John Innes Centre

Biotechnology and Biological Sciences Research Council

# Optimising the Diamond experience from a user's perspective

Dave Lawson

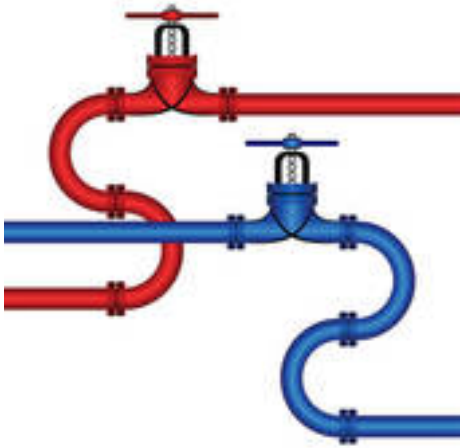
Department of Biological Chemistry

John Innes Centre, Norwich, UK

# 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 (get a life!)

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

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

## Secondary goal (get a life!)

- Try to answer these questions **during** 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…

# 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?*

Be prepared!

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

# MX data collection has become faster…

## 20th century data collection



## 21st century data collection

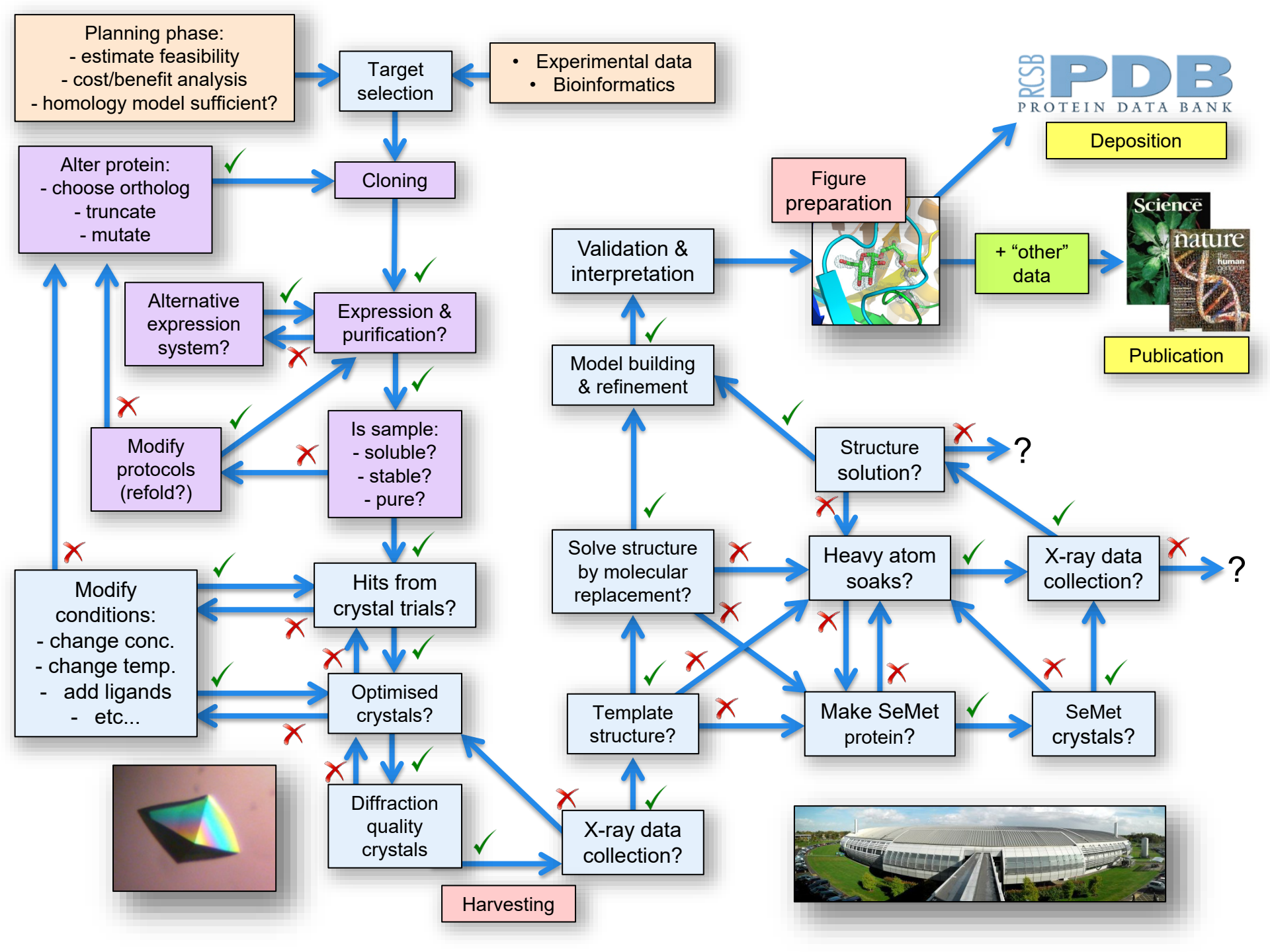
# MX data collection has become much faster…

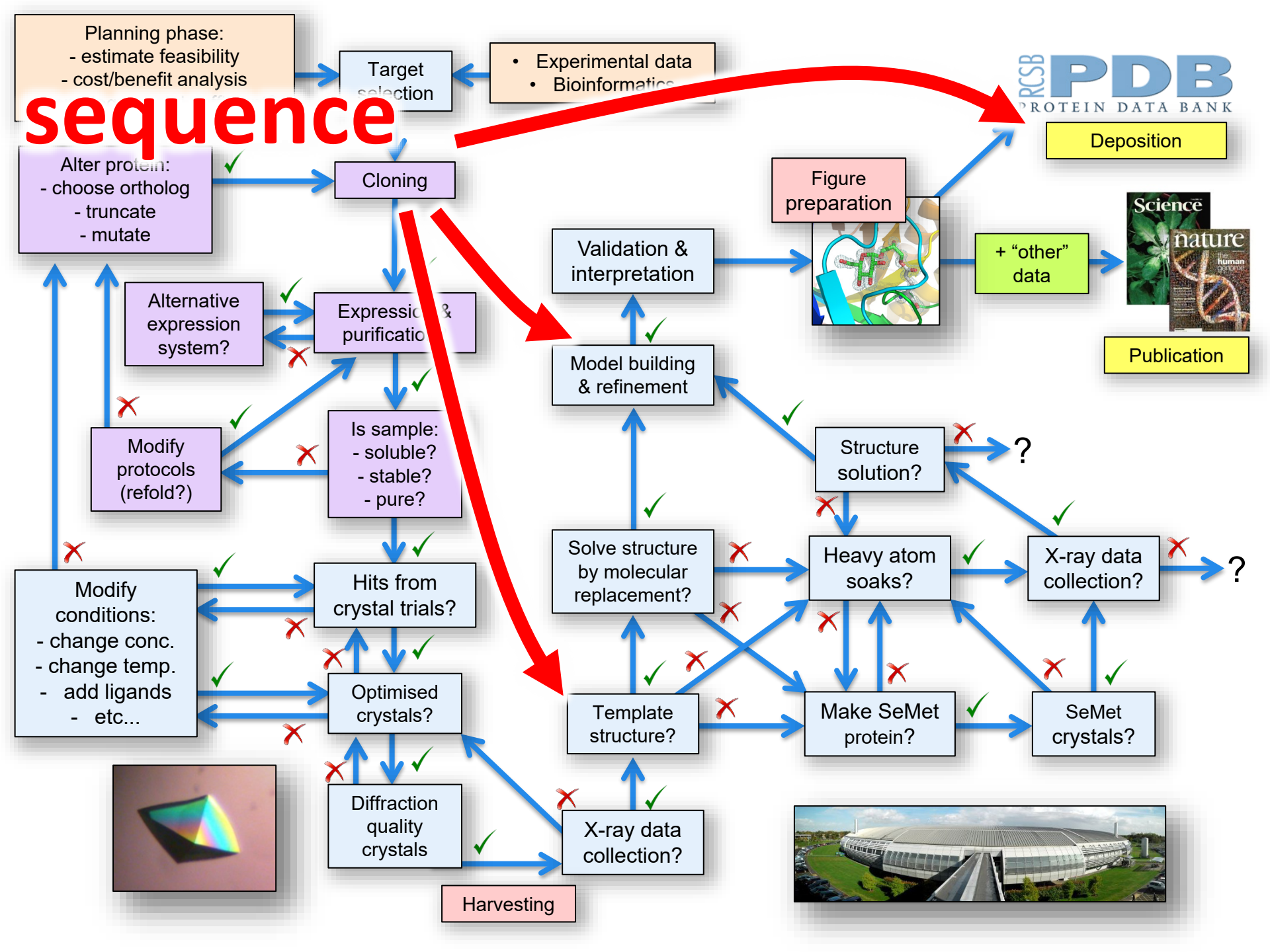
- In theory, could collect one dataset every 5 mins

  (mounting + alignment + 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...



“Take back control”


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


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

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

Target selection

- Experimental data
- Bioinformatics

Alter protein:
- choose ortholog
- truncate
- mutate

Cloning

Alternative expression system?

Expression & purification?

Modify

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

Hits from crystal tri...?

- change ...
- change temp.
- add ligands
- etc...

Optimised crystals?

Diffraction quality crystals

Harvesting

X-ray data collection?

Template structure?

S... structure molecular ...placement?

Make SeMet protein?

SeMet crystals?

Heavy atom soaks?

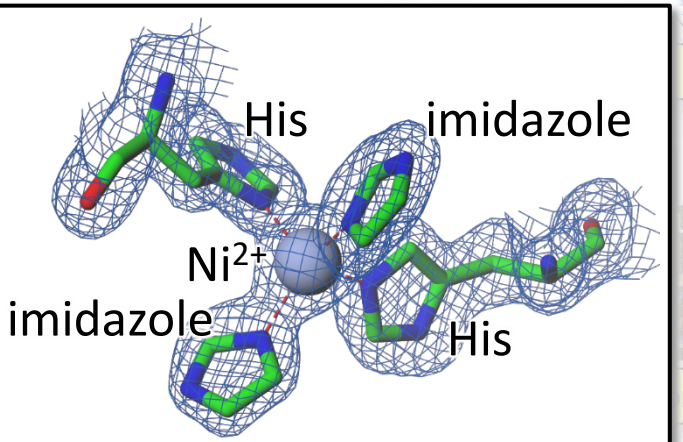
X-ray data collection?

