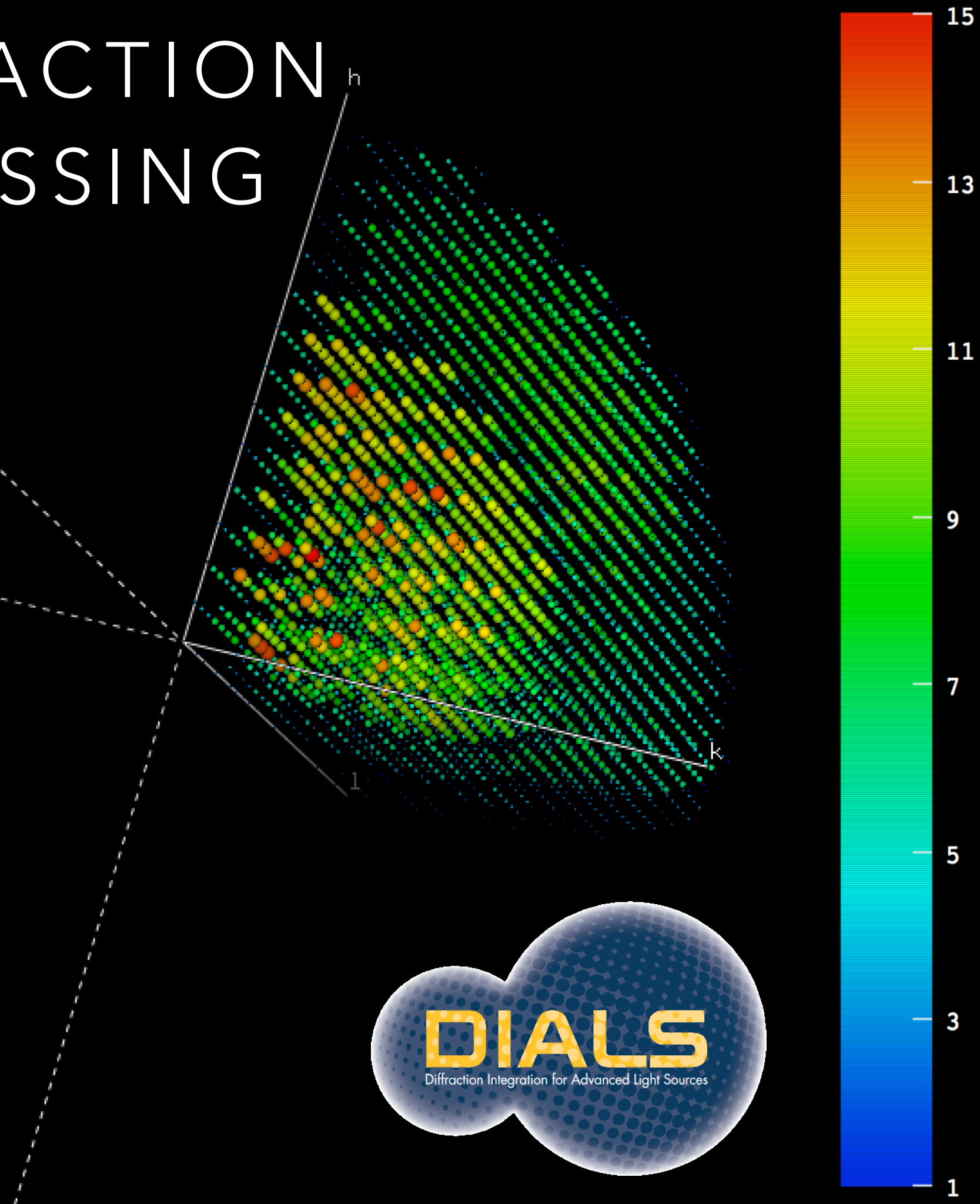


X-RAY DIFFRACTION DATA PROCESSING WITH DIALS & XIA2

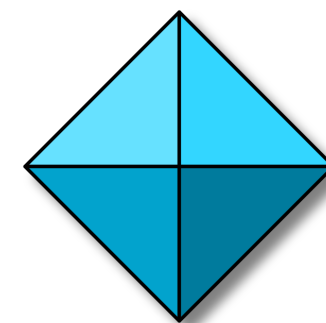
DLS / CCP4 WORKSHOP 2020
GRAEME WINTER
DIAMOND LIGHT SOURCE



PROJECT SUPPORT



Science & Technology
Facilities Council



CCP4

EU FP7: #283570 NIH: GM095887 & GM102520

ACKNOWLEDGEMENTS

research papers

Acta Crystallographica Section D
**Biological
Crystallography**
ISSN 0907-4449

XDS

Wolfgang Kabsch

Max-Planck-Institut für Medizinische Forschung,
Abteilung Biophysik, Jahnstrasse 29,
69120 Heidelberg, Germany

Correspondence e-mail:
wolfgang.kabsch@mpimf-heidelberg.mpg.de

The usage and control of recent modifications of the program package *XDS* for the processing of rotation images are described in the context of previous versions. New features include automatic determination of spot size and reflecting range and recognition and assignment of crystal symmetry. Moreover, the limitations of earlier package versions on the number of correction/scaling factors and the representation of pixel contents have been removed. Large program parts have been restructured for parallel processing so that the quality and completeness of collected data can be assessed soon after measurement.

Received 19 August 2009
Accepted 9 November 2009

A version of this paper will be published as a chapter in the new edition of Volume F of *International Tables for Crystallography*.

1. Functional specification

research papers

Acta Crystallographica Section D
**Biological
Crystallography**
ISSN 0907-4449

The integration of macromolecular diffraction data

Andrew G. W. Leslie

MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH, England

Correspondence e-mail:
andrew@mrc-lmb.cam.ac.uk

The objective of any modern data-processing program is to produce from a set of diffraction images a set of indices (*hkl*s) with their associated intensities (and estimates of their uncertainties), together with an accurate estimate of the crystal unit-cell parameters. This procedure should not only be reliable, but should involve an absolute minimum of user intervention. The process can be conveniently divided into three stages. The first (autoindexing) determines the unit-cell parameters and the orientation of the crystal. The unit-cell

Received 19 May 2005
Accepted 24 November 2005

Centre National de la Recherche Scientifique
Université Paris-Sud

research papers

Acta Crystallographica Section D
**Biological
Crystallography**
ISSN 0907-4449

The finer things in X-ray diffraction data collection

J. W. Pflugrath

Molecular Structure Corporation, 9009 New
Trails Drive, The Woodlands, TX 77381, USA

Correspondence e-mail: jwp@msc.com

X-ray diffraction images from two-dimensional position-sensitive detectors can be characterized as thick or thin, depending on whether the rotation-angle increment per image is greater than or less than the crystal mosaicity, respectively. The expectations and consequences of the processing of thick and thin images in terms of spatial overlap, saturated pixels, *X*-ray background and *I*₀(*θ*) are discussed. The *STAR*

Received 6 May 1999
Accepted 5 July 1999

computer programs

Journal of
**Applied
Crystallography**
ISSN 0021-8898

xia2: an expert system for macromolecular crystallography data reduction

G. Winter

Diamond Light Source, Harwell Science and Innovation Campus, Oxfordshire, UK, STFC Daresbury Laboratory, Warrington, Cheshire, UK, and University of Manchester, Manchester, UK

An expert system for macromolecular crystallography data reduction is presented, which builds on existing software to automate the complete data reduction process from images to merged structure factor amplitudes. This can automatically identify multi-wedge, multi-pass and multiwavelength data sets and includes explicit procedures to test for crystallographic special cases. With



research papers

Acta Crystallographica Section D
**Biological
Crystallography**
ISSN 0907-4449

How good are my data and what is the resolution?

Philip R. Evans* and Garib N. Murshudov

MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 0QH, England

Following integration of the observed diffraction spots, the process of 'data reduction' initially aims to determine the point-group symmetry of the data and the likely space group. This can be performed with the program *POINTLESS*. The scaling program then puts all the measurements on a common scale, averages measurements of symmetry-related reflections (using the symmetry determined previously) and produces

Received 7 September 2012
Accepted 2 January 2013



SOLUTION TO MOST PROCESSING PROBLEMS - COLLECT BETTER DATA!

- With a pixel array detector (pilatus, eiger) no readout noise so be sparing with photons (Andreas F talk)
- Collect more, weaker data in preference - lousy R_{merge} but great R_{pim}
- Mount your crystals carefully - make sure most of the atoms in the beam are your crystal not loop, drop, ... (Elspeth G talk)
- Cryo-cool your samples carefully (also Elspeth's talk)

RUNNING DIALS

- Tutorials will have running individual DIALS steps because this is good to know
- In reality I run DIALS with xia2
- You can also use DUI for interactive processing

DATA INTEGRATION AND SCALING (IN SIMPLE TERMS)

FOR THE TUTORIAL WE HAVE IMAGES LIKE THIS

DIFFRACTION SPOTS...

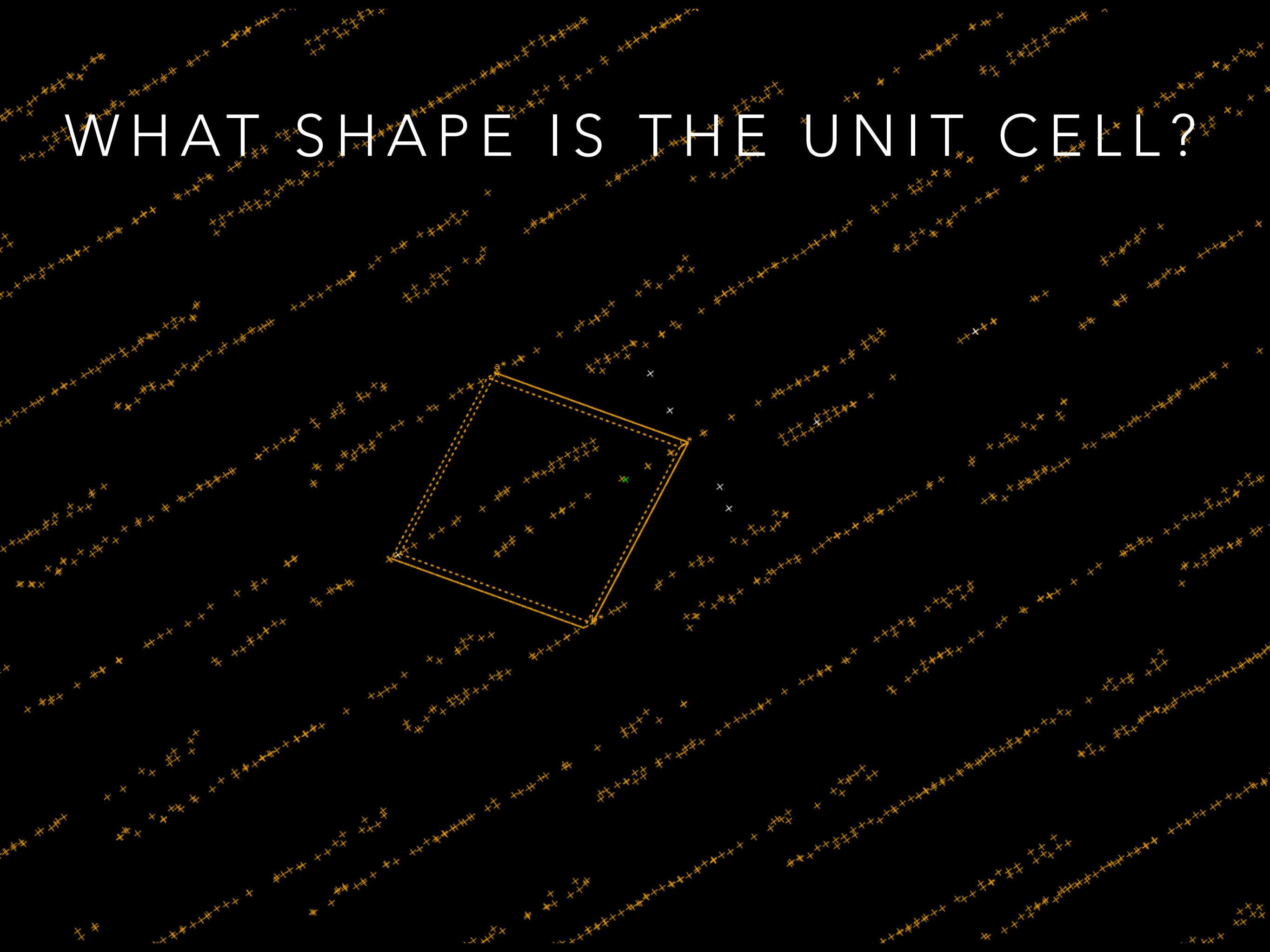


NAME THAT SPOT (MILLER INDICES)

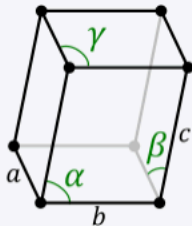
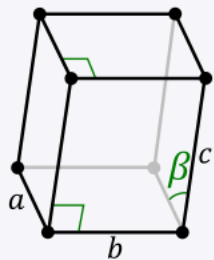
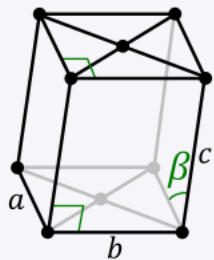
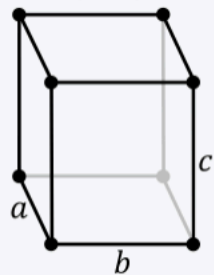
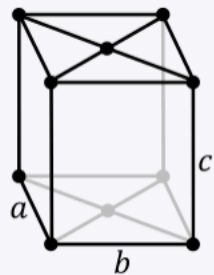
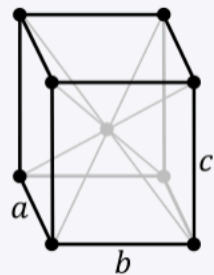
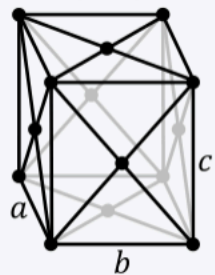
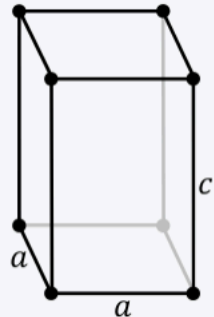
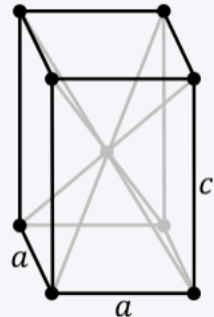
WHICH SPOT IS WHICH?



WHAT SHAPE IS THE UNIT CELL?



WHAT IS THE CORRECT BRAVAIS LATTICE?

Crystal family	Lattice system	Schönflies	14 Bravais Lattices			
			Primitive	Base-centered	Body-centered	Face-centered
triclinic		C_i				
monoclinic		C_{2h}	$\beta \neq 90^\circ$ $a \neq c$ 	$\beta \neq 90^\circ$ $a \neq c$ 		
orthorhombic		D_{2h}	$a \neq b \neq c$ 	$a \neq b \neq c$ 	$a \neq b \neq c$ 	$a \neq b \neq c$ 
tetragonal		D_{4h}	$a \neq c$ 		$a \neq c$ 	

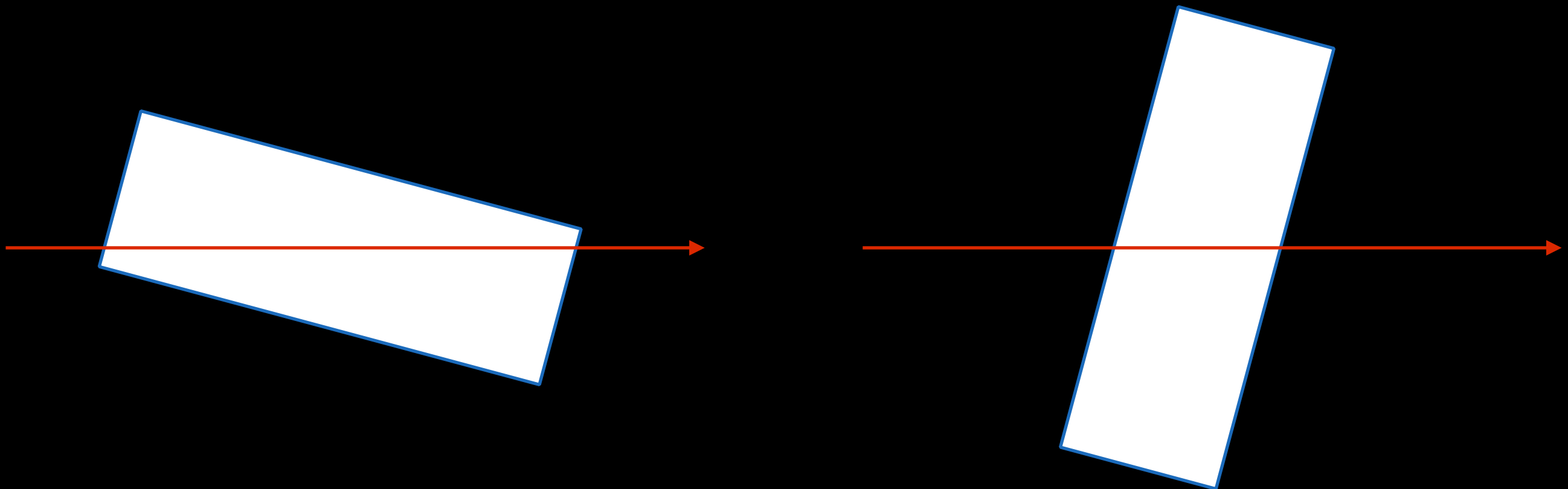
DATA INTEGRATION

HOW MANY COUNTS?



