# Twinning and other pathologies

Andrey Lebedev

CCP4

OD-structures

Twinning by (pseudo)merohedry

Statistics of one observation

Statistics of two observations

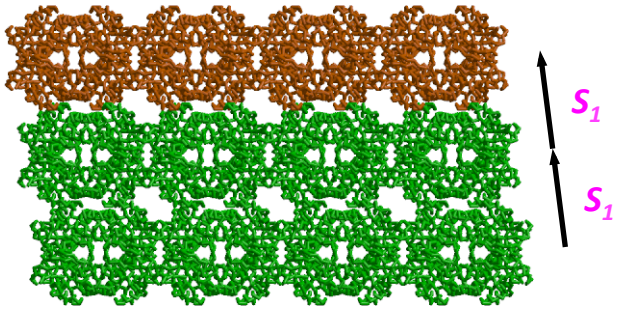
Twinning tests summary

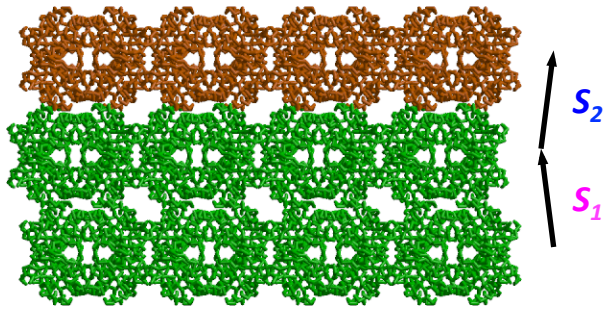
Space group validation

# OD-structures

- identical layers
- identical interfaces between the layers
- but: two or more ways of packing three adjacent layers

*) MX: "identical" means Ca r.m.s.d. < 1 A



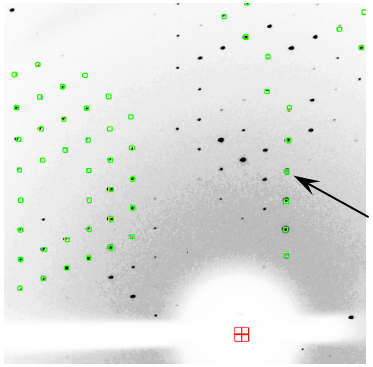




*) $S_1$ and $S_2$. are called stacking vectors

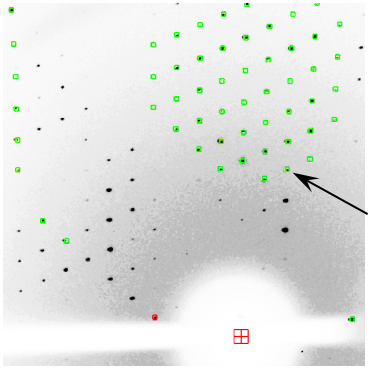
- two-dimensional periodicity
- a potential for disorder in the third dimension

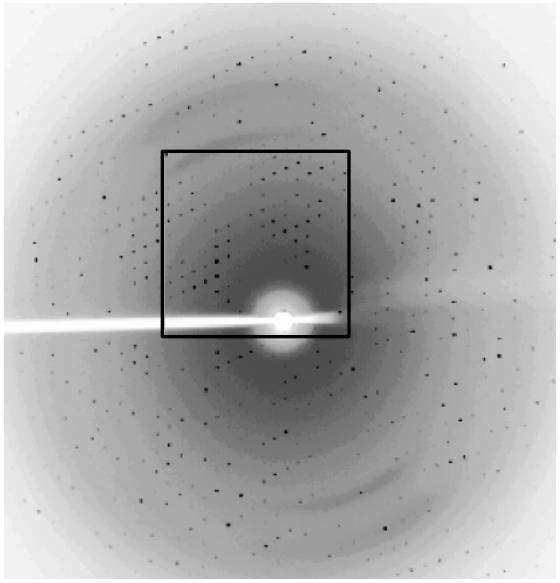
# Example 1: OD-twin (twin by lattice pseudomerohedy)

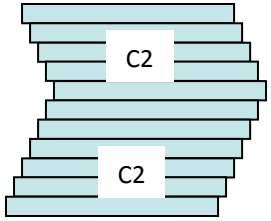
Indexing in C2



Indexing in C2





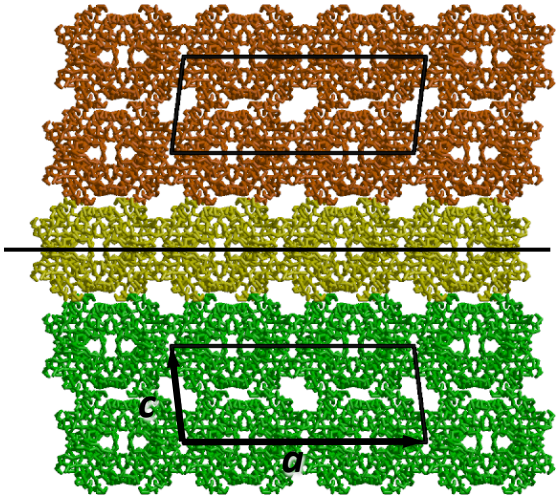


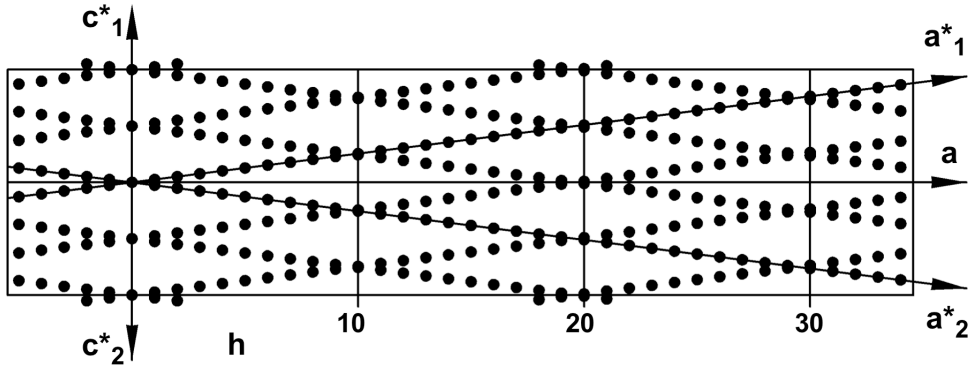

L-2-haloacid dehalogenase
from *Sulfolobus tokodaii*
Rye *et al.* (2007) *Acta Cryst.* **D**67

The diffraction images can be indexed in C2 with two different orientation of the crystal

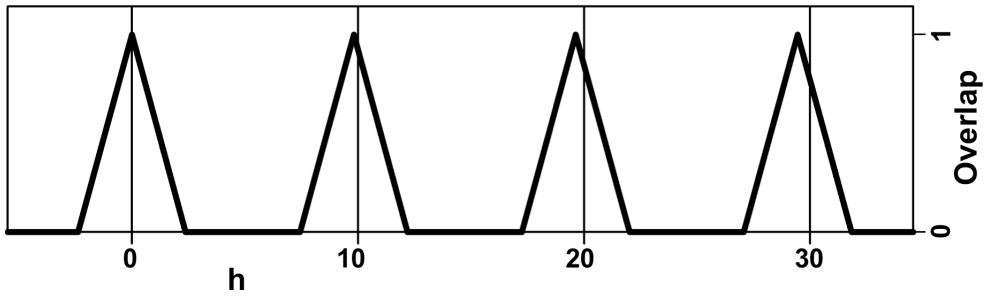
Some reflections from two lattices overlap.

# OD-twin: real and reciprocal lattices



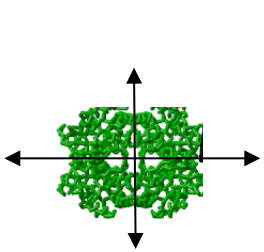

Maximum overlap is not at exactly integer *h*.

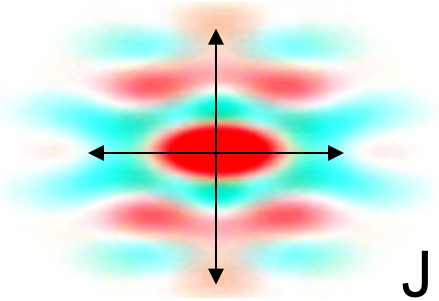



Twinning by reticular merohedry with twin index 10 and obliquity 0.1°

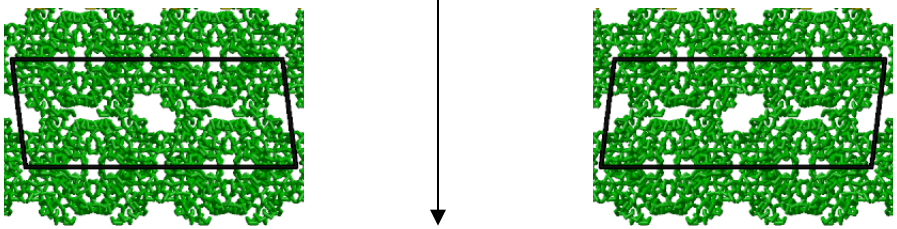
Integration of a single lattice: in effect, twinning coefficient depends on h

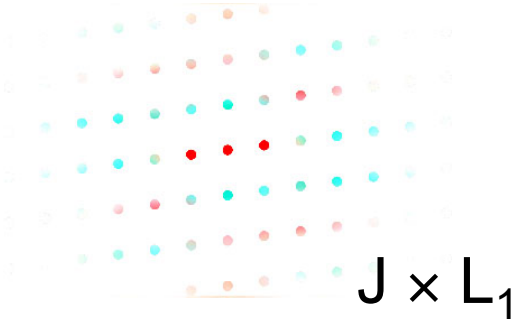
# Intensities of the overlapping reflections



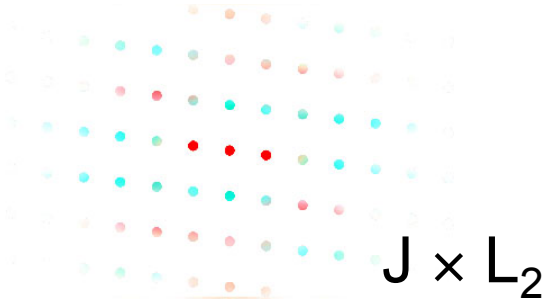



Fourier transform
of the tetramer






Diffraction pattern
of domain 1



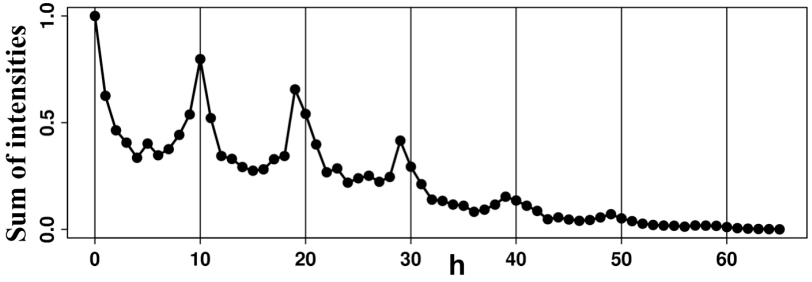

Diffraction pattern
of domain 2

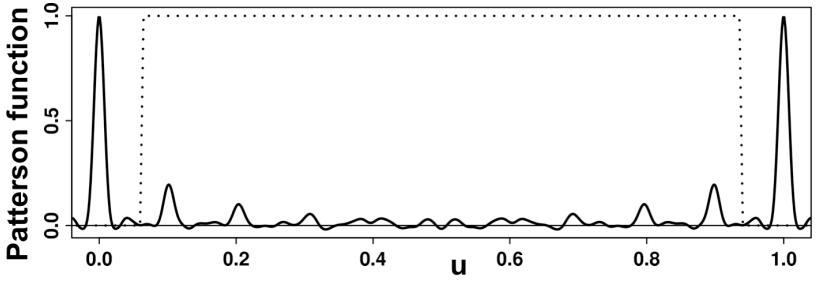
Tetramers in different twin domains are in the same orientation

Therefore, if reflections of the two lattices overlap, they have close intensities. The stronger the overlap, the closer the intensities are.

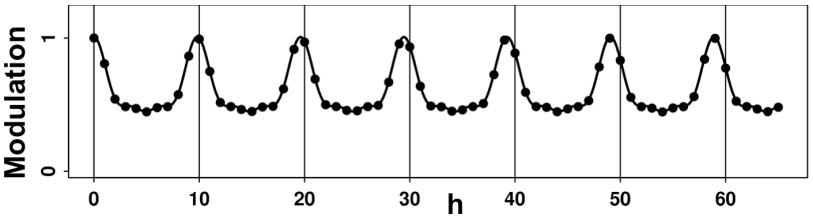
# Demodulation

Original data: R / R-free = 0.21 / 0.27



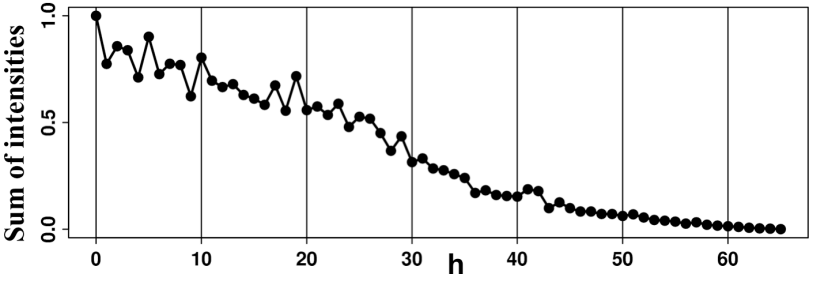



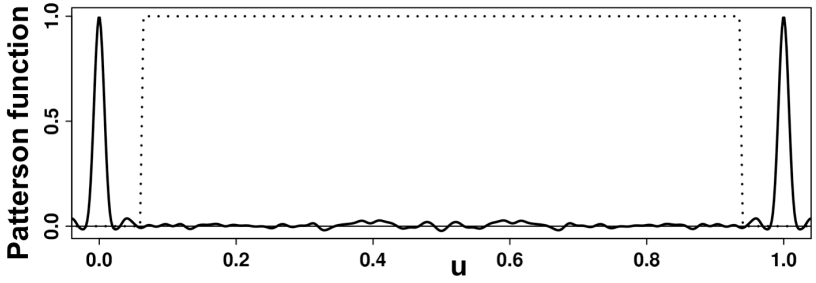

Modulation function




$$q'(h) = p_0 + p_1 \cos(2\pi th) + p_2 \cos(4\pi th) + \ldots$$

Corrected data: R / R-free = 0.16 / 0.23



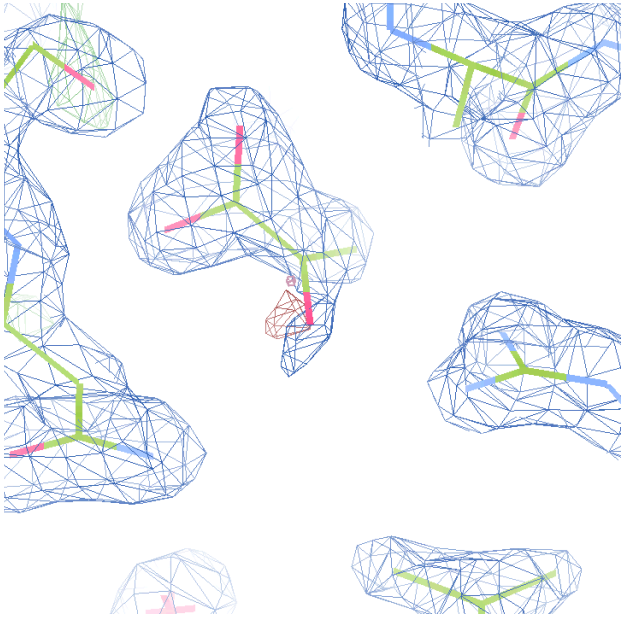



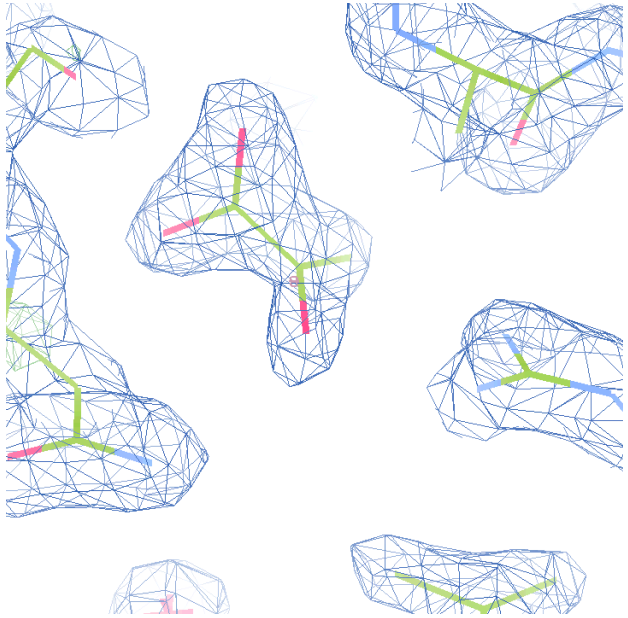
# OD-twin: Improvement in the electron density

