John Innes Centre

BBSRC bioscience for the future

Optimising the Diamond experience from a user's perspective

Dave Lawson

Department of Biological Chemistry

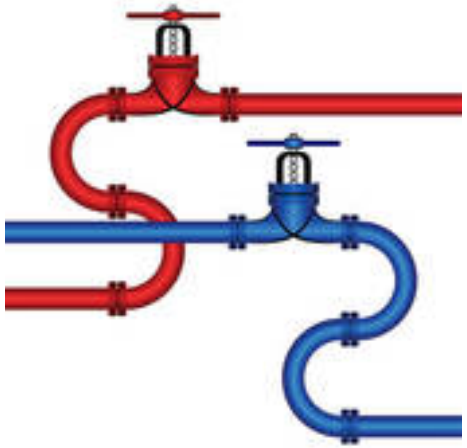
John Innes Centre, Norwich, UK

Data collection philosophy…

Strategy

Tools

Pipelines



Remote

- Maximizing efficiency
- Minimizing time commitment
- Managing your data
- How to collect data

“Routine” data collection:

- on non-hazardous samples
- at cryogenic temperatures
- on pre-cooled crystals in pucks
- at conventional wavelengths

Disclaimer: the thoughts, views, and opinions expressed here are entirely my own and do not reflect the opinions of my employer, my wife, my kids, my…

Primary goals (data quantity **AND** quality…)

- To collect as much data as possible…
- To collect the best possible data…
- To collect the data that will enable me to:
  - Solve my structure
  - Extend the resolution of my structure
  - Show that my ligand is bound

Secondary goal (get a life!)

Primary goals (data quantity **AND** quality…)

- To collect as much data as possible…
- To collect the best possible data…
- To collect the data that will enable me to:
  - Solve my structure
  - Extend the resolution of my structure
  - Show that my ligand is bound

Secondary goal (get a life!)

- Try to answer these questions **during** the visit:
  - Can I solve my structure?
  - Can I extend the resolution of my structure?
  - Is my ligand bound?
- Make best use of time and minimise the amount of follow-up work…

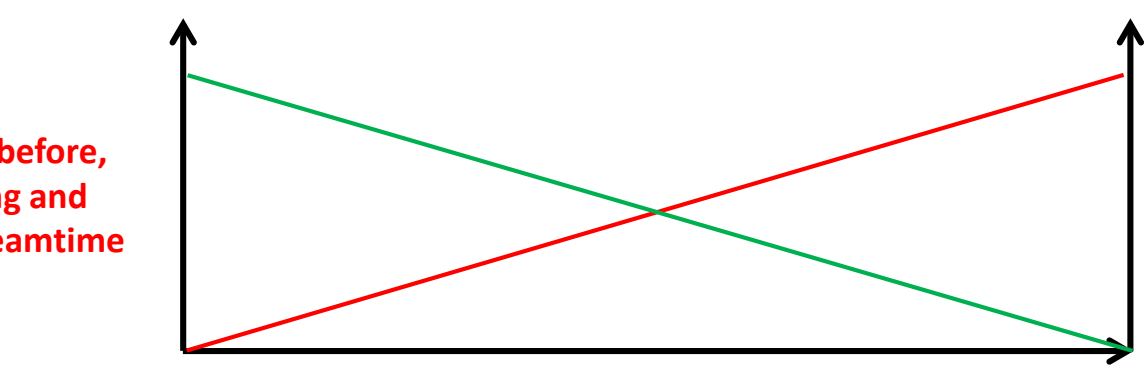
How many samples to take/send per project…

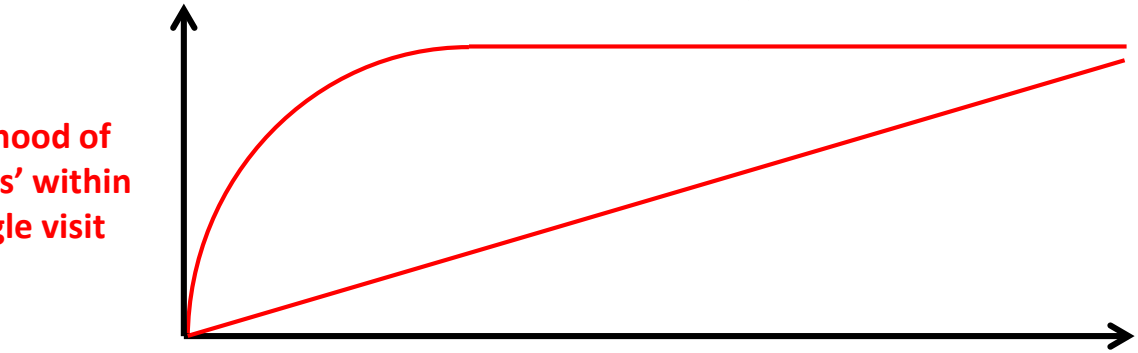
One crystal

High visit frequency

Low visit frequency

A dewar full of crystals (112)





Data collection is always a compromise…

How much data do I need?

What resolution do I need?

What are the data for?

How much time do I have?

How many crystals do I have?

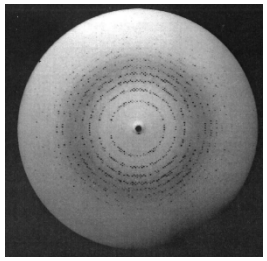
Do as much of this thinking **before** your beamtime as possible!

Detector types

hours

readout time

1 ms



Film

Image plate

CCD (charge-coupled device)

newest detectors at Diamond can collect at 500 Hz: 360° possible in 10 s! (3,600 x 0.1° images)

HPC (hybrid photon counting)

1990

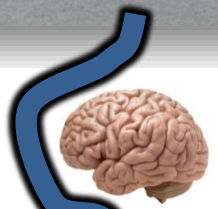
2000

2010

MX data collection has become faster…

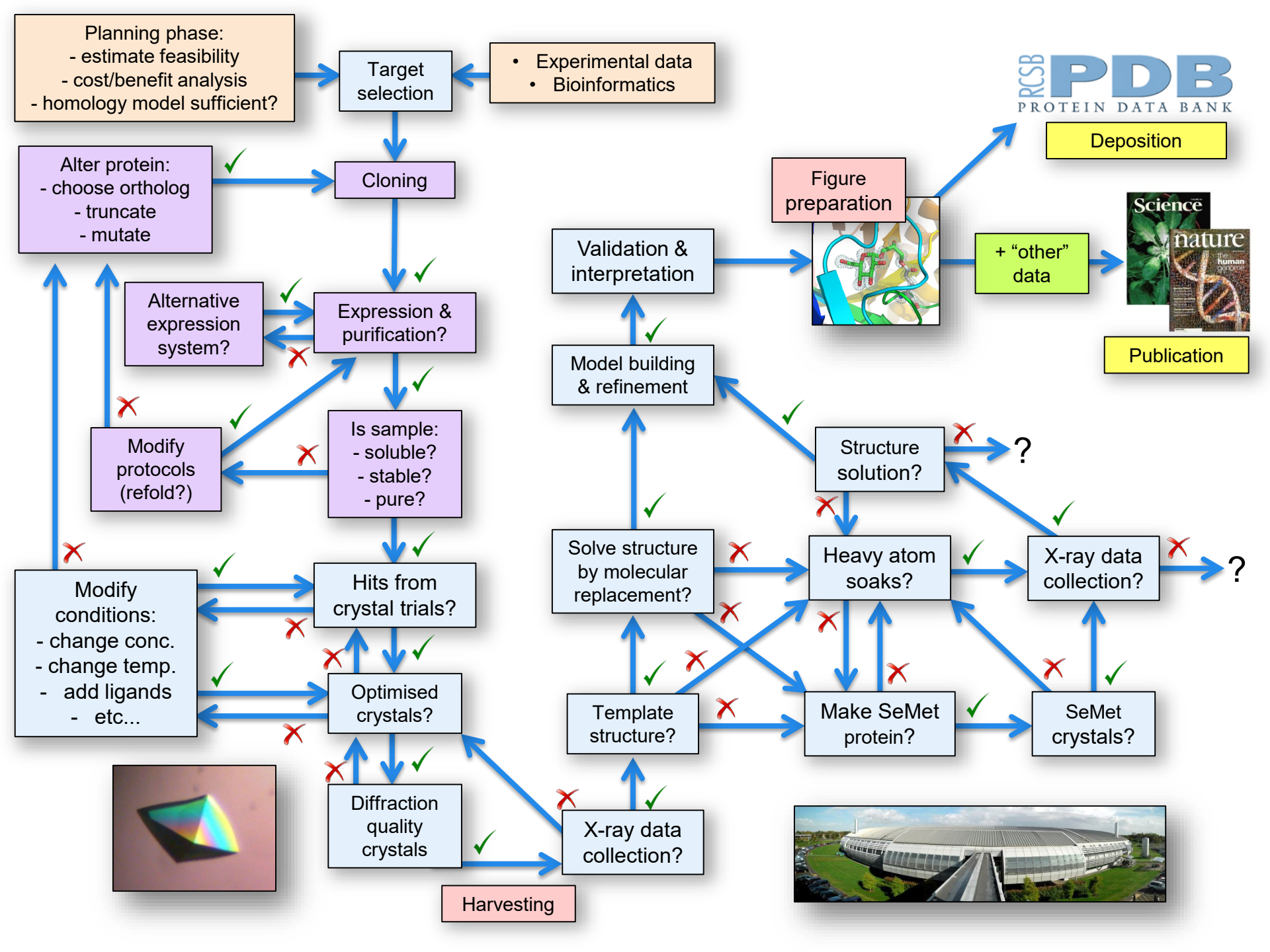
20th century data collection

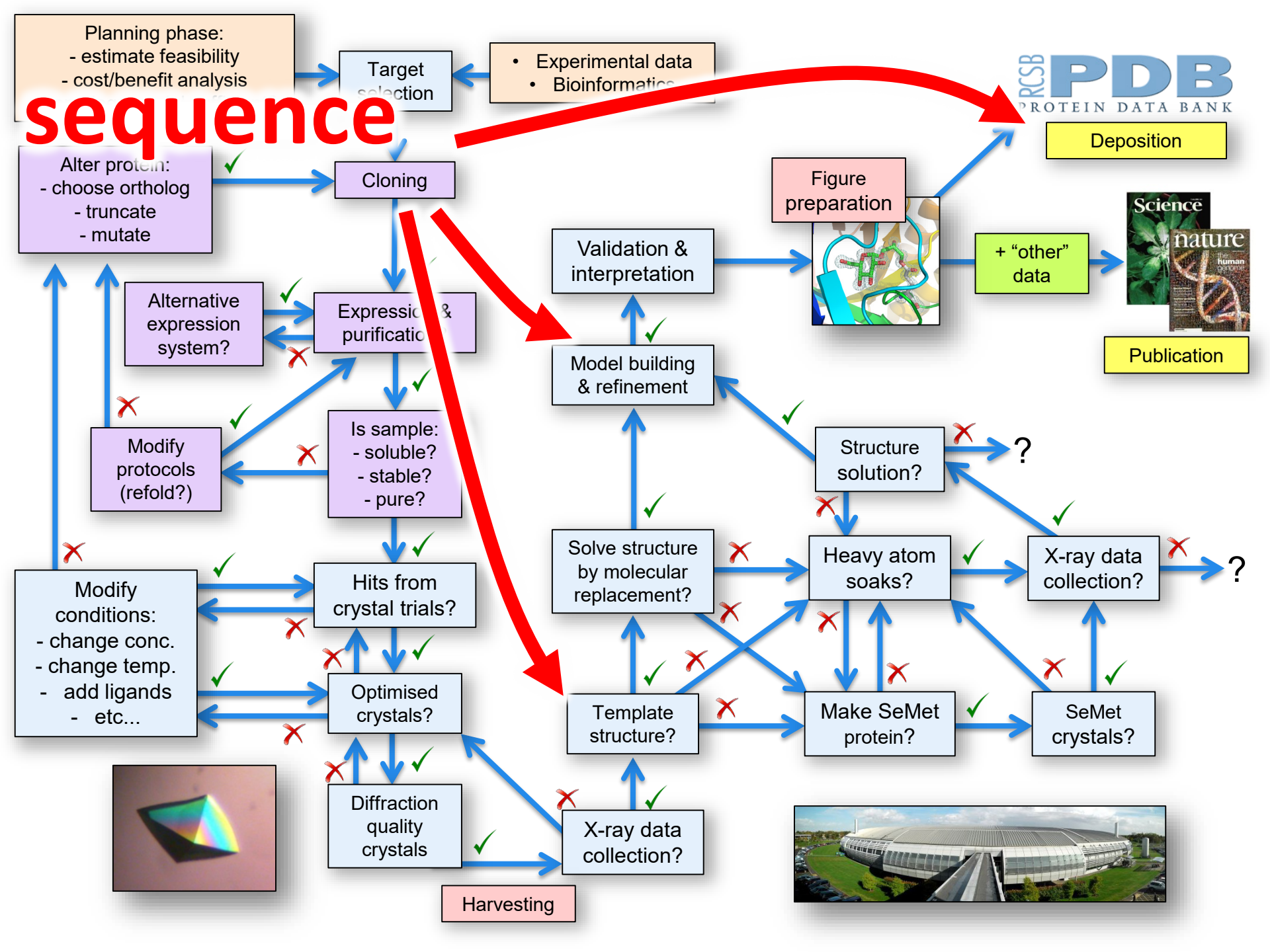
21st century data collection

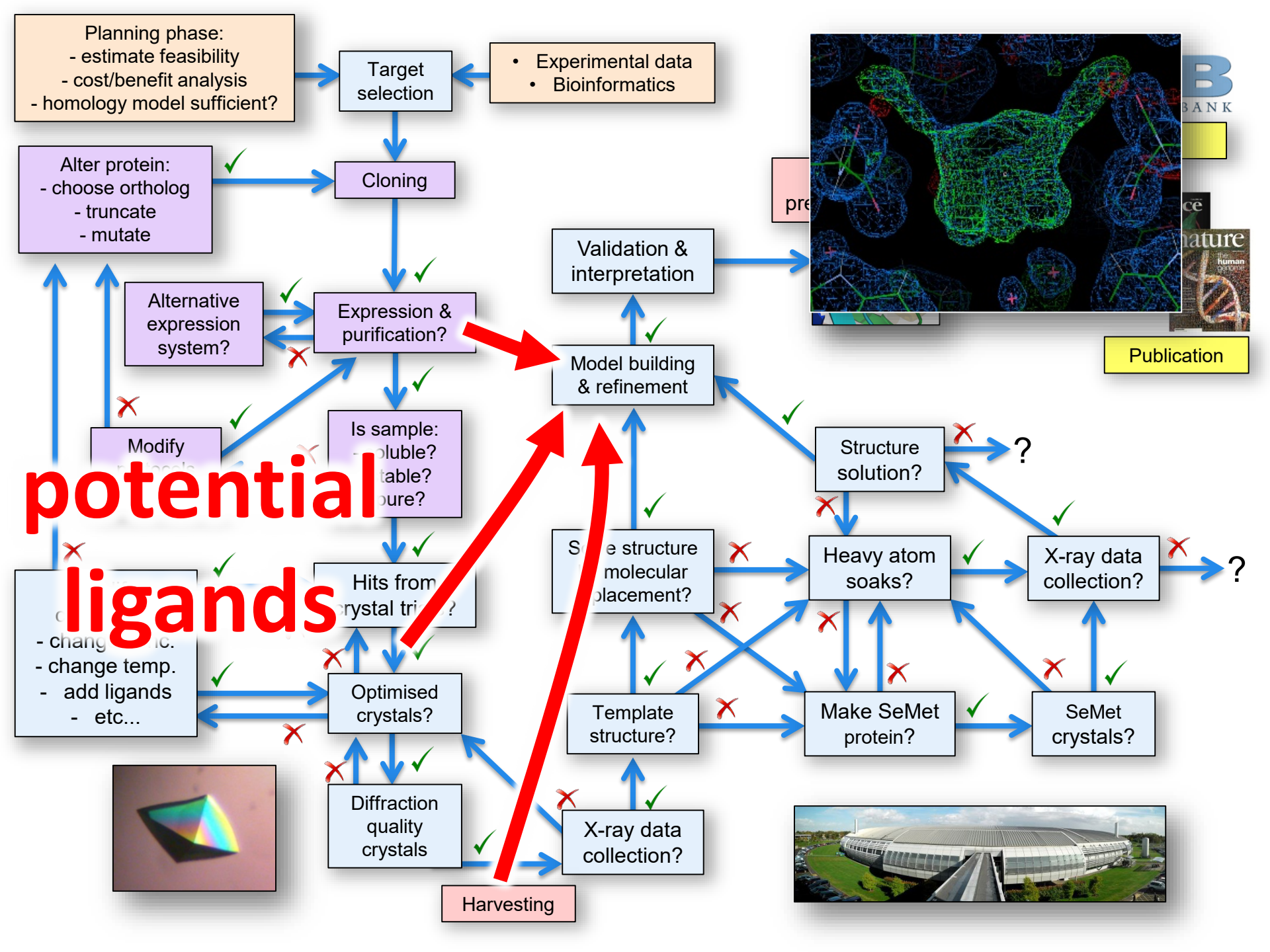


MX data collection has become faster…

- Typical dataset takes <1 min to collect
  - 10 - 20 GB raw data per data set!
  - >1.0 TB data possible in 24 hrs!
- In practice, 4 “useful” data sets per hour is good
  - includes screening for “best” crystal
  - …and thinking!
- But - beamtime is still in high demand
  - users are generating crystals more rapidly
  - can get away with “marginal” samples
- Each session is often split between several groups
  - therefore need to be efficient and organised...