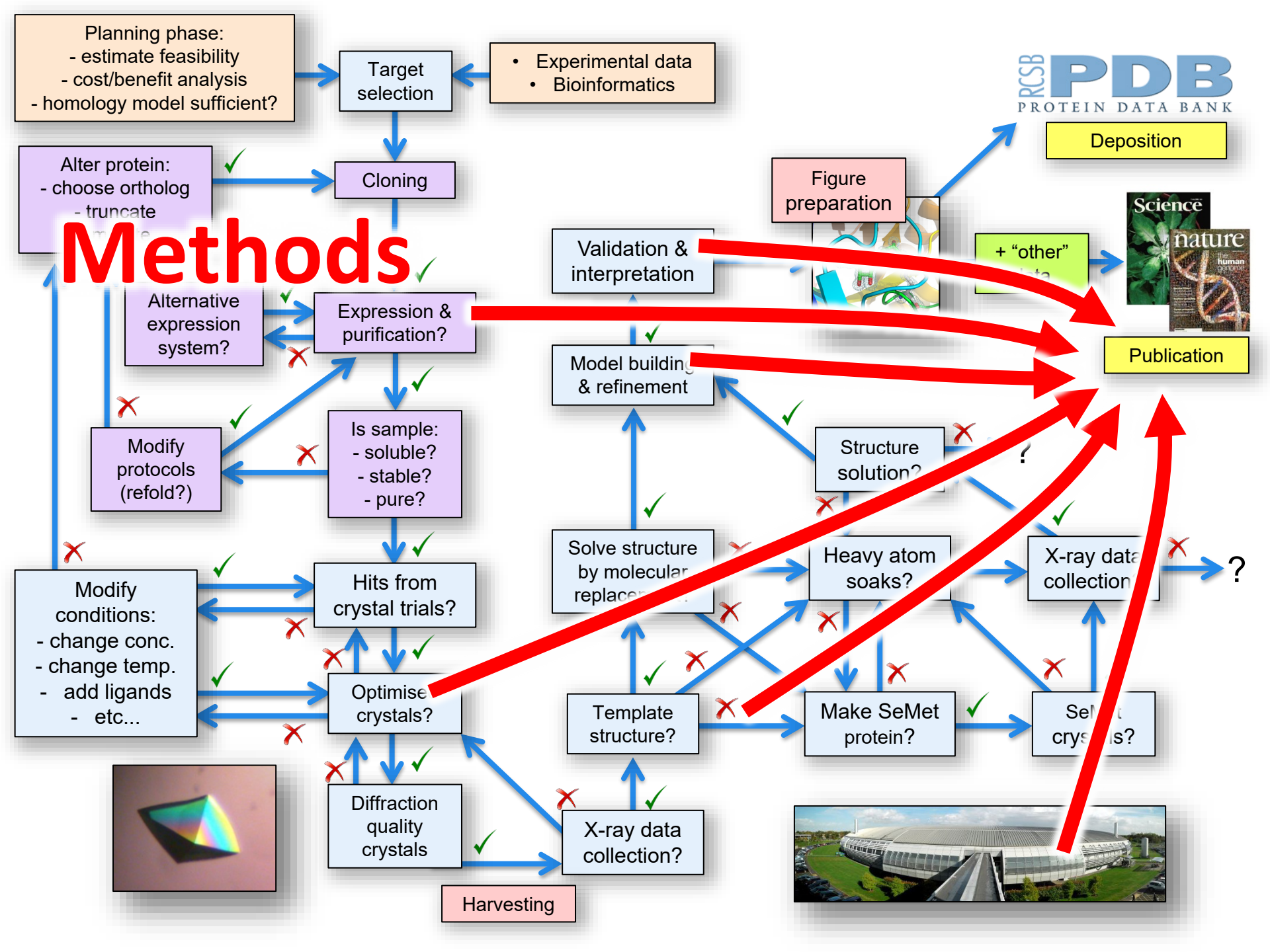
Structure solution?

Model building & refinement

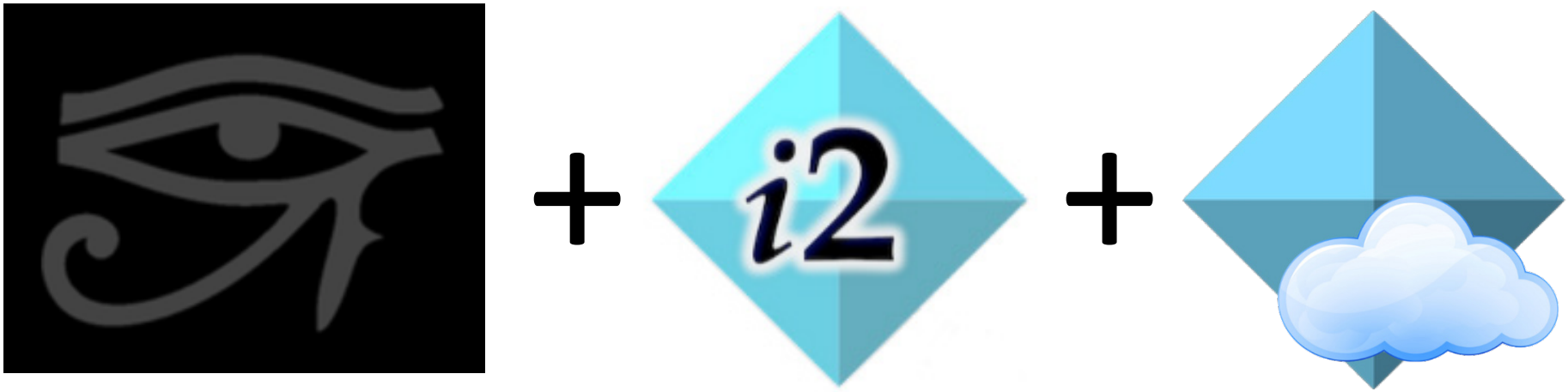
Validation & interpretation

potential ligands



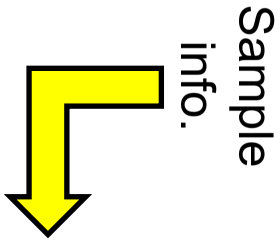


Planning phase:
- estimate feasibility
- cost/benefit analysis
- homology model sufficient?
Target selection
• Experimental data
• Bioinformatics
RCSB PDB
PROTEIN DATA BANK
Deposition
Alter protein:
- choose ortholog
- truncate
Cloning
Figure preparation
Methods
Validation & interpretation
+ "other"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

Taking back control...

