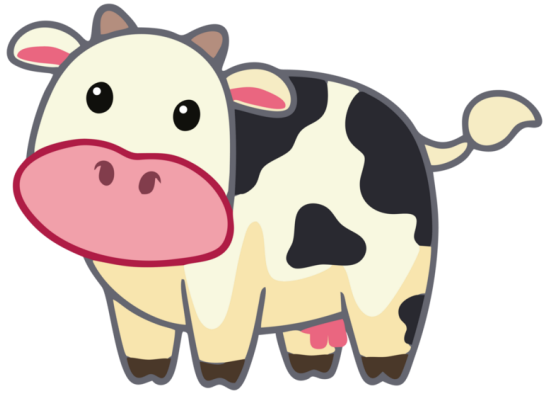
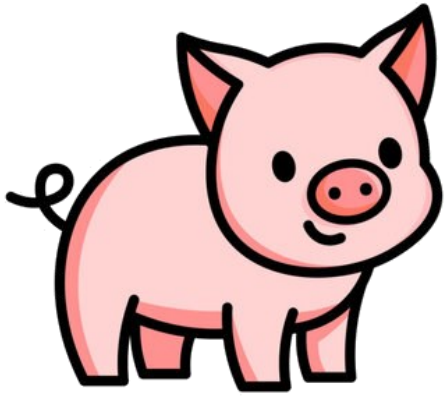




Multi-Crystal Methods

Dr Amy Thompson

PDRA – VMXi Beamline – Diamond Light Source

xia2

diamond

DIALS
Diffraction Integration for Advanced Light Sources

Why use Multi-Crystal Strategies for MX?

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
- Flexibility may show multiple states when large numbers of samples are collected
- *In situ* data collection

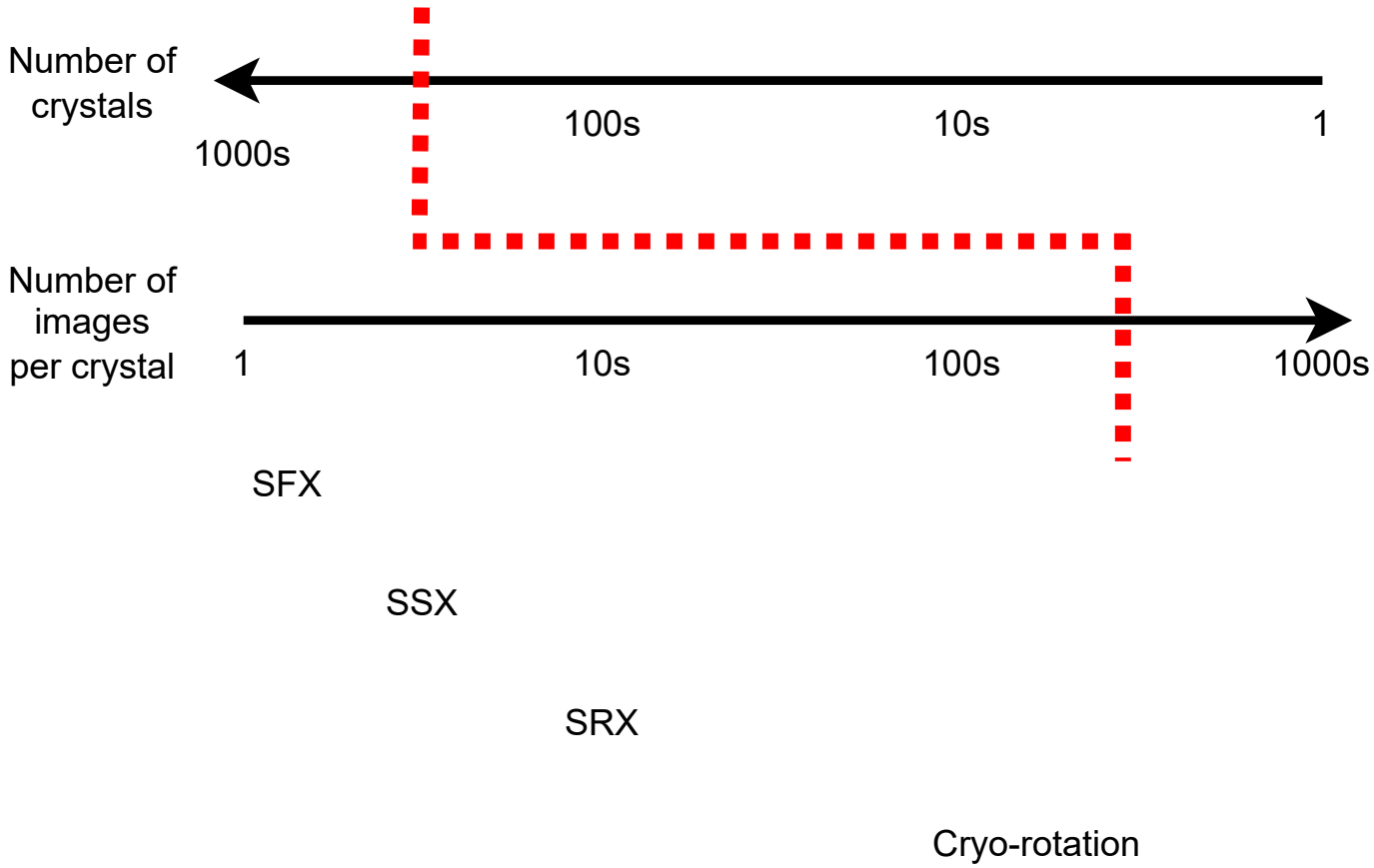
The VMXi Beamline specializes in room temperature data collection in-situ

diamond

Multi-Crystal Approaches

Three MX Beamlines at Diamond that make extensive use of multi-crystal data collections:

- I24
- VMXi
- VMXm



Additional Benefits

- Improved signal-to-noise (but beware statistics pitfalls!)
- Better CC1/2 at higher resolution

Pitfalls with "Standard" Data Reduction Statistics

- 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 $\langle I/\sigma \rangle_{mrgd}$

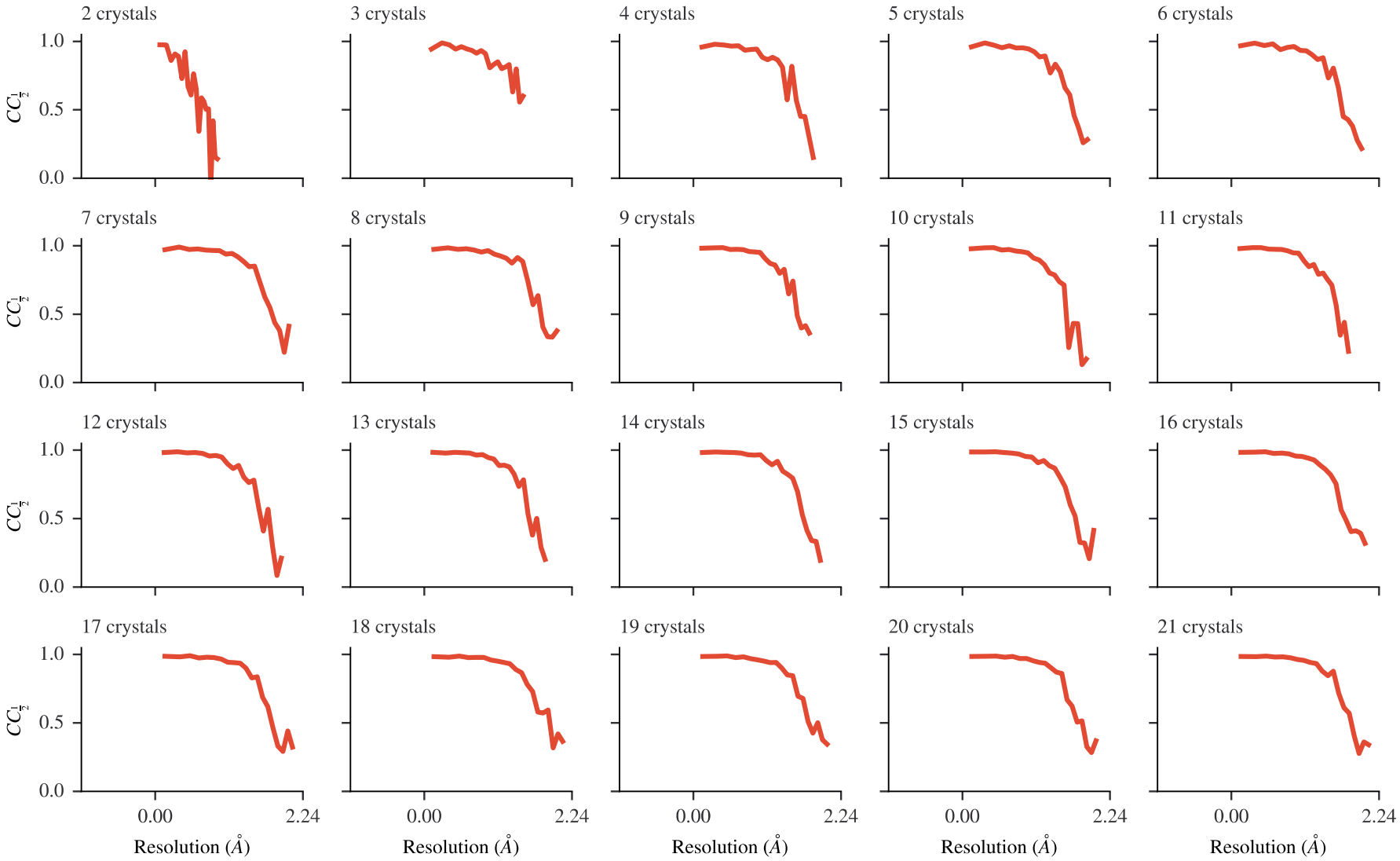
Dataset	*Big*	*Tiny*	*T100*	*Big+T100*	*Big2*
Multiplicity	2	2	200	202	4
$\langle I/\sigma \rangle_{ind}$	2.0	0.2	0.2	0.22	2.0
R_{merge}	28%	280%	399%	395%	35%
R_{meas}	40%	400%	400%	396%	40%
R_{pim}	28%	280%	28%	28%	20%
$\langle I/\sigma \rangle_{mrgd}$	2.8	0.28	2.8	4.0	4.0
$CC_{1/2}$	0.66	0.04	0.66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

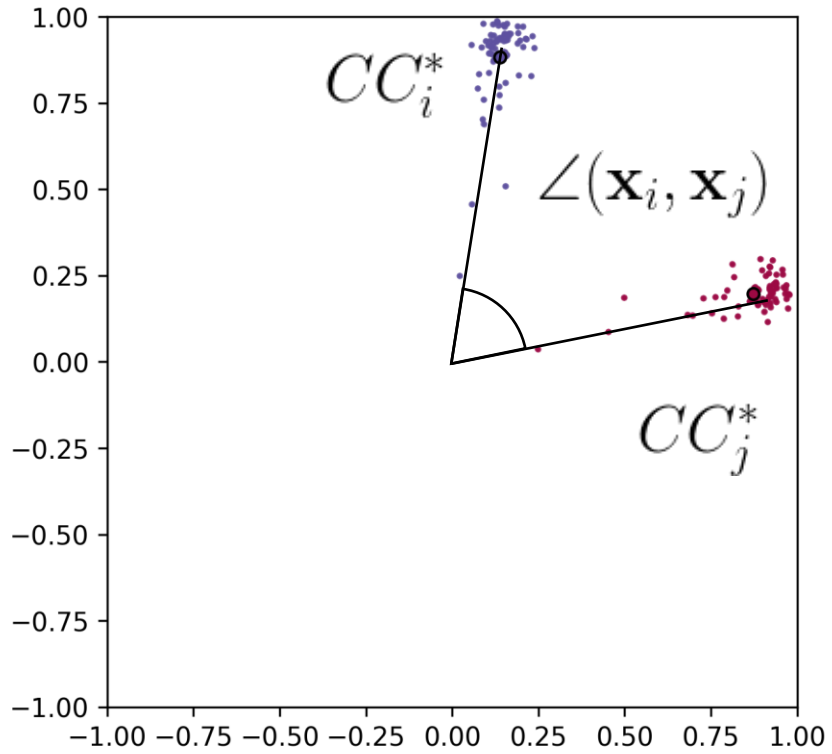


Challenges

- 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)

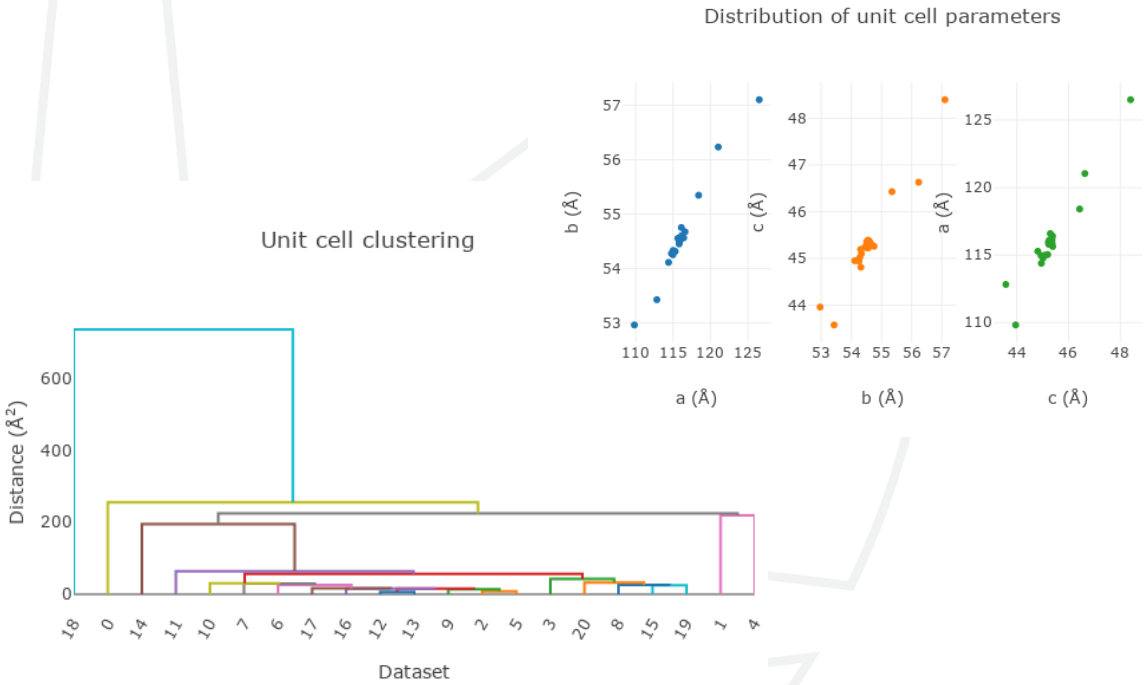


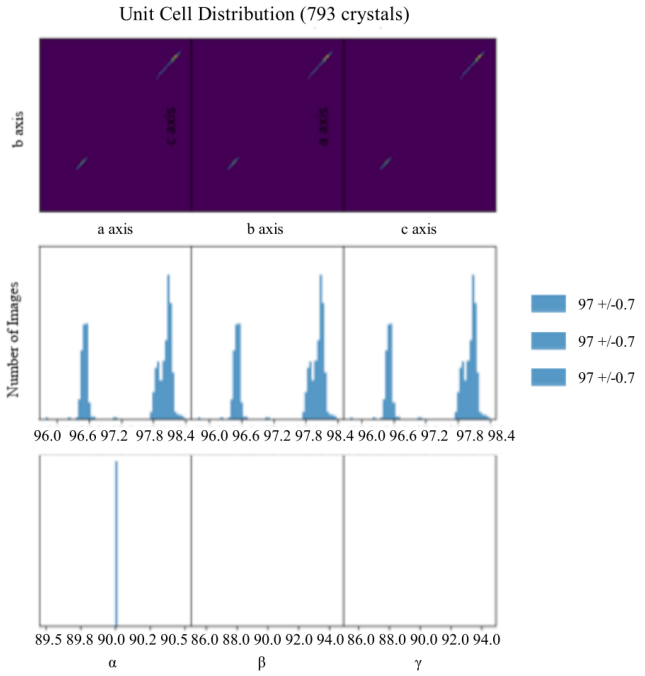
Brehm, W. & Diederichs, K. (2014). Acta Cryst. D70, 101–109
Gildea, R. J. & Winter, G. (2018). *Acta Cryst.* D**74**, 405–410

diamond

Non-Isomorphism

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

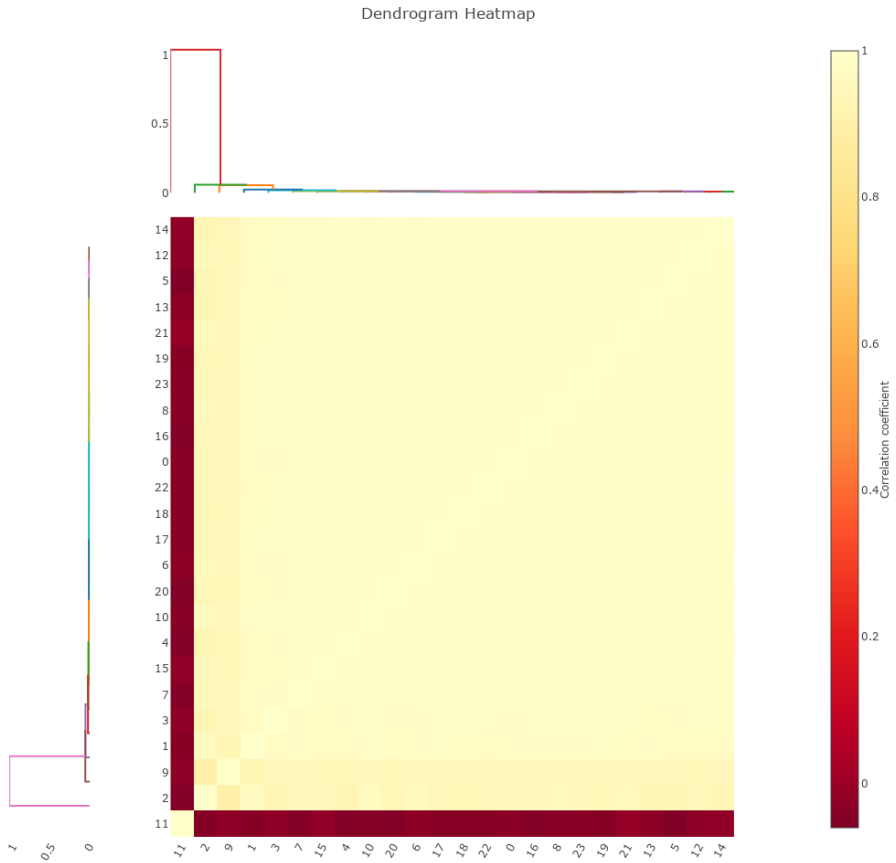




Non-Isomorphism

Clustering on pairwise correlation coefficients may help identify outliers

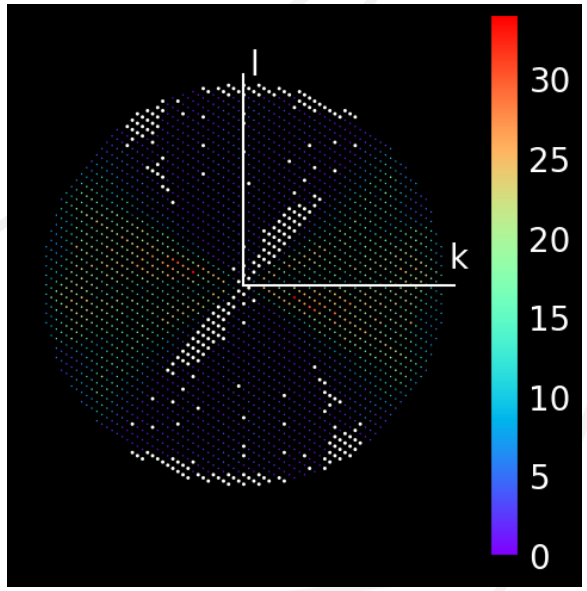
(more on this soon! :))

