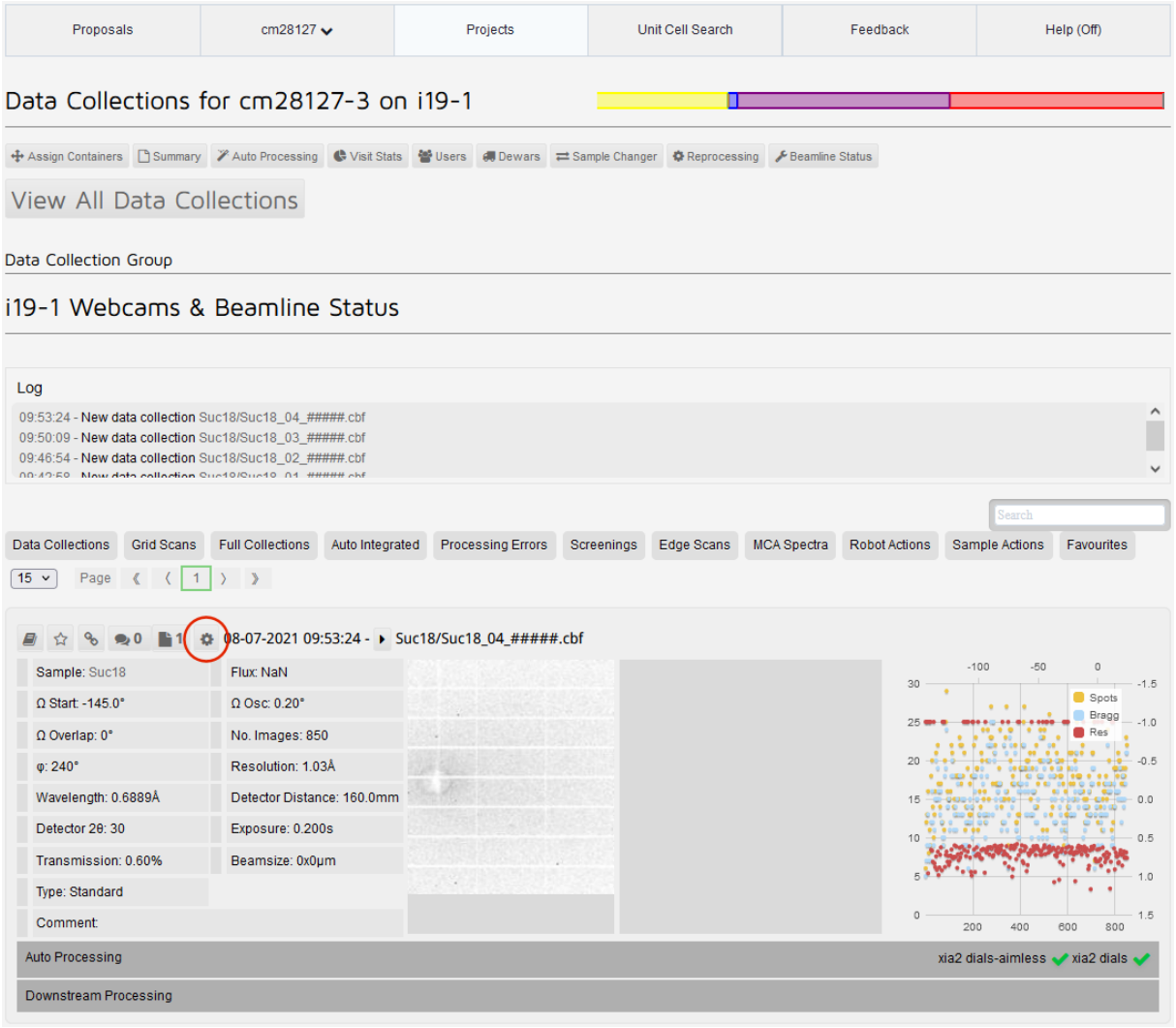
Reprocessing

ISPyB

It is possible to reprocess the data with xia2 from within ISPyB with some (but not all) additional parameters. These jobs are run in addition to the auto processing.

