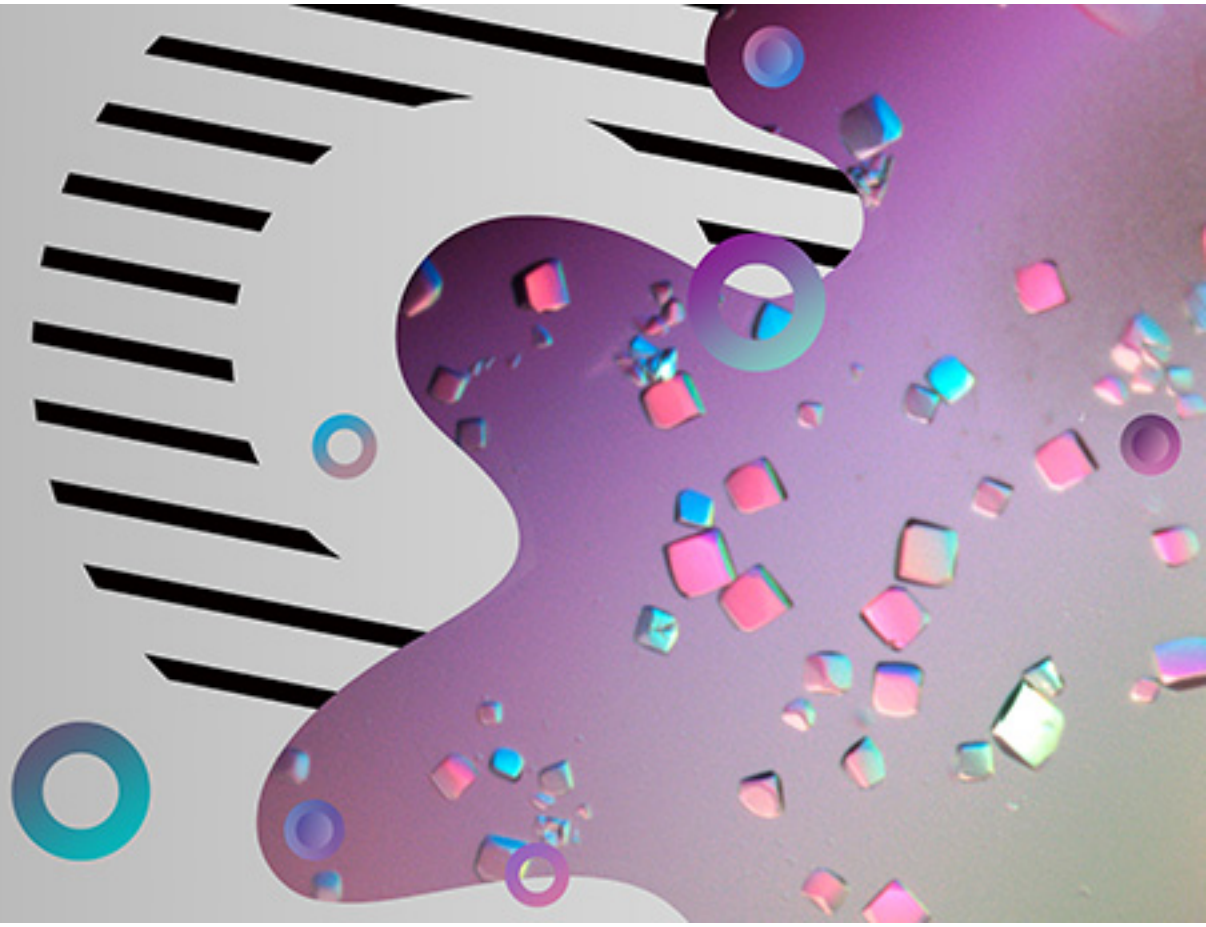
Multi-crystal approaches



CCP4 Workshop

02.12.2020

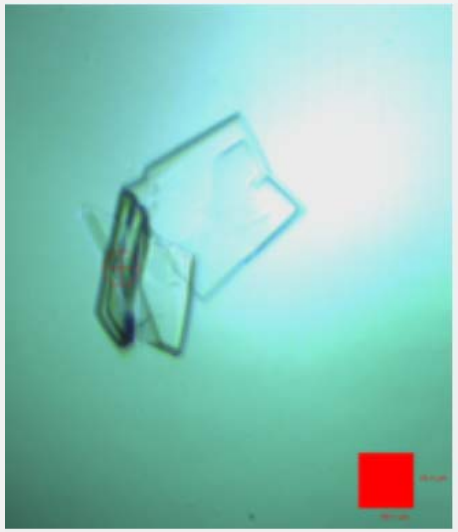
Selina Storm

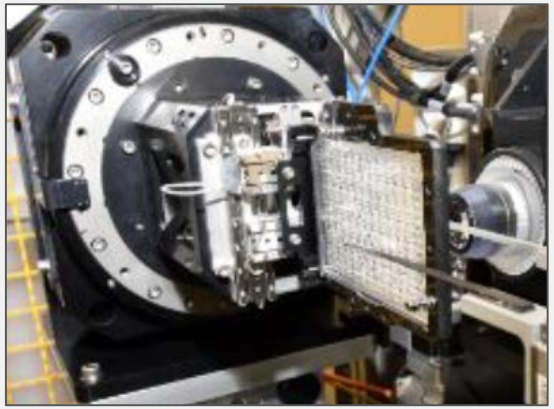
Project Manager for EMBL@PETRA IV

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

EMBL

Multiple crystal approaches

hundreds

tenths

one

- rotational crystallography (standard loops, grids, *in situ)*
- ‘rotational serial’: a few degrees of data with serial methodology, e.g. rotating chips [1]
- serial crystallography (on chips, in liquid jets …)

tenths

hundreds

thousands

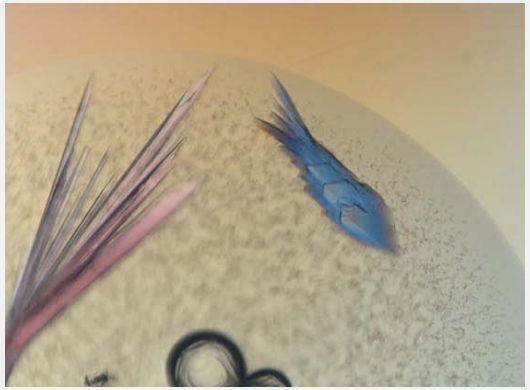
[1] Wierman et al,, IUCrJ (2019). 6, 305–316

EMBL

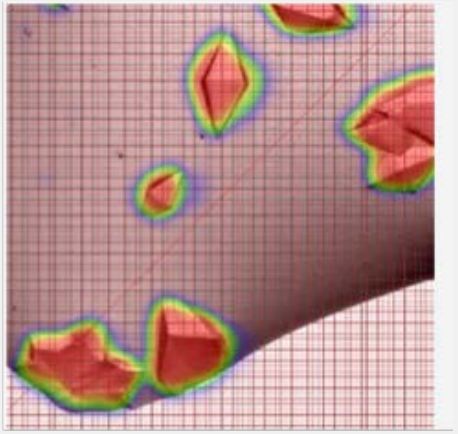
What is the aim of your experiment?

Choose your weapon accordingly to the aim/situation!

- lots of big, badly diffracting crystals:
  collect small sweeps in loops and try to get higher resolution



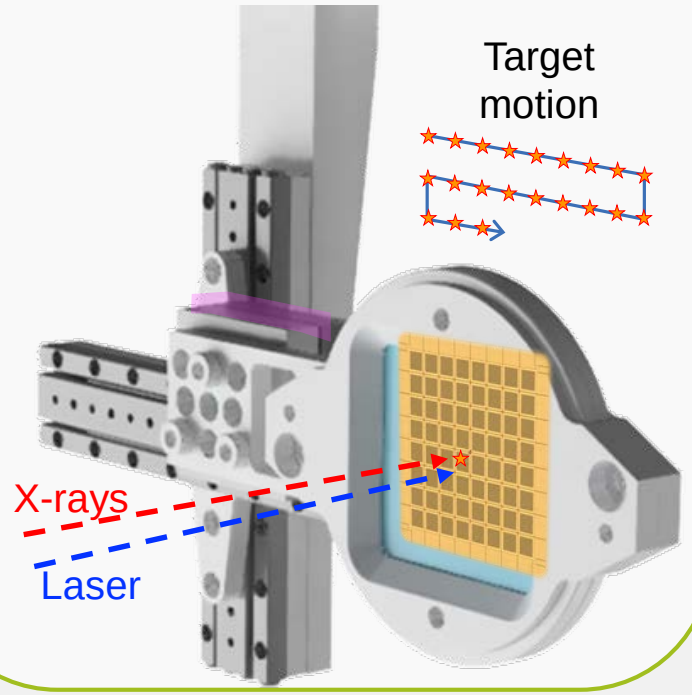
- brittle crystals which don't like to be taken out of the plate:
  grow them in plates which can be mounted on the beamline
  (VMXi, I24@Diamond, P14@PETRA III, X06DA@SLS...)



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

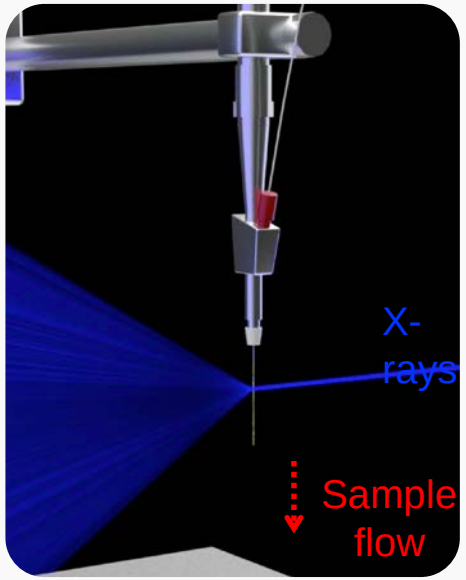
What is the aim of your experiment?

