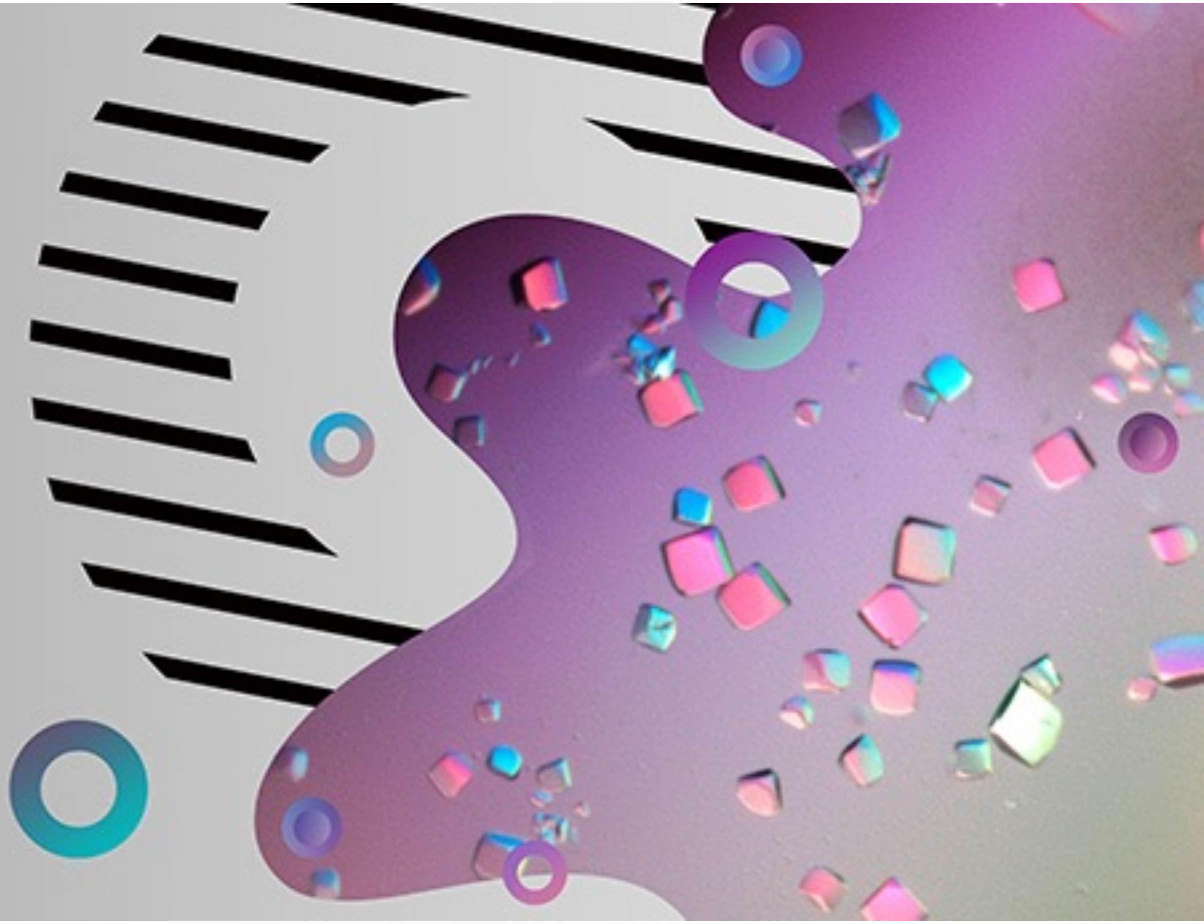
Multi-crystal approaches

CCP4 Workshop

30.11.2021



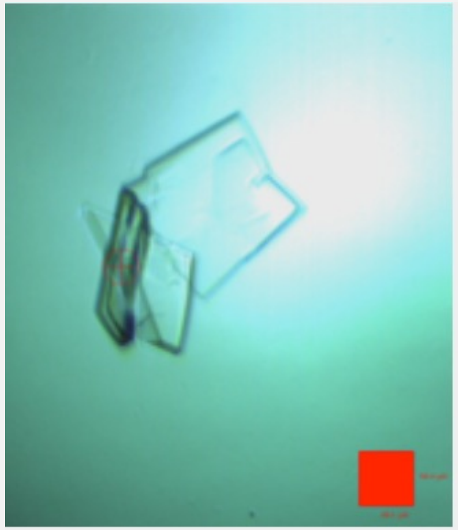
Selina Storm

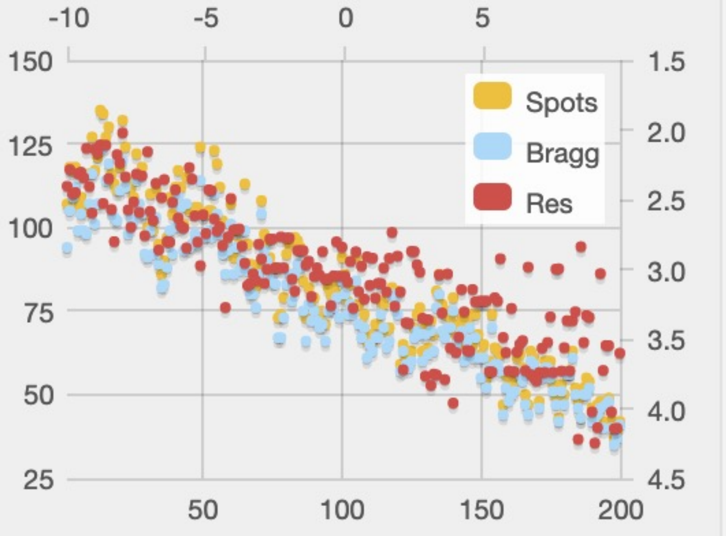
Project Manager for EMBL@PETRA IV

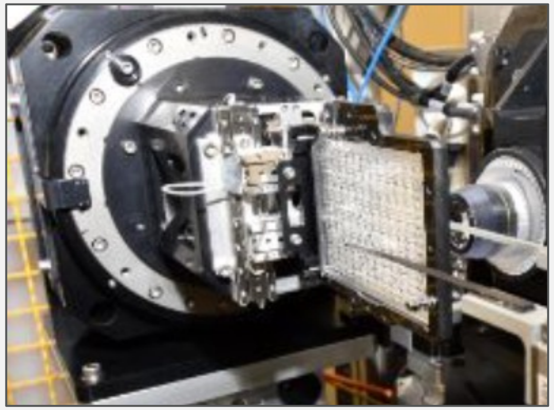
Outline

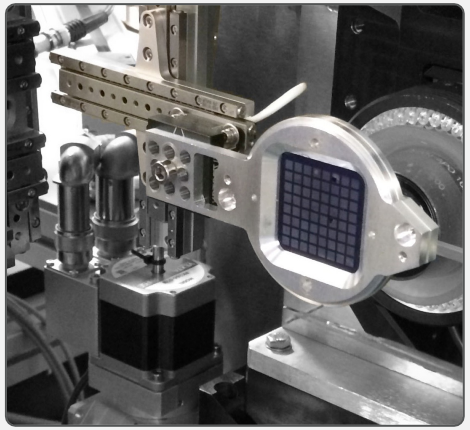
- motivation
- multi-crystal approaches and serial crystallography
- data collection of multiple crystals - things to consider
- processing data of multiple crystals
  - an introduction to dials.multiplex
  - *in situ* data collection on the main protease (MPro) of COVID-19

Why would one want to use multiple crystals?









- radiation damage
  - especially an issue with small crystals
- to achieve
  - higher resolution and better anomalous signal
  - higher completeness (limitations due to sample mount)
- requirement for time-resolved (serial) experiments

Multiple crystal approaches

frames per crystal		number of crystals
hundreds	➢ rotational crystallography (standard loops, grids, *in situ)*	tenths
tenths	➢ 'rotational serial': a few degrees of data with serial methodology, e.g. rotating chips [1] or helical scans [2]	hundreds
one	➢ serial crystallography (on chips, in liquid jets …)	thousands

[1] Wierman et al,, IUCrJ (2019). 6, 305–316

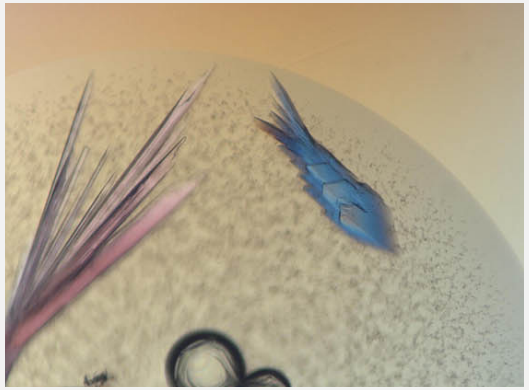
[2] Gati et al., *IUCrJ* 1.2 (2014): 87-94

Choose your weapon – what are your crystals like?

Adapt your experimental approach accordingly to the aim/situation!

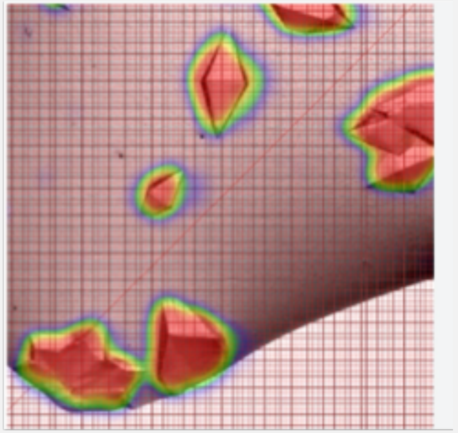
- lots of big, badly diffracting crystals:

  collect small sweeps in loops and try to get higher resolution



- brittle crystals which don’t like to be taken out of the plate:

  grow them in plates which can be mounted on the beamline (VMXi, I24@Diamond, P14@PETRA III, X06DA@SLS...)

