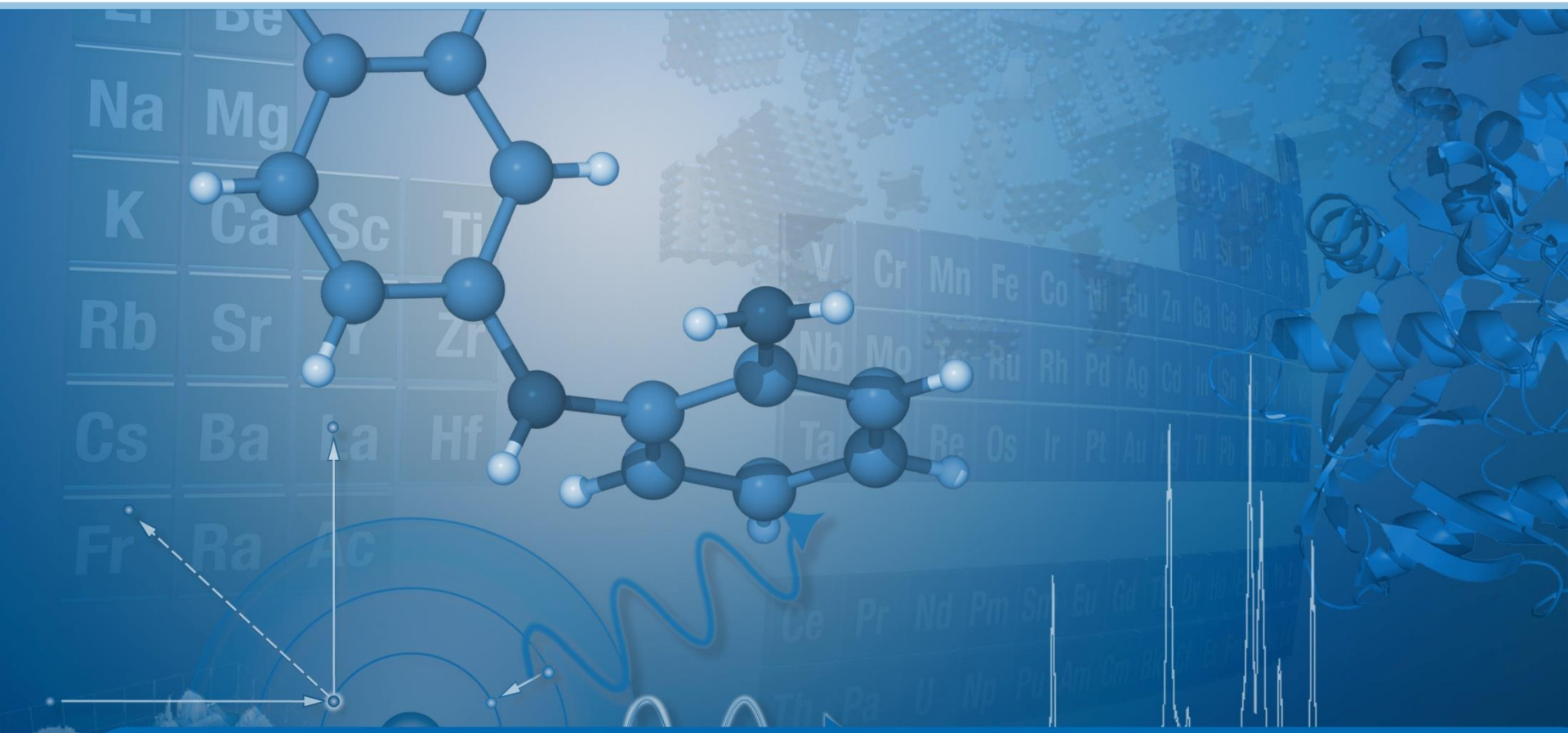
Data Reduction and Evaluation with PROTEUM3



PROTEUM2 Suite

The PROTEUM2 suite has a completely new approach on how a user interacts with a crystallographic experiment. The Graphical User Interface guides the user through the complete experiment with minimal user input and maximal graphical feedback. PROTEUM2 is easy to use for the novice but has all the features required by expert crystallographers.

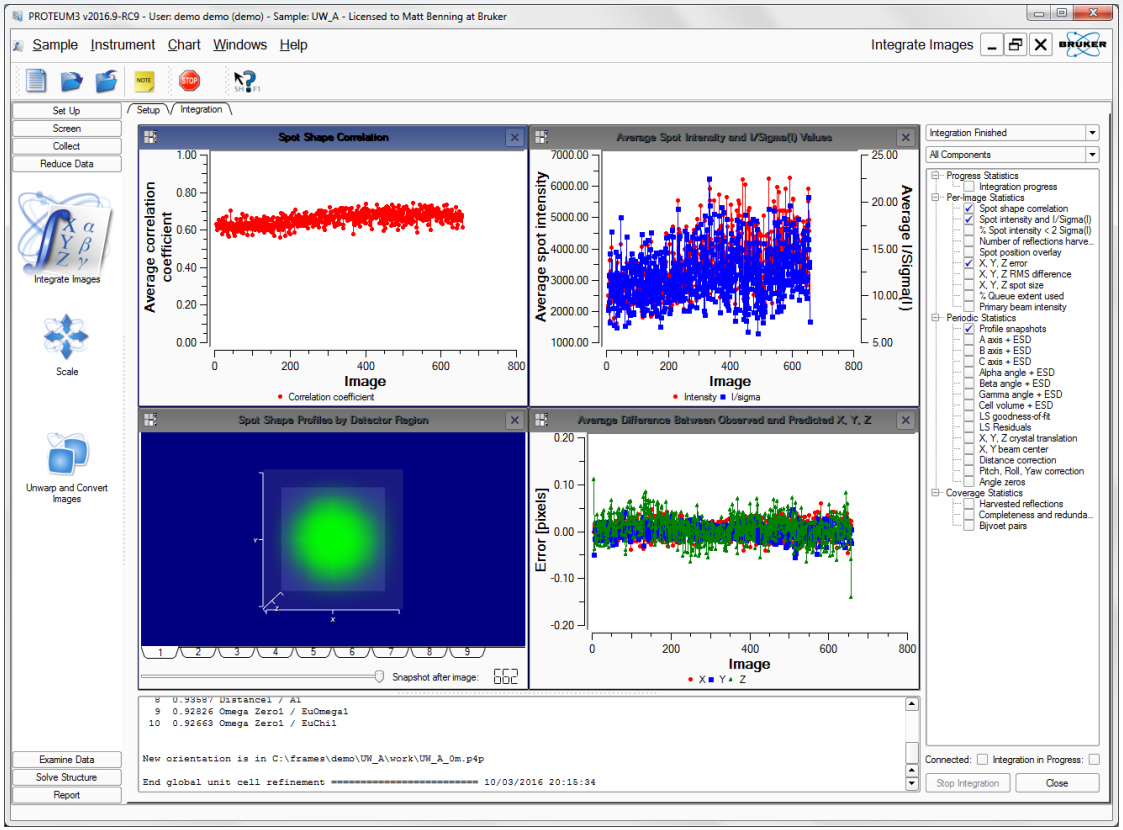
Some of the software's included in PROTEUM2 suite are:

- SAINT – 3D profile integration
- SADABS – data scaling with absorption correction
- XPREP - space group determination and data analysis
- Pointless/Aimless – data analysis and create MTZ

SAINT
Integration

Determine the raw intensities of the reflections

- True 3D profile fitting
  - Creates reflection profiles
  - No partial reflections
- Extended Graphical feedback
  - 3D profile display
  - Spot overlays
- Automatic, manual modes
- Easily handles fine sliced data
- Handles twinned data

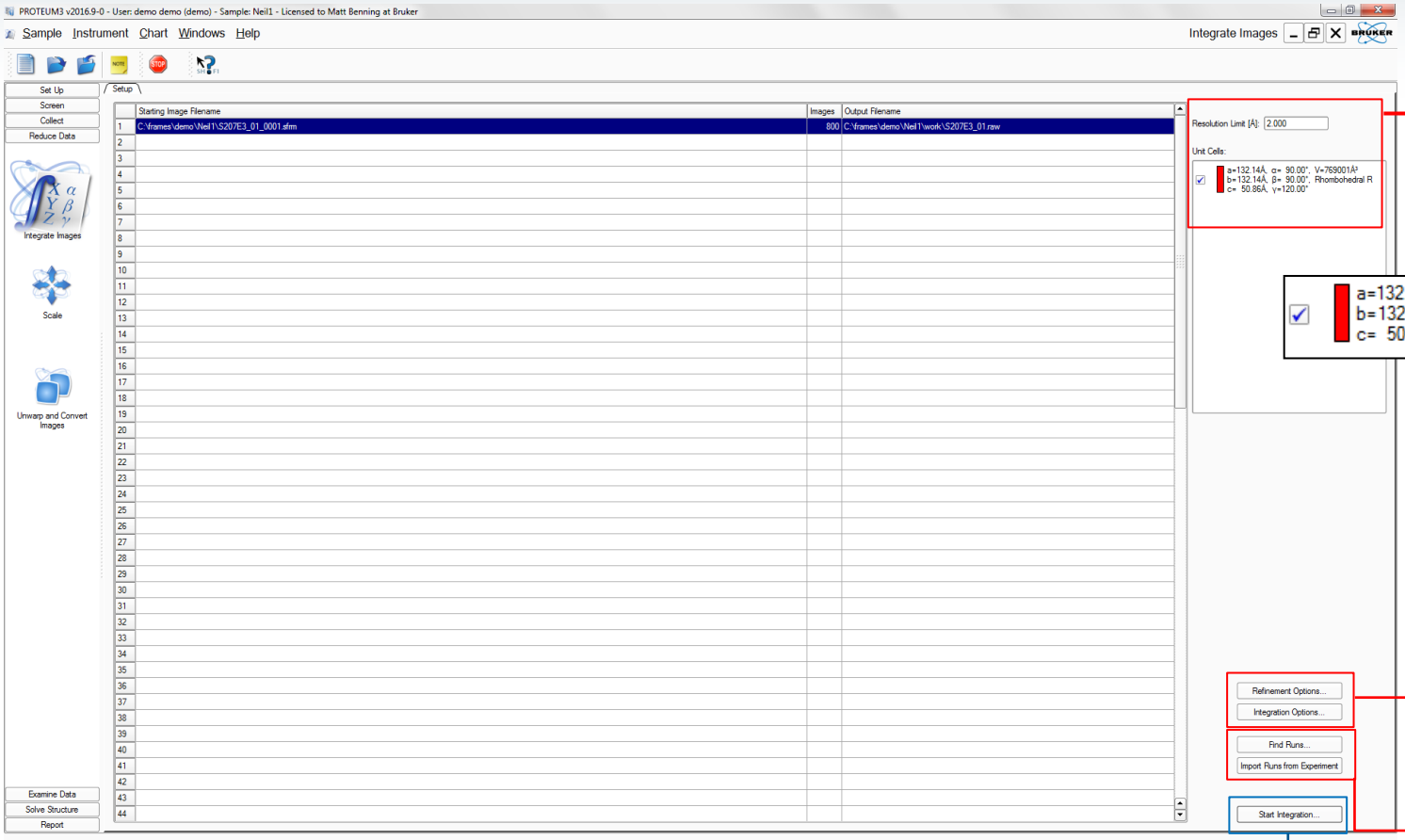


SAINT
Integration

Steps during integration:

- Determination of an initial background
- Determination of active pixel mask (for marking reflections which are outside the detector active area, behind the beam stop or the shadow of the low temp device)
- Read-in the orientation matrix
- Determination of initial spot shape profiles, with concurrent refinement of the starting orientation matrix and initial background
- Integration of each defined run; output intensities are corrected for Lorentz factor, polarisation, air absorption and absorption due to the variation of the path length through the detector faceplate
- Elimination of spots whose shapes correlate poorly with model profile shapes, relative to other spots of similar I/σ(I)

SAINT
Integration



Check resolution limit and unit cell constants

a=132.14Å, α= 90.00°, V=769001Å³
b=132.14Å, β= 90.00°, Rhombohedral R
c= 50.86Å, γ=120.00°

