# Things you don't want to see in your diffraction data

(Why we do want to see diffraction patterns in 3D)

Andrey Lebedev, CCP4

Do images indicate crystal pathologies?

If yes:

- successful structure solution is less likely
- even if structure is solved, bad refinement stats are very likely
  - explain in the manuscript (e.g. picture from dials.reciprocal_lattice_viewer)

A few examples are presented in this talk

- Graphical facilities in DIALS
- Low resolution and anisotropy
- Inter-grown crystals
- OD-structures
- Partially disordered OD-structures
- Pseudo-translation
- Non-commensurate modulated structures

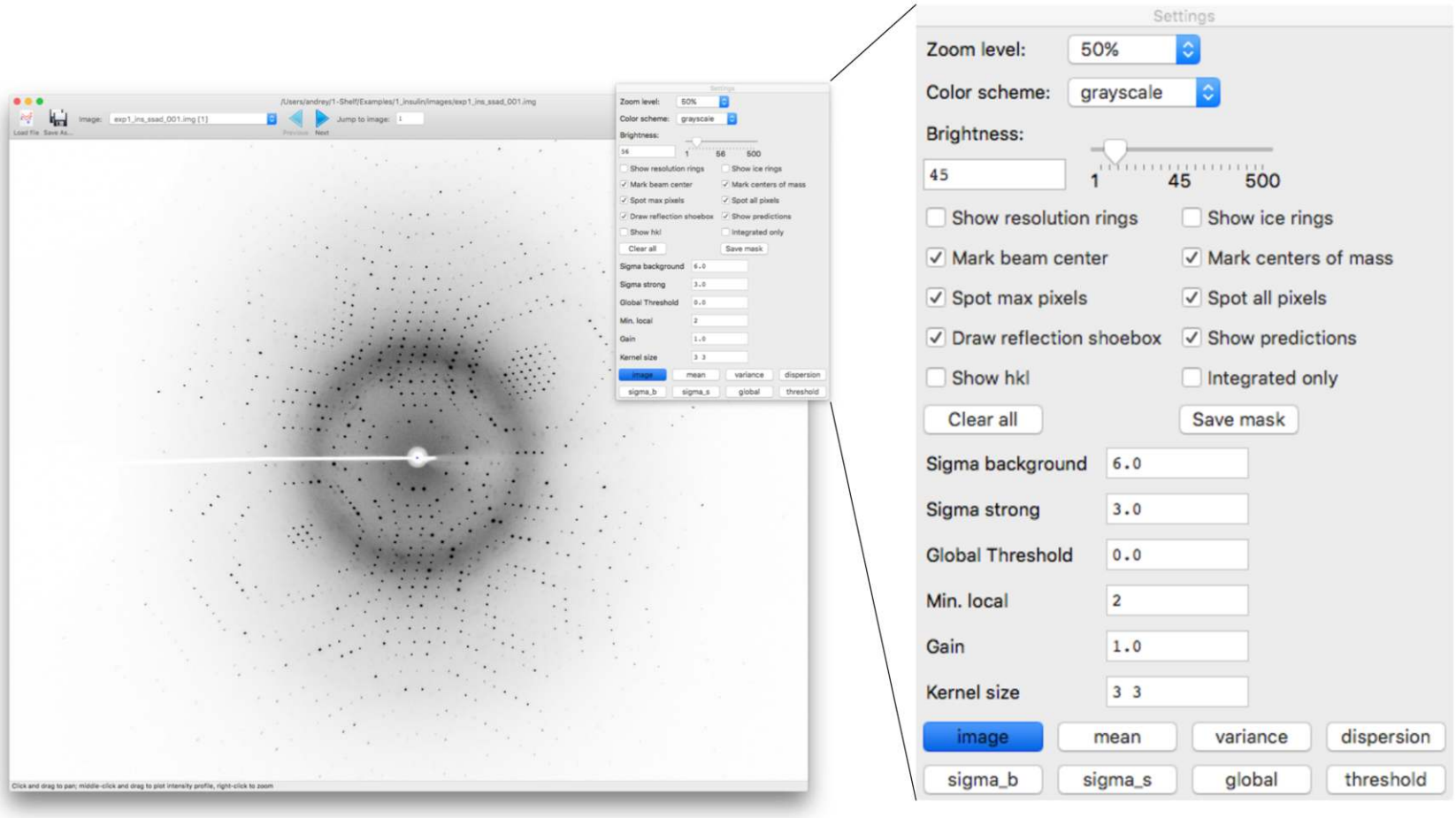
# Graphical facilities in DIALS

- A simple example
- Operating DIALS and viewers from the command line

# Import and image viewer

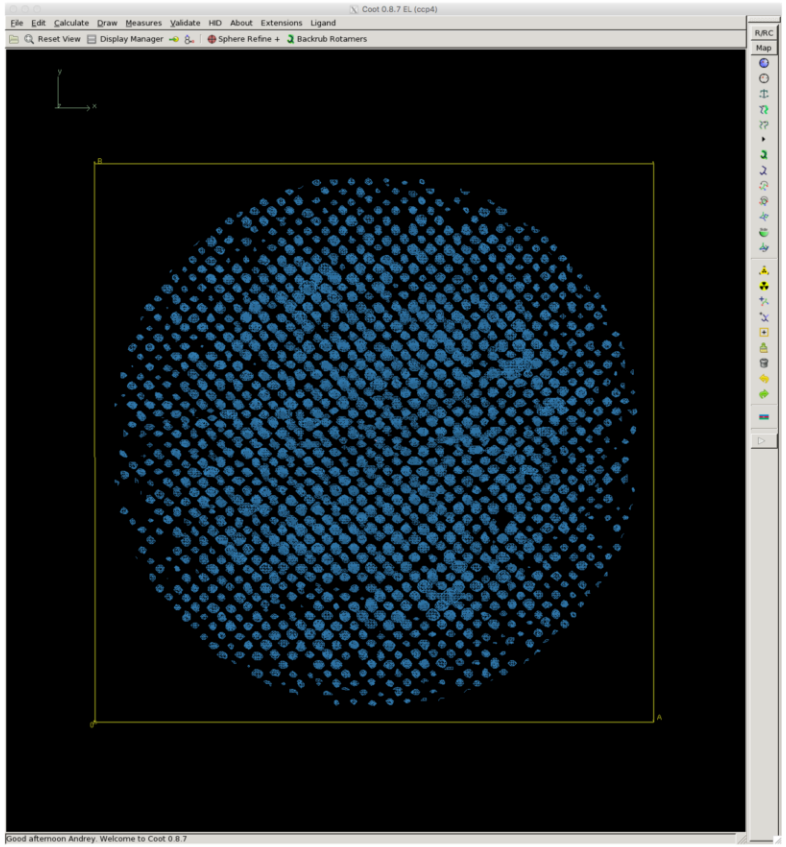
Cubic insulin, the experiment 1 from HZB MX tutorial

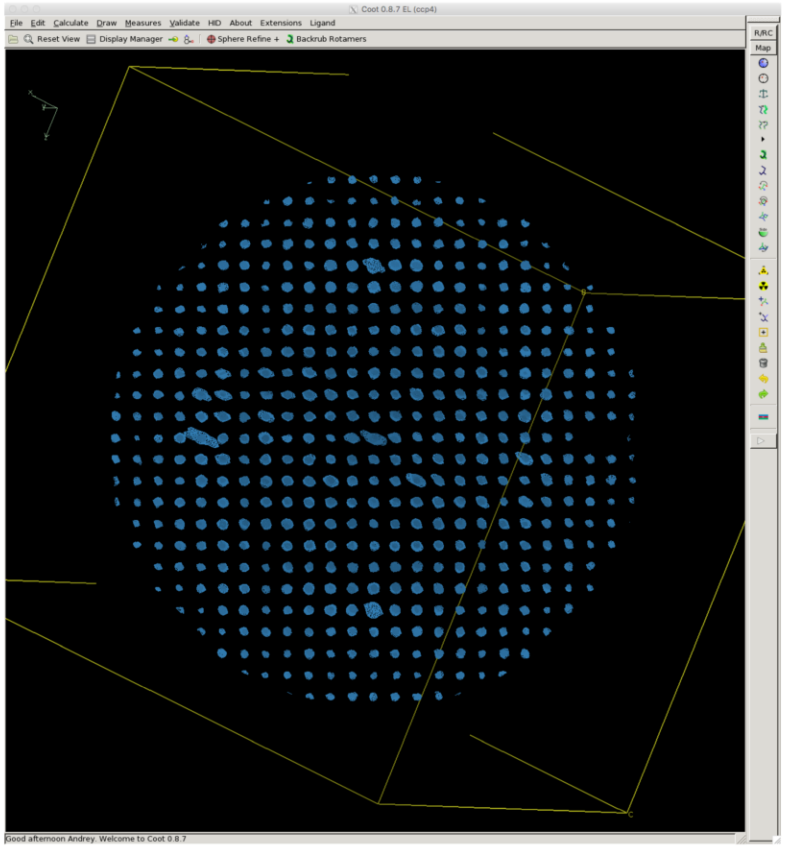
```
dials.import template=images/exp1_ins_ssad_###.img
dials.image_viewer datablock.json
```



# Sweep of images as 3D map

```
dials.rs_mapper map_file=output.ccp4 datablock.json
coot --map output.ccp4
```

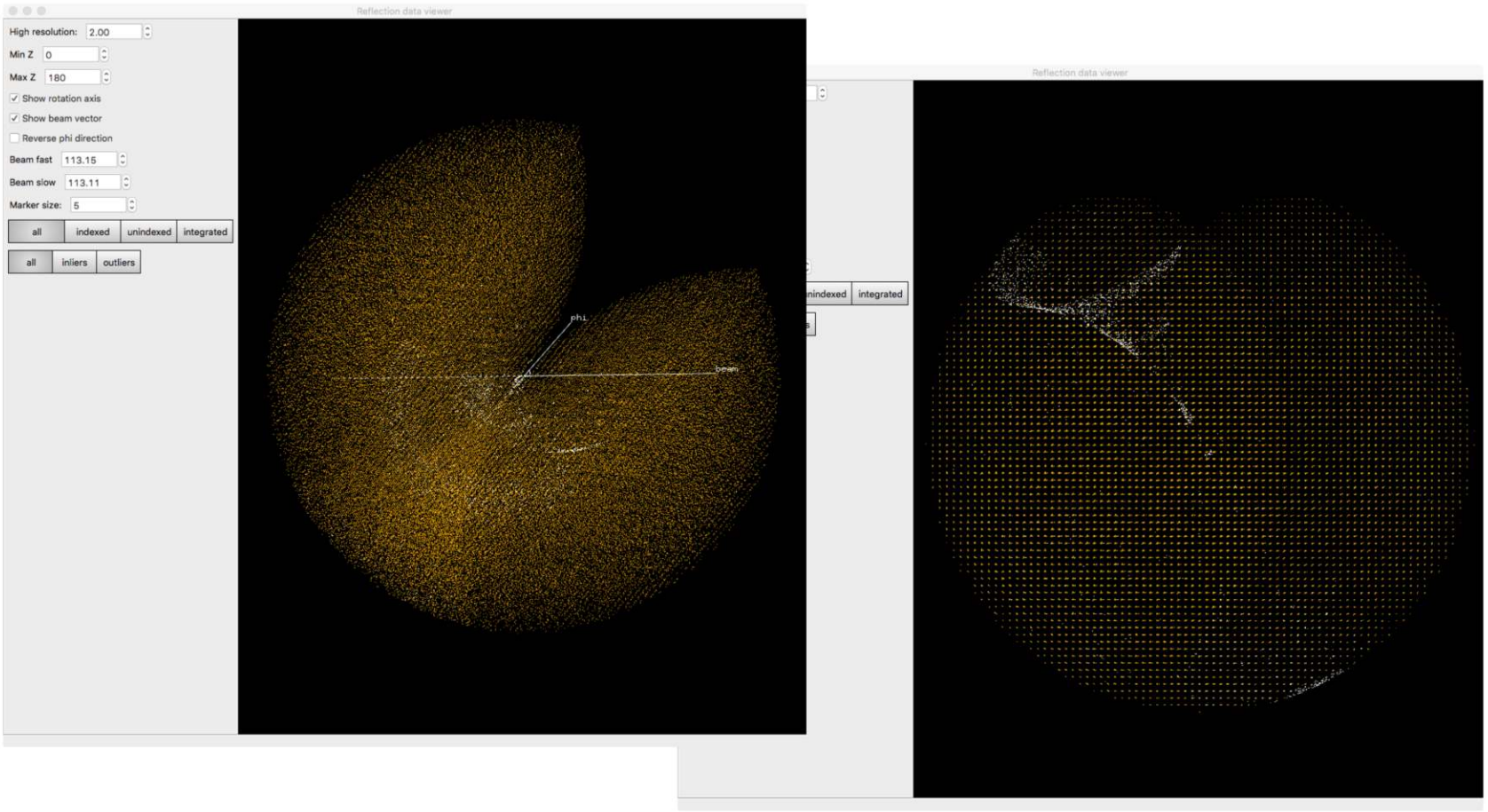



PyMol can be used as well

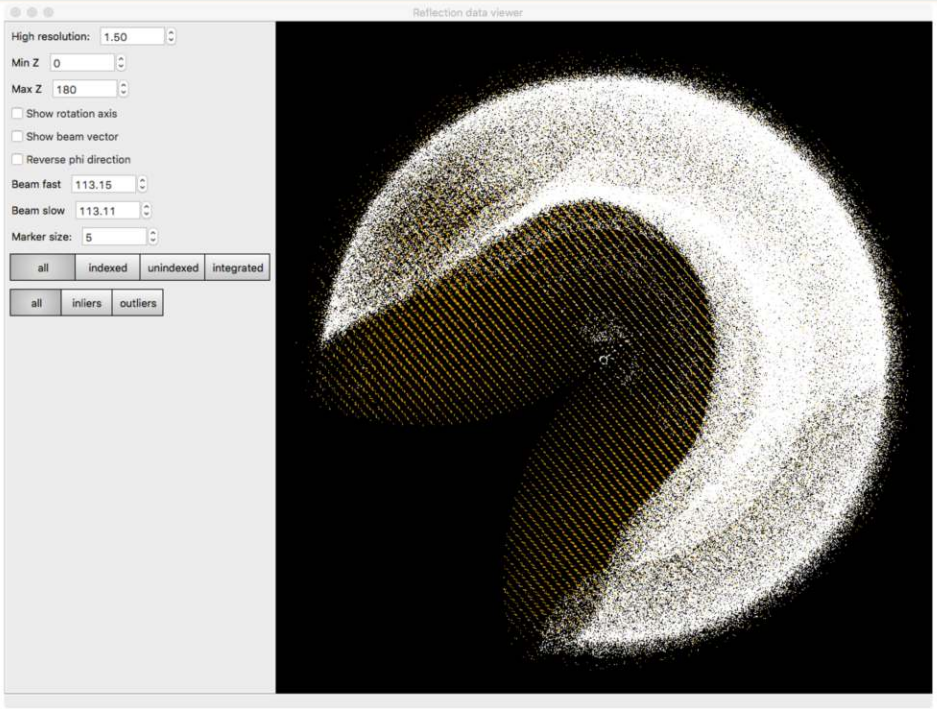
# Sweep of images: spots positions in 3D

```
dials.find_spots datablock.json
dials.reciprocal_lattice_viewer datablock.json strong.pickle
```



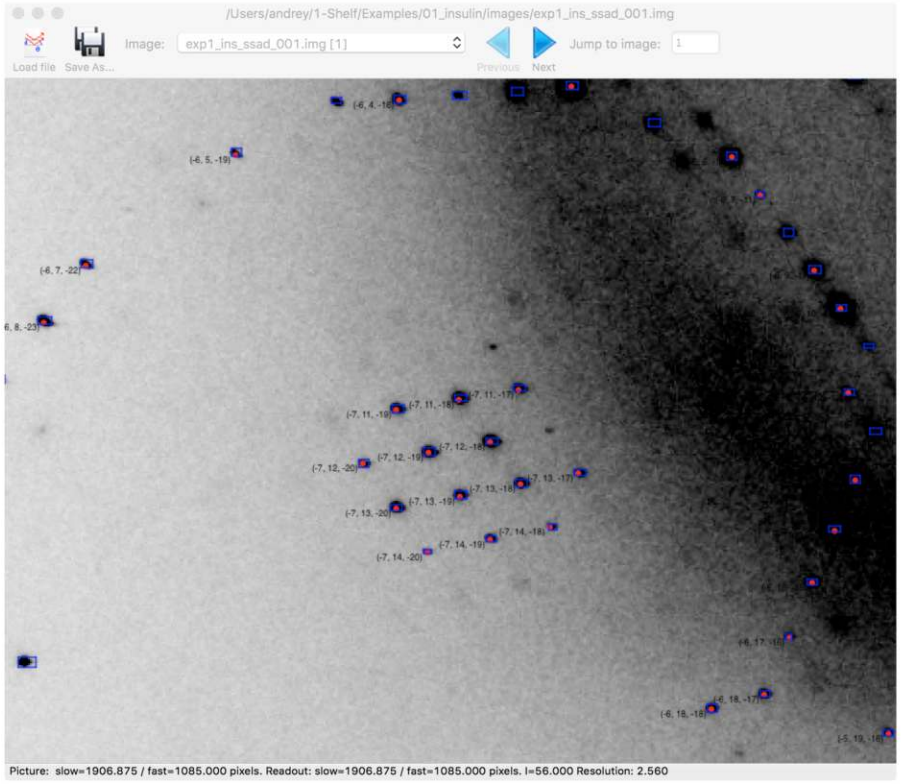
# More details are available after indexing

```
dials.index datablock.json strong.pickle
dials.refine experiments.json indexed.pickle scan_varying=True
dials.reciprocal_lattice_viewer refined.pickle refined_experiments.json
dials.image_viewer datablock.json
```




orange: indexed
white: not indexed

shoeboxes, predictions



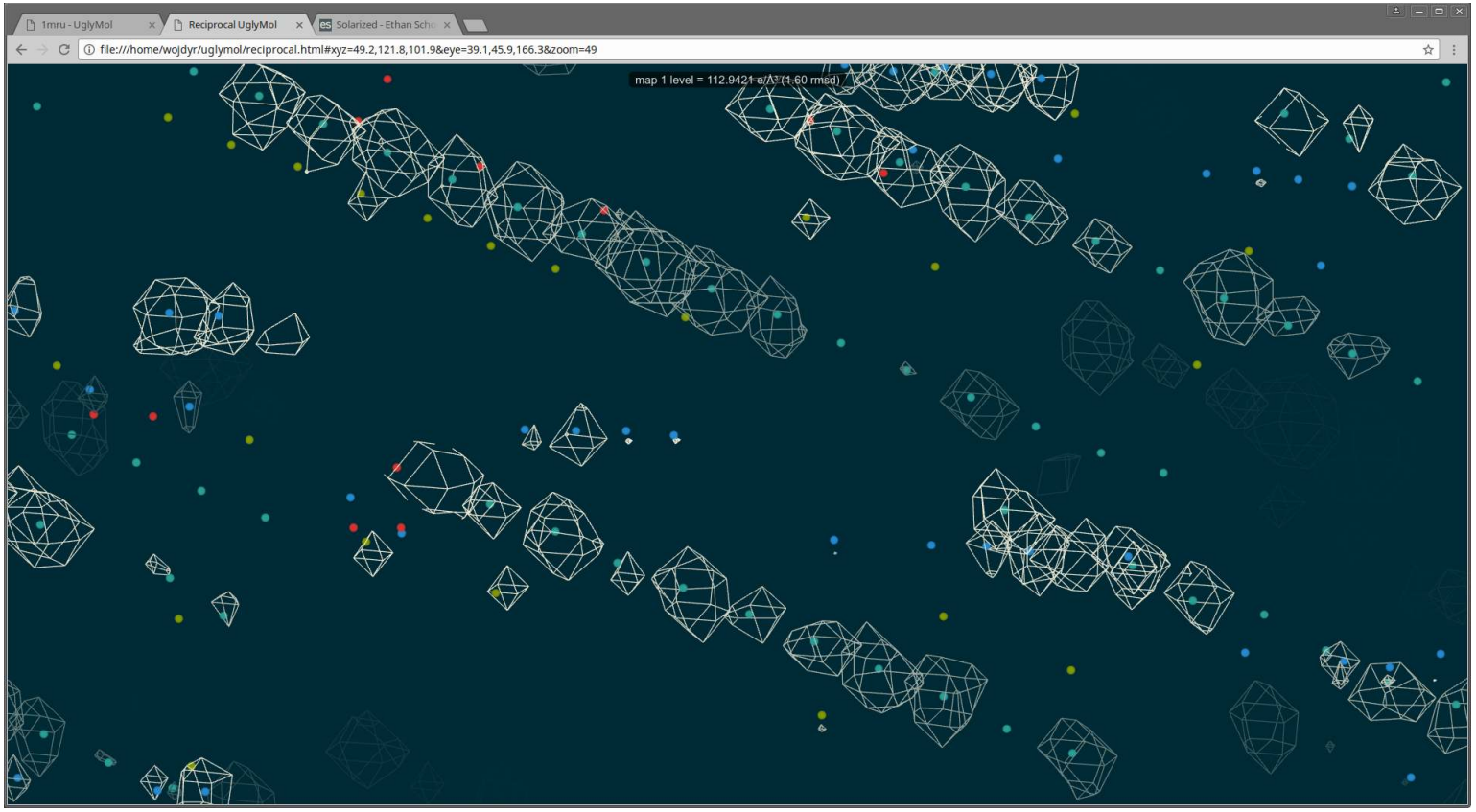
# Visualising XDS results

- XDS results can be imported to DIALS and visualised in 3D

```
dials.import_xds xds/
dials.import_xds method=reflections xds/SPOT.XDS
dials.reciprocal_lattice_viewer experiments.json spot_xds.pickle
```

# UglyMol

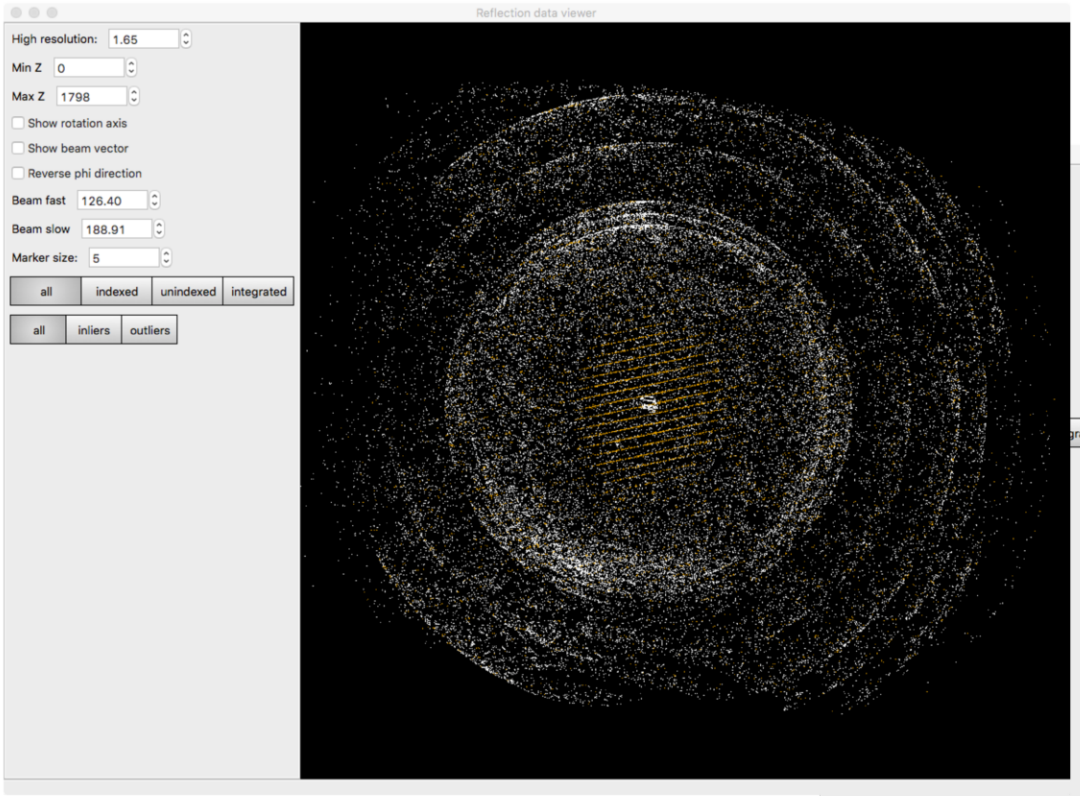
- Viewer for web-browsers combining both views, intensities as maps and spots as dots (by Marcin Wojdyr, CCP4, https://github.com/uglymol)
  - xia2 task in ccp4cloud