Planning phase:
- estimate feasibility
- cost/benefit analysis
- homology model sufficient?
Target selection
• Experimental data
• Bioinformatics
RCSB PDB
PROTEIN DATA BANK
Deposition
Alter protein:
- choose ortholog
- truncate
- mutate
Cloning
Figure preparation
Validation & interpretation
+ "other" data
Science
nature
Publication
Alternative expression system?
Expression & purification?
Model building & refinement
Structure solution?
?
Modify protocols (refold?)
Is sample:
- soluble?
- stable?
- pure?
Solve structure by molecular replacement?
Heavy atom soaks?
X-ray data collection?
?
Modify conditions:
- change conc.
- change temp.
- add ligands
- etc...
Hits from crystal trials?
Optimised crystals?
Template structure?
Make SeMet protein?
SeMet crystals?
Diffraction quality crystals
X-ray data collection?
Harvesting


Planning phase:
- estimate feasibility
- cost/benefit analysis
Target selection
• Experimental data
• Bioinformatics
sequence
RCSB PDB
PROTEIN DATA BANK
Deposition
Alter protein:
- choose ortholog
- truncate
- mutate
Cloning
Figure preparation
Validation & interpretation
+ "other" data
Science
nature
the human genome
Publication
Alternative expression system?
Expressio & purificatio
Model building & refinement
Modify protocols (refold?)
Is sample:
- soluble?
- stable?
- pure?
Structure solution?
?
Solve structure by molecular replacement?
Heavy atom soaks?
X-ray data collection?
?
Modify conditions:
- change conc.
- change temp.
- add ligands
- etc...
Hits from crystal trials?
Optimised crystals?
Template structure?
Make SeMet protein?
SeMet crystals?
Diffraction quality crystals
X-ray data collection?
Harvesting


Planning phase:
- estimate feasibility
- cost/benefit analysis
- homology model sufficient?
Target selection
• Experimental data
• Bioinformatics
Alter protein:
- choose ortholog
- truncate
- mutate
Cloning
Alternative expression system?
Expression & purification?
Modify
Is sample:
luble?
table?
ure?
potential
ligands
- change temp.
- add ligands
- etc...
Hits from
rystal tri
Optimised crystals?
Diffraction quality crystals
Harvesting
X-ray data collection?
Template structure?
e structure
molecular
placement?
Model building & refinement
Validation & interpretation
Publication
Structure solution?
?
Heavy atom soaks?
X-ray data collection?
?
Make SeMet protein?
SeMet crystals?

Methods

Planning phase:
- estimate feasibility
- cost/benefit analysis
- homology model sufficient?

Target selection

- Experimental data
- Bioinformatics

Alter protein:
- choose ortholog
- truncate

Cloning

Alternative expression system?

Expression & purification?

Modify protocols (refold?)

Is sample:
- soluble?
- stable?
- pure?

Modify conditions:
- change conc.
- change temp.
- add ligands
- etc...

Hits from crystal trials?

Optimise crystals?

Diffraction quality crystals

Harvesting

X-ray data collection?

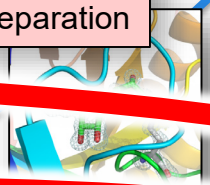
Template structure?

Solve structure by molecular replacement

Model building & refinement

Validation & interpretation

Figure preparation



RCSB PDB PROTEIN DATA BANK

Deposition

+ "other"

Publication

Structure solution?

Heavy atom soaks?

Make SeMet protein?

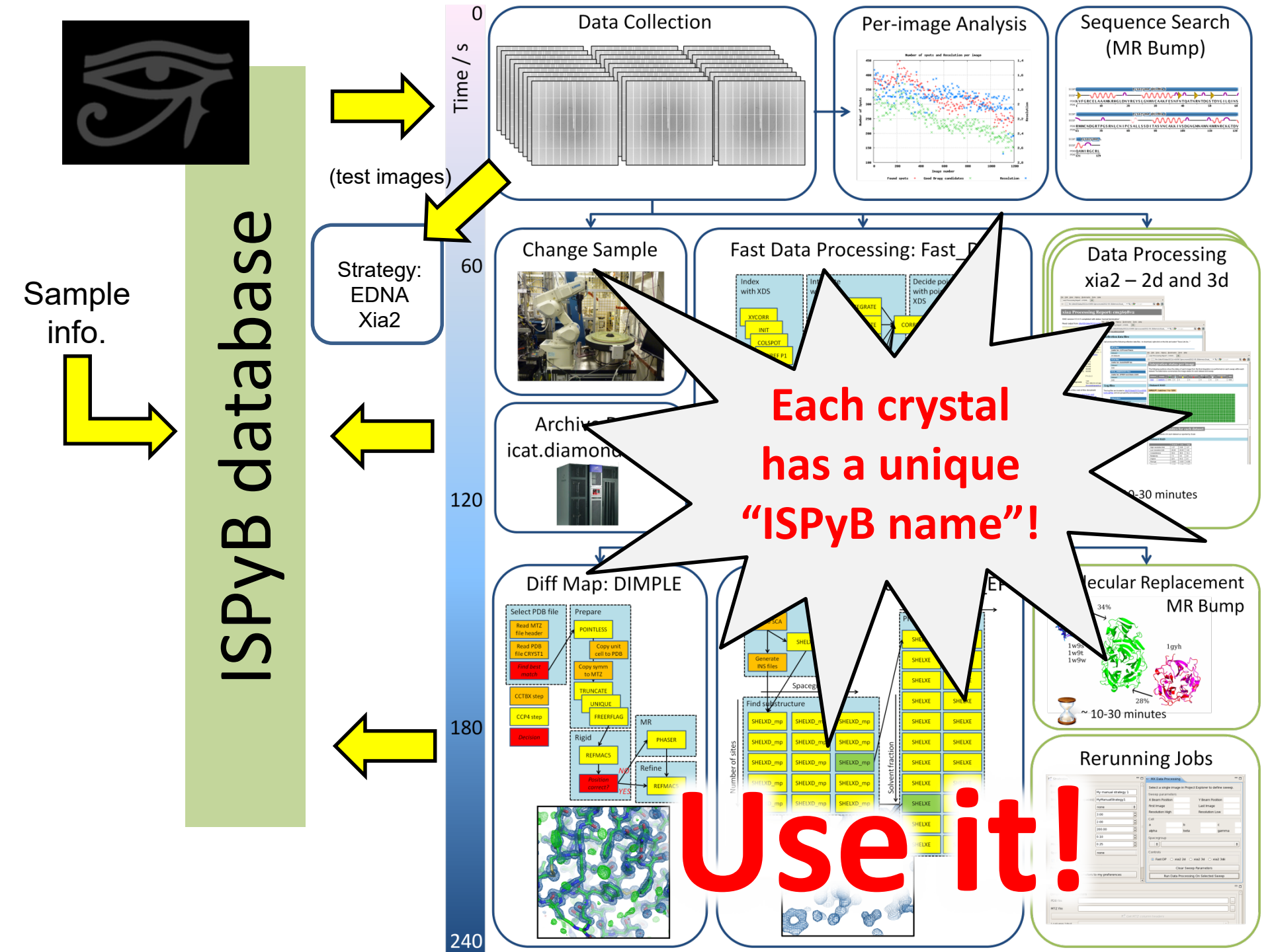
X-ray data collection

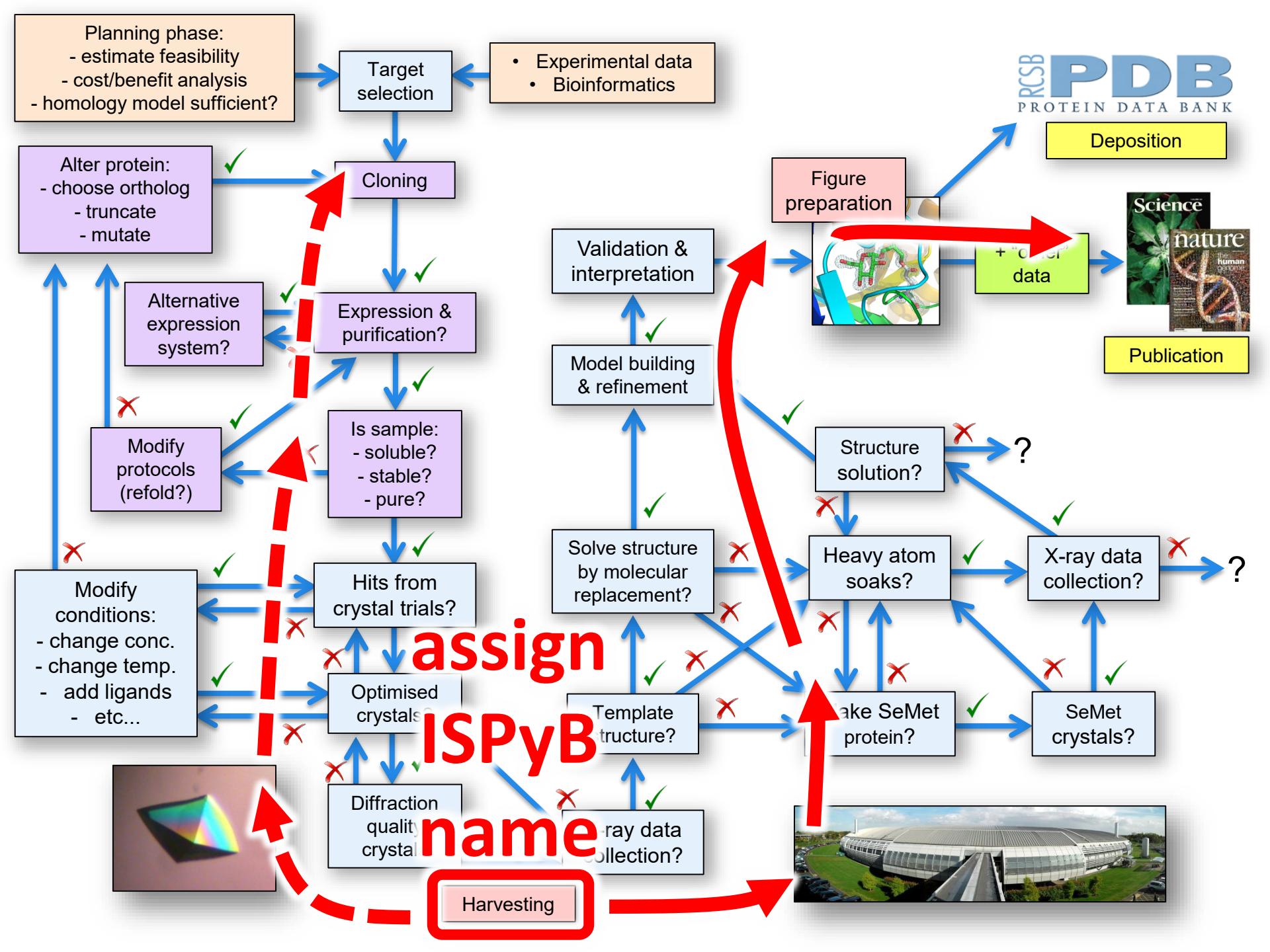
?

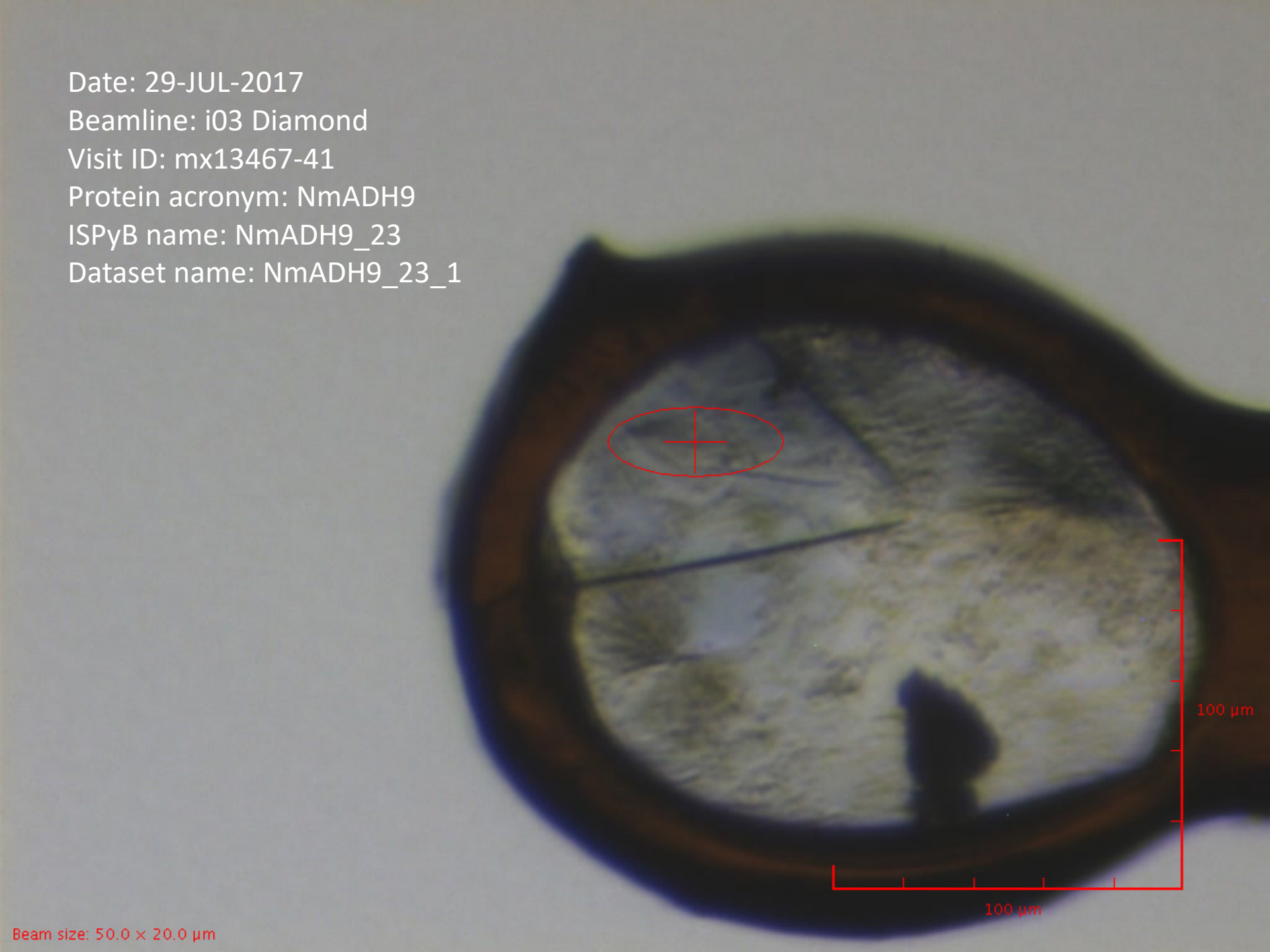
Taking back control...

+

ISPyB *CCP4i2*


ISPyB database
Sample info.
(test images)
Strategy: EDNA Xia2
Time / s
0
60
120
180
240
Data Collection
Per-image Analysis
Sequence Search (MR Bump)
Change Sample
Fast Data Processing: Fast_
Data Processing xia2 – 2d and 3d
Archiv
icat.diamond
Diff Map: DIMPLE
Molecular Replacement MR Bump
Rerunning Jobs
Each crystal has a unique "ISPyB name"!
Use it!


