Recent developments in TWINABS

Göttingen, April 19th 2007

George M. Sheldrick, *Göttingen University*

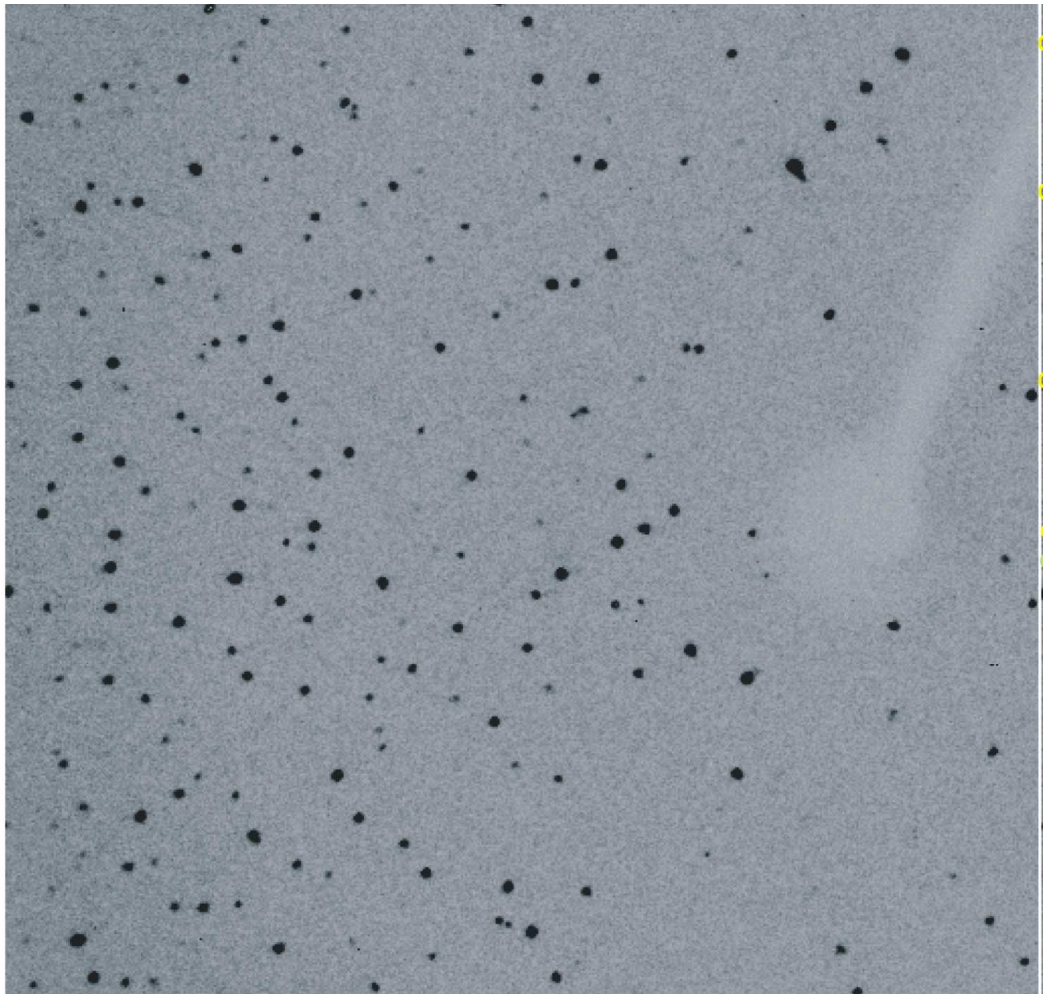
http://shelx.uni-ac.gwdg.de/SHELX/

Strategy for twinned crystals

1. **Find orientation matrices for all components, e.g. using CELL_NOW.**
2. **Verify interpretation using RLATT.**
3. **Multicomponent integration using SAINT (or EVAL_CCD), write .mul (or .sam) file (multicomponent .raw file).**
4. **Process the data using TWINABS. This writes a standard HKLF 4 format .hkl file for structure solution and initial (isotropic) refinement plus a special HKLF 5 format file for the final refinement in which all reflections that contribute to the same 'observation' are grouped together. The reflections can be selected and merged in a variety of ways.**
5. **Structure solution and initial refinement in the usual way using the HKLF 4 format file.**
6. **Final refinement including twin fractions (BASF) using the HKLF 5 format file.**

