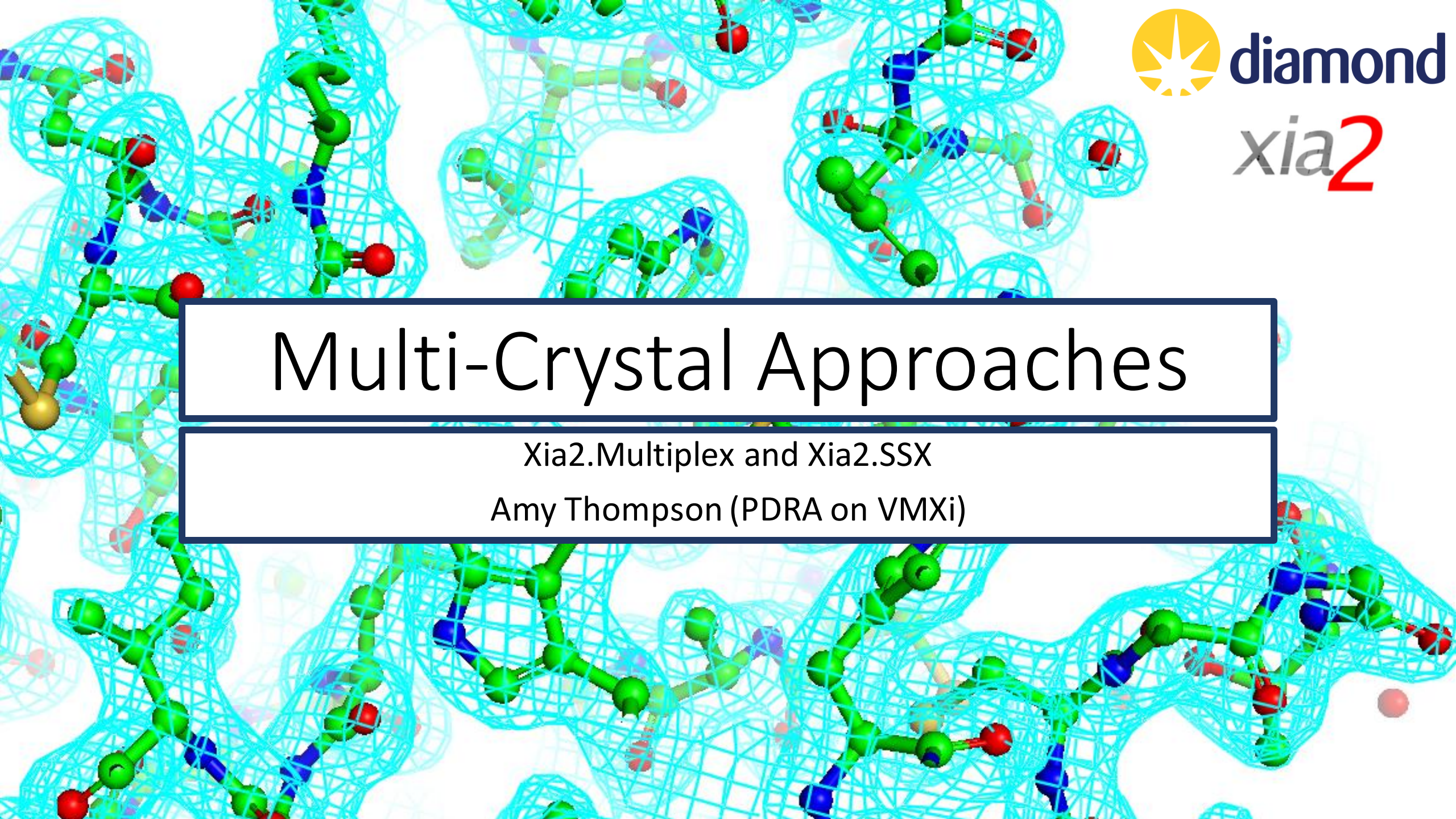
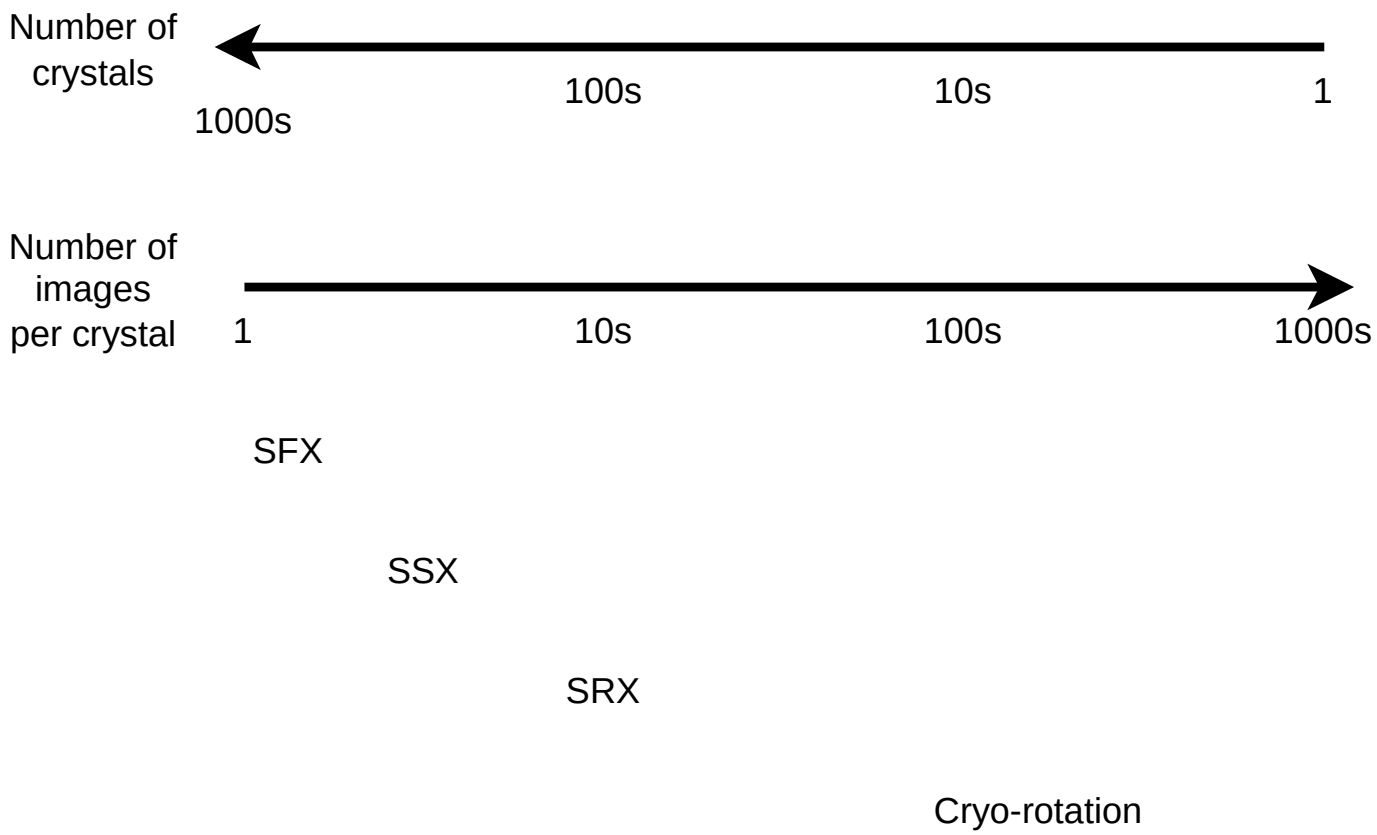


Multi-Crystal Approaches

Xia2.Multiplex and Xia2.SSX

Amy Thompson (PDRA on VMXi)

Multi-Crystal Approaches

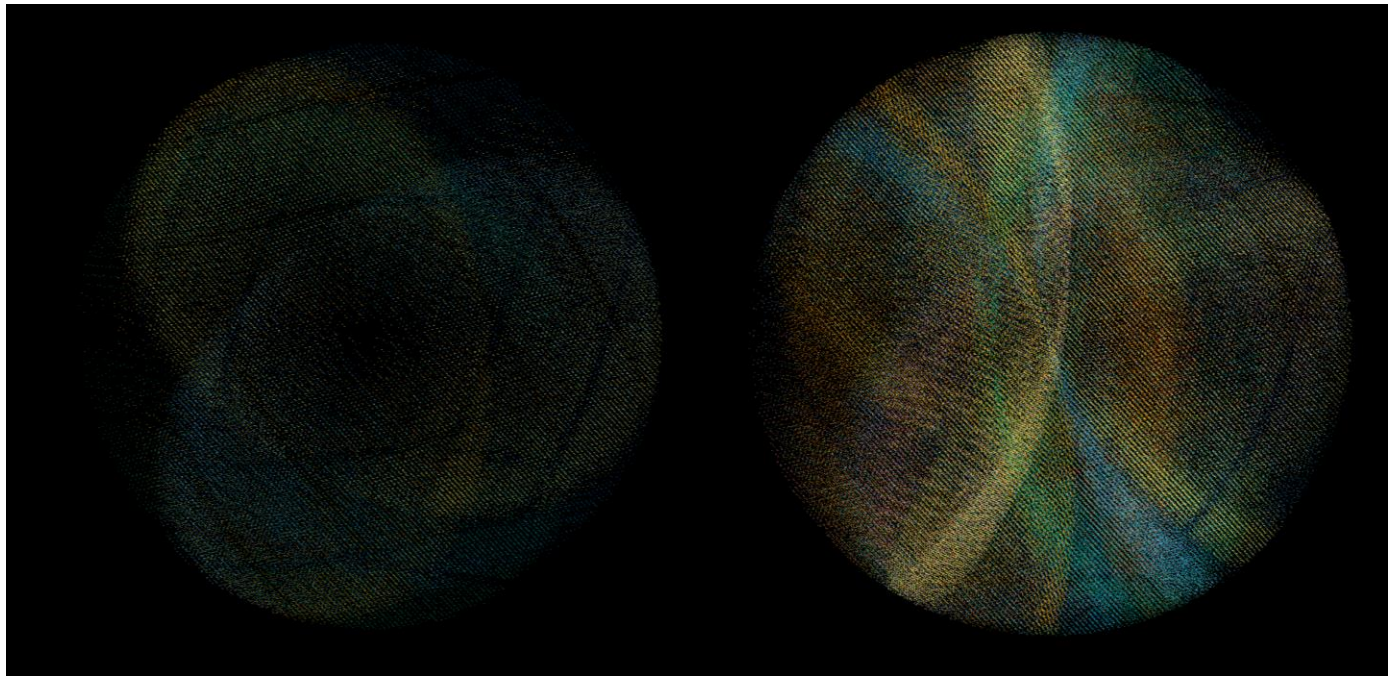


Why Use Multi-Crystal Strategies?