Planning phase:
- estimate feasibility
- cost/benefit analysis
- homology model sufficient?
Target selection
• Experimental data
• Bioinformatics
RCSB PDB
PROTEIN DATA BANK
Deposition
Alter protein:
- choose ortholog
- truncate
- mutate
Cloning
Figure preparation
Validation & interpretation
data
Science
nature
Publication
Alternative expression system?
Expression & purification?
Model building & refinement
Modify protocols (refold?)
Is sample:
- soluble?
- stable?
- pure?
Structure solution?
?
Solve structure by molecular replacement?
Heavy atom soaks?
X-ray data collection?
?
Modify conditions:
- change conc.
- change temp.
- add ligands
- etc...
Hits from crystal trials?
assign
Optimised crystals
Template structure?
ake SeMet protein?
SeMet crystals?
ISPyB
Diffraction quality crystal
ray data collection?
name
Harvesting


Date: 29-JUL-2017
Beamline: i03 Diamond
Visit ID: mx13467-41
Protein acronym: NmADH9
ISPyB name: NmADH9_23
Dataset name: NmADH9_23_1
100 µm
100 µm
Beam size: 50.0 × 20.0 µm

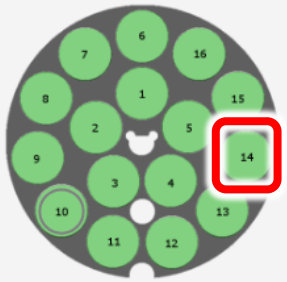
Harvest your crystals and enter sample info into ISPyB

Container: DLS-442

This page shows the contents of the selected container. Samples can be added and edited by clicking the pencil icon, and removed by clicking the x

This container is currently assigned and in use on a beamline sample changer. Unassign it to make it editable

Shipment	JIC_260717_I03
Dewar	DLS-MX-0002
Container Type	Puck
Registered Container	DLS-442 [View]
Barcode	Click to edit
Automated Collection	+ Queue this container for Auto Collect
Comments	Click to edit



Location History

Date	Status	Location	Beamline
08-09-2017 10:49	at facility		
04-08-2017 15:58	at DLS		
29-07-2017 11:26	processing	7	i03
21-07-2017 12:04	at DLS		

10 ▾ Page « < 1 > »

this is **UNIQUE**

+ Extra Fields

Location	Protein Acronym	Abundance	Components	Name	Spacegroup	Barcode	Comment	Status	
9	NmADH9			NmADH9_18					
10	NmADH9			NmADH9_19					
11	NmADH9			NmADH9_20					
12	NmADH9			NmADH9_21					
13	NmADH9			NmADH9_2					
14	NmADH9			NmADH9_23					
15	NmADH9			NmADH9_24					
16	NmADH9			NmADH9_25					

Puck number – DLS442

Posi-tion	Person	Protein name	IspyB_name	Plate	Well	Conditions	Soak/cryo	Space group	Test	Reso-lution	Comments
9	Benjy	NmADH9	NmADH9_18	MCBL0004	A11.1	"	Cis cis nepetalactol+ 20%EG				10mM soak for approx 1 h
10	Benjy	NmADH9	NmADH9_19	MCBL0004	A11.1	"	8 oxogeranial+ 20%EG				10mM soak for approx 1 h
11	Benjy	NmADH9	NmADH9_20	MCBL0004	A11.1	"	8 oxogeranial+ 20%EG				10mM soak for approx 1 h
12	Benjy	NmADH9	NmADH9_21	HD01	B6	PEG 4k 29%, 0.1M Mes pH 6.5, 1 mM NAD	8-oxocitronellal+ 20%EG				5mM soak for approx 1.5 h
13	Benjy	NmADH9	NmADH9_22	HD01	B4	"	8-oxocitronellal+ 20%EG				5mM soak for approx 1.5 h
14	Benjy	NmADH9	NmADH9_23	HD01	B3	"	Cis cis nepetalactone+ 20%EG				5mM soak for approx 1.5 h
15	Benjy	NmADH9	NmADH9_24	HD01	B2	"	Cis cis nepetalactol+ 20%EG				5mM soak for approx 1.5 h
16	Benjy	NmADH9	NmADH9_25	HD01	B1	"	8 hydroxygeraniol+ 20%EG				5mM soak for approx 1 h

annotate during data collection – helps decision making

MX13467-41 7 hrs I03 from 12:00 Sat 29 Jul 2017 to 19:00 Sat 29 Jul 2017

First Local Contact: Dr Katherine McAuley (01235 778405)

Beamline I03: 01235 778707

Experimental Hall Coordinator: 01235 778787

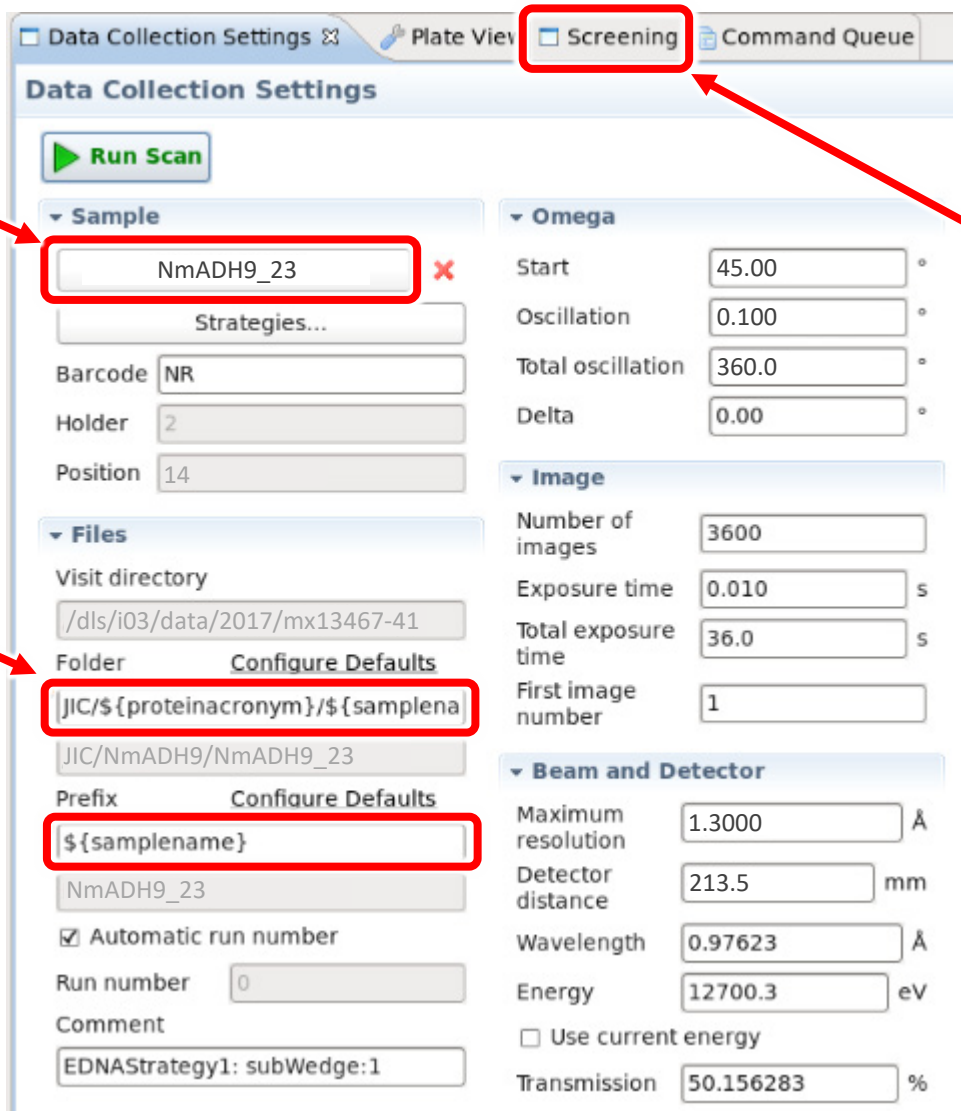
Institution	JIC	Group(s)	Lawson
Contact name(s) and number(s)	Dave Lawson: 01603 123456 / 07778 123456		
Approx. timings	12:00 – 15:00	On-site/remote?	Remote
Dewar ID	Puck ID (# samples)	Users and comments	
DLS-MX-0002	DLS0086 (16 samples)	Sue	
	DLS0087 (12 samples)	JC	
	DLS0090 (12 samples)	Sue	
	DLS442 (16 samples)	Benjy	
Anomalous elements:	Zn		
Other requirements:	MCA scans		
Comments/ questions:			

Decide on a running order for users and samples – share your plans with LC and EHC

Data collection setup in GDA

Select required sample from drop-down menu:

- no need to enter sample information or specify sample location
- less likely to get the wrong sample



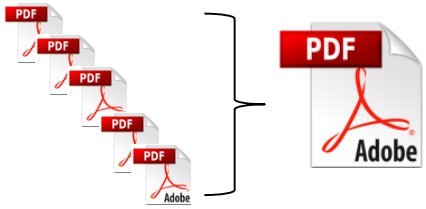
use the "screening" tab for test images – keeps them separate from datasets

Put your data into your own directory....

...especially important if there are multiple users from several institutions – simplifies backing up too....

Gives: /dls/{beamline}/data/{year}/{visit-ID}/**{your-name}**/{proteinacronym}/{samplename}/{samplename}_{run_no.}_{image_no.}.cbf

e.g.: /dls/i03/data/2017/mx13467-41/**JIC**/NmADH9/NmADH9_23/NmADH9_23_1_3600.cbf



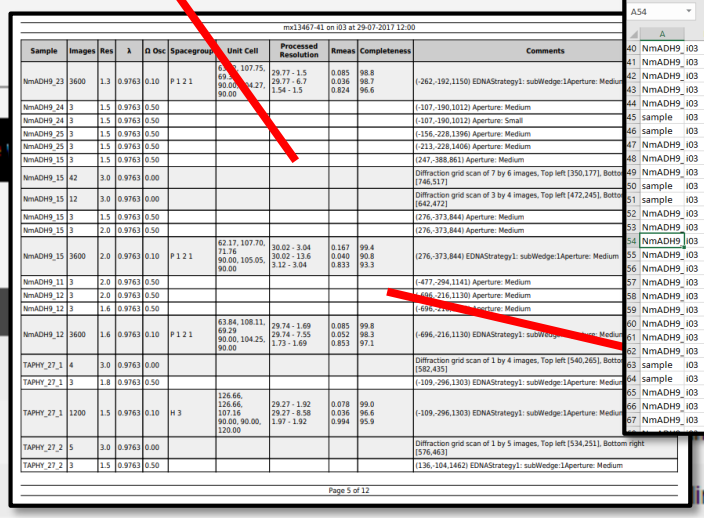
⇦ simple to search this for an ISPyB name

Visit List

This page lists the

Start						
19:00 15-09-2017						
02:00 11-09-2017						
10:00 05-08-2017					395	
12:00 29-07-2017	19:00 29-07-2017	41	i03	Dr Katherine McAuley	136	Compulsorily Remote
19:00 26-07-2017	02:00 27-07-2017	40	i03	Mr Mark Williams	71	Compulsorily Remote
02:0			i03	Dr Neil Paterson	38	Compulsorily Remote
17:0			i03	Dr Neil Paterson	76	
19:0			i04	Dr Melanie Vollmar	69	Compulsorily Remote
12:0			i04	Dr Dave Hall	106	Compulsorily Remote
02:0			i03	Dr Neil Paterson	82	Compulsorily Remote

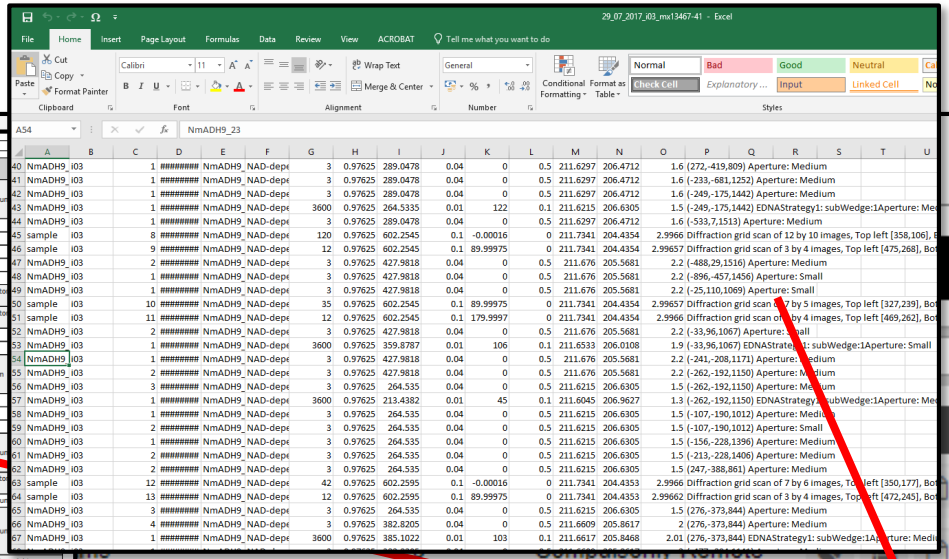
10 5 6 7 › »



mx13467-41 on i03 at 29-07-2017 12:00

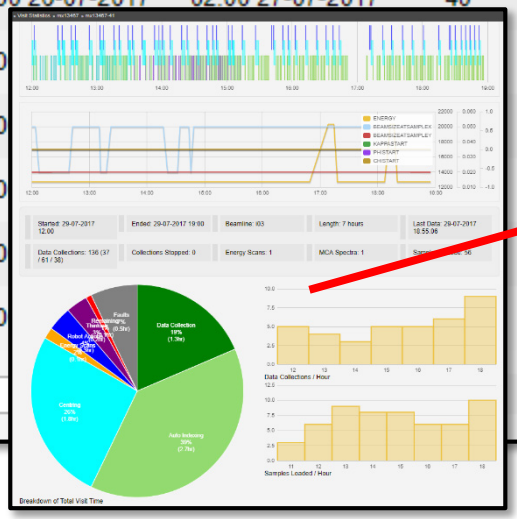
Sample	Images	Res	λ	Ω Osc	Spacegroup	Unit Cell	Processed Resolution	Rmeas	Completeness	Comments
NmADH9_23	3600	1.3	0.9763	0.10	P 1 2 1	64.77, 107.75, 69.29, 90.00, 104.27, 90.00	29.77 - 1.5 29.77 - 6.7 1.54 - 1.5	0.085 0.036 0.824	98.8 98.7 96.6	(-262,-192,1150) EDNAStrategy1: subWedge:1Aperture: Medium
NmADH9_24	3	1.5	0.9763	0.50						(-107,-190,1012) Aperture: Medium
NmADH9_24	3	1.5	0.9763	0.50						(-107,-190,1012) Aperture: Small
NmADH9_25	3	1.5	0.9763	0.50						(-156,-228,1396) Aperture: Medium
NmADH9_25	3	1.5	0.9763	0.50						(-213,-228,1406) Aperture: Medium
NmADH9_15	3	1.5	0.9763	0.50						(247,-388,861) Aperture: Medium
NmADH9_15	42	3.0	0.9763	0.00						Diffraction grid scan of 7 by 6 images, Top left [350,177], Bottom right [746,517]
NmADH9_15	12	3.0	0.9763	0.00						Diffraction grid scan of 3 by 4 images, Top left [472,245], Bottom right [642,472]
NmADH9_15	3	1.5	0.9763	0.50						(276,-373,844) Aperture: Medium
NmADH9_15	3	2.0	0.9763	0.50						(276,-373,844) Aperture: Medium
NmADH9_15	3600	2.0	0.9763	0.10	P 1 2 1	62.17, 107.70, 71.76, 90.00, 105.05, 90.00	30.02 - 3.04 30.02 - 13.6 3.12 - 3.04	0.167 0.040 0.833	99.4 90.8 93.3	(276,-373,844) EDNAStrategy1: subWedge:1Aperture: Medium
NmADH9_11	3	2.0	0.9763	0.50						(-477,-294,1141) Aperture: Medium
NmADH9_12	3	2.0	0.9763	0.50						(-696,-216,1130) Aperture: Medium
NmADH9_12	3	1.6	0.9763	0.50						(-696,-216,1130) Aperture: Medium
NmADH9_12	3600	1.6	0.9763	0.10	P 1 2 1	63.84, 108.11, 69.29 90.00, 104.25, 90.00	29.74 - 1.69 29.74 - 7.55 1.73 - 1.69	0.085 0.052 0.853	99.8 98.3 97.1	(-696,-216,1130) EDNAStrategy1: subWedge:1Aperture: Medium
TAPHY_27_1	4	3.0	0.9763	0.00						Diffraction grid scan of 1 by 4 images, Top left [540,265], Bottom right [582,435]
TAPHY_27_1	3	1.8	0.9763	0.50						(-109,-296,1303) EDNAStrategy1: subWedge:1Aperture: Medium
TAPHY_27_1	1200	1.5	0.9763	0.10	H 3	126.66, 126.66, 107.16 90.00, 90.00, 120.00	29.27 - 1.92 29.27 - 8.58 1.97 - 1.92	0.078 0.036 0.994	99.0 96.6 95.9	(-109,-296,1303) EDNAStrategy1: subWedge:1Aperture: Medium
TAPHY_27_2	5	3.0	0.9763	0.00						Diffraction grid scan of 1 by 5 images, Top left [534,251], Bottom right [576,463]
TAPHY_27_2	3	1.5	0.9763	0.50						(136,-104,1462) EDNAStrategy1: subWedge:1Aperture: Medium