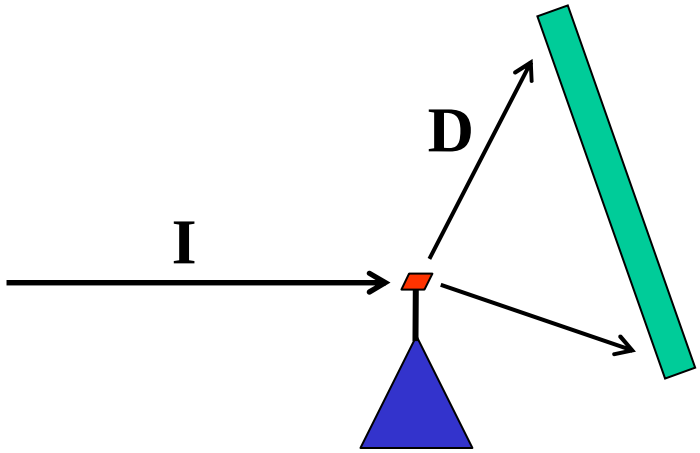
Diffraction pattern of glucose isomerase triplet

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. SAINT then uses the orientation matrices (3 in this case) to integrate the data. Some reflections can be integrated independently, in other cases there are two or more overlapping reflections Contibuting to the same observed intensity. SAINT also attempts to make a rough patitioning of the intensities within an integrateion box, but this is much less accurate.



Madhumati Sevvana

Scaling using equivalent reflections in TWINABS



Scaling is based on the following approximation:

$$I_c = I_o \cdot S(n) \cdot P(u,v,w)$$

S **= Scale factor for frame** ***n*** **(incident beam I only)**

P **= Absorption factor (diffracted beam D only)**

u,v,w **= Direction cosines relative to a*, b* and c***

S **and** ***P*** **are refined alternately to minimize** $\Sigma w(\langle I_c \rangle - I_c)^2$**, where** $\langle I_c \rangle$ **is the mean of a group of equivalents. Scaling requires a high redundancy (MoO) to work well.**

Incident beam scale factors

The scale factors *S* correct for the following systematic errors (amongst others):

- **Absorption of the primary beam by the crystal (and support)**
- **Crystal decomposition**
- **Intensity variation of the primary beam (e.g. synchrotron)**
- **Changes in the volume irradiated. This is often caused by the centers of the different twin components being separated in space, it is then impossible to center all components accurately and simultaneously!**
- **Beam inhomogeniety**

Diffracted beam absorption factor

The diffracted beam absorption factor *P*(*u*,*v*,*w*) is a sum of *spherical harmonics* for the direction cosines *u*, *v* and *w* as suggested by Blessing (1995). For example:

$$P_{0,1,2} = b_1 + b_2u + b_3v + b_4w +$$

$$b_5(3w^2-1)/2 + b_6(3uw) + b_7(3vw) + b_8(3u^2-3v^2) + b_9(6uv)$$

Since this is a linear function, only one refinement iteration is required. *S* and *b* are determined in alternate cycles; the incident beam scale factor depends linearly on *S*, so rapid convergence is assured and no initial values are needed for *S* and *b*.