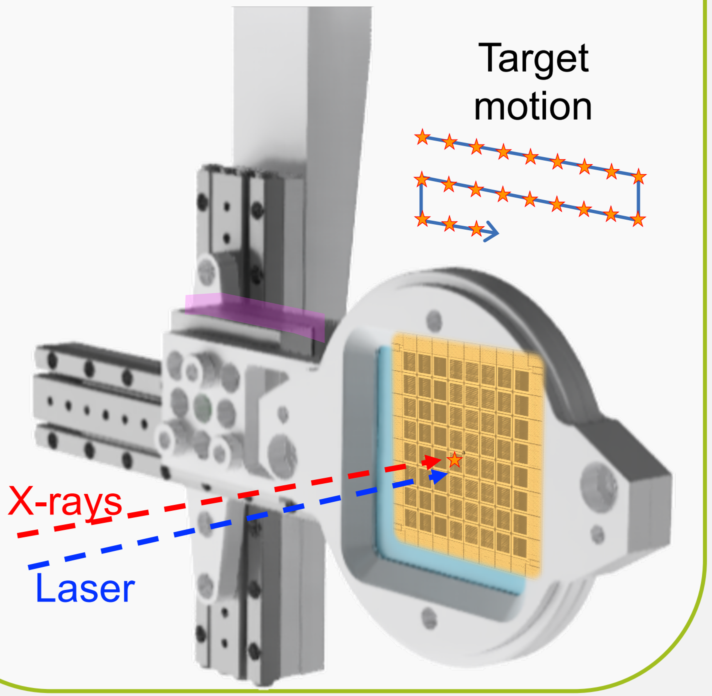


https://askabiologist.asu.edu/explore/making-life-crystal-clear

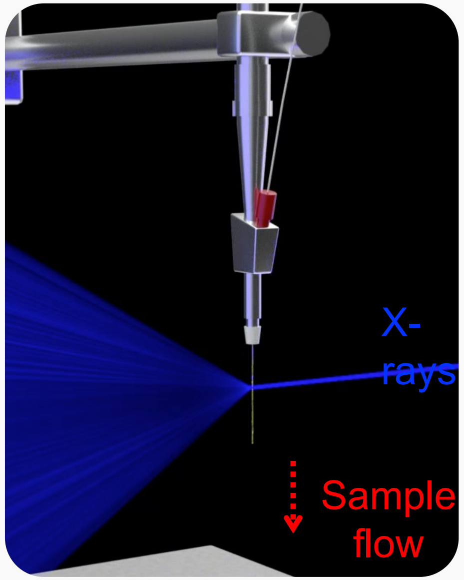
What is the aim of your experiment?

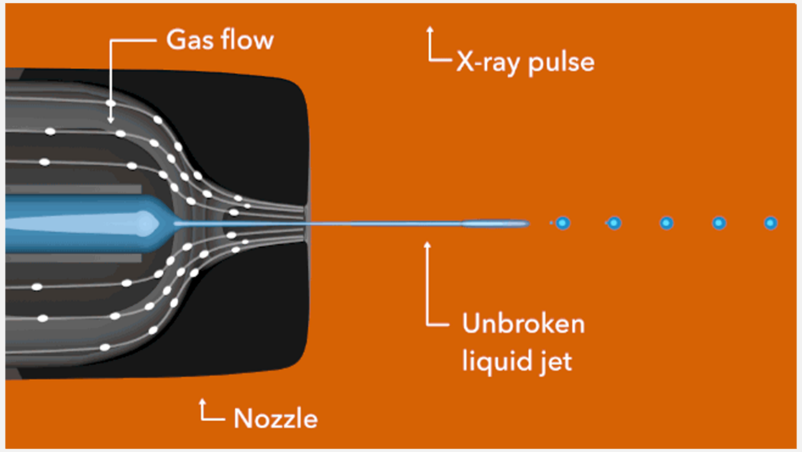
- low-dose / time-resolved structures: serial crystallography (! 'still' images !)

Fixed targets



LCP extruder





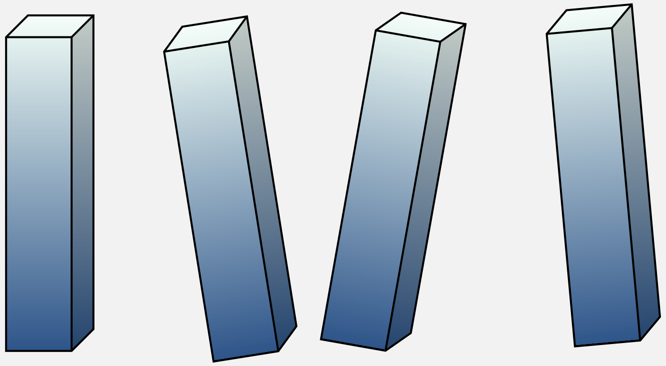
Courtesy Robin Owen, Pierre Aller

Challenges when using multiple crystals

- strictly speaking, each dataset is an experiment in itself
  - variation of the X-ray beam, temperature, humidity, solvent can play a role

- there are no two crystals alike
  - different unit cell parameters
  - different mosaicity
  - different resolution

- preferred orientation (rods, plates)



Things to consider for multi-crystal data collection

- take extra care to keep things as constant as possible (crystallization conditions, temperatures, beamsizes etc)

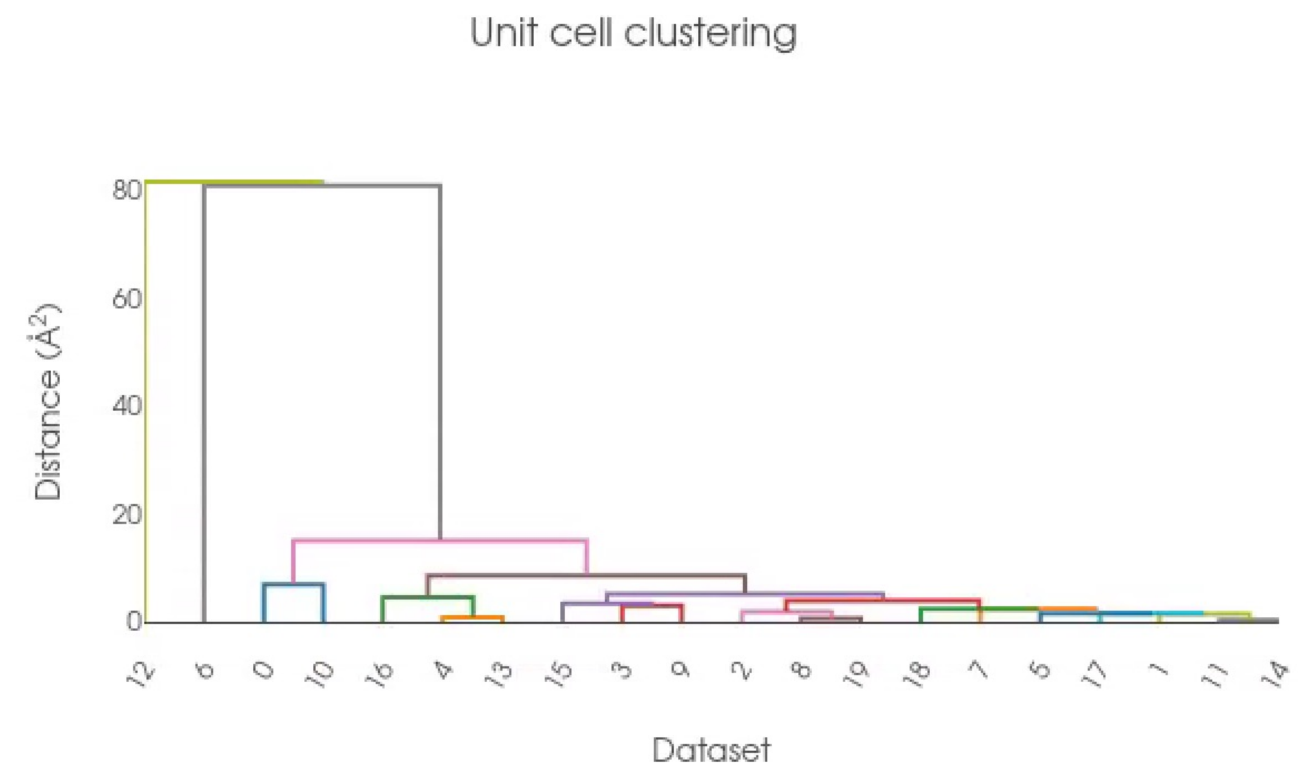
- vary the starting angle, especially when you have preferred orientation

- key parameter : completeness!
  - ideally, data are processed on the fly

- name your files consistently

Things to consider for multi-crystal analysis

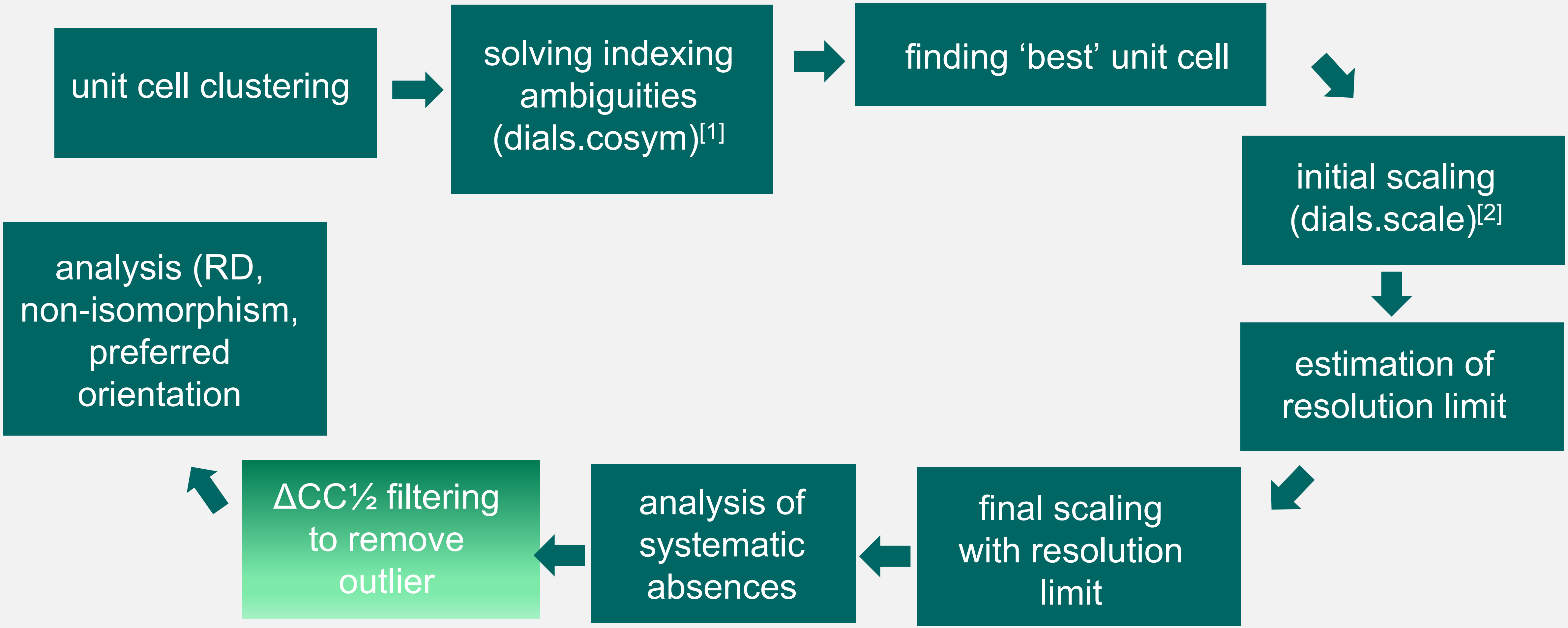
- all data sets need to be processed with the same space group (can be tricky for very small sweeps)
- unit cells need to be quite similar
- data will be scaled together



