



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

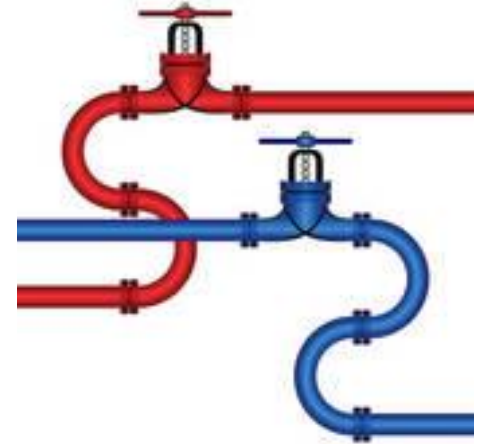
Strategy



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
- not unattended data collection (UDC)

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!)

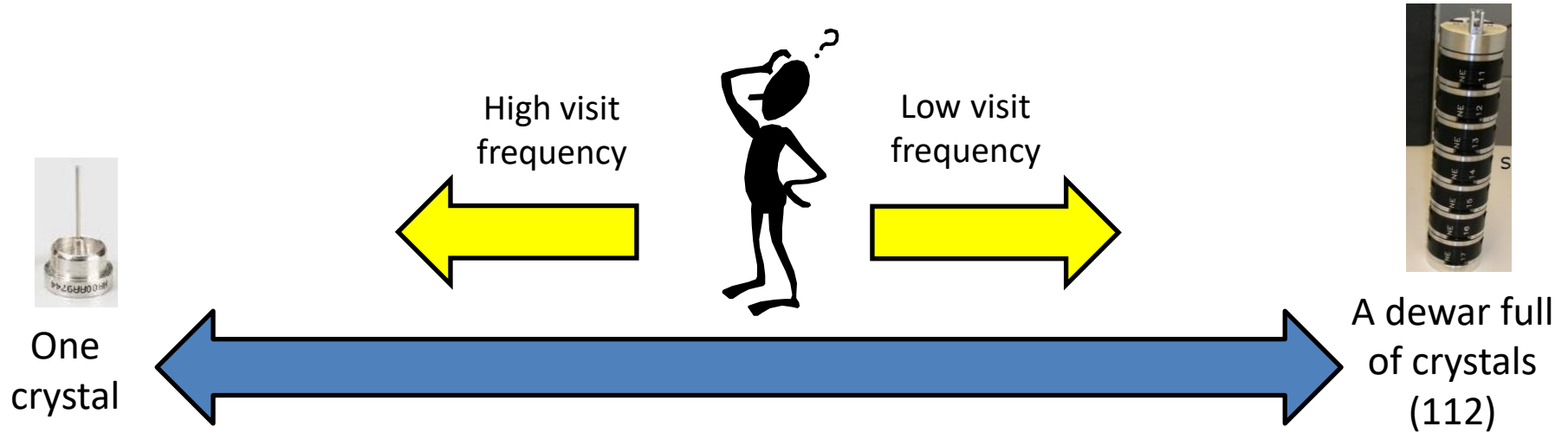
Primary goals (data quantity AND quality...)

- 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...



Effort before, during and after beamtime

Time available for other projects/users

Number of crystals

Likelihood of 'success' within a single visit

Number of crystals

Data collection is always a compromise...

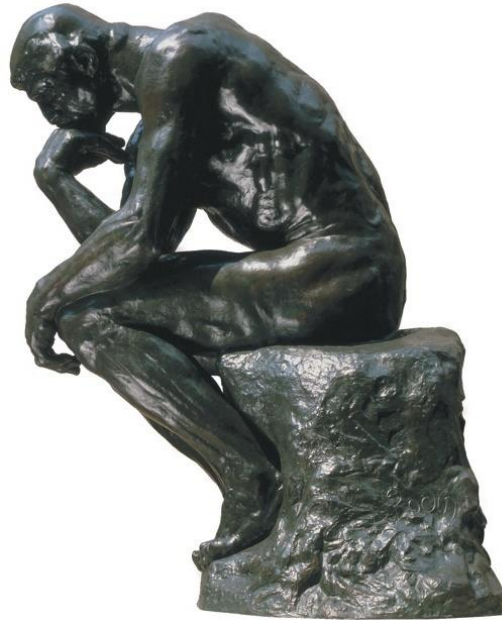
*How much
data do I
need?*

*What
resolution
do I need?*

*What are
the data
for?*

*How much
time do I
have?*

*How many
crystals do
I have?*

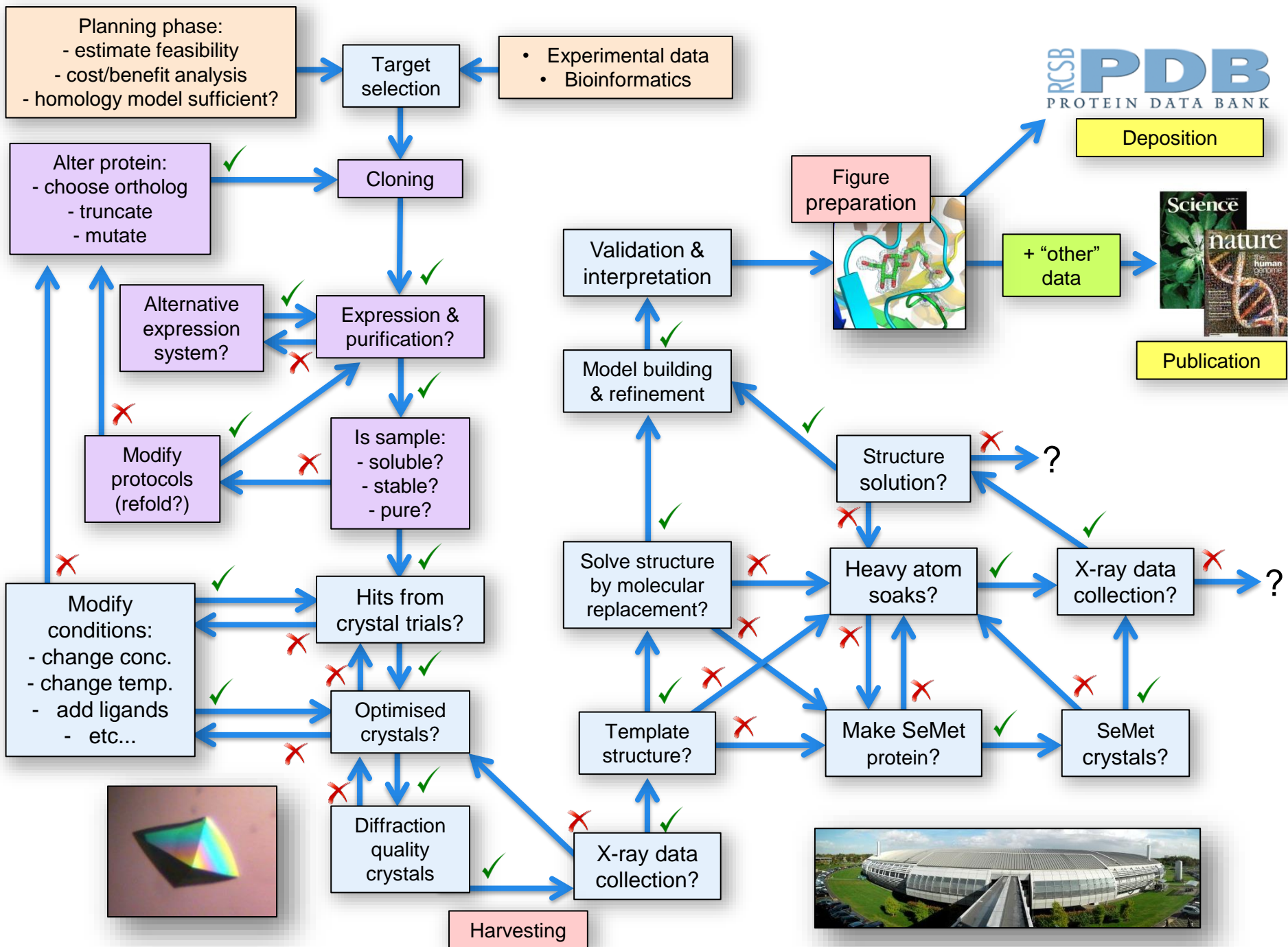


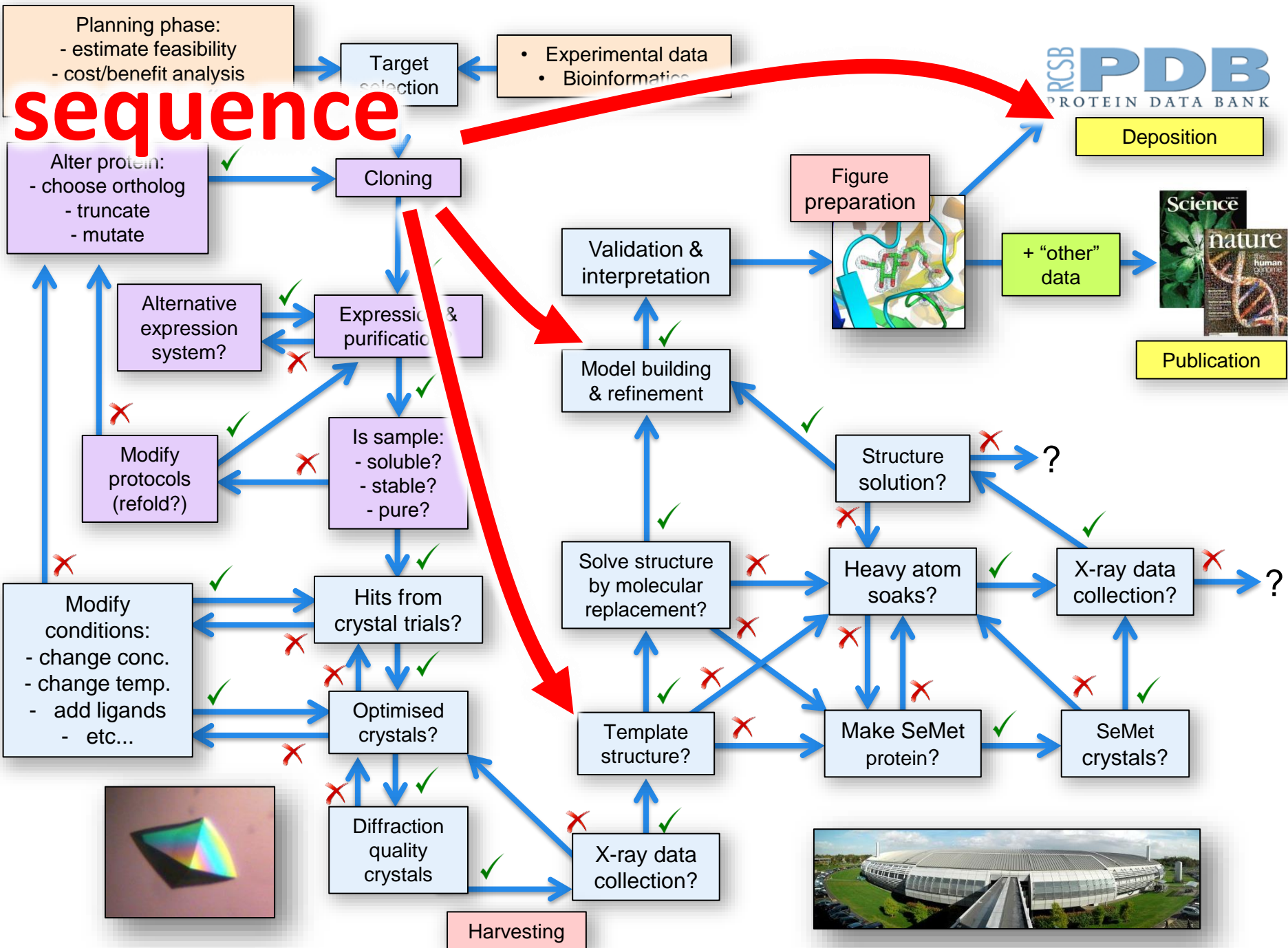
When is it my turn?