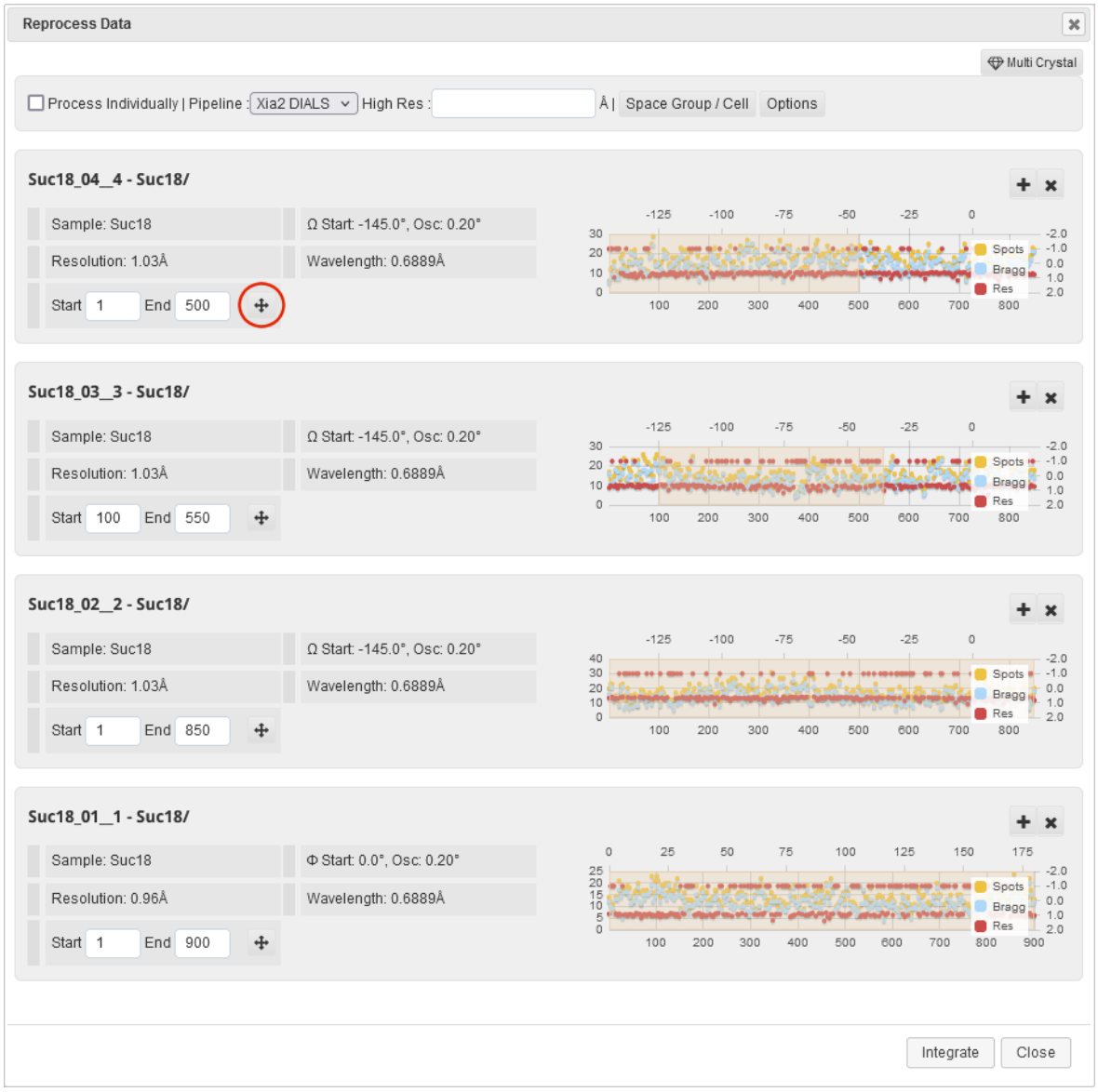
From the Data Collections page, first click on the entry next to Group to access all sweeps for that data collection, and then click on the cog for the first sweep you want to include, to open the reprocessing window.



Now click on the cog wheels for any other sweeps of data to be included – these will be added to the list in the Reprocess Data window.

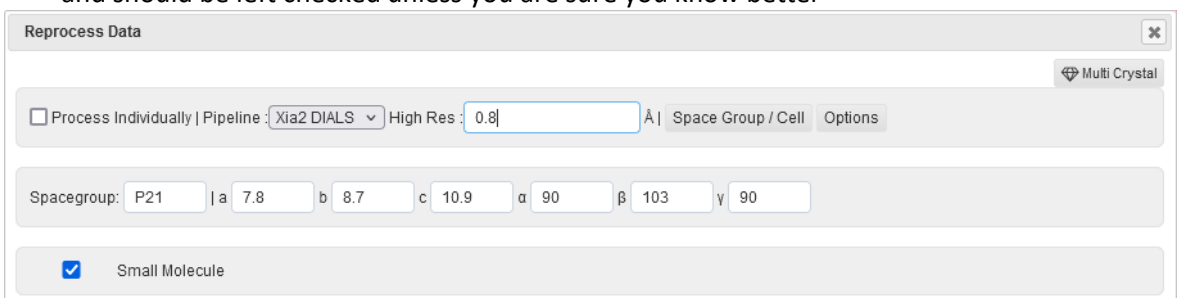
By opening the sweeps in reverse order, e.g., 4,3,2,1, the results will be included with the results for the run at the top of the list, i.e., run 4 which allows for easier comparison with the auto processing results (for all data collected).

It is necessary to select which images are to be included from each sweep – either highlight (from left to right) on the plot or type the image numbers in the boxes. Clicking on the "**+**" button will autofill with the entire sweep.