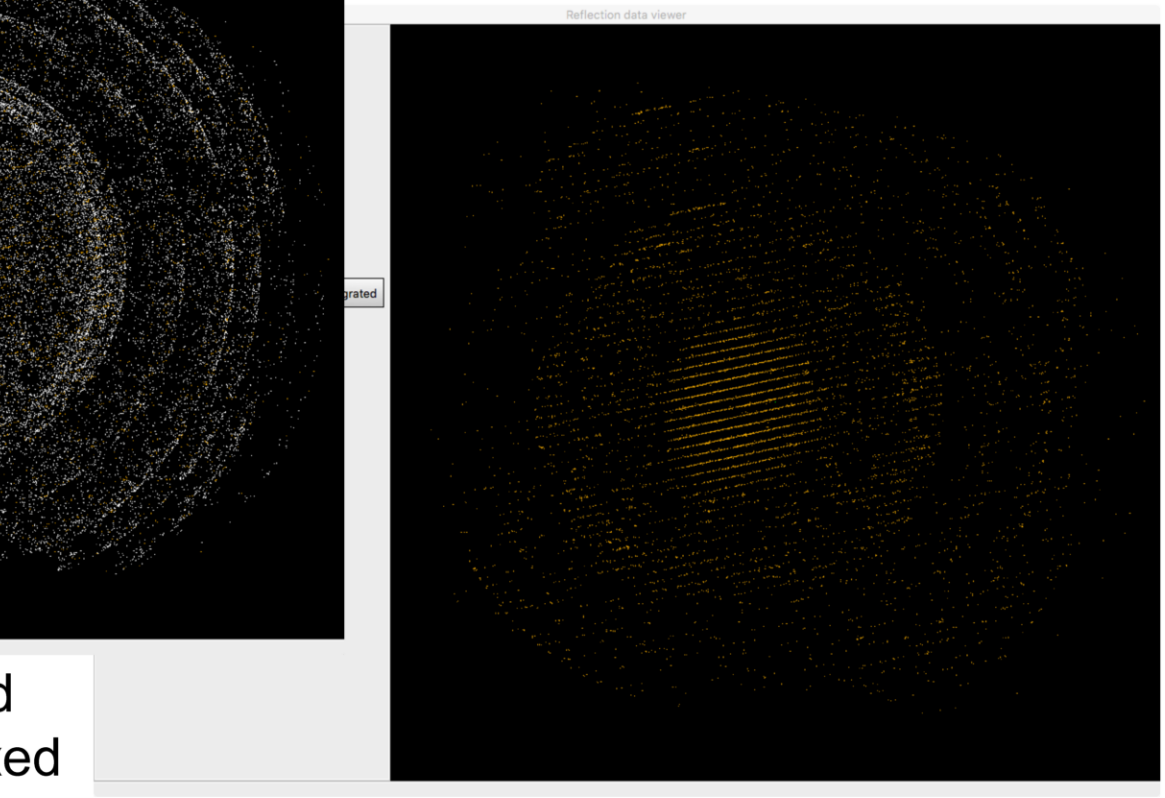
# Low resolution data and anisotropy

# Low resolution data

```
dials.reciprocal_lattice_viewer experiments.json spot_xds.pickle
```

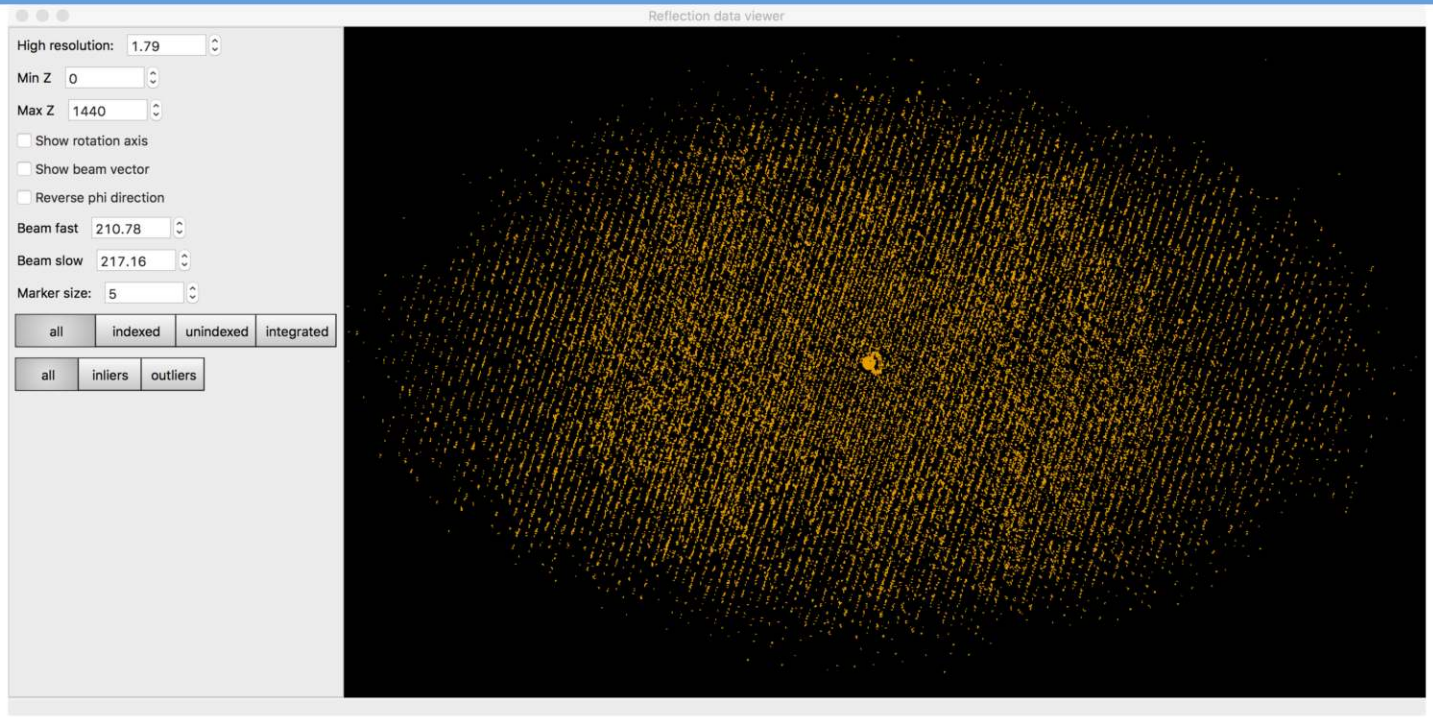



These are not 1.5 A data



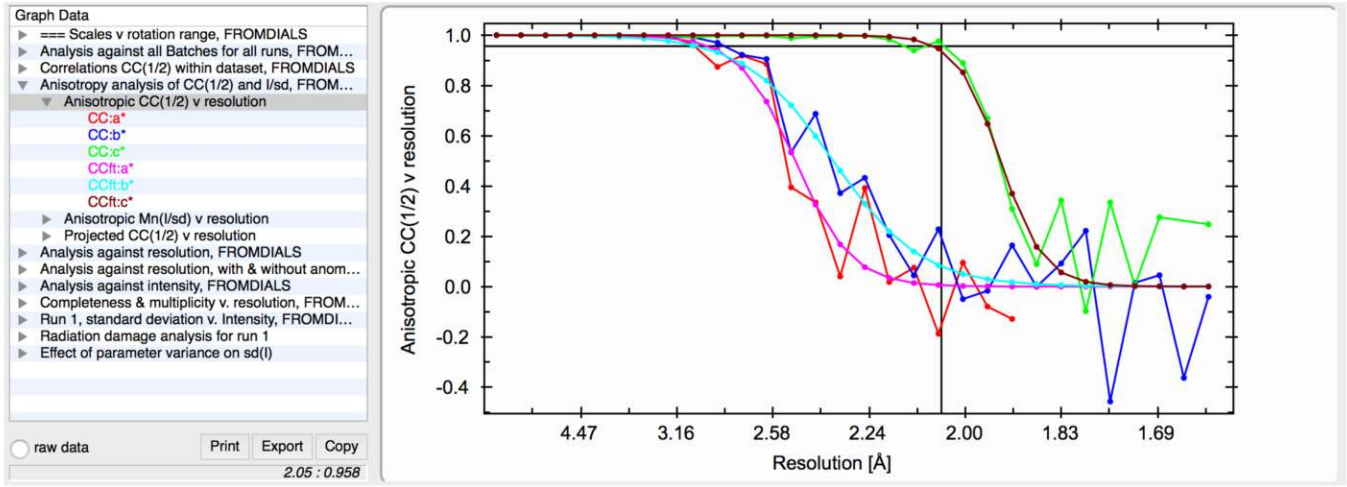

orange: indexed
white: not indexed

# Anisotropy: viewer, $CC_{1/2}$ plots

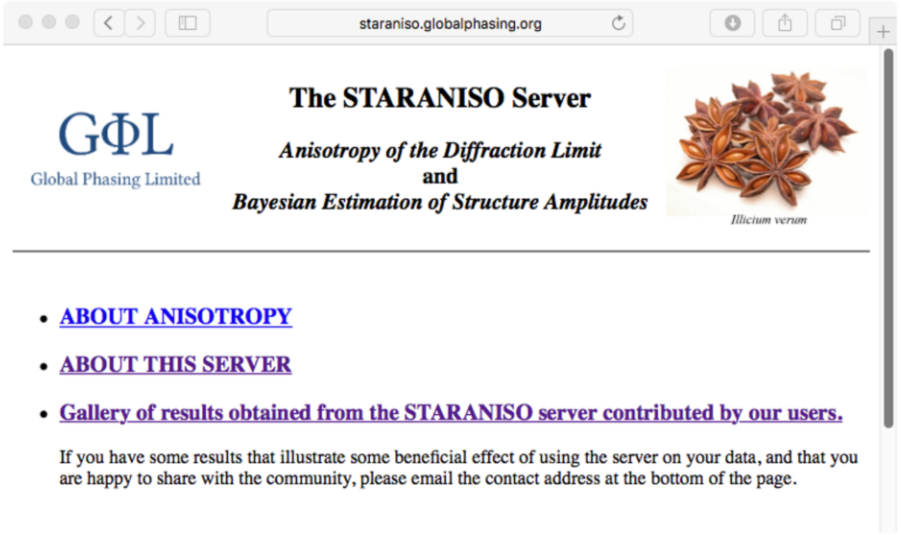


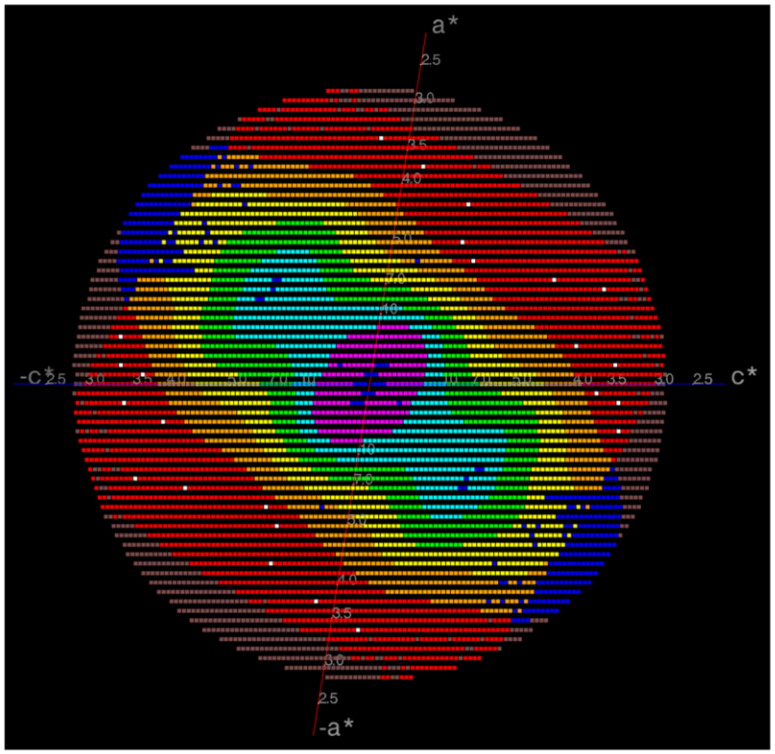

spot representation in DIALS viewer

CC(1/2) plots for three orthogonal directions in Aimless



# STARANISO Server



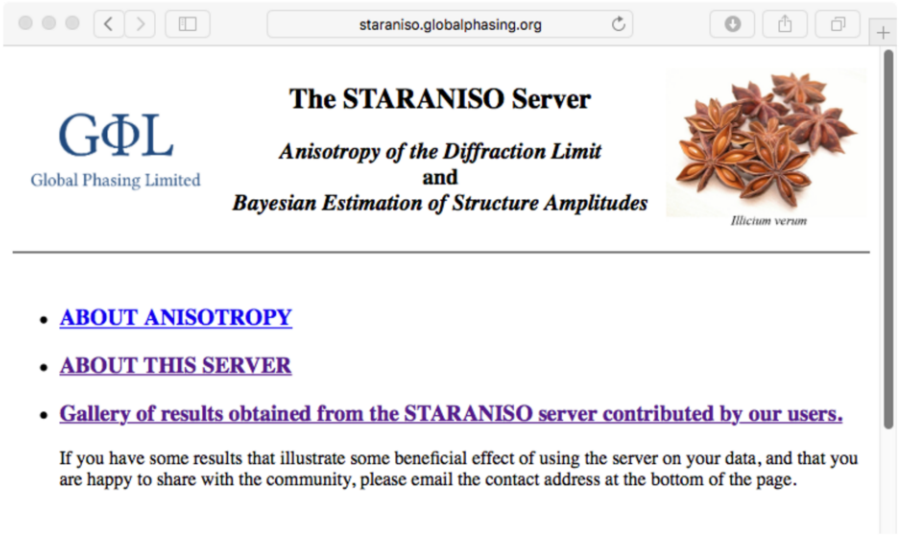



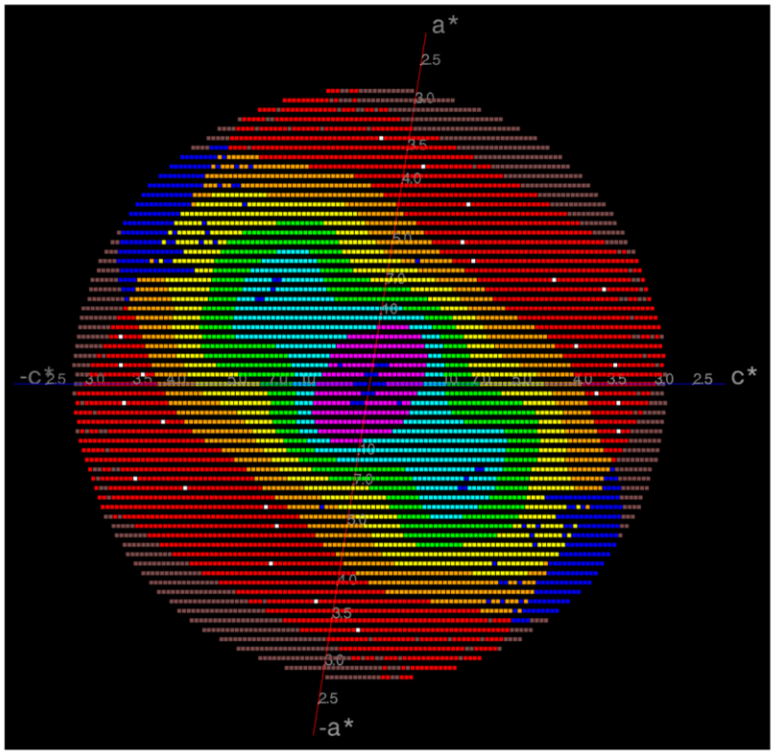

Non-spherical data truncation

- Removes noise
  - Better refinement stats
- Keeps (and optionally corrects) all useful data
  - In some cases is critical for structure solution, model building and ligand fitting
  - Some programs (e.g. Phaser) deal with anisotropy OK

- Deposit original intensities!
  - along with Fs or Is you have used in refinement

# STARANISO Server




Non-spherical data truncation

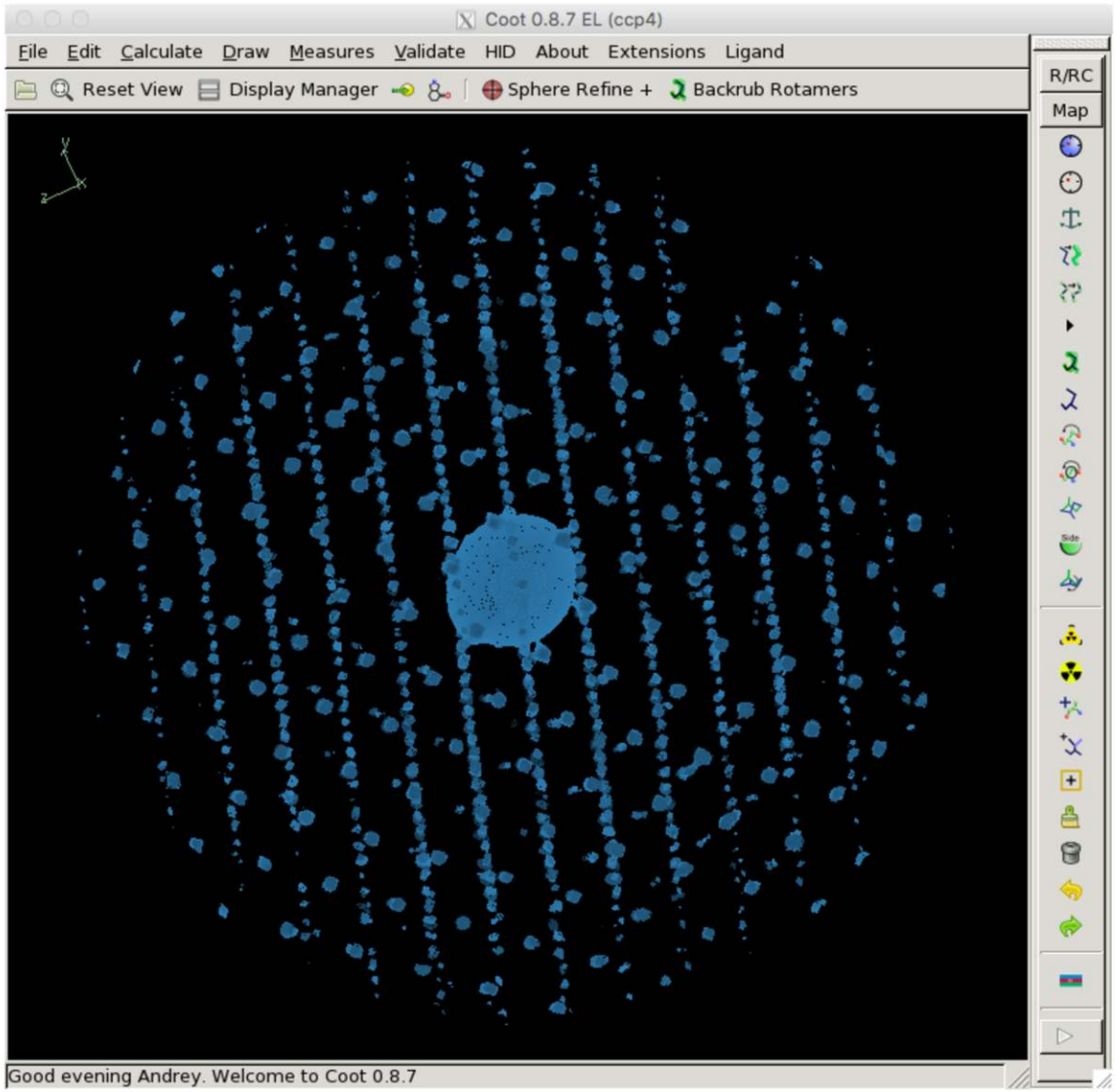
- Removes noise
  - Better refinement stats
- Keeps (and optionally corrects) all useful data
  - In some cases is critical for structure solution, model building and ligand fitting
  - Some programs (e.g. Phaser) deal with anisotropy OK

- Deposit original intensities! (along with Fs or Is used in refinement)

# Inter-grown crystals (multi-lattice data)

# Example of random crystal inter-growth

```
dials.rs_mapper ...
coot --map output.ccp4
```



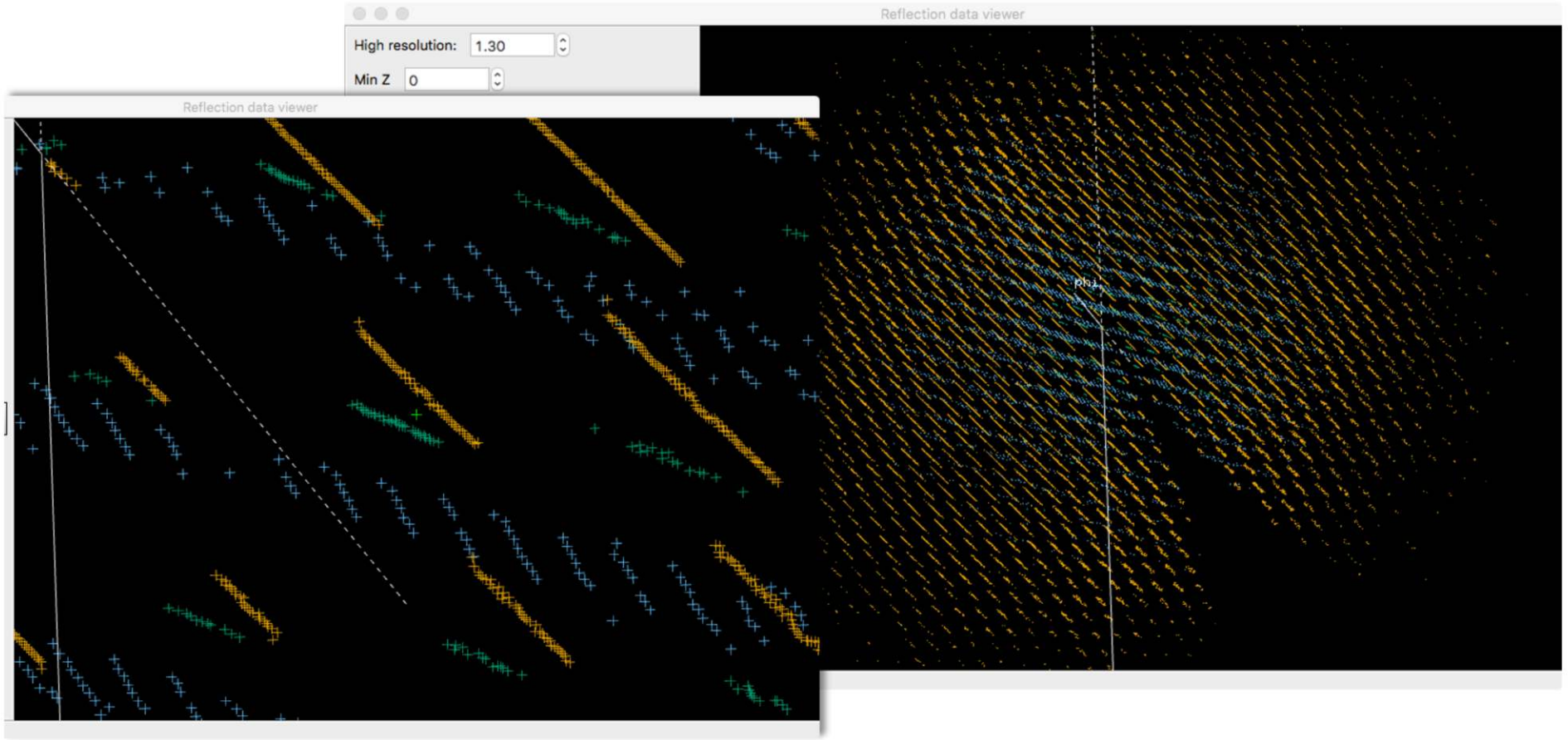

beta-lactamase OXA-163
PDB ID 4s2m

Data from Vlatko Stojanoski
Baylor College of Medicine

# Example of random crystal inter-growth

```
dials.index datablock.json strong.pickle max_lattices=3 hkl_tolerance=0.1
dials.reciprocal_lattice_viewer refined.pickle refined_experiments.json
```

- different colour means different lattice
- individual lattices can be switched off and on



# Example of random crystal inter-growth

Easy case:

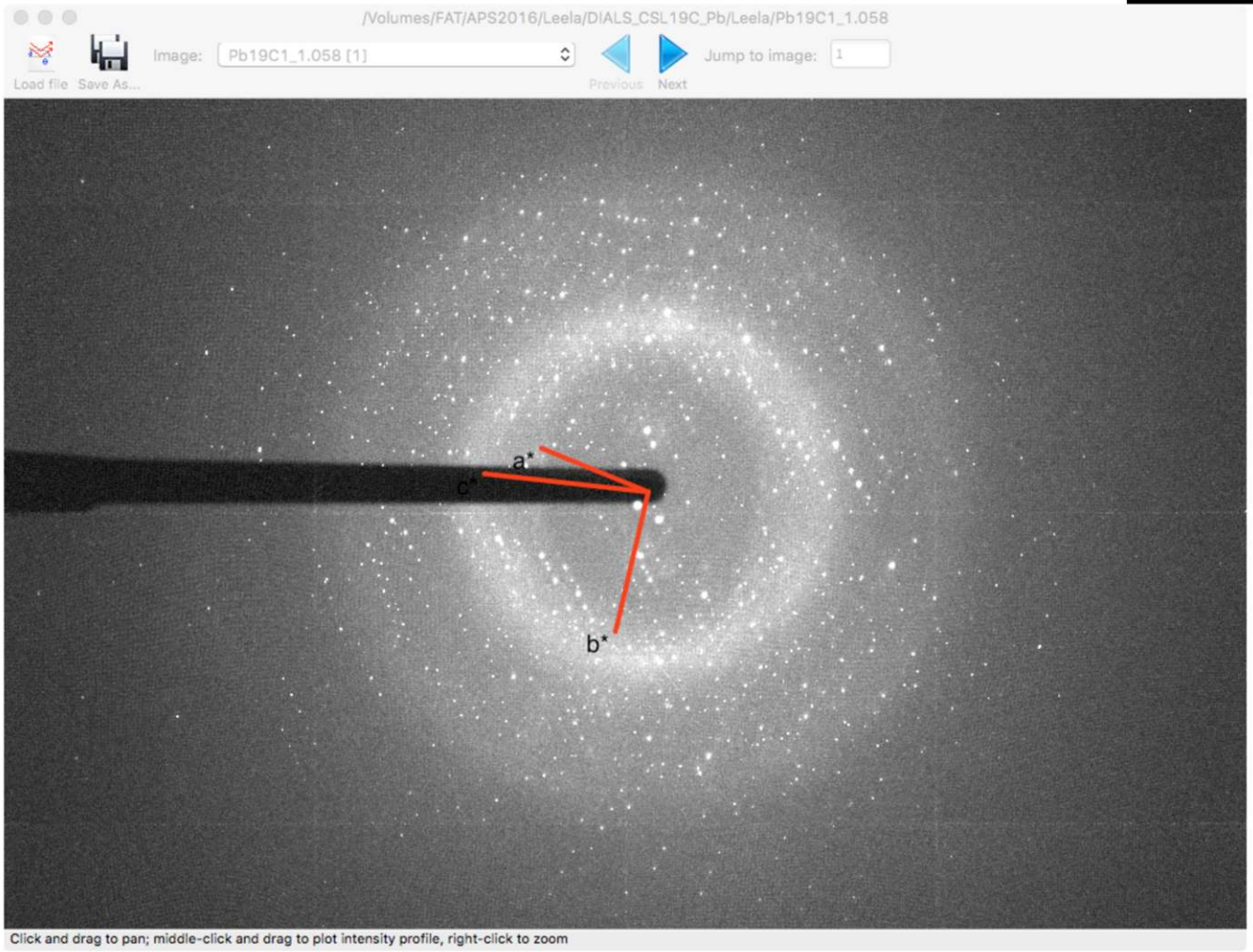
- Lattices are mainly separated, with only very few reflection overlapping
- Signal from one lattice is substantially higher than from others

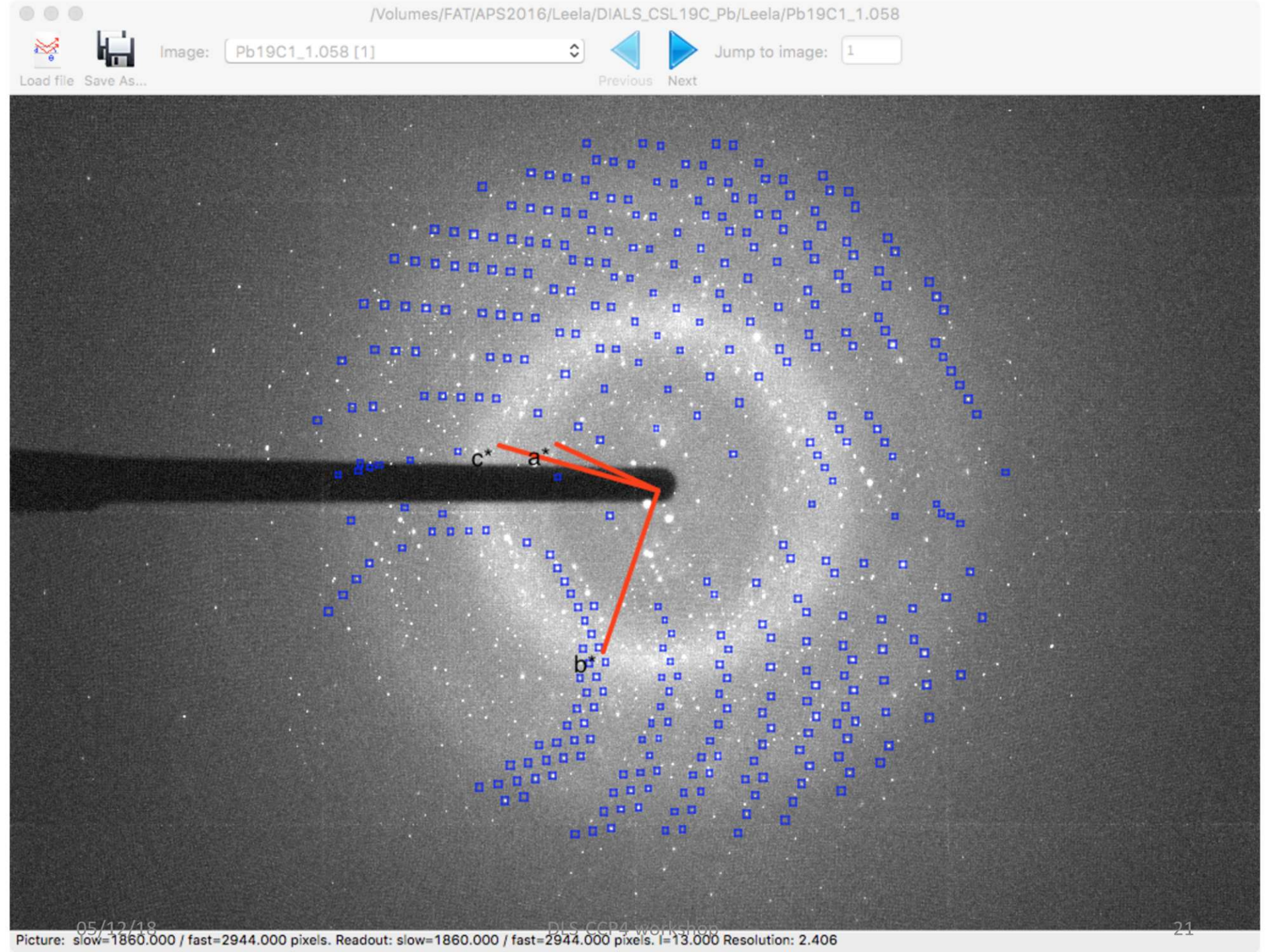
The intensities for the strongest lattice were processed,
structure solved and refined to R=0.20 R-free=0.26