CORRECTIONS FOR EXPERIMENTAL EFFECTS

PLACE ON COMMON SCALE



TERMINOLOGY (UNIVERSAL)

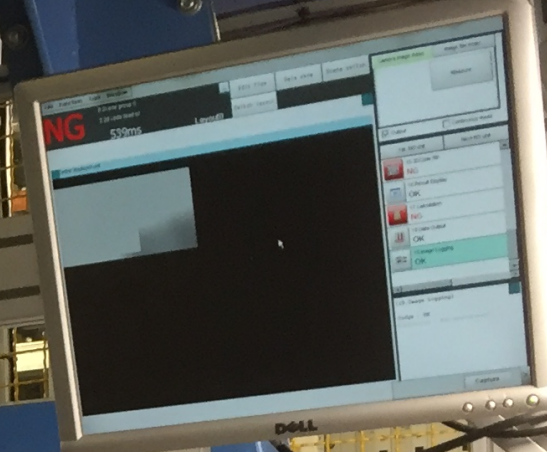
- Spot finding: locate spots on some or all of the data
- Indexing: looking at the positions of the spots and calculating a model which explains those positions (may involve choices about lattice or space group, *optional*)
- Refinement: improving the model
- Integration: record how many counts in each spot, subtract background, usually fit reflection profiles
- Symmetry determination: determine which intensities are related to one another by symmetry which is critical for ...
- Scaling: correction for experimental effects i.e. beam intensity, illuminated volume, sample absorption...