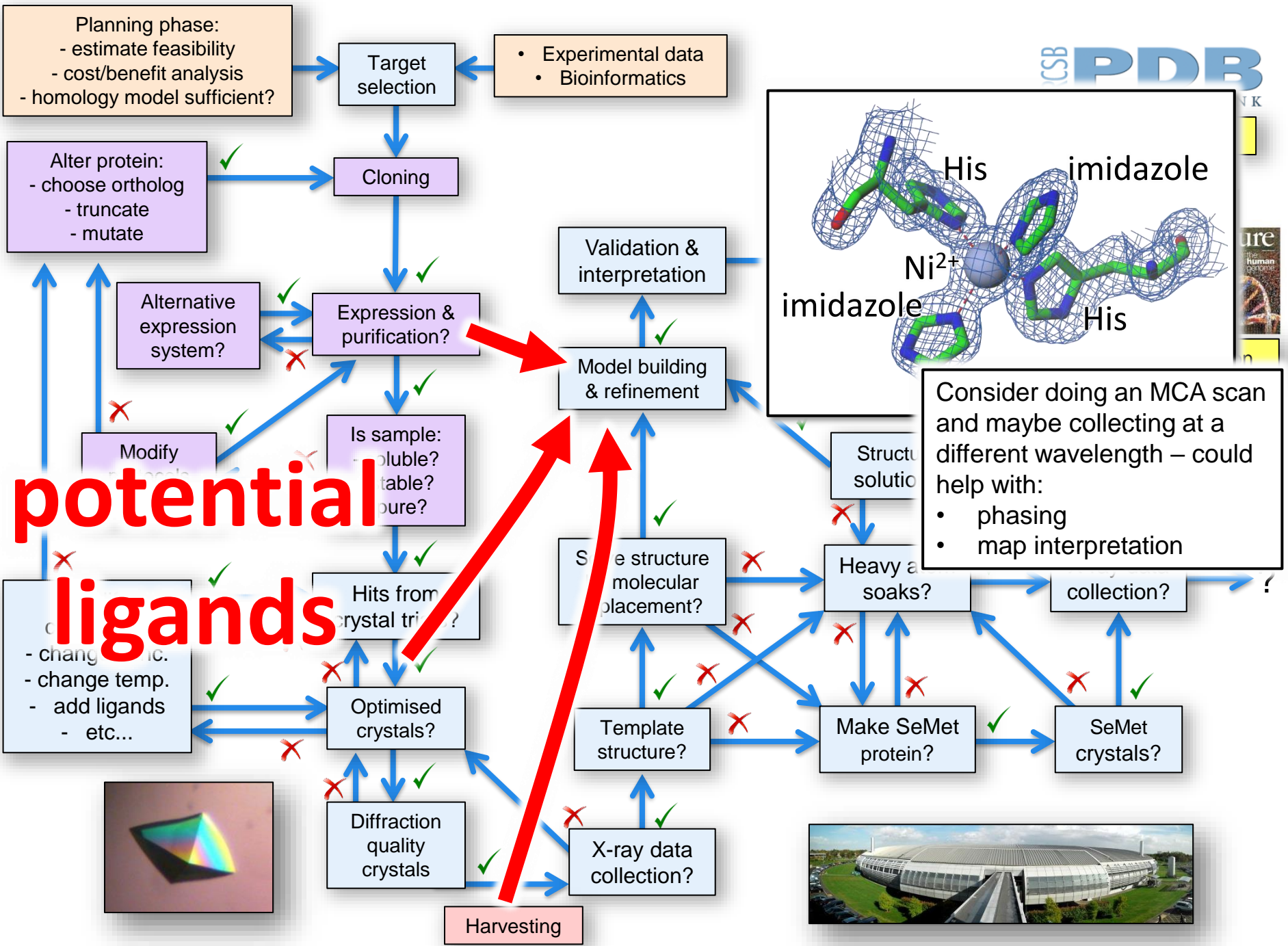


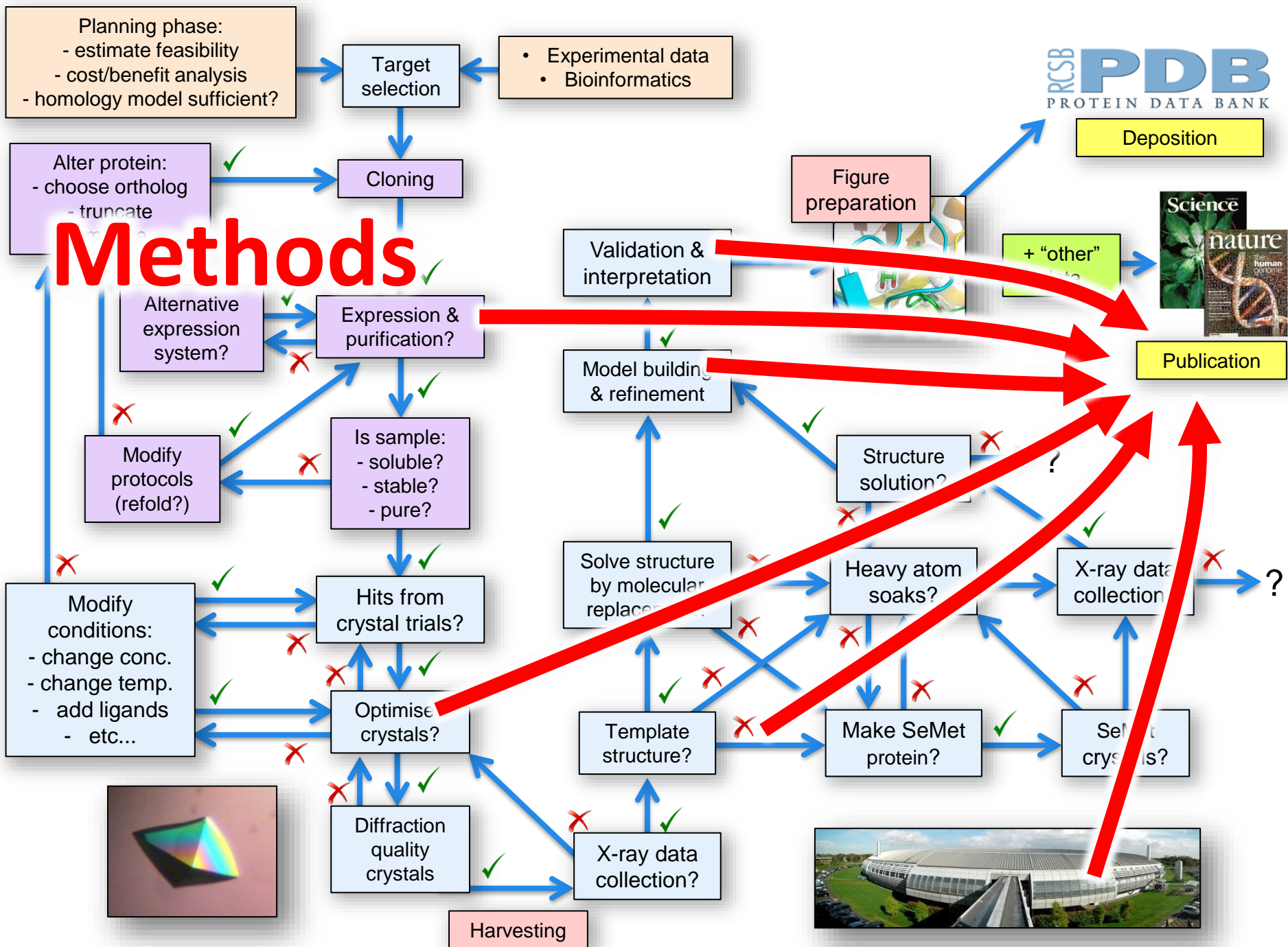
Be prepared!

Think about this **before** your beamtime!

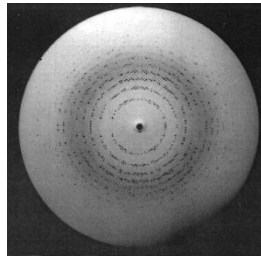








Detector types



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

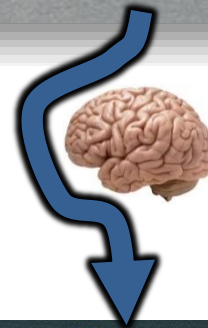
2020

MX data collection has become faster...

20th century data collection



21st century data collection



MX data collection has become much faster...

- In principle, could collect one dataset every few mins
(mounting/aligning + evaluating test images + collection)
- In practice, ~5 “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
- The session might be shared between several groups
 - therefore need to be efficient and organised...



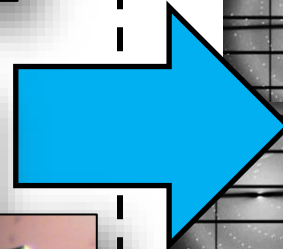
Crystals

*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"*

Help is at hand...