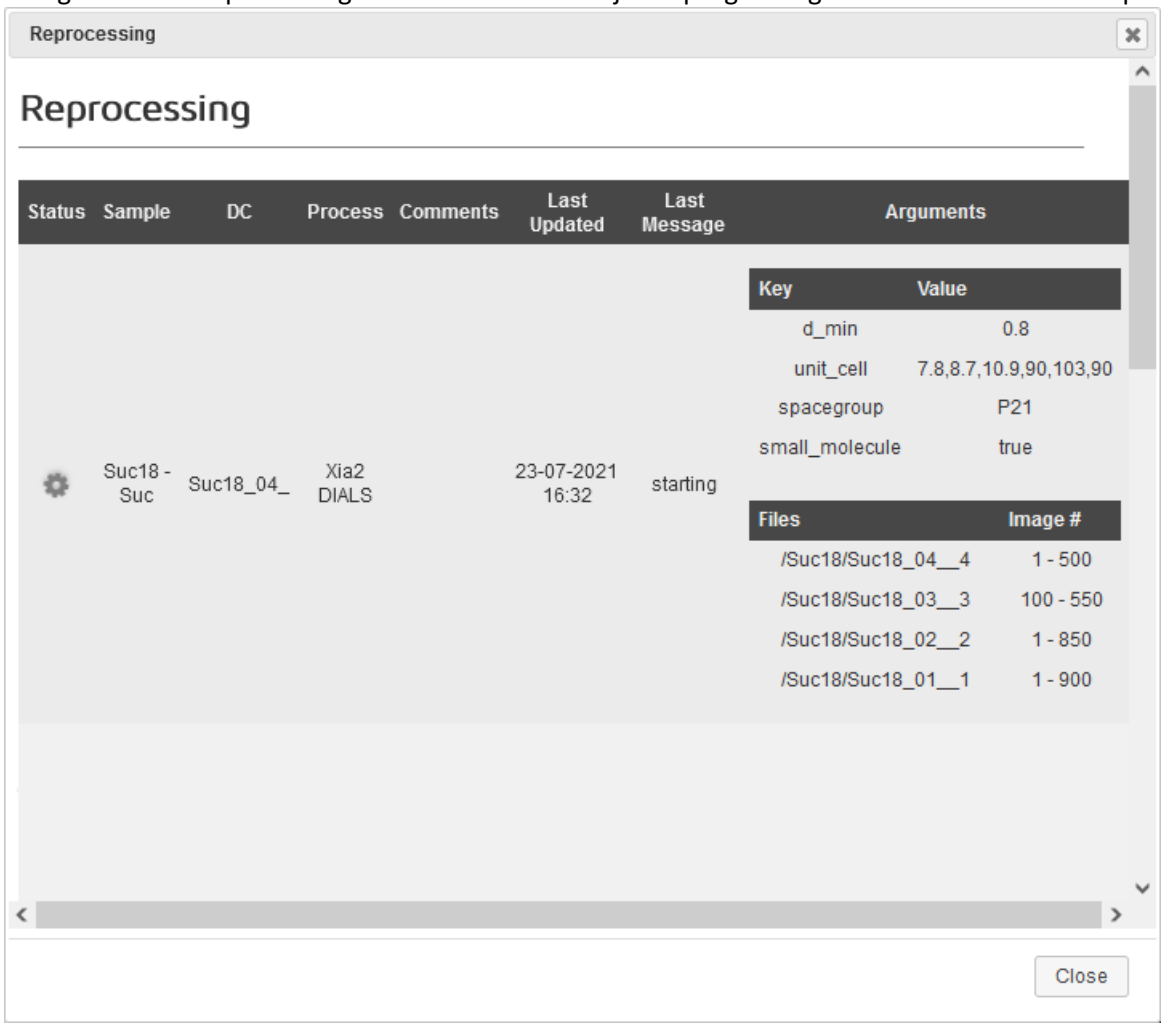
Notes:

- Process Individually should be left unchecked
- Xia2-dials is the best pipeline but if it isn't working, the others can be tried
- The high resolution limit can be set in the High Res box, probably based on inspection of the auto processing results (see xia2.txt and merging-statistics.txt)
- Clicking on the "Space Group/Cell" button opens the option to input the known or expected information
- The "Options" button has one check box for small molecule – this is checked by default for I19 data and should be left checked unless you are sure you know better



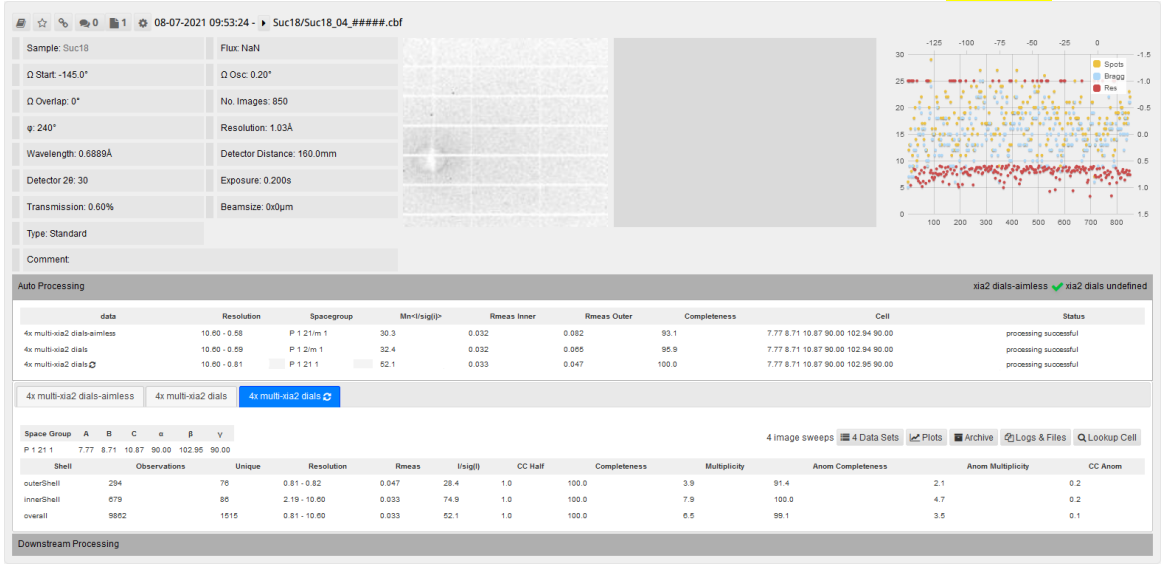
At the bottom of the Reprocess Data window click on the "Integrate" button and a red message box should pop-up to show the job has been successfully submitted.

Navigate to the Reprocessing window to check the job is progressing and see when it has completed.