Visually, improvement occurred only for the electron density for solvent molecules

(Poor density for solvent was the original reason for data revision)

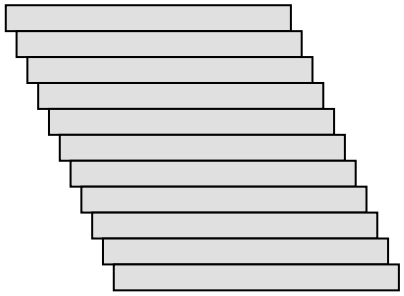
The electron density maps (2-1 at 1.5σ and 1-1 at 3σ)
around the pyruvate molecule before and after demodulation



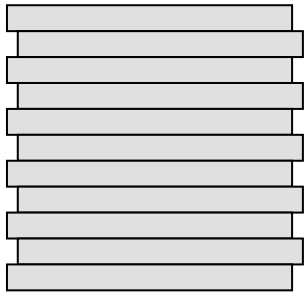


# OD-structures

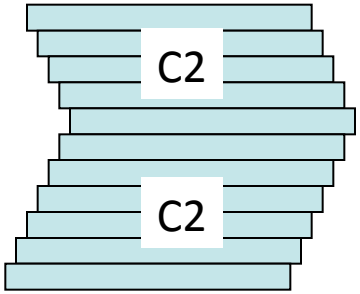
Single crystal



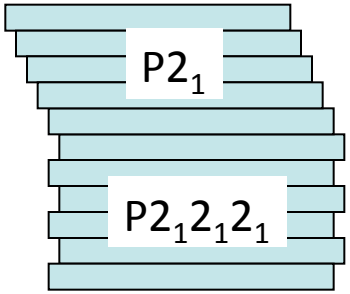
Single crystal



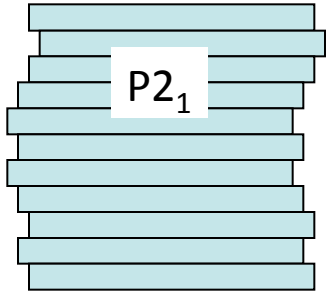
OD-twin




Allotwin




Partially disordered OD-structure




*Examples 1 & 2*

*Example 3*

*Examples 4 & 5*

# Classification: OD-structures *vs.* twins

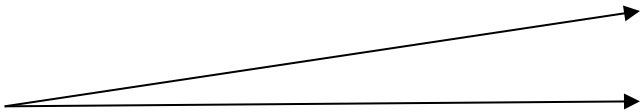
OD-structures:

Single crystals

allotwin

OD-twin

(partially) disordered OD-structure



Twinning:

by (pseudo)merohedry

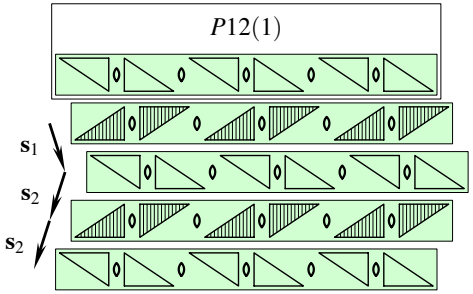
by reticular (pseudo)merohedry

...

This is structure based classification of a specific class of structures

This is geometry based classification accounting for crystal and lattice symmetries.

# Symbols for groupoid symmetry




$P \quad 1 \quad 2 \quad (1)$

$\{ \quad 2_P \quad 1 \quad (2_2) \quad \}$

In $2_P$, P is a non-integer subscript.

Special values of P correspond to space group symmetry or specialised groupoid symmetry

The following types are possible
(I) two surfaces of a single layer are identical;
(II) two surfaces of a single layer are different and contacts are made by different surfaces.
(III) two surfaces of a single layer are different but contacts are made by identical surfaces.

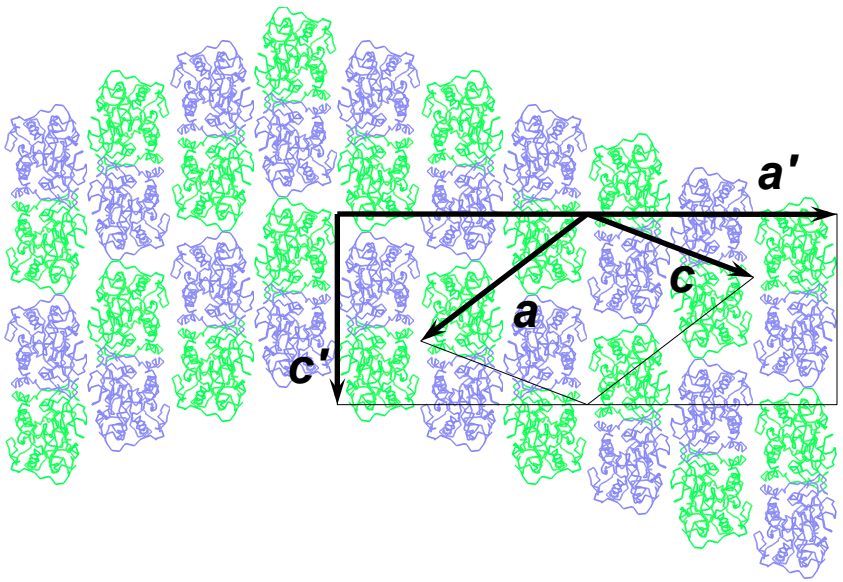
An example of symbol for groupoid of type (III):

$P \quad 1 \quad 1 \quad (4) \quad 1 \quad 1$

$\{ \quad 2_P \quad 2_Q \quad (1) \quad 2_U \quad 2_V \quad \}$

$\{ \quad 2_{P'} \quad 2_{Q'} \quad (1) \quad 2_{U'} \quad 2_{V'} \quad \}$

# Example 2: OD-twin with zero obliquity




Uppenberg *et al*. (1995).
*Biochemistry* **34**, 16838-51.

Molecule: Lipase B from *Candida antarctica*

PDB code 1lbs

Space group: C2

a = 95.9 Å, b = 95.6 Å, c = 81.8 Å

β =122.2°

OD layer: $P(2)2_12_1$

- The data were processed in C2 but in the twin lattice (twin index = 3)

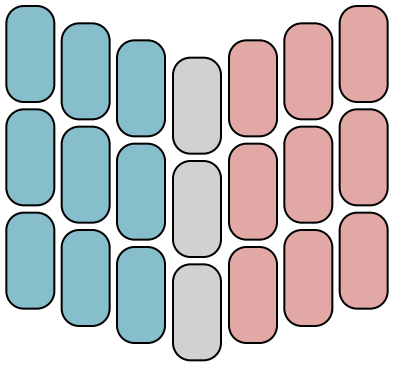
  a'=229.5 Å, c'=86.8Å, β =90°

- non-overlapping reflections from the minor twin component were removed

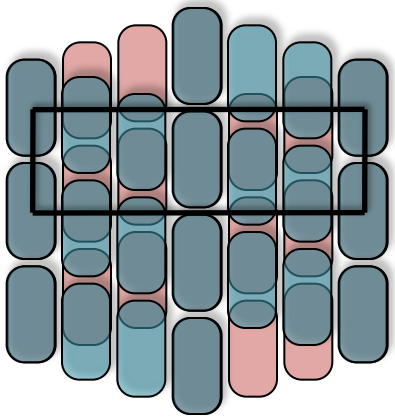
- overlapping reflections were detwinned

# Example 2: OD-twin with zero obliquity

This packing could be assumed by similarity with the previous example



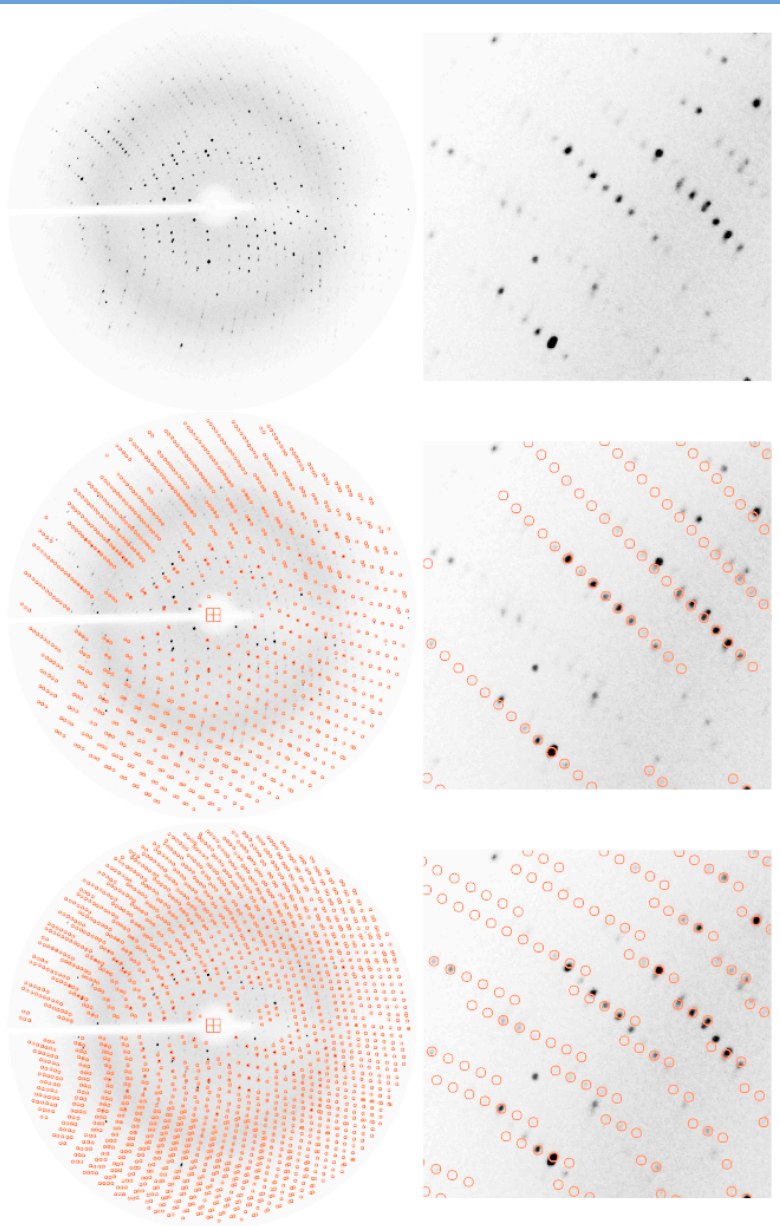
This packing is more likely to occur as it explains the exactly orthorhombic twin lattice

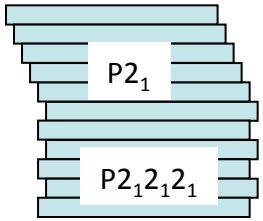


| | | |
|---|---|---|
| The previous example: | twin index 10 | obliquity 0.1° |
| This example: | twin index 3 | obliquity 0° |

In general, protein OD-twins frequently have zero obliquity (twins by metric merohedry)

# Example 3: allotwin






Crystals of Lon protease
Resolution 3Å

Dauter *et al.* (2005).
*Acta Cryst.* D**61**, 967-975.

$P2_1$

a = 48.5 Å
b = 86.3 Å
c = 138.0 Å
β = 92.3°

$P2_12_12_1$

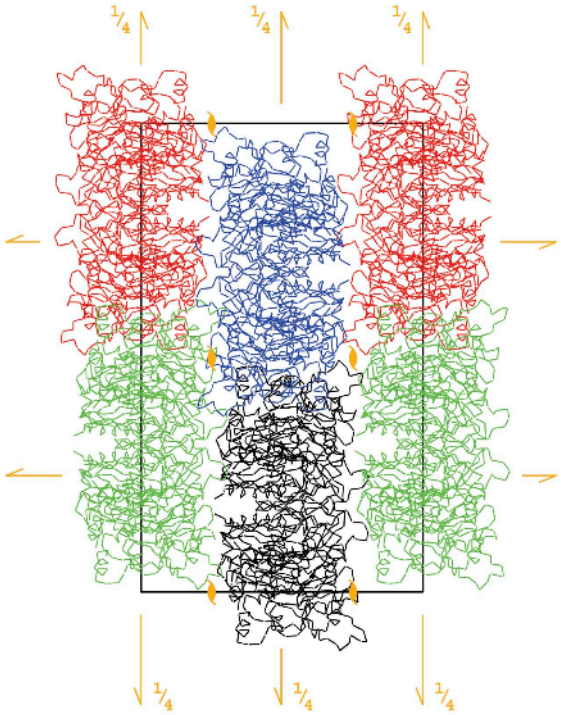
a = 86.3 Å
b = 90.6 Å
c = 148.0 Å

# Example 3: allotwin

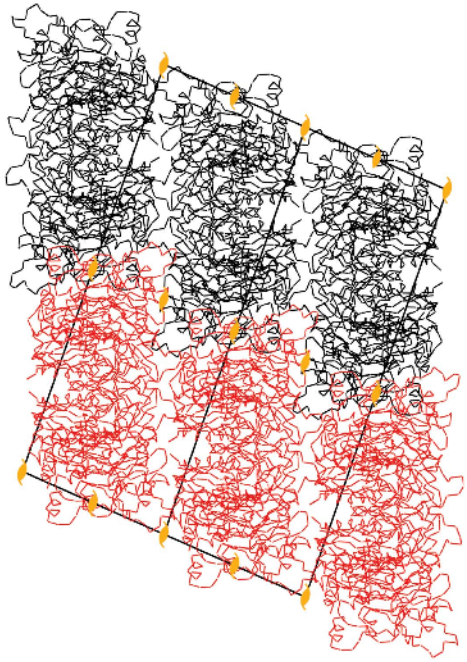
Crystals of Lon protease
Resolution 3Å

Dauter *et al.* (2005).
*Acta Cryst.* D**61**, 967-975.

