Data processing

mosflm, imosflm, pointless, scala

Judit É Debreczeni
data processing flowchart

inspect images
spot search and index
predict spots, estimate mosaicty
define dead areas
refine cell
strategy
integrate
analyse
scale, merge
convert I to F
analyse
detwin
<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inspect images</td>
<td>is this my best crystal? Worth collecting/processing?</td>
</tr>
<tr>
<td>Spot search and index</td>
<td>determine lattice</td>
</tr>
<tr>
<td>Predict spots, estimate mosaicity</td>
<td>check solution</td>
</tr>
<tr>
<td>Define dead areas</td>
<td>exclude shadows, icering etc</td>
</tr>
<tr>
<td>Refine cell</td>
<td>refine solution (multiple wedges)</td>
</tr>
<tr>
<td>Strategy</td>
<td>get total rotation, oscillation, distance</td>
</tr>
<tr>
<td>Integrate</td>
<td>determine intensities and σ</td>
</tr>
<tr>
<td>Analyse</td>
<td>determine Laue group and space group</td>
</tr>
<tr>
<td>Scale, merge</td>
<td>scale all data together and analyse</td>
</tr>
<tr>
<td>Convert I to F</td>
<td>calculate structure factors</td>
</tr>
<tr>
<td>Analyse</td>
<td>any bad parts to omit? etc</td>
</tr>
<tr>
<td>Detwin</td>
<td>detwin if necessary</td>
</tr>
</tbody>
</table>
data processing flowchart

1. Inspect images
2. Spot search and index
3. Predict spots, estimate mosaicity
4. Define dead areas
5. Refine cell
6. Strategy
7. Integrate
8. Analyse
9. Scale, merge
10. Convert I to F
11. Analyse
12. Detwin
scattering and diffraction

ARRANGEMENT OF MOLECULES IN THE CRYSTAL

DIFFRACTION PATTERN

ARRANGEMENT OF MOLECULES IN THE CRYSTAL

DIFFRACTION PATTERN

(a)

(b)

(c)

(d)

(e)

(f)
scattering and diffraction

- diffraction images: not only diffraction, also scattering from the crystal's environment:
  - solvent, air in the beam
  - loop
  - ice
  - zingers, cosmic radiation
<table>
<thead>
<tr>
<th>Crystal System</th>
<th>Minimum Symmetry</th>
<th>Bravais Lattices</th>
<th>Unit Cell Geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Triclinic</td>
<td>None</td>
<td>1. Primitive (P)</td>
<td>$a \neq b \neq c$; $\alpha \neq \beta \neq \gamma$</td>
</tr>
<tr>
<td>2. Monoclinic</td>
<td>One 2fold axis</td>
<td>2. Primitive (P)</td>
<td>$a \neq b \neq c$; $\alpha = \gamma = 90^\circ \neq \beta$</td>
</tr>
<tr>
<td>3. Orthorhombic</td>
<td>Three orthogonal 2fold axes</td>
<td>4. Primitive (P) 5. Base-Centered (C) 6. Body-Centered (I) 7. Face-Centered (F)</td>
<td>$a \neq b \neq c$; $\alpha = \beta = \gamma = 90^\circ$</td>
</tr>
<tr>
<td>4. Tetragonal</td>
<td>One 4fold axis</td>
<td>8. Primitive (P) 9. Body-Centered (I)</td>
<td>$a = b \neq c$; $\alpha = \beta = \gamma = 90^\circ$</td>
</tr>
<tr>
<td>5. Trigonal</td>
<td>One 3fold axis</td>
<td>10. Primitive (P)</td>
<td>$a = b \neq c$; $\alpha = \beta = 90^\circ, \gamma = 120^\circ$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11. Rhombohedral (R)</td>
<td>$a = b = c$; $\alpha = \beta = \gamma \neq 90^\circ$</td>
</tr>
<tr>
<td>6. Hexagonal</td>
<td>One 6fold axis</td>
<td>10. Primitive (P)</td>
<td>$a = b \neq c$; $\alpha = \beta = 90^\circ, \gamma = 120^\circ$</td>
</tr>
<tr>
<td>7. Cubic</td>
<td>Four 3fold axes</td>
<td>12. Primitive (P)</td>
<td>$a = b = c$; $\alpha = \beta = \gamma = 90^\circ$</td>
</tr>
</tbody>
</table>
crystal systems

primitive

body centered

all face centered

guest: I

face centered
<table>
<thead>
<tr>
<th>Crystal System</th>
<th>Minimum Symmetry</th>
<th>Bravais Lattices</th>
<th>Unit Cell Geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Triclinic</td>
<td>None</td>
<td>1. Primitive (P)</td>
<td>a ≠ b ≠ c; α ≠ β ≠ γ</td>
</tr>
<tr>
<td>2. Monoclinic</td>
<td>One 2fold axis</td>
<td>2. Primitive (P)</td>
<td>a ≠ b ≠ c; α = γ = 90° ≠ β</td>
</tr>
<tr>
<td>3. Orthorhombic</td>
<td>Three orthogonal 2fold axes</td>
<td>4. Primitive (P) 5. Base-Centered (C) 6. Body-Centered (I) 7. Face-Centered (F)</td>
<td>a ≠ b ≠ c; α = β = γ = 90°</td>
</tr>
<tr>
<td>4. Tetragonal</td>
<td>One 4fold axis</td>
<td>8. Primitive (P) 9. Body-Centered (I)</td>
<td>a = b ≠ c; α = β = γ = 90°</td>
</tr>
<tr>
<td>5. Trigonal</td>
<td>One 3fold axis</td>
<td>10. Primitive (P) 11. Rhombohedral (R)</td>
<td>a = b ≠ c; α = β = γ ≠ 90°</td>
</tr>
<tr>
<td>6. Hexagonal</td>
<td>One 6fold axis</td>
<td>10. Primitive (P)</td>
<td>a = b ≠ c; α = β = γ = 120°</td>
</tr>
<tr>
<td>7. Cubic</td>
<td>Four 3fold axes</td>
<td>12. Primitive (P) 13. Body-Centered (I) 14. Face-Centered (F)</td>
<td>a = b = c; α = β = γ = 90°</td>
</tr>
</tbody>
</table>

- 7 crystal systems + P, I, F, A, B, C -> 14 Bravais lattices
- many combinations reduce to one another
spot positions on a diffraction image are distorted projections of the reciprocal lattice
fulls and partials

- Fully recorded spot spans one image
- Fine sliced data with spots sampled in 3D

Start of image | End of image
-----|-----
Start of images | End of images

Detector

Intensity vs image number

Start of images | End of images
data processing flowchart

- inspect images
- spot search and index
- predict spots, estimate mosaicity
- define dead areas
- refine cell
- strategy
- integrate
- analyse
- scale, merge
- convert I to F
- analyse
- detwin
- if we know the main beam position on the image, we can count spots from the origin
- tedious, as we have to work in 3D and crystal is not aligned with rotation axis, detector etc
- find spots on diffraction images
- map back spot positions onto Ewald sphere
- rotate back to zero-phi position
autoindexing