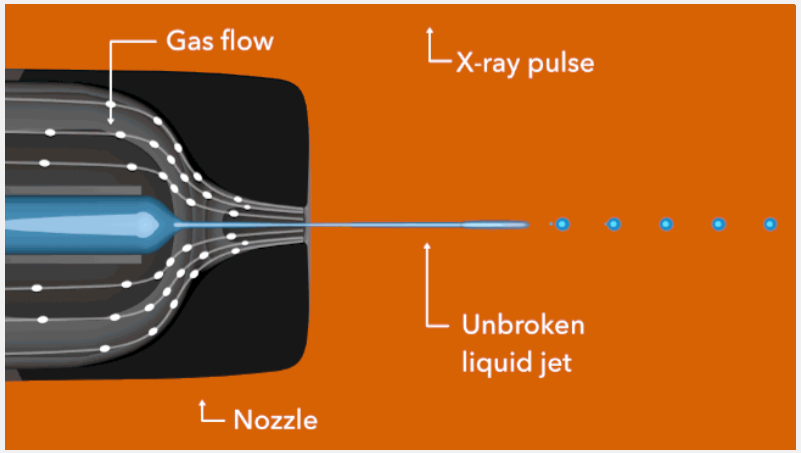
Fixed targets



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

LCP extruder





Courtesy Robin Owen, Pierre Aller

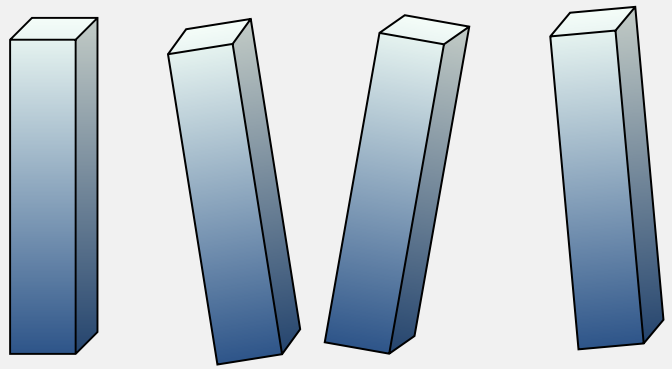
EMBL

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)



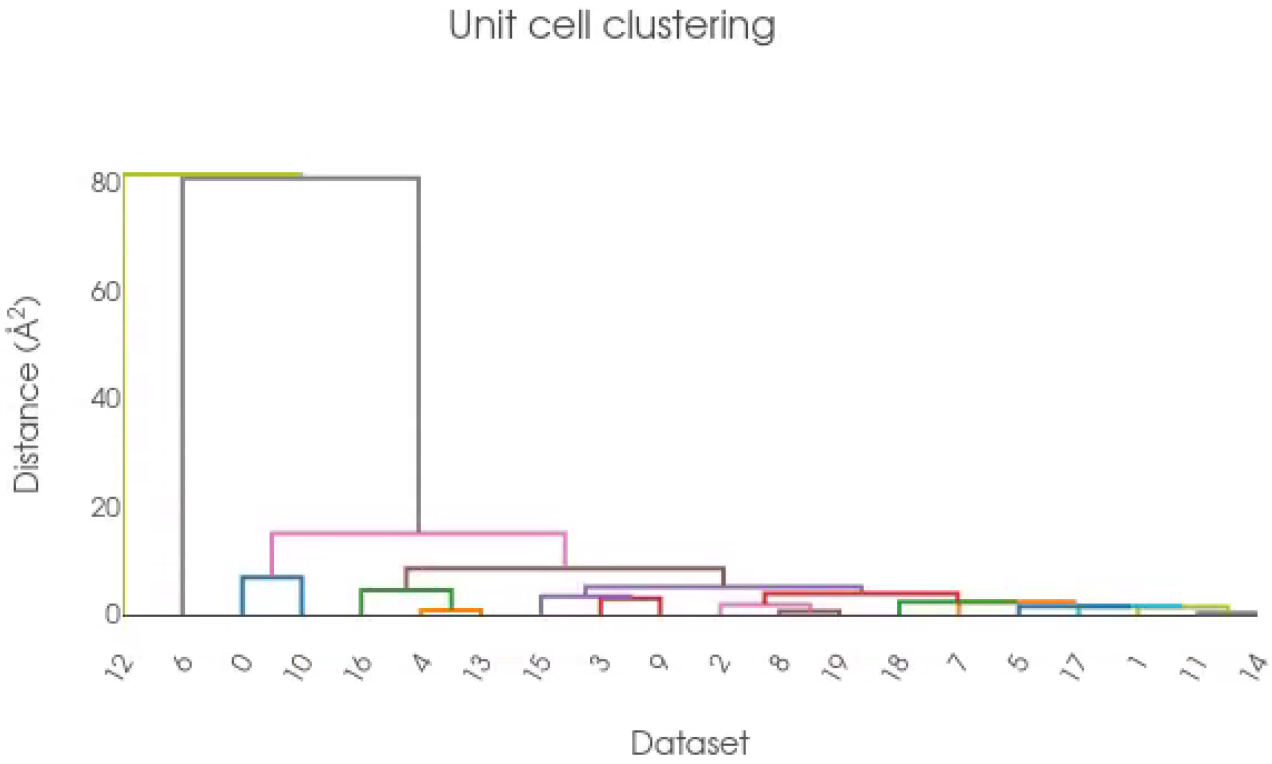
EMBL

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

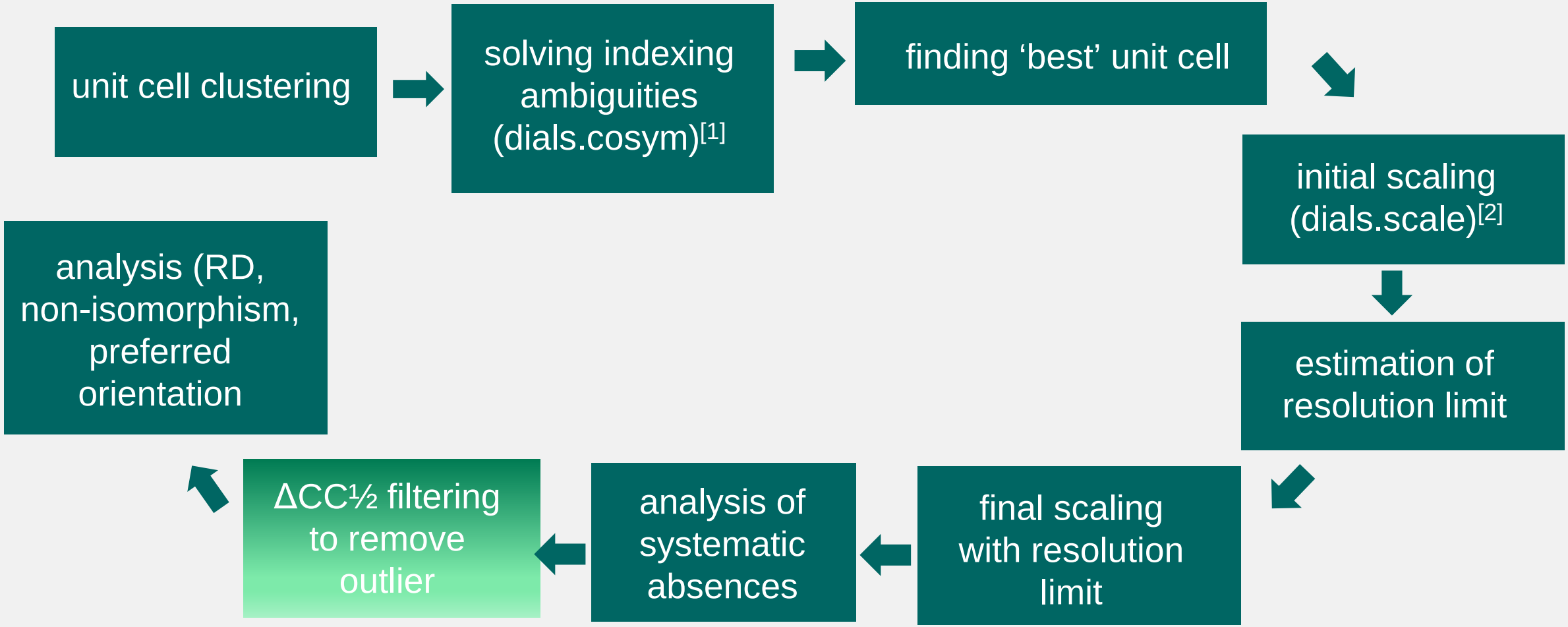


EMBL

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?

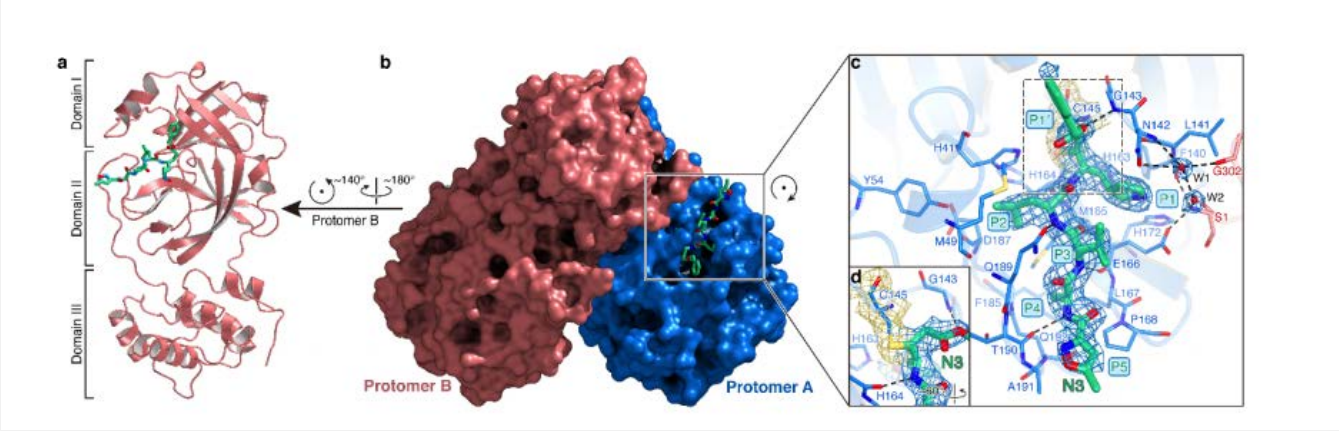


[1] Gildea & Winter, Acta Cryst. (2018). D74, 405-410
[2] Beilsten-Edmands et al., Acta Cryst. (2020). D76, 385-399

