

Processing non-merohedral twins and refinement with SHELX

ACA Meeting, Toronto, July 25th 2009

George M. Sheldrick, *Göttingen University*

<http://shelx.uni-ac.gwdg.de/SHELX/>

Non-merohedral twins



Cubic insulin twin



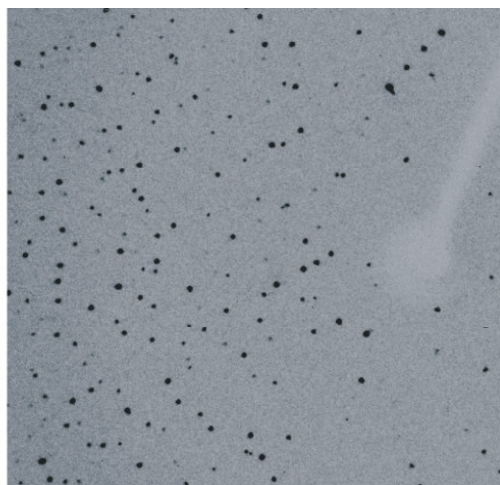
Glucose isomerase triple crystal

Note that whereas the two components of the cubic insulin twin are interpenetrant and so have approximately the same center, the three components of the GI 'drilling' have well separated centers.

Madhumati Sevana

Diffraction pattern of glucose isomerase *drilling*

To index a multiple crystal, first a cell and orientation matrix are found to index as many reflections as possible, subject to the cell being as small as possible. This cell is then rotated (twice in this case) until most of the remaining reflections have been fitted too. The resulting orientation matrices (3 in this case) are used to integrate the data. Some reflections can be integrated independently, in other cases there are two or more overlapping reflections contributing to the same observed intensity.



Madhumati Sevana

Twin refinement with SHELXL

Merohedral and pseudomerohedral twins

In these cases it is usually possible to integrate and scale the data as if the crystal was not twinned. The data should be processed for the true (lower) symmetry, e.g. as P4 when the apparent but incorrect space group is P422. In most cases the data can be read in using SHELX HKLF 4 format and the twinning is specified by a TWIN matrix (which may be applied more than once) and the starting values of the twin mass fractions for the second and subsequent components are given on BASF instructions. For perfect 50:50% twinning the BASF instruction can be omitted, the twin fractions are then held fixed.

Non-merohedral twins

SHELXL HKLF 5 format may require special file preparation, but is a very general format for all eventualities. It may also be used for merohedral and pseudomerohedral twinning when more than one TWIN matrix would be required.

Merohedral twinning

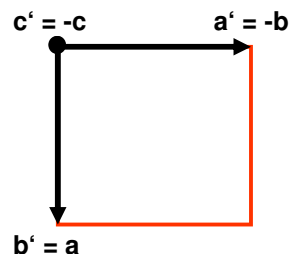
For merohedral twins the reciprocal lattices of the two components overlap exactly; there is no way of resolving the reflections. A typical example is the rotation of the crystal about a two-fold axis at right angles to a fourfold axis. This would cause the space group P4 to appear to be P422. It does not make any difference whether the twofold is along b, corresponding to the reflection index transformation specified by the SHELXL instruction:

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

or whether it is at 45° to a and b:

TWIN 0 1 0 1 0 0 0 0 -1

because combining one of these operators with the 90° rotation operator of the true Laue group will generate the other.



TWIN matrices for merohedral twinning

For merohedral twinning of (chiral) macromolecules in the cubic, hexagonal and tetragonal systems, the TWIN matrix is *always*:

TWIN 0 1 0 1 0 0 0 0 -1

If the twinning is not perfect, one could start with

BASF 0.3

which will be updated by the refinement. This corresponds to a twin fraction of 0.7 for the major component (indices as input) and 0.3 for the component generated from the input indices by applying the TWIN matrix.

Even for trigonal crystals, where there is a choice of TWIN matrices for merohedral twinning, the above matrix is the most common and should be the first one tried! However trigonal merohedral twins with four components (tetrahedral twins) require HKLF 5 format.

Pseudomerohedral twinning

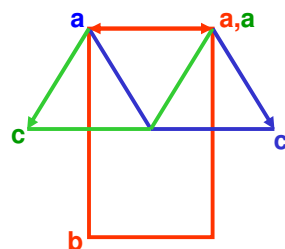
Pseudomerohedral twinning can arise when two axes have accidentally almost exactly the same length or when e.g. angles very close to 90° happen by chance. For example, if a crystal in P2₁2₁2₁ has equal a and b axes, the two twin components can be related by a 90° rotation about c:

TWIN 0 1 0 -1 0 0 0 0 1

and the refinement can be performed using this matrix. This should lead to systematic absences that would look like the space group P4₂2₁2, but cases have been reported in which it looked like P4₃2₁2!