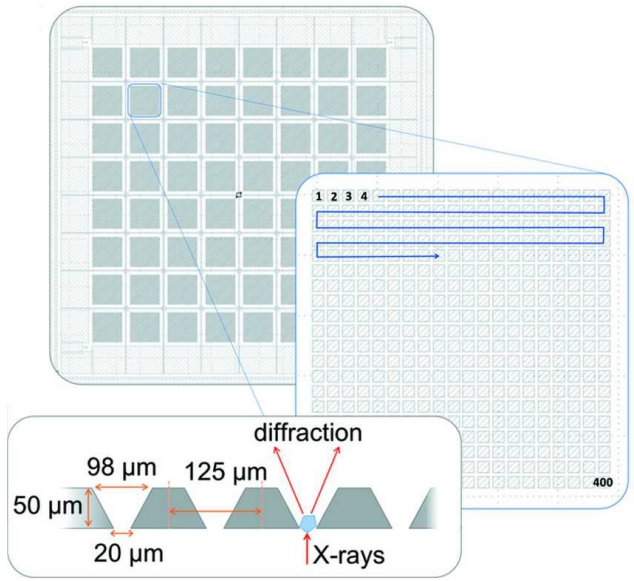


Giordano, R., Leal, R. M. F., Bourenkov, G. P. et al. (2012). Acta Cryst. D68, 649–658
Santoni, G., Zander, U., Mueller-Dieckmann, C. et al. (2017). J. Appl. Cryst. 50, 1844–1851

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

Rotational Multi-Crystal Approaches

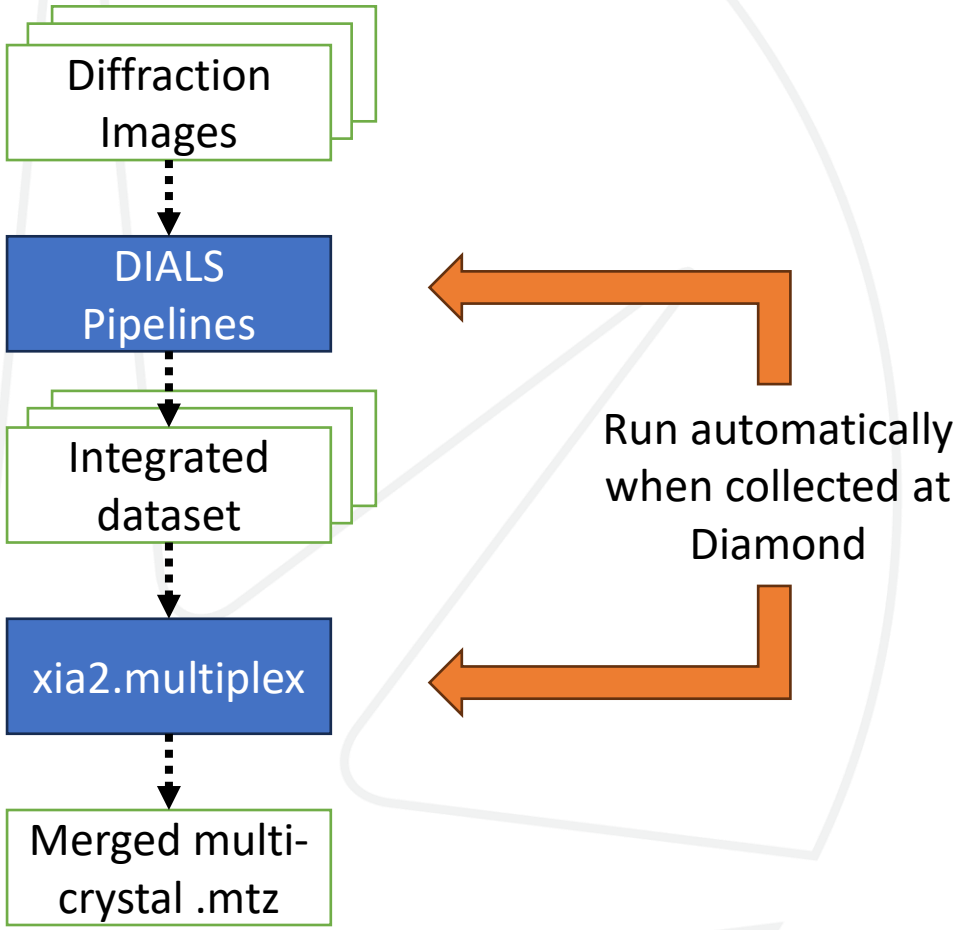
Why rotation?

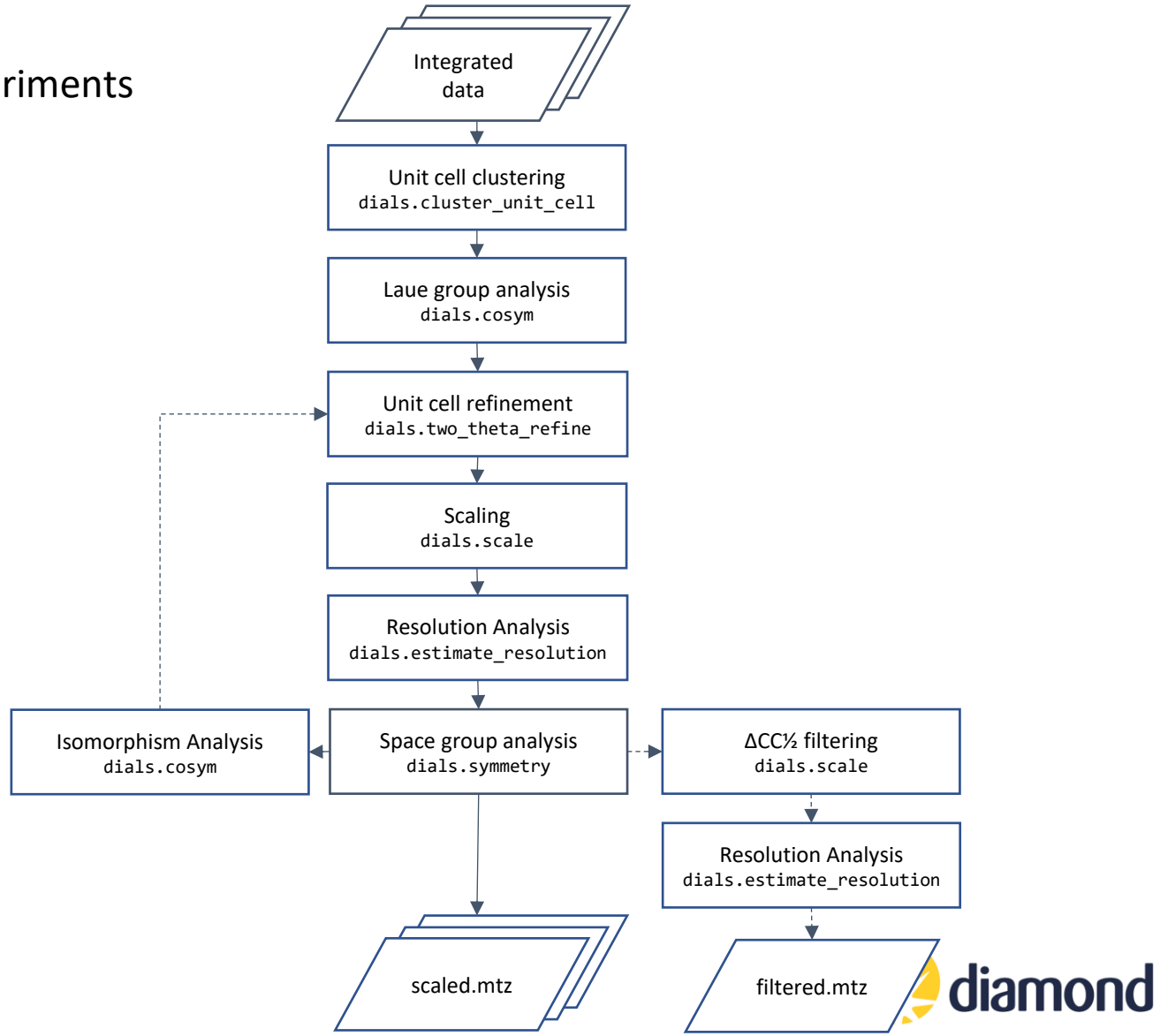
- More straight-forward data processing compared to still-shot data
- Radiation sensitive sample only allows small wedge of data prior to deteriorating
- Identify sub-populations within your sample easily
- Collect room temperature structures
  - No damaging cryo-protectants

diamond

Processing Multi-Crystal Data at Diamond

Auto-processing Pipeline for Rotation Multi-Crystal Experiments

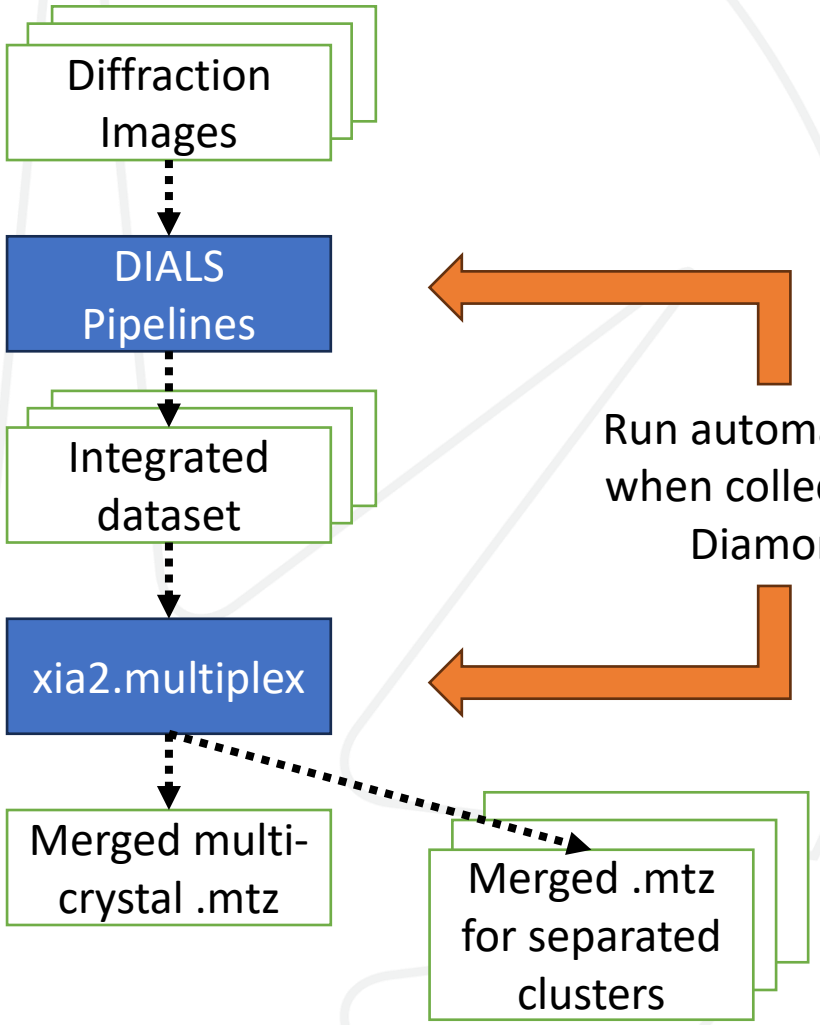




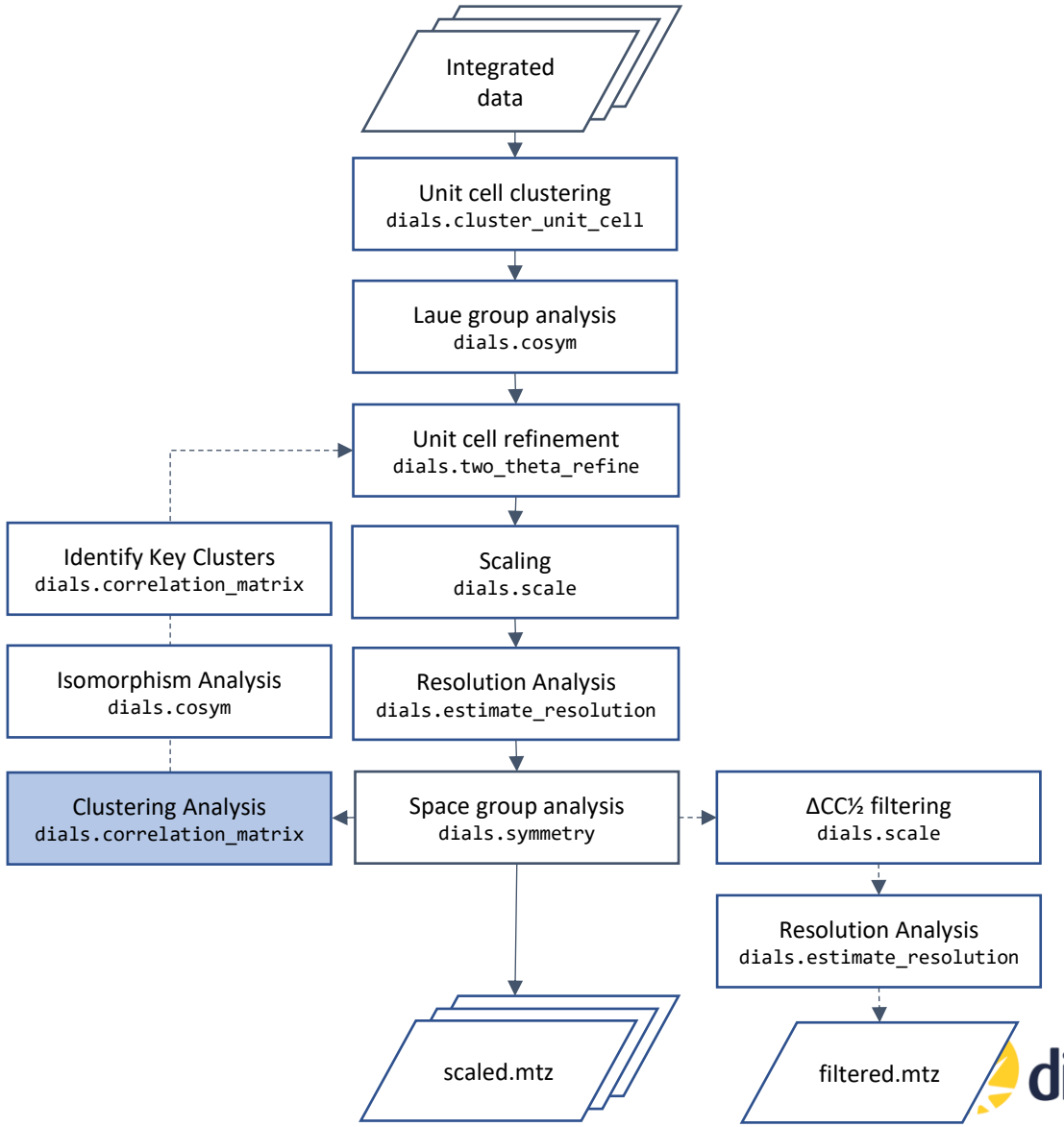
diamond

Processing Multi-Crystal Data at Diamond

Auto-processing Pipeline



Run automatically when collected at Diamond



diamond

Hierarchical Clustering in xia2.multiplex

After consistent symmetry determination using dials.cosym, there are three types of clustering available in xia2.multiplex:

1. Unit Cell Clustering
2. Correlation Clustering (intensity-based)
3. Cosine Angle Clustering (intensity-based)

Correlation Clustering

- Comparison of pairwise correlation coefficients
- Confused by different dataset qualities (random error)

$$r_{ij} = \frac{\sum_h (I_i(h) - \bar{I}_i)(I_j(h) - \bar{I}_j)}{\sqrt{\sum_h (I_i(h) - \bar{I}_i)^2 \sum_h (I_j(h) - \bar{I}_j)^2}}$$

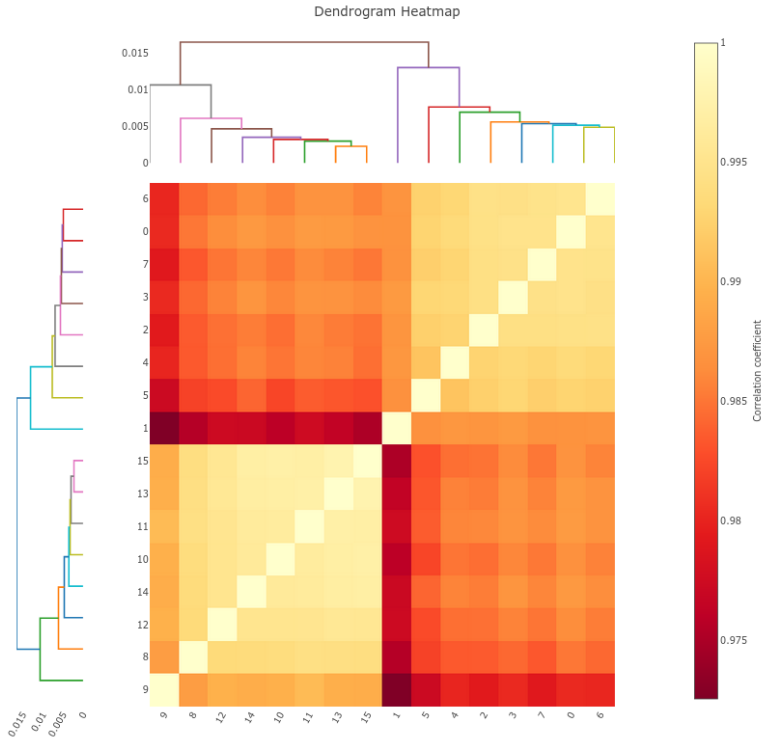
Gildea, R. J. (2022). *Acta. Cryst., D78, 752-769.*

Cosine Angle Clustering

- Extension of correlation clustering
- Separates systematic and random error
- Same methods as dials.cosym

$$\Phi = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (r_{ij} - \boldsymbol{x}_i \cdot \boldsymbol{x}_j)^2$$

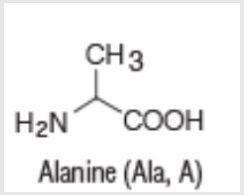
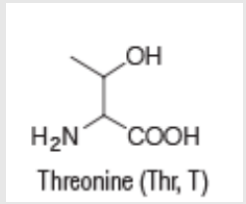
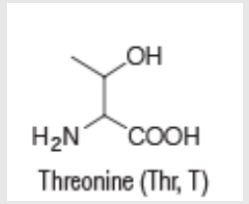
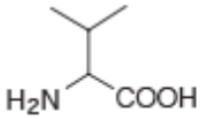
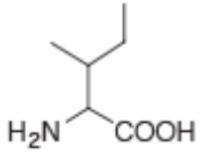
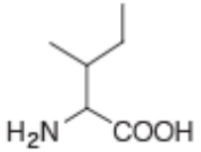
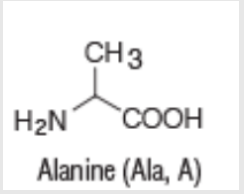
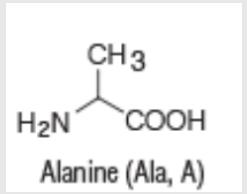
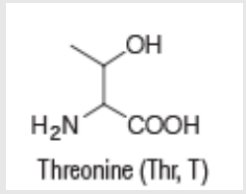
Diederichs, K. (2017). *Acta. Cryst., D73, 286-293.*
Gildea, R. J., Winter, G. (2018). *Acta. Cryst., D74, 405-410.*

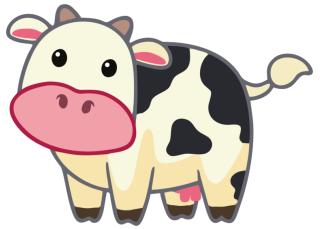


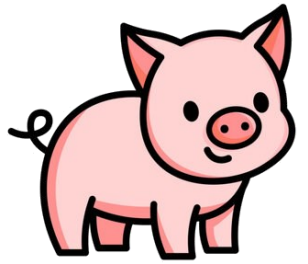
Introducing a Test Case… Cows, Pigs and People

