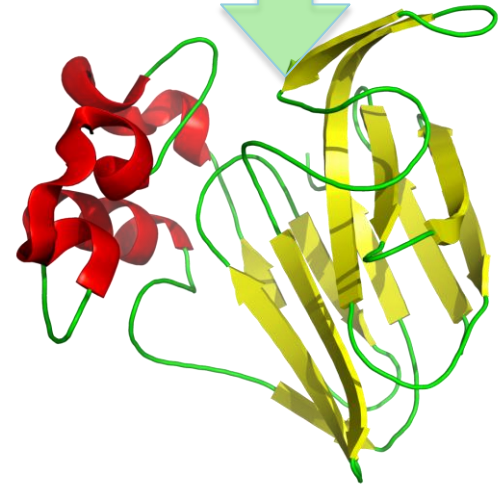
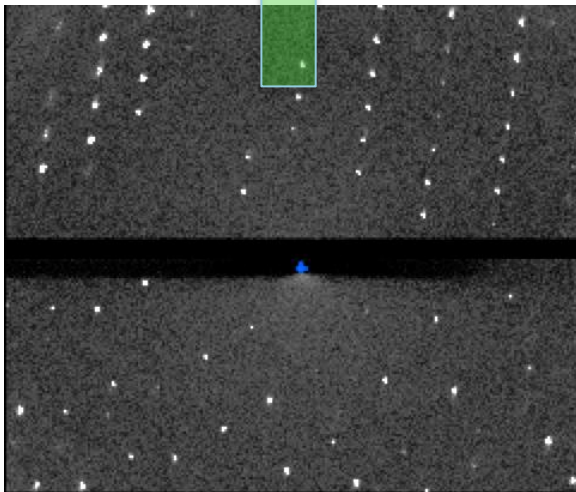


MX Pipelines and Automated Feedback at Beamlines

Joining the dots for good results

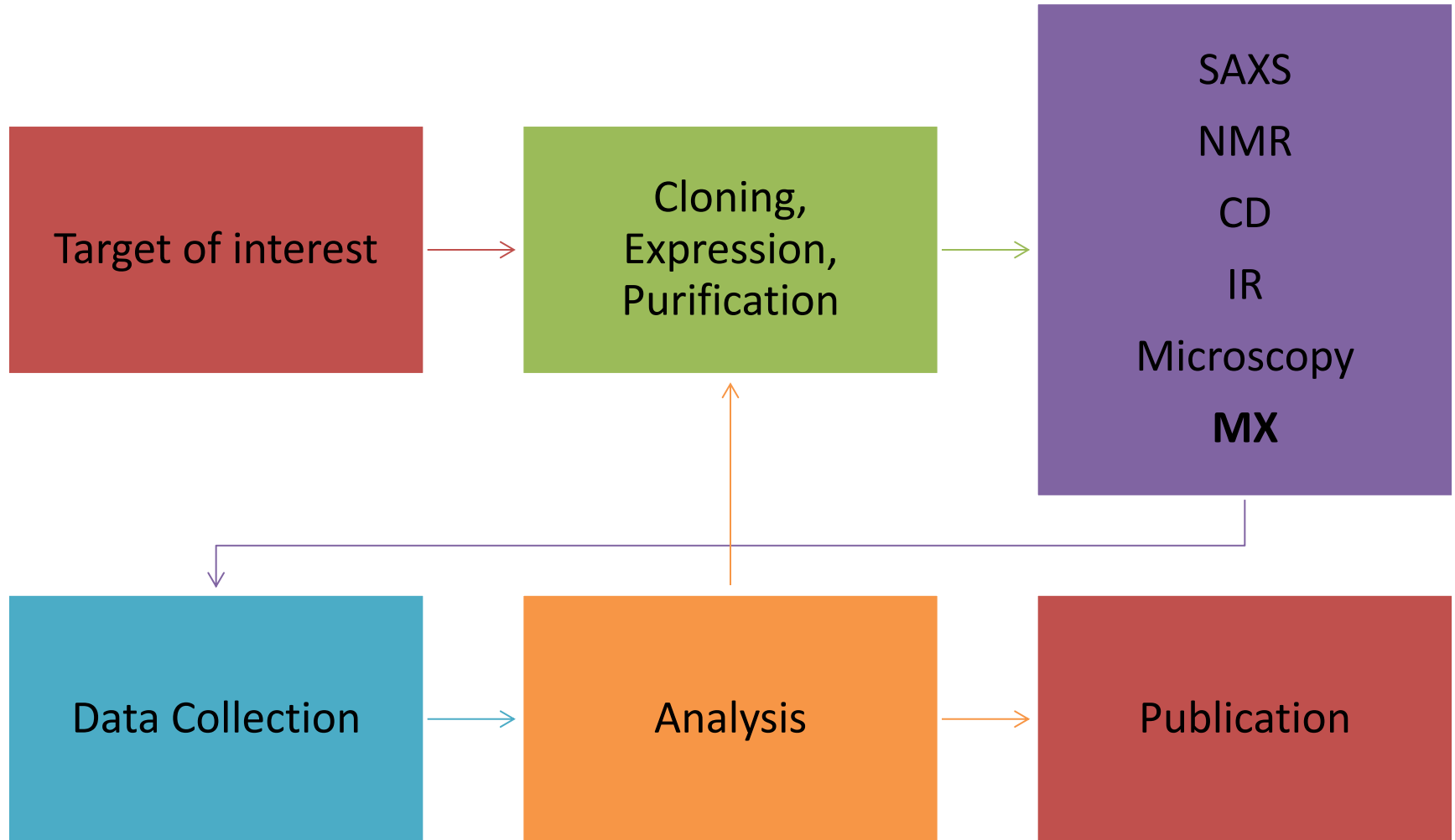
Dave Hall



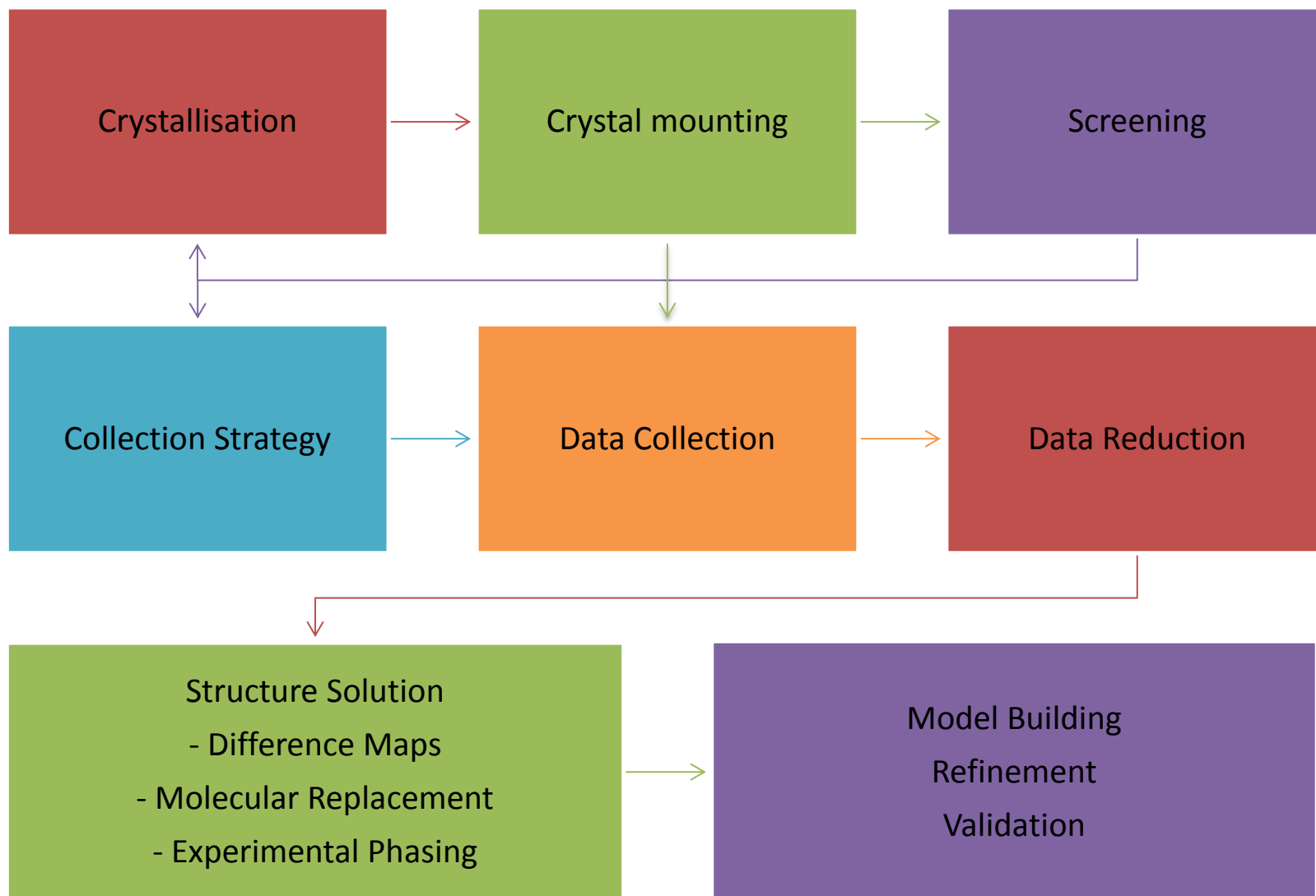
Acknowledgements

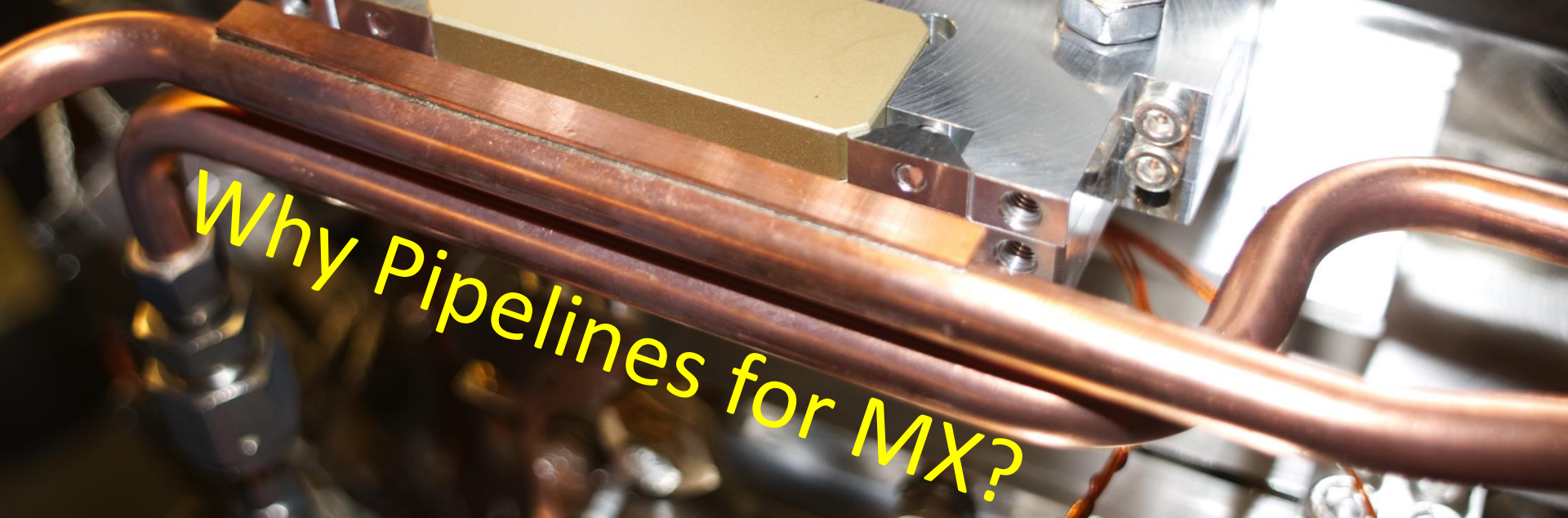
- Diamond MX Staff & Users
- Diamond scientific software and data acquisition groups
- IT / sys admins (clusters, file systems)
- DIALS development team
- Collaborations e.g. EDNA, CCTBX, DIALS, ISPyB, ...
- External software developers - XDS, Mosflm, CCP4, SHELX, ...