The Reprocessing window is accessed from the line of Option buttons and shows a list of current and completed reprocessing jobs with any input commands set, i.e., resolution, unit cell information and which images were used. Click on the relevant line (the top line has the results for all included sweeps) under Files to be taken to the results of the processing job.

This will appear as another result in the Auto Processing section for that data collection. The  indicates that this is a manually started reprocessing job rather than an auto processing job. At this point, all functions for the reprocessed job match those of the auto processing (see ???)



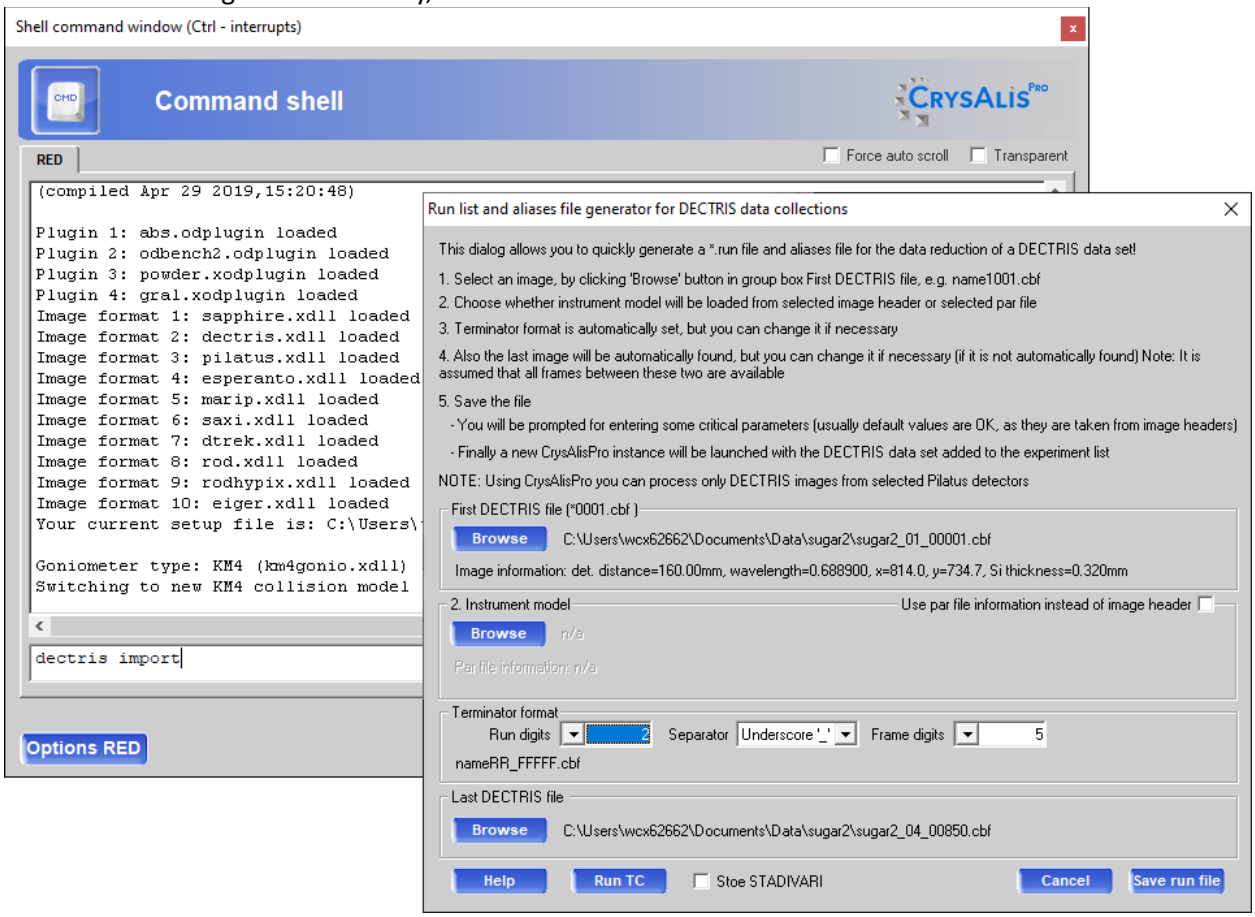
CrysAlis Pro

Download the raw images (a link to this information was sent out by the user office)

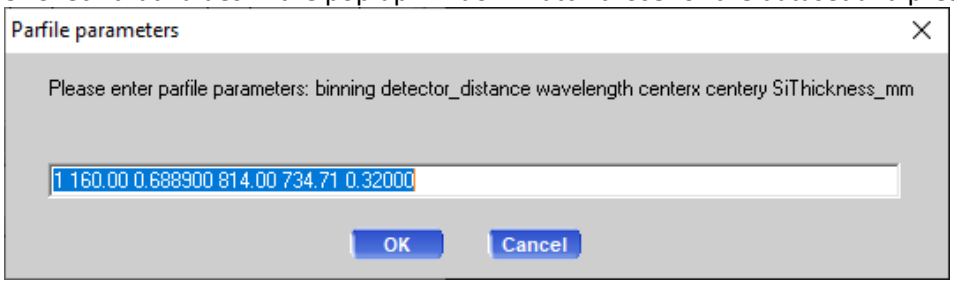
1. Open CrysAlisPro (version 38.41 or later) and any random project from the list