Human insulin, porcine insulin and bovine insulin differ subtly in their amino acid sequences and readily grow high-quality, cubic, isomorphous crystals

~Similar scale of differences to a bound ligand!~

	Cows	Pigs	People
Chain A Residue 8	 Alanine (Ala, A)	 Threonine (Thr, T)	 Threonine (Thr, T)
Chain A Residue 10	 Valine (Val, V)	 Isoleucine (Ile, I)	 Isoleucine (Ile, I)
Chain B Residue 30 (terminus)	 Alanine (Ala, A)	 Alanine (Ala, A)	 Threonine (Thr, T)

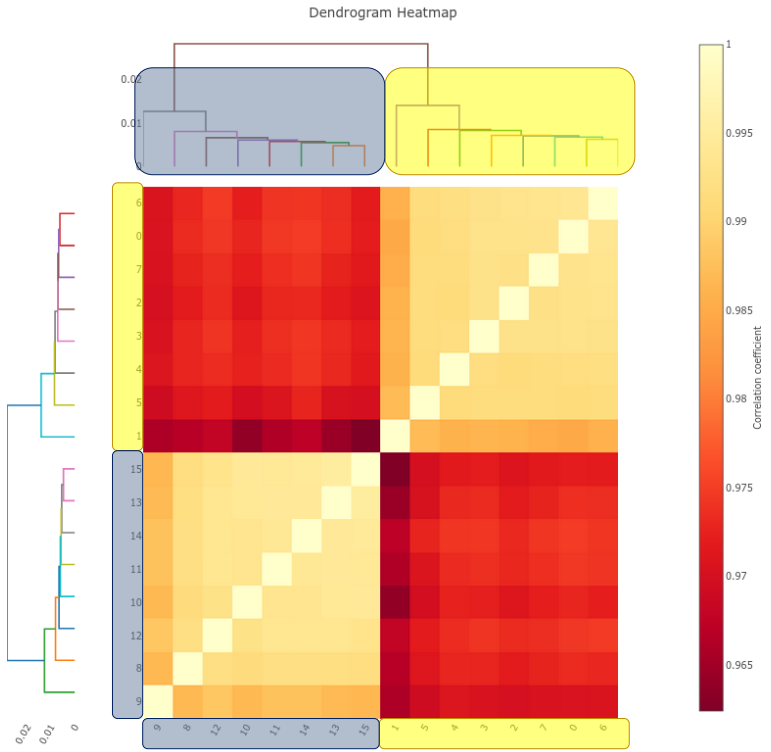




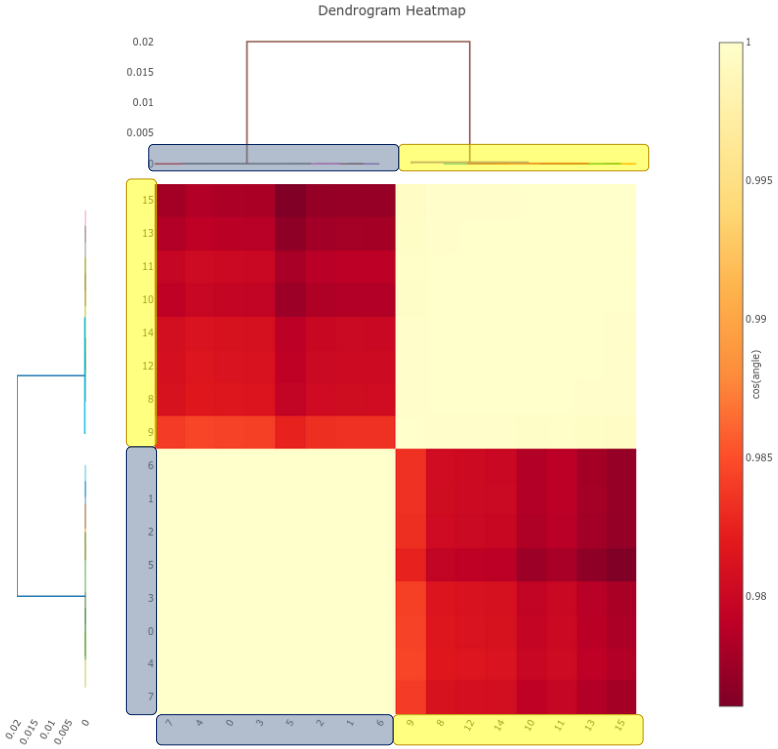
diamond

Cows and People – Room Temperature

- Correlation and cosine angle clustering can correctly separate cows and people
- Tighter clustering when able to separate random and systematic error (cosine angle clustering)



Correlation Clustering

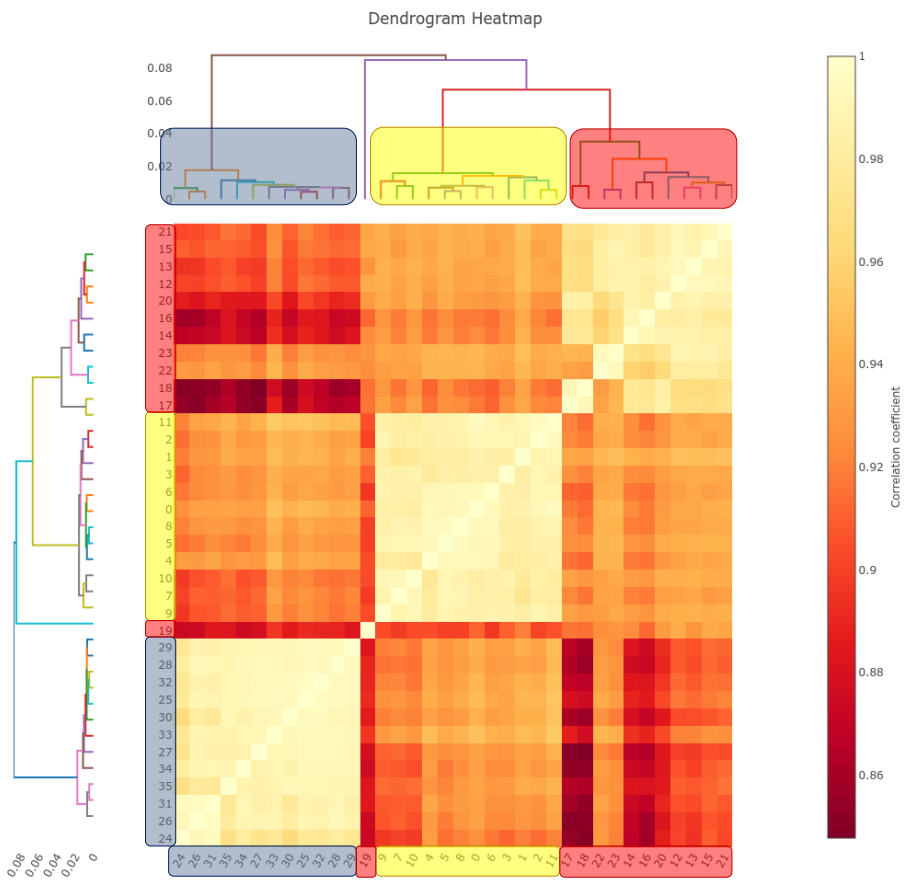


Cosine Angle Clustering

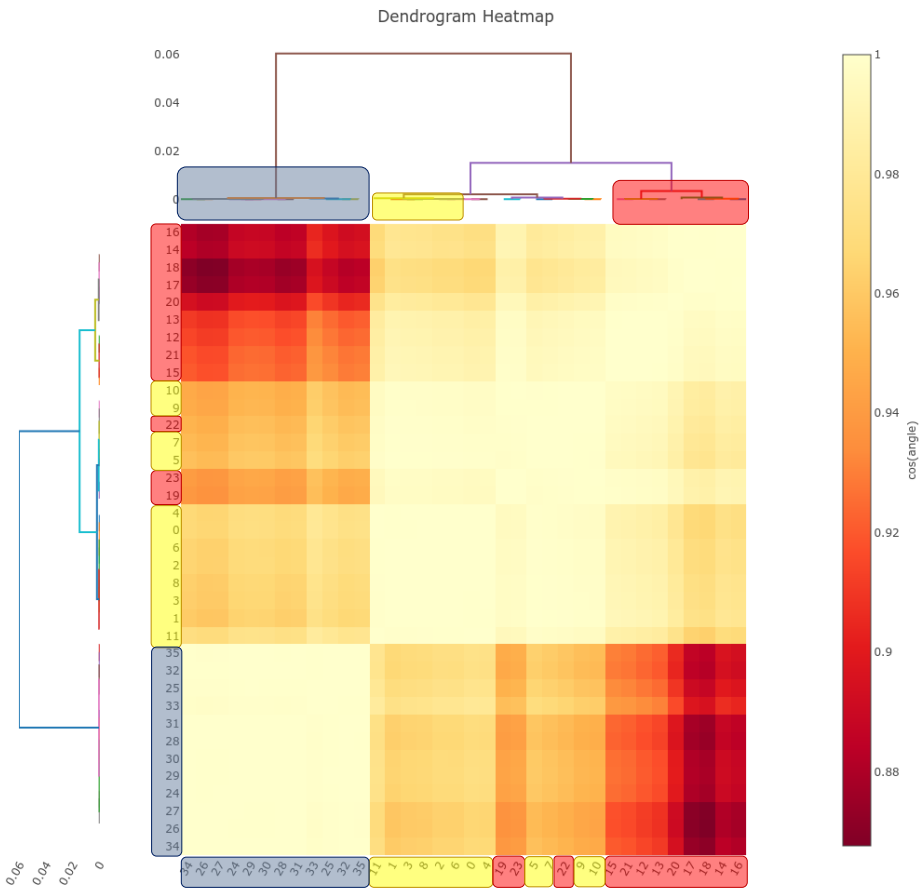
diamond

Cows, Pigs and People – Cryo Temperature

- When add a third type of insulin, correlation clustering still succeeds (although one outlier)
- Cosine angle clustering can no longer reliably distinguish all three groups



Correlation Clustering



Cosine Angle Clustering

Dimension Optimisation for Cosine Angle Clustering

As xia2.multiplex uses dials.cosym methods for cosine angle clustering, it was optimized to determine a consensus symmetry, not cluster different groups. Therefore, the initial implementation was strictly performed in 2-dimensions.

However, the theory that these methods were based upon (Diederichs, K., 2017) mention that this can be tailored:

- **As more types of systematic differences are present, higher dimensions need to be used when minimizing**

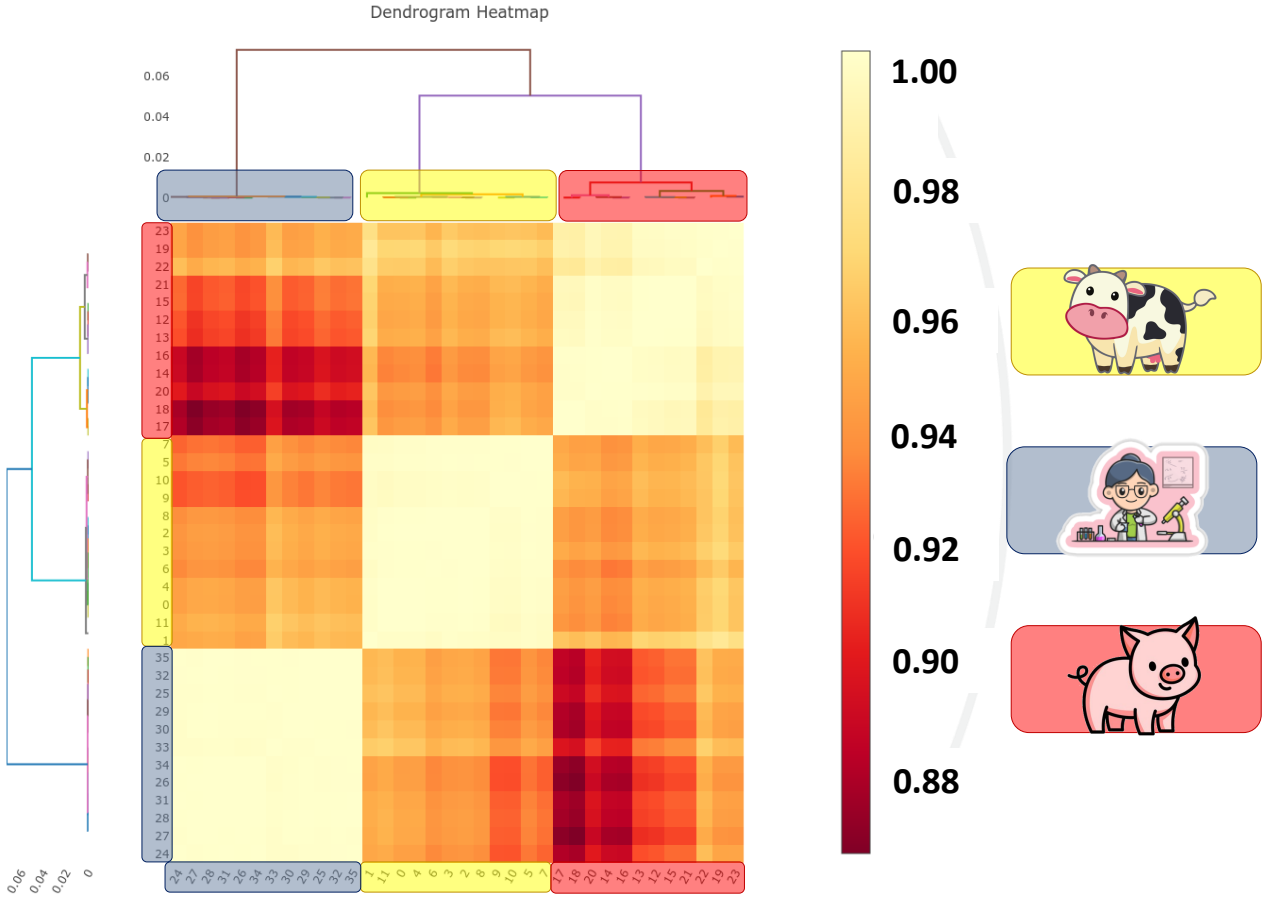
$$\Phi = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (r_{ij} - \boldsymbol{x}_i \cdot \boldsymbol{x}_j)^2$$

For this dataset of cows, pigs and people, three were needed!!!

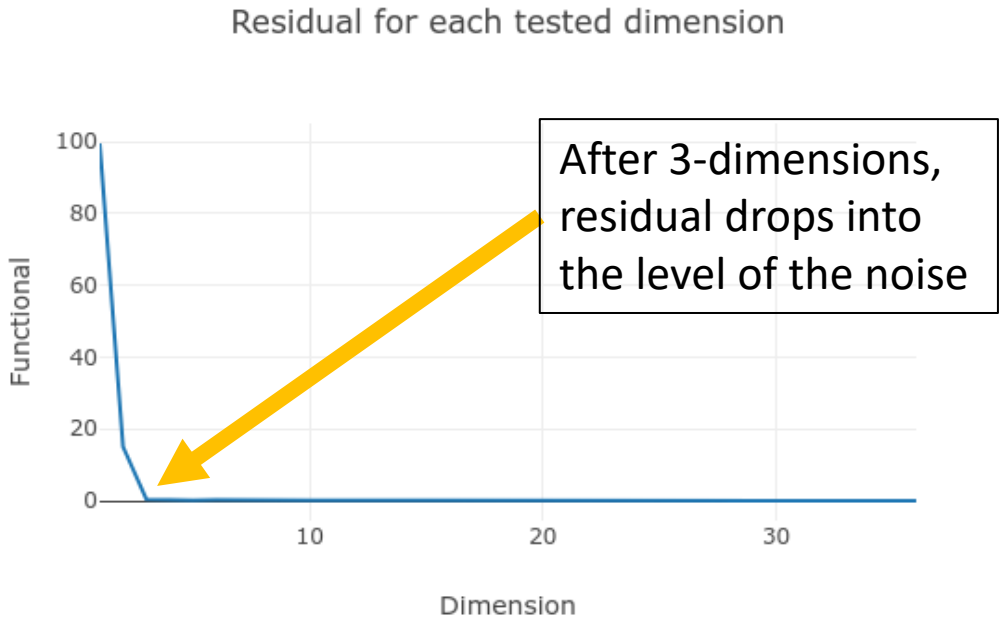
Diederichs, K. (2017). *Acta. Cryst., D73, 286-293.*

Cows, Pigs and People – Cryo Temperature

- Performing the cosine angle clustering in three dimensions now resolves cows, pigs and people (including the outlier dataset identified in the correlation clustering)



3D Cosine Angle Clustering



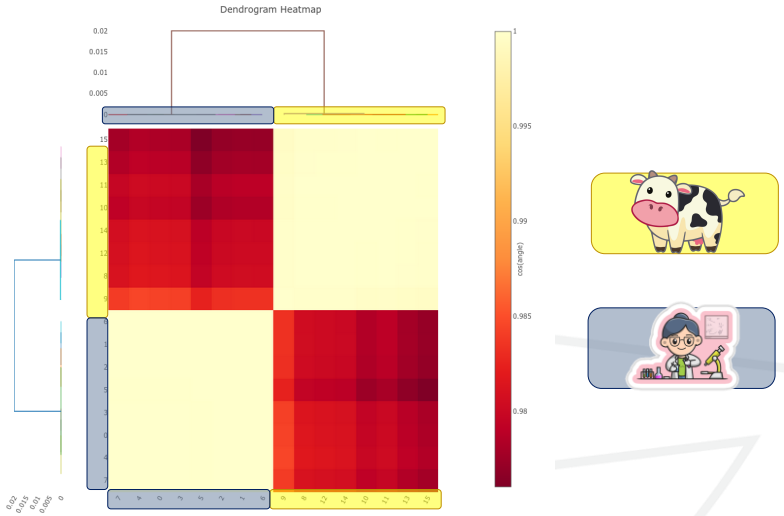
But how do we get these clusters out automatically?