Refinement, integration options

Importing runs

Start integration

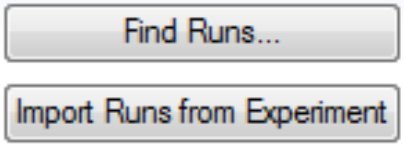
SAINT
Importing runs

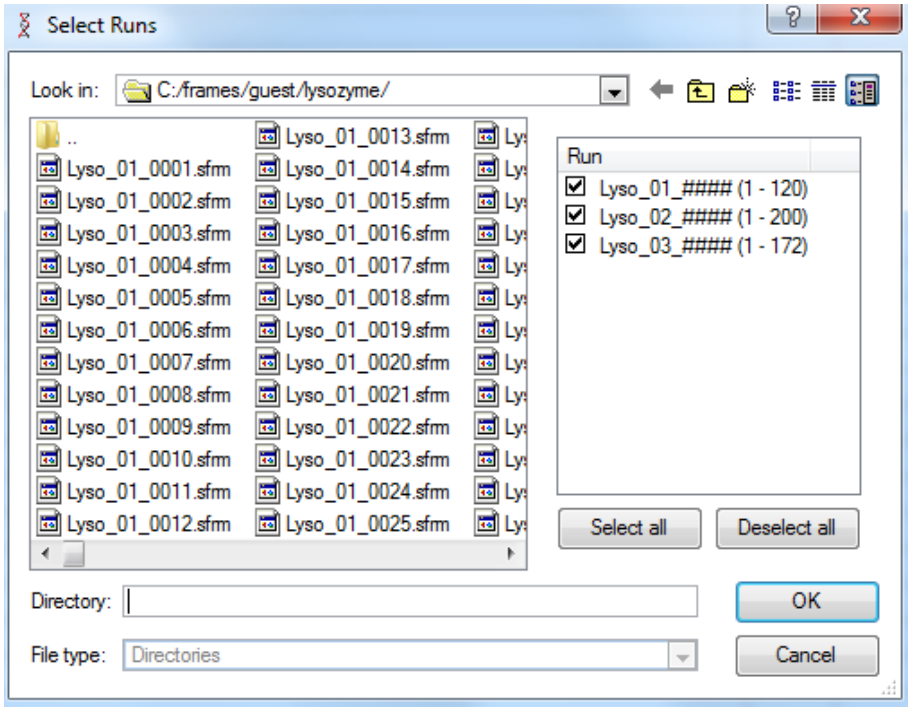
Find Runs

- Looks in the entry folder for the number of runs and images
- You can select all or just the runs you want and click “OK”

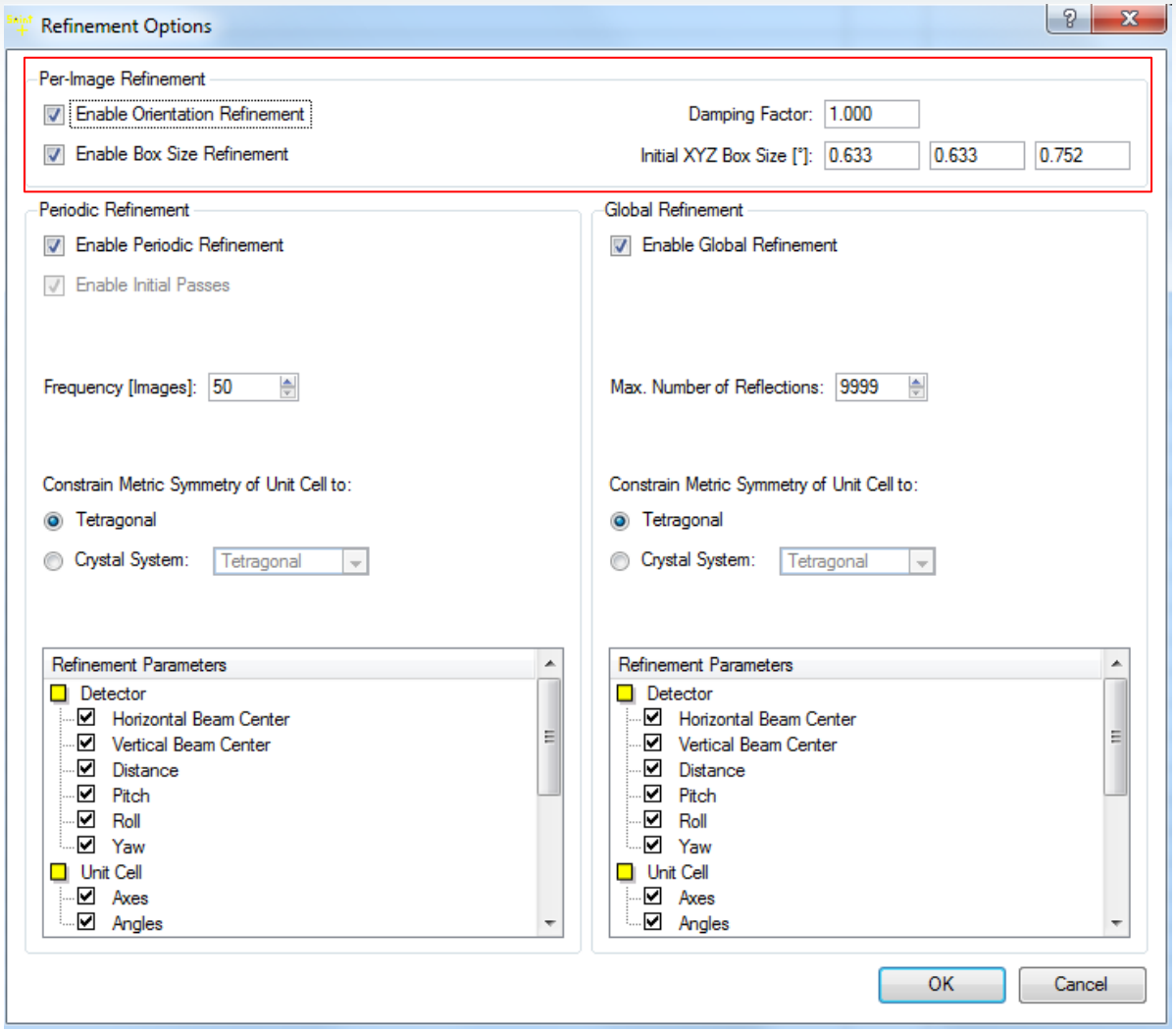
Import Runs from Experiment

- Gets scan information directly from BIS
- Need to be connected to the instrument to import from BIS