2. Click on the  button, type dectris import in the box and press *enter*

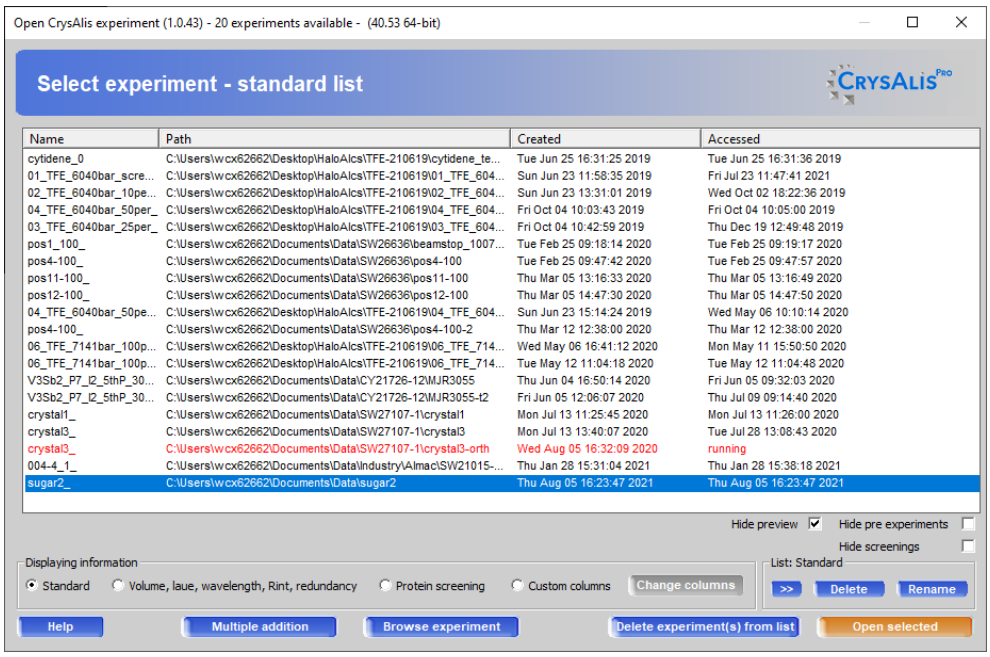
In the window that pops up, browse to the dataset and select the first image.
Then set the file format: run digits should be 2, the separator is an underscore, and frame digits should be 5.
Check the last image is set correctly, then click on “Save run file”.



3. Check that values in the pop up window match those for the dataset and press OK

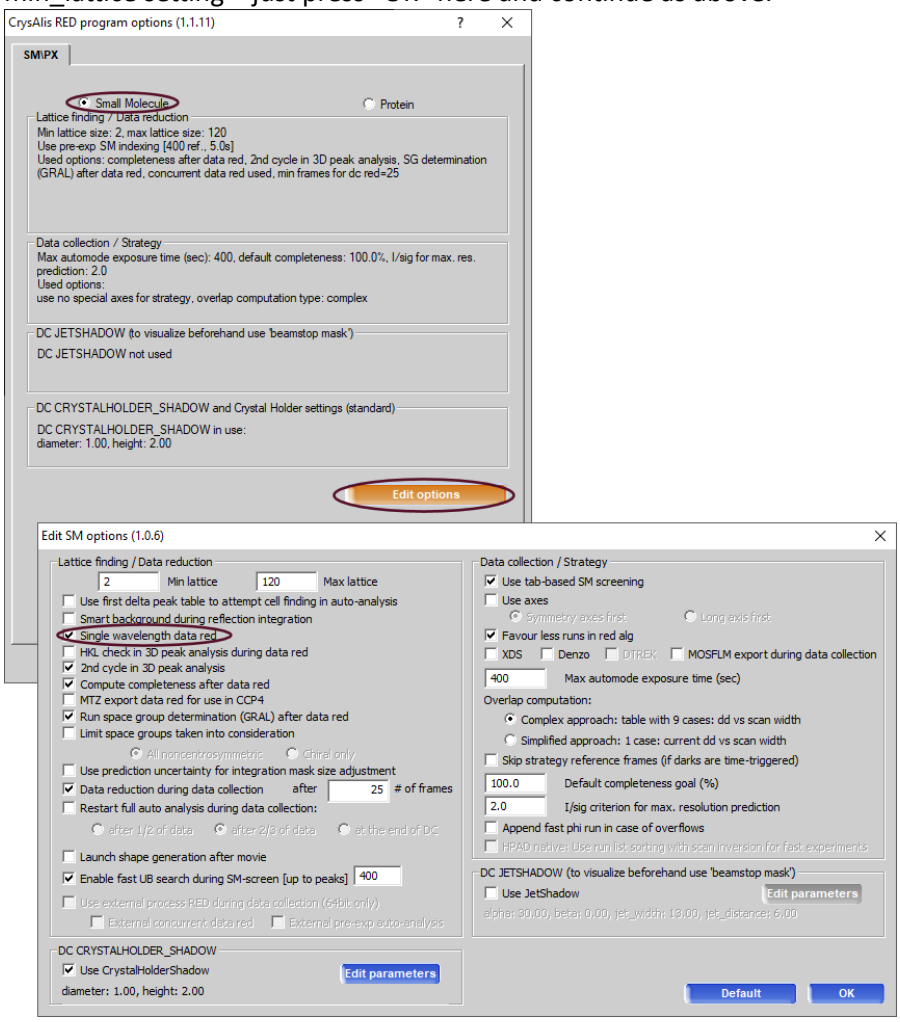


4. Once the program has finished setting up the run file, select this experiment (it should be highlighted at the bottom of the list) and click on “Open selected” to open the dataset in a new instance of CrysAlis.