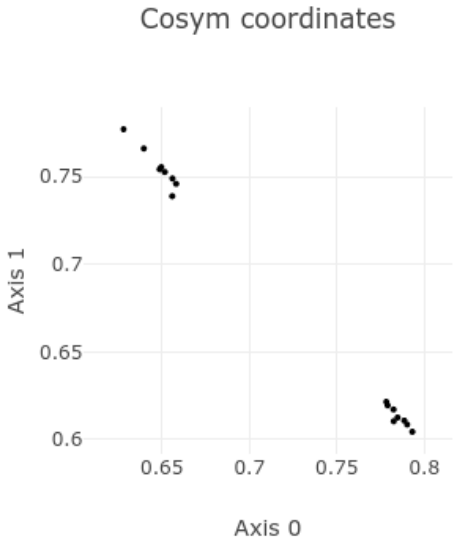
Introducing the OPTICS algorithm

- To perform the cosine-angle clustering, we represent each dataset as a point in some-dimensional space based on its correlation with other datasets

$$\Phi = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (r_{ij} - \boldsymbol{x}_i \cdot \boldsymbol{x}_j)^2$$

- This produces coordinate plots for visualisation
  - Points further away from the origin have less random error
  - Pairs of points with smaller angles have less systematic error between them



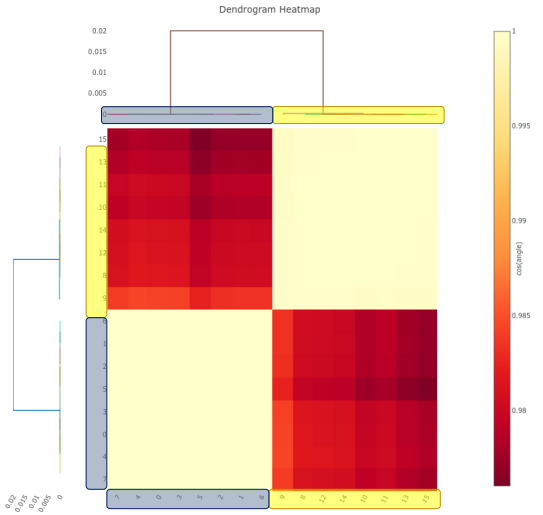


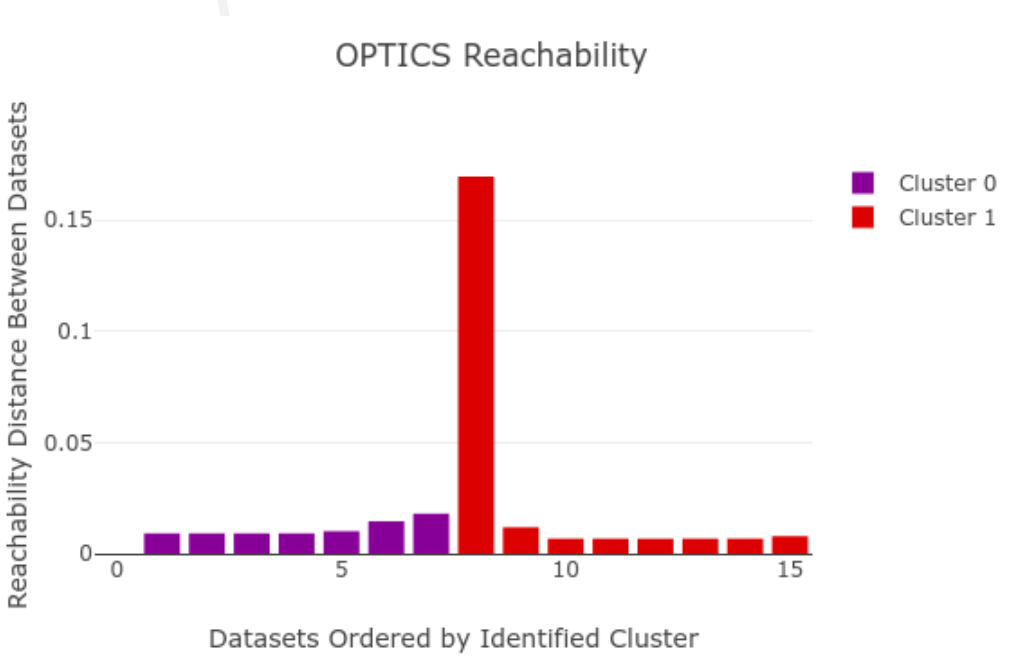
Introducing the OPTICS algorithm

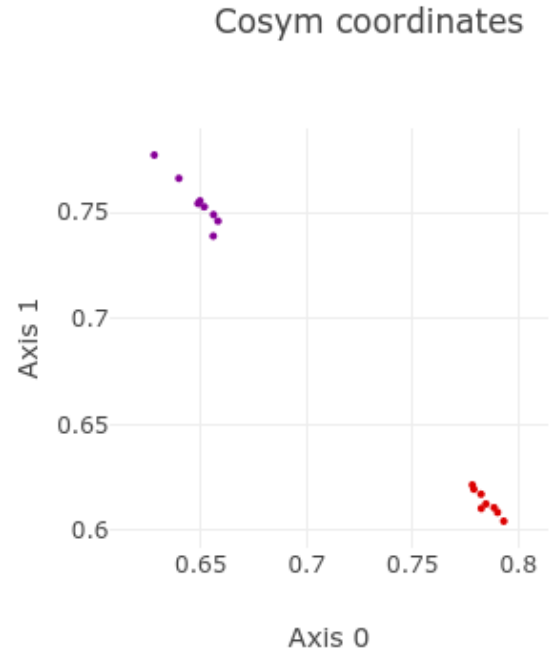
OPTICS is a density-based algorithm that automatically defines clusters with spatial relations (which we have calculated using cosine-angle clustering!)

- Requires no global density parameter (can detect clusters of varying density)
- Does not assume any specific number of clusters
- Can detect noise (does not require every dataset to be a part of a cluster)

OPTICS orders the datasets to best represent the spatial density structure, visualised using a “Reachability Plot”

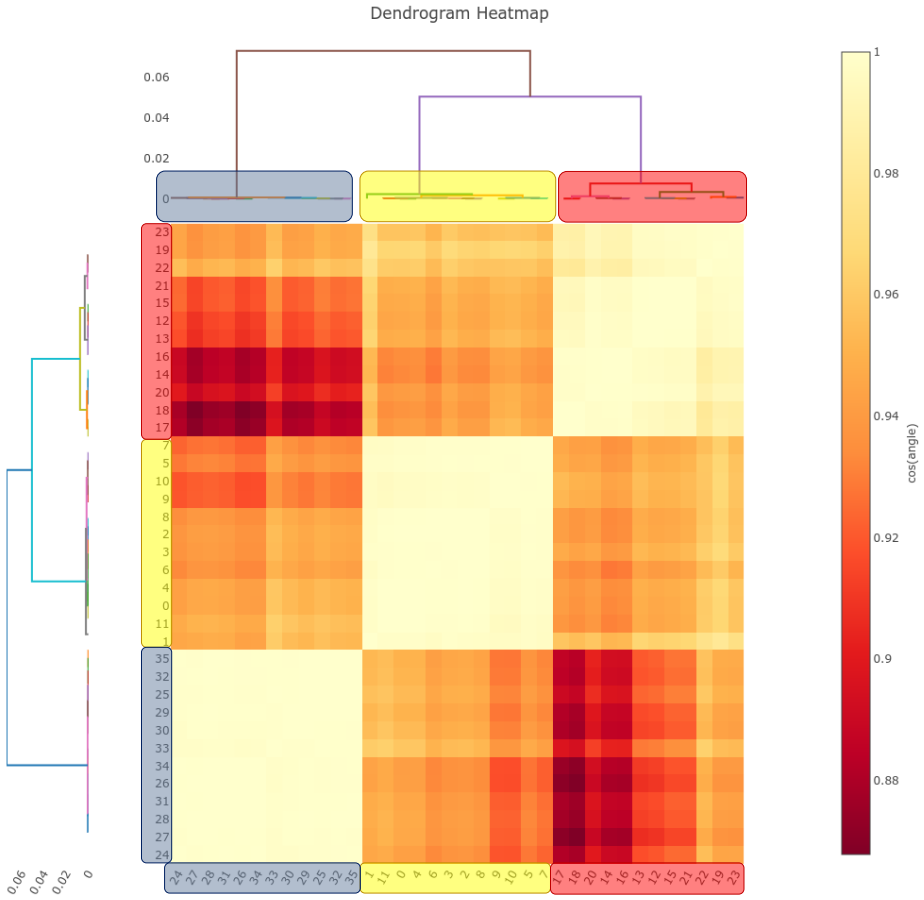


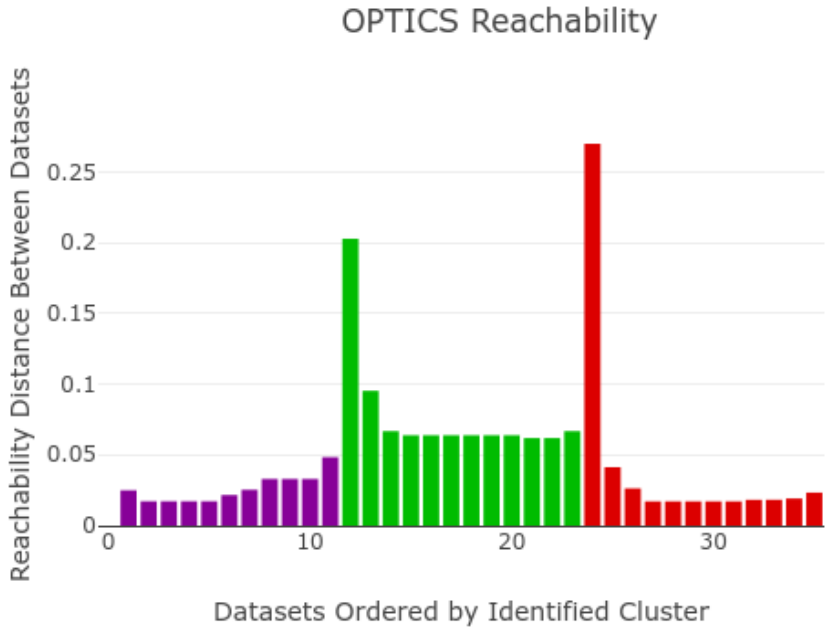


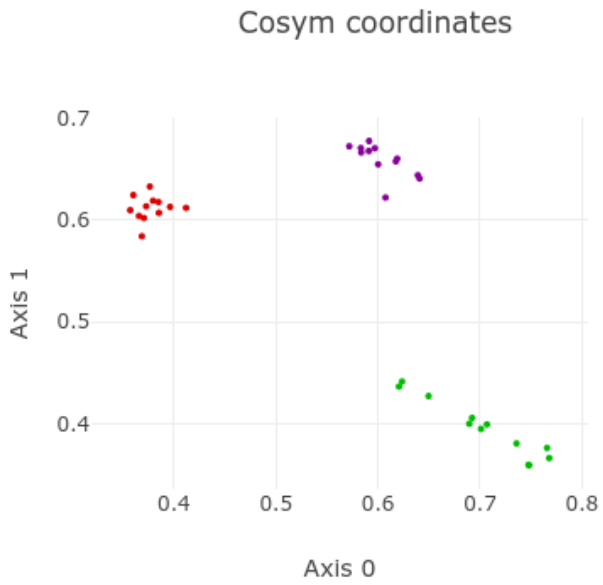


Introducing the OPTICS algorithm

For the 3-Dimensional cows, pigs and people measured under cryogenic conditions:

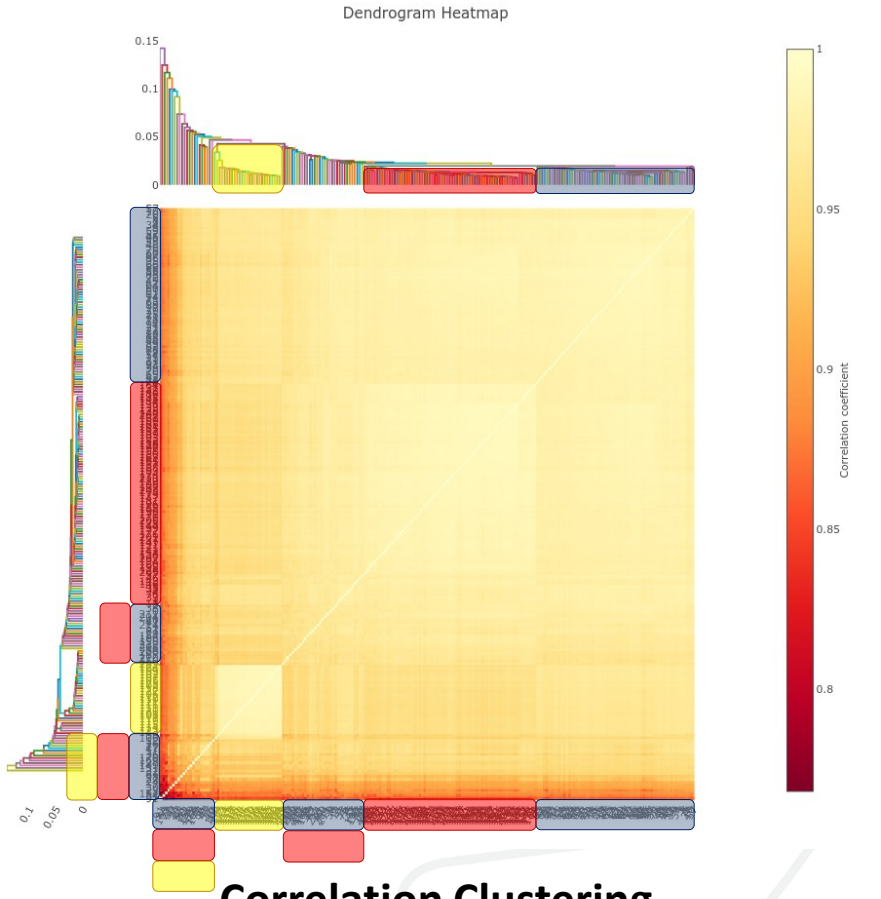




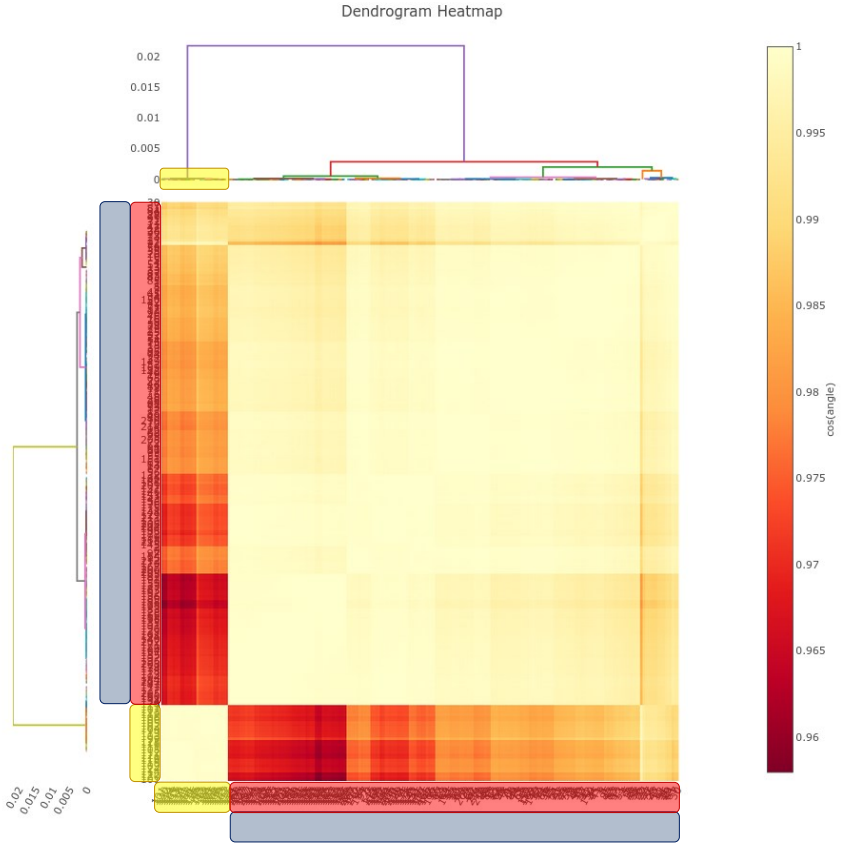


Cows, Pigs and People – Room Temperature

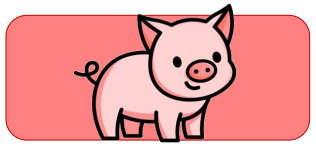
- Can also distinguish all three at room temperature!
- People and pigs mix in the dendrogram before cows – consequence of terminal residue position being less well defined at room temperature
- 2D-clustering once again not using high enough dimensions to resolve all three groups in cosine angle



Correlation Clustering



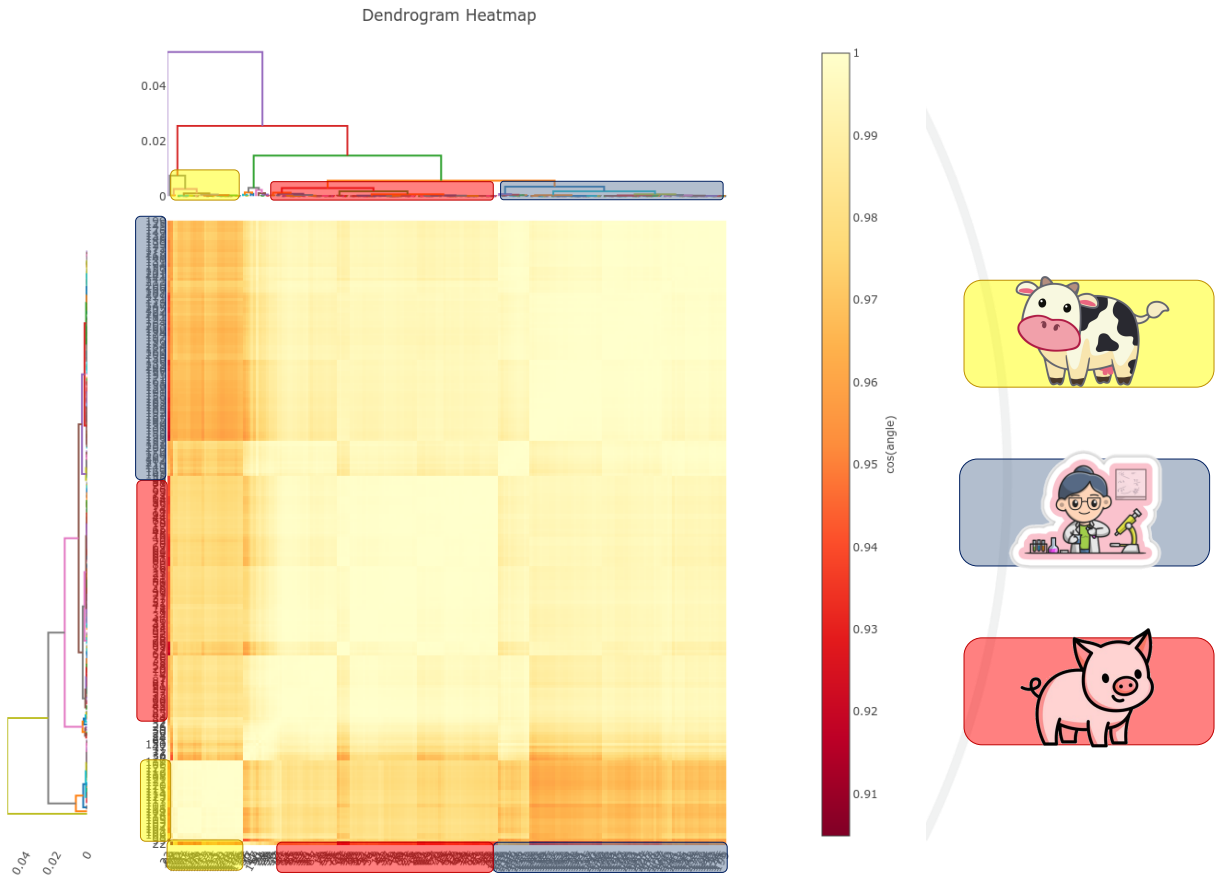
2D Cosine Angle Clustering



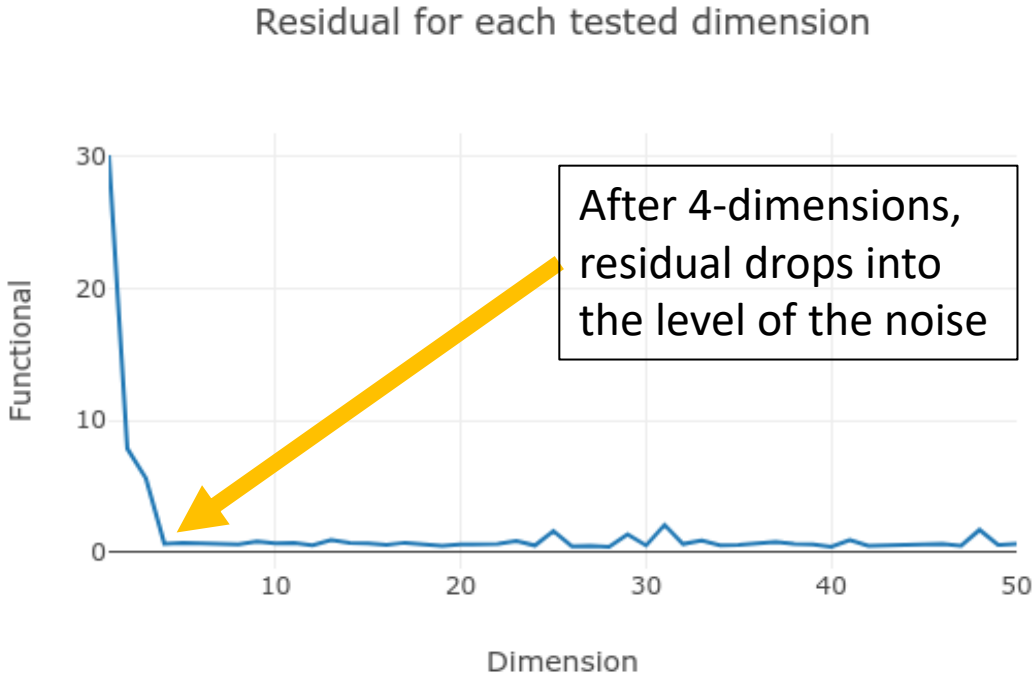
diamond

Cows, Pigs and People – Room Temperature

- Optimised 4D clustering has well separated pigs and people clusters, and an outlier group of datasets are identified