The Structural Biology Pipeline



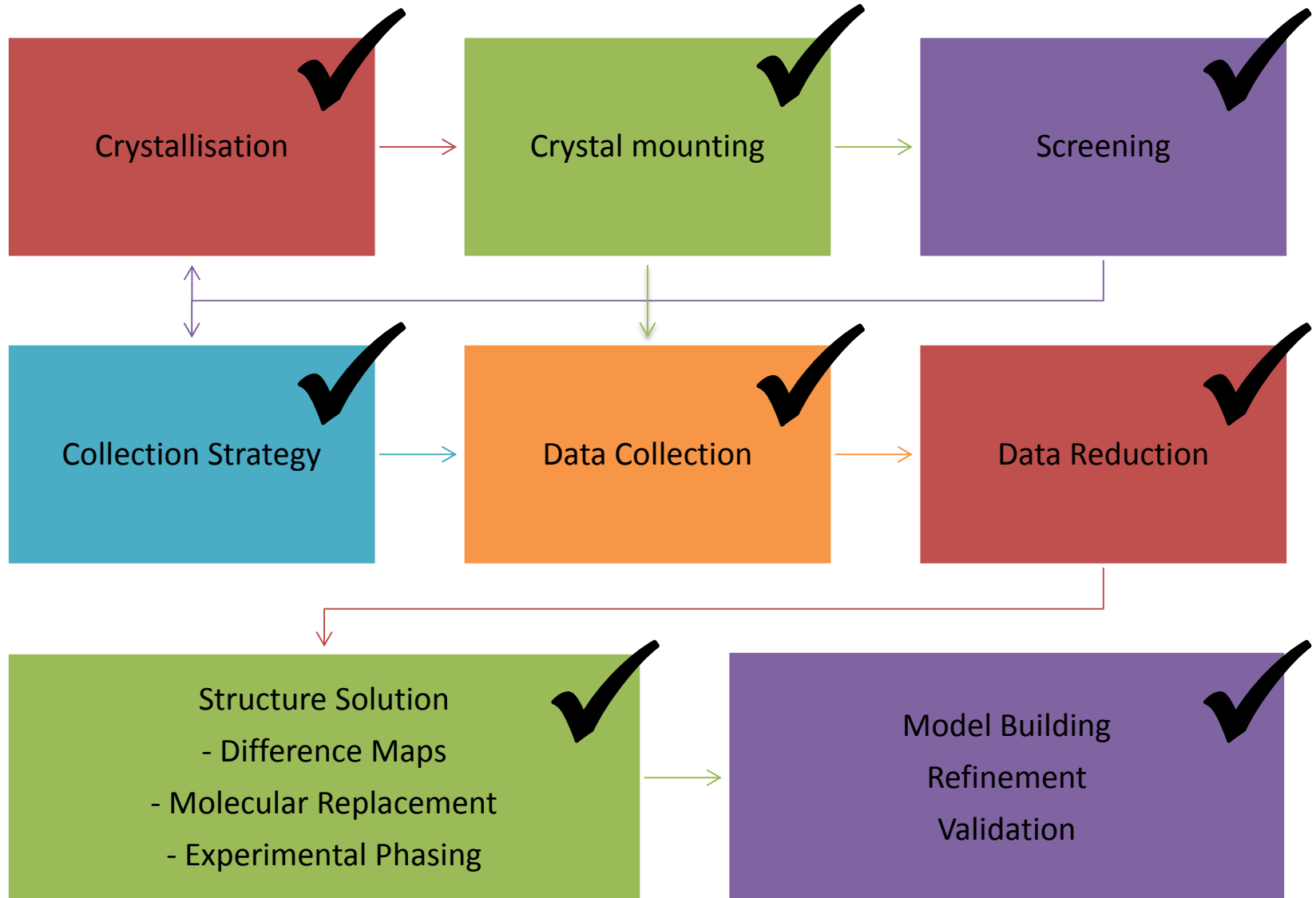
MX – A Pipeline within a Pipeline



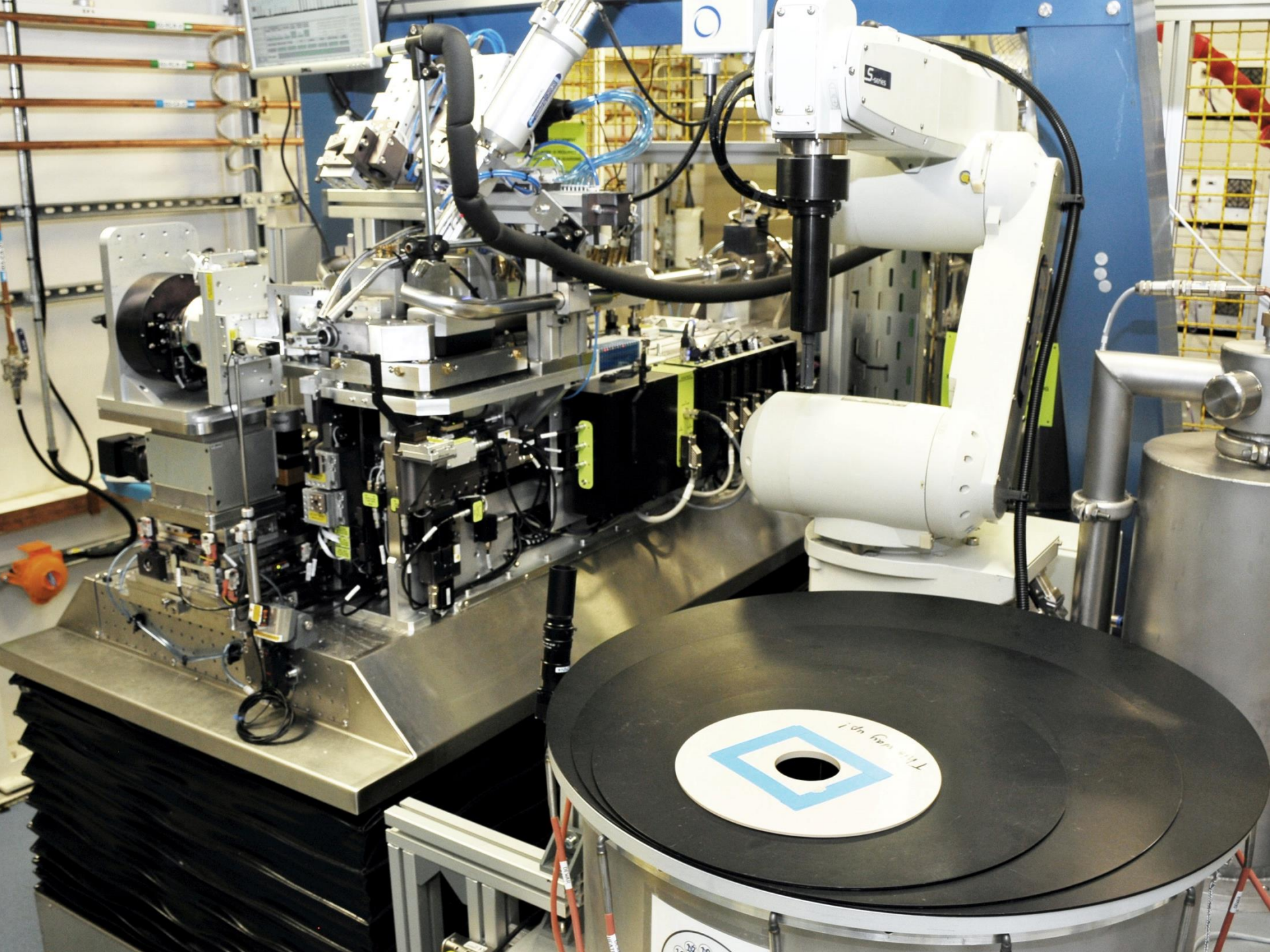


- Defined processes
- Data Deluge
- Experimenter overload
- Better than you?
- Faster than you?
- Valuable feedback (near real time?)
- Focus on science
- Doesn't sleep/need a life?

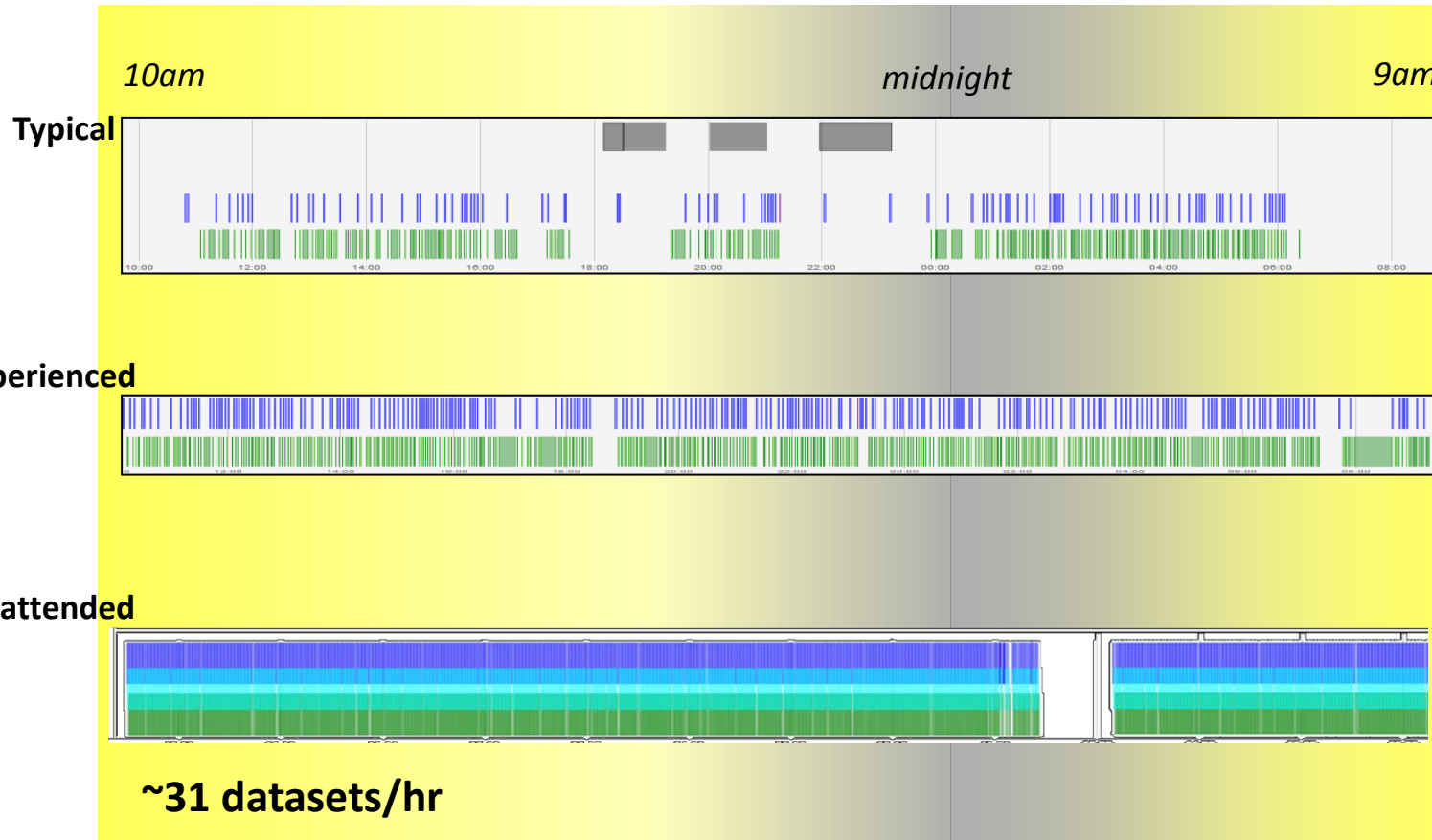
What can be automated if money and time no object



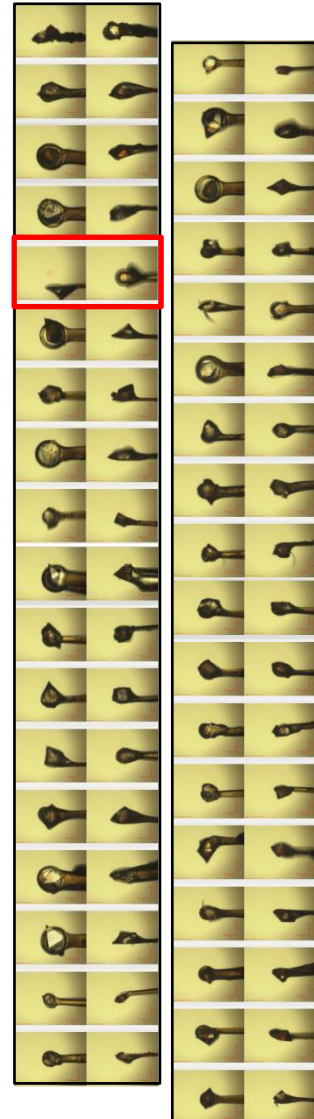
BEAMLINE AUTOMATION HELPS



How fast can you go?



Sample auto-centring: >97%



NB Faster is not always better



08-05-2016 00:48:11 - ▶ SPIN4/SPIN4-x0131/SPIN4-x0131_1_####.cbf

00:48:11

Sample: SPIN4-x0131

Flux: NaN

 Ω Start: 0.0° Ω Osc: 0.12° Ω Overlap: 0°

No. Images: 1500

Resolution: 2.10Å

Wavelength: 0.9282Å

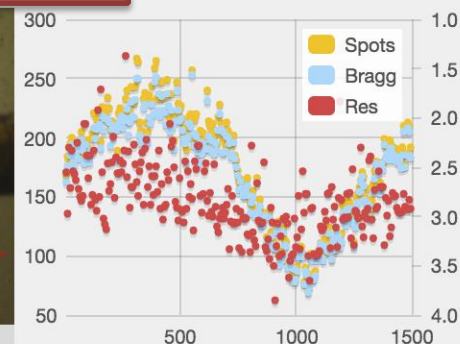
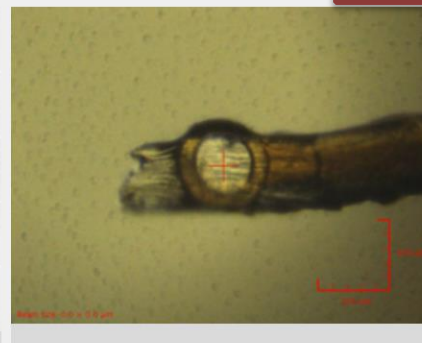
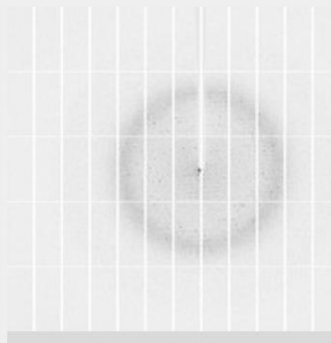
Exposure: 0.060s

Transmission: 100.00%

Beamsize: 0x0µm

Type: SAD

Comment: (119923,-109,100) Aperture: 70µm



Auto Processing

Fast DP: ✓ Xia2: ? ✓ ✓

Downstream Processing

Fast EP: ? Dimple: ? MrBUMP: ? Big EP: ?

08-05-2016 00:47:17 - Robot loading puck 16 pin 7 (Barcode: DF075C1002) Status: SUCCESS - OK (Took 22s) Sample: SPIN4-x0131

08-05-2016 00:45:42 - ▶ SPIN4/SPIN4-x0130/SPIN4-x0130_1_####.cbf

00:45:42

Sample: SPIN4-x0130

Flux: NaN

 Ω Start: 90.0° Ω Osc: 0.12° Ω Overlap: 0°

No. Images: 1500

Resolution: 2.10Å

Wavelength: 0.9282Å

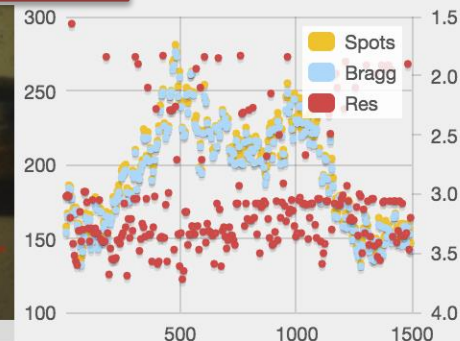
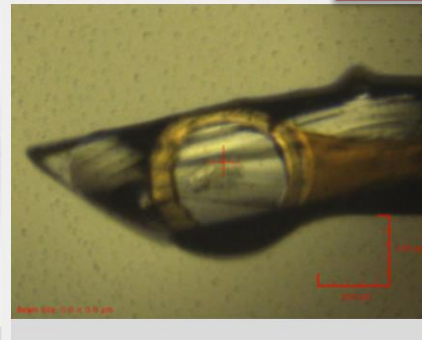
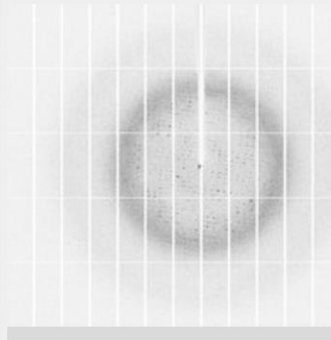
Exposure: 0.060s