In some cases, it is only possible to collect incomplete data from a single crystal due to radiation damage: obtain a complete dataset by combining data from multiple crystals

This strategy allows for the use of:

- small crystals
- room-temperature data collection



Room Temperature Advantages

- Cryo-cooling may hide biologically significant structural features
- Virus crystals often suffer on cryo-cooling
- *In situ* data collection

Fraser, J., Clarkson, M., Degnan, S. et al. (2009) Nature 462, 669–673
Sanchez-Weatherby, J., Sandy, J., Mikolajek, H. et al. (2019) J. Synchrotron Rad. 26, 291–301
Nave, C. & Garman, E. F. (2005) J. Synchrotron Rad. 12, 257–260

diamond

Merging Multiple Datasets Increases Signal-to-Noise

- Given only random errors, the standard error σ in a measurement is reduced by $\sqrt{n}$ if the measurement is repeated n times
- Beware R_{merge} and R_{meas}!
- Use $CC_{½}$ and $\langle I/\sigma \rangle_{mrgd}$

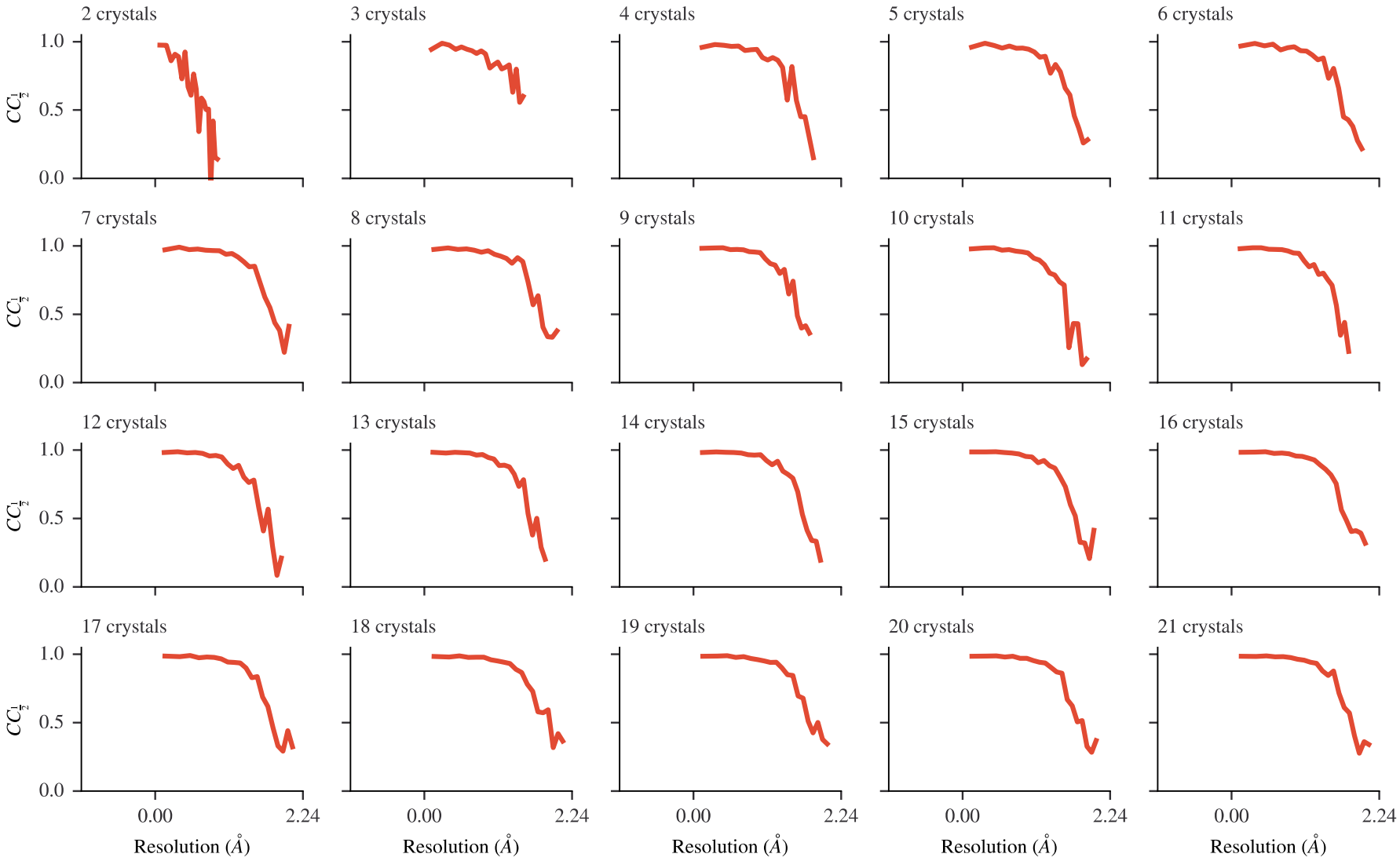
Dataset	*Big*	*Tiny*	*T100*	*Big+T100*	*Big2*
Multiplicity	2	2	200	202	4
$\langle I/\sigma \rangle_{ind}$	2.0	0.2	0.2	0.22	2.0
R_{merge}	28%	280%	399%	395%	35%
R_{meas}	40%	400%	400%	396%	40%
R_{pim}	28%	280%	28%	28%	20%
$\langle I/\sigma \rangle_{mrgd}$	2.8	0.28	2.8	4.0	4.0
$CC_{1/2}$	0.66	0.04	0.66	0.80	0.80

Karplus, P. A., & Diederichs, K. (2015) Current opinion in structural biology, 34, 60-68

diamond

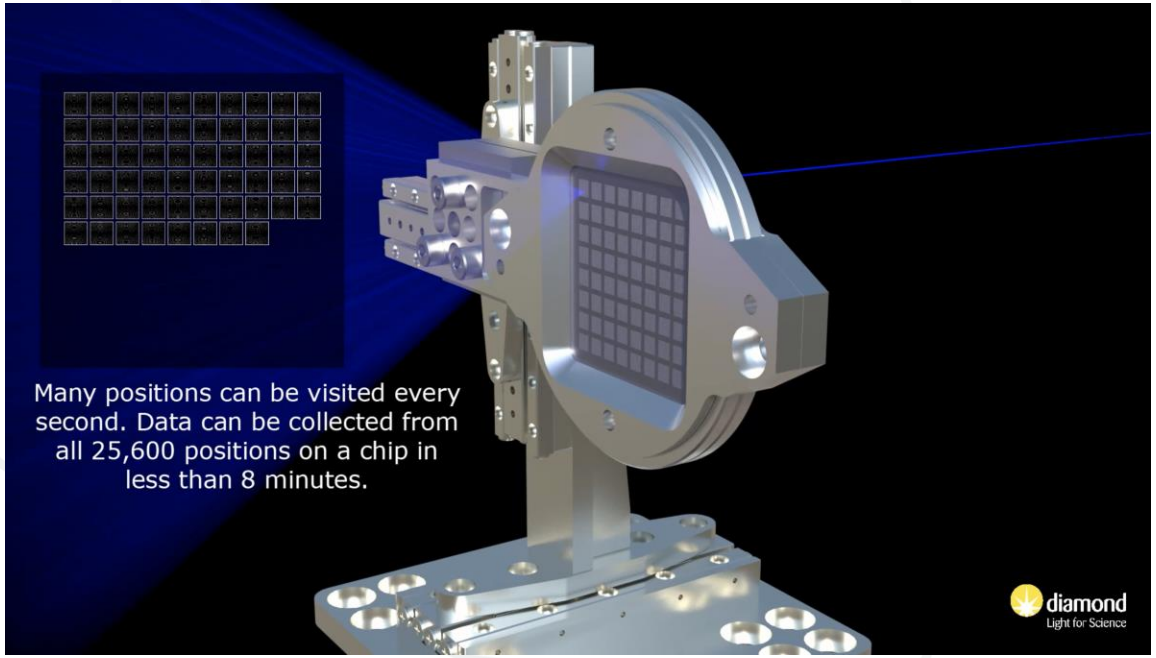
Better CC1/2 at higher resolution

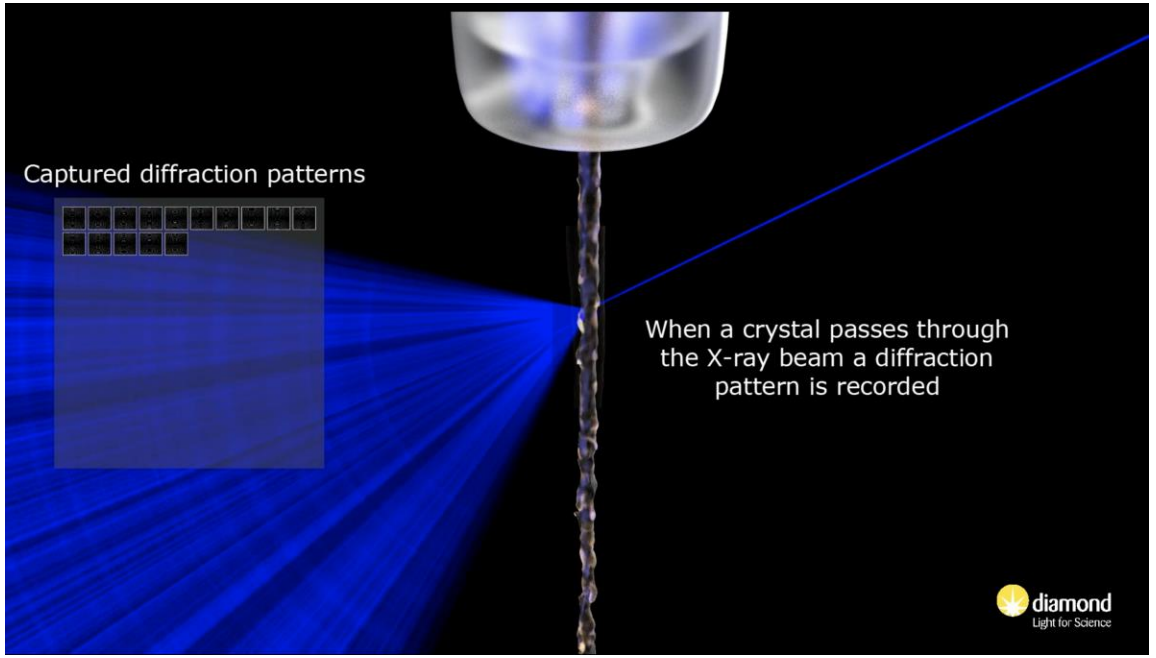
```
xia2.compare_merging_stats unmerged_1.mtz
unmerged_2.mtz [...] small_multiples=True
```



Time-Resolved Crystallography

- Study dynamics of light- or chemical-induced reactions using serial femtosecond (SFX) or synchrotron (SSX) crystallography





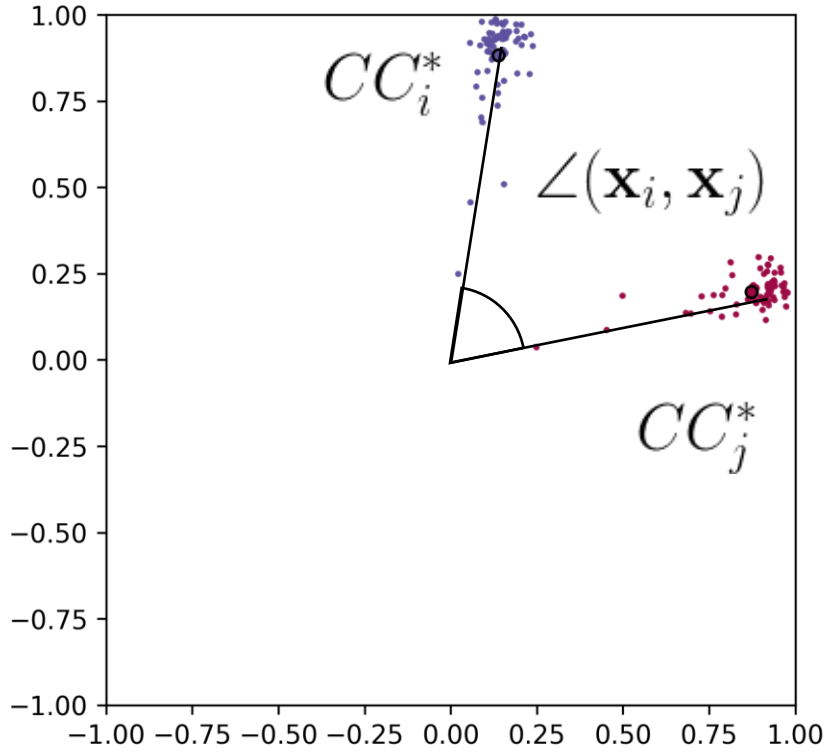
diamond

Challenges

- Symmetry determination
- Non-isomorphism
- Preferential orientation

Symmetry Determination

- Identification of consensus symmetry from narrow wedges or stills can be challenging
- Complicated by presence of potential indexing ambiguity
- New algorithms have been developed to help in symmetry determination from narrow wedges and stills (dials.cosym)