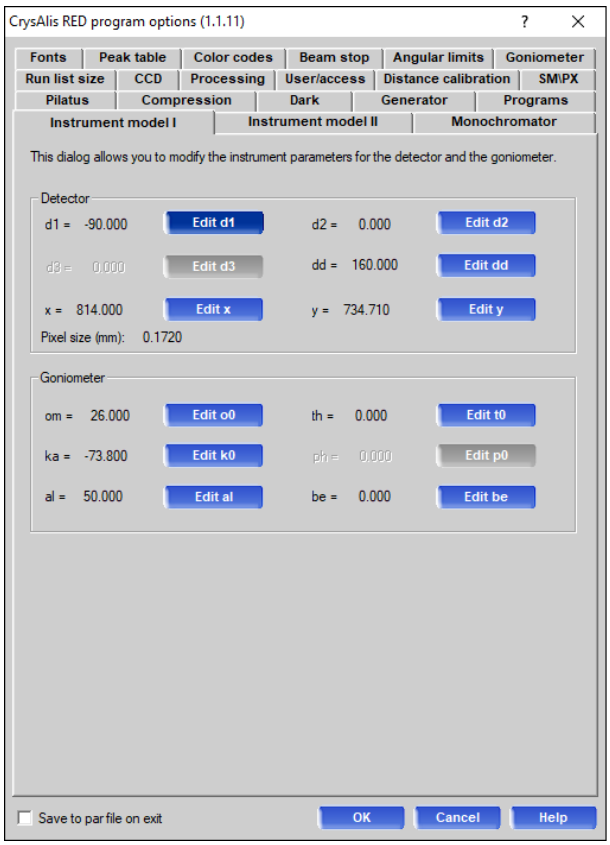
5. Click OK in the pop-up error box to set the basic settings

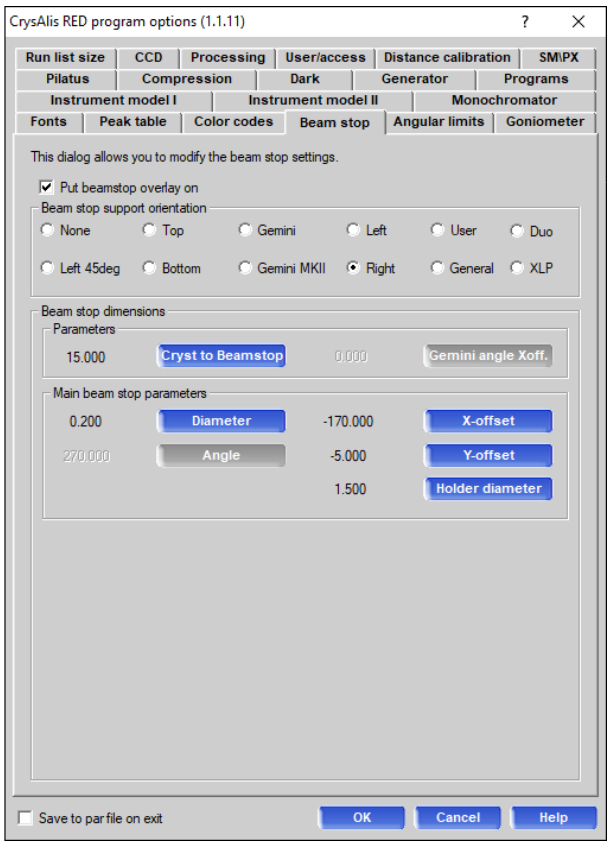
6. In the window that opens up, make sure Small Molecule is checked and then click on “Edit options”. Check Single wavelength data red in the new window.
Click “OK” in both windows
Note: For newer versions (42.49 onwards) there is an additional pop-up window which queries the min_lattice setting – just press “OK” here and continue as above.



7. Click on  and then “Options RED” in the bottom left of this window and edit the values in the Instrument Model I and Beamstop tabs.

	Older versions	More recent versions
Instrument Model I	**Beamstop**	**Beamstop**
d1 = -90	support orientation = right	support orientation = leave as default
om = 26	diameter = 0.2	diameter = 0.5
ka = -73.8	x-offset = -170	x-offset = -170
	y-offset = -5	y-offset = -2
		holder diameter=0.5





Click “OK” in this window and then close the CMD window.

8. Click forward through an image using  along the bottom row and you should see the changes take effect.
Edit the beamstop parameters as required.

Use CrysAlis Pro as normal.

Notes:

- Do make sure you edit the instrument model correctly, all 3 values need updating, otherwise you won’t get nice lattices.
- The general advice is to work up your best-looking dataset first and refine the instrument model parameters on that. Then use this par file (with this instrument model) on the less good datasets.
- NOTE: It may be necessary to use CrysAlis Pro index the data, but better processing results are often obtained by using this information and reprocessing with DIALS.

DIALS

For more information about running DIALS, check the website: https://dials.github.io/index.html

Reprocessing jobs using xia2 / DIALS can be submitted directly using the command line.
This is a more manual procedure, and a suitable directory structure needs to be created by hand.

The first step is to navigate to the *processing* directory within the visit folder as this is the only place with write access.

Open a terminal window and, if not already done, type module load i19 and press *enter*.
This should direct you to your current visit directory, e.g.: /dls/i19-1/data/2024/cm37266-1 as well as enable DIALS to work.