Transmission: 100.00%

Beamsize: 0x0µm

Type: SAD

Comment: (119777,123,-119) Aperture: 70µm

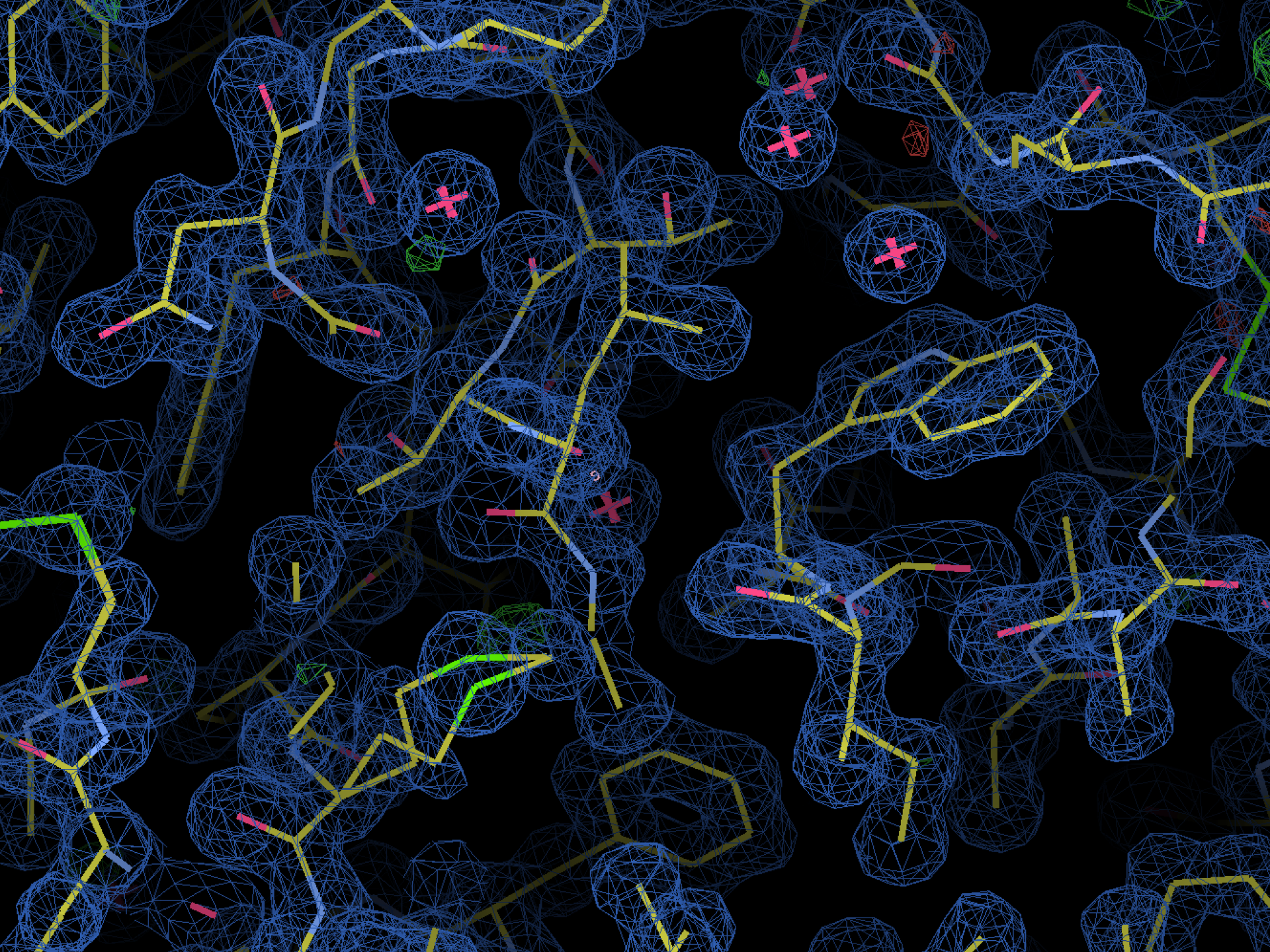


Auto Processing

Fast DP: ✓ Xia2: ? ✓ ✓

Downstream Processing

Fast EP: ? Dimple: ? MrBUMP: ? Big EP: ?



What are you looking for?

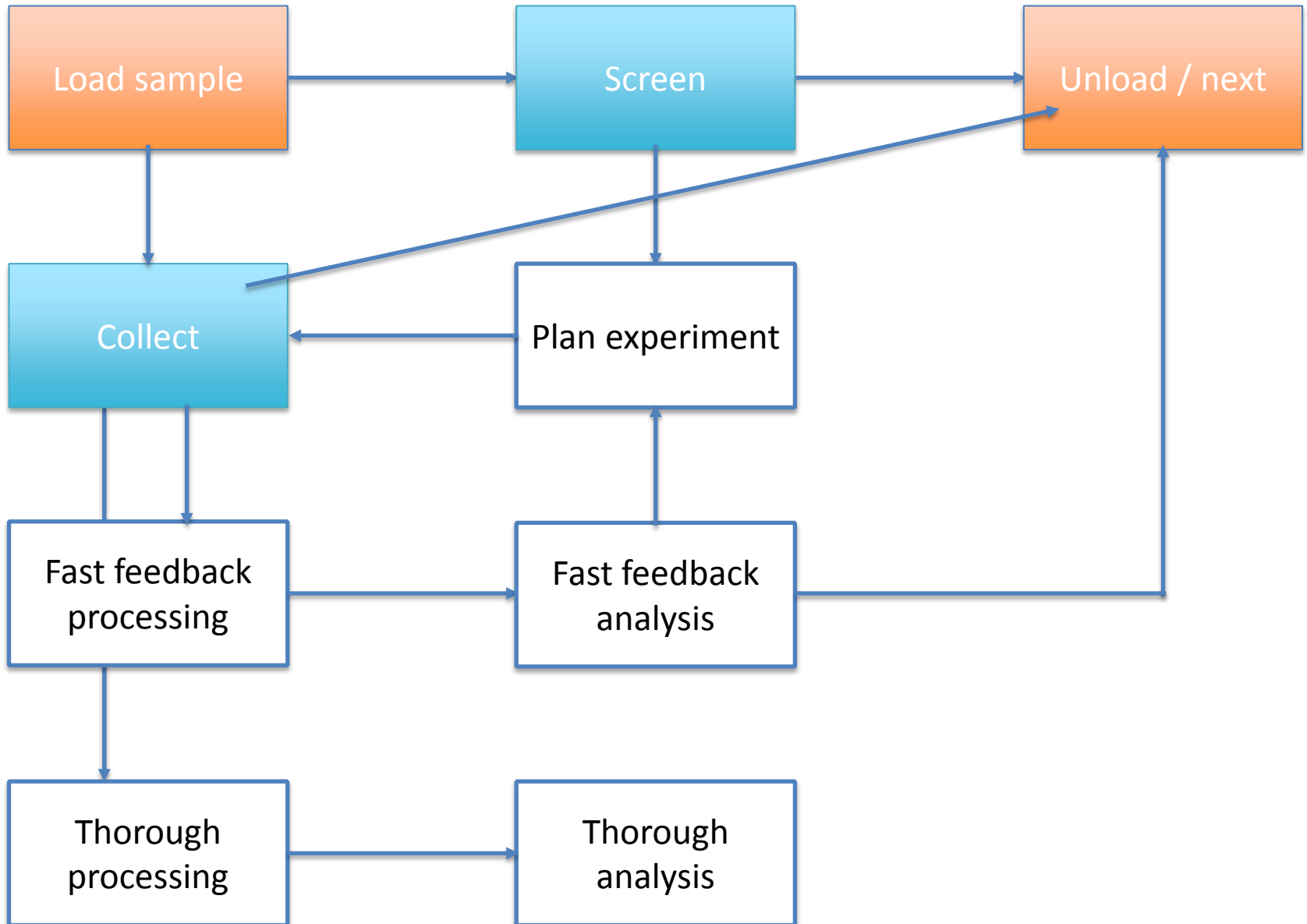
Diffraction data? No!

Science!

- Novel structure
- Higher resolution
- Ligand binding
- Effect of mutation
- Chemistry
- Biology
- ...

But to get this you need *good data*
To see this during experiment - *automation*

Data collection pipeline



EDNA: Collaboration With ESRF

Mosflm: Andrew Leslie

XOAlign: Pierre Legrand

PLANNING YOUR EXPERIMENT

“Screening”



Standard Native Dataset Multiplicity=3 I/sig=2 Maxlifespan=807 s

Gentle: Target Multiplicity=2 and target I/Sig 2 and Maxlifespan=81

Standard Anomalous Dataset Multiplicity=3 I/sig=2 Maxlifespan=807 s

strategy with target multiplicity=16, target I/sig=2 Maxlifespan=807 s

Strategy	Description	Ω Start	Ω Osc	Res (Å)	Rel Trn (%)	Abs Trn (%)	Exposure (s)	No. Images
native	null	109	1.40	1.45	0.0	0	0.000	65
anomalous	null	109	1.40	1.45	0.0	0	0.000	65

Kappa Alignment Suggestion

EDNA

Strategies

Mosflm:

EDNA:

XOalign

Aligned Axes

Kappa

Phi

[(a*, c*), (a*, b*)]

93.912

284.211

EDNA

Space Group

A

B

C

α

β

γ

P3

92.39

92.39

128.11

90.00

90.00

120.00

Strategy

Description

Ω Start

Ω Osc

Res (Å)

Rel Trn (%)

Abs Trn (%)

Exposure (s)

No. Images

Strategy4

Gentle: Target Multiplicity=2 and target I/Sig 2 and Maxlifespan=10

165

0.15

2.06

34.2

34.2

0.040

615

Strategy2

Standard Anomalous Dataset Multiplicity=3 I/sig=2 Maxlifespan=98 s

84

0.15

1.92

100.0

100

0.060

1260

Strategy1

Standard Native Dataset Multiplicity=3 I/sig=2 Maxlifespan=98 s

86

0.15

1.87

100.0

100

0.116

628

Strategy3

strategy with target multiplicity=16, target I/sig=2 Maxlifespan=98 s

0

0.15

1.87

76.4

76.4

0.040

2400

Mosflm

Space Group

A

B

C

α

β

γ

P3

92.20

92.20

128.23

90.00

90.00

120.00

Strategy

Description

Ω Start

Ω Osc

Res (Å)

Rel Trn (%)

Abs Trn (%)

Exposure (s)

No. Images

native

null

82

0.70

1.60

0.0

0

0.000

129

anomalous

null

82

0.70

1.60

0.0

0

0.000

129

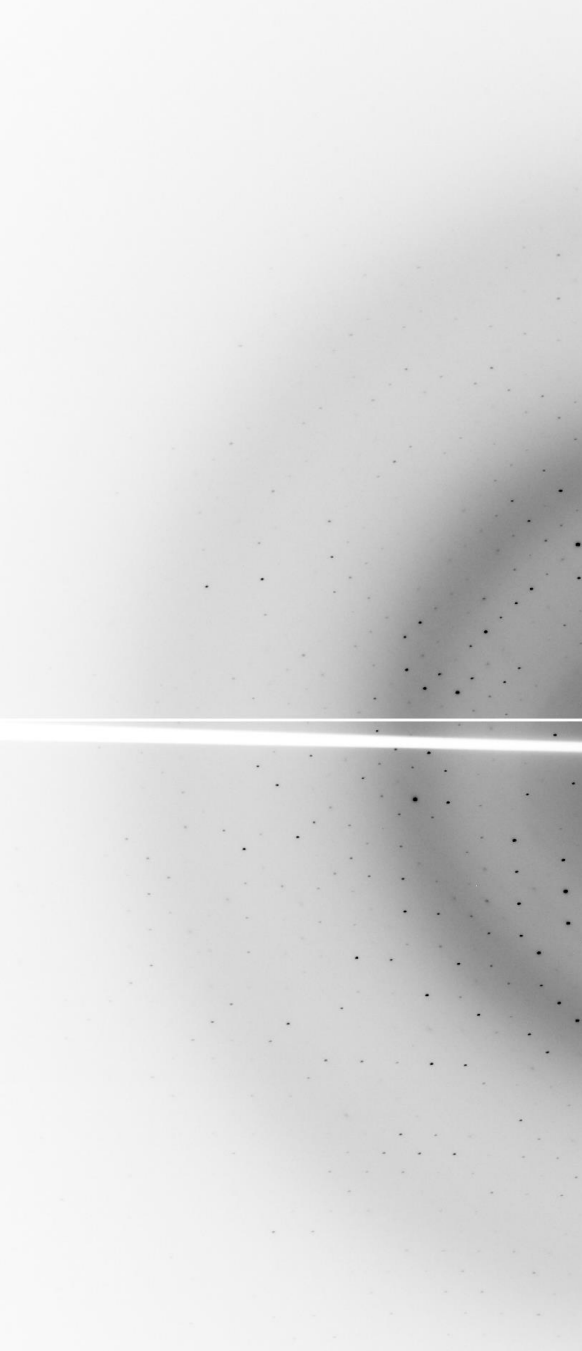
Mosflm “index strategy autocomplete go”