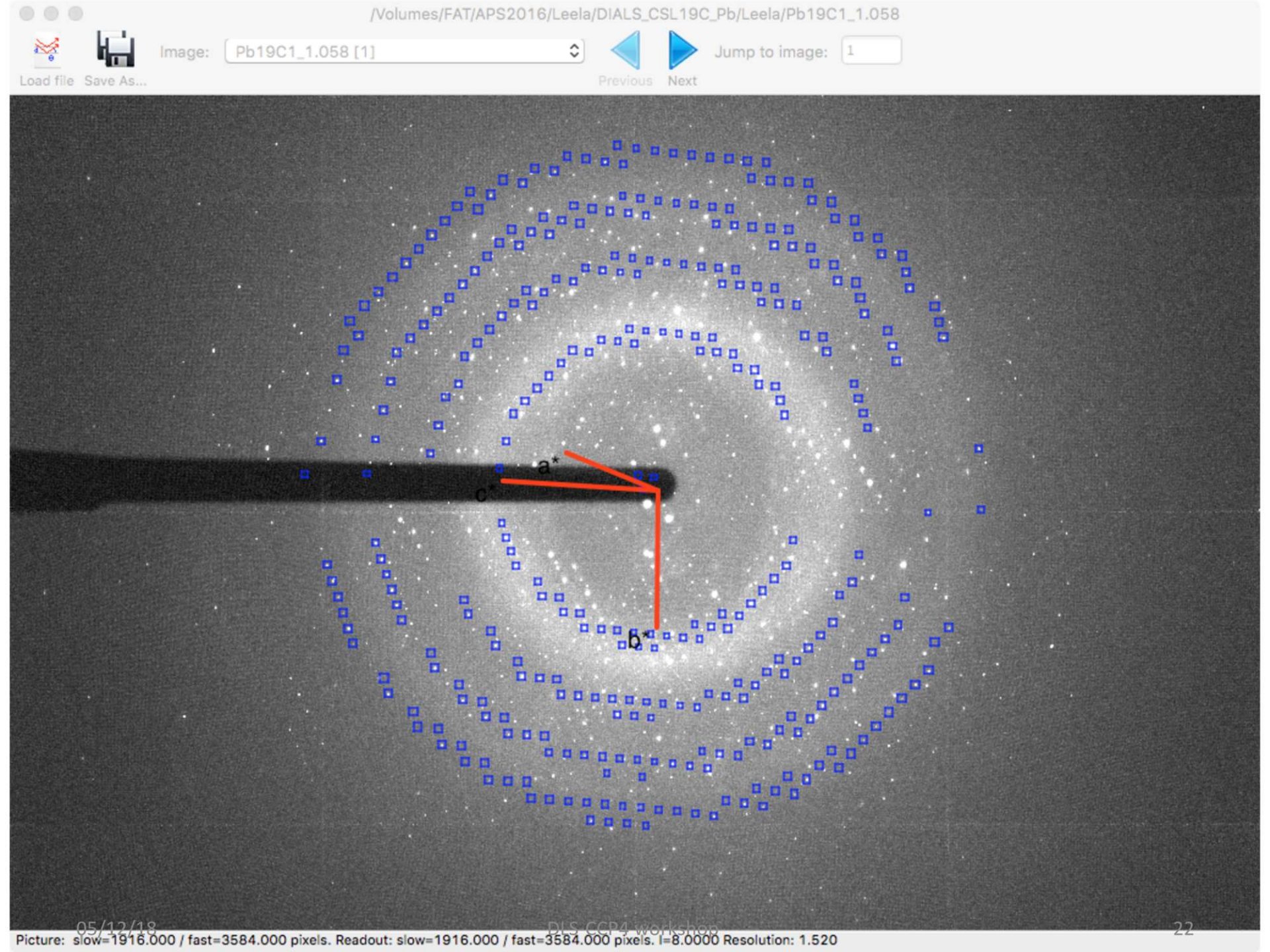
# An extreme case

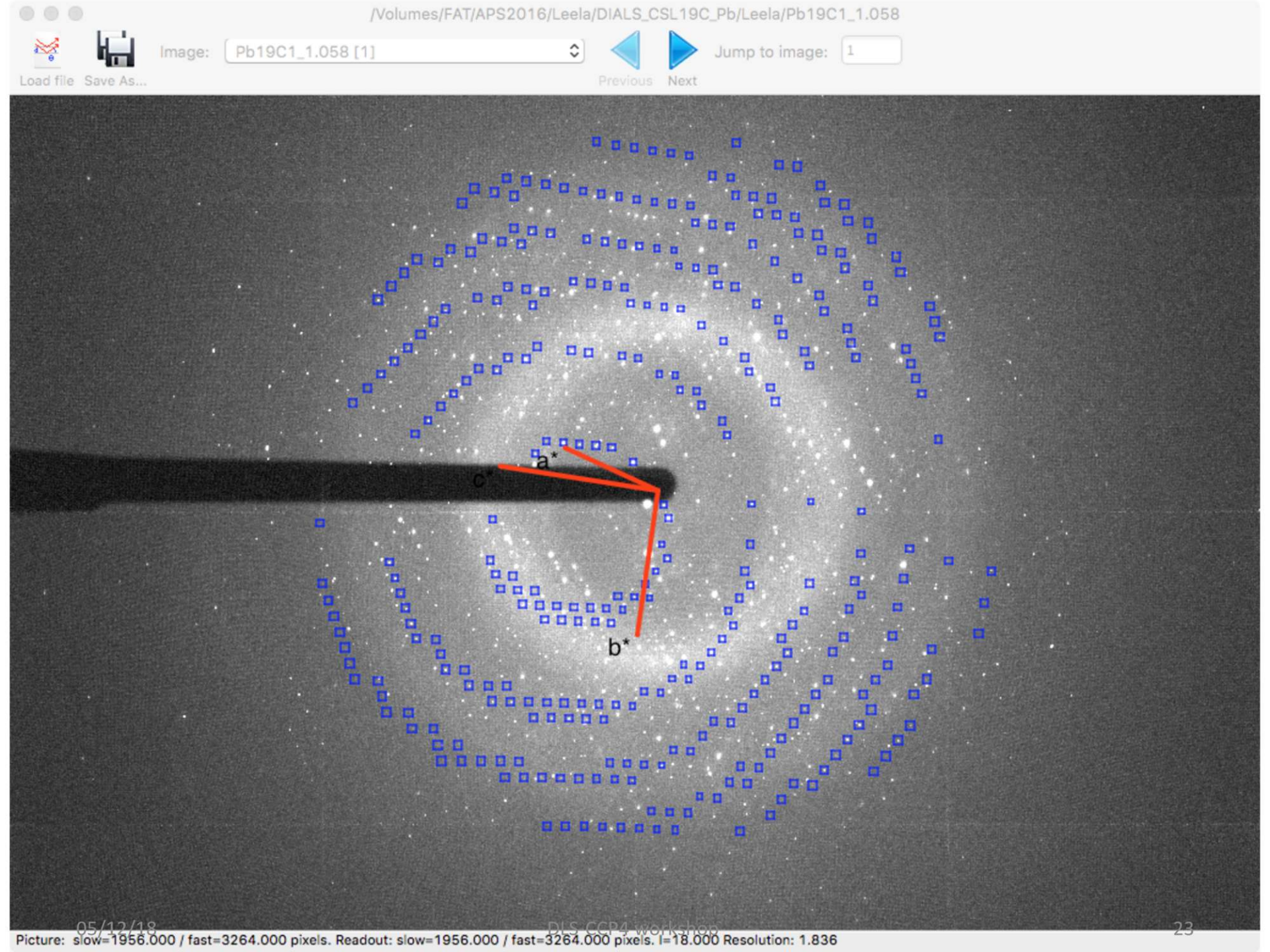
Example from Leela Ruckthong

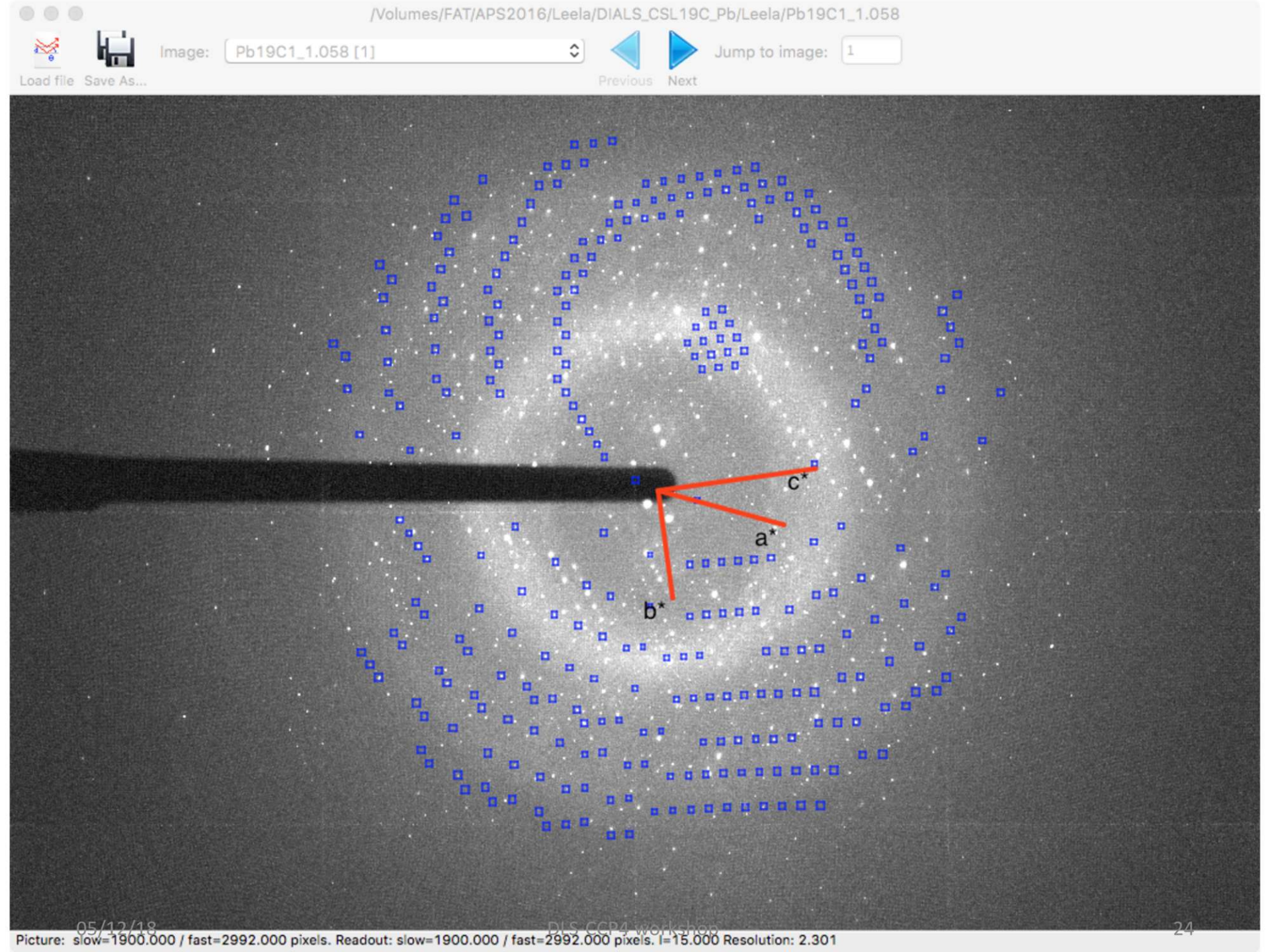
- How many lattices you can spot here?

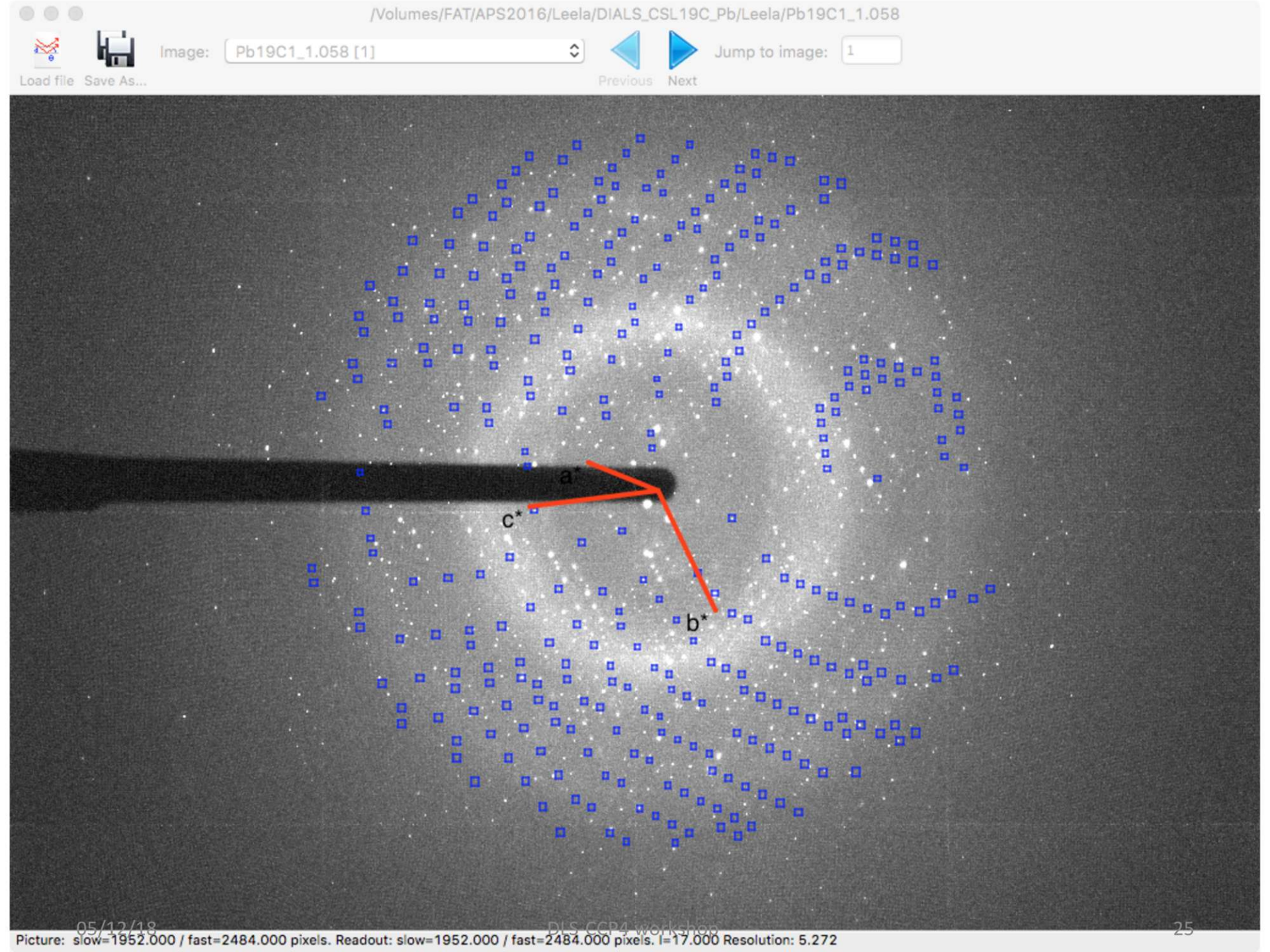


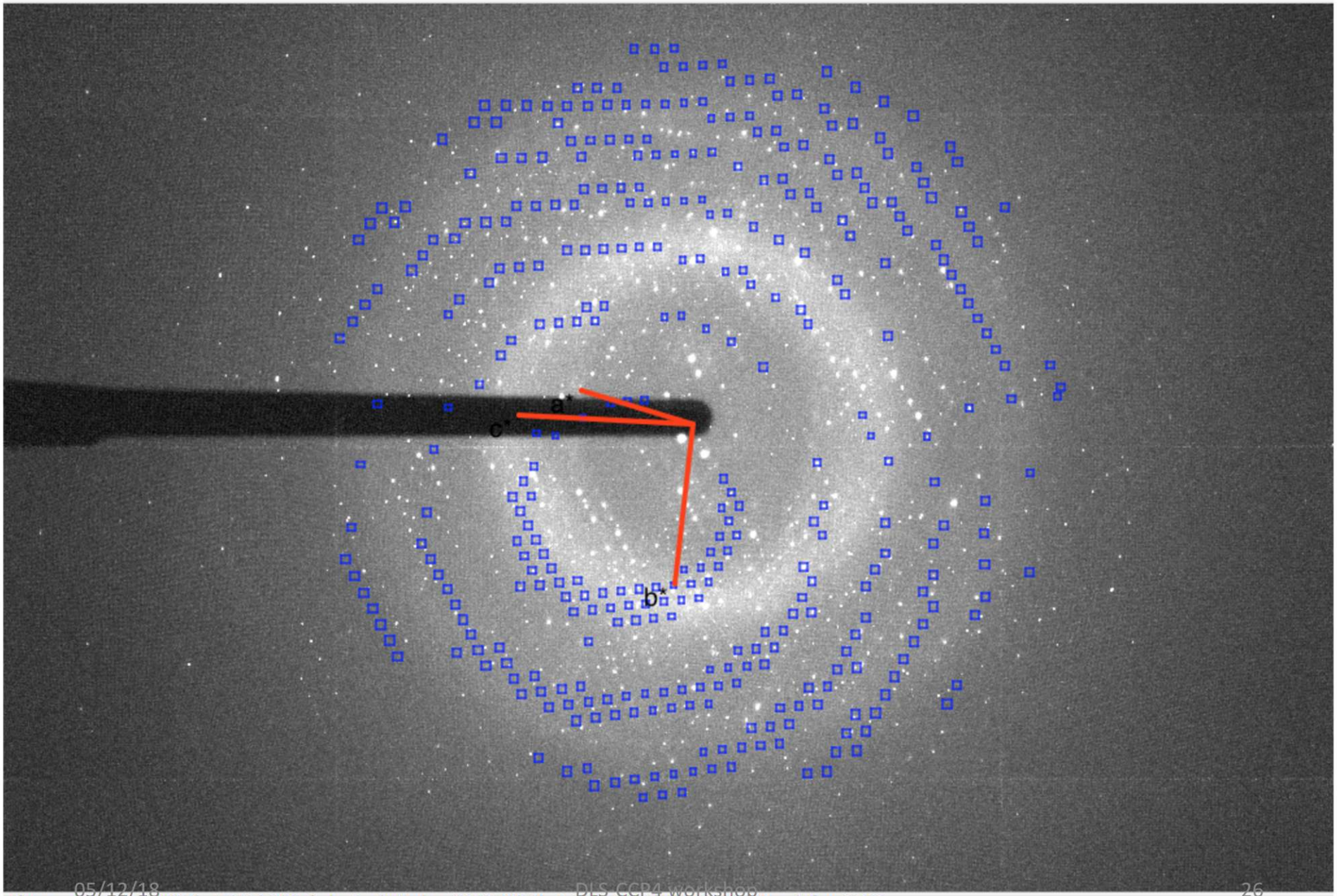

/Volumes/FAT/APS2016/Leela/DIALS_CSL19C_Pb/Leela/Pb19C1_1.058
Load file
Save As...
Image:
Pb19C1_1.058 [1]
Previous
Next
Jump to image:
1
c*
a*
b*
Picture: slow=1860.000 / fast=2944.000 pixels. Readout: slow=1860.000 / fast=2944.000 pixels. I=13.000 Resolution: 2.406

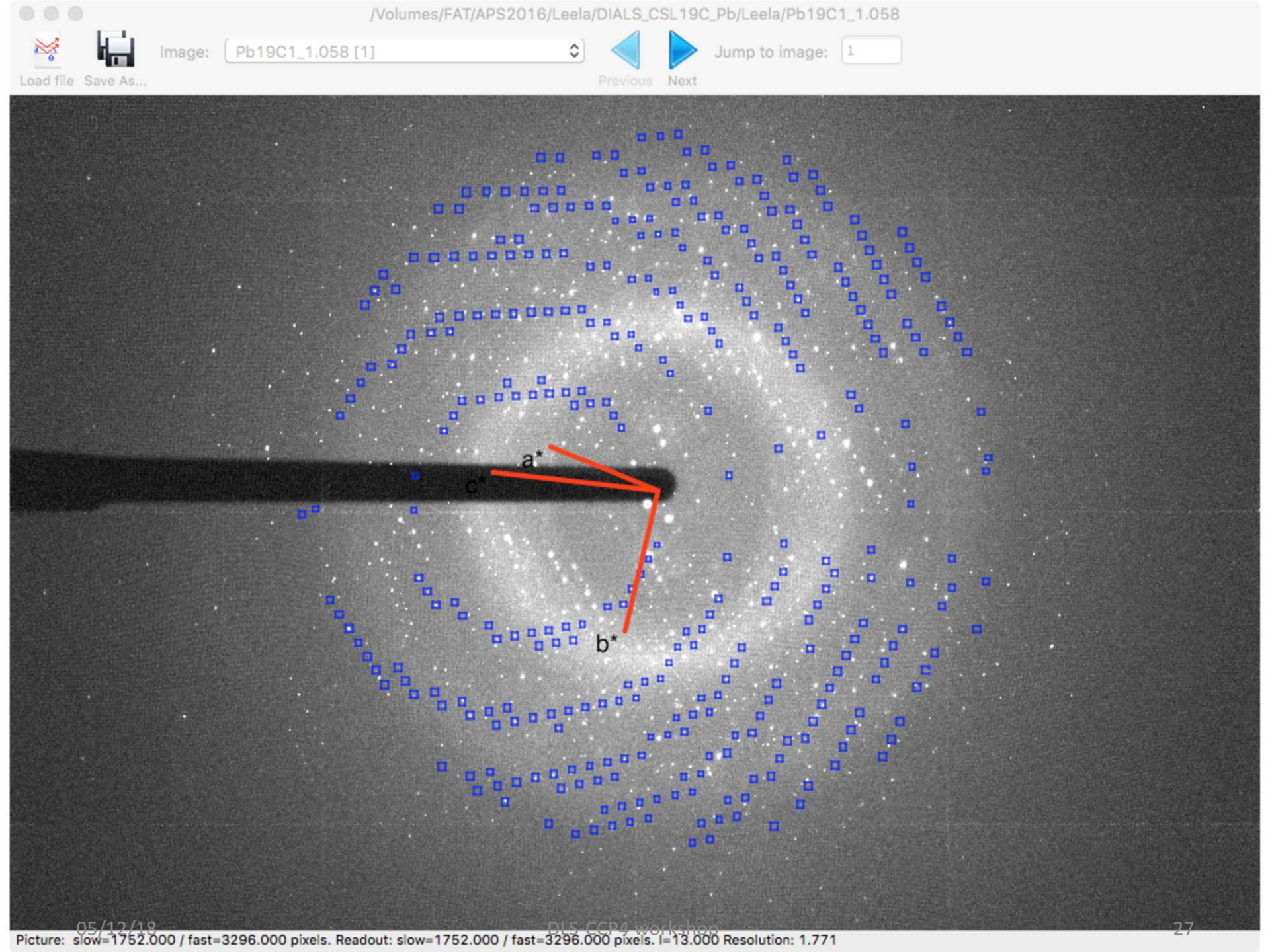

/Volumes/FAT/APS2016/Leela/DIALS_CSL19C_Pb/Leela/Pb19C1_1.058
Load file
Save As...
Image:
Pb19C1_1.058 [1]
Previous
Next
Jump to image:
1
a*
c*
b*
Picture: slow=1916.000 / fast=3584.000 pixels. Readout: slow=1916.000 / fast=3584.000 pixels. I=8.0000 Resolution: 1.520

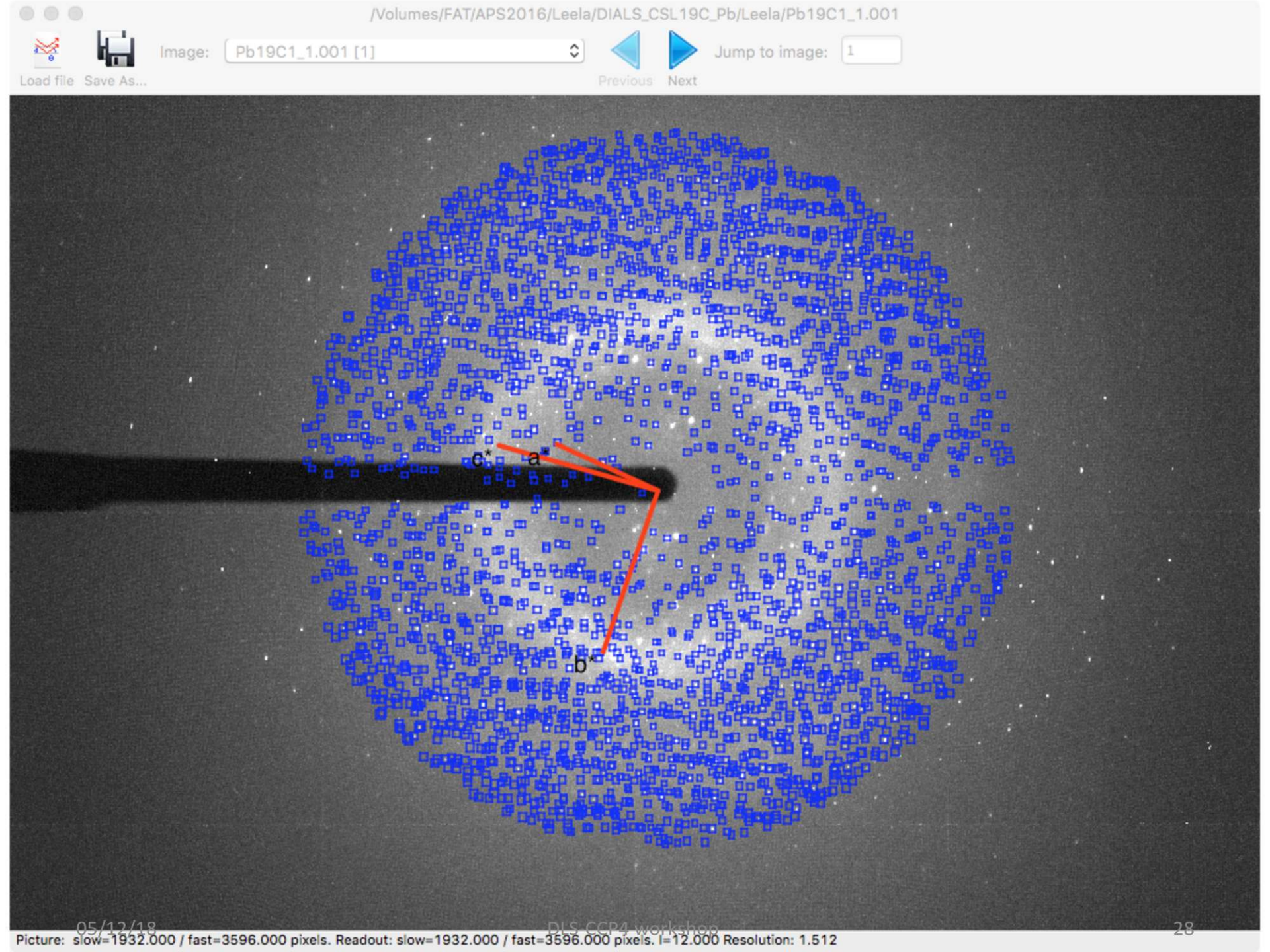

/Volumes/FAT/APS2016/Leela/DIALS_CSL19C_Pb/Leela/Pb19C1_1.058
Load file
Save As...
Image:
Pb19C1_1.058 [1]
Previous
Next
Jump to image:
1
a*
c*
b*
Picture: slow=1956.000 / fast=3264.000 pixels. Readout: slow=1956.000 / fast=3264.000 pixels. I=18.000 Resolution: 1.836

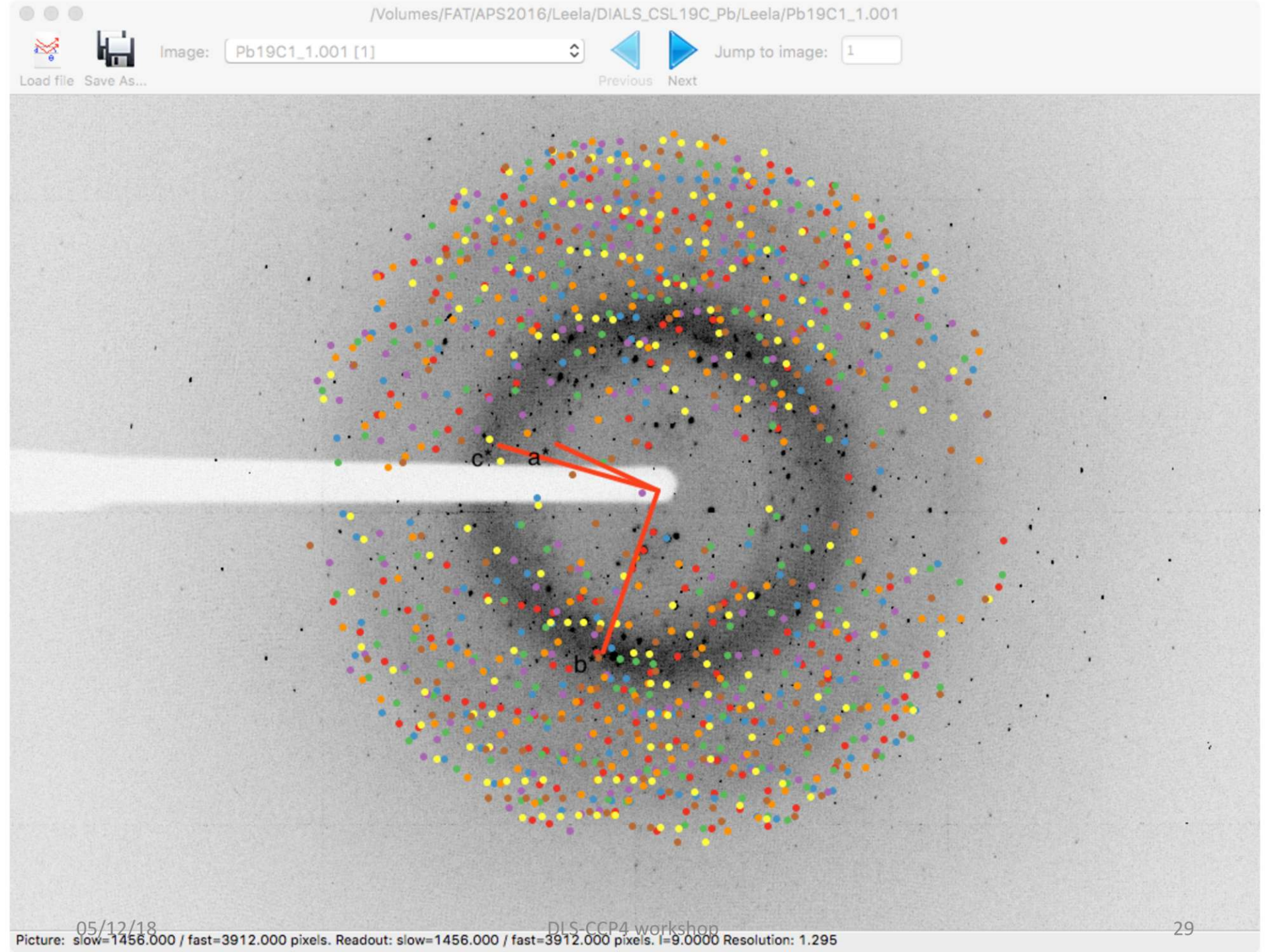

/Volumes/FAT/APS2016/Leela/DIALS_CSL19C_Pb/Leela/Pb19C1_1.058
Load file
Save As...
Image:
Pb19C1_1.058 [1]
Previous
Next
Jump to image:
1
c*
a*
b*
Picture: slow=1900.000 / fast=2992.000 pixels. Readout: slow=1900.000 / fast=2992.000 pixels. I=15.000 Resolution: 2.301


/Volumes/FAT/APS2016/Leela/DIALS_CSL19C_Pb/Leela/Pb19C1_1.058
Load file
Save As...
Image:
Pb19C1_1.058 [1]
Previous
Next
Jump to image:
1
a*
c*
b*
Picture: slow=1952.000 / fast=2484.000 pixels. Readout: slow=1952.000 / fast=2484.000 pixels. I=17.000 Resolution: 5.272


/Volumes/FAT/APS2016/Leela/DIALS_CSL19C_Pb/Leela/Pb19C1_1.058
Load file
Save As...
Image:
Pb19C1_1.058 [1]
Previous
Next
Jump to image:
1
a*
c*
b*
Picture: slow=2152.000 / fast=3312.000 pixels. Readout: slow=2152.000 / fast=3312.000 pixels. I=7.0000 Resolution: 1.761


/Volumes/FAT/APS2016/Leela/DIALS_CSL19C_Pb/Leela/Pb19C1_1.058
Load file
Save As...
Image:
Pb19C1_1.058 [1]
Previous
Next
Jump to image:
1
a*
c*
b*
Picture: slow=1752.000 / fast=3296.000 pixels. Readout: slow=1752.000 / fast=3296.000 pixels. I=13.000 Resolution: 1.771


/Volumes/FAT/APS2016/Leela/DIALS_CSL19C_Pb/Leela/Pb19C1_1.001
Load file
Save As...
Image:
Pb19C1_1.001 [1]
Previous
Next
Jump to image:
1
c*
a*
b*
Picture: slow=1932.000 / fast=3596.000 pixels. Readout: slow=1932.000 / fast=3596.000 pixels. I=12.000 Resolution: 1.512


/Volumes/FAT/APS2016/Leela/DIALS_CSL19C_Pb/Leela/Pb19C1_1.001
Load file
Save As...
Image:
Pb19C1_1.001 [1]
Previous
Next
Jump to image:
1
c
a
b
Picture: slow=1456.000 / fast=3912.000 pixels. Readout: slow=1456.000 / fast=3912.000 pixels. I=9.0000 Resolution: 1.295

# Seven lattices: too many overlapping spots?

- Only the strongest single lattice gave reasonable merged data
  - » all others were incomplete or had much lower I/sig(I)
  - » merging data from several lattices did not work well

- Unfortunately, the merged data were not good enough for modelling the protein residues of interest
  - » possibly because of too many overlapping reflections from different lattices.