PDB code 1z0t




PDB code 1z0v



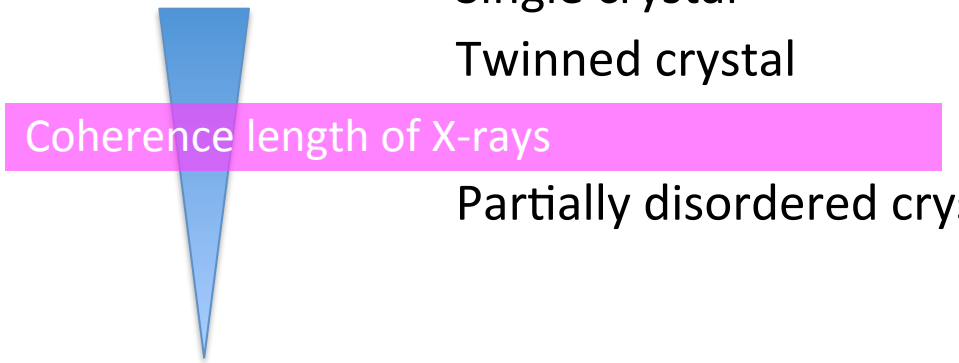
Structures of both crystal forms were solved

| | $P2_12_12_1$ | $P2_1$ |
|---|---|---|
| R / R-free | 0.19 / 0.35 | 0.21 / 0.31 |

# Crystal disorder

Twinning, partial disorder: Missing global periodicity

size of ordered domains

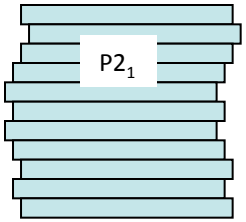



Single crystal (Single ordered domain)

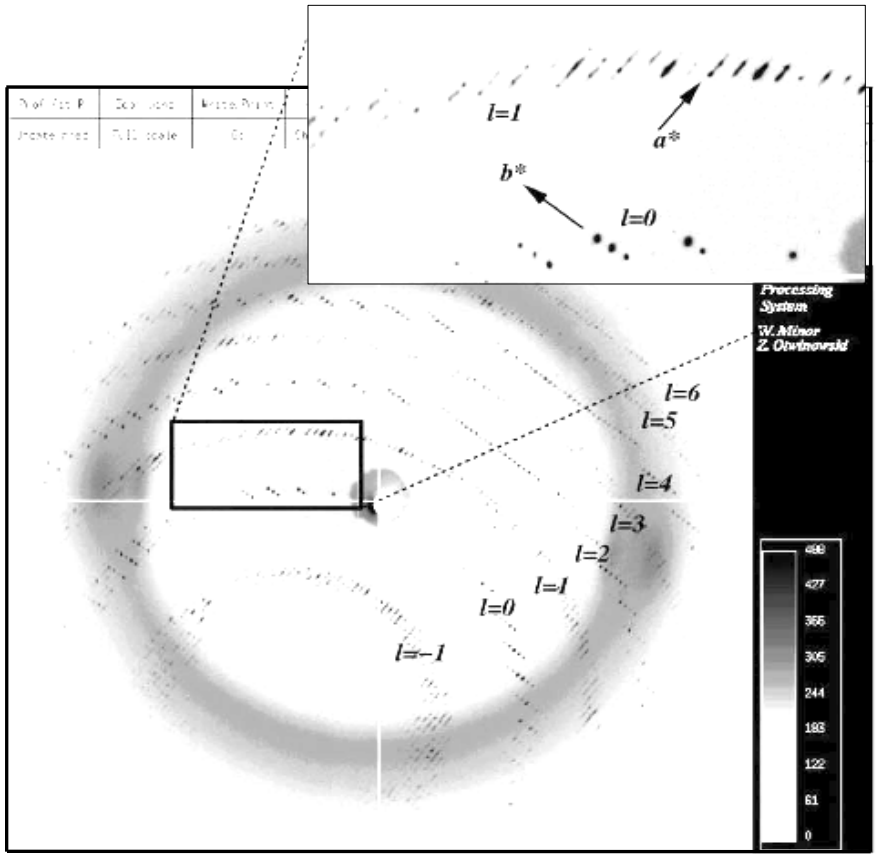
Twinned crystal (Two or more ordered domains)

Partially disordered crystal (Many ordered domains)

# Example 4: partially disordered OD-structure




***a**** Wang *et al.* (2005). *Acta Cryst.* D**61**, 67-74.




Crystals of Phi29 DNA polymerase
Resolution 2.2Å

The translation symmetry is not global in the direction ***a****.

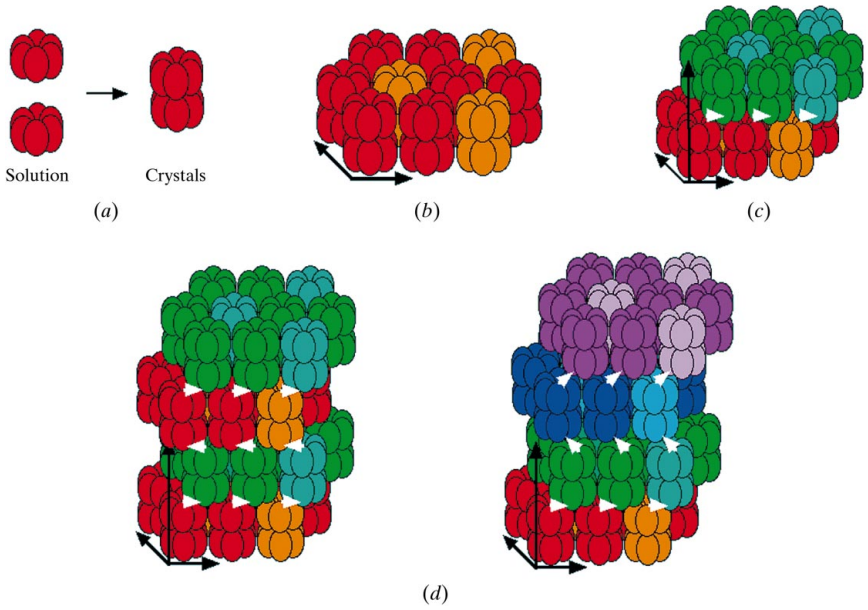
The diffraction pattern is characterized by the presence of the diffuse streaks along ***a****.

The structure was solved using demodulated data and experimental phasing

Refinement against corrected data: R=0.28

# Example 5: Partial disorder with several stacking vectors

Trame, C. B. & McKay, D. B. (2001).
*ActaCryst. D***57**, 1079–1090.




model of $P222_1$ single crystal

model of disordered crystal

Heat-shock locus U protein from *Haemophilus influenzae* and its complexes

Several crystal forms,
all partially disordered OD
belonging to different OD-families.

Data:

| | |
|---|---|
| Resolution | 2.3Å |
| Processed in | P622 |

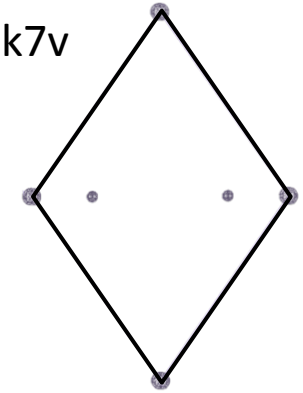
a = 110.6, c = 335.8

OD layer: P(6)22

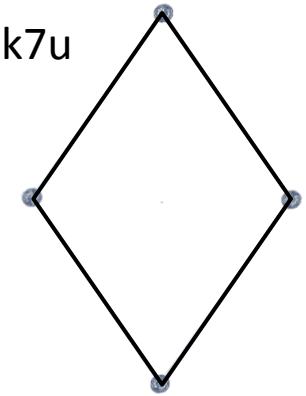
# Four types of domains

Patterson maps at W=0

1k7v

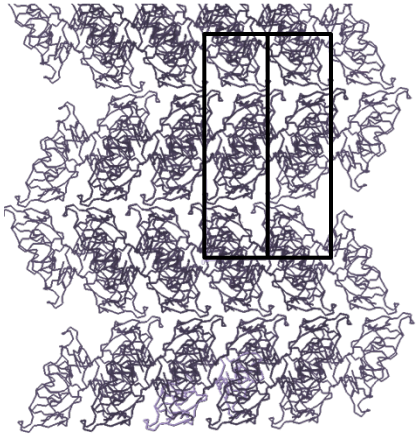


1k7u

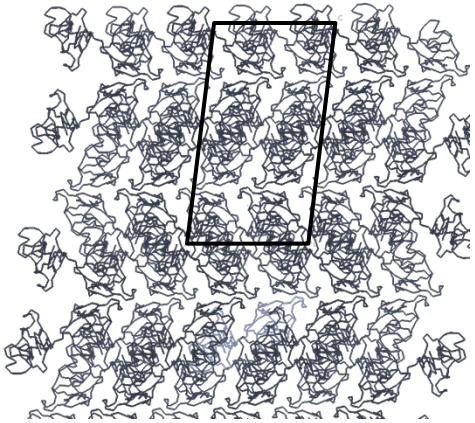


R / R-free
1k7u: 24.2 / 30.6
1k7v: 23.2 / 31.0

$P2_1$ structure (1k7u, 1k7v)



Putative C2 structure



Interpretation of the Patterson map for 1k7v: four types of domains

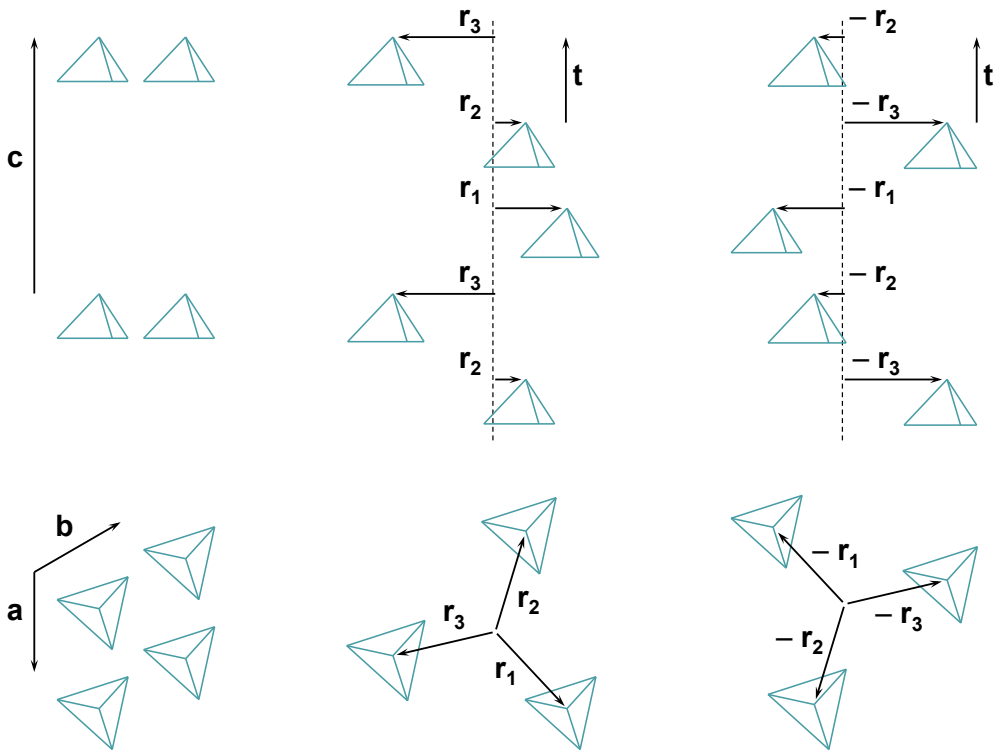
- $P2_1$ (orientation 1)
- $P2_1$ (orientation 2)

} Twinned $P2_1$ data

- C2 (orientation 1)
- C2 (orientation 2)

} contribute to some of the $P2_1$ spots, hence non-origin Patterson peaks

# Enantiomorphic stacking vectors




(1) (2)

Structures (1) and (2)

- belong to different space groups:

  (1) $P3_1$ (2) $P3_2$

- are not necessarily related by inversion
- but have the same structure amplitudes:

  $F(1) = F(2)$

- and belong to the same OD family

# Enantiomorphic stacking vectors

Gulbis et al. (1996). Structure of the C-terminal region of p21WAF1/CIP1 complexed with human PCNA. *Cell* **87**, 297–306.

Space group: $P3_221$
a = 83.5 Å, c = 233.9Å

OD layer: P(3)21

PDB code 1axc

| Structure: | from PDB | generated |
|---|---|---|
| Spacegroup: | $P3_221$ | $P3_121$ |
| R (%): | 22.09 | 22.35 |
| R-free (%): | 29.15 | 30.02 |

Asymmetry of OD layer is within 0.2Å, but it helps choosing the right space group

OD-structures

Twinning by (pseudo)merohedry

Statistics of one observation

Statistics of two observations

Twinning tests summary

Space group validation

# Twinning by (pseudo)merohedry

Twins by reticular merohedry (inc some OD-twins), allotwins, disordered structures

- Can be readily seen in images with predictions

---

Important special case: twinning by (pseudo)merohedry

- All spots overlap with related spots from another individual crystal

- Detection requires analysis of intensity statistics

- More significant effect on model if ignored

- Space group determination may be a problem

# Monoclinic OD-twin (twin by pseudomerohedry)

Au et al. (2006).
*Acta Cryst*. D**62**, 1267-1275.

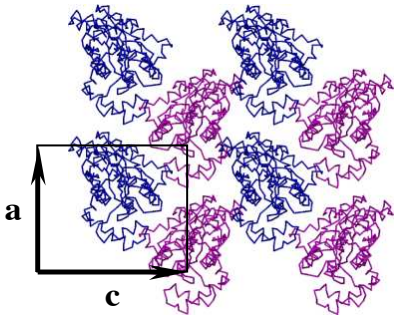
Ferrochelatase-1 from *B. anthracis*

PDB code 2c8j

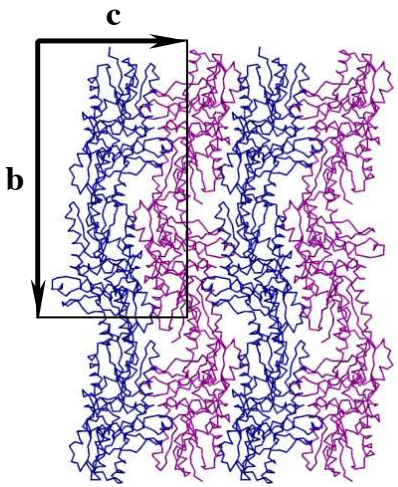
Space group: $P2_1$
Resolution 2.2Å

a = 49.9, b = 109.9, c = 59.4 Å
α = β = γ = 90°

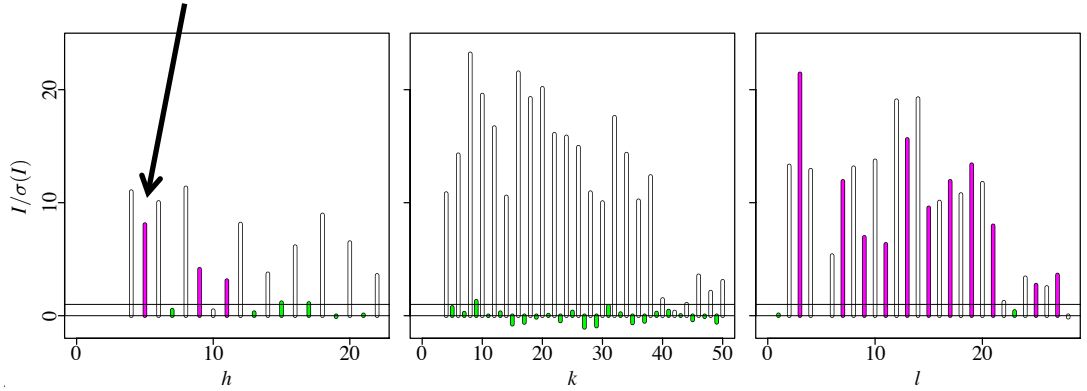
OD layer: P2(1)1




(*a*)



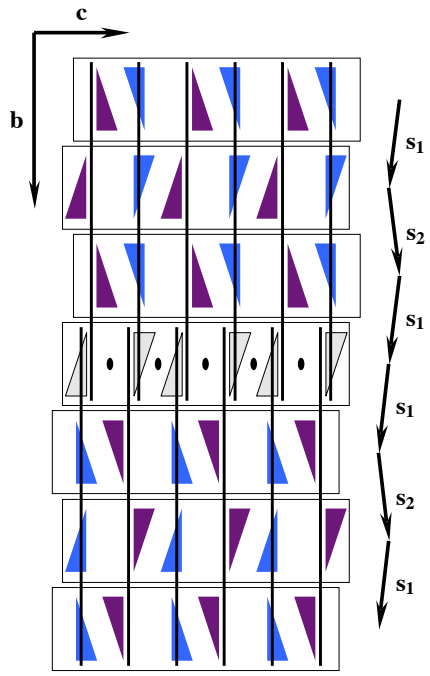

The only reflection with h = 2n, k=l=0 and I > 3 sig(I)



# Monoclinic OD-twin (twin by pseudomerohedry)

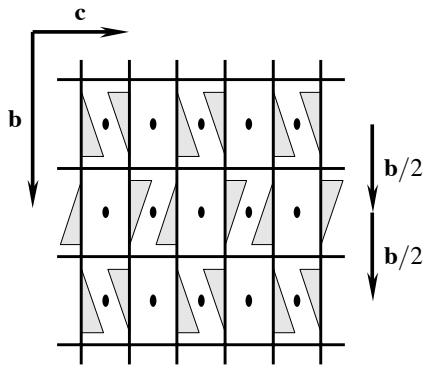
$P2_1$ true structure

The lattice is exactly orthorhombic
(twin by metric merohedry)




$P2_12_12$ reference
fully ordered structure

molecules shifted along **c** by 2.5Å




Twinning was suspected only after several unsuccessful attempts at solving structure in an orthorhombic space group

# Tutorial

Ferrochelatase-1 Tutorial:

Space group assignment in the presence of pseudosymmetry and twinning

Data:

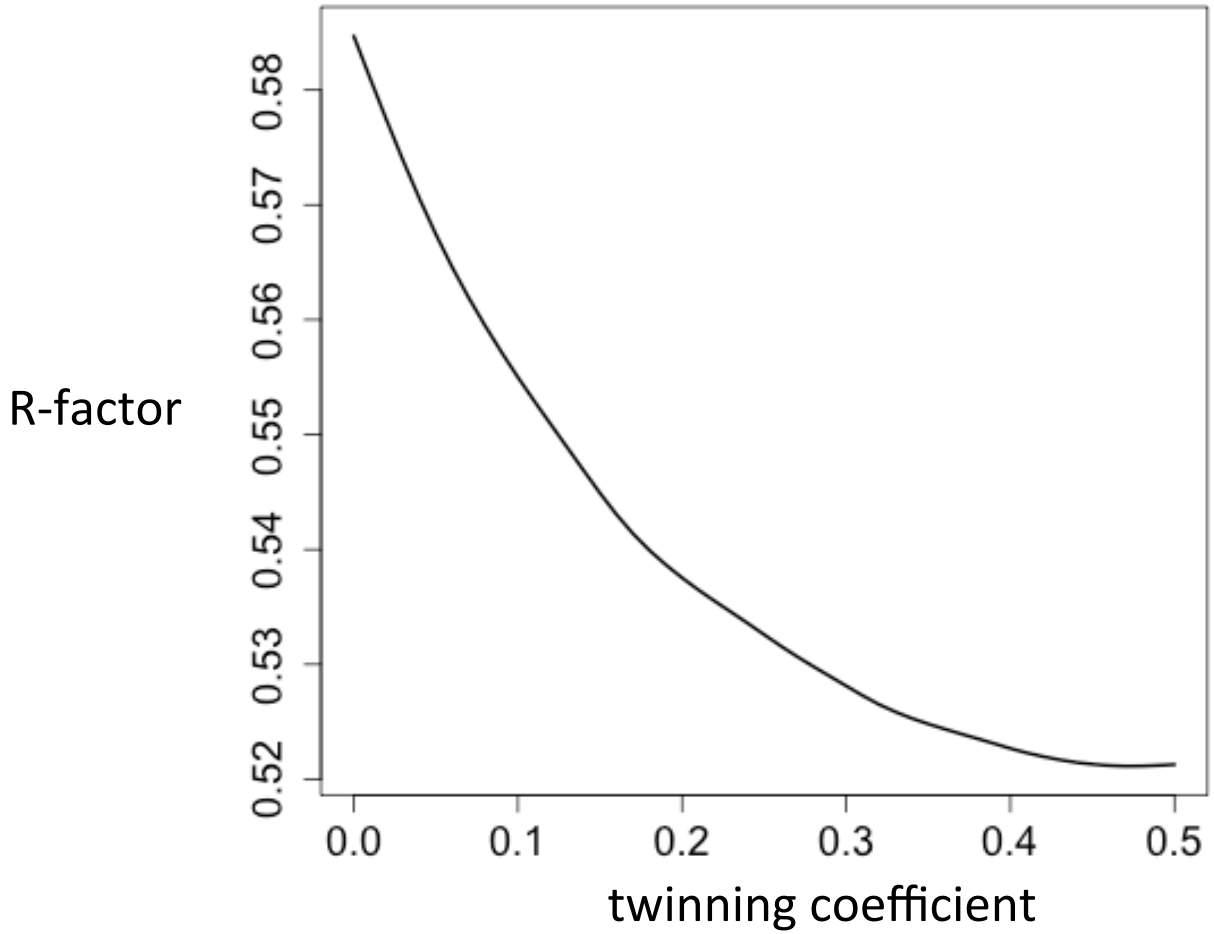
**http://www.ysbl.york.ac.uk/mxstat/andrey/hemh.html**

- OD-twin by pseudomerohedry
- use of pointless for point group detemination in a relatively difficult case
- use of molecular replacement

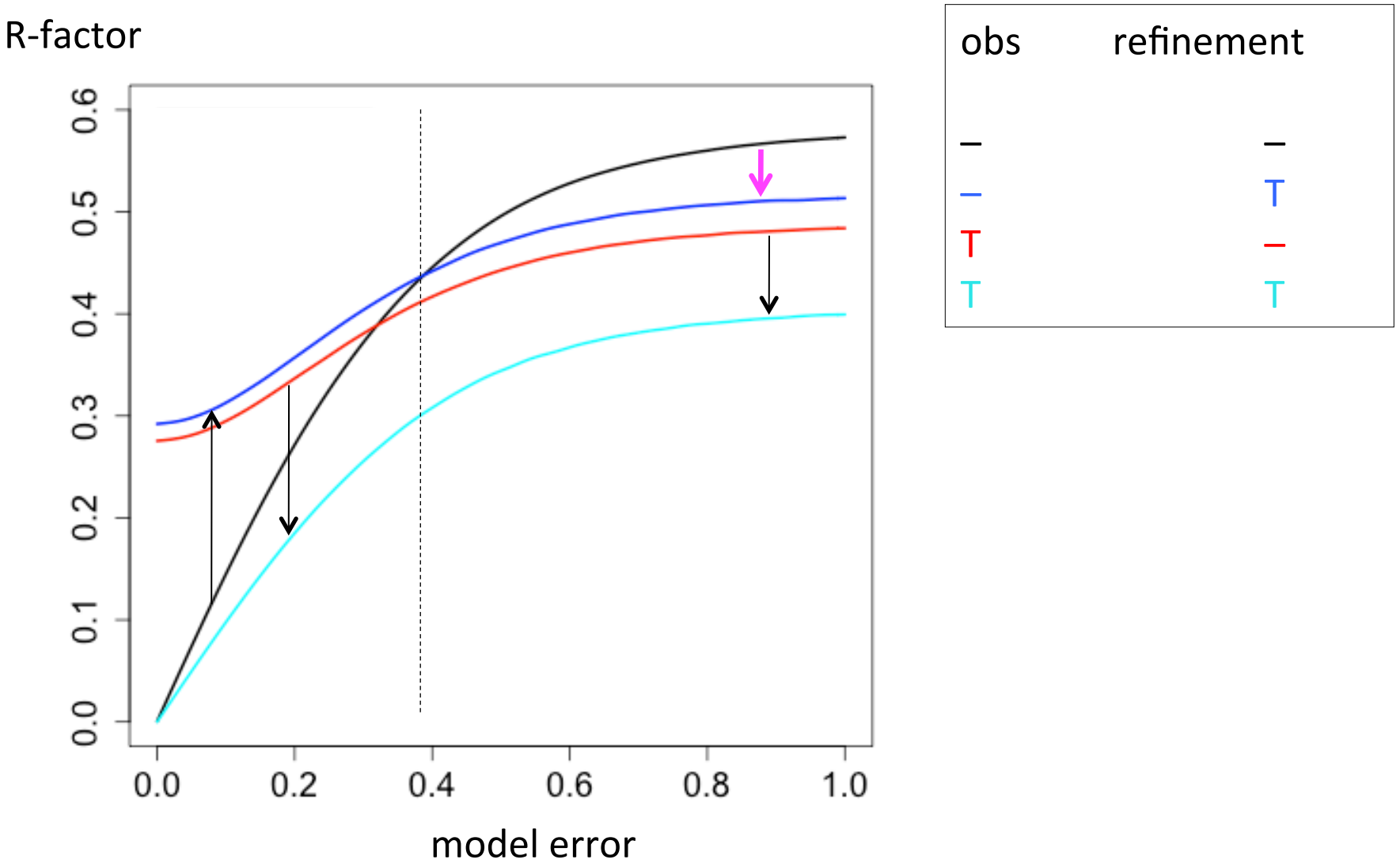
# Twinned refinement against non-twinned data

Beginning of refinement:

- Two unrelated structures, one is twinned
- Twinning coefficient would converge to 0.5



# Switching to twin refinement



Examples of crystal pathologies

Twinning by (pseudo)merohedry

Statistics of one observation

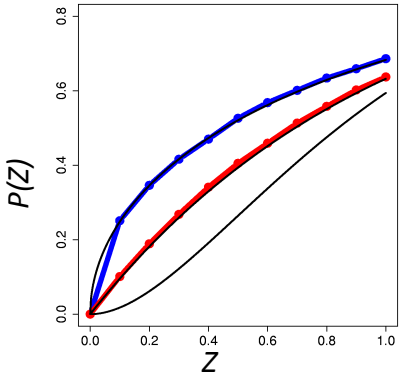
Statistics of two observations

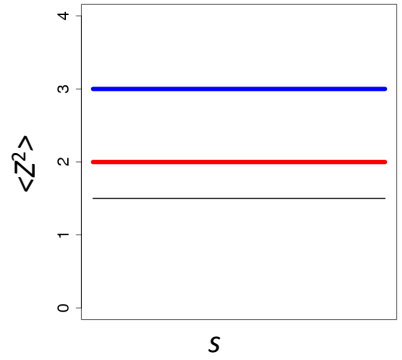
Twinning tests summary

Space group validation

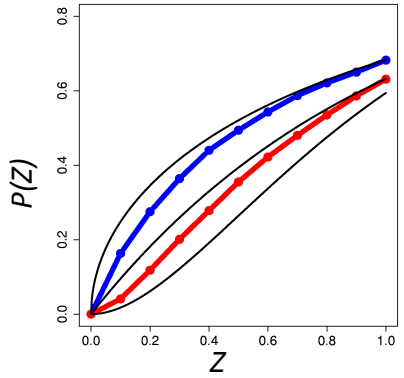
# Theoretical distribution of intensities

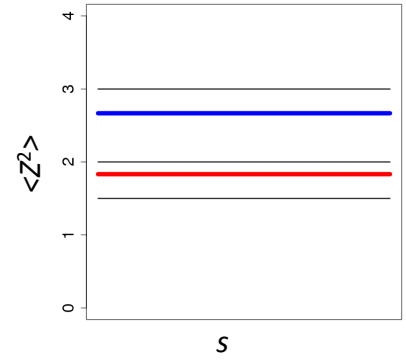
Single crystal



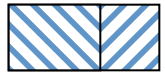



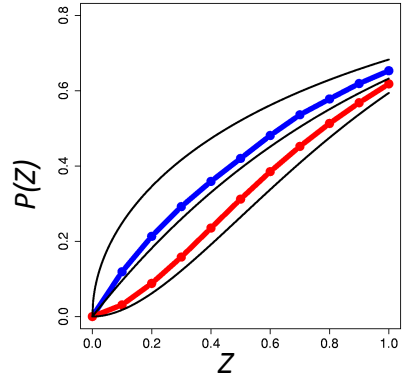

Partial twin

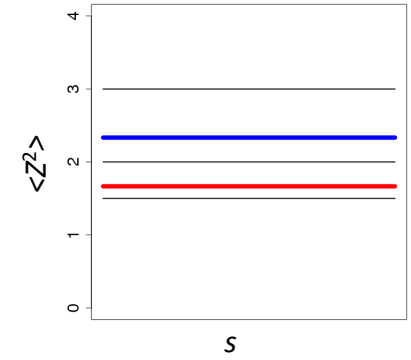


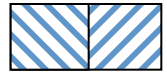

Partial twin

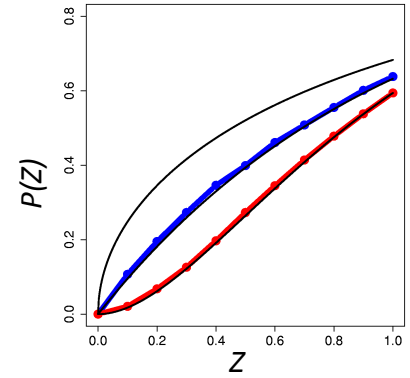


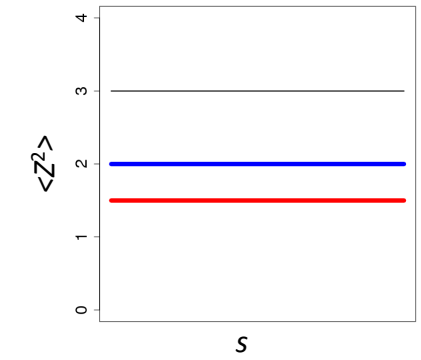



Perfect twin



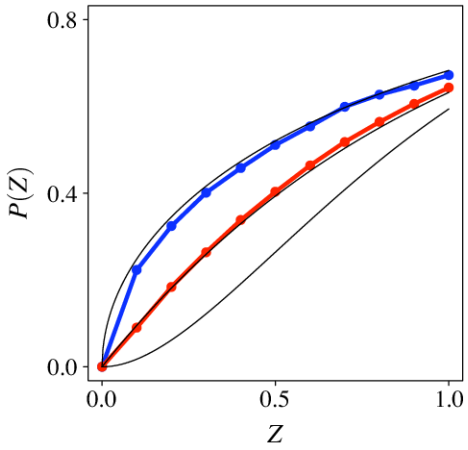


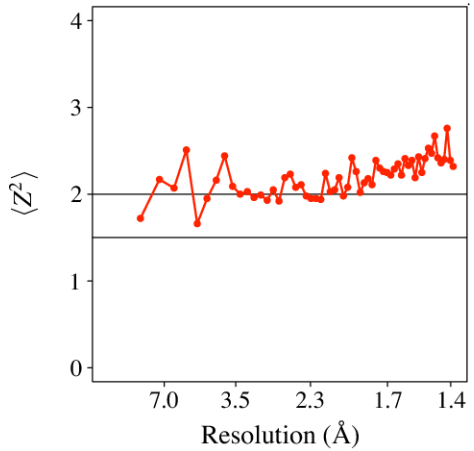




Acentric reflections
Centric reflections

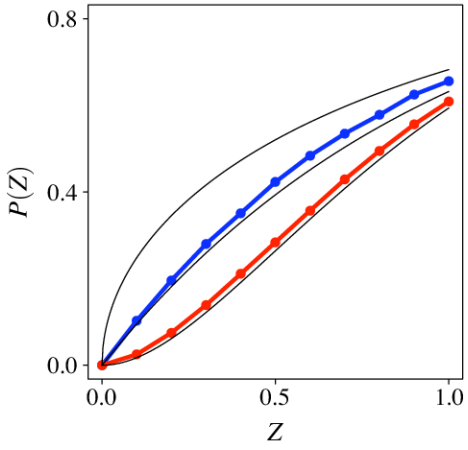
(works for incomplete data set)

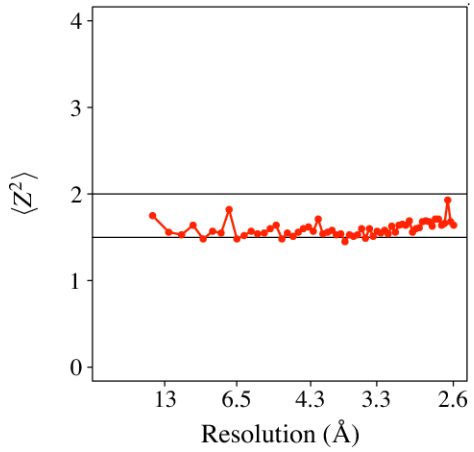
# Two good, two bad



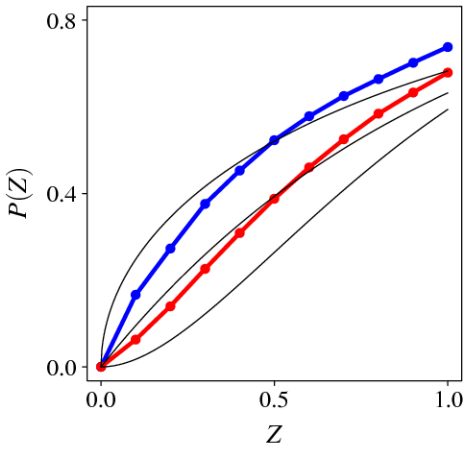

PDB entry 1i1j
single crystal



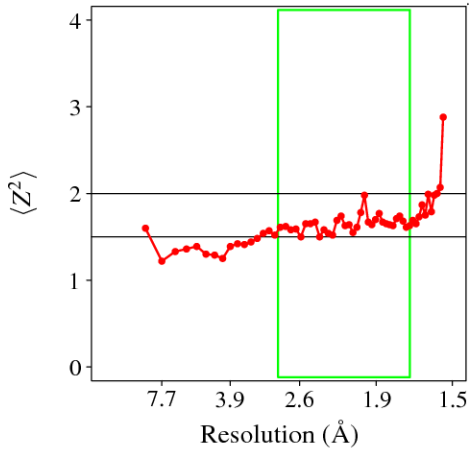




C-terminal domain of gp2 protein from phage SPP1
perfect twin

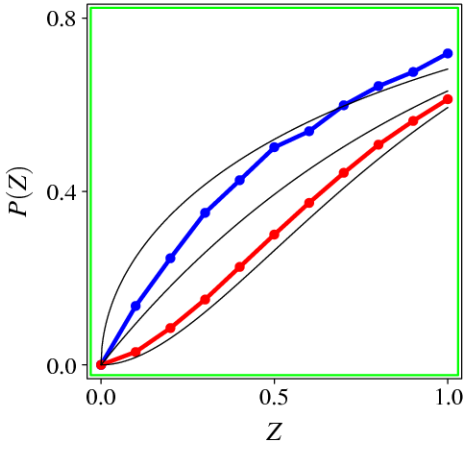
# Bad example 1



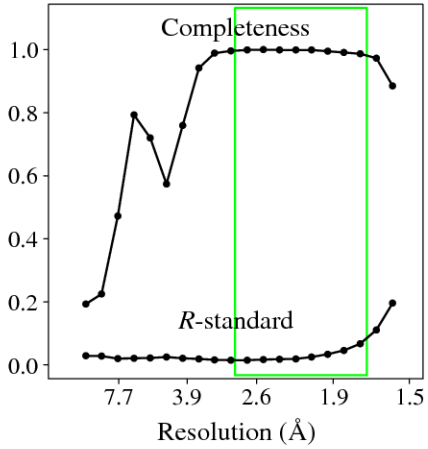

(a)