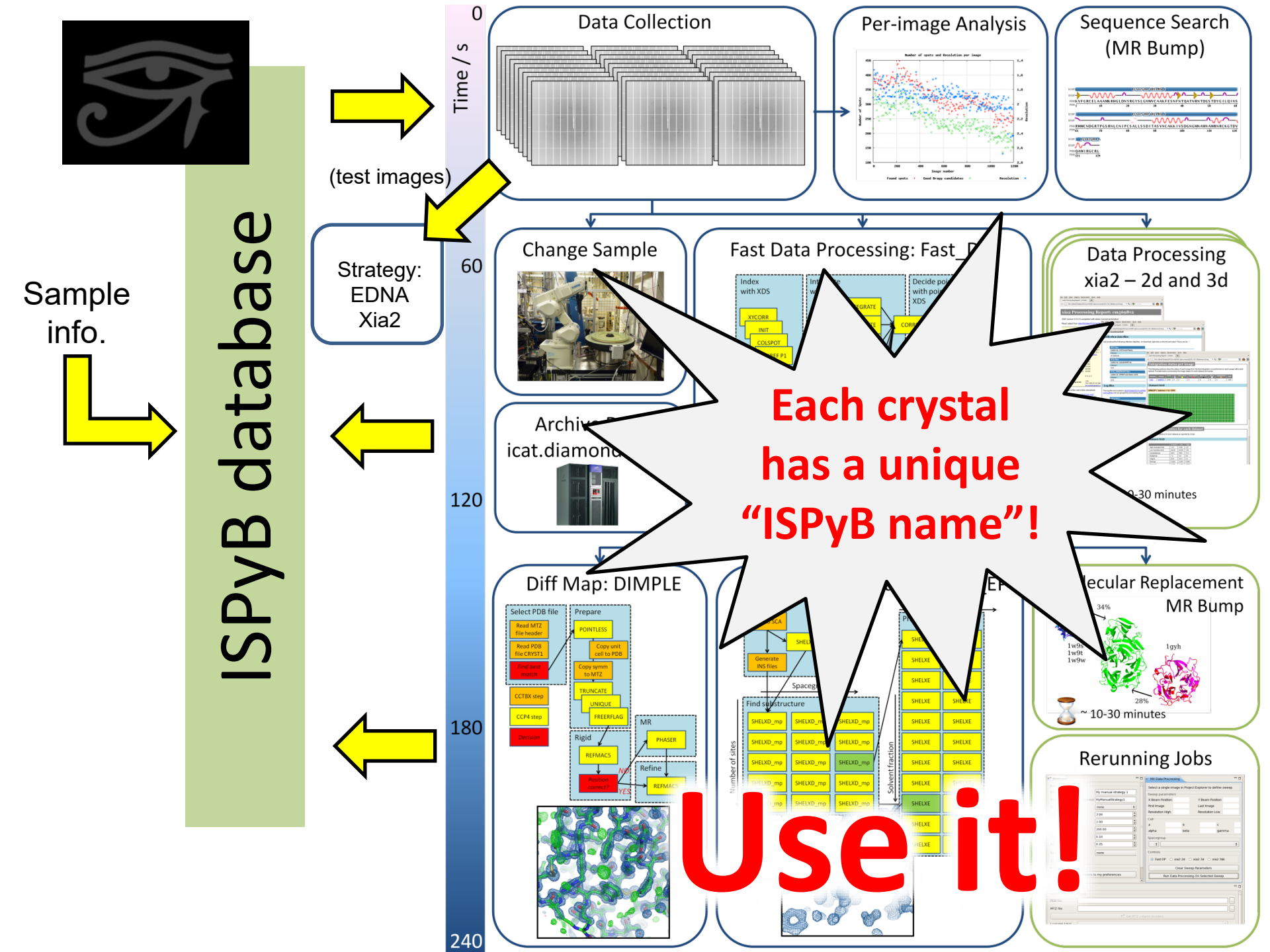


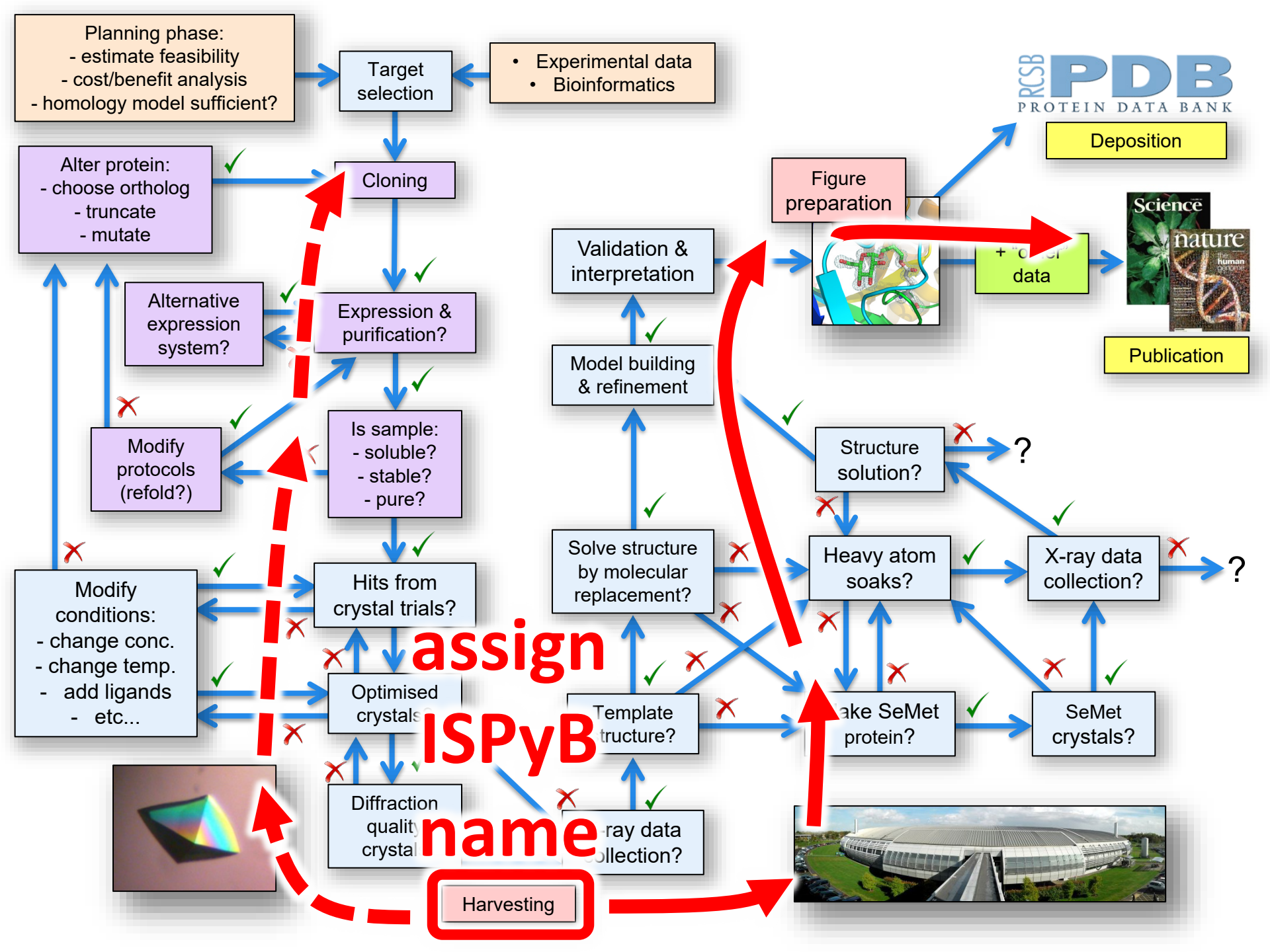

*ISPyB* *CCP4i2* *CCP4 Cloud*

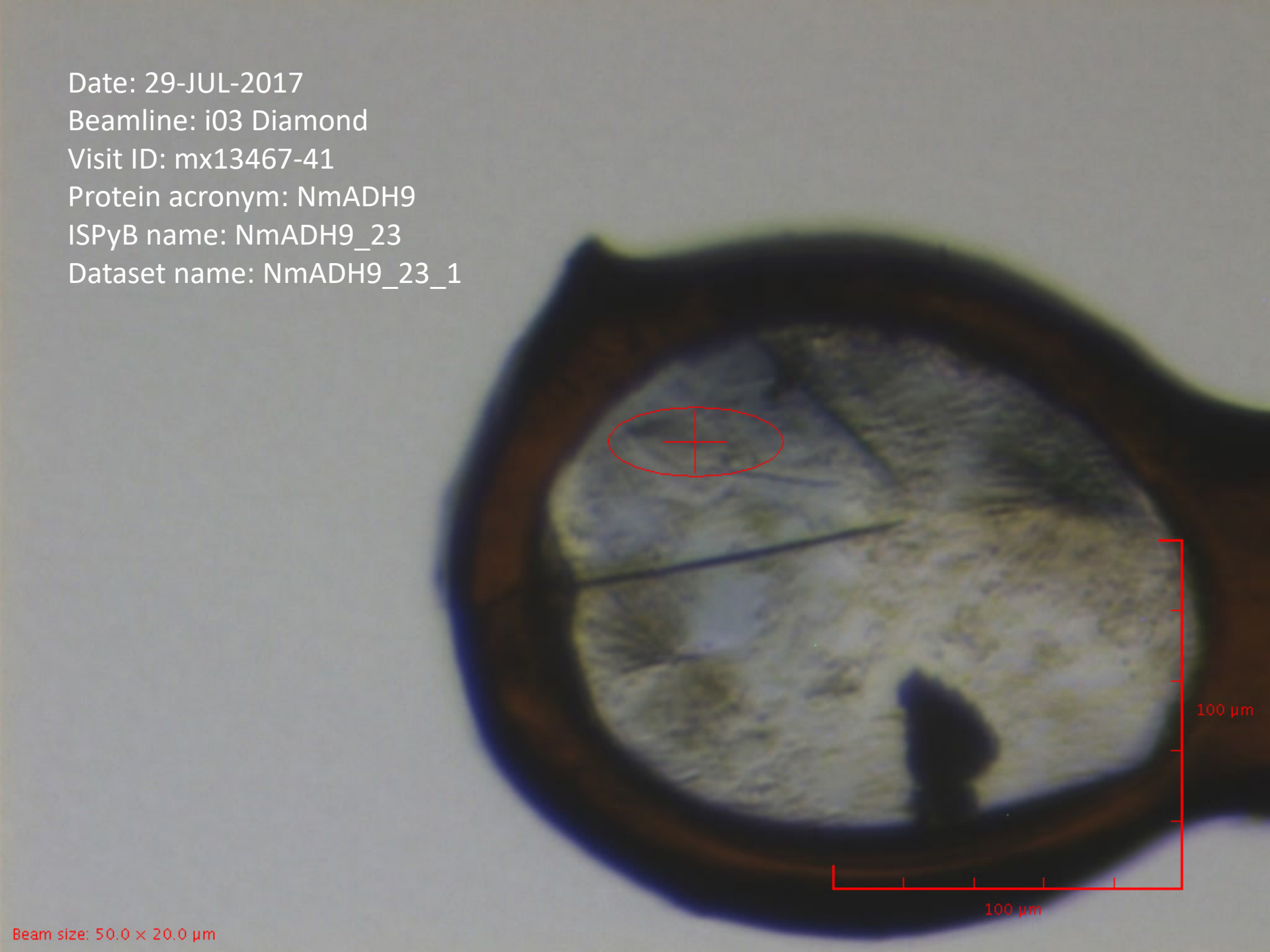
F·R·I·E·N·D·S




# ISPyB database


Time / s
0
60
120
180
240
Data Collection
Per-image Analysis
Sequence Search (MR Bump)
(test images)
Strategy: EDNA Xia2
Sample info.
ISPyB database
Change Sample
Fast Data Processing: Fast_
Data Processing xia2 – 2d and 3d
Archiv
icat.diamond
Each crystal has a unique "ISPyB name"!
Diff Map: DIMPLE
Select PDB file
Prepare
Rigid
MR
Refine
Find substructure
Generate INS files
Spacegr
Number of sites
Solvent fraction
ecular Replacement
MR Bump
~ 10-30 minutes
Rerunning Jobs
Use it!


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


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

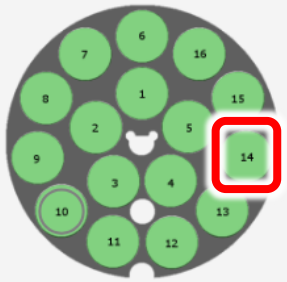
# Harvest your crystals and enter sample info into ISPyB

## Container: DLS-442

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

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

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




Location History

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

10 ▾ Page « ‹ 1 › »

this is **UNIQUE**

+ Extra Fields

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

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

turns on EP pipelines

Required for UDC

**Puck number – DLS442**

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

annotate during data collection – helps decision making

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

First Local Contact: Dr Katherine McAuley (01235 778405)

Beamline I03: 01235 778707

Experimental Hall Coordinator: 01235 778787

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

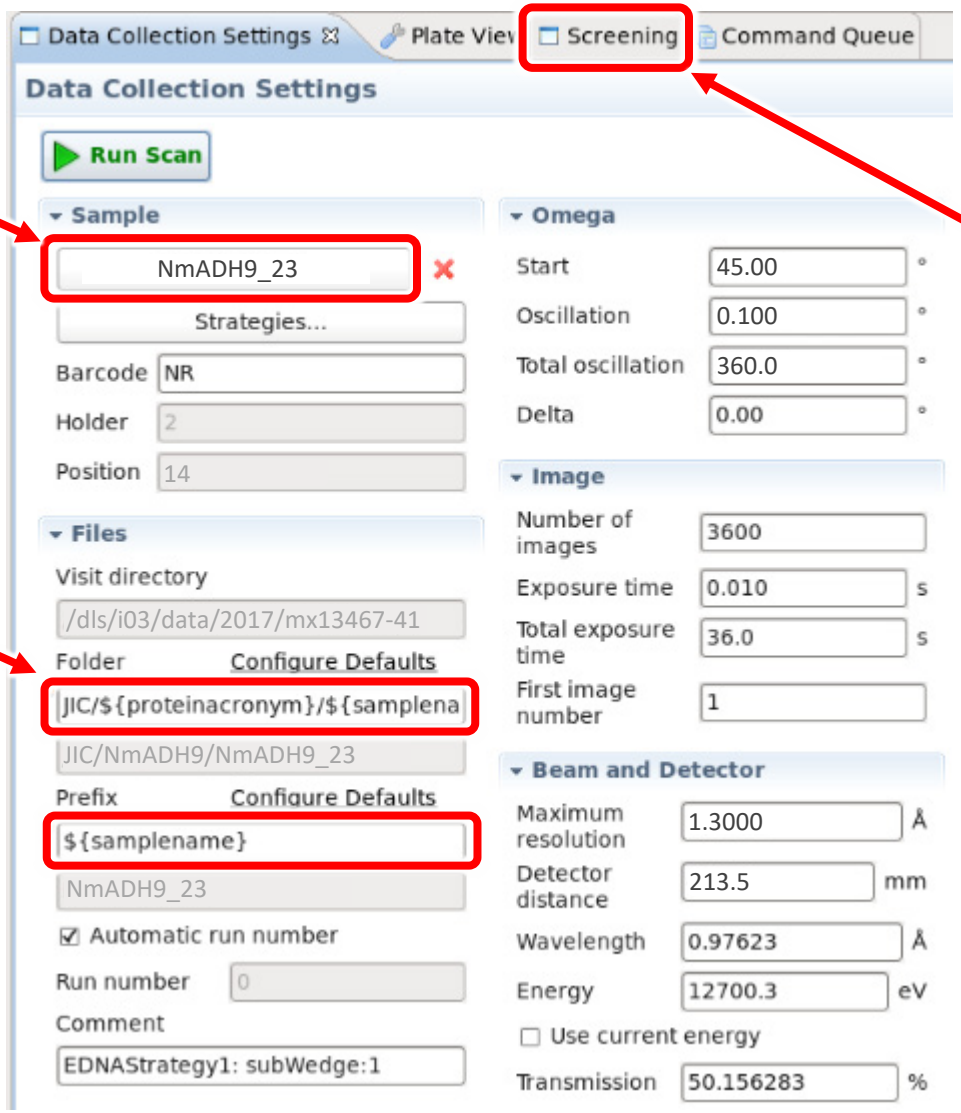
| | | | |
|---|---|---|---|
| Institution | UEA | Group(s) | Hemmings |
| Contact name(s) and number(s) | Andrew Hemmings: 01603 123456/07799 123456 | | |
| Approx. timings | 15:00 – 19:00 | On-site/remote? | Remote |
| Dewar ID | Puck ID (# samples) | Users and comments | |

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

# Data collection setup in GDA

Select required sample from drop-down menu:

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



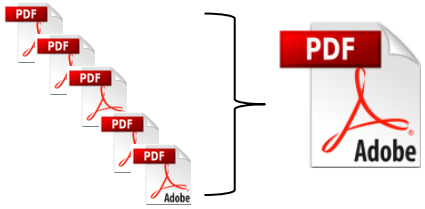

use the "screening" tab for test images – keeps them separate from datasets

Put your data into your own directory....

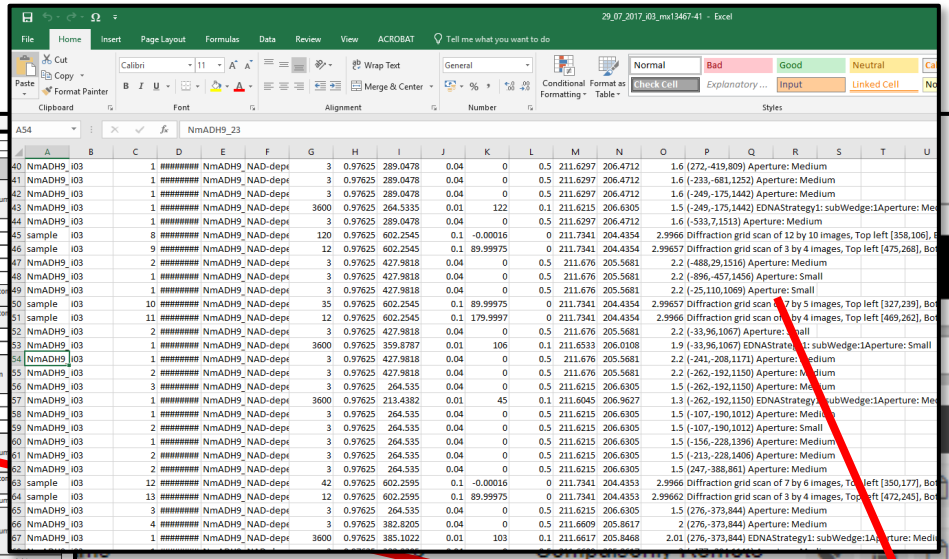
...especially important if there are multiple users from several institutions – simplifies backing up too....