SADABS and TWINABS strategy

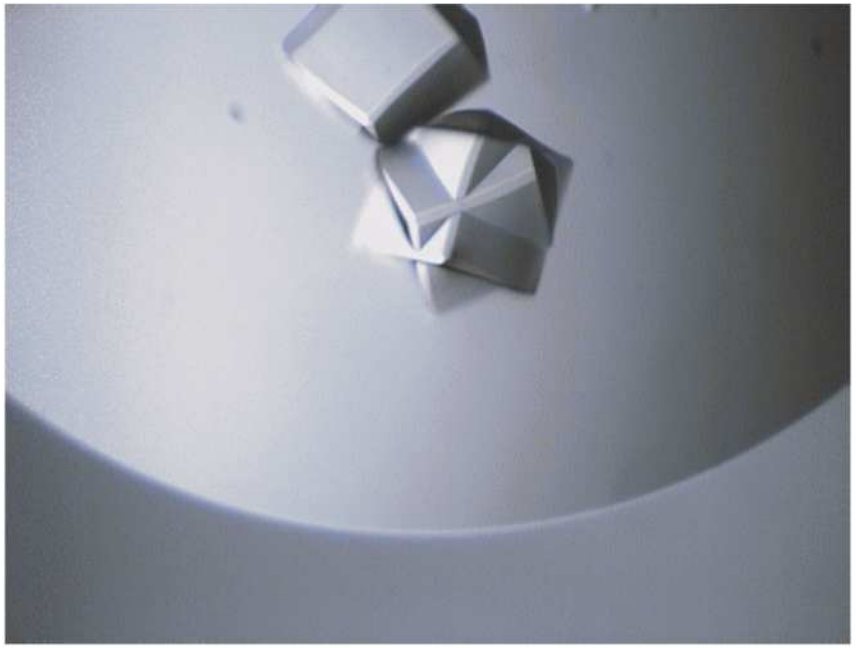
These programs run in 3 stages that may be repeated if necessary:

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. For TWINABS, the 'equivalents' may be single reflections or groups of overlapping reflections that contribute to a single integrated intensity.

2. Delete a small number of 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) and corrected data. TWINABS writes both HKLF 4 format files for structure solution and initial refinement and HKLF 5 files for final accurate refinement.

Non-merohedral protein 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 Sevvana

Equivalent reflections and groups

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	} not equivalent to the other groups shown here
1	-2	-3	…..	-3	
-1	1	2	…..	1	

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 options

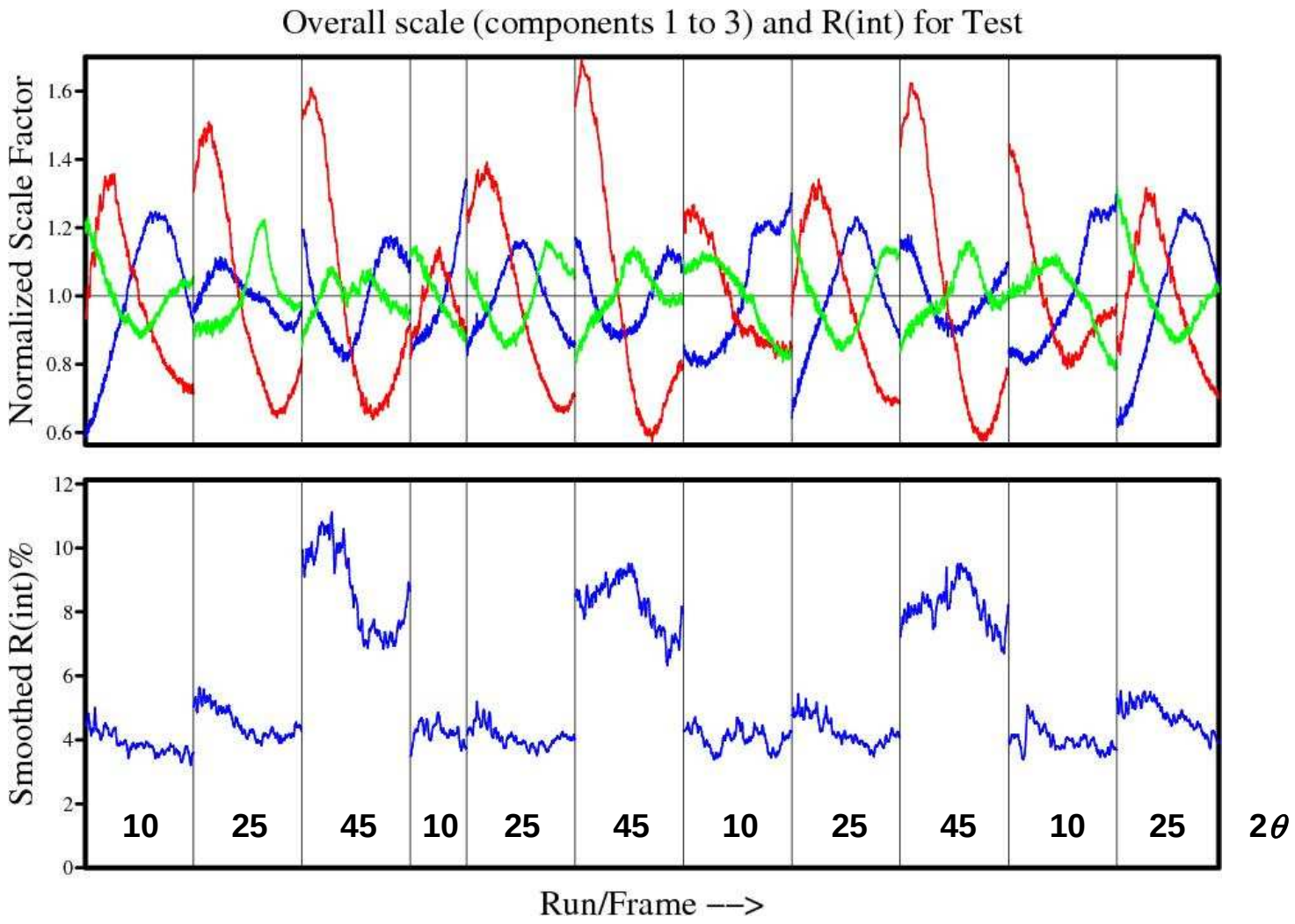
The selection of reflections to be used for scaling is controlled by a code number:

0: Each component is scaled separately using singles, the resulting scale factors are then also used to scale the compositie reflections. This is almost always the best option, even when one component dominates, except when there are very few singles.

N: Derive a single set of scaling parameters using all singles and composites that contain component N, then apply to all singles and composites. This can be used when this component dominates and almost all reflections overlap.

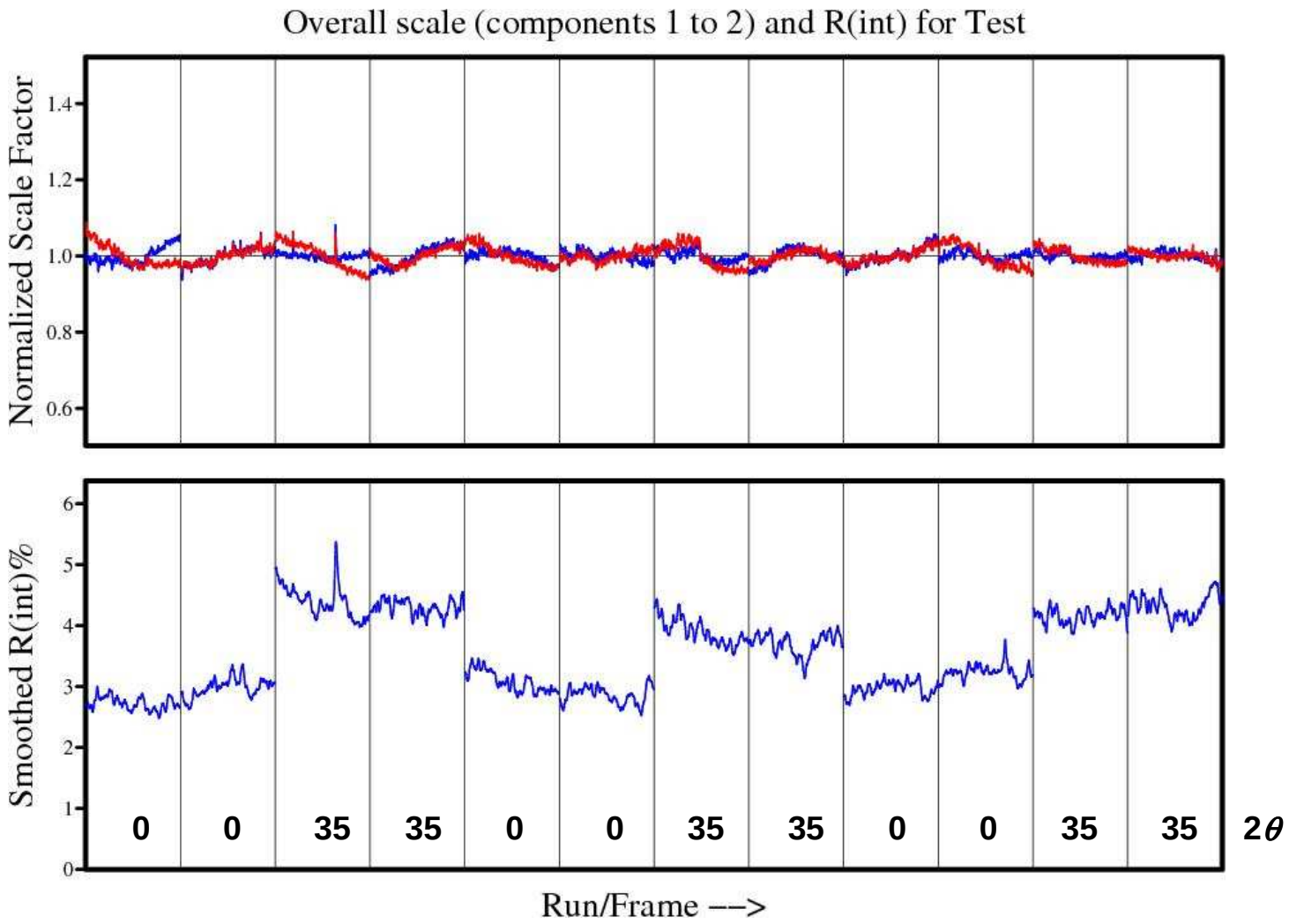
–N: Derive a single set of scaling parameters using only singles and composites that contain at least one of the components 1…N, then apply to all singles and composities. Useful for e.g. a four component twin where components 1 and 2 dominate and virtually all reflections overlap.