(b)




(c)

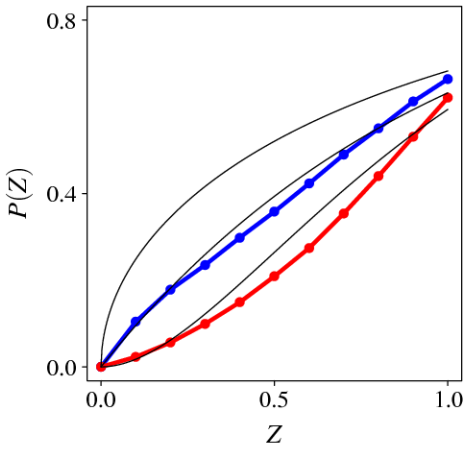



(d)

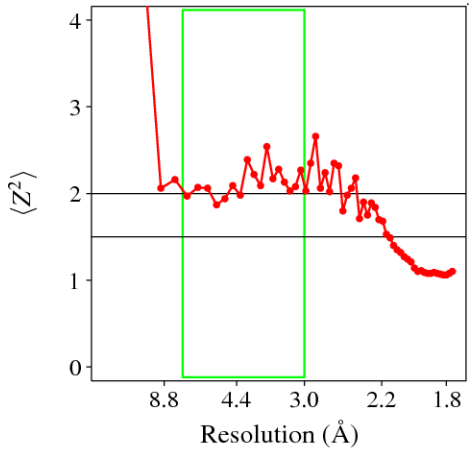
PDB code 1l2h

partial twin

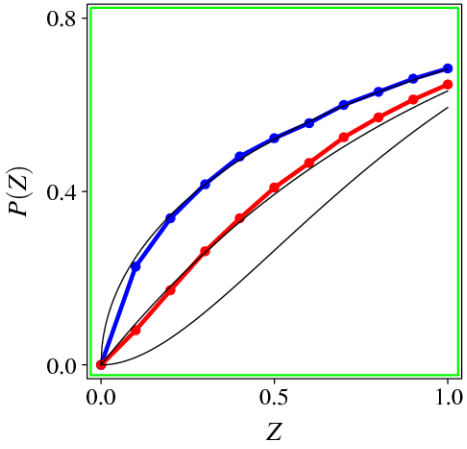
# Bad example 2



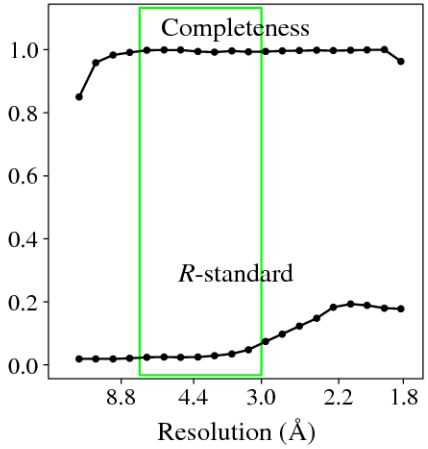

(a)




(b)



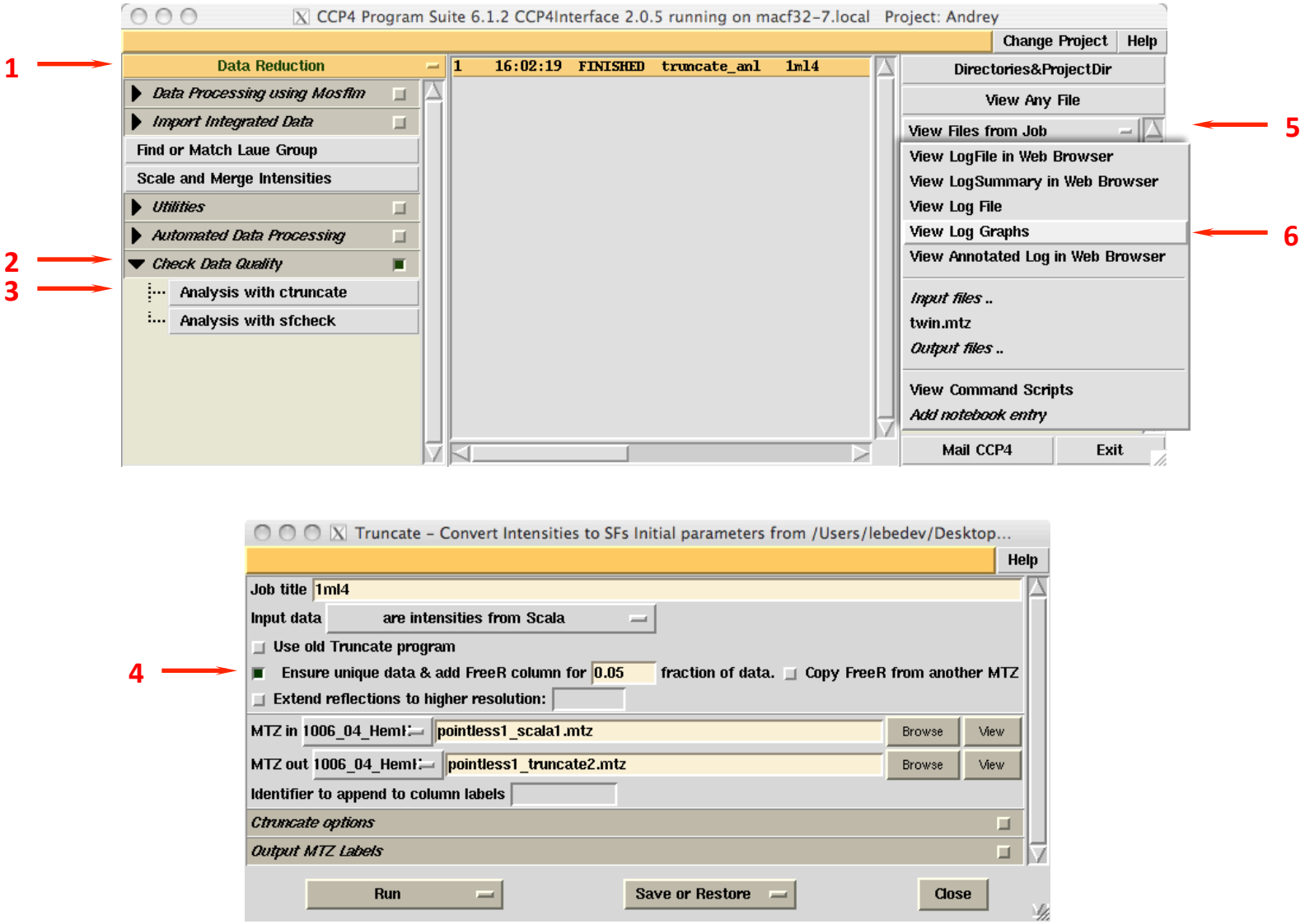

(c)




(d)

human deoxycytidine kinase single crystal

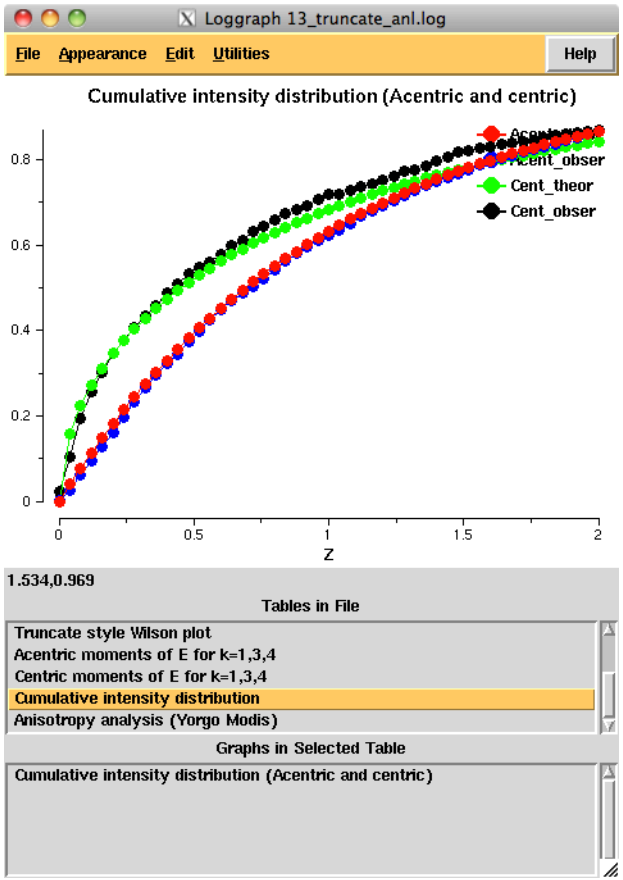
# Twinning tests in CCP4I (ctruncate)



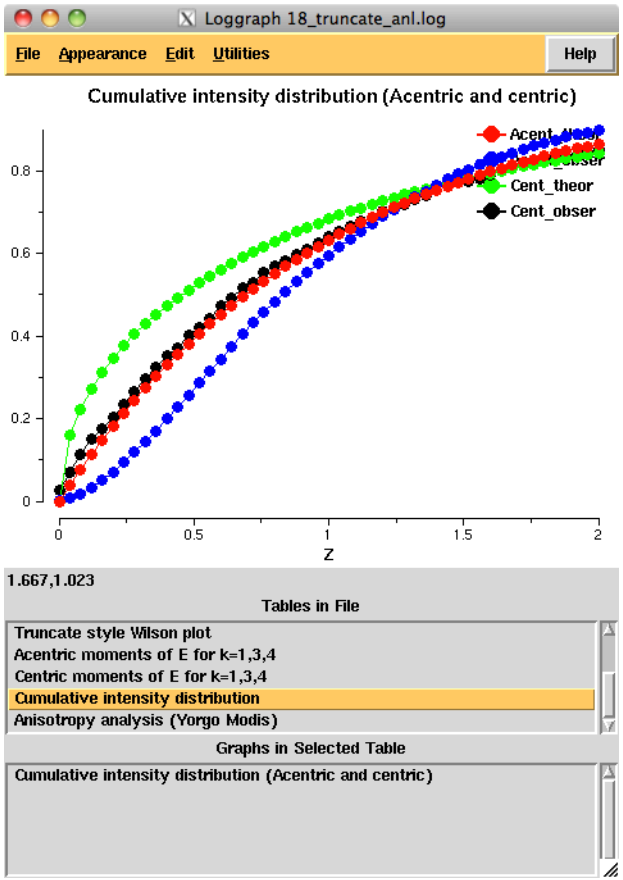
# Cumulative intensity distribution

To compare: Red: Acentric theoretical, Blue: Acentric observed $Z \approx |E|^2$

## Untwinned data




## Twinned data

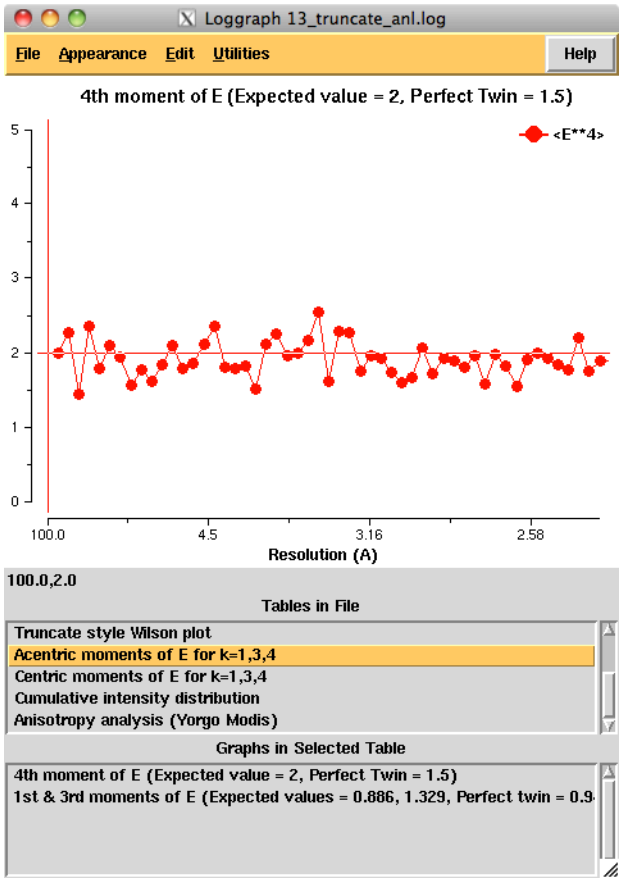



> Cumulative intensity distribution
> Cumulative ... (Centric and acentric)

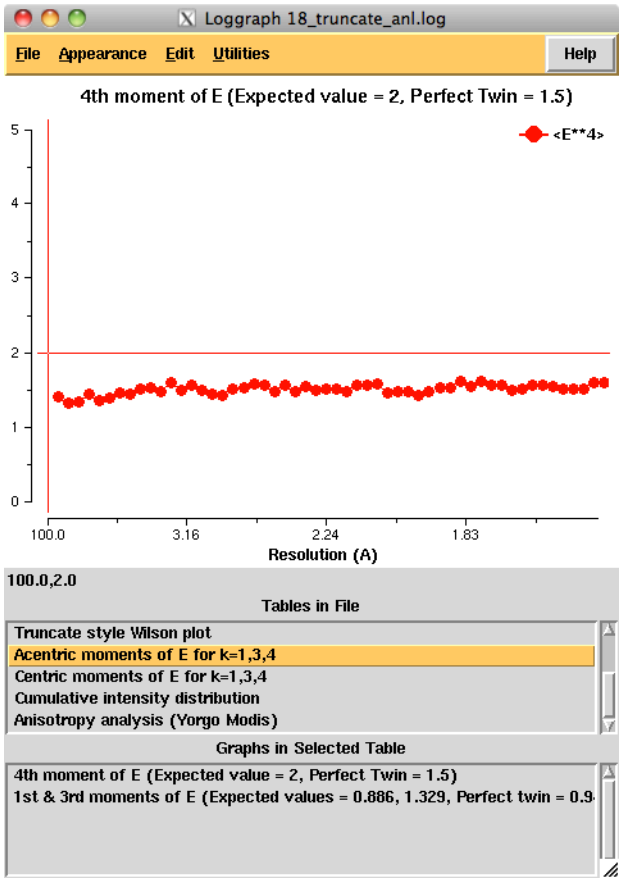
# Second moments of Z (fourth moments of |E|)

Compare the experimental curve with the line $\langle E^4 \rangle = 2$

## Untwinned data




## Twinned data




> Acentric moments of E for k=1,3,4

> 4th moments of E ...

