Twinning in protein crystals

Zbigniew Dauter
Twinning

when reflections from more than one lattice (crystal) overlap

1. merohedral
   - crystal symmetry is the subgroup of the lattice symmetry

2. pseudomerohedral
   - coincidence of cell dimensions

3. non-merohedral
   - coincidence of the supercell

4. epitaxial
   - overlap in one or two dimensions

4. crystal cracking or splitting
   - this is not twinning!
Crystal consisting of several domains

- Pyramid
- Bipyramid
Crystal consisting of several domains

Domain 1

Domain 2

Twin
Crystal consisting of several domains

Domain 1

Domain 2

Twin
Crystal consisting of several domains

If size of domains is larger than the coherence length of X-rays, addition of intensities,

\[ \text{= twinning} \]

\[ \gg \lambda \]

If size of domains is smaller than the coherence length of X-rays, addition of amplitudes,

\[ \text{= disorder} \]
Reciprocal lattices
symmetry 4

Only four-fold axis exists in both cases but the polar axis is in opposite directions
Reciprocal lattices overlapped mimicking symmetry 422

Superposition of two lattices
Reciprocal lattices overlapped mimicking symmetry 422

Superposition of two lattices has higher symmetry
Merohedral twinning

is possible when lattice symmetry is higher than crystal symmetry

(for macromolecules, without center of symmetry)

crystal point group  lattice symmetry
4  422 (4/mmm)  
3 > 321, 312, 6 > 622 (6/mmm)  
32  622 (6/mmm)  
6  622 (6/mmm)  
23  432 (m3m)
**Twinning operators**

Any operation existing in lattice symmetry but absent in crystal point group

<table>
<thead>
<tr>
<th>crystal</th>
<th>lattice</th>
<th>twinning operator</th>
<th>operator</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>422</td>
<td>k,h,-l</td>
<td>2-fold</td>
</tr>
<tr>
<td>3</td>
<td>622</td>
<td>-h,-k,l</td>
<td>2-fold</td>
</tr>
<tr>
<td></td>
<td></td>
<td>k,h,-l</td>
<td>2-fold</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-k,-h,-l</td>
<td>2-fold</td>
</tr>
<tr>
<td>321</td>
<td>622</td>
<td>-h,-k,l</td>
<td>2-fold</td>
</tr>
<tr>
<td>312</td>
<td>622</td>
<td>-h,-k,l</td>
<td>2-fold</td>
</tr>
<tr>
<td>6</td>
<td>622</td>
<td>k,h,-l</td>
<td>2-fold</td>
</tr>
<tr>
<td>23</td>
<td>432</td>
<td>k,h,-l</td>
<td>2-fold</td>
</tr>
</tbody>
</table>

NCI, Macromolecular Crystallography Laboratory, Synchrotron Radiation Research Section @ ANL
Pseudomerohedral twinning

when cell dimensions are special
and lattice has higher metric symmetry

eamples:

<table>
<thead>
<tr>
<th>cell</th>
<th>lattice</th>
</tr>
</thead>
<tbody>
<tr>
<td>monoclinic P β~90°</td>
<td>orthorhombic P</td>
</tr>
<tr>
<td>monoclinic P a~b</td>
<td>orthorhombic C</td>
</tr>
<tr>
<td>or 2acosβ+c = 0</td>
<td>&quot;</td>
</tr>
<tr>
<td>rhombohedral α~60°</td>
<td>cubic F</td>
</tr>
<tr>
<td>rhombohedral α~90°</td>
<td>cubic P</td>
</tr>
<tr>
<td>rhombohedral α~109.5°</td>
<td>cubic I</td>
</tr>
</tbody>
</table>
Cubic F cell - Rhombohedral R with $\alpha = 60^\circ$

Cubic F cell can accommodate:
4 orientations of R32 crystals
8 orientations of R3 crystals
Pseudomerohedry

\[ 2a \cos \beta + c = 0 \]
Pseudomerohedry

\[ 2a \cos \beta + c = 0 \]
Non-merohedry (Reticular twinning)
Non-merohedry (Reticular twinning)

8a \cos \beta + c = 0
Non-merohedry (Reticular twinning)

almo0

Reticular twinning occurs when only subset of reflections coincides.

