



John Innes Centre

*Unlocking Nature's Diversity*

# Optimising the Diamond experience from a user's perspective

Dave Lawson

Department of Biochemistry and Metabolism

John Innes Centre, Norwich, UK



Biotechnology and  
Biological Sciences  
Research Council

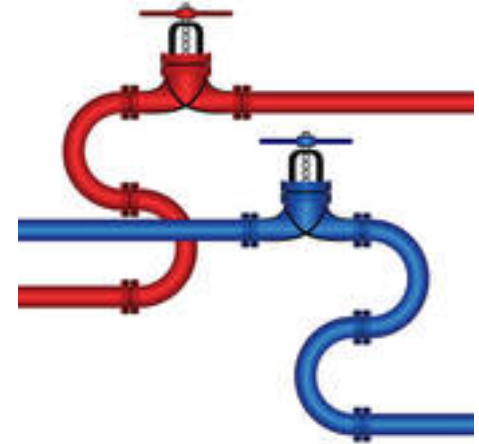
# Strategy



# 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

# 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 (make your life easier!)

# Primary goals (data quantity AND quality...)

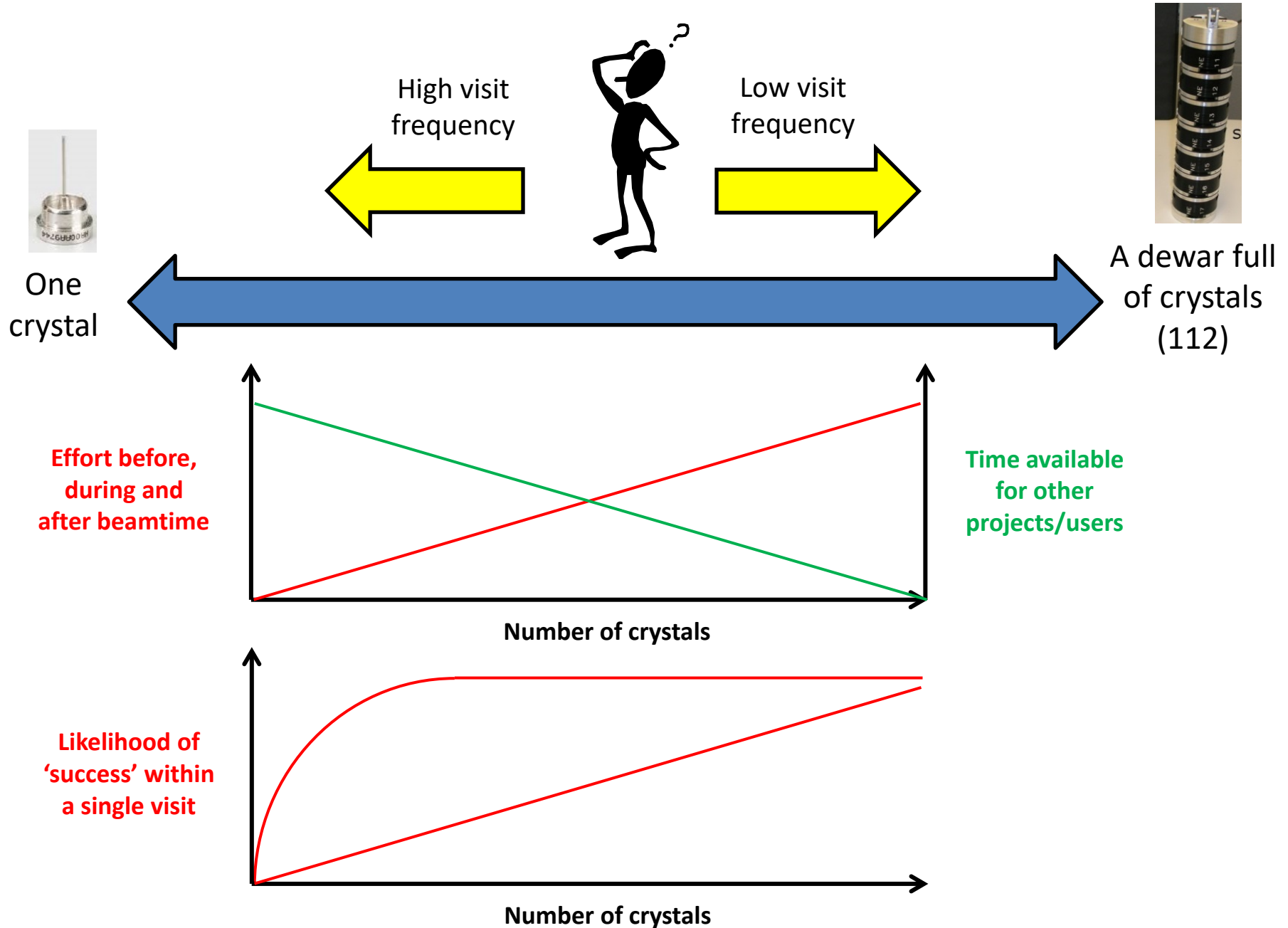
- To collect as much data as possible...
- To collect the best possible data...
- To collect the data that will enable me to:
  - Solve my structure
  - Extend the resolution of my structure
  - Show that my ligand is bound

## Secondary goal (make your life easier!)

- Try to answer these questions during your beamtime:
  - 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 prepare per project...



# 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?*



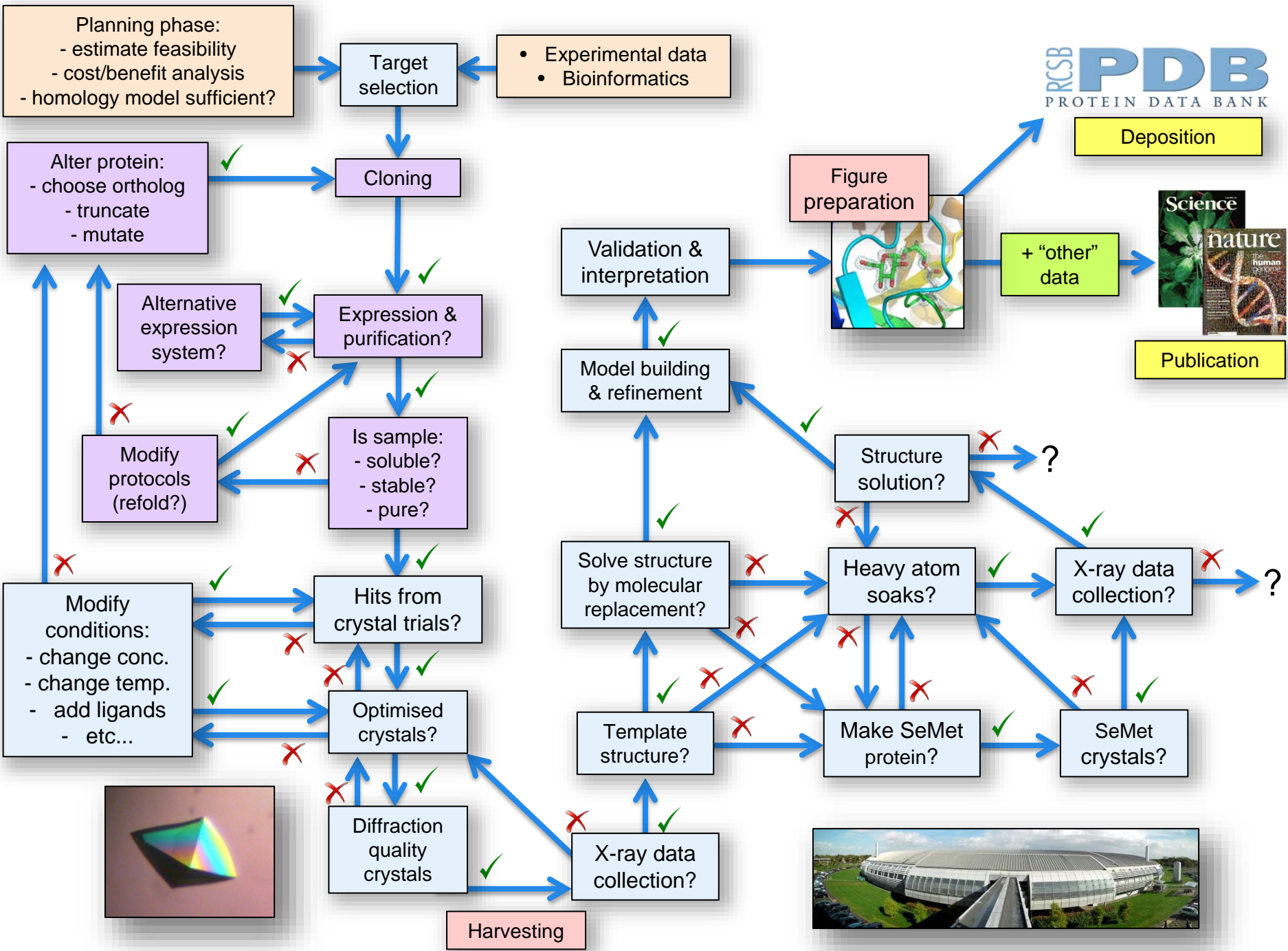
*How many  
crystals do  
I have?*

*When is it my turn?*



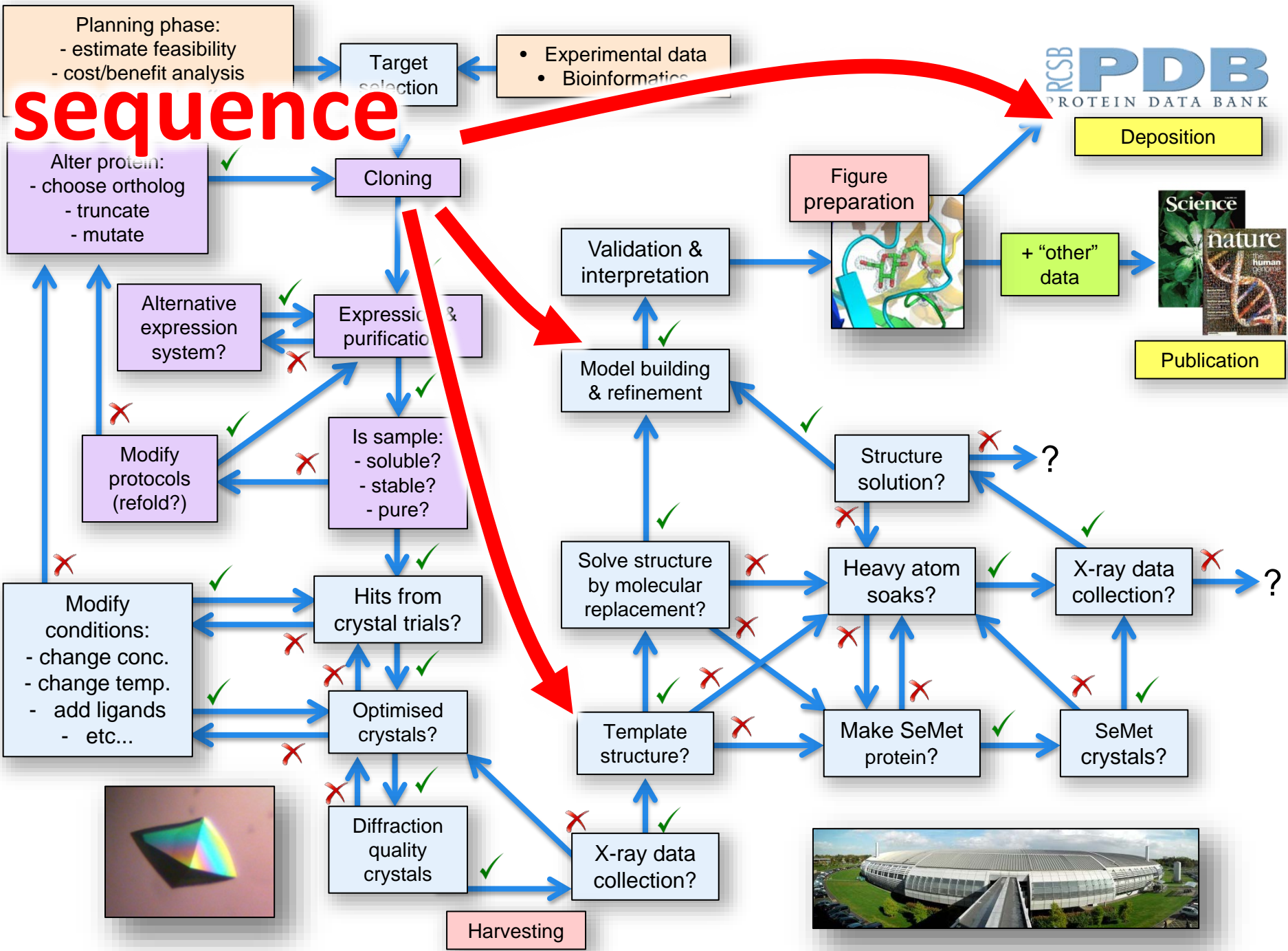
Be prepared!