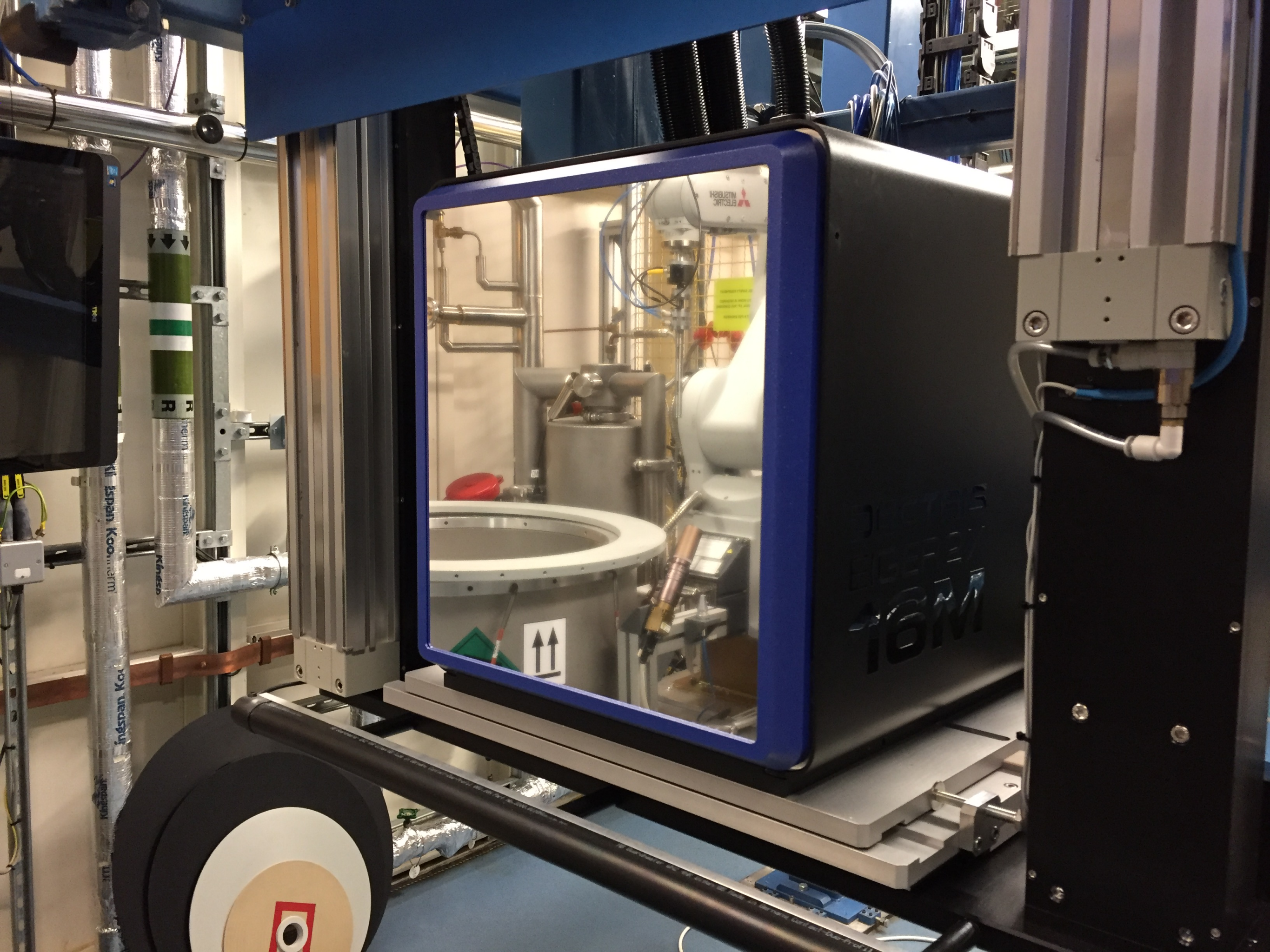
EXPERIMENT

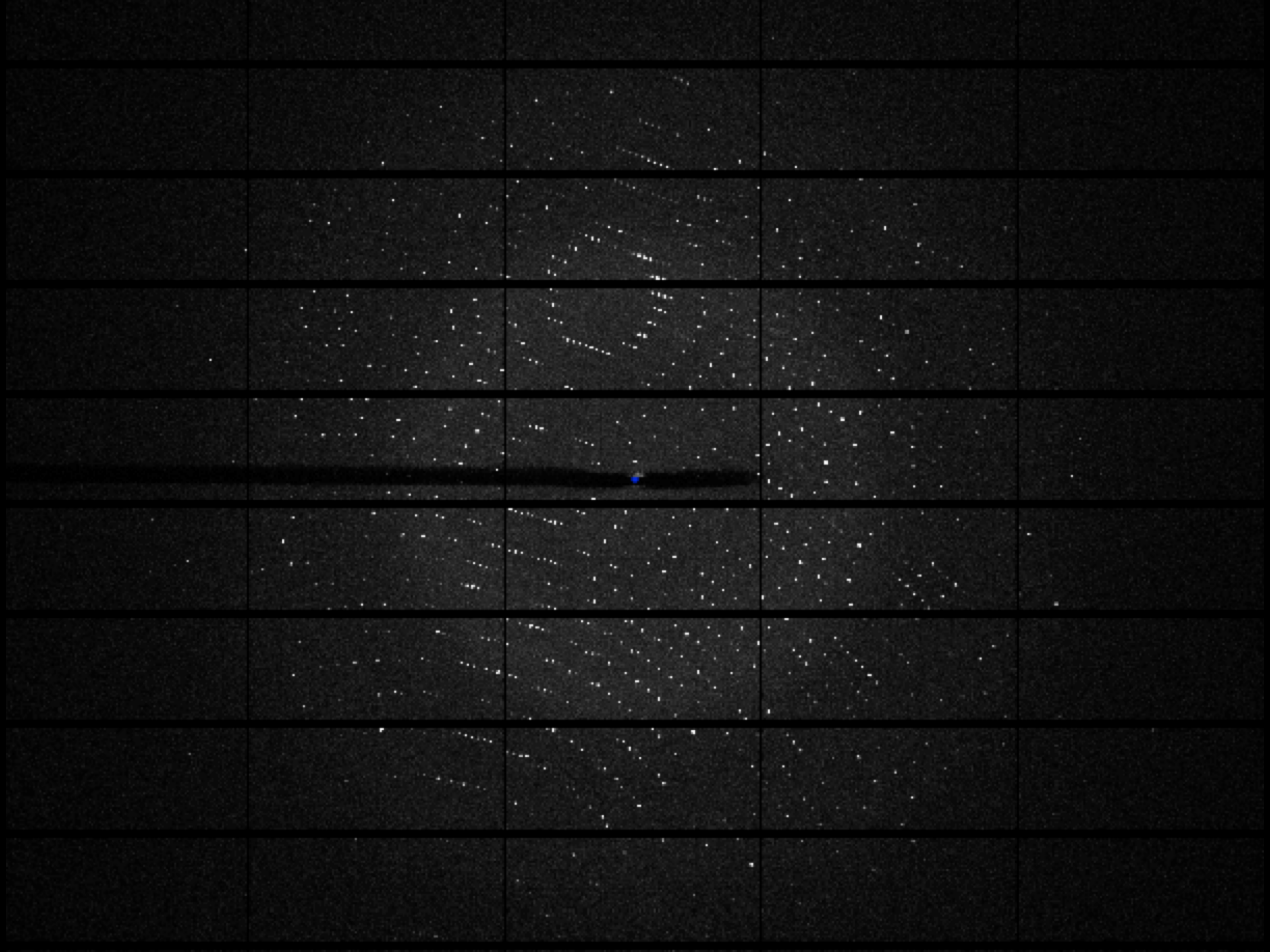
CAUTION
TRAP

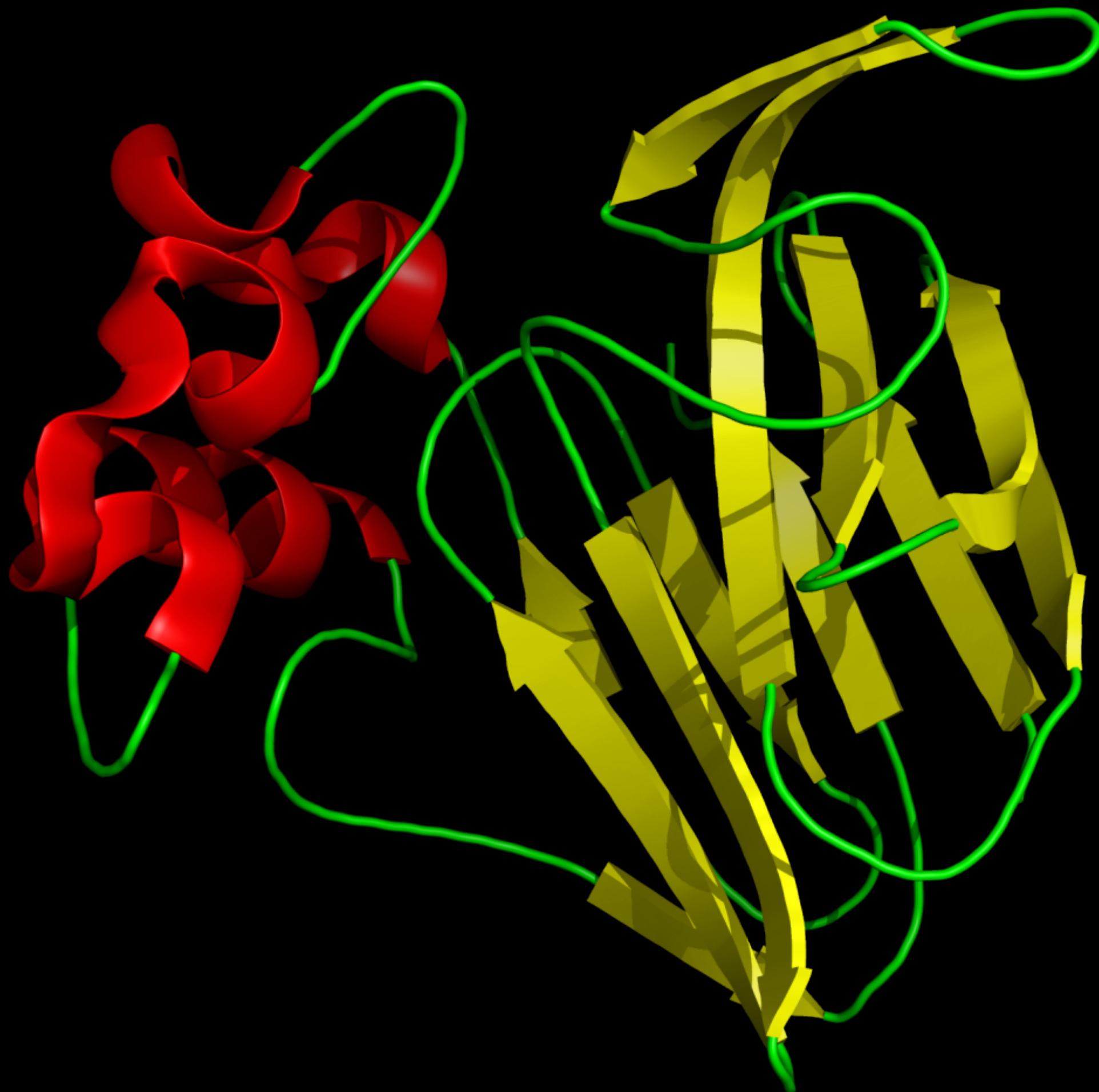
CAUTION
TRAP
HAZARD

NODE 31





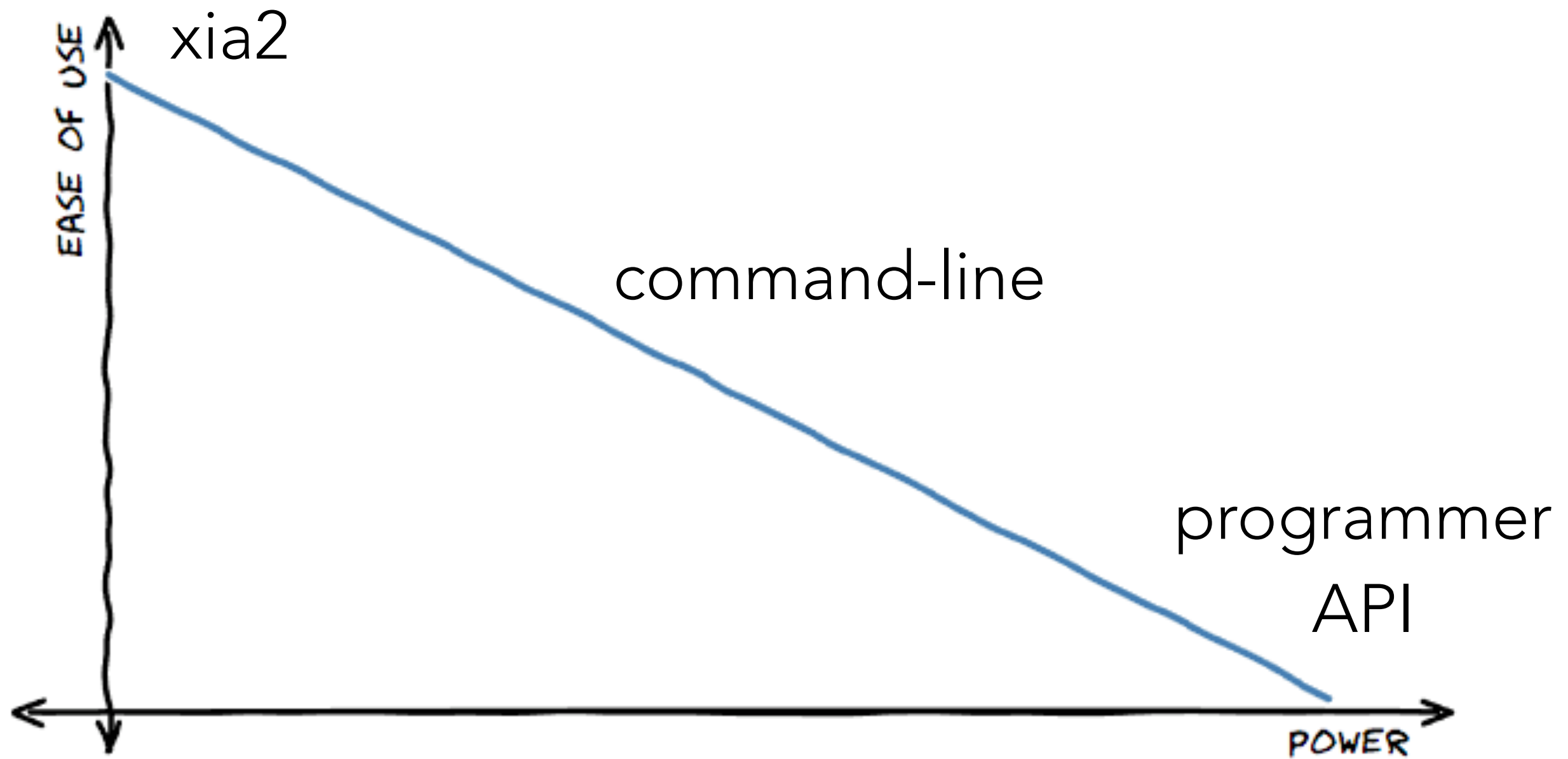


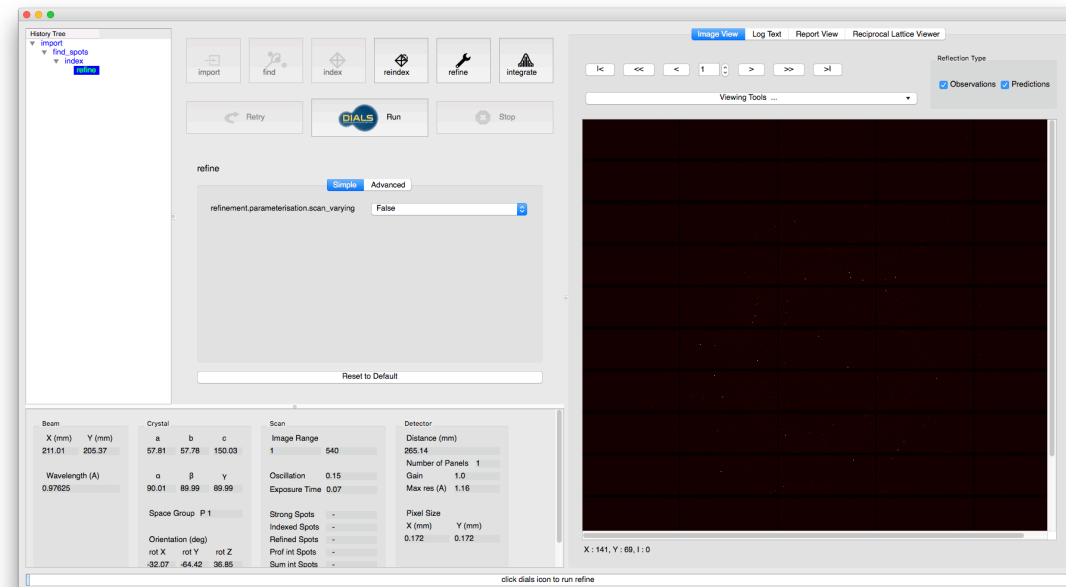
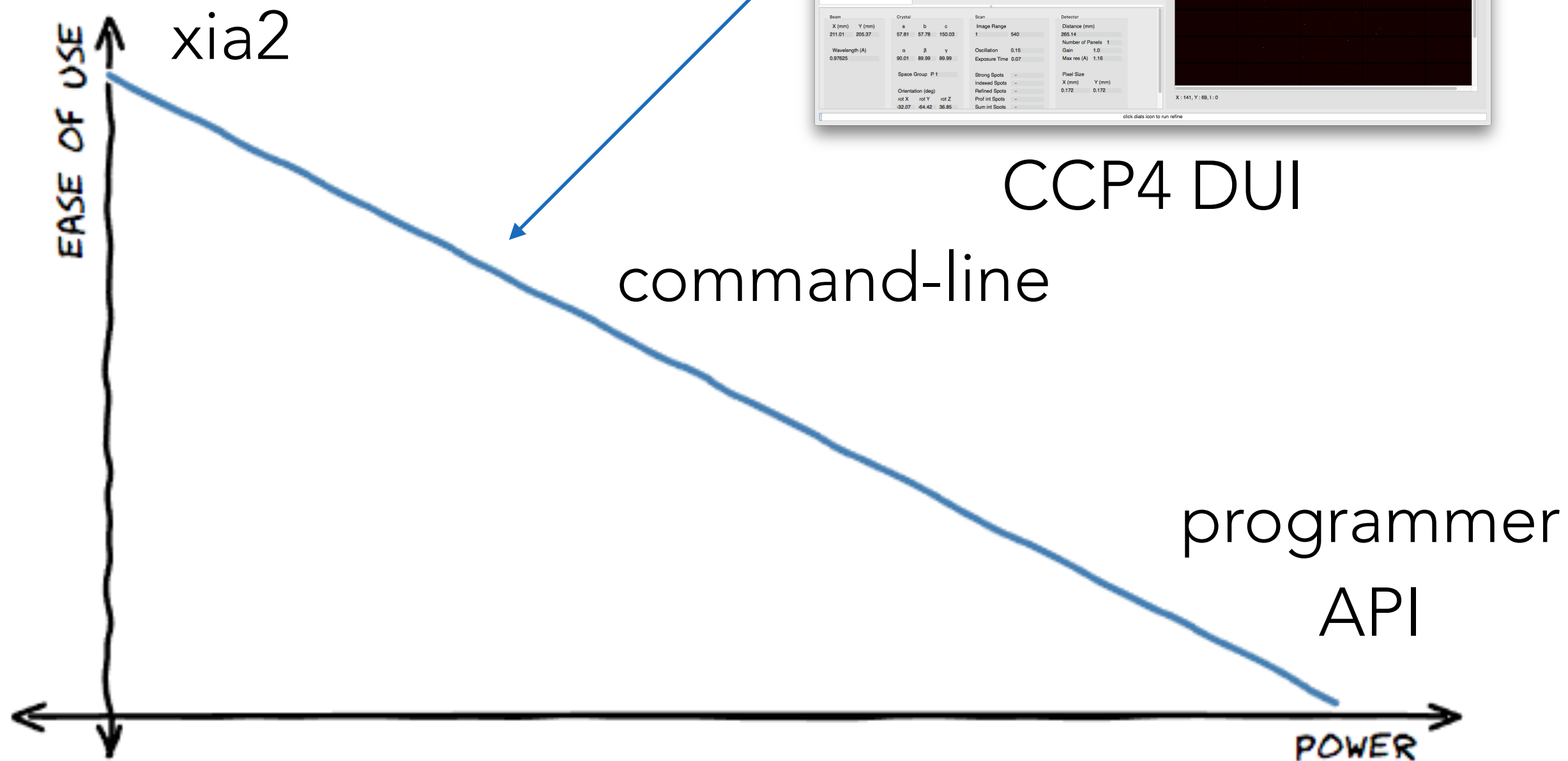


FASTEST ROUTE FROM IMAGES TO INTENSITIES

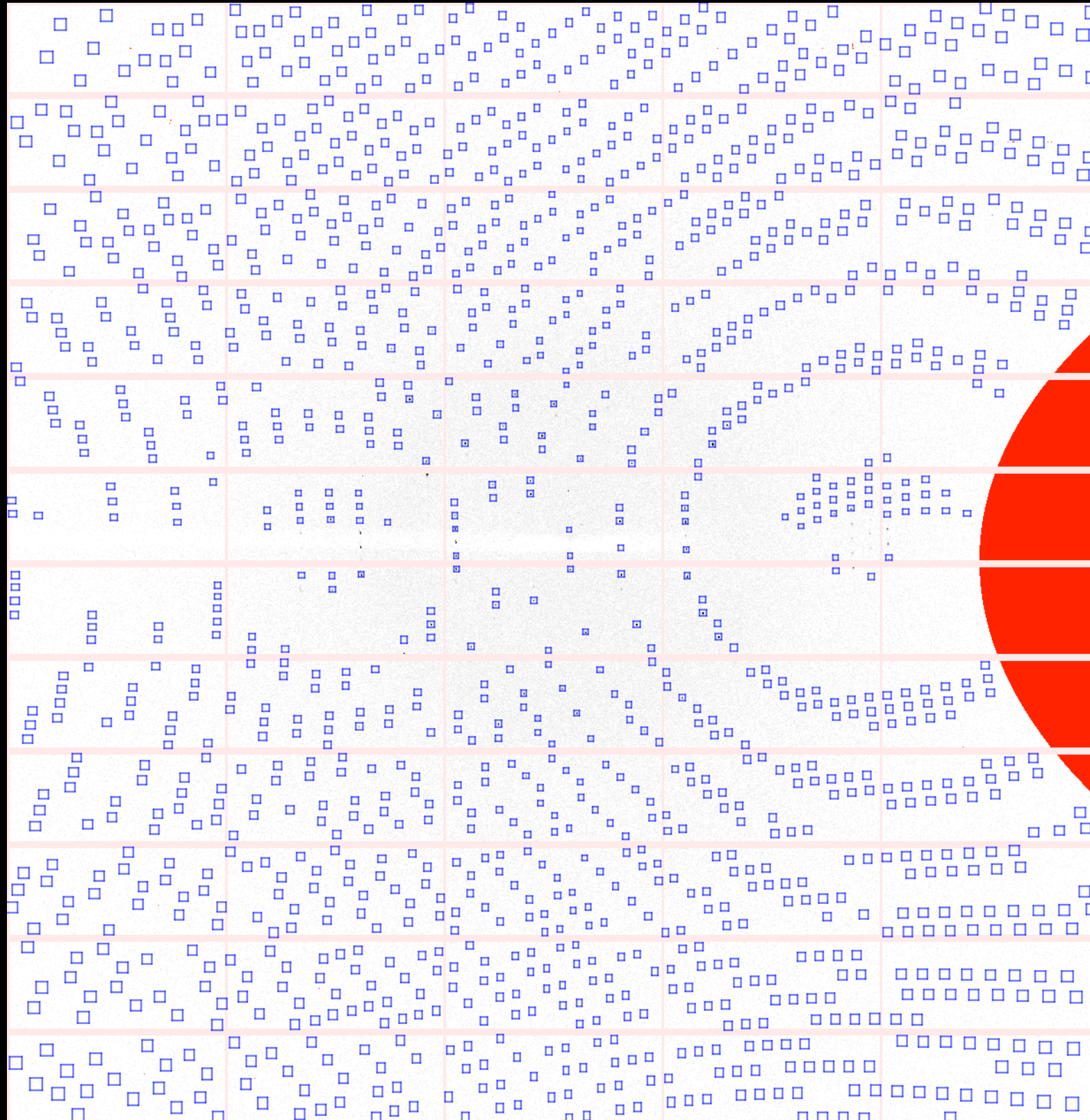
- Automation!
- generate_XDS.INP / autoPROC / xia2 / fast_dp /
most facilities have *something*
- For dials: xia2 =>

xia2 atom=Se image=../se-thau-mad/SeThau_hrm_3_master.h5
- You would learn *nothing*

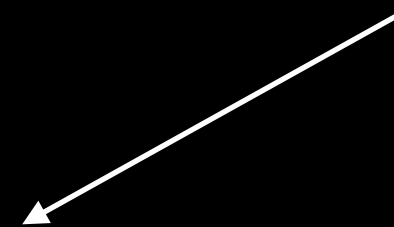




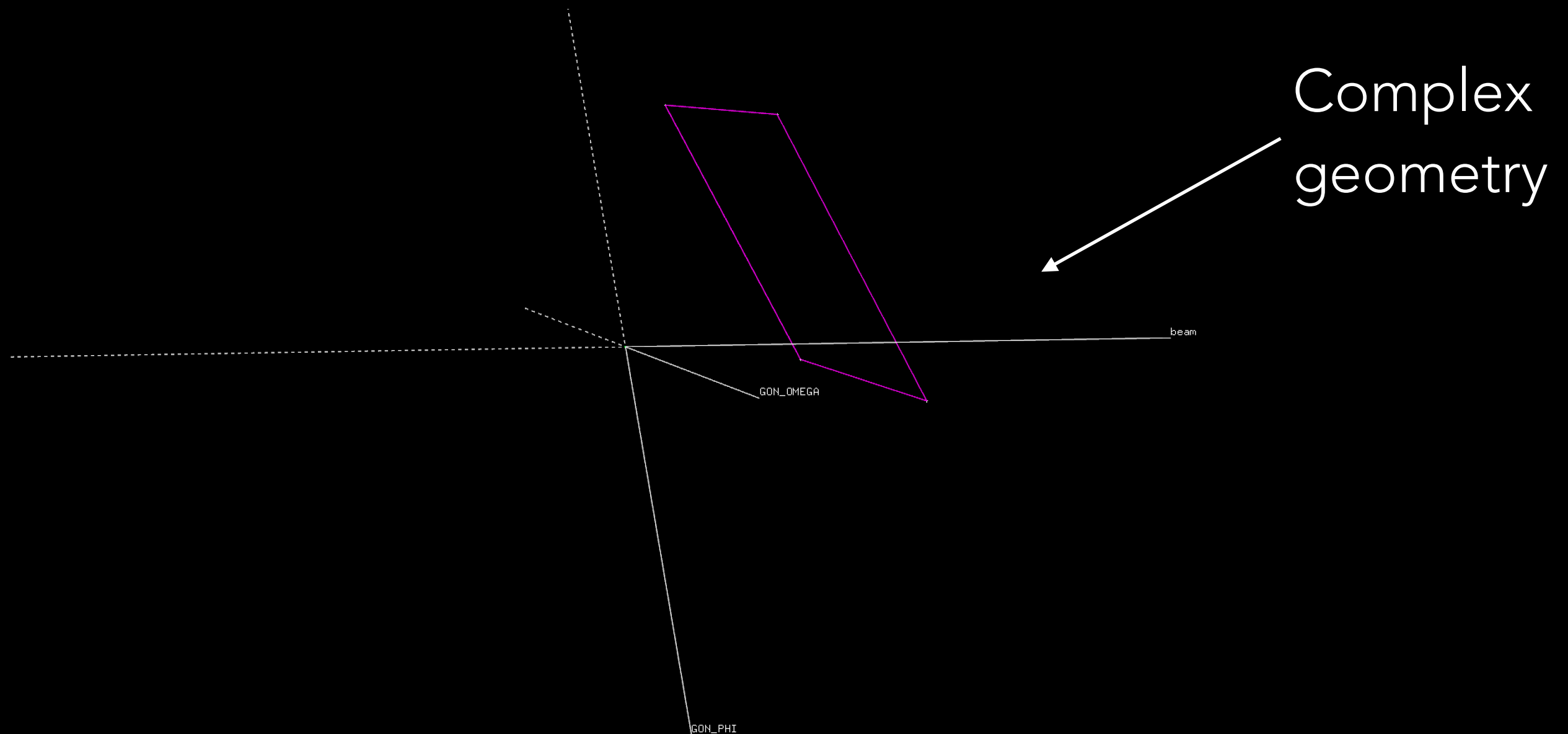
WHAT CAN DIALS COPE WITH?



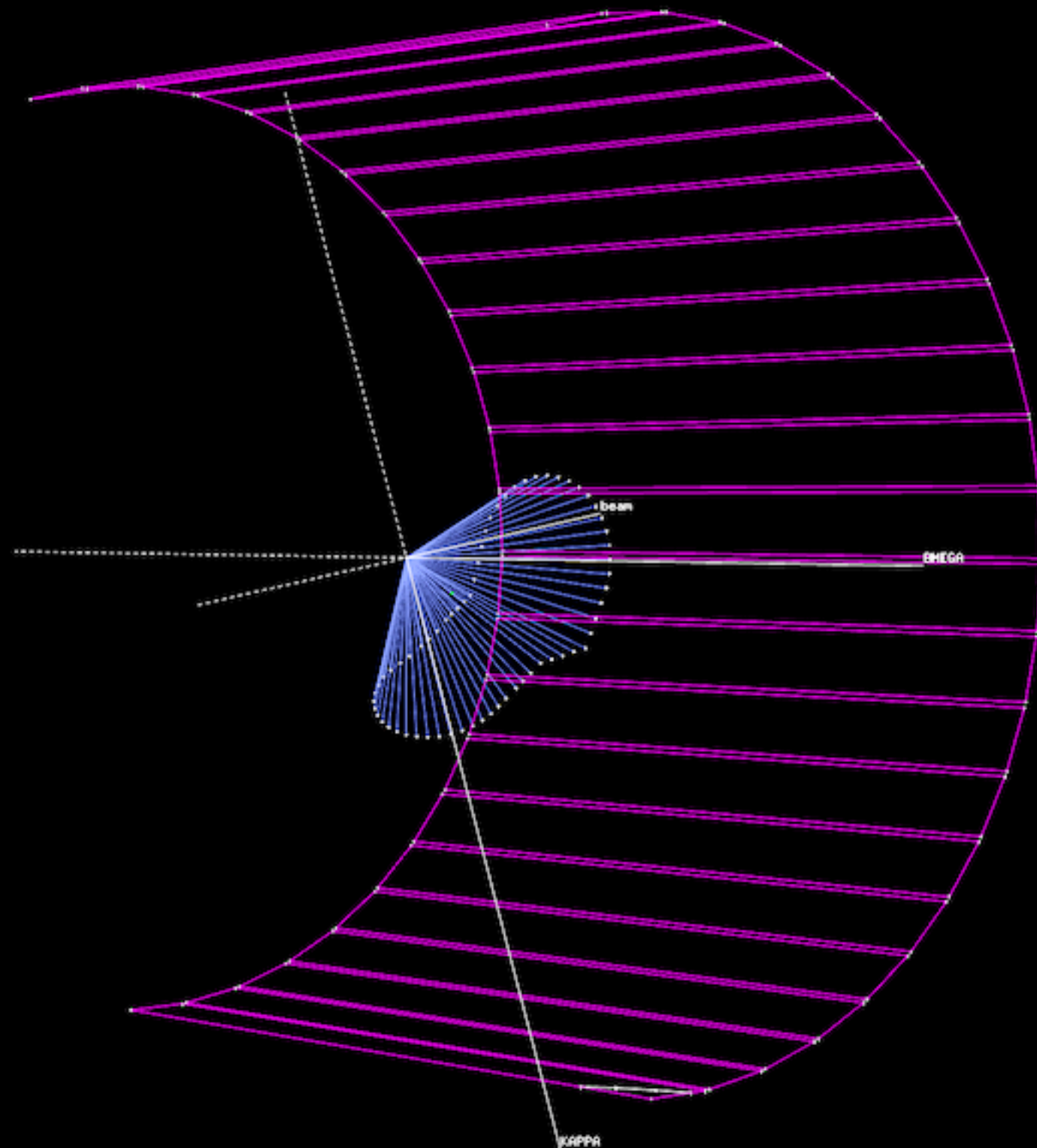
Moving
shadows!



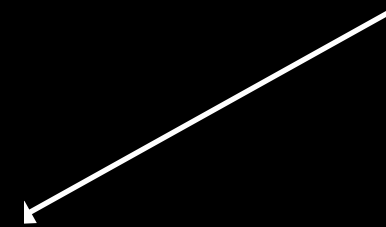
WHAT CAN DIALS COPE WITH?