EMBL

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



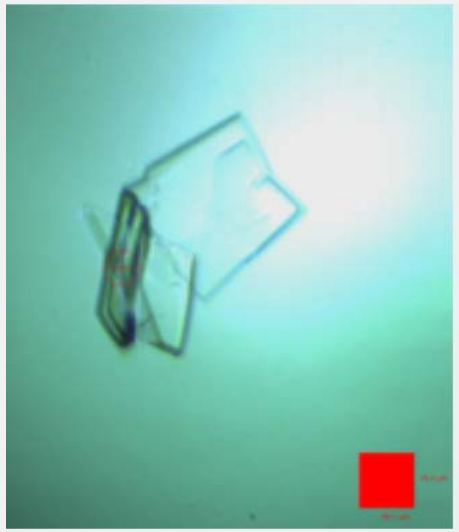
- active site empty and solvent accessible - perfect for fragment screening
- more than 2000 fragments screened, including Moonshot compounds

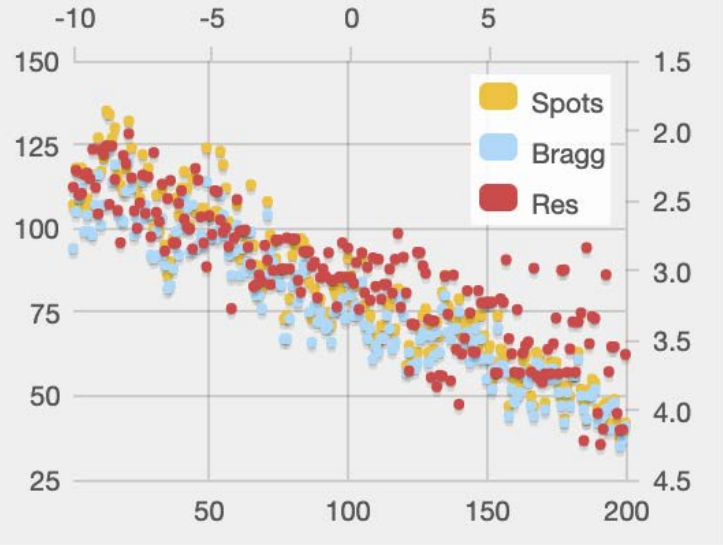
[1] Zhenming Jin et al., Nature volume 582, pages 289–293 (2020)
[2] Douangamath et al *Nature Communications* **volume 11**, Article number: 5047 (2020)

EMBL

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

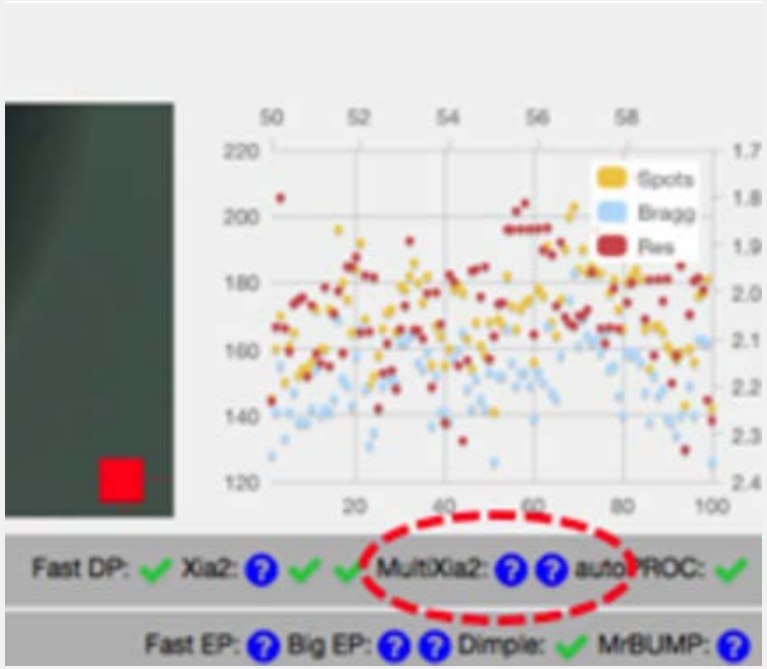




EMBL

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
- raw data all need to go into the same folder
- two multiplex auto-processing runs in ISPyB:



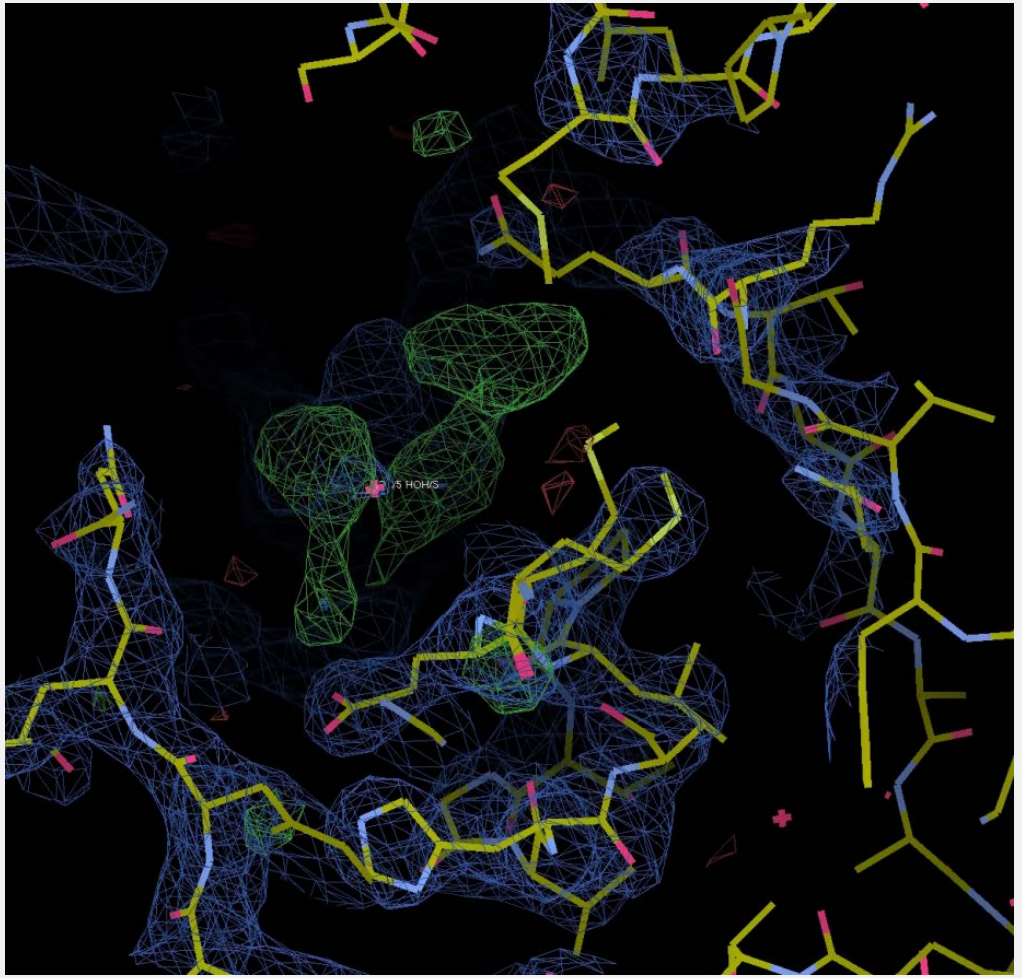
https://www.diamond.ac.uk/Instruments/Mx/I24/Status/multi-xtal-proc.html

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



EMBL

Processing via command line

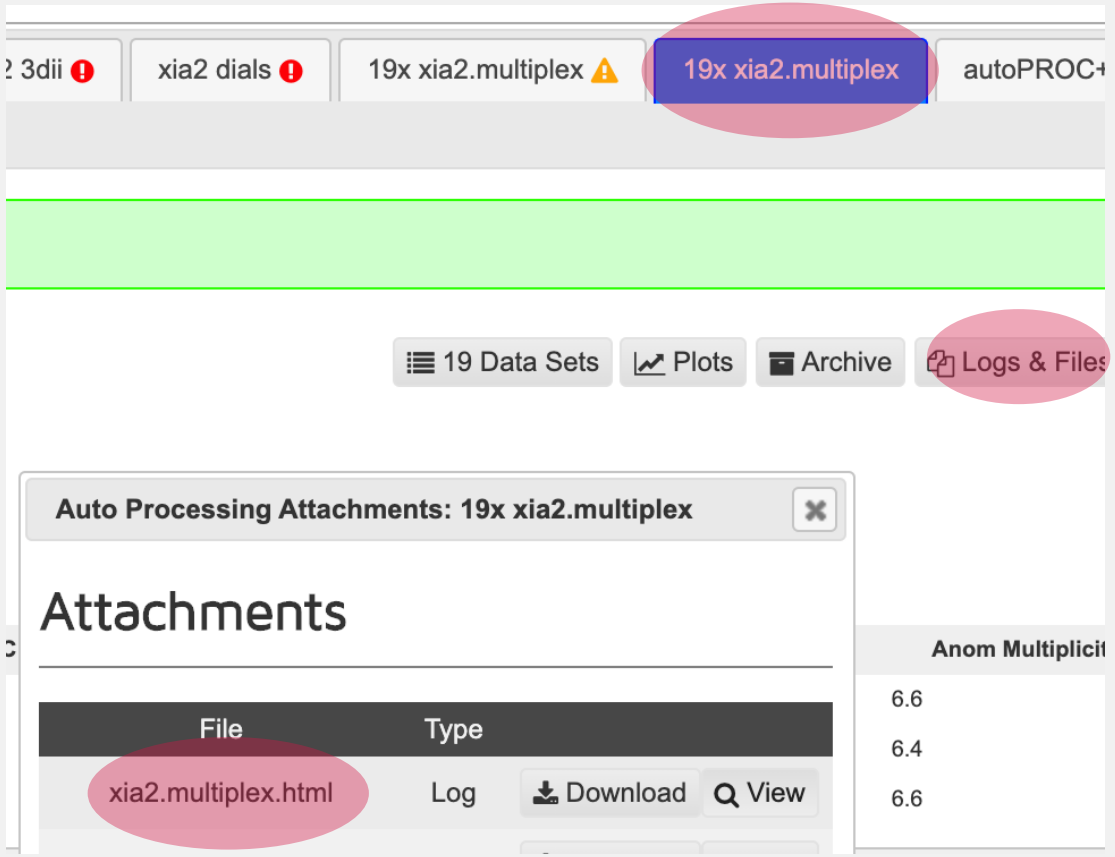
- required: path to integrated dials results