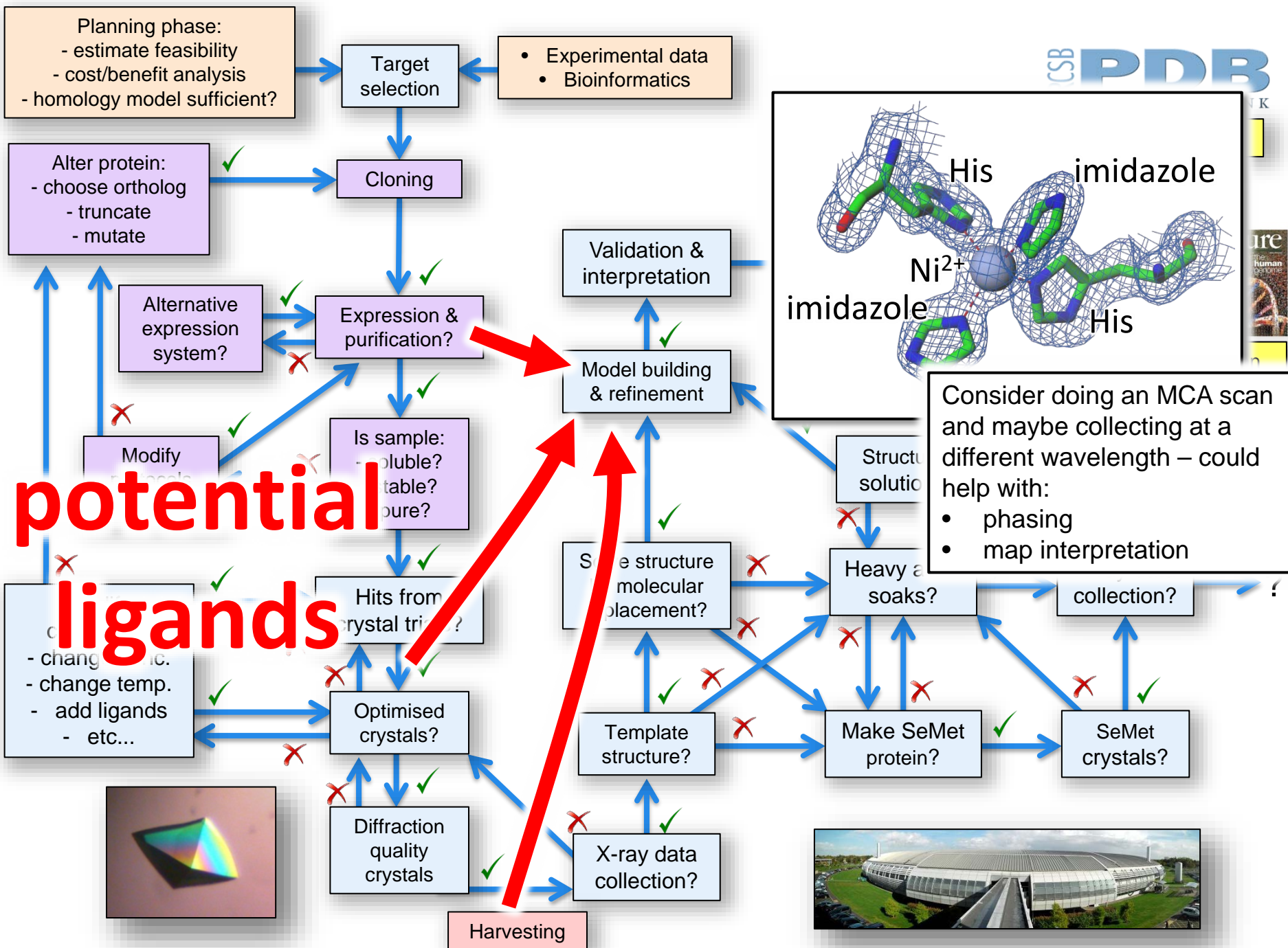
Think about this **before** your beamtime!

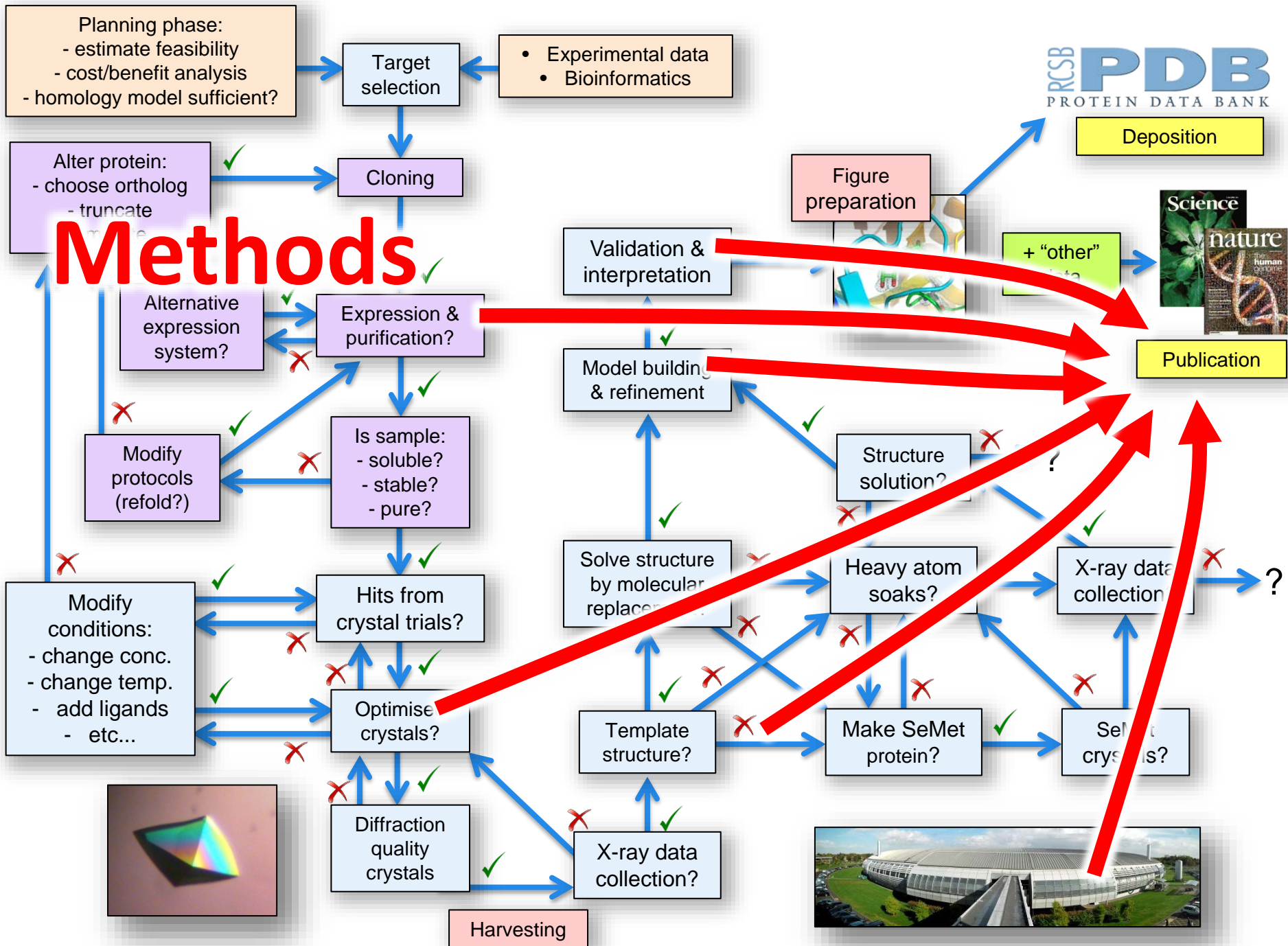




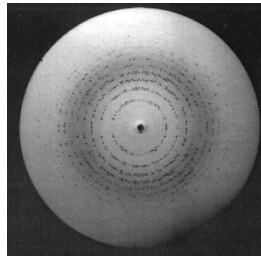
# sequence







# Detector types



Film



Image plate



CCD (charge-coupled device)



HPC (hybrid photon counting)

**360° data sets possible in <10 s!**

hours

readout time

1 ms

1990

2000

2010

2020

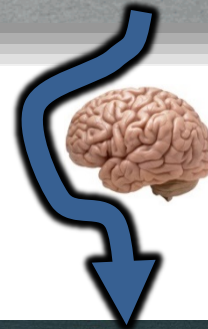


# MX data collection has become faster...

20<sup>th</sup> century data collection



21<sup>st</sup> century data collection





# MX data collection has become much faster...

- In principle, could collect one dataset every 5 mins  
(mounting/aligning + evaluating test images + collection)
- 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...

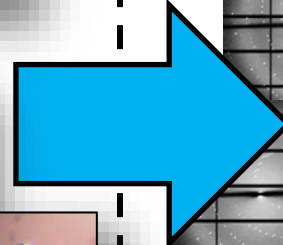
# Crystals

*Which  
protein did I  
crystallize?*

*Was it the  
wild-type or  
mutant?*

*Did I add  
any ligands?*

*What else  
was in that  
crystal?*



# Data

*Which  
crystal gave  
this dataset?*

*How did I  
collect the  
data?*

*Is this the best  
dataset for  
the sample?*

*Are the data  
"good enough"*

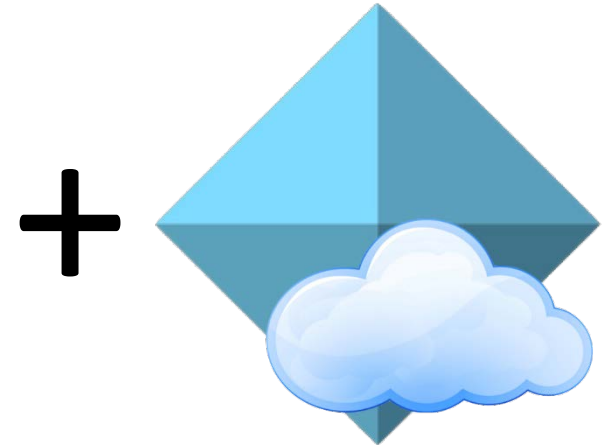
Taking back control...



*ISPyB*



*CCP4i2*



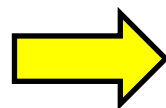
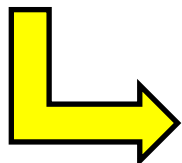
*CCP4 Cloud*

**F•R•I•E•N•D•S**



# ISPyB database

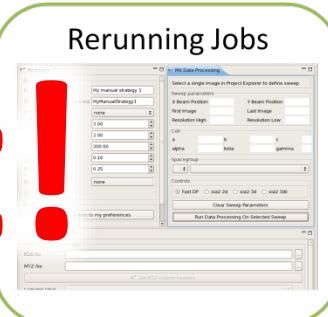
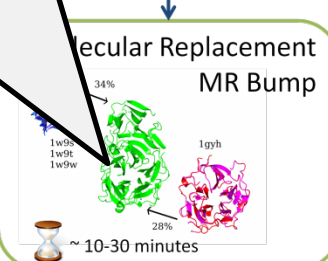
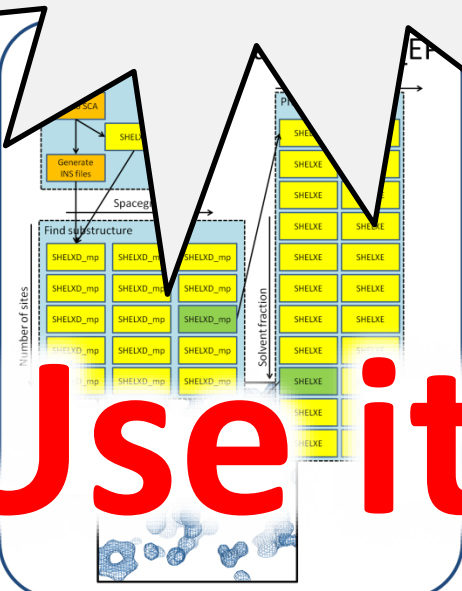
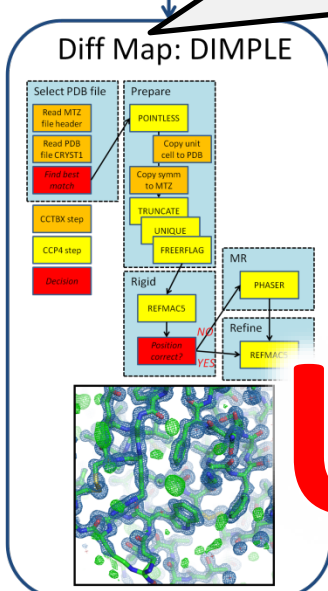
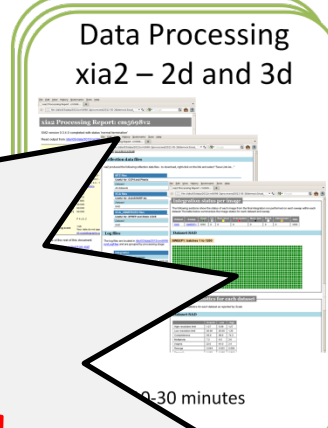
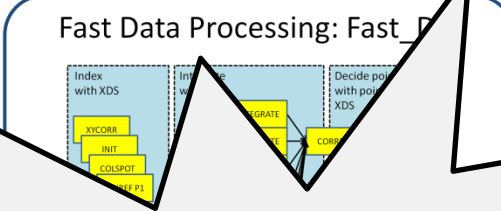
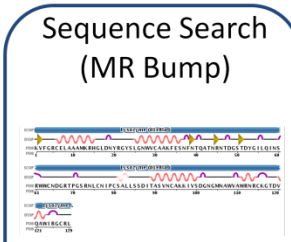
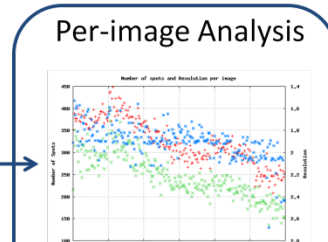
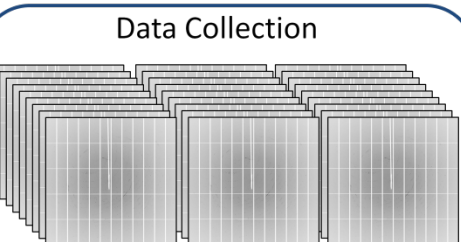
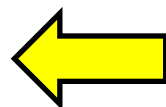
Sample info.



