3 + \alpha_2 I_4 \quad (1c)$$
$$J_4 = \alpha_4 I_1 + \alpha_3 I_2 + \alpha_2 I_3 + \alpha_1 I_4. \quad (1d)$$

The I_n indicate true crystallographic intensities and the J_n represent observed intensities. α_m refers to the twin fraction for domain m; the twin fractions have a range between 0 and 1 and sum to unity. The indices on I and J refer to a particular twin-related reflection within a quadruplet. For example, if I_1 refers to $I(hkl)$, then I_2 refers to the reflection related to hkl by the twin operation that relates twin domains 1 and 2 (Fig. 1). The assignment of subscripts to twin operations is arbitrary, but according to the assignments made in Fig. 1, I_1, I_2, I_3 and I_4 would refer to $I(h, k, l)$, $I(h, -h-k, -l)$, $I(-h, h+k, -l)$ and $I(-h, -k, l)$, respectively.

During the later stages of structure determination, for example after a partial model is in hand, applications of (1) above should be straightforward. For example, once estimates of the true crystallographic intensities can be obtained from the model then the twin fractions in (1) can be treated as unknowns in a large system of linear equations and good estimates for the twin fractions can be obtained easily by linear least squares (Sheldrick & Schneider, 1997). Similarly, in the later stages of structure determination the calculated crystallographic intensities, together with the twin fractions, can be used to obtain relatively unbiased estimates of the true crystallographic intensities to be used as targets in atomic refinement. For example, assuming the α_m are known, J_1 has been measured and estimates for I_2, I_3 and I_4 can be calculated from the current model, (1a) can be rearranged to obtain a value for the crystallographic intensity I_1 that is not strongly biased by the value of I_1 calculated from the current model. This would parallel one of the approaches that has been taken in the refinement of protein structures from hemihedrally twinned specimens (Redinbo & Yeates, 1993; Ito *et al.*, 1995). Such applications of (1) require the use of intensities calculated from a model structure. However, at the earliest stages of structure determination this information is usually not available and obtaining estimates of the tetartohedral twin fractions and the underlying crystallographic intensities is then a considerably more challenging problem.

The problem we address here is to obtain statistical estimates for the α_m based on a large set of observed intensities (J), knowing only the expected probability distribution for the true crystallographic intensities (I) but not their individual values. For the simpler problem of hemihedral twinning, analyses of the differences between pairs of twin-related observed intensities lead to expressions for the single twin fraction α (Yeates, 1988; Rees, 1982). In the case of tetartohedral twinning, taking differences between pairs of observed intensities (1) leads to complex equations involving all four true intensities and all four twin fractions. Instead, if properly chosen sums and differences involving all four twin-related observed intensities are evaluated, simpler equations are obtained which can then be rearranged to give statistical estimates for the twin fractions, as explained below:

$$\begin{aligned}(J_1 + J_2) - (J_3 + J_4) &= (\alpha_1 I_1 + \alpha_2 I_2 + \alpha_3 I_3 + \alpha_4 I_4) \\ &+ (\alpha_2 I_1 + \alpha_1 I_2 + \alpha_4 I_3 + \alpha_3 I_4) \\ &- (\alpha_3 I_1 + \alpha_4 I_2 + \alpha_1 I_3 + \alpha_2 I_4) \\ &- (\alpha_4 I_1 + \alpha_3 I_2 + \alpha_2 I_3 + \alpha_1 I_4) \\ &= I_1(\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4) \\ &+ I_2(\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4) \\ &- I_3(\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4) \\ &- I_4(\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4) \\ &= (\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4)[(I_1 + I_2) - (I_3 + I_4)]. \end{aligned} \quad (2)$$

Then, making use of the condition that $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$,

$$(J_1 + J_2) - (J_3 + J_4) = [2(\alpha_1 + \alpha_2) - 1][(I_1 + I_2) - (I_3 + I_4)]. \quad (3)$$