Strategies can be found and applied in the data collection UI

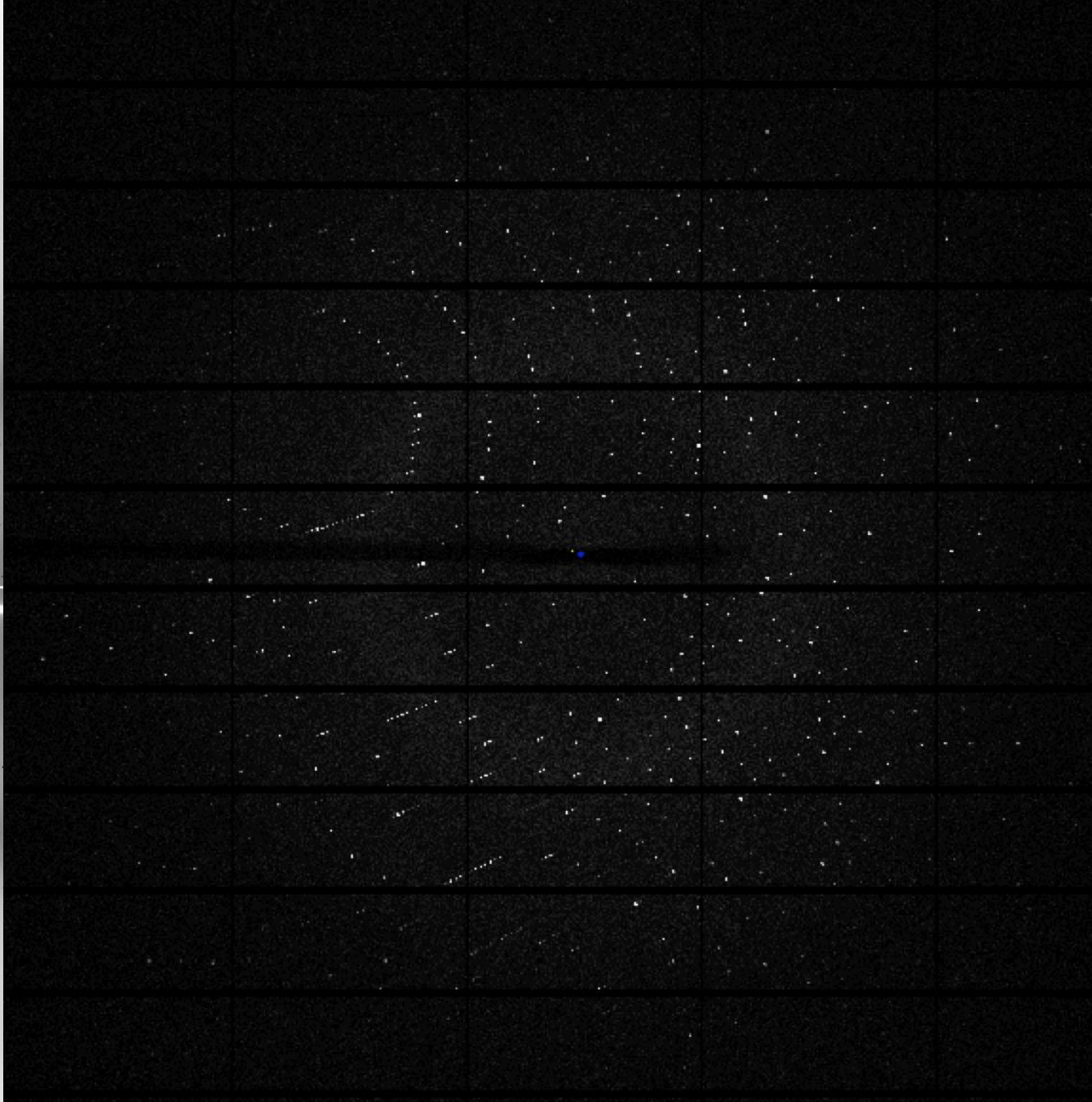
DISTL: SSRL / LBNL

DIALS: Collaboration with LBNL & CCP4

**LOOKING AT THE IMAGES (FOR SIGNS
OF RADIATION DAMAGE AND OTHER
PATHOLOGIES)**



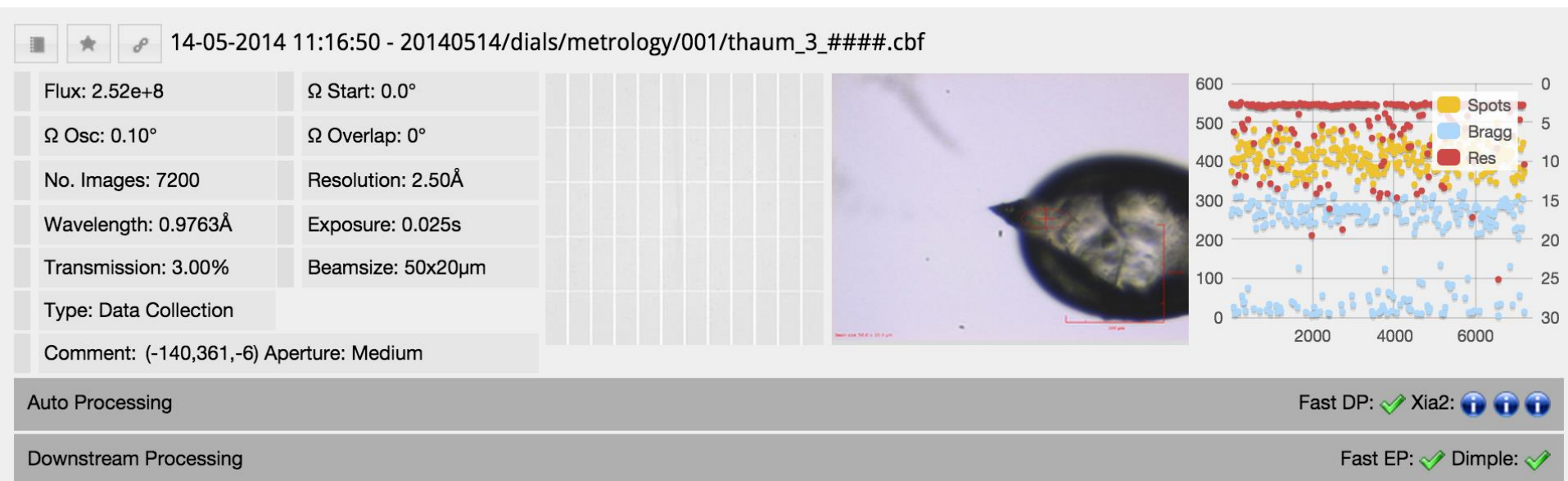
2005: 90 - 180 x 5-30s

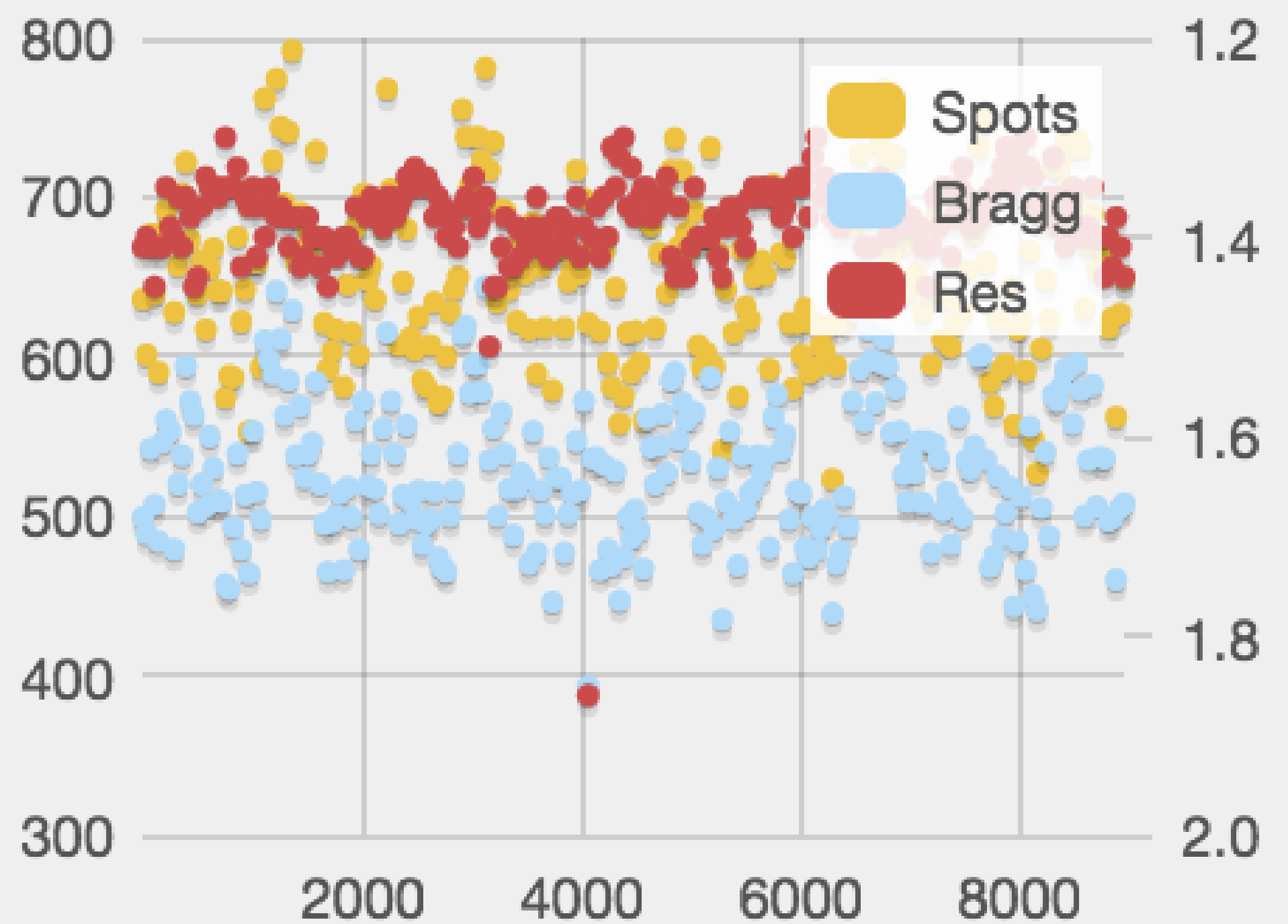


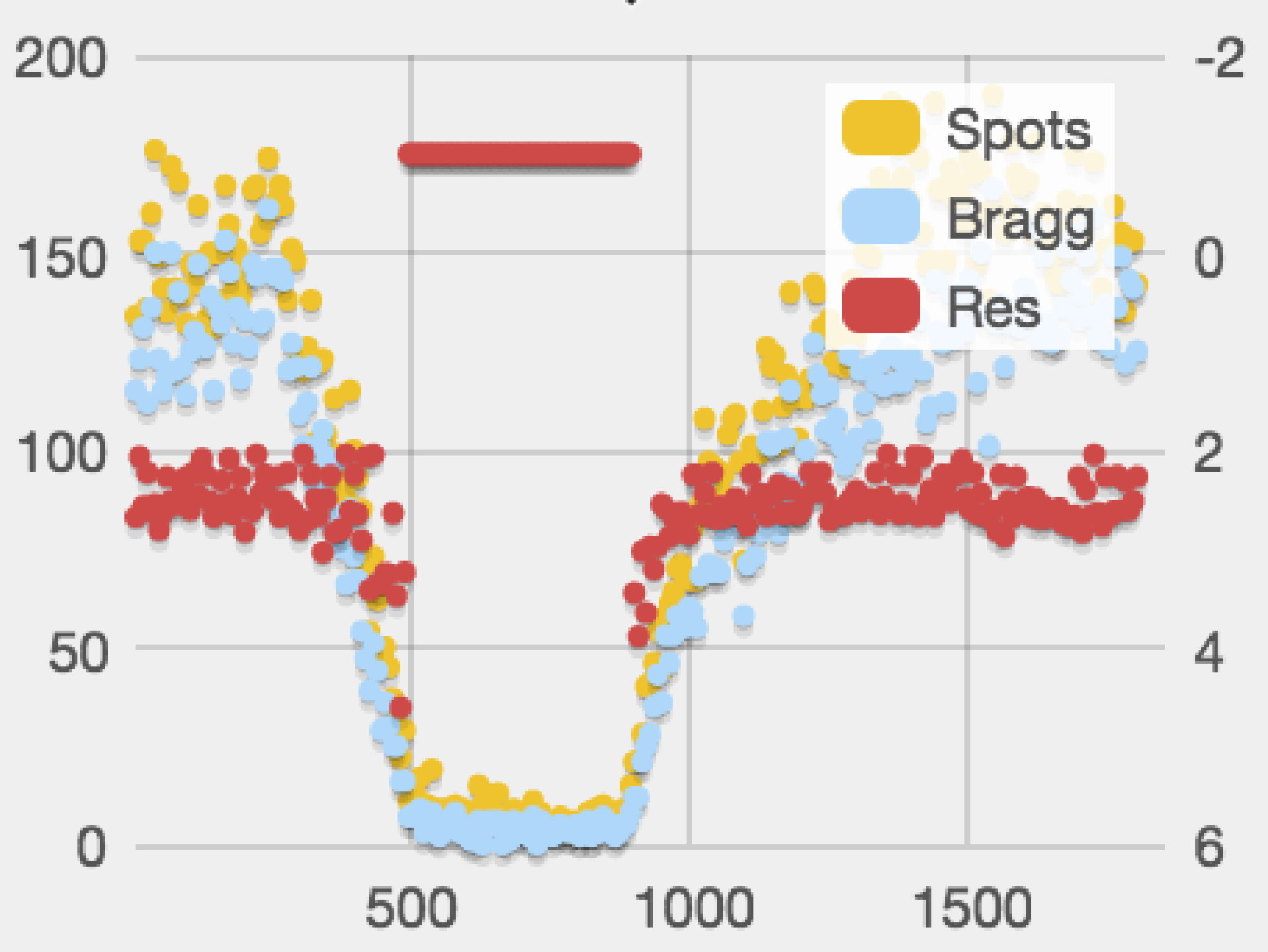
2015: 900 - 36,000 x 5-100/s

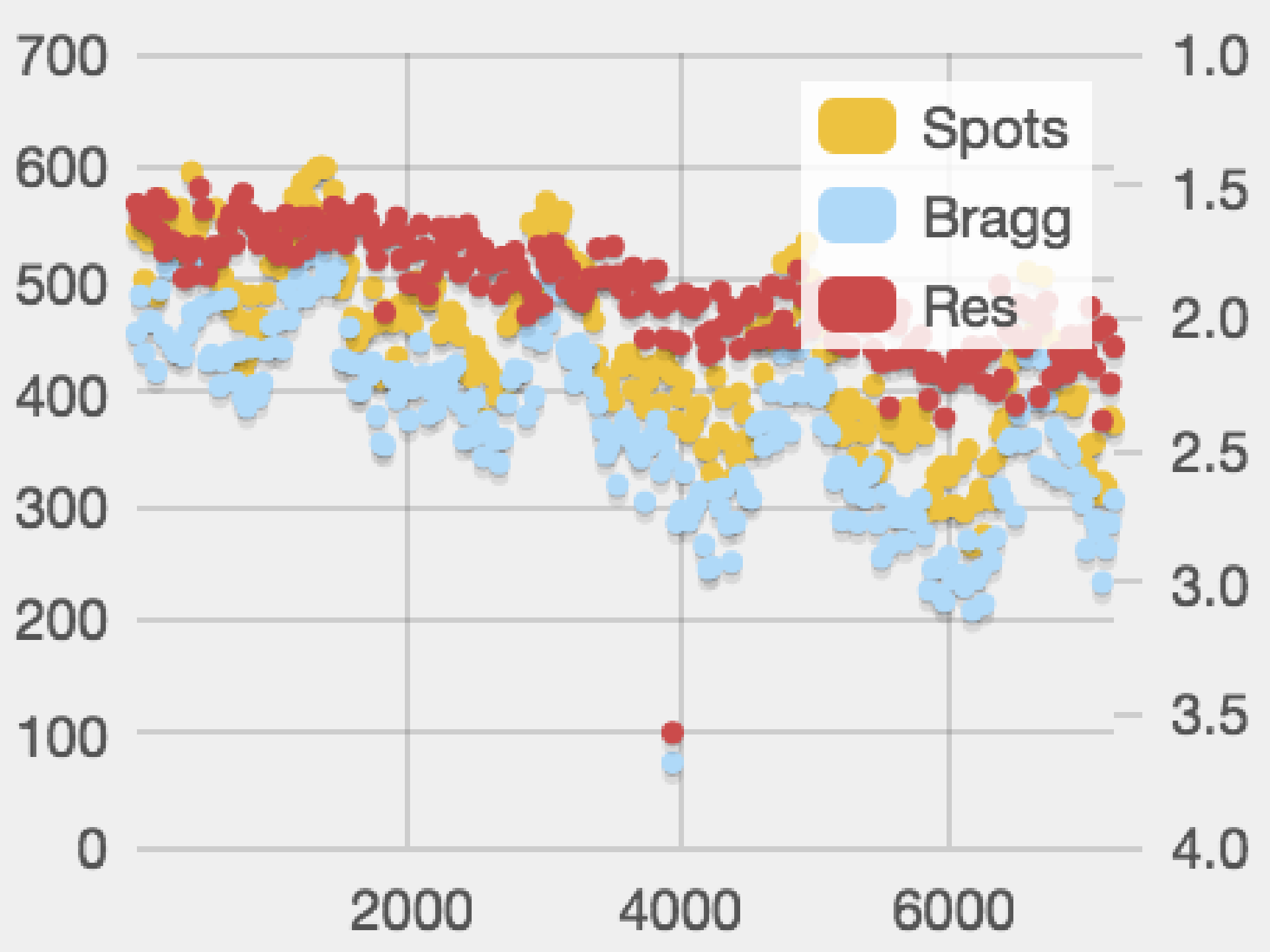
Data deluge

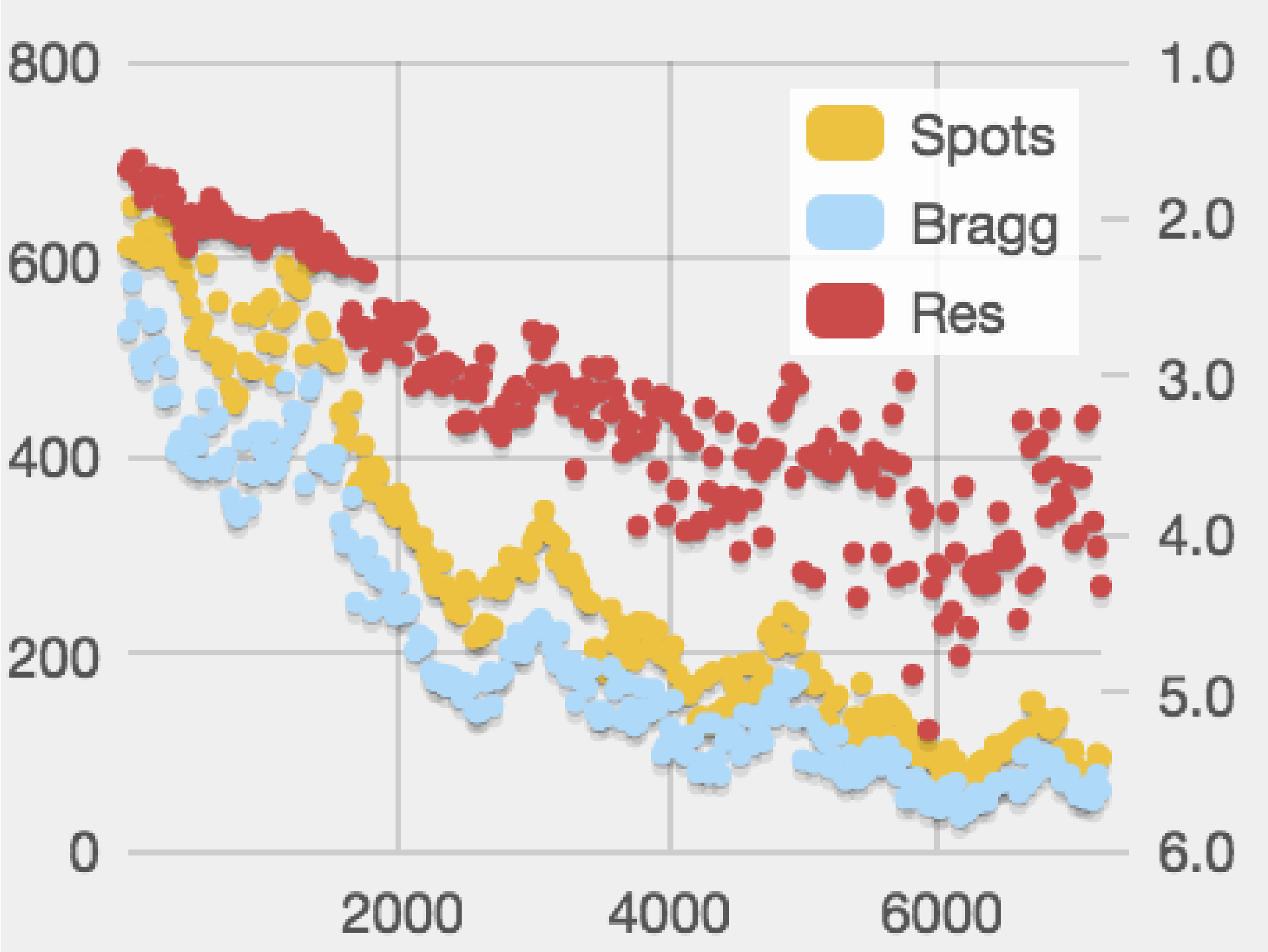
- Inspect individual images?
 - Challenging – may be necessary if all else fails
- Automated image analysis
 - Looks for spots on the images; estimates resolution; powered by DIALS
 - Over whole data set builds up (useful) story of what happened









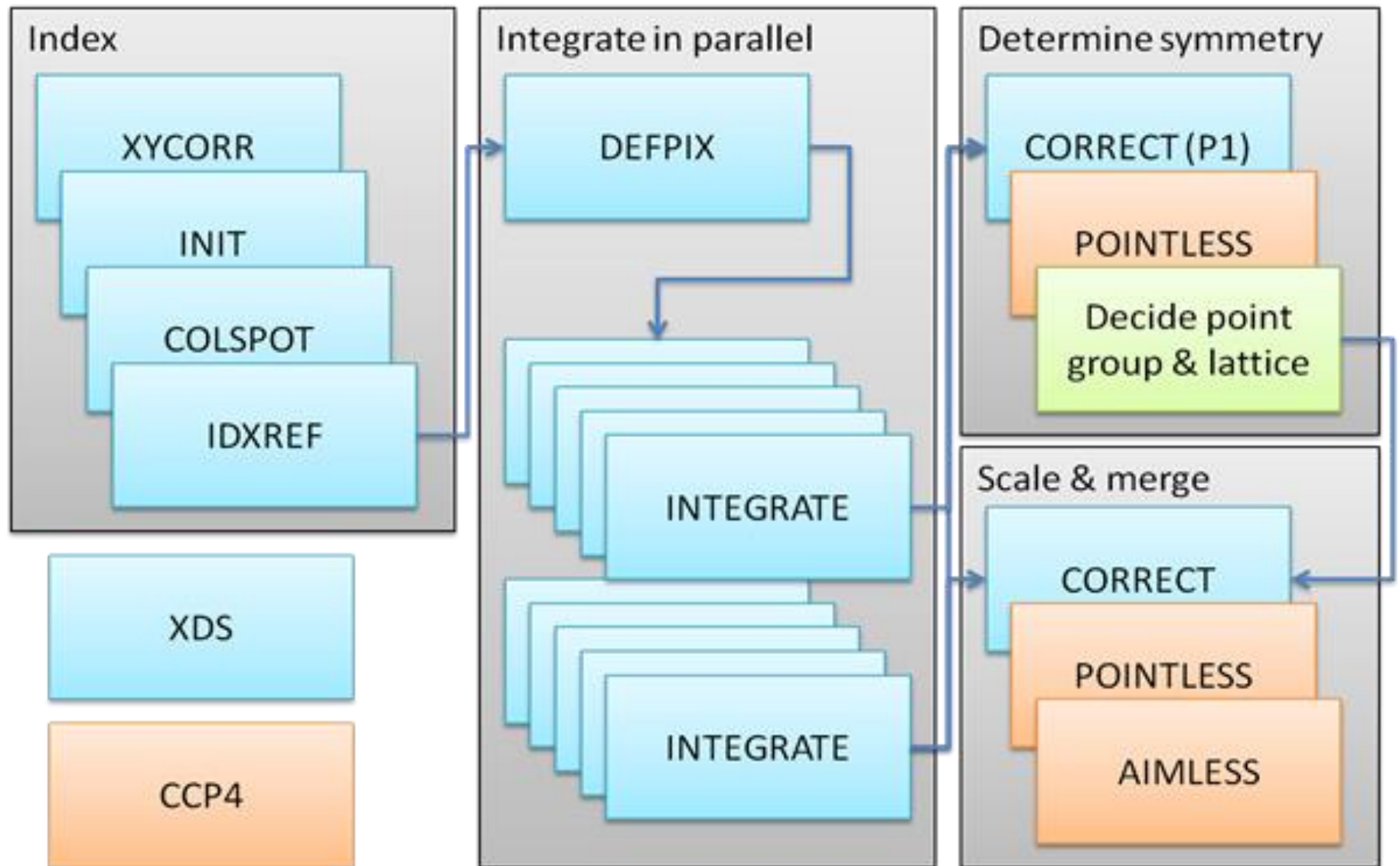


fast_dp: uses XDS extensively