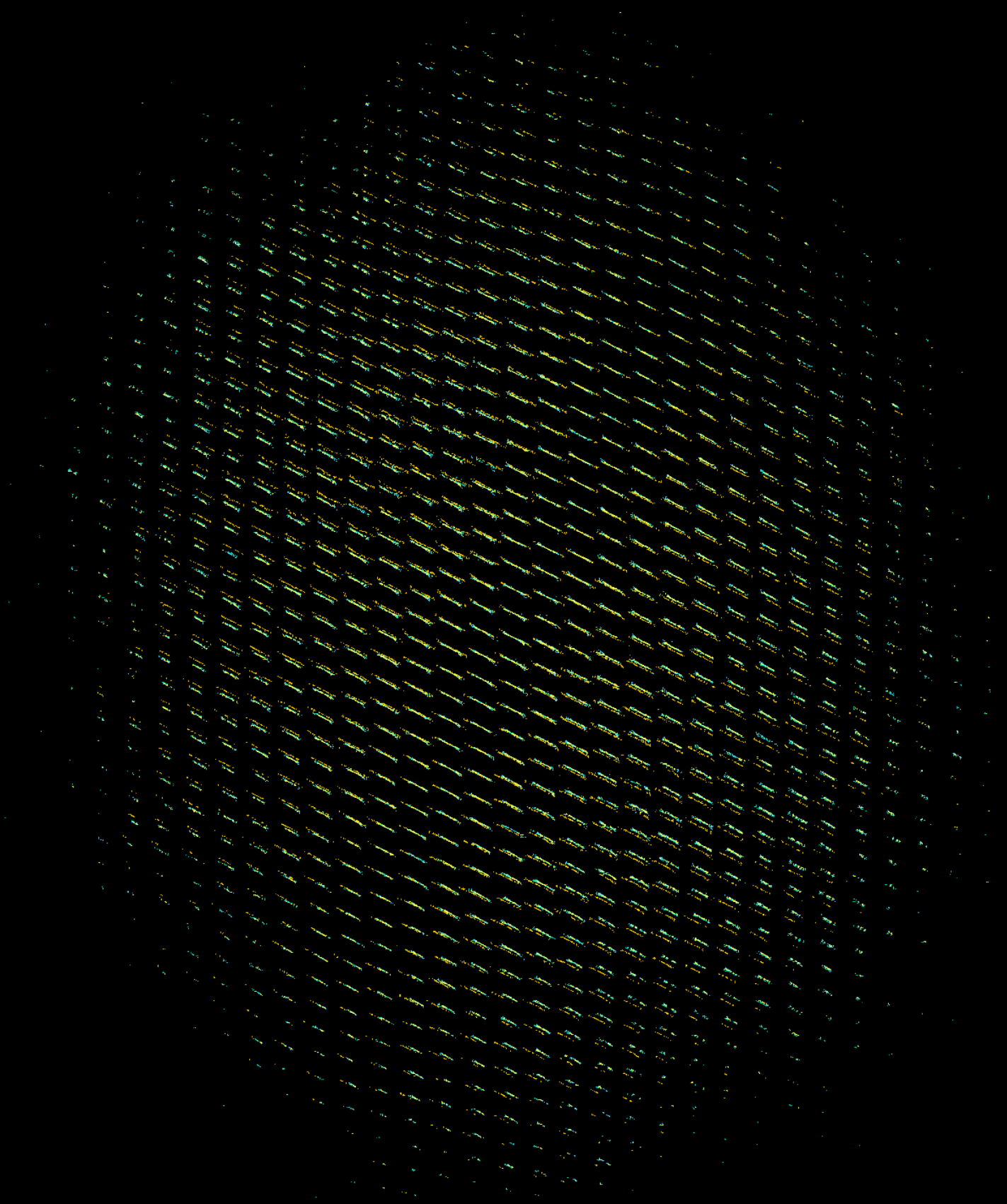
WHAT CAN DIALS COPE WITH?



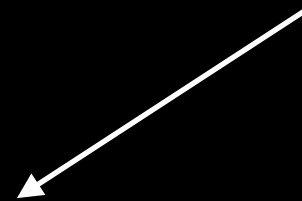
Complex
detectors



WHAT CAN DIALS COPE WITH?



Multiple
orientations



WEAK DIFFRACTION



KEY STEPS IN DATA PROCESSING

- Look at your images <—— people often miss this one
- Spot finding
- Indexing (optionally selecting lattice)
- Refinement (may be part of indexing)
- Integration
- Post integration refinement
- Point group identification
- Scaling

KEY STEPS WITH DIALS

dials.import

dials.find_spots imported.expt [nproc=8]

dials.index imported.expt strong.refl [space_group=..] [unit_cell=..]

[dials.refine_bravais_settings indexed.refl indexed.expt]

dials.refine

dials.integrate

dials.symmetry

dials.scale

dials.export

dials.image_viewer imported.expt

dials.reciprocal_lattice_viewer imported.expt strong.refl

KEY STEPS WITH DIALS

dials.import

dials.find_spots imported.expt [nproc=8]

dials.index imported.expt strong.refl [space_group=..] [unit_cell=..]

[dials.refine_bravais_settings indexed.refl indexed.expt]

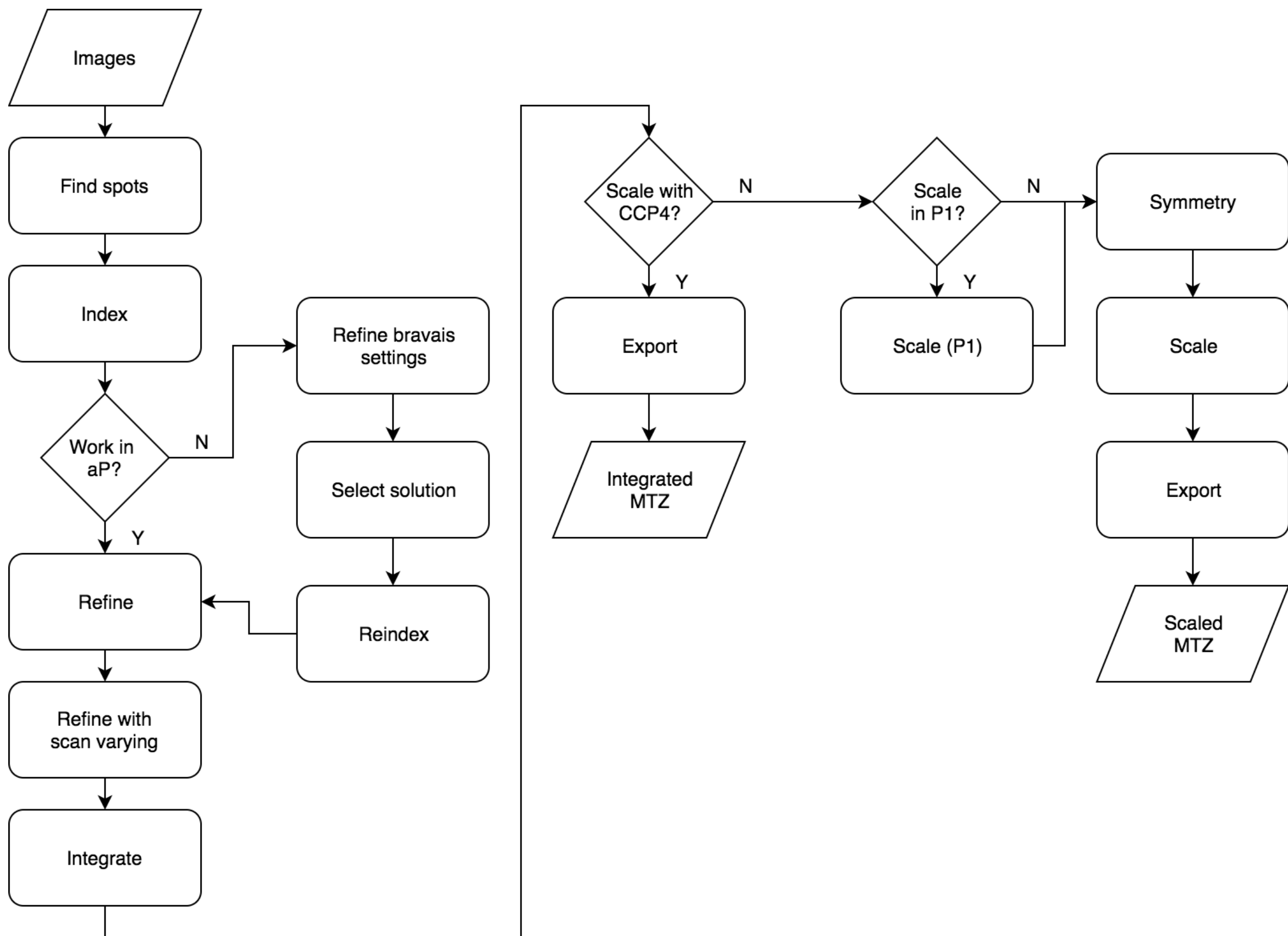
dials.refine

dials.integrate

dials.symmetry

dials.scale

dials.export



RUNNING DIALS

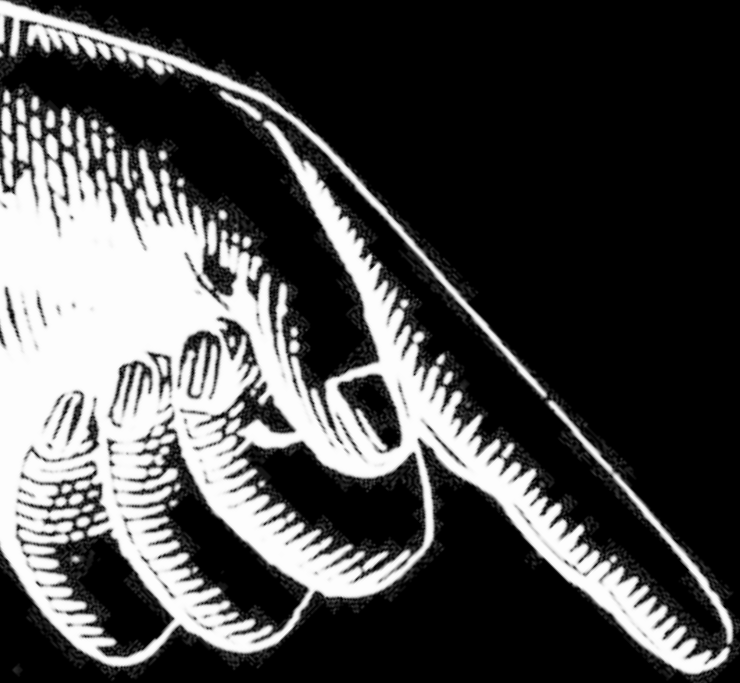
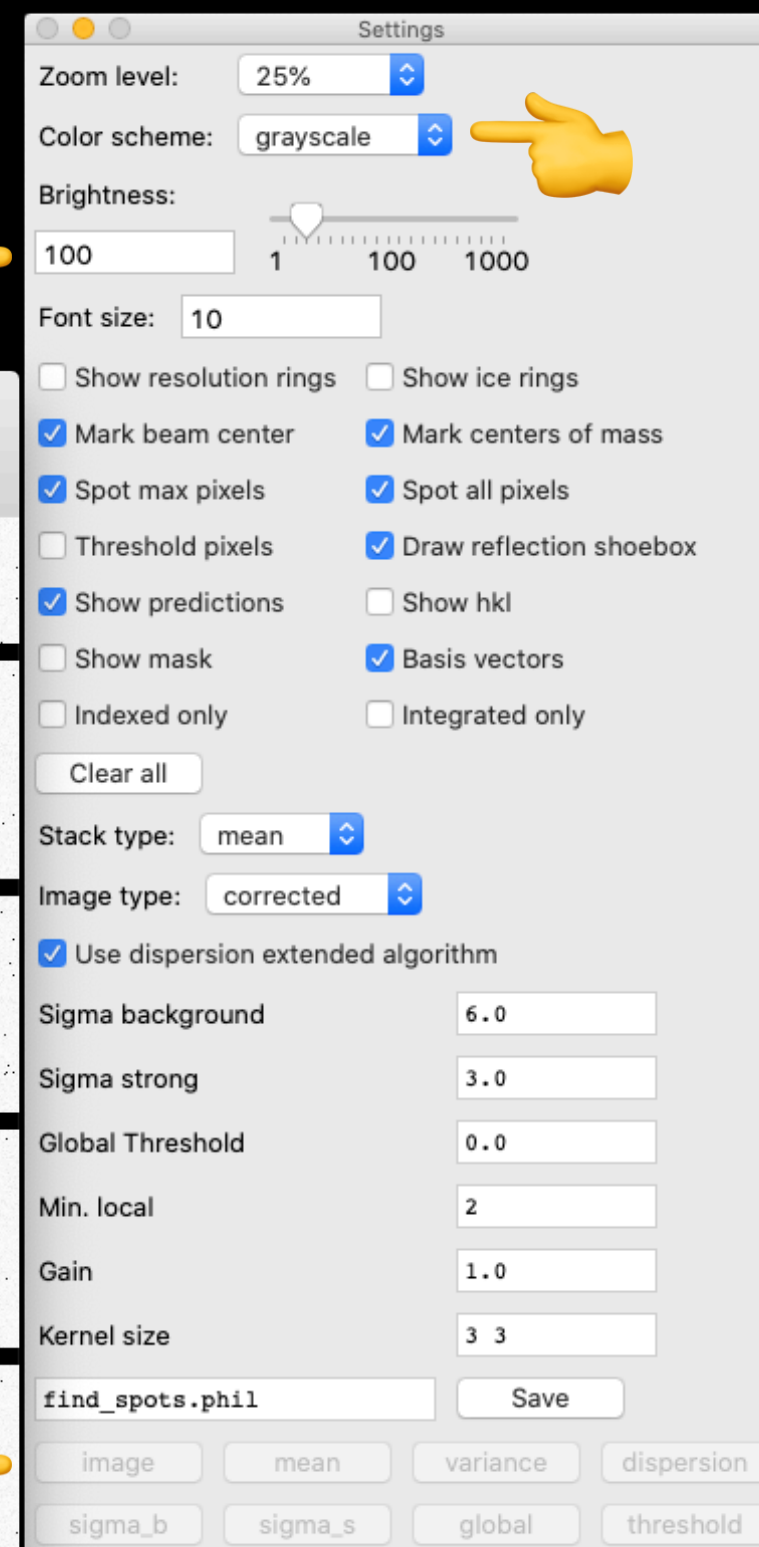
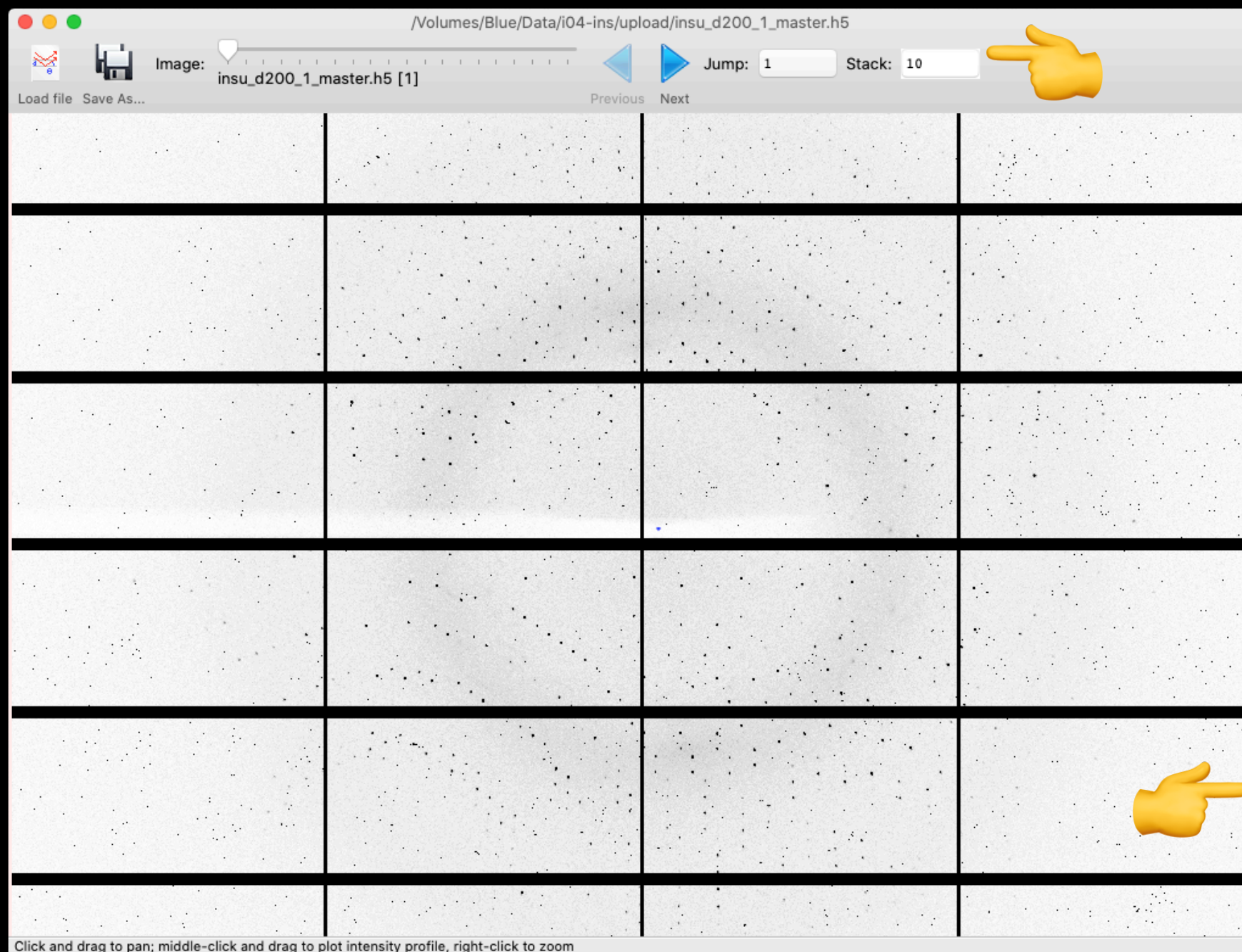