ISPyB/SynchWeb

+



CCP4i2

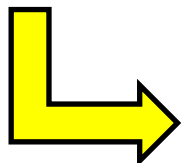
+



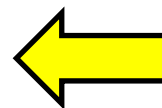
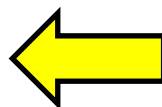
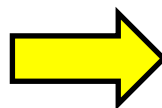
CCP4 Cloud



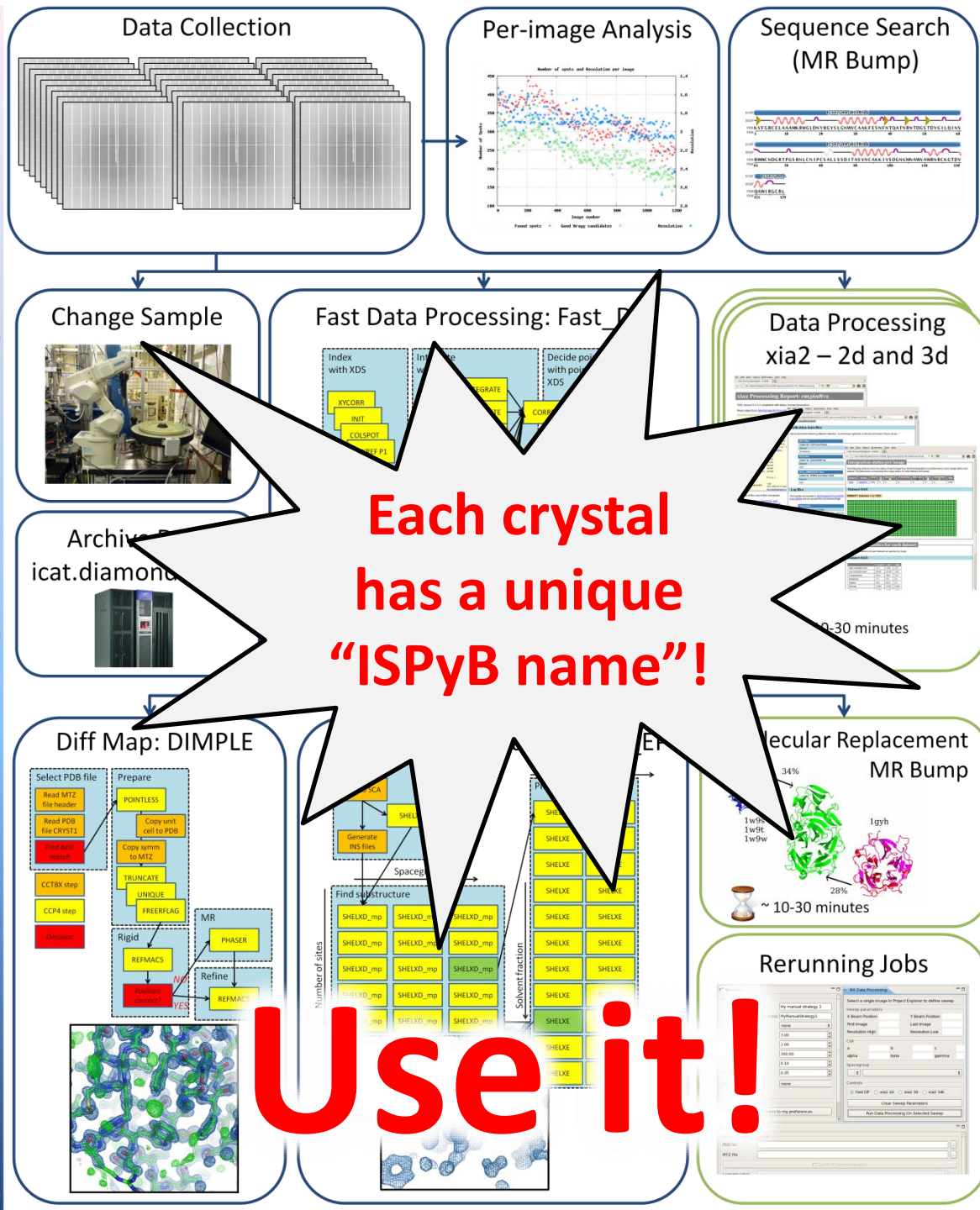
Sample info.

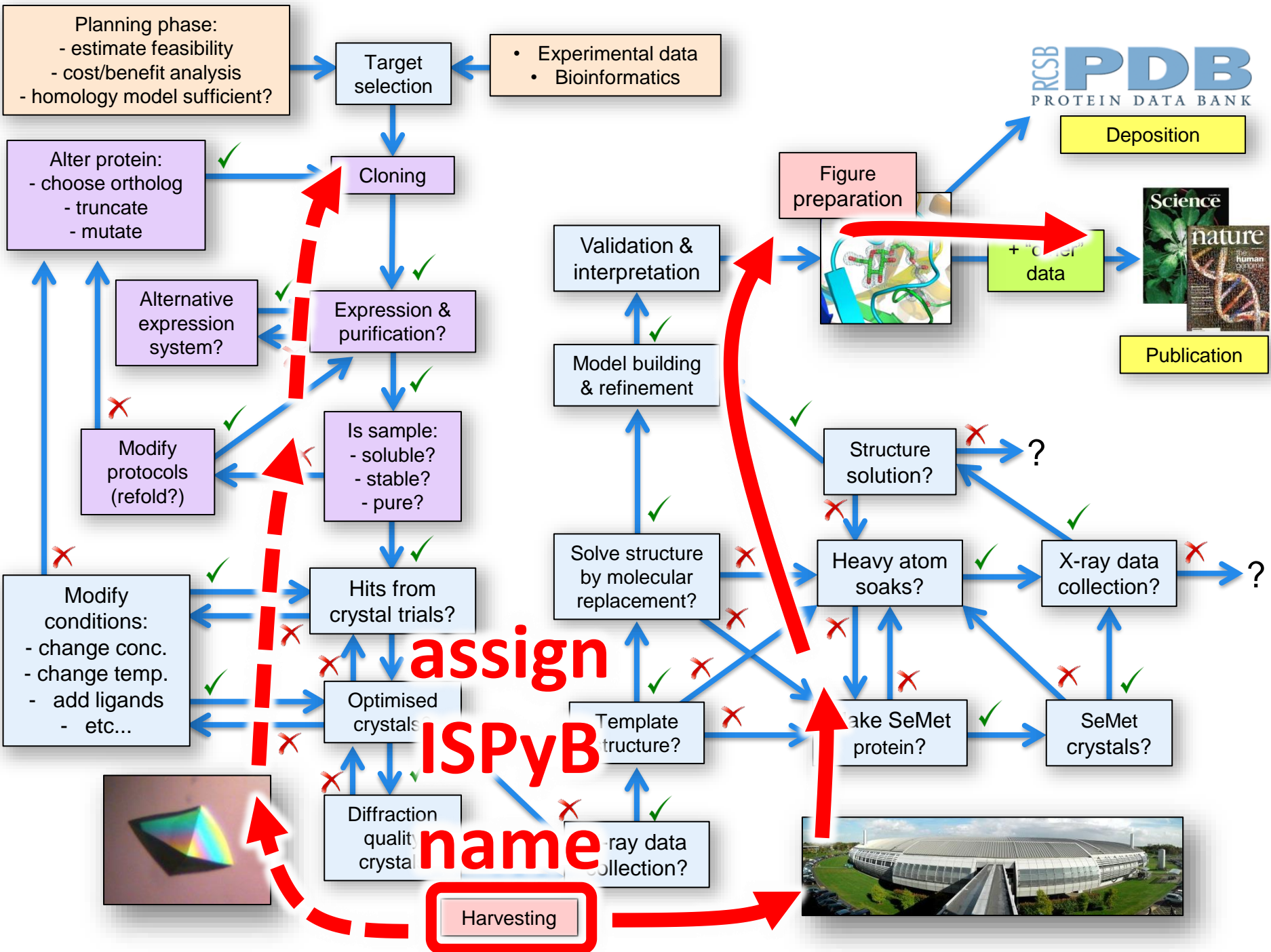


ISPyB database



Time / s
0
60
120
180
240





Date: 29-JUL-2017

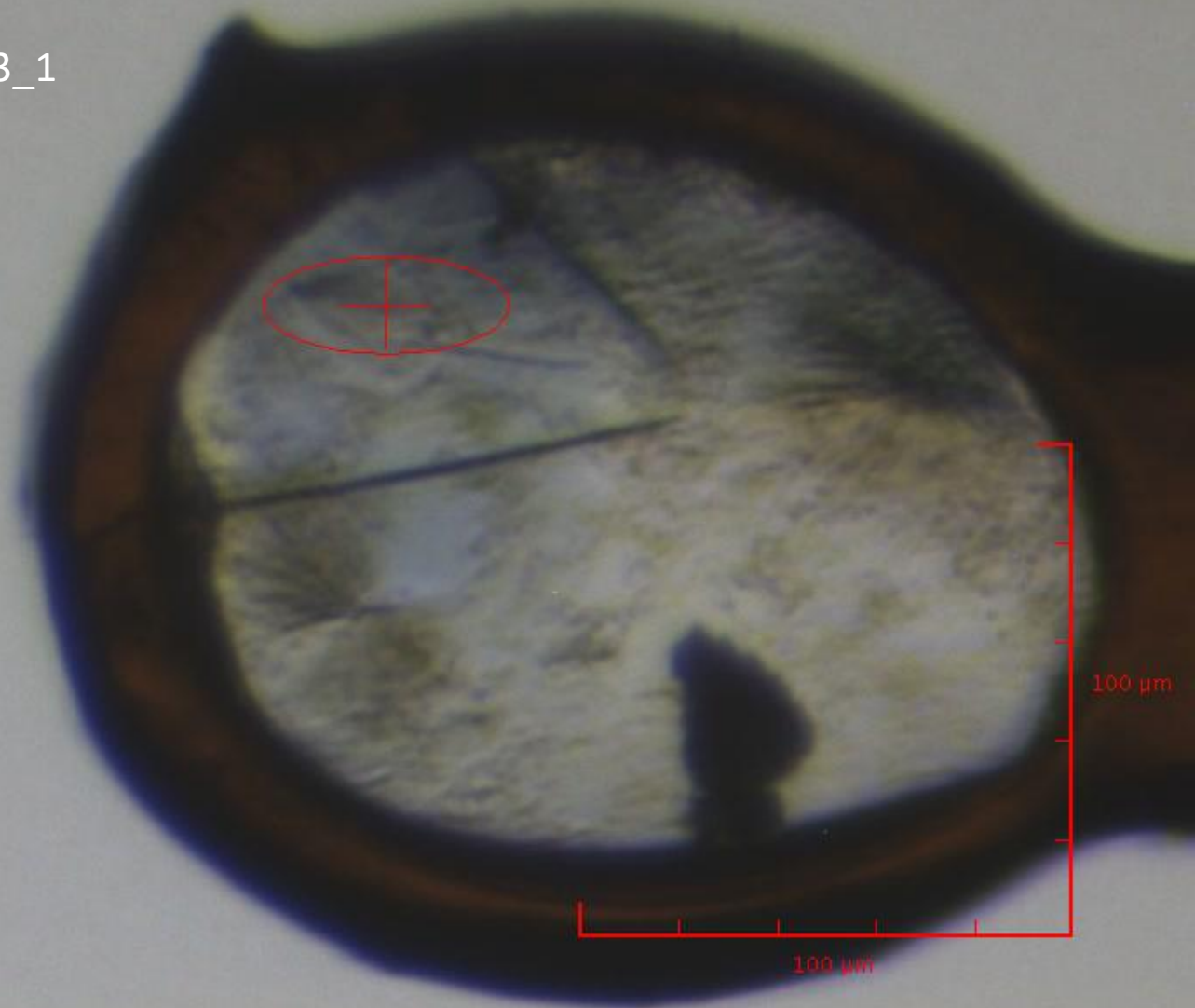
Beamline: i03 Diamond

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



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				Q
10	NmADH9			NmADH9_19				Q
11	NmADH9			NmADH9_20				Q
12	NmADH9			NmADH9_21				Q
13	NmADH9			NmADH9_22				Q
14	NmADH9			NmADH9_23				Q