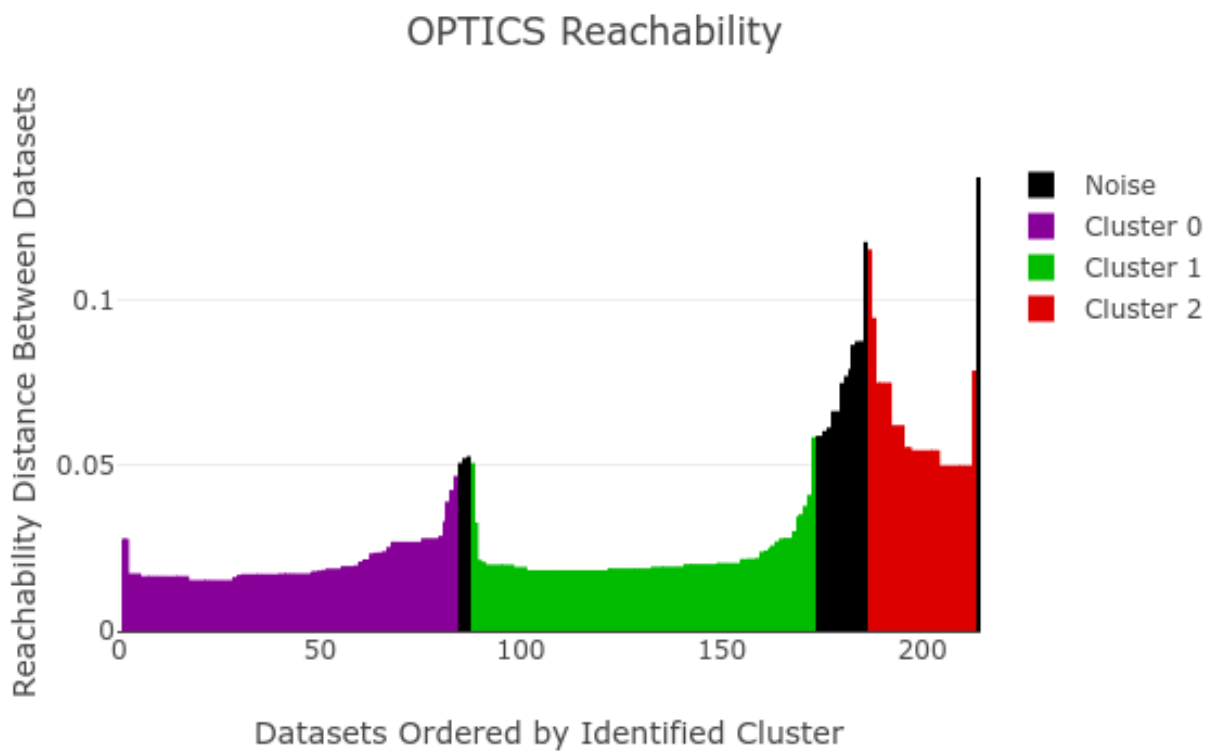
4D Cosine Angle Clustering

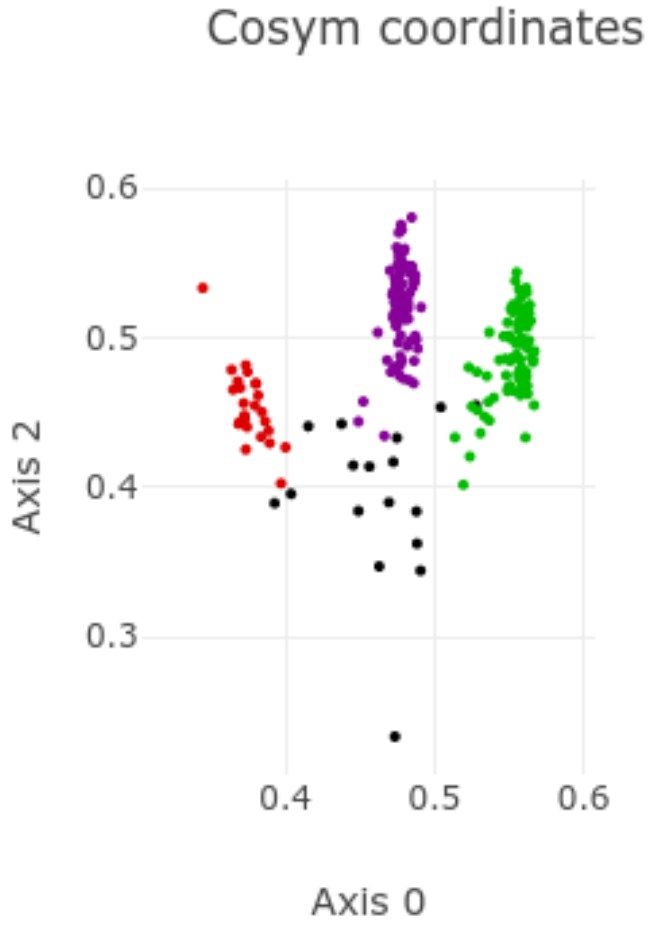


diamond

Cows, Pigs and People – Room Temperature

- 4D clustering has well separated pigs and people clusters, and an outlier group of datasets are identified





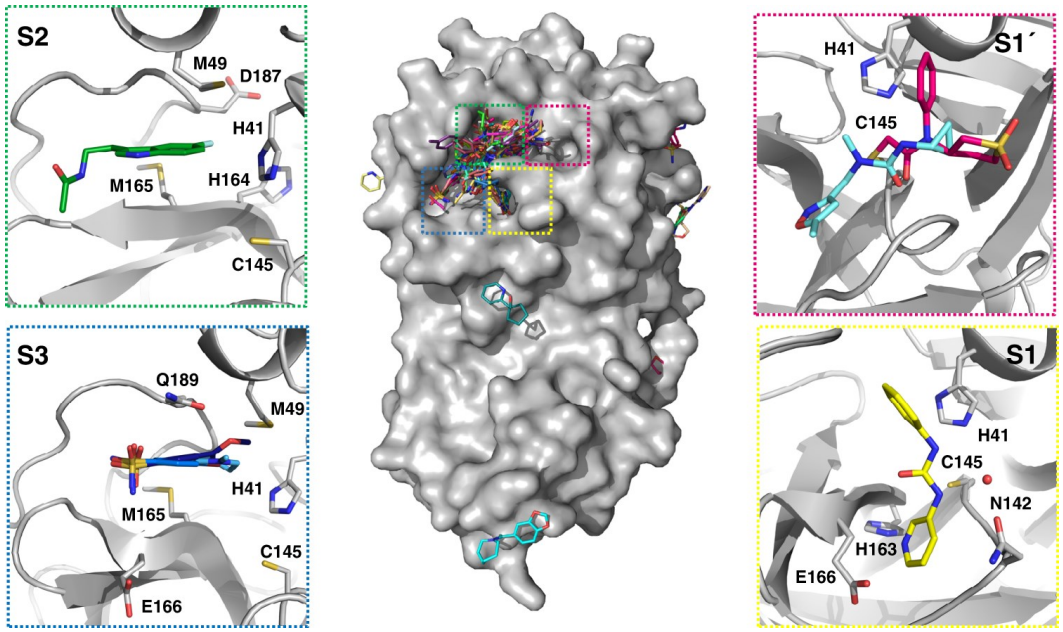
- Classification of outliers (ie steep zones in the plot) controlled by ξ-parameter, and by setting the minimum number of datasets

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



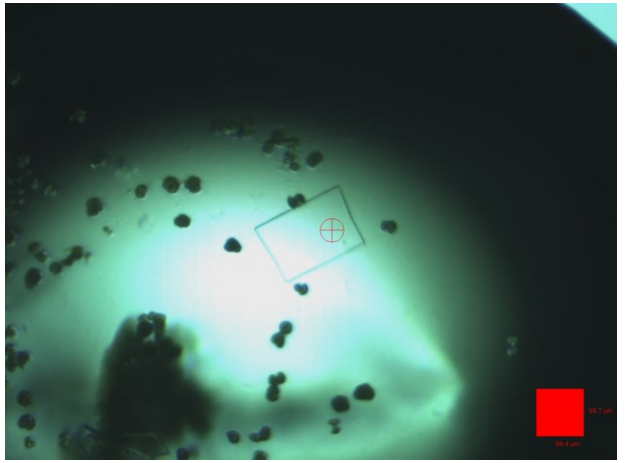
Douangamath, A., Fearon, D., Gehrtz, P. et al. (2020). Nat. Commun. 11, 5047

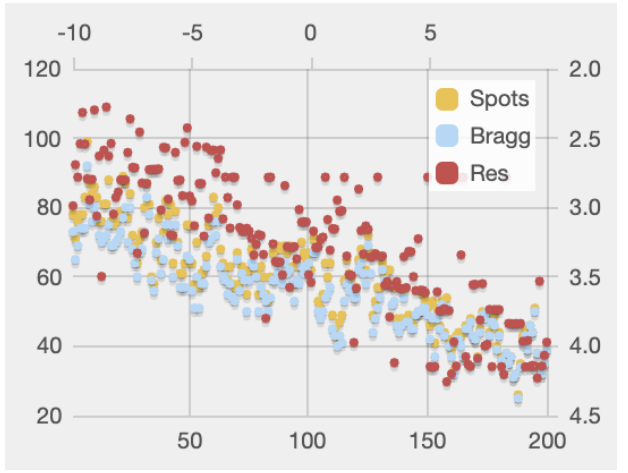
diamond

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



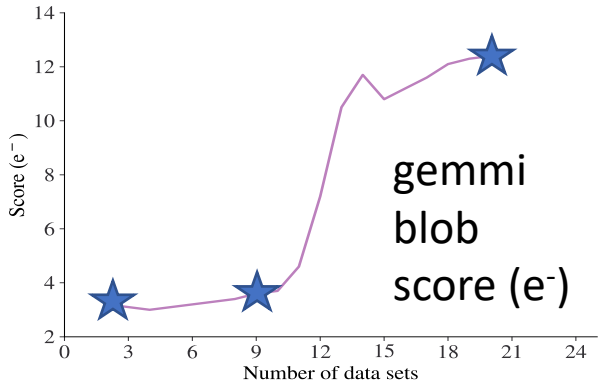


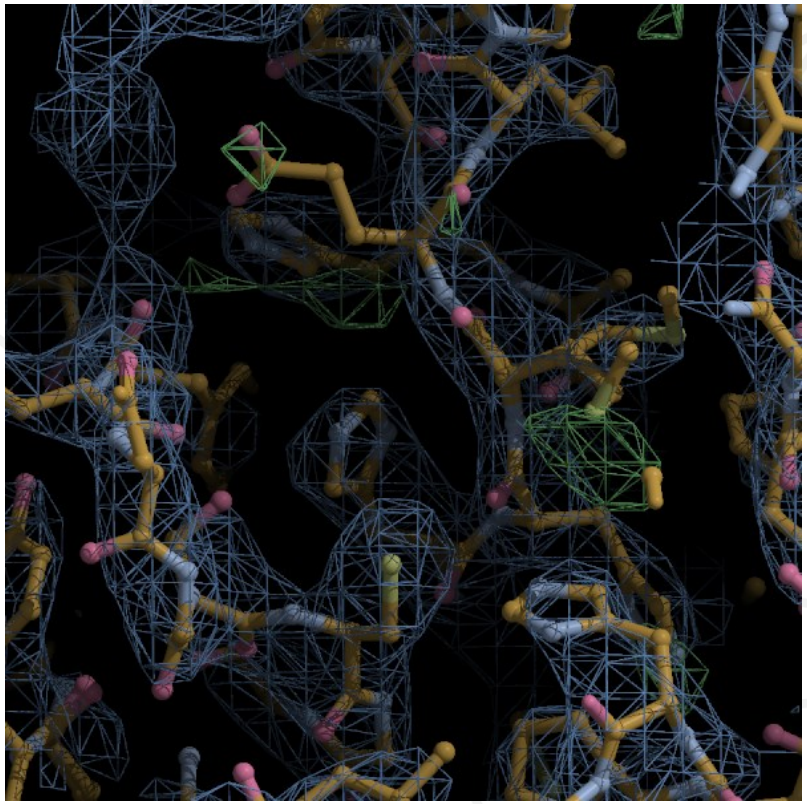
Gildea, R. J., Beilsten-Edmands, J., Axford, D., et al. (2022) Acta Cryst. D78, 752-769

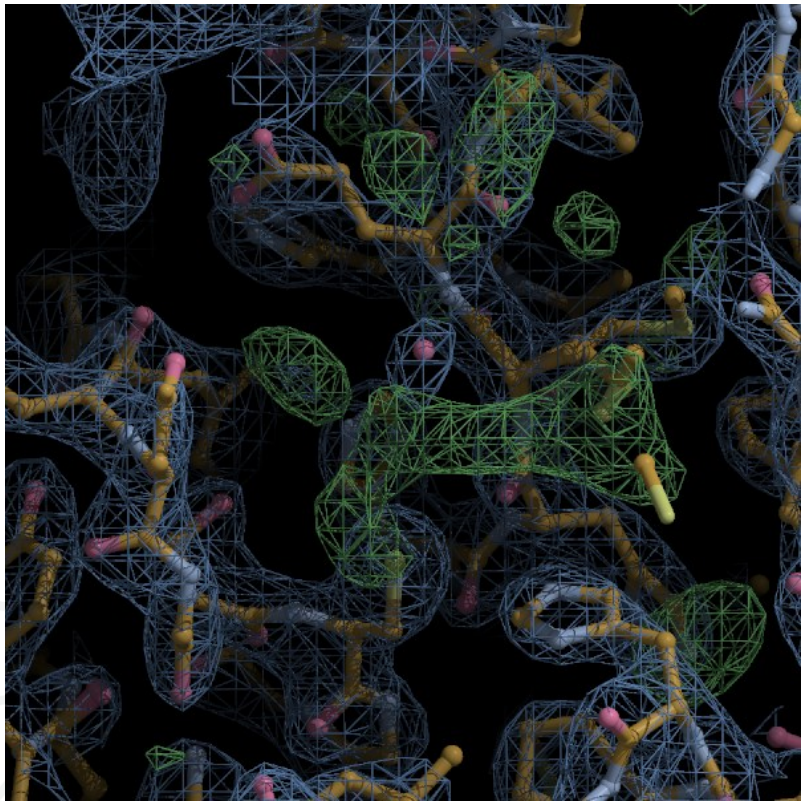
diamond

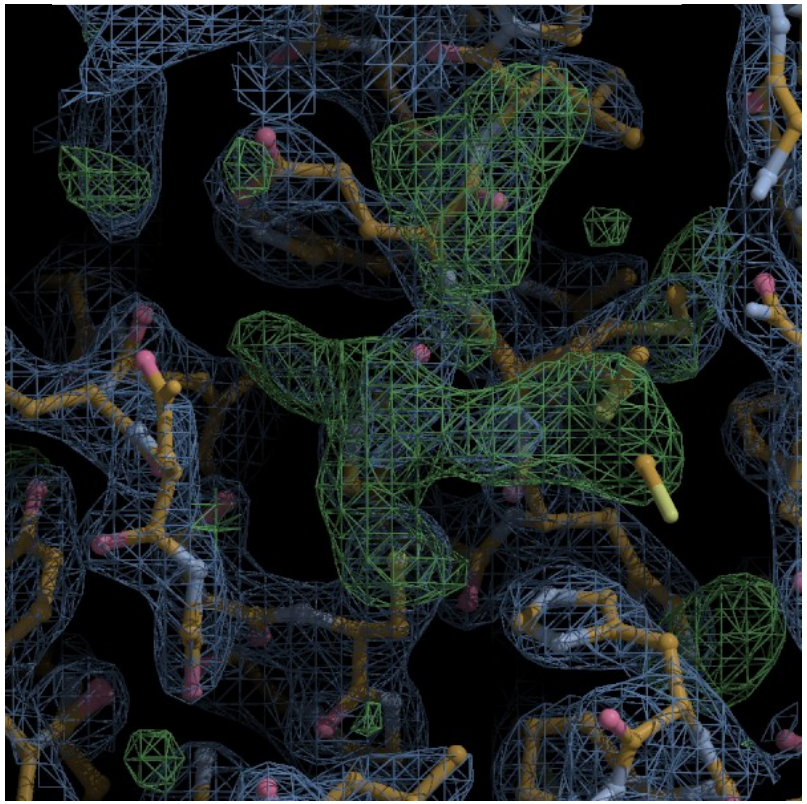
Example: SARS-CoV-2 M^{pro} ligand screening

Automatic dimple maps



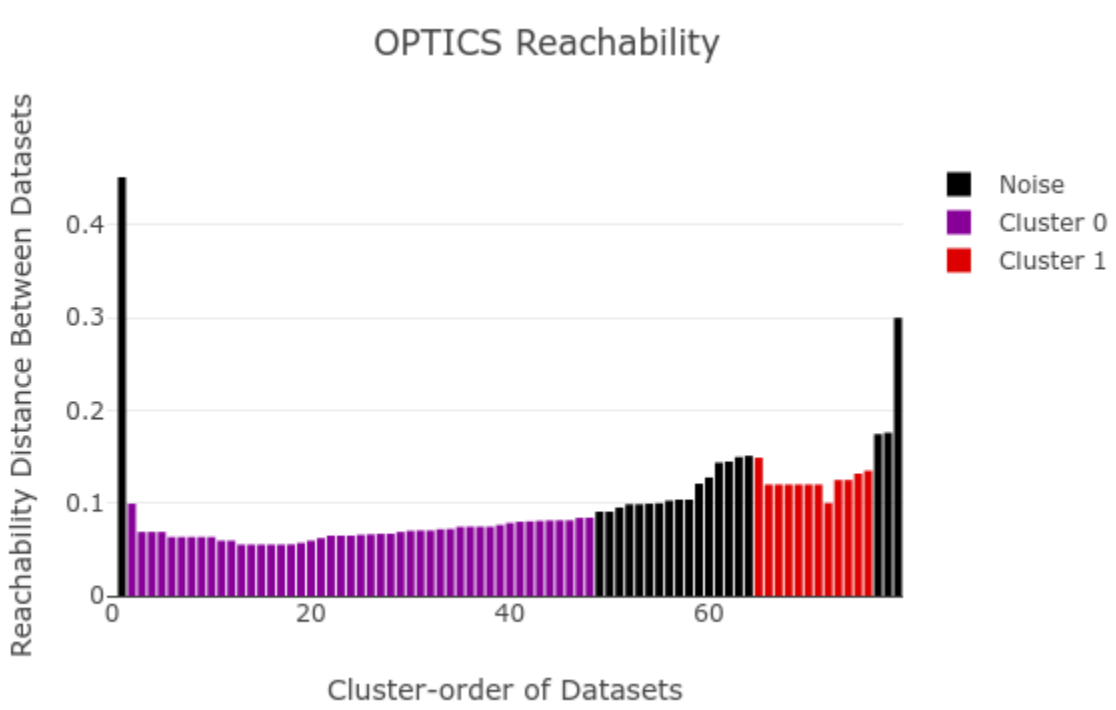


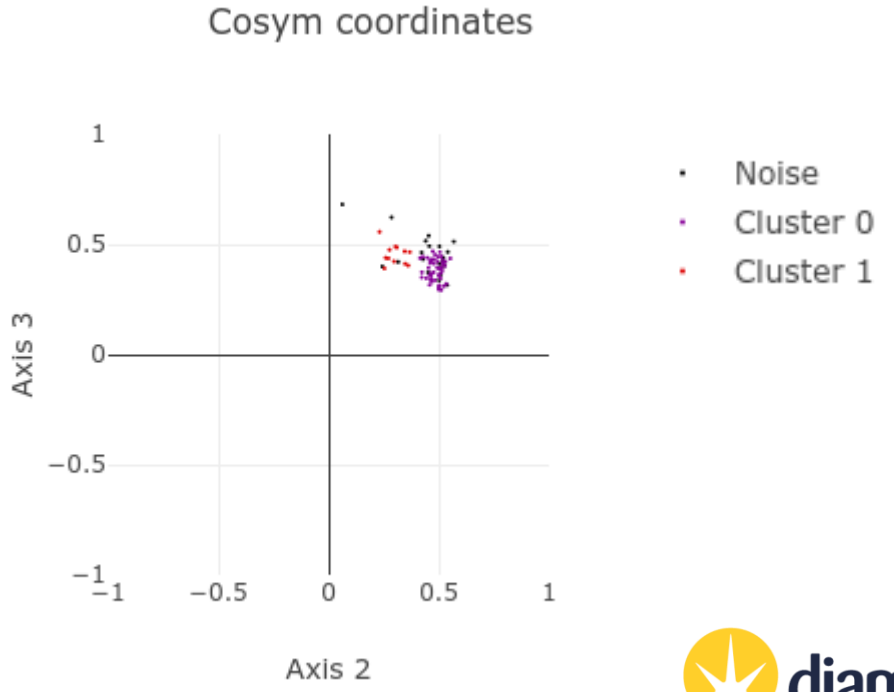




Example: Mac1 ligand screening

- Follow up compounds developed based on initial hits from a previous fragment screen
- Measured multi-crystal fragment screen on VMXi using new clustering updates to further analyse
- Data first separated by which ligand was dispensed (sample-groups)
- Each sample-group run through xia2.multiplex with new clustering algorithms
- One fragment gave a very interesting result:

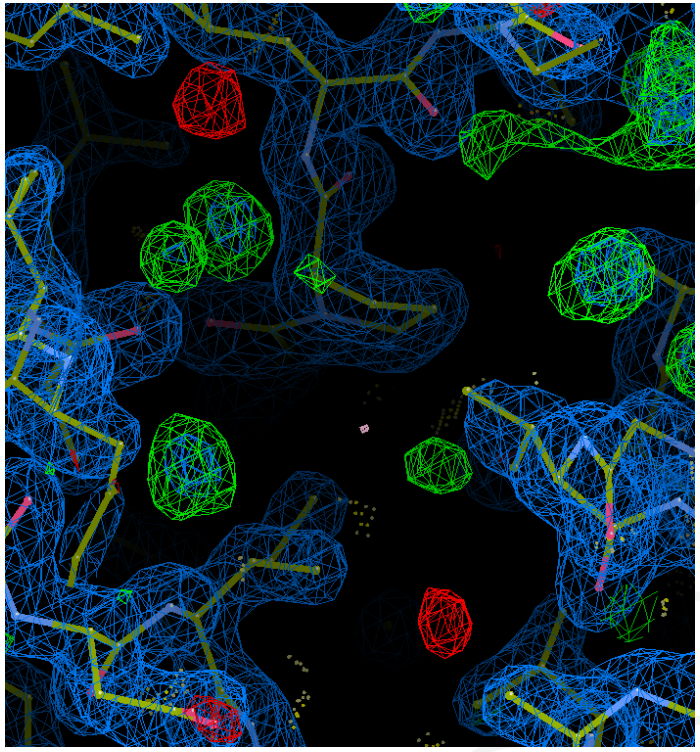




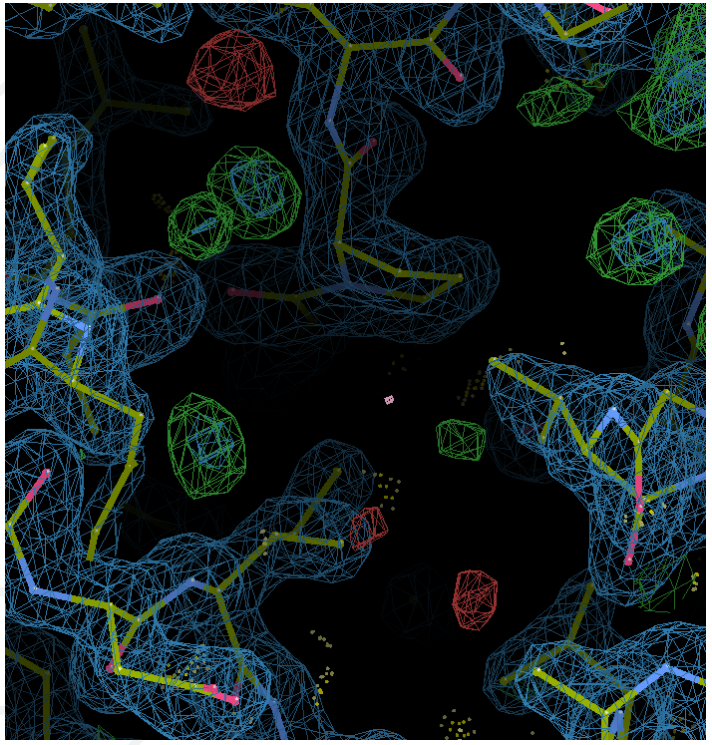
Example: Mac1 ligand screening

- The entire dataset, and both identified clusters were scaled, merged, and refined against the reference PDB for Mac1

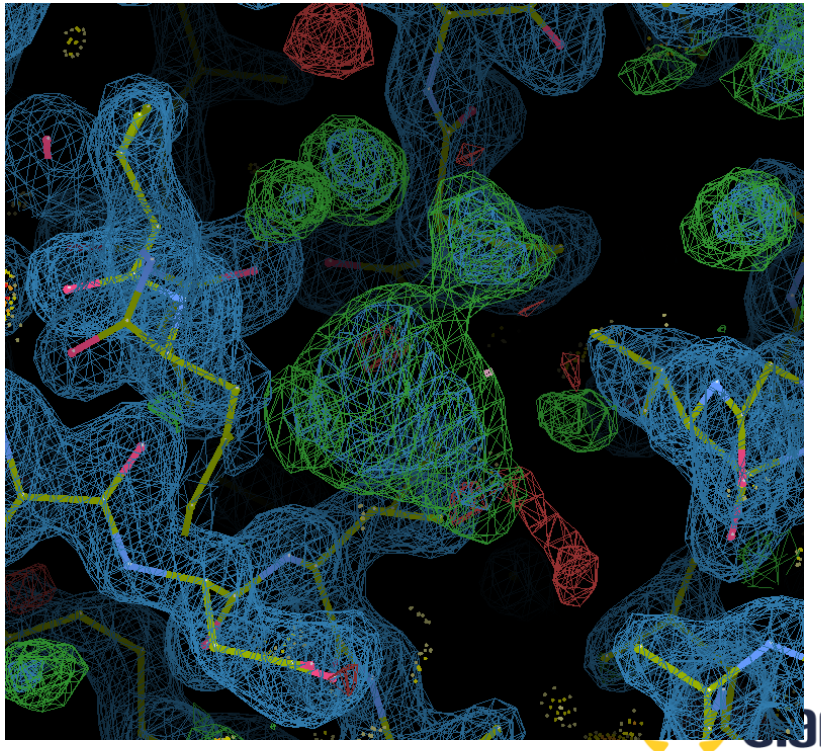
All Data



Coordinate Cluster 0



Coordinate Cluster 1



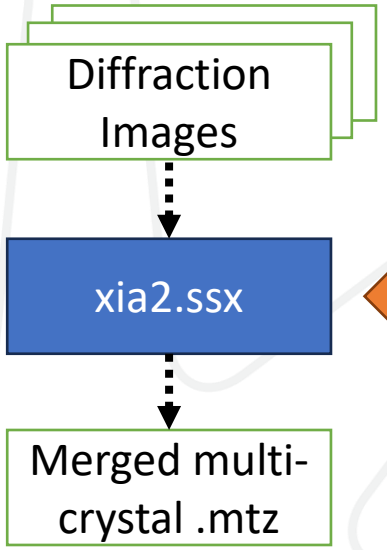
Still-Shot Multi-Crystal Approaches

Why still shots?

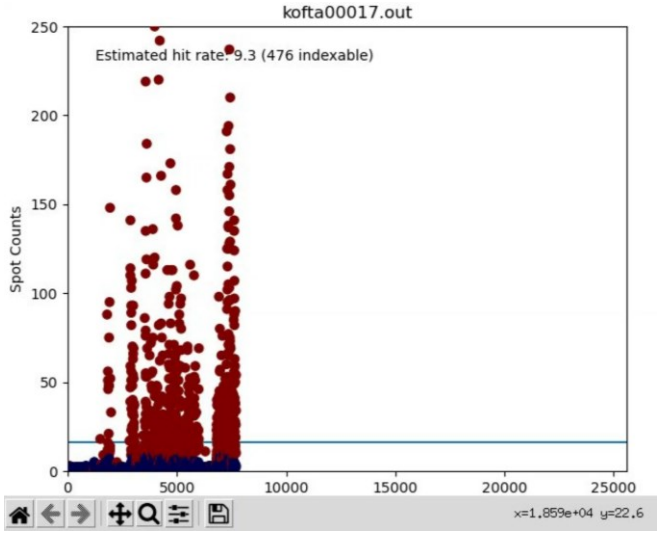
- Crystals extremely radiation sensitive
- Pump-probe experiments better with very small samples to ensure entire crystal affected by change
  - Ligand binding
  - Light activated reactions
- Time-resolved experiments
- Can use extremely high flux for weakly diffracting samples (diffraction before destruction)

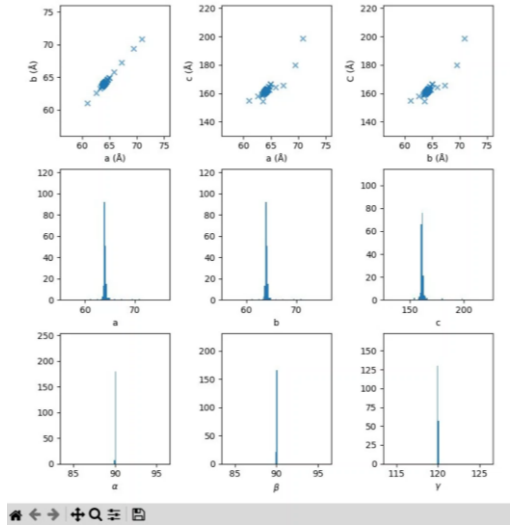
Processing Multi-Crystal Data at Diamond

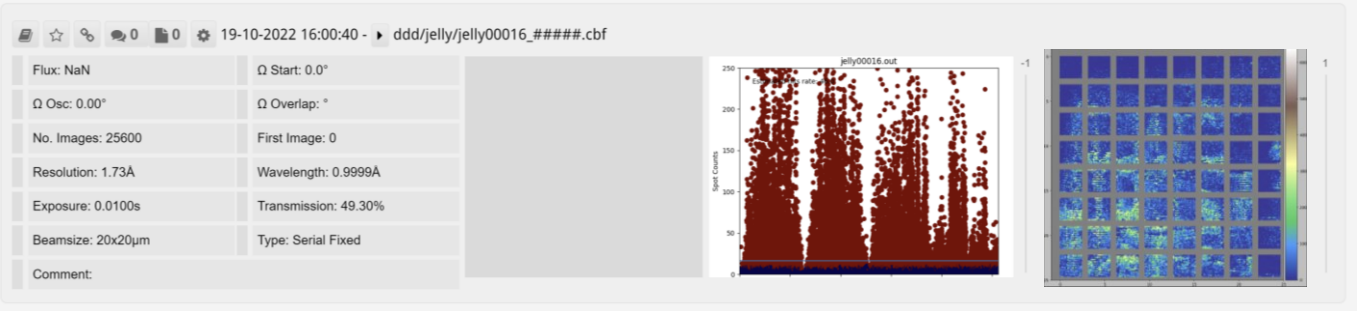
Auto-processing Pipeline for still-shot Multi-Crystal Experiments



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



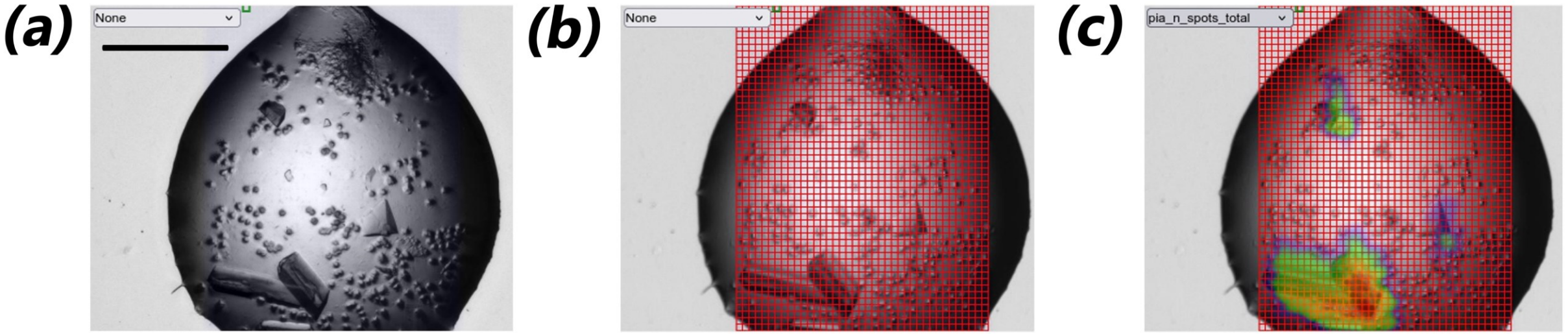




diamond

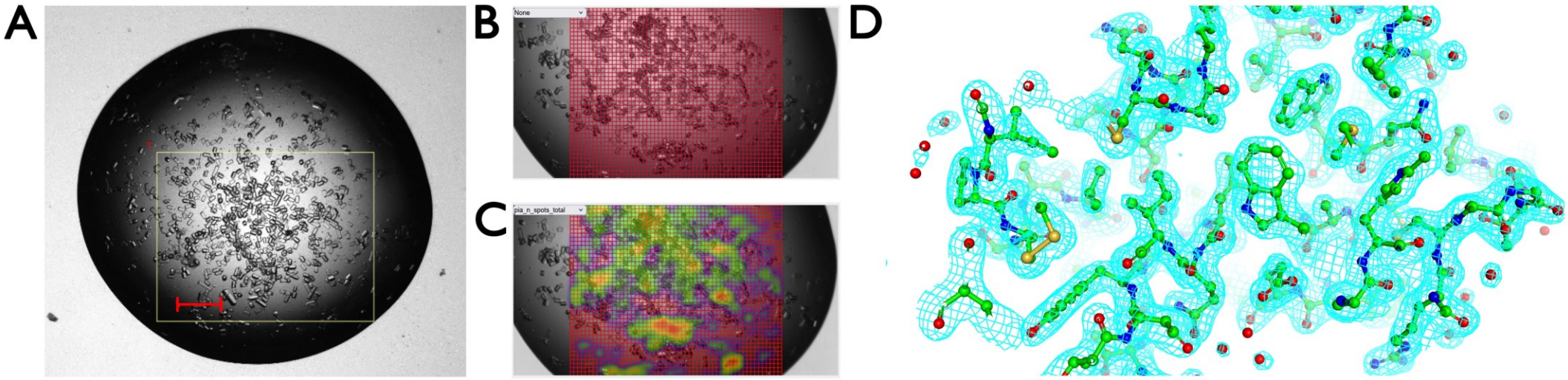
Using VMXi to screen for serial experiments

- Most common use-case of raster scanning (aka grid-scanning) -> identify crystalline material
- This technique is fundamentally a still-shot serial crystallography experiment – we can process these as serial data!



Using VMXi to screen for serial experiments

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

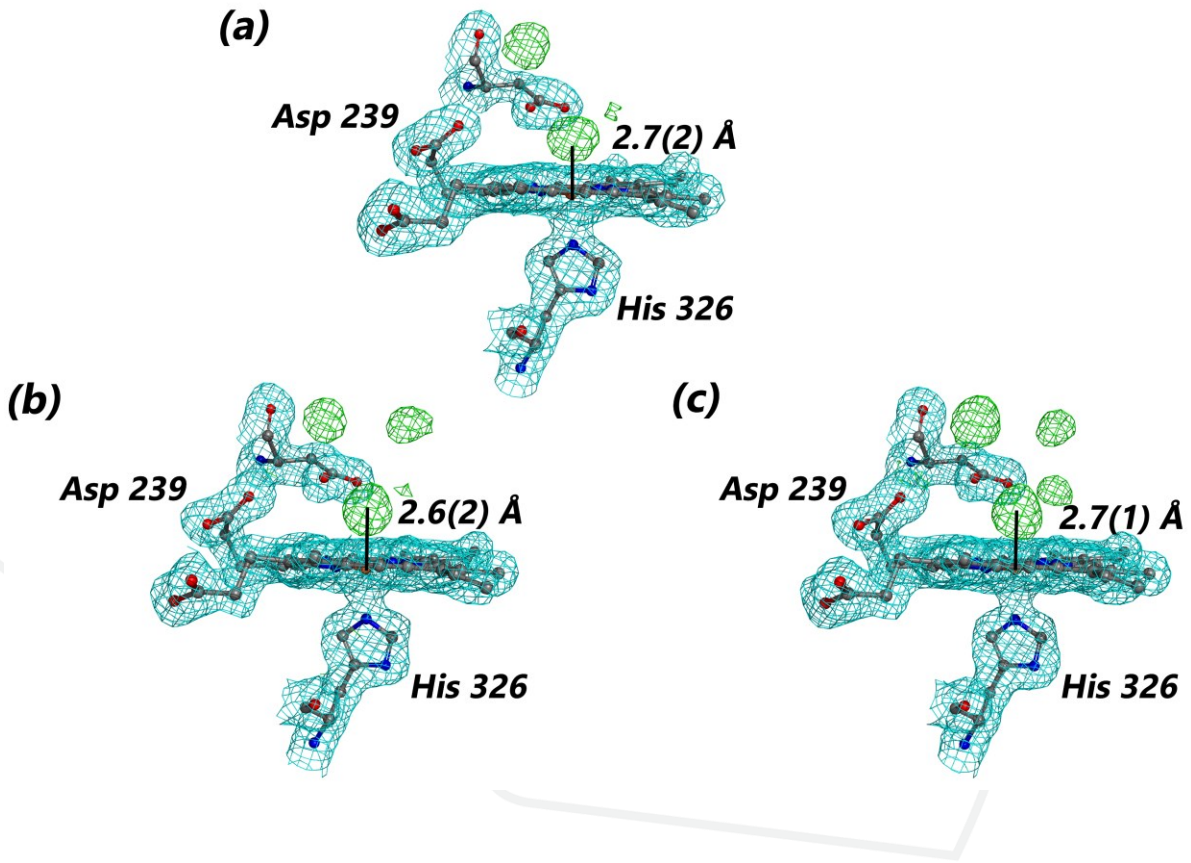


Statistics for eight drops merged together	Completeness (%)	Multiplicity	I/σ(I)	R_{split}	$CC_{1/2}$	Unique Observations	Crystals merged
Overall (55.56 – 1.88)	99.27	110.9	19.1	0.083	0.997	10076	9891
High Resolution (1.95 – 1.88)	99.90	65.1	1.2	1.143	0.398	973	

diamond

Using VMXi to screen for serial experiments

DtpAa: heme-containing dye-decolourising peroxidase (Space Group $P2_1$)