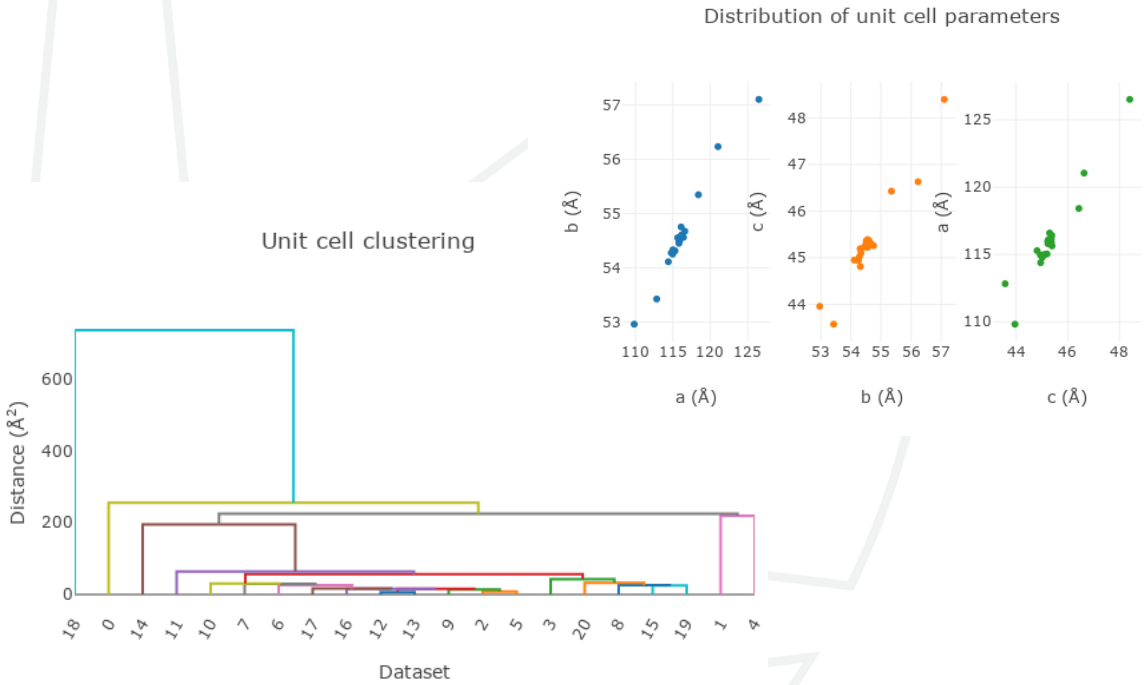


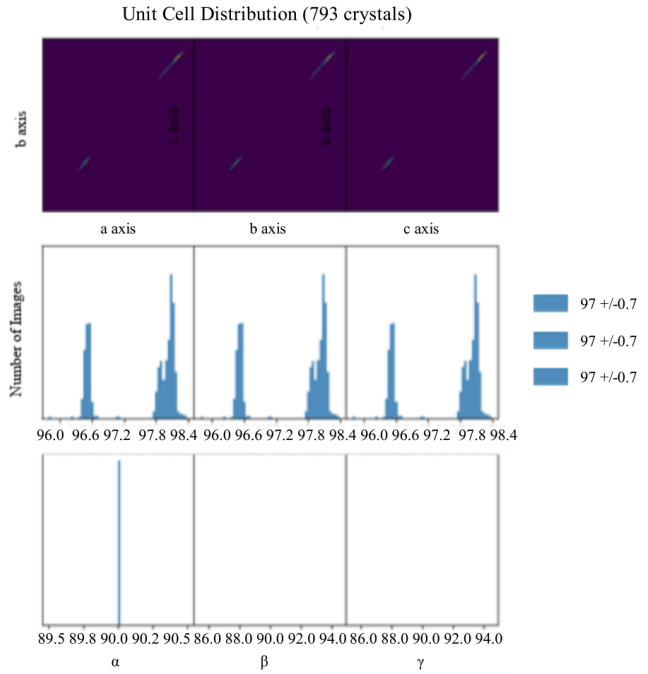
Brehm, W. & Diederichs, K. (2014). Acta Cryst. D70, 101–109
Gildea, R. J. & Winter, G. (2018). *Acta Cryst.* D**74**, 405–410

diamond

Non-Isomorphism

- Inclusion of non-isomorphous crystals may degrade the final data set
- Unit cell clustering may help identify outliers or different populations

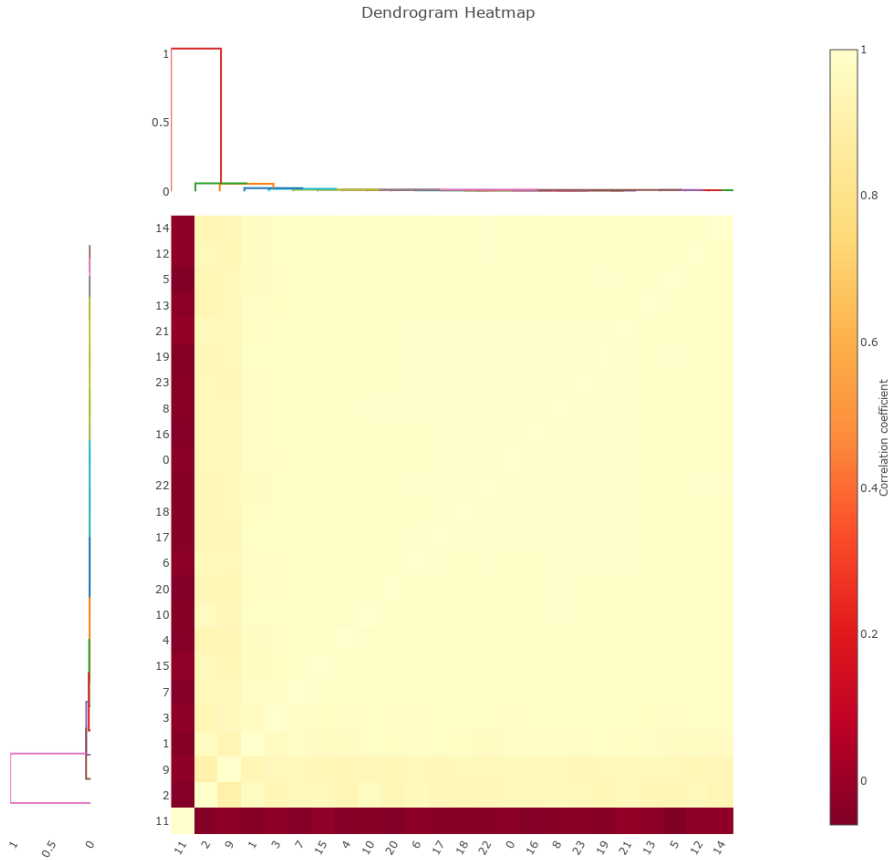




Non-Isomorphism

Clustering on pairwise correlation coefficients may help identify outliers

$$r_{i,j} = \frac{\sum_h \left[I_i(h) - \overline{I}_i\right]\left[I_j(h) - \overline{I}_j\right]}{\left\{\sum_h \left[I_i(h) - \overline{I}_i\right]^2 \sum_h \left[I_j(h) - \overline{I}_j\right]^2\right\}^{1/2}}$$

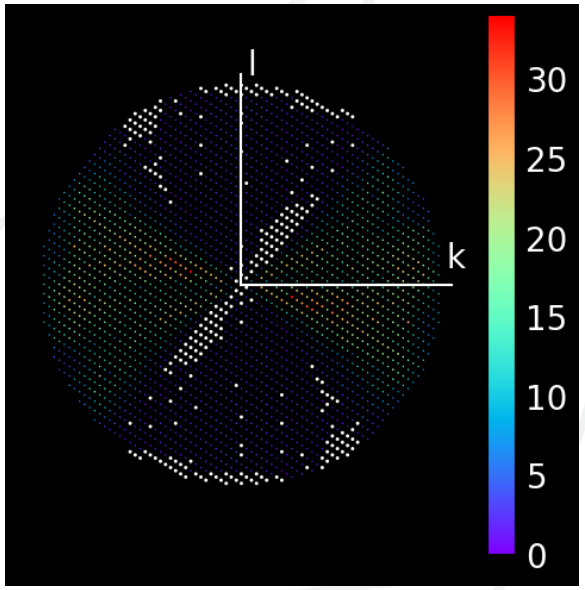


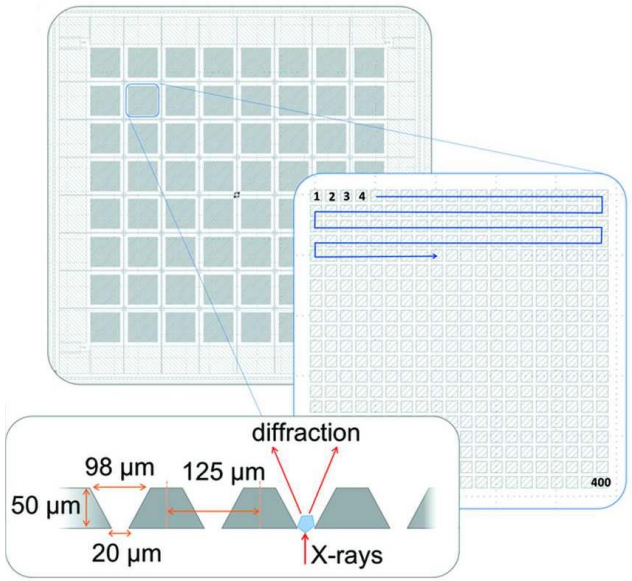
Giordano, R., Leal, R. M. F., Bourenkov, G. P. et al. (2012). Acta Cryst. D68, 649–658
Santoni, G., Zander, U., Mueller-Dieckmann, C. et al. (2017). J. Appl. Cryst. 50, 1844–1851

diamond

Preferential Orientation

- Crystal symmetry and morphology combined with data collection conditions may lead to preferential crystal orientation
- May result in under-sampled regions of reciprocal space
- Check stereographic projection and multiplicity plots

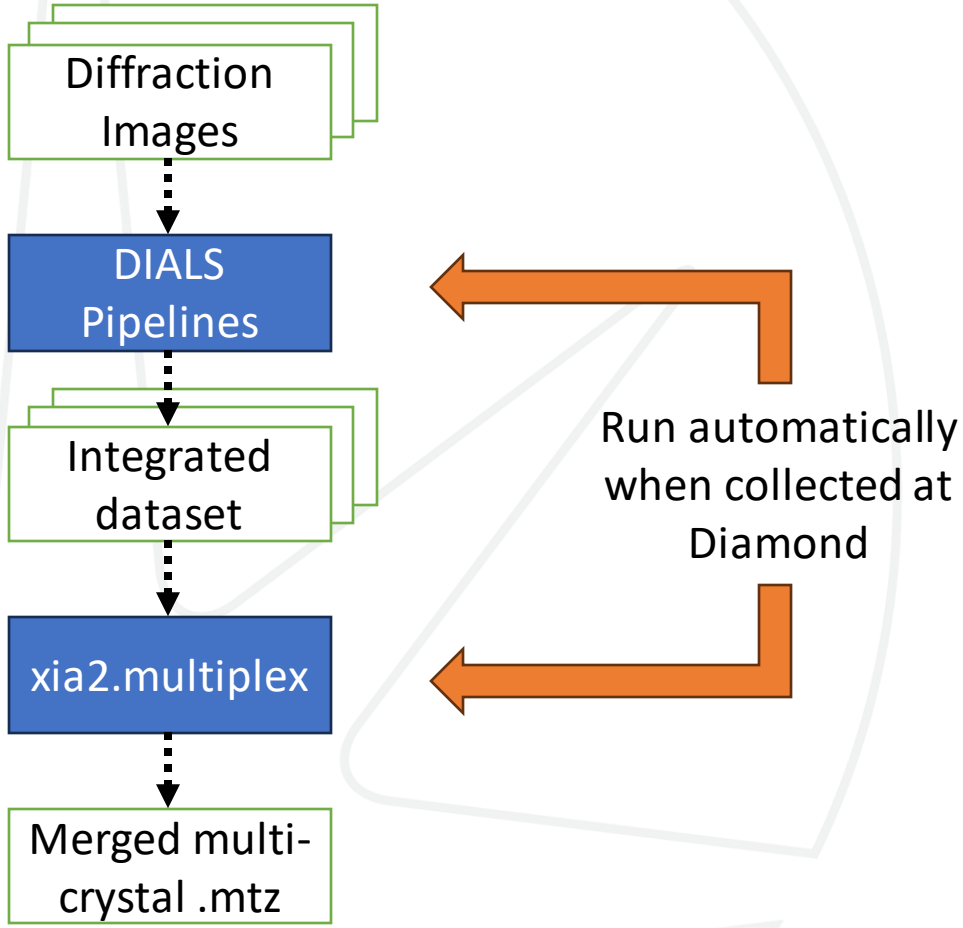




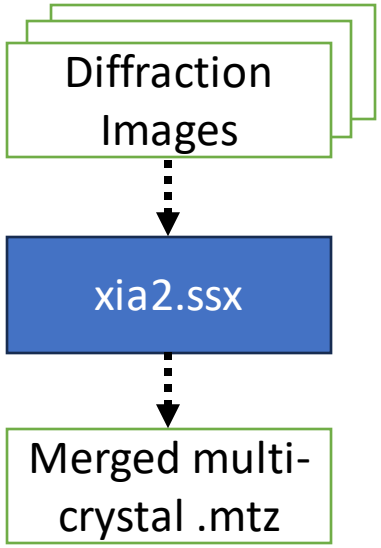
Horrell, S., Axford, D., Devenish, N. E., et al. (2021) J. Vis. Exp. (168), e62200