- this gives a set of scattering vectors
now consider every direction in the Ewald sphere (sample the whole hemisphere with ~8000 directions)

- take the projection of scattering vectors onto these directions

set of projections
if the direction in question is a real space cell axis (perpendicular to a reciprocal space lattice plane), the projections will cluster at lengths which are multiples of lattice spacing

otherwise lengths of projections will be random, no clustering
- Evenness of spacing is evaluated by a 1D Fourier transform of the projection.
- If peaks are evenly spaced, a few low frequency terms will be large.
- Periodicity is related to the unit cell (real space) dimension in that direction.
- Three such directions describe the unit cell.
- pick three non-coplanar directions with large peaks in the Fourier
- the lattice found this way is not necessarily the simplest cell:

black: reduced cell
red, blue: possible autoindexing solutions
autoindexing - in practice

- multiple solutions with penalties: how well does the cell obey the constraints of that lattice type
- sharp drop between solutions and non-solutions
- solution's penalty < 10 (20 okay for dodgy crystal)
- pick the highest symmetry from the low-penalty bunch
## Solutions:

<table>
<thead>
<tr>
<th>Solution</th>
<th>Lat.</th>
<th>Pen.</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>r(x,y)</th>
<th>r(Φ)</th>
<th>≤ beam</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (ref)</td>
<td>aP</td>
<td>0</td>
<td>67.1</td>
<td>67.4</td>
<td>67.5</td>
<td>90.1</td>
<td>90.1</td>
<td>90.0</td>
<td>0.14</td>
<td>0.49</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>2 (ref)</td>
<td>mP</td>
<td>0</td>
<td>67.1</td>
<td>67.4</td>
<td>67.5</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>0.14</td>
<td>0.49</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>3 (ref)</td>
<td>mP</td>
<td>0</td>
<td>67.1</td>
<td>67.5</td>
<td>67.4</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>0.14</td>
<td>0.49</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>4 (ref)</td>
<td>mP</td>
<td>0</td>
<td>67.4</td>
<td>67.1</td>
<td>67.5</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>0.14</td>
<td>0.49</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>5 (ref)</td>
<td>aP</td>
<td>0</td>
<td>67.1</td>
<td>67.4</td>
<td>67.5</td>
<td>89.9</td>
<td>89.9</td>
<td>90.0</td>
<td>0.14</td>
<td>0.49</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>6 (ref)</td>
<td>mC</td>
<td>1</td>
<td>95.3</td>
<td>95.4</td>
<td>67.1</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>0.14</td>
<td>0.59</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>7 (ref)</td>
<td>oP</td>
<td>1</td>
<td>67.1</td>
<td>67.4</td>
<td>67.5</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>0.14</td>
<td>0.49</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>8 (ref)</td>
<td>tP</td>
<td>1</td>
<td>67.4</td>
<td>67.4</td>
<td>67.1</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>0.14</td>
<td>0.49</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>9 (ref)</td>
<td>oC</td>
<td>1</td>
<td>95.4</td>
<td>95.4</td>
<td>67.1</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>0.14</td>
<td>0.59</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>10 (ref)</td>
<td>mC</td>
<td>1</td>
<td>95.3</td>
<td>95.4</td>
<td>67.1</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>0.14</td>
<td>0.59</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>11 (ref)</td>
<td>mC</td>
<td>2</td>
<td>95.1</td>
<td>95.0</td>
<td>67.5</td>
<td>90.0</td>
<td>90.3</td>
<td>90.0</td>
<td>0.16</td>
<td>0.66</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>12 (ref)</td>
<td>oC</td>
<td>2</td>
<td>95.1</td>
<td>95.0</td>
<td>67.5</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>0.19</td>
<td>0.65</td>
<td>0.33 (0.1)</td>
</tr>
<tr>
<td>13 (ref)</td>
<td>mC</td>
<td>2</td>
<td>95.1</td>
<td>95.0</td>
<td>67.5</td>
<td>90.0</td>
<td>90.3</td>
<td>90.0</td>
<td>0.16</td>
<td>0.66</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>14 (ref)</td>
<td>tP</td>
<td>3</td>
<td>67.2</td>
<td>67.2</td>
<td>67.5</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>0.18</td>
<td>0.50</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>15 (ref)</td>
<td>hR</td>
<td>3</td>
<td>95.1</td>
<td>95.1</td>
<td>116.9</td>
<td>90.0</td>
<td>90.0</td>
<td>120.0</td>
<td>0.17</td>
<td>0.58</td>
<td>0.35 (0.2)</td>
</tr>
<tr>
<td>16 (ref)</td>
<td>oP</td>
<td>3</td>
<td>67.3</td>
<td>67.3</td>
<td>67.3</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>0.21</td>
<td>0.55</td>
<td>0.34 (0.1)</td>
</tr>
<tr>
<td>17 (ref)</td>
<td>hR</td>
<td>3</td>
<td>95.2</td>
<td>95.2</td>
<td>116.5</td>
<td>90.0</td>
<td>90.0</td>
<td>120.0</td>
<td>0.20</td>
<td>0.46</td>
<td>0.34 (0.2)</td>
</tr>
<tr>
<td>18 (reg)</td>
<td>oC</td>
<td>249</td>
<td>67.1</td>
<td>150.7</td>
<td>67.4</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>19 (reg)</td>
<td>mC</td>
<td>249</td>
<td>150.7</td>
<td>67.1</td>
<td>67.4</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>20 (reg)</td>
<td>mC</td>
<td>250</td>
<td>67.1</td>
<td>150.6</td>
<td>67.5</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>21 (reg)</td>
<td>mC</td>
<td>250</td>
<td>67.1</td>
<td>150.7</td>
<td>67.4</td>
<td>90.0</td>
<td>90.1</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>22 (reg)</td>
<td>oC</td>
<td>250</td>
<td>67.1</td>
<td>150.5</td>
<td>67.5</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>23 (reg)</td>
<td>mC</td>
<td>251</td>
<td>150.8</td>
<td>67.4</td>
<td>67.1</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>24 (reg)</td>
<td>mC</td>
<td>251</td>
<td>67.4</td>
<td>150.8</td>
<td>67.1</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>25 (reg)</td>
<td>oC</td>
<td>251</td>
<td>67.4</td>
<td>150.8</td>
<td>67.1</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>26 (reg)</td>
<td>mC</td>
<td>252</td>
<td>67.4</td>
<td>150.8</td>
<td>67.1</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>27 (reg)</td>
<td>hP</td>
<td>252</td>
<td>67.2</td>
<td>67.2</td>
<td>67.5</td>
<td>90.0</td>
<td>90.0</td>
<td>120.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>28 (reg)</td>
<td>hP</td>
<td>252</td>
<td>67.4</td>
<td>67.4</td>
<td>67.1</td>
<td>90.0</td>
<td>90.0</td>
<td>120.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>29 (reg)</td>
<td>mC</td>
<td>499</td>
<td>150.5</td>
<td>67.1</td>
<td>95.4</td>
<td>90.0</td>
<td>129.3</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>30 (reg)</td>
<td>oI</td>
<td>499</td>
<td>67.1</td>
<td>95.4</td>
<td>116.5</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>31 (reg)</td>
<td>oI</td>
<td>500</td>
<td>67.1</td>
<td>67.4</td>
<td>165.0</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>32 (reg)</td>
<td>tI</td>
<td>500</td>
<td>67.2</td>
<td>67.2</td>
<td>165.0</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>33 (reg)</td>
<td>mC</td>
<td>502</td>
<td>95.1</td>
<td>95.1</td>
<td>95.1</td>
<td>90.0</td>
<td>119.9</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>34 (reg)</td>
<td>oP</td>
<td>502</td>
<td>95.1</td>
<td>95.1</td>
<td>165.0</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