xia2: uses XDS, Mosflm, CCP4 & DIALS

DID I GET WHAT I NEED?

Data reduction pipelines – Fast DP



Low resolution	28.91	28.91	1.40
High resolution	1.36	6.08	1.36
Rmerge	0.052	0.017	0.673
I/sig	0.10	155.50	5.10
Completeness	99.9		99.4
Multiplicity			24.2
Anom.		92.9	
Anom. Mult.		99.4	
Anom. Correlat.		14.4	12.9
Nr		7.8	4.9
Nuni		5	13840
Mid-sl	8174		572
	1.748		
	0.059		
dI/sig(I)	2.888		

~ 60 s after data collection

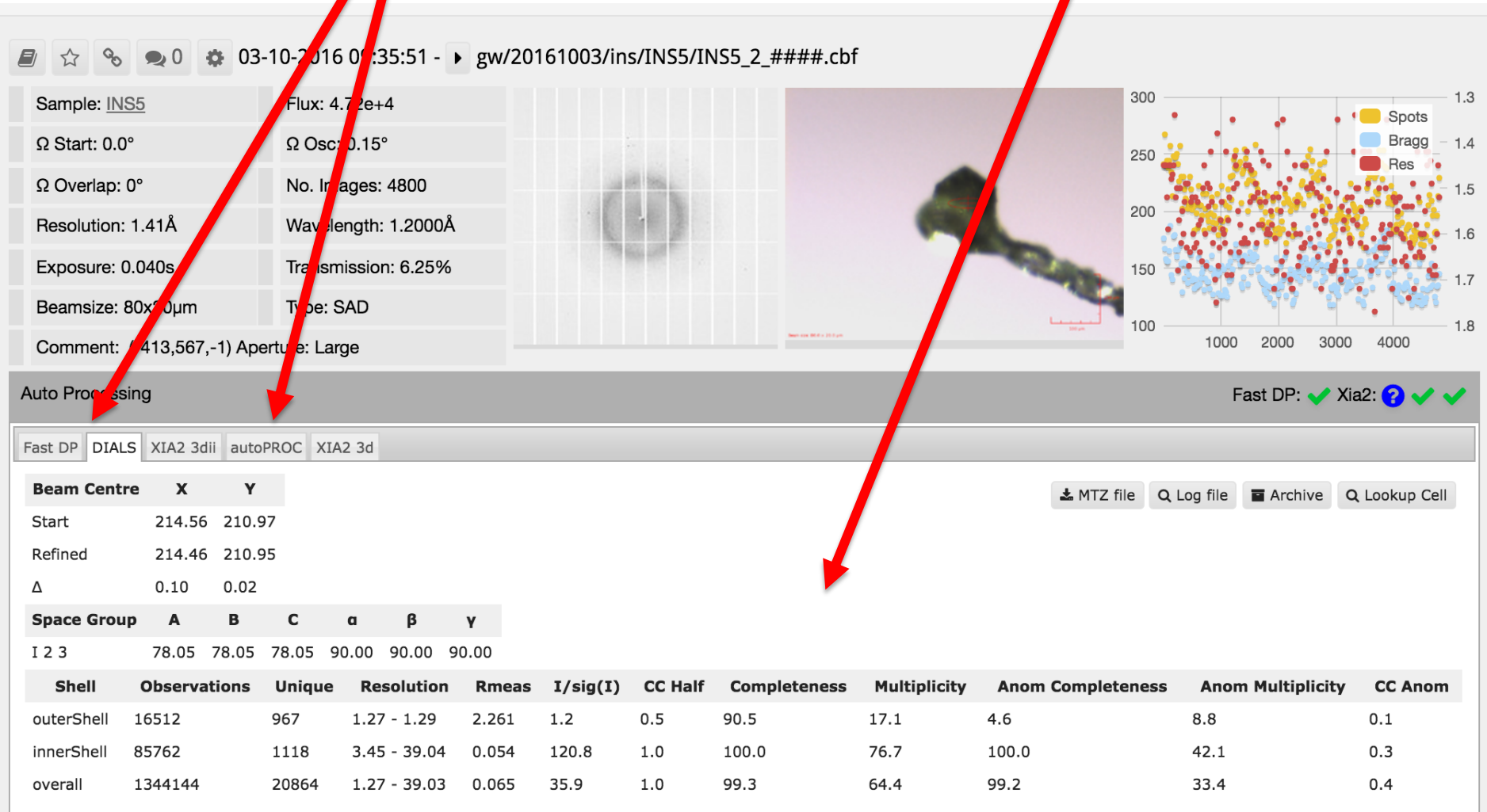
Thorough processing pipelines

- xia2 using XDS / XSCALE
- xia2 using DIALS / AIMLESS
- Multi-xia2 tasks taking all sets from sample (inverse beam, MAD, ...)
- AutoPROC using XDS
- 10 – 30 minutes after data collection

Where to find feedback

Range of processing programs

Merging statistics



DOWNSTREAM PROCESSING

Automatic Downstream Processing

Fast feedback:

- Fast EP
- DIMPLE

Thorough feedback:

- MrBUMP
- Big EP

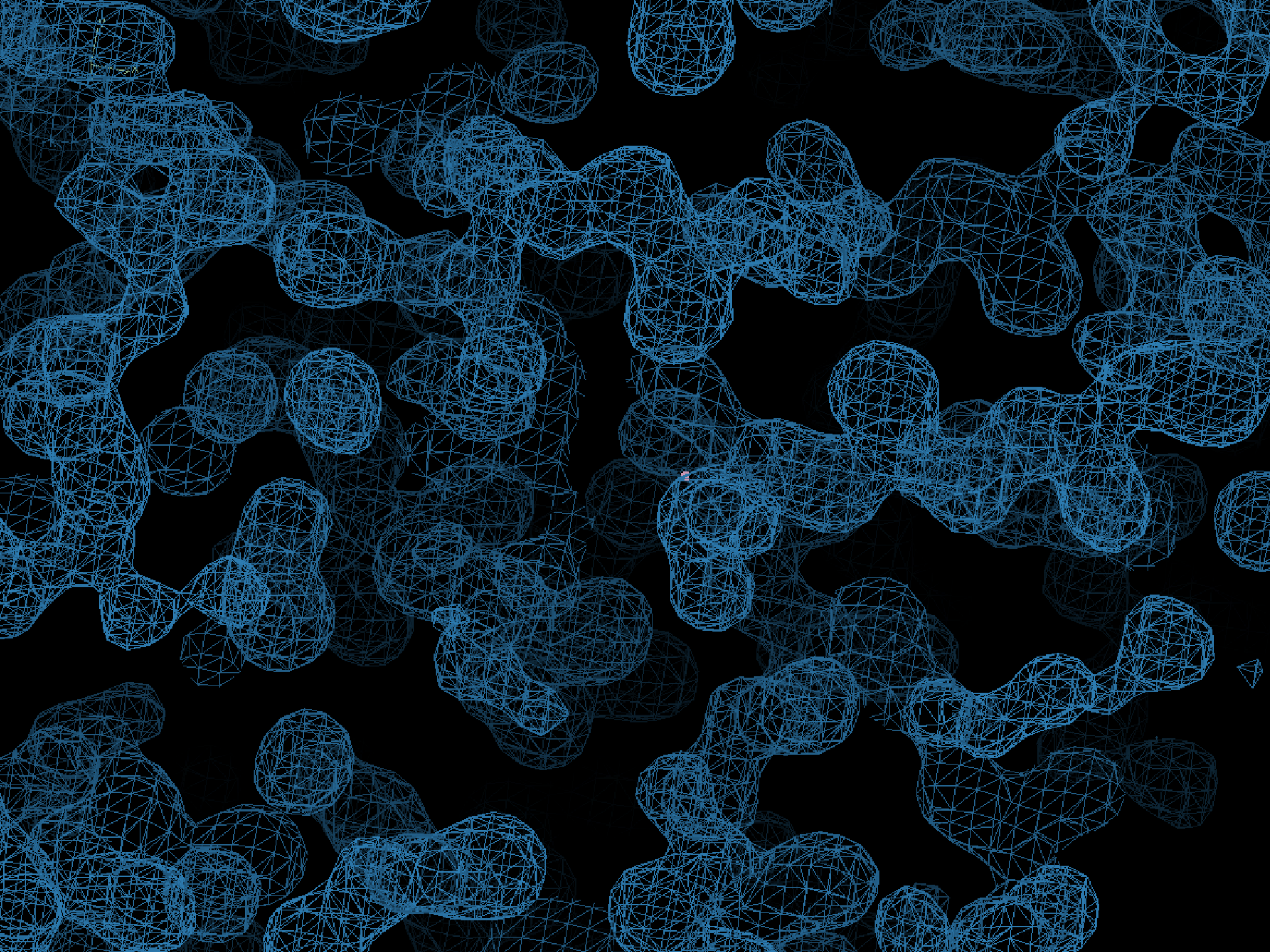
Fast EP

Using 12 cpus / 40 machines

```

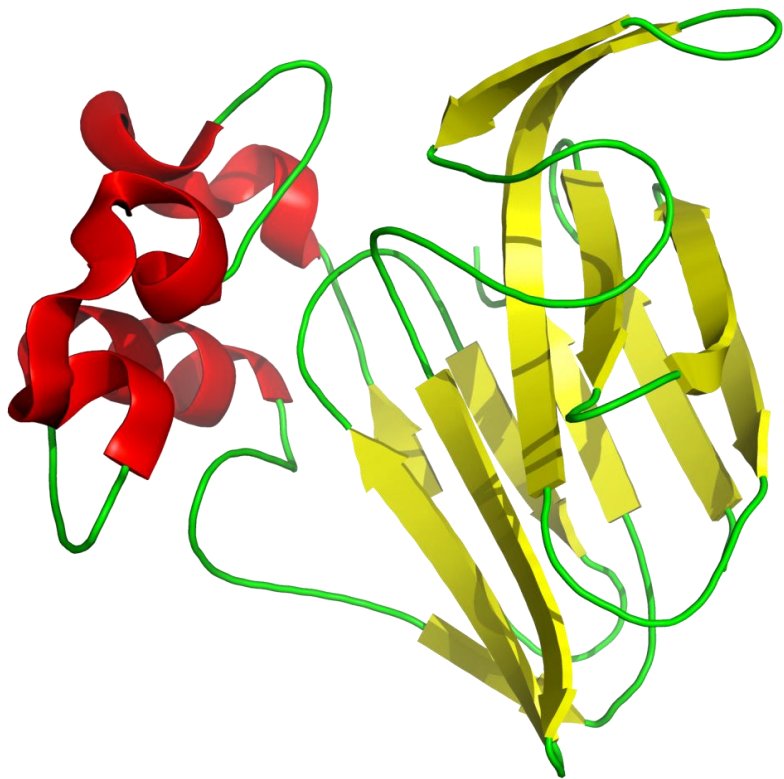
Input: /d1s/104/data/2015/mx13273-3/processed/bag_training/dna1130/PNAS_M3S5_1_/fast_dp/fa
Columns: I( Running 12 x shelxd_mp jobs
Unit cell: 42 Spacegroup: P422
N try: 20 No. CCall CCweak CFOM No. found
Pointgroup: P 1 20.82 6.53 27.36 0
Resolution: 28 2 33.56 21.24 54.80 4
Nrefl: 81 Spacegroup: P4212
DF/F: 0. No. CCall CCweak CFOM No. found
dI/sig(dI): 2. 1 18.92 5.58 24.50 2
Spacegroups: P4 2 26.69 15.36 42.05 4
SHELXC summary: Spacegroup: P4122
Dmin <I/sig> No. CCall CCweak C Running 22 x shelxe jobs
3.02 161.50 1 32.45 16.83 4 Solv. Orig. Inv.
2.40 91.60 2 33.50 20.18 5 0.25 0.482 0.797 (best)
2.10 64.40 Spacegroup: P41212 ( 0.30 0.497 0.796
1.91 40.10 No. CCall CCweak C 0.35 0.496 0.794
1.77 29.00 1 44.87 26.79 7 0.40 0.505 0.793
1.66 20.90 2 45.26 26.20 7 0.45 0.478 0.791
1.58 15.00 Spacegroup: P4222 0.50 0.425 0.789
1.51 11.10 No. CCall CCweak C 0.55 0.306 0.783
1.45 8.80 1 19.27 8.69 2 0.60 0.237 0.776
1.40 6.80 2 34.78 21.35 5 0.65 0.228 0.764
1.36 5.30 Spacegroup: P42212 0.70 0.236 0.753
No. CCall CCweak C 0.75 0.242 0.739
1 22.65 15.65 3
2 26.50 18.41 4 Best solvent: 0.25
Time: 9.01 Best hand: inverted
Best spacegroup: P41 Best spacegroup: P43212
Best nsites: 1 Time: 8.07
Best CC / weak: 44.87 / 26.79

```

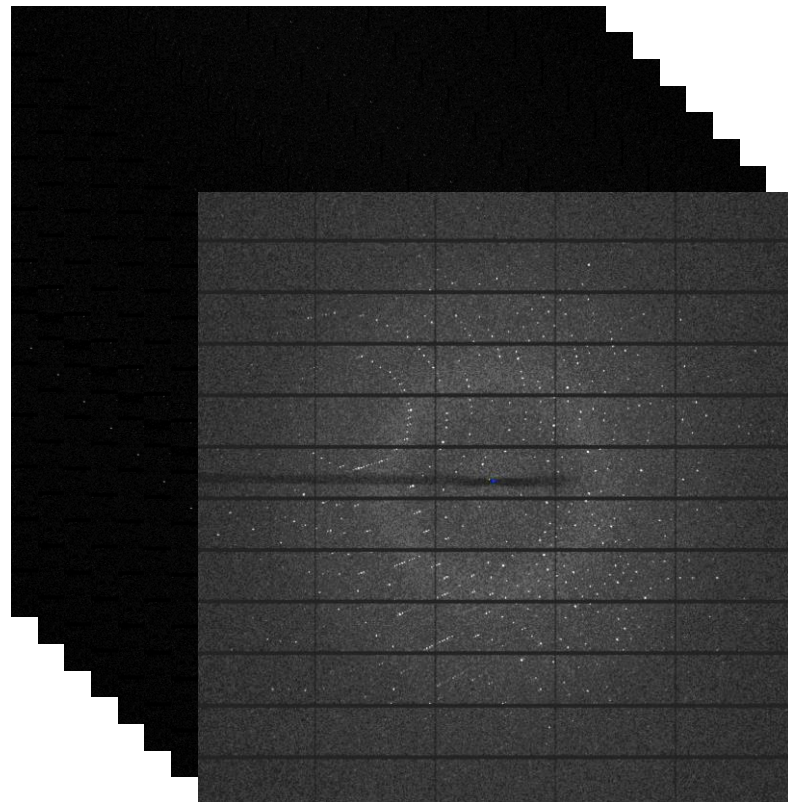


DIMPLE: CCP4 Collaboration

DO I HAVE A LIGAND?

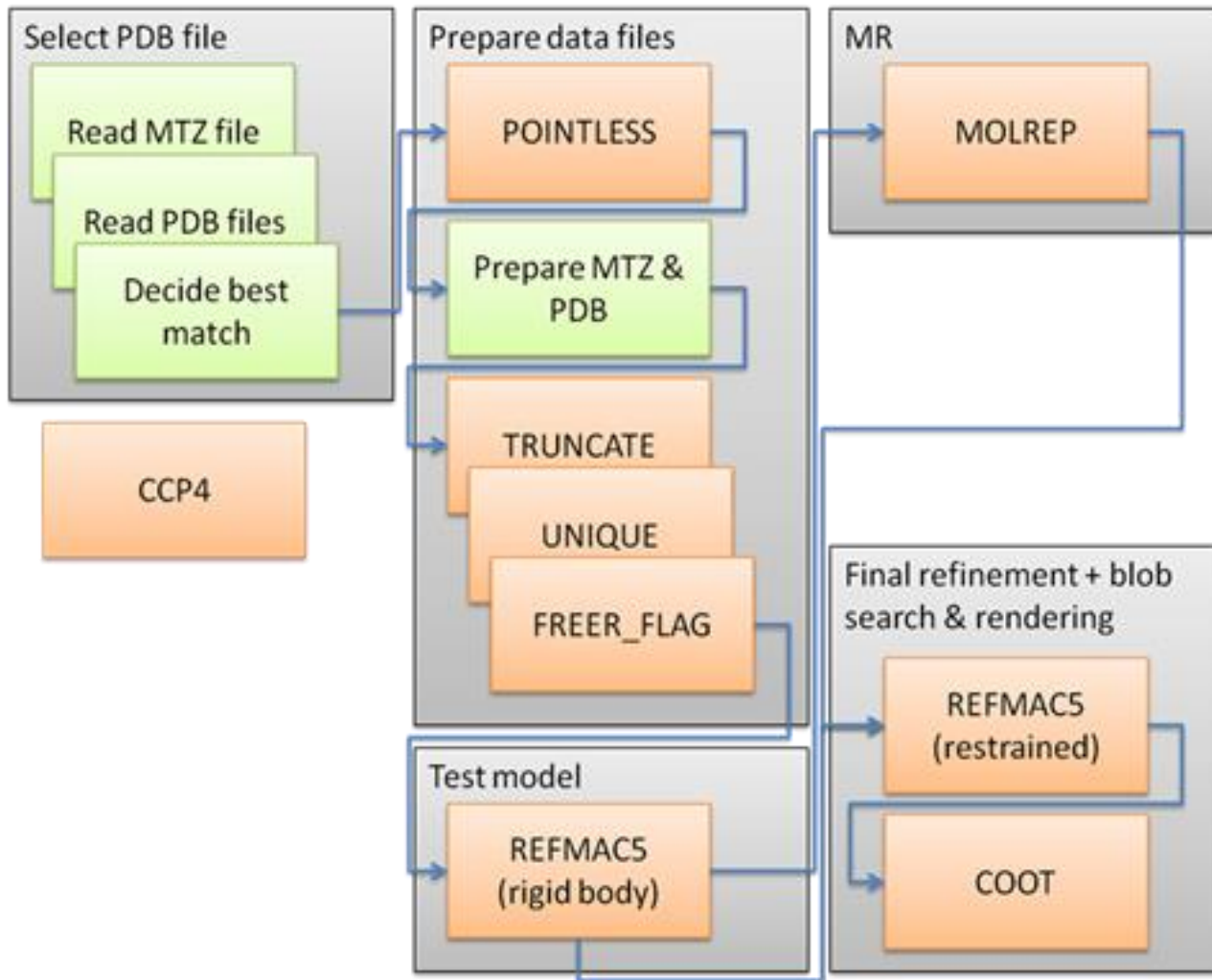


+

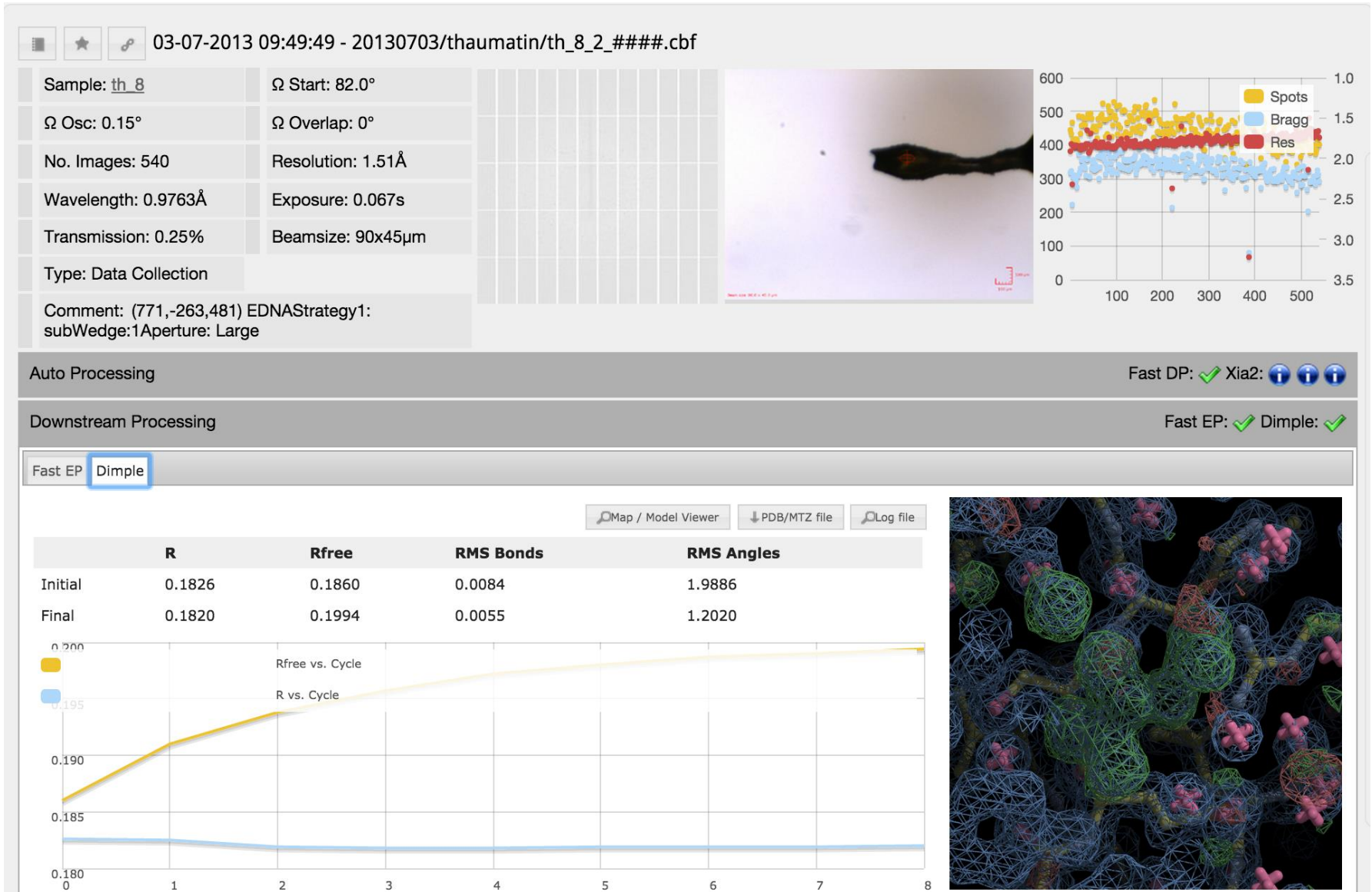


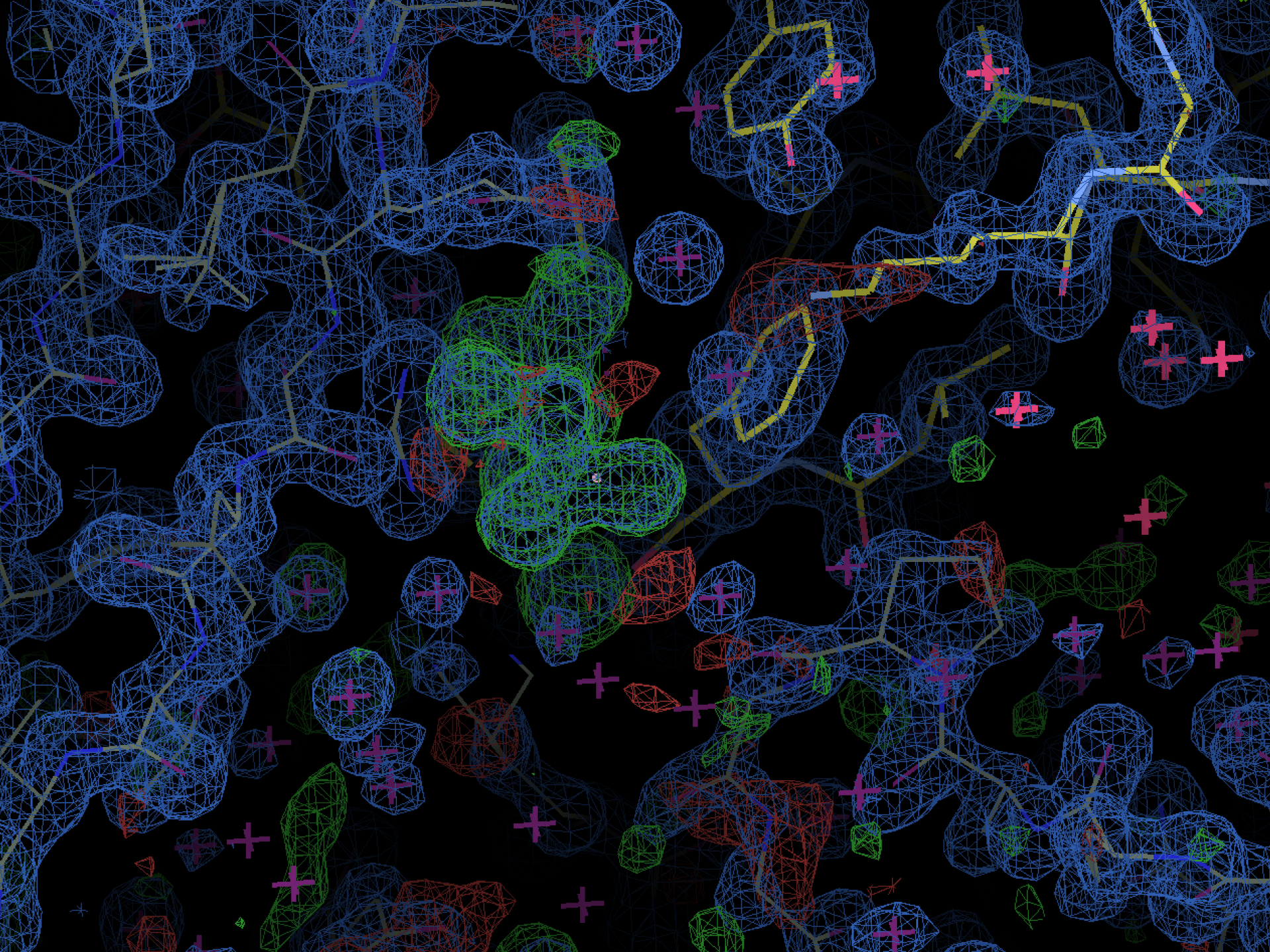
= ?