Processing Multi-Crystal Data at Diamond

xia2.multiplex – for rotation data



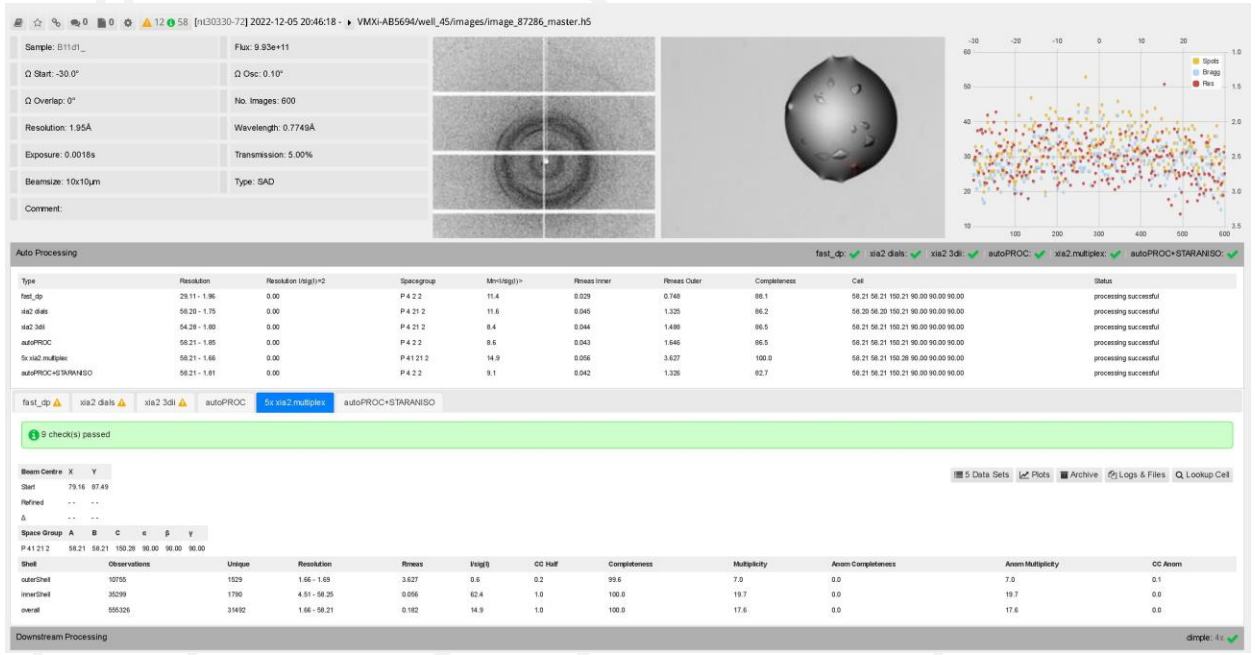
xia2.ssx – for still data

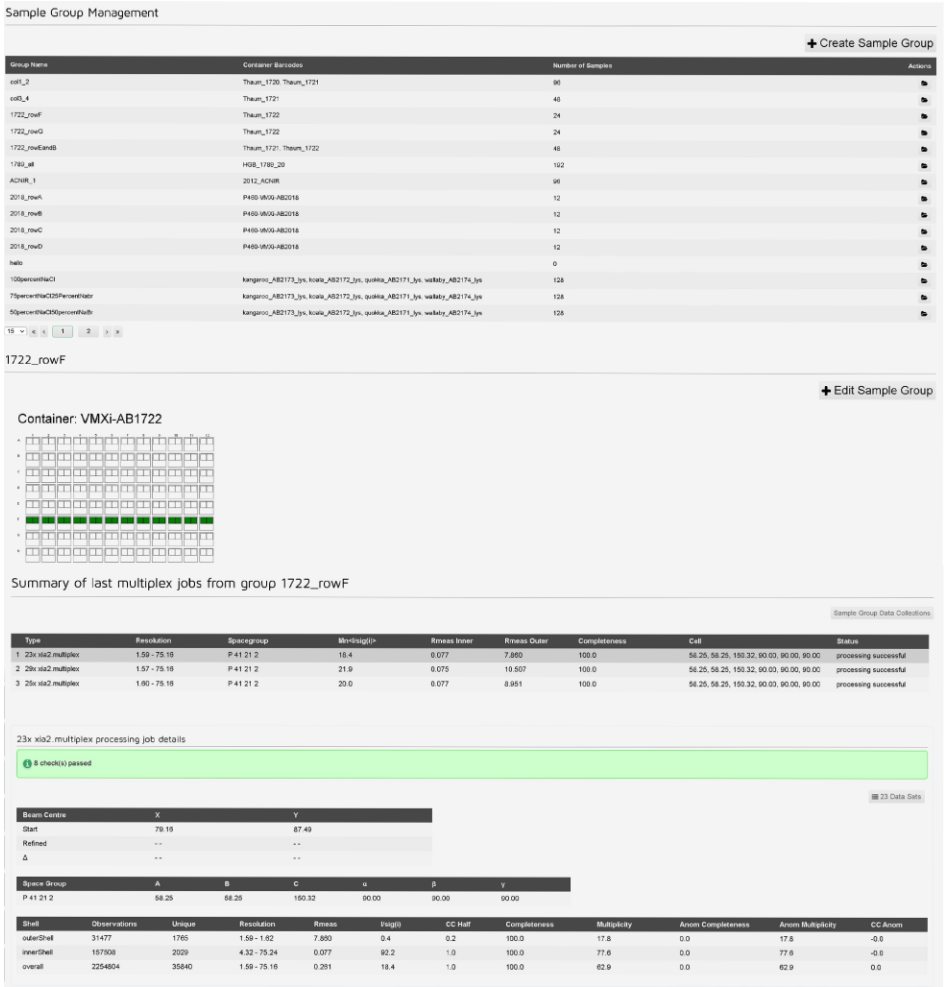


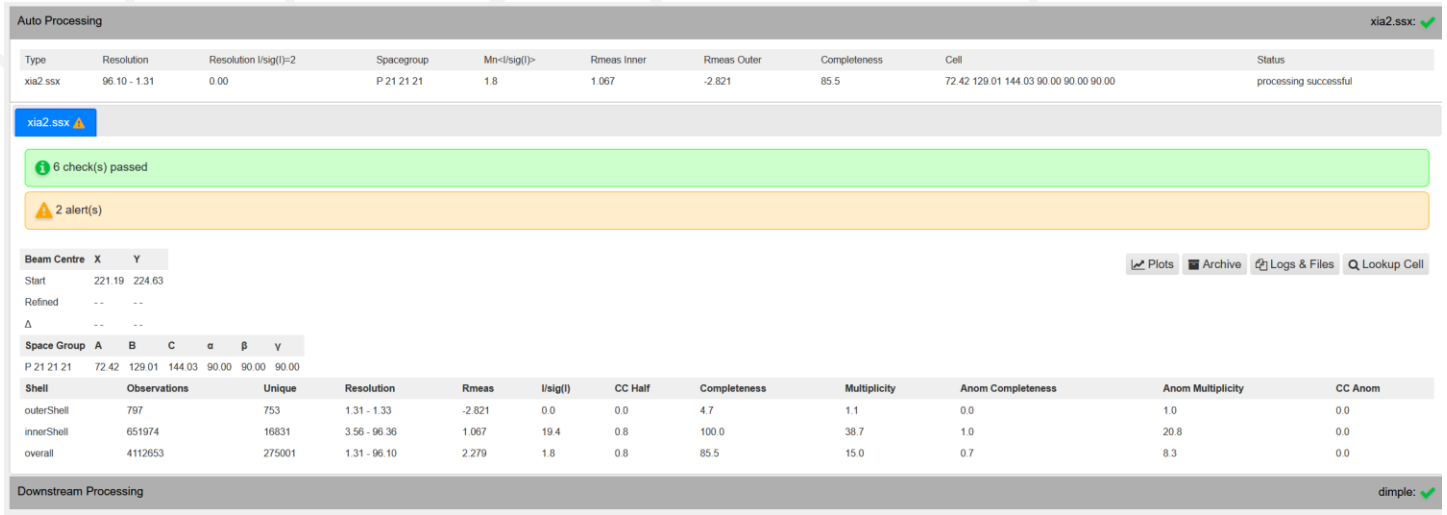
Runs automatically on I24 – otherwise need to run manually via command line

diamond

Availability through ISPyB at Diamond



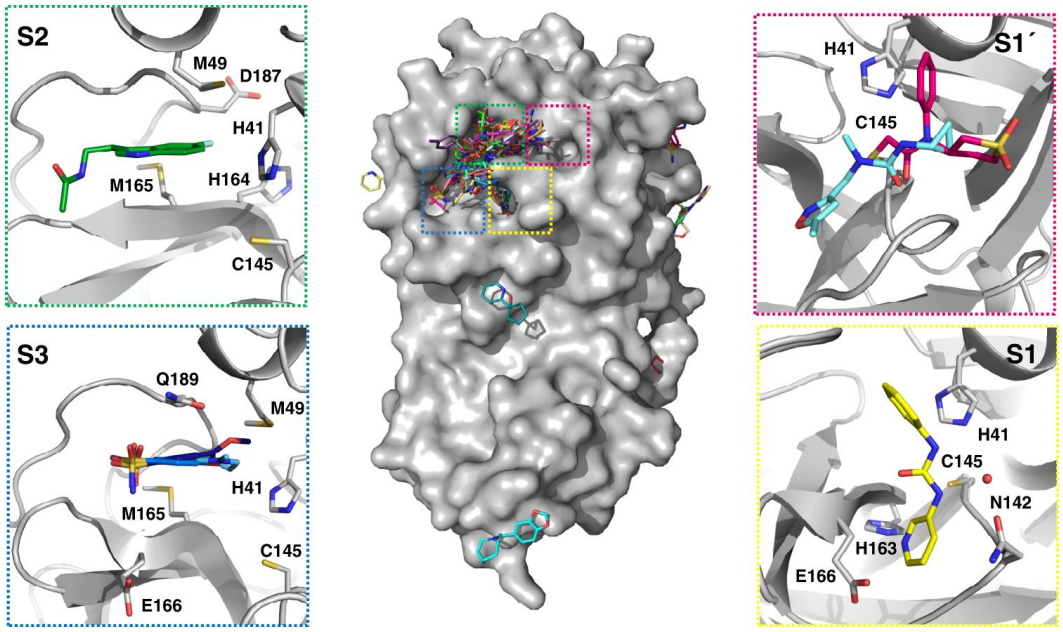




Example: SARS-CoV-2 M^{pro} ligand screening

Conventional Collection at CT

- SARS-CoV-2 main protease
- Central role in viral replication
- Key antiviral drug target
- Conventional fragment-screening campaign performed on I04-1
- Over 1250 unique fragments, identifying 74 high-value fragment hits



