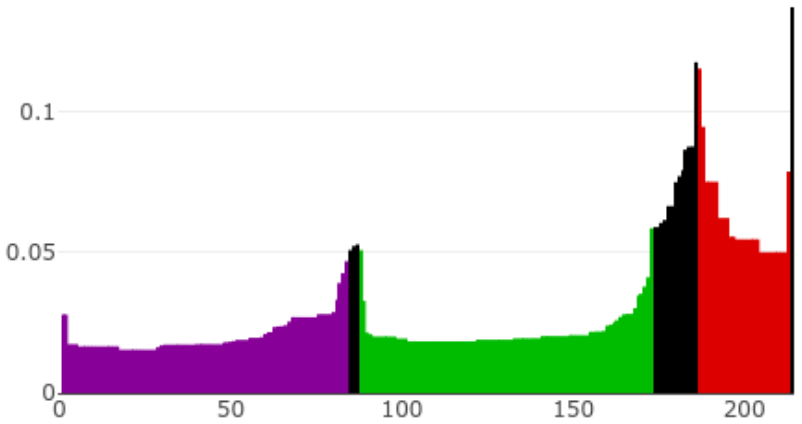


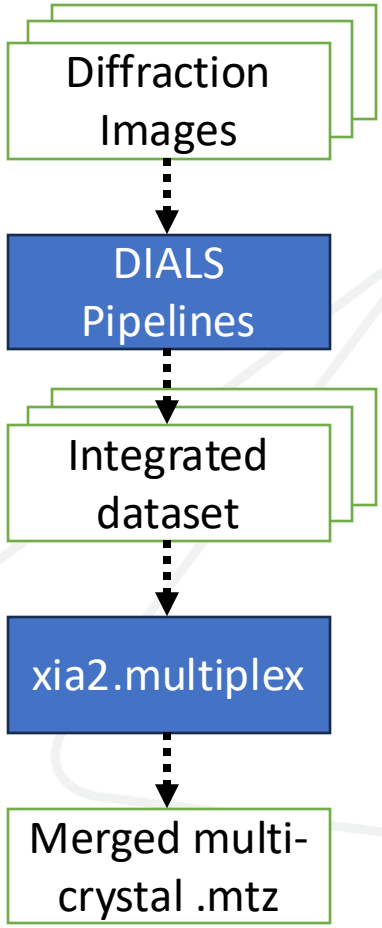
Multi-dataset DIALS / xia2.multiplex

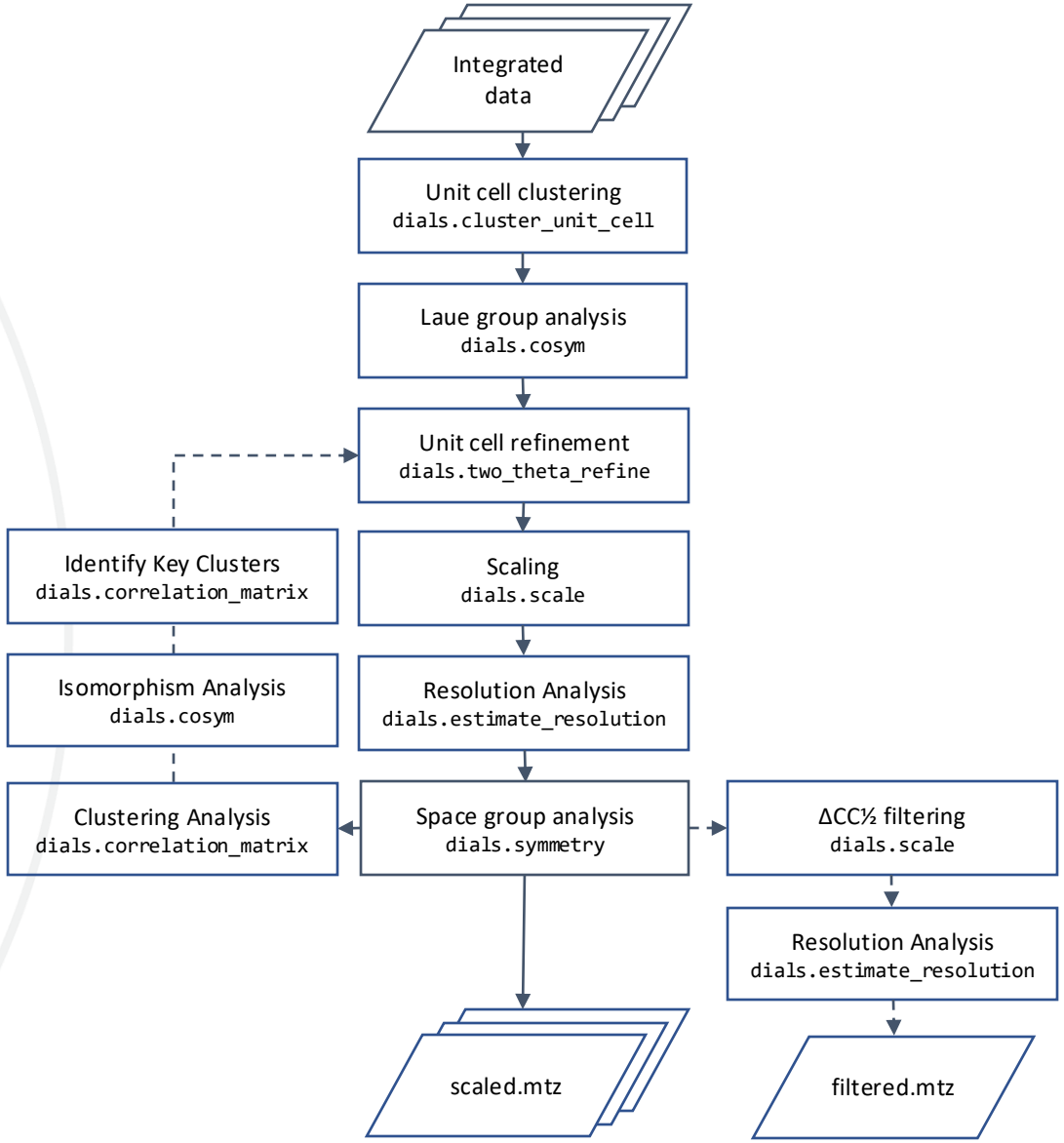
Dr Amy Thompson

PDRA – VMXi Beamline – Diamond Light Source

Processing Multi-Crystal Data with xia2.Multiplex

Auto-processing Pipeline





diamond

Clustering Methods to Achieve Isomorphous Data

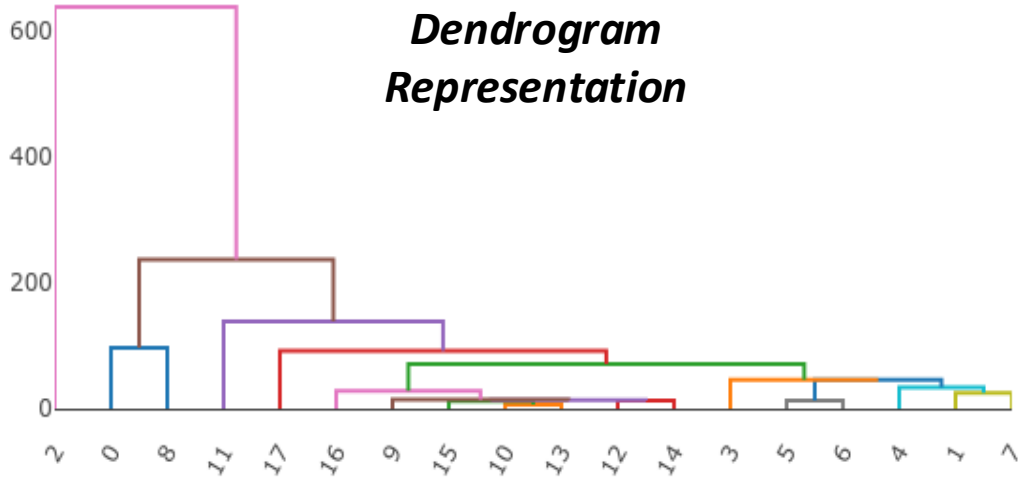
Achieving isomorphous data from multi-crystal collections is critical, although differences can still exist within structurally isomorphous crystals (ligand binding, alternative confirmations, etc).
First requires a known and consistent symmetry (methods from Brehm and Diederichs)

Unit Cell Clustering

- Good to detect structural non-isomorphism
- Clusters on unit cell dimensions

Dendrogram Representation

Height (proportional to difference)



Dataset Number

Intensity-based Clustering

- Good to detect differences in structurally isomorphous datasets
- Clusters on pair-wise correlation coefficients

Brehm, W., Diederichs (2014). *Acta. Cryst., D70, 101-109.*

Hierarchical Clustering in xia2.multiplex

After consistent symmetry determination, there are three types of clustering available:

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

Correlation Clustering

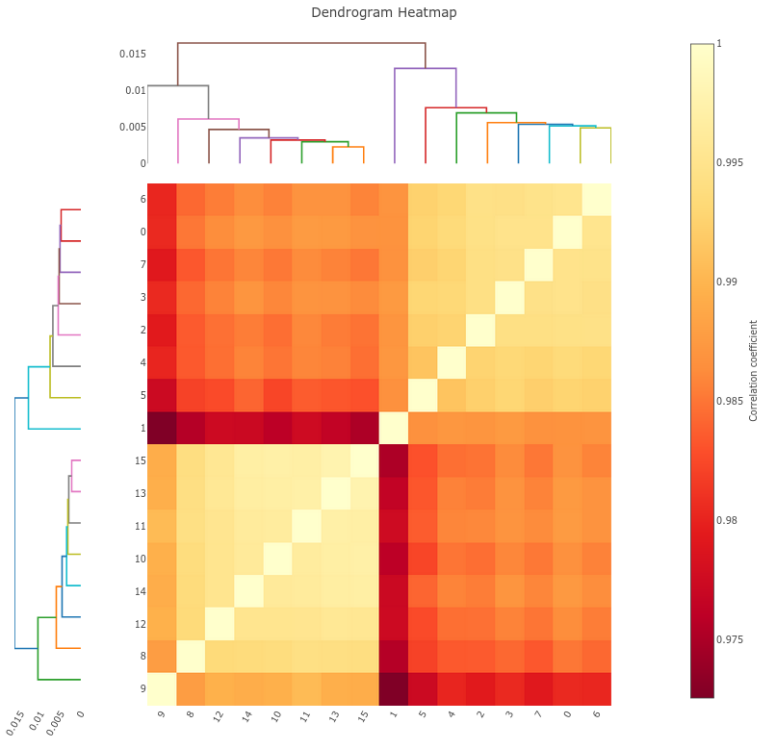
- Comparison of pairwise correlation coefficients
- Does not distinguish types of error resulting in differences in intensities

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

Cosine Angle Clustering

- Extension of correlation clustering
- Separates systematic and random error
- Same methods as indexing ambiguity resolution (dials.cosym)

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



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

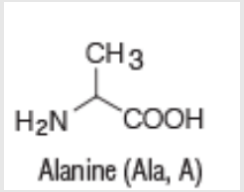
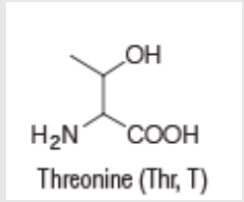
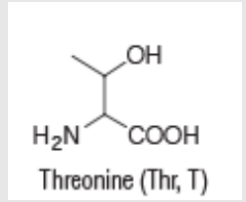
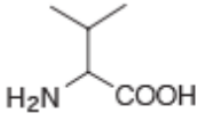
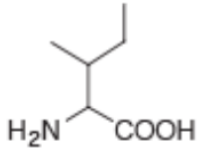
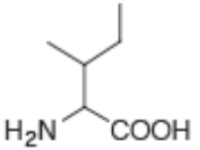
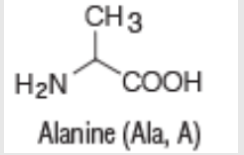
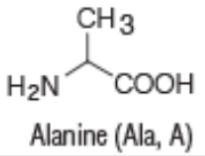
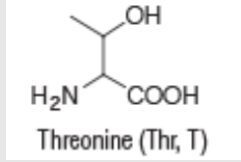
Diederichs, K. (2017). *Acta. Cryst., D73, 286-293.*

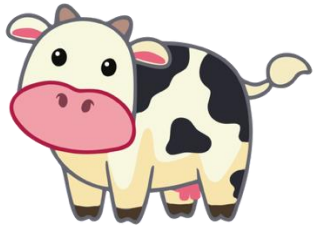
Gildea, R. J., Winter, G. (2018). *Acta. Cryst., D74, 405-410.*

diamond

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

	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)