This shows that particular additive and subtractive combinations of four twin-related observed intensities are equal to that same combination of true intensities multiplied by a factor that is constant over all reflections and which reflects a sum of two twin fractions, namely $\alpha_1 + \alpha_2$ in the equation above. The statistical behavior of the quantities $(J_1 + J_2 - J_3 - J_4)$ can be measured, while the quantities $(I_1 + I_2 - I_3 - I_4)$ have a certain expected statistical behavior (*e.g.* based on Wilson statistics; Wilson, 1949). A comparison of the statistical behavior of the two quantities therefore yields information about the quantity $\alpha_1 + \alpha_2$. Adopting a strategy parallel to that introduced for hemihedral twinning (Yeates, 1988), we first divide by the sum of the quartet of twin-related intensities (noting that $J_1 + J_2 + J_3 + J_4 = I_1 + I_2 + I_3 + I_4$); this leads to a simplification with practical advantages in terms of making resolution-dependent normalization of intensities unnecessary,

$$\frac{(J_1 + J_2) - (J_3 + J_4)}{(J_1 + J_2 + J_3 + J_4)} = [2(\alpha_1 + \alpha_2) - 1]\frac{(I_1 + I_2) - (I_3 + I_4)}{(I_1 + I_2 + I_3 + I_4)}. \quad (4)$$

One way to proceed from this point is to square the equation above and evaluate the variances of the probability distributions on both sides. However, simpler results are obtained by taking absolute values; sign ambiguities (discussed later) arise in either case. Next, taking the expected values (or the average over all reflections) on both sides and rearranging, one obtains

$$|[2(\alpha_1+\alpha_2)-1]| = \frac{\left\langle\left|\dfrac{(J_1+J_2)-(J_3+J_4)}{J_1+J_2+J_3+J_4}\right|\right\rangle}{\left\langle\left|\dfrac{(I_1+I_2)-(I_3+I_4)}{I_1+I_2+I_3+I_4}\right|\right\rangle}. \quad (5)$$

In the context of a different problem (Padilla & Yeates, 2003), the expected value of the expression in the denominator above was determined to be 3/8 for acentric reflections, giving

$$|[2(\alpha_1+\alpha_2)-1]| = \frac{8}{3}\left\langle\left|\frac{(J_1+J_2)-(J_3+J_4}{J_1+J_2+J_3+J_4}\right|\right\rangle \quad \text{or}$$

$$[2(\alpha_1+\alpha_2)-1] = \pm\frac{8}{3}\left\langle\left|\frac{(J_1+J_2)-(J_3+J_4)}{J_1+J_2+J_3+J_4}\right|\right\rangle. \quad (6a)$$

By a similar process, related expressions are obtained for the other pairs of twin fractions,

$$|[2(\alpha_1+\alpha_3)-1]| = \frac{8}{3}\left\langle\left|\frac{(J_1+J_3)-(J_2+J_4)}{J_1+J_2+J_3+J_4}\right|\right\rangle \quad \text{or}$$

$$[2(\alpha_1+\alpha_3)-1] = \pm\frac{8}{3}\left\langle\left|\frac{(J_1+J_3)-(J_2+J_4)}{J_1+J_2+J_3+J_4}\right|\right\rangle. \quad (6b)$$

$$|[2(\alpha_1+\alpha_4)-1]| = \frac{8}{3}\left\langle\left|\frac{(J_1+J_4)-(J_2+J_3)}{J_1+J_2+J_3+J_4}\right|\right\rangle \quad \text{or}$$

$$[2(\alpha_1+\alpha_4)-1] = \pm\frac{8}{3}\left\langle\left|\frac{(J_1+J_4)-(J_2+J_3)}{J_1+J_2+J_3+J_4}\right|\right\rangle. \quad (6c)$$

Solutions for the twin fractions are obtained by taking appropriate linear combination of the equations. For example, one allowable solution is obtained by first choosing positive values for all three ambiguous signs in (6*a*)–(6*c*) and then adding the three equations together. On the left side, $[2(\alpha_1+\alpha_2)-1]+[2(\alpha_1+\alpha_3)-1]+[2(\alpha_1+\alpha_4)-1] = 4\alpha_1-1$. Setting this equal to the sum of the right-hand sides and rearranging gives a solution for α_1. The other twin fractions are obtained likewise from other linear combinations of (6*a*)–(6*c*). Despite the lengths of the expressions (below), they are trivial to compute.