(test images)

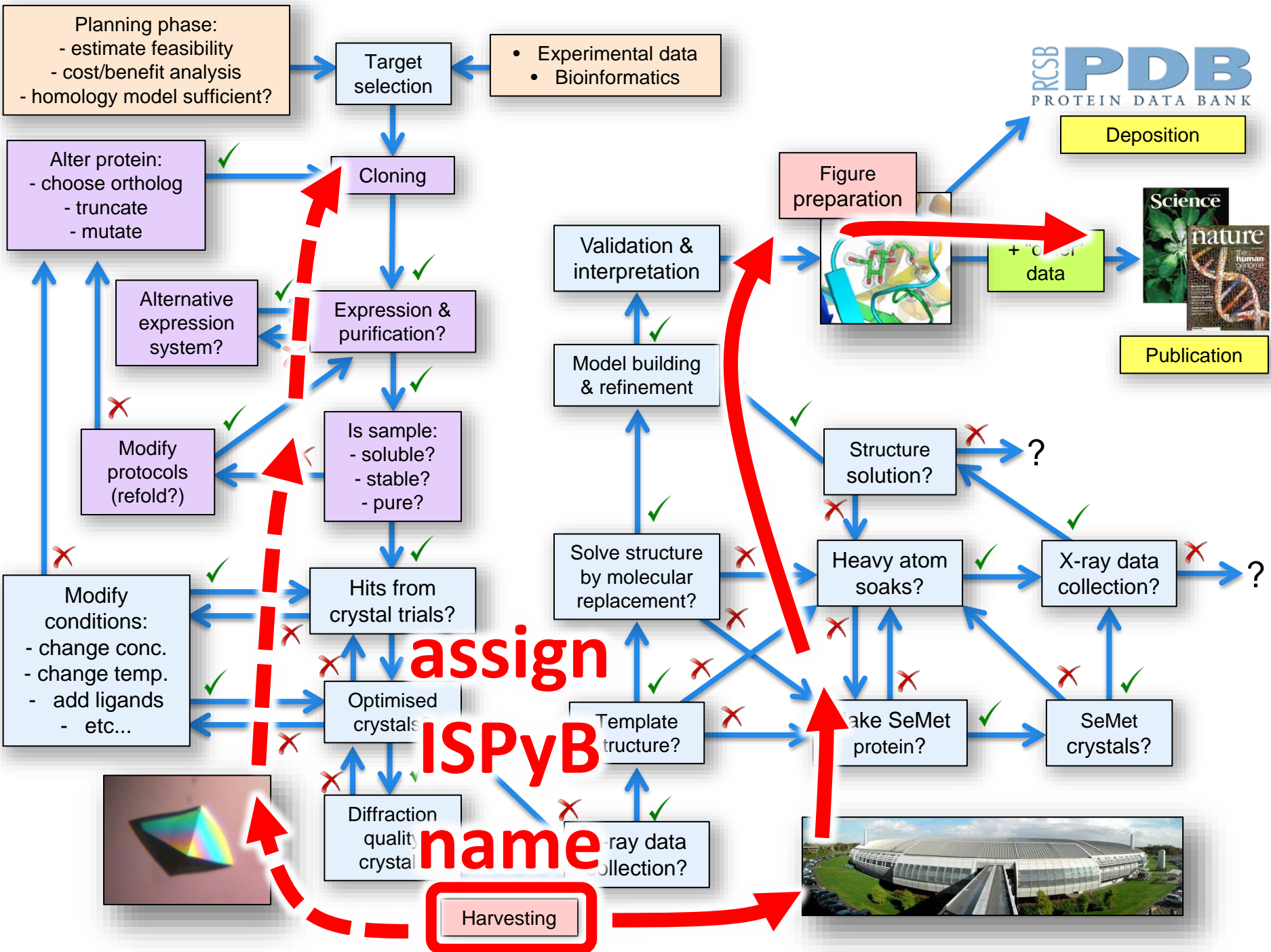


Strategy:  
EDNA  
Xia2



**Each crystal has a unique "ISPyB name"!**

**Use it!**





Date: 29-JUL-2017

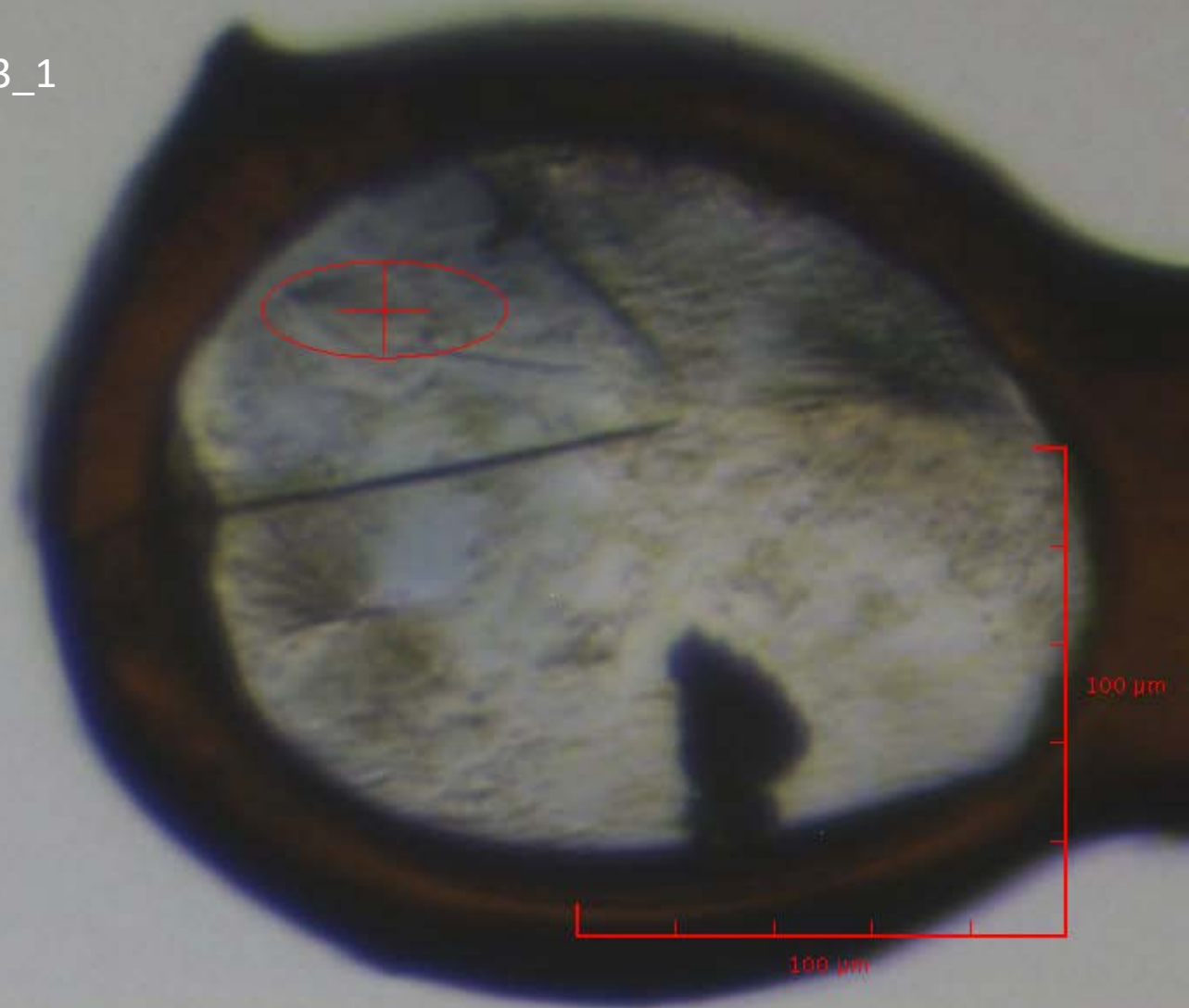
Beamline: i03 Diamond

Visit ID: mx13467-41

Protein acronym: NmADH9

ISPyB name: NmADH9\_23

Dataset name: NmADH9\_23\_1



Beam size: 50.0 × 20.0 μm

# Harvest your crystals and enter sample info into ISPyB

Container: DLS-442

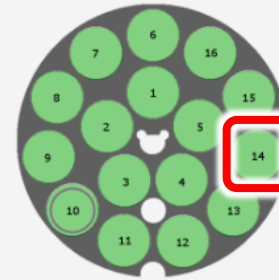
This page shows the contents of the selected container. Samples can be added and edited by clicking the pencil icon, and removed by clicking the x

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**

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_22				
14	NmADH9			NmADH9_23				
15	NmADH9			NmADH9_24				
16	NmADH9			NmADH9_25				

Comment	Anomalous	Required Res	Status
	Zn	1.5	Auto Integrated
	Zn	1.5	Auto Integrated

turns on EP  
pipelines

Required  
for UDC

# Use a spreadsheet...

crystals\_MX13467-41\_i03\_29-July-2017

## Puck number – DLS442

Position	Person	Protein name	IspeyB_name	Plate	Well	Conditions	Soak/cryo	Space group	Test	Resolution	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 hydroxygeranial+ 20%EG				5mM soak for approx 1 h



annotate hardcopy  
during data collection  
– helps decision  
making

# Data collection setup in GDA

Select required sample from drop-down menu:

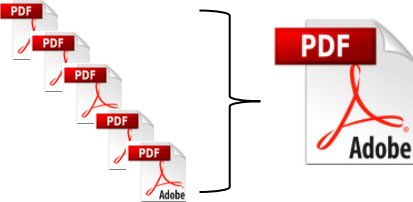
- no need to enter sample information or specify sample location
- less likely to get the wrong sample

Put your data into your own directory....

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

The screenshot shows the 'Data Collection Settings' window in GDA. The 'Screening' tab is selected and highlighted with a red box. A red arrow points from the text 'use the "screening" tab for test images – keeps them separate from datasets' to this tab. In the 'Sample' section, the sample name 'NmADH9\_23' is selected in a drop-down menu, highlighted with a red box, and a red arrow points from the text 'Select required sample from drop-down menu:'. Below this, the 'Files' section shows the 'Folder' field containing the path 'JIC/\${proteinacronym}/\${samplename}', which is also highlighted with a red box. A red arrow points from the text 'Put your data into your own directory....' to this field. The 'Visit directory' field shows '/dls/i03/data/2017/mx13467-41'. Other fields include 'Barcode' (NR), 'Holder' (2), 'Position' (14), 'Prefix' (JIC/NmADH9/NmADH9\_23), and 'Run number' (0). The 'Image' section shows 'Number of images' (3600), 'Exposure time' (0.010 s), 'Total exposure time' (36.0 s), and 'First image number' (1). The 'Beam and Detector' section shows 'Maximum resolution' (1.3000 Å), 'Detector distance' (213.5 mm), 'Wavelength' (0.97623 Å), 'Energy' (12700.3 eV), and 'Transmission' (50.156283 %).