Scale factors and R_{int} for GI-triplet



In view of the large scale variations, option '0' is essential for scaling. 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.

The HKLF 4 format output file

In previous versions of TWINABS this file was created using the SAINT partitioning of composite reflections and hoping that averaging over many equivalents would compensate for the approximations involved. Such an HKLF 4 file can still be created by asking SADABS to process one component in a .mul file (new in SADABS 2007-2).

The new TWINABS makes this file by 'solving' the almost linear set of equations in which the unknown parameters are the intensities of the unique reflections and the twin ratios, and the observations are the (total) intensities of the single and composite reflections. Since this system of equations may be ill-conditioned, the SAINT partitioning is applied in the form of weak restraints (extra observational equations). The algorithm is robust, converges fast and can process several million reflections in a few seconds. The twin ratios obtained are close to those from the HKLF 5 refinement.

There is an option to use only reflections containing particular components for this analysis when one (or two) components dominate. It may be necessary to generate and test two or more HKLF 4 format files, but often the option '0' (use all) is best.

Inconsistent component indexing

The new HKLF 4 algorithm has revealed a subtle and unexpected elefant trap. In the cases where the Laue symmetry is lower than the metric symmetry of the lattice, the component may be indexed inconcistently, even though CELL_NOW obtains the second and subsequent orientation matrices by rotating the first!

For the insulin twin, CELL_NOW had rotated the cell by about 180º about an axis parallel to a face diagonal of the cubic cell [1 1 0], leading 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 HKLF 4 deconvolution, the scaling is not affected by the inconsistent indexing!! A re-indexing option has been added that can also be used when the components have different hands (a form of racemic twinning).

The HKLF 5 format output file

Despite the improvements in the 'detwinned' HKLF 4 format file, refinement tests show that refinement using HKLF 5 is still almost always slightly more precise. 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 number of reflections are used.

TWINABS provides a comprehensive set of options for preparing this file and merging equivalent groups of overlapping reflections, and some trial and error will often be required. When one component dominates it is best to use only singles and composites involving it, and the option to delete certain singles to reduce 'twin pairing errors' (a feature of SAINT) is often but not always better.

However the currently popular (especially amongst Acta Editors) 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 (e.g. when the option '0' is used to select all reflections for HKLF 5) 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 file minus the number of reflections in the HKLF 4 file!