Every fourth line of reflections coincides perfectly
Distribution of intensities differs from ordinary (Wilson) statistics

Because intensities, not amplitudes add up from different domains

There are less very weak and very strong reflections – smaller probability that two extreme intensities combine
Identifiers of twinning

based on

> statistics of all intensities
  (data can be processed in low or high symmetry)

> comparison of twin-related intensities
  (data have to be processed in proper. Low symmetry)

- Wilson ratios, $<I^2>/<I>^2$
- higher moments of E
- $N(z)$ cumulative intensity test
- $H$ test of Yeates
- negative intensity Britton test
- $L$ test of Padilla & Yeates
Wilson ratios

\[ \frac{\langle I^2 \rangle}{\langle I \rangle^2} = 2 \quad \text{for not twinned} \]

\[ \frac{\langle I^2 \rangle}{\langle I \rangle^2} = 1.5 \quad \text{for 50 \% twinned} \]

Moments of E

\[ \langle |E| \rangle = 0.866 \quad \text{for not twinned} \]

\[ \langle |E| \rangle = 0.940 \quad \text{for 50 \% twinned} \]

\[ \langle |E|^3 \rangle = 1.329 \quad \text{for not twinned} \]

\[ \langle |E|^3 \rangle = 1.175 \quad \text{for 50 \% twinned} \]
capsid protein gpD from bacteriophage λ

trimer, 3 * (93 a.a. + 2 SeMet)

1.7 Å MAD data from X9B @ NSLS

monoclinic, P2_1

a = 45.51, b = 68.52, c = 45.52 Å
β = 104.4°

45 % solvent

Refined by SHELXL with native data

@ 1.3 Å, R = 13 %, α = 36 %
gpD diffraction

NCI, Macromolecular Crystallography Laboratory, Synchrotron Radiation Research Section @ ANL
cumulative intensity N(z) test
(look at TRUNCATE output)
data can be processed in wrong symmetry

- number of reflections > z, percent of average intensity

it is not highly probable that both very weak or very strong reflections will overlap after twinning
- so for twinned crystal there are less weak reflections
- sigmoidal N(z) curve

$$\frac{\langle I^2 \rangle}{\langle I \rangle^2} = 2 \quad \text{for not twinned}$$

$$\frac{\langle I^2 \rangle}{\langle I \rangle^2} = 1.5 \quad \text{for 50 % twinned}$$
gpD SeMet peak, cumulative intensity $N(z)$ test

$\alpha = 0$

$\alpha = 50\%$
negative intensity (Britton) test
data have to be processed in low symmetry

\[ I_{\text{obs}}^1 = (1-\alpha)I_{\text{detw}}^1 + \alpha I_{\text{detw}}^2 \]

\[ I_{\text{obs}}^2 = \alpha I_{\text{detw}}^1 + (1-\alpha)I_{\text{detw}}^2 \]

detwinning equations:

\[ I_{\text{detw}}^1 = [(1-\alpha)I_{\text{obs}}^1 - \alpha I_{\text{obs}}^2]/(1-2\alpha) \]

\[ I_{\text{detw}}^2 = [(1-\alpha)I_{\text{obs}}^2 - \alpha I_{\text{obs}}^1]/(1-2\alpha) \]

if \( \alpha \) is too high, there will be many negative estimations
gpD SeMet peak, negative intensity (Britton) test
gpD SeMet peak, negative intensity (Britton) test
Yates S(H) test

Data have to be processed in low symmetry

\[ H = \frac{|I_1^{\text{obs}} - I_2^{\text{obs}}|}{(I_1^{\text{obs}} + I_2^{\text{obs}})} \]

\( I^1 \) and \( I^2 \) are twin-related

Cumulative distribution of \( H \) is linear

\[ S(H) = \frac{H}{(1 - 2\alpha)} \]

\[ \langle H \rangle = \frac{1}{2} - \alpha \]

\[ \langle H^2 \rangle = (1 - 2\alpha)^2/3 \]