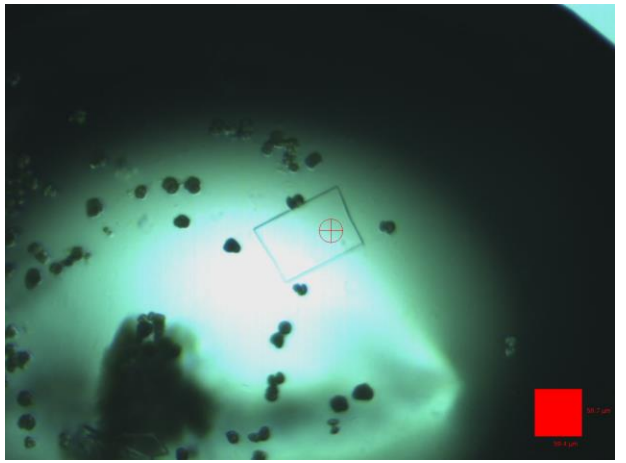
Douangamath, A., Fearon, D., Gehrtz, P. et al. (2020). Nat. Commun. 11, 5047

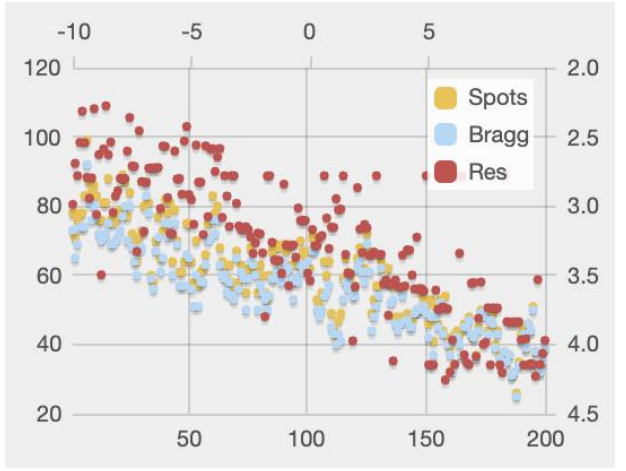
diamond

Example: SARS-CoV-2 M^{pro} ligand screening

Multi-Crystal Collection at RT

- Initial fragment screening performed at cryo-temperatures (100 K)
- Are room temperature structures identical?
- RT in situ data collections on known ligand hits performed on I24 and VMXi
- Preferred orientation (plate-like crystals): vary starting angle
- xia2.multiplex provided near real time feedback during the experiment



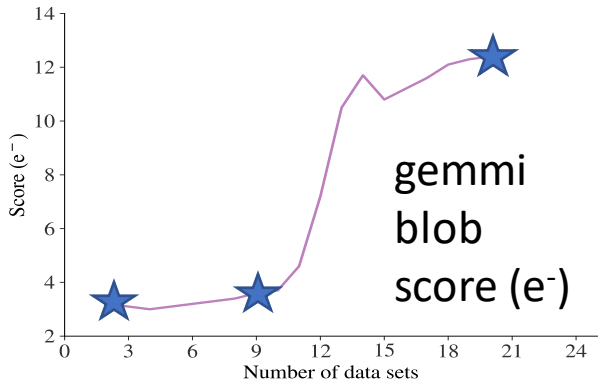


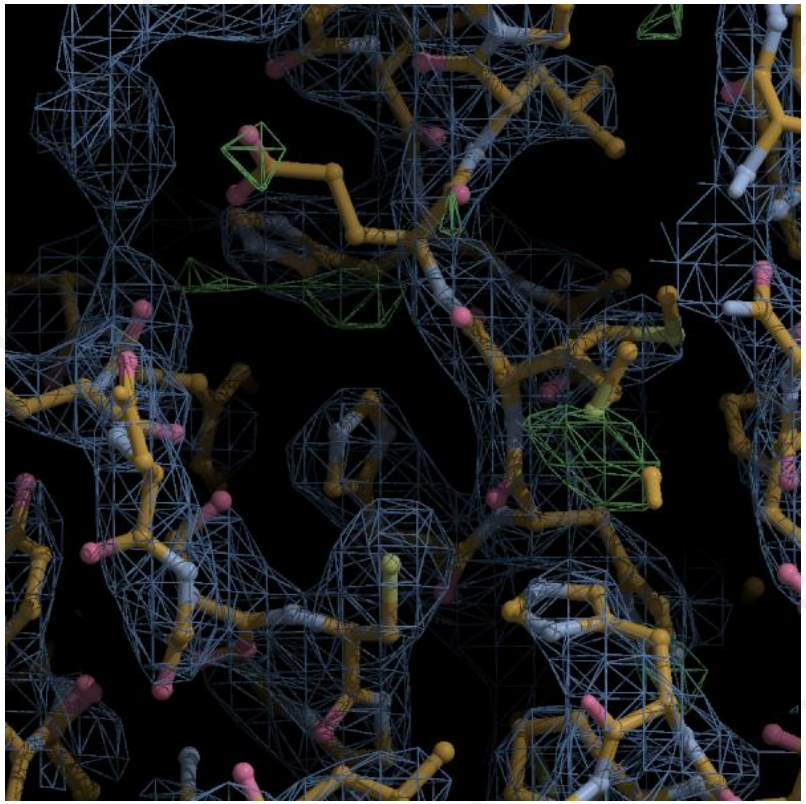
Gildea, R. J., Beilsten-Edmands, J., Axford, D., et al. (2022) Acta Cryst. D78, 752-769

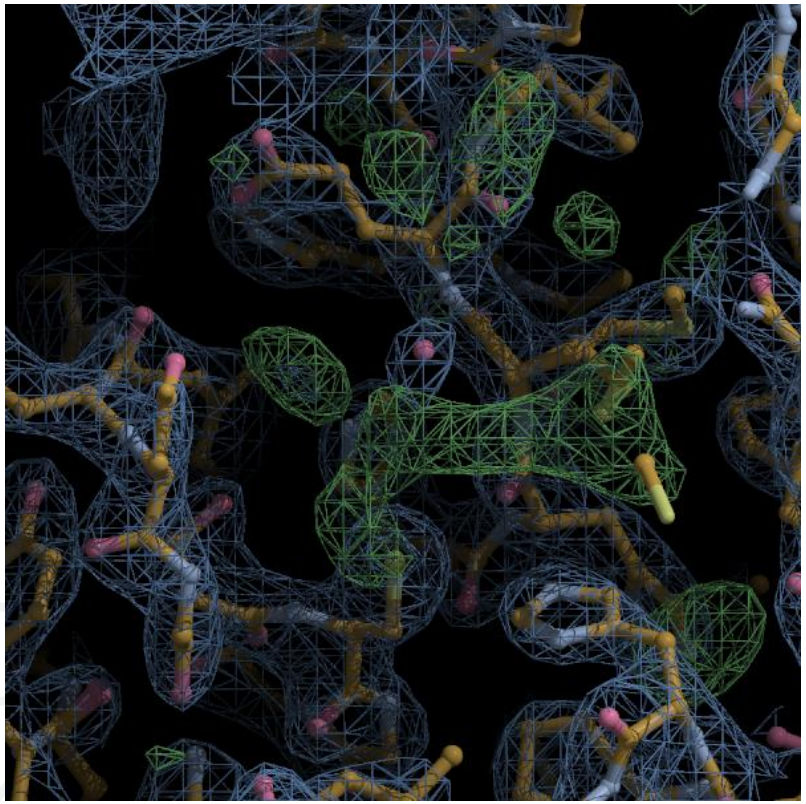
diamond

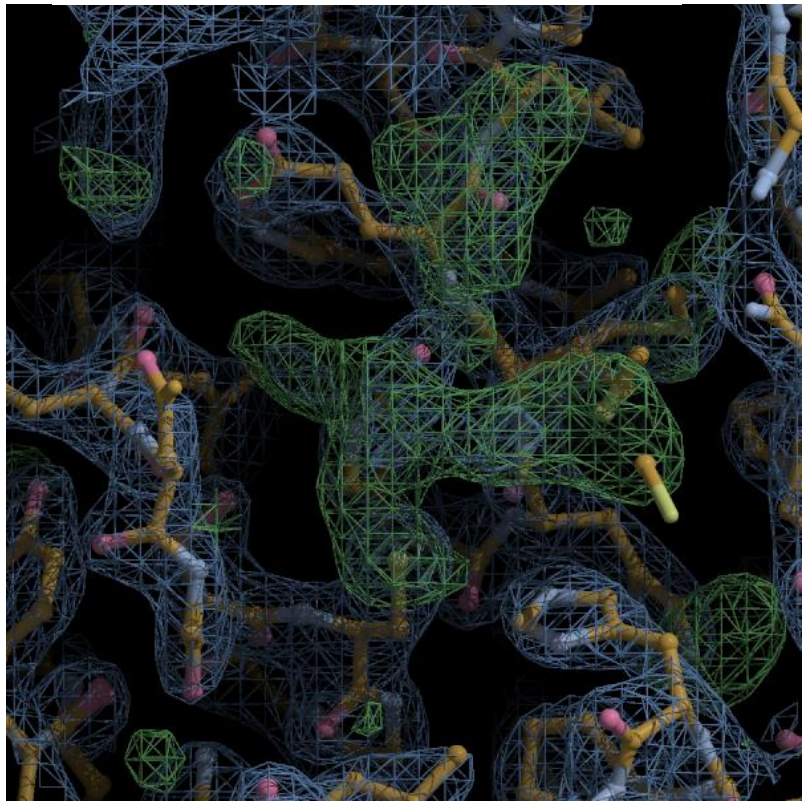
Example: SARS-CoV-2 M^{pro} ligand screening