Page 5 of 12



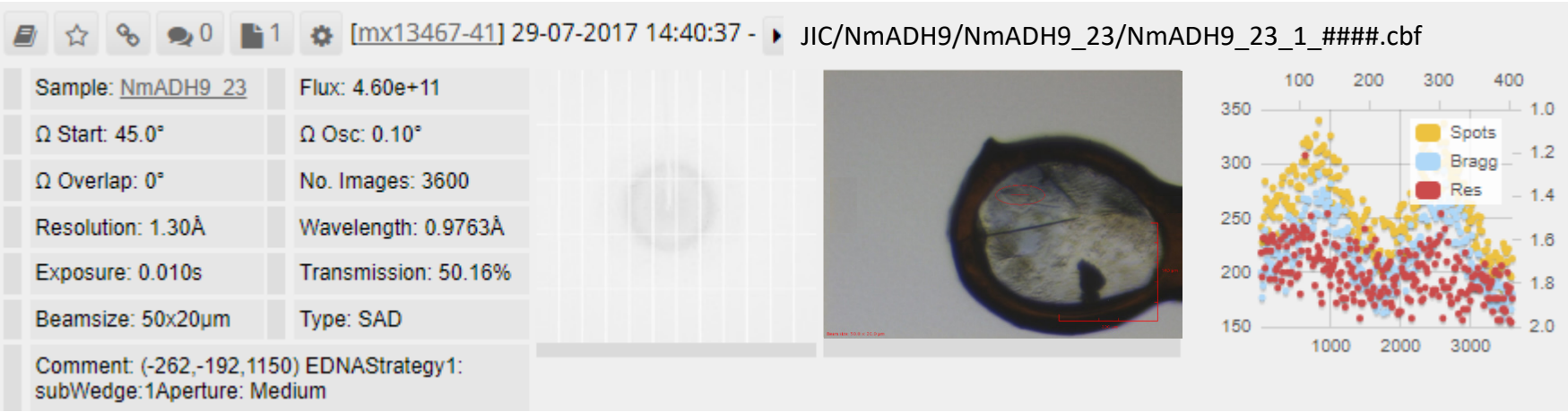
29_07_2017_i03_mx13467-41 - Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
40	NmADH9_i03		1	########	NmADH9_NAD-dep		3	0.97625	289.0478	0.04	0	0.5	211.6297	206.4712	1.6 (272,-419,809) Aperture: Medium						
41	NmADH9_i03		1	########	NmADH9_NAD-dep		3	0.97625	289.0478	0.04	0	0.5	211.6297	206.4712	1.6 (-233,-681,1252) Aperture: Medium						
42	NmADH9_i03		1	########	NmADH9_NAD-dep		3	0.97625	289.0478	0.04	0	0.5	211.6297	206.4712	1.6 (-249,-175,1442) Aperture: Medium						
43	NmADH9_i03		1	########	NmADH9_NAD-dep		3600	0.97625	264.5335	0.01	122	0.1	211.6215	206.6305	1.5 (-249,-175,1442) EDNAStrategy1: subWedge:1Aperture: Me						
44	NmADH9_i03		1	########	NmADH9_NAD-dep		3	0.97625	289.0478	0.04	0	0.5	211.6297	206.4712	1.6 (-533,7,1513) Aperture: Medium						
45	sample	i03	8	########	NmADH9_NAD-dep		120	0.97625	602.2545	0.1	-0.00016	0	211.7341	204.4354	2.9966 Diffraction grid scan of 12 by 10 images, Top left [358,106], B						
46	sample	i03	9	########	NmADH9_NAD-dep		12	0.97625	602.2545	0.1	89.99975	0	211.7341	204.4354	2.99657 Diffraction grid scan of 3 by 4 images, Top left [475,268], Bot						
47	NmADH9_i03		2	########	NmADH9_NAD-dep		3	0.97625	427.9818	0.04	0	0.5	211.676	205.5681	2.2 (-488,29,1516) Aperture: Medium						
48	NmADH9_i03		1	########	NmADH9_NAD-dep		3	0.97625	427.9818	0.04	0	0.5	211.676	205.5681	2.2 (-896,-457,1456) Aperture: Small						
49	NmADH9_i03		1	########	NmADH9_NAD-dep		3	0.97625	427.9818	0.04	0	0.5	211.676	205.5681	2.2 (-25,110,1069) Aperture: Small						
50	sample	i03	10	########	NmADH9_NAD-dep		35	0.97625	602.2545	0.1	89.99975	0	211.7341	204.4354	2.99657 Diffraction grid scan of 7 by 5 images, Top left [327,239], Bot						
51	sample	i03	11	########	NmADH9_NAD-dep		12	0.97625	602.2545	0.1	179.9997	0	211.7341	204.4354	2.9966 Diffraction grid scan of 3 by 4 images, Top left [469,262], Bot						
52	NmADH9_i03		2	########	NmADH9_NAD-dep		3	0.97625	427.9818	0.04	0	0.5	211.676	205.5681	2.2 (-33,96,1067) Aperture: Small						
53	NmADH9_i03		1	########	NmADH9_NAD-dep		3600	0.97625	359.8787	0.01	106	0.1	211.6533	206.0108	1.9 (-33,96,1067) EDNAStrategy1: subWedge:1Aperture: Small						
54	NmADH9_i03		1	########	NmADH9_NAD-dep		3	0.97625	427.9818	0.04	0	0.5	211.676	205.5681	2.2 (-241,-208,1171) Aperture: Medium						
55	NmADH9_i03		2	########	NmADH9_NAD-dep		3	0.97625	427.9818	0.04	0	0.5	211.676	205.5681	2.2 (-262,-192,1150) Aperture: Medium						
56	NmADH9_i03		3	########	NmADH9_NAD-dep		3	0.97625	264.535	0.04	0	0.5	211.6215	206.6305	1.5 (-262,-192,1150) Aperture: Medium						
57	NmADH9_i03		1	########	NmADH9_NAD-dep		3600	0.97625	213.4382	0.01	45	0.1	211.6045	206.9627	1.3 (-262,-192,1150) EDNAStrategy1: subWedge:1Aperture: Me						
58	NmADH9_i03		1	########	NmADH9_NAD-dep		3	0.97625	264.535	0.04	0	0.5	211.6215	206.6305	1.5 (-107,-190,1012) Aperture: Medium						
59	NmADH9_i03		2	########	NmADH9_NAD-dep		3	0.97625	264.535	0.04	0	0.5	211.6215	206.6305	1.5 (-107,-190,1012) Aperture: Small						
60	NmADH9_i03		1	########	NmADH9_NAD-dep		3	0.97625	264.535	0.04	0	0.5	211.6215	206.6305	1.5 (-156,-228,1396) Aperture: Medium						
61	NmADH9_i03		2	########	NmADH9_NAD-dep		3	0.97625	264.535	0.04	0	0.5	211.6215	206.6305	1.5 (-213,-228,1406) Aperture: Medium						
62	NmADH9_i03		2	########	NmADH9_NAD-dep		3	0.97625	264.535	0.04	0	0.5	211.6215	206.6305	1.5 (247,-388,861) Aperture: Medium						
63	sample	i03	12	########	NmADH9_NAD-dep		42	0.97625	602.2595	0.1	-0.00016	0	211.7341	204.4353	2.9966 Diffraction grid scan of 7 by 6 images, To left [350,177], Bot						
64	sample	i03	13	########	NmADH9_NAD-dep		12	0.97625	602.2595	0.1	89.99975	0	211.7341	204.4353	2.99662 Diffraction grid scan of 3 by 4 images, Top ft [472,245], Bot						
65	NmADH9_i03		3	########	NmADH9_NAD-dep		3	0.97625	264.535	0.04	0	0.5	211.6215	206.6305	1.5 (276,-373,844) Aperture: Medium						
66	NmADH9_i03		4	########	NmADH9_NAD-dep		3	0.97625	382.8205	0.04	0	0.5	211.6609	205.8617	2 (276,-373,844) Aperture: Medium						
67	NmADH9_i03		1	########	NmADH9_NAD-dep		3600	0.97625	385.1022	0.01	103	0.1	211.6617	205.8468	2.01 (276,-373,844) EDNAStrategy1: subWedge:1Aperture: Medi						



What you did...

(1) ISPyB interface



(2) Visit PDF (combine with others...)

mx13467-41 on i03 at 29-07-2017 12:00

Sample	Images	Res	λ	Ω Osc	Spacegroup	Unit Cell	Processed Resolution	Rmeas	Completeness	Comments
NmADH9_23	3600	1.3	0.9763	0.10	P 1 2 1	63.92, 107.75, 69.36 90.00, 104.27, 90.00	29.77 - 1.5 29.77 - 6.7 1.54 - 1.5	0.085 0.036 0.824	98.8 98.7 96.6	(-262,-192,1150) EDNAStrategy1: subWedge:1Aperture: Medium

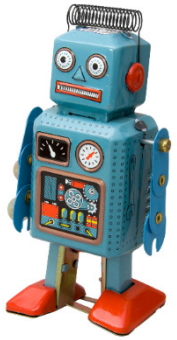
(3) Visit Excel sheet (annotate...)

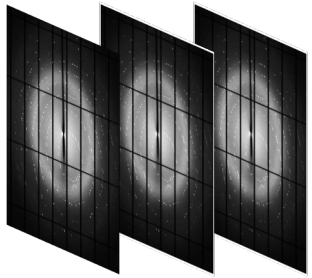
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1		mx13467-41_i03_29-JUL-2017 - 7 hr - Pilatus3 6M detector (100 Hz)																		
2																				
3	data	Image prefix	Run	Sta	Sam	Prot	# imag	Wavele	Dista	Exp.	Phi st	Phi ra	Xbea	Ybea	Det	auto/m	resoln	space gr	cell	twinnin	comments
54		NmADH9_23	2	##	NmA	NAD	3	0.9763	428	0.04	0	0.5	212	206	2.2						
55		NmADH9_23	3	##	NmA	NAD	3	0.9763	265	0.04	0	0.5	212	207	1.5						
56	12	NmADH9_23	1	##	NmA	NAD	3600	0.9763	213	0.01	45	0.1	212	207	1.3	a/3dii	1.37	P21	64 108 69 / 90 104 90	2.09	Binary complex with NAD - best data so far
57		NmADH9_24	1	##	NmA	NAD	3	0.9763	265	0.04	0	0.5	212	207	1.5						
58		NmADH9_24	2	##	NmA	NAD	3	0.9763	265	0.04	0	0.5	212	207	1.5						

Ideal scenario during session:

- Load first sample
- Collect test images
- Based on these, decide:
  - to collect...
  - not to collect...
  - to revisit later...
- For a “suitable” sample:
  - devise a data collection strategy
  - collect data set
- Analyse data as they are collected
- Based on this analysis, revise plans if appropriate
- Move onto next sample…

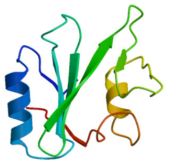
– not practical without automation





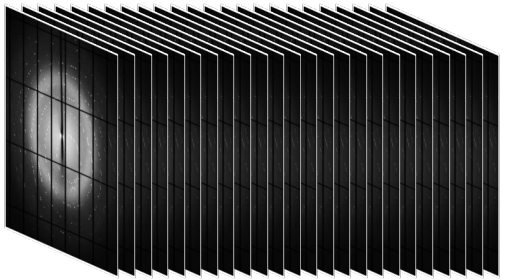
test images

before session...



template

sequence



dataset

< 1 min

Strategies:
Mosflm
EDNA
Xia2

ISPyB

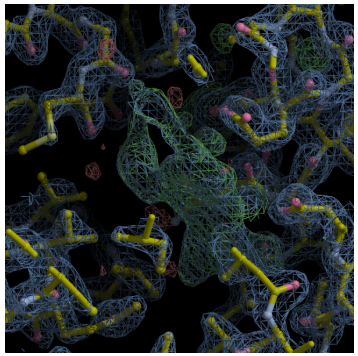
< 2 min

Fast DP

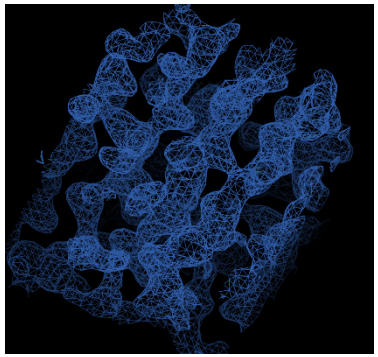
h	k	l	Iplus	SIGIplus	Iminus	SIGIminus
-45	0	4	-1.00	1.20	-1.00	1.20
-45	0	5	-0.03	1.82	-0.03	1.82
-45	0	6	2.17	2.01	2.17	2.01
-45	0	7	-0.22	1.24	-0.22	1.24
-45	0	8	0.63	1.33	0.63	1.33
-45	0	9	1.46	1.40	1.46	1.40
-45	0	10	0.11	1.34	0.11	1.34
-45	0	11	2.02	1.41	2.02	1.41

< 5 min

Quick feedback

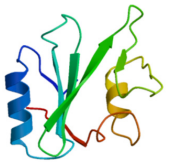


Dimple

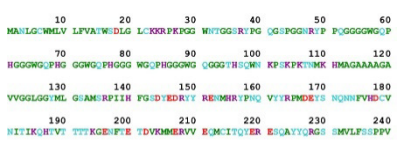


Fast EP

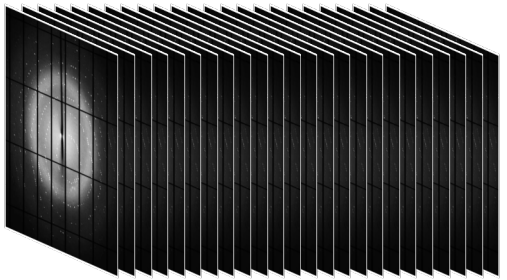
before session...



template



sequence

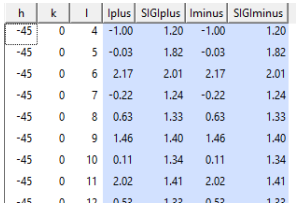


dataset

ISPyB

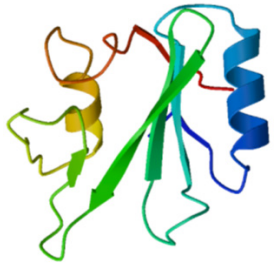
20-30 min

Xia2
DIALS
MultiXia2
autoPROC

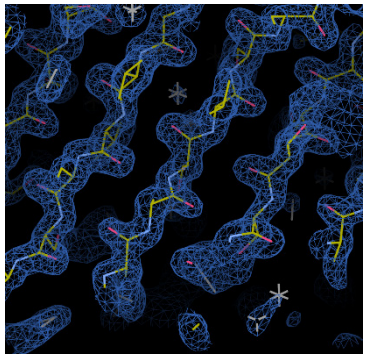


h	k	l	Iplus	SIGIplus	Iminus	SIGIminus
-45	0	4	-1.00	1.20	-1.00	1.20
-45	0	5	-0.03	1.82	-0.03	1.82
-45	0	6	2.17	2.01	2.17	2.01
-45	0	7	-0.22	1.24	-0.22	1.24
-45	0	8	0.63	1.33	0.63	1.33
-45	0	9	1.46	1.40	1.46	1.40
-45	0	10	0.11	1.34	0.11	1.34
-45	0	11	2.02	1.41	2.02	1.41

hours



MrBUMP

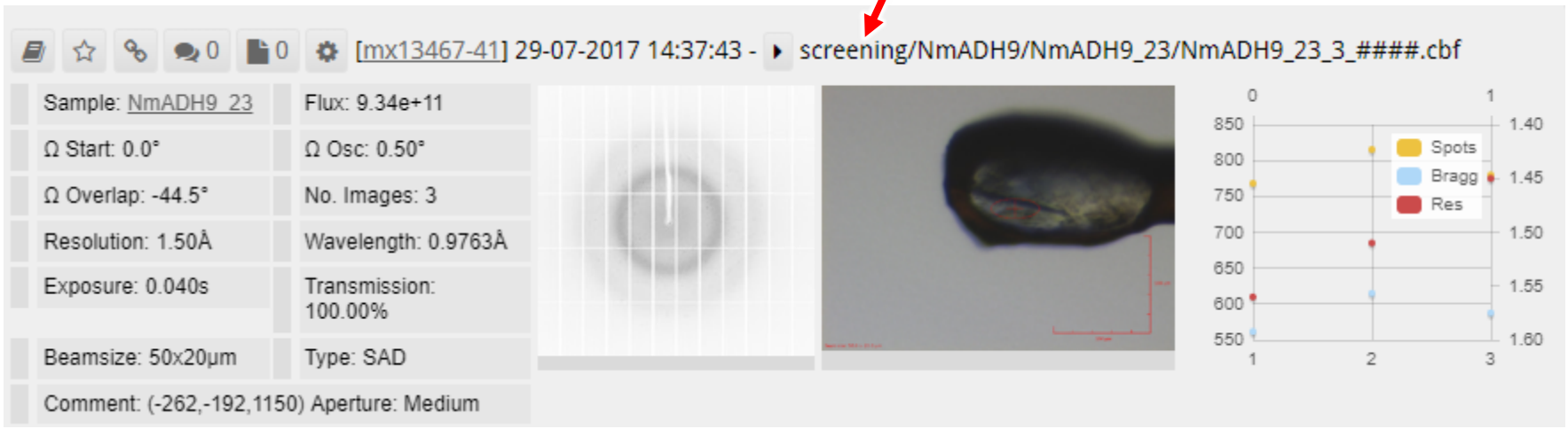


Big EP

Slow feedback

Screening…

(using the “screening” tab...)