Weak reflections can be discarded
gpD SeMet peak, Yates S(H) test
Padilla & Yates L test

data can be processed in any symmetry

\[ L = \frac{(I_{obs}^{1} - I_{obs}^{2})}{(I_{obs}^{1} + I_{obs}^{2})} \]

\( I^1 \) and \( I^2 \) are close in reciprocal space
(they indices differ by a small vector)

cumulative distribution of \( L \) is linear

\[ N(|L|) = |L| \]

\[ \langle |L| \rangle = 1/2 \]
\[ \langle L^2 \rangle = 1/3 \]

NCI, Macromolecular Crystallography Laboratory, Synchrotron Radiation Research Section @ ANL
Padilla & Yates L test

Data can be processed in any symmetry

$I^1$ and $I^2$ are close in reciprocal space
their indices differ by a small vector

This statistic depends on the local differences of intensities

If the NCS translation is close to $\frac{1}{2}, 0, 0$
reflections $h=2n+1$ are weak,
but if only reflections with the same parity of $h$ are compared ($|h_1-h_2| = 2, 0, 0$)
then the L-test is not biased

also with anisotropic diffraction
gpD anomalous Patterson
gpD anomalous Patterson

original data

detwinned data

scrambled data

NCI, Macromolecular Crystallography Laboratory, Synchrotron Radiation Research Section @ ANL
Molecular replacement of twinned structures

MR is based on the overlap of Pattersons (from the known model and unknown data) therefore two solutions can be expected (but the contribution of other twinned intensities introduce noise)

MAD and SAD phasing

mixing of twinned intensities diminishes the inherently small anomalous signal (only the same Friedel mates are mixed)
Phasing results of gpD

\[ \alpha \quad \text{orig} \quad 10\% \quad 20\% \quad 35\% \quad 50\% \quad \text{scrambled} \]

**SHELXD results (SAD peak data)**

<table>
<thead>
<tr>
<th>solutions %</th>
<th>98</th>
<th>87</th>
<th>100</th>
<th>99</th>
<th>97</th>
</tr>
</thead>
<tbody>
<tr>
<td>best CC all</td>
<td>31.0</td>
<td>32.9</td>
<td>34.9</td>
<td>31.1</td>
<td>20.8</td>
</tr>
<tr>
<td>best PATFOM</td>
<td>34.9</td>
<td>36.5</td>
<td>38.4</td>
<td>32.3</td>
<td>22.8</td>
</tr>
</tbody>
</table>

**SHELXE SAD phasing**

<table>
<thead>
<tr>
<th>FOM</th>
<th>0.72</th>
<th>0.71</th>
<th>0.69</th>
<th>0.63</th>
<th>0.64</th>
</tr>
</thead>
<tbody>
<tr>
<td>map CC</td>
<td>0.86</td>
<td>0.86</td>
<td>0.84</td>
<td>0.86</td>
<td>0.79</td>
</tr>
</tbody>
</table>
SHELXD peaks for peak data

(6 Se atoms expected)

<table>
<thead>
<tr>
<th>peak</th>
<th>orig</th>
<th>10%</th>
<th>20%</th>
<th>35% scrambl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>0.88</td>
<td>0.86</td>
<td>0.86</td>
<td>0.85</td>
</tr>
<tr>
<td>3</td>
<td>0.82</td>
<td>0.83</td>
<td>0.83</td>
<td>0.81</td>
</tr>
<tr>
<td>4</td>
<td>0.78</td>
<td>0.79</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>5</td>
<td>0.70</td>
<td>0.72</td>
<td>0.71</td>
<td>0.69</td>
</tr>
<tr>
<td>6</td>
<td>0.62</td>
<td>0.60</td>
<td>0.59</td>
<td>0.57</td>
</tr>
<tr>
<td>7</td>
<td>0.14</td>
<td>0.11</td>
<td>0.13</td>
<td>0.16</td>
</tr>
</tbody>
</table>
**Phasing results**

*original*  \( \alpha = 35\% \)  *scrambled*

NCI, Macromolecular Crystallography Laboratory, Synchrotron Radiation Research Section @ ANL
Phasing results

original $\alpha = 35\%$ scrambled
Conclusion

Do not give up too early
Try various options
If a model can be built,
It can be properly refined with SHELXL, PHENIX or CNS