Basic Extra Fields Unattended (UDC)

Loc	Protein Acronym	Name	Sample Group	Anomalous	Barcode	Comment	Status
1	LYS	JIC_LYS_OCT	-				

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 hydroxygeraniol+ 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. At the top, there are three tabs: 'Data Collection Settings', 'Plate View', and 'Screening'. The 'Screening' tab is highlighted with a red box and a red arrow pointing to it from the right. Below the tabs, there is a 'Run Scan' button. The main area is divided into several sections: 'Sample', 'Omega', 'Image', and 'Beam and Detector'. In the 'Sample' section, a dropdown menu is open, showing 'NmADH9_23' as the selected sample, which is highlighted with a red box. Below this, there are fields for 'Barcode' (NR), 'Holder' (2), and 'Position' (14). In the 'Files' section, the 'Visit directory' is set to '/dls/i03/data/2017/mx13467-41'. The 'Folder' field is highlighted with a red box and contains the text 'JIC/\${proteinacronym}/\${samplena'. Below this, there is a 'Prefix' field with the text '\${samplename}' and a 'Run number' field with the value '0'. In the 'Image' section, the 'Number of images' is set to '3600', 'Exposure time' is '0.010 s', 'Total exposure time' is '36.0 s', and 'First image number' is '1'. In the 'Beam and Detector' section, 'Maximum resolution' is '1.3000 Å', 'Detector distance' is '213.5 mm', 'Wavelength' is '0.97623 Å', 'Energy' is '12700.3 eV', and 'Transmission' is '50.156283 %'. A red arrow points from the 'Screening' tab to the 'Image' section, and another red arrow points from the 'Sample' dropdown to the 'Folder' field.

Data Collection Settings

Sample

Run Scan

NmADH9_23

Barcode NR

Holder 2

Position 14

Files

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

Folder JIC/\${proteinacronym}/\${samplena

Prefix \${samplename}

Run number 0

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

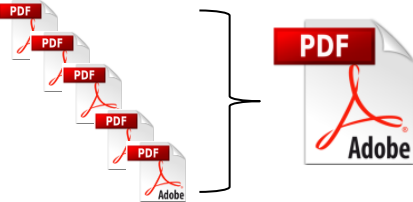
Detector distance 213.5 mm

Wavelength 0.97623 Å

Energy 12700.3 eV

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	λ	Q Osc	Spacegroup	Unit Cell	Processed Resolution	Rmerge	Completeness	Comments
NmADH9_23	3600	1.5	0.9763	0.10	P 1 2 1	63.84, 107.75, 90.00, 104.25, 90.00	29.77 - 1.5, 29.77 - 6.7, 1.54 - 1.5	0.085, 98.8, 0.036, 98.9, 0.024, 96.0		(-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						(-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, 102.02, 136.90, 100.105, 90.00	30.02 - 3.04, 29.02 - 13.6, 3.12 - 3.04	0.167, 99.4, 0.040, 99.8, 0.033, 93.3		(276, -373, 844) EDNAStrategy1: subWedge:1Aperture: Medium
NmADH9_11	3	2.0	0.9763	0.50						(-477, -294, 1143) 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, 95.29, 107.16, 90.00, 104.25, 90.00	29.74 - 1.69, 29.74 - 7.55, 1.73 - 1.69	0.085, 98.8, 0.036, 98.9, 0.024, 96.0		(-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 [592,405]
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					(-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 [534,253], 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

Sample	Images	Res	λ	Q Osc	Spacegroup	Unit Cell	Processed Resolution	Rmerge	Completeness	Comments
NmADH9_103	3	0.97625	289.0478	0.04	0	0.5	211.6297	206.4712	1.6	(-272, -419, 809) Aperture: Medium
NmADH9_103	3	0.97625	289.0478	0.04	0	0.5	211.6297	206.4712	1.6	(-231, -461, 1252) Aperture: Medium
NmADH9_103	3	0.97625	289.0478	0.04	0	0.5	211.6297	206.4712	1.6	(-249, -175, 1442) Aperture: Medium
NmADH9_103	3600	0.97625	264.5335	0.01	122	0.1	211.6215	206.6305	1.5	(-249, -175, 1442) EDNAStrategy1: subWedge:1Aperture: Medium
NmADH9_103	3	0.97625	289.0478	0.04	0	0.5	211.6297	206.4712	1.6	(-533, 7, 1513) Aperture: Medium
NmADH9_103	8	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], Bottom right [746,517]
NmADH9_103	120	0.97625	602.2545	0.1	89.99975	0	211.7341	204.4354	2.9967	Diffraction grid scan of 3 by 4 images, Top left [475,268], Bottom right [642,472]
NmADH9_103	3	0.97625	427.9818	0.04	0	0.5	211.676	205.5681	2.2	(-488, 28, 1516) Aperture: Medium
NmADH9_103	3	0.97625	427.9818	0.04	0	0.5	211.676	205.5681	2.2	(-896, 457, 1456) Aperture: Small
NmADH9_103	3	0.97625	427.9818	0.04	0	0.5	211.676	205.5681	2.2	(-25, 110, 1069) Aperture: Small
NmADH9_103	10	0.97625	602.2545	0.1	89.99975	0	211.7341	204.4354	2.9967	Diffraction grid scan of 7 by 5 images, Top left [327,239], Bottom right [642,472]
NmADH9_103	11	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], Bottom right [642,472]
NmADH9_103	2	0.97625	427.9818	0.04	0	0.5	211.676	205.5681	2.2	(-33, 96, 1067) Aperture: Small
NmADH9_103	3600	0.97625	359.8787	0.01	106	0.1	211.6533	206.0108	1.9	(-33, 96, 1067) EDNAStrategy1: subWedge:1Aperture: Small
NmADH9_103	3	0.97625	427.9818	0.04	0	0.5	211.676	205.5681	2.2	(-241, -208, 1171) Aperture: Medium
NmADH9_103	2	0.97625	427.9818	0.04	0	0.5	211.676	205.5681	2.2	(-262, -192, 1150) Aperture: Medium
NmADH9_103	3	0.97625	264.5335	0.04	0	0.5	211.6215	206.6305	1.5	(-262, -192, 1150) Aperture: Medium
NmADH9_103	3600	0.97625	213.4382	0.01	45	0.1	211.6045	206.9627	1.3	(-262, -192, 1150) EDNAStrategy1: subWedge:1Aperture: Medium
NmADH9_103	3	0.97625	264.5335	0.04	0	0.5	211.6215	206.6305	1.5	(-107, -190, 1012) Aperture: Medium
NmADH9_103	2	0.97625	264.5335	0.04	0	0.5	211.6215	206.6305	1.5	(-107, -190, 1012) Aperture: Small
NmADH9_103	2	0.97625	264.5335	0.04	0	0.5	211.6215	206.6305	1.5	(-156, -228, 1396) Aperture: Medium
NmADH9_103	2	0.97625	264.5335	0.04	0	0.5	211.6215	206.6305	1.5	(-215, -228, 1406) Aperture: Medium
NmADH9_103	3	0.97625	264.5335	0.04	0	0.5	211.6215	206.6305	1.5	(-247, -388, 861) Aperture: Medium
NmADH9_103	12	0.97625	602.2595	0.1	-0.00016	0	211.7341	204.4353	2.9966	Diffraction grid scan of 7 by 6 images, Top left [350,177], Bottom right [746,517]
NmADH9_103	13	0.97625	602.2595	0.1	89.99975	0	211.7341	204.4353	2.9966	Diffraction grid scan of 3 by 4 images, Top left [472,245], Bottom right [642,472]
NmADH9_103	3	0.97625	264.5335	0.04	0	0.5	211.6215	206.6305	1.5	(276, -373, 844) Aperture: Medium
NmADH9_103	4	0.97625	385.1022	0.04	0	0.5	211.6609	205.8617	2	(276, -373, 844) Aperture: Medium
NmADH9_103	3600	0.97625	385.1022	0.01	103	0.1	211.6617	205.8668	2.01	(276, -373, 844) EDNAStrategy1: subWedge:1Aperture: Medium

19:00 26-07-2017

02:00 27-07-2017

02:00

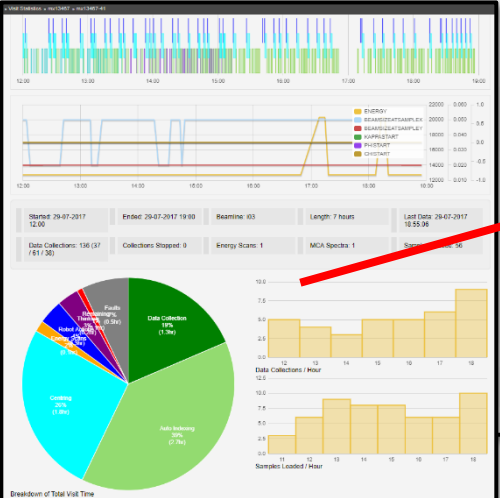
17:00

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	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 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:

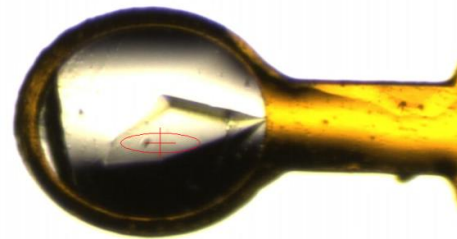
- Load first sample
- Collect test images
- Based on these, decide:
 - to collect...
 - not to collect...
 - to revisit later...
- For a “suitable” sample:
 - devise a data collection strategy
 - collect data set
- Analyse data as they are collected
- Based on this analysis, revise plans if appropriate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

