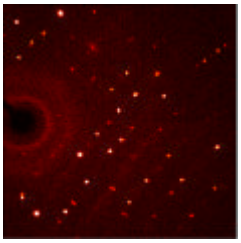
Data processing



h	*k*	*l*	F^2	s(F^2)
1	0	0	187.41	4.84
2	0	0	16.29	3.20
3	0	0	1868.63	31.73
4	0	0	99.41	2.16
5	0	0	99.83	2.32
7	1	0	22.65	2.64
6	1	0	6.60	1.74
5	1	0	1.43	0.76
4	1	0	307.10	14.35
3	1	0	1390.84	28.38

Data reduction

Overview

- extract integrated intensities from raw frames
- computationally intensive
- requires a valid and precise orientation matrix
- matrix from initial indexing *may* be adequate …

  … then can run data acquisition/reduction together

Data reduction

Overview …

- extract intensities
- apply fixed instrumental corrections (spatial, etc)
- apply variable corrections (e.g., dark time)
- apply corrections for absorption if required
- apply decay correction if this is significant

Data reduction – best matrix

Safest procedure?

- harvest the entire frameset for reflections
- re-index
- re-refine the matrix using more/all reflections
- repeat the Bravais lattice determination
- re-check the indexing (e.g. using *hkl* overlay)
- if cell has changed, re-check for known cells

Data reduction

Integration

- software uses the orientation matrix
  ➔ determines the reflection positions
- estimation of intensities
  ➔ uses three-dimensional profile fitting (“box”)

Data reduction

Integration – some options

can update or refine the orientation matrix

- to allow for uncertainties in peak positions
- to allow for gradual drift in crystal orientation

can refine the size of the integration box

- may be better when this changes with orientation
- but need to be certain the result is sensible

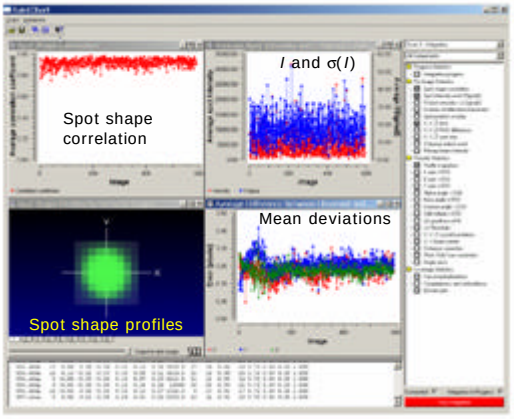
Data reduction

Reflection widths – refined or fixed

- from trial integration runs
- better to err on the wide side
  ➔ to make sure you get all the peak
- but too wide will lead to overlap with other peaks
  ➔ this will also give incorrect intensities
- problems with large unit cells and broad peaks

Integration program output

- monitors how the integration is proceeding
- gives diagnostic information
- warns about problems
- output can be voluminous
- instruction manual or other documentation
- refer to it if there are problems later
- graphical display of information is very powerful



Unitcell refinement

- suitably constrained cell parameters
- part of the data reduction or separately
- should include most of the significant reflections
- some systems use all reflections
- absolute number of reflections is always high, but what is a safe proportion?

Unitcell refinement

```
Input file contains      5746 reflections for this component
Maximum allowed reflections =     25000
Wavelength, relative uncertainty:  0.7107300,  0.0000089
Orientation ('UB') matrix:
  0.0375682 -0.0244249  0.0626351
 -0.0522494  0.0170492  0.0547843
 -0.0979531 -0.0184620 -0.0322734

   A       B       C        Alpha   Beta     Gamma    Vol
  8.8205 28.535  11.582   90.000  104.686  90.000   2820.0
  Standard uncertainties:
  0.0009  0.003   0.002    0.000    0.002    0.000     0.9

Range of reflections used:
  Worst res   Best res  Min 2Theta  Max 2Theta
    8.8119     0.7685     4.622      55.084

                Crystal system constraint:     monoclinic b-unique
```

Corrections to intensity data

- Lorentz – time in the diffracting position
- polarisation – varies as a function of q
- spatial distortion; non-uniformity; dead time
- incident beam intensity variations (esp. SR)
- absorption – depends on elements and l
- any other sources of systematic error?