OD-structures

Twinning by (pseudo)merohedry

Statistics of one observation

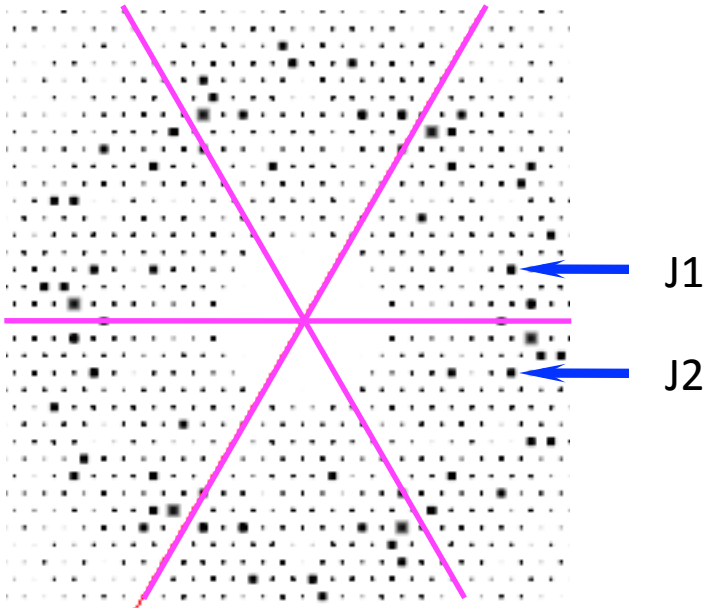
Statistics of two observations

Twinning tests summary

Space group validation

# H-test and L-test

$H = | J1 - J2 | / ( J1 + J2 )$



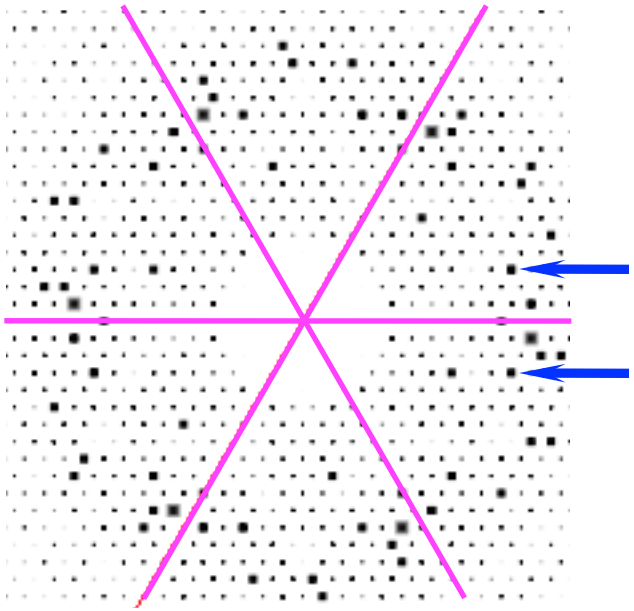

twin axes

$L = | J1 - J2 | / ( J1 + J2 )$

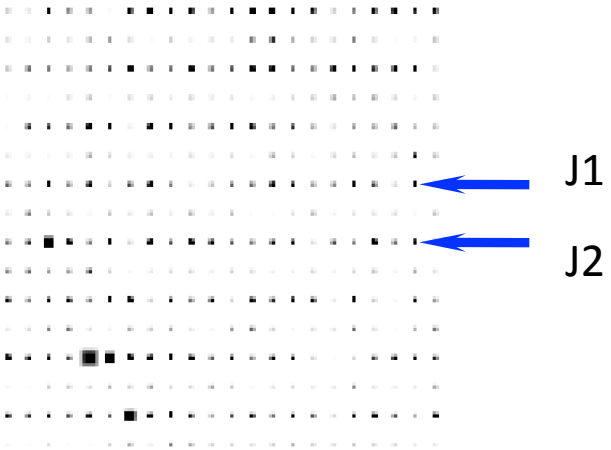

sublattices with strong and weak reflections (pseudotranslation)

# H-test and L-test

$H = | J1 - J2 | / ( J1 + J2 )$



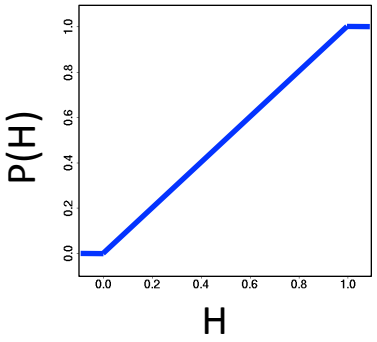

$L = | J1 - J2 | / ( J1 + J2 )$



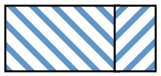

sublattices with strong and weak reflections (pseudotranslation)

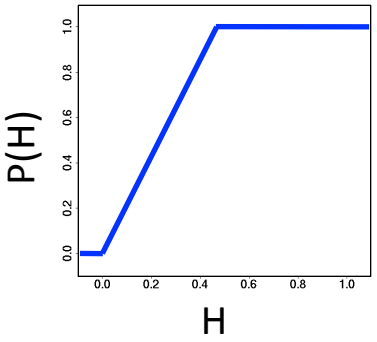
# Theoretical distribution of H

Single crystal

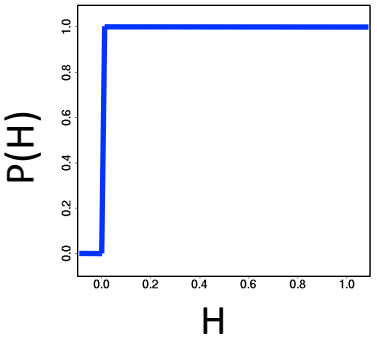



Partial twin

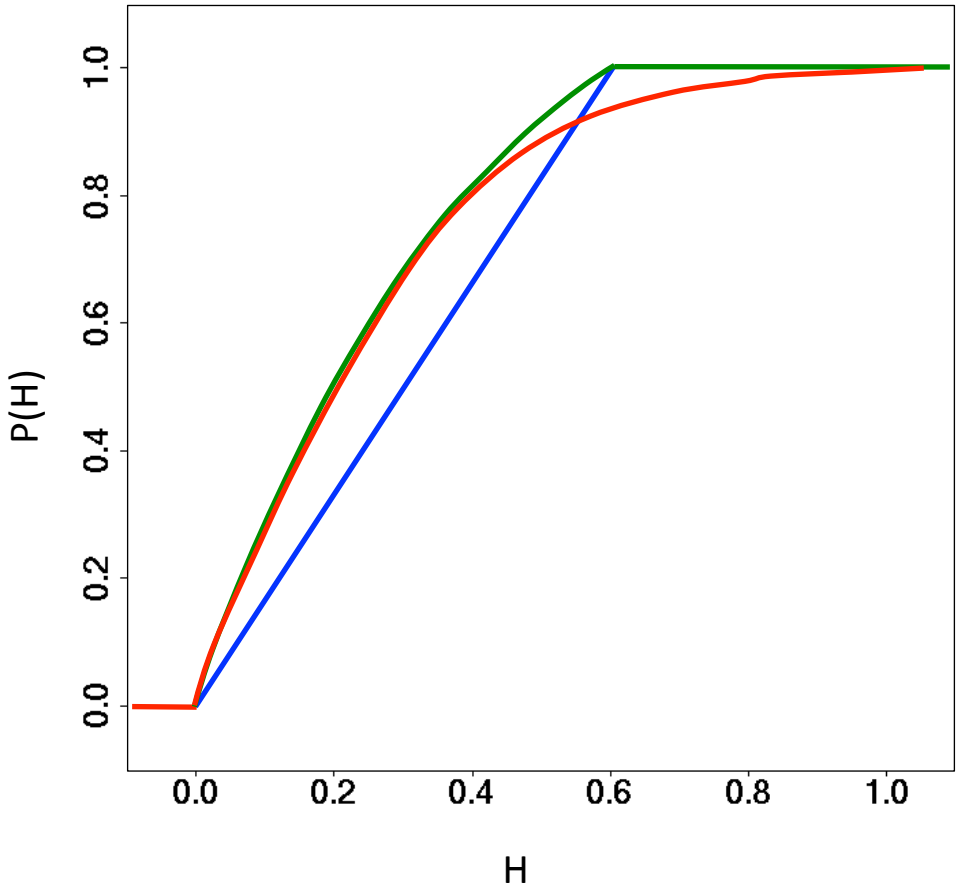





Perfect twin



# Distribution of H can be perturbed by NCS and weak observations

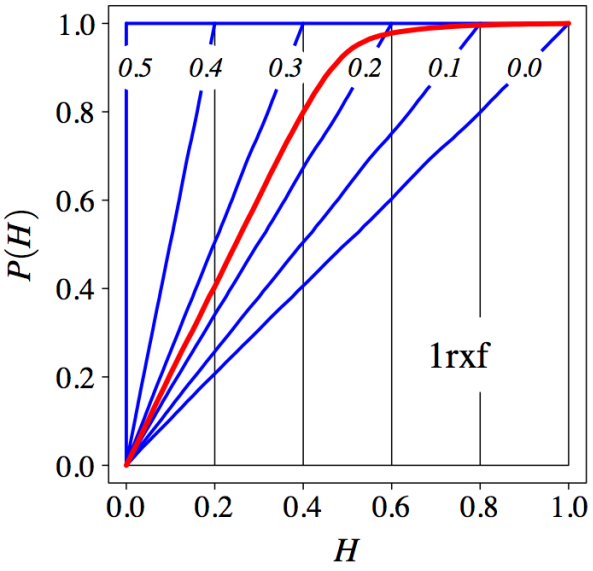


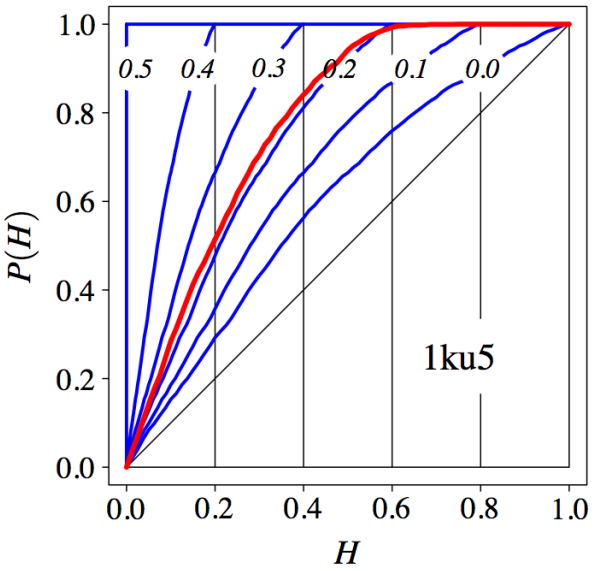

Blue:
ideal distribution for partial twin

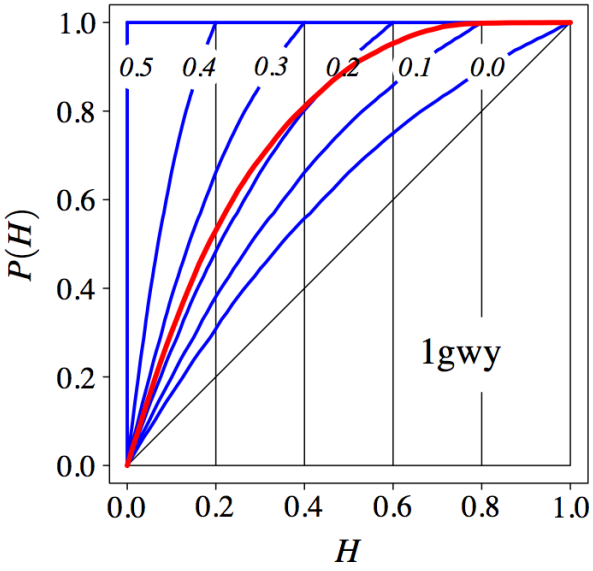
Green:
blue + effect of NCS axis || twin axis

Red:
green + effect of intensities with small I/ sig(I)

# Examples of experimental P(H)



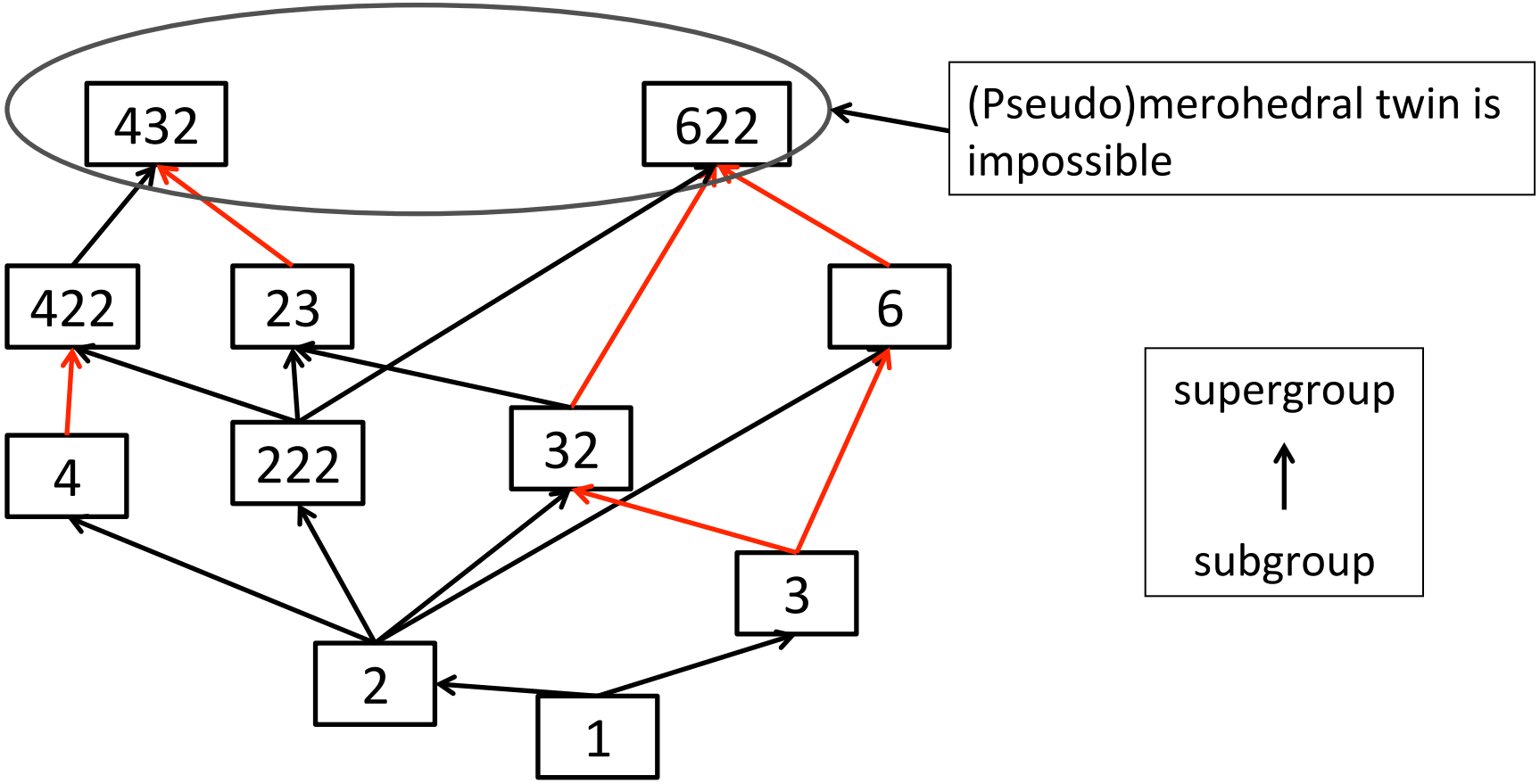




An almost ideal case

+ effect of
NCS axis || twin axis

+ effect of
intensities with
small I/ sig(I)

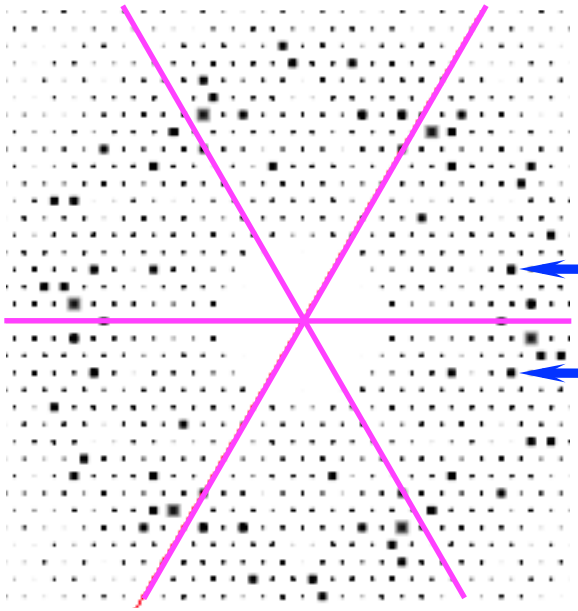
# Relations between point groups




Red arrows: No constraints are needed, merohedral twin could happen
Black arrows: Additional constraints on cell parameters are needed, psedo merohedral twinning can happen