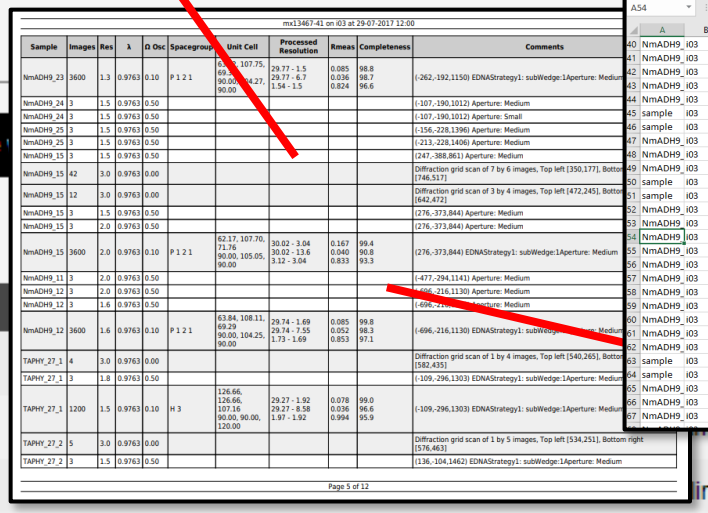
Gives: /dls/{beamline}/data/{year}/{visit-ID}/**{your-name}**/{proteinacronym}/{samplename}/{samplename}_{run_no.}_{image_no.}.cbf

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



simple to search this for an ISPyB name



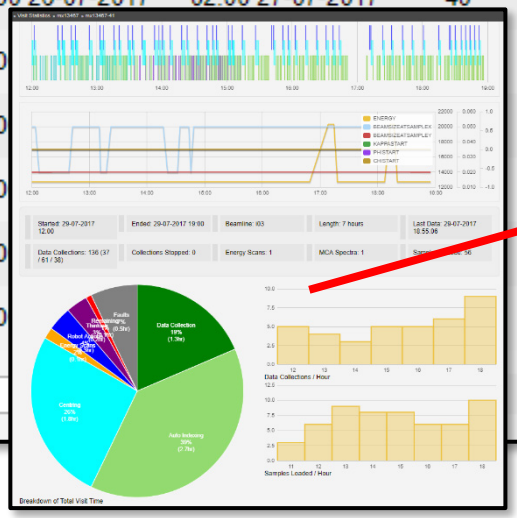


# Visit List

This page lists the

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

10 5 6 7



# 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<br>90.00, 104.27, 90.00 | 29.77 - 1.5<br>29.77 - 6.7<br>1.54 - 1.5 | 0.085<br>0.036<br>0.824 | 98.8<br>98.7<br>96.6 | (-262,-192,1150) EDNAStrategy1: subWedge:1Aperture: Medium |

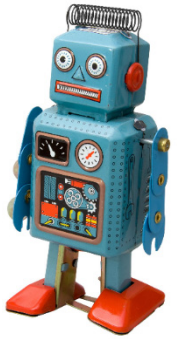
## (3) Visit Excel sheet (annotate...)

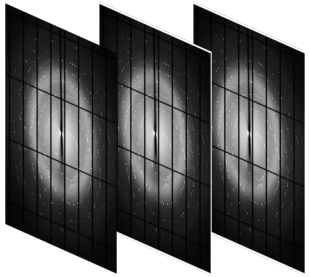
| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | mx13467-41_i03_29-JUL-2017 - 7 hr - Pilatus3 6M detector (100 Hz) | | | | | | | | | | | | | | | | | | |
| 2 | | | | | | | |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

# Ideal scenario during session:

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

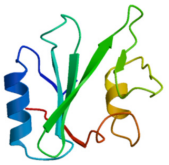
– not practical without automation



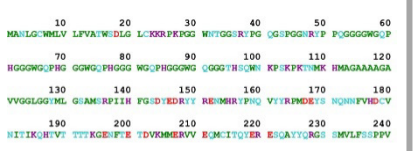


test images

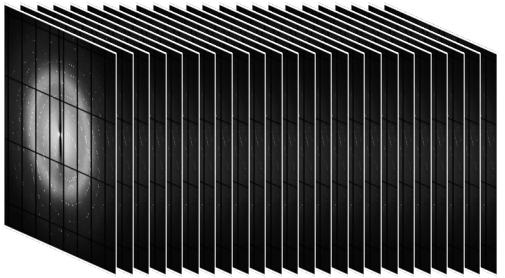
before session...



template



sequence



dataset

< 1 min

Strategies:
Mosflm
EDNA
Xia2

*ISPyB*

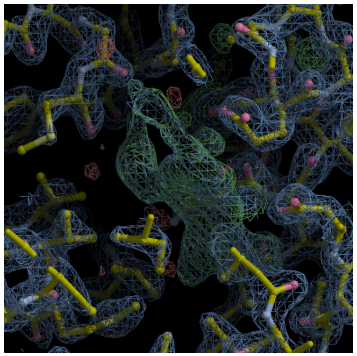
< 2 min

Fast DP

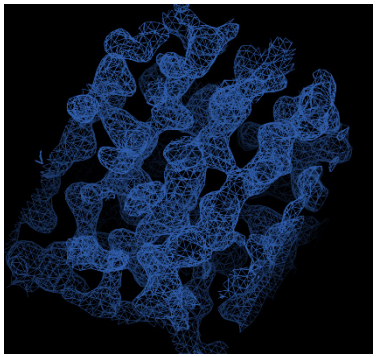
| h | k | l | Iplus | SIGIplus | Iminus | SIGIminus |
|---|---|---|---|---|---|---|
| -45 | 0 | 4 | -1.00 | 1.20 | -1.00 | 1.20 |
| -45 | 0 | 5 | -0.03 | 1.82 | -0.03 | 1.82 |
| -45 | 0 | 6 | 2.17 | 2.01 | 2.17 | 2.01 |
| -45 | 0 | 7 | -0.22 | 1.24 | -0.22 | 1.24 |
| -45 | 0 | 8 | 0.63 | 1.33 | 0.63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

< 5 min

# Quick feedback

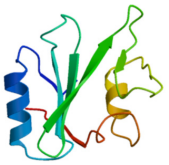


Dimple

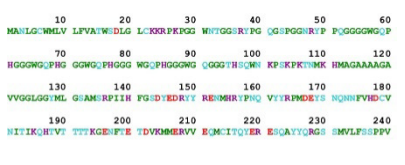


Fast EP

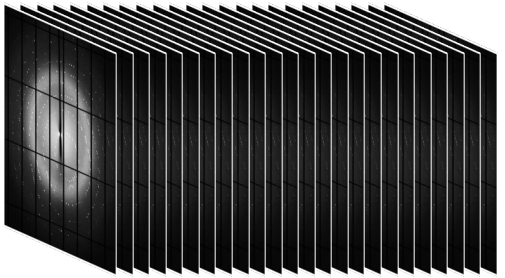
before session...



template



sequence

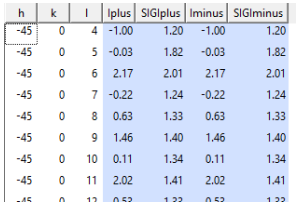


dataset

*ISPyB*

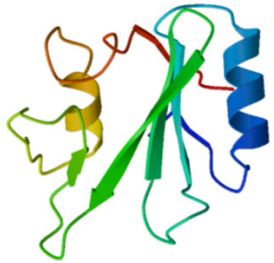
20-30 min

Xia2
DIALS
MultiXia2
autoPROC

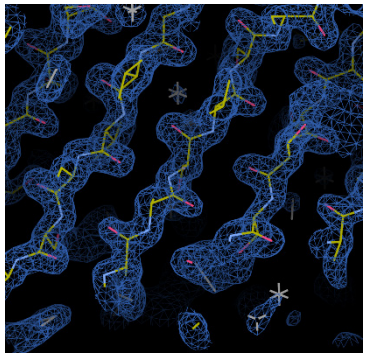


| h | k | l | Iplus | SIGIplus | Iminus | SIGIminus |
|---|---|---|---|---|---|---|
| -45 | 0 | 4 | -1.00 | 1.20 | -1.00 | 1.20 |
| -45 | 0 | 5 | -0.03 | 1.82 | -0.03 | 1.82 |
| -45 | 0 | 6 | 2.17 | 2.01 | 2.17 | 2.01 |
| -45 | 0 | 7 | -0.22 | 1.24 | -0.22 | 1.24 |
| -45 | 0 | 8 | 0.63 | 1.33 | 0.63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

hours



MrBUMP



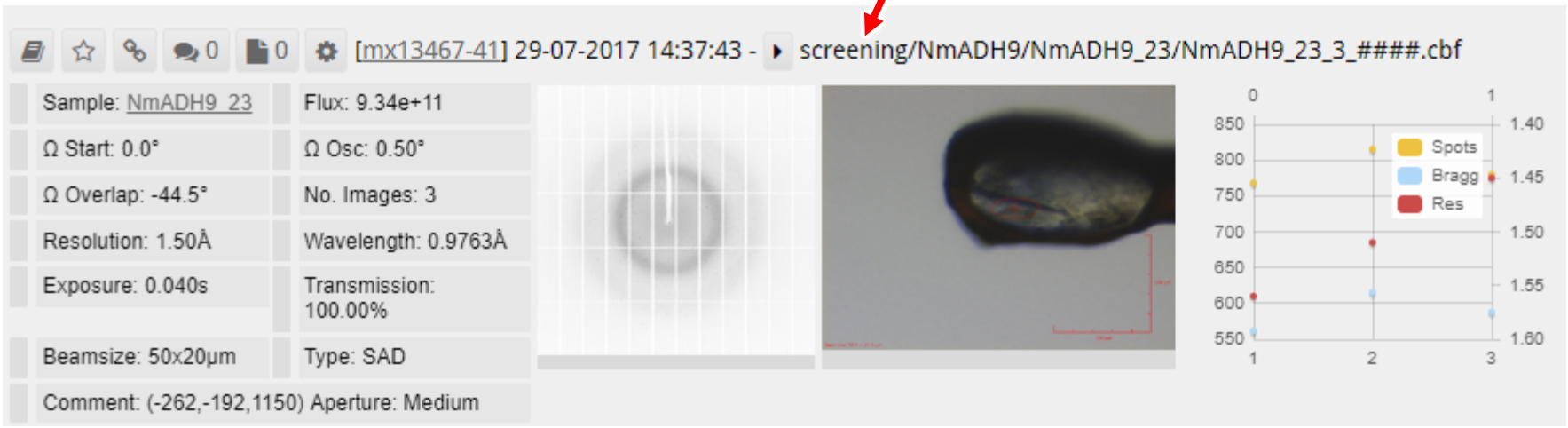
Big EP

# Slower feedback

# More detail on ISPyB output

## Screening…

(using the “screening” tab...)