An even more common case for macromolecules is that the apparent space group C22₁ is actually P2₁ and twinned. To refine this with SHELX and HKLF 4, first the data must be reindexed from the red (centered) to the blue (primitive) cell, then a TWIN instruction to express the green cell in terms of the blue:

TWIN -1 0 0 0 1 0 1 0 1



Equivalent reflections and groups (HKLF 5 format)

h k l component (assuming point group mmm)

1 -2 3 1 } equivalent singles
 -1 -2 -3 1

-1 -2 -3 2 — not equivalent to the above singles

-1 -2 -3 -2

2 0 -4 1 } equivalent groups

1 2 -3 -2

-2 0 -4 1

4 1 1 -2

1 -2 -3 -3

-1 1 2 1

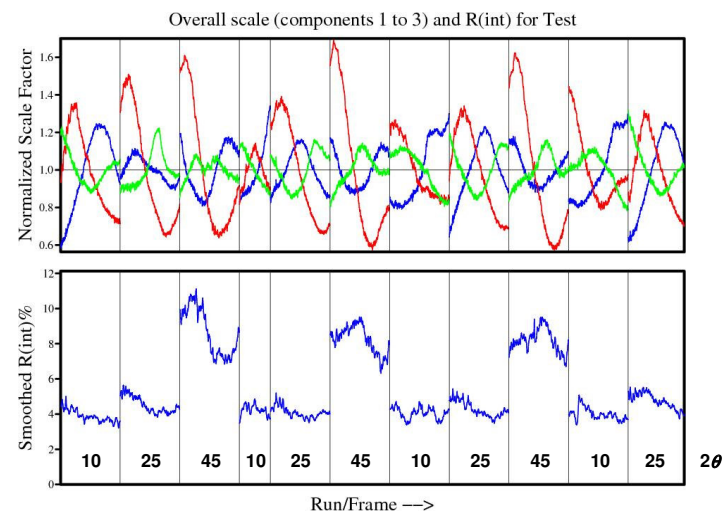
} not equivalent to the other groups shown here

In SHELX HKLF 5 format, a group of overlapping reflections is defined by negative component numbers for all but the last reflection in the group. For scaling purposes the component numbers MUST match.

Scaling non-merohedrally twinned data

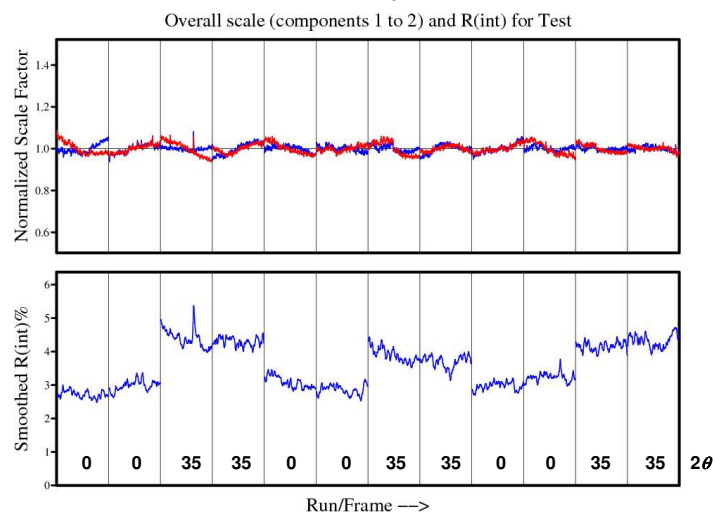
1. Determine scaling and absorption parameters by fitting individual intensities to the mean corrected intensities (averaged over equivalents). Outliers are downweighted but not rejected in this stage. For parameter determination, Friedel opposites should be treated as equivalent (i.e. **Laue group** symmetry imposed). For the remaining calculations it may be better to use the **point group**. **The equivalents may be single reflections or groups of overlapping reflections with the same pattern of contributors.**
2. Delete reflections that are completely incompatible with their equivalents, e.g. reflections blocked by the beam stop etc. Then determine an error model for the remaining reflections by fitting χ^2 to unity to put $\sigma(I)$ onto an absolute scale.
3. Output diagnostic statistics (graphically).
4. Generate SHELX HKLF 4 format files for structure solution and initial refinement and HKLF 5 files for final accurate refinement.

Scale factors and R_{int} for GI-triplet



The smallest crystal (shown in red) was furthest from the center.

Scale factors and R_{int} for cubic insulin



For this interpenetrant twin, the two crystals have approximately the same center and so show little variation in scale.

Generating a HKLF 4 file from a non-merohedral twin

The measured total intensities I_m of the single or composite reflections are given by a set of equations that can be solved to obtain the relative twin masses k_n and the intensities of the unique reflections I_h :

$$I_m = k_1 I_h + k_2 I_h' + \dots + k_n I_h''$$

Since this system of equations may well be ill-conditioned, an approximate partitioning of the overlapped reflections (obtained by profile fitting in the integration stage) is applied in the form of weak restraints (extra observational equations).

This algorithm is robust, converges fast and (by using sparse matrix techniques) can process several million reflections in a few seconds. The twin ratios obtained are close to those from the HKLF 5 refinement with SHELXL.

Inconsistent component indexing

The algorithm to generate HKLF 4 format data revealed a subtle and unexpected elephant trap. In the cases where the Laue symmetry is lower than the metric symmetry of the lattice, the component may be indexed inconsistently, even when the second and subsequent orientation matrices were obtained by rotating the first!