- Solved and published

# Summary on multiple lattices

- Usually it is reasonable to use the data derived from one singe lattice
- Sometimes completeness can be improved by merging datasets derived from two or more lattices
- Sometimes the best lattice is not the first found by Dials
- DIALS: Indexing all the lattices together facilitates refinement of the parameters for each individual lattice

# Order-Disorder structures (OD-structures)

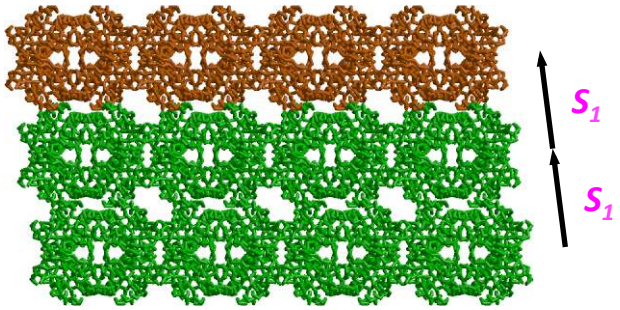
- Definition
- Example of an OD twin
- Example of allotwin

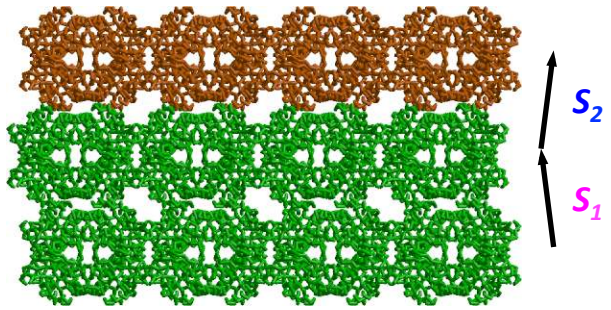
- Relatively frequent
- Data can sometimes be useful
- Good for illustration purpose

# Order-disorder structures (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



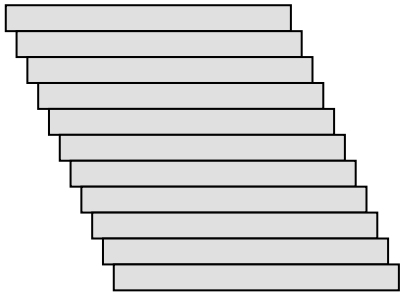




*) $S_1$ and $S_2$. are called stacking vectors

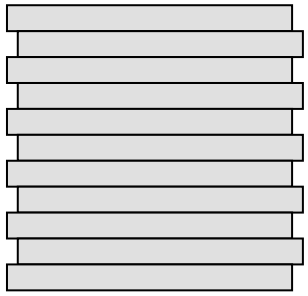
- two-dimensional periodicity
- a potential for disorder in the third dimension

# OD-structures

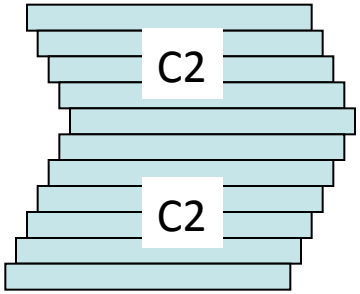
Single crystal



Single crystal

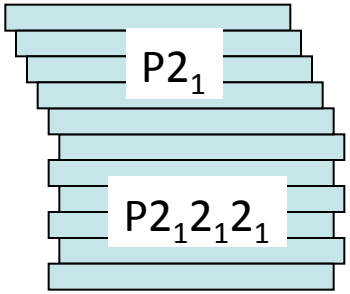


OD-twin



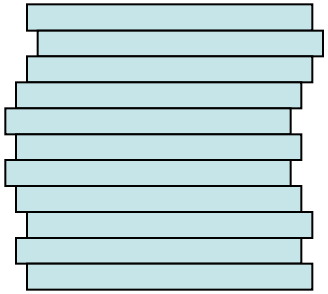

*Example 1*

Allotwin




*Example 2*

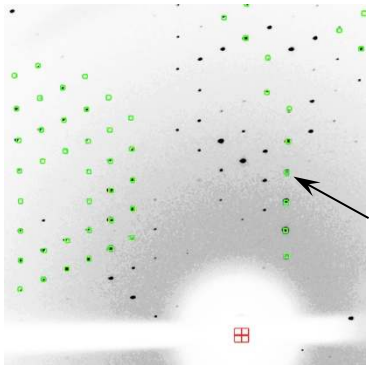
Partially disordered OD-structure



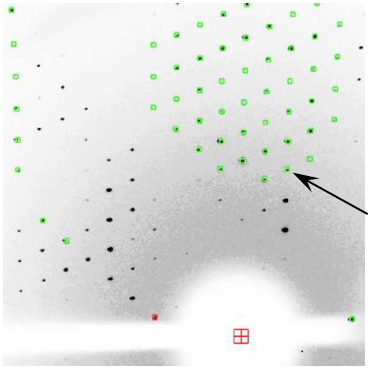
*Examples in the next section*

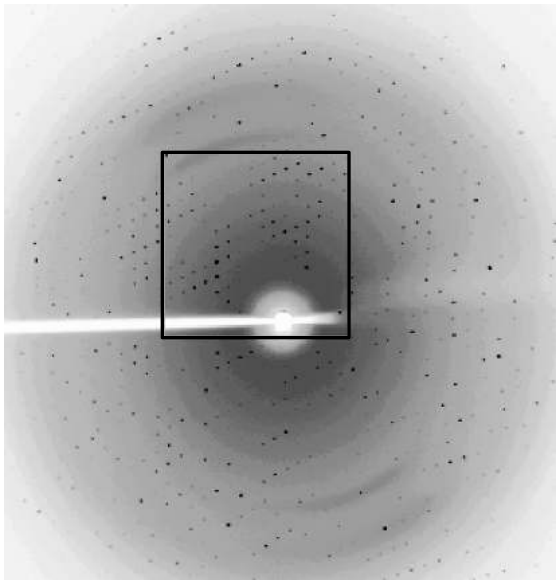
# Example 1: OD-twin

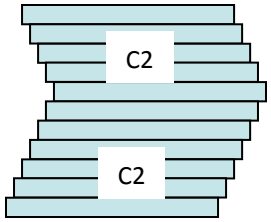
Indexing in C2



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 (arrows).

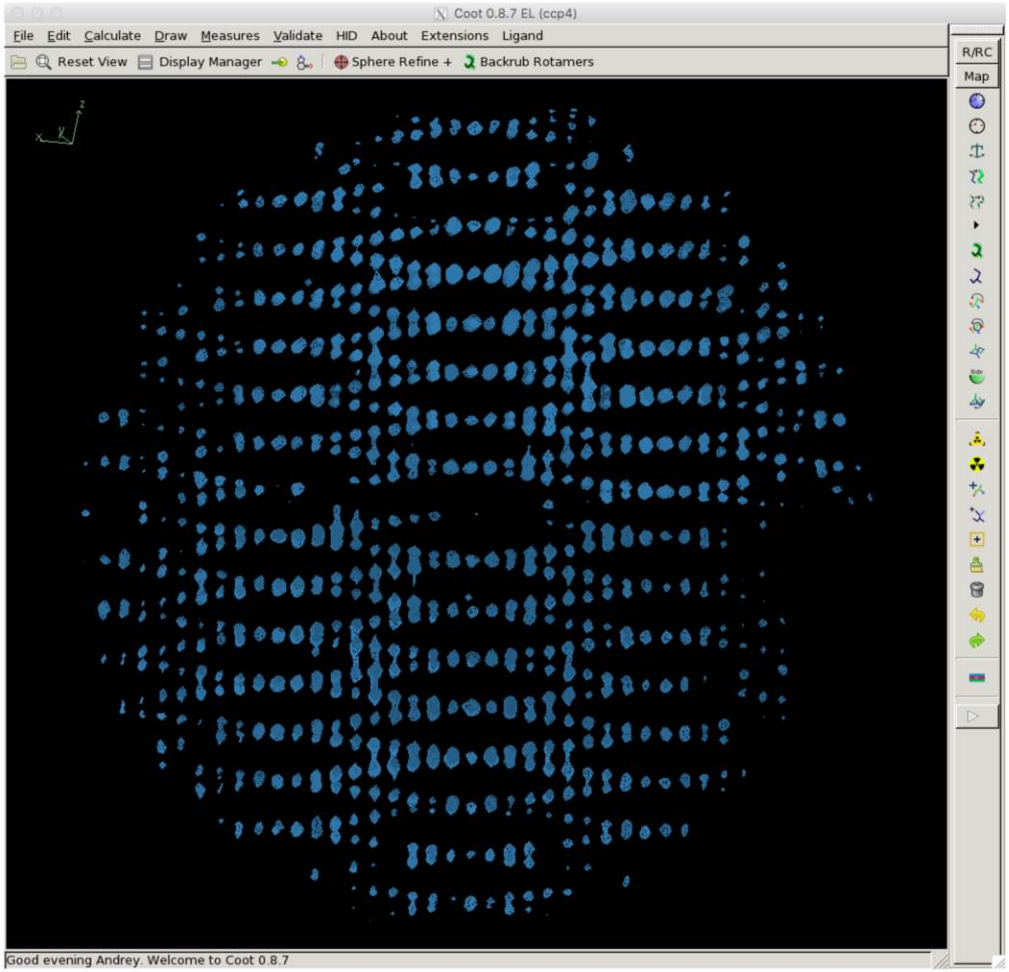
# Example 1: OD-twin

*Morphological classification*
OD-twin

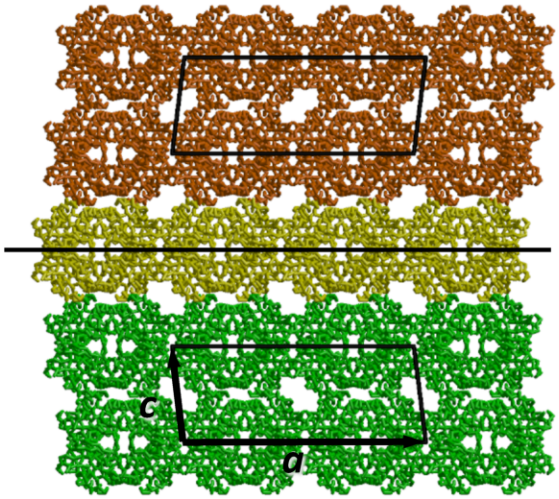
*Geometrical classification*
Twinning by reticular pseudo-merohedry

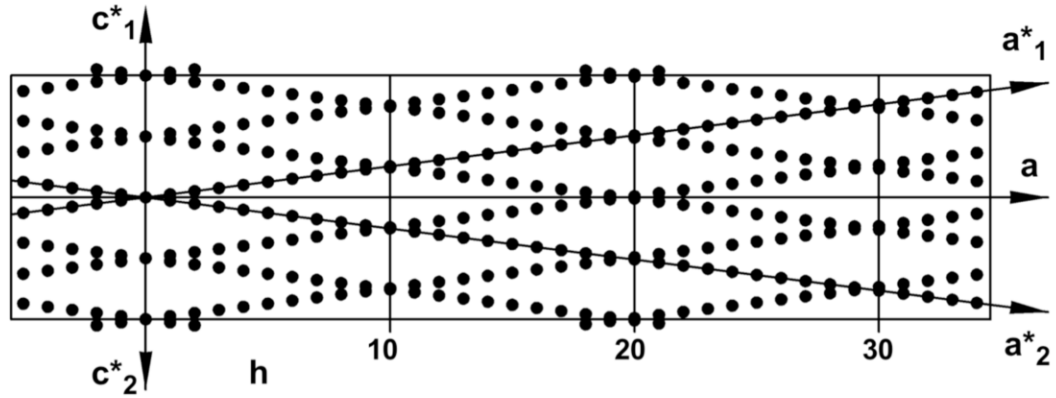
*Synonym*
Non-merohedral twinning

dials.rs_mapper + coot



# Real and reciprocal lattices



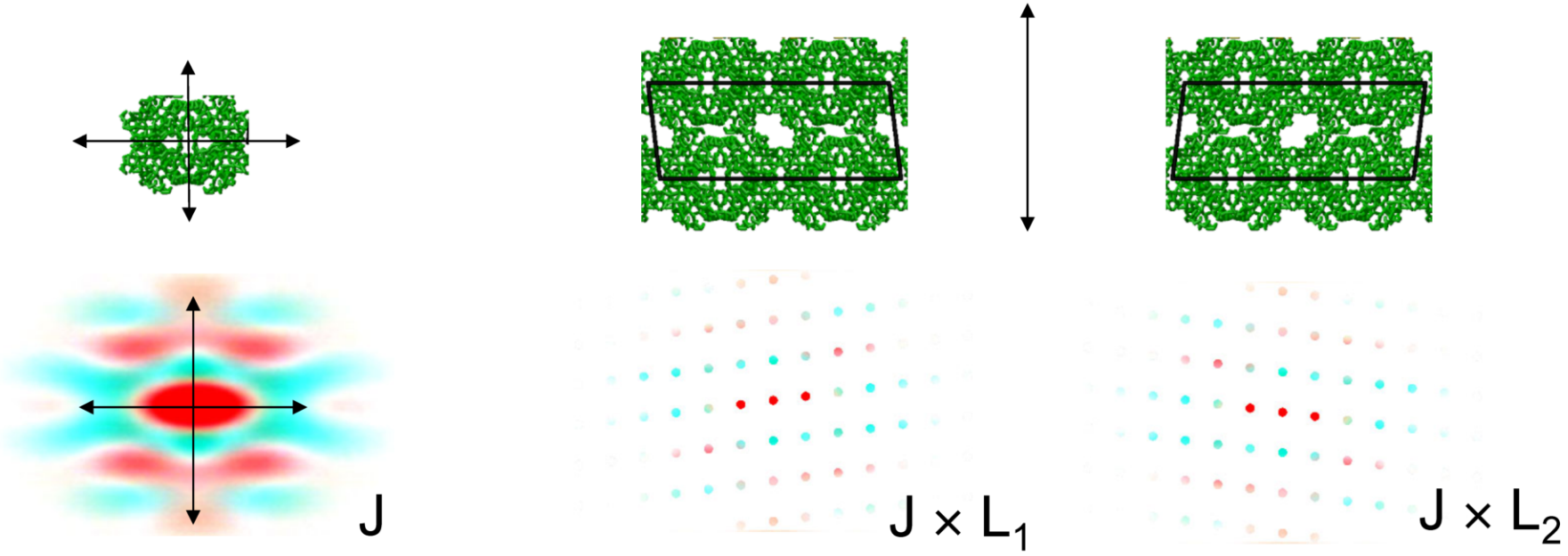




Twinning by reticular pseudo-merohedry (Non-merohedral twinning)

- Process data from one lattice and ignore twinning
- Process data from one lattice and demodulate the data

- Record total intensity of overlapping spots (SAINT, iMosflm) and deal with it at refinement (SHELXL)

# Intensities of the overlapping reflections




Fourier transform of the tetramer

Diffraction pattern of domain 1

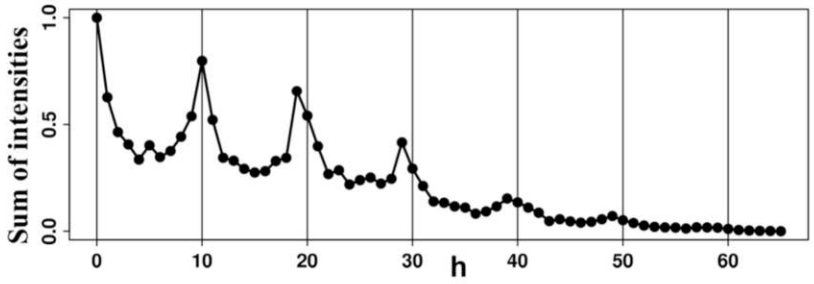
Diffraction pattern of domain 2

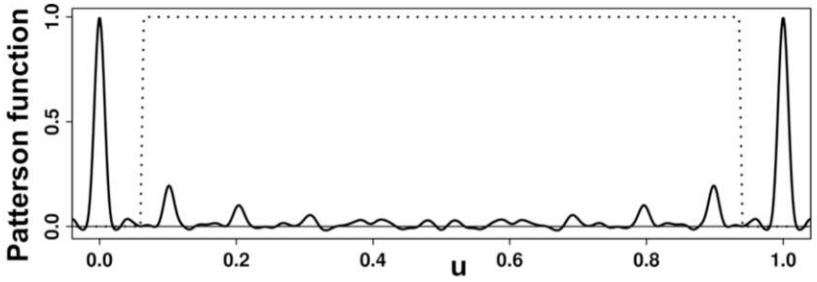
Tetramers in different twin domains are in the same orientation

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

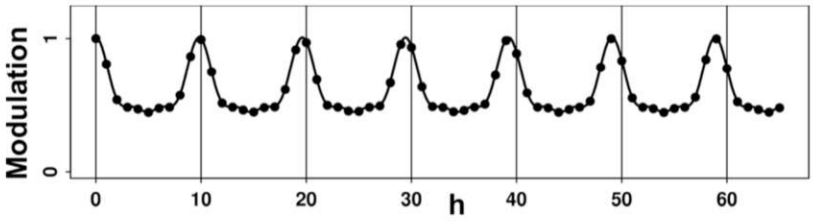
# Demodulation

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



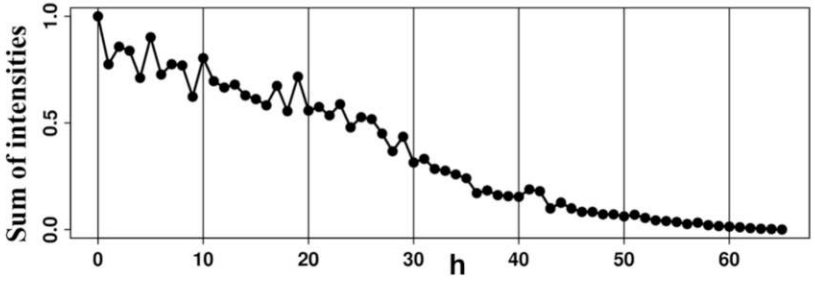



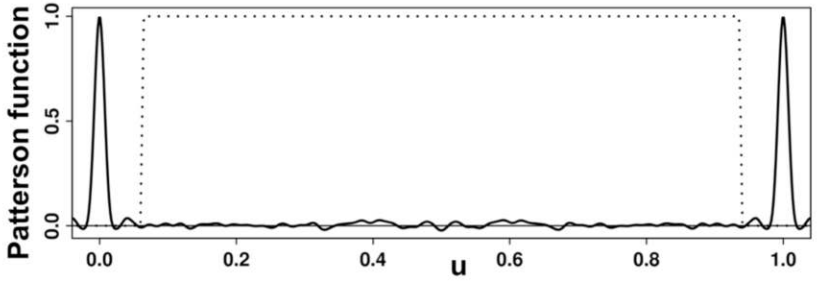

Modulation function




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

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

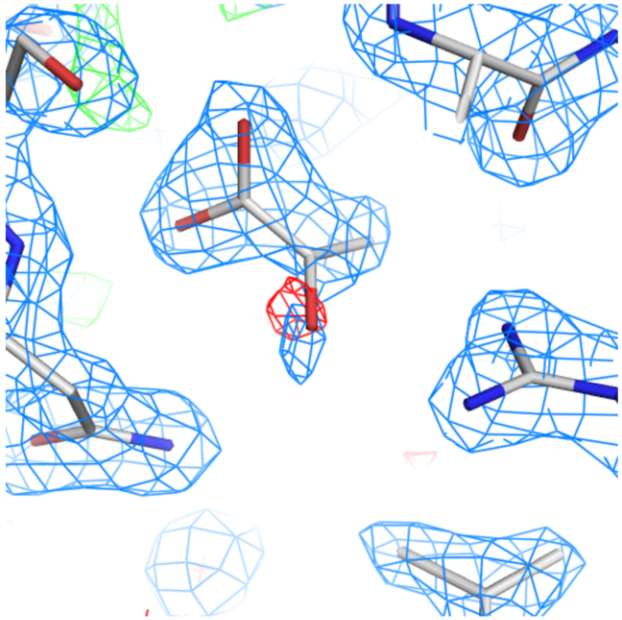


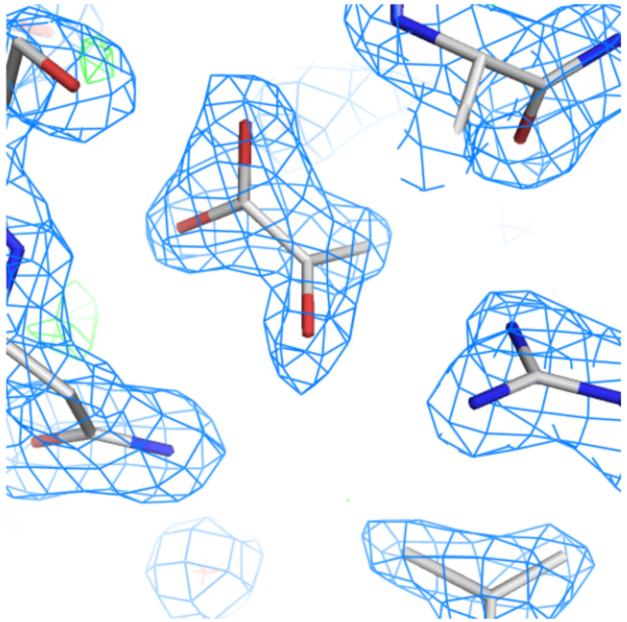
# Improvement in the electron density

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 L-lactate molecule before and after demodulation

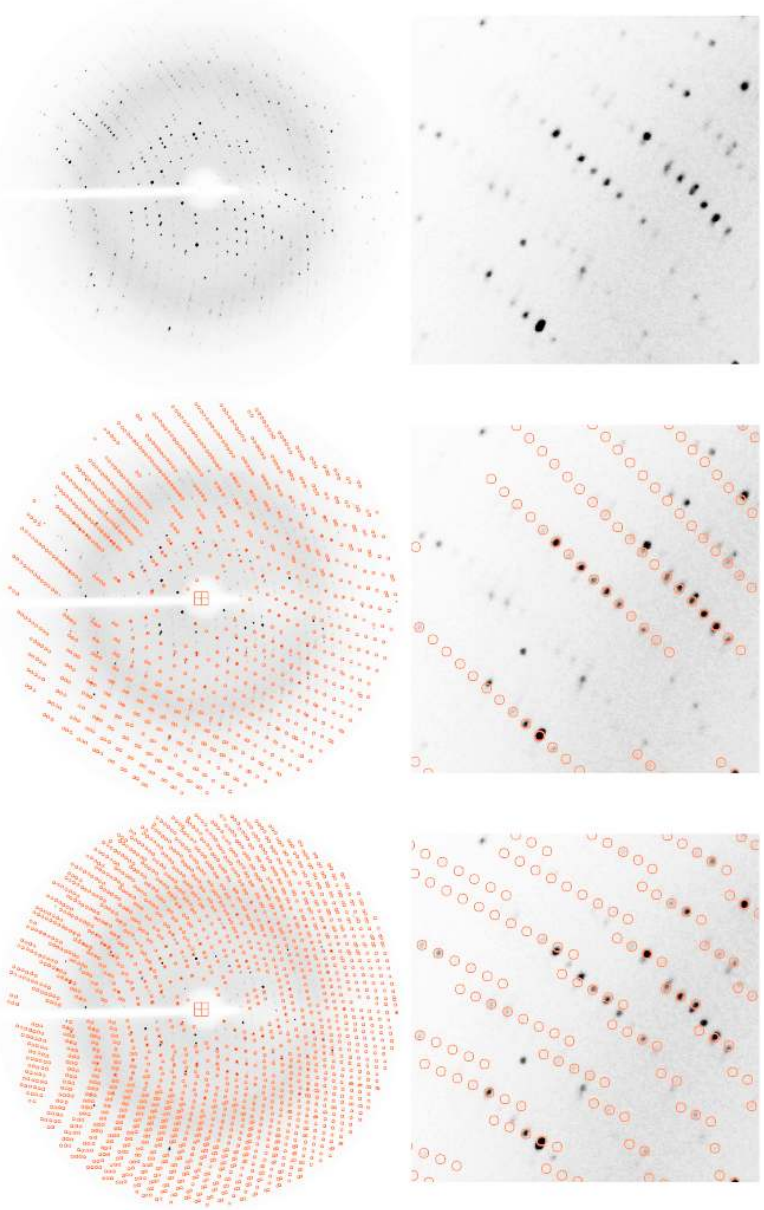


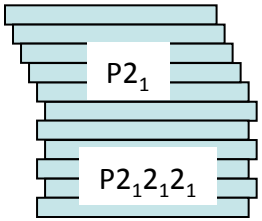
R / R-free = 0.21 / 0.27



R / R-free = 0.16 / 0.23

# Example 2: 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°

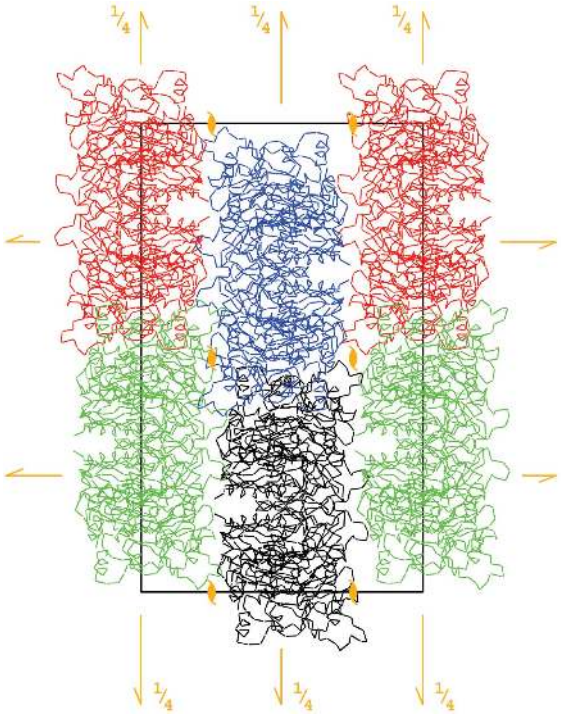
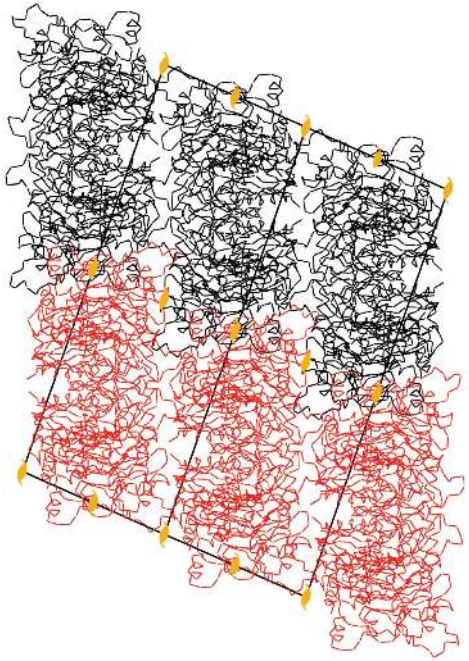
$P2_12_12_1$

a = 86.3 Å
b = 90.6 Å
c = 148.0 Å

# Example 2: 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 |

# Example 2: allotwin

- More frequently, the presence of very different indexing solutions means that the indexing program is struggling rather than domains belonging to different space groups actually exist.