– not efficient
without automation



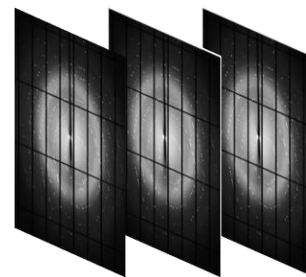


robotic sample exchange

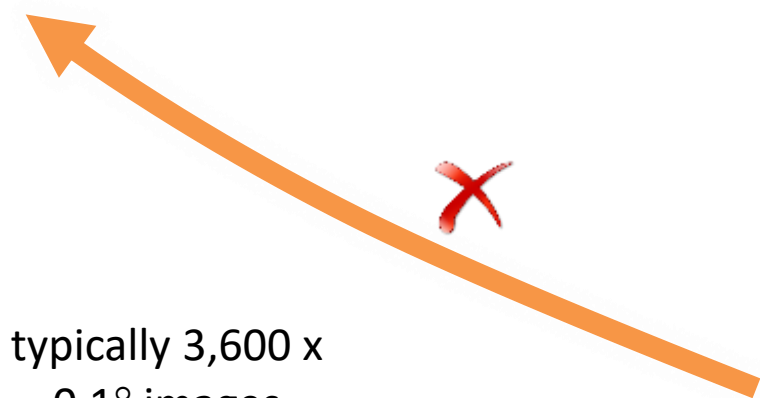


align crystal in X-ray beam

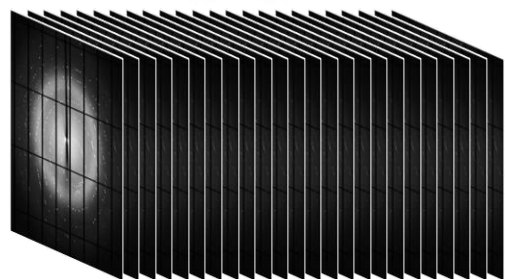
analyse data
on-the-fly



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)



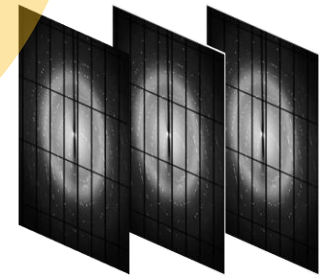
Don't
forget to
think!



robotic sample exchange



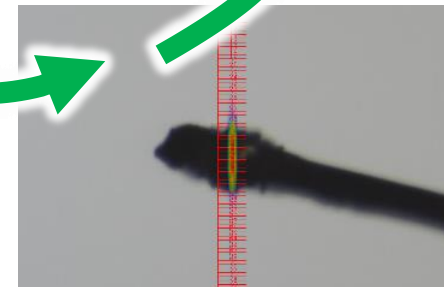
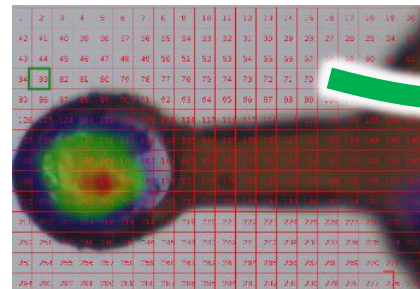
align crystal in X-ray beam

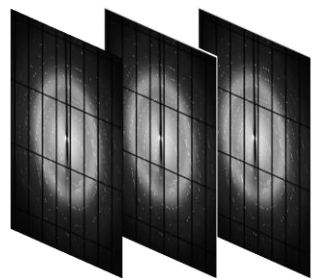


collect test images

May be more efficient to screen your samples in batches, then decide what to do...

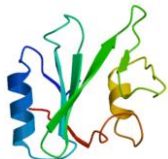
- do this automatically with X-ray centring
- also an opportunity to grab a coffee!





test images

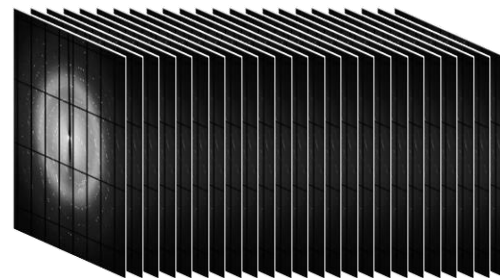
before session...



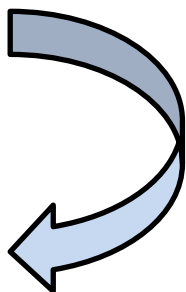
template

```
10      20      30      40      50      60
MANLGNELV LFVAWEDLG LKKRKPKGG MFGGGRTPG QGPGGRYP PGGGGGGQP
70      80      90     100     110     120
HGGGNGPHG GGGGPHGG WGPFGGGG QGGGTHGNN KFKPKPKNK IMAGAAAAGA
130     140     150     160     170     180
VVGGLGGML GMAERPIIH FGSDTEDRYT KEIMERTPG VTRPMDEYS NNNFVIDCV
190     200     210     220     230     240
NITIKQHYT TTKGKSPTE TDVQRRRVV KMCITQVER EGGATYGRG SMVLFPPFV
```

sequence



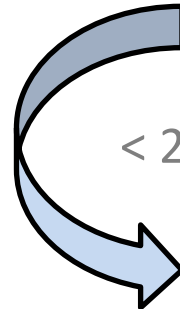
dataset



Look at images
(e.g. in ADXV)
- can they be
indexed?



ISPyB

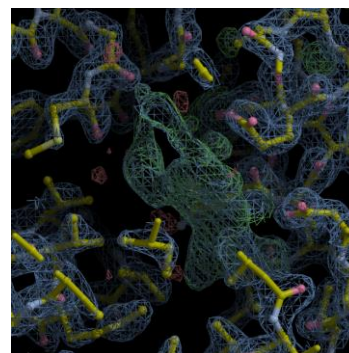
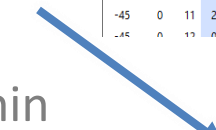
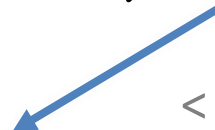


< 2 min

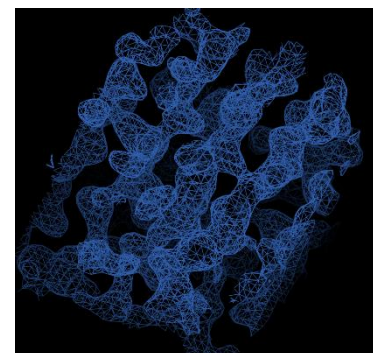
Fast DP

h	k	l	p _{plus}	SIG p _{plus}	p _{minus}	SIG p _{minus}
-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