Example: cubic insulin twin

206779 singles of component 1, 206734 of 2 and 18849 composites.

R_{int} = 3.47% for scaling and 3.26% (HKLF 4) based on 21955 unique data, truncated to 1.60Å, completeness = 100.00%, <I/σ> 60.6(9.5).

Twin factors 58.4:41.6% (HKLF 4 option 0/0), 56.7:43.3% from HKLF 5 refinement.

The six sulfurs could be found easily by S-SAD (truncated to 1.90Å) with CC 50.5%, weakCC 28.1%, and an excellent experimental map obtained.

HKLF 4 (0/0) refinement (isotropic): R1 18.94% $R1_{free}$ 20.65%

anisotropic: R1 15.02% $R1_{free}$ 18.14%

all data: R1 15.03%

HKLF 5 (0/1): R1 12.50% (14.18% after merging for Fourier)

HKLF 5 (0/1, no pairing): R1 11.72% (11.97% after merging for Fourier)

Example: glucose isomerase triplet

220819 singles of component 1, 221371 of 2, 221749 of 3, and 146778 composites.

R_{int} = 4.73% for scaling and 5.90% (HKLF 4) based on 124260 unique data, truncated to 1.65Å, completeness = 99.61%, <I/σ> 27.5(4.9).

Twin factors 55.9:14.1:30.0% (HKLF 4 deconvolution), 53.0:17.5:29.5% from HKLF 5 refinement.

The two cations could be found by SAD (truncated to 2.10Å) with CC 33.1%, weakCC 21.2%, and an excellent experimental map obtained.

HKLF 4 refinement (isotropic): R1 17.64% $R1_{free}$ 20.71%

all data: R1 17.67%

HKLF 5 refinement: R1 15.13% (16.02% after merging for Fourier)

HKLF 5, no pairing: R1 14.08% (14.48% after merging for Fourier)

Small molecule tests

Ganges ($P2_1$) 2 components, one dominant: 23609 (1), 22709 (2) and 19846 (composite). R_{int} 4.70% (scaling) and 4.78% (HKLF 4). Twin ratio 94.6:5.4% (HKLF 4) and 95.5:4.5% (HKLF 5 refinement).

HKLF 4: (0/0) R1 3.49%, (0/1) 2.94%, (1/1) 2.81%

HKLF 5: (0/1) R1 2.80%, (0/1np): 2.97%, (1/1) 2.66%, (1/1np) 2.81%

Feoxet2 ($P2_1/c$) 4 components, 2 strong: 2902 (1), 3827(2), 4112(3), 4085 (4), 1298 (composite). R_{int} 1.67% (scaling) and 2.56% (HKLF 4). Twin ratio 67.2:30.8:1.1:0.9% (HKLF 4) and 67.6:30.4:1.7:0.3% (HKLF 5 refinement).

HKLF 4: (0/0) R1 2.24%, (0/-2) 2.52%. HKLF 5: (0/1) 2.09%, (0/1np) 2.02%.

JL55 ($P2_1$): 2 components, strong overlap. 1624 (1), 1793 (2) 19825 (composite). Rint 3.69% (scaling), 3.91% (HKLF 4). Twin ratio 83.5:16.5% (HKLF 4) and 86.0:14.0% (HKLF 5 refinement).

HKLF 4: (0/0) R1 6.81%, (1/0) 5.87%, (1/1) 4.66%.

HKLF 5: (0/1np) R1 3.91%, (1/1) 3.29%, (1/1np) 3.15%.

Conclusions

The improvement in the HKLF 4 treatment means that a full HKLF 5 refinement – with testing of the different options for preparing the HKLF 5 file – only needs to be performed right at the end of the refinement (if at all). In the case of proteins, this means that free R flags are only required for the HKLF 4 file, neatly sidestepping the problem of how to select really independent reflections or groups of them in the HKLF 5 file. It is in any case normal to perform the final refinement job with all data.

Collecting data from 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!

Acknowledgements

I am very grateful to my group in Göttingen and many TWINABS users for their help in optimizing the program, and in particular to Regine Herbst-Irmer, Madhumati Sevvana, Victor Young and Ina Dix for their help in testing TWINABS.