Screening…


0
0
[mx13467-41] 29-07-2017 14:37:43 - screening/NmADH9/NmADH9_23/NmADH9_23_3_####.cbf
Sample: NmADH9_23
Flux: 9.34e+11
Ω Start: 0.0°
Ω Osc: 0.50°
Ω Overlap: -44.5°
No. Images: 3
Resolution: 1.50Å
Wavelength: 0.9763Å
Exposure: 0.040s
Transmission: 100.00%
Beamsize: 50x20µm
Type: SAD
Comment: (-262,-192,1150) Aperture: Medium
0
1
850
800
750
700
650
600
550
1.40
1.45
1.50
1.55
1.60
1
2
3
Spots
Bragg
Res
Strategies
Mosflm:
EDNA:
Xia2:

Screening…

Strategies — Mosflm: ✔ EDNA: ✔ Xia2: ✔

xia2.strategy

Space Group	A	B	C	α	β	γ
P 1 2 1	63.87	107.69	69.31	90.00	104.25	90.00

Lookup Cell

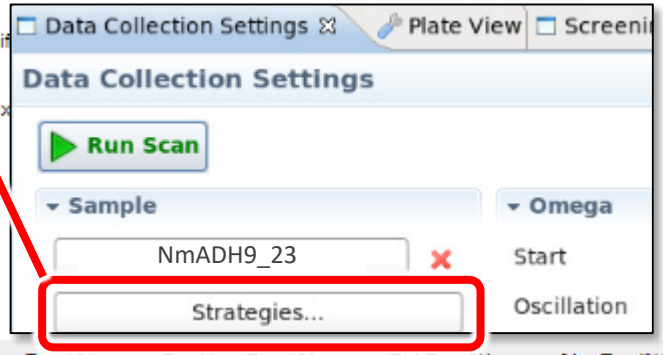
Strategy	Description	Ω Start	Ω Osc	Res (Å)	Ranking Res (Å)	Rel Trn (%)	Abs Trn (%)	Exposure (s)	No. Images
anomalous Wedge1	Standard Anomalous Dataset Multiplicity=3 I/sig=2 202s	0	0.15	1.51	1.17	4.0	4	0.010	2214
gentle Wedge1	Gentle: Target Multiplicity=2 I/sig=2 Maxlifespan=20s	6	0.15	1.51	1.29	4.0	4	0.010	2000
high multiplicity Wedge1	Strategy with target multiplicity=16 I/sig=2 202s	0	0.15	1.51	1.17	4.1	4.1	0.010	2400
native Wedge1	Standard Native Dataset Multiplicity=3 I/sig=2 Maxlifespan=202s	162	0.15	1.51	1.14	2.2	2.2	0.010	1947

EDNA MXv1

Space Group	A	B	C	α	β	γ
P2	63.99	107.49	69.28	90.00	104.37	90.00

Lookup Cell

Strategy	Description	Ω Start	Ω Osc	Res (Å)	Ranking Res (Å)	Rel Trn (%)	Abs Trn (%)	Exposure (s)	No. Images
Strategy1 Wedge1	Standard Native Dataset Multiplicity=3 I/sig=2 Maxlifespan=202 s	45	0.10	1.51	1.21	11.0	11	0.010	1630
Strategy2 Wedge1	Standard Anomalous Dataset Multiplicity=3 I/sig=2 Maxlifespan=202 s	78	0.10	1.51	1.24	12.4	12.4	0.010	2740
Strategy3 Wedge1	strategy with target multiplicity=16, target I/sig=2 Maxlif s						5.1	0.010	3600
Strategy4 Wedge1	Gentle: Target Multiplicity=2 and target I/Sig 2 an Max s						14.8	0.010	1110
Strategy5 Wedge1	UnderDEV Anomalous Dataset, RadDamage of stand						12.4	0.010	2740



mosflm

Space Group	A	B	C	α	β	γ
P2	63.80	107.65	69.44	90.00	104.13	90.00

Lookup Cell

Strategy	Description	Ω Start	Ω Osc	Res (A)	Ranking Res (A)	Rel Trn (%)	Abs Trn (%)	Exposure (s)	No. Images
anomalous Wedge1		196	0.20	1.46	0.00	0.0	0	0.000	525
native Wedge1		211	0.20	1.46	0.00	0.0	0	0.000	525

Collecting a dataset…

Space Group	A	B	C	α	β	γ
P2	63.99	107.49	69.28	90.00	104.37	90.00

Lookup Cell

Strategy	Description	Ω Start	Ω Osc	Res (Å)	Ranking Res (Å)	Rel Trn (%)	Abs Trn (%)	Exposure (s)	No. Images
Strategy1 Wedge1	Standard Native Dataset Multiplicity=3 I/sig=2 Maxlifespan=202 s	45	0.10	1.51	1.21	11.0	11	0.010	1630

Data Collection Settings | Plate View | Screening | Command Queue

Data Collection Settings

Run Scan

Sample

NmADH9_23

Strategies...

Barcode: NR

Holder: 2

Position: 14

Files

Visit directory: /dls/i03/data/2017/mx13467-41

Folder: Configure Defaults

JIC/${proteinacronym}/${samplena

JIC/NmADH9/NmADH9_23

Prefix: Configure Defaults

${samplename}

NmADH9_23

☑ Automatic run number

Run number: 0

Comment: EDNAStrategy1: subWedge:1

Omega

Start	45.00	°
Oscillation	0.100	°
Total oscillation	360.0	°
Delta	0.00	°

Image

Number of images	3600	
Exposure time	0.010	s
Total exposure time	36.0	s
First image number	1	

Beam and Detector

Maximum resolution	1.3000	Å
Detector distance	213.5	mm
Wavelength	0.97623	Å
Energy	12700.3	eV

☐ Use current energy

Transmission: 50.156283 %

make sure you collect enough data – still could be P1...

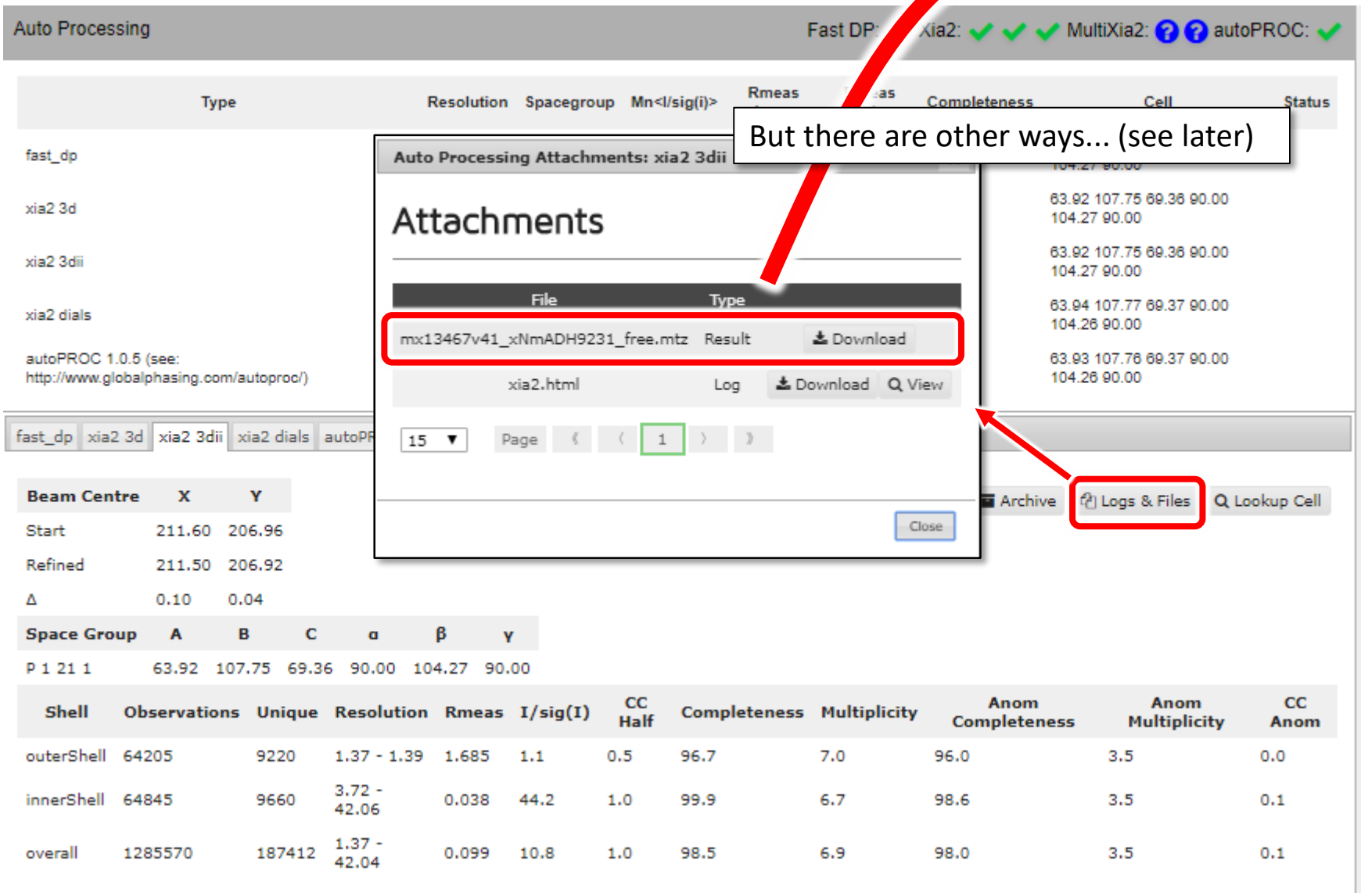
push the resolution a little...

increased transmission...

Checking the results…

Checking the results…

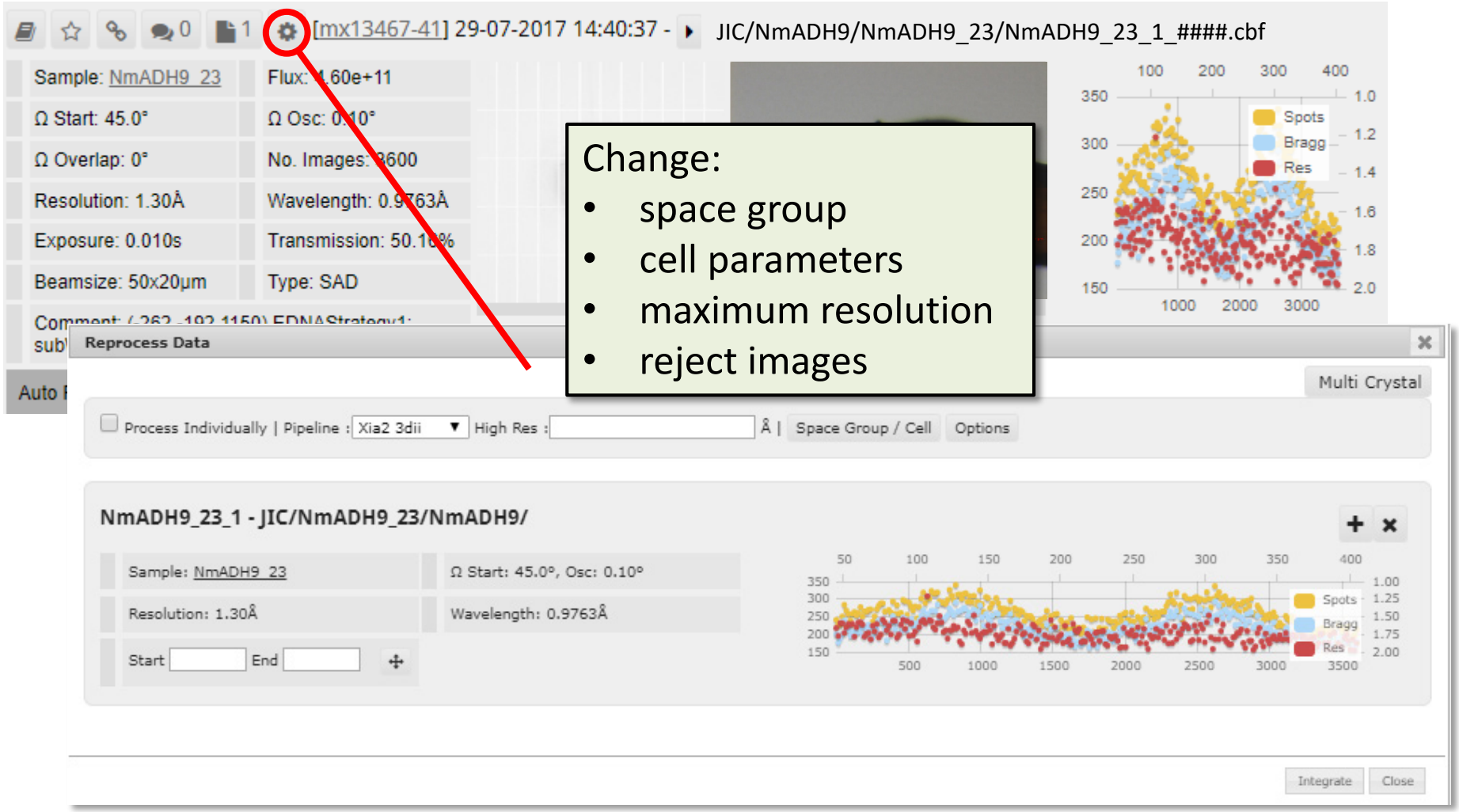
Import directly into CCP4i2



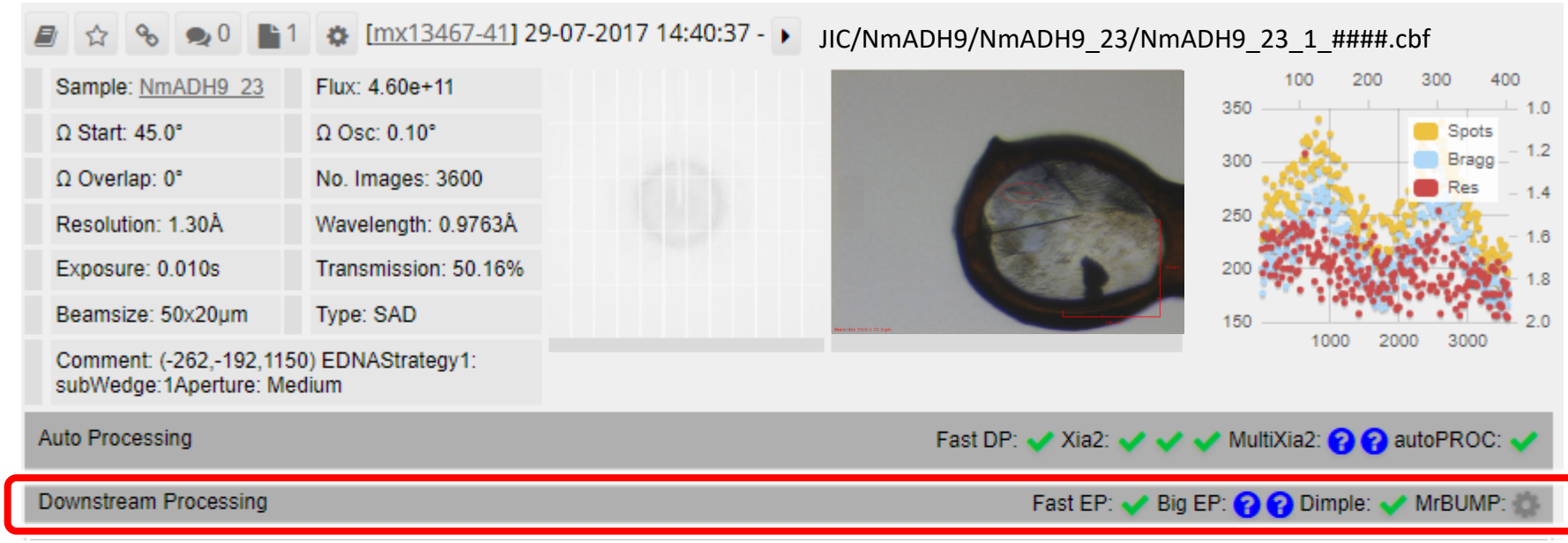
N.B. space group is only a hypothesis at this stage! – see talk about data processing

Re-running jobs:

- Most pipelines will run from the command line (Terminal window)
- Also through ISPyB interface...