# Screening…


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

# Screening…

## Strategies

Mosflm: ✔ EDNA: ✔ Xia2: ✔

xia2.strategy

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

Lookup Cell

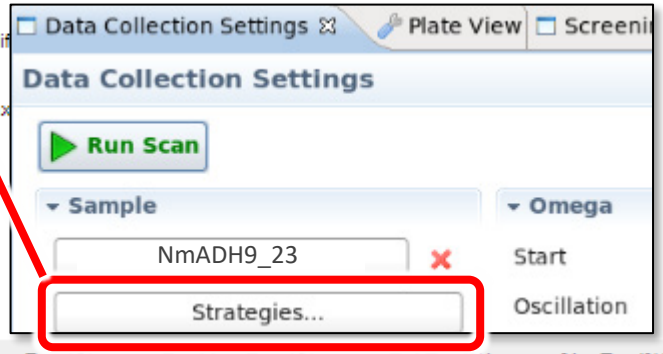
| Strategy | Description | Ω Start | Ω Osc | Res (Å) | Ranking Res (Å) | Rel Trn (%) | Abs Trn (%) | Exposure (s) | No. Images |
|---|---|---|---|---|---|---|---|---|---|
| anomalous Wedge1 | Standard Anomalous Dataset Multiplicity=3 I/sig=2 202s | 0 | 0.15 | 1.51 | 1.17 | 4.0 | 4 | 0.010 | 2214 |
|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

EDNA MXv1

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

Lookup Cell

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




mosflm

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

Lookup Cell

| Strategy | Description | Ω Start | Ω Osc | Res (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 |

# Collecting a dataset…

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

Lookup Cell

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

Data Collection Settings | Plate View | Screening | Command Queue

**Data Collection Settings**

Run Scan

**Sample**

NmADH9_23

Strategies...

Barcode: NR

Holder: 2

Position: 14

**Files**

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

Folder — Configure Defaults: JIC/${proteinacronym}/${samplena

JIC/NmADH9/NmADH9_23

Prefix — Configure Defaults: ${samplename}

NmADH9_23

☑ Automatic run number

Run number: 0

Comment: EDNAStrategy1: subWedge:1

**Omega**

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

**Image**

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

**Beam and Detector**

| | | |
|---|---|---|
| Maximum resolution | 1.3000 | Å |
| Detector distance | 213.5 | mm |
| Wavelength | 0.97623 | Å |
| Energy | 12700.3 | eV |
| ☐ Use current energy | | |
| Transmission | 50.156283 | % |

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 | 0.036 | 0.824 | 98.8 | 63.92 107.75 69.36 90.00 104.27 90.00 | |
| xia2 3d | 28.52 - 1.37 | P 1 21 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 21 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 21 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: http://www.globalphasing.com/autoproc/) | 107.76 - 1.50 | P 1 21 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: http://www.globalphasing.com/autoproc/)

| Beam Centre | X | Y |
|---|---|---|
| Start | 211.60 | 206.96 |
| Refined | 211.50 | 206.92 |
| Δ | 0.10 | 0.04 |

Plots | Archive | Logs & Files | Lookup Cell

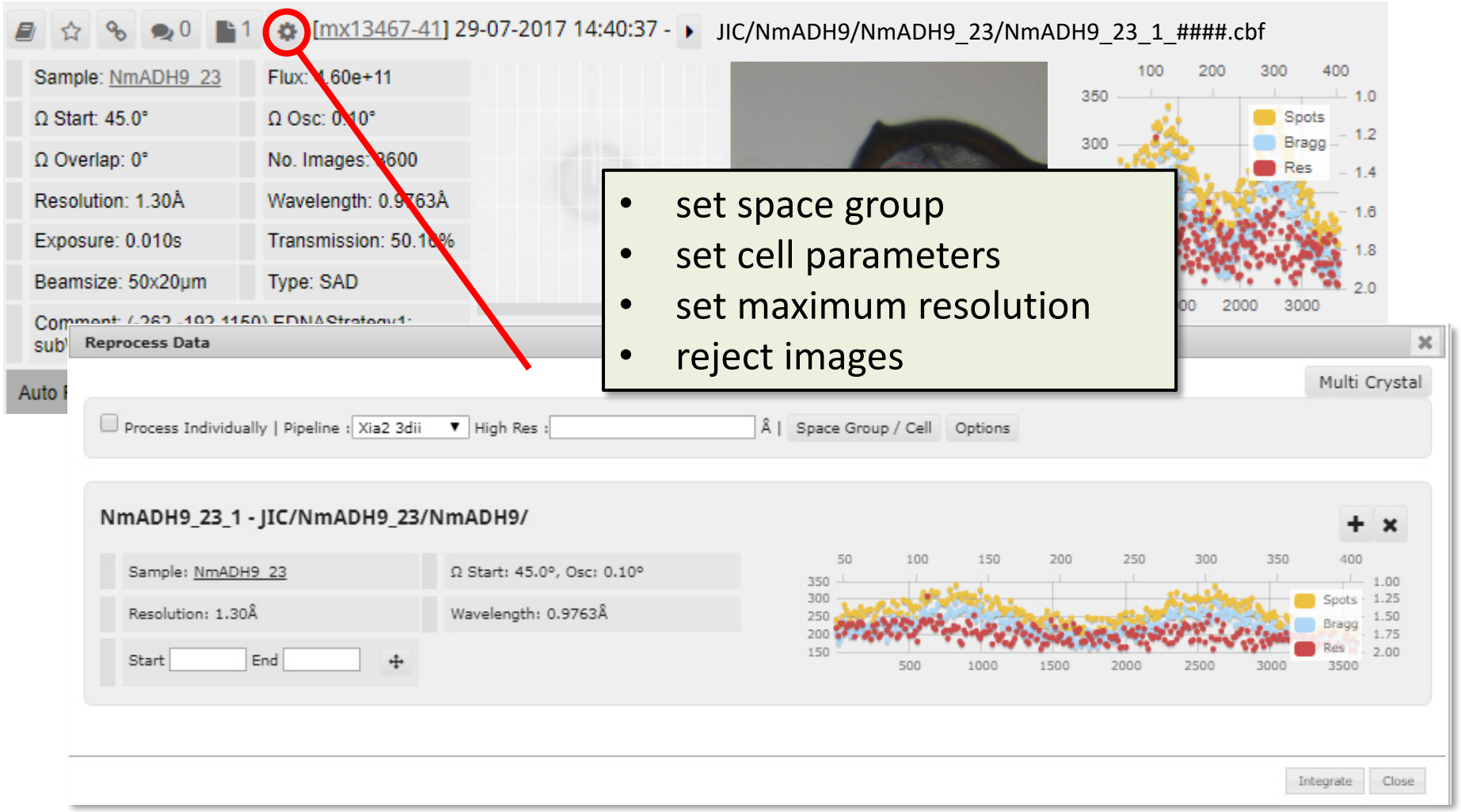
| Space Group | A | B | C | α | β | γ |
|---|---|---|---|---|---|---|
| P 1 21 1 | 63.92 | 107.75 | 69.36 | 90.00 | 104.27 | 90.00 |

| 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

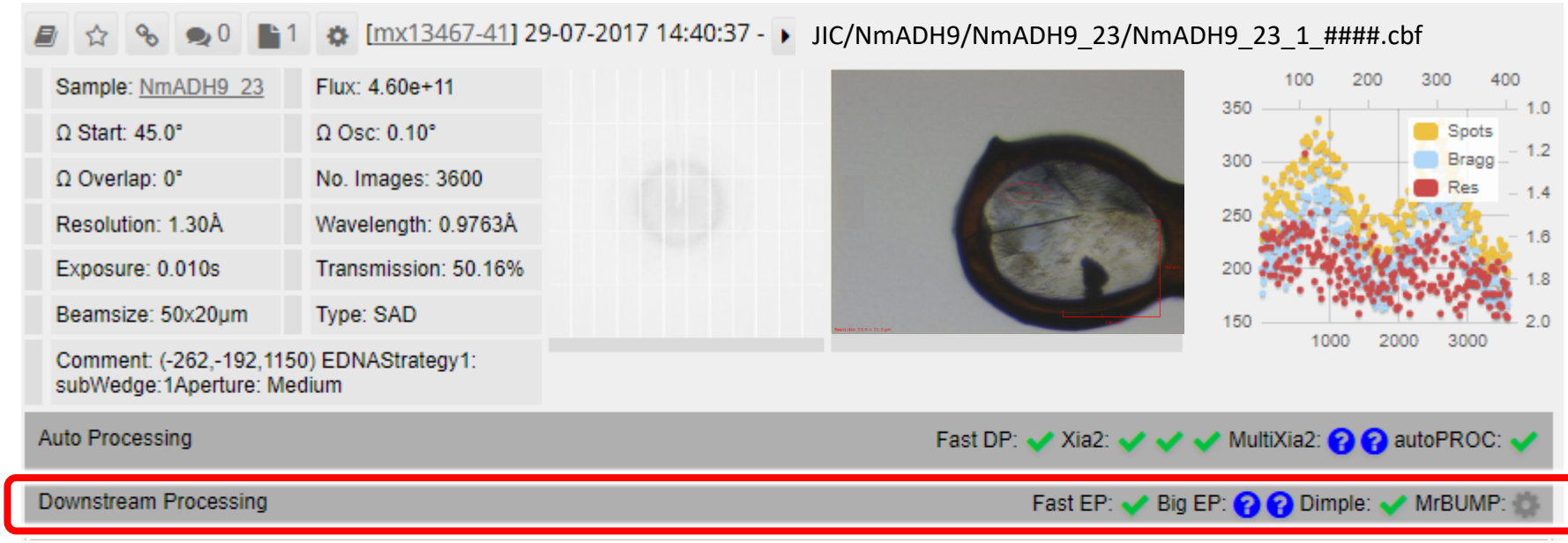
# Reprocessing:

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

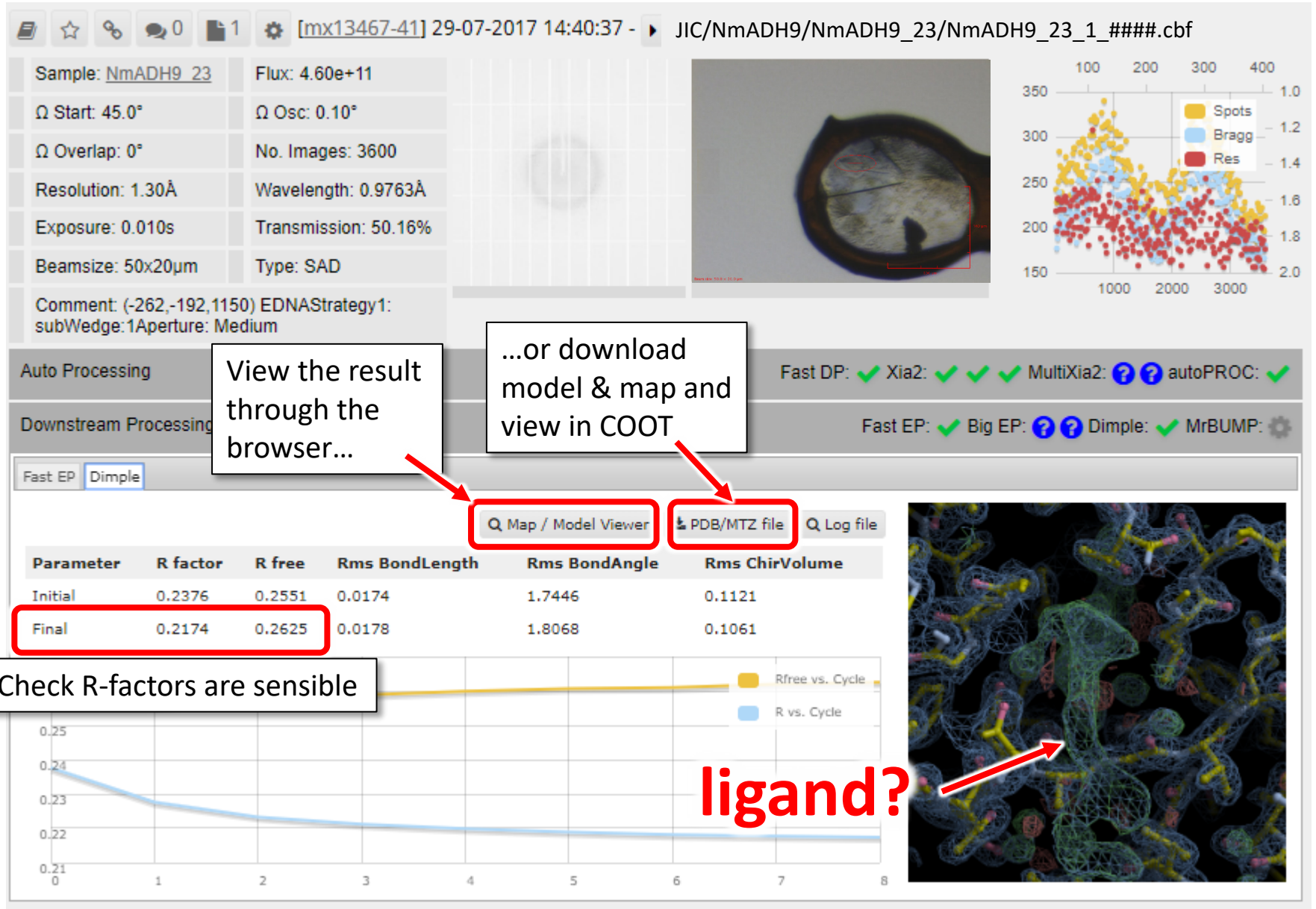



- set space group
- set cell parameters
- set maximum resolution
- reject images

# 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)**
– deleted after 40 days
– still available through TopCAT (tape archive)


**"Meta" data** (all the other "stuff")
– deleted after 40 days
– not backed up


>90% of useful datasets start 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)

Slow...

globus

Faster?