- 3D viewers will help to check what is actually  happening.

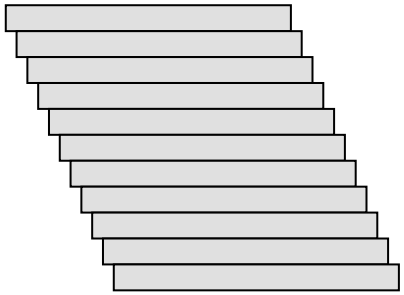
- Merging several fine-sliced images together may help indexing

# Partially disordered OD-structures

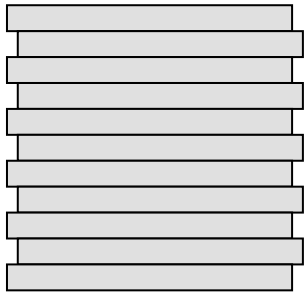
- Visualisation (detection)
- Ghost density
- Indexing
- Effect on structure solution

# OD-structures

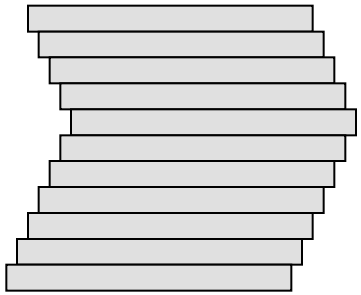
Single crystal



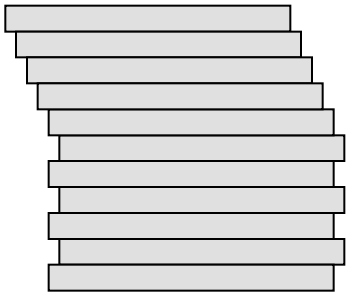
Single crystal



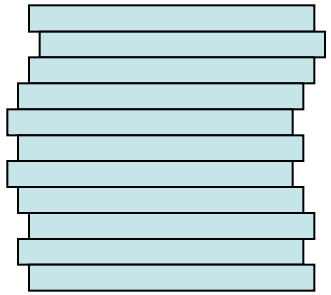
OD-twin



Allotwin

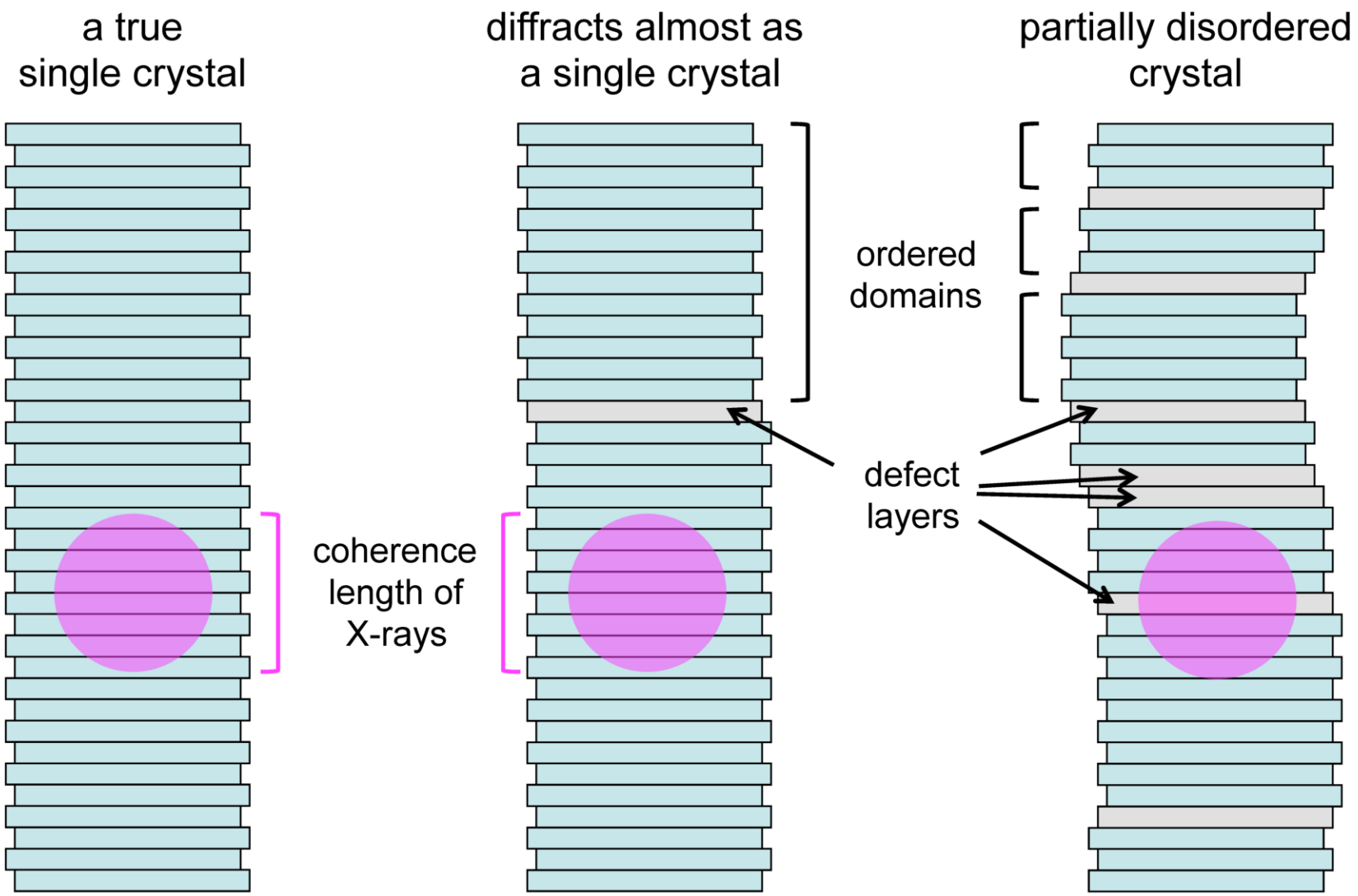


Partially disordered OD-structure



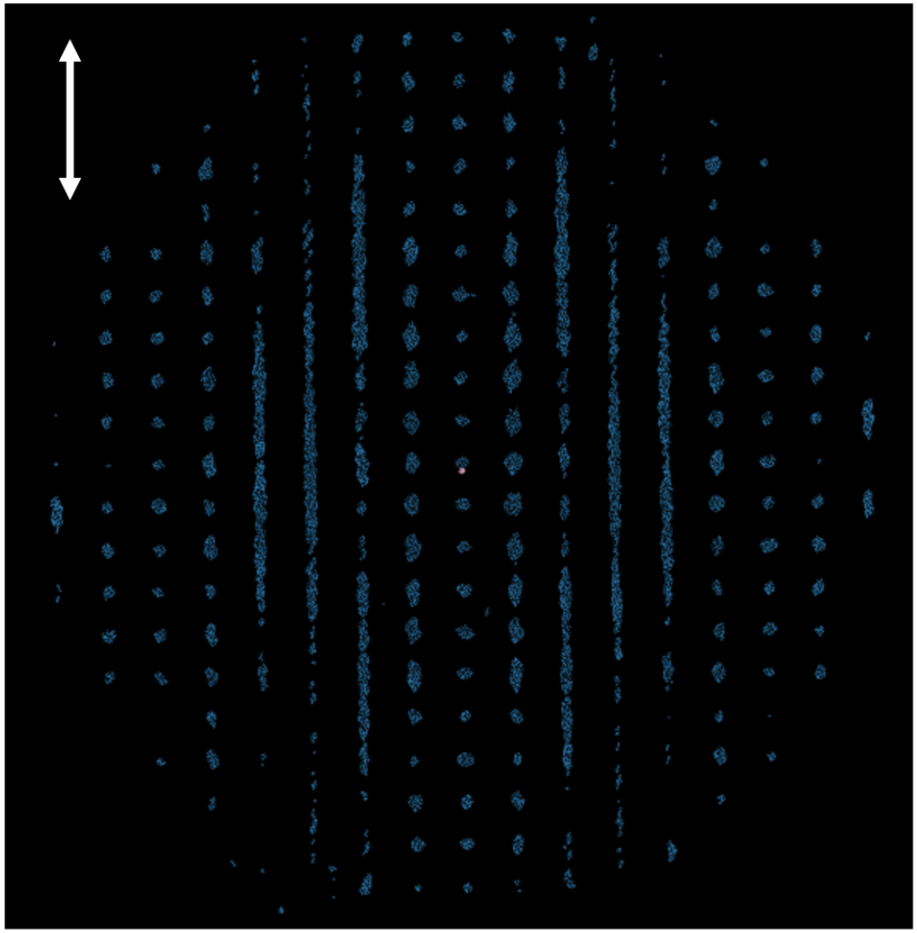
*Examples 1,2 & 3*

# Partially disordered OD structures

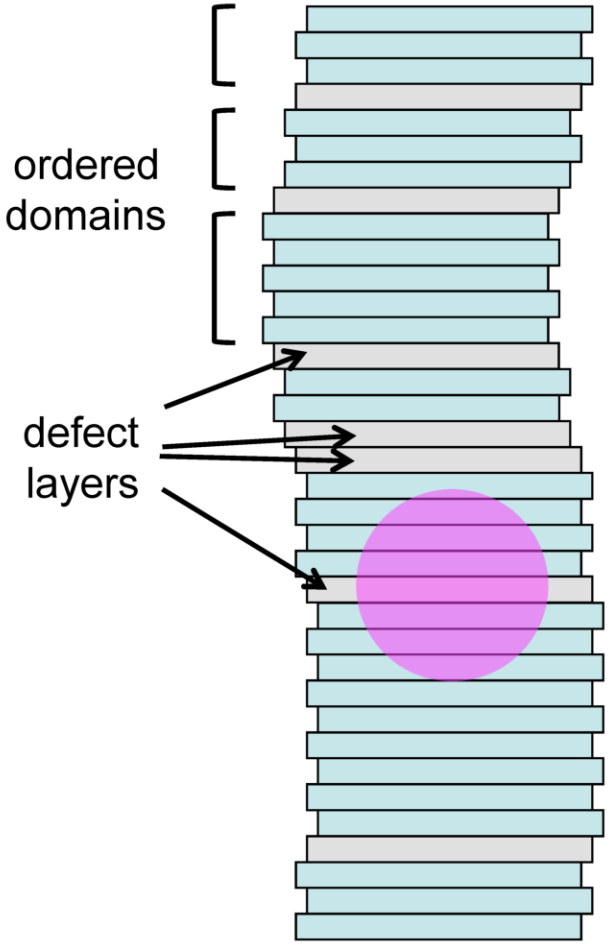


# Diffraction of partially disordered structures

White arrow - direction in which global periodicity is missing

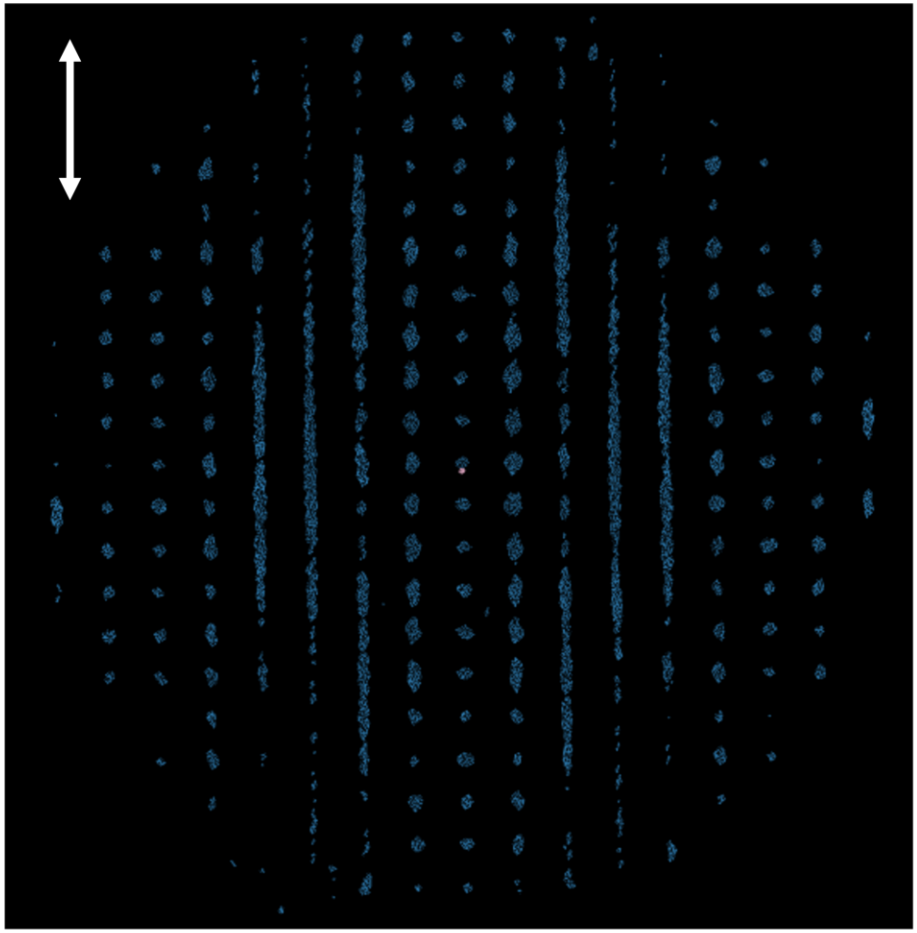


dials.rs_mapper + coot



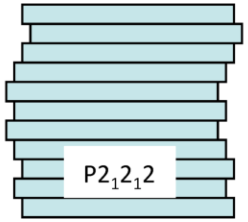
# Example 1: ghost density

White arrow - direction in which global periodicity is missing



dials.rs_mapper + coot

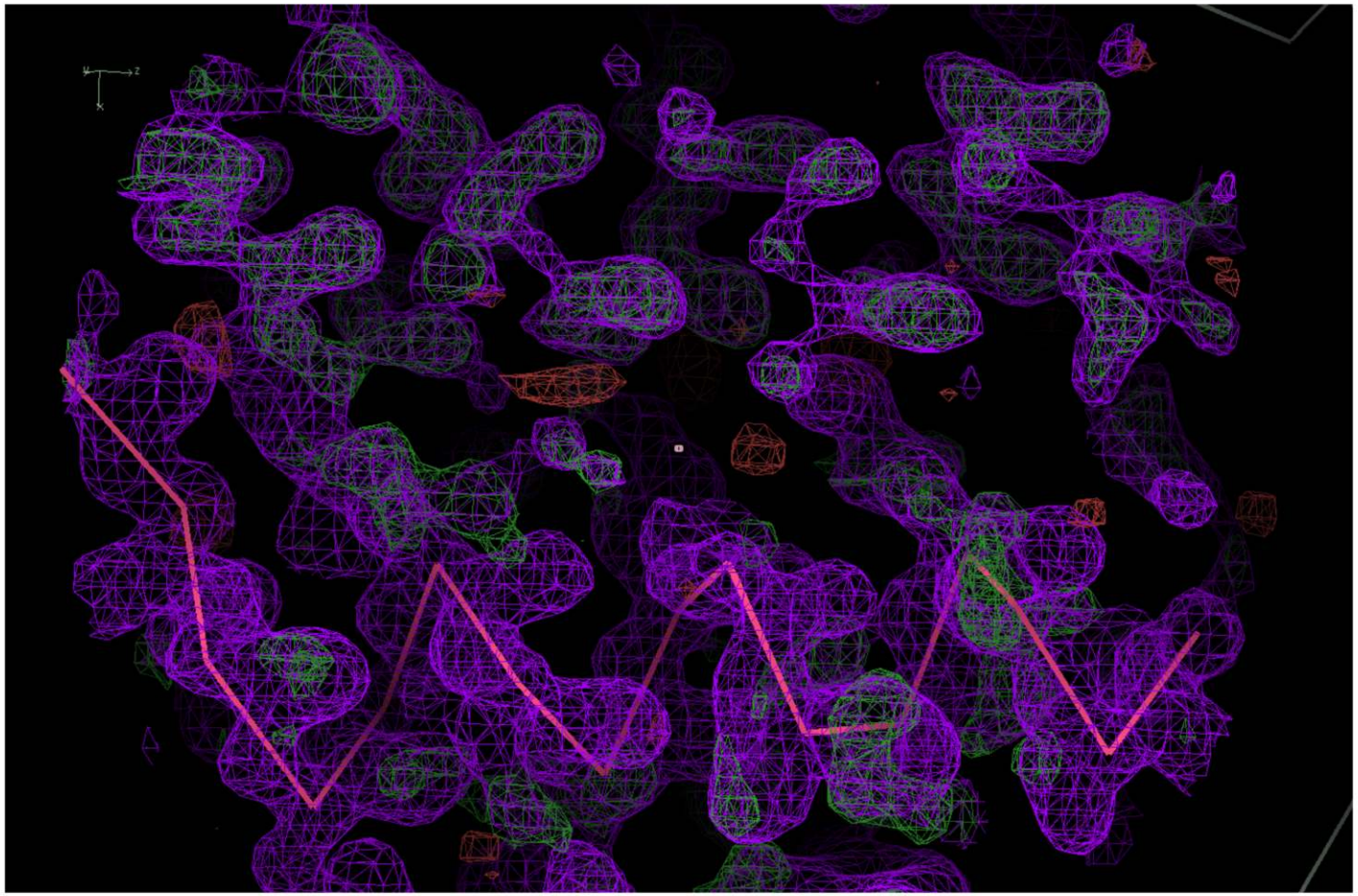
An example from **Rafael Ciges**, Biomedical Institute of Valencia



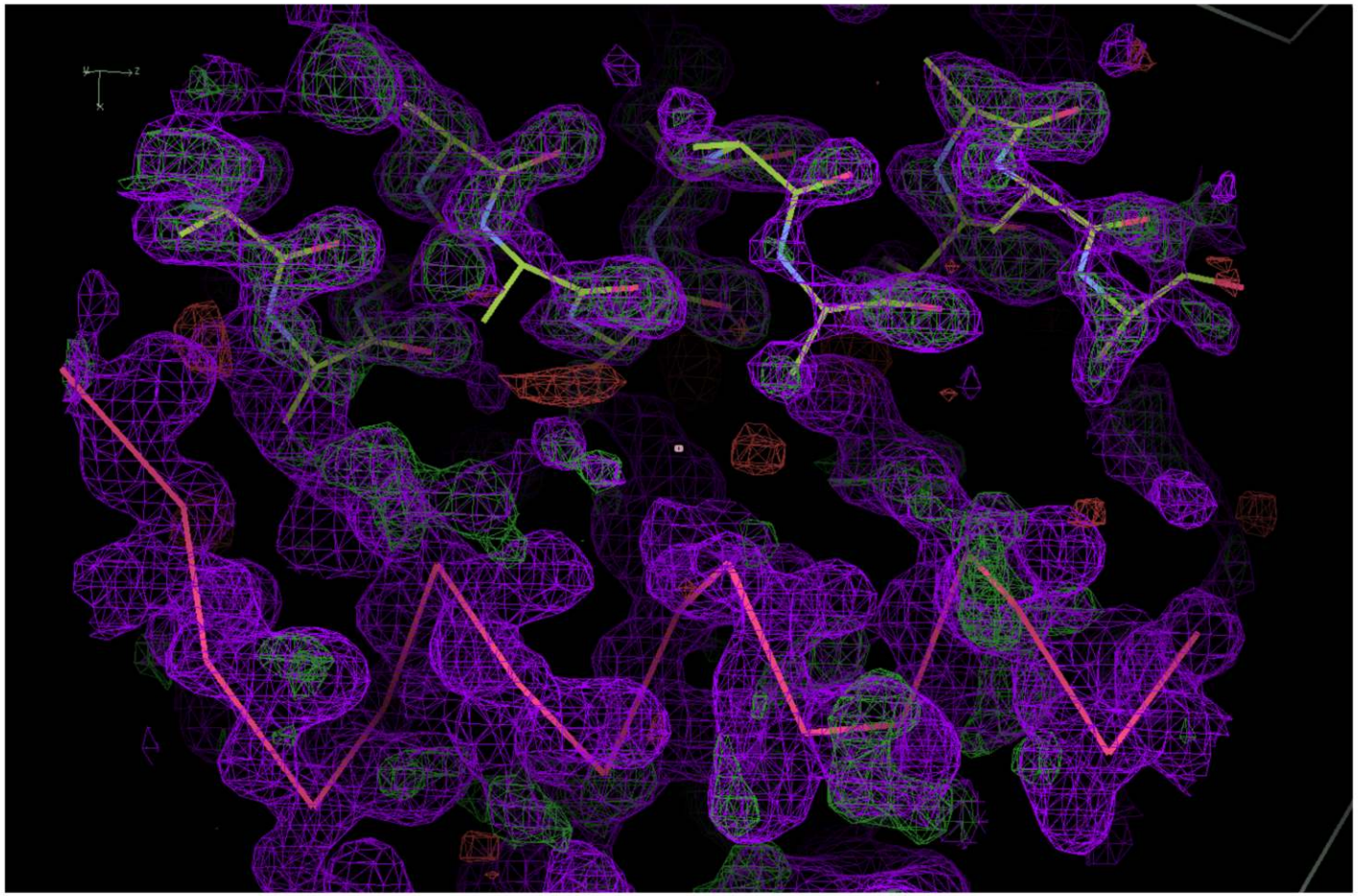

- Space group $P2_12_12$
- Resolution 1.2Å

- The diffraction images were processed with XDS
- Structure was solved with MR
- Preliminary refinement $R_{free}$ = 0.35
- Extra residues were expected compared to MR model