Automatic dimple maps



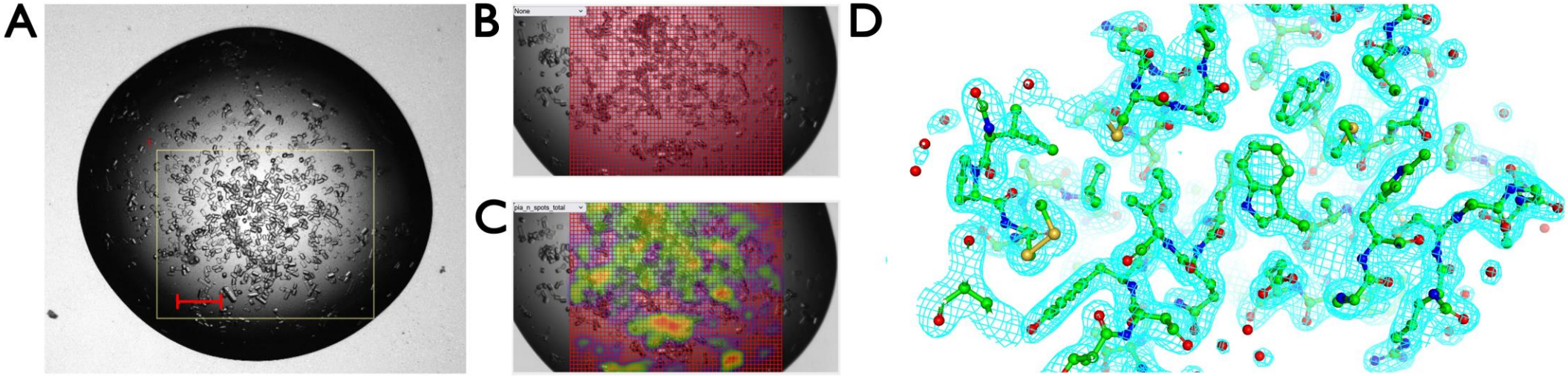






Example: Using VMXi to screen for serial experiments

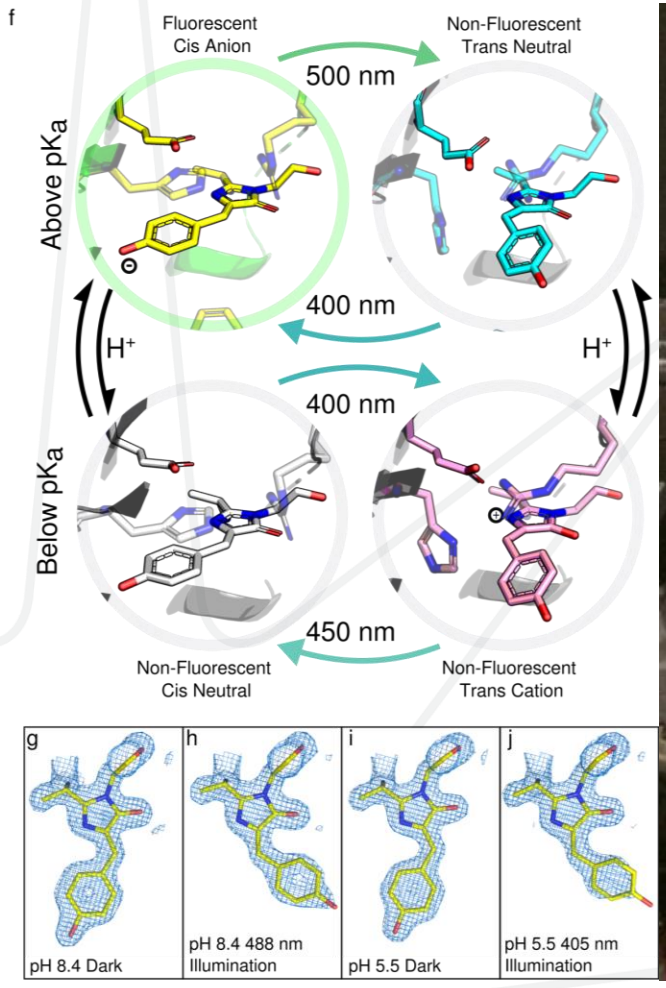
You can get decent quality serial data from grid scans on VMXi – example: lysozyme

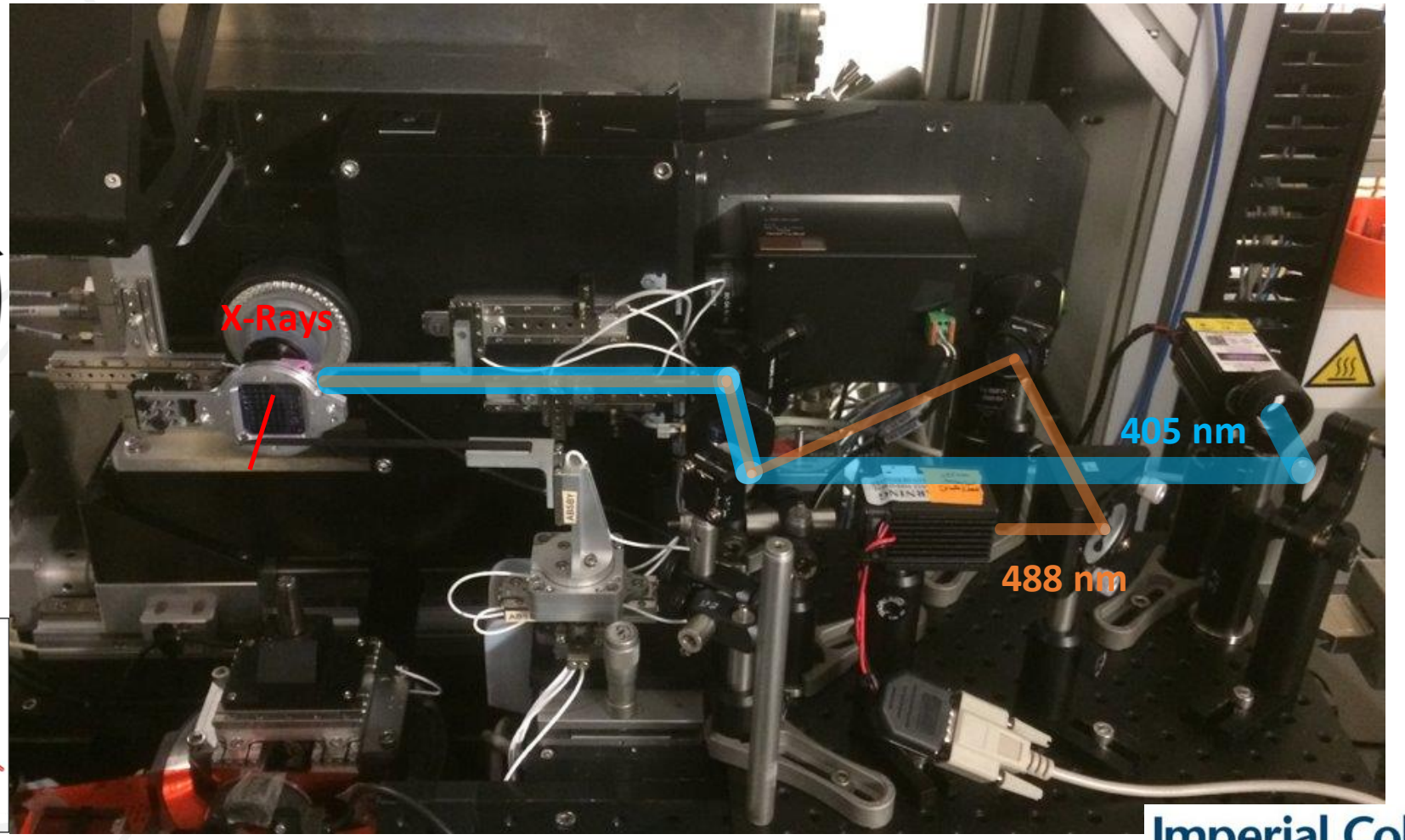


Statistics for eight wells merged together	Completeness (%)	Multiplicity	I/σ(I)	R_{split}	$CC_{1/2}$	Unique Observations	Indexed Images
Overall	100	95.5	20.8	0.063	0.998	8422	9891 / 25906 (38.18 %)
Low (55.58 – 5.43)	100	147.1	81.7	0.028	0.999	488	
High (2.03 – 2.00)	100	75.3	1.2	1.092	0.410	411	

diamond

Example: I24 SSX (Van Thor Group – Imperial College London)





Baxter, J. M., Hutchison, C. D., Maghlaoui, K. et al. (2022) J. Phys. Chem. B, 126, 45, 9288–9296

Imperial College London

Furthering Output of Multi-Crystal Collections

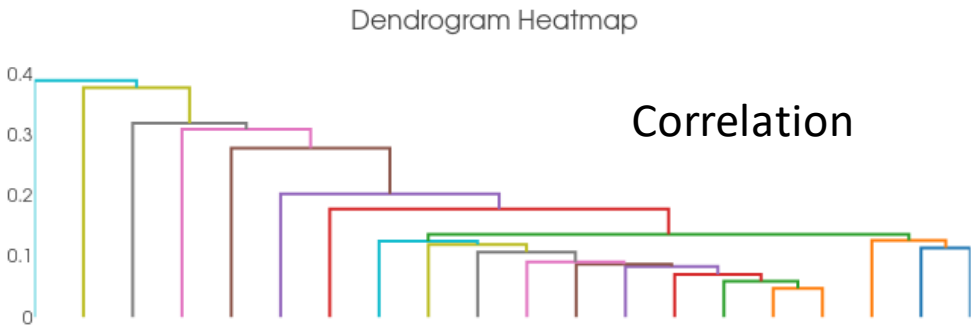
Can we do more than just structure solution with multi-crystal data?

Clustering of Multi-Crystal Data

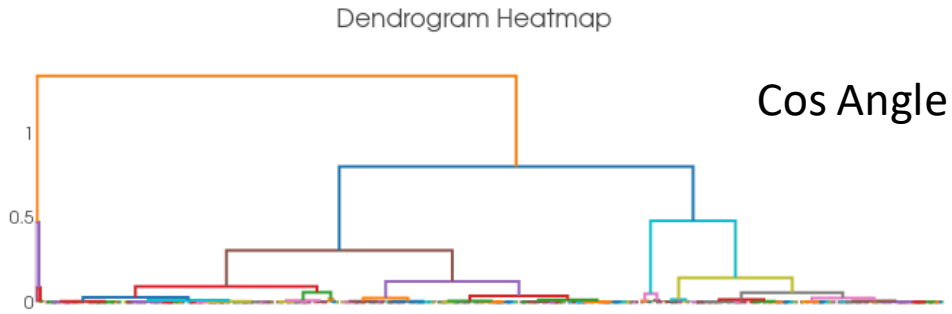
When merging many datasets together, there are going to be differences: especially when collected at room temperature!

Multiplex includes two different intensity-based clustering tools to group datasets:

- Correlation Coefficient
- Cos Angle



Correlation

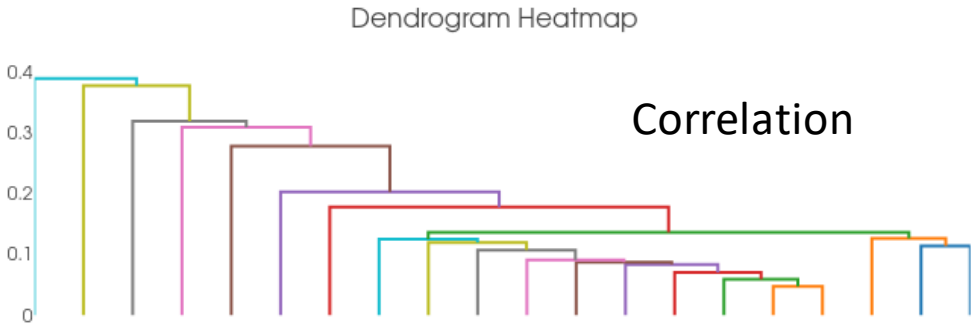


Cos Angle

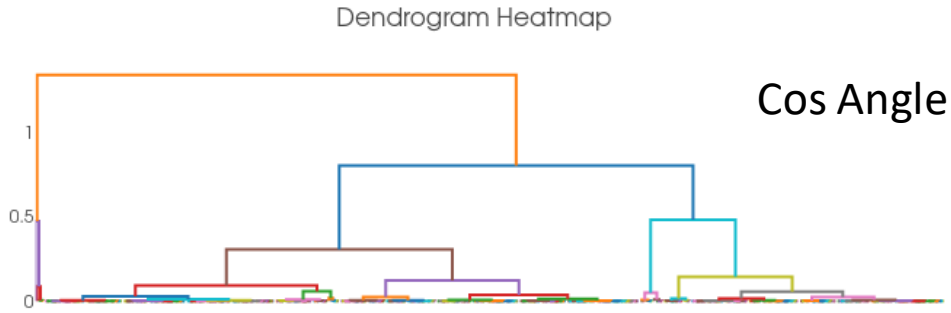
Clustering of Multi-Crystal Data

By analysing differences between clusters, can we extract more scientific outcomes from multi-crystal datasets?

- Structural dynamics?
- Ligand binding dynamics?