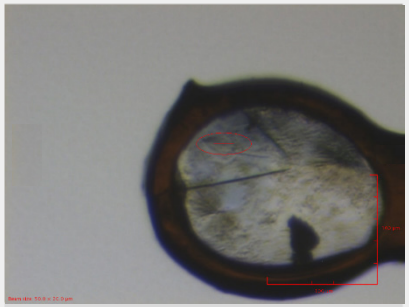
Checking the results…

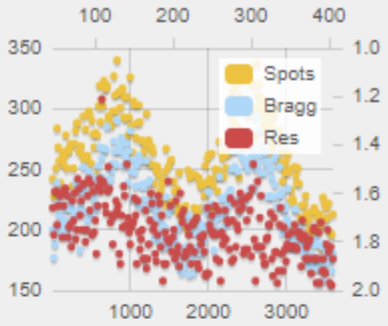


Checking the results…

0 1 [mx13467-41] 29-07-2017 14:40:37 - JIC/NmADH9/NmADH9_23/NmADH9_23_1_####.cbf

Sample: NmADH9_23	Flux: 4.60e+11
Ω Start: 45.0°	Ω Osc: 0.10°
Ω Overlap: 0°	No. Images: 3600
Resolution: 1.30Å	Wavelength: 0.9763Å
Exposure: 0.010s	Transmission: 50.16%
Beamsize: 50x20µm	Type: SAD
Comment: (-262,-192,1150) EDNAStrategy1: subWedge:1Aperture: Medium	





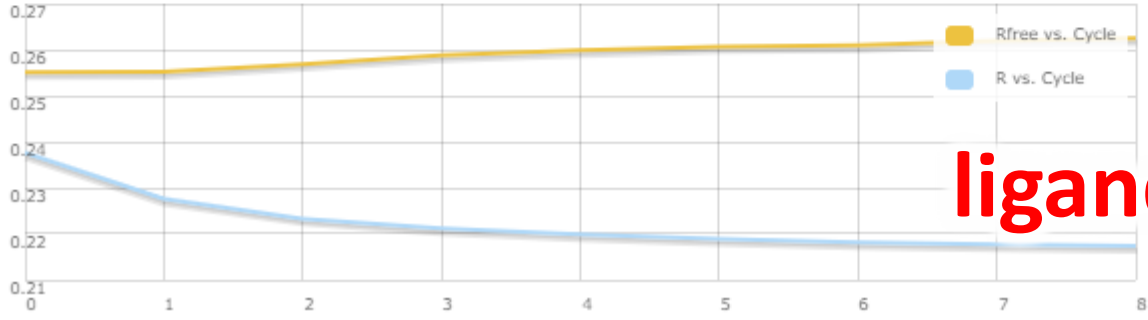
Auto Processing — Fast DP: ✔ Xia2: ✔ ✔ ✔ MultiXia2: ? ? autoPROC: ✔

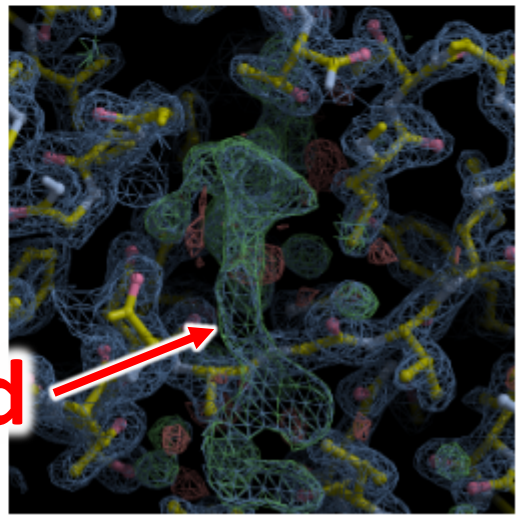
Downstream Processing — Fast EP: ✔ Big EP: ? ? Dimple: ✔ MrBUMP: ⚙

Fast EP | Dimple

Map / Model Viewer | PDB/MTZ file | Log file

Parameter	R factor	R free	Rms BondLength	Rms BondAngle	Rms ChirVolume
Initial	0.2376	0.2551	0.0174	1.7446	0.1121
Final	0.2174	0.2625	0.0178	1.8068	0.1061





Looking at maps in Coot (on DLS computer):

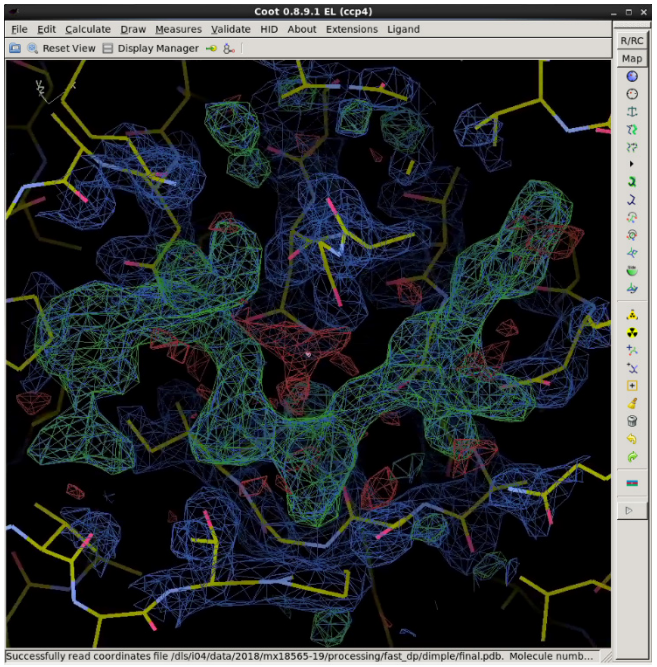
Using Terminal…

...go to the appropriate dimple directory in “processed”:

>module load coot

>coot

read in “final.pdb” and “final.mtz”
Validate... Unmodelled blobs...



On the basis of a “positive” result, may decide to move onto a different project...

Top tip: setup directory aliases (shortcuts) to avoid too much typing:

- e.g. alias m41=”cd /dls/i03/data/2017/mx13467-41” (put in .bashrc file)

Ideal scenario after session:

- Each data set characterised as:
  - useful/may be useful/not useful
- Some datasets processed to your satisfaction
- You may have interpretable experimental maps
- You may have preliminary structures
- You **will** have less follow-up work to do!

What to do with all the data...

Raw data (images)
- removed from disk after 40 days – still available through TopCAT (tape archive)

"Meta" data – all the other "stuff" – removed from disk after 40 days – not backed up

>90% of useful datasets derived from **Meta data** rather than going back to **Raw data**

...just be thankful it's not cryo-EM!

Getting your data home…

- FTP data home (can use a script)

Slow...

globus

Faster?

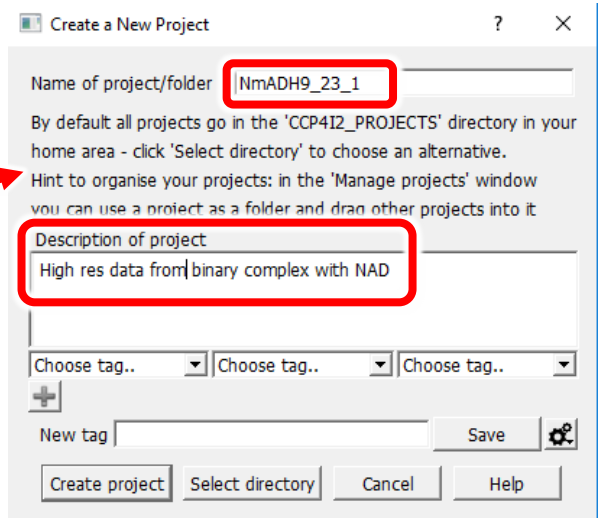
In the meantime:

- Use autoprocessing output or...
- (re)process data remotely on DLS computers and FTP output only
- archive (compressed) raw data later

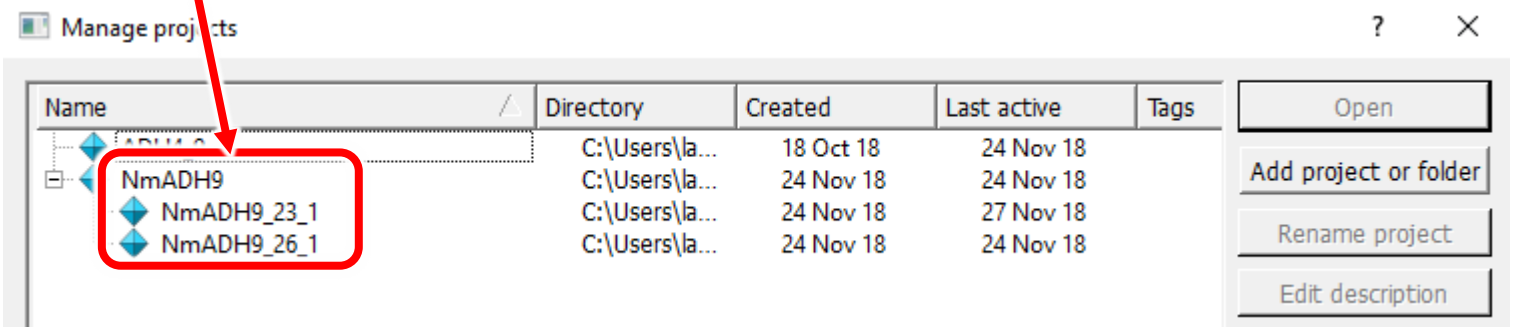
Retain ISPyB name in CCP4i2...

CCP4i2

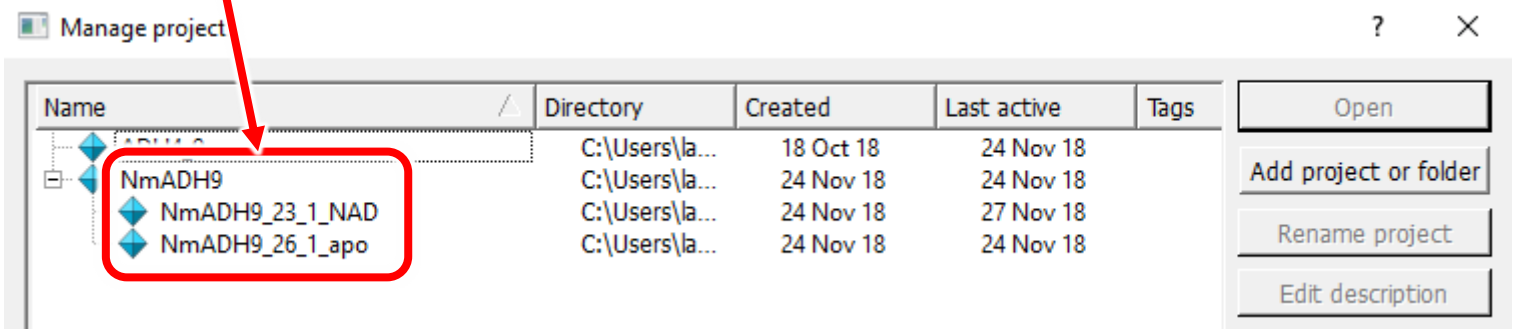
Create a separate "project" for each dataset



If you have more than one dataset for a given protein – create a separate "dummy project" and group the "dataset projects" under this:



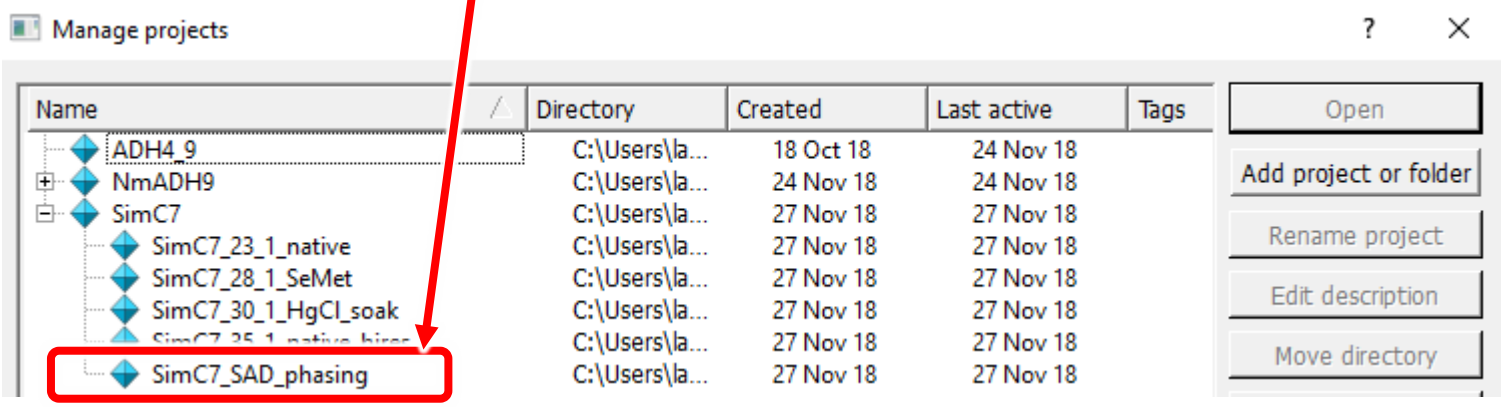
Can rename "projects" later... (when you know more...)*



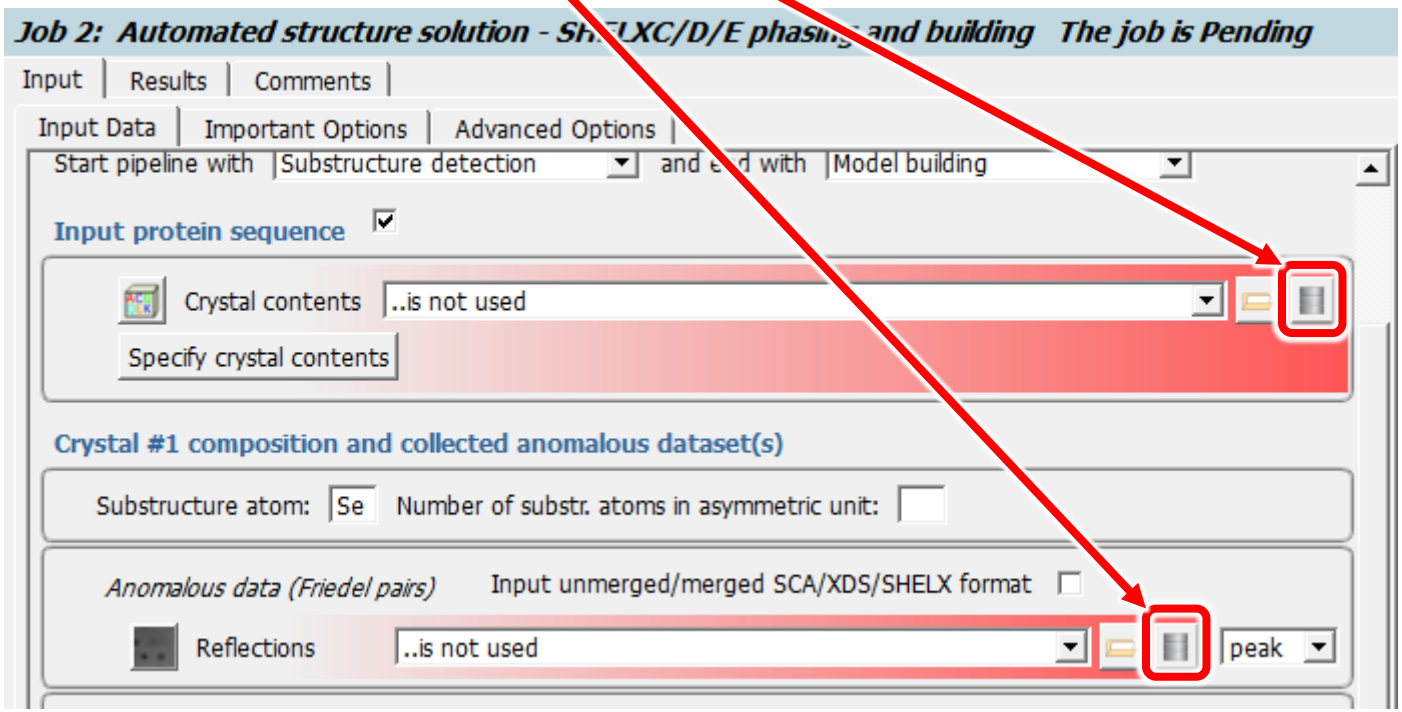
*but directory names remain the same

CCP4i2

Could create “non-dataset projects” for specific activities such as phasing...