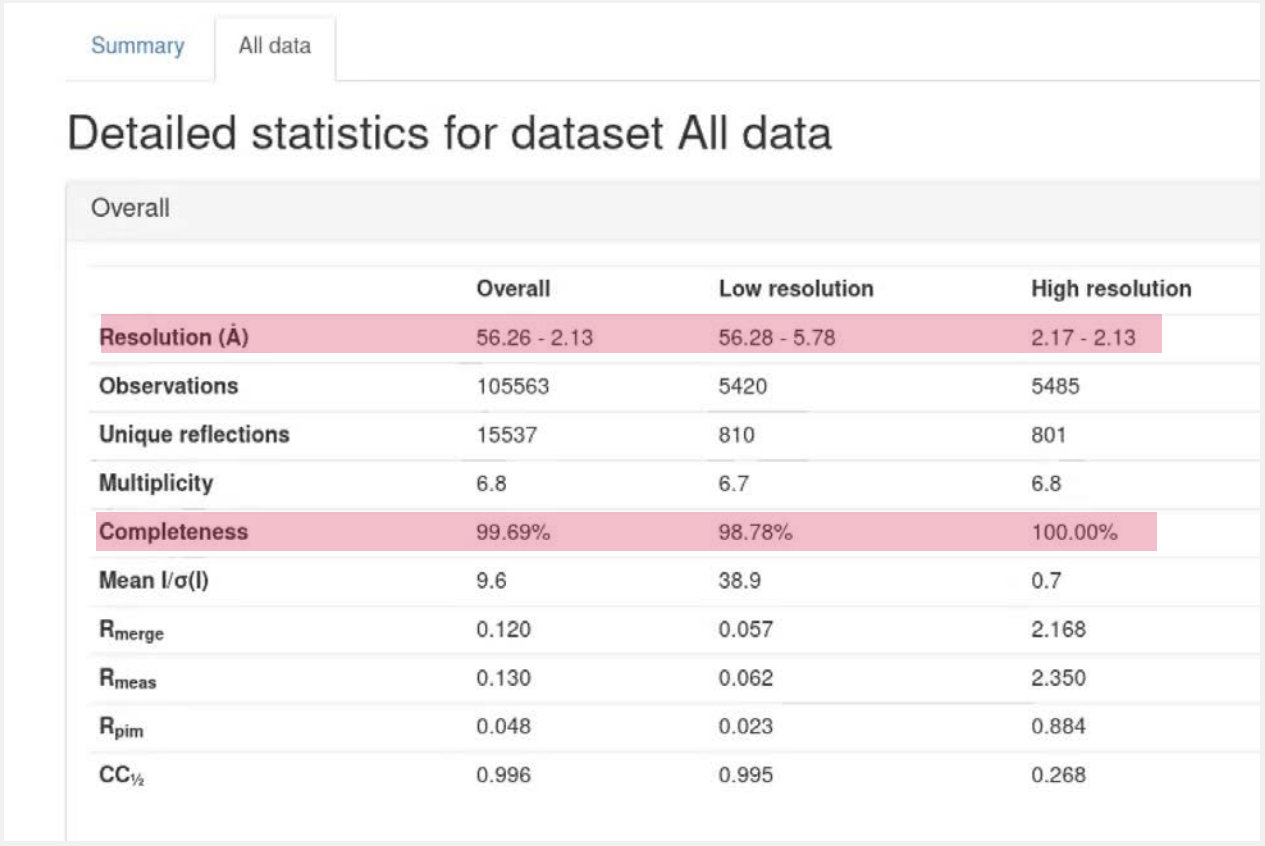
```
xia2.multiplex $path_to_dials_results.{refl,expt} options
```

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

EMBL

Understanding the results: the multiplex html report

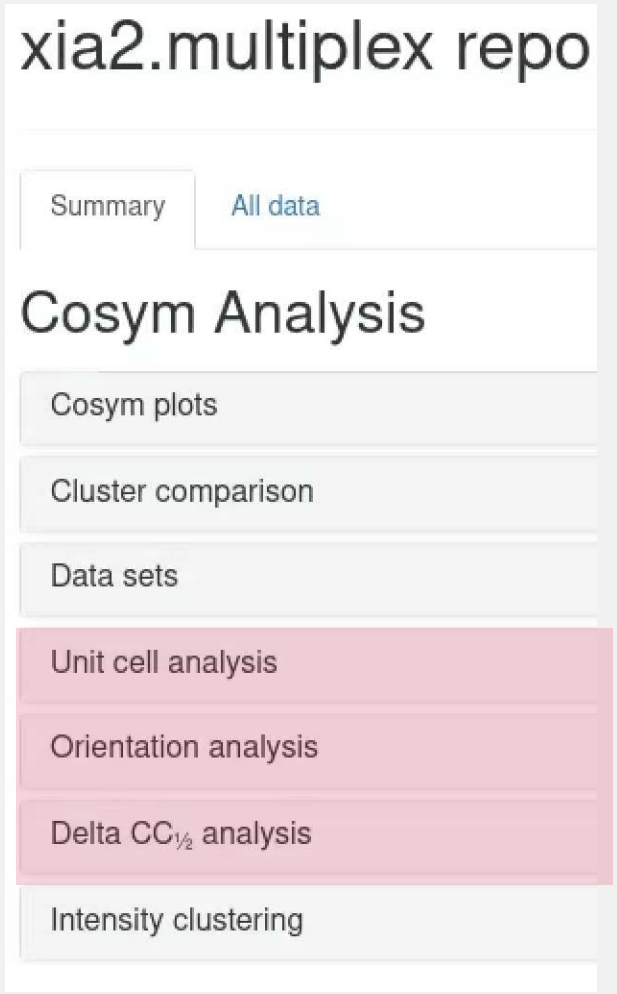


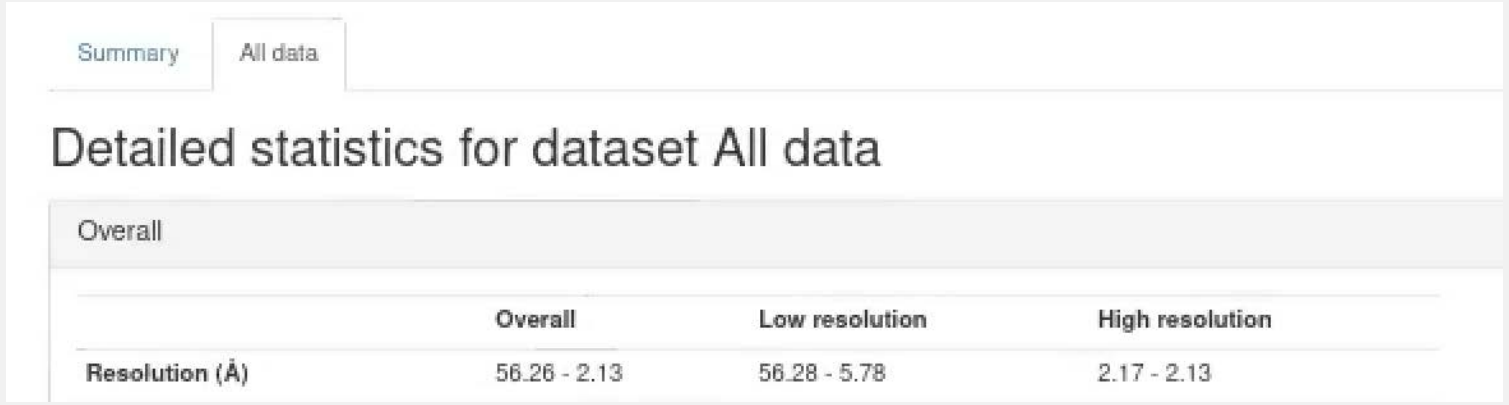


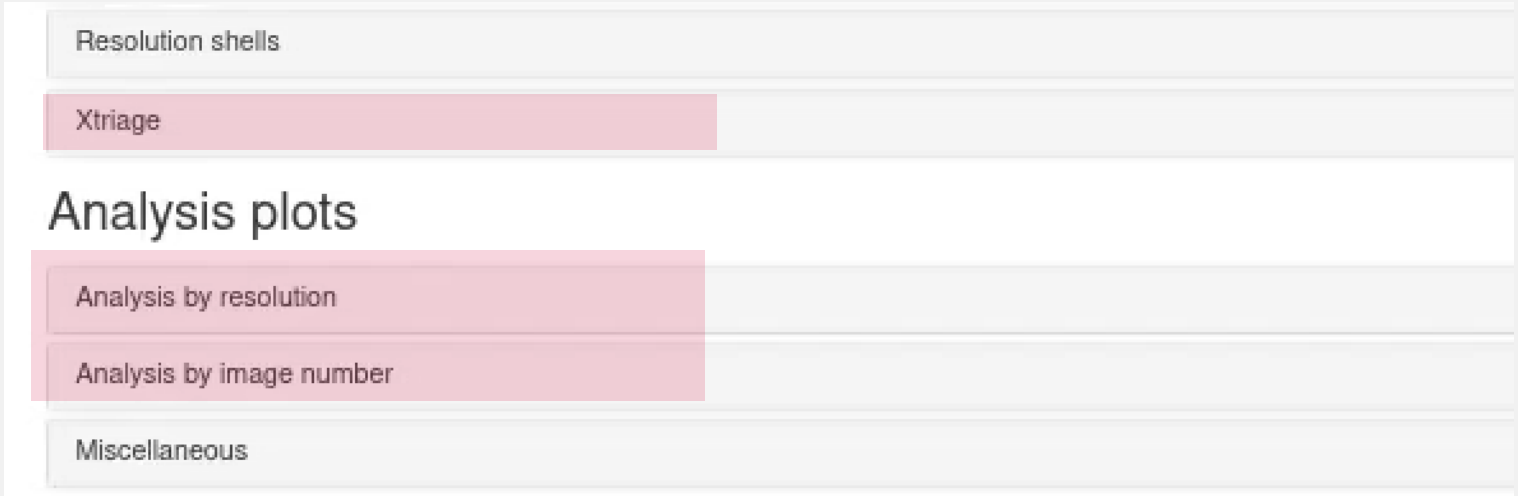
	Overall	Low resolution	High resolution
Resolution (Å)	56.26 - 2.13	56.28 - 5.78	2.17 - 2.13
Observations	105563	5420	5485
Unique reflections	15537	810	801
Multiplicity	6.8	6.7	6.8
Completeness	99.69%	98.78%	100.00%
Mean I/σ(I)	9.6	38.9	0.7
R_{merge}	0.120	0.057	2.168
R_{meas}	0.130	0.062	2.350
R_{pim}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