DIMPLE



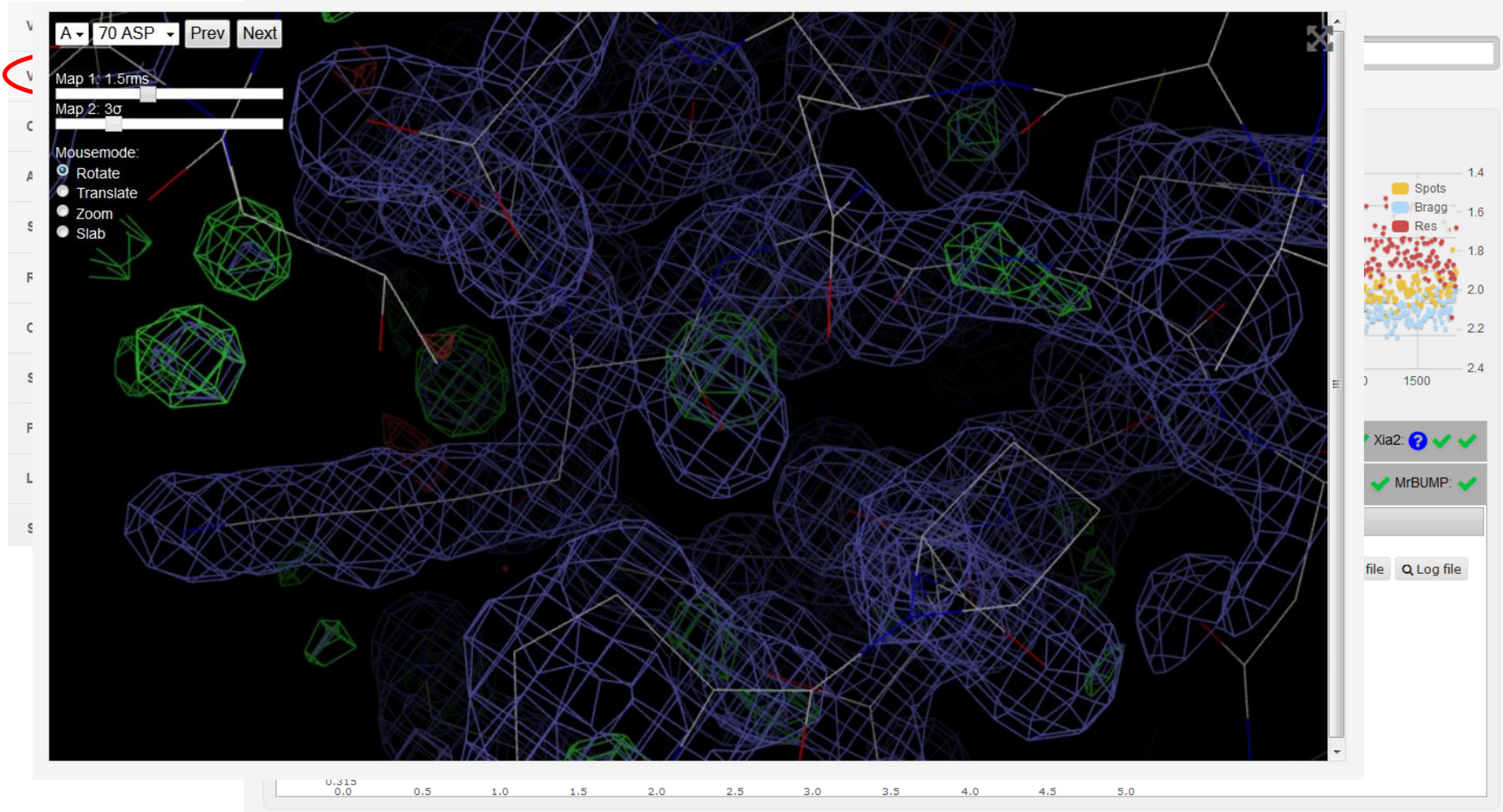
DIMPLE in SynchWeb





MrBUMP

This page shows all data collections for the selected visit. If the visit is ongoing the page will automatically update as new data is collected. Auto processing results will be displayed



- ⇒ Download MTZ/PDB files
- ⇒ See the log file

Big EP

- BIG_EP (developer Irakli Sikharulidze)
 - Fast_EP +++
 - Slow and accurate c/w fast and furious
- Automatically runs autoSHARP, AutoSol and Crank2
- Phases and attempts model building on SAD data
- If sequence provided will build side chains
- Many user structures solved

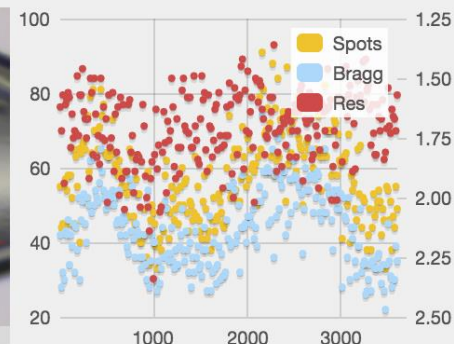
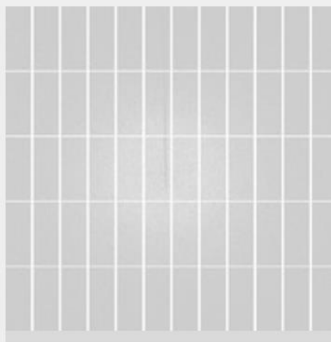
Zn-SAD β -lactamase

PHASING EXAMPLE

Fast DP

12-05-2016 16:15:14 - C2_s4/data2/C2_s4_1_####.cbf

Sample: s_4 Flux: 0.00e+0
 Ω Start: 0.0° Ω Osc: 0.10°
 Ω Overlap: 0° No. Images: 3600
Resolution: 1.44Å Wavelength: 1.2399Å
Exposure: 0.100s Transmission: 1.00%
Beamsize: 32x20µm Type: SAD
Comment: (-1647,-232,326) from DLS GraemeAperture: Large



Auto Processing

Fast DP: ✓ Xia2: ? ✗ ✓

Fast DP XIA2 3dii autoPROC DIALS

Beam Centre	X	Y
Start	218.70	209.20
Refined	218.65	209.27
Δ	0.05	-0.07

Space Group	A	B	C	α	β	γ
C 1 2 1	53.22	61.28	69.28	90.00	93.02	90.00

MTZ file Log file Archive Radiation Damage Lookup Cell

Shell	Observations	Unique	Resolution	Rmeas	I/sig(I)	CC Half	Completeness	Multiplicity	Anom Completeness	Anom Multiplicity	CC Anom
outerShell	16707	2742	1.44 - 1.47	0.540	2.3	86.8	91.9	6.1	88.3	3.2	3.0
innerShell	3109	481	6.43 - 28.01	0.028	54.2	99.8	98.7	6.5	98.2	3.5	58.2
overall	260102	39914	1.44 - 28.01	0.055	15.9	99.9	98.7	6.5	97.4	3.3	37.9

Downstream Processing

Fast EP: ✓ Dimple: ✓ MrBUMP: ✗ Big EP: ✓

Processing images: 1 -> 3600

Phi range: 0.00 -> 360.00

Template: C2_s4_1_####.cbf

Wavelength: 1.23985

Working in: /dls/i04/data/2016/nt14493-7/processed/C2_s4/data2/C2_s4_1_/fast_dp

All autoindexing results:

Lattice a b c alpha beta gamma

mC 53.20 61.30 69.30 90.00 93.00 90.00

mP 40.60 69.30 40.60 90.00 98.10 90.00

aP 40.60 40.60 69.30 92.00 91.90 98.10

Mosaic spread: 0.07 < 0.08 < 0.09

Happy with sg# 5

53.20 61.30 69.30 90.00 93.00 90.00

Low resolution 28.01 28.01 1.47

High resolution 1.44 6.43 1.44

Rmerge 0.055 0.028 0.540

I/sigma 15.90 54.20 2.30

Completeness 98.7 98.7 91.9

Multiplicity 6.5 6.5 6.1

CC 1/2 99.9 99.8 86.8

Anom. Completeness 97.4 98.2 88.3

Anom. Multiplicity 3.3 3.5 3.2

Anom. Correlation 37.9 58.2 3.0

Nrefl 260102 3109 16707

Nunique 39914 481 2742

Mid-slope 1.206

dF/F 0.064

dI/sig(dI) 1.157

Merging point group: C 1 2 1

Unit cell: 53.22 61.28 69.28 90.00 93.02 90.00

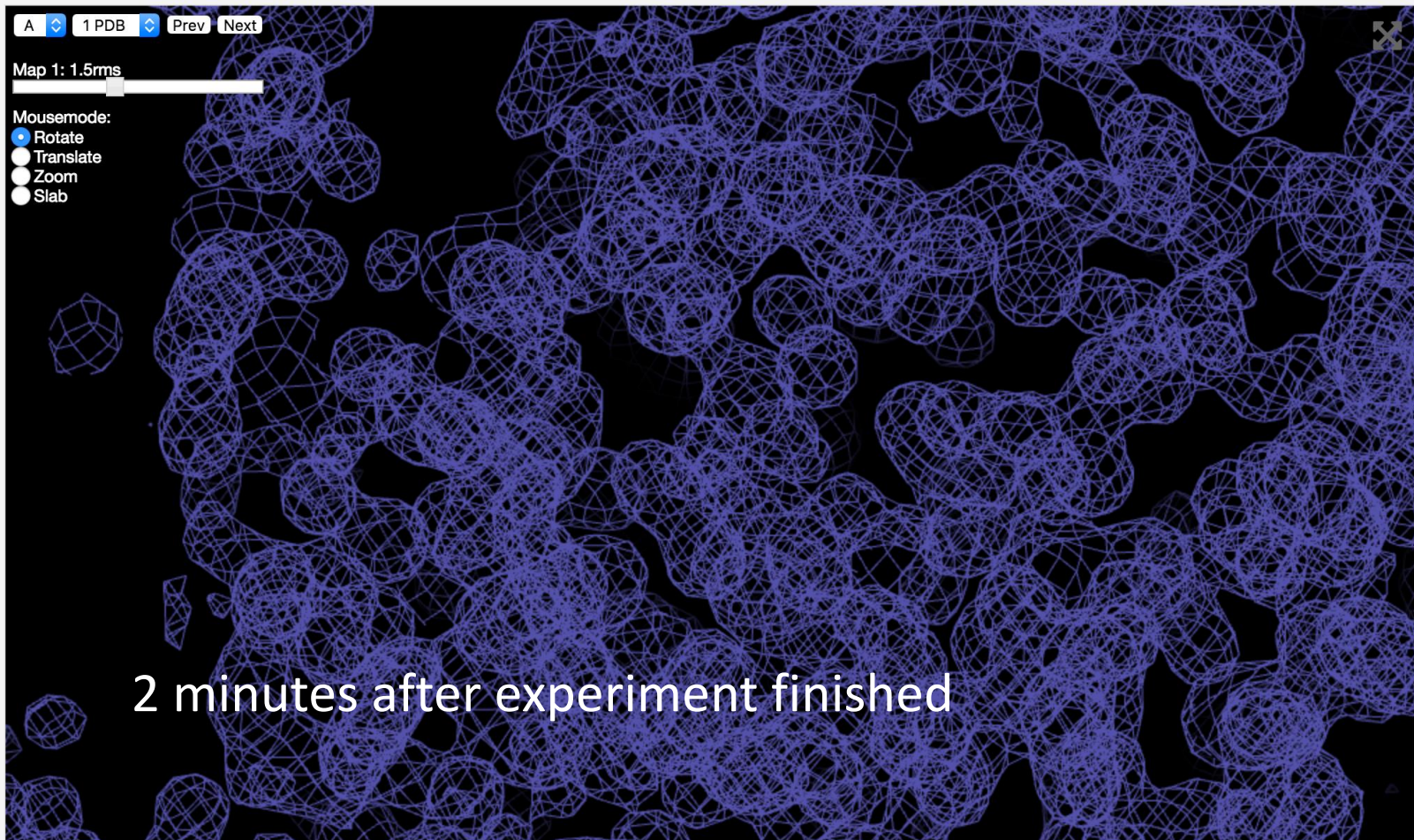
Processing took 00h 00m 53s (53 s) [260102 reflections]

RPS: 4842.1

Fast EP

nt14493-7: C2_s4/data2/C2_s4_1_####.cbf

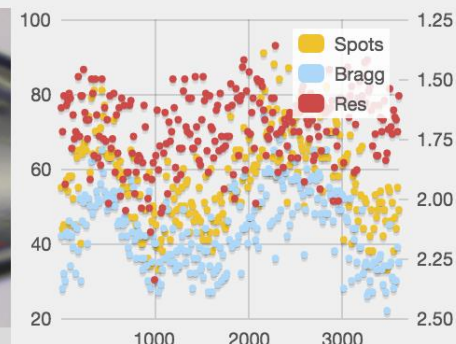
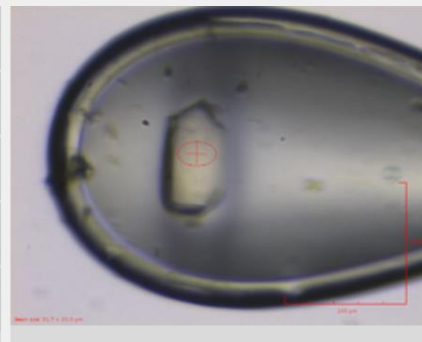
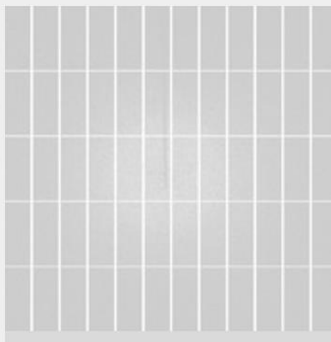
This page is an interactive map and model viewer. You can rotate the view using left mouse. Mousewheel zooms in and out. You can navigate through residues by pressing space or shift+space. Toggle fullscreen with the icon in the top right. CTRL+ left click to translate



xia2 DIALS

12-05-2016 16:15:14 - C2_s4/data2/C2_s4_1_####.cbf

Sample: s_4 Flux: 0.00e+0
 Ω Start: 0.0° Ω Osc: 0.10°
 Ω Overlap: 0° No. Images: 3600
Resolution: 1.44Å Wavelength: 1.2399Å
Exposure: 0.100s Transmission: 1.00%
Beamsize: 32x20µm Type: SAD
Comment: (-1647,-232,326) from DLS GraemeAperture: Large