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



← 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

12:00 29-07-2017

Sample	Images	Res	k	Q Dec	Spacegroup	Unit Cell	Processed Resolution	Rmean	Completeness	Comments
NmADH9_23	3600	1.3	0.9763	0.50	P 1 2 1	63.04, 100.11, 80.00, 104.25, 90.00	29.77 - 1.5, 29.77 - 0.7, 1.34 - 1.5	0.085, 0.036, 0.024	98.8, 98.7, 98.9	(-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	1.0	0.9763	0.00						Diffraction grid scan of 7 by 6 images, Top left (130,177), Bottom right (176,517)
NmADH9_15	12	2.0	0.9763	0.50						Diffraction grid scan of 3 by 4 images, Top left (147,245), Bottom right (147,245)
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: Small
NmADH9_15	3600	2.6	0.9763	0.50	P 1 2 1	62.17, 107.78, 75.76, 80.00, 105.05, 90.00	30.02 - 3.04, 30.02 - 3.04, 3.12 - 3.04	0.167, 0.040, 0.033	99.4, 98.8, 99.3	(-276, -373, 844) EDNAStrategy1: subWedge:1Aperture: Medium
NmADH9_11	3	2.0	0.9763	0.50						(-477, -294, 1142) Aperture: Medium
NmADH9_12	3	2.0	0.9763	0.50						(-216, -216, 1130) Aperture: Medium
NmADH9_12	3	1.6	0.9763	0.50						(-694, -724, 1000) Aperture: Medium
NmADH9_12	3600	1.6	0.9763	0.50	P 1 2 1	63.04, 100.11, 80.00, 104.25, 90.00	29.74 - 1.69, 29.74 - 1.69, 1.73 - 1.69	0.085, 0.036, 0.033	99.8, 99.3, 97.1	(-694, -724, 1000) EDNAStrategy1: subWedge:1Aperture: Medium
TAPHY_27_1	4	3.0	0.9763	0.50						Diffraction grid scan of 1 by 4 images, Top left (540,265), Bottom right (540,265)
TAPHY_27_1	3	1.6	0.9763	0.50						(-109, -296, 1303) EDNAStrategy1: subWedge:1Aperture: Medium
TAPHY_27_1	1200	1.5	0.9763	0.50	H 3	126.66, 126.66, 107.16, 90.00, 90.00, 120.00	30.27 - 1.62, 30.27 - 1.62, 1.97 - 1.92	0.078, 0.036, 0.094	99.0, 96.6, 95.9	(-109, -296, 1303) EDNAStrategy1: subWedge:1Aperture: Medium
TAPHY_27_2	5	2.0	0.9763	0.00						Diffraction grid scan of 1 by 5 images, Top left (134,251), Bottom right (134,251)
TAPHY_27_2	3	1.5	0.9763	0.50						(-136, -104, 1462) EDNAStrategy1: subWedge:1Aperture: Medium

Page 5 of 12

Sample	Images	Res	k	Q Dec	Spacegroup	Unit Cell	Processed Resolution	Rmean	Completeness	Comments
NmADH9_23	3600	1.3	0.9763	0.50	P 1 2 1	63.04, 100.11, 80.00, 104.25, 90.00	29.77 - 1.5, 29.77 - 0.7, 1.34 - 1.5	0.085, 0.036, 0.024	98.8, 98.7, 98.9	(-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	1.0	0.9763	0.00						Diffraction grid scan of 7 by 6 images, Top left (130,177), Bottom right (176,517)
NmADH9_15	12	2.0	0.9763	0.50						Diffraction grid scan of 3 by 4 images, Top left (147,245), Bottom right (147,245)
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: Small
NmADH9_15	3600	2.6	0.9763	0.50	P 1 2 1	62.17, 107.78, 75.76, 80.00, 105.05, 90.00	30.02 - 3.04, 30.02 - 3.04, 3.12 - 3.04	0.167, 0.040, 0.033	99.4, 98.8, 99.3	(-276, -373, 844) EDNAStrategy1: subWedge:1Aperture: Medium
NmADH9_11	3	2.0	0.9763	0.50						(-477, -294, 1142) Aperture: Medium
NmADH9_12	3	2.0	0.9763	0.50						(-216, -216, 1130) Aperture: Medium
NmADH9_12	3	1.6	0.9763	0.50						(-694, -724, 1000) Aperture: Medium
NmADH9_12	3600	1.6	0.9763	0.50	P 1 2 1	63.04, 100.11, 80.00, 104.25, 90.00	29.74 - 1.69, 29.74 - 1.69, 1.73 - 1.69	0.085, 0.036, 0.033	99.8, 99.3, 97.1	(-694, -724, 1000) EDNAStrategy1: subWedge:1Aperture: Medium
TAPHY_27_1	4	3.0	0.9763	0.50						Diffraction grid scan of 1 by 4 images, Top left (540,265), Bottom right (540,265)
TAPHY_27_1	3	1.6	0.9763	0.50						(-109, -296, 1303) EDNAStrategy1: subWedge:1Aperture: Medium
TAPHY_27_1	1200	1.5	0.9763	0.50	H 3	126.66, 126.66, 107.16, 90.00, 90.00, 120.00	30.27 - 1.62, 30.27 - 1.62, 1.97 - 1.92	0.078, 0.036, 0.094	99.0, 96.6, 95.9	(-109, -296, 1303) EDNAStrategy1: subWedge:1Aperture: Medium
TAPHY_27_2	5	2.0	0.9763	0.00						Diffraction grid scan of 1 by 5 images, Top left (134,251), Bottom right (134,251)
TAPHY_27_2	3	1.5	0.9763	0.50						(-136, -104, 1462) EDNAStrategy1: subWedge:1Aperture: Medium

19:00 26-07-2017

02:00 27-07-2017

02:00

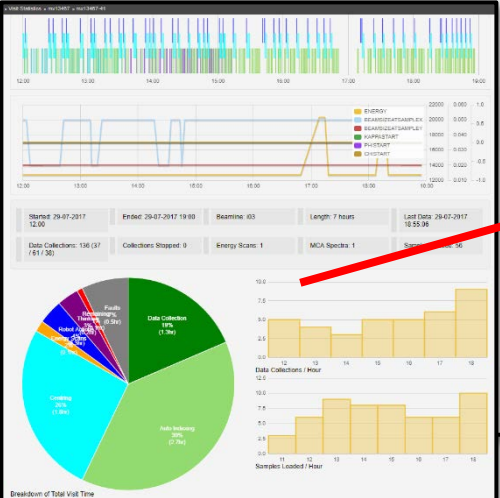
17:00

19:00

12:00

02:00

10



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
02:00			i03	Dr Neil Paterson	38	Compulsorily Remote
17:00			i03	Dr Neil Paterson	76	Compulsorily Remote
19:00			i04	Dr Melanie Vollmar	69	Compulsorily Remote
12:00			i04	Dr Dave Hall	106	Compulsorily Remote
02:00			i03	Dr Neil Paterson	82	Compulsorily Remote

5 6 7 > >>



# 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	$\lambda$	$\Omega$ 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 s	Phi ra	Xbe	Ybe	Det	auto/m	resoln	space gr	cell	twinn	in comments
54		NmADH9_23	2	##	NmADH9_23	3	0.9763	428	0.04	0	0.5	212	206	2.2							
55		NmADH9_23	3	##	NmADH9_23	3	0.9763	265	0.04	0	0.5	212	207	1.5							
56	12	NmADH9_23	1	##	NmADH9_23	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	##	NmADH9_24	3	0.9763	265	0.04	0	0.5	212	207	1.5							
58		NmADH9_24	2	##	NmADH9_24	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

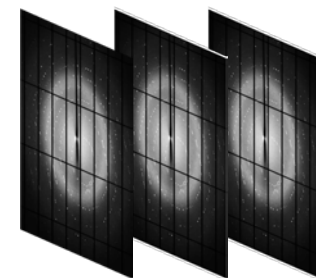


robotic sample exchange



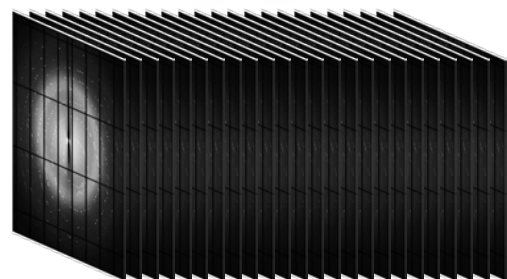
align crystal in X-ray beam

analyse data  
on-the-fly



collect test images

typically 3,600 x  
0.1° images



collect dataset



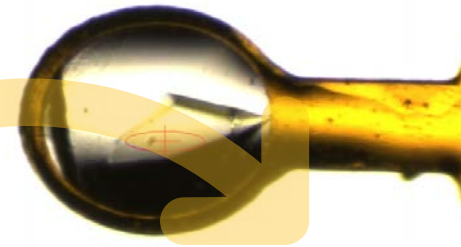
devise data collection strategy based on test  
images (and what you want to use the data for)



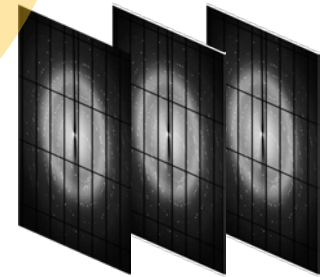
Don't  
forget to  
think!



robotic sample exchange



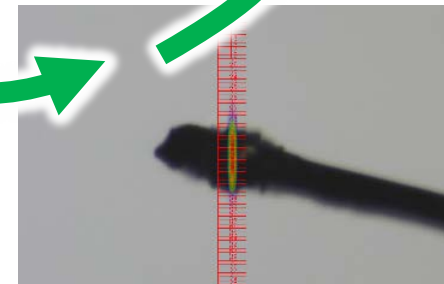
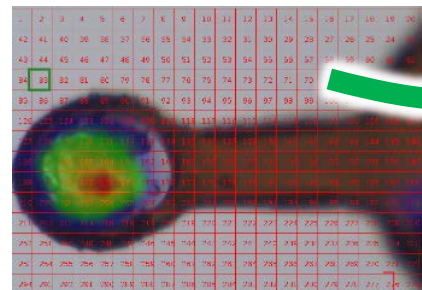
align crystal in X-ray beam

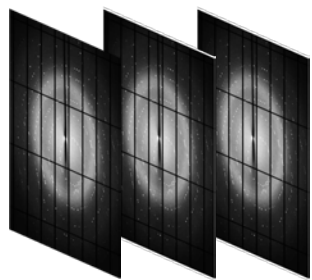


collect test images

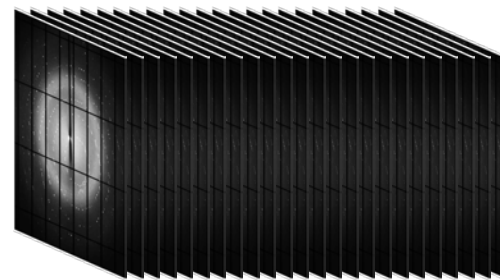
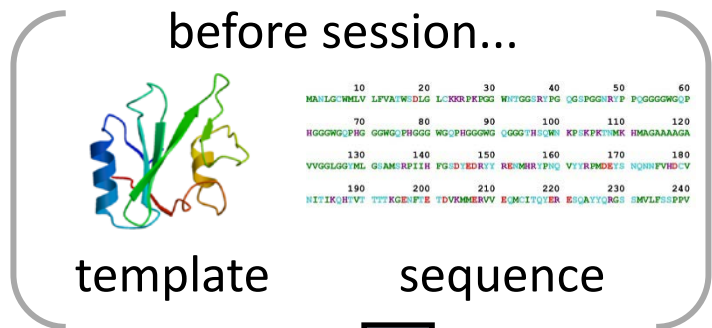
More efficient to screen your samples in batches,  
then decide what to do...

- do this automatically with X-ray centring
- also an opportunity to grab a coffee!





test images



dataset



Strategies:  
Mosflm  
EDNA  
Xia2

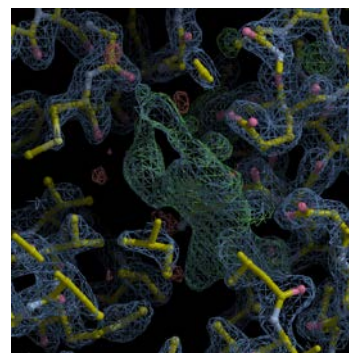


*ISPyB*

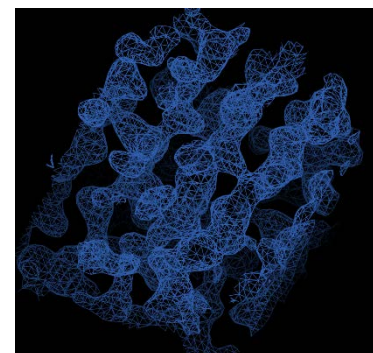


Fast DP

h	k	l	lplus	SIGlplus	lminus	SIGlminus
-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
-45	0	12	0.63	1.33	0.63	1.33



Dimple

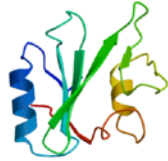


Fast EP

Quick  
feedback



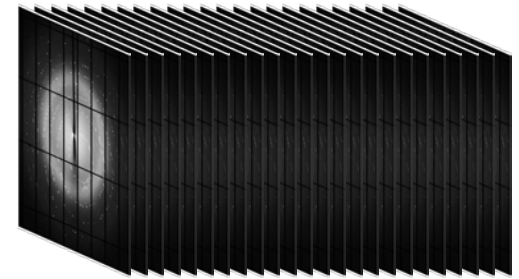
before session...