Programs for data processing – rotational data

- XSCALE (for data processed with XDS)
- BLEND (for data processed with MOSFLM, XDS or DIALS)
- dials.multiplex (for data processed with DIALS)
  - generates an html file with an excellent overview
  - autoprocessing on Diamond beamlines to give feedback on-the/fly

Dials.multiplex: how does it work?



unit cell clustering

solving indexing ambiguities (dials.cosym)[1]

finding ‘best’ unit cell

initial scaling (dials.scale)[2]

estimation of resolution limit

final scaling with resolution limit

analysis of systematic absences

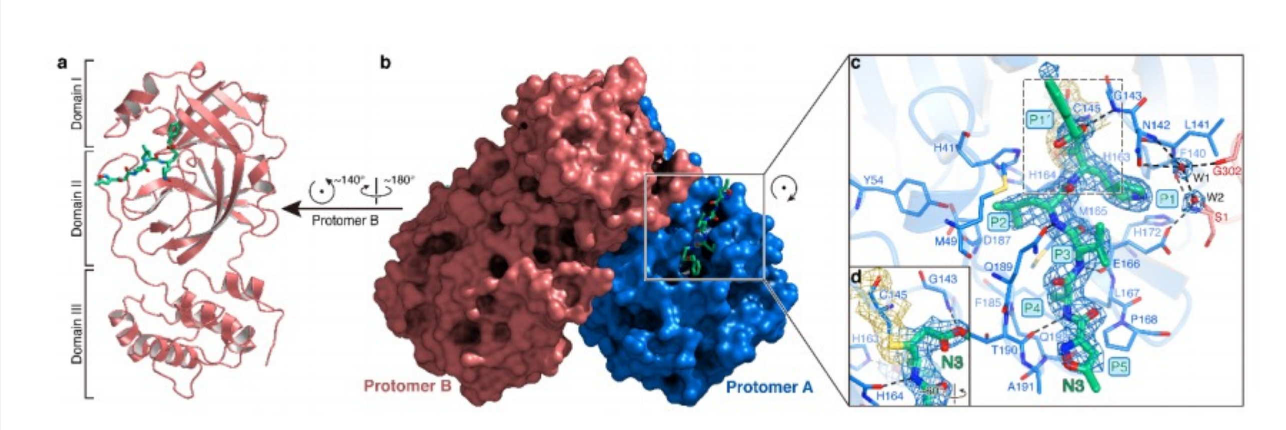
ΔCC½ filtering to remove outlier

analysis (RD, non-isomorphism, preferred orientation

[1] Gildea & Winter, Acta Cryst. (2018). D74, 405-410
[2] Beilsten-Edmands et al., Acta Cryst. (2020). D76, 385-399

Example: structure of M^{pro} binding ligands

- M^{pro}: main protease of SARS-CoV-2
- key enzyme of coronaviruses having a central role in mediating viral replication and transcription

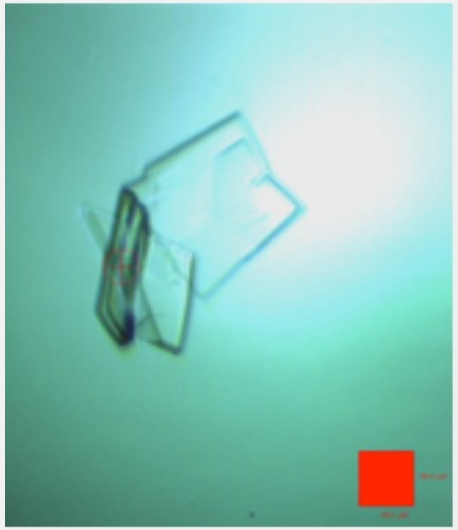


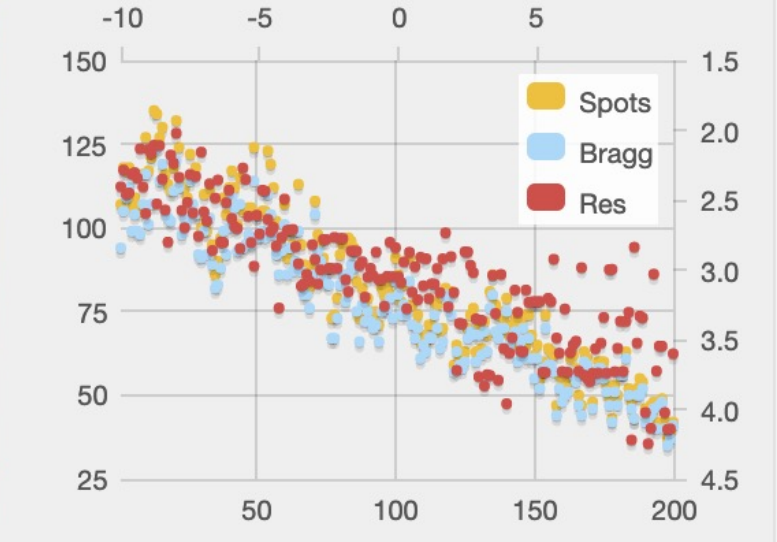
- active site empty and solvent accessible - perfect for fragment screening
- more than 2000 fragments screened, including Moonshot compounds

[1] Zhenming Jin et al., Nature volume 582, pages 289–293 (2020)
[2] Douangamath et al *Nature Communications* **volume 11**, Article number: 5047 (2020)

Example: RT structure of M^{pro} binding ligands

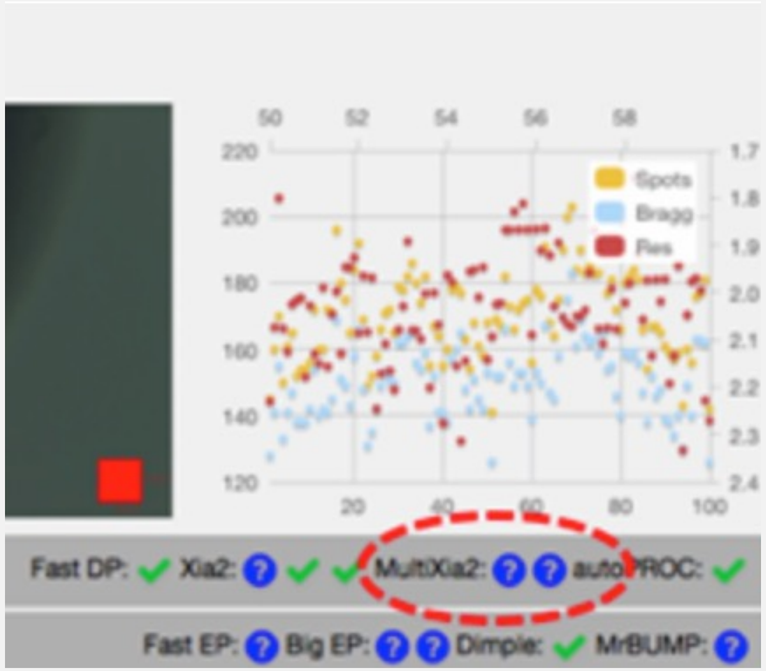
- all initial screens performed on I04-1 at 100 K - are the structures the same at room temperature?
- *in situ* room temperature data collection in plates
- a few drops with the same compound
- strong radiation damage (20 degree)
- preferred orientation, SG C121
  - varying starting angles





Processing on-the-fly

- to answer the question: do we have enough data to get high completeness and sufficient resolution to see the ligand?
- ‘force’ the correct space group for auto-processing by entering it into ISPyB
- raw data all need to go into the same folder
- two multiplex auto-processing runs in ISPyB:



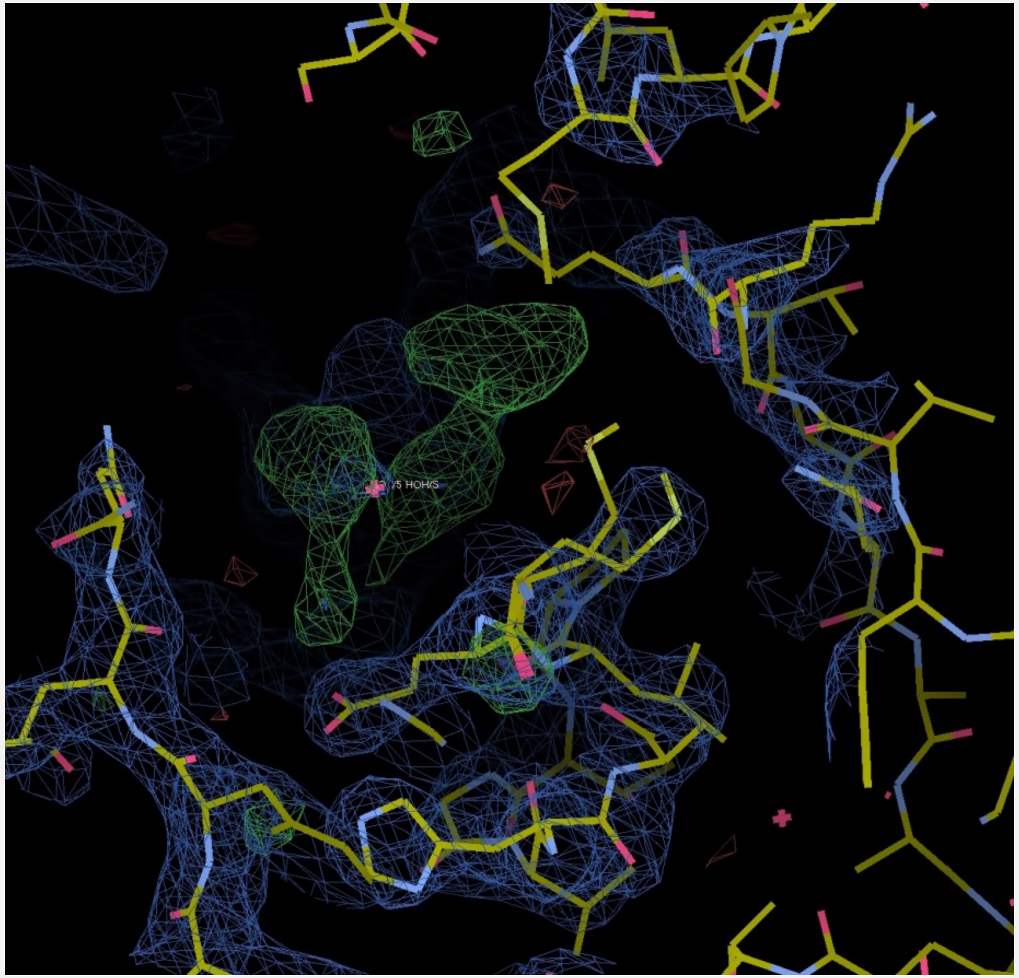
https://www.diamond.ac.uk/Instruments/Mx/I24/Status/multi-xtal-proc.html