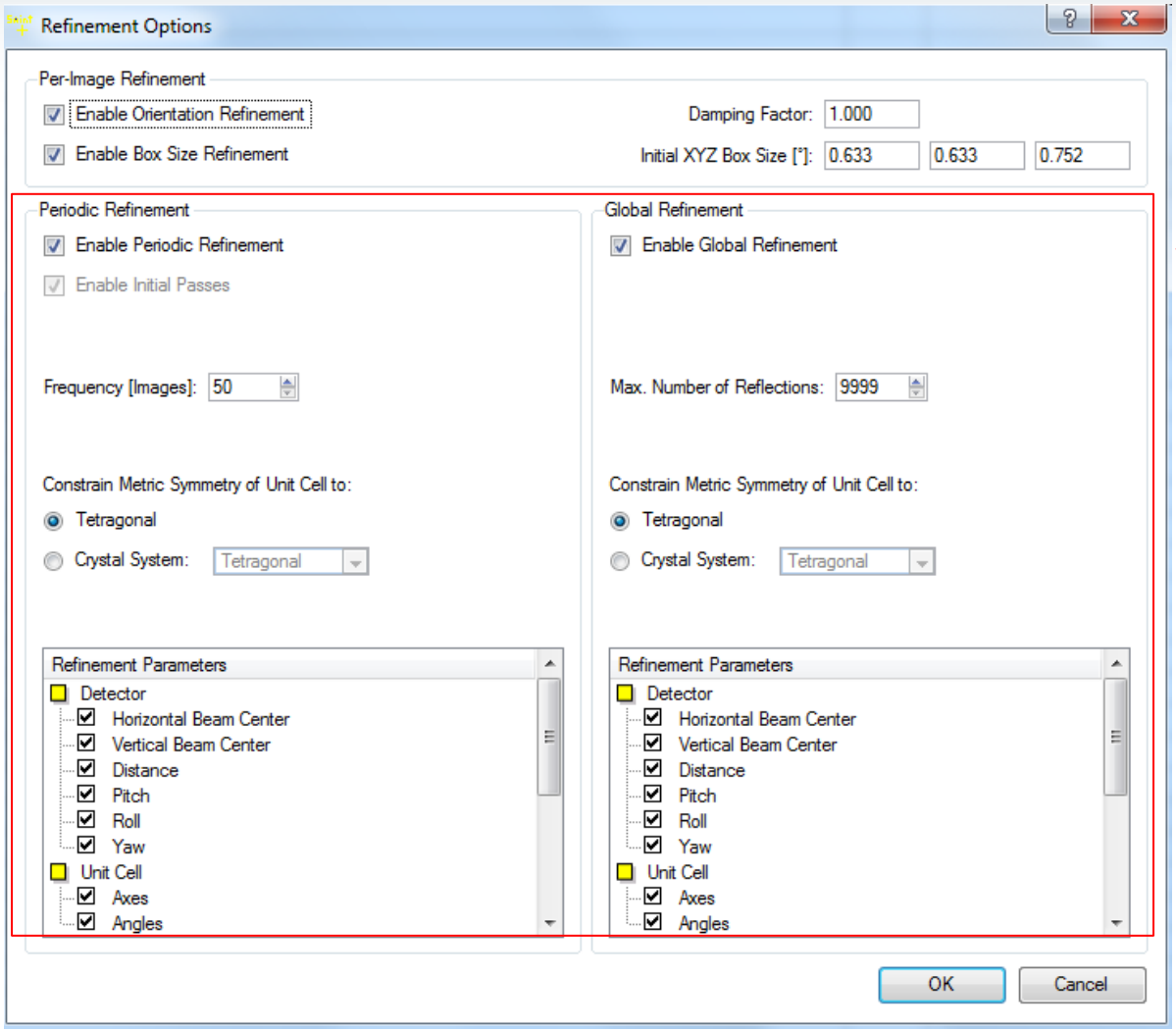
SAINT
Refinement options



← Initial box size

- Determined automatically by the program and refined during integration.
- If the mosaic spread is very high (> 1.5°), you may want to turn off the refinement and set the box size based on the initial profiles.

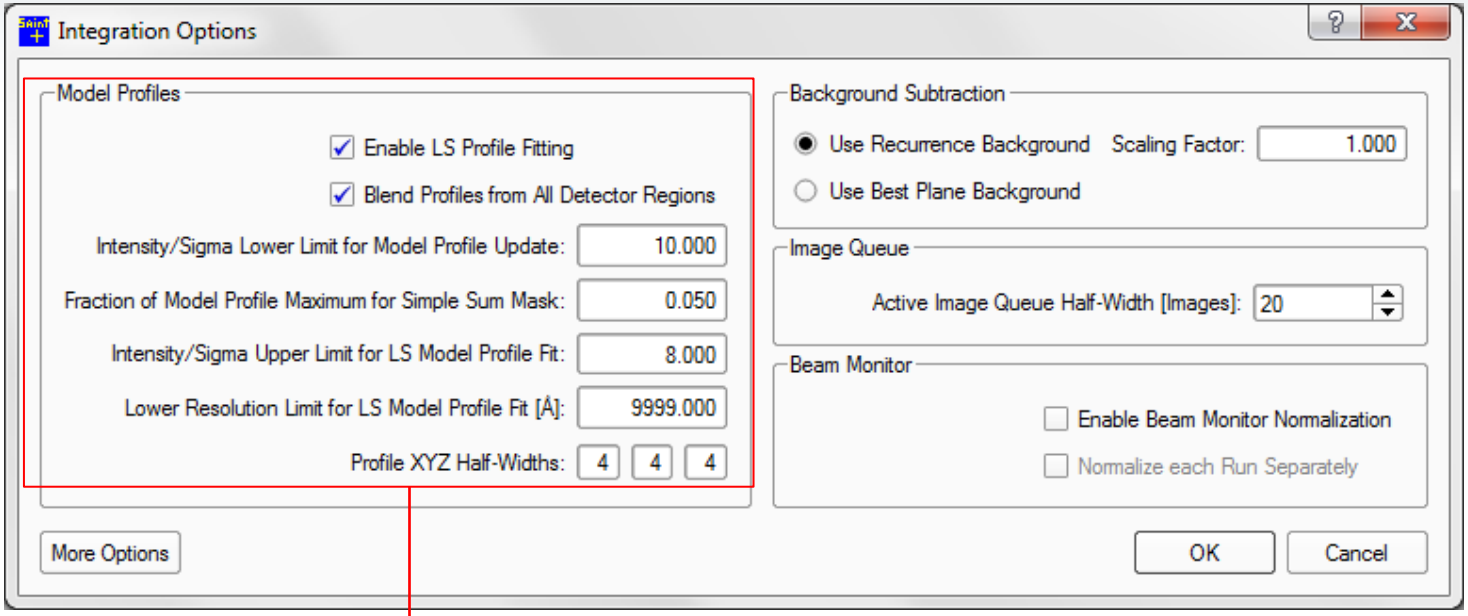
SAINT
Refinement options



← Cell refinement

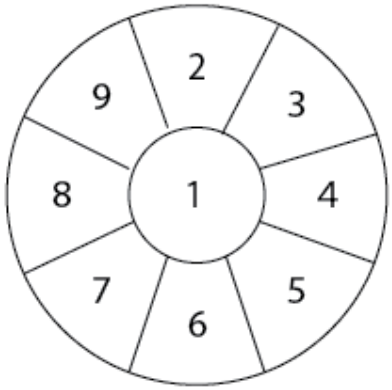
- Periodic LS refinement during integration after a set of images.
- Global LS refinement takes reflections from the whole data set and produces the final unit cell constants
- Refinement Parameters assigns the offsets updated during refinement. To add or subtract parameters, click the box next to the offset.