IMAGE VIEWER



DIALS FILES

- CBF / HDF5 files contain X-ray diffraction images
- refl files contain reflection data
- expt files contain model information
- MTZ files are probably what you want at the end for structure determination and refinement

ALGORITHMS

SPOT FINDING

SPOT FINDING



SPOT FINDING



SPOT FINDING



SPOT FINDING



SPOT FINDING



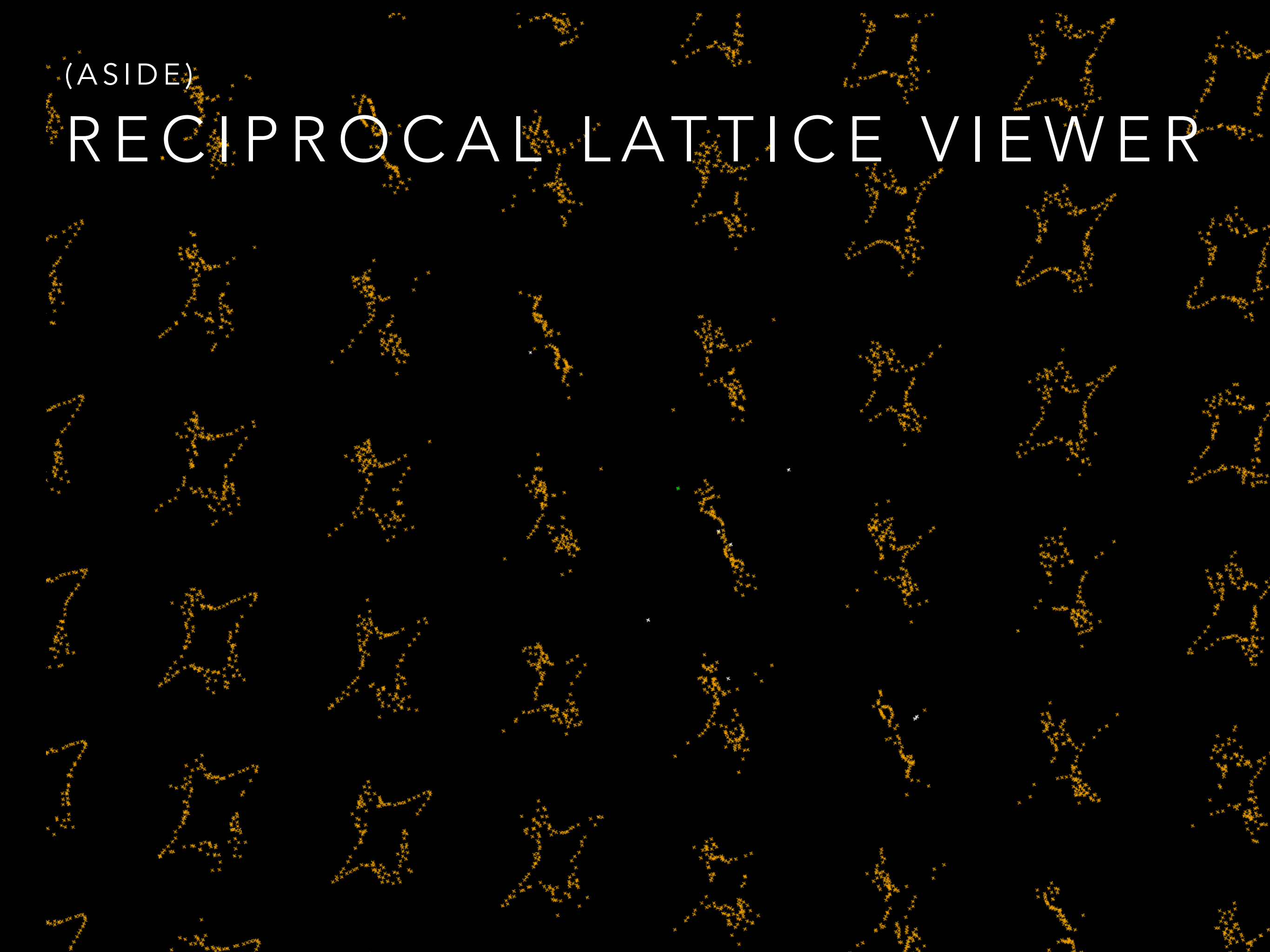
SPOT FINDING



MAP TO RECIPROCAL SPACE

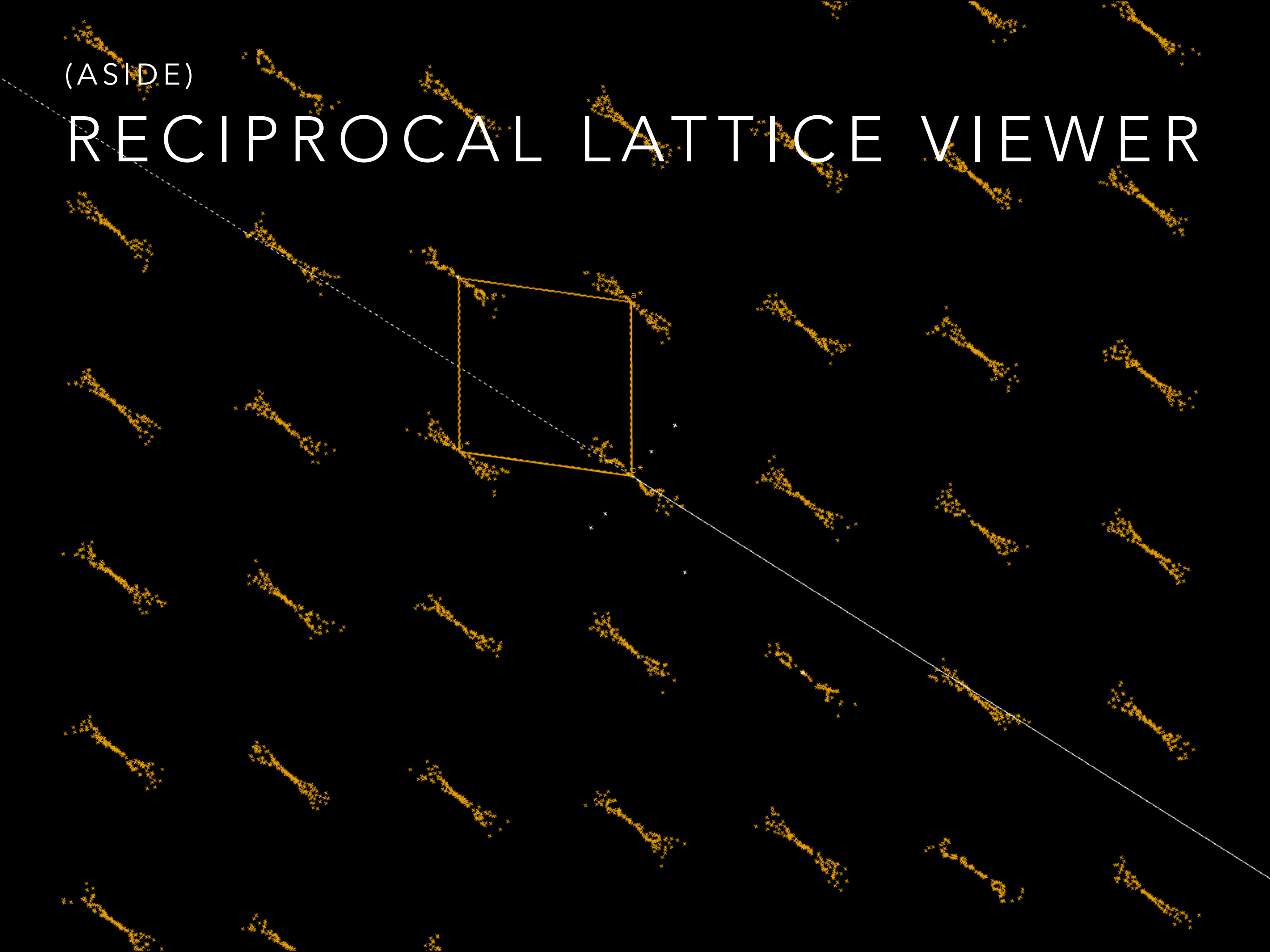


RECIPROCAL LATTICE VIEWER

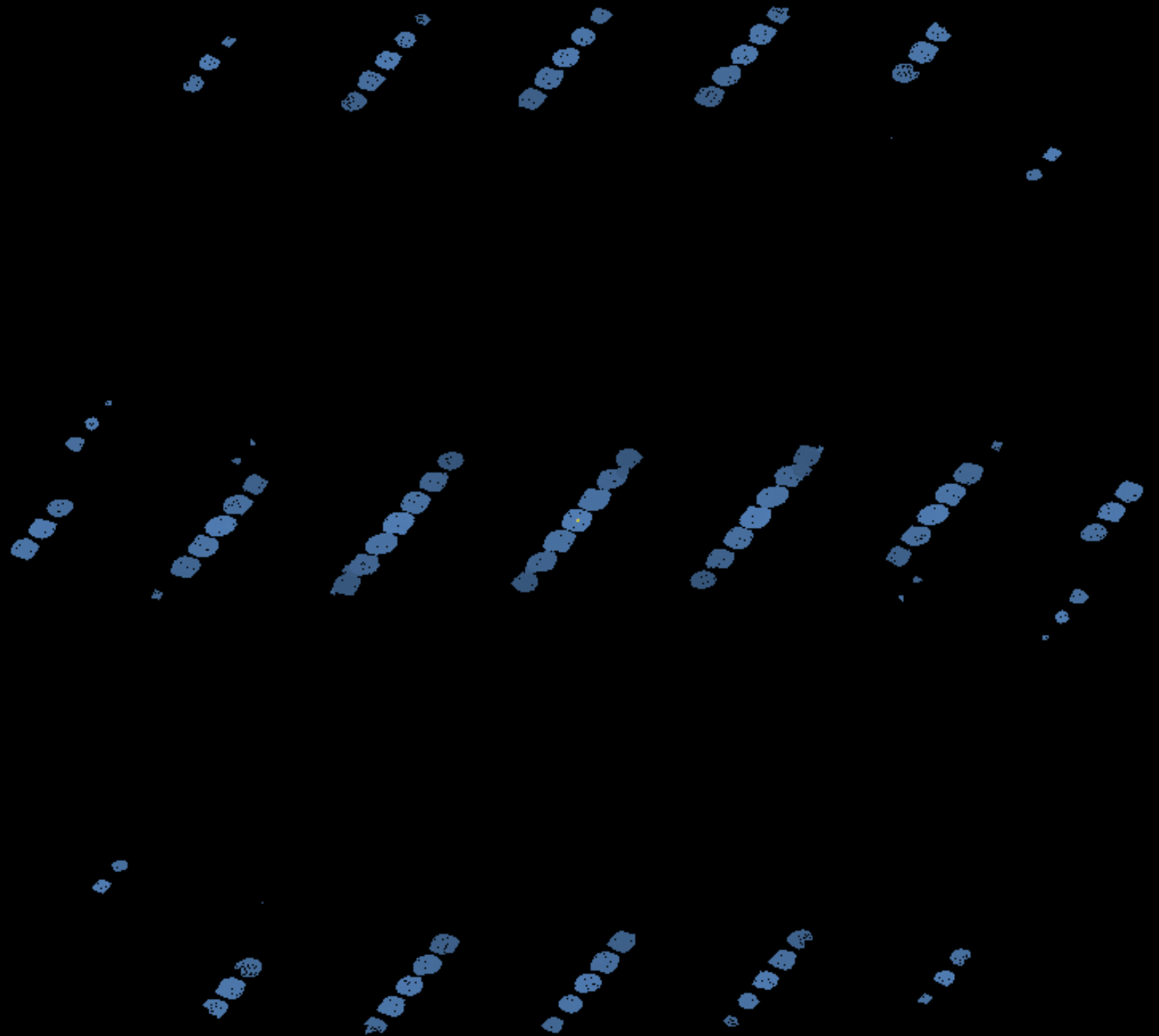


(ASIDE)

RECIPROCAL LATTICE VIEWER



INDEXING (3D FFT)



SOLUTION PICKING

Solution	Metric	fit	rmsd	min/max	cc	#spots	lattice			unit_cell			volume	cb_op

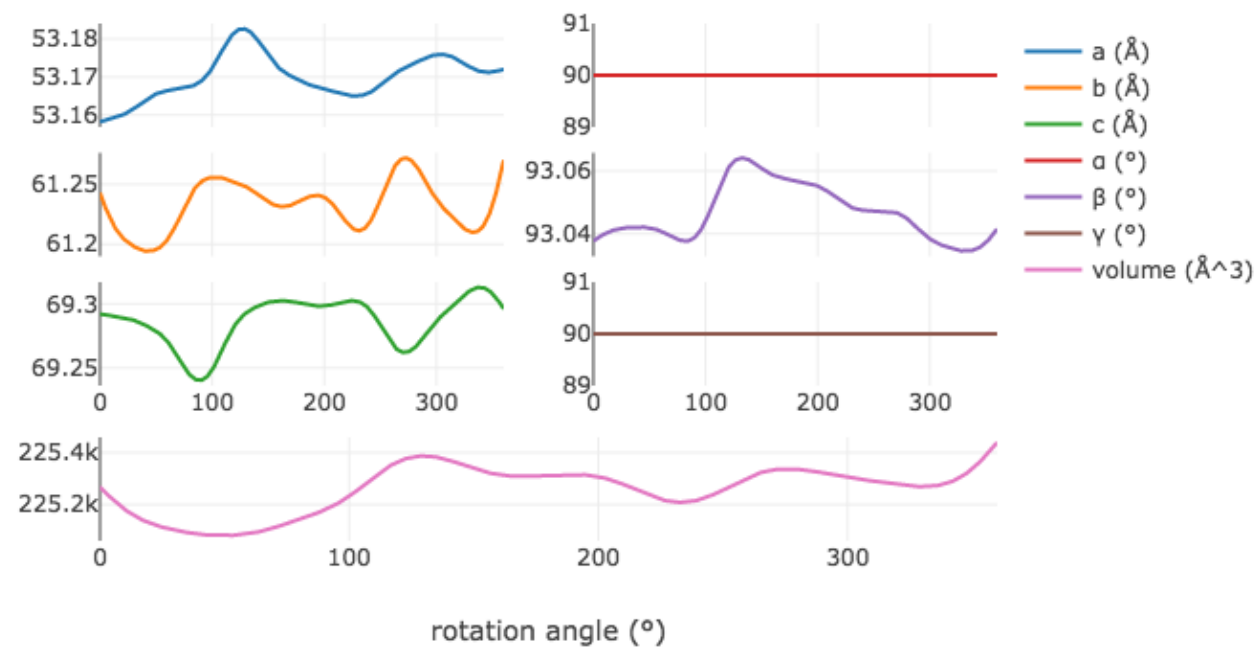
5	3.0470	2.331	0.619/0.965	36000	oC	52.67	60.74	68.55	90.00	90.00	90.00	219311	a+b, -a+b, c	
4	3.0469	2.325	0.621/0.621	36000	mC	60.77	52.69	68.57	90.00	90.04	90.00	219554	a-b, a+b, c	
3	3.0470	2.346	0.619/0.619	35999	mP	40.27	68.59	40.25	90.00	98.17	90.00	110071	b, c, a	
* 2	0.0313	0.072	0.965/0.965	36000	mC	53.18	61.25	69.29	90.00	93.05	90.00	225383	a+b, -a+b, c	
* 1	0.0000	0.070	-/-	36000	aP	40.56	40.55	69.30	91.97	92.02	98.07	112701	a, b, c	