Processing on-the-fly

- two ISPyB outputs:
  - one processes with the user-defined space group
  - the other determines the space group itself using dials.cosym and dials.symmetry

- typically 10 data sets needed for a complete data set, more used to increase resolution

- automated molecular replacement with DIMPLE possible



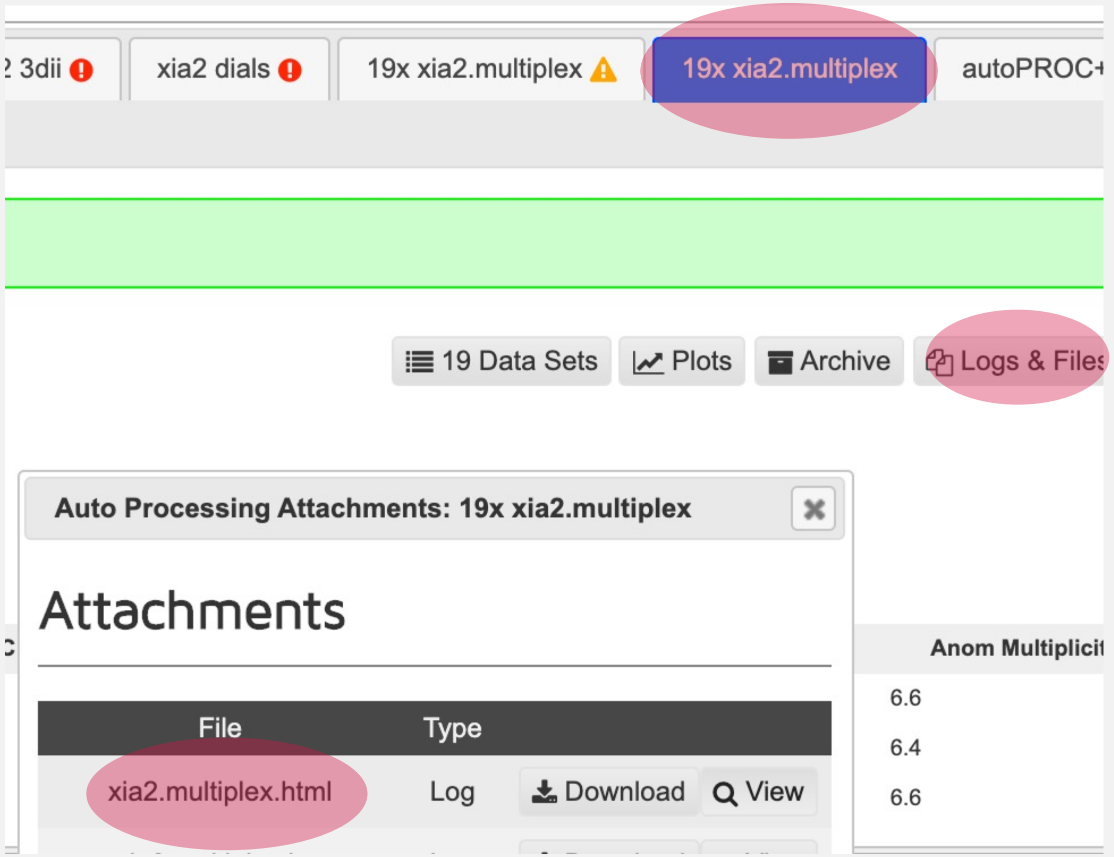
Processing via command line

- required: path to integrated dials results

```
xia2.multiplex $path_to_dials_results.{refl,expt} options
```

- common options:
  - `symmetry.space_group=C2`
  - `resolution.d_min=2.7`
  - `filtering.method=deltacchalf filtering.deltacchalf.stdcutoff=1.5`
  - `dose=1,100`

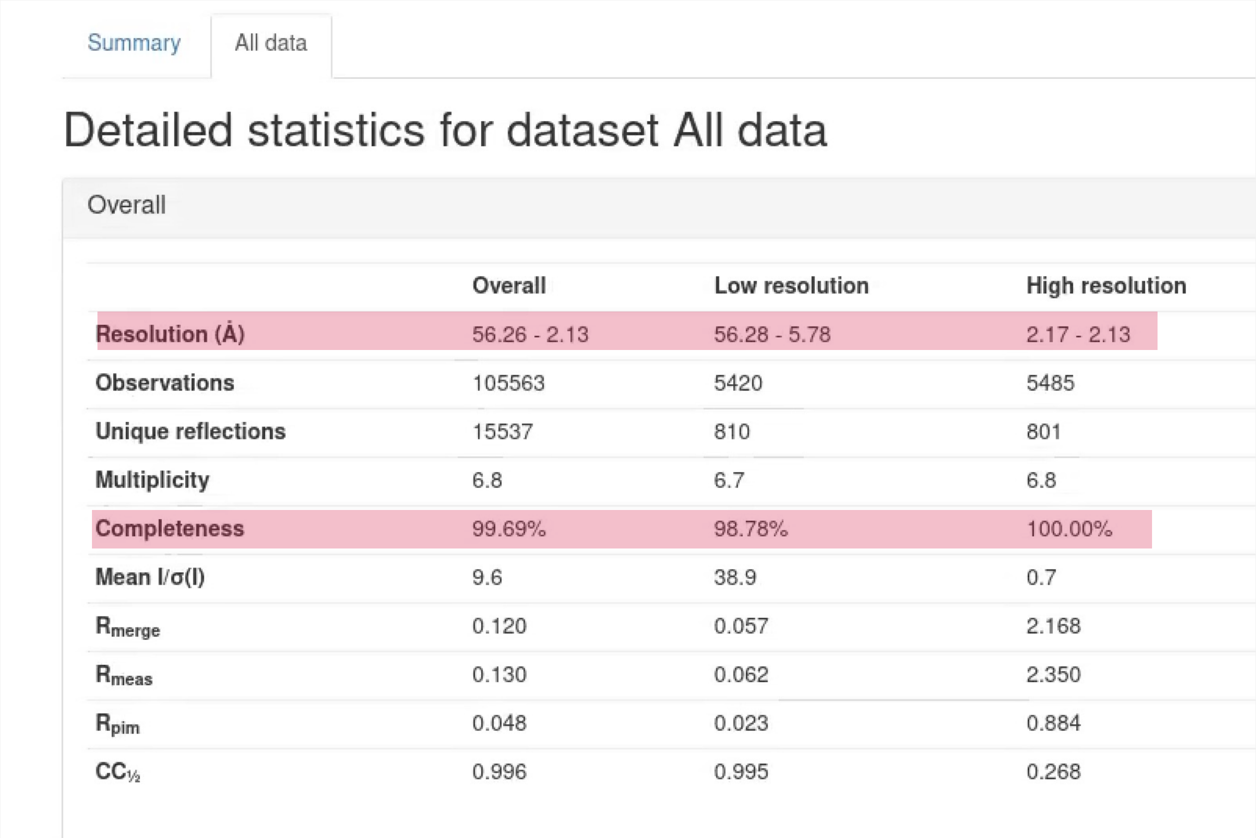
Understanding the results: the multiplex html report



Auto Processing Attachments: 19x xia2.multiplex

Attachments

File	Type	
xia2.multiplex.html	Log	Download View

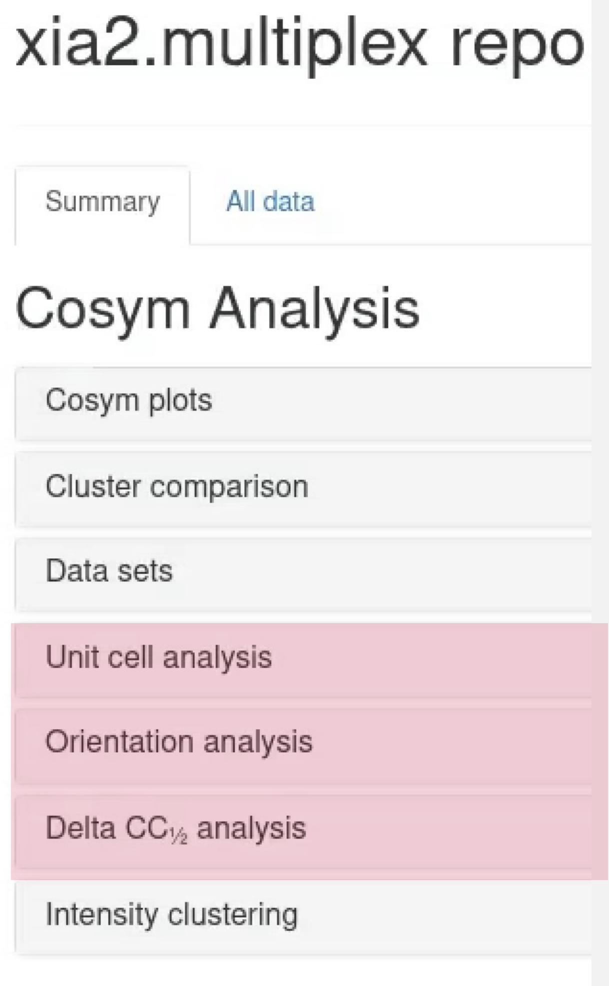


Detailed statistics for dataset All data

Overall

	Overall	Low resolution	High resolution
Resolution (Å)	56.26 - 2.13	56.28 - 5.78	2.17 - 2.13
Observations	105563	5420	5485
Unique reflections	15537	810	801
Multiplicity	6.8	6.7	6.8
Completeness	99.69%	98.78%	100.00%
Mean I/σ(I)	9.6	38.9	0.7
R_{merge}	0.120	0.057	2.168
R_{meas}	0.130	0.062	2.350
R_{pim}	0.048	0.023	0.884
$CC_{½}$	0.996	0.995	0.268