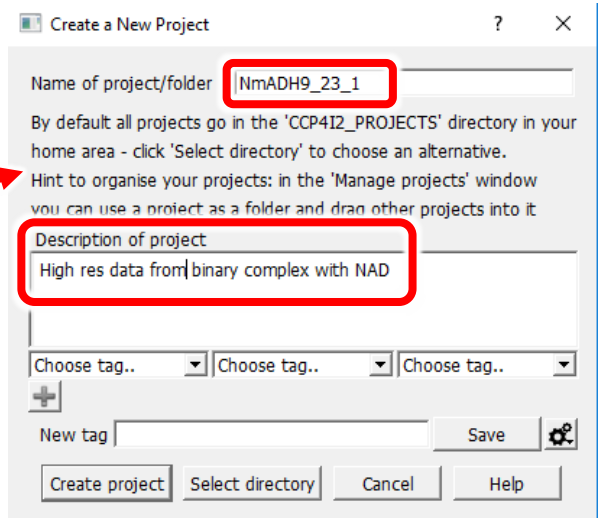
In the meantime:

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

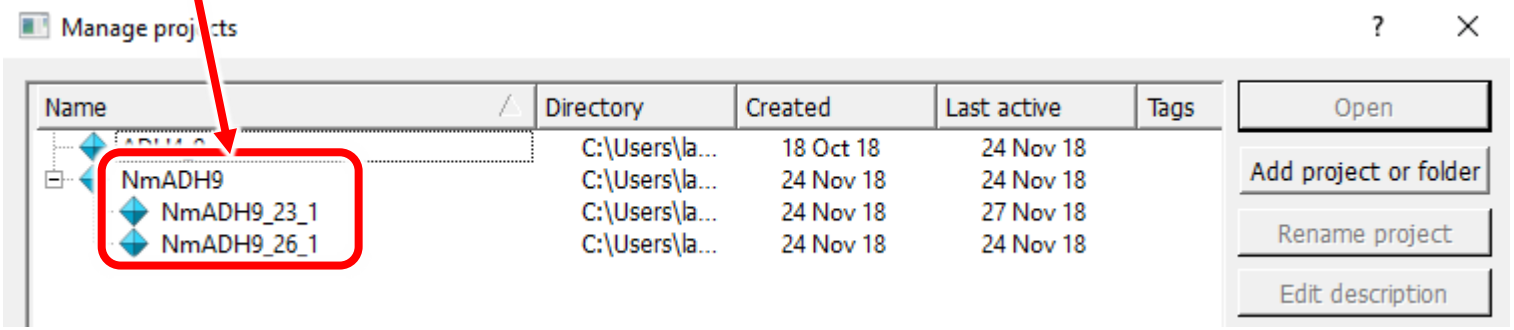
# Retain ISPyB name in CCP4i2...

CCP4i2

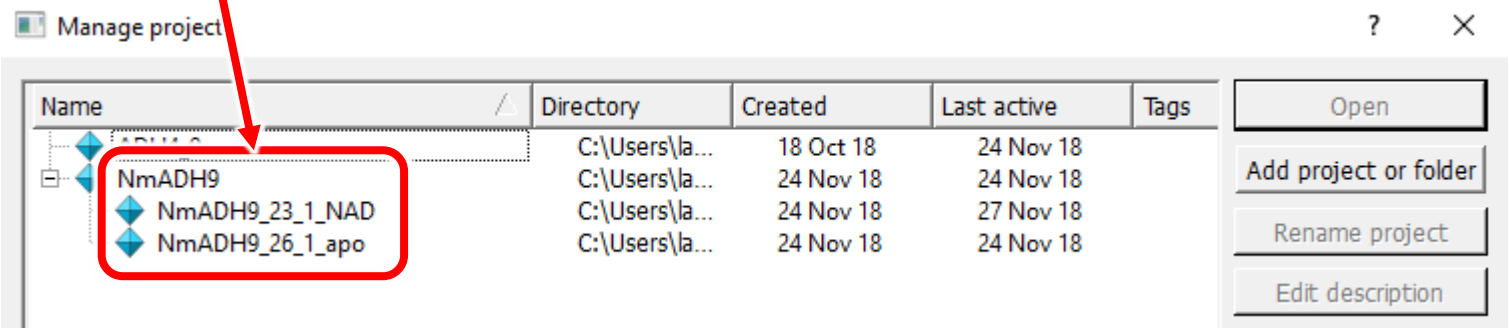
Create a separate "project" for each dataset




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



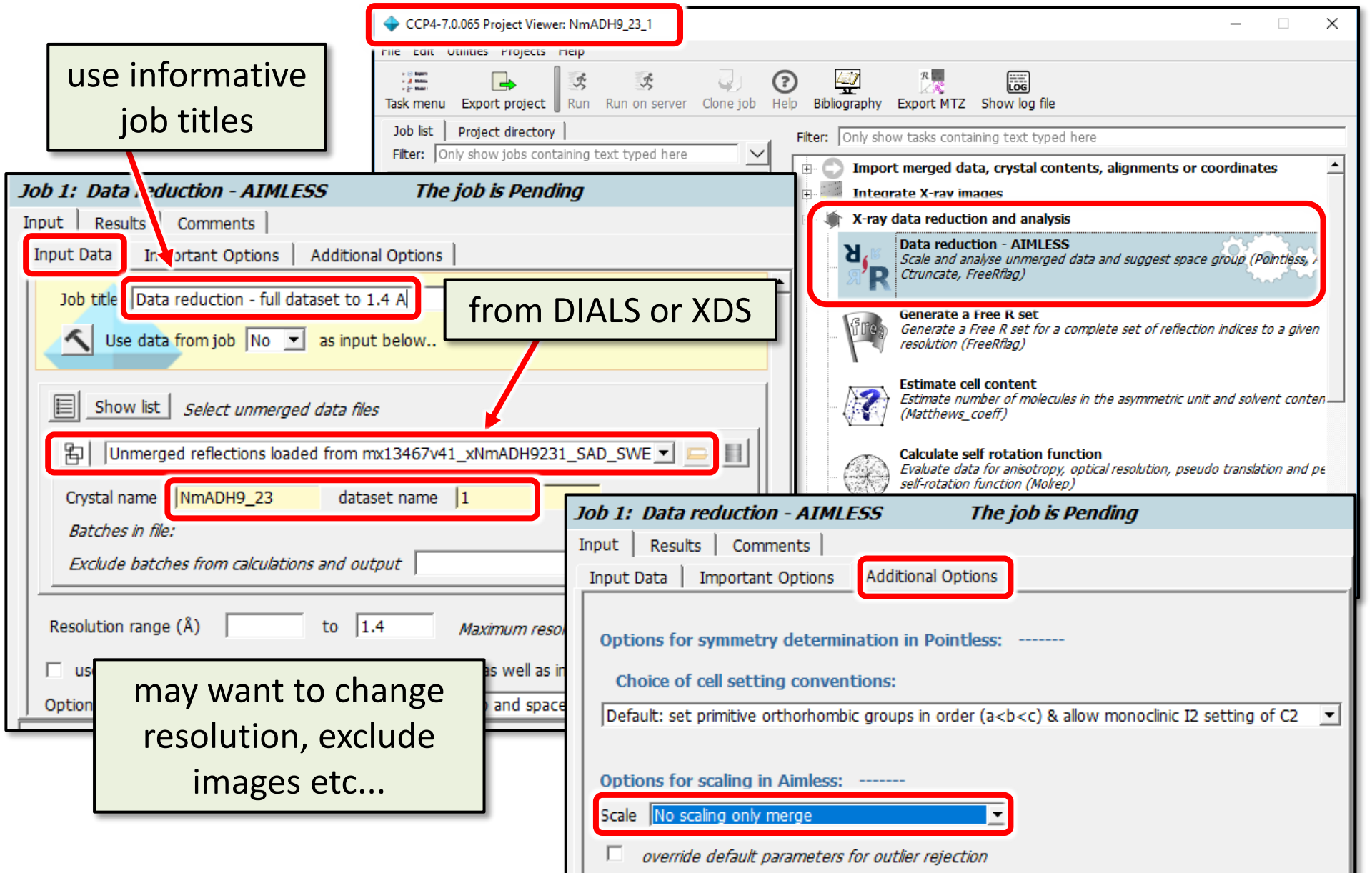

Can rename "projects" later... (when you know more...)*




*but directory names remain the same

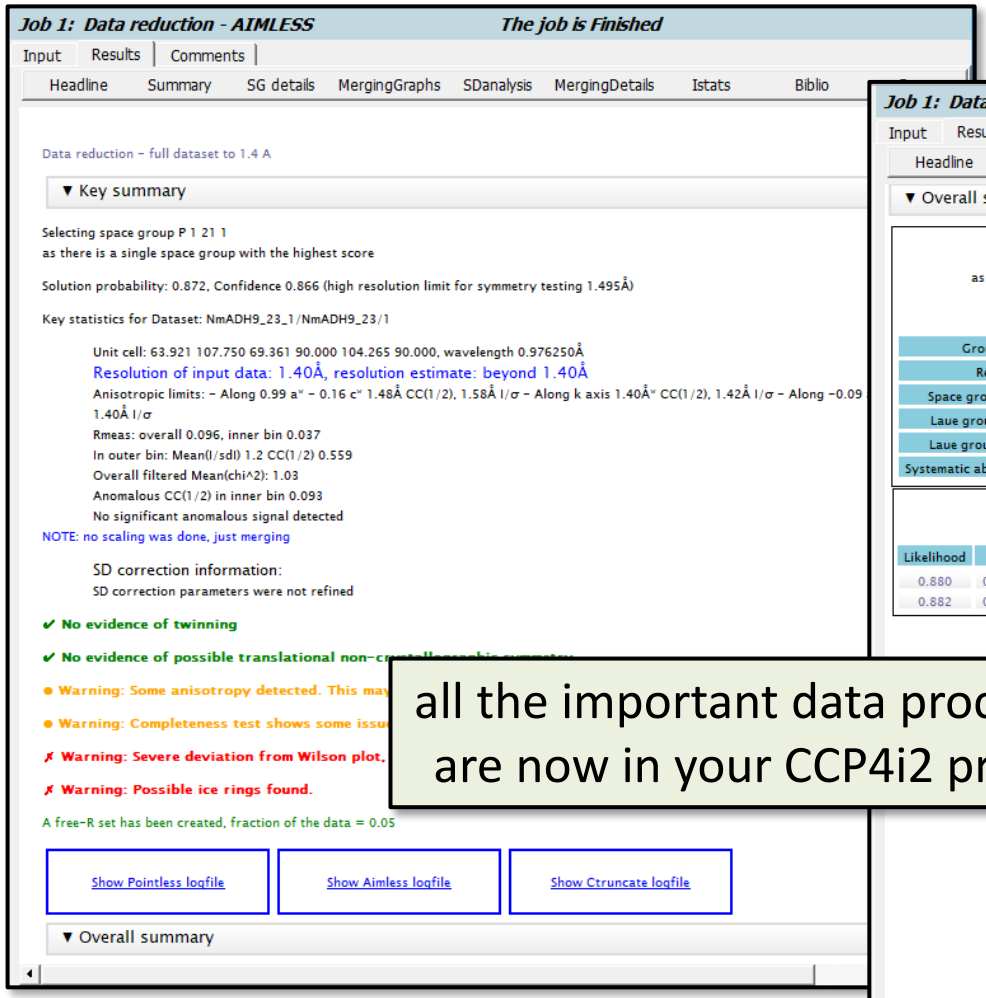
My preferred option for processing...