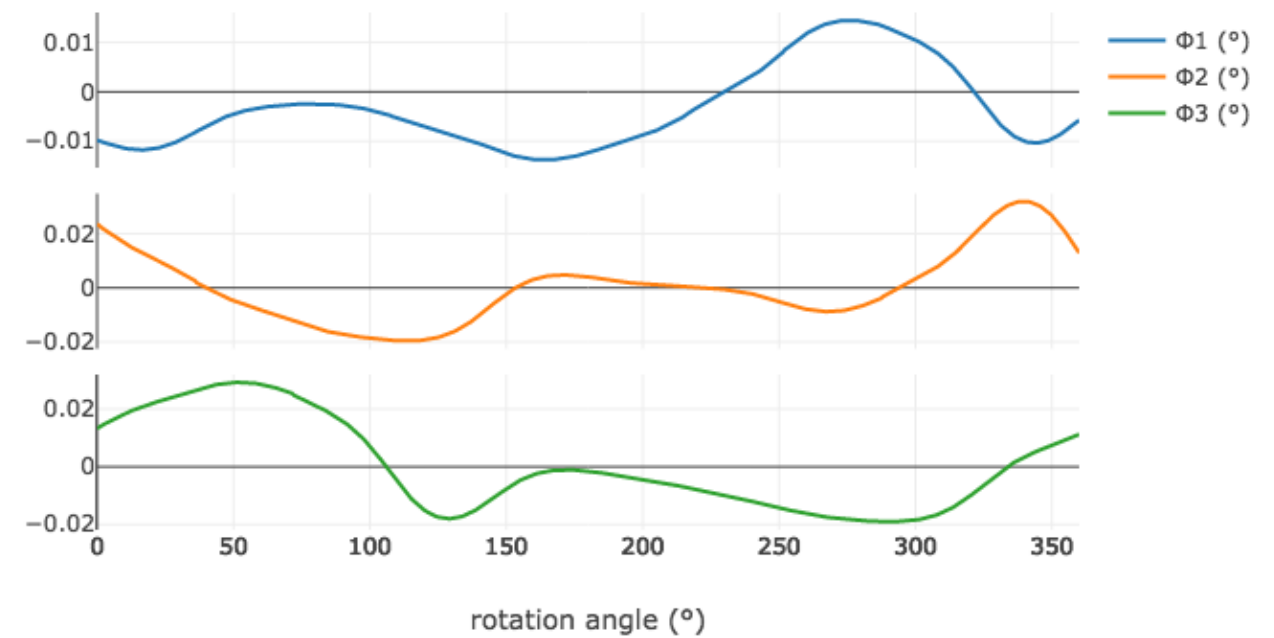
REFINEMENT

REFINEMENT

Scan-varying cell parameters



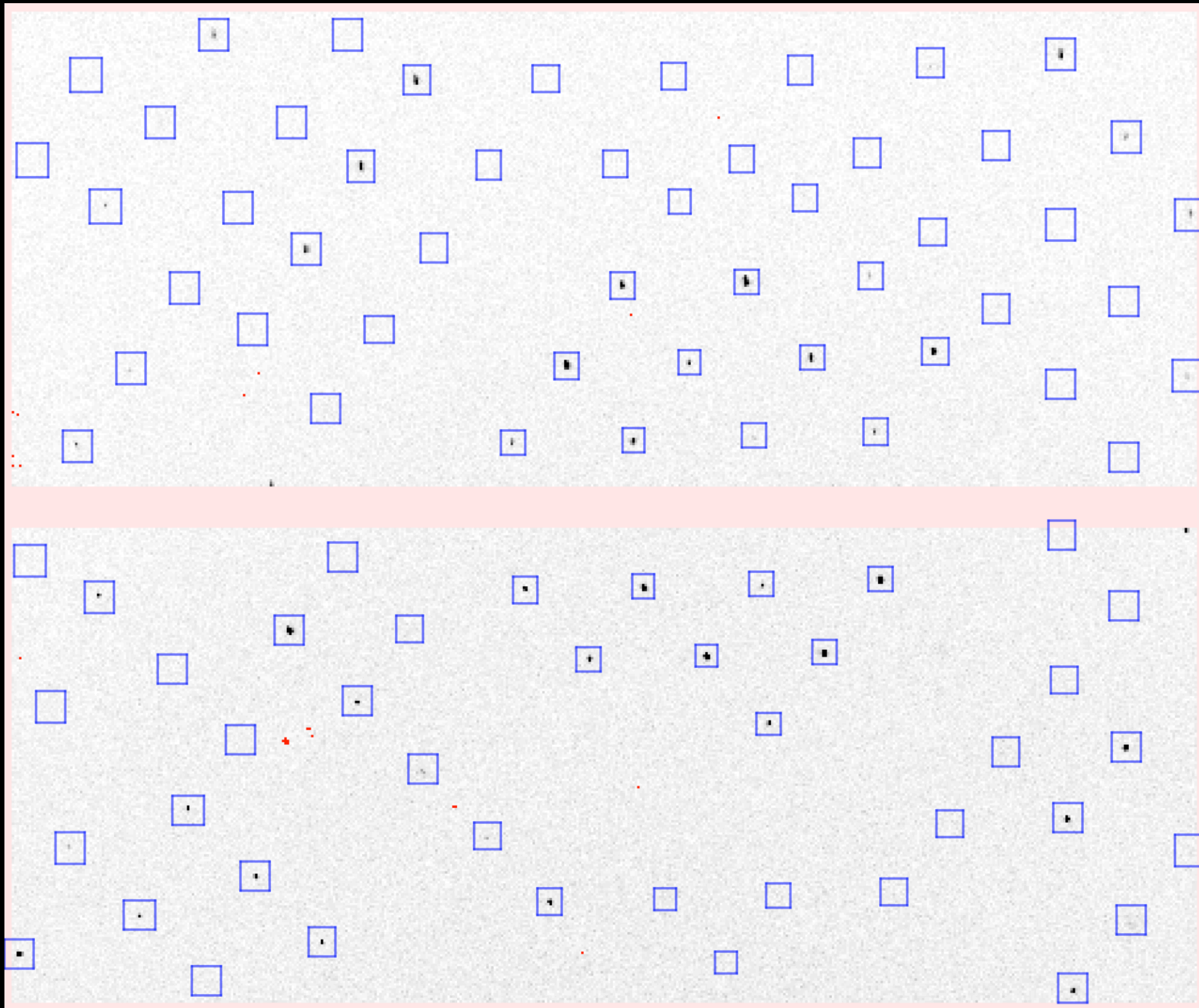
Scan-varying orientation parameters



REFINEMENT FEATURES

- Possible to do *joint refinement* - one model for geometry for e.g. 50 data sets
- True errors on unit cell parameters - critical for assessing isomorphism

INTEGRATION



SCALING

- `dials.symmetry` - determine the Patterson symmetry of the data
- `dials.scale` - correct for experimental effects using symmetry-related experiments

SYMMETRY

Scoring individual symmetry elements

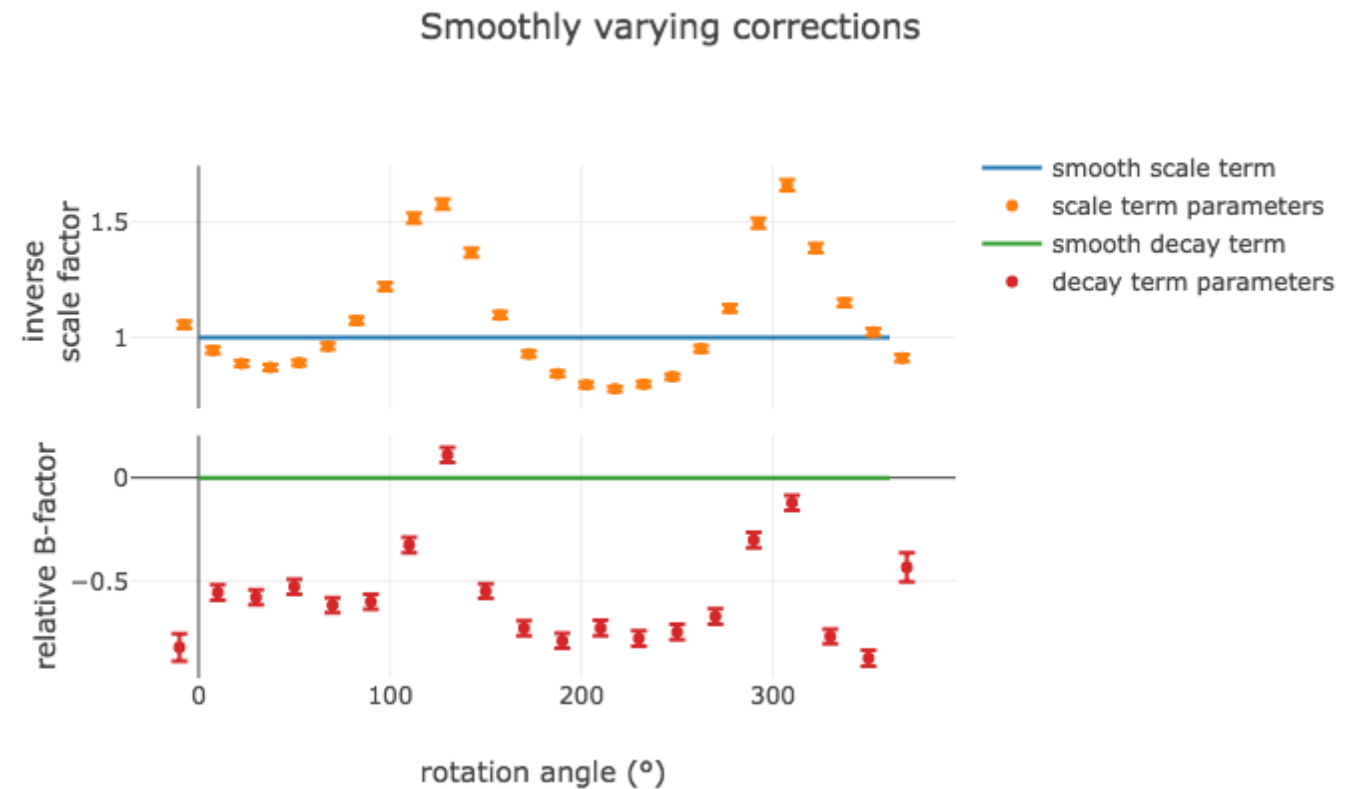
likelihood	Z-CC	CC	N		Operator
0.916	9.81	0.98	297147	***	1 (0, 0, 0)
0.913	9.72	0.97	294871	***	2 (-1, 1, 0)

Scoring all possible sub-groups

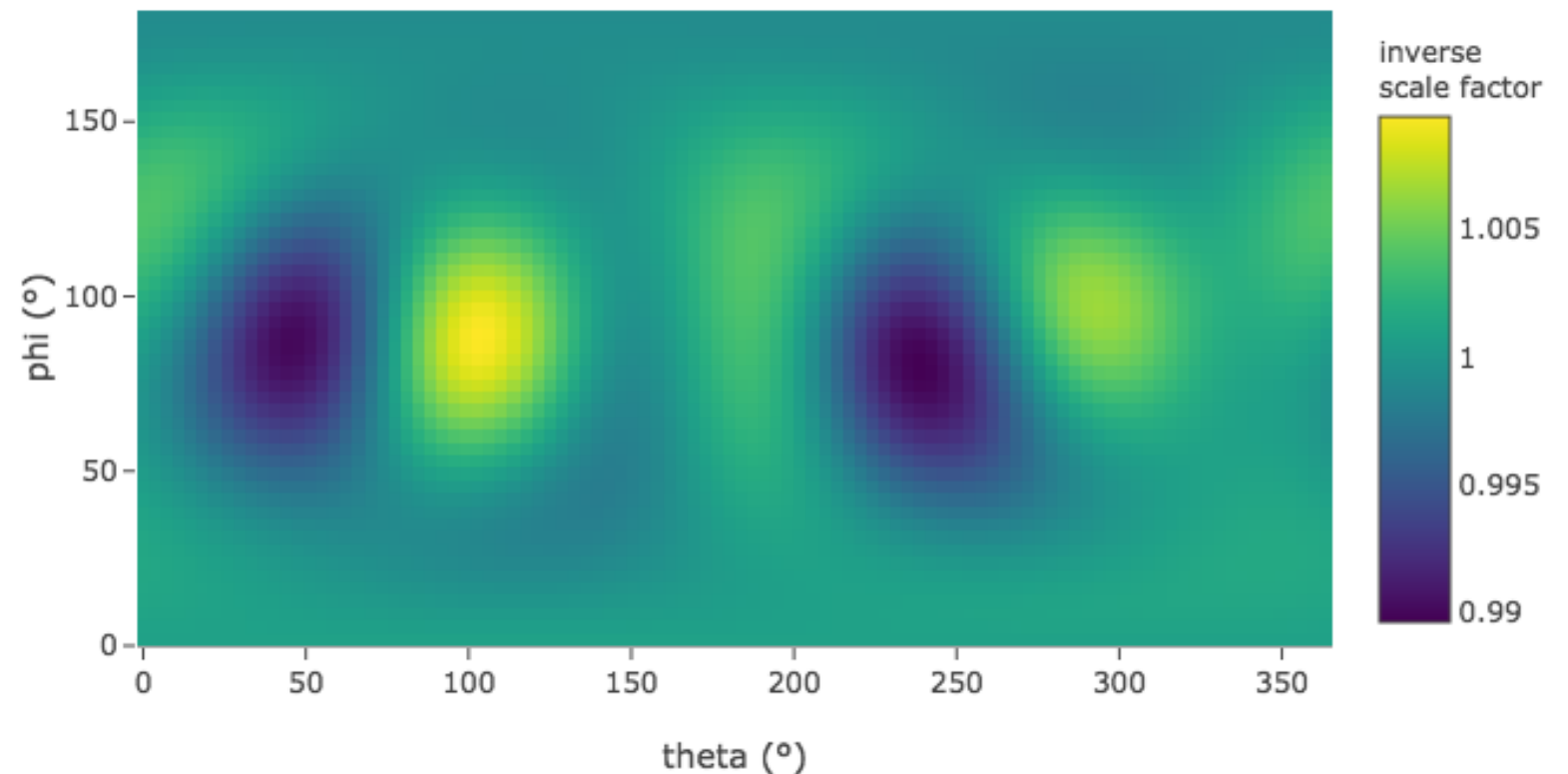
Patterson group		Likelihood	NetZcc	Zcc+	Zcc-	CC	CC-	delta	Reindex
C 1 2/m 1	***	0.913	9.76	9.76	0.00	0.98	0.00	0.0	-a,b,-c
P -1		0.087	0.09	9.81	9.72	0.98	0.97	0.0	-x-y,-x+y,-z

SCALING

kB (smoothed)

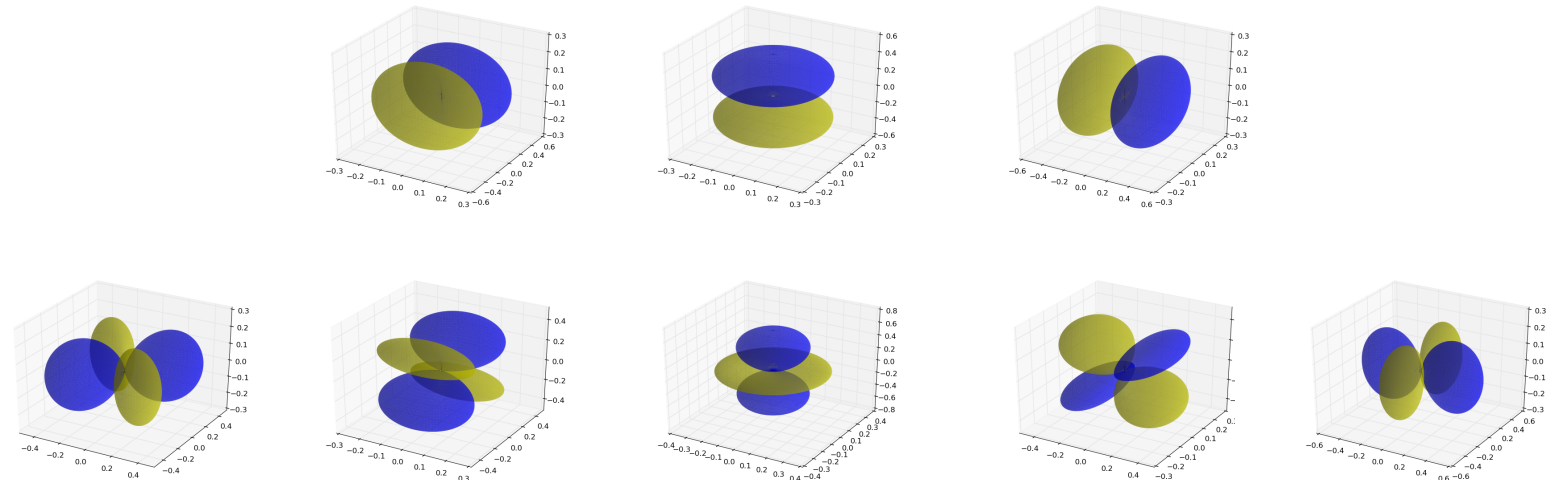


Absorption correction surface



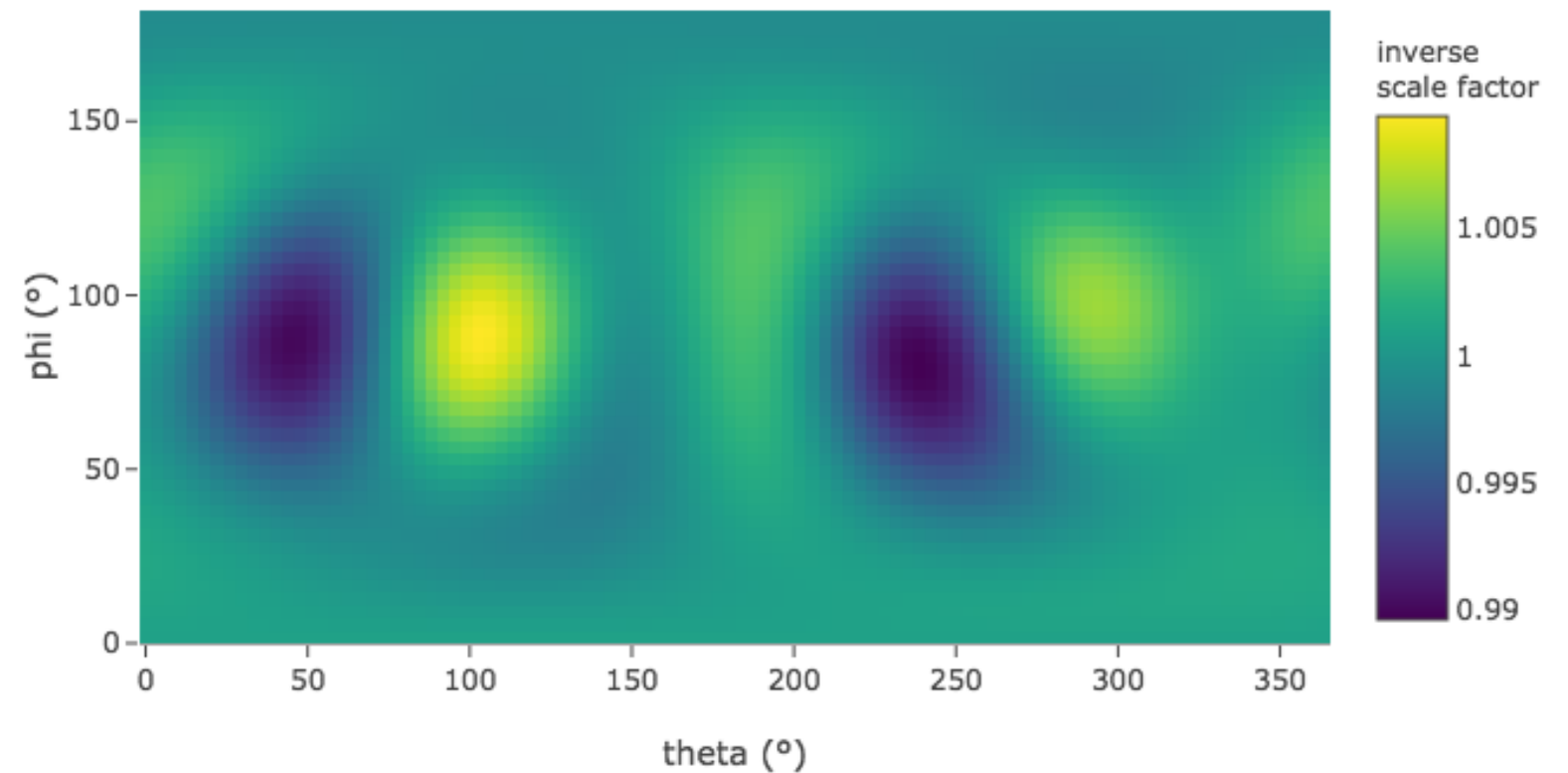
Absorption surface
(as spherical
harmonics)