Absorption corrections

Methods for absorption correction

- numerical by face indexing
- no separate ψ scan measurements
  instead exploit redundancy in the data
  low redundancy for triclinic crystals

  correction factors include other effects
  so may not be as predicted
- apply an empirical correction if all else fails?

Data reduction output

Always worth a close look, includes analyses of:

- data significance (should be high)
- data coverage (should be close to 100%)
- redundancy (should ideally be high)
- consistency with *assumed* symmetry

Data reduction output

Coverage Statistics

Resolution	%Complete	Redund	Rsym	Shell Rshell	Shell #Sigma	Shell %<2s
to 1.656	97.92	4.70	0.054	0.054	80.93	5.9
to 1.315	99.28	4.63	0.049	0.057	31.97	10.9
to 1.149	99.85	4.50	0.044	0.061	23.65	11.0
to 1.044	99.94	4.38	0.042	0.086	19.49	17.4
to 0.969	99.97	4.27	0.040	0.096	12.62	21.3
to 0.912	99.95	4.18	0.040	0.141	7.72	21.5
to 0.867	99.48	4.10	0.041	0.151	5.98	27.2
to 0.829	99.42	3.84	0.045	0.159	4.46	20.6
to 0.797	98.07	3.68	0.050	0.177	4.36	34.2
to 0.770	96.10	3.42	0.051	0.218	3.27	41.6

Data reduction output

Can indicate problems with:

- assumed diffraction symmetry
- the orientation matrix
- quality of your crystal
- data completeness or resolution

Possible actions:

- re-process the frames – with different input
- collect more data – perhaps on a new crystal
- proceed to structure solution

A typical experiment

1. Mount, orient and optically centre the crystal (1 minute)
2. Assess crystal quality using still or limited-oscillation exposures (1 minute)
3. Collect some frames and harvest reflections for indexing (several minutes)
4. Index, refine orientation matrix and determine Bravais lattice (a few minutes)
5. Check whether the unitcell is known and whether its volume is sensible (1 minute)
6. Survey frames visually to check indexing and assess quality (a few minutes)

Steps 2—6 are decision points

A typical experiment

7. Determine the exposure time, frame width and fraction to collect (1 minute)
8. Record the crystal colour, shape and size (1 minute); index faces? (several minutes)
9. Collect the data (several minutes to hours)
10. Redetermine the matrix using all available/significant reflections (a few minutes)
11. Survey frames visually as a check on the orientation matrix (a few minutes)
12. Process the data, applying corrections as required (several minutes)

More problematic cases

- can collect from very unpromising crystals
- challenge is to to obtain a useable dataset

- problems with
  the inherent quality of the crystal
  the techniques used
  instrumental factors

- resulting in the lack of
  either a single accurate, valid orientation matrix
  or a generally valid description of peak shape

Example 1

A frameset would not index as a whole despite the appearance of the frames suggesting no problems.

- each run of frames could be indexed separately
- give the same unit cell for each run
- different orientation matrices were obtained
- no decay or crystal movement detectable

- problem was mechanical
- slippage of the φ circle while slewing between runs
- separate orientation matrices allowed processing

Example 2

Symptoms very like those for Example 1, but only an approximate matrix could be defined for each run.

- looked at frames with predicted *hkl* overlaid
- found a systematic drift in reflection positions
- either crystal was not securely mounted
  poor bond between the crystal and fibre
  poor attachment between fibre and support
- or changes were occurring in the crystal

- data processed with a separate matrix for each run
- but each matrix was updated through the run
- this accommodated the "movement" of the crystal
- otherwise recollect with crystal securely mounted

Example 3

Crystals were coated with microcrystals that could not be removed, contributing a pervasive background of weak reflections to the main diffraction pattern.