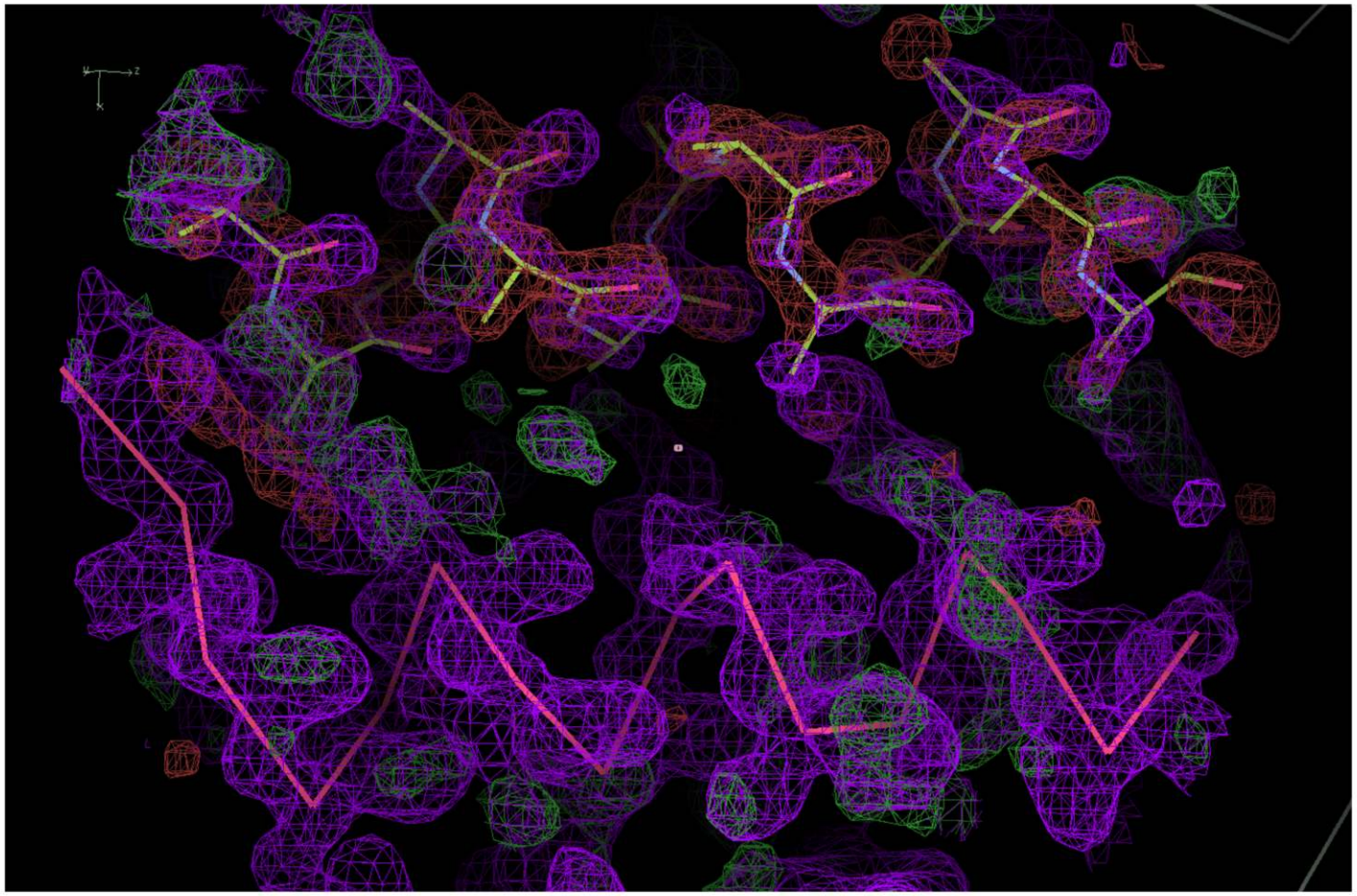
# Example 1: after initial refinement



# Example 1: helix added



# Example 1: after refinement with extra helix

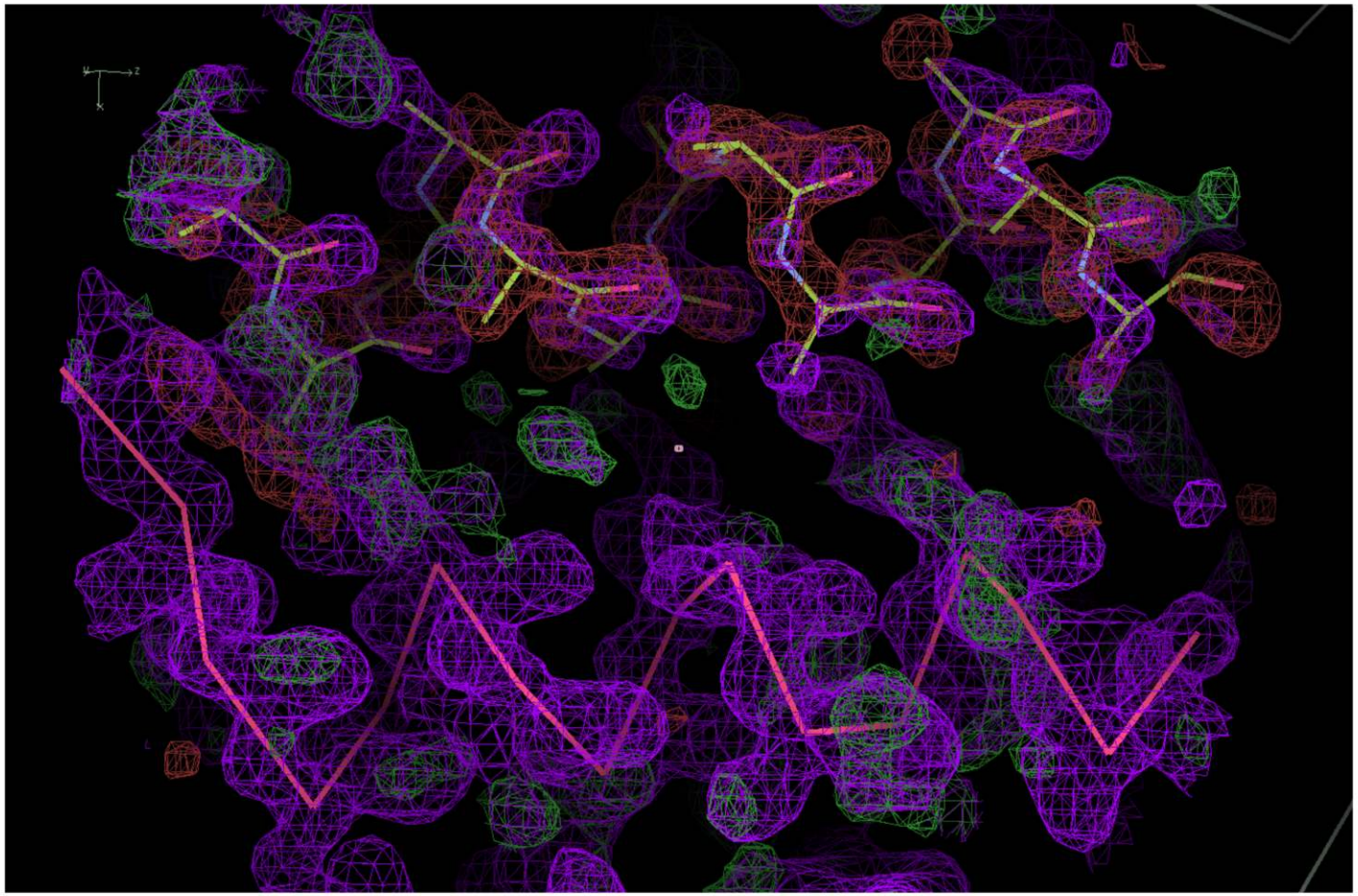


# Example 1: demodulation of intensities

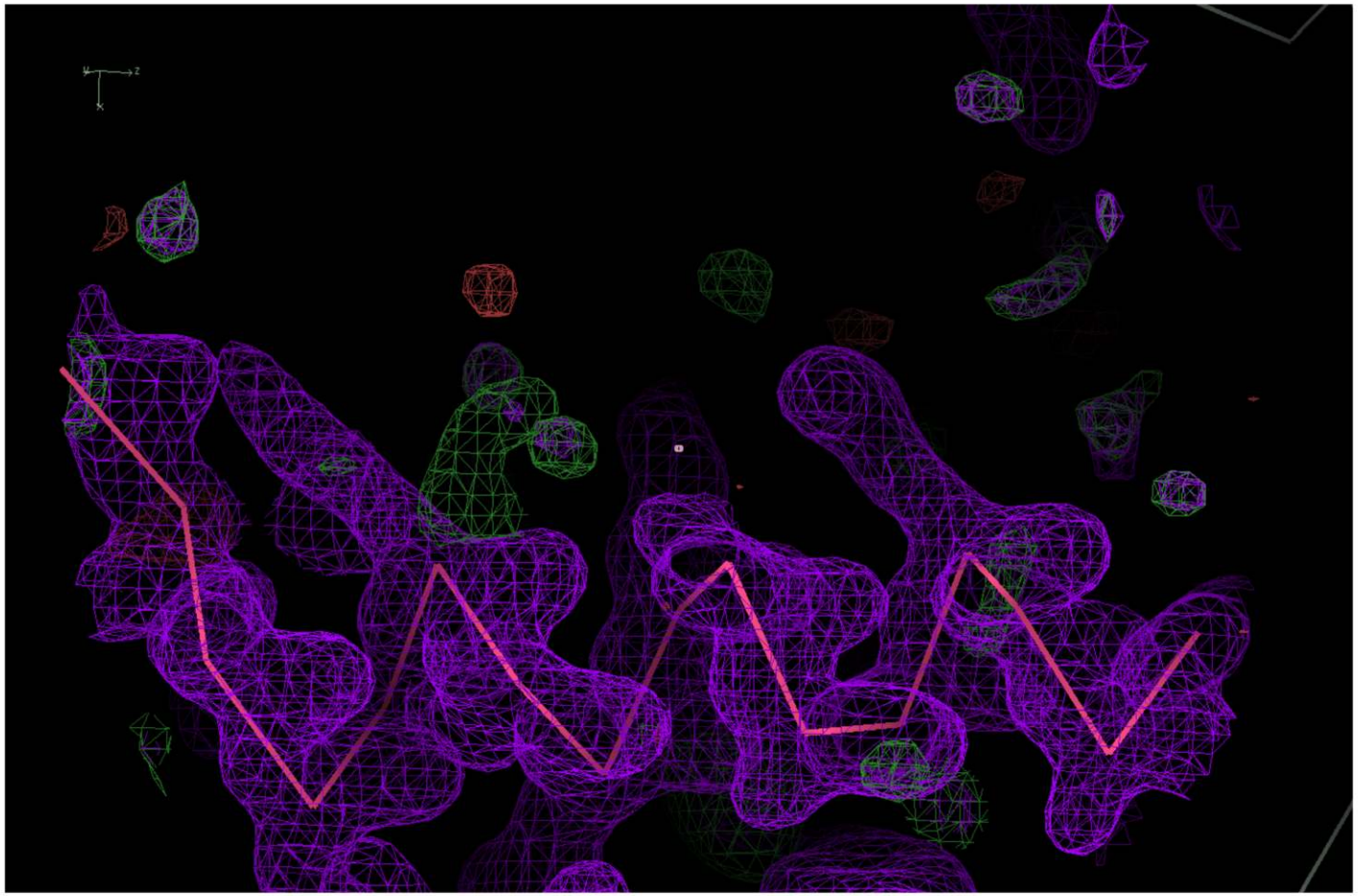
- Data were demodulated and structure re-refined
  - » demodulation procedure was conceptually similar to the one used in the OD-twin example

| | R | R-free |
|---|---|---|
| Original data | 0.33 | 0.34 |
| Corrected data | 0.25 | 0.26 |

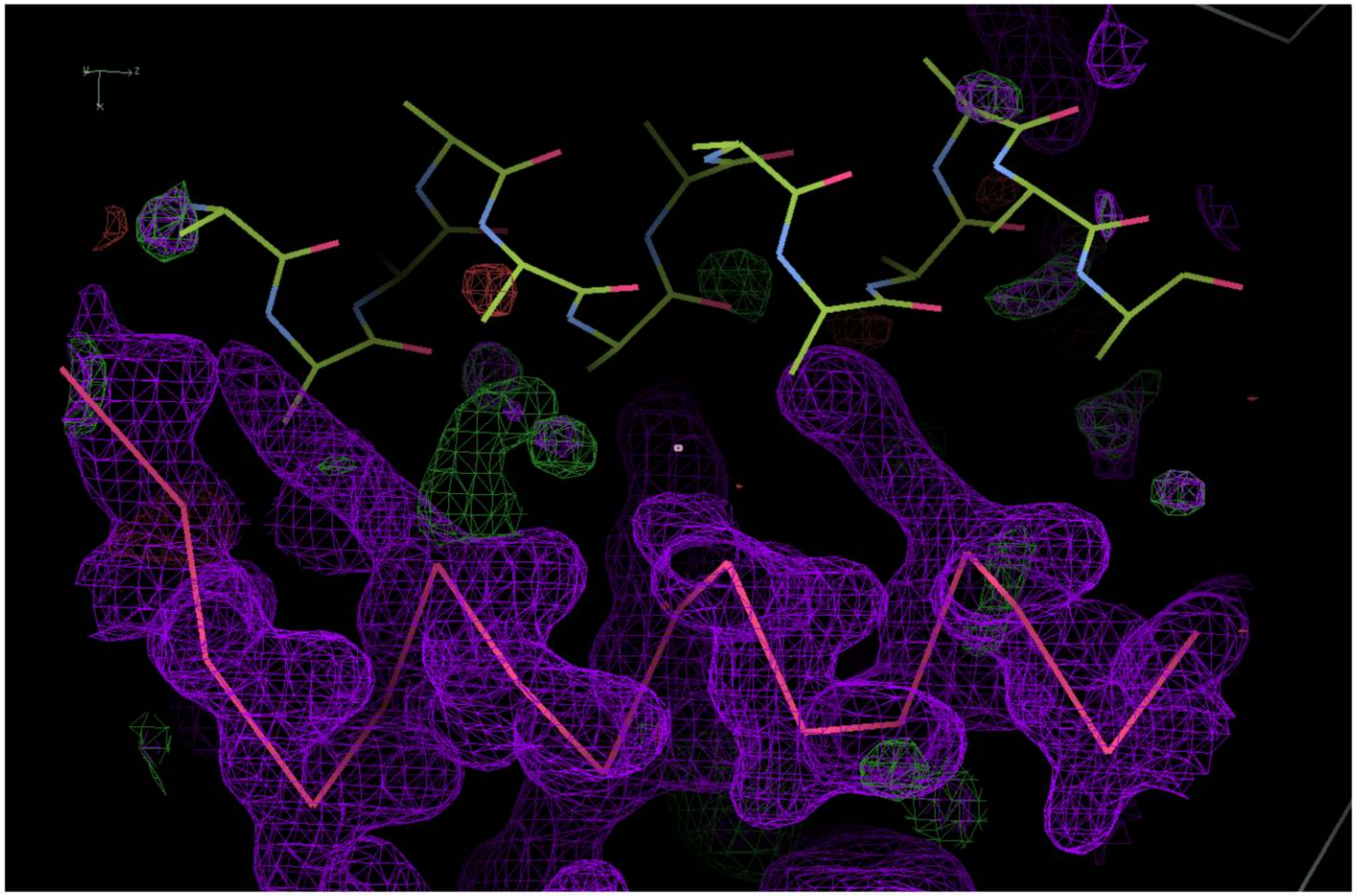
# Example 1: after refinement with extra helix



# Example 1: after refinement against demodulated data ...



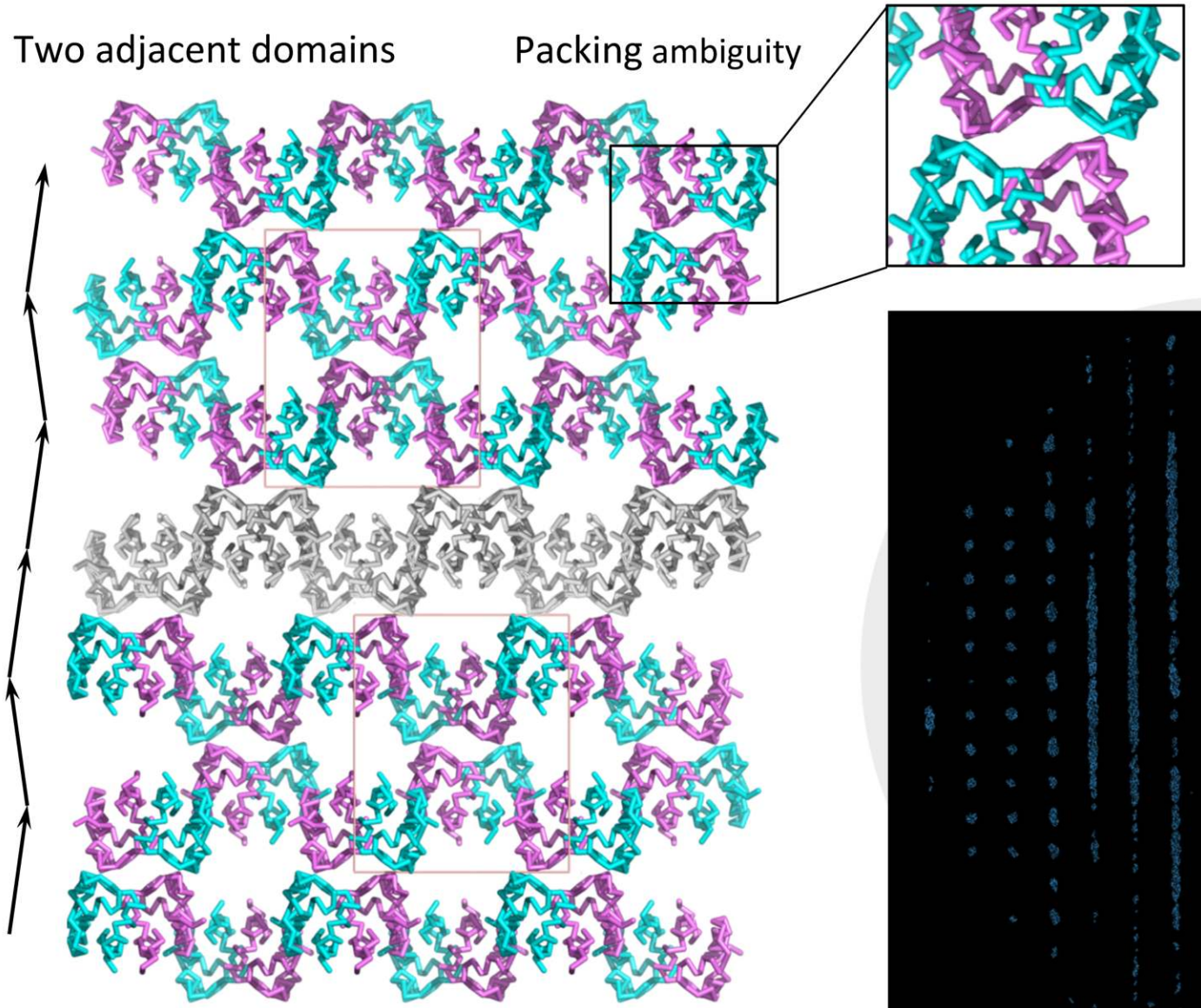
# Example 1: ... there is no ED for the extra helix



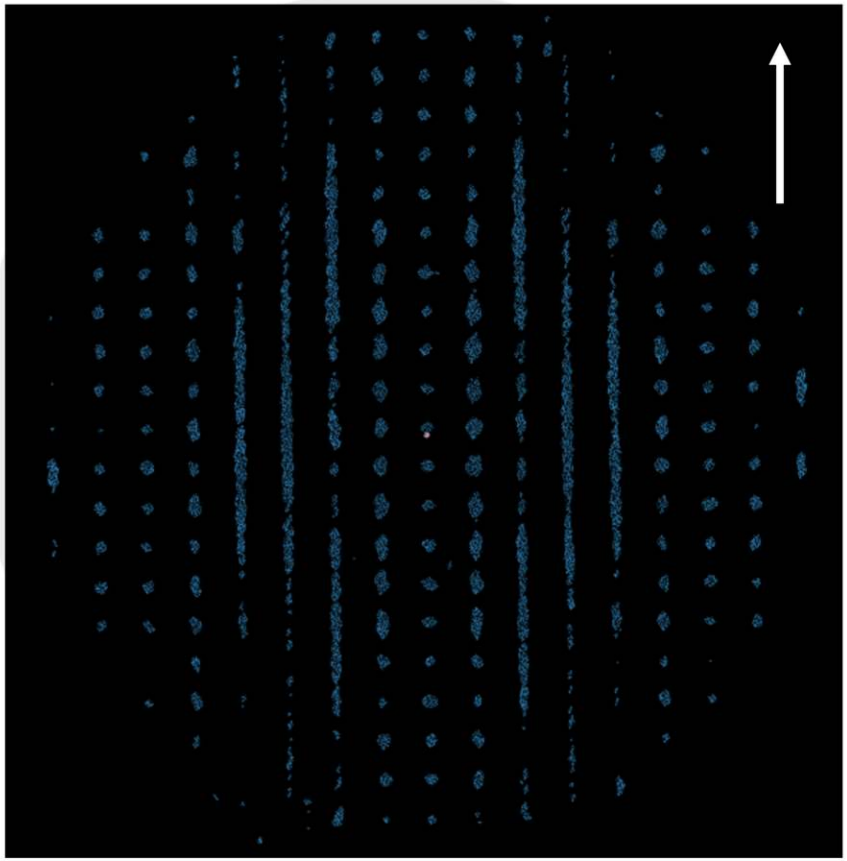
# Example 1: ghost density

Two adjacent domains

Packing ambiguity



White arrow - direction in which global periodicity is missing



# Example 1: Summary

- Partial disorder in OD structures results in a ghost density
- Structure can be solved and refined ignoring partial disorder
- Demodulation procedure may remove ghost density and therefore help with interpretation of the ED maps
  - » Not always badly needed and not always works
  - » There are several bespoke scripts around
  - » A general automated software solution is needed

# Example 2: auto-indexing failure

Fast DP @ DIAMOND

$R_{meas}$ = 0.12 ?

CC(1/2) = 0.3 at 1.56 Å

Refinement

$R_{cryst}$ = 0.33 ?

$R_{free}$ = 0.36

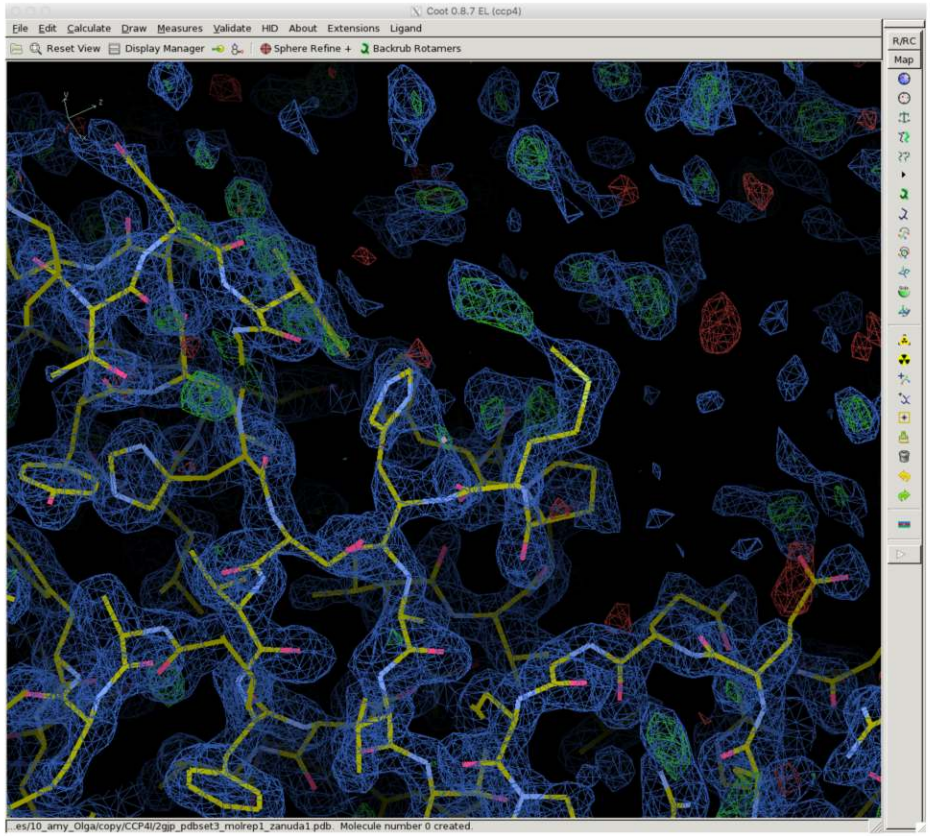
Molecular Replacement ✓

--- Peaks of Rotation Function ---

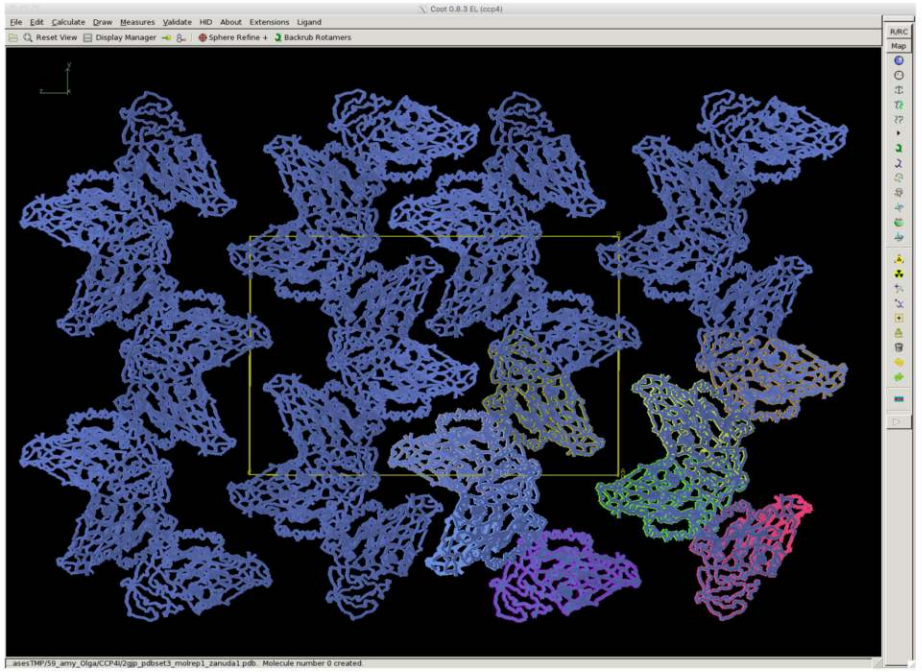
| | theta | phi | chi | Rf/sigma |
|---|---|---|---|---|
| 1 | 63.62 | 174.24 | 148.98 | 13.70 |
| 2 | 80.19 | -58.05 | 61.61 | 13.63 |
| 3 | 149.48 | -148.30 | 170.26 | 13.34 |
| 4 | 107.22 | 84.22 | 129.22 | 13.04 |
| 5 | 87.46 | 75.99 | 136.16 | 12.18 |
| 6 | 111.97 | -14.20 | 175.28 | 12.10 |
| 7 | 157.20 | 173.73 | 153.99 | 11.25 |
| 8 | 58.77 | -96.16 | 51.96 | 11.24 |
| 9 | 75.76 | -63.11 | 54.46 | 6.21 |
| 10 | 102.46 | 82.67 | 133.90 | 5.83 |

# Example 2: evidences of wrong indexing

Maps ?




Crystal Packing ✗

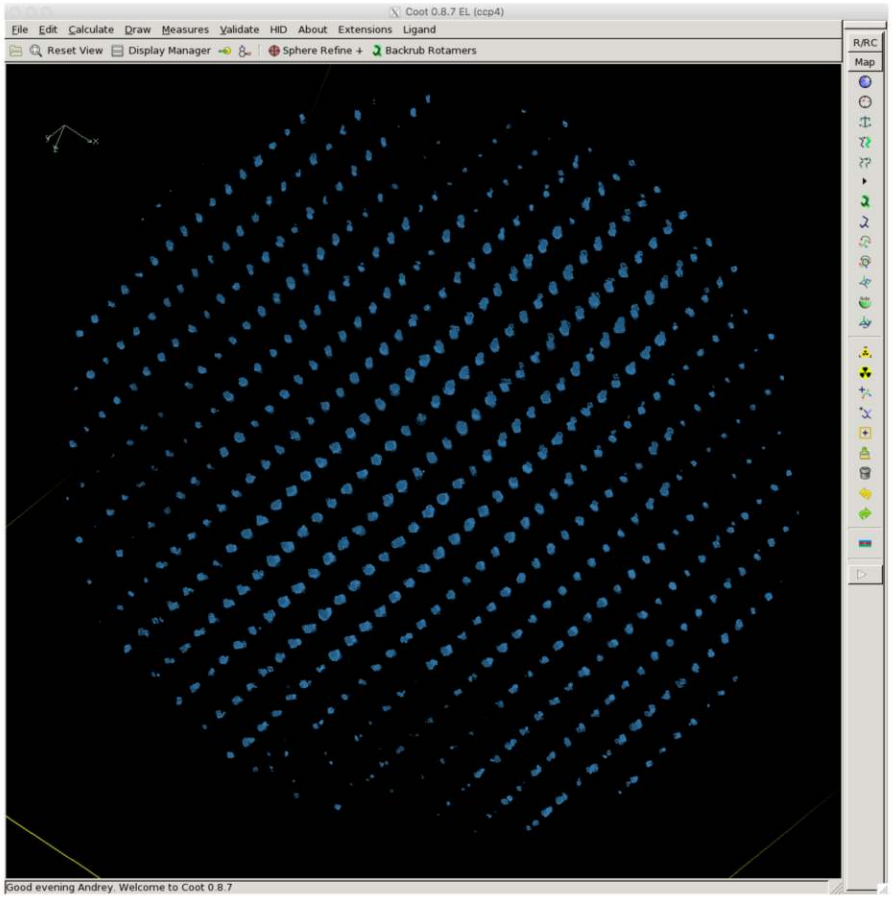



Wrong indexing?

# Example 2: evidences of partial disorder

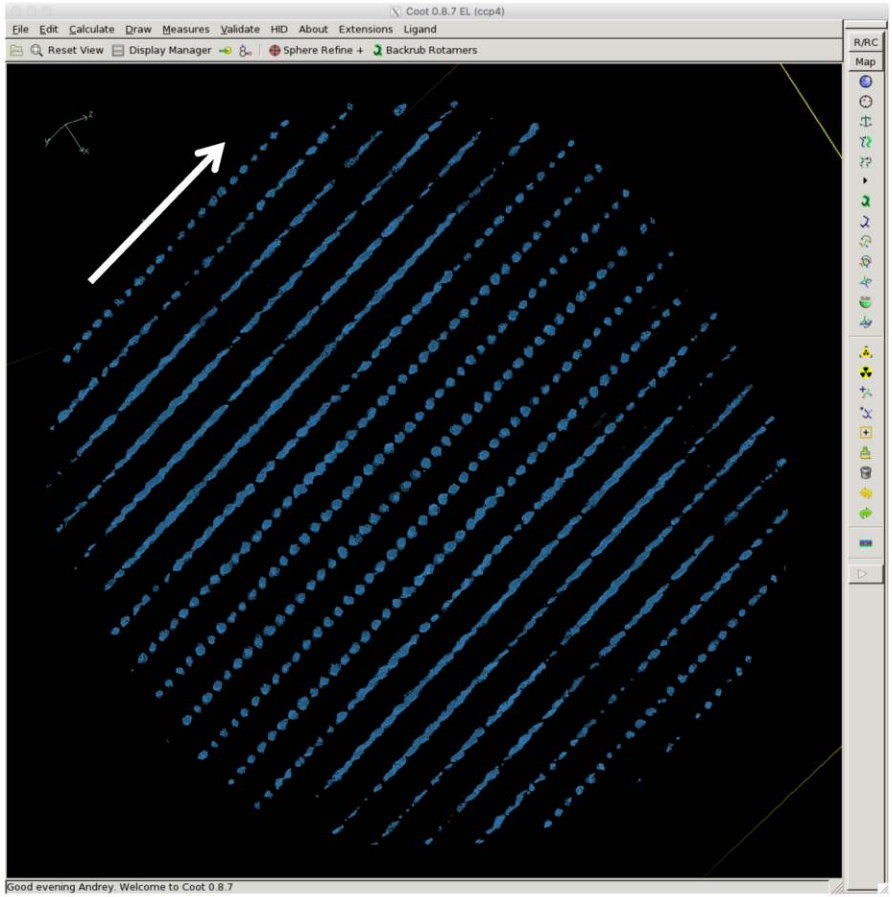
## front view

There is global 2D translational symmetry in the plane of figure

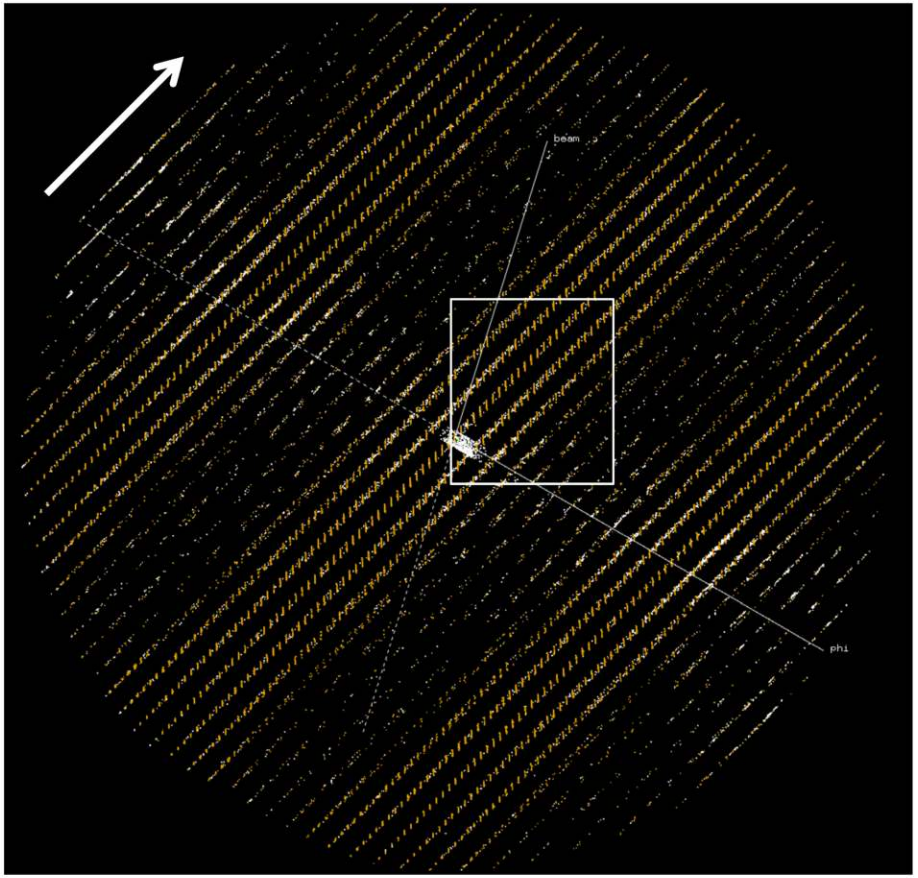



## side view

White arrow indicates direction in which translational symmetry is not global (only within individual domains)