EMBL

Xia2.multiplex.html

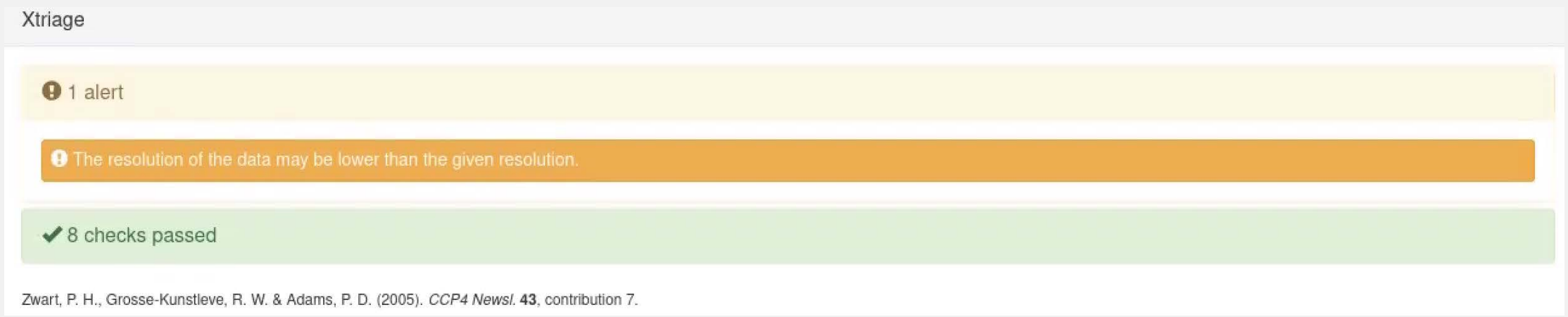


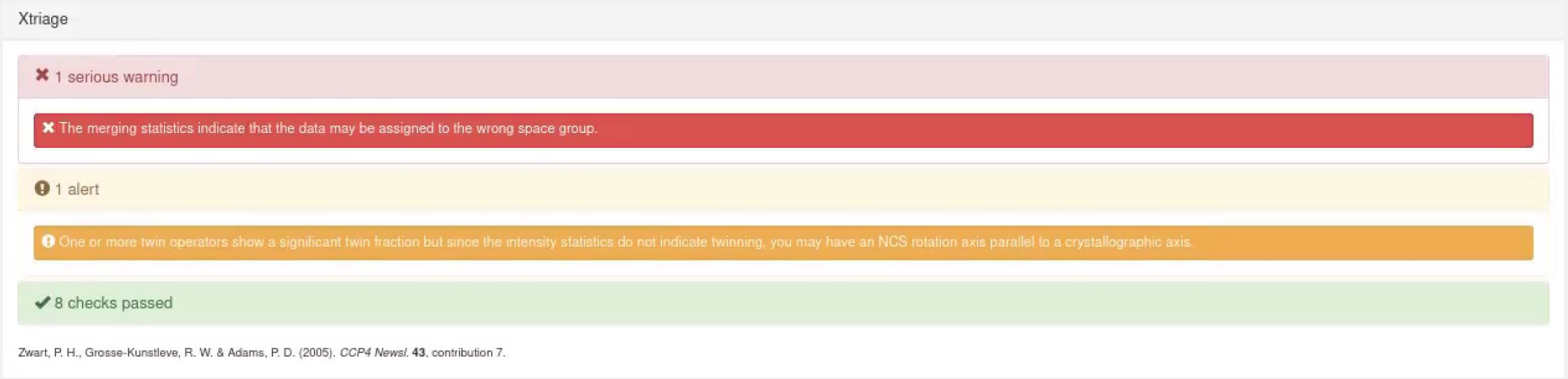




EMBL

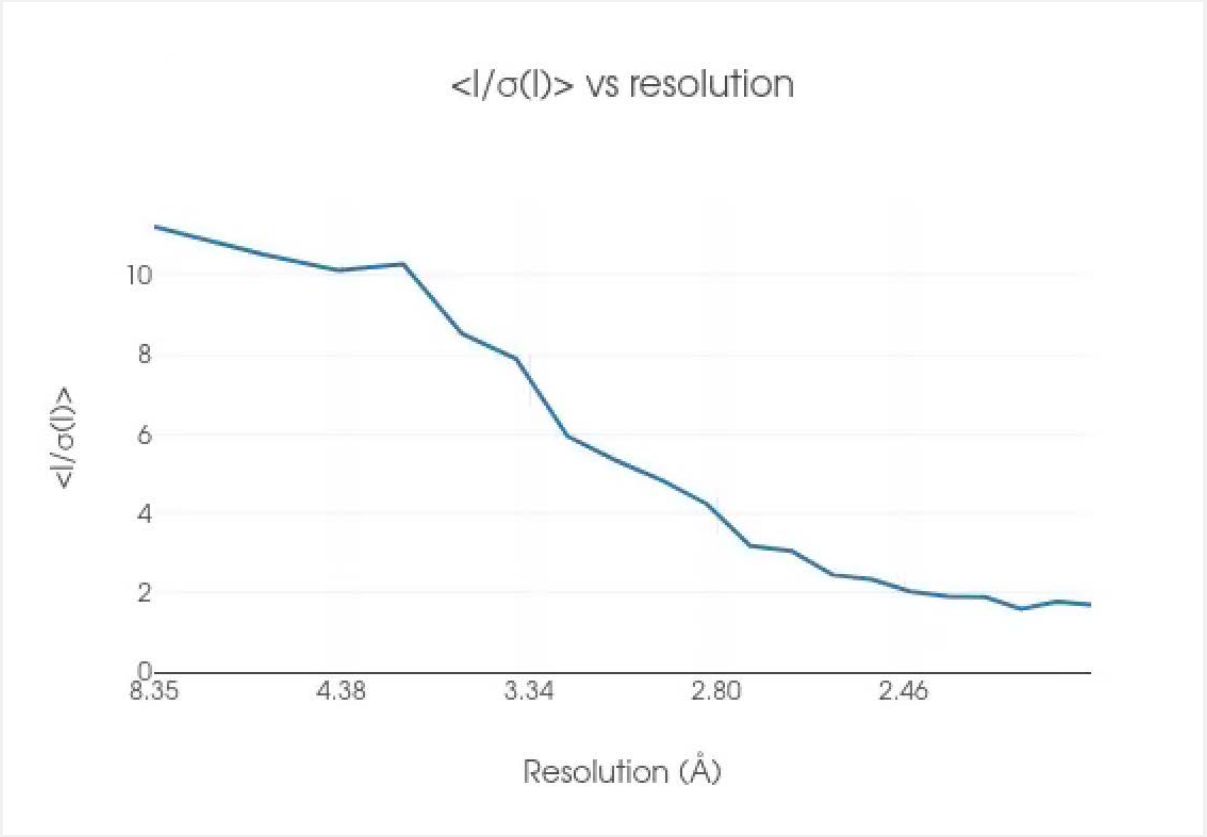
Xtriage (all data)

