

6. Refinement of Twinned Structures; Absolute Structure

A typical definition of a twinned crystal is the following: "Twins are regular aggregates consisting of crystals of the same species joined together in some definite mutual orientation" (Giacovazzo, 1992). So for the description of a twin two things are necessary: a description of the orientation of the different species relative to each other (twin law) and the fractional contribution of each component. The *twin law* can be expressed as a matrix that transforms the *hkl* indices of one species into the other.

6.1 Twin refinement method

In SHELXL the twin refinement method of Pratt, Coyle & Ibers (1971) and Jameson, Schneider, Dubler & Oswald (1982) has been implemented. F_c^2 values are calculated by:

$$(F_c^2)^* = \text{osf}^2 \sum_{m=1}^n k_m F_{c_m}^2$$

where *osf* is the overall scale factor, k_m is the fractional contribution of twin domain *m* and F_c is the calculated structure factor of twin domain *m*. The sum of the fractional contributions k_m must be unity, so (n-1) of them can be refined and k_1 is calculated by:

$$k_1 = 1 - \sum_{m=2}^n k_m$$

In SHELXL two kinds of twins are distinguished:

(a) For twins in which the reciprocal lattices exactly coincide (twinning by merohedry or pseudo-merohedry), the procedure is relatively simple. The command `TWIN r11 r12 r13 r21 r22 r23 r31 r32 r33 n` defines the twin law. **R** as the matrix that transforms the *hkl* indices of one component into the other and *n* is the number of twin domains. **R** is applied (n-1) times; the default value of *n* is 2.

(b) In cases where only some reflections have contributions from more than one domain (non-merohedral twins or twinning by reticular merohedry) the *.hkl* file must be edited and the index transformations applied to individual contributors, which are also assigned component numbers. The code `HKLF 5` is used to read in this file; no `TWIN` command should be used.

In both cases, starting values of the fractional contributions are input with the instruction `BASF k2 ... kn`; the k_m values will be refined. Note that (in the new version of SHELXL) linear restraints may be applied to these *k* values by means of `SUMP` instructions; this can be very useful to prevent instabilities in the early stages of refinement. For this purpose $k_2 \dots k_n$ are assigned parameter numbers immediately following the free variables.

6.2 Absolute structure

Even if determination of absolute configuration is not one of the aims of the structure determination, it is important to refine **every** non-centrosymmetric structure as the correct *absolute structure* in order to avoid introducing systematic errors into the bond lengths etc. In some cases the absolute structure will be known with certainty (e.g. proteins), but in others it has to be deduced from the X-ray data. Generally speaking, a single phosphorus or heavier atom suffices to determine an absolute structure using Cu-K α radiation, and with accurate high-resolution low-temperature data including Friedel opposites such an atom may even suffice for Mo-K α .

In the course of the final structure factor calculation, the program estimates the absolute structure parameter x (Flack, 1983) and its esd. x is the fractional contribution of the inverted component of a 'racemic twin'; it should be zero if the absolute structure is correct, unity if it has to be inverted, and somewhere between 0 and 1 if racemic twinning is really present. Thus the above formulas apply with $n=2$ and $\mathbf{R} = (-1\ 0\ 0, 0\ -1\ 0, 0\ 0\ -1)$.

It is a bonus of the refinement against F^2 that this calculation is a 'hole in one' and doesn't require expensive iteration. A comparison of x with its esd provides an indication as to whether the refined absolute structure is correct or whether it has to be 'inverted'; the program prints a suitable warning should this be necessary. This attempt to refine x 'on the cheap' is reliable when the true value of x is close to zero, but may produce a (possibly severe) underestimate of x for structures which have to be inverted, because x is correlated with positional and other parameters which have not been allowed to vary. Effectively these parameters have adapted themselves to compensate for the wrong (zero) value of x in the course of the refinement, and need to be refined with x to eliminate the effects of correlation. These effects will tend to be greater when the correlation terms are greater, e.g. for polar space groups and for poor data to parameter ratios (say less than 8:1). x can be refined at the same time as all the other parameters using the TWIN instruction with the default matrix $\mathbf{R} = (-1\ 0\ 0, 0\ -1\ 0, 0\ 0\ -1)$ and BASF with one parameter (x); this implies racemic twinning and so is refined exactly as for other simple cases of twinning. Refinement of racemic twinning should normally only be attempted towards the end of the refinement after all non-hydrogen atoms have been located. If racemic twinning is refined in this way, the automatic calculation of the Flack x parameter in the final structure factor cycle is suppressed, since the BASF parameter is x .