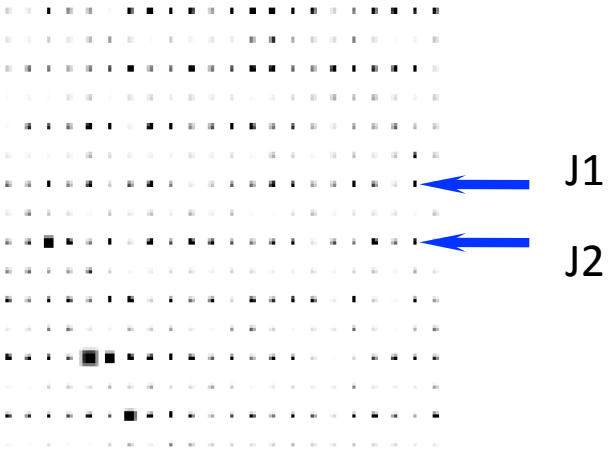
# H-test and L-test

$H = | J1 - J2 | / ( J1 + J2 )$




twin axes

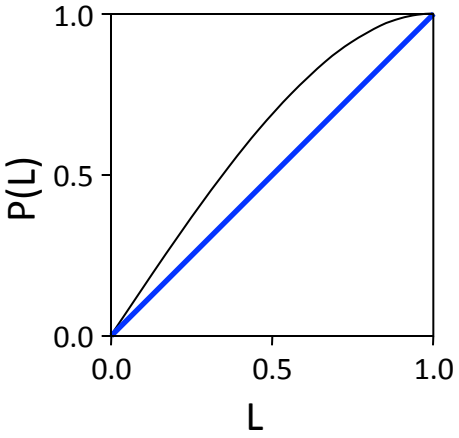
$L = | J1 - J2 | / ( J1 + J2 )$



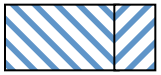

sublattices with strong and weak reflections (pseudotranslation)

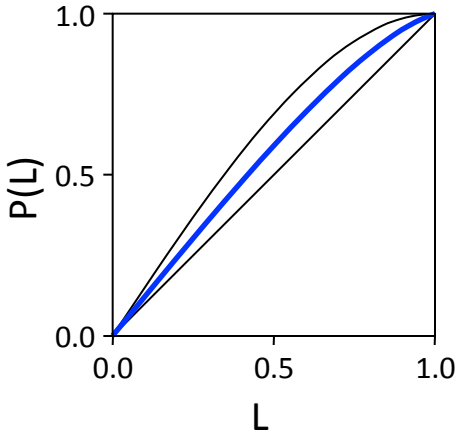
# Theoretical distribution of L

Single crystal

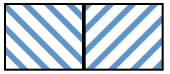


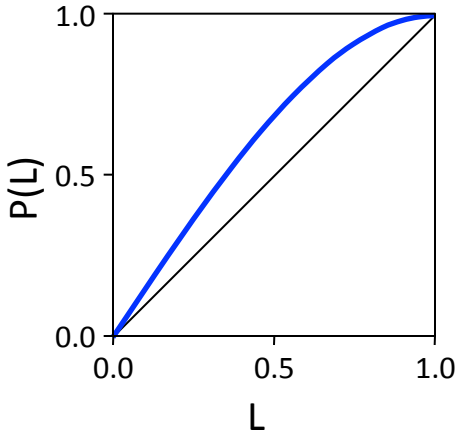

Partial twin





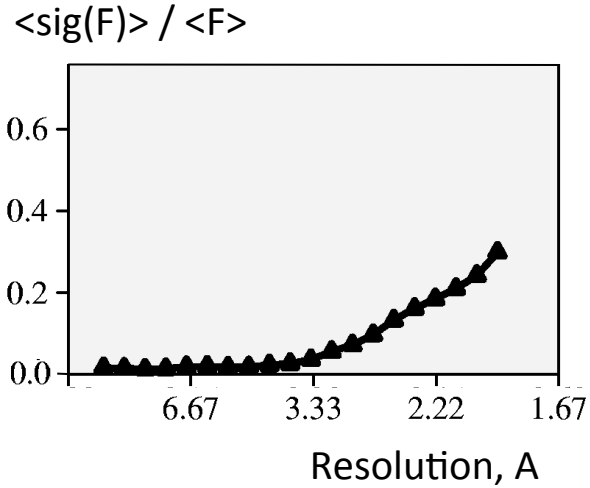

Perfect twin



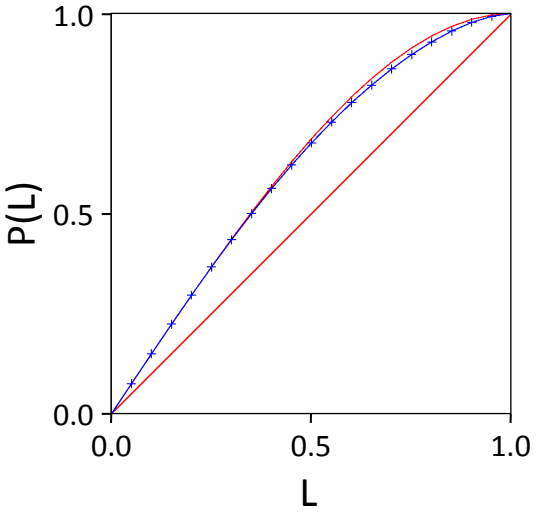


# Distribution of L can be strongly perturbed by weak observations

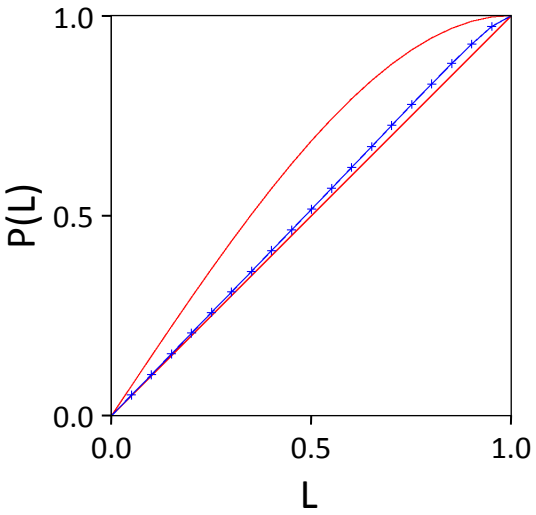
Cell: 64.2 109.2 100.2 90 93.8 90
Space group: $P2_1$
No pseudo symmetry
Pseudomerohedral twinning is impossible




All data:
as if a perfect twin



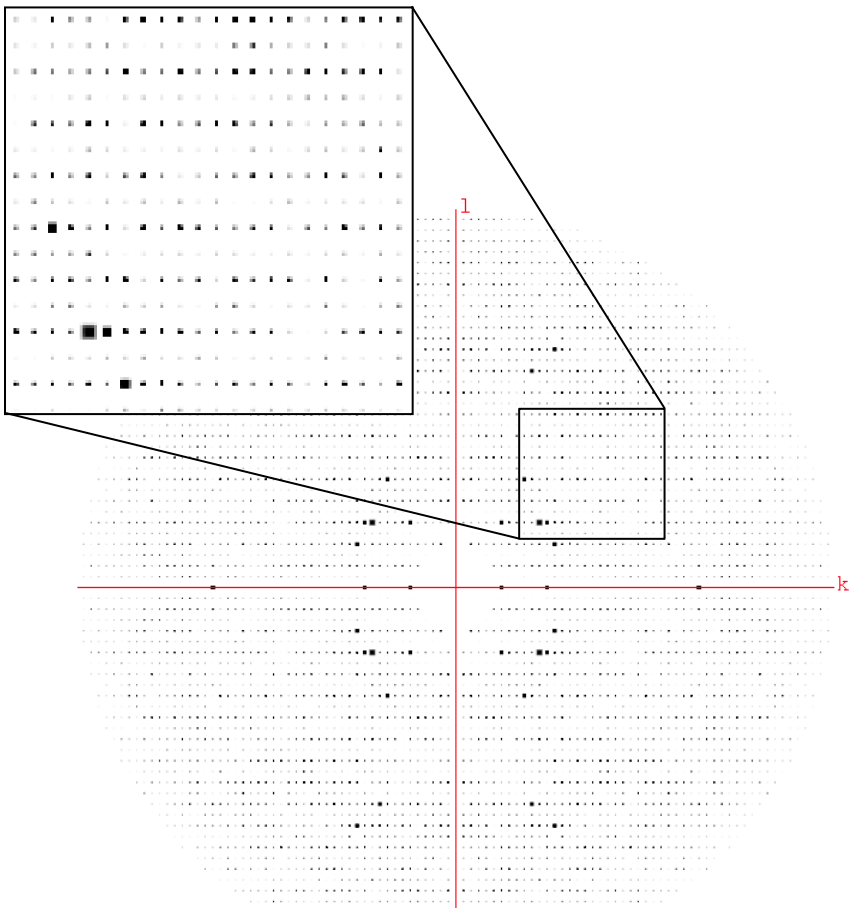

Data below 3A:
untwinned

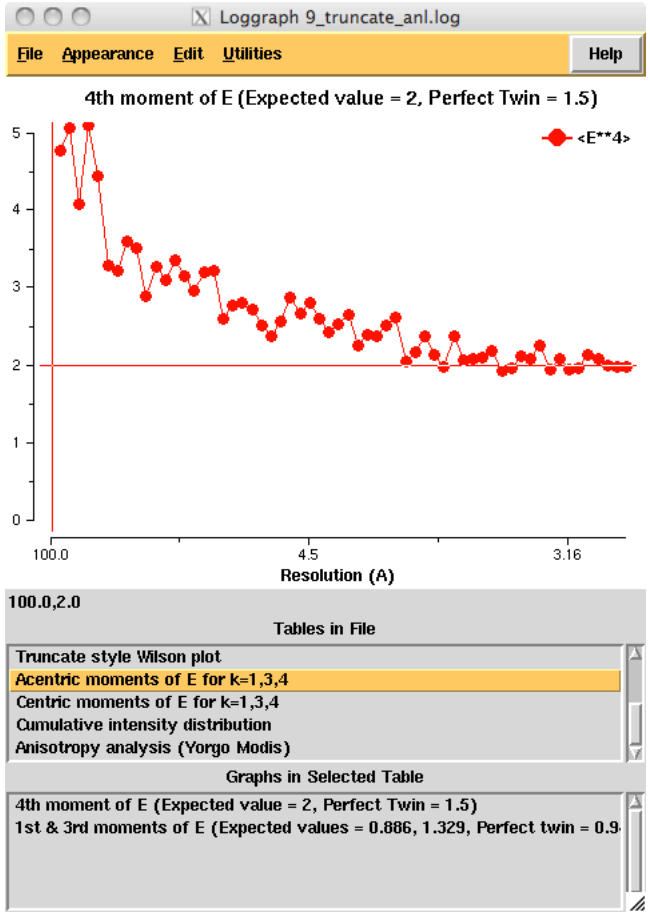



Nevertheless the L-test is very useful when performed with right resolution range (or with several ranges)

# Statistics of one intensity are strongly affected by pseudotranslation

1jjk: Pseudotranslation results in alteration of strong and weak reflections

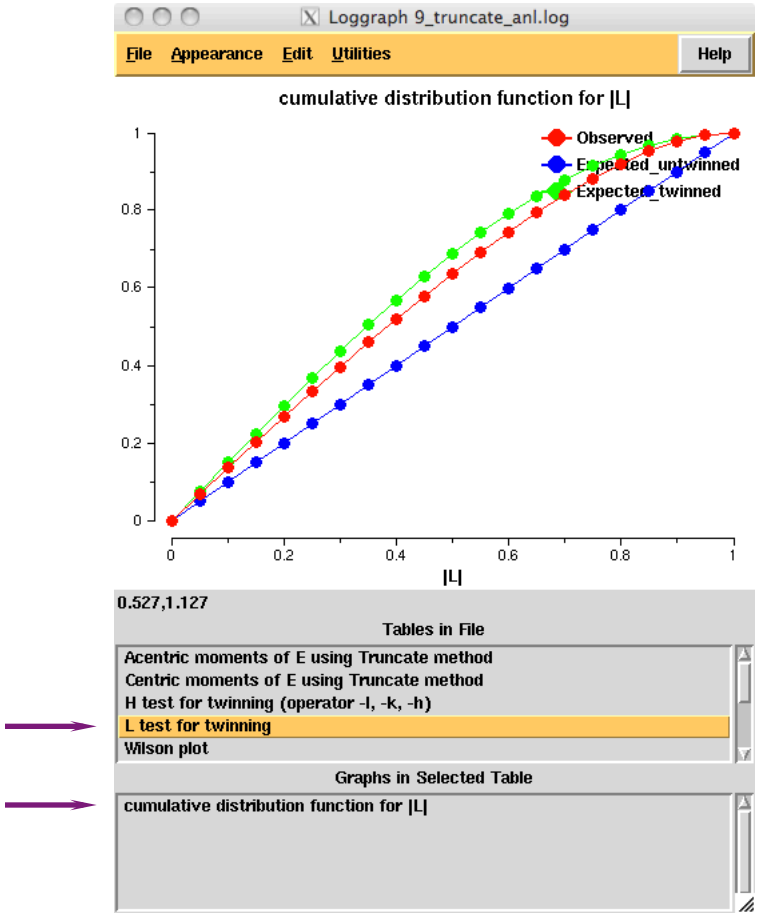


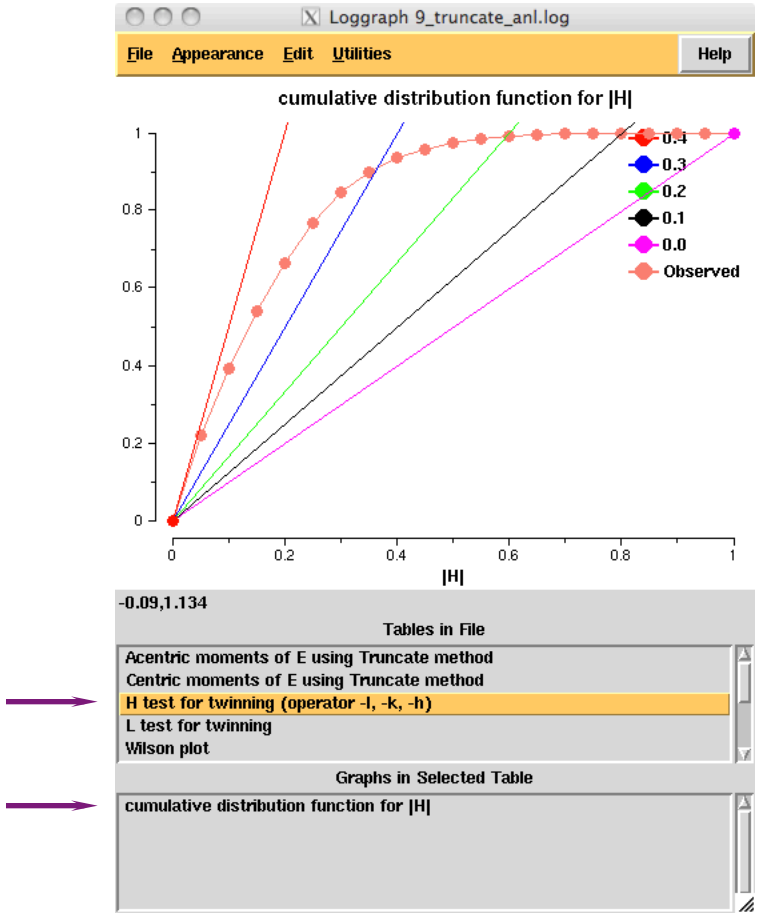




> Acentric moments of E for k=1,3,4

> 4th moments of E ...