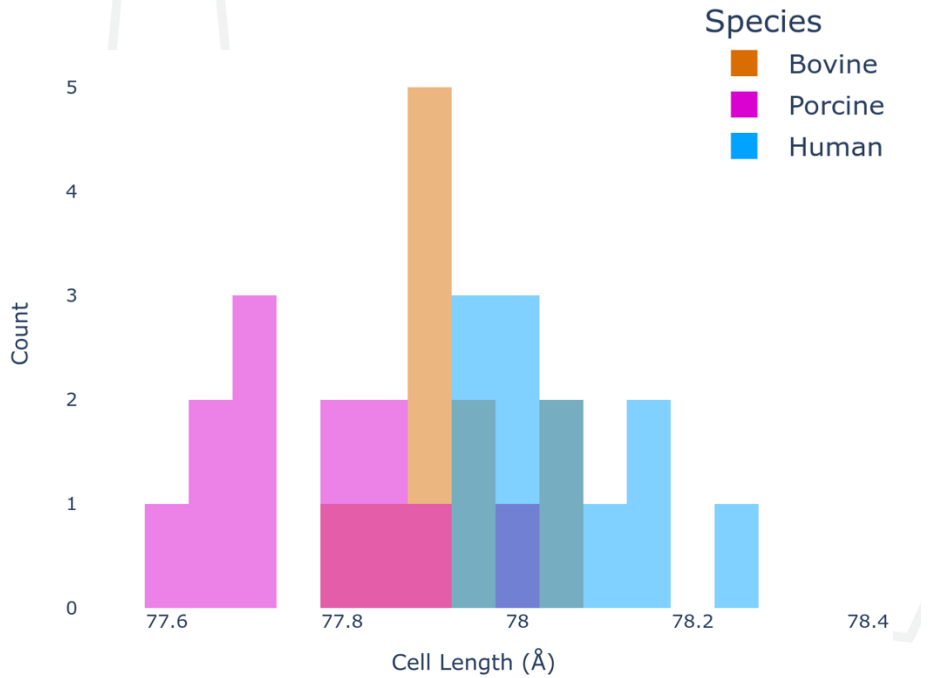
diamond

Thompson, A. J. et al. (2025) Acta Cryst. D81, 278-290.

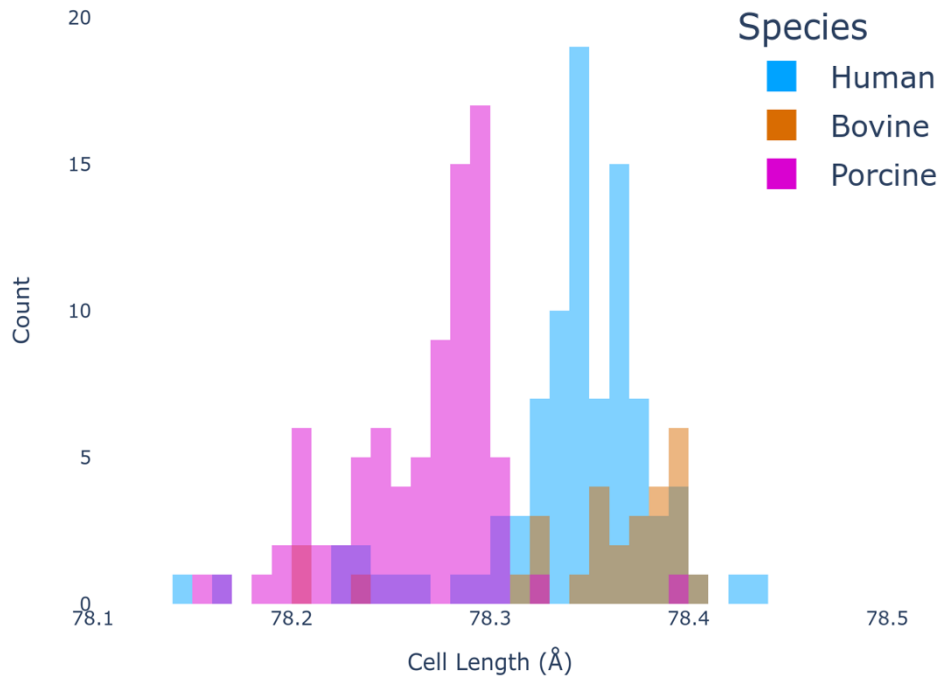
Introducing a Test Case… Cows, Pigs and People

Isomorphous nature of these samples evident at both cryogenic and room temperature

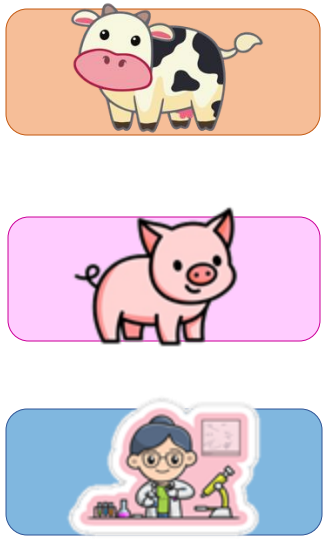
- While subtle differences in unit cell are evident, natural spread across many crystals provides overlap, making them impossible to distinguish with unit cell-based clustering



100 K



293 K



Thompson, A. J. et al. (2025) Acta Cryst. D81, 278-290.