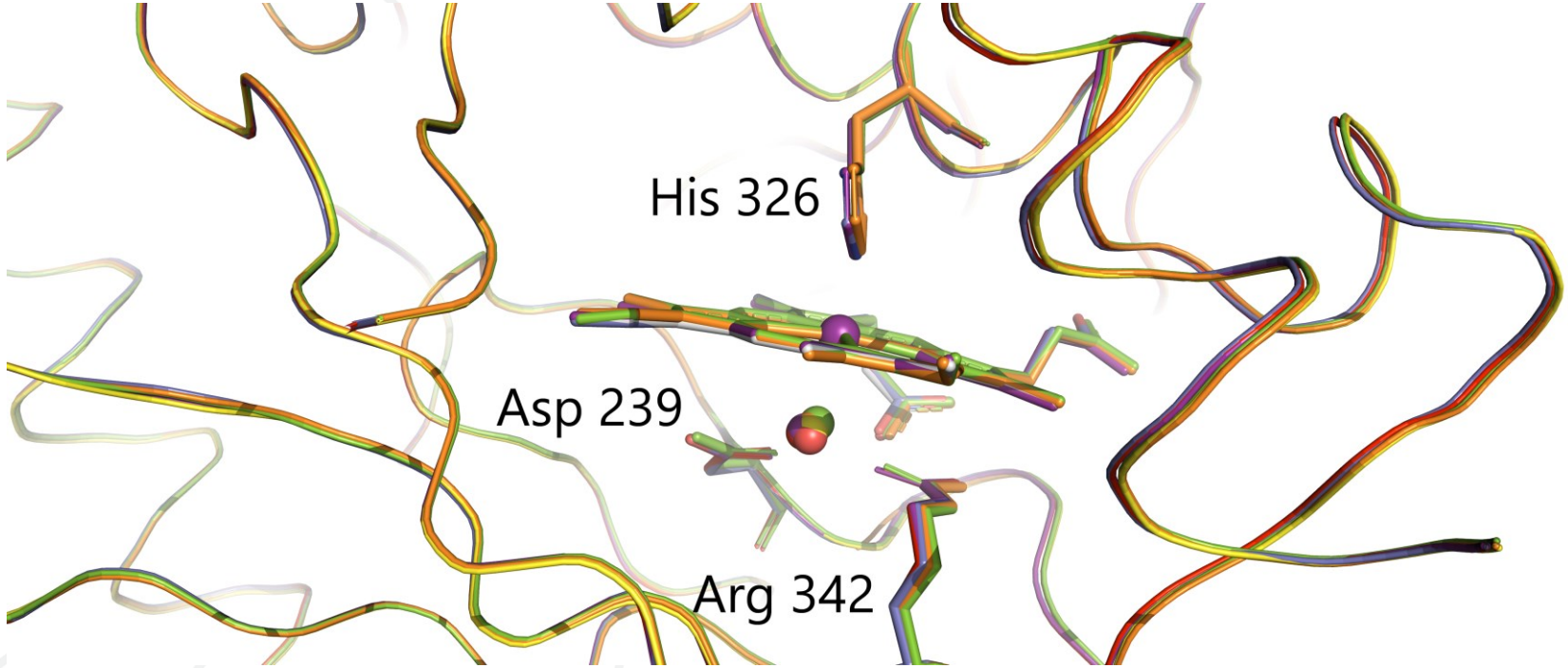


	(a)	*(b)*	*(c)*
	DtpAa (8 drops)	DtpAa (12 drops)	DtpAa (35 drops)
Resolution Range	69.76 – 2.07 (2.14 – 2.07)	69.73 – 1.88 (1.95 – 1.88)	69.77 – 1.79 (1.85 – 1.79)
Crystals Merged	5360	10054	22854
Volume Dispensed	0.8 µL	1.2 µL	3.5 µL
Unique Reflections	42538 (4203)	56729 (5631)	65821 (6543)
Multiplicity	28.7 (22.9)	48.0 (25.7)	98.9 (45.8)
Completeness	97.74 (84.49)	99.73 (99.36)	99.75 (99.69)
I/σ(I)	12.6 (2.0)	19.7 (2.2)	21.8 (2.0)
$CC_{1/2}$	0.927 (0.329)	0.958 (0.302)	0.990 (0.315)
R_{work}	0.2233 (0.3855)	0.2115 (0.2927)	0.1885 (0.3265)
R_{free}	0.2748 (0.4277)	0.2420 (0.3491)	0.2281 (0.3893)

- Lower symmetry necessitates more drops for high quality maps
- High quality data with small volumes possible

diamond

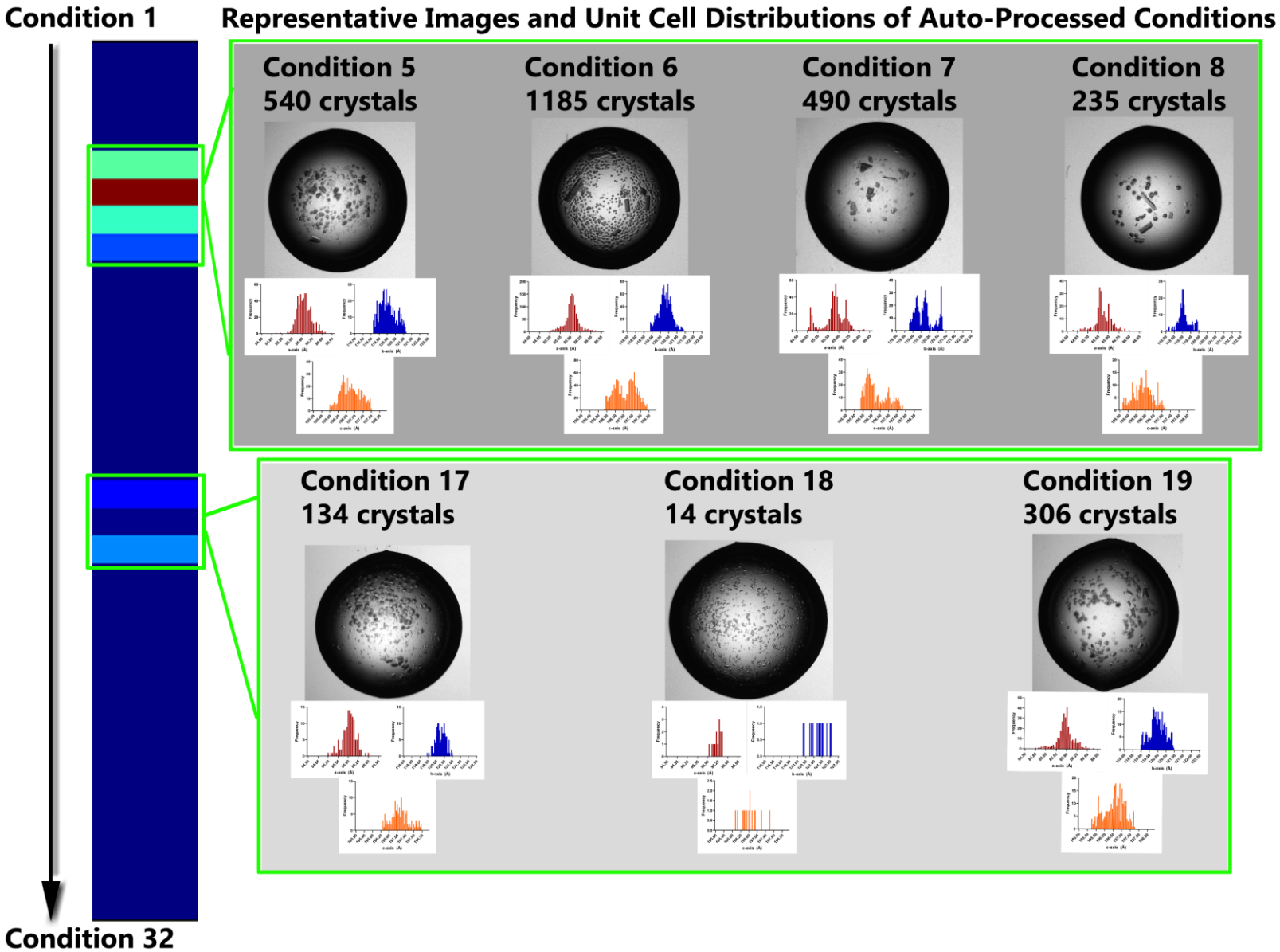
Using VMXi to screen for serial experiments



- Comparison to previously published XFEL and synchrotron serial structures revealed minimal differences
- Simple, in-plate serial collections are very viable for early stages of a serial project for initial structure determination

diamond

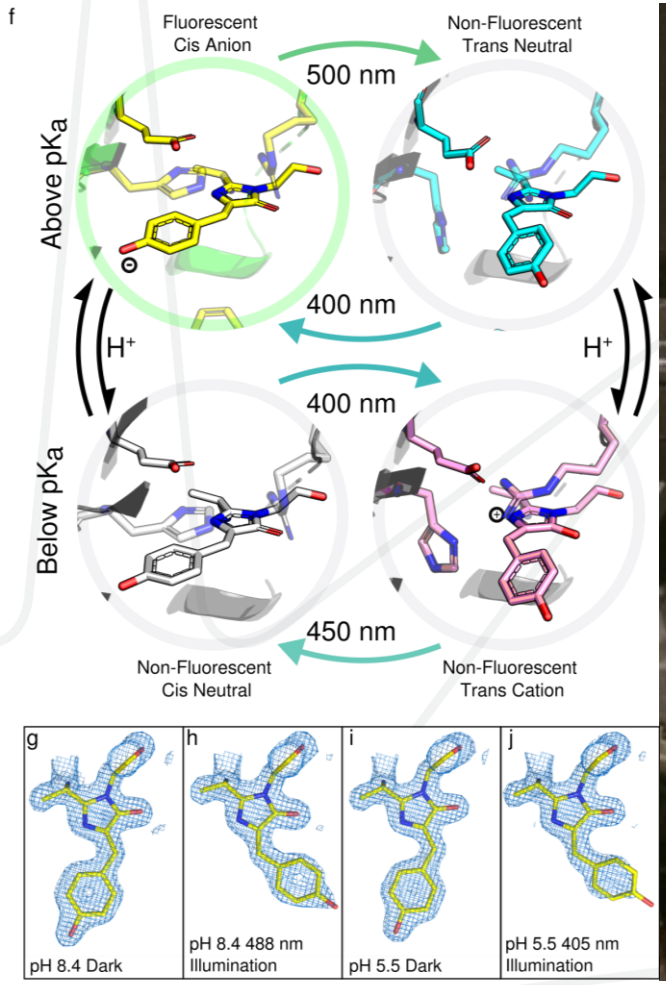
Using VMXi to screen for serial experiments

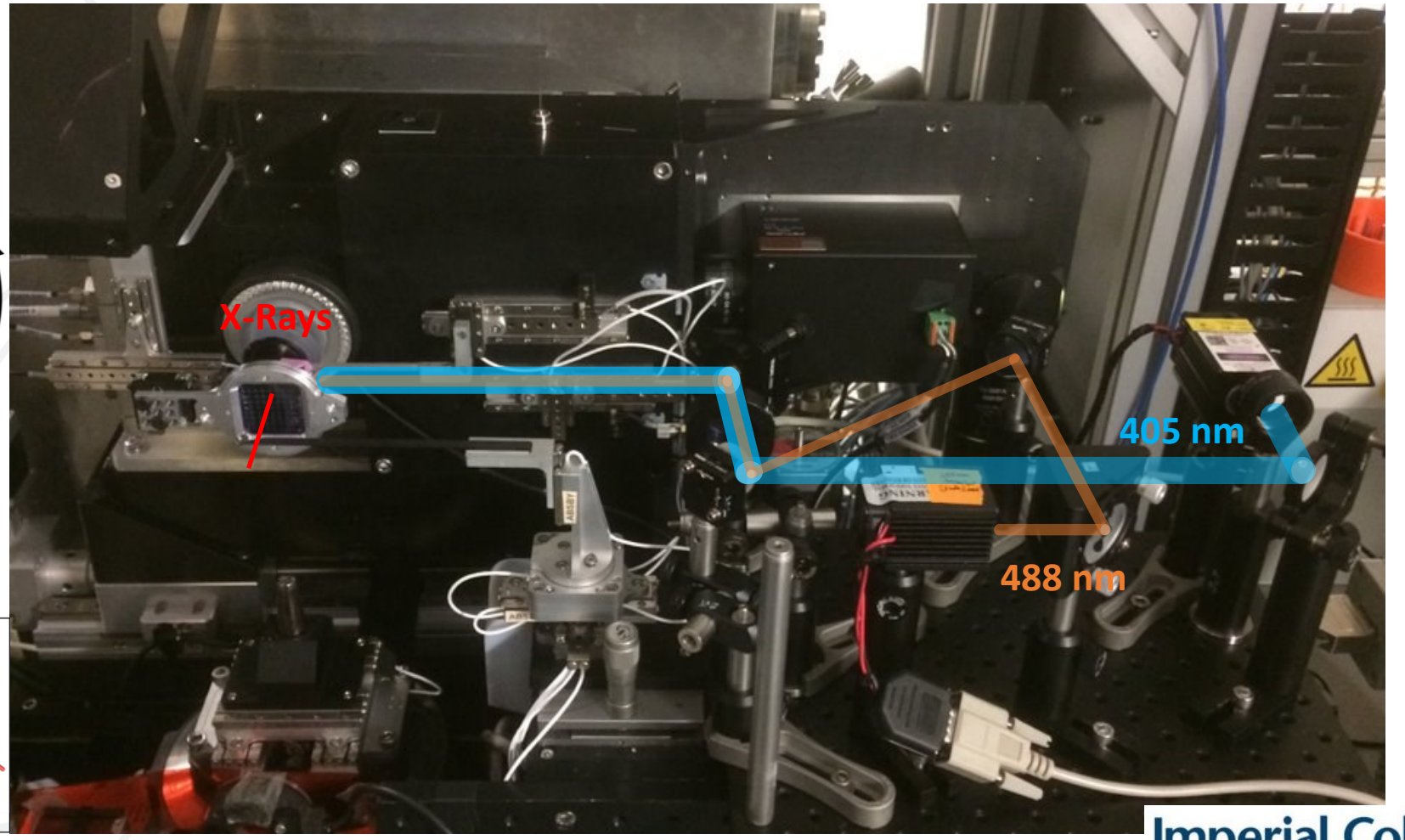


- Very quickly determine diffraction quality of micro-crystallisation slurries with minimal loss of sample
- Get insight into homogeneity, isomorphism, crystal size, etc

diamond

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

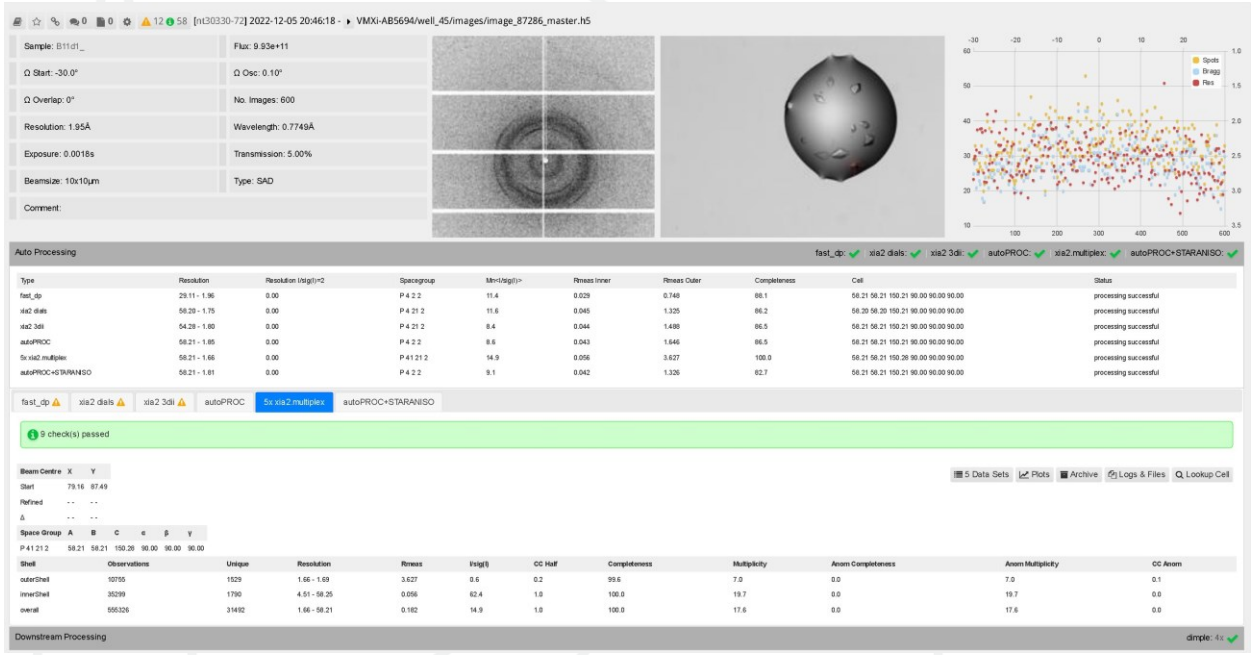


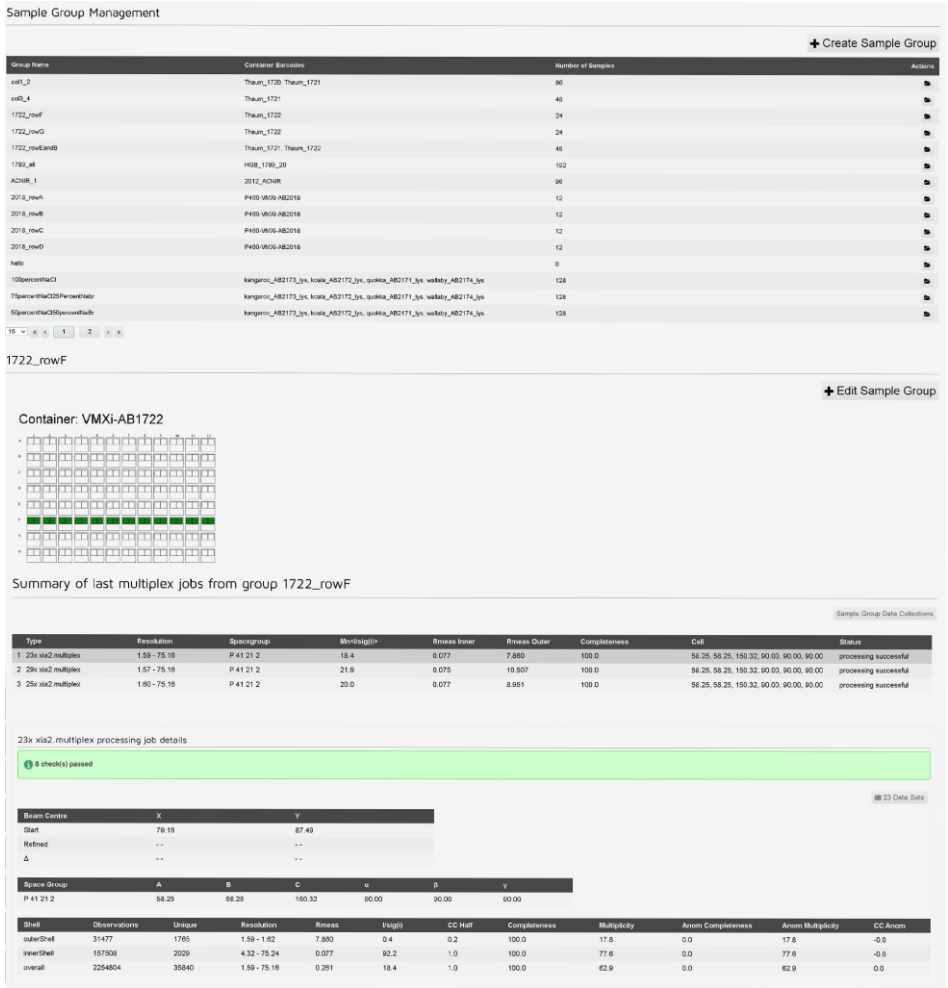


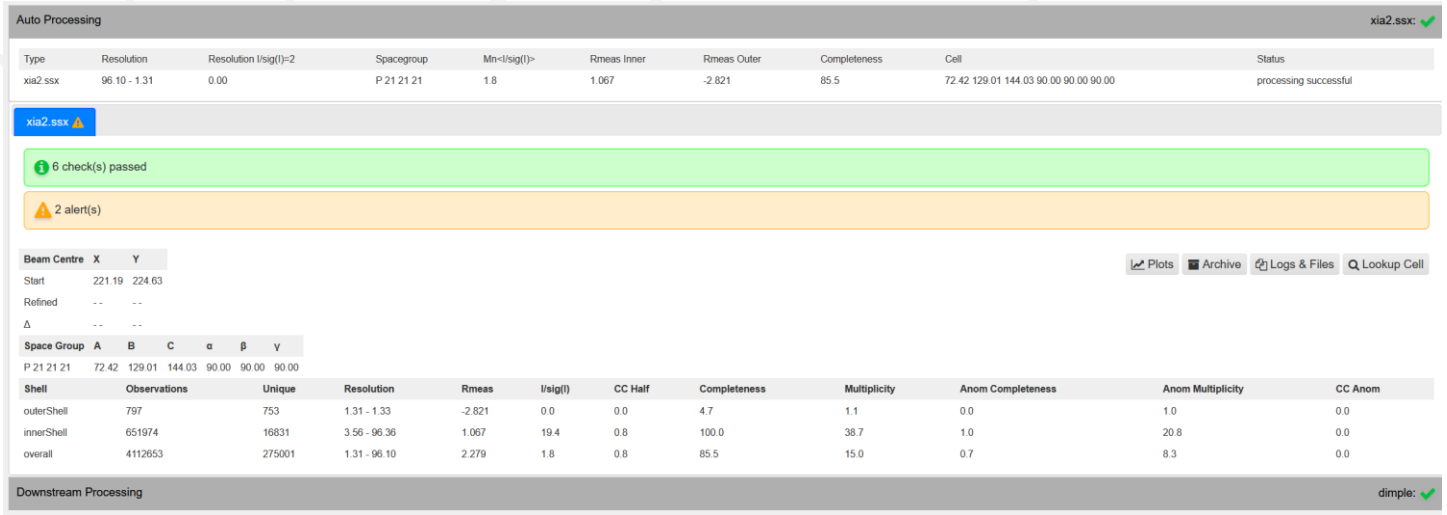
Baxter, J. M., Hutchison, C. D., Maghlaoui, K. et al. (2022) J. Phys. Chem. B, 126, 45, 9288–9296