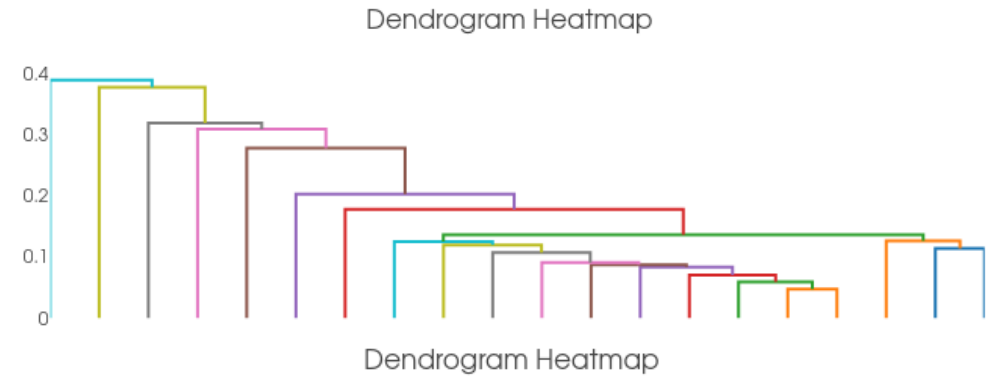
Correlation

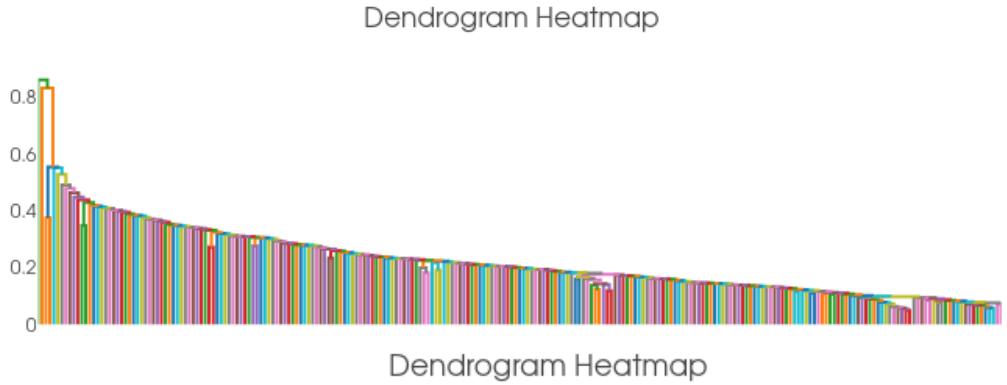


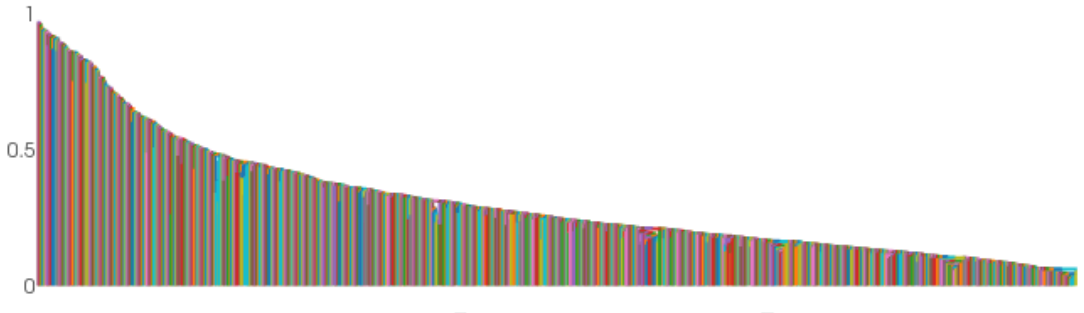
Cos Angle

Correlation Clustering with xia2.multiplex

Clustering based on correlation coefficients can improve data quality for small datasets – but could it identify distinct states for large datasets?







Issues with current method of cluster output!

Which ones are scaled and merged is based on completeness – for cubic crystals, 100% can be achieved with as little as 2 datasets. If you have thousands of datasets, this is thousands of clusters!

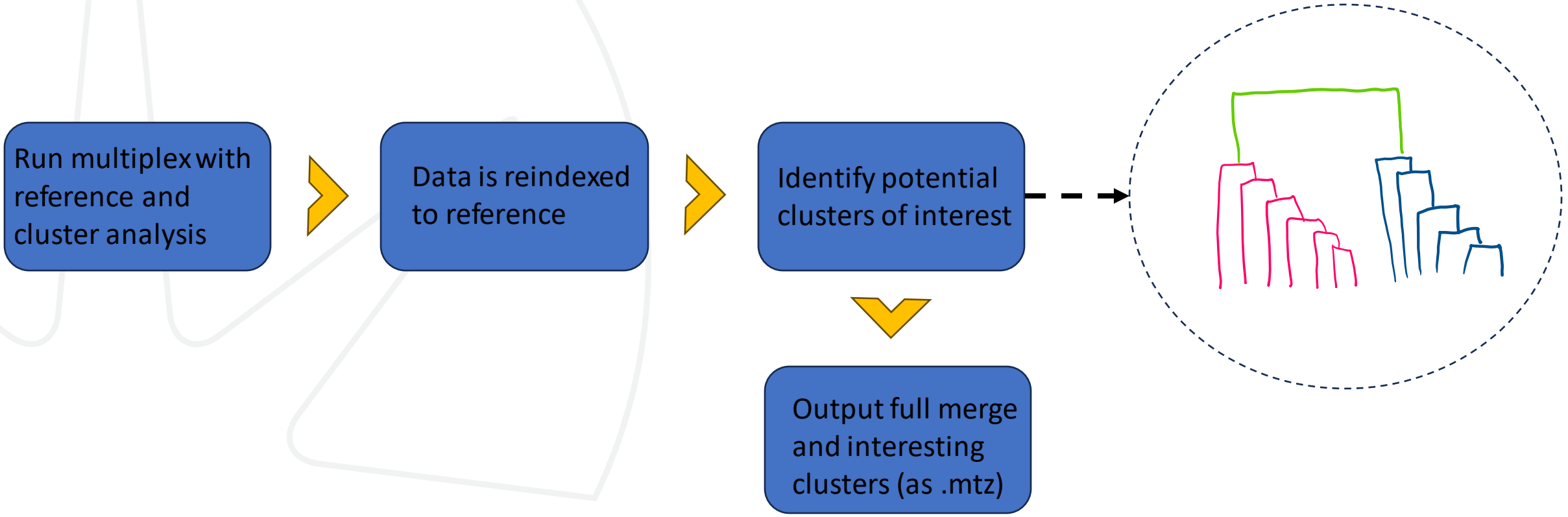
Computationally expensive and time wasting when you may only want to analyse a couple of clusters!

1500 x 20° wedges of insulin on VMXi

diamond

Improvements to Multiplex Towards Cluster Analysis

- Inclusion of a cluster picking algorithm to automatically identify clusters that may be of interest

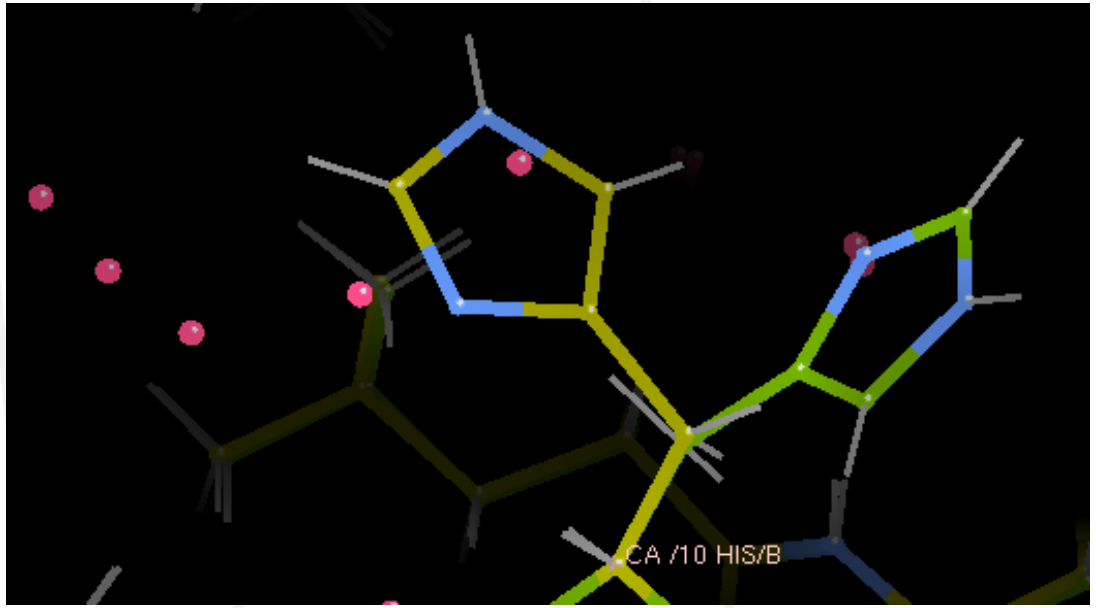


Consistent Rfree flags too!

diamond

Initial Findings – Insulin

Do notice some subtle, but distinct conformational changes between pairs of clusters, despite all clusters starting from the same initial model



Main example: flip of a well-defined HIS side-chain and corresponding water movement

From the merging of 500 datasets

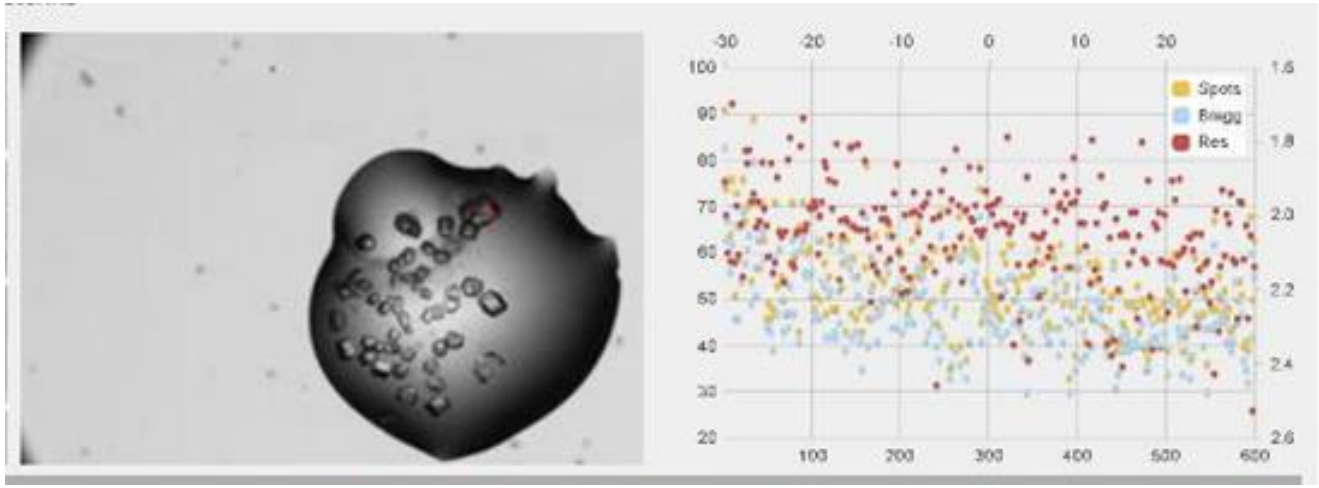
Applications to Ligand Binding?

Hypothetical situation: drop with multiple crystals, all datasets merged together, see a ligand binding. Great right?