$$\alpha_1 = \left[1+\frac{8}{3}\left\langle\frac{|J_1+J_2-J_3-J_4|+|J_1+J_3-J_2-J_4|+|J_1+J_4-J_2-J_3|}{J_1+J_2+J_3+J_4}\right\rangle\right]/4 \quad (7a)$$

$$\alpha_2 = \left[1+\frac{8}{3}\left\langle\frac{|J_1+J_2-J_3-J_4|-|J_1+J_3-J_2-J_4|-|J_1+J_4-J_2-J_3|}{J_1+J_2+J_3+J_4}\right\rangle\right]/4 \quad (7b)$$

$$\alpha_3 = \left[1+\frac{8}{3}\left\langle\frac{-|J_1+J_2-J_3-J_4|+|J_1+J_3-J_2-J_4|-|J_1+J_4-J_2-J_3|}{J_1+J_2+J_3+J_4}\right\rangle\right]/4 \quad (7c)$$

$$\alpha_4 = \left[1+\frac{8}{3}\left\langle\frac{-|J_1+J_2-J_3-J_4|-|J_1+J_3-J_2-J_4|+|J_1+J_4-J_2-J_3|}{J_1+J_2+J_3+J_4}\right\rangle\right]/4. \quad (7d)$$

The equations above represent just one possible solution. The ambiguity that arises is best illustrated with a hypothetical example. If an unknown specimen had (α_1, α_2, α_3, α_4) = (0.40, 0.15, 0.20, 0.25), the three values that would appear on the left side of (6*a*)–(6*c*) would be [$|2\cdot(0.40+0.15)-1|$, $|2\times(0.40+0.20)-1|$, $|2\times(0.40+0.25)-1|$] = (0.10, 0.20, 0.30). There are eight sets of twin fractions that will produce this same set of values in (6*a*)–(6*c*) (based on the eight ways of choosing the ambiguous signs in the three equations). They are

(α_1, α_2, α_3, α_4) =

(0.40, 0.15, 0.20, 0.25)		(0.10, 0.35, 0.30, 0.25)
(0.15, 0.40, 0.25, 0.20)	or	(0.35, 0.10, 0.25, 0.30)
(0.20, 0.25, 0.40, 0.15)		(0.30, 0.25, 0.10, 0.35)
(0.25, 0.20, 0.15, 0.40)		(0.25, 0.30, 0.35, 0.10).

The eight solutions occur in two groups as illustrated above. One of the groups represents the correct solution and the other represents a false solution. The four solutions within a group represent equivalent solutions that arise from a different arbitrary assignment of domain numbers to the different domain orientations. There is no distinction regarding which domain is designated number 1, although the subsequent domain assignments must be consistent with the twin operations (*i.e.* domain 2 is related to domain 1 by the first twin operation as in Fig. 1). The four equivalent solutions can be viewed simply as different assignments of domain 1. Interestingly, the alternate group of four solutions is related to the first group by subtraction of every element from 1/2. Evidently, such complementary sets of twin fractions give observed intensity distributions that are indistinguishable, at least by the kind of intensity comparisons described here. There are some instances where one of the solutions would contain a negative value for one of the twin fractions, which would rule out that solution. In other cases, such as that illustrated above, both are plausible. We show subsequently how the correct solution can be distinguished.

It should be noted that the treatment presented here assumes that the quartet of twin-related reflections are all crystallographically distinct from each other. There are zones of reflections for which this is not the case; these need to be discarded in order to avoid systematically overestimating the degree of twinning. For example, a reflection in the *hk*0 zone is equivalent by Friedel's law (*i.e.* has the same intensity aside from anomalous scattering effects) to the $-h$, $-k$, 0 reflection, but that reflection is also one of those related to the given reflection by a twin operation, namely the twofold twin operation along the *c* axis in the *P*3 space group. Similar arguments apply to reflections falling in zones that contain the origin and are perpendicular to any of the real or reciprocal unit-cell axes in the *ab* plane. Essentially, any reflection that would be centric in 622 must be excluded.

Alternatively, we note that a second approach for estimating tetartohedral twin fractions could be developed based exclusively on the classes of reflections that were discarded in the previous paragraph. For a reflection in one of the zones indicated there are not four distinct reflections, but only two,

so that the four equations in (1) collapse to two. For instance, if only reflections in the $hk0$ zone are considered (and Friedel's law is obeyed), then $I_1 = I_4$, $I_2 = I_3$, $J_1 = J_4$ and $J_2 = J_3$. Based on these special zones of reflections, equations for the twin fractions can be obtained that are somewhat simpler than those developed here, but that minor advantage is outweighed by a significant loss in statistical power owing to the lower number of reflections considered.