**Spacegroup:** h3
autoindexing - why it goes wrong

- weak diffraction: enough spots? -- needs 50+
- experimental parameters:
  - beam position (never trust the header!)
  - wavelength
  - detector distance
  these are almost always indicated by no separation between solution and noise (in penalties)
- multiple lattices
- overlapping spots
- true symmetry lower than metric symmetry
- indexing can be misleading for low symmetry if only from one image
autoindexing - is it OK?

check predicted pattern at different angles:
- does it match?
- are all reflections predicted?

- unpredicted spots:
  - incorrect mosaicity
  - multiple lattices
  - pseudosymmetry

- solutions given by mosflm are the same, but imposing different lattice symmetries. In other words, if the P1 solution is wrong, all the others will be wrong too...
data processing flowchart

- inspect images
- spot search and index
- predict spots, estimate mosaicity
- define dead areas
- refine cell
- strategy
- integrate
- analyse
- scale, merge
- convert I to F
- analyse
- detwin
estimating mosaicity

- relevant directly after autoindexing, improve estimate later
- for a series of mosaicity values (0.0, 0.05, 0.1 etc) integrate the total intensity under all predicted spots
- surprisingly accurate (don't be mislead by tails of strong spots)
data processing flowchart

inspect images
spot search and index
predict spots, estimate mosaicty
define dead areas
refine cell
strategy
integrate
analyse
scale, merge
convert I to F
analyse
detwin
mask dead areas

- otherwise shadowed areas will be processed as active areas!
  - beamstop
  - cryo nozzle
data processing flowchart

- inspect images
- spot search and index
- predict spots, estimate mosaicity
- define dead areas
- refine cell
- strategy
- integrate
- analyse
- scale, merge
- convert I to F
- analyse
- detwin
cell refinement

- **idea:**
  - first obtain a good estimate of cell parameters
  - then integrate with fixed cell

- **cell refinement**
  - on well separated segments
  - each segment should ideally span > 2 x mosaicity
    e.g. 1-5, 91-95
  - collect these before start of data collection for better strategy

- **monitor spot profile, residuals**
  
  two kinds of residuals for independent refinement of quantities
  - based on spot coordinates
  - based on spot position in phi

- **other programs have different philosophy:**
  - HKL2000: postrefinement with all spots
  - XDS: refine cell as we go in integration
Parameters to refine:

- **Crystal parameters**
  - cell dimensions
  - orientation
  - mosaic spread

- **Detector parameters**
  - position
  - orientation
  - distortion parameters

- **Beam parameters**
  - orientation
  - beam divergence
cell refinement: spot coordinates

- Refinement against spot positions:
  - crystal to detector distance
  - beam centre
  - detector tilts and other miscalibrations
  - X-vsY fudge factor

- what to expect:
  < 0.03 : excellent
  ~ 0.08 : still good
  ~ 0.12 : okay if high mosaicity explains it

- residual: \[ \Omega_1 = \sum_i \omega_{ix} (X_i^{\text{calc}} - X_i^{\text{obs}})^2 + \omega_{iy} (Y_i^{\text{calc}} - Y_i^{\text{obs}})^2 \]
cell refinement: spot \( \phi \) position

- Refinement against spot position
  - cell
  - crystal orientation
  - mosaicity

- residual: \( \Omega = \sum \left( \frac{Robs - Rcalc}{d^*} \right)^2 \)

- partial reflections:
  e.g. spreading over two images with intensity on the two images: \( I_1 \) and \( I_2 \)

To determine the observed position on the first image from the fraction of the total intensity on that image: \( F = \frac{I_1}{(I_1 + I_2)} \)

requires a model for the rocking curve of the spot

\[
F = \frac{1}{2} \left( 1 \pm \sin \left( \frac{\pi \Delta R}{2 \epsilon} \right) \right)
\]

(epsilon: radius of a reciprocal lattice point)
First orientations are refined for all specified images:
- Angular and positional residuals printed for each image
- Spot profile printed

Finally, cell parameters and mosaicity are refined from all images.

If parameters not converged, whole cycle is repeated.

How to judge refinement:
- Values drifted a lot?
- Negative mosaicity? (misindexed?)
- Predictions, profiles
cell refinement flow
data processing flowchart

- inspect images
- spot search and index
- predict spots, estimate mosaicity
- define dead areas
- refine cell
- strategy
- integrate
- analyse
- scale, merge
- convert I to F
- analyse
- detwin
Two steps:
- Predict the position in the digitised image of each Bragg reflection
- Estimate intensity (subtract X-ray background) and determine an error estimate of the intensity

1. Predicting reflection positions:
   - Accuracy in prediction is crucial. Ideally, cell parameters should be known to better than 0.1%.
   - Errors in prediction will introduce systematic errors in profile fitting.
   - Refined during integration:
     - detector parameters
     - crystal orientation
     - mosaicity
   - Cell parameters not refined.

2. Estimating intensity and error:
   - profile fitting
   - summation integration
Summation integration:
- Sum the pixel values of all pixels in the peak area of the mask.
- Subtract the sum of the background values calculated from the background plane for the same pixels.
Profile fitting:
- Assume that the shape or profile (2D or 3D) of the spots is known.
- Assume that spot shape does not depend on intensity.
- Determine the scale factor, which, when applied to the standard spot profile, gives the best fit to the observed spot profile.
- This scale factor is then proportional to the profile fitted intensity for the reflection.
- Minimise difference

More reliable intensity estimate of weak reflections (down-weighting peripheral peak pixels)
- Assumption #2 not true for saturated spots.
How do we know what the standard profile looks like?