Imperial College London

Availability through ISPyB at Diamond







Conclusions

Multi-Crystal Methods at Diamond:

- I24, VMXi and VMXm use multi-crystal strategies to achieve high-quality data from small, sensitive samples
- Both rotational and still-shot multi-crystal experiments are highly automated from data collection through to data processing

Using the new DIALS module dials.correlation_matrix, multi-crystal clustering methods have been improved by:

- Implementing a weighted algorithm to calculate pairwise correlation coefficients
- Optimising the number of dimensions calculated for the cosine-angle clustering
- Using the OPTICS algorithm for automated identification and output of significant clusters

diamond

Acknowledgements

VMXi:

- Michael Hough
- Juan Sanchez-Weatherby
- James Sandy
- Halina Mikolajek
- Megan Lambert
- Cicely Tam

Software Team:

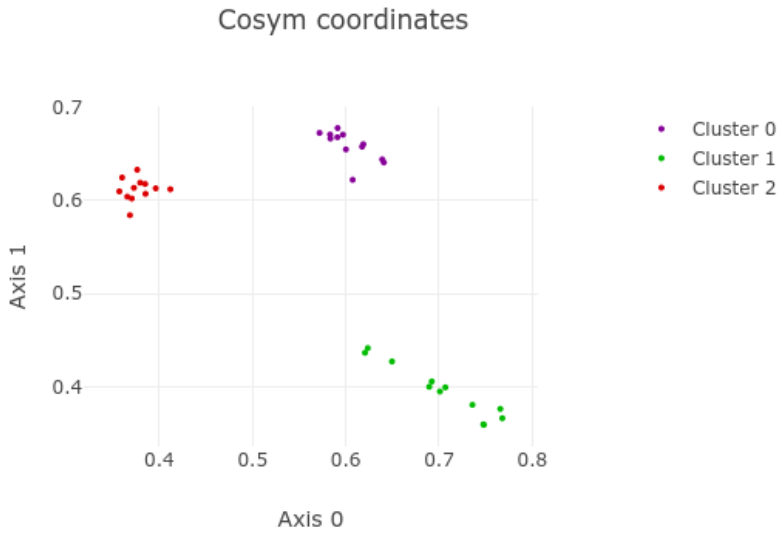
- Graeme Winter
- James Beilsten-Edmands

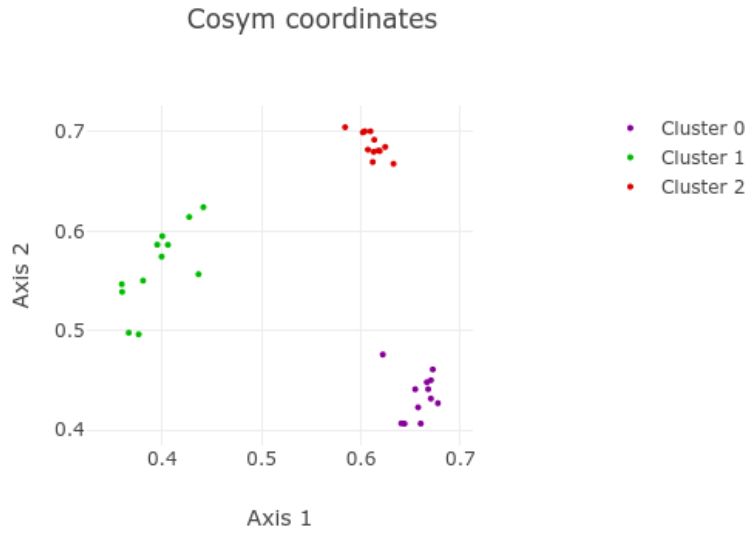
I24:

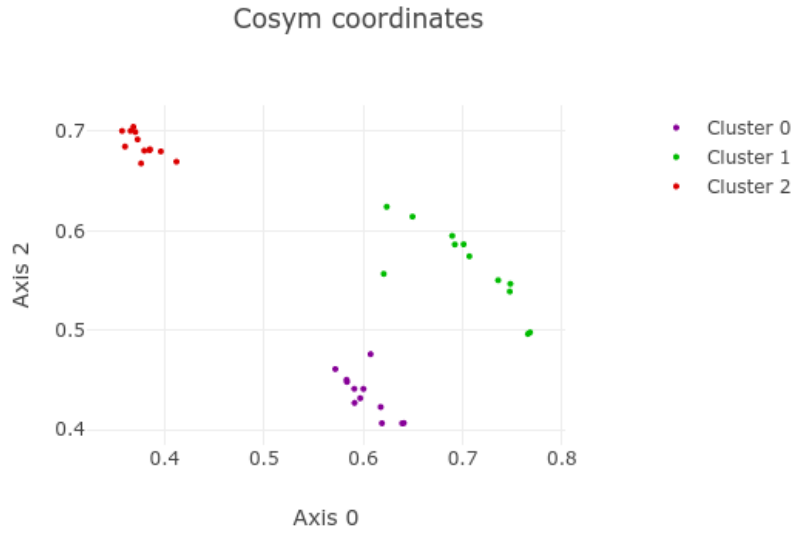
- Sofia Jaho
- Danny Axford
- Robin Owen

Questions???

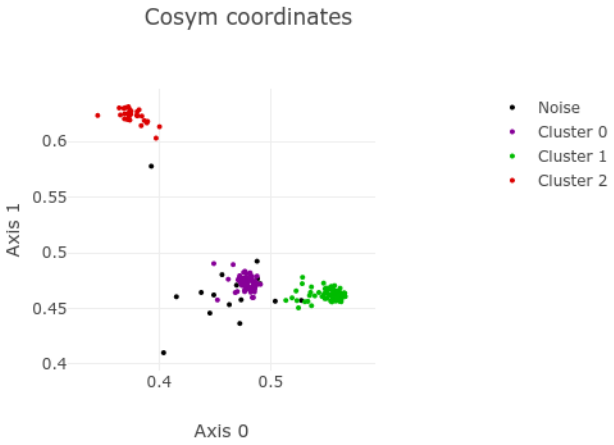
All Projections of I24 3D Dataset

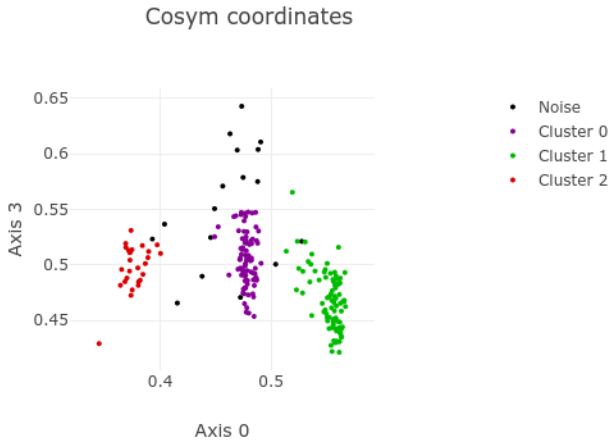


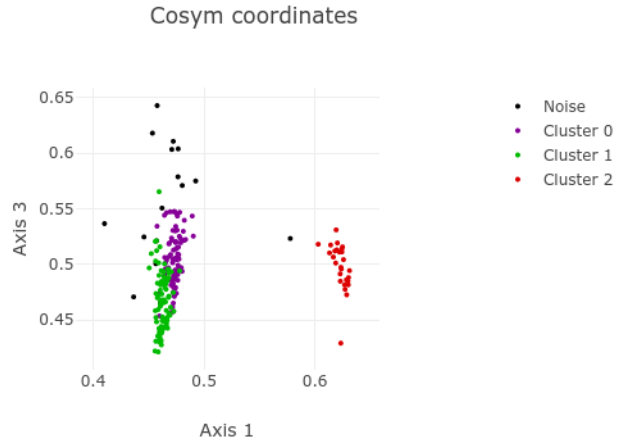


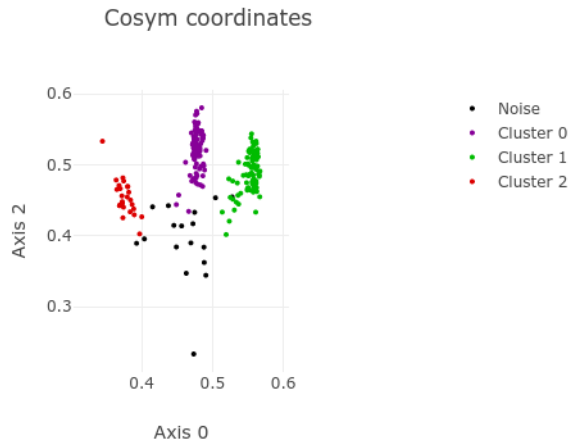


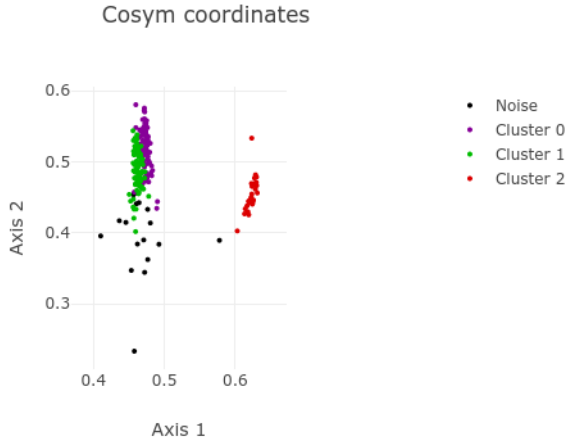
All Projections of VMXi 4D Dataset

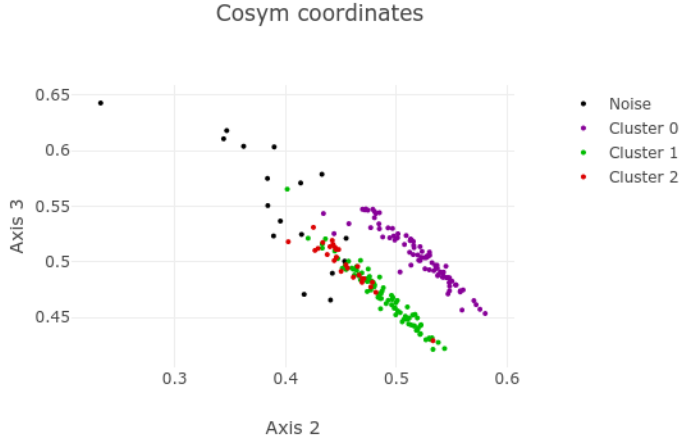






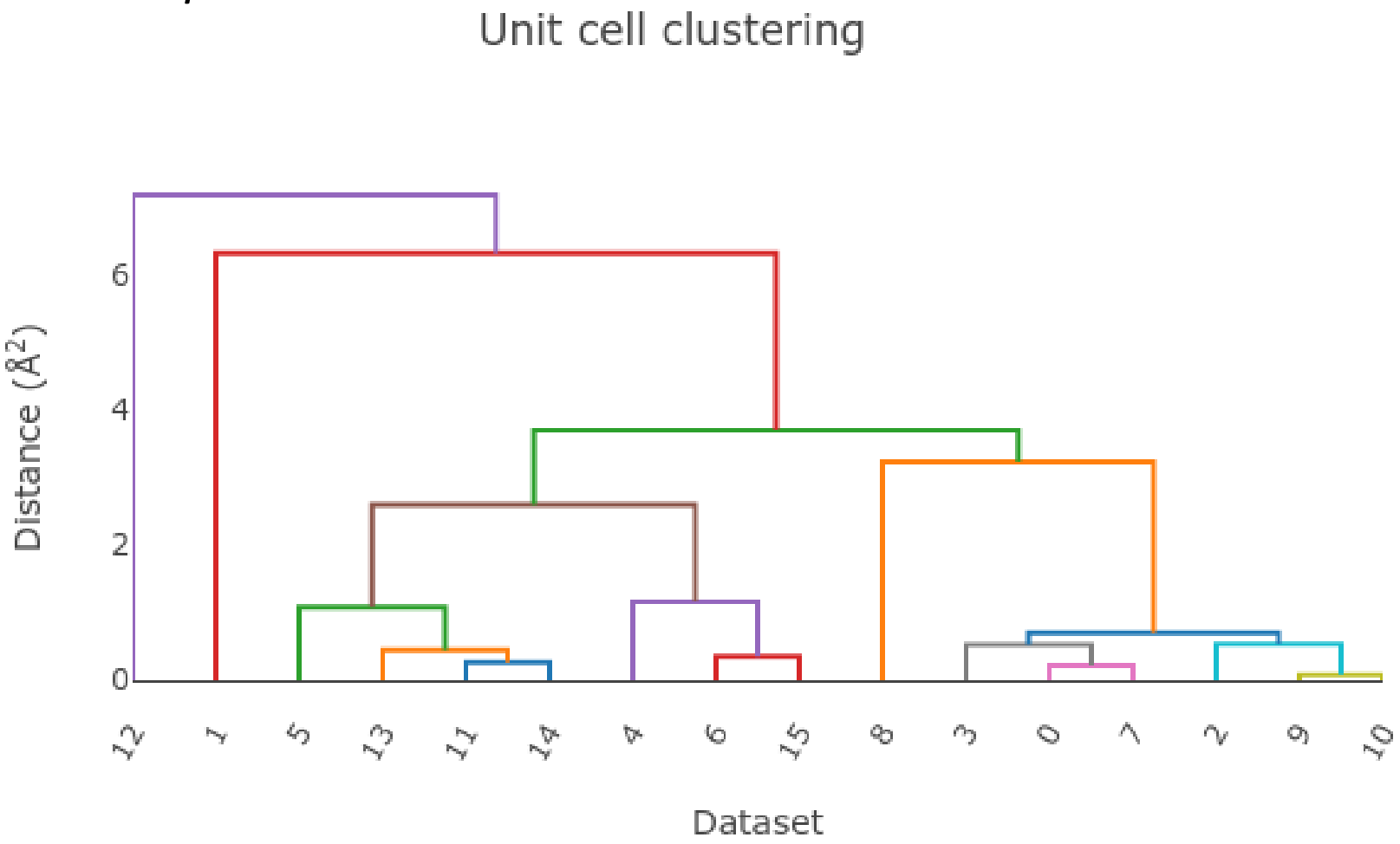






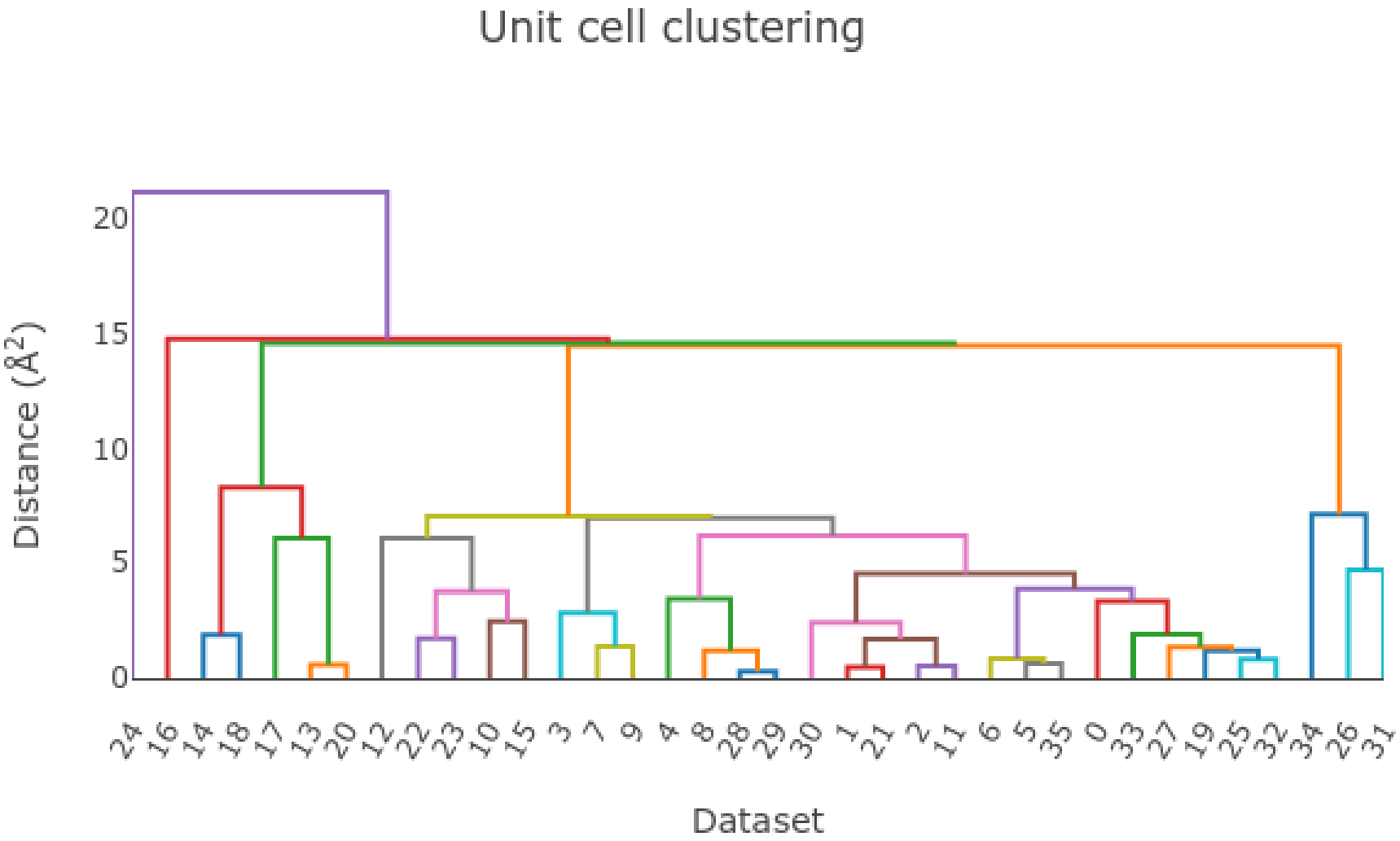
Unit Cell Clustering VMXi 2D Dataset

0 -> 7 Cows, 8 -> 15 People



Unit Cell Clustering I24 3D Dataset

0 -> 11 Cows, 12 -> 23 Pigs, 24 -> 35 People



Unit Cell Clustering VMXi 4D Dataset