SAINT
Integration options



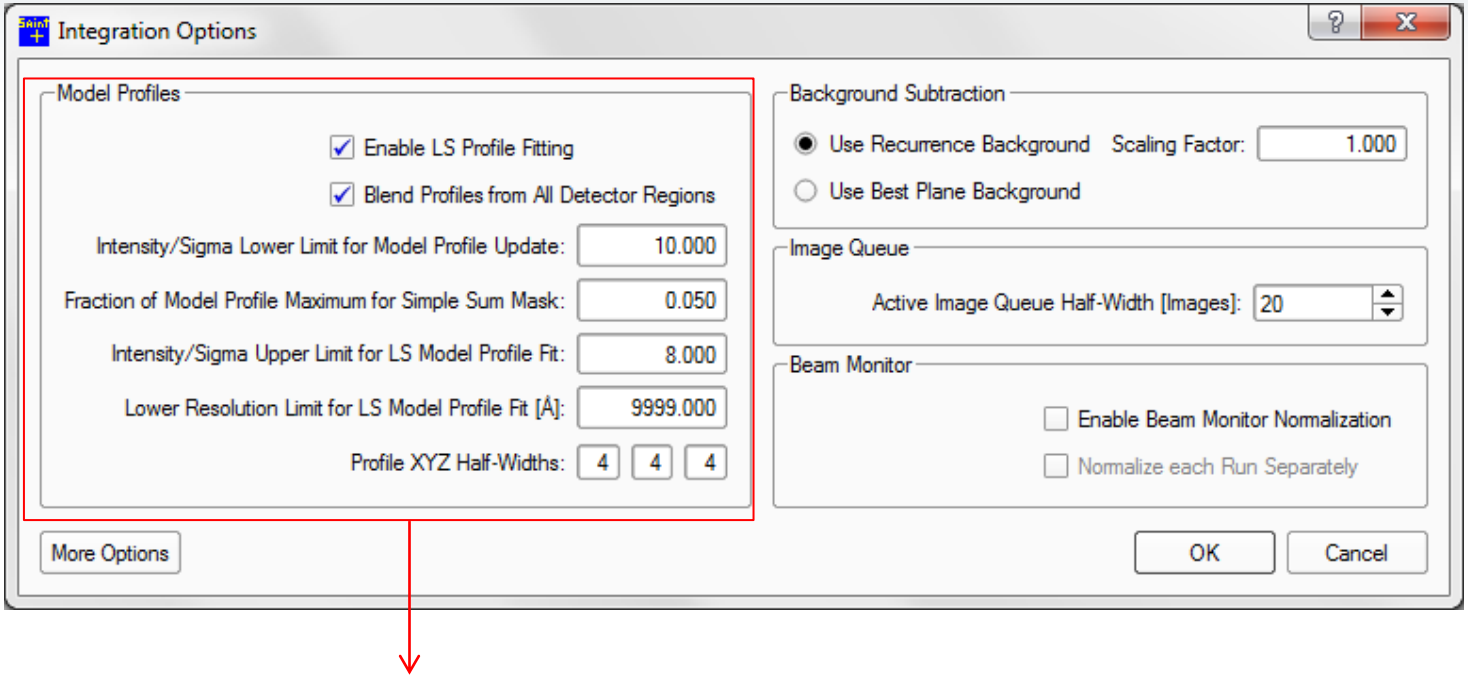
Model profile determination

- Use either one profile for the entire detector (Blend) or split the detector into 9 different regions.
- If the detector has enough strong reflections in each region then using nine profiles will probably improve things.
- If the data is weaker, blend the profiles into one global model