Xia2.multiplex.html

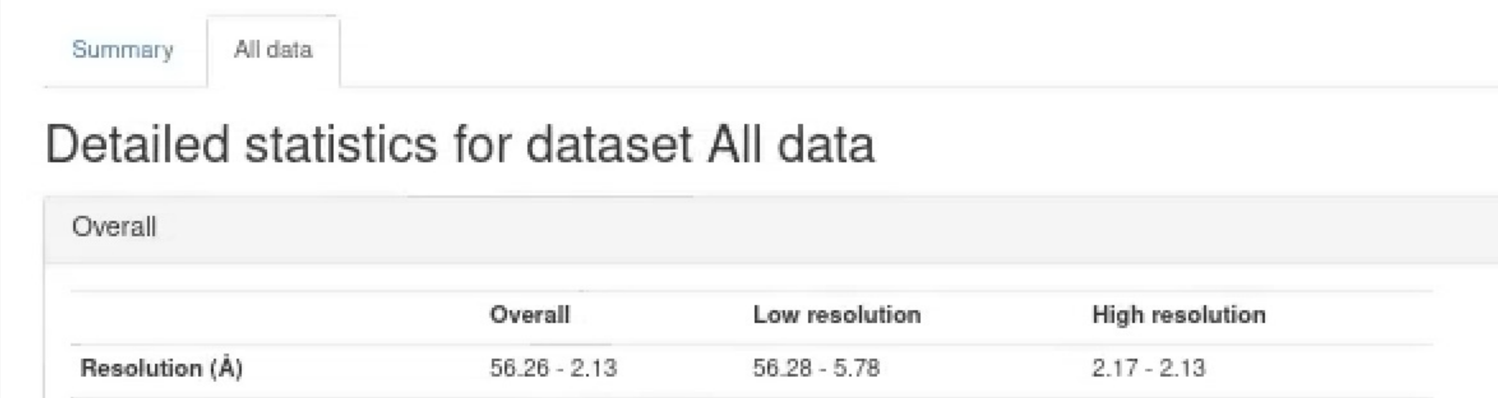


xia2.multiplex repo

Summary | All data

Cosym Analysis

- Cosym plots
- Cluster comparison
- Data sets
- Unit cell analysis
- Orientation analysis
- Delta $CC_{1/2}$ analysis
- Intensity clustering

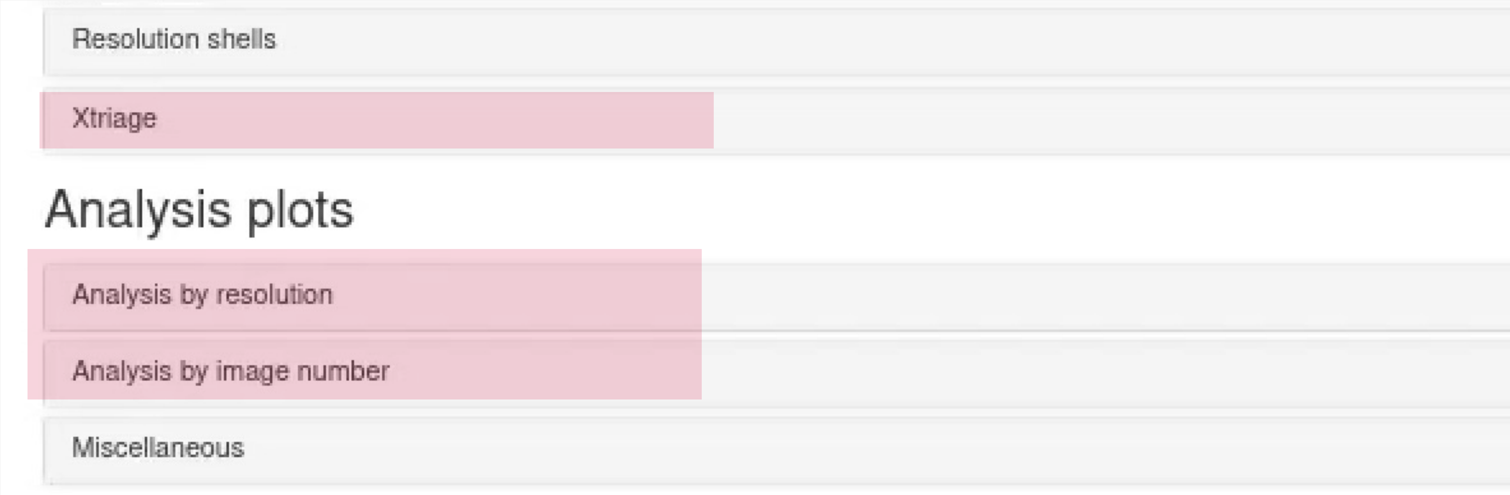


Summary | All data

Detailed statistics for dataset All data

Overall

	Overall	Low resolution	High resolution
Resolution (Å)	56.26 - 2.13	56.28 - 5.78	2.17 - 2.13

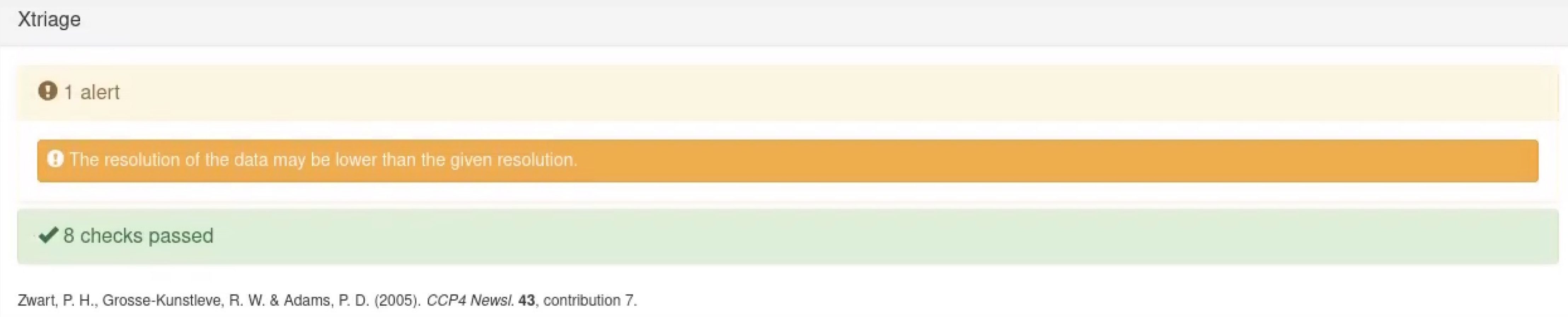


- Resolution shells
- Xtriage

Analysis plots

- Analysis by resolution
- Analysis by image number
- Miscellaneous

Xtriage (all data)



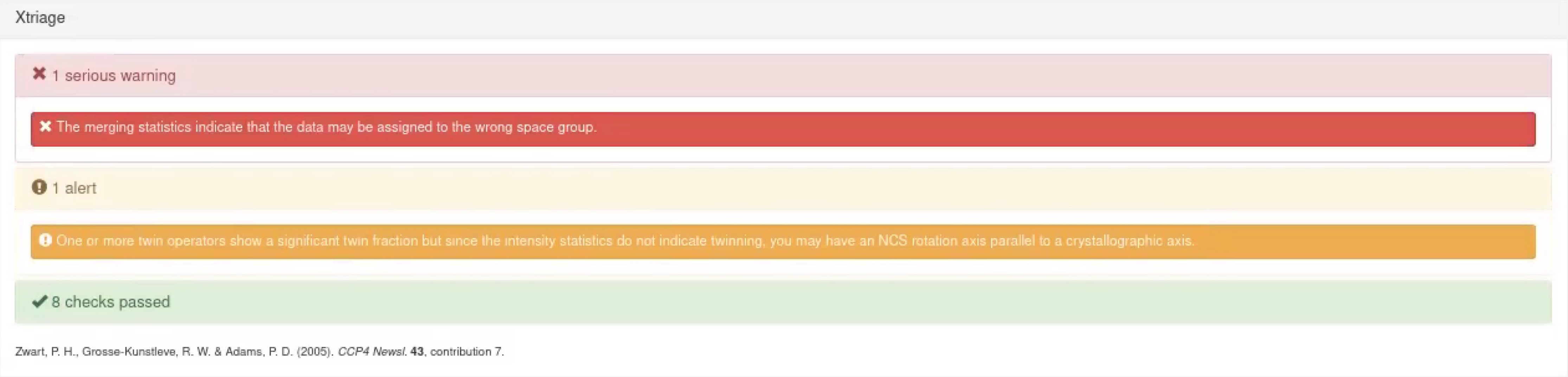
Xtriage

❗ 1 alert

❗ The resolution of the data may be lower than the given resolution.

✓ 8 checks passed

Zwart, P. H., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl.* **43**, contribution 7.



Xtriage

✖ 1 serious warning

✖ The merging statistics indicate that the data may be assigned to the wrong space group.