- Split the detector surface into tiles
- Average strong reflections within tile (i.e. different detector areas will have different profiles -- other programs work differently)

For each reflection a new profile is calculated as a weighted mean of profiles in adjacent regions (weight: scales with distance)
- Used for both full and partial reflections, although not really valid for partials, but in practice just works fine.
Estimation of standard deviations:

- For summation integration or profile fitted partials: based on Poison statistics.
- For profile fitted intensities: goodness of fit of the scaled standard profile to the true reflection profile

BUT: these underestimate the true errors, and should be modified accordingly at the merging step so that reflect the actual differences between multiple measurements.
integration in practice

- images integrated in blocks (typically 10 images/block)
- first all images are postrefined with fixed cell
  then images in block are integrated

- look at:
  - standard profiles
  - I/\sigma per resolution bin
  - warning messages at end of run
  - mosaicity refinement (not on per default)

- most likely warning messages:
  - gain must be adjusted (detector-dependent)
  - YSCALE -- sensitive, can ignore
  - poor profiles, averaging
    -- can be genuine if reflections are weak at edge
  - crystal slippage --
integration in practice
data processing flowchart

- inspect images
- spot search and index
- predict spots, estimate mosaicty
- define dead areas
- refine cell
- strategy
- integrate
- analyse
- scale, merge
- convert I to F
- analyse
- detwin
The space group is only a hypothesis until the structure is solved!
- It is hard to distinguish between true (crystallographic) and approximate (NCS) symmetry
- Metric symmetry can be higher than symmetry of the space group

It would be great, however, to know the symmetry early to determine the data collection strategy... hmm.
space group determination

- Stages in space group determination

1. Lattice symmetry (symmetry of crystal class)
   - Crystal class imposes restrictions on cell dimensions (spot positions), and this is needed for accurate prediction.
   - Determined at indexing
   - BUT: pseudo-symmetry (metric symmetry higher than true symmetry). E.g. a=b=c \(\alpha=\beta=\gamma=90\) is necessarily cubic.

2. Laue group symmetry
   - determined by the symmetry of the diffraction pattern (spot positions and intensities)
   - corresponds to the space group without translations with an added centre of symmetry (from Friedel's law)
Stages in space group determination (cont.)

3. Point group symmetry
   - For chiral space groups (like macromolecular crystals), there is only one point group corresponding to each Laue group.
   - It is the space group without any translations.

4. Space group symmetry
   - point group + translations (screw axes etc)
   - translations are only visible as systematic absences, usually along axes, which are not very reliable indicators -- there are only a few axial reflections and chances are that we don't record them. Also: accidental absences.
Systematic absences:

- For a $P_q$ axis along e.g. c (index l), axial reflections are only present if
  $l = n(p/q)$ where $n$ is an integer

- examples:

  $\begin{array}{ccc}
  2^1 & 2n & 2, 4, 6, \ldots \\
  3^1 & 3n & 3, 6, 9, \ldots \\
  4^1, 4^3 & 4n & 4, 8, 12, \ldots \\
  4^2 & 2n & 2, 4, 6, \ldots \\
  6^1, 6^5 & 6n & 6, 12, 18, \ldots \\
  6^2, 6^4 & 3n & 3, 6, 9, \ldots \\
  6^3 & 2n & 2, 4, 6, \ldots \\
  \end{array}$

But we may not record many axial reflections!
Systematic absences (cont.)

- centering causes systematic absences too, but these conditions apply to all $h, k, l$ (i.e. not to axial reflections only)

A centered \[ k + l = 2n \]
B centered \[ h + l = 2n \]
C centered \[ h + k = 2n \]
F centered \[ k + l = 2n, h + l = 2n, h + k = 2n \]
I centered \[ h + k + l = 2n \]
R (obverse) \[ -h + k + l = 3n \]
R (reverse) \[ h - k + l = 3n \]
Program to determine Laue group from unmerged intensities.

Strategy used in pointless:

- From unit cell dimensions, find the highest compatible lattice symmetry.

- Score each symmetry element (rotation) belonging to lattice symmetry using all pairs of observations related by that element. Score based on correlation coefficient, which is relatively independent of unknown scales.

- Score combinations of symmetry elements of all possible sub-groups (Laue-groups) of lattice symmetry group.

- Score possible space groups from axial systematic absences.
pointless example

Example: a false diamond

- data integrated in P1

- cell: 67.01 67.31 67.43 90.11 89.96 90.05
  whoa, looks like cubic....
  except that integration is difficult and scaling fails

- ... and indeed, pointless says:

  Lattice point group: P 4 3 2

- OK, let's score the symmetry elements!
Scoring each symmetry element for P432:

<table>
<thead>
<tr>
<th>Nelmt</th>
<th>Lklhd</th>
<th>Z-cc</th>
<th>CC</th>
<th>N</th>
<th>Rmeas</th>
<th>Symmetry &amp; operator (in Lattice Cell)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.933</td>
<td>8.93</td>
<td>0.89</td>
<td>23366</td>
<td>0.148</td>
<td>identity</td>
</tr>
<tr>
<td>2</td>
<td>0.044</td>
<td>0.77</td>
<td>0.08</td>
<td>45821</td>
<td>0.943</td>
<td>2-fold (1 0 1) {+l,-k,+h}</td>
</tr>
<tr>
<td>3</td>
<td>0.044</td>
<td>1.00</td>
<td>0.10</td>
<td>45091</td>
<td>0.906</td>
<td>2-fold (1 0-1) {-l,-k,-h}</td>
</tr>
<tr>
<td>4</td>
<td>0.055</td>
<td>-0.08</td>
<td>-0.01</td>
<td>45251</td>
<td>0.998</td>
<td>2-fold (0 1-1) {-h,-l,-k}</td>
</tr>
<tr>
<td>5</td>
<td>0.050</td>
<td>0.11</td>
<td>0.01</td>
<td>45053</td>
<td>0.979</td>
<td>2-fold (0 1 1) {-h,+l,+k}</td>
</tr>
<tr>
<td>6</td>
<td>0.045</td>
<td>0.53</td>
<td>0.05</td>
<td>45050</td>
<td>0.943</td>
<td>2-fold (1-1 0) {-k,-h,-l}</td>
</tr>
<tr>
<td>7</td>
<td>0.898</td>
<td>8.62</td>
<td>0.86</td>
<td>44891</td>
<td>0.248</td>
<td>2-fold k (0 1 0) {-h,+k,-l}</td>
</tr>
<tr>
<td>8</td>
<td>0.045</td>
<td>0.47</td>
<td>0.05</td>
<td>45125</td>
<td>0.959</td>
<td>2-fold (1 1 0) {+k,+h,-l}</td>
</tr>
<tr>
<td>9</td>
<td>0.866</td>
<td>8.45</td>
<td>0.84</td>
<td>44834</td>
<td>0.260</td>
<td>2-fold h (1 0 0) {+h,-k,-l}</td>
</tr>
<tr>
<td>10</td>
<td>0.883</td>
<td>8.53</td>
<td>0.85</td>
<td>44874</td>
<td>0.273</td>
<td>2-fold l (0 0 1) {-h,-k,+l}</td>
</tr>
<tr>
<td>11</td>
<td>0.056</td>
<td>-0.12</td>
<td>-0.01</td>
<td>91238</td>
<td>1.568</td>
<td>3-fold (1 1 1) {+l,+h,+k}</td>
</tr>
<tr>
<td>12</td>
<td>0.058</td>
<td>-0.21</td>
<td>-0.02</td>
<td>91364</td>
<td>1.614</td>
<td>3-fold (1-1-1) {-l,-h,-k}</td>
</tr>
<tr>
<td>13</td>
<td>0.053</td>
<td>-0.01</td>
<td>-0.00</td>
<td>91229</td>
<td>1.556</td>
<td>3-fold (1-1 1) {-l,+h,-k}</td>
</tr>
<tr>
<td>14</td>
<td>0.056</td>
<td>-0.14</td>
<td>-0.01</td>
<td>91231</td>
<td>1.565</td>
<td>3-fold (1 1-1) {-l,+h,-k}</td>
</tr>
<tr>
<td>15</td>
<td>0.055</td>
<td>-0.10</td>
<td>-0.01</td>
<td>91251</td>
<td>1.132</td>
<td>4-fold h (1 0 0) {+h,-l,+k}</td>
</tr>
<tr>
<td>16</td>
<td>0.044</td>
<td>0.65</td>
<td>0.07</td>
<td>91270</td>
<td>1.066</td>
<td>4-fold k (0 1 0) {+l,+k,-h}</td>
</tr>
<tr>
<td>17</td>
<td>0.046</td>
<td>0.37</td>
<td>0.04</td>
<td>91221</td>
<td>1.093</td>
<td>4-fold l (0 0 1) {-k,+h,+l}</td>
</tr>
</tbody>
</table>