For most space groups 'inversion' of the structure simply involves inserting an instruction 'MOVE 1 1 1 -1' before the first atom. Where the space group is one of the 11 enantiomorphous pairs [e.g. P3₁ and P3₂] the translation parts of the symmetry operators need to be inverted as well to generate the other member of the pair. There are seven cases for which, if the standard setting of the International Tables for Crystallography has been used, inversion in the origin does **not** lead to the inverted absolute structure. This problem was probably first described in print by Parthe & Gelato (1984) and Bernardinelli & Flack (1985), but had been investigated previously by D. Rogers (personal communication to GMS, ca. 1980).

The offending space groups and corresponding correct MOVE instructions are:

```
Fdd2      MOVE .25 .25 1 -1 I41cd      MOVE 1 .5 1 -1
I41      MOVE 1 .5 1 -1      I42d      MOVE 1 .5 .25 -1
I4122    MOVE 1 .5 .25 -1    F4132    MOVE .25 .25 .25 -1
I41md    MOVE 1 .5 1 -1
```

6.3 Refinement against powder data

Refinement of twinned crystals and refinement against F^2 -values derived from powder data are similar in that several reflections with different indices may contribute to a single F^2 observation. For powder data this requires some small adjustments to the format of the *.hkl* file; the batch number becomes the multiplicity *m*, and where several reflections contribute to the same observation the multiplicity is made positive for the last reflection in the group and negative for the rest.

Although SHELXL may be useful for some high symmetry and hence reasonably well resolved powder and fibre diffraction patterns - the various restraints and constraints should be exploited in full to make up for the poor data/parameter ratio - for normal powder data a Rietveld refinement program would be much more appropriate.

For powder data the least-squares refinement fits the overall scale factor (osf^2 where *osf* is given on the FVAR instruction) times the multiplicity weighted sum of calculated intensities to F_o^2 :

$$(F_c^2)^* = osf^2 [m_1 F_{c1}^2 + m_2 F_{c2}^2 + m_3 F_{c3}^2 + \dots]$$

where the multiplicities of the contributors are given in the place of the batch numbers in the *.hkl* file. Since it is then not possible to define batch numbers as well, BASF cannot be used with powder data.

6.4 Frequently encountered twin laws

The following cases are relatively common:

(a) Twinning by merohedry. The lower symmetry trigonal, tetragonal, hexagonal or cubic Laue groups may be twinned so that they look (more) like the corresponding higher symmetry Laue groups (assuming the *c*-axis unique except for cubic):

```
TWIN 0 1 0 1 0 0 0 0 -1
```

plus one BASF parameter if the twin components are not equal in scattering power. If they are equal, i.e. the twinning is perfect, as indicated by the R_{int} for the higher symmetry Laue group, then the BASF instruction can be omitted and k_1 and k_2 are fixed at 0.5.

(b) Orthorhombic with **a** and **b** approximately equal in length may emulate tetragonal:

```
TWIN 0 1 0 1 0 0 0 0 -1
```

plus one BASF parameter for unequal components.

(c) Monoclinic with beta approximately 90° may emulate orthorhombic:

```
TWIN 1 0 0 0 -1 0 0 0 -1
```

plus one BASF parameter for unequal components.

(d) Monoclinic with **a** and **c** approximately equal and beta approximately 120 degrees may emulate hexagonal [P2₁/c would give absences and possibly also intensity statistics corresponding to P6₃]. There are three components, so n must be specified on the TWIN instruction and the matrix is applied once to generate the indices of the second component and twice for the third component. In German this is called a 'Drilling' as opposed to a 'Zwilling' (with two components):

```
TWIN 0 0 1 0 1 0 -1 0 -1 3
```

plus TWO BASF parameters for unequal components. If the data were collected using an hexagonal cell, then an HKLF matrix would also be required to transform them to a setting with b unique:

```
HKLF 4 1 1 0 0 0 0 1 0 -1 0
```

(e) Rhombohedral obverse/reverse twinning on hexagonal axes.

```
TWIN -1 0 0 0 -1 0 0 0 1
```