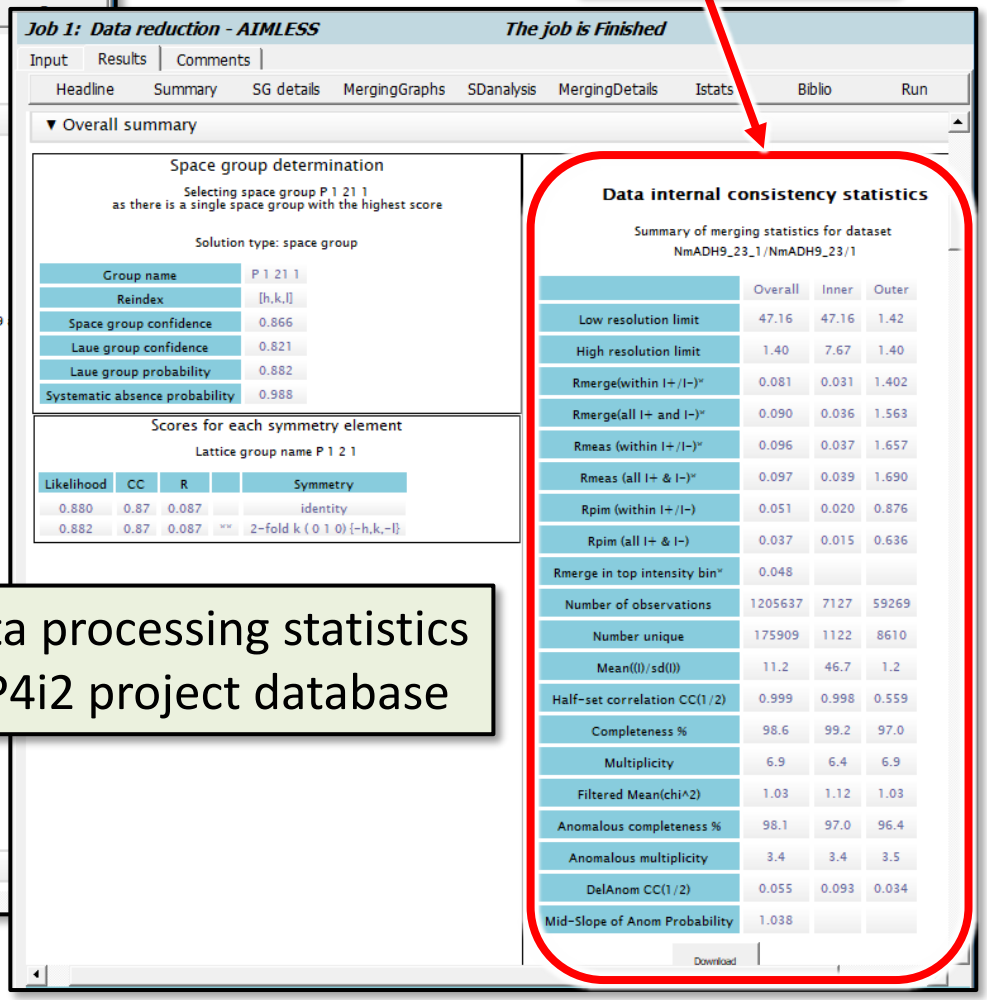
- where the pipelines have done a good job – take part-processed dataset (from the "Meta" data) and re-run the merging step in CCP4i2



# Gives lots of output – useful for troubleshooting...

for your paper/thesis




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

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

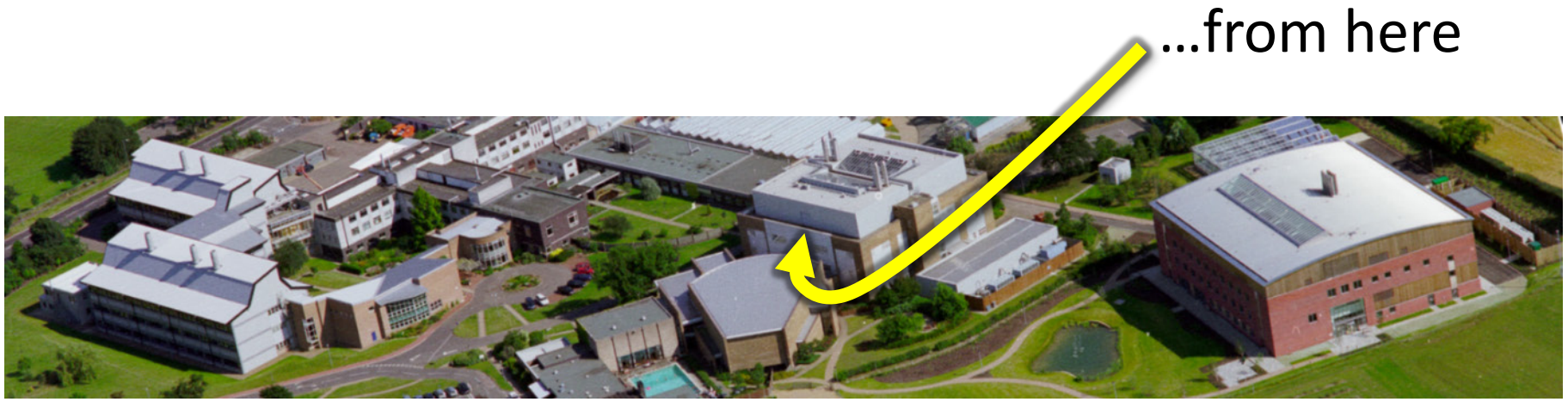
**Integrate images with Dials**
*Launch DUI and capture output*

see DIALS talk...

# 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

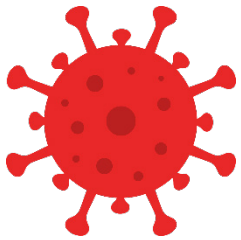


*it's 2020…*
*remote data collection?*
*it's a no-brainer!*

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

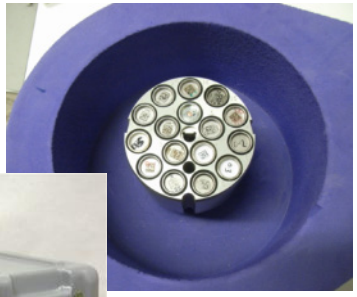
# Remote data collection saves you time…