< 5 min

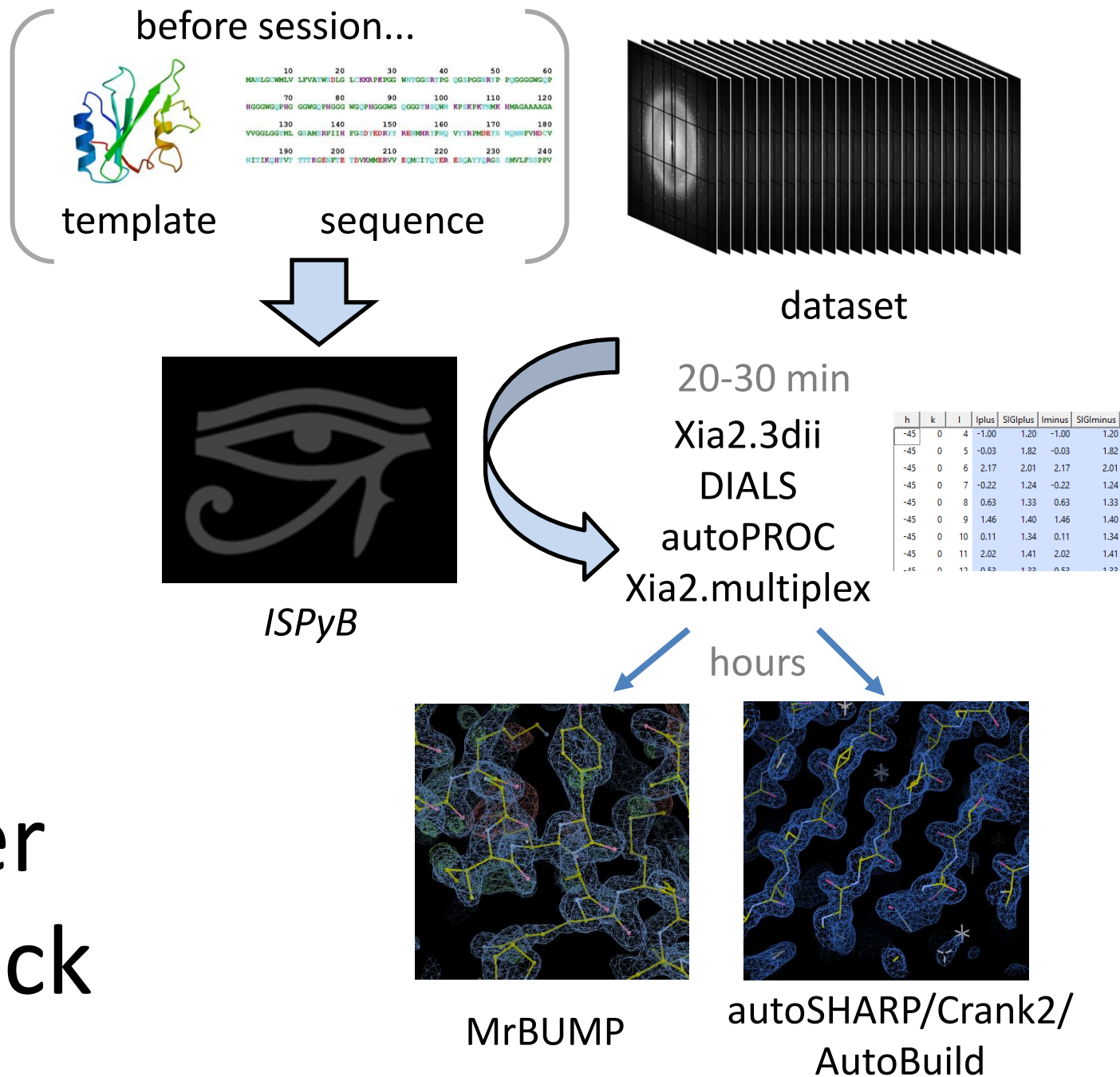


Dimple



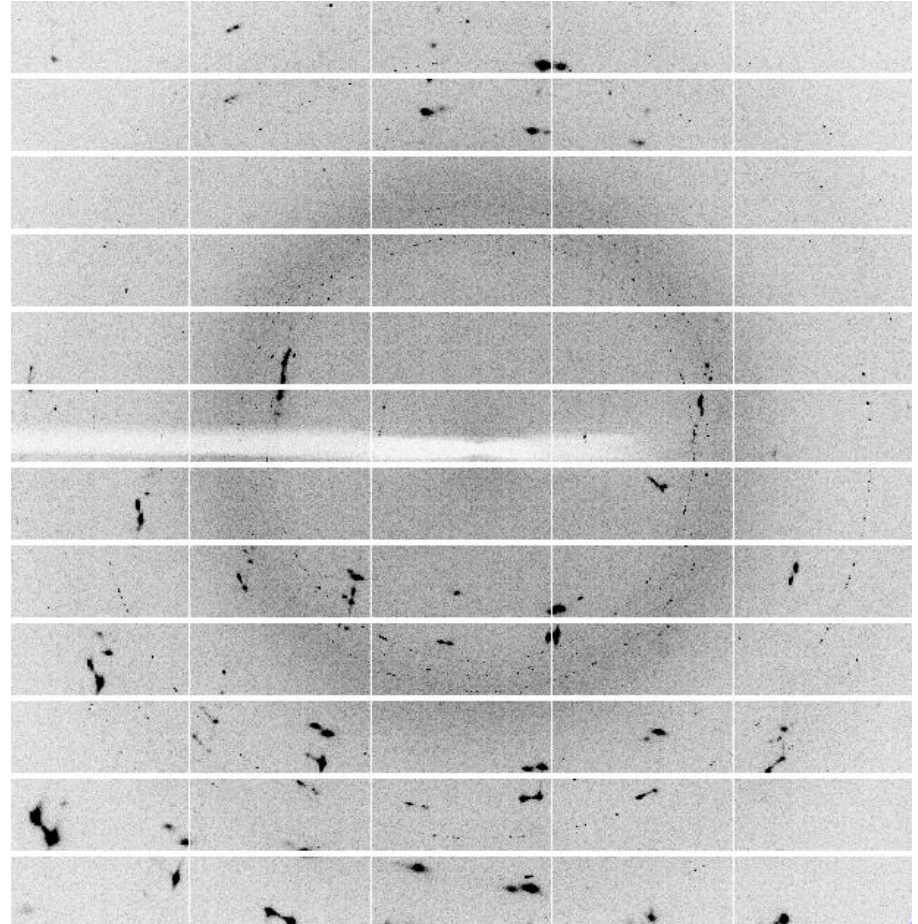
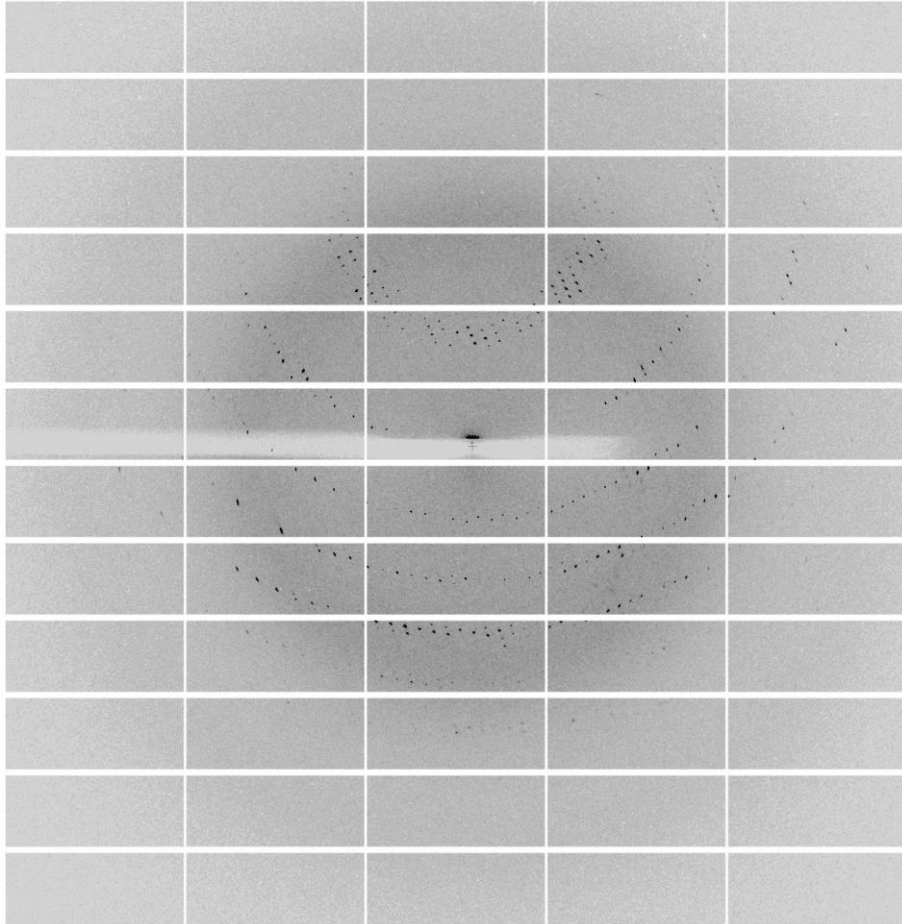
Fast EP

Quick
feedback

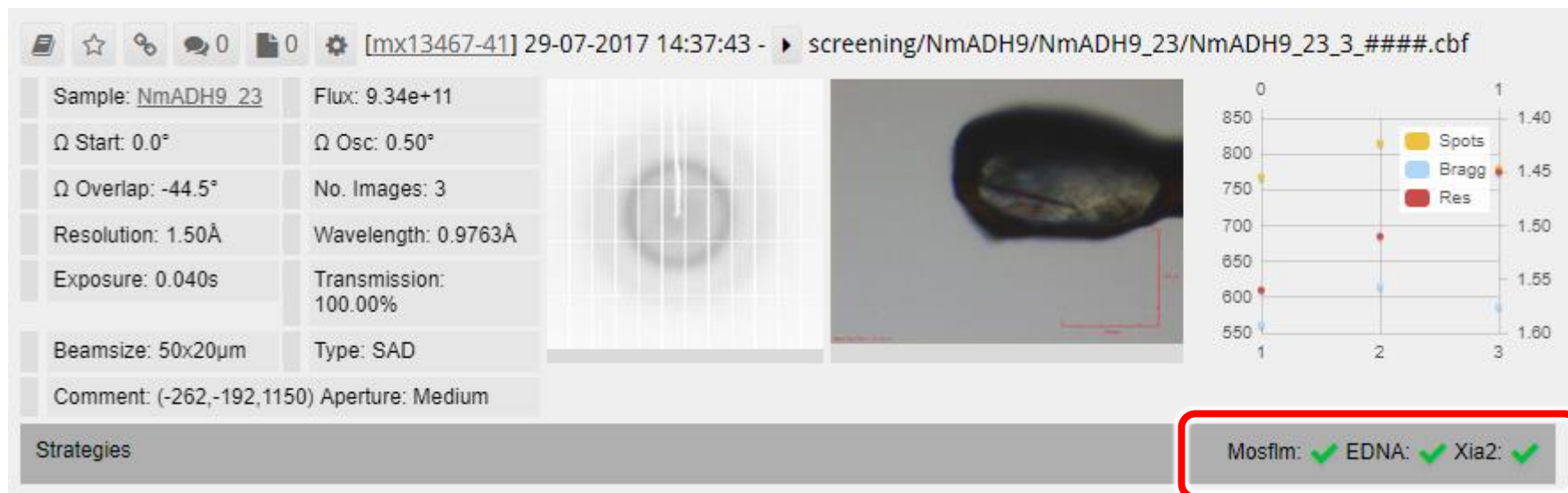


Slower
feedback

LOOK at your images!!!!



Screening...



Screening...

Have I seen
this crystal
form before?

Mosfilm: ☒ EDNA: ☒ Xia2: ☒

Q Lookup Cell

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

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	1.1	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	Native Dataset Multiplicity=3 I/sig=2 202s	162	0.15	1.51	1.1	2.2	2.2	0.010	1947

EDNA MXV

Space Group	A	B	C	α	β	γ
P2	63.99	107.46	69.31	90.00	104.25	90.00

Q Lookup Cell

Strategy

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 202s	0	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	6	0.10	1.51	1.24	12.4	12.4	0.010	2740
Strategy3 Wedge1	strategy with target multiplicity=1 s	0	0.10	1.51	1.24	5.1	5.1	0.010	3600
Strategy4 Wedge1	Gentle: Target Multiplicity=2 and s	6	0.10	1.51	1.24	14.8	14.8	0.010	1110
Strategy5 Wedge1	UnderDEV Anomalous Dataset, I	162	0.10	1.51	1.24	2.2	2.2	0.010	2740

mosfilm

Space Group	A	B	C	α	β	γ
P2	63.80	107.65	69.44	90.00	104.25	90.00

Q Lookup Cell

Strategy

Description

Ω Start	Ω 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

Strategies are currently not reliable



Notes about strategies...

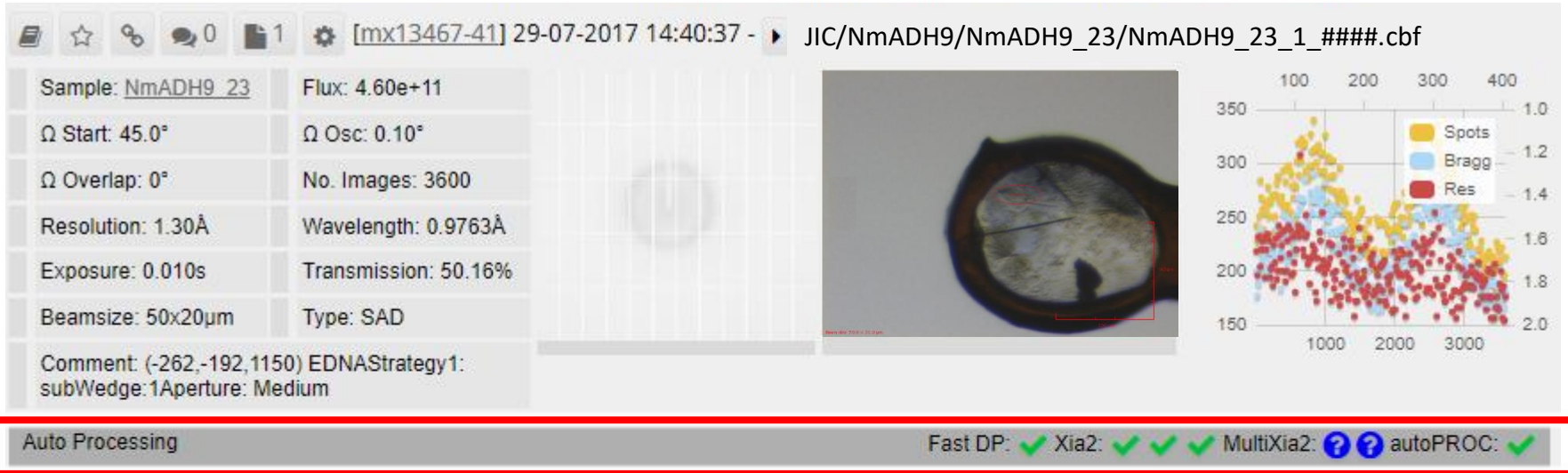
Despite what your supervisor or colleague advises...

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

If your test images don't index, but the crystal is diffracting reasonably well...