6.5 Combined general and racemic twinning

If general and racemic twinning are to be refined simultaneously, n (the last parameter on the TWIN instruction) should be doubled and given a negative sign, and there should be |n|-1 BASF twin component factors (or none, in the unlikely event that all are to be fixed as equal). The inverted components follow those generated using the TWIN matrix, in the same order. Sometimes it is necessary to use this approach to distinguish between possible twin laws for non-centrosymmetric structures, when they differ only in an inversion operator. In a typical example (an organocesium compound), when the TWIN instruction was input as:

```
TWIN 0 1 0 1 0 0 0 0 -1 -4
```

The BASF parameters refined to:

```
BASF 0.33607 0.00001 0.00455
```

Which means that the last two components (the ones involving inversion) can be ignored, and the final refinement performed with the '-4' deleted from the end of the TWIN instruction, and a single BASF parameter. The introduction of twinning reduced the R1-value from 18% to 1.8% in this example. Note that the program does not allow the BASF parameters to become negative, since this would be physically meaningless (this explains the 0.00001 above).

6.6 Processing of twinned and powder data

The HKLF 5 and 6 instructions force MERG 0, i.e. neither a transformation of reflection indices into a standard form nor a sort-merge is performed before refinement. If twinning is specified using the TWIN instruction, any MERG instruction may be used and the default remains MERG 2. Although this is always safe for racemic twinning, there may be other forms of twinning for which it is not permissible to sort-merge first. Whether or not MERG is used, the program ignores all systematically absent contributions, with the result that a reflection is excluded from the data if it is systematically absent for all components.

For both powder (HKLF 6) and twinned data (HKLF 5 or TWIN with HKLF 4), the reflection data are reduced to the 'prime' component, by multiplying F_o^2 by the ratio of the F_c^2 for the prime reflection divided by the total F_c^2 , before performing the analysis of variance and the Fourier calculations. Similarly 'OMIT h k l' refers to the indices of the prime component. The prime component is the one for which the indices have not been transformed by the TWIN instruction (i.e. $m = 1$), or in the case of HKLF 5 or HKLF 6 the component given with positive m (i.e. the last contributor to a given intensity measurement, not necessarily the one with $|m| = 1$).

6.7 The warning signs for twinning

Experience shows that there are a number of characteristic warning signs for twinning. Of course not all of them can be present in any particular example, but if one finds several of them the possibility of twinning should be given serious consideration.

- (a) The metric symmetry is higher than the Laue symmetry.
- (b) The R_{int} -value for the higher symmetry Laue group is only slightly higher than for the lower symmetry Laue group.
- (c) The mean value for $|E^2-1|$ is much lower than the expected value of 0.736 for the non-centrosymmetric case. If we have two twin domains and every reflection has contributions from both, it is unlikely that both contributions will have very high or that both will have very low intensities, so the intensities will be distributed so that there are fewer extreme values.
- (d) The space group appears to be trigonal or hexagonal.
- (e) There are impossible or unusual systematic absences.
- (f) Although the data appear to be in order, the structure cannot be solved.
- (g) The Patterson function is physically impossible.

The following points are typical for non-merohedral twins, where the reciprocal lattices do not overlap exactly and only some of the reflections are affected by the twinning:

- (h) There appear to be one or more unusually long axes, but also many absent reflections.
- (i) There are problems with the cell refinement.
- (j) Some reflections are sharp, others split.
- (k) $K = \text{mean}(F_o^2) / \text{mean}(F_c^2)$ is systematically high for the reflections with low intensity.
- (l) For all of the 'most disagreeable' reflections, F_o is much greater than F_c .

6.8 Conclusions

Twinning usually arises for good structural reasons. When the heavy atom positions correspond to a higher symmetry space group it may be difficult or impossible to distinguish between twinning and disorder of the light atoms; see Hoenle & von Schnering (1988). Since refinement as a twin usually requires only two extra instructions and one extra parameter, in such cases it should be attempted first, before investing many hours in a detailed interpretation of the 'disorder'! Indeed, it has been suggested by G.B. Jameson that all structures (including proteins) that are solved in space groups (such as $P3_1$) that could be merohedrally twinned without changing the systematic absences should be tested for such twinning (possible only present to a minor extent) by:

```
TWIN 0 1 0 1 0 0 0 0 -1
BASF 0.1
```

Refinement of twinned crystals often requires the full arsenal of constraints and restraints, since the refinements tend to be less stable, and the effective data to parameter ratio may well be low. In the last analysis chemical and crystallographic intuition may be required to distinguish between the various twinned and disordered models, and it is not easy to be sure that all possible interpretations of the data have been considered.

I should like to thank Regine Herbst-Irmer who wrote most of this chapter.