Auto Processing

Fast DP: ✓ Xia2: ? ✗ ✓

Fast DP XIA2 3dii autoPROC DIALS

Beam Centre	X	Y
Start	218.70	209.20
Refined	218.57	209.18
Δ	0.13	0.02

MTZ file Log file Archive Lookup Cell

Space Group	A	B	C	α	β	γ
C 1 2 1	53.21	61.27	69.28	90.00	93.03	90.00

Shell	Observations	Unique	Resolution	Rmeas	I/sig(I)	CC Half	Completeness	Multiplicity	Anom Completeness	Anom Multiplicity	CC Anom
outerShell	5289	1622	1.31 - 1.33	0.865	1.0	0.7	61.6	3.3	2.4	1.8	0.0
innerShell	18344	2739	3.56 - 40.16	0.046	39.4	1.0	100.0	6.7	99.6	3.5	0.3
overall	306381	50790	1.31 - 40.14	0.076	10.8	1.0	95.2	6.0	93.2	3.1	0.3

Downstream Processing

Fast EP: ✓ Dimple: ✓ MrBUMP: ✗ Big EP: ✓

Big EP

Downstream Processing

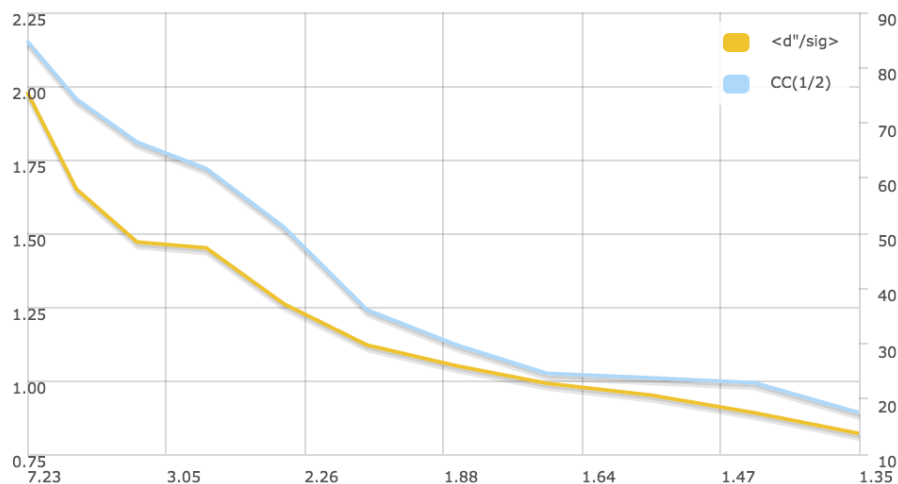
Fast EP: ☒ Dimple: ☒ MrBUMP: ☒ Big EP: ☒

Fast EP ☐ Dimple ☐ Big EP/XDS:0 ☒

Settings

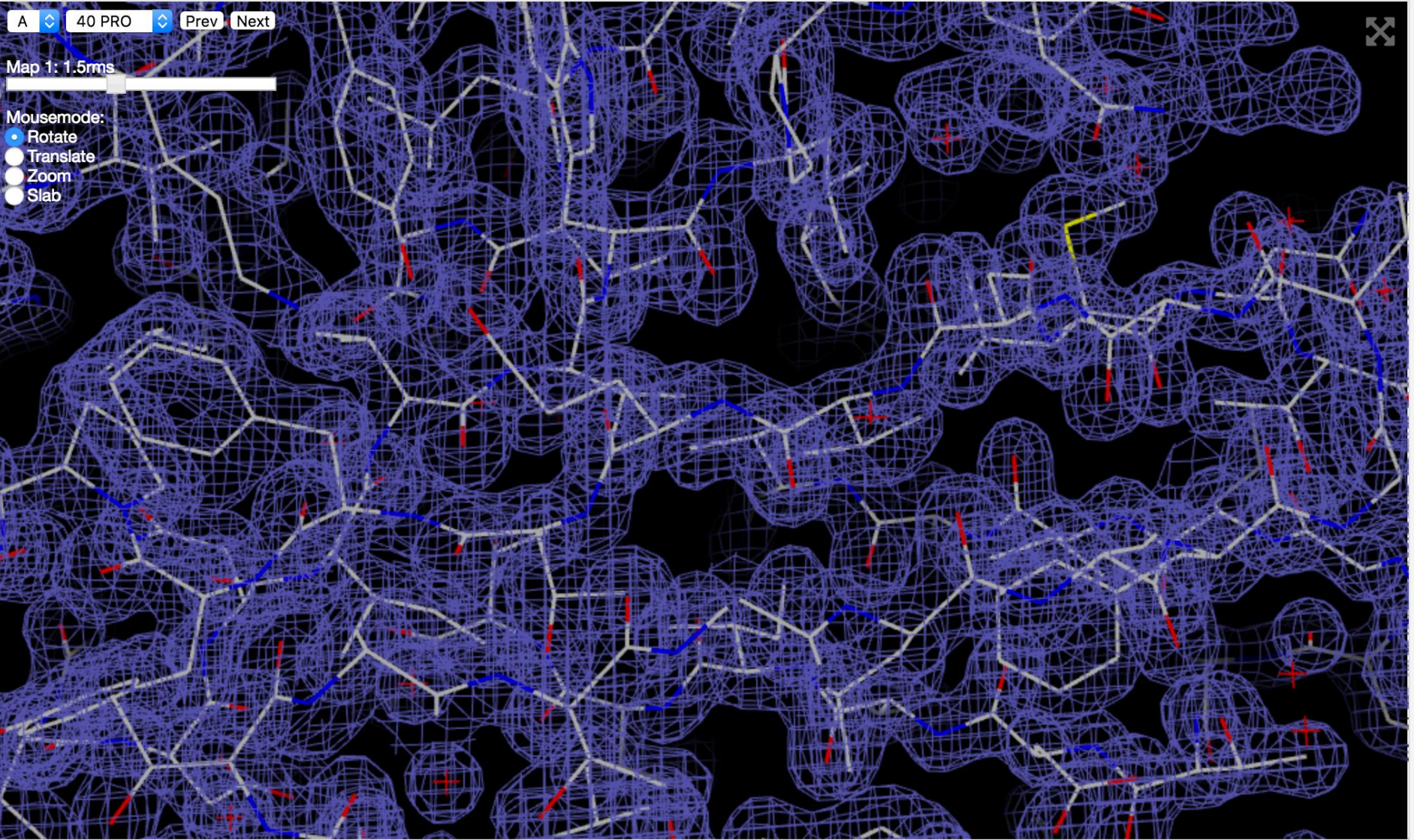
Processed	Anom. scatterer	Space group	Num. scatterers	Dataset type	Compound	Sequence
XDS	Zn	C121	2	hrem	Protein	> SQKVEKTVIKNETGTISISQLNKNVWVHTELGSFNGEAVPSNGLVLNTSKGLVLVDSSWDDKLTKELIEMVEKKFQ KRVT DVIITHAHADRIGGIKTLKERGIKAHSTALTAELAKKNGYEELGDLQTVTNLKFNGMKVETFPYPGKHTEDNIVV WLPQ YNILVGGCLVKSTSADLGNVADAYVNEWSTSIENVLKRYRNINAVVPGHGEVGDKGLLLHTDLLK

XDS



Pipeline	Resid. / Frag / Max. Frag.	Best MapCC (Resol.)	Q Map / Model Viewer	PDB/MTZ file	Log file
AUTOSHARP	49 / 11 / 8	0.63 (1.37)	Q Map / Model Viewer	PDB/MTZ file	Log file
AUTOBUILD	183 / 2 / 170	0.76 (1.53)	Q Map / Model Viewer	PDB/MTZ file	Log file
CRANK2	131 / 10 / 18	0.35 (1.37)	Q Map / Model Viewer	PDB/MTZ file	Log file

This page is an interactive map and model viewer. You can rotate the view using left mouse. Mousewheel zooms in and out. You can navigate through residues by pressing space or shift+space. Toggle fullscreen with the icon in the top right. CTRL+ left click to translate



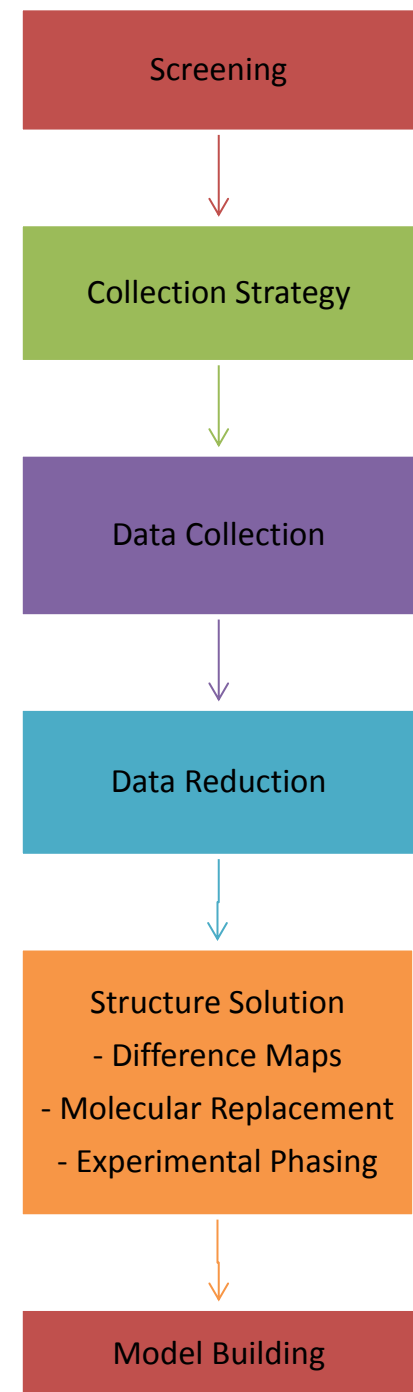
CONCLUSIONS

Feedback from your experiment

- Screening: strategy / alignment suggestions
- Near-real-time: per-image analysis
- Fast feedback: fast_dp / fast_ep / dimple
- Collection is fast so can take a cautious approach to collection
- Thorough processing: xia2, Big EP, MrBUMP
- Find everything: SynchWeb
- Find details:
 - E.g. **/dls/I04/2016/mx12345-6/processed/**

Taking this to conclusion

- Can automate processing and analysis
- But can also automate collection
- Beamlines have these capabilities
 - MASSIF, ESRF
 - I04-1 XChem, Diamond
- Or elements of these capabilities
 - Read the webpages
 - Speak to your local contacts
 - Before your visit

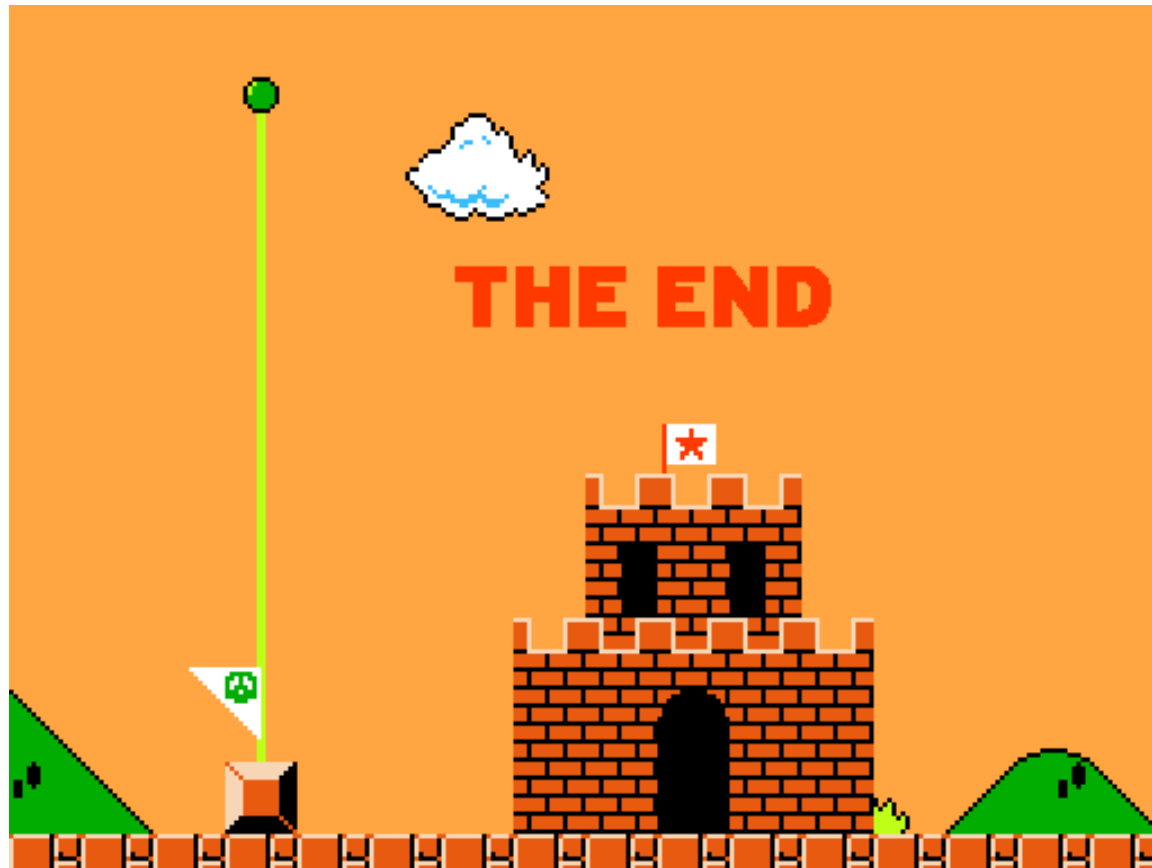


Automated Pipelines Summary

	Diamond	ESRF
Strategies	EDNA Mosflm XOAlign	EDNA
Data Reduction	Fast DP (XDS) Xia2 (XDS/DIALS) Autoproc (XDS)	GrenADES (XDS) Xia2 (DIALS) EDNA Autoprocessing (XDS) Autoproc (XDS)
Difference Maps	DIMPLE	DIMPLE
Molecular Replacement	MrBUMP	Autorickshaw
Experimental Phasing	Fast EP (SHELX) Big EP	Autorickshaw SHELX CDE
Automated Beamline	I04-1 XChem VMXi	MASSIF

The Good, the Bad (and no Ugly?) of automation and pipelines

- Bad
 - Black box – analysis of failure!
 - Crystals, beamline, software, hardware...
 - Personal preferences not always an option
 - Not always the solution to difficult problems
 - Focussed, supported effort on a few crystals can often solve the problem
 - What has happened?
 - This course should give you the tools to know
- Good
 - Efficient – you, beamtime, results
 - Better decision making
 - Better data
 - Better results
 - Focus on your science



Thank you