Once the twin fractions have been estimated, they can be used to obtain values for the true crystallographic intensities (I) from the observed intensities (J). For each quadruplet of twin-related reflections, (1a)–(1d) give four linear equations in four unknowns. (1) can be rewritten as a matrix multiplication relating a vector of four observed intensities to four unknown crystallographic intensities,

$$\begin{pmatrix} J_1 \\ J_2 \\ J_3 \\ J_4 \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ \alpha_2 & \alpha_1 & \alpha_4 & \alpha_3 \\ \alpha_3 & \alpha_4 & \alpha_1 & \alpha_2 \\ \alpha_4 & \alpha_3 & \alpha_2 & \alpha_1 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{pmatrix} \quad \text{or} \quad \mathbf{J} = \mathbf{TI} \qquad (8)$$

and

$$\mathbf{I} = \mathbf{T}^{-1}\mathbf{J}, \qquad (9)$$

where $\mathbf{T}$ is a matrix comprised of the tetartohedral twin fractions. For each quadruplet of twin-related reflections, the I_n can be obtained from the J_n by inverting the matrix $\mathbf{T}$, assuming $\mathbf{T}$ is not degenerate. The properties of $\mathbf{T}$ determine whether or not accurate values can be obtained for the true crystallographic intensities. The eigenvalues of $\mathbf{T}$ describe the extent to which the four-dimensional space of crystallographic intensities is compressed by the effects of tetartohedral twinning. If any of the eigenvalues approaches zero then the determinant of $\mathbf{T}$ approaches zero and (8) cannot be inverted without unacceptably large magnifications of the measurement errors associated with the observed intensities. As a rough approximation, the largest eigenvalue of $\mathbf{T}^{-1}$ may be taken as a measure of the error amplification that might arise by detwinning a quadruplet of reflections in the worst-case scenario (*i.e.* depending on the particular values of the intensities involved). Conditions under which $\mathbf{T}$ is degenerate include all cases where the sum of any two twin fractions equals 1/2. Therefore, detwinning observed intensities is problematic for any crystal specimen for which the sum of two twin fractions approaches 1/2.

Finally, it is important to note that estimating tetartohedral twin fractions by comparing twin-related intensities is subject to the same caveats that apply to the case of hemihedral twinning. In order to obtain accurate estimates of the twin fractions, true crystallographic intensities that are related by the twin operation(s) must be statistically independent; this assumption was implicit in the substitution of the denominator of (5) by the quantity 3/8. The presence of noncrystallographic symmetry (NCS) can cause the assumption of statistical independence to be false, particularly for reflections in the lower resolution range. In practice, NCS often interferes with analyses of twinning, as twinning is frequently associated with an NCS axis that is nearly coincident with a potential twin operation; this is true of all four of the previously reported cases of tetartohedral twinning. Cautionary steps should therefore be taken, in hemihedral as well as tetartohedral cases, in order to avoid confusing NCS with twinning. Firstly, NCS can lead to similarity between pairs (or quadruplets) of true intensities related by potential twinning, but this effect tends to diminish at higher resolution (*i.e.* smaller Bragg spacing), while the effects of twinning do not. Estimates of twin fractions should therefore be made on the basis of reflections in increasing resolution ranges. In a case where NCS is present but twinning is not, the estimated twin fractions should typically decrease towards zero as the resolution range increases. Secondly, if twinning is indeed present and the