Regions for unblended profiles

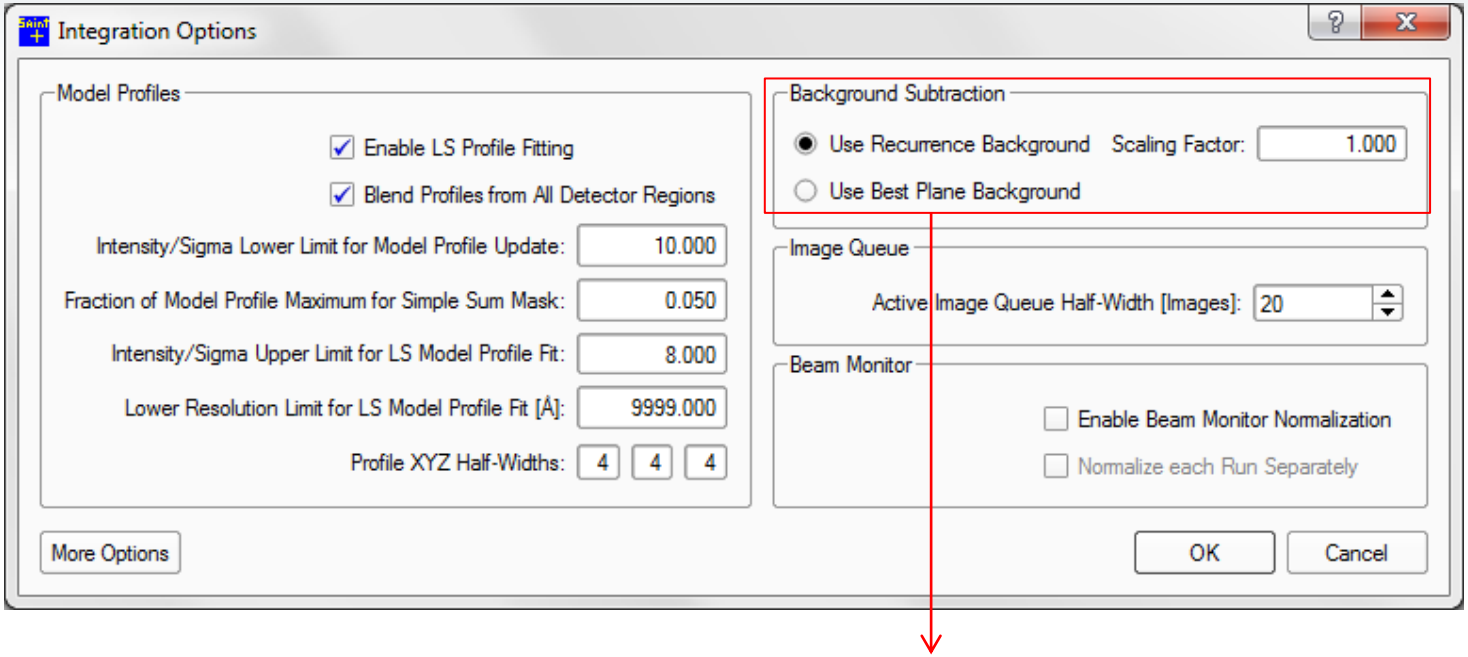
SAINT
Integration options



Model profile determination

- Enable LS profile fitting to help model the weak data better.
- Intensity/Sigma Lower Limit is the signal to noise cutoff for reflections used in the model profile determination.
- Profile XYZ Half-Widths – if using very fine slicing (ex 0.1 – 0.2°), try increasing the profile widths. The widths in each direction are 2N + 1, for 0.2° try 8, 8, 8.

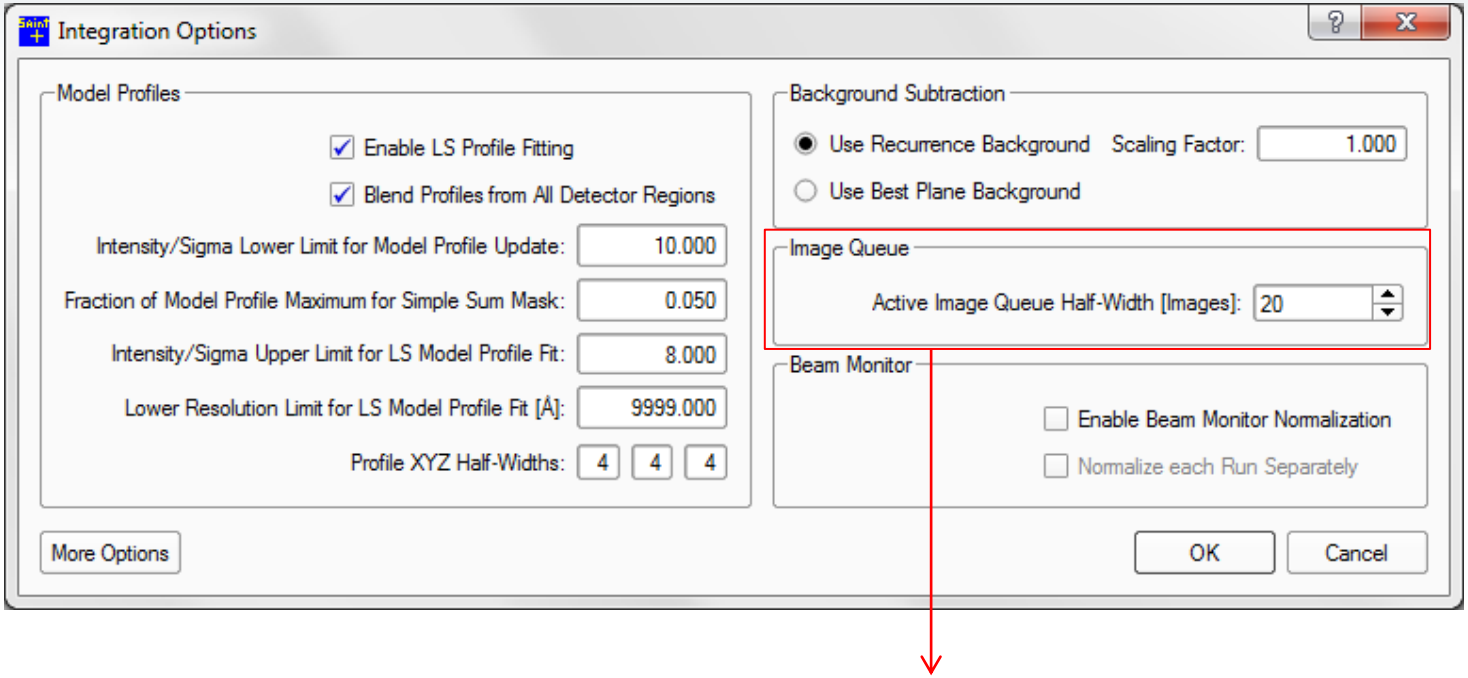
SAINT
Integration options



The background scatter is subtracted to increase the signal noise of the reflection

- Recurrence method – Calculates average local background over several frames
- Best Plans method – Determines local background by pixels around the reflection on the current frame only (like HKL, denzo)
- Try both to see which gives the better result

SAINT
Integration options



The image queue defines the angular range over which a spot is integrated. Spots that are very wide, like those in the Lorentz region, can be rejected.

- Defines the queue half-width (2N+1). For example, if you are collecting 0.2° rotations and have set the image queue to 7, the angular range is: 0.2° x 2(15) = 6°.
- Decrease the queue to allow more reflections to be rejected, increase it to integrate more of the data.