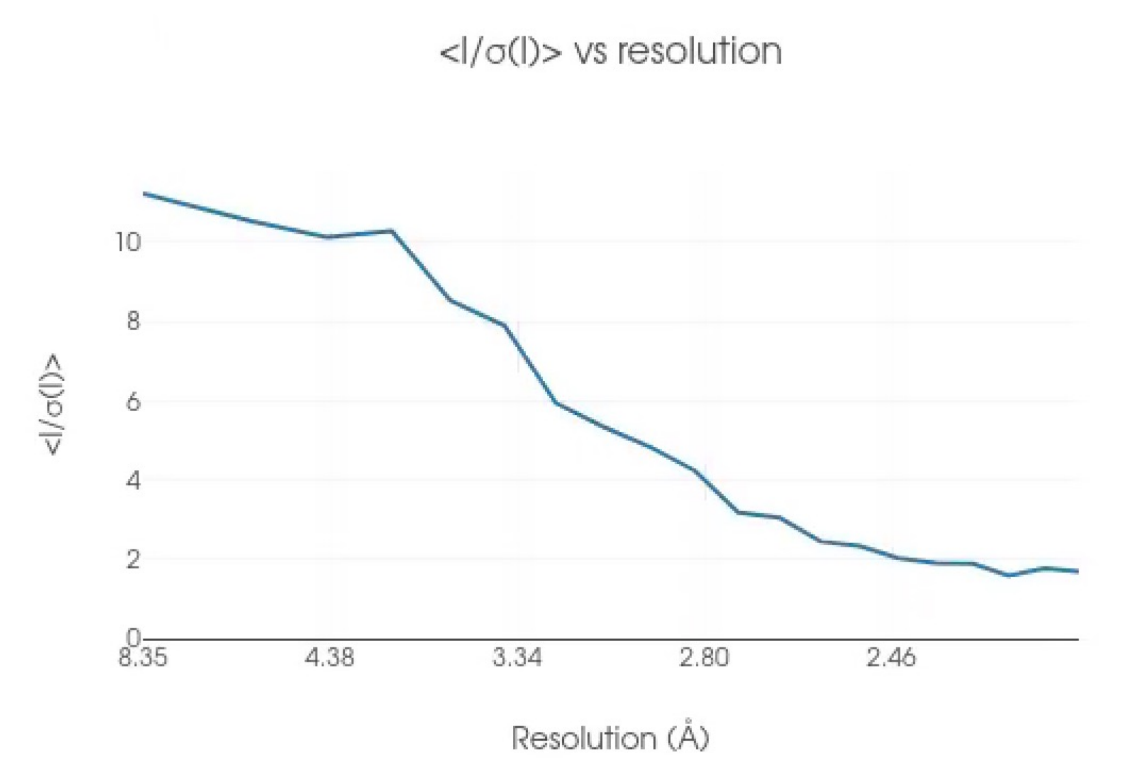
❗ 1 alert

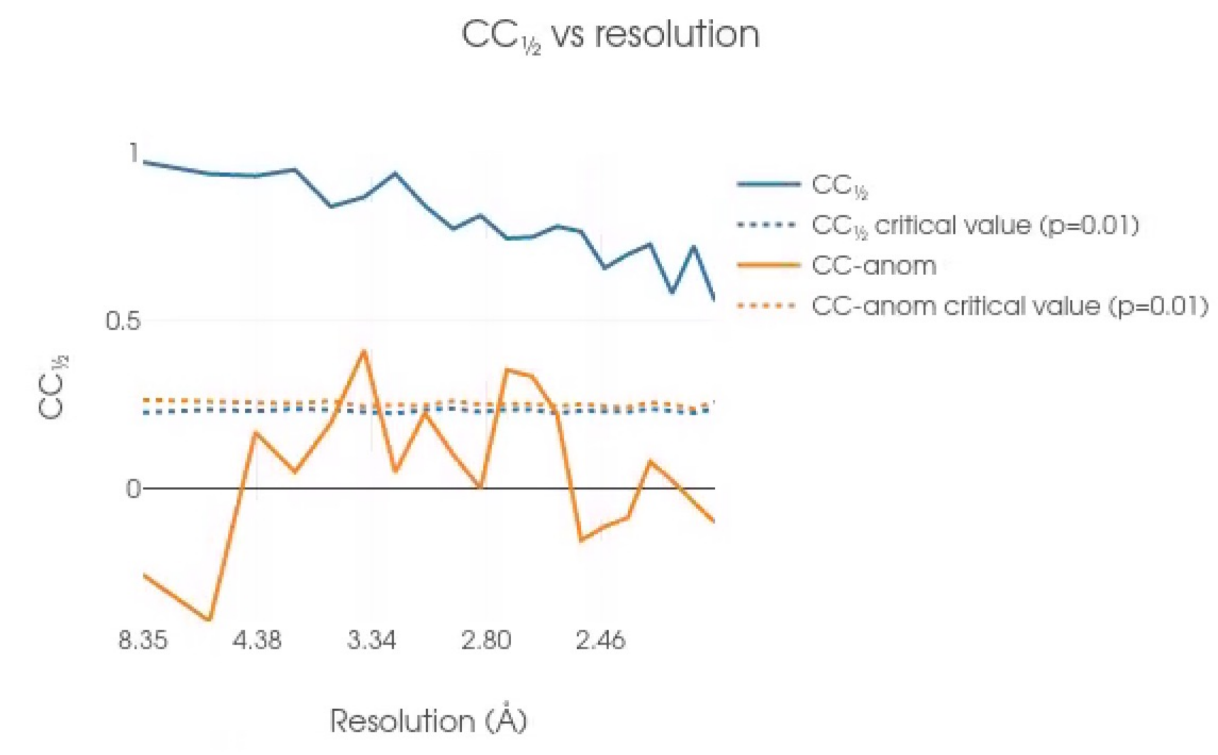
❗ One or more twin operators show a significant twin fraction but since the intensity statistics do not indicate twinning, you may have an NCS rotation axis parallel to a crystallographic axis.

✓ 8 checks passed

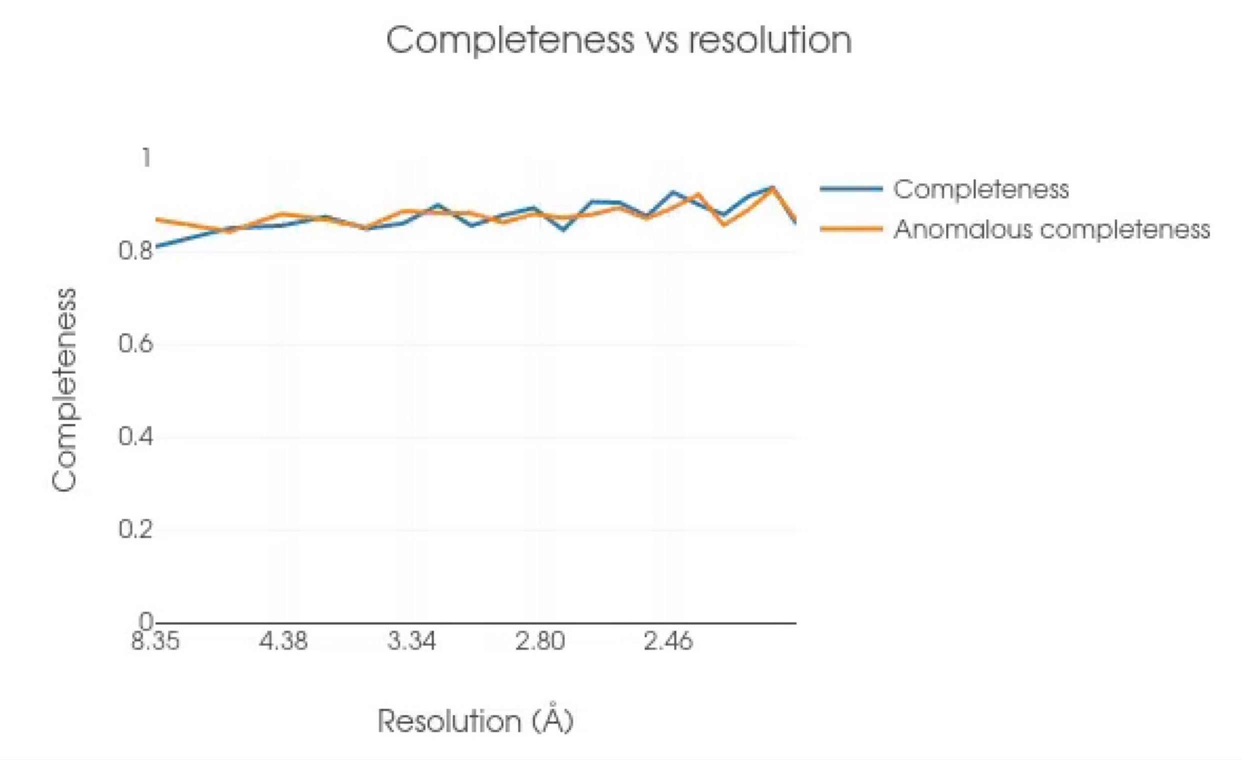
Zwart, P. H., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl.* **43**, contribution 7.

Analysis by resolution (all data)

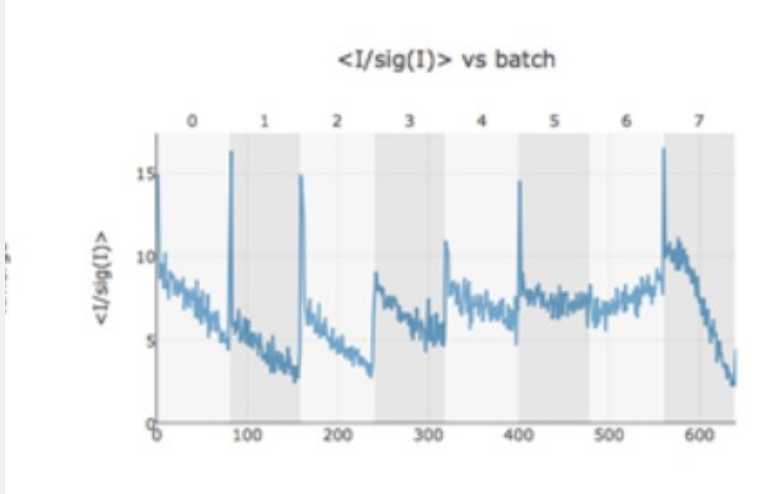




More analysis by resolution (all data)



Analysis by image number - looking for radiation damage



- ➢ Scale and Rmerge as well as I/sig(I) vs batch plots help to decide which data sets or frames to exclude because of radiation damage

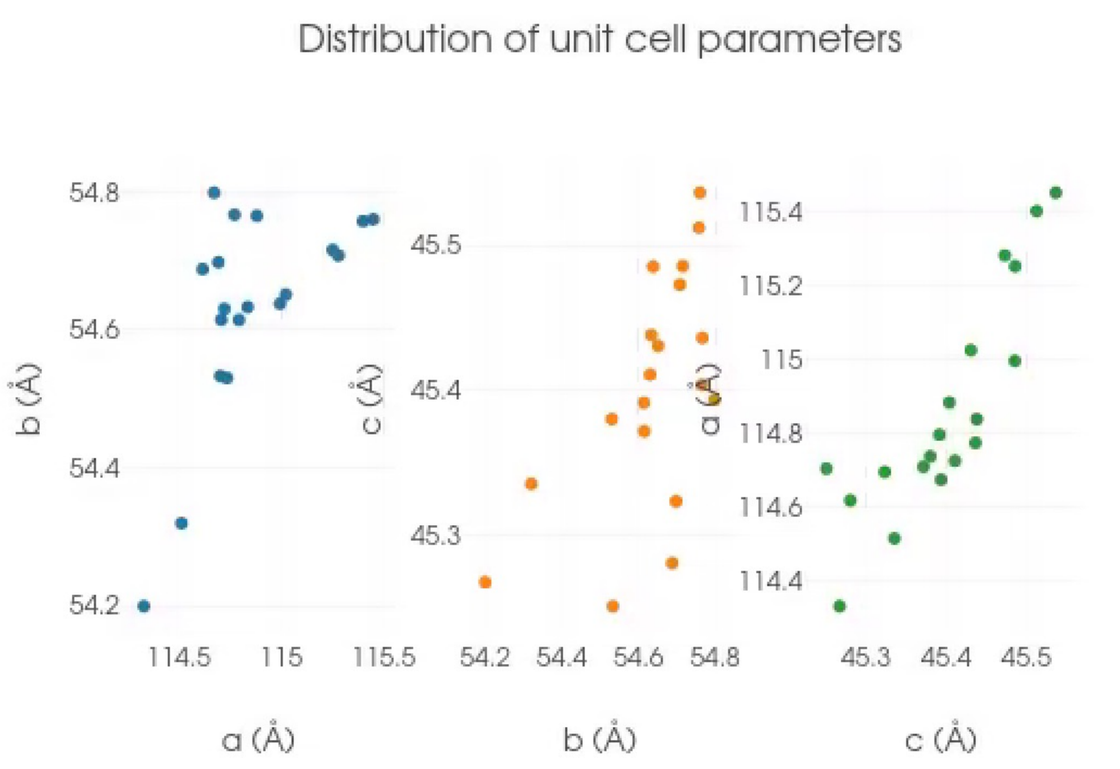
```
xia2.multiplex
$path_to_dials_results.{refl,expt}
dose=1,50
```