template

```
10      20      30      40      50      60
MANLGNNELV LIPVATWEDLG LKKRPPKGG WFFGGRTFG QGPGGNRYF PGGGGWGQP
70      80      90     100     110     120
HGGGNGPHG GGGGPHGG WGPPIGGWG QGGTHSGNI KFKPKTNKK IMAGAAAGA
130     140     150     160     170     180
VVGGLGGML GQMERPIIH PGDIEDRYT REMMERTPGQ VTRPMDEYS NQNFVYDLY
190     200     210     220     230     240
NITIKQRYT TTKGKSPR TDVQRRRVV KMCITQYR ESQATYGRG SMVLFSPFV
```

sequence



dataset



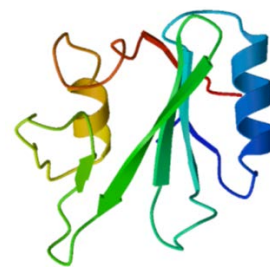
ISPyB

20-30 min

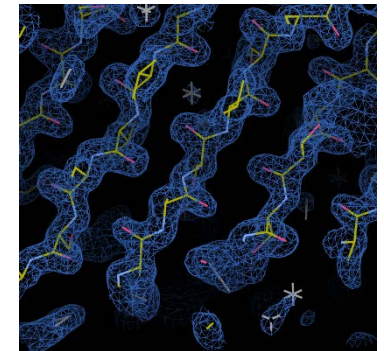
Xia2  
DIALS  
MultiXia2  
autoPROC

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
-45	0	12	0.52	1.22	0.52	1.22

hours



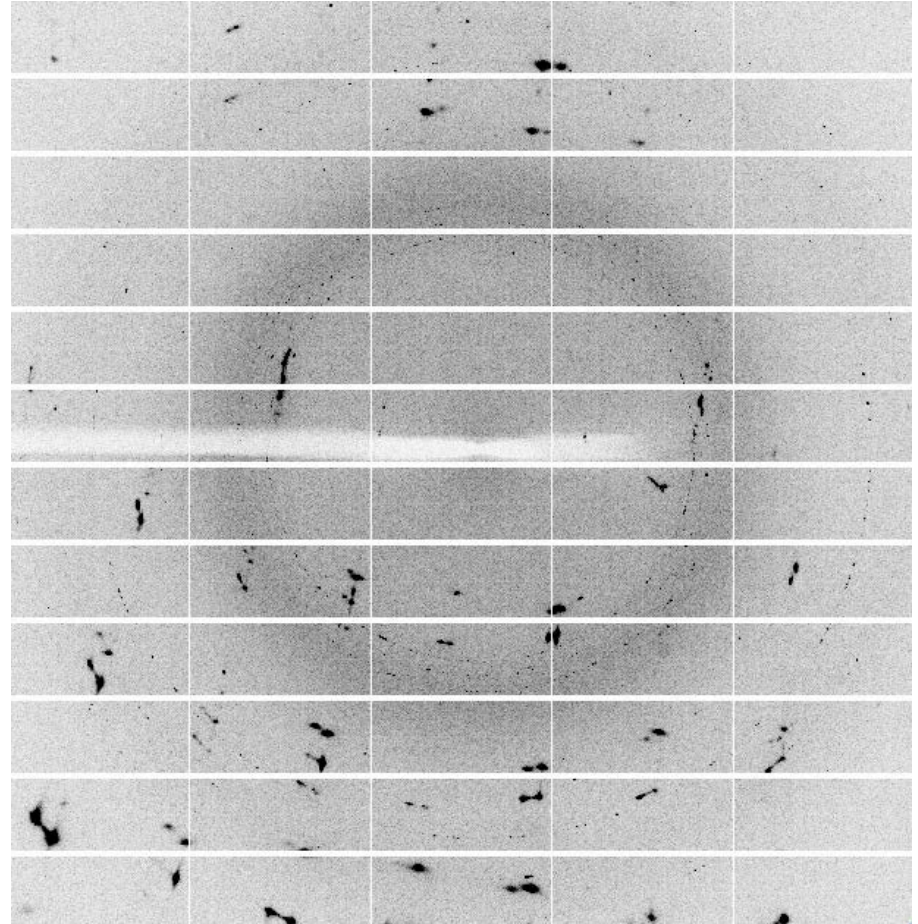
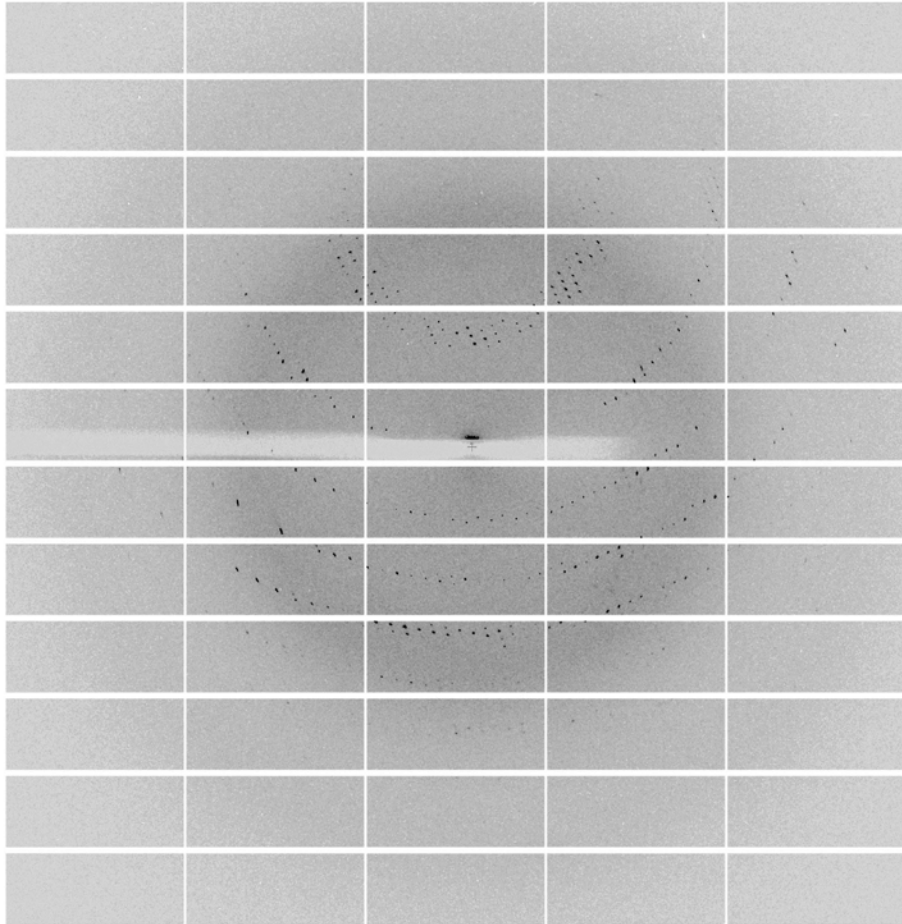
MrBUMP



Big EP

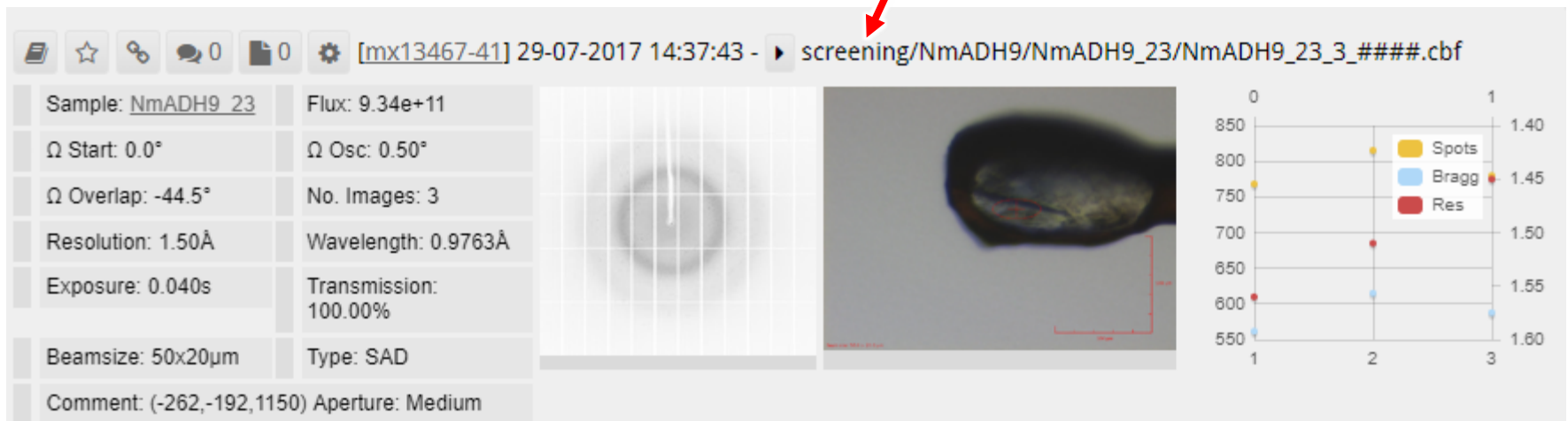
Slower  
feedback

# LOOK at your images!!!!

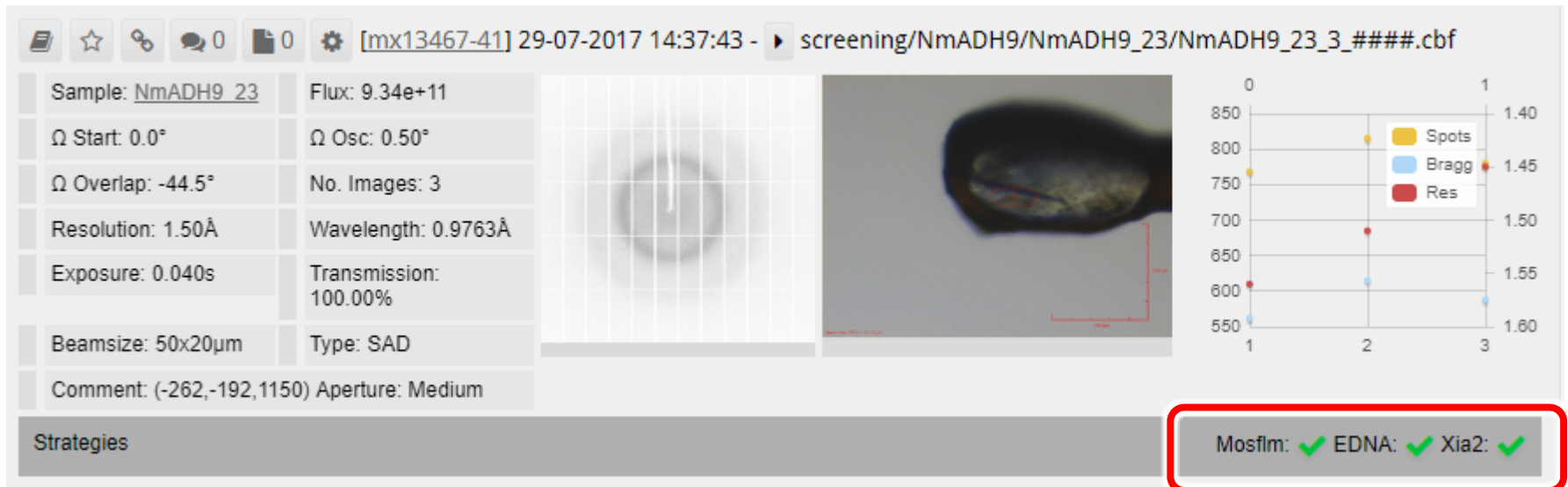


# Screening...

(using the “screening” tab...)



# Screening...



# Screening...

Have I seen  
this crystal  
form before?

## Strategies

Mosfilm: ☒ EDNA: ☒ Xia2: ☒

### xia2.strategy

Space Group	A	B	C	$\alpha$	$\beta$	$\gamma$
P 1 2 1	63.87	107.69	69.31	90.00	104.25	90.00

Q Lookup Cell

Strategy	Description	$\Omega$ Start	$\Omega$ 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	1.1	4.0	4	0.010	2214
gentle Wedge1	Gentle: Target Multiplicity=2 I/sig=2 Maxlifespan=20s	6	0.15	1.51	1.2	4.0	4	0.010	2000
high multiplicity Wedge1	Strategy with target multiplicity=16 I/sig=2 202s	0	0.15	1.51	1.1	4.1	4.1	0.010	2400
native Wedge1	Standard Native Dataset Multiplicity=3 I/sig=2 Maxlifespan=202s	162	0.15	1.51	1.1	2.2	2.2	0.010	1947

### EDNA MXv1

Space Group	A	B	C	$\alpha$	$\beta$	$\gamma$
P2	63.99	107.49	69.28	90.00	104.37	90.00

Q Lookup Cell

Strategy	Description	$\Omega$ Start	$\Omega$ 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 Maxlifespan=202 s					5.1		0.010	3600
Strategy4 Wedge1	Gentle: Target Multiplicity=2 and target I/Sig 2 and target I/sig=2 Maxlifespan=202 s					14.8		0.010	1110
Strategy5 Wedge1	UnderDEV Anomalous Dataset, RadDamage of standard					12.4		0.010	2740

### mosfilm

Space Group	A	B	C	$\alpha$	$\beta$	$\gamma$
P2	63.80	107.65	69.44	90.00	104.13	90.00

Q Lookup Cell

☐ Data Collection Settings ☒ Plate View ☐ Screening

### Data Collection Settings

Sample: NmADH9\_23

Strategies...

Omega: Start Oscillation

Strategy	Description	$\Omega$ Start	$\Omega$ Osc	Res (Å)	Ranking Res (Å)	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



# Notes about strategies...

Despite what the strategy programs may suggest (or your supervisor/colleague!)