Ooops, only the orthorhombic symmetry operators are present!
### Laue group determination:

<table>
<thead>
<tr>
<th>Laue Group</th>
<th>Lkhd</th>
<th>NetZc</th>
<th>Zc+</th>
<th>Zc-</th>
<th>CC</th>
<th>CC-</th>
<th>Rmeas</th>
<th>R-</th>
<th>Delta</th>
<th>ReindexOperator</th>
</tr>
</thead>
<tbody>
<tr>
<td>P m m m</td>
<td>0.944</td>
<td>8.38</td>
<td>8.62</td>
<td>0.24</td>
<td>0.86</td>
<td>0.02</td>
<td>0.23</td>
<td>1.14</td>
<td>0.1</td>
<td>[h,k,l]</td>
</tr>
<tr>
<td>2 P 1 2/m 1</td>
<td>0.019</td>
<td>7.46</td>
<td>8.78</td>
<td>1.31</td>
<td>0.88</td>
<td>0.13</td>
<td>0.19</td>
<td>0.96</td>
<td>0.1</td>
<td>[-h,-k,l]</td>
</tr>
<tr>
<td>3 P 1 2/m 1</td>
<td>0.017</td>
<td>7.40</td>
<td>8.73</td>
<td>1.33</td>
<td>0.87</td>
<td>0.13</td>
<td>0.20</td>
<td>0.96</td>
<td>0.1</td>
<td>[-h,-l,-k]</td>
</tr>
<tr>
<td>4 P 1 2/m 1</td>
<td>0.014</td>
<td>7.36</td>
<td>8.68</td>
<td>1.33</td>
<td>0.87</td>
<td>0.13</td>
<td>0.20</td>
<td>0.96</td>
<td>0.1</td>
<td>[-k,-h,-l]</td>
</tr>
<tr>
<td>-5 P -1</td>
<td>0.002</td>
<td>7.17</td>
<td>8.93</td>
<td>1.77</td>
<td>0.89</td>
<td>0.18</td>
<td>0.15</td>
<td>0.89</td>
<td>0.0</td>
<td>[-h,k,-l]</td>
</tr>
<tr>
<td>6 P 4/m</td>
<td>0.001</td>
<td>4.66</td>
<td>6.03</td>
<td>1.36</td>
<td>0.60</td>
<td>0.14</td>
<td>0.36</td>
<td>0.95</td>
<td>0.4</td>
<td>[l,h,k]</td>
</tr>
<tr>
<td>7 P 4/m</td>
<td>0.001</td>
<td>4.27</td>
<td>5.70</td>
<td>1.43</td>
<td>0.57</td>
<td>0.14</td>
<td>0.38</td>
<td>0.95</td>
<td>0.1</td>
<td>[k,l,h]</td>
</tr>
<tr>
<td>8 P 4/m</td>
<td>0.001</td>
<td>4.46</td>
<td>5.87</td>
<td>1.40</td>
<td>0.59</td>
<td>0.14</td>
<td>0.38</td>
<td>0.95</td>
<td>0.3</td>
<td>[h,k,l]</td>
</tr>
<tr>
<td>9 P 4/m m m</td>
<td>0.000</td>
<td>4.56</td>
<td>4.88</td>
<td>0.32</td>
<td>0.49</td>
<td>0.03</td>
<td>0.47</td>
<td>1.18</td>
<td>0.1</td>
<td>[k,l,h]</td>
</tr>
<tr>
<td>10 H -3</td>
<td>0.000</td>
<td>2.41</td>
<td>4.31</td>
<td>1.90</td>
<td>0.43</td>
<td>0.19</td>
<td>0.53</td>
<td>0.86</td>
<td>0.4</td>
<td>[-k+l,h+k,-h+k+l]</td>
</tr>
<tr>
<td>11 H -3</td>
<td>0.000</td>
<td>2.44</td>
<td>4.34</td>
<td>1.90</td>
<td>0.43</td>
<td>0.19</td>
<td>0.52</td>
<td>0.86</td>
<td>0.4</td>
<td>[-h+k,h+l,h+k-l]</td>
</tr>
<tr>
<td>12 H -3</td>
<td>0.000</td>
<td>2.46</td>
<td>4.36</td>
<td>1.89</td>
<td>0.44</td>
<td>0.19</td>
<td>0.52</td>
<td>0.86</td>
<td>0.4</td>
<td>[k-l,-h+l,h+k+l]</td>
</tr>
<tr>
<td>13 C 1 2/m 1</td>
<td>0.000</td>
<td>2.53</td>
<td>4.42</td>
<td>1.89</td>
<td>0.44</td>
<td>0.19</td>
<td>0.42</td>
<td>0.88</td>
<td>0.1</td>
<td>[k+l,-k+l,h]</td>
</tr>
<tr>
<td>14 H -3</td>
<td>0.000</td>
<td>2.52</td>
<td>4.41</td>
<td>1.89</td>
<td>0.44</td>
<td>0.19</td>
<td>0.52</td>
<td>0.86</td>
<td>0.4</td>
<td>[h-l,k+l,h-k+l]</td>
</tr>
</tbody>
</table>

Clear preference for primitive orthorhombic.
(Likelihood allows for the possibility of pseudo-symmetry.)
Translational symmetry: from systematic absences