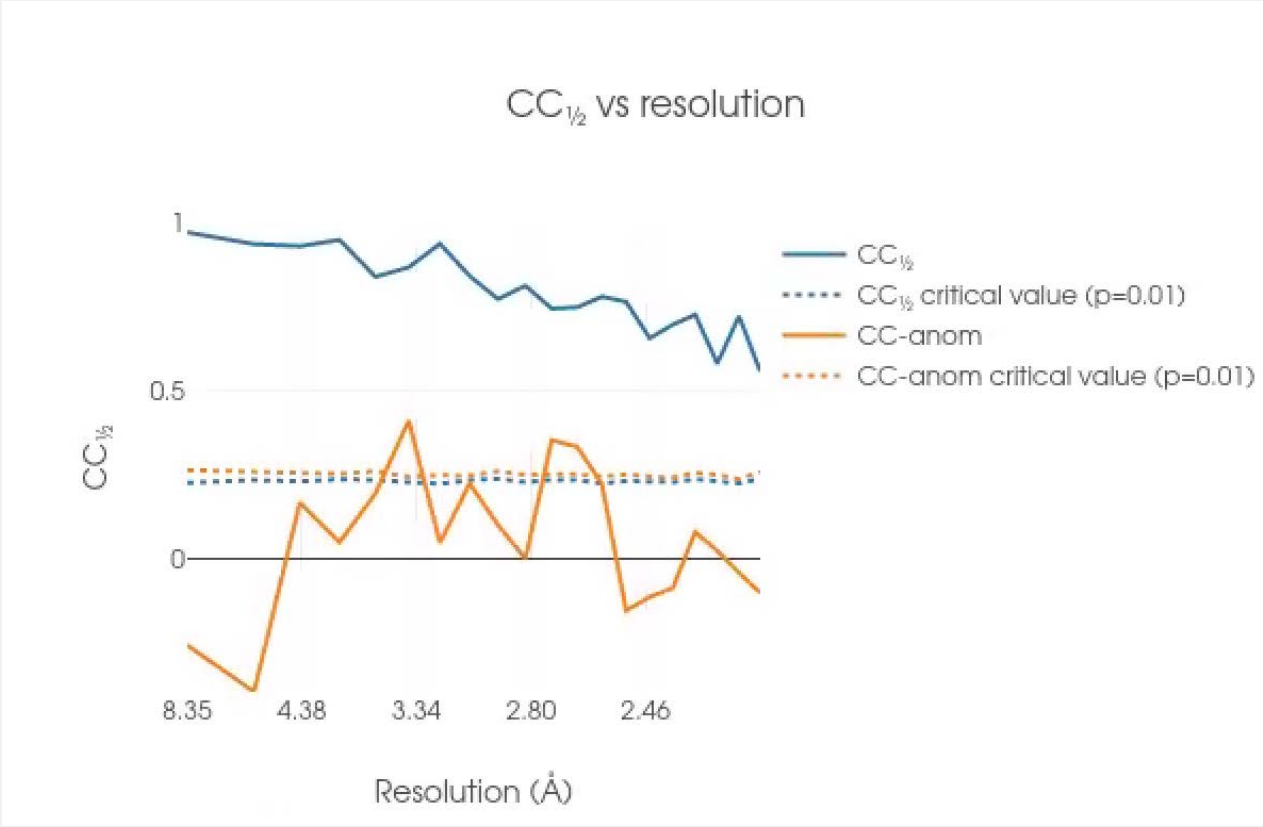




EMBL

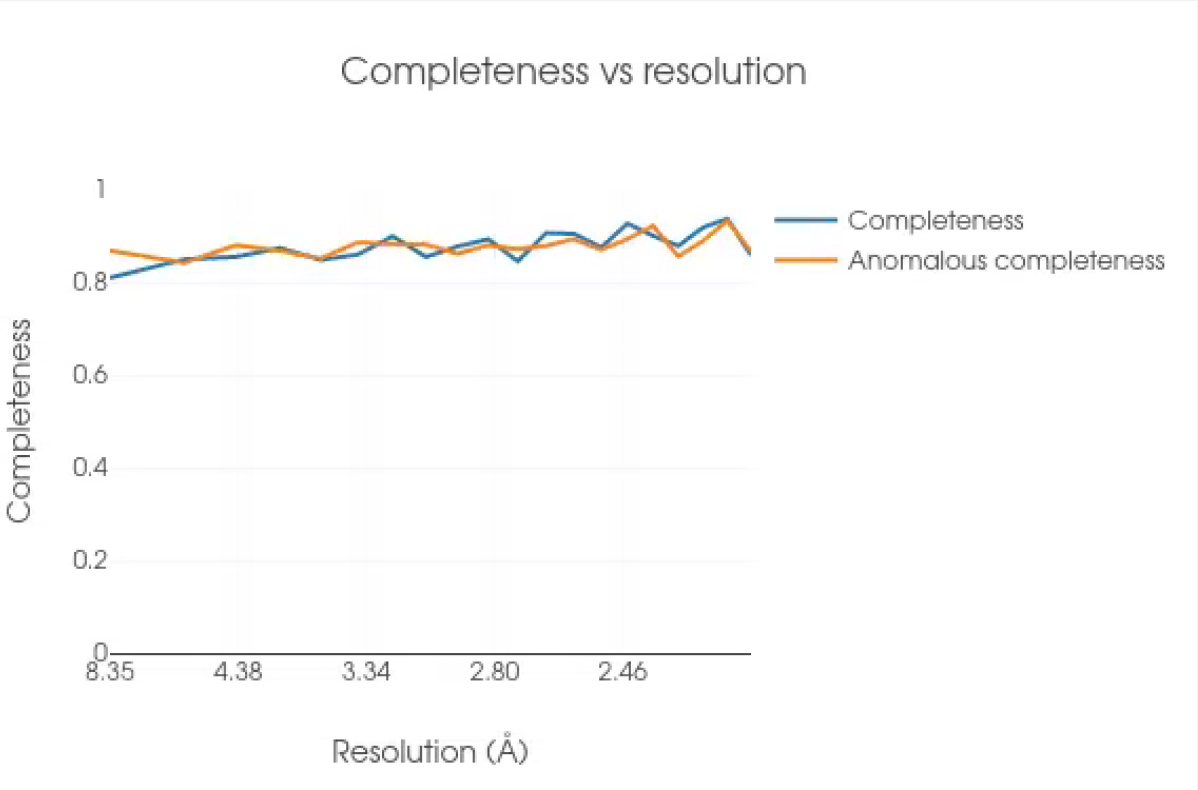
Analysis by resolution (all data)





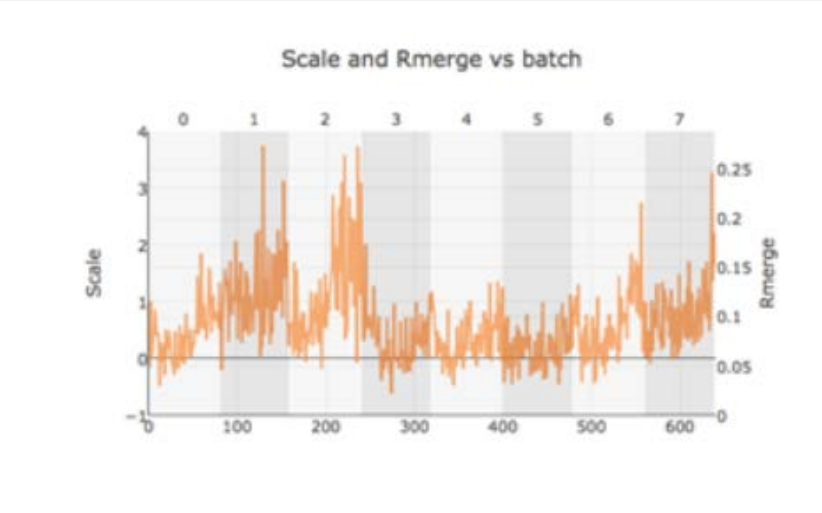
EMBL

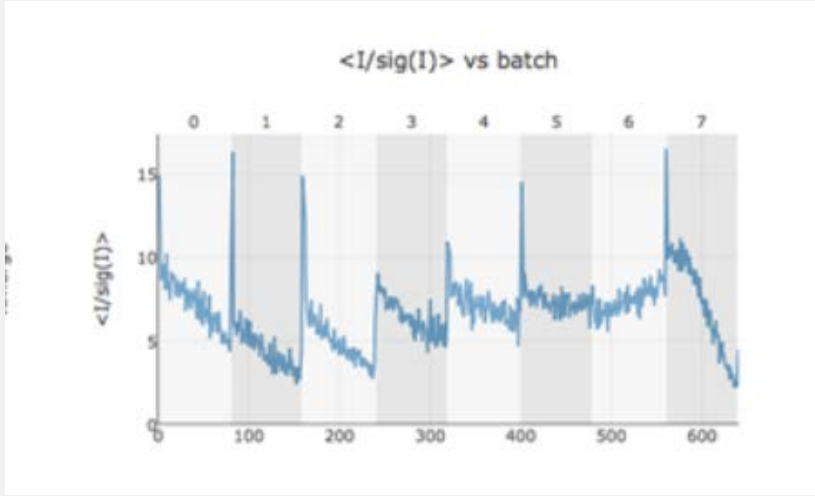
More analysis by resolution (all data)