- always collect a minimum of  $360^\circ$  (unless you have a very good reason not to)
- consider multiple  $360^\circ$  passes if you have low symmetry (ideally rotating around a different axis for each – see data collection talks...)

If you can't get a sensible strategy and the crystal is diffracting reasonably well...

- collect a data set anyway – it might be useful!
  - inspect screening images in ADXV (may need to sum several adjacent images)
  - look for highest resolution spots (NOT ice or salt spots!)
  - subtract 0.5 from this resolution value. E.g. spots at  $2.5 \text{ \AA}$  → collect to  $2.0 \text{ \AA}$
  - collect  $3600 \times 0.1^\circ$  images with a total dose of 1-2 MGy

# Collecting a dataset...

Space Group	A	B	C	$\alpha$	$\beta$	$\gamma$
P2	63.99	107.49	69.28	90.00	104.37	90.00

Q Lookup Cell

Strategy	Description	$\Omega$ Start	$\Omega$ 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	1830

☐ 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}/\${samplename}  
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 %

data collection is fast so always collect at least 360°

push the resolution a little...

# Checking the results...



# Checking the results...

Auto Processing									
					Fast DP: ✓ Xia2: ✓ ✓ ✓ MultiXia2: ? ? autoPROC: ✓				
Type	Resolution	Spacegroup	Mn<I/sig(I)>	Rmeas Inner	Rmeas Outer	Completeness	Cell	Status	
fast_dp	29.77 - 1.50	P 1 2 1	15.2	0.038	0.824	98.8	63.92 107.75 69.36 90.00 104.27 90.00		
xia2 3d	28.52 - 1.37	P 1 2 1 1	10.9	0.038	1.688	98.4	63.92 107.75 69.36 90.00 104.27 90.00		
xia2 3dii	42.04 - 1.37	P 1 2 1 1	10.8	0.038	1.685	98.5	63.92 107.75 69.36 90.00 104.27 90.00		
xia2 dials	107.77 - 1.30	P 1 2 1 1	8.9	0.039	1.753	98.2	63.94 107.77 69.37 90.00 104.26 90.00		
autoPROC 1.0.5 (see: <a href="http://www.globalphasing.com/autoproc/">http://www.globalphasing.com/autoproc/</a> )	107.76 - 1.50	P 1 2 1 1	13.3	0.038	0.877	98.8	63.93 107.76 69.37 90.00 104.26 90.00		

fast_dp	xia2 3d	xia2 3dii	xia2 dials	autoPROC 1.0.5 (see: <a href="http://www.globalphasing.com/autoproc/">http://www.globalphasing.com/autoproc/</a> )
---------	---------	-----------	------------	--

Beam Centre	X	Y
Start	211.60	206.96
Refined	211.50	206.92
$\Delta$	0.10	0.04

Space Group	A	B	C	$\alpha$	$\beta$	$\gamma$
P 1 2 1 1	63.92	107.75	69.36	90.00	104.27	90.00

Shell	Observations	Unique	Resolution	Rmeas	I/sig(I)	CC Half	Completeness	Multiplicity	Anom Completeness	Anom Multiplicity	CC Anom
outerShell	64205	9220	1.37 - 1.39	1.685	1.1	0.5	96.7	7.0	96.0	3.5	0.0
innerShell	64845	9660	3.72 - 42.06	0.038	44.2	1.0	99.9	6.7	98.6	3.5	0.1
overall	1285570	187412	1.37 - 42.04	0.099	10.8	1.0	98.5	6.9	98.0	3.5	0.1

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

# Re-running jobs:

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

The screenshot displays the ISPyB interface. At the top, a toolbar includes a gear icon (settings) circled in red. Below it, a table lists job parameters for 'NmADH9\_23':

Sample: NmADH9_23	Flux: 1.60e+11
$\Omega$ Start: 45.0°	$\Omega$ Osc: 0.10°
$\Omega$ Overlap: 0°	No. Images: 3600
Resolution: 1.30Å	Wavelength: 0.9763Å
Exposure: 0.010s	Transmission: 50.16%
Beamsize: 50x20µm	Type: SAD

A comment field contains: 'Comment: (262, 102, 1150) EDNA Strategy 1: sub'. Below this is a 'Reprocess Data' button. A green box with a red arrow points to the gear icon and contains the text:

Change:

- space group
- cell parameters
- maximum resolution
- reject images

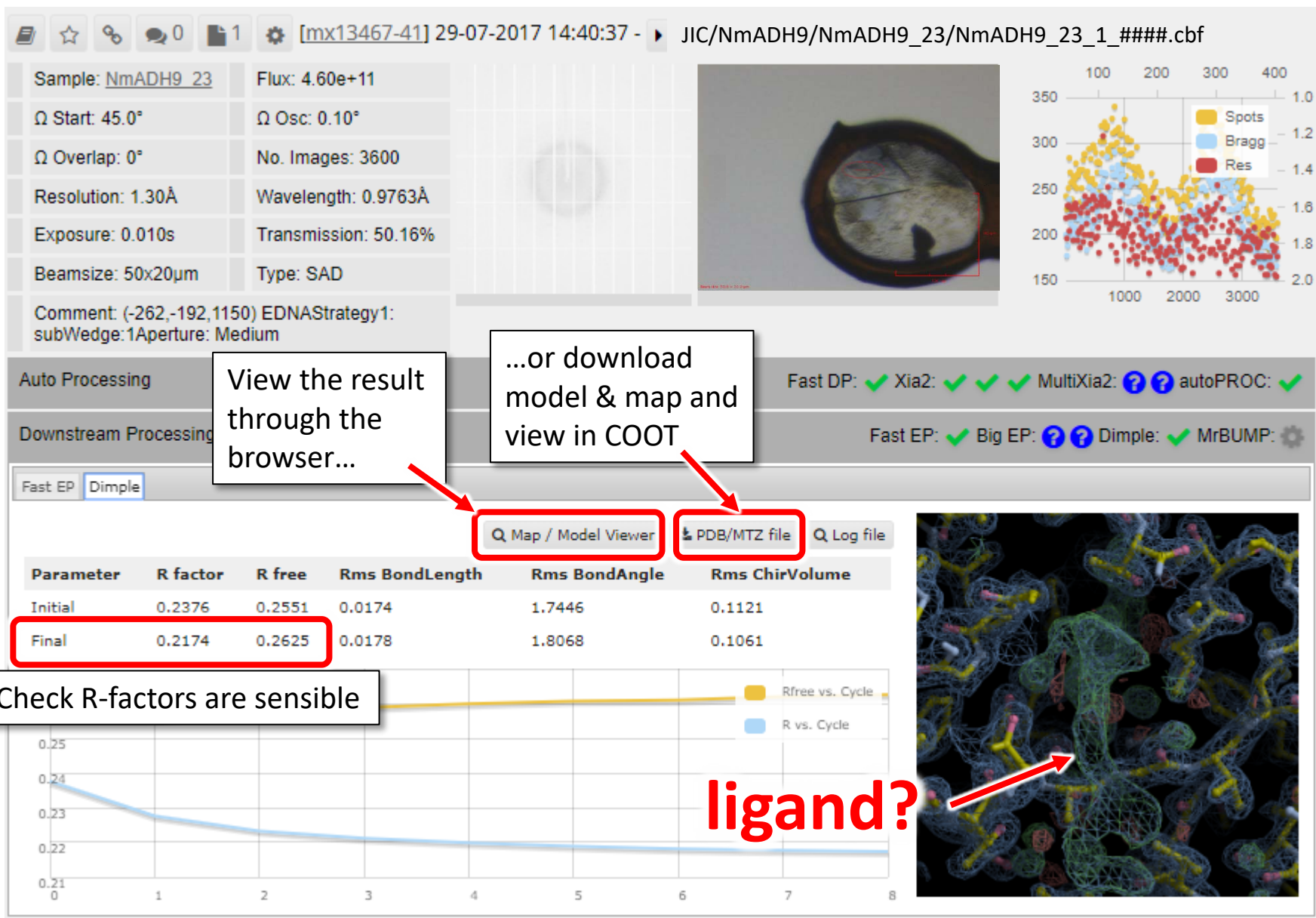
The bottom section shows a 'Multi Crystal' window for 'NmADH9\_23\_1 - JIC/NmADH9\_23/NmADH9/'. It includes input fields for 'Sample: NmADH9\_23', ' $\Omega$  Start: 45.0°, Osc: 0.10°', 'Resolution: 1.30Å', and 'Wavelength: 0.9763Å'. There are also 'Start' and 'End' fields with a '+' button. To the right is a scatter plot of 'Spots' (yellow), 'Bragg' (blue), and 'Res' (red) data points. The plot has two x-axes (0-400 and 0-3500) and two y-axes (150-350 and 1.00-2.00). At the bottom right are 'Integrate' and 'Close' buttons.



# Checking the results...



# Checking the results...



# 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...

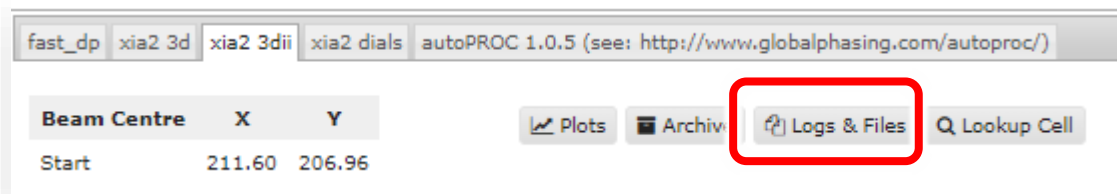
- FTP data home (use an App or a script)



Quite slow...



Faster...



In the meantime:

- use autoproccessing output or...
- (re)process data remotely on DLS computers and transfer output only
- copy and archive 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:

Create a New Project

Name of project/folder: NmADH9\_23\_1

By default all projects go in the 'CCP4I2\_PROJECTS' directory in your home area - click 'Select directory' to choose an alternative.  
Hint to organise your projects: in the 'Manage projects' window you can use a project as a folder and drag other projects into it

Description of project: High res data from binary complex with NAD

Choose tag.. Choose tag.. Choose tag..

New tag

Save

Create project Select directory Cancel Help

Manage projects

Name	Directory	Created	Last active	Tags
NmADH9	C:\Users\la...	18 Oct 18	24 Nov 18	
NmADH9_23_1	C:\Users\la...	24 Nov 18	24 Nov 18	
NmADH9_26_1	C:\Users\la...	24 Nov 18	27 Nov 18	
	C:\Users\la...	24 Nov 18	24 Nov 18	

Open

Add project or folder

Rename project

Edit description

Can rename “projects” later... (when you know more...)\*

Manage project

Name	Directory	Created	Last active	Tags
NmADH9	C:\Users\la...	18 Oct 18	24 Nov 18	
NmADH9_23_1_NAD	C:\Users\la...	24 Nov 18	24 Nov 18	
NmADH9_26_1_apo	C:\Users\la...	24 Nov 18	27 Nov 18	
	C:\Users\la...	24 Nov 18	24 Nov 18	