For the insulin twin, a rotation of the cell by ca. 180° about an axis parallel to a face diagonal of the cubic cell [1 1 0] led to inconsistently indexed components. This is similar to the generation of a *merohedral* twin. The only warning sign was a high R_{int} for the least-squares deconvolution.

The scaling is not affected by the inconsistent indexing!

The HKLF 5 format output file

Refinement using HKLF 5 is still almost always slightly more precise than using the deconvoluted HKLF 4 data. It is best to compare the $R1$ value after 'merging for Fourier' at the end of the HKLF 5 refinement so that approximately the same total number of reflections are used. Despite this, for macromolecules it is strongly recommended that the data are detwinned in this way because then all the standard procedures can be used without having to take the twinning into account.

However the currently popular (especially amongst Acta C and E Editors for small) argument that it is essential to use only an 'independent' set of reflections in the HKLF 5 refinement is a red herring. In such a case all that is necessary to get the correct parameter esds is to set the third L.S. parameter for SHELXL to the number of 'observations' in the HKLF 5 file minus the number of reflections in the HKLF 4 file!

The phase problem for merohedral twins

Non-merohedral twins can be solved by standard methods using the HKLF 4 format file obtained by deconvoluting the data. For merohedral twins the situation appears to be more difficult, and as the twin fraction k approaches 0.5, attempts to *detwin* the data become increasingly unstable.

Twinned structures are often successfully solved by molecular replacement. To understand this, remember that the Patterson of a twin is simply the sum of the Pattersons of the components, and MR may be considered as finding solutions that fit the Patterson.

Amazingly, experimental phasing is also often successful. Both the insulin twin and the gi-drilling crystals shown in the pictures could be easily solved by SAD using in-house data and SHELXC/D/E!

In all cases the true (usually lower symmetry) space group should be used for the data.

The 'resolution' of twinned structures

A perfect merohedral twin has half the data to parameter ratio of a normal crystal collected to the same resolution. This is like multiplying the diffraction limit by $(2)^{1/3} = 1.26$.

Thus a twinned 2.4Å dataset actually behaves like a 3.0Å untwinned crystal from the point of view of refinement!

This argument does not apply to non-merohedral twins unless the overlap of the reflections of the twin components is severe.

The choice of program for refining twins

SHELXL was written for small molecule refinement and although it has many nice features – like the sophisticated treatment of *disorder*, the many ways of employing *free variables* and the estimation of *standard deviations* – the program also suffers from limitations for macromolecular refinement. In particular it needs:

1. A better solvent model;
2. Torsion angle restraints (for the side-chains, one would still wish to leave the Ramachandran plot angles unrestrained so that they can be used for verification).
3. A better force field, the current antibumping model is too naive.
4. A better treatment of restraints for the displacement parameters.

So, except possibly at high resolution, now that there are at last good alternatives available it may well be better to use them.

The free *R*-factor and twinned crystals

Some care is needed in selecting the free-*R* reflections for merohedral and pseudomerohedral twins so that twin-related reflections are not split between the two sets.

For non-merohedral twins it would be possible to apply the flags to groups of reflections, but unfortunately the SHELX syntax does not allow this because the minus sign normally used for flagging the free *R* set in HKLF 4 format is being used for something else in HKLF 5 format. The only remedy is to set up separate files for the two sets and run a separate SHELXL job with 0 refinement cycles to calculate the free *R*. However in many cases the same reflection will appear in different singles and/or composites, so it is difficult to avoid contamination the free *R* set completely. This is another good reason to deconvolute the data first to make a HKLF 4 format file.

Difference maps for twinned crystals

In refinement, the sum of the intensities of the twin components is fitted to the observed intensity. But how do we partition the differences to get the 'best' ($|F_o - F_c|$, ϕ_c) map?

In SHELXL, the observed F_o^2 is partitioned according to F_c^2 for the twin components, i.e.

$$(F_o^2)_1 = F_o^2 (1-k) (F_c^2)_1 / F_c^2; \quad (F_o^2)_2 = F_o^2 k (F_c^2)_2 / F_c^2$$

Where F_o^2 is the observed intensity and F_c^2 is the weighted sum of the calculated intensities of the components.

The resulting difference maps improve as the refinement progresses, but these maps are sub-optimal, especially if k is close to 0.5.

Conclusions and Acknowledgements

Collecting data from non-merohedrally twinned crystals leads to an *increase* in the number of reflections that can be collected in a given time and to *improvements in the redundancy and completeness of the data* – minor components do not suffer from overloads – and so *should become normal practice*, both for small and macromolecules!

Unfortunately merohedral and pseudomerohedral twins do not have this advantage, and result in a poorer data to parameter ratio than for an untwinned crystal diffracting to the same resolution, making the structure refinement more difficult.

I am very grateful to many SHELX users and in particular to Regine Herbst-Irmer, Madhumati Sevvana, Peter Müller, Chuck Campana, Michael Ruf, Jörg Kärcher, Victor Young and Ina Dix for test data and many useful discussions.