EMBL

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
```

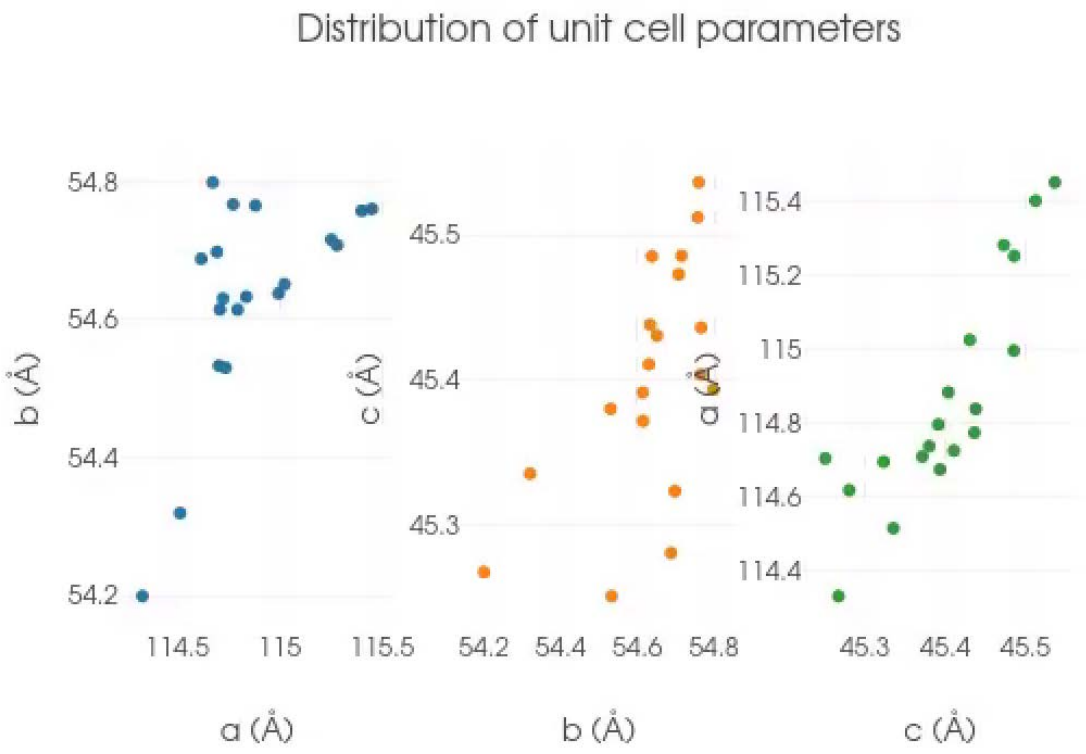
(to include the first 50 frames per data set)

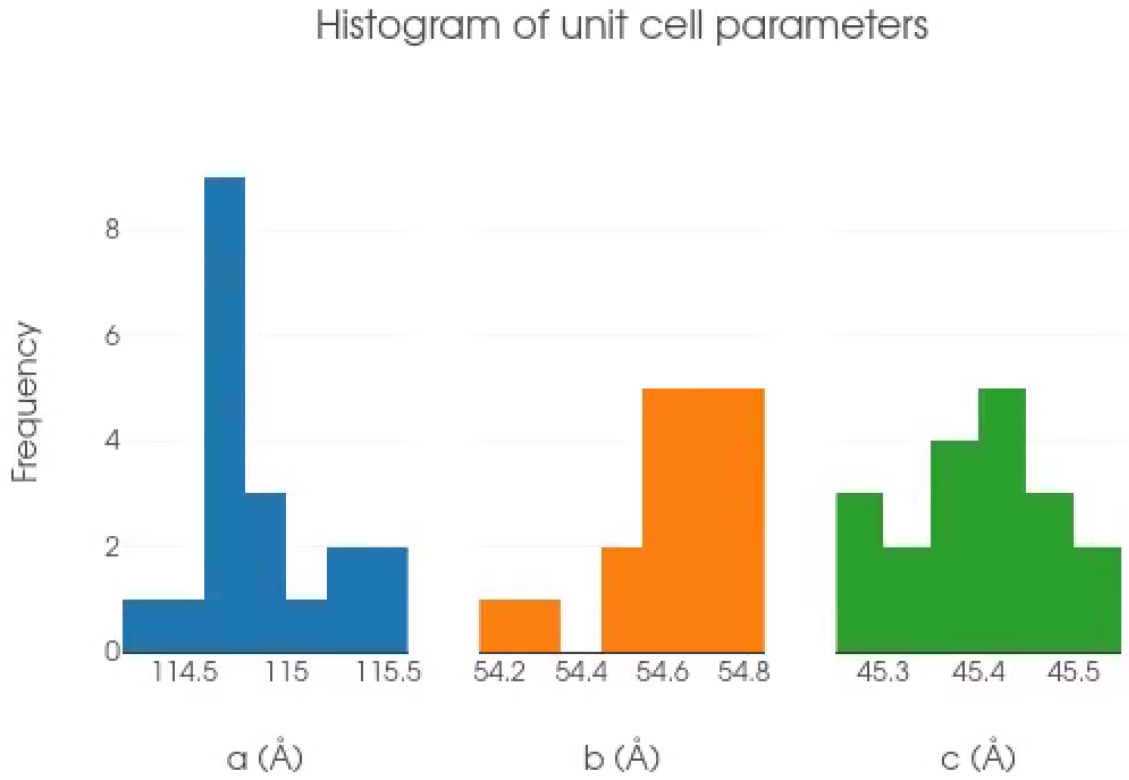
```
xia2.multiplex
$path_to_dials_results.{refl,expt}
exclude_images=7:600:650
```

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

EMBL

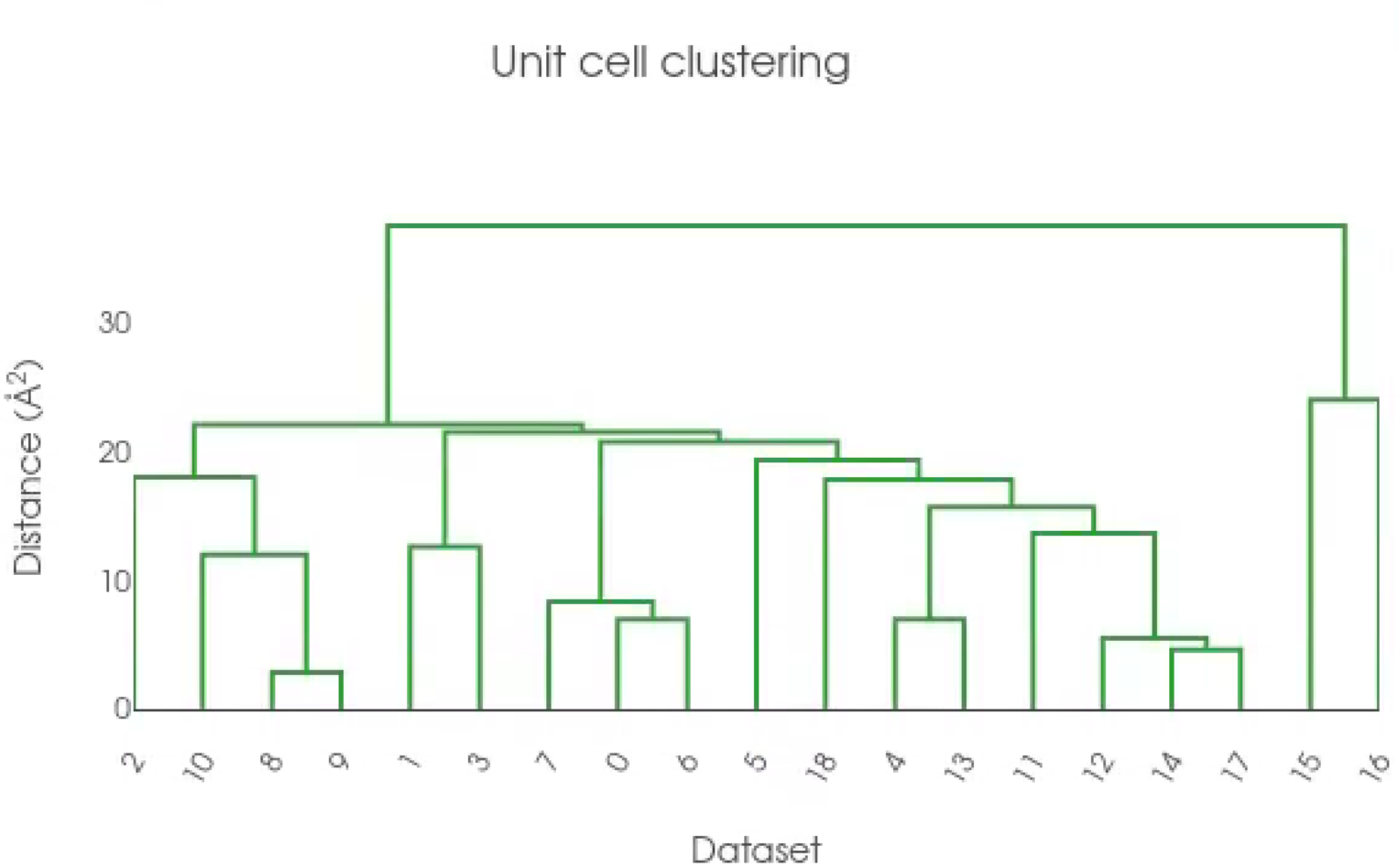
Unit cell analysis (summary)





EMBL

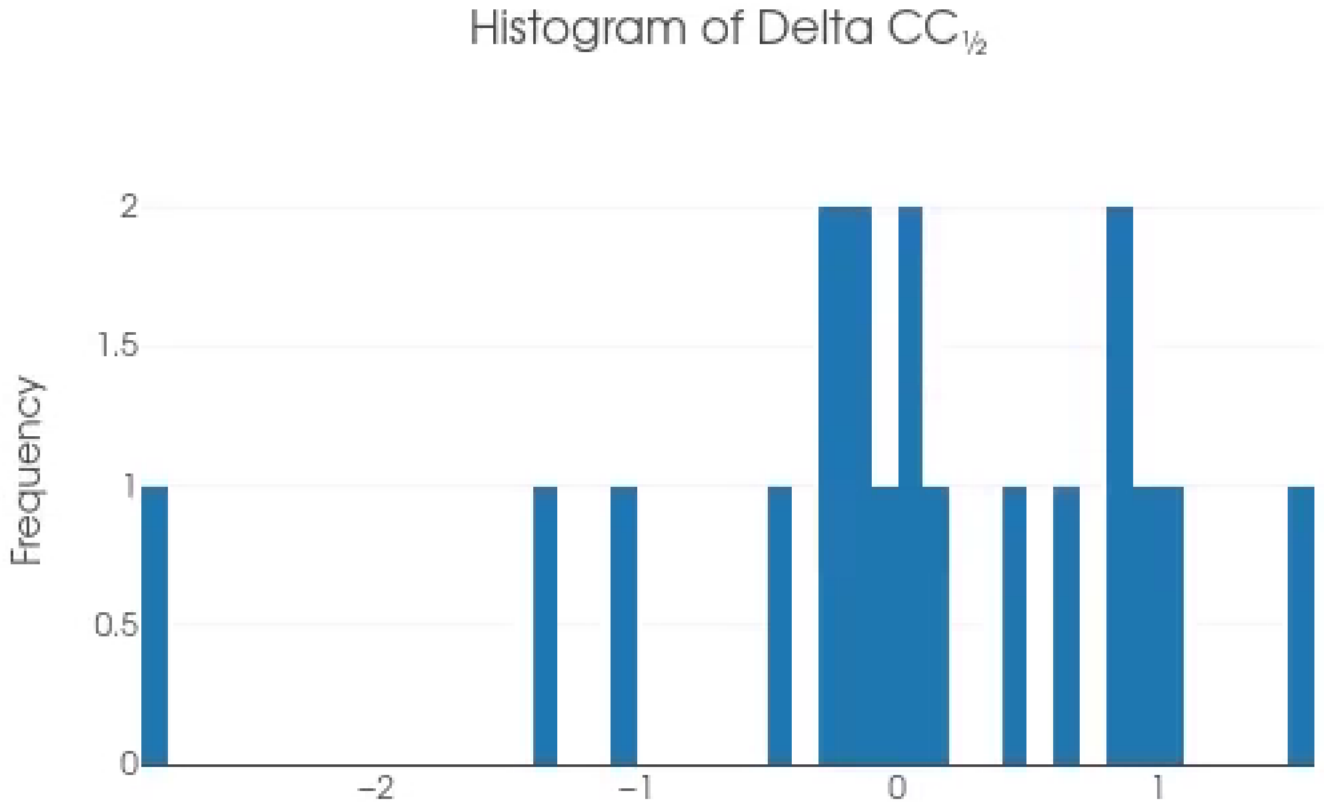
Unit cell analysis (summary)



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

EMBL

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