SCALING

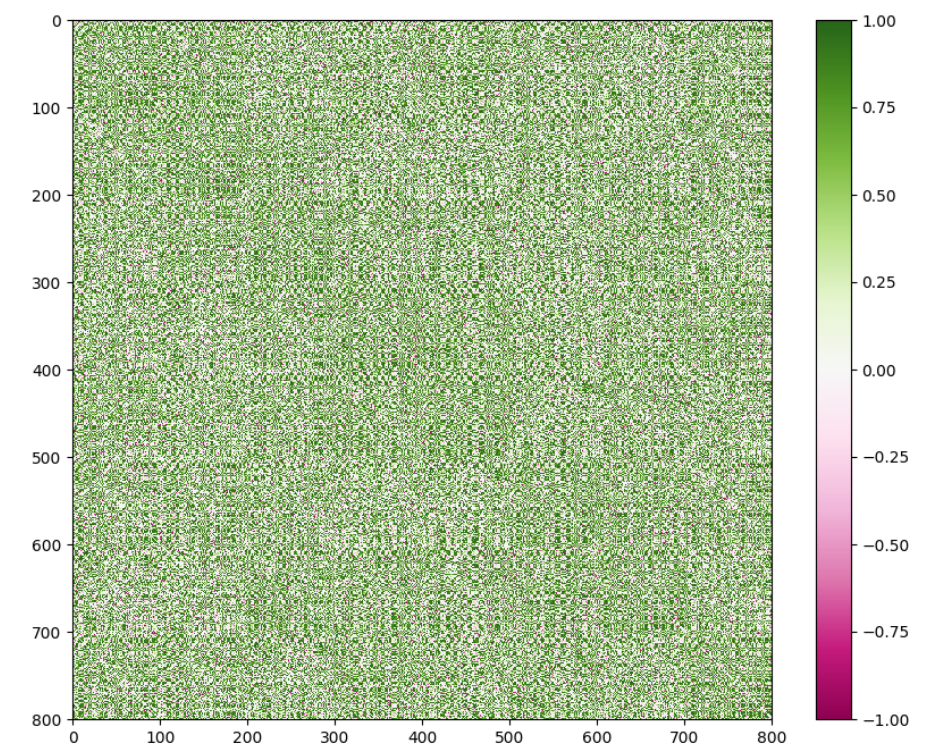
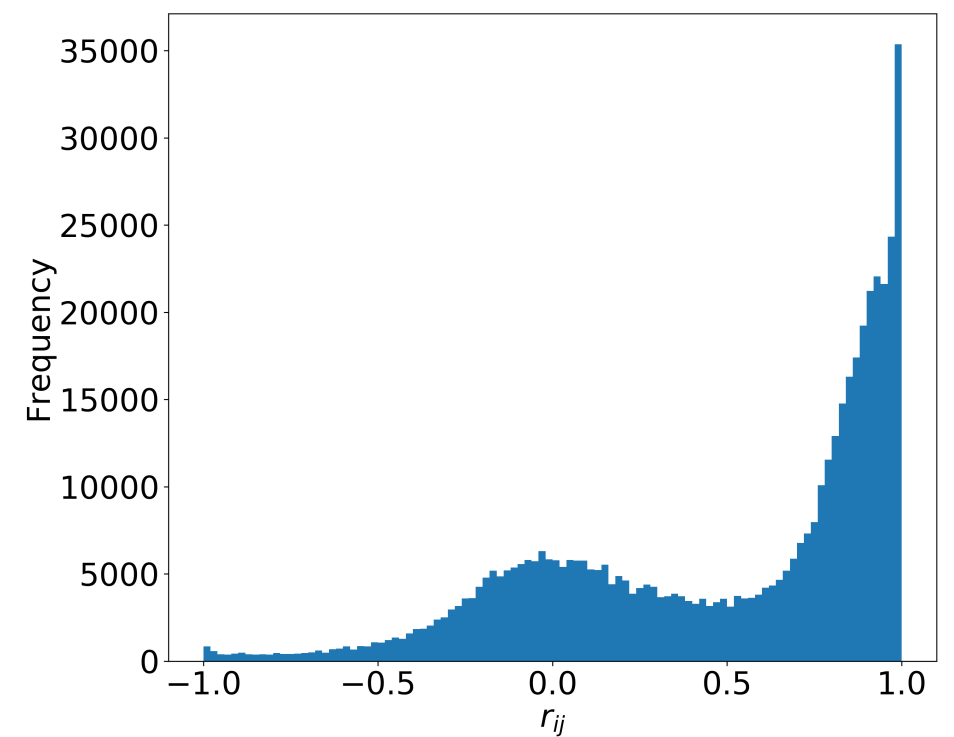
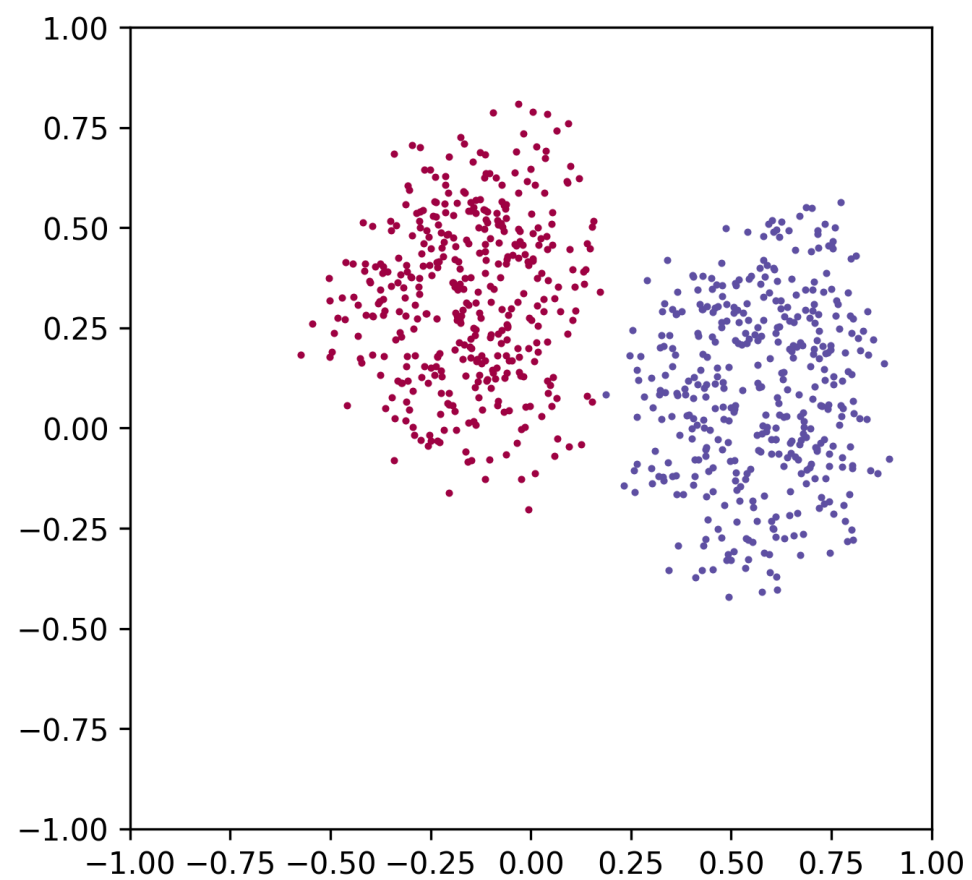


Absorption correction surface

Absorption surface
(as spherical
harmonics)



SYMMETRY & ISOMORPHISM - COSYM



COMMENTARY

- DIALS is open source, so you can look at the code to find out exactly what it is doing
- Everything here is applied mathematics and nothing more - there are no black boxes
- If ever you really want to understand X-ray diffraction, write an integration program 😊



Diffraction Integration for Advanced Light Sources

[Home](#)

[About](#)

[Installation](#)

[Documentation](#)

[Tutorials](#)

[How-to](#)

[Workshops](#)

[Publications](#)

[Links](#)

[License](#)

Processing in Detail - CCP4 Workshop Edition

Introduction

DIALS processing may be performed by either running the individual tools (spot finding, indexing, refinement, integration, exporting to MTZ) or you can run `xia2 pipeline=dials`, which makes informed choices for you at each stage. In this tutorial we will run through each of the steps in turn, checking the output as we go. We will also enforce the correct lattice symmetry.

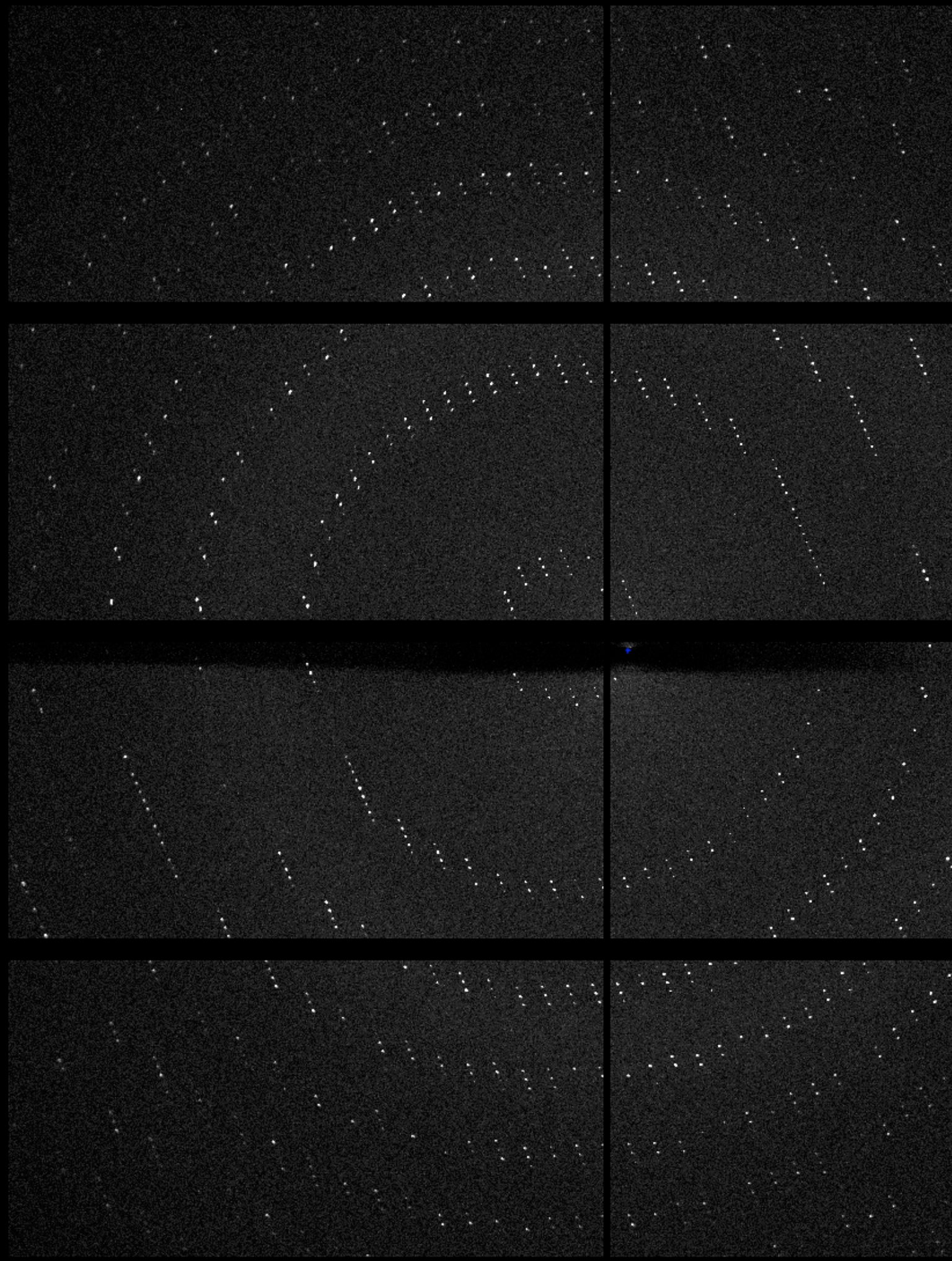
Tutorial data

The following example uses a Beta-Lactamase dataset collected using beamline I04 at Diamond Light Source, and reprocessed especially for these tutorials. The data is available for download from [DOI 10.5281/zenodo.1014387](https://doi.org/10.5281/zenodo.1014387), but is also available for the workshop in the beamline data directory of `/dls/i03/data/2017/mx19576-1/tutorial_data/summed/`. We'll only be using the first run of data in this tutorial.

TUTORIAL

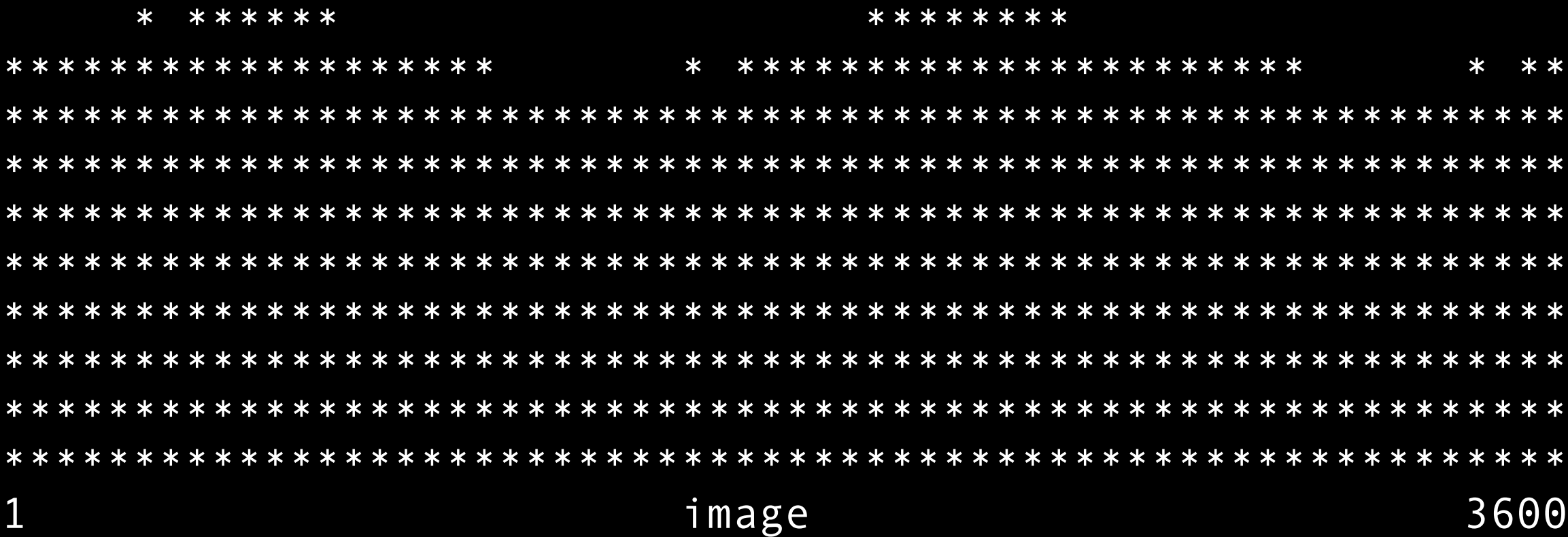
SE THAUMATIN

- Single sweep of data measured for routine testing
- 3600 x 0.1° images
- Simple to process through to refined structure
- Good for learning how to drive DIALS



SPOT FINDING

Histogram of per-image spot count for imageset 0:
299032 spots found on 3600 images (max 5558 / bin)

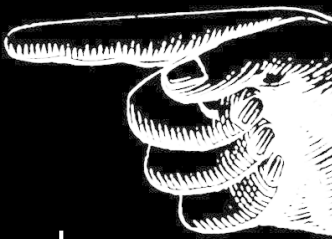


RECIPROCAL LATTICE VIEW

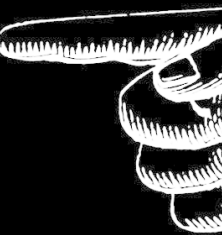


INDEXING

Exp id	Nref	RMSD_X (px)	RMSD_Y (px)	RMSD_Z (images)
0	36000	0.36113	0.28824	0.53236