<table>
<thead>
<tr>
<th>Zone</th>
<th>Reflection Condition</th>
<th>Number</th>
<th>PeakHeight</th>
<th>SD</th>
<th>Probability</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>screw axis 2(1) [a]</td>
<td>40</td>
<td>1.081</td>
<td>0.124</td>
<td>*** 0.968</td>
<td>h00: h=2n</td>
</tr>
<tr>
<td>2</td>
<td>screw axis 2(1) [b]</td>
<td>41</td>
<td>1.050</td>
<td>0.089</td>
<td>*** 0.984</td>
<td>0k0: k=2n</td>
</tr>
<tr>
<td>3</td>
<td>screw axis 2(1) [c]</td>
<td>37</td>
<td>1.086</td>
<td>0.143</td>
<td>*** 0.960</td>
<td>00l: l=2n</td>
</tr>
</tbody>
</table>

There's only one primitive orthorhombic space group with these elements:

<table>
<thead>
<tr>
<th>Spacegroup</th>
<th>TotProb</th>
<th>SysAbsProb</th>
<th>Reindex</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;P 21 21 21&gt; (19)</td>
<td>0.863</td>
<td>0.914</td>
<td></td>
<td>h00: h=2n, 0k0: k=2n, 00l: l=2n (zones 1,2,3)</td>
</tr>
<tr>
<td>.........................</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;P 21 21 2&gt; (18)</td>
<td>0.036</td>
<td>0.038</td>
<td></td>
<td>h00: h=2n, 0k0: k=2n (zones 1,2)</td>
</tr>
<tr>
<td>.........................</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;P 2 21 21&gt; (18)</td>
<td>0.029</td>
<td>0.030</td>
<td></td>
<td>0k0: k=2n, 00l: l=2n (zones 2,3)</td>
</tr>
<tr>
<td>.........................</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;P 21 2 21&gt; (18)</td>
<td>0.014</td>
<td>0.015</td>
<td></td>
<td>h00: h=2n, 00l: l=2n (zones 1,3)</td>
</tr>
</tbody>
</table>
Alternative indexing

- Alternative, valid but non-equivalent indexing schemes are possible in space groups where the point group symmetry is lower than that of the lattice group. These are related by symmetry operators present in the crystal system but not in the point group.

- Tetragonal, trigonal, hexagonal, cubic

E.g. in P3 there are four different ways to index:
- \((h, k, l)\)
- \((-h, -k, -l)\)
- \((k, h, -l)\)
- \((-k, -h, -l)\)

For the first dataset, you can choose any of these.

For subsequent datasets, autoindexing might pick an alternative one and you have to fix it in order to scale them together. Pointless will do this for you by comparing the unmerged test data to a merged reference dataset.
data processing flowchart

1. Inspect images
2. Spot search and index
3. Predict spots, estimate mosaicity
4. Define dead areas
5. Refine cell
6. Strategy
7. Integrate
8. Analyse
9. Scale, merge
10. Convert I to F
11. Analyse
12. Detwin
scaling, merging - why?

- Various physical factors lead to observed intensities being on different scales.
  - Some corrections are known (e.g. Lorentz and polarisation corrections.
  - Others can only be determined from the data.

- Scaling parameters should match the experiment

- Factors influencing scaling:
  1. Incident beam and camera
  2. Crystal and diffracted beam
  3. Detector

- Merging: all observations of each unique reflection to a single $I_{\text{mean}}$
  - reject outliers
  - estimate standard deviations
  - calculate statistics
beam intensity:
- variable on synchrotrons (beam decay, esp short half life beams)
- warming-up effects of in-house optics

illuminated volume might change with phi if beam smaller than crystal (synchrotrons)

absorption of primary beam by crystal

rotation speed variations
- difficult to detect, therefore we assume that rotation speed is constant.
- can be disastrous for data quality

shutter synchronisation errors
scaling - 2. crystal and diffraction

- absorption in secondary beam
  - significant at long wavelengths (CuKα or long wavelength SAD/MAD)

- radiation damage
  - synchrotron, or even high intensity in-house sources
  - not easily correctable, as structural changes are involved.
  - approaches: zero-dose extrapolation
    low-dose plus high redundancy data collection
scaling - 3. detector

- detector errors are difficult to detect

- detectors should be properly calibrated:
  - spacial distortion
  - sensitivity of response (floodfield)

- Dead area should be excluded at integration stage
  - including bad pixels, ideally
  - shadows of beamstop, cryo nozzle
scales are determined by comparison of symmetry-related reflections.
- gives internal consistency of intensities.
- However, we do no know the true intensities and hence internally consistent data does not mean that it is correct.
- This might lead to systematic errors.

Scaling function:

\[ g = g \text{ (rotation/image number)} \cdot g \text{ (time)} \cdot g \text{ (s)} \ldots \text{and others} \]

- primary beam
- B-factor
- absorption
- e.g. tails

- scale is smooth function of spindle rotation
- time dependent radiation damage correction

A discontinuous function of is usually less appropriate
scaling - scaling models

- Batch scaling
  - each frame gets separate scale
  - slow for many images
- Smooth scaling
  - extrapolate between images
- Local scaling
  - if reference dataset exists
  - higher reliability of outlier rejection
  - e.g. scale all wavelengths of MAD experiment together
    helps to extract weak signal
  - approximation of f" and relative Δf'

Single B-factor
- Smooth B-factor
  - equivalent of smooth scales
- Anisotropic B-factor

- Secondary beam correlation
  - in spherical harmonics
- Absorption correction
scaling - smooth scaling

- determine scale factor for every $n^{th}$ image
- interpolate between images
- fast (few scale factors to compute)

batch scales

smooth scales
scaling - B-factors

- dataset with little radiation decay
- small B-factor variation

- dataset with severe radiation decay
- large B-factor drop (larger than 10Å²)
scaling - in practice

For single input mtz

Smooth scaling

Create scaling protocol for which run (or for all) (SCALES RUN)

MTZ in: rapidata.peak_3-s.wtz

MTZ out: rapidata.peak_3-s.scale1.mtz

Convert to SFs & Wilson Plot

Estimated number of residues in the asymmetric unit: 200

Data Harvesting

Create harvest file in project directory

Define Output Datasets

The input file contains a single dataset, which will be transferred to the output file.

Project name: Unspecified and dataset name: Unspecified

Scaling Protocol

Scale: on rotation axis

Define constant

Index: on rotation axis

Apply on rotation axis & detector

Other: on rotation axis with secondary beam correction

Scaling Details

Run

Save or Restore

Close
Determination of scaling parameters depends on symmetry related reflections having different scales. If all observations of a reflection have the same value of the scale component, then there is no information about that component and it remains as a systematic error in the merged data (e.g. absorption).

Thus, to get intensities with the lowest absolute error, the symmetry-related observations should be measured in as different way as possible (e.g. rotation about multiple axes, different slicing). This will increase the merging $R$, but improve the estimate of $I$.