SAINT
Image queue

Image queue = set the number of frames used for determine the profile

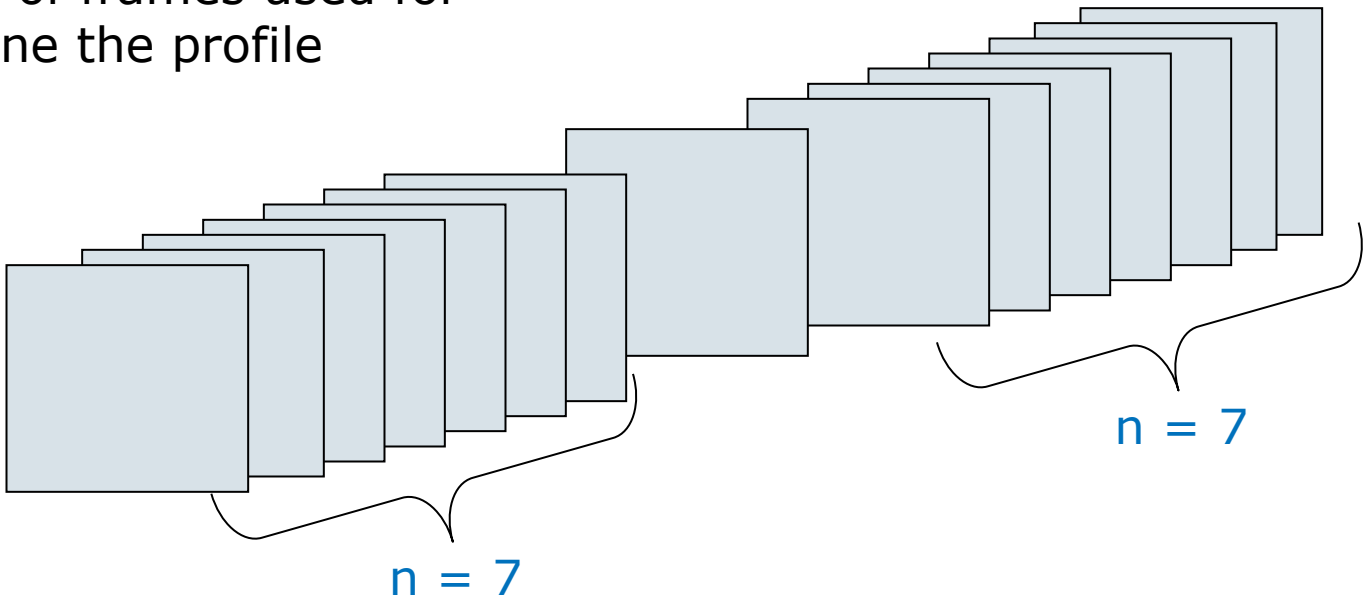
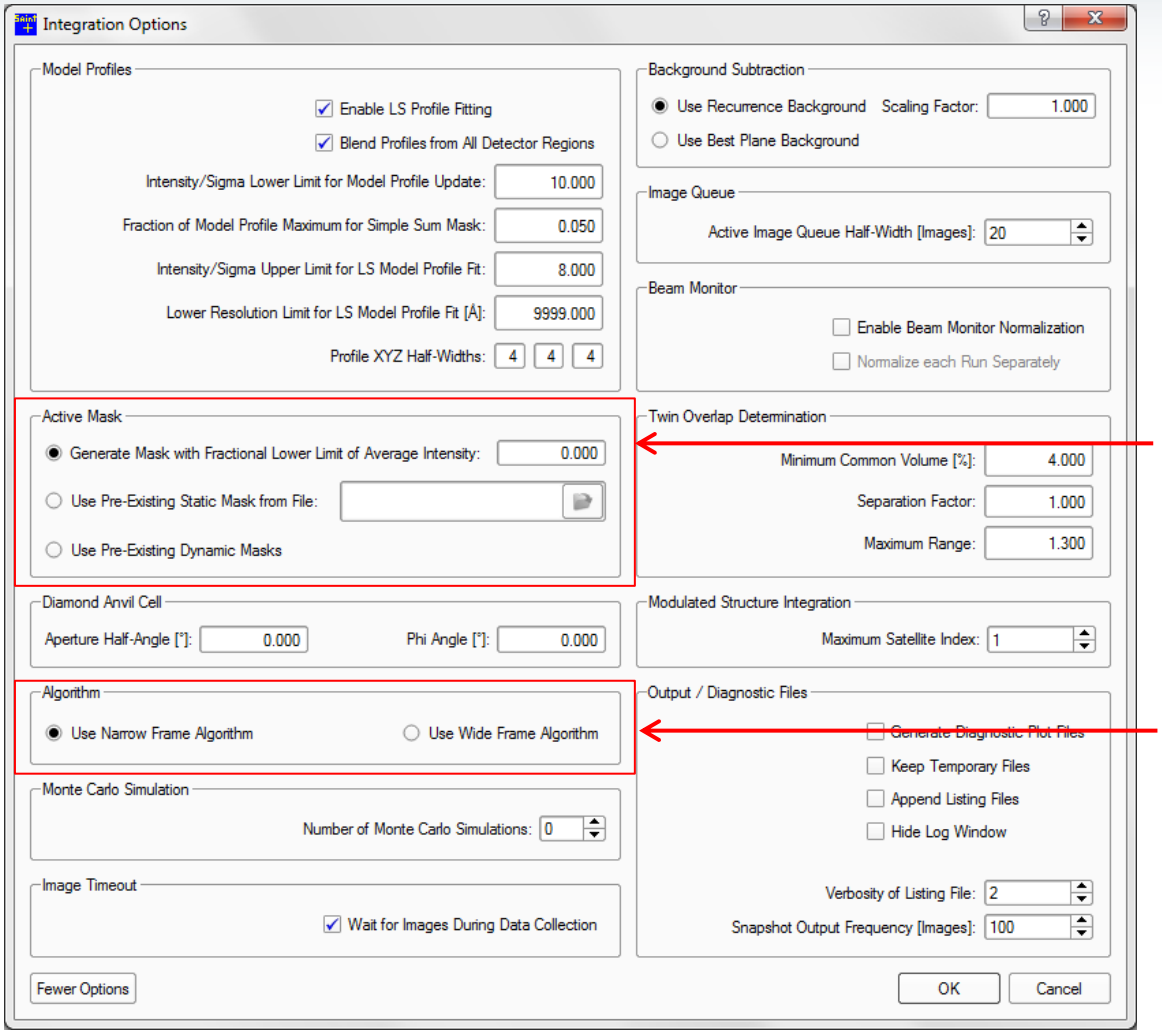