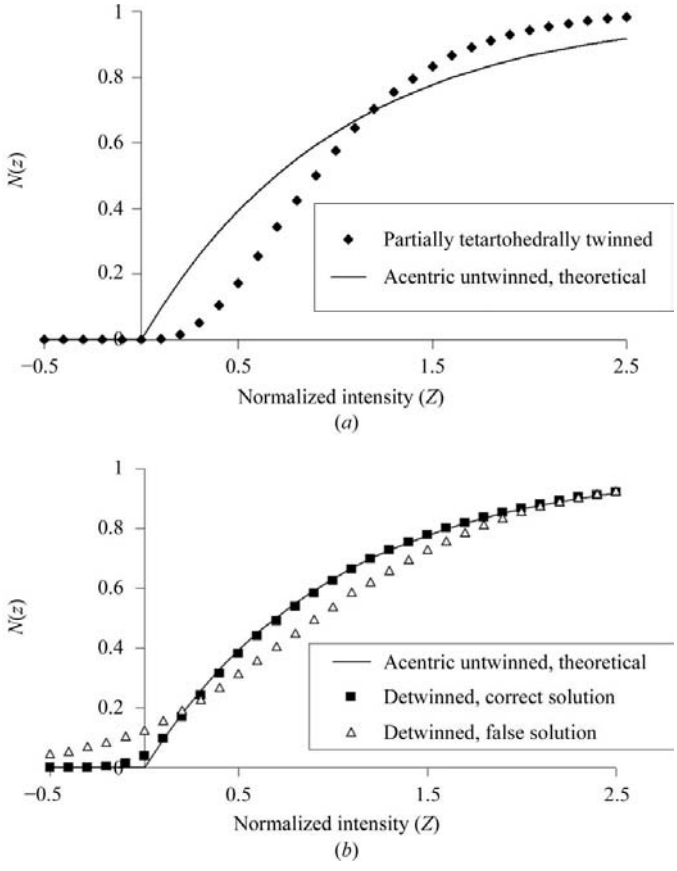

Figure 2
An illustration of the cumulative intensity statistics for the simulated data used as a test case. (*a*) The intensity distribution for the tetartohedrally twinned data is shown by the diamond-shaped data points. The twin fractions used in simulating the tetartohedral twinning were (0.40, 0.30, 0.25, 0.05). The theoretical curve expected for ordinary (untwinned acentric) data according to Wilson statistics is shown as a thin solid curve. The curve for the synthetic data is shifted in the direction expected for twinning. (*b*) A single ambiguity arises in the determination of twin fractions from the statistical method presented here; the correct solution (0.40, 0.30, 0.25, 0.05) and a false solution (0.10, 0.20, 0.25, 0.45) are identified as possibilities. The intensity distributions are shown for data detwinned according to the two potential solutions. The data obtained under the correct solution are shown as square data points, while the data obtained under the false solution are shown as triangular data points. As in (*a*), the theoretical curve for ordinary data is shown as a thin solid curve. The incorrect choice for the twin fractions gives a significant number of negative intensities and a distribution that does not obey Wilson statistics.

twin fractions are significant, then this should be evident in analyses of overall intensity statistics (Padilla & Yeates, 2003; Stanley, 1972; Rees, 1980), which do not rely on comparisons of potentially twin-related reflections and which are therefore not affected severely by NCS. These steps are important to avoid systematic error or misinterpretation of twinning. Furthermore, random errors in measured intensities also lead to errors in estimating twin fractions. Mathematical treatments of errors have been developed for the case of hemihedral twinning (Lunin *et al.*, 2007; Rees, 1982). This has not been attempted here for the case of tetartohedral twinning, but it remains true that measurement errors tend to increase the differences between twin-related measurements. This leads to a slight underestimation of the amount of twinning present.

2.2. Tests using synthetic intensity data

Synthetic intensity data were generated in order to test the equations above. 10 000 quadruplets of twin-related reflection intensities were generated according to an exponential probability distribution as expected for acentric reflections according to Wilson statistics. These were then mixed together to simulate tetartohedral twinning according to (1) using the following test values for the twin fractions: (α_1, α_2, α_3, α_4) = (0.40, 0.30, 0.25, 0.05). In order to simulate the effects of small measurement errors, an error term, uniformly distributed from −5% to +5%, was added to each simulated observed intensity to give an average error of 2.5%. These values were then used in (7) to obtain estimates of the twin fractions. The following estimates were obtained: (α_1, α_2, α_3, α_4) = (0.400, 0.301, 0.251, 0.048) [plus three equivalent permutations (0.301, 0.400, 0.048, 0.251), (0.251, 0.048, 0.400, 0.301) and (0.048, 0.251, 0.301, 0.400)] or (α_1, α_2, α_3, α_4) = (0.100, 0.199, 0.249, 0.452) [plus three equivalent permutations (0.199, 0.100, 0.452, 0.249), (0.249, 0.452, 0.100, 0.199) and (0.452, 0.249, 0.199, 0.100)]. One of the solutions from the first set listed matches the correct values, with small deviations arising from the errors that were introduced into the twinned intensities.

How can the correct group of solutions be discriminated from the incorrect group? Equation (9) can be used to calculate values for the true crystallographic intensities from the observed intensities. The correct solution set for the twin fractions should generate a set of I_n that obey ordinary (*i.e.* untwinned) Wilson statistics, while the incorrect set of twin fractions should not. This assertion was tested by calculating values for the I_n under the two groups of solutions; only one solution from each group needs to be evaluated because equivalent solutions within one group give the same intensities permuted according to the twin operations and so have the same intensity distributions. The cumulative intensity distributions under the two possible solutions are shown in Fig. 2(*b*). Note that the incorrect set of twin fractions gives a distribution that is inconsistent with Wilson statistics and that in particular a significant number of negative calculated intensities are obtained. As expected, the correct solution gives a set of calculated crystallographic intensities whose distribution matches Wilson statistics.