However, to measure the most accurate difference for phasing (e.g. anomalous experiments), observations should be measured in as similar way as possible...
Questions you might want to ask:

- Overall quality -- How does this dataset compare to others in the project?
- What is the resolution?
- Are there bad batches that should be excluded?
- Is there significant radiation damage?
- Is outlier rejection working well? Are outliers arranged in any systematic way?
- Is there anomalous signal?
Scaling - R-factors

R-factors: traditional measures of overall quality

- \( R_{\text{merge}} = \frac{\sum |I_h - <I_h>|}{\sum |<I_h>|} \)

The traditional measure of agreement. But: it increases with higher multiplicity even though data are better

- \( R_{\text{meas}} = R_{r.i.m} = \frac{\sum \sqrt{(n/n-1)} |I_h - <I_h>|}{\sum |<I_h>|} \)

The multiplicity-weight R-factor allows for the improvement of data with higher multiplicity. Useful when comparing different point-pointgroups.

- \( R_{p.i.m.} = \frac{\sum \sqrt{(1/n-1)} |I_h - <I_h>|}{\sum |<I_h>|} \)

Precision-indicating R-factor gets better with increasing multiplicity, estimates the precision of the merged \(<I>\).
Calculate standard deviation based on sigma from profile fitting for each reflection
- needs calibration against scatter in observations of that reflection
- error model adjusts sigma vs scatter

Standard deviation vs. intensity plots:
\[ \delta = \frac{I - \langle I \rangle}{\sigma} \]
- should be around 1
- similar to chi-squared test in HKL2000 or d*trek, but plotted by intensity

If there is a significant variation from 1, one should adjust the error model:
- SDFAC automatically modified (typically 1.2-1.3)
- increase SDADD in steps of 0.01

\[ \sigma(I) = SDFAC \times \sqrt{\sigma^2(I) + (SDADD \times I)^2} \]

Should not have too many rejections per image.
- mild outliers mean large standard deviations, not rejection
scaling - error model
scaling - Rmerge plot

- Radiation decay: steady increase in Rmerge

- Suboptimal initial batches: can be fixed by reindexing and reprocessing
scaling - completeness

- data completeness vs resolution

anomalous completeness!
scaling - the tails correction

- It there is significant Thermal Diffuse Scatter (TDS, spots have halos), full and summed partial reflections may not behave the same in a normal probability plot.

- Tails correction: correct this discrepancy.
outlier rejection

- reliable outlier rejection requires high multiplicity
- remember the backstop shadow, ideally at integration stage, but scala has mechanisms to exclude regions too.
- inspect the ROGUEPLOT

- what are outliers?
  - spots on dead areas
  - iceringgs
  - zingers
  - bad prediction (spot not there)
  - spot overlap
    - high mosaicity
    - multiple lattices

- position of outliers mapped on detector surface: ice rings...
data processing flowchart

- inspect images
- spot search and index
- predict spots, estimate mosaicty
- define dead areas
- refine cell
- strategy
- integrate
- analyse
- scale, merge
- convert I to F
- analyse
- detwin
1. classification of twins
2. twin detection
3. detwinning
4. structure solution and refinement (examples)
what is a twin?
- two (or more) crystals of the same kind
- joined together in a well defined orientation

2D twin crystal:

Twin operator (twin law):
\[
\begin{pmatrix}
1 & 0 \\
0 & -1
\end{pmatrix}
\]

Twin fraction:
\[\alpha_1 = \frac{14}{20}\]
\[\alpha_2 = \frac{6}{20}\]

Perfect twin: \(\alpha = 0.5\)
Partial twin: \(\alpha \neq 0.5\)

Diffraction intensities:
\[I_{\text{twin}} = \alpha_1 I + \alpha_2 I\]

Detwinning: deconvolution of twin components
### Classification

<table>
<thead>
<tr>
<th>twin</th>
<th>twin operator</th>
</tr>
</thead>
<tbody>
<tr>
<td>merohedral twin</td>
<td>belongs to the crystal system but not the point group</td>
</tr>
<tr>
<td>pseudo-merohedral twin</td>
<td>belongs to a higher crystal system</td>
</tr>
<tr>
<td>reticular merohedral twin</td>
<td>obverse-reverse twinning in rhombohedral</td>
</tr>
<tr>
<td>non-merohedral twin</td>
<td>arbitrary operator</td>
</tr>
</tbody>
</table>

**Epitaxial twinning:** non-merohedrals, split crystals  
lattice overlap not perfect  
**Hemihedral twins:** 2 domains are present
merohedral twins

twin operator:
belongs to the crystal system but not the Laue group

- exact overlap of lattices
- all reflections affected by twinning
- determination of the space group and Laue group difficult
- low $|E^2-1|$ value
- possible in the following crystal systems:
  - tetragonal
  - trigonal/hexagonal
  - cubic
pseudo-merohedral twins

twin operator:
belongs to a higher crystal system

- if the metric symmetry is higher than the symmetry of the crystal
  typical examples:
  - monoclinic with $\beta \approx 90^\circ$ mimics orthorhombic
  - monoclinic with $a \approx c$
  - monoclinic with $\beta \approx 120^\circ$ and 3 domains mimics $6/m$
- overlap of lattices not exact, depending on how well the higher metric symmetry is fulfilled
- all reflections affected by twinning
obverse-reverse twins

twin operator:
1. twofold axis $||$ to the threefold
2. twofold axis $||$ to a-b

- exact overlap of lattices
- lattice centering not easy to detect because
  - $h+k+l=3n$ (obverse)
  - $h-k+l=3n$ (reverse) are overlapping
- reciprocal lattice plot:
  - $l=3n$ layers: every third reflection observed
  - $l \neq 3n$ layers: every third reflection absent
- high ratio of non-overlapping reflections $\rightarrow$ easy to solve/refine
non-merohedral twins

- twin operator: arbitrary operator, e.g. a twofold rotation about a cell edge or diagonal

- overlap of lattices not exact; some spots split, some sharp
- twinning is obvious from the diffraction pattern
- indexing, cell refinement and integration problematic (non-protein or commercial software)
- majority of reflections not affected
- refinement problems (HKLF 5 format, free R-set selection - twin pairing error)
1. **Indexing**
   - split reflections (pseudo- and non-merohedral)
   - metric symmetry higher than Laue symmetry (pseudo-merohedral)

2. **Data reduction, data statistics**
   - systematic absences not consistent with any space groups
   - $R_{int}$ for the higher symmetry Laue group is just a bit higher than for the lower one
   - $|E^2-1|<<0.736$
   - the second moment of $<I>$ ($<I^2>/<I>^2$) for acentric reflections

   (perfect twins)
   - twin: **1.5**
   - no twinning: **2.0**

   - H-test: $H=(l_{1,twin}-l_{2,twin})/(l_{1,twin}+l_{2,twin})$
   - $\alpha = \frac{1}{2}(1-2<H>)$

   - intensity distribution tests affected by anisotropy and pseudo-symmetry!
   - estimation of $\alpha$ with ± 5% error
detection of twins