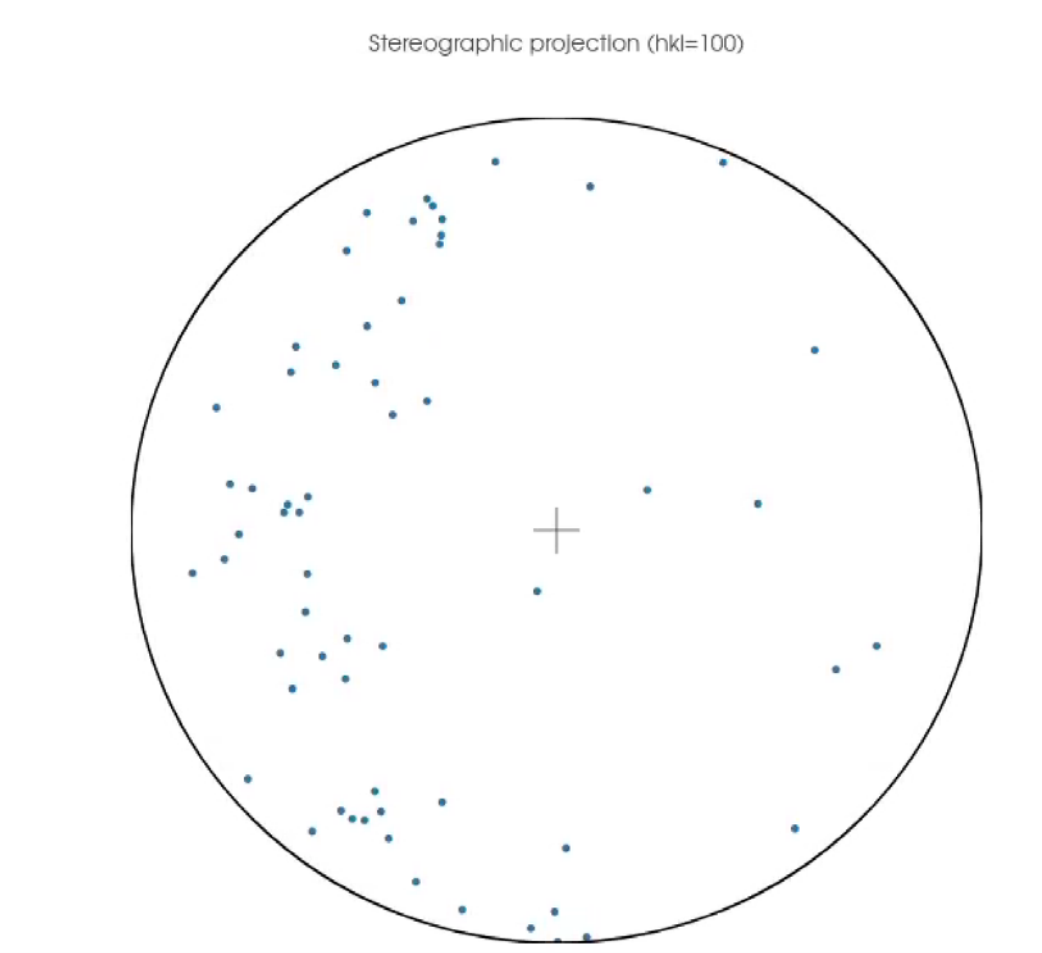
$$\Delta CC_{1/2\_i} = CC_{1/2\_i} - CC_{1/2\_overall}.$$

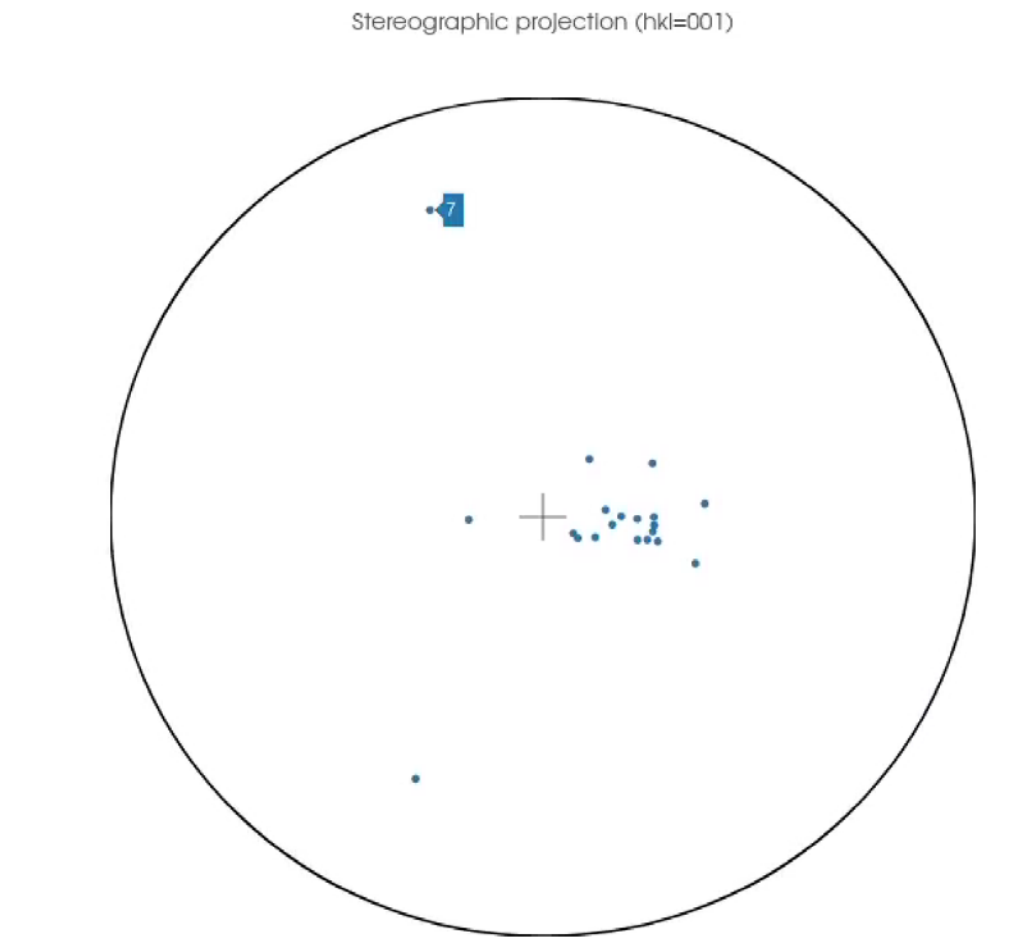
- excluding the worst negative $\Delta CC_{1/2}$ can improve the overall result

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

EMBL

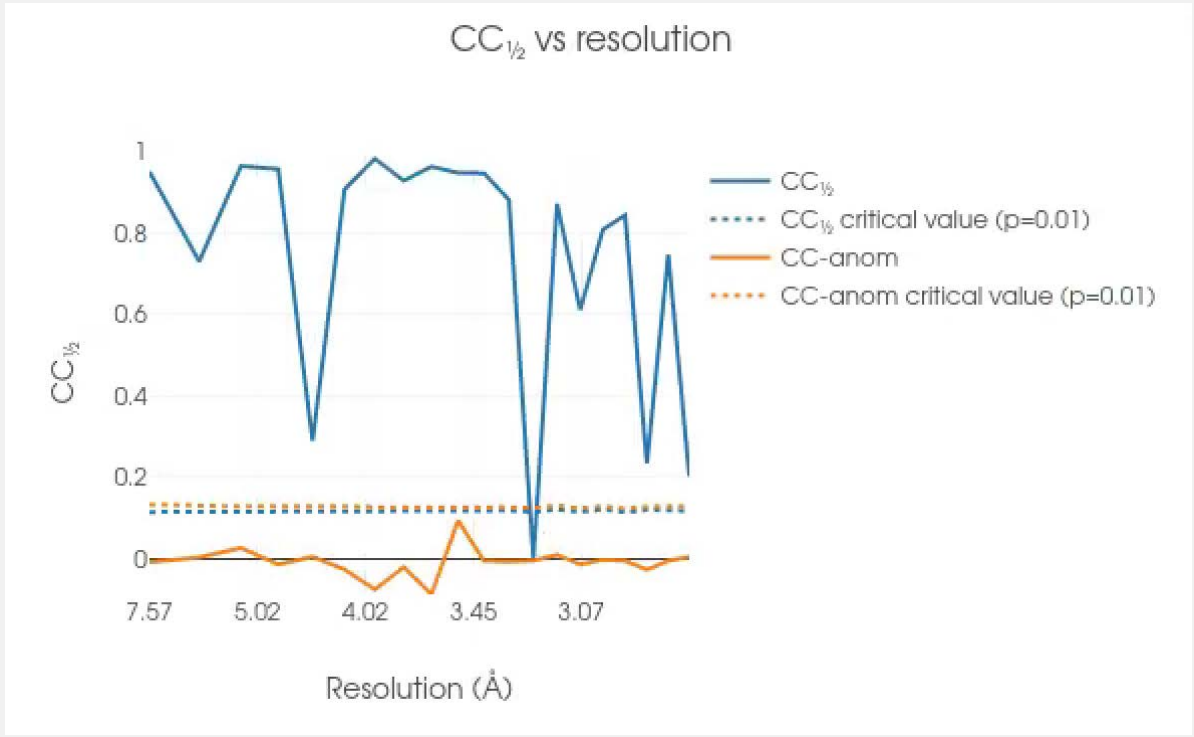
Orientation analysis (in summary)

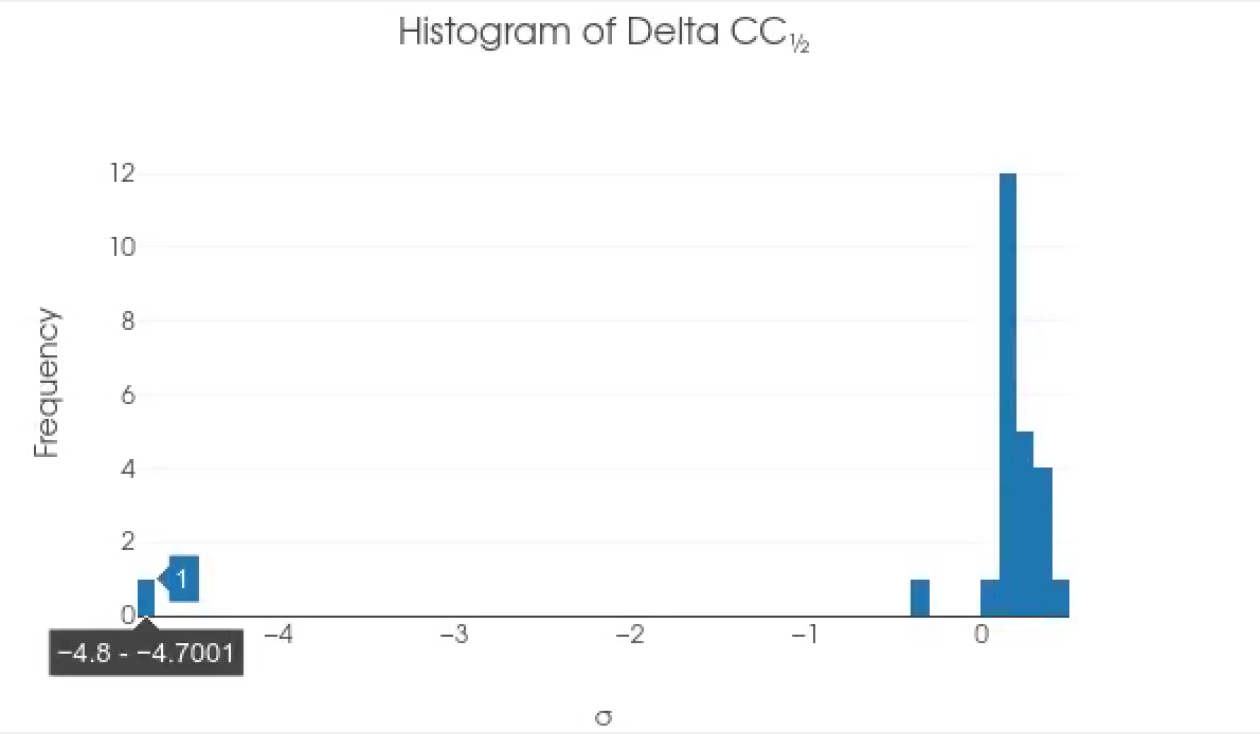




EMBL

Example of something going wrong





EMBL

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

Post Doctoral Research Associate, I24

Application deadline: 10/01/2021

Thank you for your attention!