# L-test and H-test are not affected by pseudotranslation




> L test for twinning
> cumulative distribution function for |L|

> H test for twinning (operator ...)
> cumulative distribution function for |H|

OD-structures

Twinning by (pseudo)merohedry

Statistics of one observation

Statistics of two observations

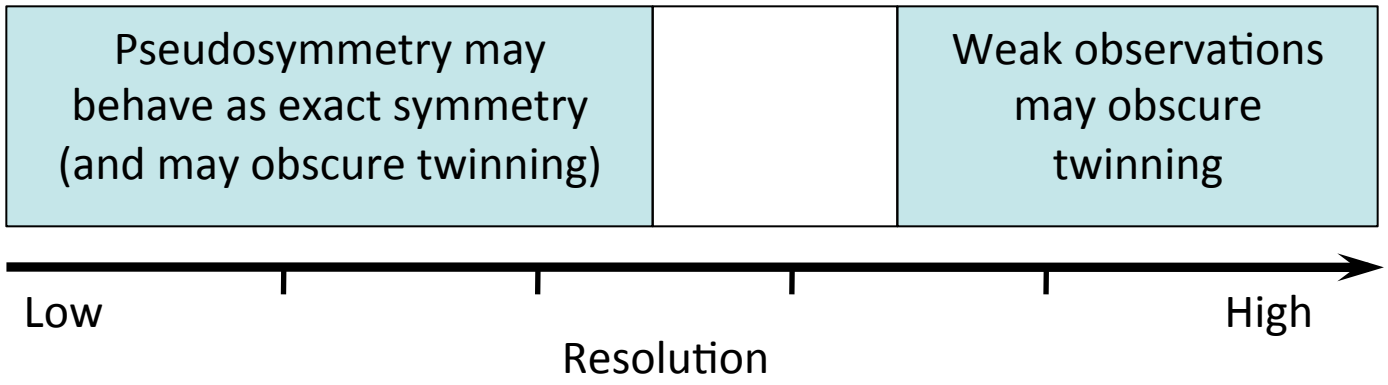
Twinning tests summary

Space group validation

# Why so many tests?

| | Statistics of one observation | | Statistics of two observations | |
|---|---|---|---|---|
| | P(Z) | <Z^2> | H-test | L-test |
| Specific for a given resolution shell | No | Yes | No | No |
| Specific for a given twin operation | No | No | Yes | No |
| Can detect perfect twinning | + | + | – | + |
| Works for incomplete data | + | + | – | + |
| Insensitive to pseudotranslation | – | – | + / – | + |
| Insensitive to anisotropy | – | – | + / – | + |
| Insensitive to weak reflections at high resolution | – | (–) | – | – |

# Are these tests always sufficient?




How to handle the cases with strong pseudosymmetry?

Validation of crystallographic symmetry instead of twinning tests:

refinement in space groups compatible with

- unit cell
- current model (considered as at least approximately correct)

OD-structures

Twinning by (pseudo)merohedry

Statistics of one observation

Statistics of two observations

Twinning tests summary

Space group validation

# An example of symmetry correction

| | | |
|---|---|---|
| PDB code: | 1yup | |
| space group (PDB): | P1 | 8 molecules per a.u. |
| space group (true): | $P2_1$ | 4 molecules per a.u. |
| Pseudo-symmetry space group:<br>(because of pseudo-translation) | C2 | 2 molecules per a.u. |

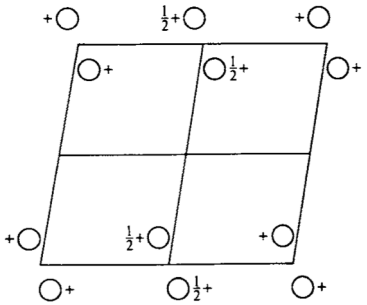
# Monoclinic structures related to 1yup

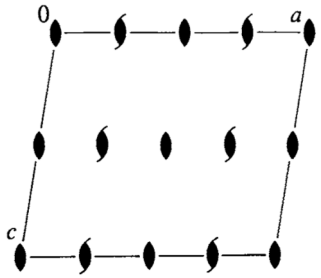
Positions of molecules

Crystallographic axes

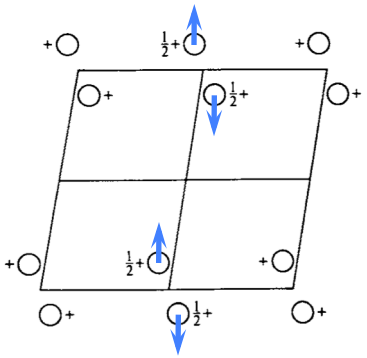
NCS axes

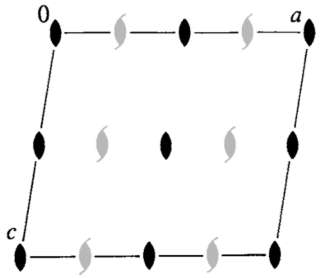
Space group and its relation to the structure 1yup



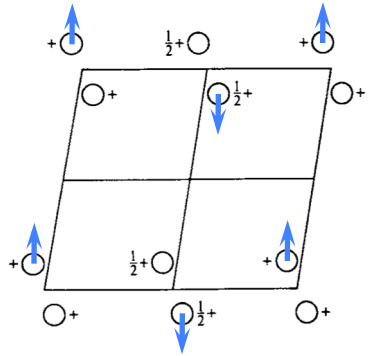


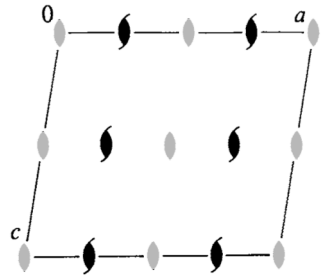

C2 Pseudo-symmetry space group





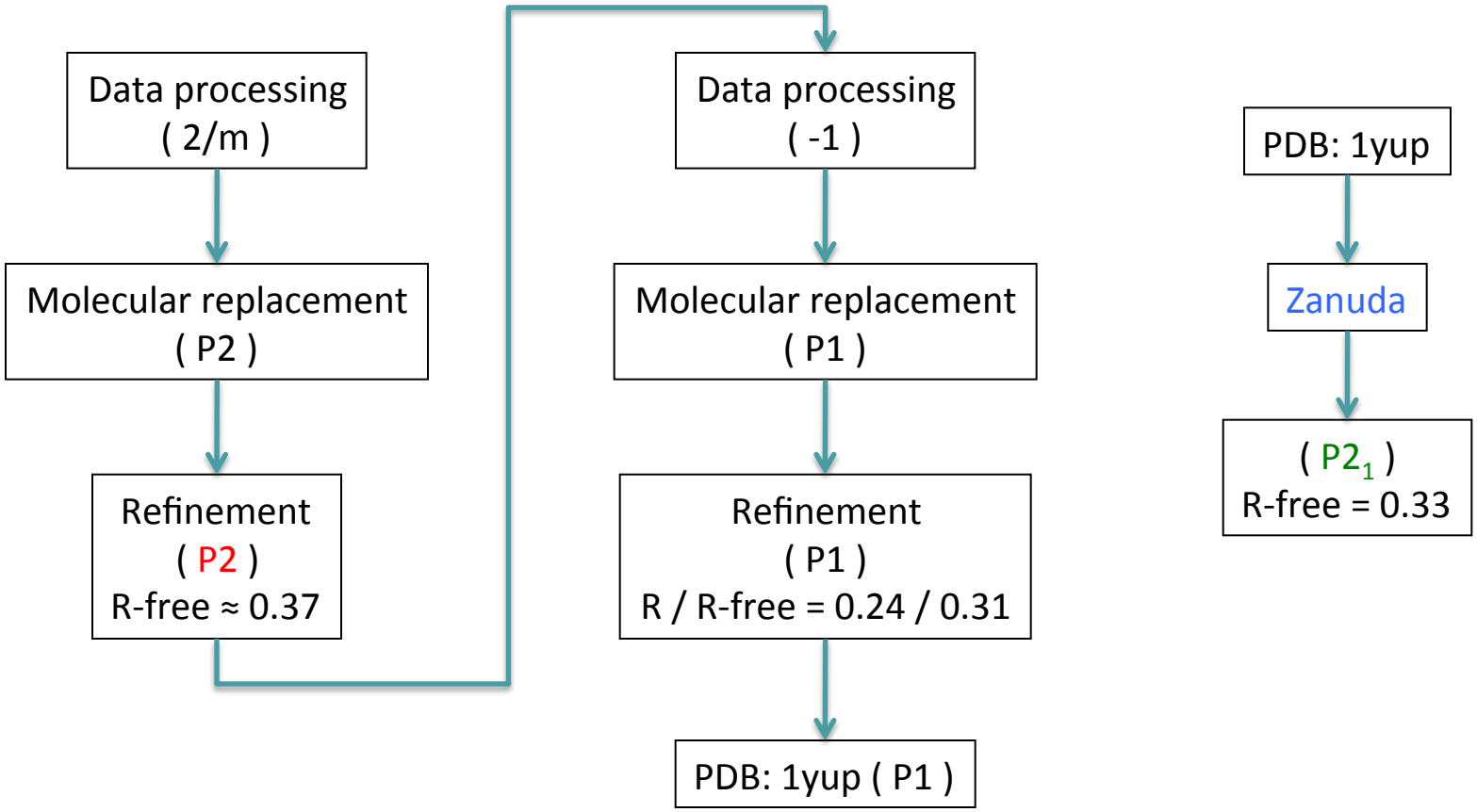

P2 False space group






$P2_1$ True space group

# Structure solution and symmetry validation



# Zanuda: space group validation

Algorithm:

- From input model: determine pseudosymmetry space group (PSSG)
- From PSSG: select subgroups with observed unit cell
- For each such subgroup:
  - Convert model and data into the subgroup
  - Restrained refinement
- Repeat refinements of the best (R-free) model
  - Starting from P1
  - Adding the best (r.m.s.d.) symmetry element at each refinement
    - » Terminate if there is no symmetry elements to be added
    - » Terminate and cancel the last symmetry element if R-free jumps

# Zanuda: limitations

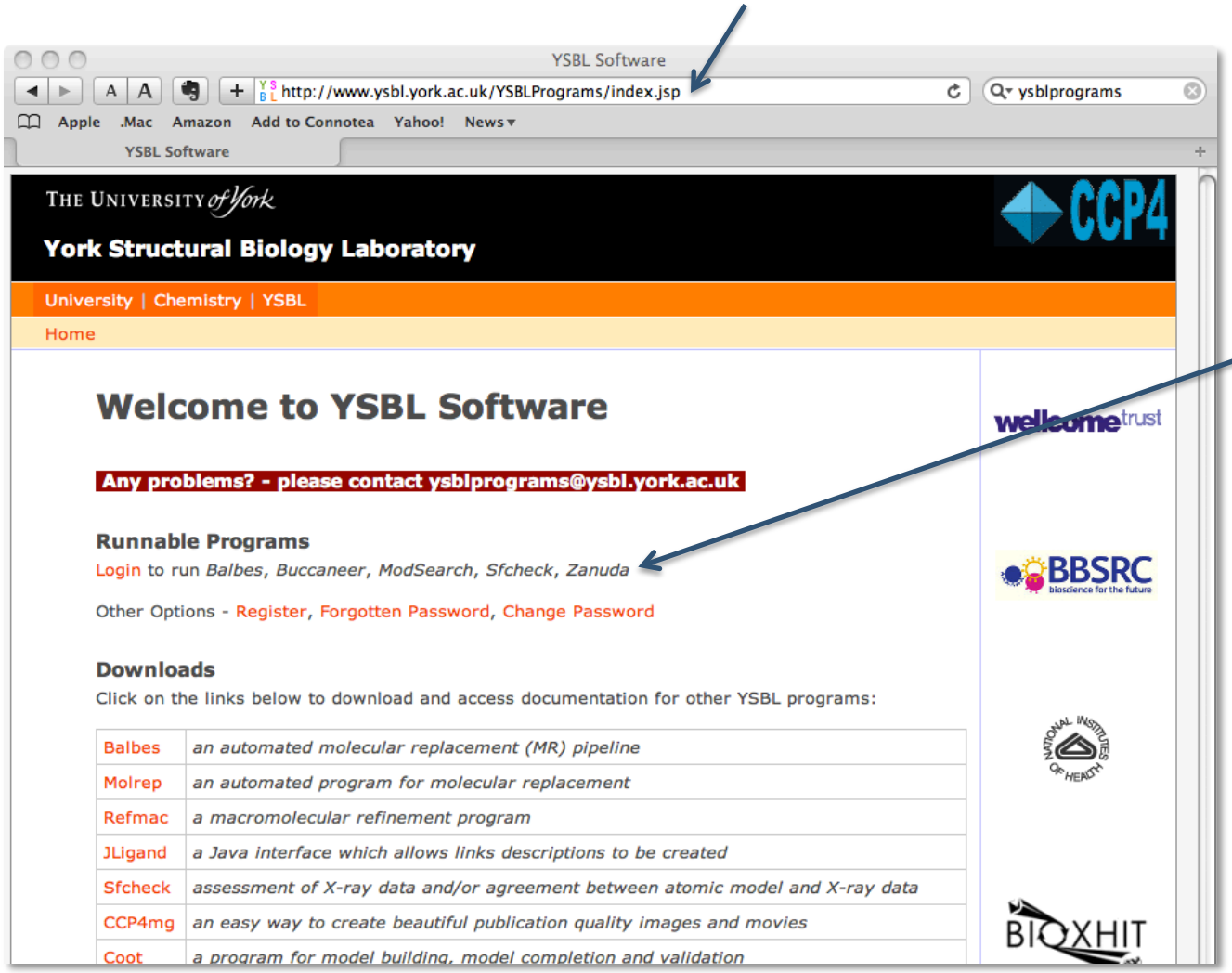
Assumptions:

- The pseudosymmetry is very strong (r.m.s.d. from exact symmetry ≈ 1A)
- The structure is almost correct
  - although it might have been refined / rebuilt in an incorrect space group

If assumptions are not satisfied, the results will likely to be wrong.

# YSBL server

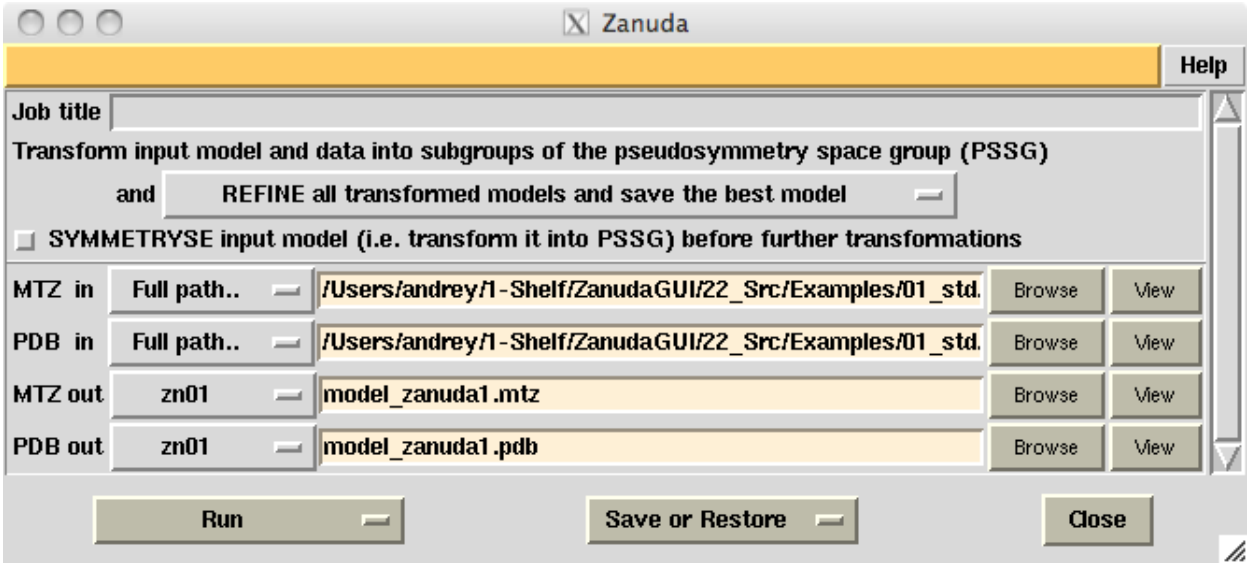
http://www.ysbl.york.ac.uk/YSBLPrograms/index.jsp




Zanuda

# CCP4I interface

CCP4I > Validation & Deposition > Validate space group




Starting from ccp4-6.3.0 (forthcoming release)

# Acknowledgements


| | |
|---|---|
| Eleanor Dodson | University of York |
| Roberto Steiner | King's College, London |
| Zbigniew Dauter | APS, Chicago |
| Gleb Bournkov | EMBL-Hamburg |
| Michail Isupov | University of Exeter |
| Garib Murshudov | University of York |

All the authors of cited papers