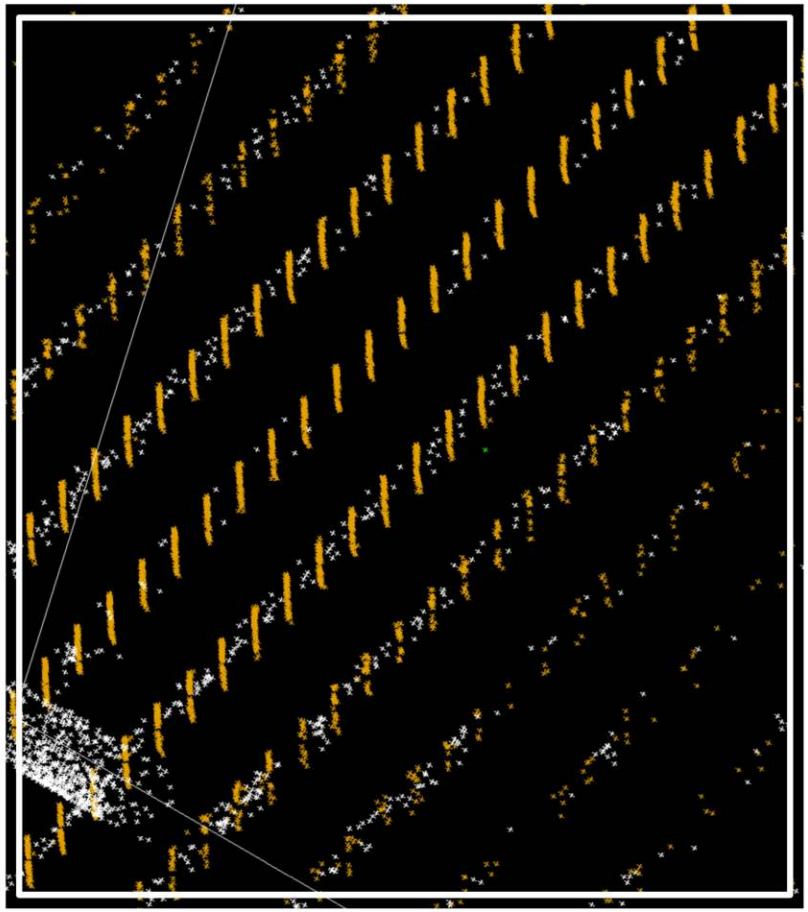
# Example 2: correct indexing



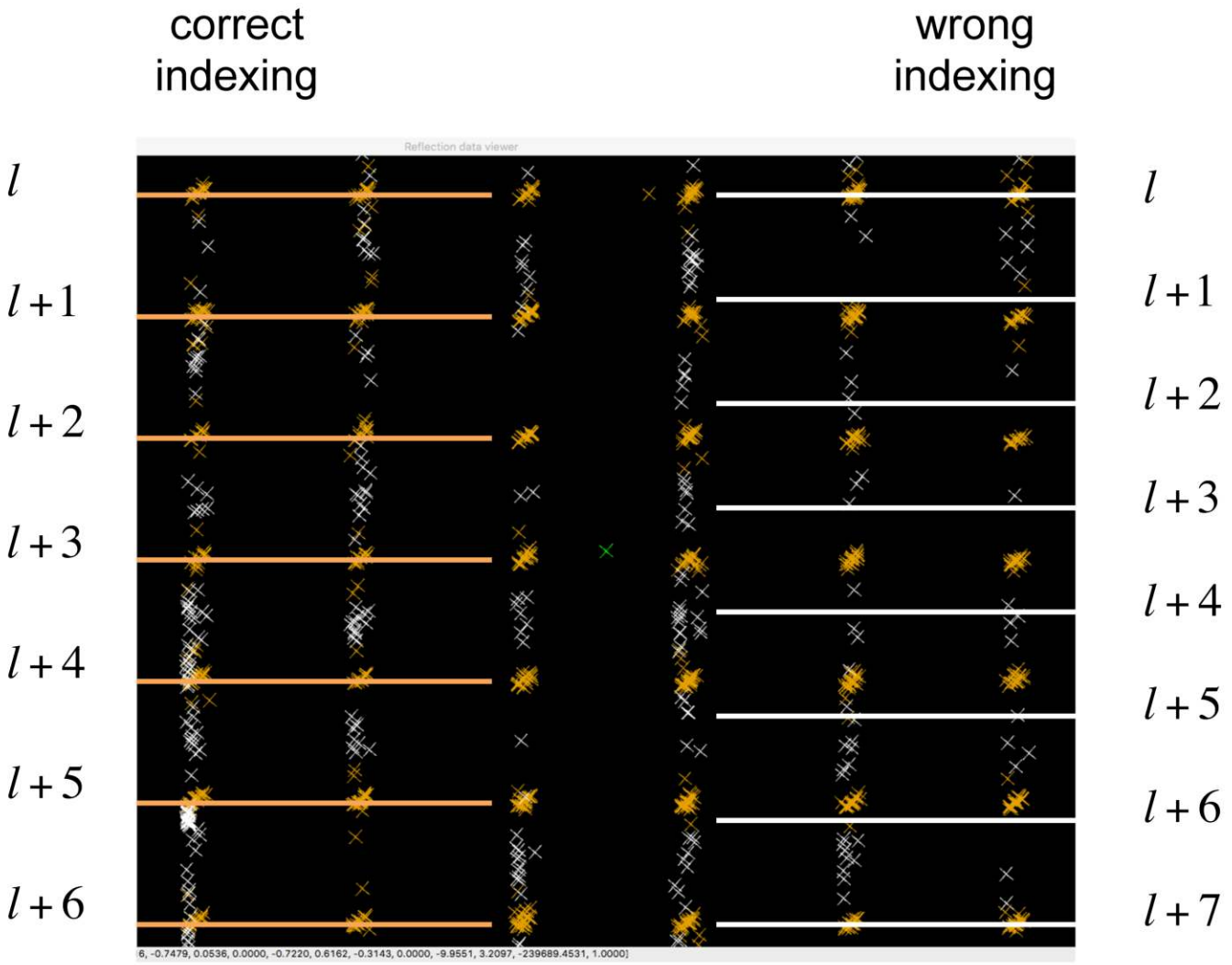

White arrow indicates direction in which translational symmetry is not global (only within individual domains)
There are also areas with less spots

White “spots” are not indexed; actually, these are tails of diffuse reflections

Indexing program may take them for real spots and produce wrong result.

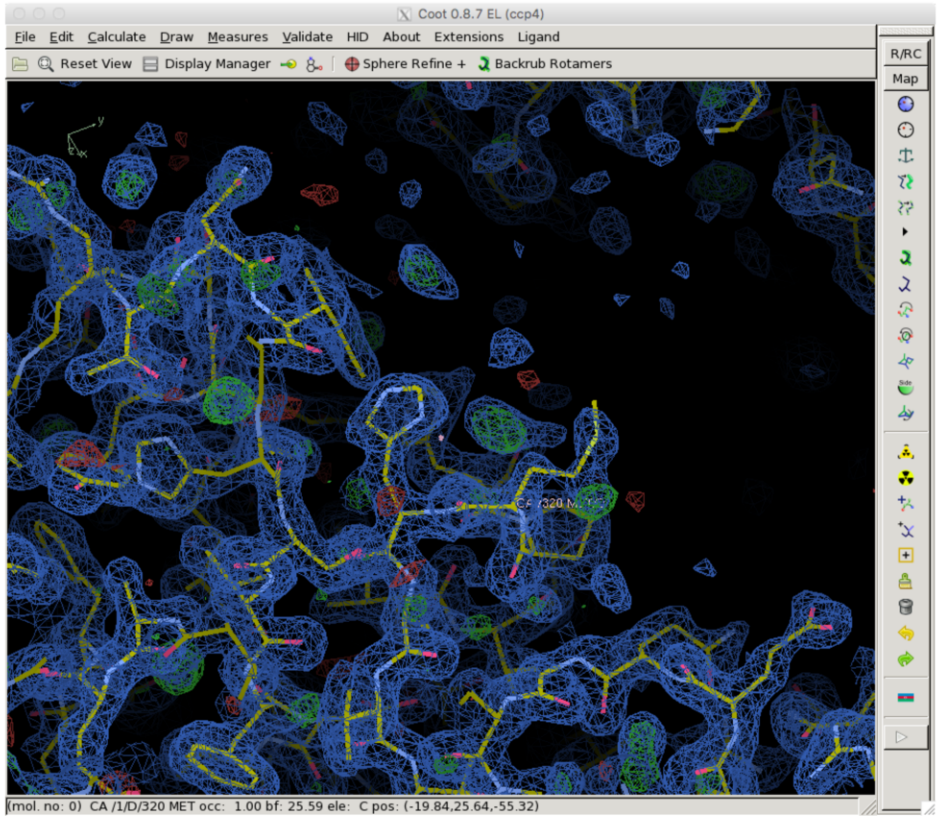


# Example 2: what initially was wrong

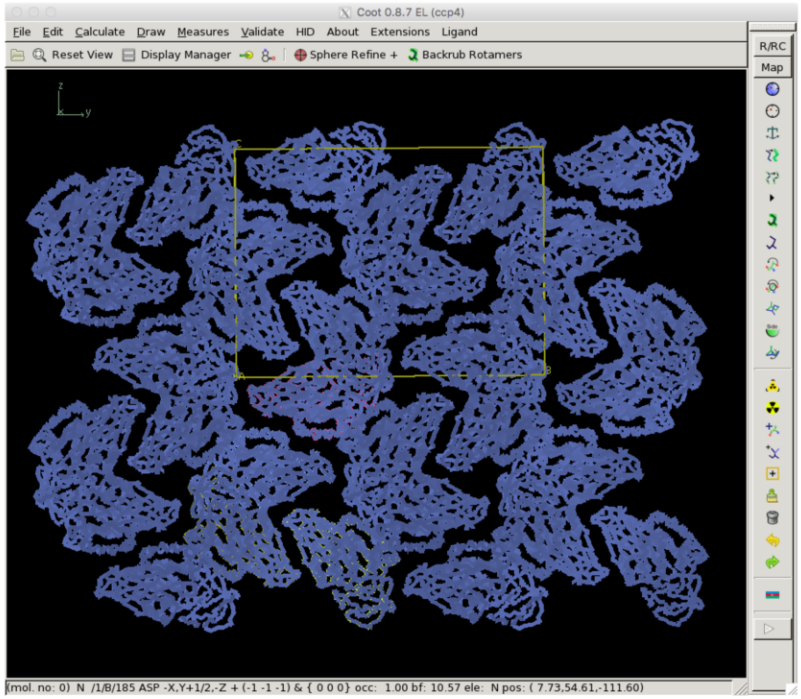


# Example 2: happy end

Maps ✓




Crystal Packing ✓

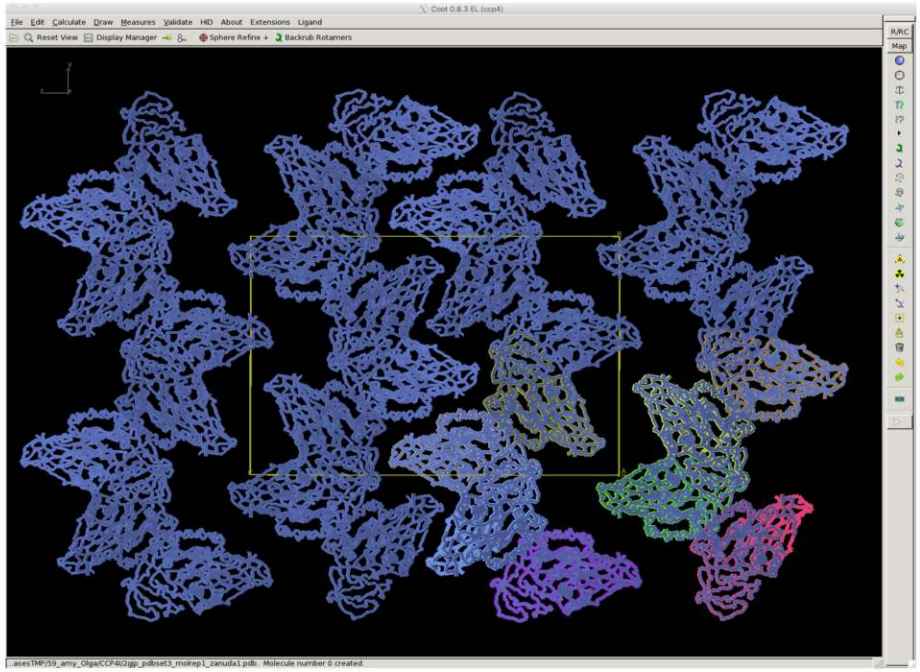


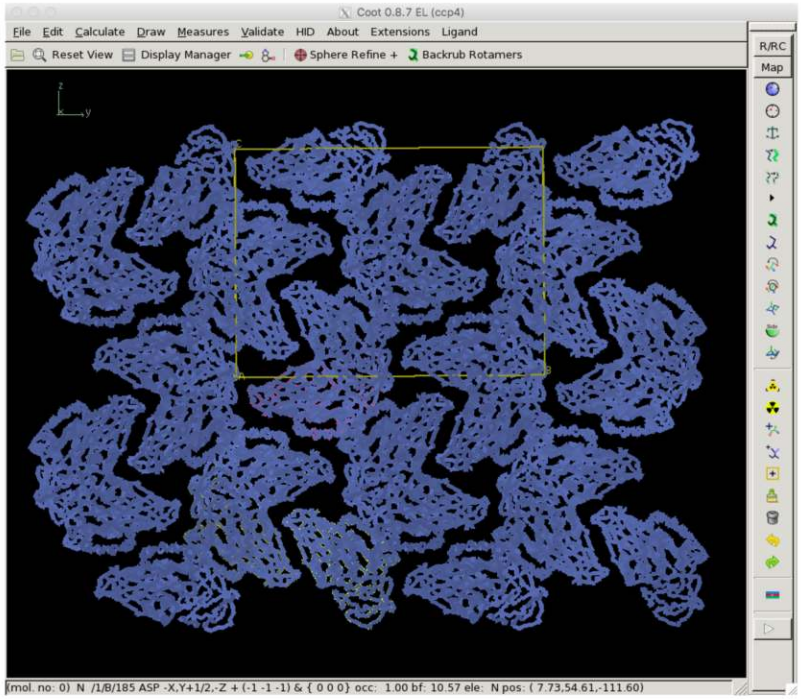

Refinement ✓

$R_{cryst}$ = 0.23

$R_{free}$ = 0.26

# Example 2: wrong and correct



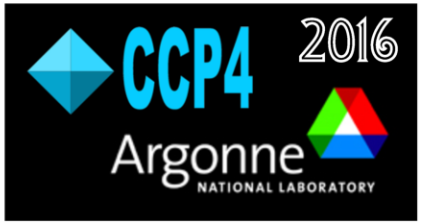
# Example 2: Summary

- Partial disorder a frequent reason of indexing failure
- Use 3d viewers for diagnostics
- Warning: high contrast in MR can be obtained even for wrongly indexed data provided that the search model is highly similar to the target
- Molecular replacement is quite tolerant to partial crystal disorder
  - » Especially RF
  - » In the next example this property of RF will be utilised

# Example 3: unsolvable structure

Input information:

Example from Rui Wu

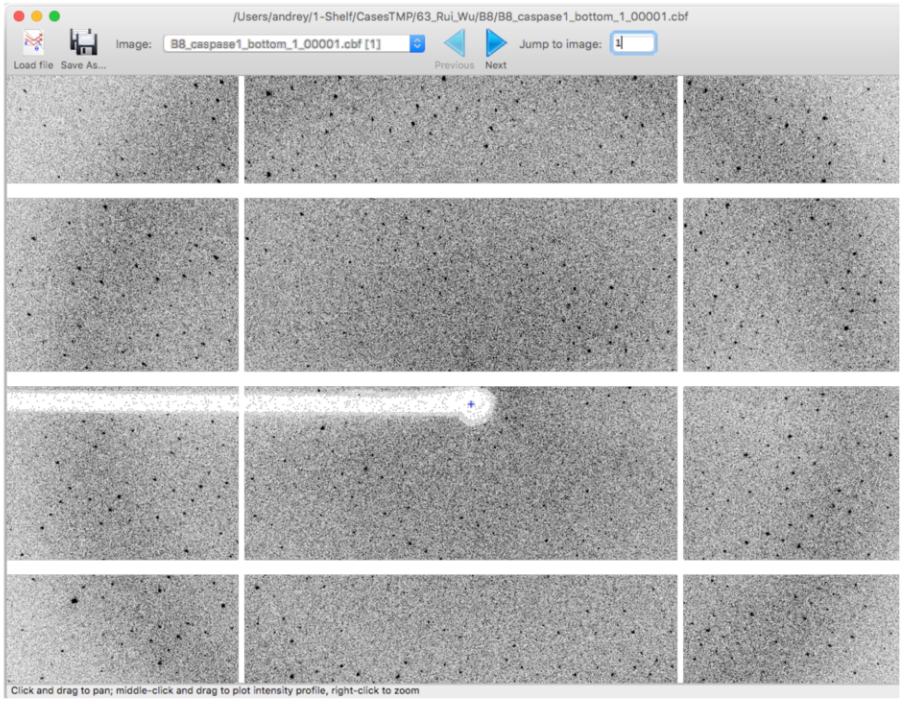



- Images are good
  - But there a several different indexing solutions
- 99% homologue for Molecular Replacement
  - But no MR solution
  - Even more, no contrast on Rotation Function
- Twinning?

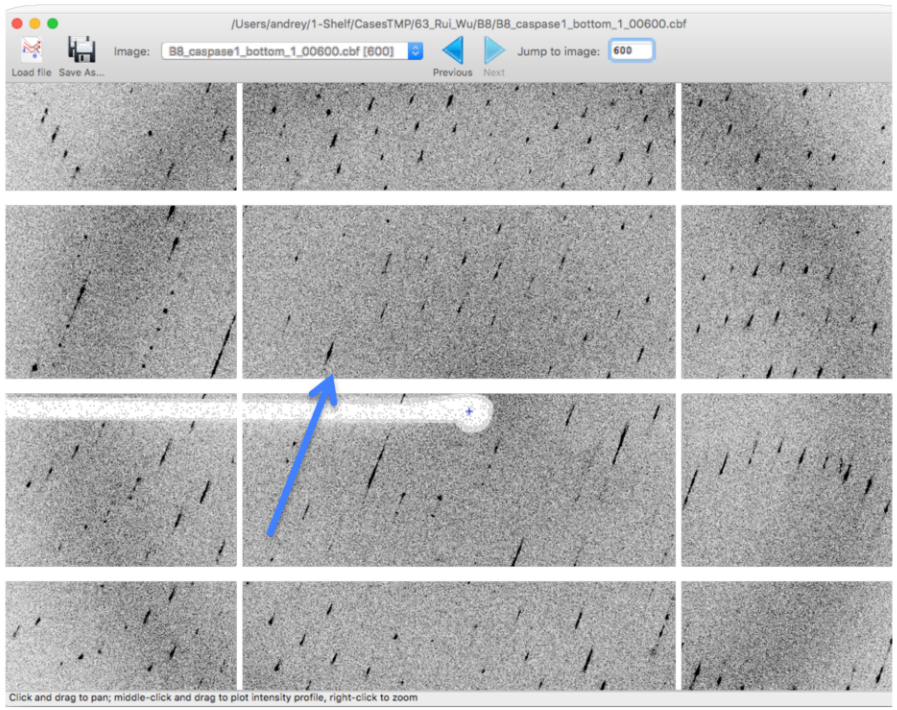
# Example 3: first and last images

Partial disorder was not detected directly from images

first image



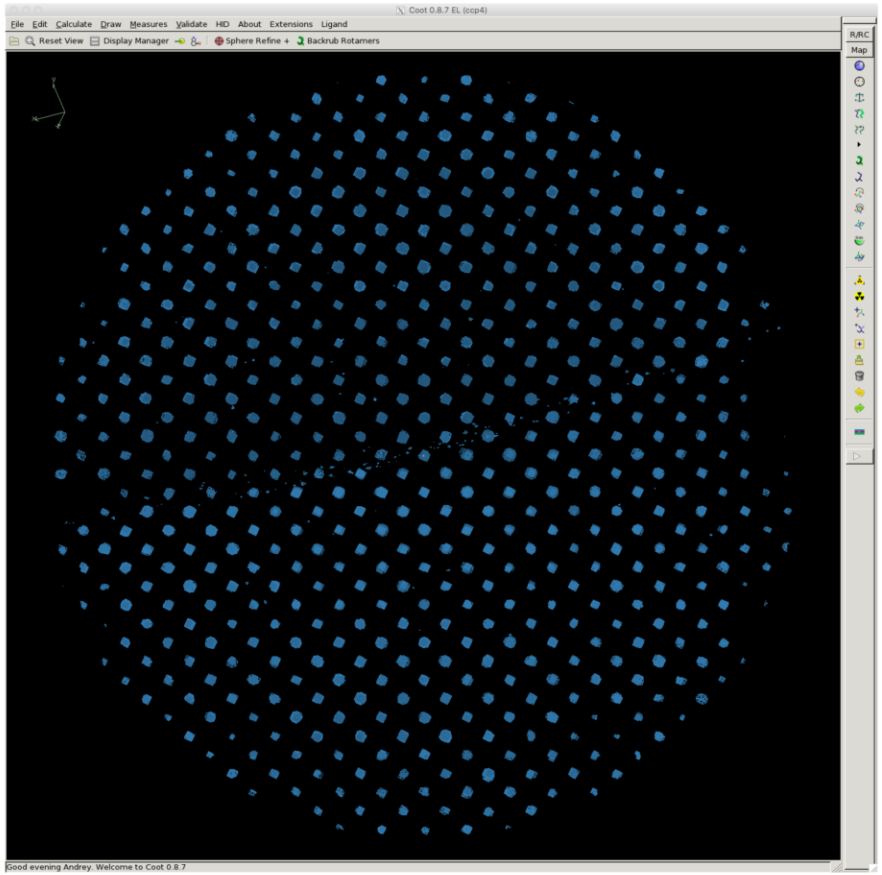

last image



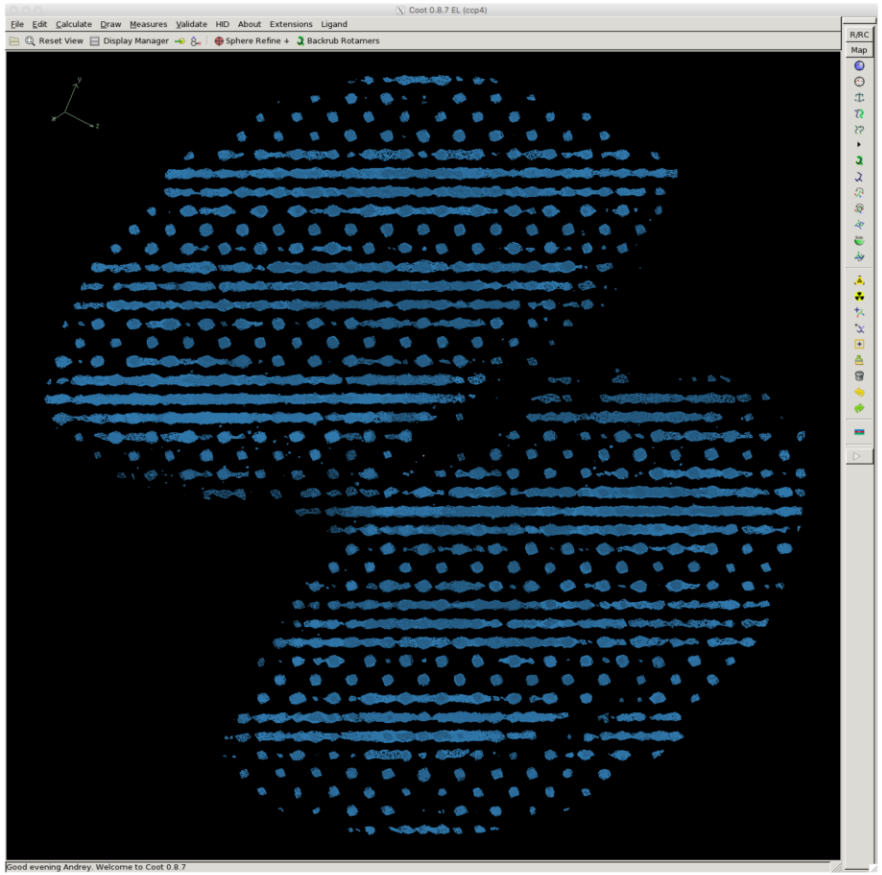

Blue arrow – direction of missing global translation

# Example 3: checking diffraction in 3D

front view



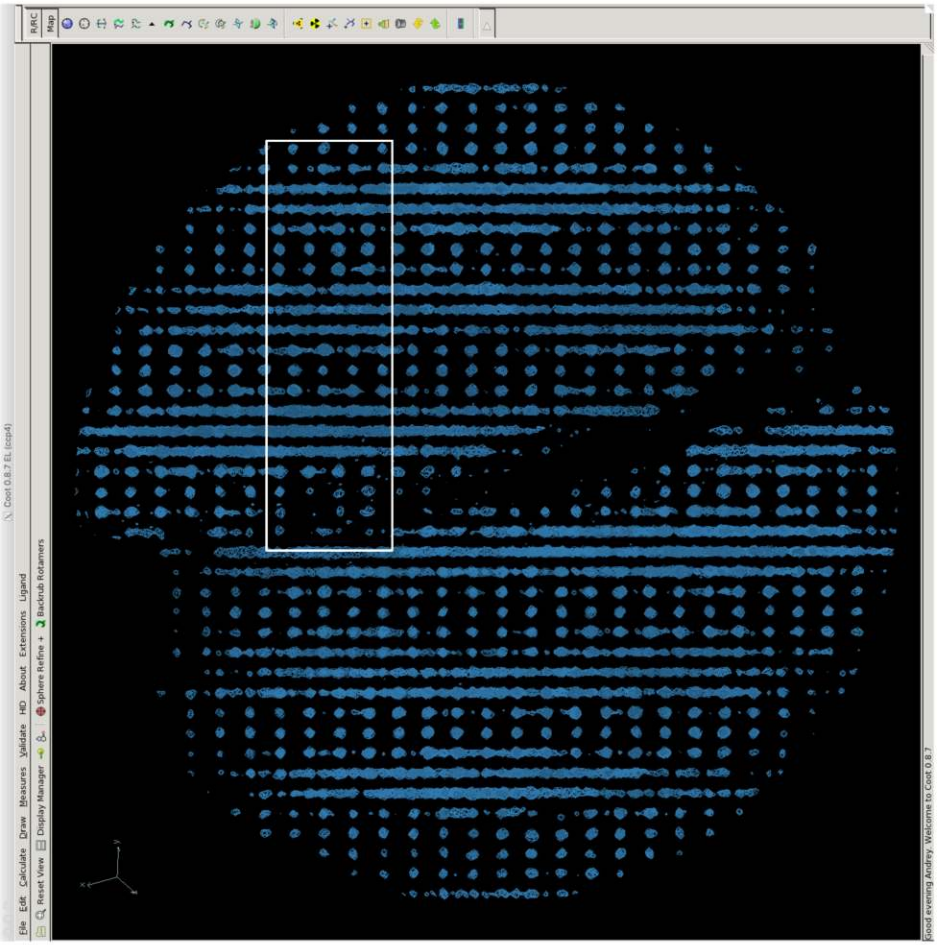

side view



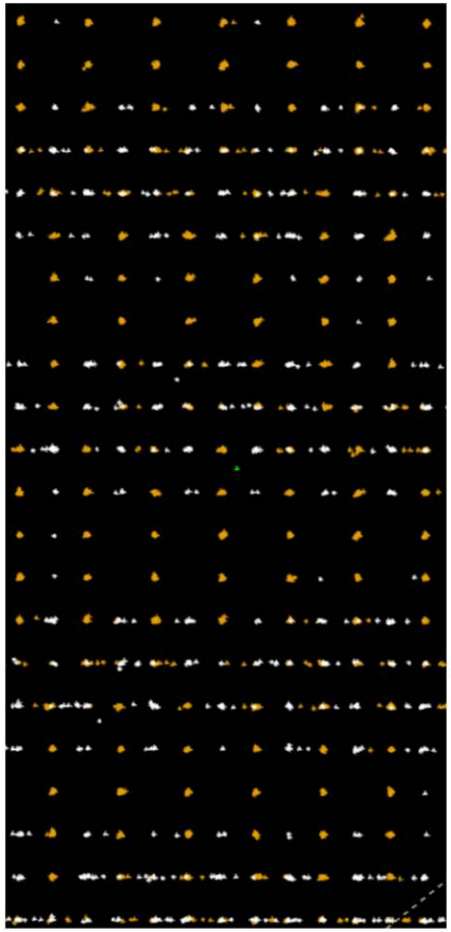

Clear partial disorder

# Example 3: wrong and correct indexing

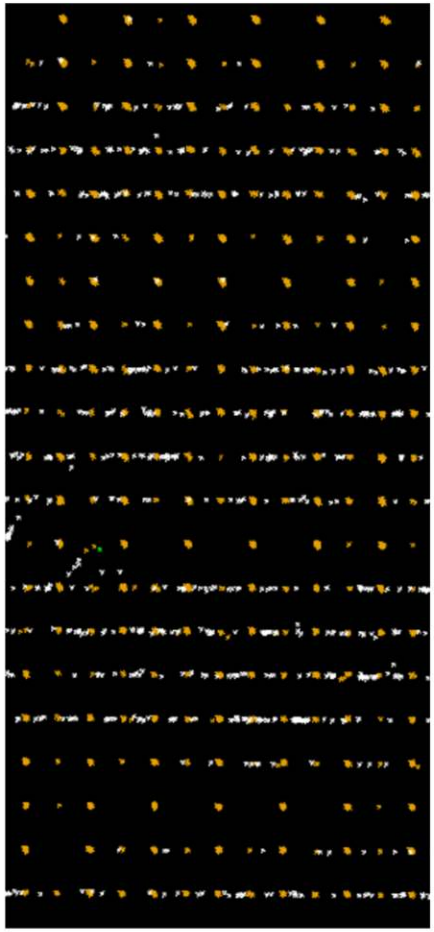
- Yellow spots are indexed, the white ones are not.