Now type: cd processing

Create a suitably named directory for the processing files to be written in, e.g.: mkdir SampleA
And then enter the directory: cd SampleA

Use of DIALS to run processing step-by-step is not covered here, but there is an example of how to do it here: https://dials.github.io/documentation/tutorials/small_molecule_tutorial.html

Useful things to know:
The up arrow
- Press this once to redo the command above
- Repeated pressing will scroll through previous commands (to save quite a lot of tedious typing)

TAB key
- If enough information is already included, TAB will autocomplete instructions (saving tedious typing)
- Double TAB will give the options available from whatever is input

Copy and Paste must be done via the right mouse button
Everything is case sensitive

xia2

The easiest way to invoke a DIALS processing job is to run xia2 as all of the steps are run sequentially automatically.

The basic xia2 command is:
xia2 small_moleule=true /dls/i19-1/data/2024/cm37266-1/SampleA

where:
xia2 – runs a xia2 processing job
small_moleule=true – tells DIALS to run a small molecule processing job, which means things like cell errors are calculated
/dls/i19-1/data/2024/cm37266-1/SampleA – path to where the images are stored.

Various additional instructions can now be added to this command – see the table for some of the most common ones.

Action	xia2 Instruction
Include space group and unit cell (both must be used together)	space_group= unit_cell="a b c α β γ" OR unit_cell=a,b,c,α,β,γ
Maximum resolution cutoff	d_min=
Choose a different indexing method	index.method= index.method=real_space_grid_search (normally the most robust method)
Set a threshold so only the strongest reflections are used for indexing	sigma_strong= (default value is 3)
absorption correction	absorption_level= choose from low/medium/high as most appropriate
Select which images should be included, for example to omit images which look to have radiation damage	image=path/to/images/image_XX_00001.cbf:first:last OR copy over an automatic.info file produced for the same dataset and edit the images required in this file - save as edit.info include xinfo=edit.xinfo
Can help integration complete	scan_varying=false
Will make the job complete even if the processing doesn't work for all sweeps	failover=true

So, an example command may look like:
xia2 small_moleule=true sigma_strong=15 index.method=real_space_grid_search space_group=P21 unit_cell="7.72 8.68 10.82 90 103 90" /dls/i19-1/data/2024/cm37266-1/SampleA

shelxt
Manually running xia2 processing does not include the structure solution step.
If desired, shelxt can be run on the xia2 output files.

Navigate to where the files are: cd DataFiles
type shelxt shelxt to run shelxt

It is possible to include the correct (expected) atom types by editing the atom types listed in the shelxt.ins file using gedit shelxt.ins

Files

The manual xia2 processing results are not incorporated into ISPyB so must be viewed (use gedit to open files from the terminal) within the file system – useful files can be found in the following locations.

Main folder:
xia2.txt
xia2.html

DataFiles subdirectory:
Shelxt.ins and shelxt.hkl, plus any output files from running shelxt
Xia2.cif

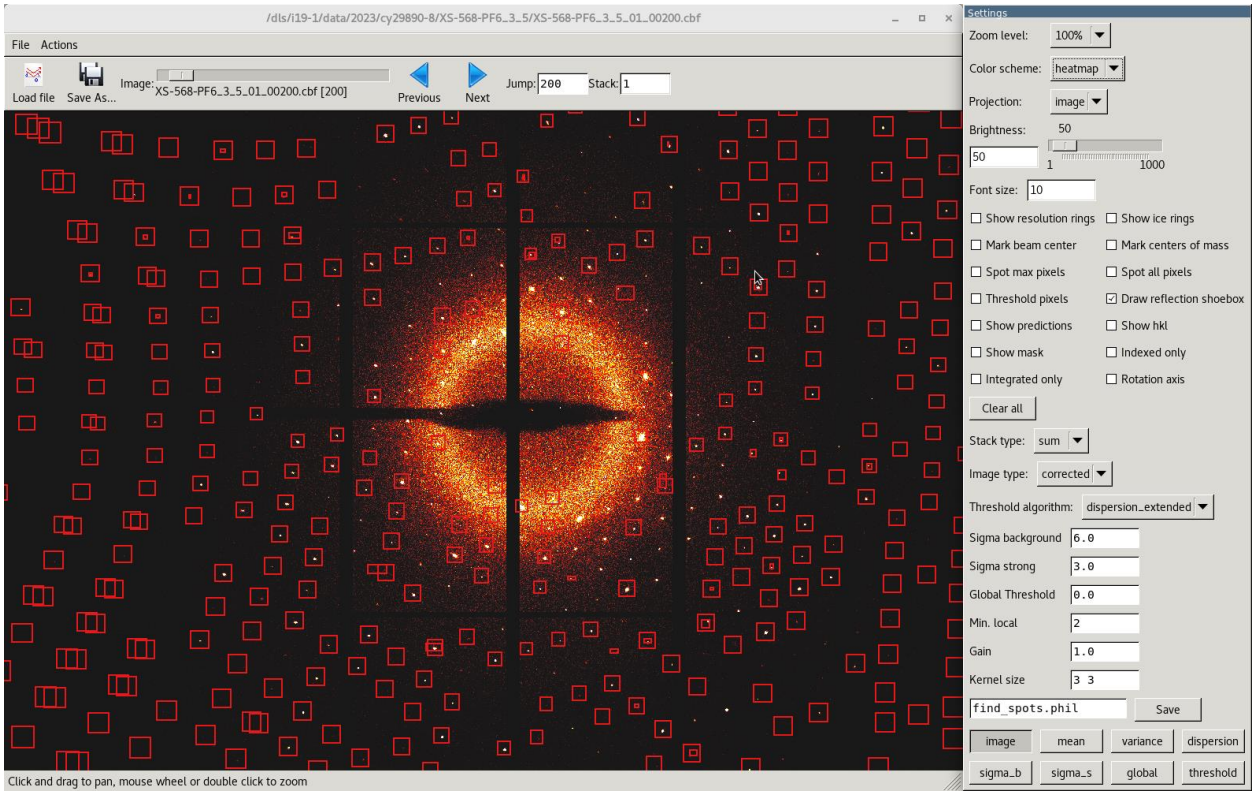
LogFiles subdirectory:
AUTOMATIC_DEFAULT_NATIVE_merging-statistics.txt

Viewing Images

dials.image_viewer can be opened using the files produced from the various processing steps.
A matching pair of .refl and .expt files are required.