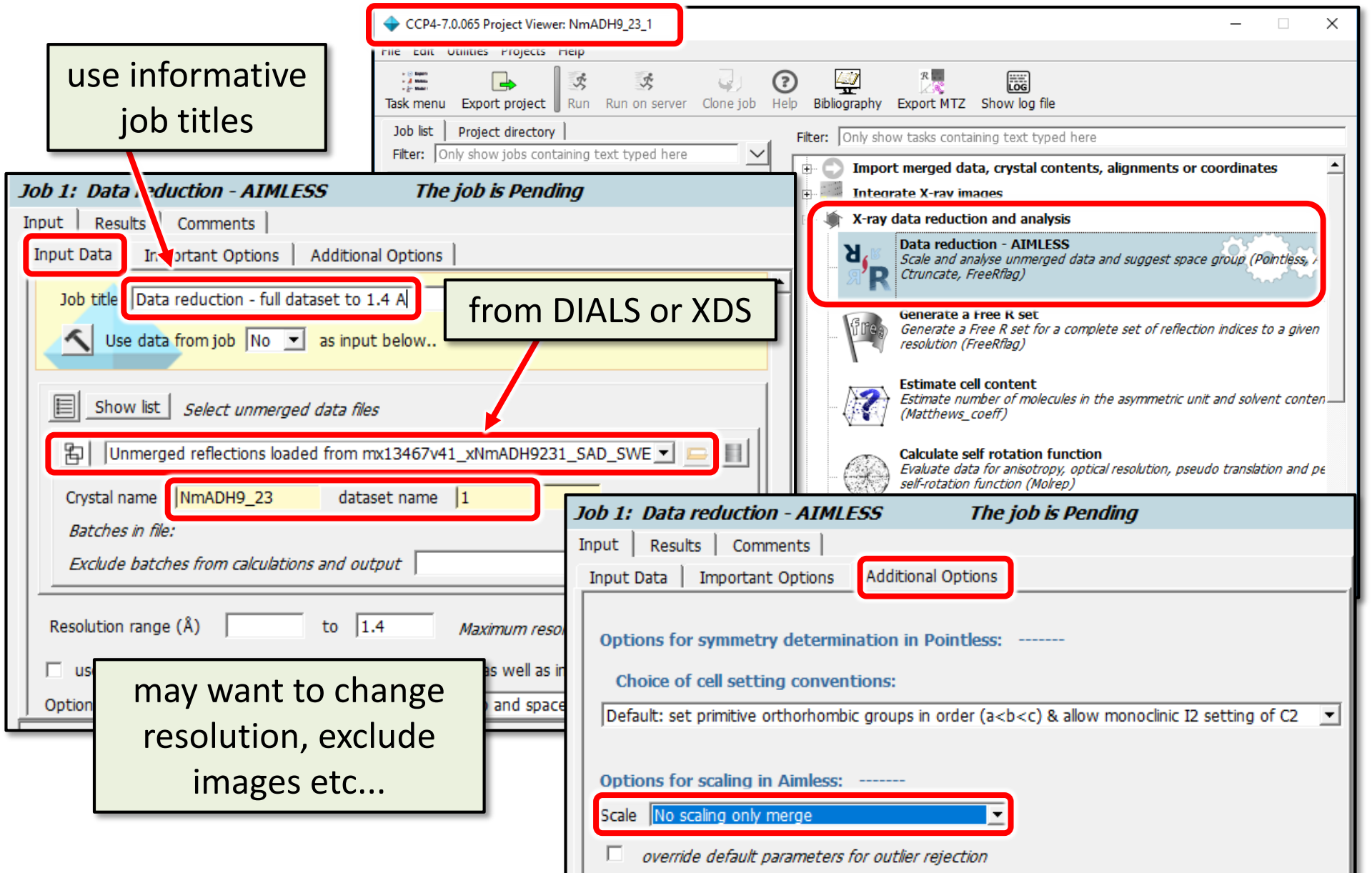


...then borrow files from other projects...



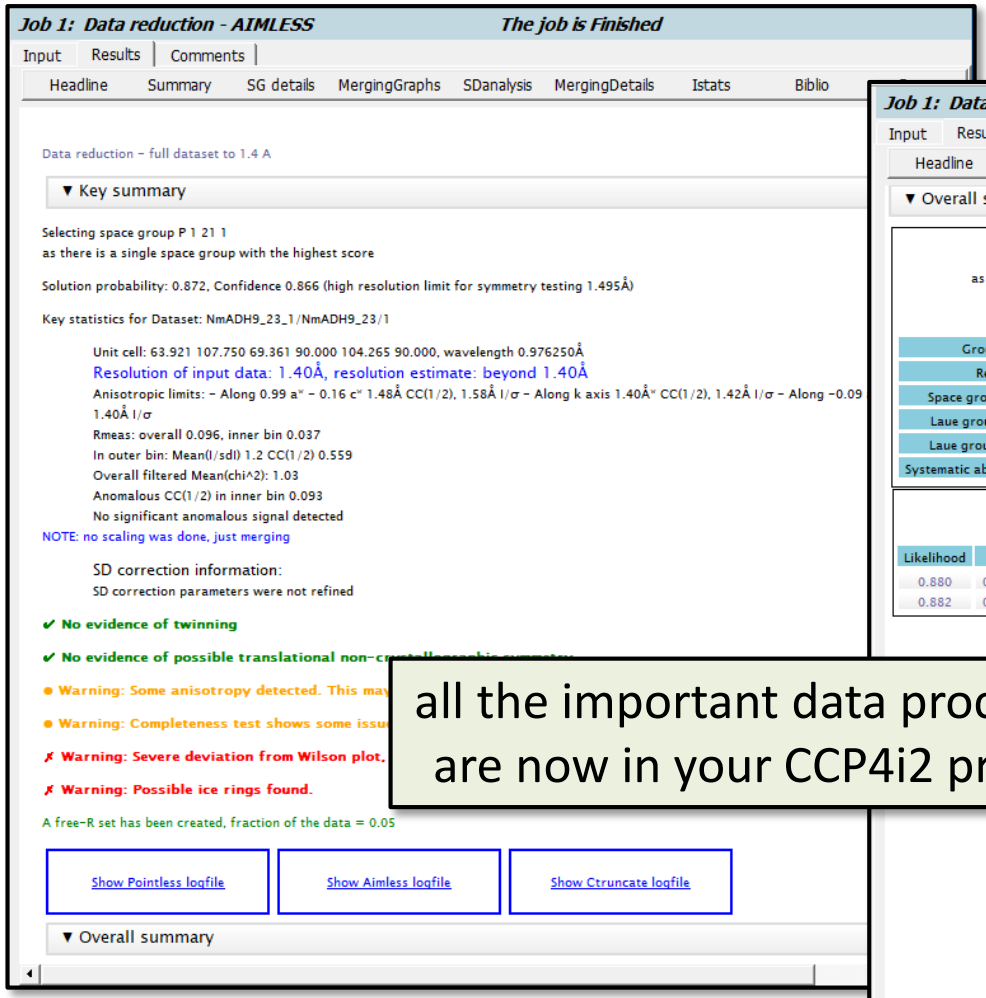
My preferred option for processing...

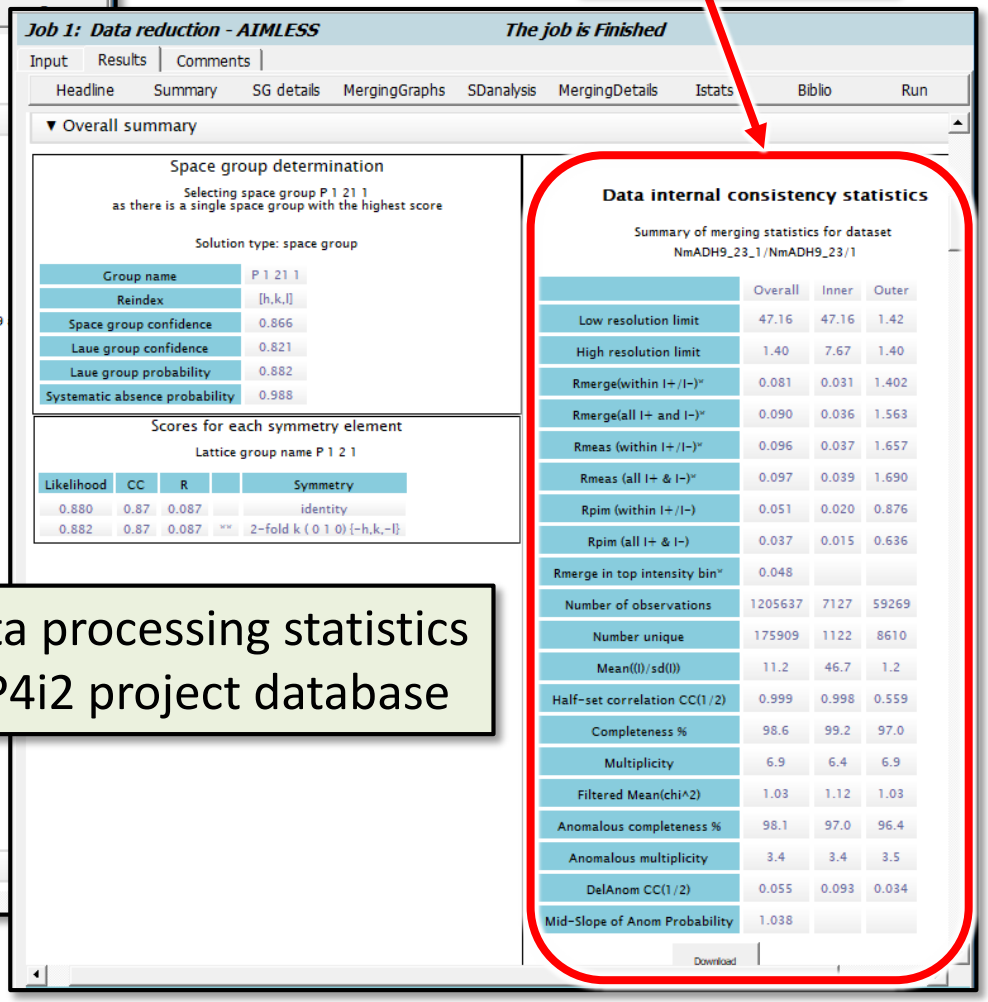
- where the pipelines have done a good job – take part-processed dataset (from the "Meta" data) and re-run the merging step in CCP4i2



Gives lots of output – useful for troubleshooting...

for your paper/thesis





all the important data processing statistics are now in your CCP4i2 project database

...or reprocess from scratch using the Dials User Interface

Integrate images with Dials
Launch DUI and capture output

DIALS talk this afternoon...

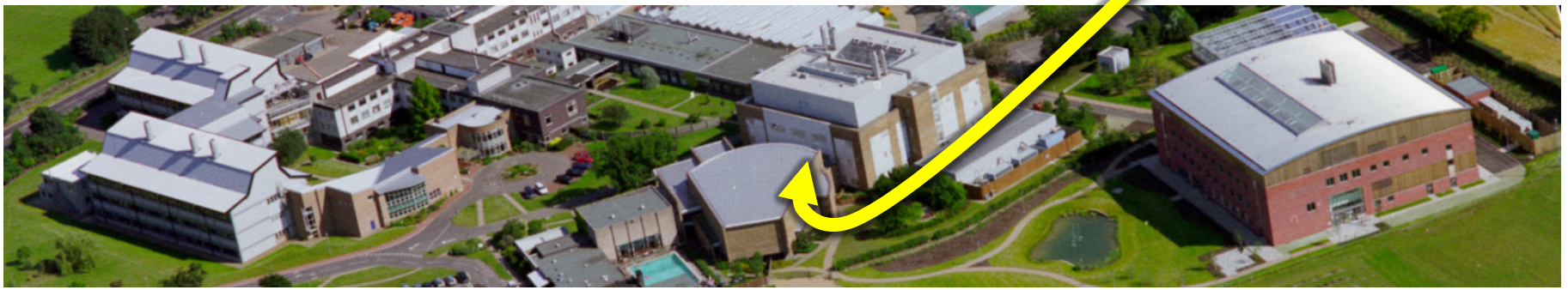
Summary - why use ISPyB & pipelines?

- Faster sample changing (select by ISPyB name)
  - essential for remote…
- Informs the decision making process
  - make decisions sooner
  - revise strategy on the fly
    - (e.g. recollect dataset x... no more data required for project y...)
  - make better overall use of beamtime
- Easier to cope with a small team
- Reduces amount of post-beamtime follow-up work
- Simple to keep track of your samples and data

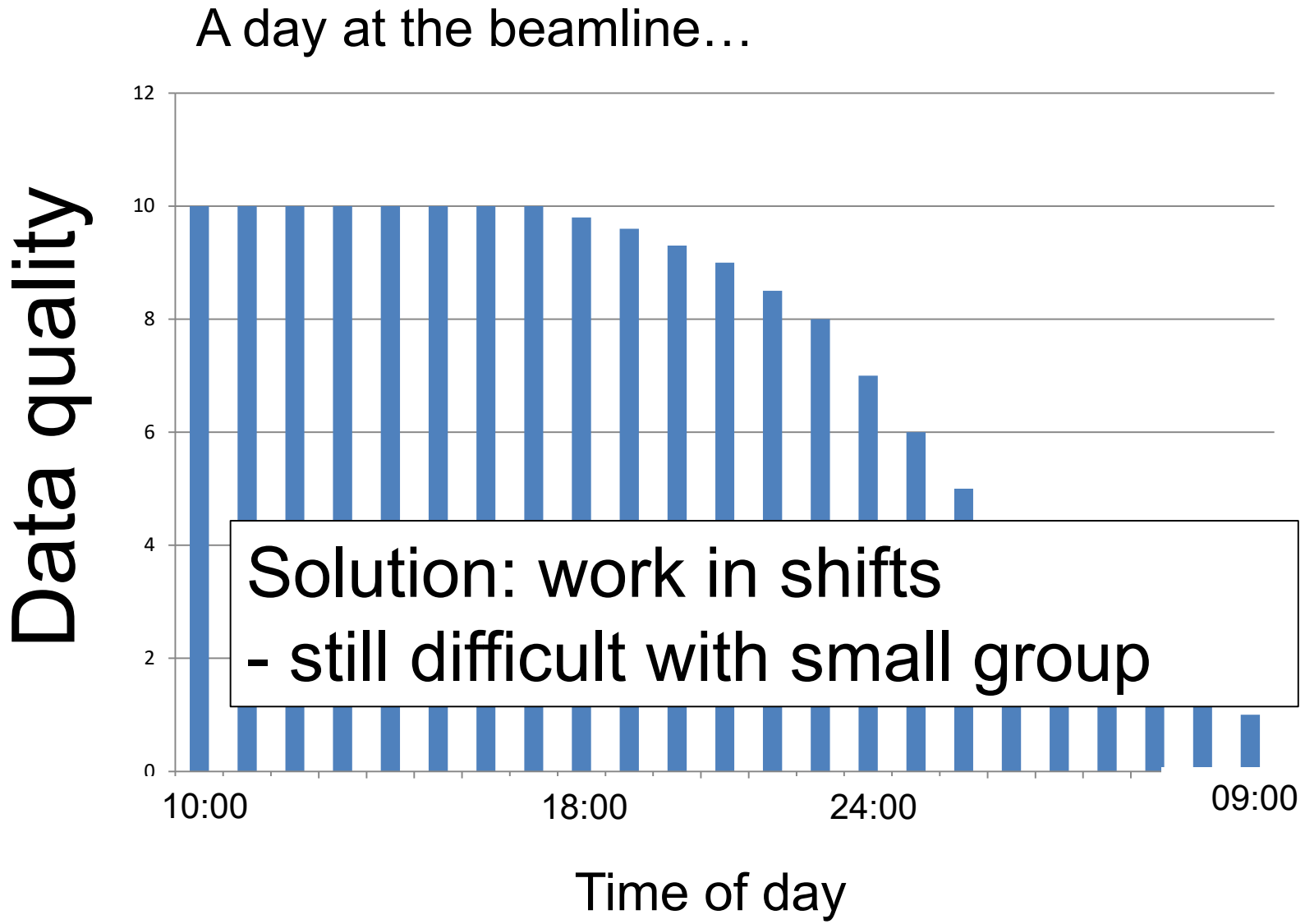
data

Remote collection

…from here



Some fictitious data…



Who should collect the data?

- Someone who knows the project
  - what data are required?
  - what space group/cell parameters/resolution to expect
- Someone who has a stake in the project
- Person who grew and mounted the crystals
- Someone who is not tired!