But what if:

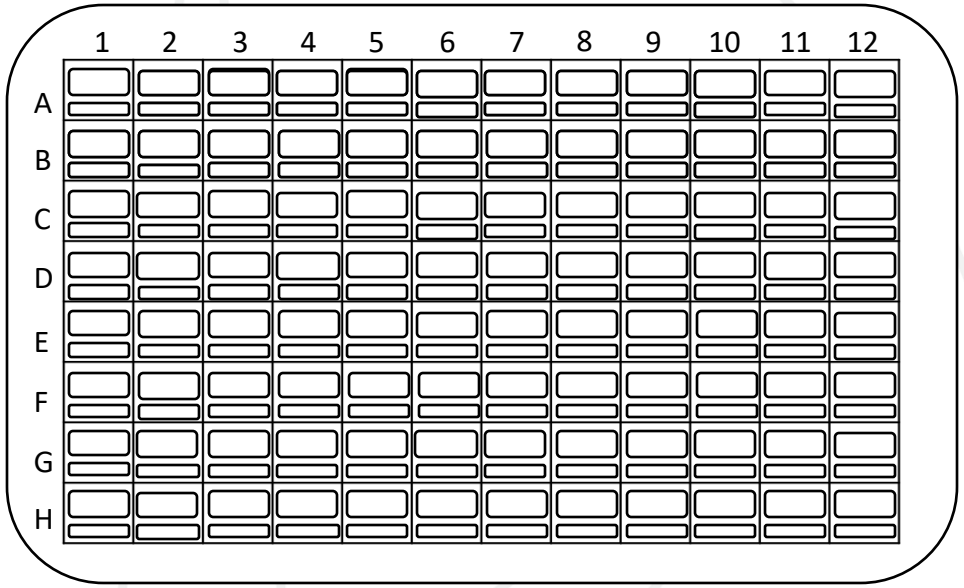
- Not all crystals actually had a ligand in them?
- Different crystals had different binding states?
- Different crystals had different ligand occupancies?

How much information can multiplex clustering provide towards answering these questions?

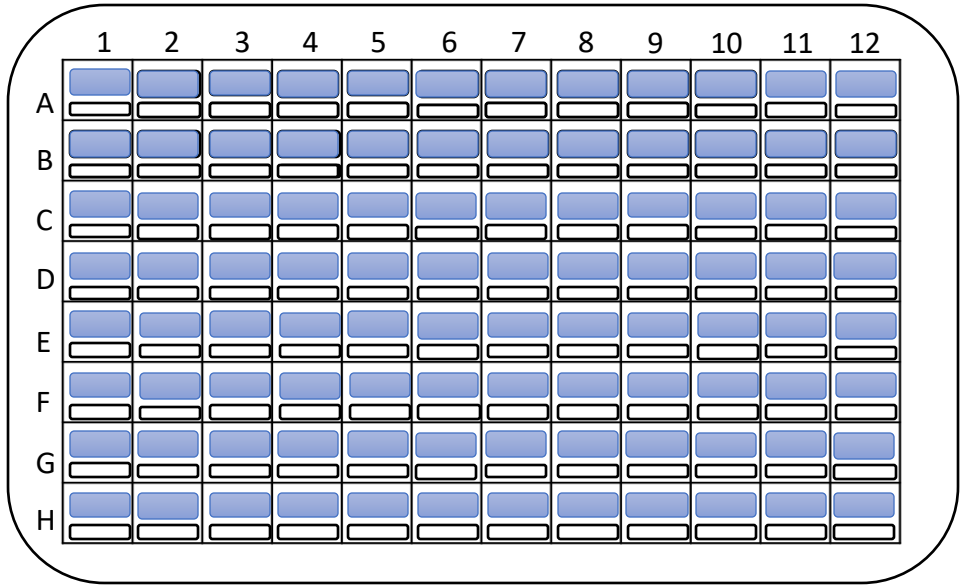


Initial Test Case - Lysozyme

Regular Lysozyme



Drop of $NaBr_{(aq)}$ on top of each position

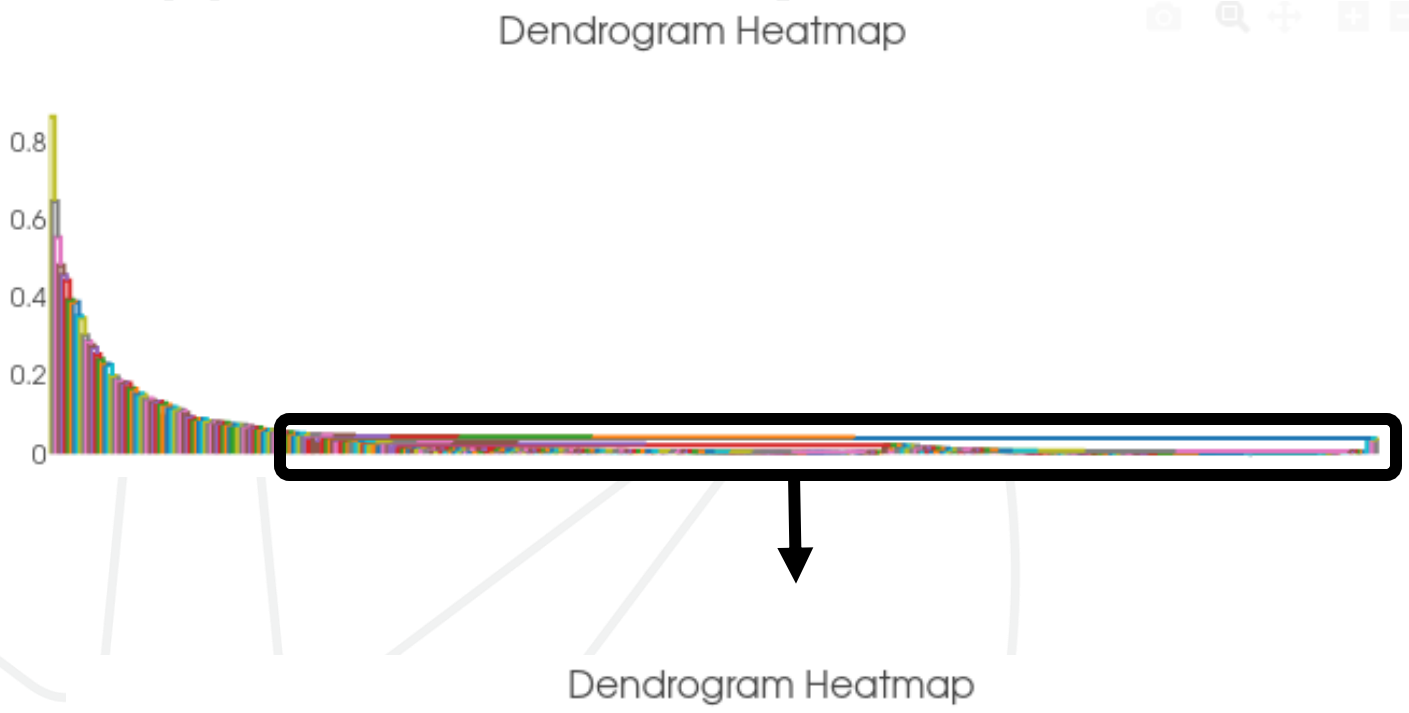


Measured whole plate on VMXi (LYS_Cl)
50 ° wedges

Measured whole plate on VMXi (LYS_Br)
50 ° wedges

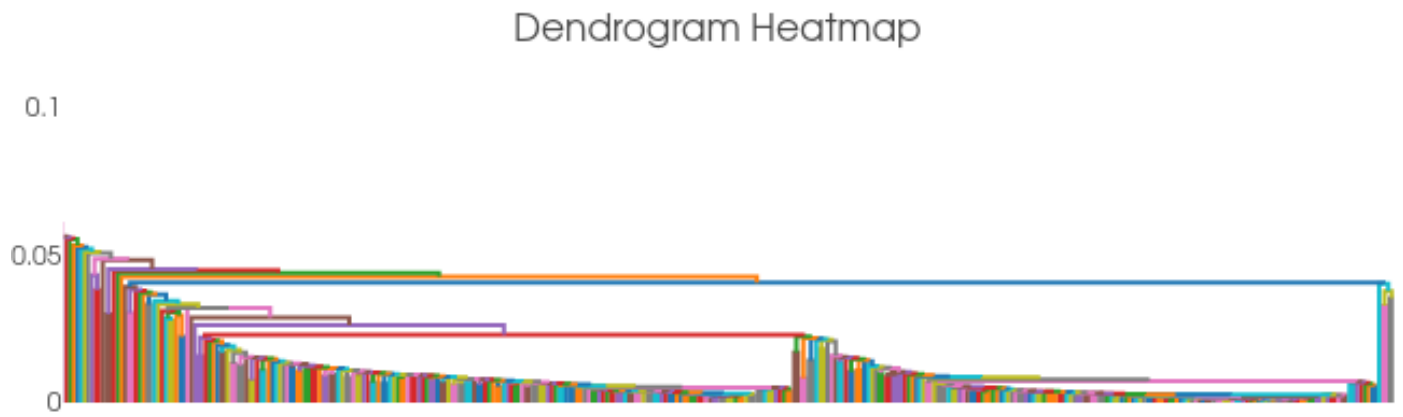
LYS_Cl + LYS_Br through xia2.multiplex
(438 crystals merged together)

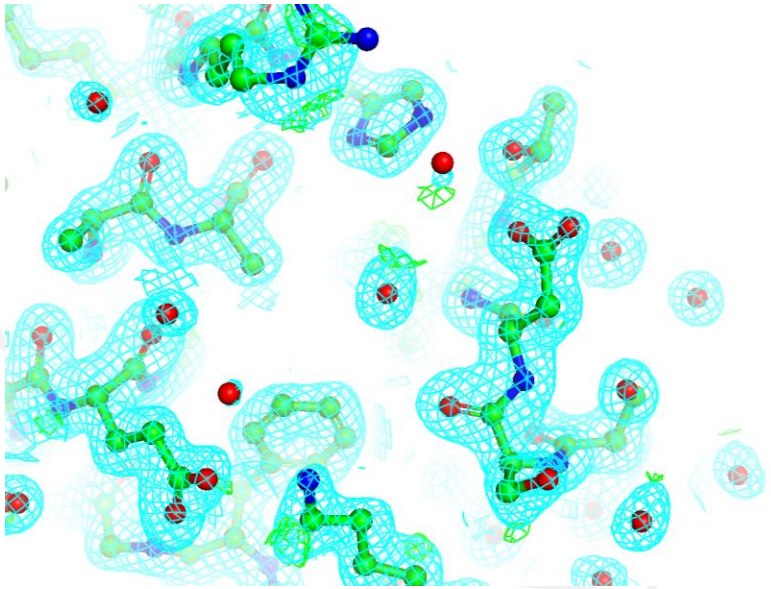
Correlation Clustering of Combined Lysozyme



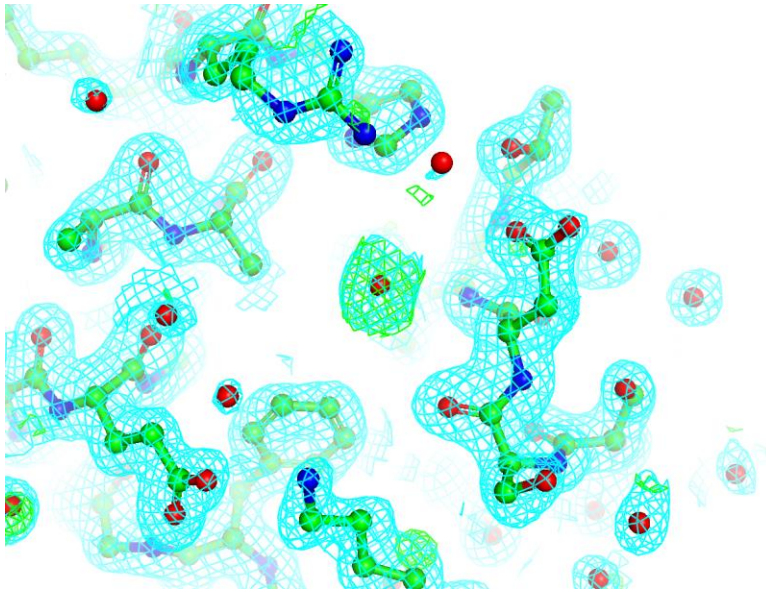
Datasets corresponding to each cluster indicates that they correspond to the two plates!

Correlation Clustering with xia2.multiplex





First cluster – only datasets from NaCl plate



Second cluster – only datasets from NaBr plate

Acknowledgements

VMXi:

- Michael Hough
- Juan Sanchez-Weatherby
- James Sandy
- Halina Mikolajek
- Megan Lambert
- Hans Pfalzgraf
- Sotaro Fujii
- Cicely Tam

Software Team:

- Graeme Winter
- James Beilsten-Edmands
- Richard Gildea

Questions?