If the auto processing has completed, then the files corresponding to the final stage reached (scaling) are written to the DataFiles folder:
AUTOMATIC_DEFAULT_scaled.refl
AUTOMATIC_DEFAULT_scaled.expt

From DataFiles, type (dials.ima TAB)
dials.image_viewer AUTOMATIC_DEFAULT_scaled.refl AUTOMATIC_DEFAULT_scaled.expt



By opening the images this way, knowledge of the indexing is read in so having the options checked will show how well the indexing has been done, for example.

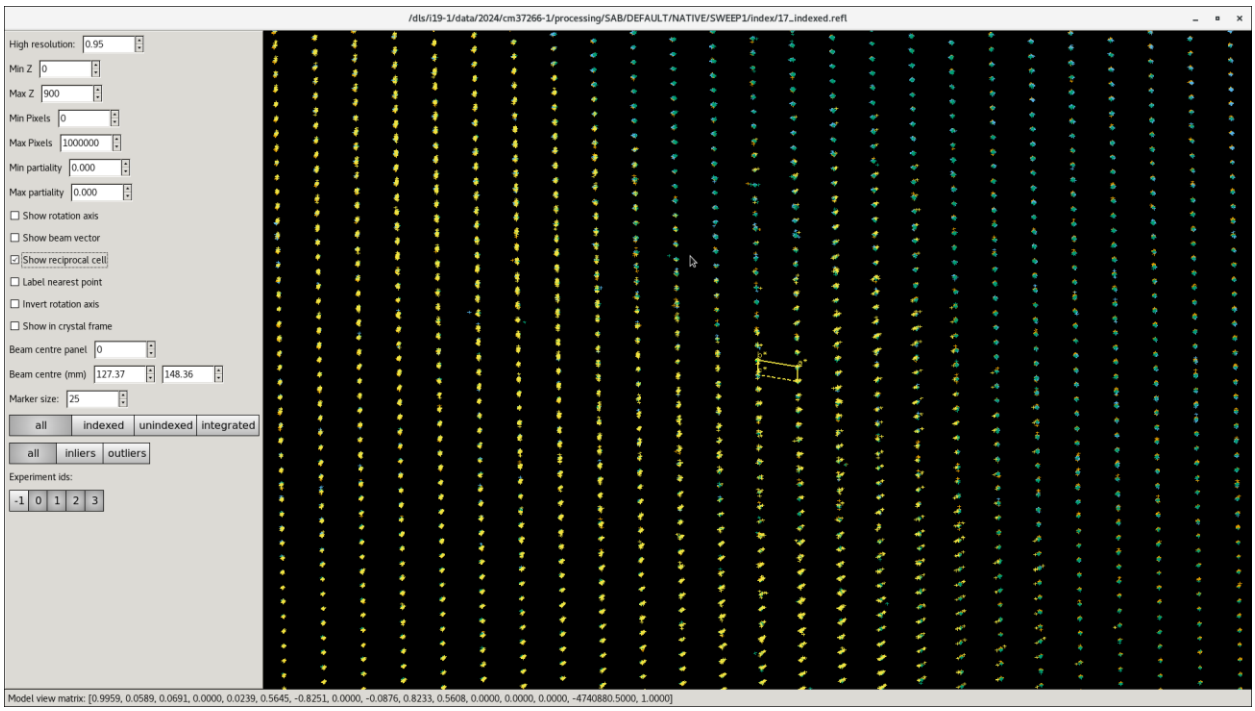
Reciprocal Lattice Viewer

It is possible, from the file system, to open the reciprocal lattice viewer to include reflections from all sweeps of data collected. Navigate (cd) to the directory DEFAULT/NATIVE/SWEEP1/index.
Although the folder is SWEEP1, the results from indexing all sweeps are actually here.

If you have collected 4 sweeps of data, then the indexing results are numbered 17, but choose whichever files are titled indexed (dials.rec TAB)
dials.reciprocal_lattice_viewer 17_indexed.expt 17_indexed.refl

It is possible see which reflections come from which sweep (they are different colours) by selecting the appropriate Experiment ID's and to see which reflections have been indexed or not.

It is also possible to overlay the unit cell and see how well that fits.



If indexing has failed then look, instead, at the matching pair of files called xx_strong.refl and xx_strong.expt as this may give some useful insight.

Radiation Damage

If you are concerned about radiation damage and want to check for evidence, there is a damage analysis tool (dials.dam TAB)
dials.damage_analysis DataFiles/AUTOMATIC_DEFAULT_scaled.{expt,refl}

Then open the resulting file by typing:
firefox dials.damage_analysis.html &

Look at the Rcp vs dose plot to check where the line veers off and then use this as a guide as to where to cut the data for the next reprocessing job.

Auto Processing

It is possible to look at the results from the auto processing in the same way as for the manual processing, but it is important to note that most of the files required are only available for **1 week** after data collection.

They can be found within tmp/zocalo and the file structure here matches the way the data was collected. The results from all auto processing jobs can be found here, including the screen19 output.