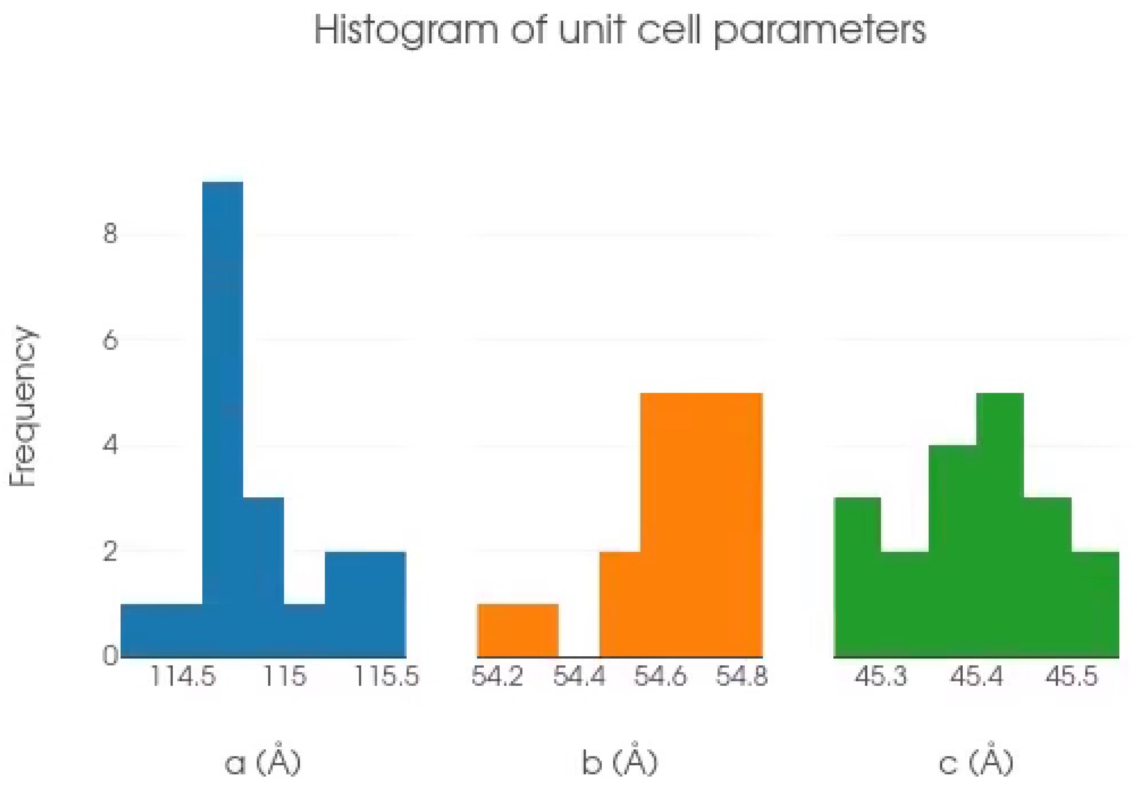
(to include the first 50 frames per data set)

```
xia2.multiplex
$path_to_dials_results.{refl,expt}
exclude_images=7:600:650
```

(to exclude the last 50 images from the 7th data set)

Unit cell analysis (summary)

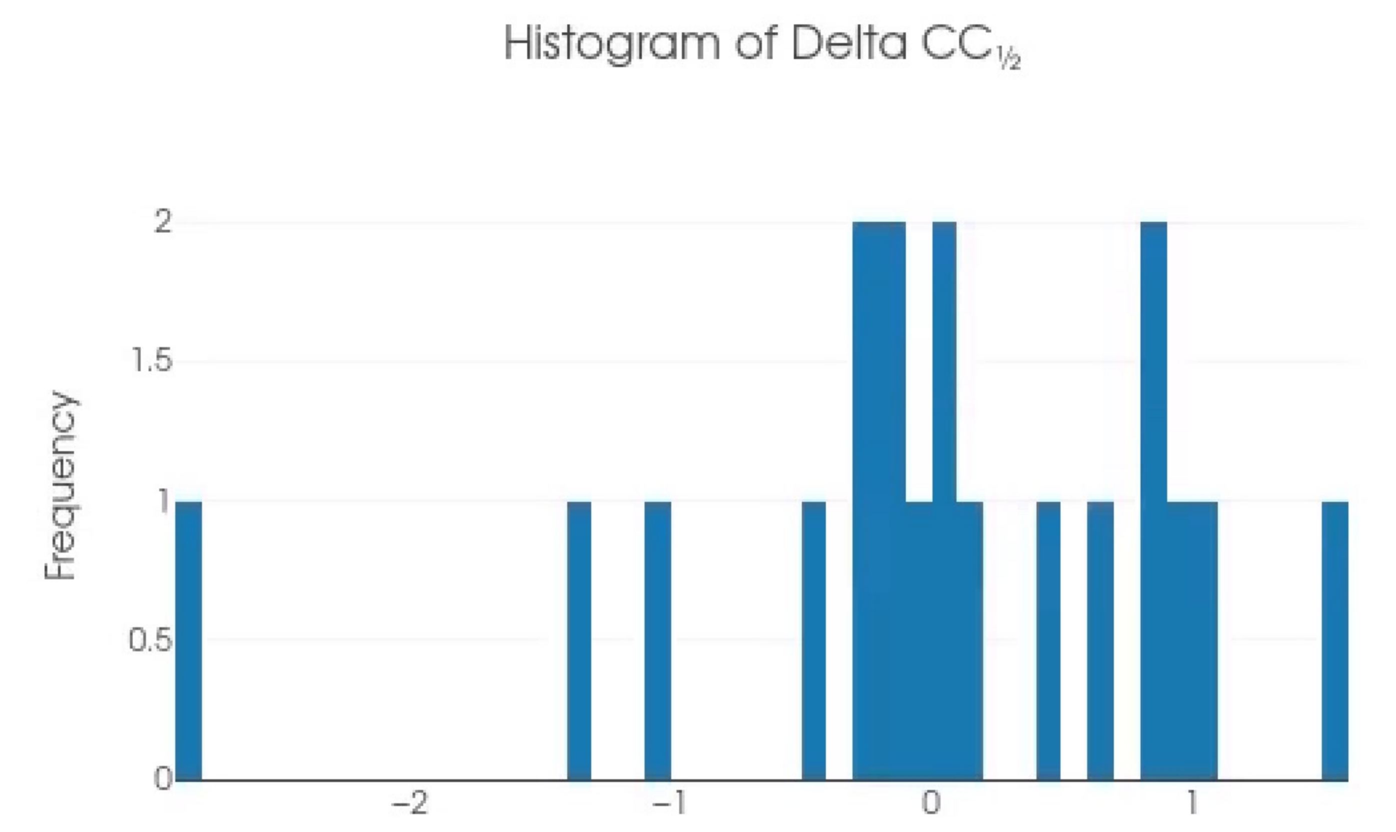




Unit cell analysis (summary)