- collect a data set anyway – it might be useful (e.g. to establish what you crystallized!)
 - inspect screening images in ADXV
 - look for highest resolution spots (NOT ice or salt spots!)
 - subtract 0.5 from this resolution value. E.g. spots at 2.5 \AA → collect to 2.0 \AA
 - collect $3,600 \times 0.1^\circ$ images

Checking the results...



Checking the results...

Auto Processing

Fast DP: ✓ Xia2: ✓ ✓ ✓ MultiXia2: ? ? autoPROC: ✓

Type	Resolution	Spacegroup	ln<I/sig(I)>	Rmeas Inner	Rmeas Outer	Completeness	Cell	Status
fast_dp	29.77 - 1.50	P 1 2 1	5.2	0.036	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	0.9	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	0.8	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	0.9	0.039	1.753	98.2	63.94 107.77 69.37 90.00 104.26 90.00	
autoPROC 1.0.5 (see: http://www.globalphasing.com/autoproc/)	107.76 - 1.50	P 1 2 1 1	3.3	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: <http://www.globalphasing.com/autoproc/>)

Beam Centre X Y

Start 211.60 206.96

Refined 211.50 206.92

Δ 0.10 0.04

Space Group A B C α β γ

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

Plots

Archive

Logs & Files

Lookup Cell

Consistent indexing gives more confidence in the results...

... but treat space group assignment as only a hypothesis at this stage! (see later talks)



Re-running jobs:

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

The screenshot displays the ISPyB web interface for managing X-ray diffraction jobs. At the top, a toolbar includes icons for file operations, a star, a link, a chat bubble, and a gear icon (highlighted with a red circle). The main header shows the job ID [mx13467-41], the date and time 29-07-2017 14:40:37, and the file path JIC/NmADH9/NmADH9_23/NmADH9_23_1_####.cbf.

On the left, a table lists job parameters:

Sample: NmADH9_23	Flux: 1.60e+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

Below the table, a comment field contains the text: Comment: (-262,-102,1150) EDNA Strategy 1. A 'Reprocess Data' button is visible.

A central text box titled 'Change:' lists the following options:

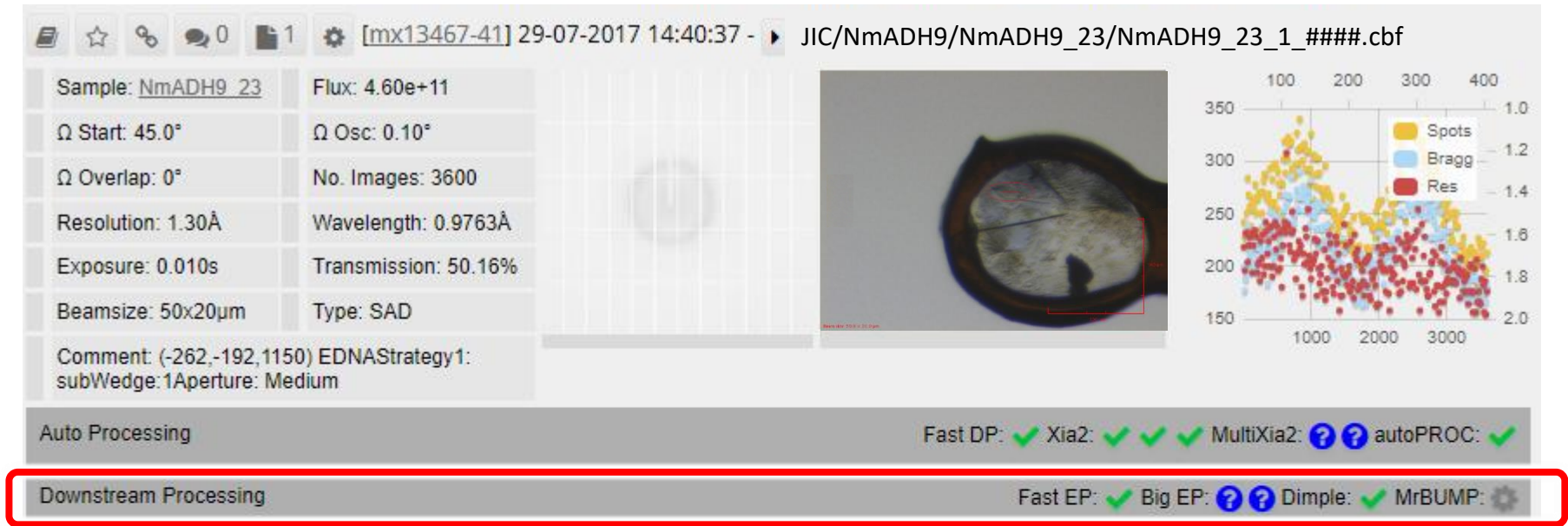
- space group
- cell parameters
- maximum resolution
- reject images

At the bottom, a 'Process Individually' checkbox is present, along with a 'Pipeline' dropdown set to 'Xia2 3dii', a 'High Res' input field, and buttons for 'Space Group / Cell' and 'Options'. The main section displays the job title 'NmADH9_23_1 - JIC/NmADH9_23/NmADH9/' and a summary of its parameters, including 'Sample: NmADH9_23', ' Ω Start: 45.0°, Osc: 0.10°', 'Resolution: 1.30Å', and 'Wavelength: 0.9763Å'. A 'Start' and 'End' input field with a '+' button is also shown.

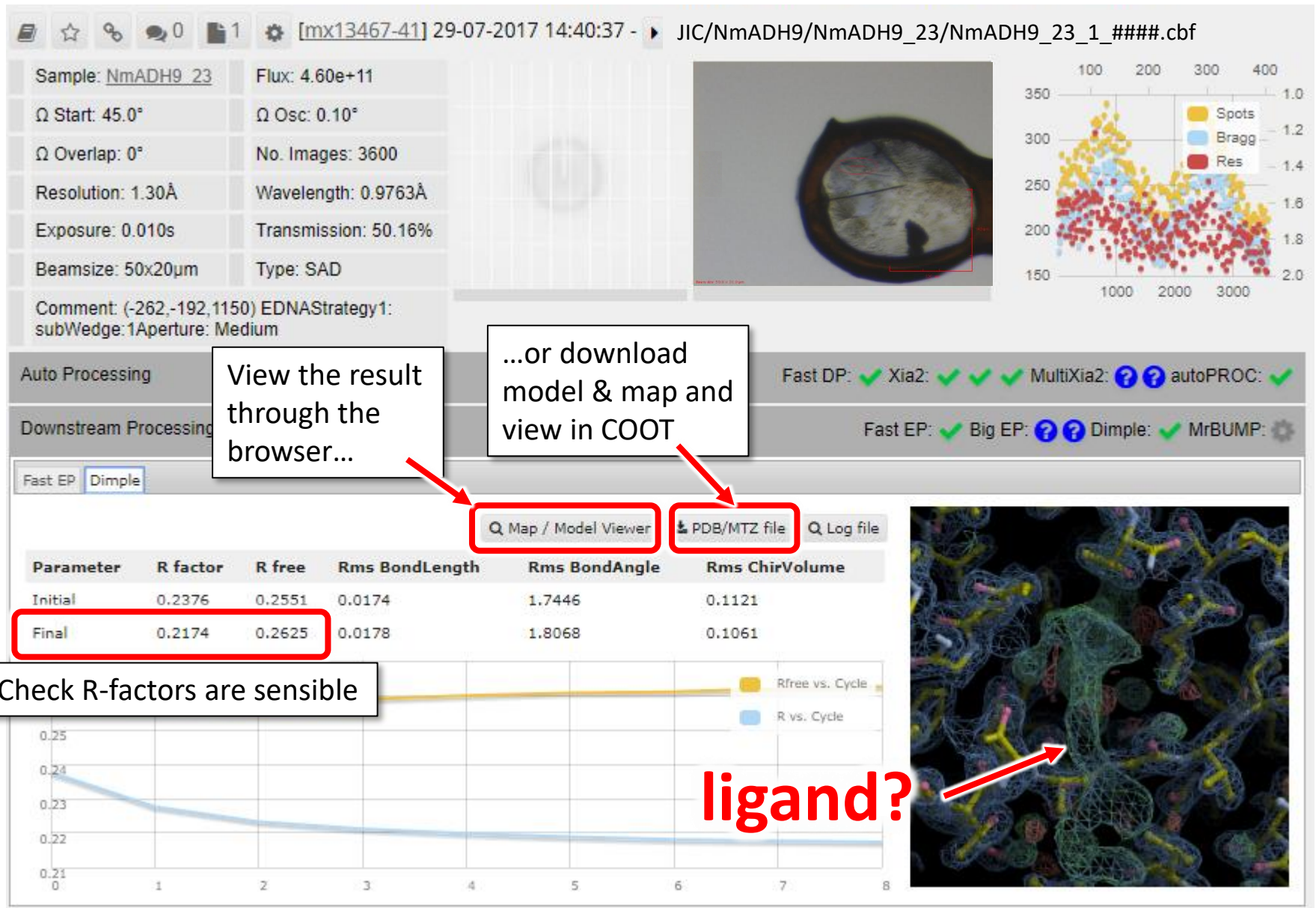
Two scatter plots are displayed on the right. The top plot shows 'Spots' (yellow), 'Bragg' (blue), and 'Res' (red) data points against a grid. The bottom plot is a similar scatter plot with a different x-axis range. Both plots include a legend and a 'Multi Crystal' button.

At the bottom right, 'Integrate' and 'Close' buttons are visible.

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!



If you used your only crystals for a project...

- still might be useful for seeding



What to do with all the data...

Raw data (images)

- removed from disk after 30 days – still available through TopCAT (tape archive)



“Meta” data – all the other “stuff” – most is removed from disk after 30 days – but important files persist for > 1 year

>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...

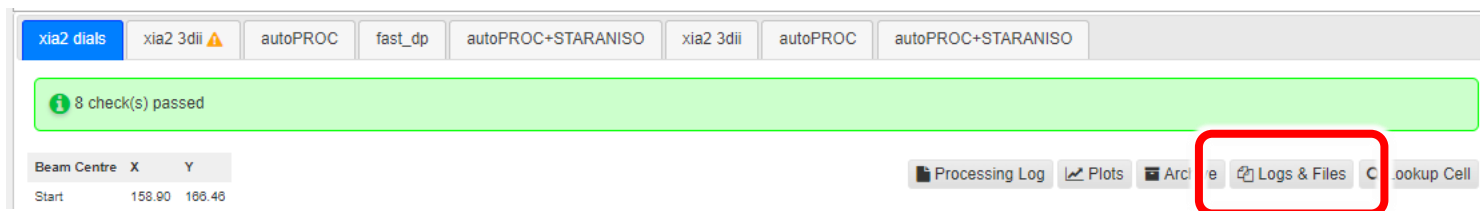
- FTP data home (use an App or a script)



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

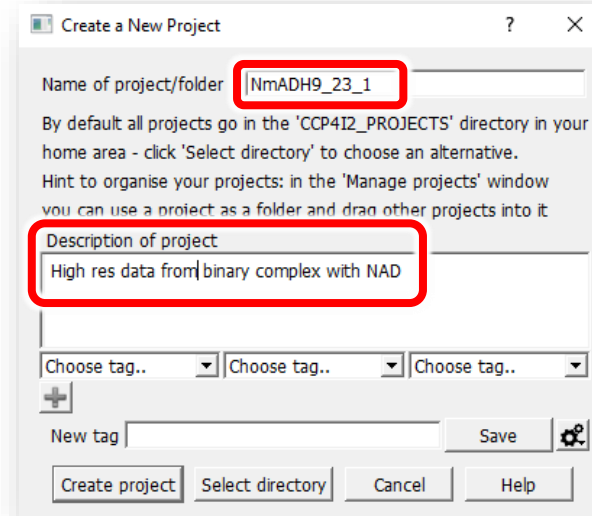
Retain ISPyB name in CCP4 downstream processing

Create a separate “project” for each dataset...

either...