- L-test: (not affected by anisotropy): $L = (I_{1,twin} - I_{2,twin})/(I_{1,twin} + I_{2,twin})$
  (on reflection close in reciprocal space)
  $<|L|> = \frac{1}{2}, \frac{1}{3}$
  $<L^2> = \frac{2}{\pi}, \frac{1}{2}$
  perfect twin, acentric $\frac{3}{8}, \frac{1}{5}$

- cumulative intensity plot

- Britton plot (# of - intensities vs $\alpha$)
3. Structure solution

- Patterson map physically impossible (for heavy atoms)
- Molecular replacement problematic even if the search model highly homologous
- Twin ghosts, messy maps (unexplained blobs in density)
- Disorders that cannot be explained or modeled reasonably
- High R-factors
detection of twins

alternate conformations vs different species

P6\textsubscript{5}22
1 mol / ASU
alternate conformations of symmetry equivalents clashing

P6\textsubscript{5}
2 mol / ASU
after 1 cycle of twin refinement

P6\textsubscript{5}
2 mol / ASU
final refined structure
detection of twins

map quality

P3₂²¹
unexplained blobs, R > 30

P₃₂
TWIN 0 1 0 1 0 0 0 0 0 -1
better maps, R ~ 20
Diffraction intensities of two overlapping reflections related by the twin operation but not by crystallographic symmetry:

\[
I_{1,\text{twin}} = (1-\alpha)I_1 + \alpha I_2 \\
I_{2,\text{twin}} = \alpha I_1 + (1-\alpha)I_2
\]

\[
I_1 = \frac{((1-\alpha)I_{1,\text{twin}} - \alpha I_{2,\text{twin}})}{(1-2\alpha)} \\
I_2 = \frac{((1-\alpha)I_{2,\text{twin}} - \alpha I_{1,\text{twin}})}{(1-2\alpha)}
\]

\[
v(I_1) = \left(\frac{(1-\alpha)}{(1-2\alpha)}\right)^2 (\sigma_{I_{1,\text{twin}}}^2) + \left(\frac{\alpha}{(1-2\alpha)}\right)^2 (\sigma_{I_{2,\text{twin}}}^2) \\
v(I_2) = \left(\frac{(1-\alpha)}{(1-2\alpha)}\right)^2 (\sigma_{I_{2,\text{twin}}}^2) + \left(\frac{\alpha}{(1-2\alpha)}\right)^2 (\sigma_{I_{1,\text{twin}}}^2)
\]

\[
v(I_1) = \left(\frac{(1-\alpha)}{(1-2\alpha)}\right)^2 \sigma_{I_{1,\text{twin}}}^2 + \left(\frac{\alpha}{(1-2\alpha)}\right)^2 \sigma_{I_{2,\text{twin}}}^2
\]

\[
v(I_2) = \left(\frac{(1-\alpha)}{(1-2\alpha)}\right)^2 \sigma_{I_{2,\text{twin}}}^2 + \left(\frac{\alpha}{(1-2\alpha)}\right)^2 \sigma_{I_{1,\text{twin}}}^2
\]

Unreliable results if \(\alpha \approx 0.5\)
## Human Hydroxyacid Oxidase 1

- **Cell:**
  - 97.300 97.300 80.540 90 90 90
  - **I4** or **I422**

<table>
<thead>
<tr>
<th>Laue group</th>
<th>4/mm</th>
<th>4/m</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&lt;sub&gt;int&lt;/sub&gt;</td>
<td>10.1%</td>
<td>7.5%</td>
</tr>
<tr>
<td>Matthews</td>
<td>1.36</td>
<td>2.72</td>
</tr>
<tr>
<td>Solvent c.</td>
<td>9.7%</td>
<td>54.8%</td>
</tr>
</tbody>
</table>

- **xprep:**
  - Mean $|E^*E^-1| = 0.586$
  - TWIN 0 1 0 1 0 0 0 0 -1
  - BASF 0.363 [NC]

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

- **BASF**
  - 0.363 [NC]

- **R/R<sub>free</sub>** = 12.6% / 15.5%

- **SHELXL in I4**
  - Twin law: 0 1 0 1 0 0 0 -1
  - 0 0 -1

- **Britton plot**

- **Cumulative intensity plot**
example: pseudo-merohedral

Human Retinal Short-Chain Dehydrogenase/Reductase 3

- cell: 167.115 98.823 167.463
  90.00 115.87 90.00
  \( P2_1 \) (Rmerge=0.08)

- or: 178.666 284.619 98.990
  90.00 90.00 90.90
  \( C222_1 \) (Rmerge=0.17)

- overlapping reflections

16 mol / ASU
R/R\text{free} = 18.2\% / 22.9\%

\( P2_1 \)

TWIN 0 0 -1 0 -1 0 -1 0 0
twin refinement not pursued as no significant improvement
Human Transplantation Specific Transplantation Antigen 3

- cell: 91.640 91.630
  429.980
  90.00 90.00 120.00

P6 or R3

- obverse/reverse test: Mean I
  for
  obv only: 81.4
  rev only: 78.0

<table>
<thead>
<tr>
<th>l=3</th>
<th>nei every 3rd observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>TWIN</td>
<td>-1 0 0 0 -1 0 0 0 1</td>
</tr>
<tr>
<td>BASF</td>
<td>0.489</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>l=2</th>
<th>every 3rd missing</th>
</tr>
</thead>
</table>
Human Adenylate Kinase 1

- **cell:** 35.184 43.919 64.742
  90.80 95.61 106.26

- **twin operation:**
  - twofold (180°) rotation around cell diagonal
- **partially overlapping reflections**

Data reduction:
EVALCCCD

Refinement:
SHELXL, HKLF 5

**free R set!**

R/R<sub>free</sub> = 20.8% / 27.8%
indexing problems
no MR solution with good homologue
iffy maps, high R-factors
etc...

is twinning possible?

twinning tests, twin law, twin fraction

non-merohedral
reprocess, refine
- EVALCCD, SAINT, ...
- SHELXL

pseudo-merohedral, obverse-reverse
refine
- REFMAC
- PHENIX.REFINE
- SHELXL
- CNS

merohedral,
partial twins: detwin & refine
*Acta Cryst* **A36**, 578-581

*Acta Cryst* **D49**, 375-380

*Acta Cryst* **D51**, 819-823

Yeates, T.O. (1997), Detecting and overcoming crystal twinning.  

*Acta Cryst* **B54**, 443-449

*Acta Cryst* **B58**, 477-481

*Acta Cryst.* **D59**, 1124-1130


Yeates twin server and tutorial:  [www.doe-mbi.ucla.edu/Services/Twinning](http://www.doe-mbi.ucla.edu/Services/Twinning)


Regine’s twin site:  [http://shelx.uni-ac.gwdg.de/~rherbst/twin.html](http://shelx.uni-ac.gwdg.de/~rherbst/twin.html)

CNS tutorials, phenix documentation etc etc etc