- can help to find data sets which should be excluded due to non-isomorphism

$\Delta CC_{1/2}$ outlier analysis

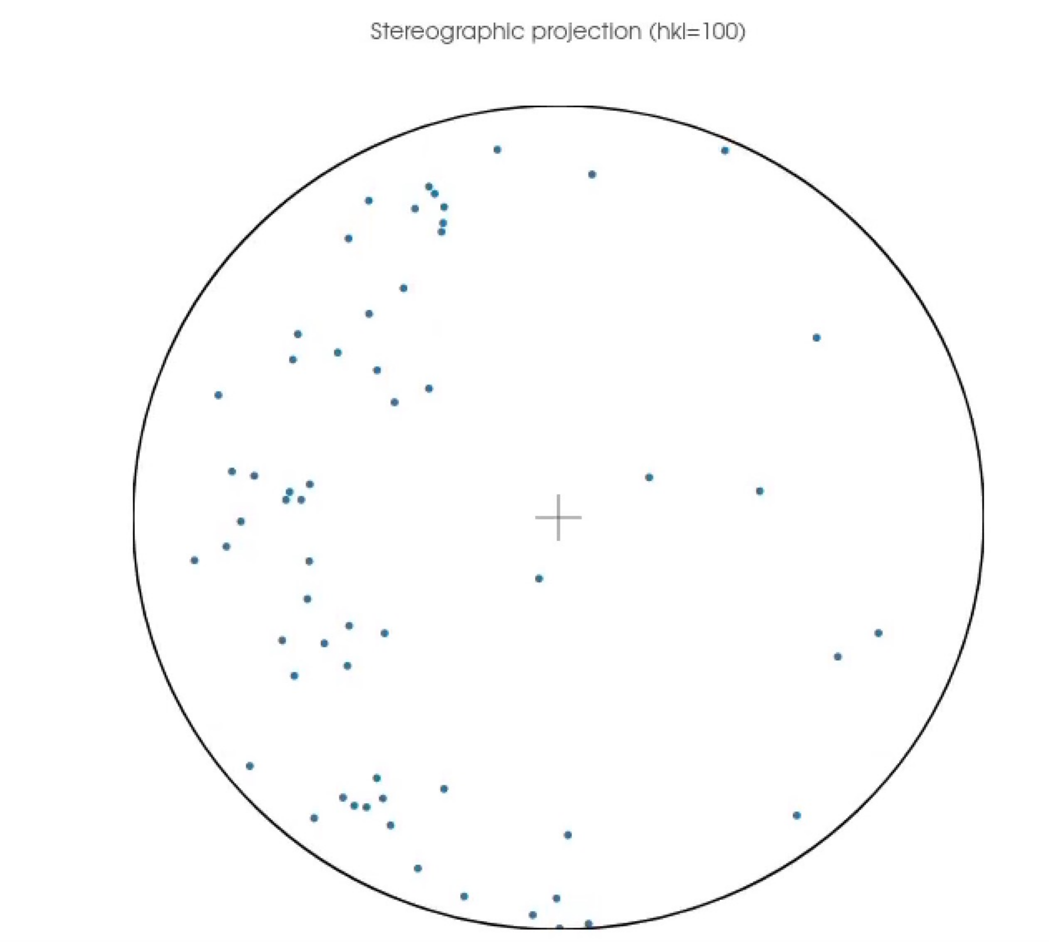


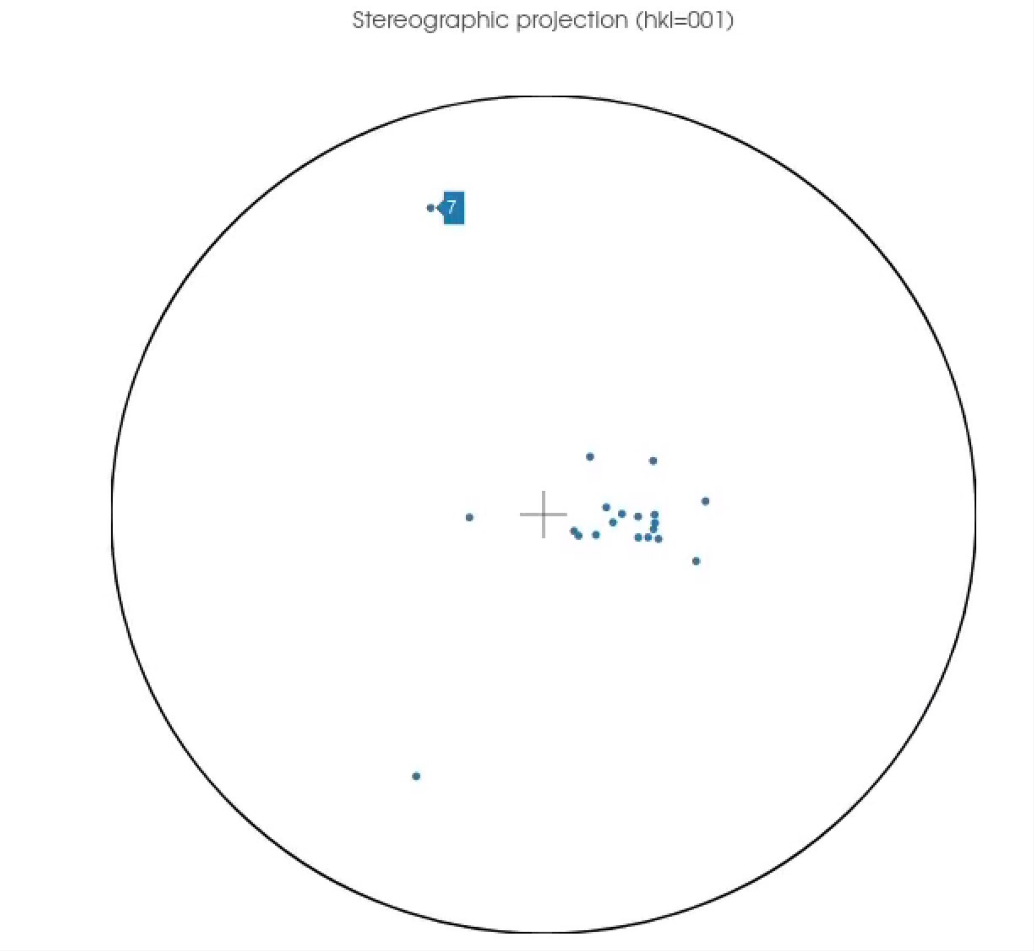
$$\Delta CC_{1/2\_i} = CC_{1/2\_i} - CC_{1/2\_\mathrm{overall}}.$$

- excluding the worst negative $\Delta CC_{1/2}$ can improve the overall result

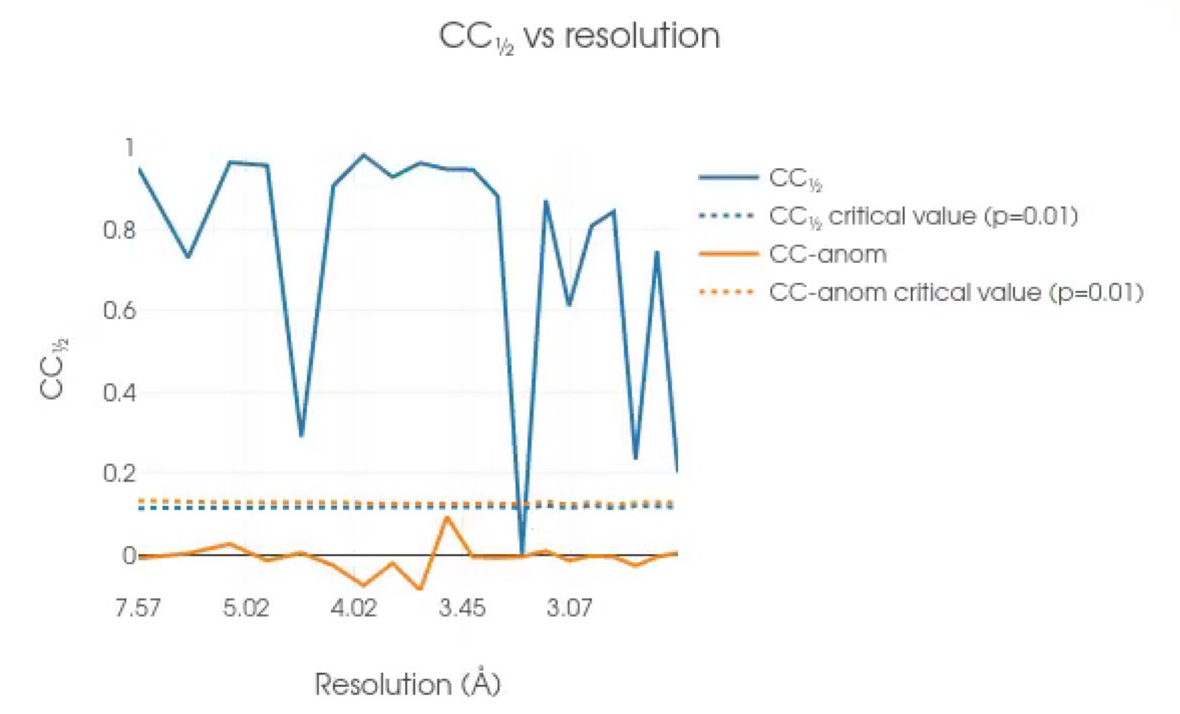
Assmann et al, . Appl. Cryst. (2016). 49, 1021–1028

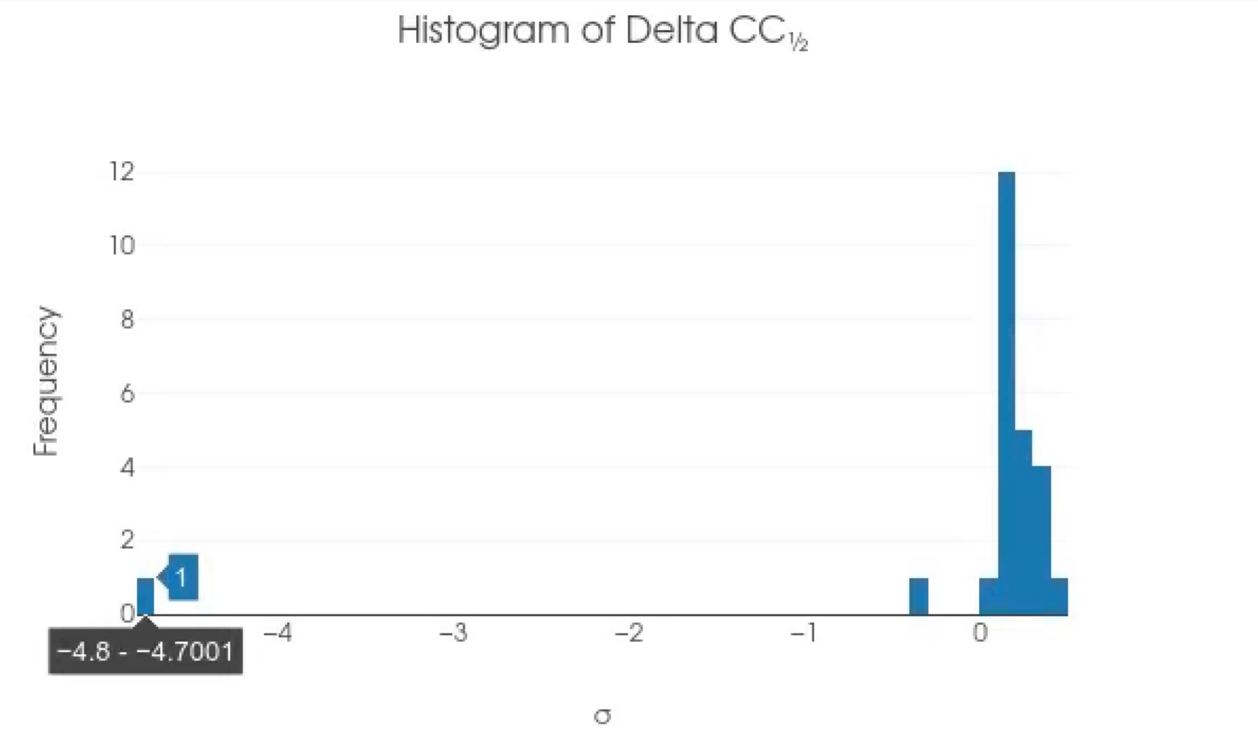
Orientation analysis (in summary)





Example of something going wrong





General advice for processing multiple data sets

- try to process data on-the-fly and keep an eye on completeness and resolution **as a first step!**
- check the analysis carefully for radiation damage and non-isomorphism and identify data sets to cut or exclude
- before you exclude data, check in the orientation analysis whether this data set is critical for completeness
- rerun xia2.multiplex manually and look for improvements

Take-home messages

- plan your experiment and choose your weapon
- collect your data carefully (parameters, path names)
- use auto-processing (xia2.multiplex), but use it wisely

https://www.snorgtees.com/shoot-first-ask-questions-later

Acknowledgements

Robin Owen
Danny Axford
Sam Horrell

Richard Gildea
James Beilsten-Edmands
Nick Devenish
Graeme Winter

Adam Crawshaw

DIALS
Diffraction Integration for Advanced Light Sources

diamond

Thank you for your attention!