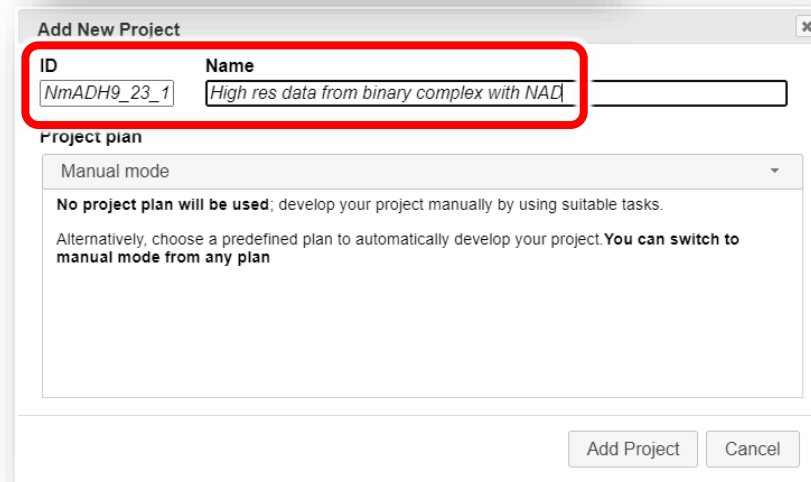
CCP4i2

A screenshot of the 'Create a New Project' dialog box. The 'Name of project/folder' field contains 'NmADH9_23_1'. The 'Description of project' field contains 'High res data from binary complex with NAD'. There are three 'Choose tag..' dropdown menus and a 'New tag' input field. At the bottom are buttons for 'Create project', 'Select directory', 'Cancel', and 'Help'.

or...



CCP4 Cloud

A screenshot of the 'Add New Project' dialog box. The 'ID' field contains 'NmADH9_23_1' and the 'Name' field contains 'High res data from binary complex with NAD'. Below is a 'Project plan' section with a dropdown menu set to 'Manual mode' and a text area containing instructions: 'No project plan will be used; develop your project manually by using suitable tasks. Alternatively, choose a predefined plan to automatically develop your project. You can switch to manual mode from any plan'. At the bottom right are buttons for 'Add Project' and 'Cancel'.

- where the pipelines have done a good job – take part-processed dataset (download this directly from SyncWeb) and re-run the merging step in CCP4i2/Cloud

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 21 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 21 1
as there is a single space group with the highest score

Solution type: space group

Group name	P 1 21 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 21 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

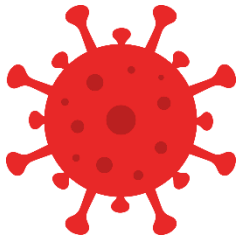


not going to cover unattended data collection (UDC)...

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



Remote data collection saves you time...



...and is covid-secure

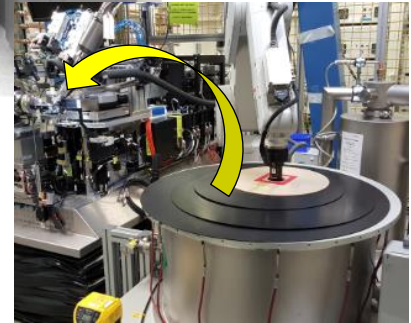


...and the planet



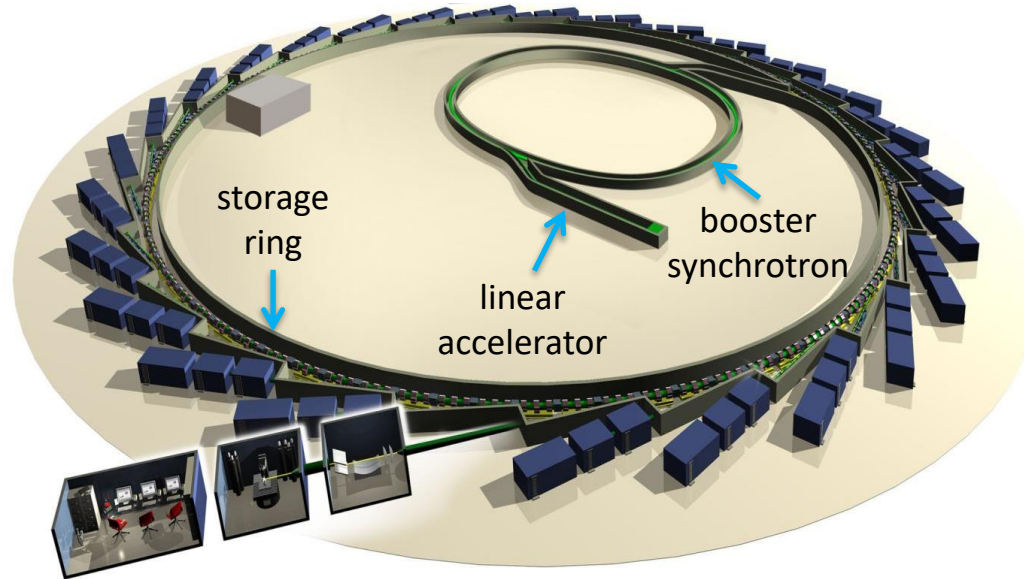
For routine data collection at 100 K:

- 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

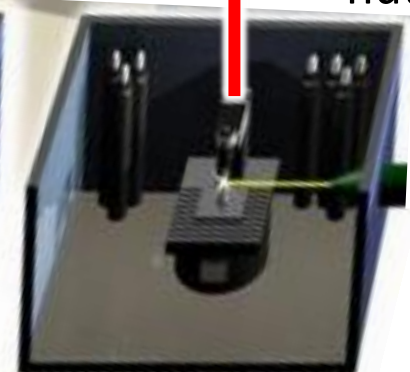
control
cabin



remote

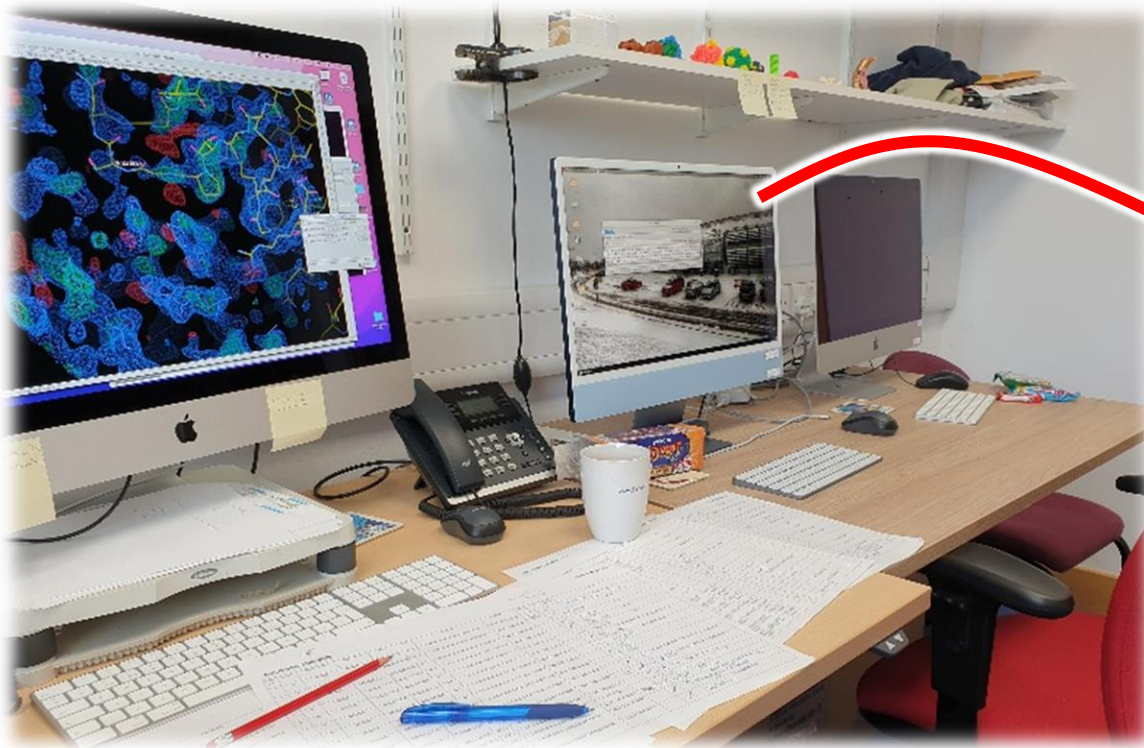


experimental
hutch



Remote data collection – in practice

...from here



share screen
with others

Do these 3 things BEFORE your session

1.

Think about the experiments that you want to run

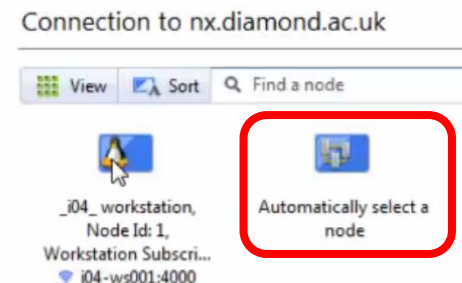
- discuss with your local contact if necessary



2.

Check that your remote NX connection works

- if you are given the option to “automatically select a node”, choose this (unless you are going to collect data)
- if intending to use a 2nd monitor – check it works



3.

While you are connected, read the beamline “message of the day”

- open a Terminal window and type:
- more /dls_sw/<beamline_name>/etc/motd

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. At the top, a status bar shows the time as 15:19 and battery level at 73%. Below the status bar, there is a search bar and a filter dropdown. The main content area features a grid of status indicators and two live webcam feeds. The status indicators include Ring Current (299.991), Refill (255.758), Hutch (Locked), Port Shutter (Open), Expt Shutter (Open), Fast Shutter (Open), and Wav (0.9). The two webcam feeds show the I04 Sample Position and the I04 Sample Changer. 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