Open

Add project or folder

Rename project

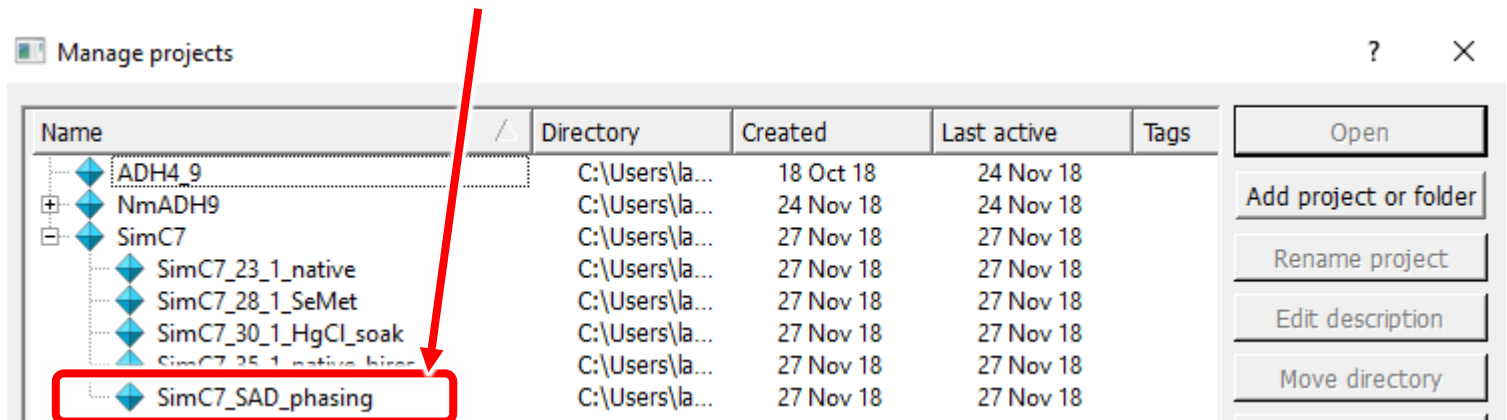
Edit description

\*but directory names remain the same

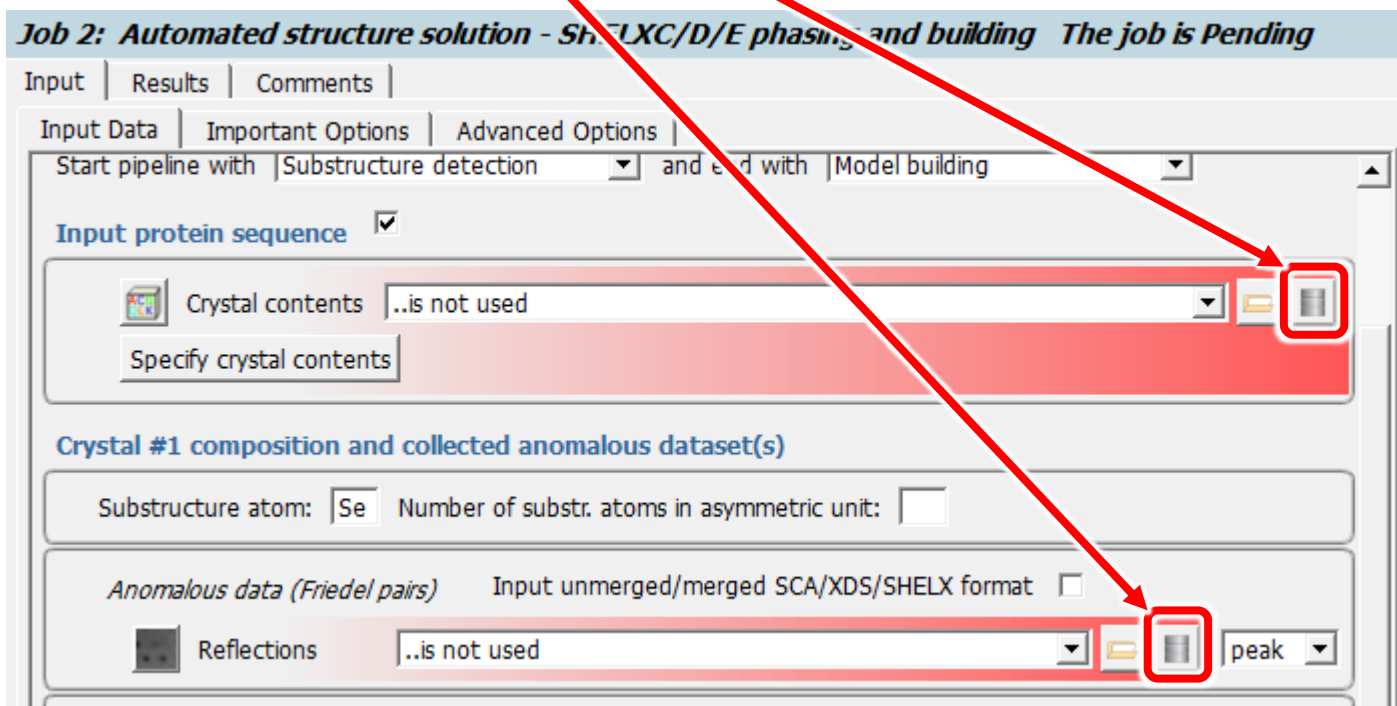


CCP4i2

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



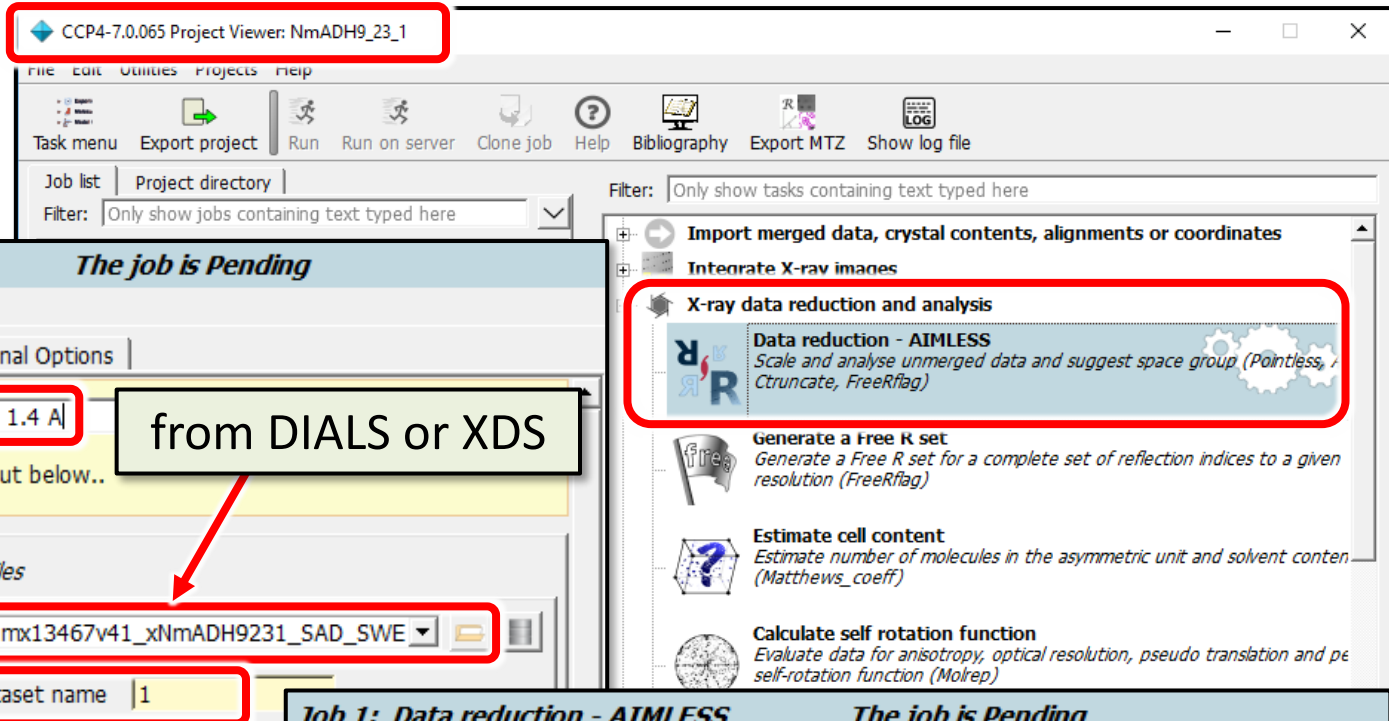
...then “borrow” files from other projects...



My preferred option for processing...

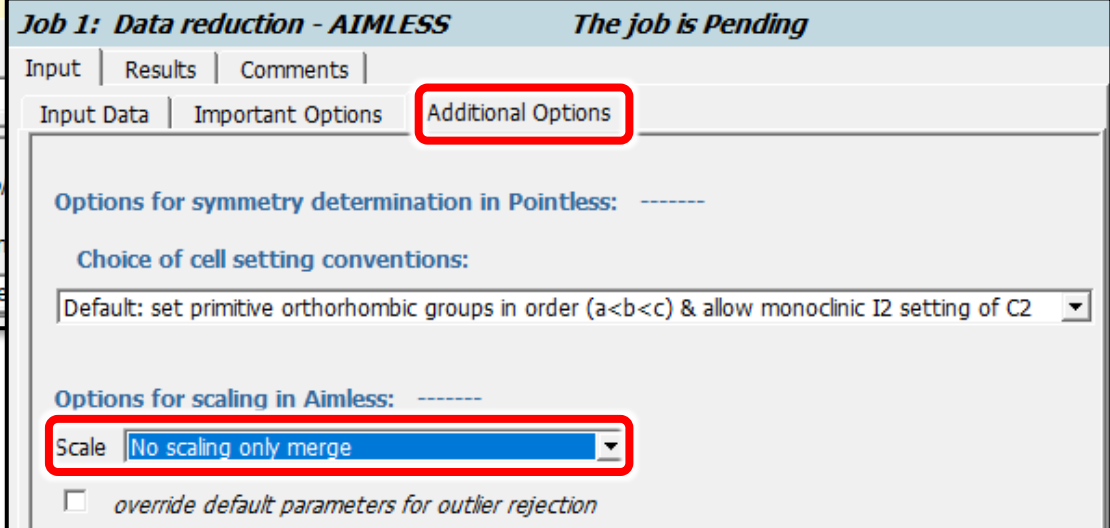
- where the pipelines have done a good job – take part-processed dataset (from the “Meta” data) and re-run the merging step in CCP4i2

use informative  
job titles



from DIALS or XDS

may want to change  
resolution, exclude  
images etc...



Gives lots of output – useful for troubleshooting...

for your  
paper/thesis

**Job 1: Data reduction - AIMLESS** *The job is Finished*

Input Results Comments

Headline Summary SG details MergingGraphs SDanalysis MergingDetails Istats Biblio

Data reduction - full dataset to 1.4 Å

▼ Key summary

Selecting space group P 1 2 1  
as there is a single space group with the highest score

Solution probability: 0.872, Confidence 0.866 (high resolution limit for symmetry testing 1.495 Å)

Key statistics for Dataset: NmADH9\_23\_1/NmADH9\_23/1

Unit cell: 63.921 107.750 69.361 90.000 104.265 90.000, wavelength 0.976250 Å  
Resolution of input data: 1.40 Å, resolution estimate: beyond 1.40 Å  
Anisotropic limits: - Along 0.99 a° - 0.16 c° 1.48 Å CC(1/2), 1.58 Å I/σ - Along k axis 1.40 Å° CC(1/2), 1.42 Å I/σ - Along -0.09 1.40 Å I/σ  
Rmeas: overall 0.096, inner bin 0.037  
In outer bin: Mean(I/sd) 1.2 CC(1/2) 0.559  
Overall filtered Mean(chi^2): 1.03  
Anomalous CC(1/2) in inner bin 0.093  
No significant anomalous signal detected  
NOTE: no scaling was done, just merging

SD correction information:  
SD correction parameters were not refined

✓ No evidence of twinning

✓ No evidence of possible translational non-crystallographic symmetry

● Warning: Some anisotropy detected. This may affect the quality of the data.

● Warning: Completeness test shows some issues.

✗ Warning: Severe deviation from Wilson plot.

✗ Warning: Possible ice rings found.

A free-R set has been created, fraction of the data = 0.05

Show Pointless logfile Show Aimless logfile Show Ctruncate logfile

▼ Overall summary

**Job 1: Data reduction - AIMLESS** *The job is Finished*

Input Results Comments

Headline Summary SG details MergingGraphs SDanalysis MergingDetails Istats Biblio Run

▼ Overall summary

Space group determination  
Selecting space group P 1 2 1  
as there is a single space group with the highest score

Solution type: space group

Group name	P 1 2 1
Reindex	[h,k,l]
Space group confidence	0.866
Laue group confidence	0.821
Laue group probability	0.882
Systematic absence probability	0.988

Scores for each symmetry element  
Lattice group name P 1 2 1

Likelihood	CC	R	Symmetry
0.880	0.87	0.087	identity
0.882	0.87	0.087	2-fold k (0 1 0) [-h,k,-l]

**Data internal consistency statistics**

Summary of merging statistics for dataset  
NmADH9\_23\_1/NmADH9\_23/1

	Overall	Inner	Outer
Low resolution limit	47.16	47.16	1.42
High resolution limit	1.40	7.67	1.40
Rmerge(within I+ /I-)*	0.081	0.031	1.402
Rmerge(all I+ and I-)*	0.090	0.036	1.563
Rmeas (within I+ /I-)*	0.096	0.037	1.657
Rmeas (all I+ & I-)*	0.097	0.039	1.690
Rpim (within I+ /I-)	0.051	0.020	0.876
Rpim (all I+ & I-)	0.037	0.015	0.636
Rmerge in top intensity bin*	0.048		
Number of observations	1205637	7127	59269
Number unique	175909	1122	8610
Mean(I) / sd(I)	11.2	46.7	1.2
Half-set correlation CC(1/2)	0.999	0.998	0.559
Completeness %	98.6	99.2	97.0
Multiplicity	6.9	6.4	6.9
Filtered Mean(chi^2)	1.03	1.12	1.03
Anomalous completeness %	98.1	97.0	96.4
Anomalous multiplicity	3.4	3.4	3.5
DelAnom CC(1/2)	0.055	0.093	0.034
Mid-Slope of Anom Probability	1.038		

Download

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

...or reprocess from scratch using DIALS or XDS...

# 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
- Reduces amount of post-beamtime follow-up work
- Simple to keep track of your samples and data



# Remote data collection



not going to cover unattended data collection (UDC)...

*During the lockdowns...  
remote data collection?  
it's a no-brainer!*

remote users  
since 2010...

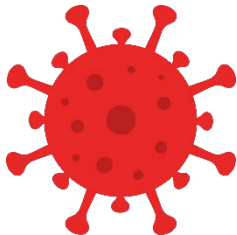
On-site data collection can be a big time commitment...



Remote data collection saves you time...