2, C (wrong)

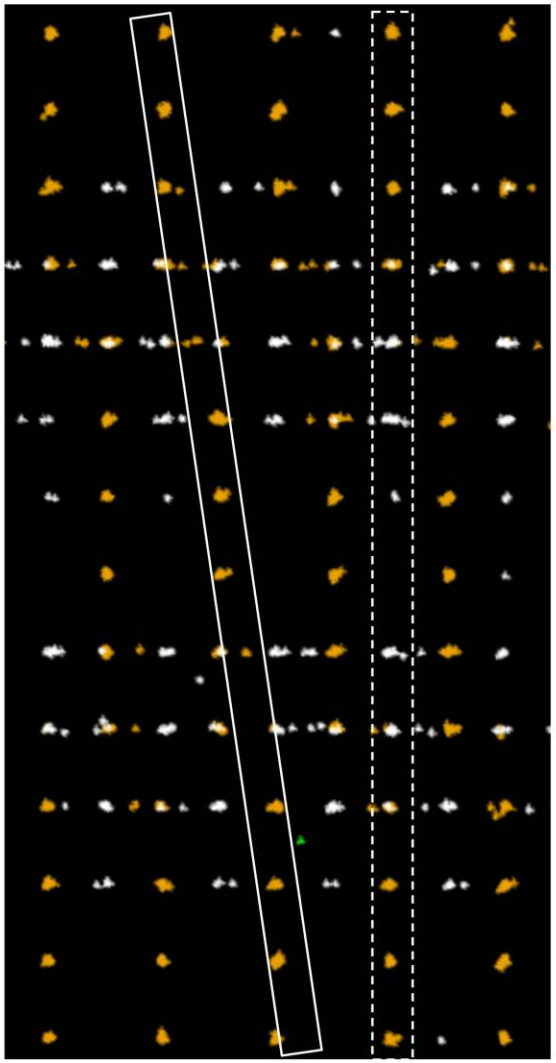


222, C (correct)

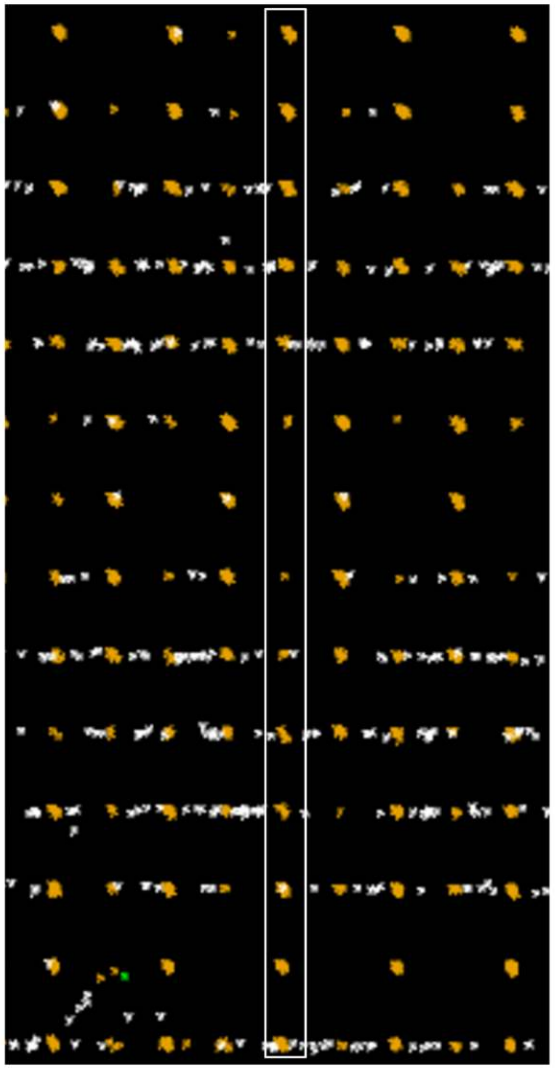


# Example 3: wrong and correct indexing

2, C (wrong)



222, C (correct)



# Example 3: unsolvable structure

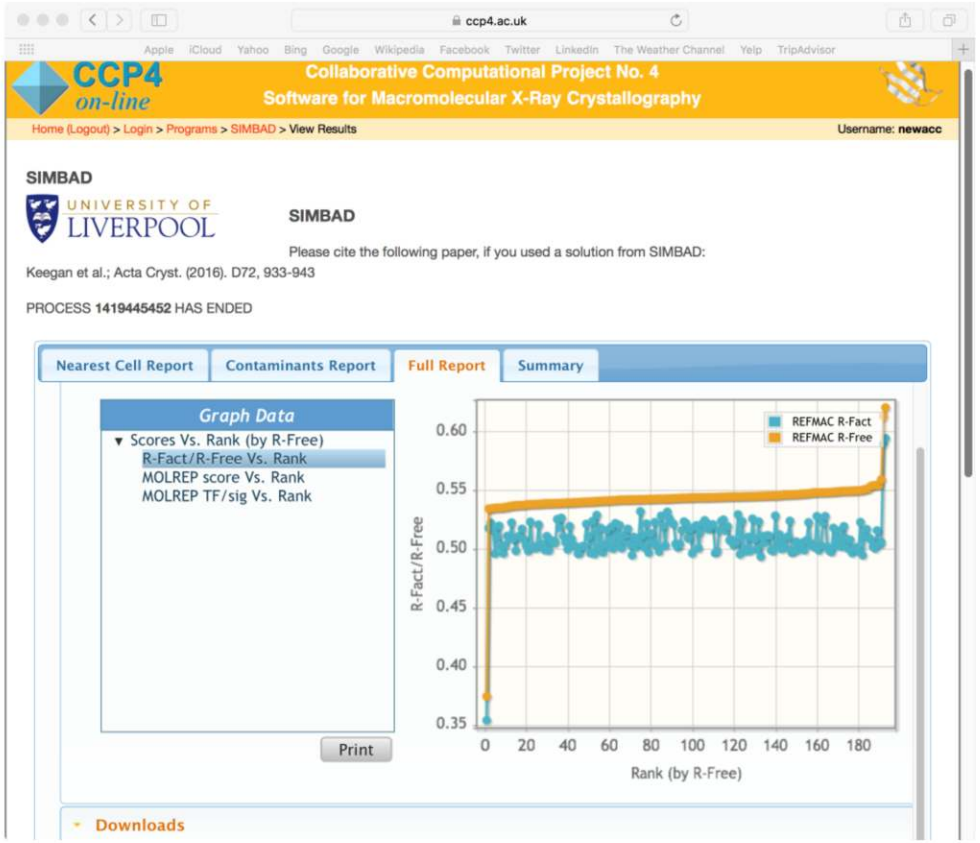
Input information:

- Images are good
  - But there a several different indexing solutions
- 99% homologue for Molecular Replacement
  - But no MR solution
  - Even more, no contrast on Rotation Function
- Twinning?

# Example 3: MR against PDB (Simbad; less happy end)

Despite very clean sample
a minor contaminant has crystallised:

Crystal pathology is not necessarily a reason for failure to solve structure!



# Partial disordered OD structures

- Data processing
  - Indexing can go wrong (use higher "gain" parameter, merge several adjacent images together etc. to get it right)

- Structure solution:
  - Molecular Replacement - yes
  - Experimental phasing – may be problematic

- Refinement / model building:
  - Some features of electron density may not be interpreted (ghost density)
  - Expect higher R-factors

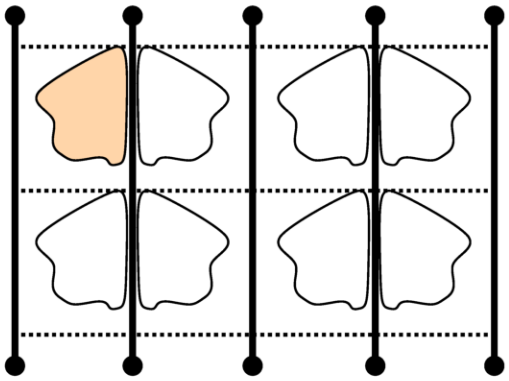
- Crystals with translocation defects
  - Term usually used in MX for partially disordered OD-structures

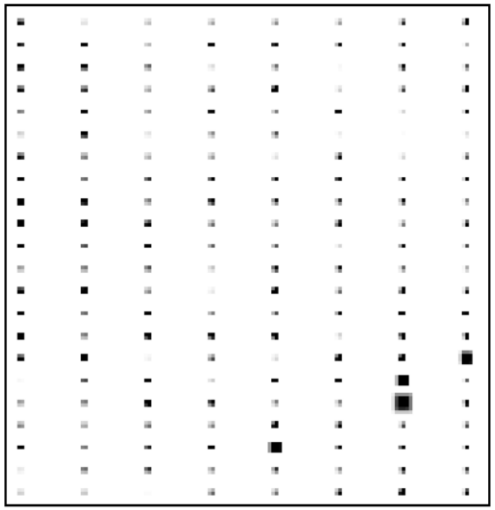
# Pseudo-translation

- Visualisation
- Effect on indexing
- Pseudo-origin MR solutions

# Pseudotranslation

Crystallographic translation





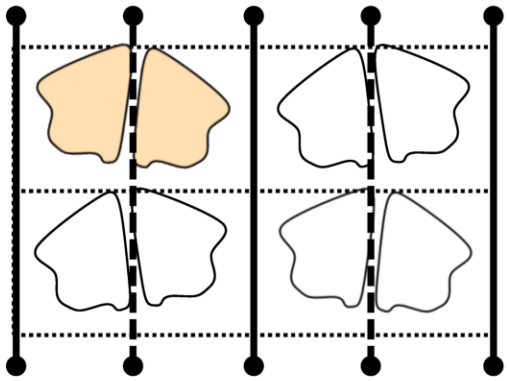
No pseudotranslation

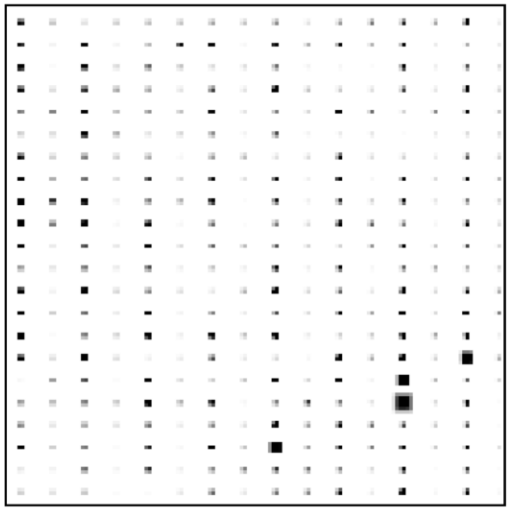
*c*

*c**

Crystallographic translation

Pseudo-translation





Pseudotranslation

*c*' = 2 *c*

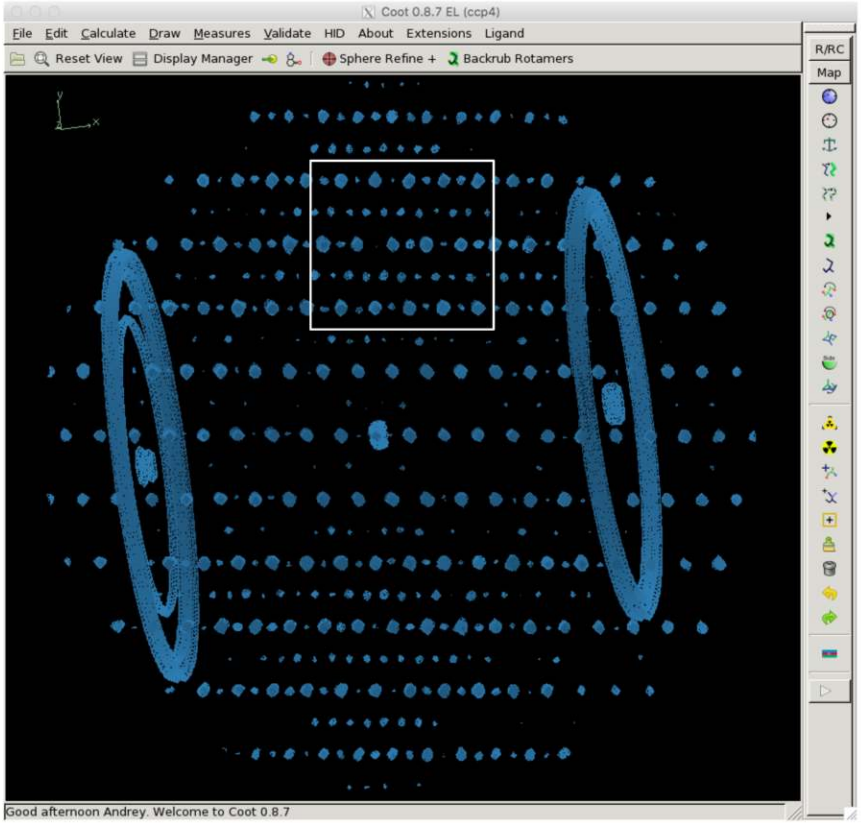
*c*'* = *c**/ 2

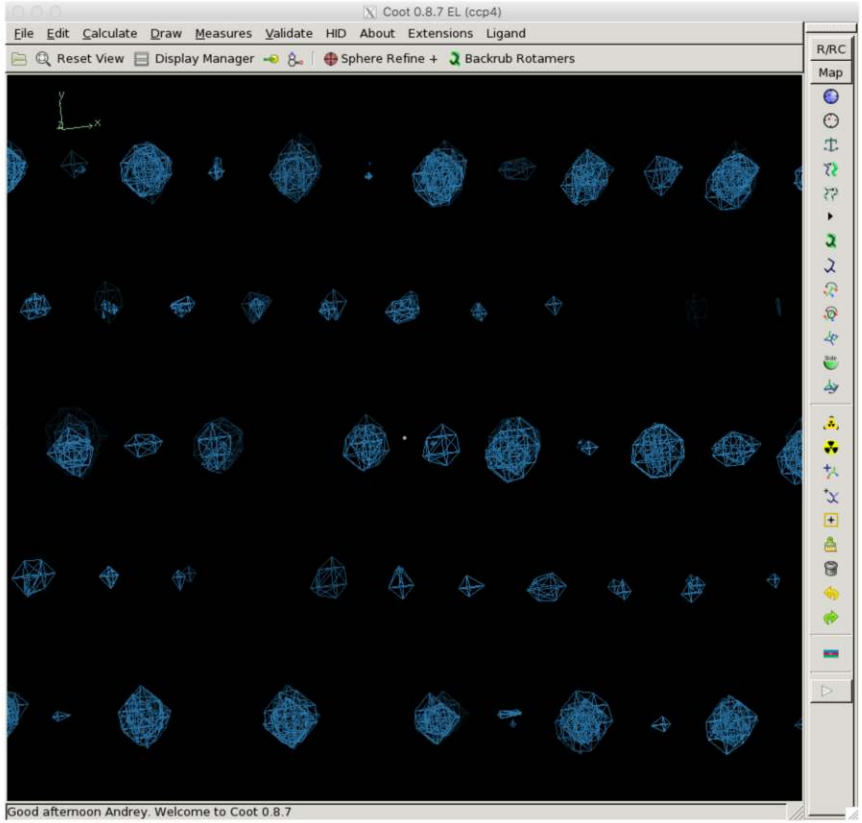
Planes 2L+1 contain weak reflections

# Example: two pseudo-translation vectors

Example from Victor Lamzin, YSBL-DESY

| | point group | lattice type | *a* (Å) | *b* (Å) | *c* (Å) |
|---|---|---|---|---|---|
| Space group | 222 | *C* | 74.9 | 122.8 | 125.0 |
| Pseudo-symmetry space group | 222 | *I* | 37.5 | 61.4 | 125.0 |



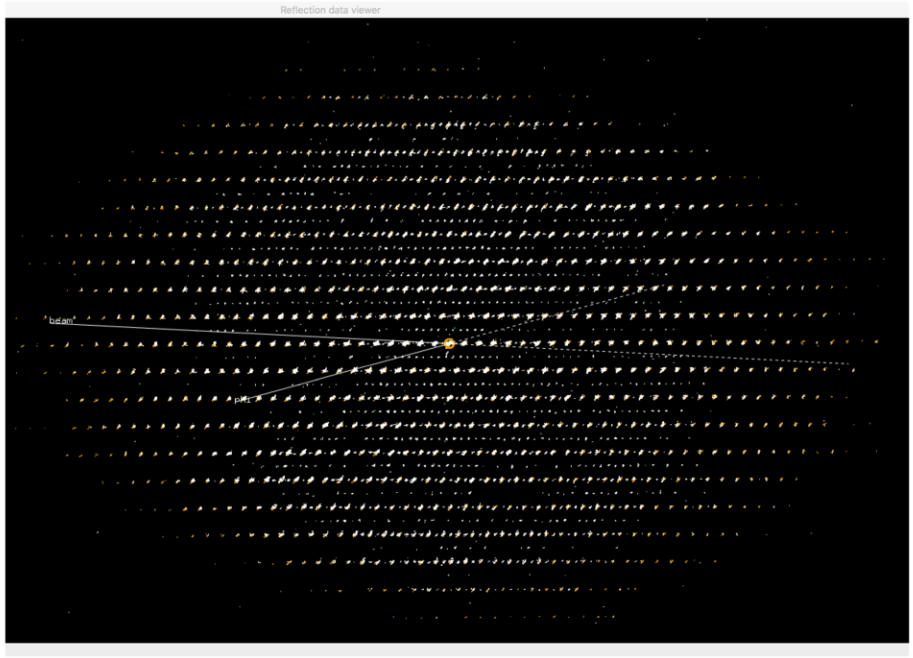



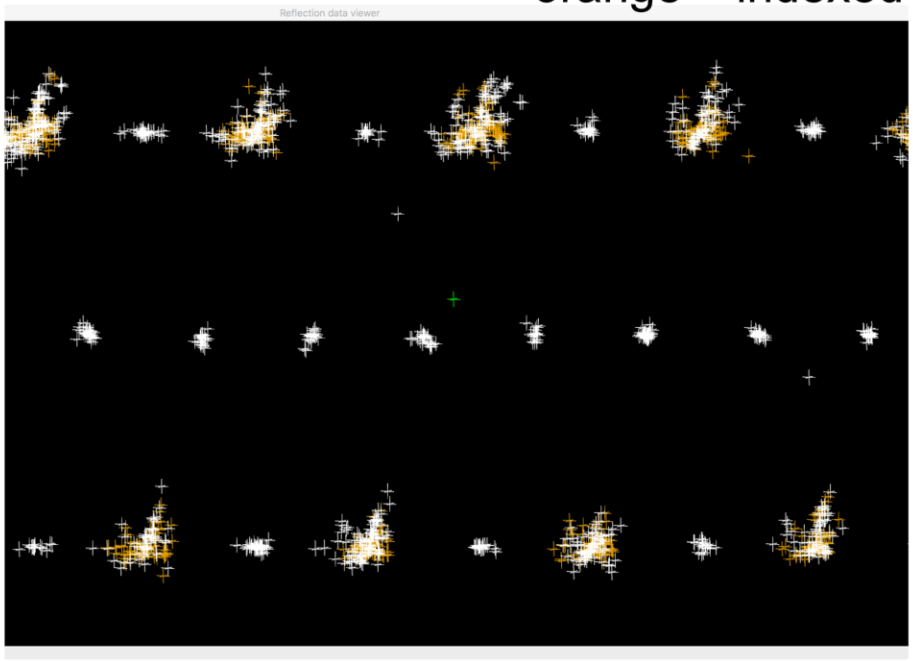
# Example: two pseudo-translation vectors

Images imported as they were, oscillation 0.1°

```
dials.import template=images/SeMet_38_04_0####.cbf
dials.find_spots ...
dials.index ...
dials.refine ...
dials.reciprocal_lattice_viewer ...
```

white – not indexed
orange – indexed



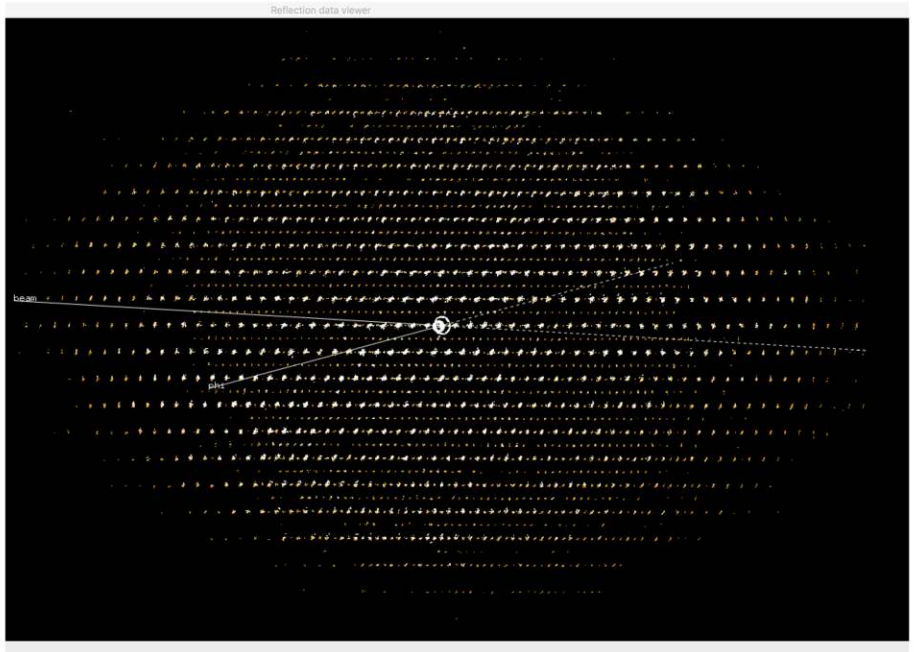



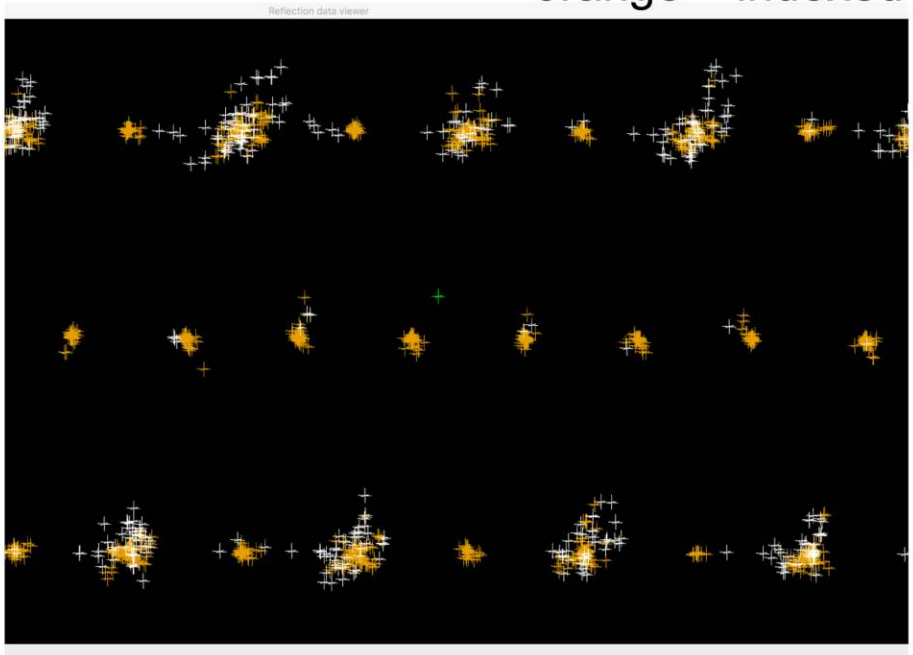
# Example: two pseudo-translation vectors

Merged each 5 adjacent images to make oscillation 0.5°, then imported

```
dials.merge_cbf images/SeMet_38_04_0####.cbf merge_n_images=5
dials.import template=sum_####.cbf
dials.find_spots ...
dials.index ...
dials.refine ...
dials.reciprocal_lattice_viewer ...
```

white – not indexed
orange – indexed



# Pseudo-translation and indexing

The last example:

- structure solved using SAD
- then native structure was solved by MR

Week reflections may confuse indexing programs

Visual control using 3D viewers is useful

- check if pseudo-translation is not overlooked
- check if pseudo-translation is not an indexing artefact

How important is to use the weak reflections?

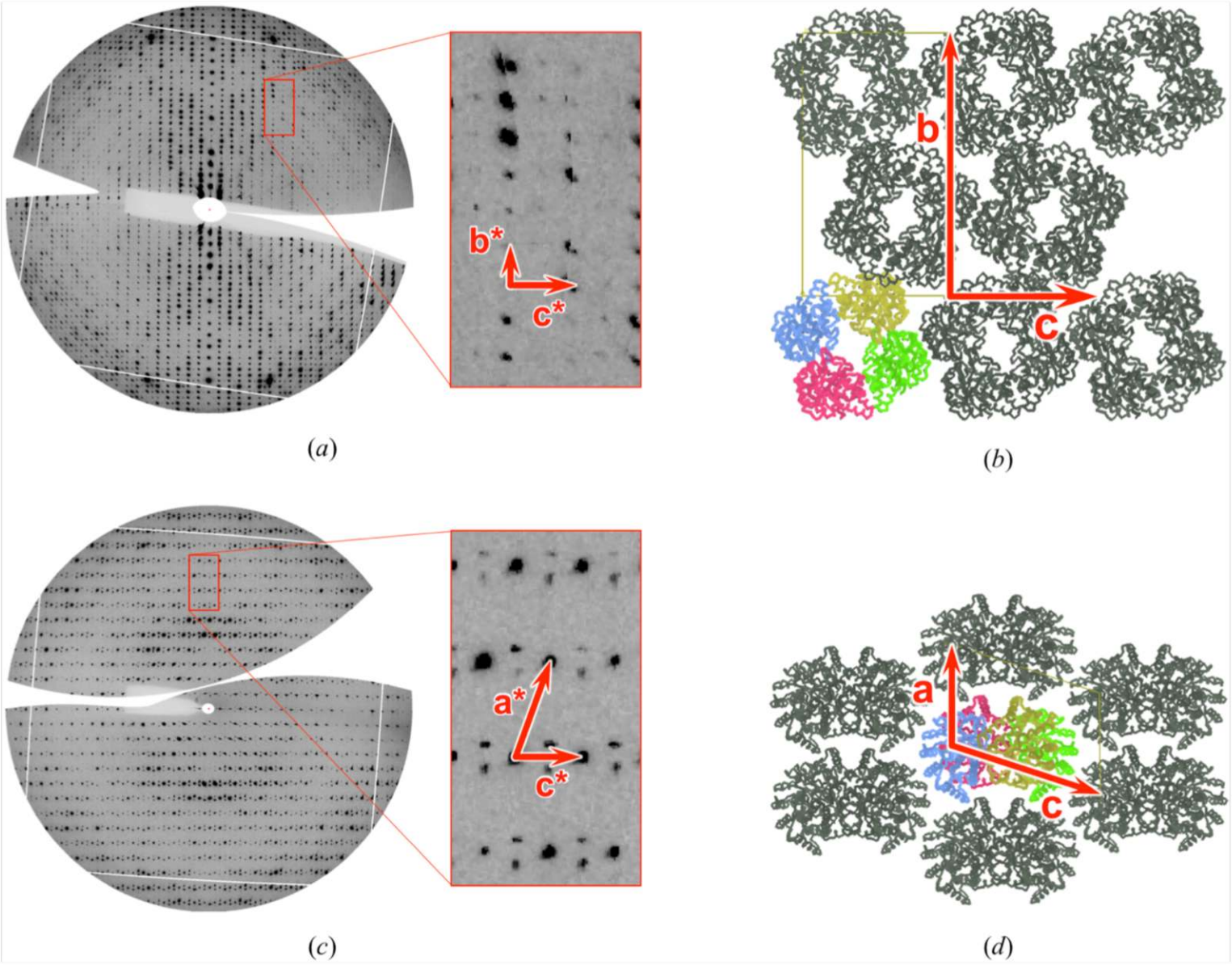
- usually improve both density and refinement stats
- there are examples when these only make refinement stats worse
- sometimes ignored to simplify the first steps of structure solution and used later

# Non-commensurate modulated structures

- Example

» from Ivan Campeotto, Oxford and Arwen Pearson, DESY (PDB id 2wnq)

# Non-commensurate modulated structure



END