Ideal scenario:

- everyone collects data from their own crystals

- work in short shifts

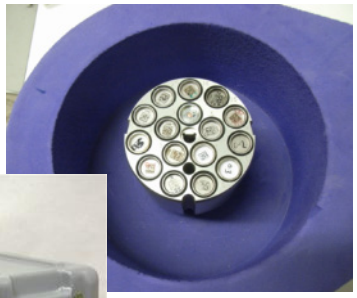
Solutions:

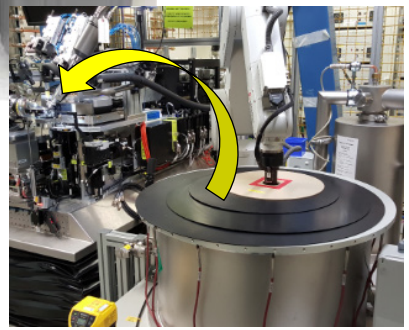
1. everyone travels to Diamond (not practical … and expensive)

2. small team travels to Diamond – everyone else collects data remotely (mixed access)

3. everyone collects data remotely (fully remote access)
   - ➢ beamtime can be shared by several groups at multiple remote sites

For routine data collection at 100 K:

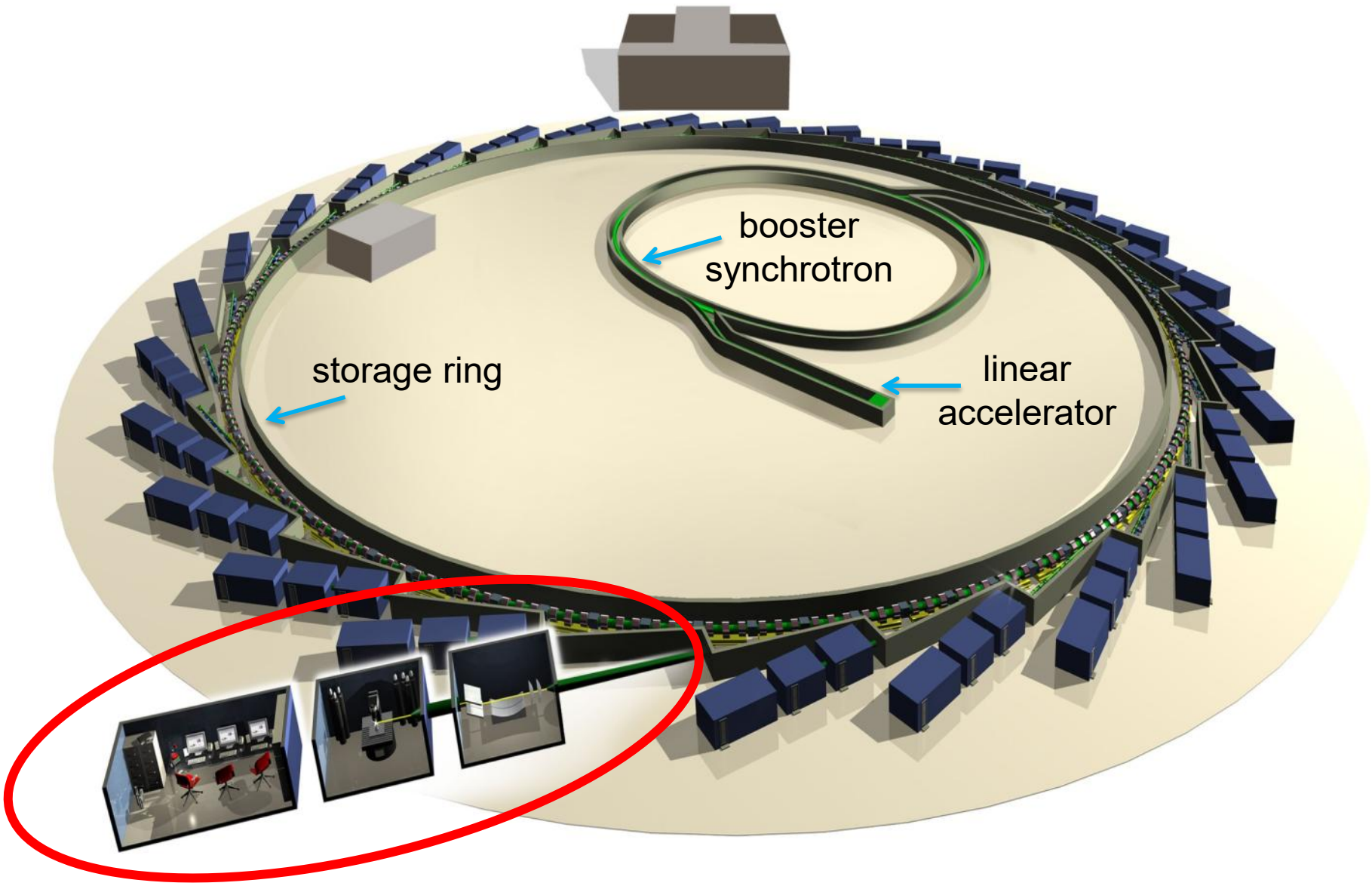


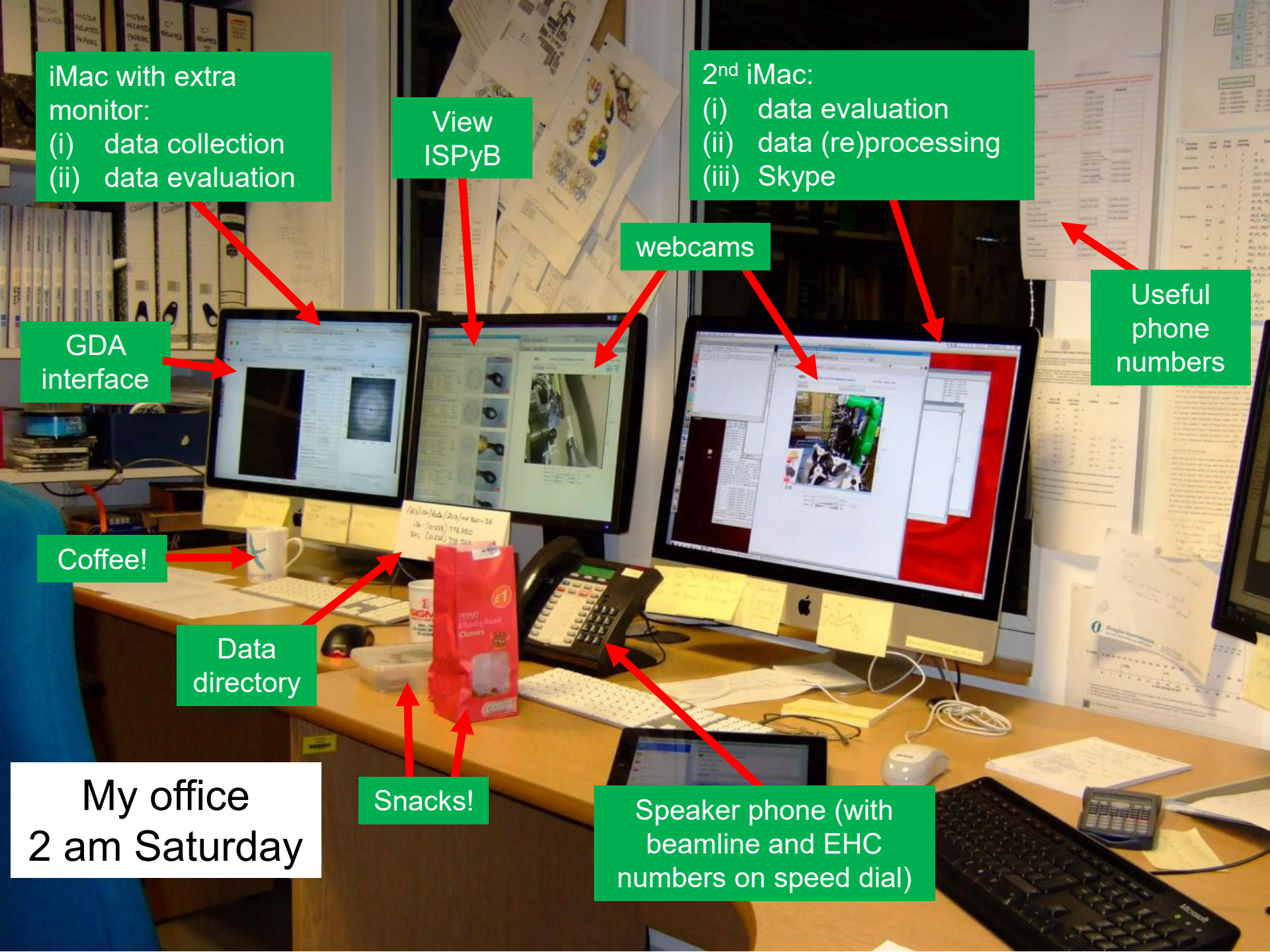


- Samples prepared and cryo-cooled in home lab
- Sample information entered into ISPyB database
- Transported to Diamond in dry shipping dewars
- Samples mounted robotically
- Data collection controlled through GDA
- Only manual operation at DLS: loading/unloading pucks
  - Up to 592 samples can be loaded at once
  - may only need to enter hutch at beginning and end of session
- Everything else computer driven…

Therefore you don’t need to be there!

What is remote data collection?

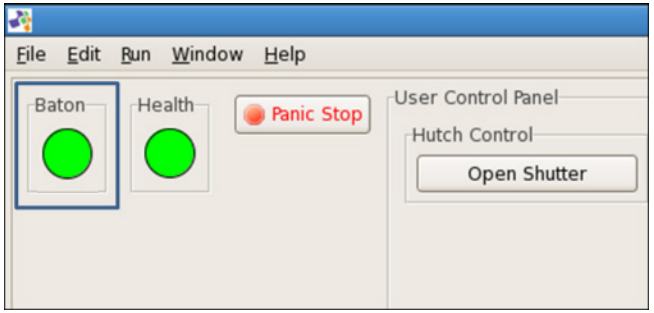



iMac with extra monitor:
(i) data collection
(ii) data evaluation
View ISPyB
2nd iMac:
(i) data evaluation
(ii) data (re)processing
(iii) Skype
webcams
Useful phone numbers
GDA interface
Coffee!
Data directory
Snacks!
Speaker phone (with beamline and EHC numbers on speed dial)
My office
2 am Saturday

Timeline for fully remote data collection

Sometime prior to shift:

- LC loads pucks into beamline dewar and enters puck positions into ISPyB database

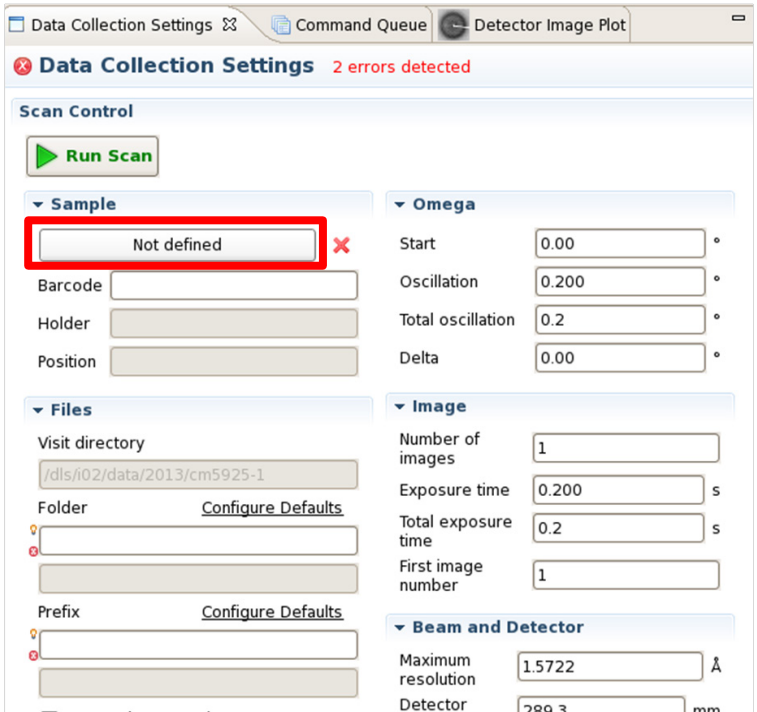


Shortly before start of shift (or slot in running order):

- connect to beamline computer using NX client
- start GDA

Zero hour:

- take the baton, select sample from menu, collect data!



When things go wrong…

Can **you** fix things?

- probably not!

On-site user:

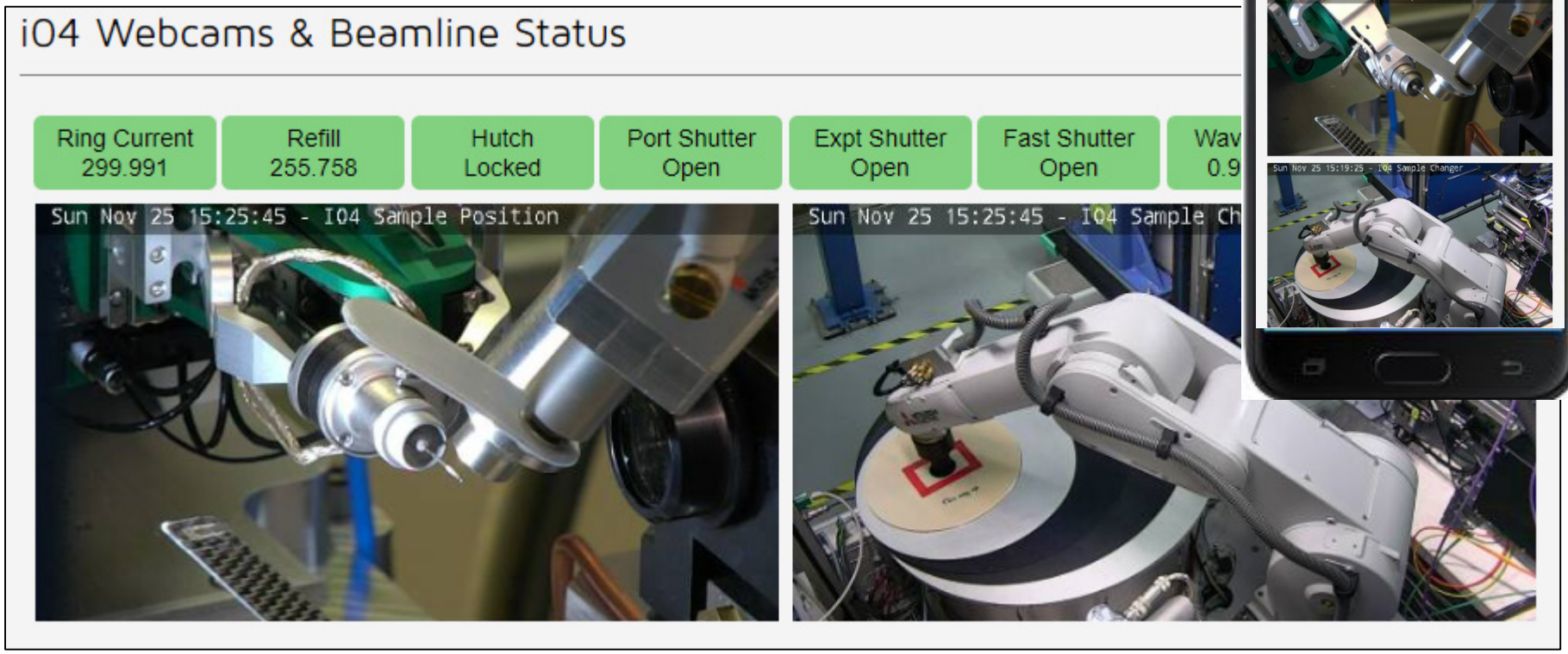
- During normal working hours – call local contact
- Out of hours – call EHC

Remote user:

- During normal working hours – call local contact
- Out of hours – call EHC

Therefore you don’t need to be there!

Check the webcams!



If you suspect a problem – call the EHC!

Advantages of remote data collection:

- Beamtime is fully used
- Users collect data on their own crystals
- Most/all of the users stay at home labs
  - time commitment is low for remote users
  - stakeholders can observe data collection
  - useful for training non-experts
- Time can be used flexibly
- Less stressful
- Everyone gets some sleep!
- Difficult to “break” the beamline

Take home messages

- If you are not using remote data collection – why not?
- If you are not using ISPyB – why not?
- If you are not using the software tools/pipelines – why not?
- If you don’t have a life – you know why….

This is your last experiment

– don’t mess it up!

Acknowledgements

- Access to MX beamlines at Diamond
- Excellent support of:
  - Beamline staff
  - EHC/Control Room staff
  - User Office
- Software developers
- BBSRC

John Innes Centre

UKRI Biotechnology and Biological Sciences Research Council