- isolated reflections from the main crystal
- using a matrix based on the strongest reflections
- then included those of medium intensity

- this matrix used for integration of main pattern
- weak, non-fitting reflections effectively ignored
- only a few intensities were affected by overlap
- effects were minor

Example 4

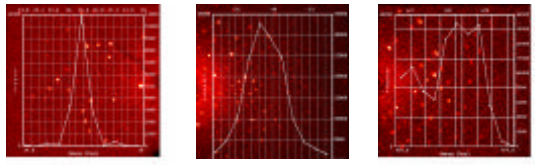
A crystal indexed with a primitive unit cell with a moderately long *b* axis of 32 Å, but with broad reflections (FWHM = 1.2°).

- reflection overlap quite likely
- so increased the crystal-to-detector distance
- then collected the frameset
- still only one detector setting was needed

- minimised reflection overlap
- could investigate larger integration box sizes

Example 5

Reflection profiles widths varied strongly as a function of goniometer angle. This anisomosaicity makes it difficult to define a single integration box size.



GOOD **BAD** **UGLY**

Example 5

- allow the box size to vary continuously?
- then it always matches the local reflection profiles
- simple, but starting values may be critical

- above option not available, or does not work well?
- another approach is to use fixed box size
- somewhat biased towards the wider reflections
- then correct the dataset using multi-scan methods

Example 6

Screening indicated problems with crystal quality, including split/badly shaped reflections, possibly >1 component. It was possible to index the main component.

- look for a better crystal, otherwise…
- collect the frameset
- extract intensities for the main component
- try to solve the structure based on this

- may be able to carry on with this dataset
- but look out for any indicators of non-singularity
- may have to treat the crystal as twinned

Example 7

Initial indexing yielded a primitive unit cell with one very long axis (85.6 Å), giving the possibility of severe overlap at the standard crystal-to-detector distance.

- increase the crystal-to-detector distance
- highest achievable 2? value may be rather low
- use two different detector ? settings
- ensure that the 2? ranges overlap
- use longer exposures for high angle data

- (same strategy may be useful elsewhere)
- with low angle data strong, high angle very weak
- use different detector ? settings and exposure times

Example 8

The output from the integration program indicated no problems with crystal quality or the indexing but the merging *R* was 0.66 under monoclinic symmetry. Unsurprisingly, the structure would not solve.

- Bravais lattice was *metrically* monoclinic *C* (2/*m*)
- but *diffraction symmetry* did not match 2/*m*
- a smaller triclinic *P* cell was chosen.
- re-processing gave a merging *R* of 0.10
- the structure solved at the first attempt

- metric symmetry *versus* diffraction symmetry
- also a problem for monoclinic cells with ß ~ 90°

Non-merohedral twinning

- can cause problems are all stages
- from indexing to structure refinement
- need to recognise it as soon as possible
- otherwise can waste time trying other things

Symptoms during refinement can include:

- stubbornly high *R* indices, with no obvious cause
- high ΔF residuals, again with no obvious cause
- individual reflections with $F(\text{obs})^2 >> F(\text{calc})^2$
- certain indices may be particularly affected
- extreme indicators in some low $F(\text{calc})^2$ ranges

Non-merohedral twinning: general outline

1. examine frameset for visible indications of twinning
   - can use various visualisation aids
2. identify major twin component - visually or software
3. index major twin component → orientation matrix 1
4. identify (first) minor twin component
5. index minor twin component → orientation matrix 2
6. repeat 4 and 5 for any other minor components

Non-merohedral twinning: general outline

7. determine and view the relationships between the different matrices
8. generate predicted patterns using matrices and check these against frames
9. export orientation matrices for use by your data reduction program
10. process the frameset using these orientation matrices
11. output separate or combined datasets for solution and refinement

Looking at the *hkl* data

Before proceeding to structure solution:

- analyse the systematic absences → space group(s)

hkl → centering?
h0l → glide plane?
0*k*0 → screw axis?

- intensity statistics → centrosymmetric or not?