A final issue is whether accurate values for the true intensities can be extracted from the twinned observations. Using the same simulated data as above, the agreement was tested between the crystallographic intensities obtained using (9) (based on the estimated twin-fraction values above) and the original values of the synthetic crystallographic intensities from which the twinned values (J_n) were computed. The average error in the extracted intensities, evaluated in the form of an R factor, was 13%. This is approximately five times larger than the average error introduced into the simulated observed intensities, which was 2.5%. This illustrates the magnification of errors expected by detwinning: the magnitudes of the eigenvalues of $\mathbf{T}^{-1}$ obtained from the twin fractions employed in this simulated test case were (1.0, 2.5, 3.3 and 10).

3. Conclusions

Tetartohedral twinning presents a particularly complex challenge for interpreting observed intensities in terms of true crystallographic intensities. This is particularly true at the earliest stages of a crystallographic investigation, before additional data (such as calculated intensities from a partial model) are available. Four twin fractions (which sum to one) govern the behavior of the observed intensities. When only the observed tetartohedrally twinned intensities are available, values for the twin fractions must be estimated by statistical methods. Equations are described here for estimating tetartohedral twin fractions based on an analysis of simple sums and differences between quadruplets of twin-related observed reflection intensities. A test case using synthetic intensity data shows that correct values for the twin fractions can be obtained and that those values are not strongly affected by random measurement errors. A single ambiguity in the twin-fraction values arises in the statistical analysis and an approach for resolving the ambiguity is illustrated.

Based on the estimated twin fractions, the observed (tetartohedrally twinned) intensities can be effectively detwinned to yield true crystallographic intensities. However, the detwinning process can significantly magnify the errors inherent in the observed data. The extent of this error magnification is discussed in terms of the twin-fraction values. In some cases it should be possible to obtain reasonably accurate values for the true crystallographic intensities by detwinning, while in other cases this will not be possible. Efforts to determine atomic structures in such cases will have to rely on the use of twinned intensities in crystallographic calculations. Knowledge of the twin fractions and an understanding of the underlying twinning equations will also be important in these calculations.

The authors thank Jian-Hua He, Doug Rees, Yingssu Tsai and George Sheldrick for critical readings of the manuscript, and the BER program of the DOE Office of Science for support.

References

Anand, K., Schulte, A., Fujinaga, K., Scheffzek, K. & Geyer, M. (2007). *J. Mol. Biol.* **370**, 826–836.

Barends, T. R. M., de Jong, R. M., van Straaten, K. E., Thunnissen, A.-M. W. H. & Dijkstra, B. W. (2005). *Acta Cryst.* D**61**, 613–621.

Britton, D. (1972). *Acta Cryst.* A**28**, 296–297.

Buerger, M. J. (1960). *Crystal Structure Analysis.* New York: John Wiley & Sons.

Dauter, Z., Botos, I., LaRonde-LeBlanc, N. & Wlodawer, A. (2005). *Acta Cryst.* D**61**, 967–975.

Fisher, R. G. & Sweet, R. M. (1980). *Acta Cryst.* A**36**, 755–760.

Gayathri, P., Banerjee, M., Vijayalakshmi, A., Azeez, S., Balaram, H., Balaram, P. & Murthy, M. R. N. (2007). *Acta Cryst.* D**63**, 206–220.

Grainger, C. T. (1969). *Acta Cryst.* A**25**, 427–434.

Ito, N., Komiyama, N. H. & Fermi, G. (1995). *J. Mol. Biol.* **250**, 648–658.

Lunin, V. Y., Lunina, N. L. & Baumstark, M. W. (2007). *Acta Cryst.* D**63**, 1129–1138.

Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* D**59**, 1124–1130.

Parsons, S. (2003). *Acta Cryst.* D**59**, 1995–2003.

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.

Rosendal, K. R., Sinning, I. & Wild, K. (2004). *Acta Cryst.* D**60**, 140–143.

Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.

Stanley, E. (1972). *J. Appl. Cryst.* **5**, 191–194.

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

Yeates, T. O. (1988). *Acta Cryst.* A**44**, 142–144.

Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.