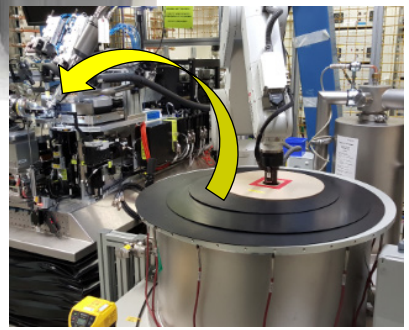
…and the planet



…and is covid-secure

# For routine data collection at 100 K:

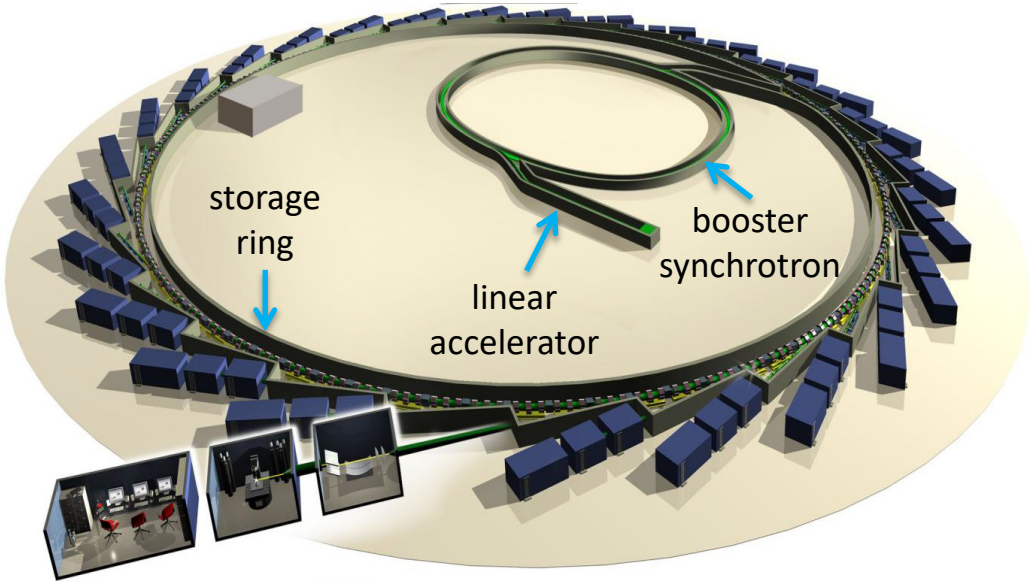


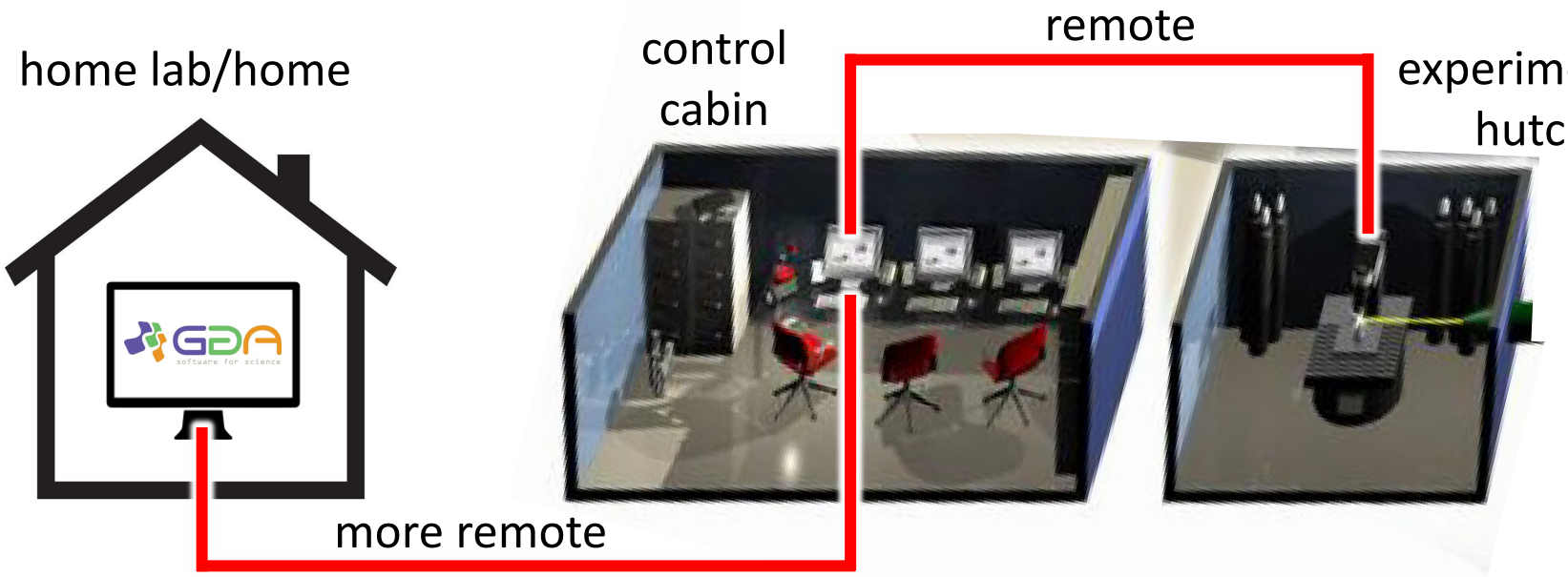


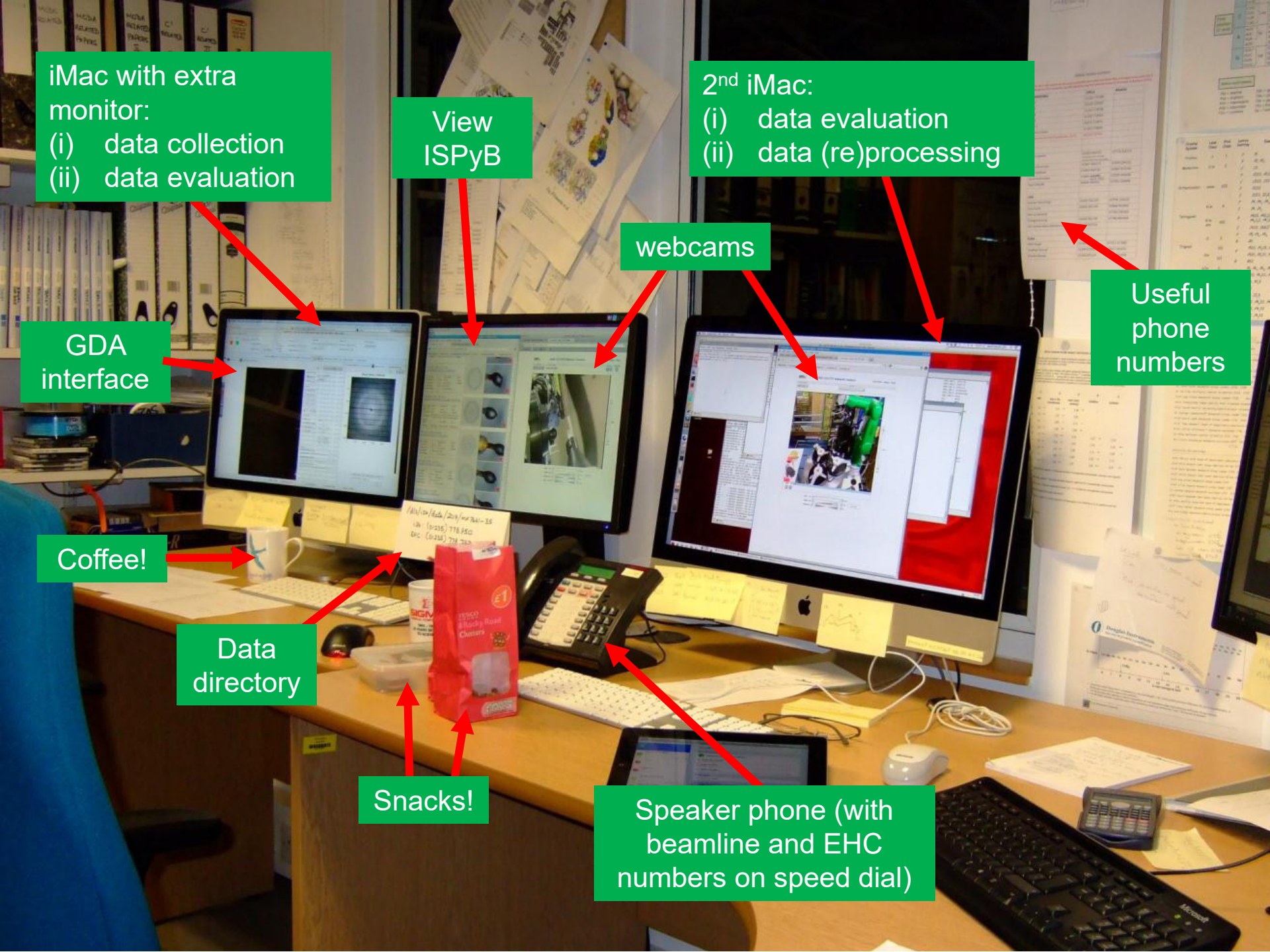
- Samples prepared and cryo-cooled in home lab
- Sample information entered into ISPyB database
- Transported to Diamond in dry shipping dewars
- Data collection controlled through GDA interface
- Samples mounted robotically
- Only manual operation at DLS: loading/unloading pucks
- Everything else computer driven…

Therefore you don’t need to be there!

# What is remote data collection?




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

# Remote data collection

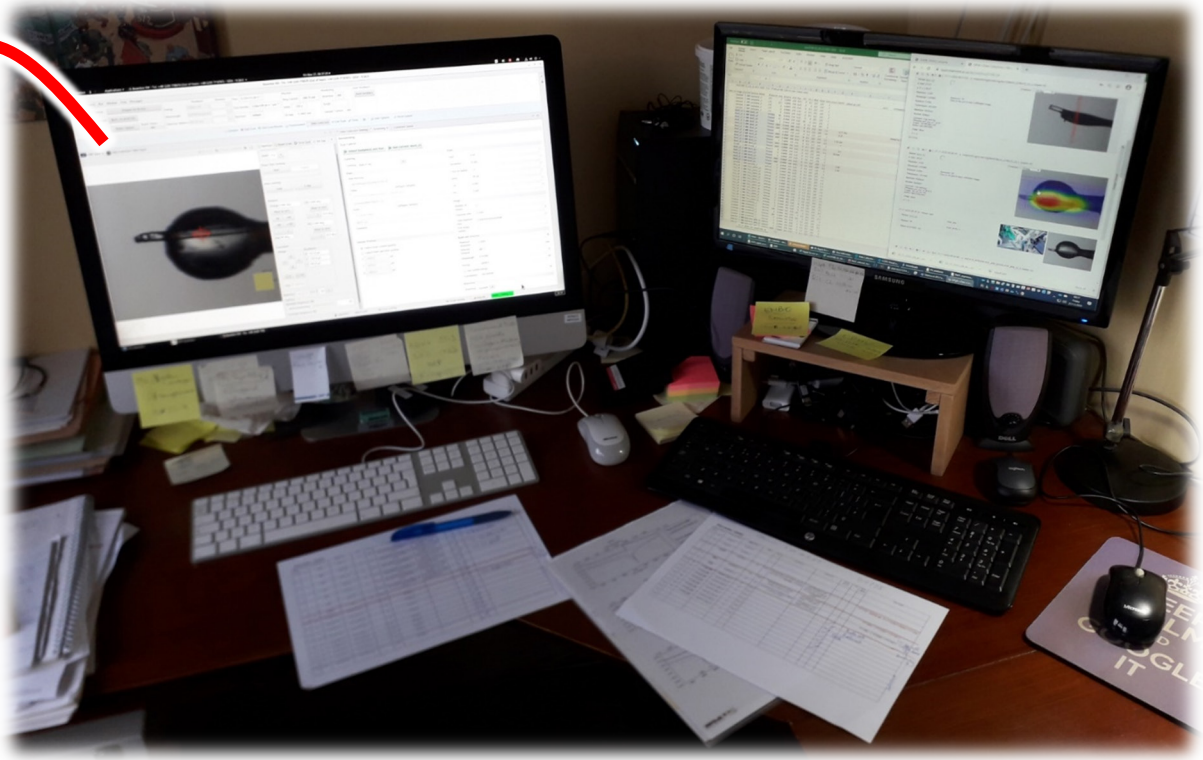
…in 2020

…from here

# Remote data collection 2020 style!

iMac connected to beamline computer via NX client

PC connected to work PC via VPN



share screen with others

# When things go wrong…

Can **you** fix things?

- probably not!

On-site user:

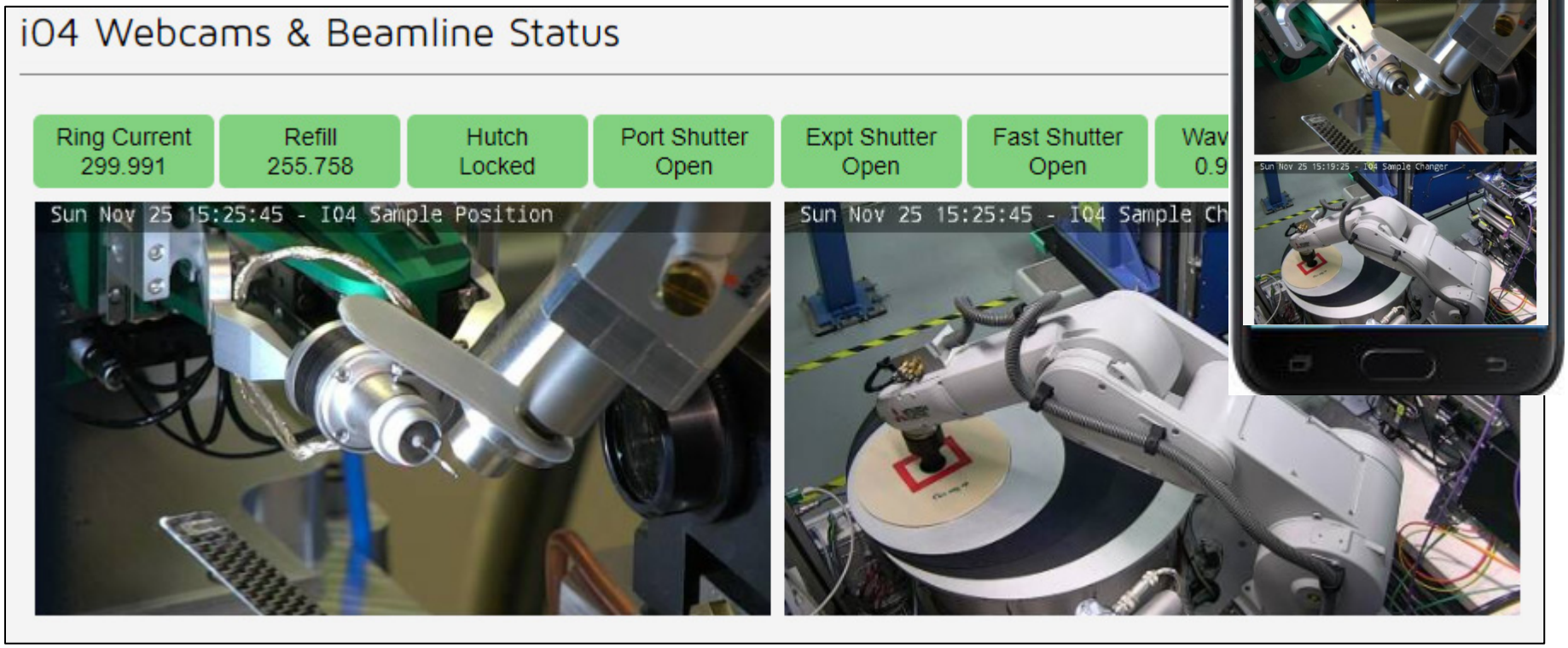
- During normal working hours – call local contact
- Out of hours – call EHC

Remote user:

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

Therefore you don’t need to be there!

# Check the webcams!







If you suspect a problem – call the EHC!

# If you see no diffraction – 3 main causes:


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


(2) you are doing something wrong/stupid

(3) your samp[illegible]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:

- Users collect data on their own crystals
- Users stay at home labs
  - 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

# 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
- use remote data collection for routine experiments
- think before and during data collection
- this is 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


John Innes Centre

UKRI Biotechnology and Biological Sciences Research Council