Imageset	# indexed	# unindexed	% indexed
0	281437	17594	94.1%

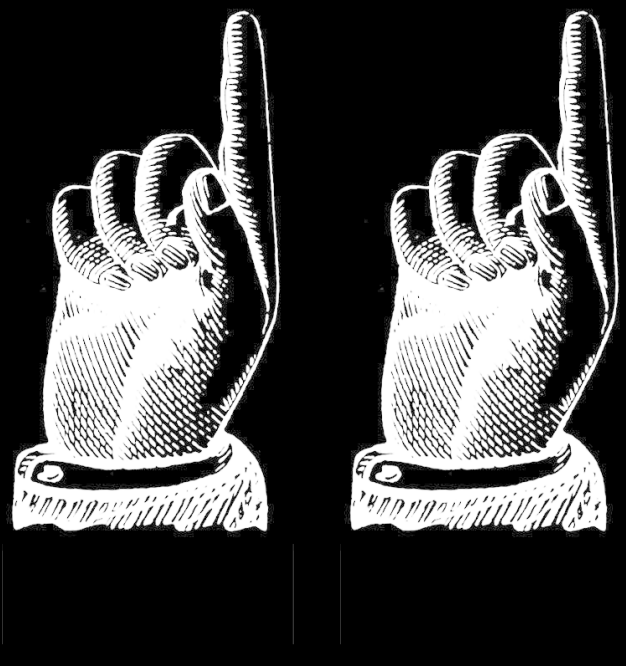


LOOK AT THE IMAGES



REFINE BRAVAIS SETTINGS

Solution		Metric fit	rmsd	min/max cc	#spots	lattice	unit_cell				
*	9	0.0307	0.038	0.829/0.865	36000	tP	57.92	57.92	150.05	90.00	90.00
*	8	0.0256	0.037	0.827/0.865	36000	oC	81.89	81.93	150.05	90.00	90.00
*	7	0.0307	0.037	0.865/0.954	36000	oP	57.92	57.90	150.03	90.00	90.00
*	6	0.0256	0.037	0.832/0.832	36000	mC	81.89	81.93	150.05	90.00	90.00
*	5	0.0307	0.037	0.866/0.866	36000	mP	57.91	57.92	150.04	90.00	90.00
*	4	0.0175	0.036	0.827/0.827	36000	mC	81.92	81.89	150.04	90.00	90.00
*	3	0.0265	0.036	0.954/0.954	36000	mP	57.91	57.90	150.01	90.00	90.00
*	2	0.0206	0.036	0.865/0.865	36000	mP	57.90	150.03	57.92	90.00	90.00
*	1	0	0.035	-/-	36000	aP	57.92	57.90	150.02	90.01	90.01



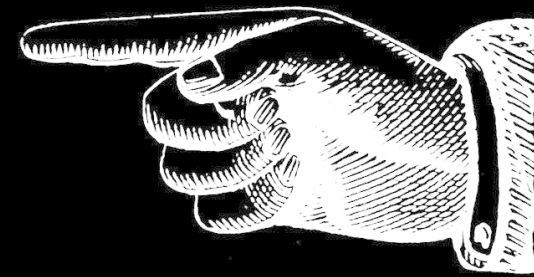
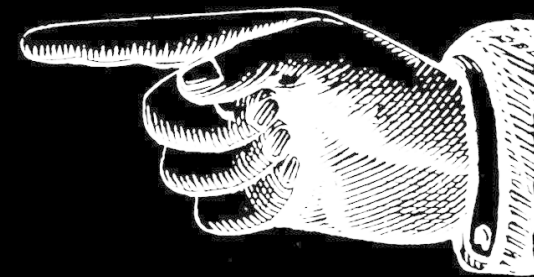
REFINEMENT (WITH SCAN VARYING)

From indexing:

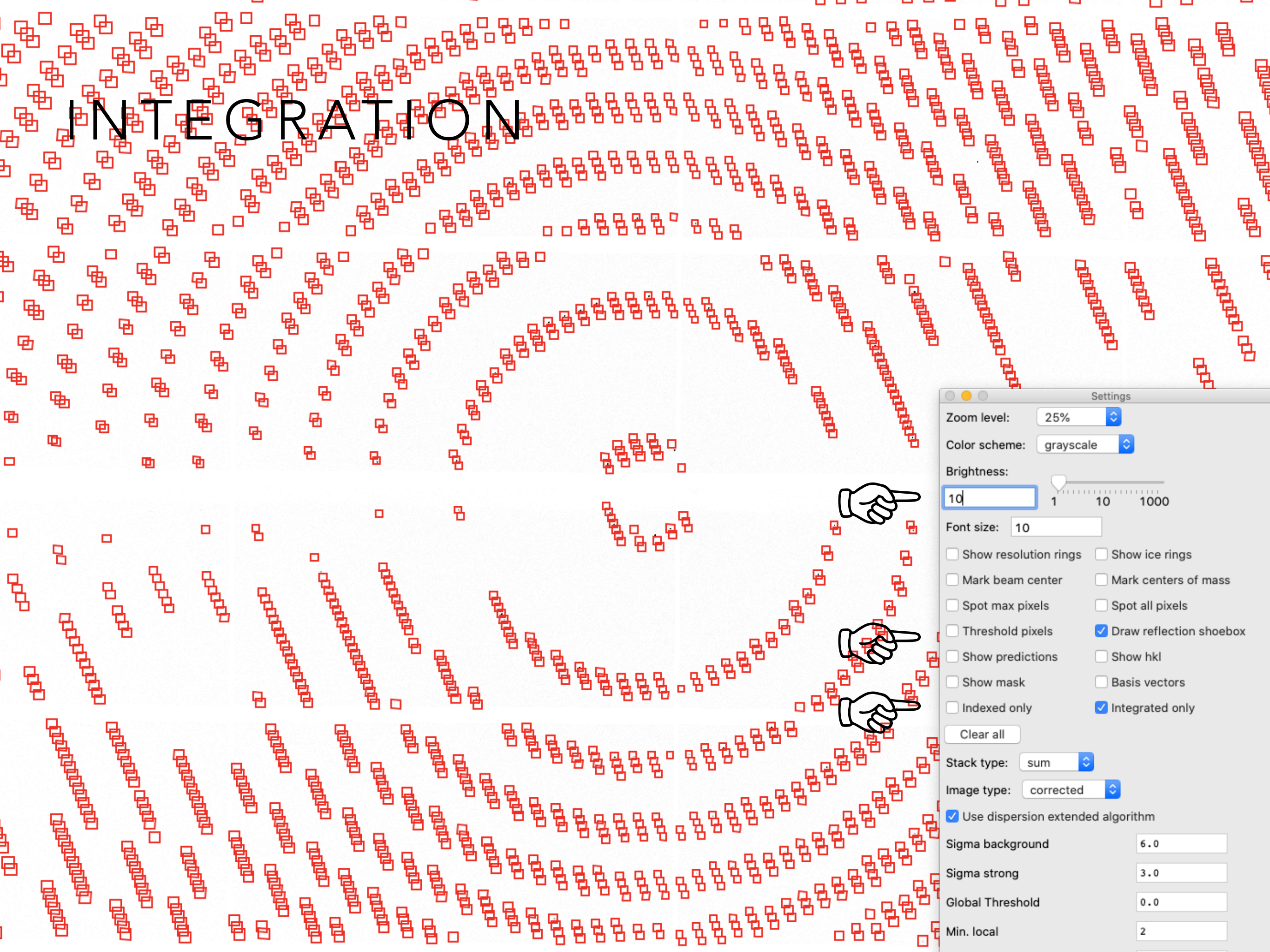
Exp id	Nref	RMSD_X (px)	RMSD_Y (px)	RMSD_Z (images)
0	36000	0.36113	0.28824	0.53236

From refinement:

Exp id	Nref	RMSD_X (px)	RMSD_Y (px)	RMSD_Z (images)
0	222068	0.24386	0.22748	0.20681



INTEGRATION



Settings

Zoom level: 25%

Color scheme: grayscale

Brightness: 10

Font size: 10

☐ Show resolution rings ☐ Show ice rings

☐ Mark beam center ☐ Mark centers of mass

☐ Spot max pixels ☐ Spot all pixels

☐ Threshold pixels ☒ Draw reflection shoebox

☐ Show predictions ☐ Show hkl

☐ Show mask ☐ Basis vectors

☐ Indexed only ☒ Integrated only

Clear all

Stack type: sum

Image type: corrected

☒ Use dispersion extended algorithm

Sigma background: 6.0

Sigma strong: 3.0

Global Threshold: 0.0

Min. local: 2

SYMMETRY

likelihood	Z-CC	CC	N		Operator
0.936	9.93	0.99	694952	***	1 (0, 0, 0)
0.936	9.92	0.99	1334196	***	4 (0, 0, 1)
0.936	9.93	0.99	689472	***	2 (1, 0, 0)
0.936	9.92	0.99	667382	***	2 (0, 1, 0)
0.936	9.92	0.99	667388	***	2 (0, 0, 1)
0.936	9.92	0.99	667136	***	2 (1, 1, 0)
0.936	9.92	0.99	667218	***	2 (-1, 1, 0)



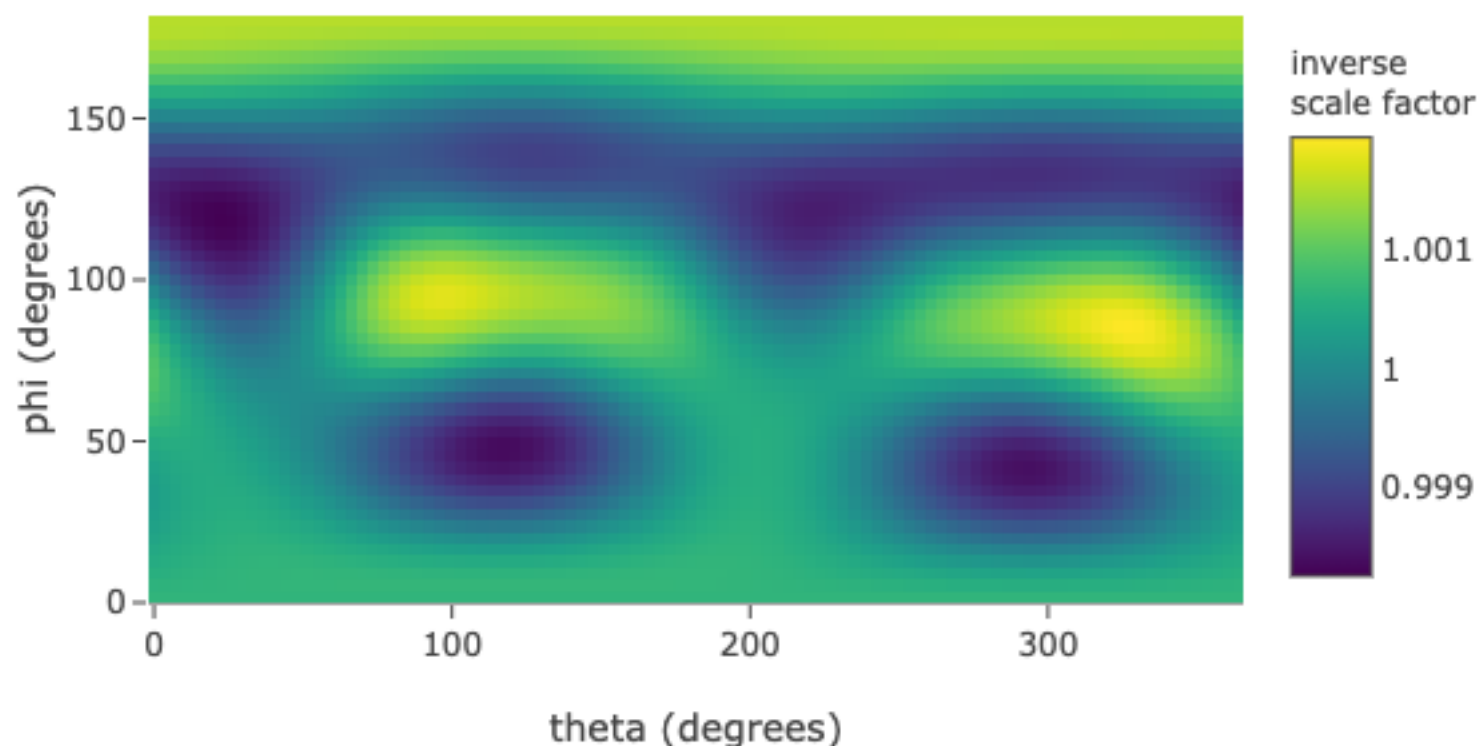
SYMMETRY

Patterson group		Likelihood	NetZcc	Zcc+	Zcc-	CC	CC-	delta	Reindex c
P 4/m m m	***	0.999	9.92	9.92	0	0.99	0	0	a,b,c
P m m m		0	0.01	9.92	9.92	0.99	0.99	0	a,b,c
C m m m		0	-0	9.92	9.92	0.99	0.99	0	-a+b,a+b,
P 4/m		0	0	9.92	9.92	0.99	0.99	0	a,b,c
P 1 2/m 1		0	0.01	9.93	9.92	0.99	0.99	0	a,b,c
P 1 2/m 1		0	0.01	9.93	9.92	0.99	0.99	0	b,a,-c
P 1 2/m 1		0	0	9.92	9.92	0.99	0.99	0	-b,c,-a
C 1 2/m 1		0	0	9.92	9.92	0.99	0.99	0	a+b,a-b,-
C 1 2/m 1		0	0	9.92	9.92	0.99	0.99	0	-a+b,a+b,
P -1		0	0.01	9.93	9.92	0.99	0.99	0	a,b,c

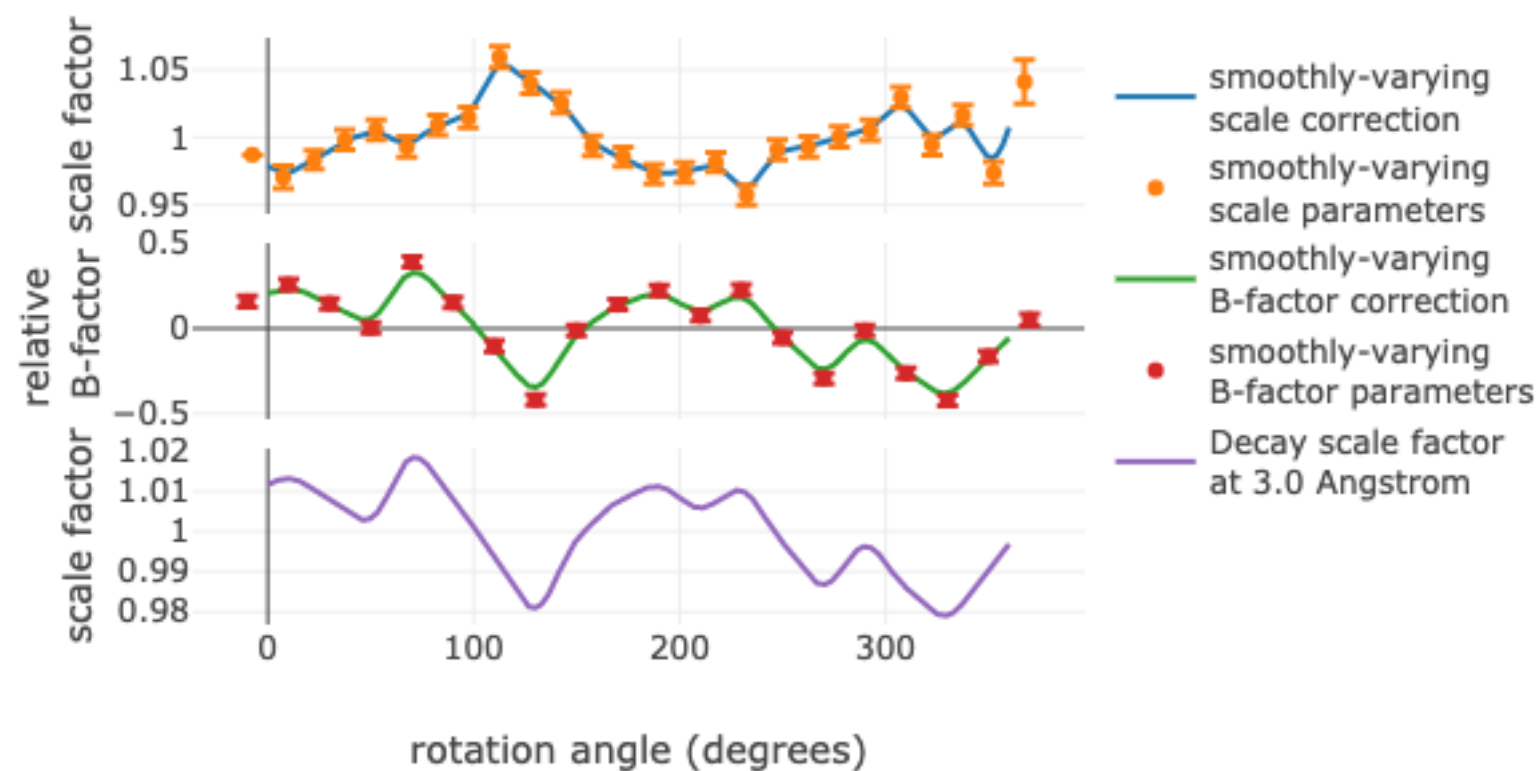


SCALING

Absorption correction surface (dataset 0)



Smoothly varying corrections (dataset 0)





MAP WITH SHELX C/D/E
(VIA FAST_EP)

DIALS.GITHUB.IO

XIA2.GITHUB.IO