...and the planet

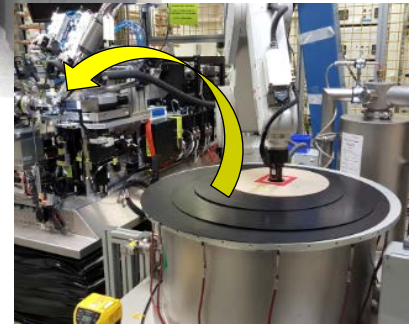


...and is covid-secure



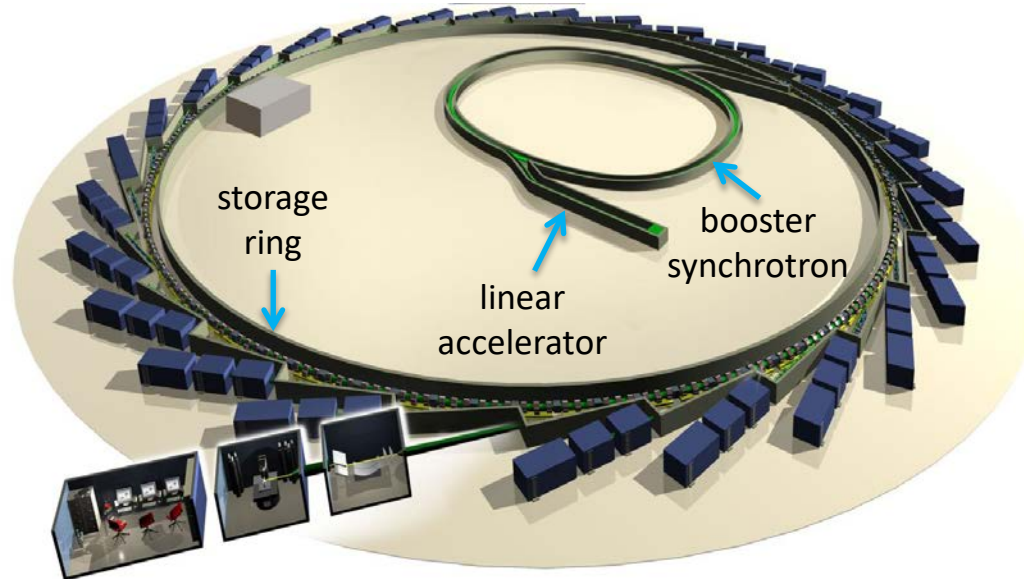
# For routine data collection at 100 K:

- 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 interface
- Only manual operation at DLS: loading/unloading pucks
  - Up to 592 samples can be loaded at once
- Everything else computer driven...



**Therefore you don't need to be there!**

# What is remote data collection?



home lab/home



more remote

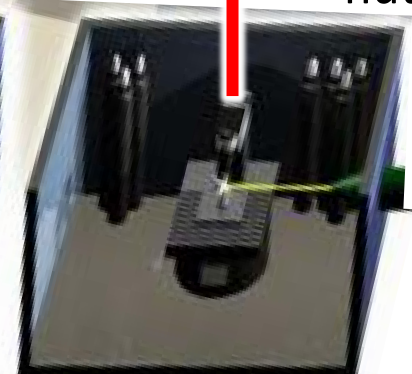
control  
cabin



remote



experimental  
hutch





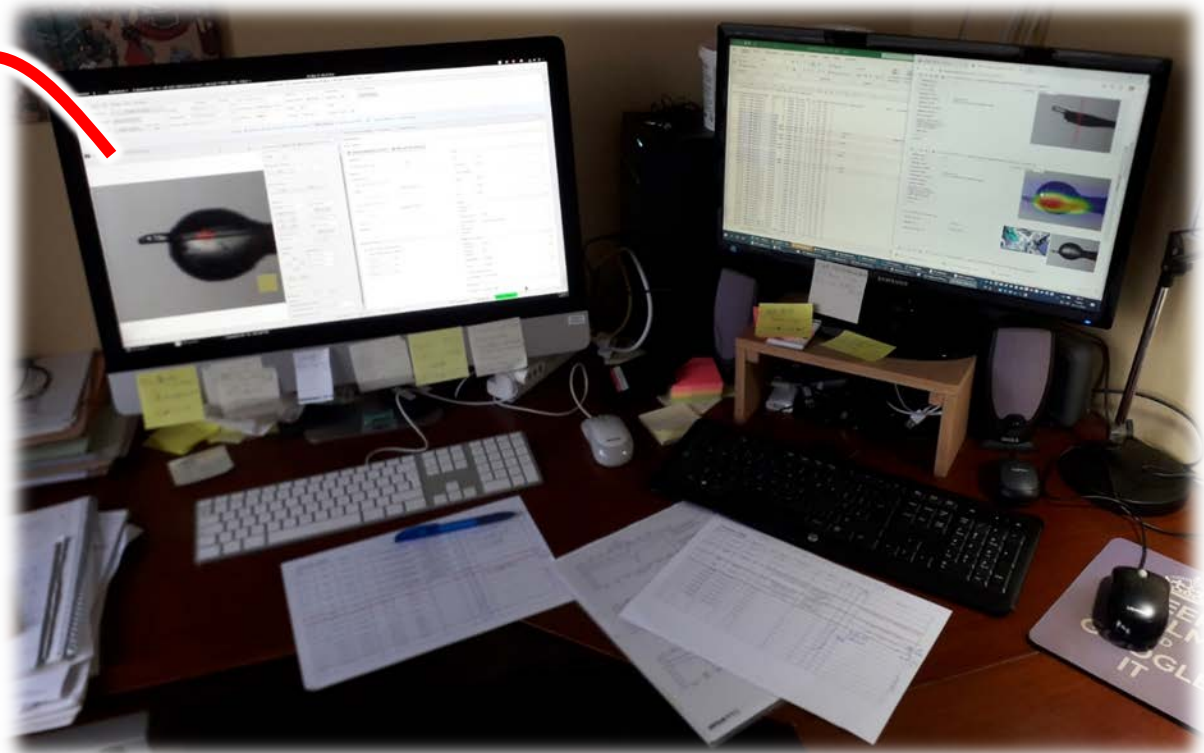
# Remote data collection lockdown style!



...from here

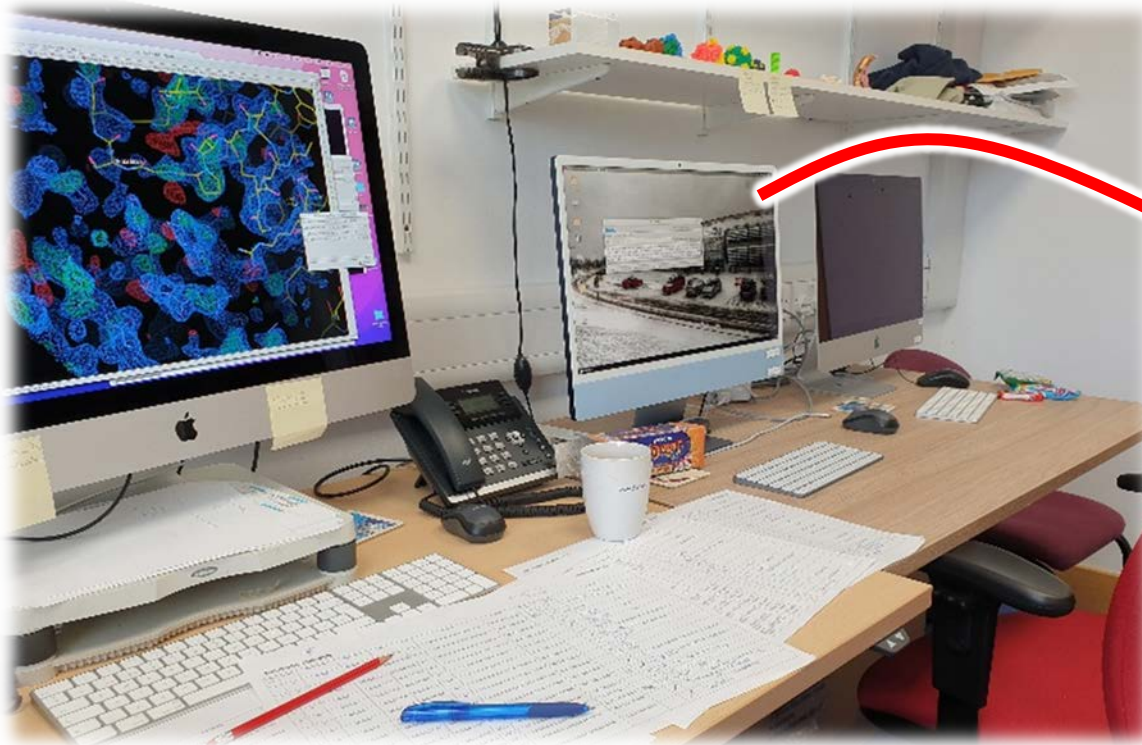


share screen  
with others



# Remote data collection – normality resumed

...from here



share screen  
with others

# Timeline for fully remote data collection

Sometime prior to shift:

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

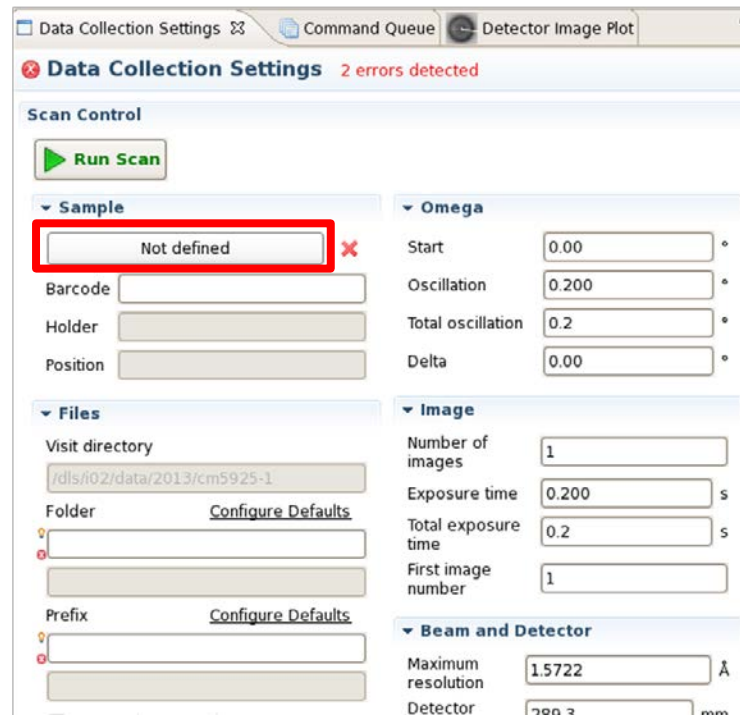
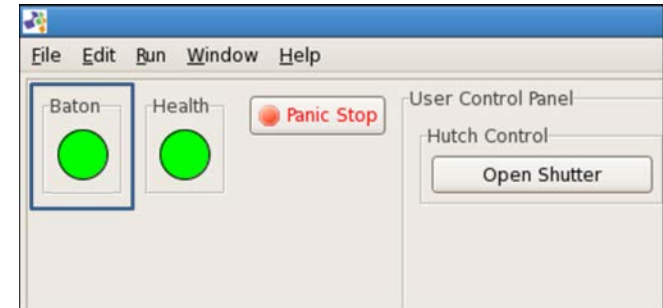


Shortly before start of shift (or slot in running order):

- 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!


i04 Webcams & Beamline Status

Ring Current 299.991	Refill 255.758	Hutch Locked	Port Shutter Open	Expt Shutter Open	Fast Shutter Open	Wav 0.9
-------------------------	-------------------	-----------------	----------------------	----------------------	----------------------	------------

Sun Nov 25 15:25:45 - I04 Sample Position



Sun Nov 25 15:25:45 - I04 Sample Ch



The image displays a mobile application interface for monitoring the i04 beamline. The interface includes a status bar at the top with a search bar, a filter dropdown, and a data collection path. Below this are four green buttons for 'Expt Shutter Open', 'Fast Shutter Open', 'Wavelength 0.979499', and 'Transmission 100'. The main content area features two large video feeds. The left feed, titled 'Sun Nov 25 15:25:45 - I04 Sample Position', shows a close-up of a green and silver mechanical assembly. The right feed, titled 'Sun Nov 25 15:25:45 - I04 Sample Ch', shows a robotic arm positioned over a circular sample stage. A smartphone is shown on the right side of the image, displaying the same application interface.

If you suspect a problem – call the EHC!

# If you see no diffraction – 3 main causes:

...could be all 3 🤔

(1) there is a problem (any number of things...)



(2) you are doing something wrong/stupid



(3) your sample is rubbish!



Pop a couple of test crystals into one of your pucks (something you know will diffract e.g. lysozyme)





# Advantages of remote data collection:

- Beamtime is fully used
- Users collect data on their own crystals
- Users stay at home labs (or at home)
  - time commitment is low
  - your boss/collaborator can observe data collection
  - useful for training non-experts
- Time can be used flexibly
- Less stressful
- Difficult to “break” the beamline

# Are there any disadvantages of remote data collection?

- miss the “wow factor” of being at a synchrotron
- lose face-to-face interactions with Diamond staff
  - do BAG training
  - go to User meeting
  - get in touch online



# Take home messages

- make full use of ISPyB (use ISPyB name!)
- exploit the MX software tools/pipelines (collect 360°!)
- use remote data collection for routine experiments
- think before and during data collection
- this is may be your last experiment – don't mess it up!

# Acknowledgements

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