Image queue = 2n + 1 = 15

SAINT
Integration options



Selecting the More Options button shows more parameters

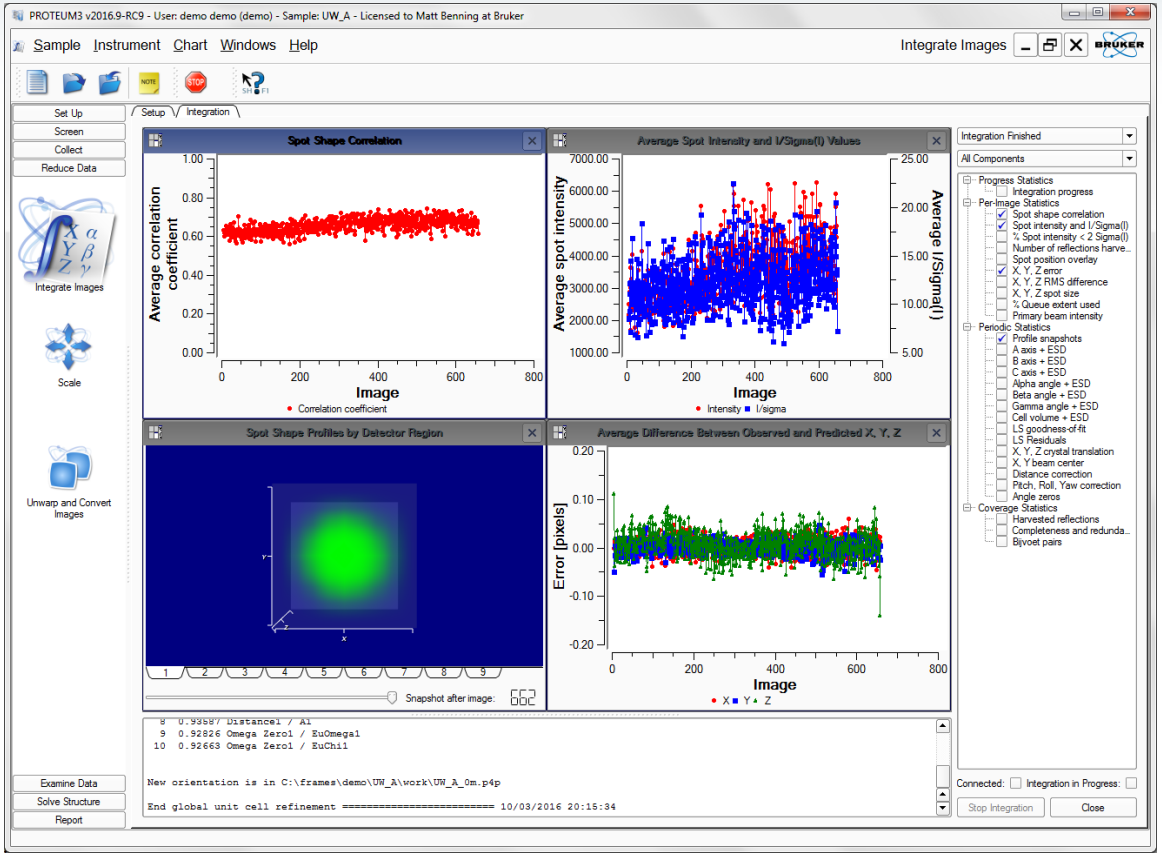
Active Pixel

- Creates a mask for the beamstop shadow
- Program automatically creates a mask if fractional lower limit is set to 0
- Can also read in a predefined mask (Synchrotron detectors)

Algorithm

- Narrow frame for rotation angle < 1°
- Try Wide frame for images > 1°

SAINT
Integration



Spot Shape Correlation

- Agreement between the model profile and reflections.
- Typically > 0.5, if too low (0.2) then the space group is not correct.
- Integrating at an incorrect resolution limit will also cause the correlation to be low.

Average Spot intensity

- Spot intensity and I/σI values per image

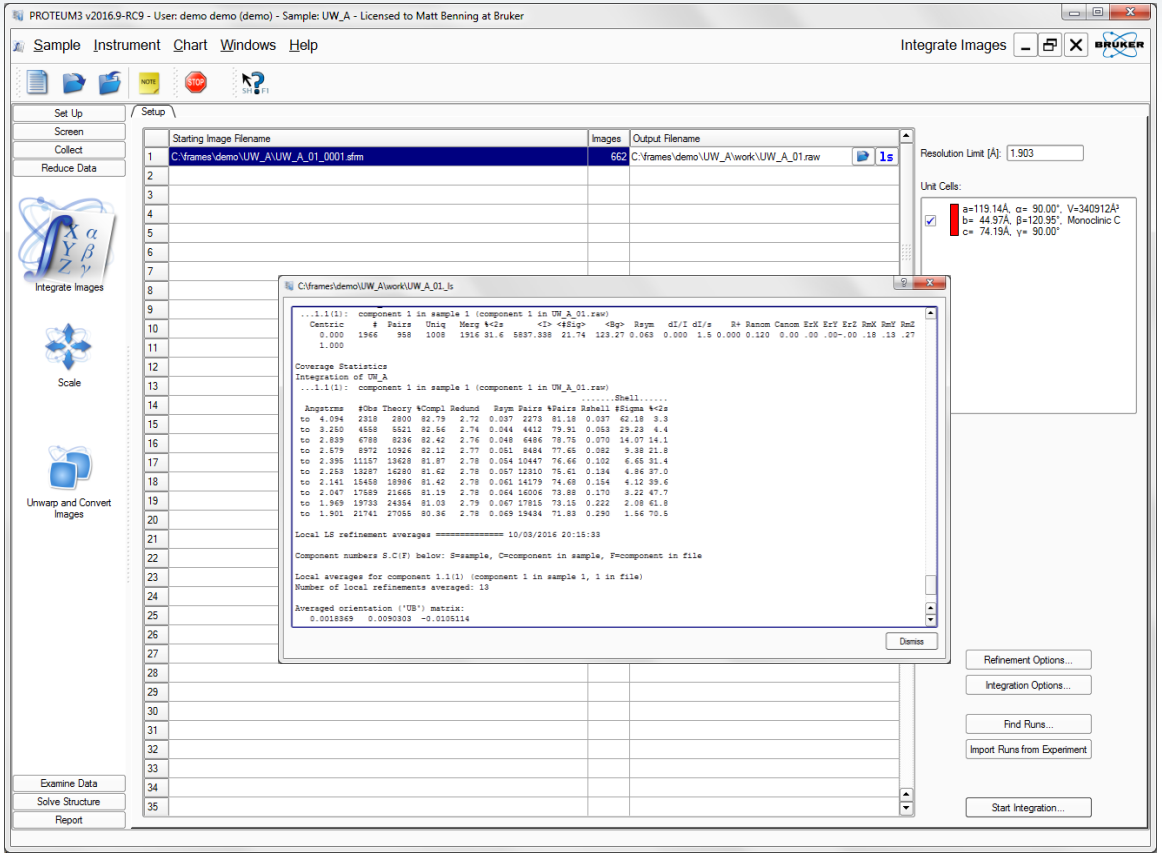
Average Difference: X,Y,Z

- Positional errors between observed and predicted reflections. Values consistently over 0.3 suggest problems

Spot Profiles

- 3-D display of the model spot profiles base on strong reflections

SAINT
Integration



Double clicking on the Output filename activates the  buttons.

- The folder button allows you to search for and update the filename.
- The “ls” button opens the log file.

SAINT
Integration

Output files in the *work* subdirectory

Integrate Images (SAINT)

Output Files	Extension	Description
Raw intensity	*.raw	Contains the raw unscaled, unmerged intensities. A separate file is created for each scan which has the filename prefix plus the scan number (*prefix*_#.raw). A merged file is also created containing all the reflections from each scan (*prefix*_0m.raw).
Log	*._ls	Contains the output from integration. A separate file is created for each scan which has the filename prefix plus the scan number (*prefix*_#._ls). A merged file is also created containing all the reflections from each scan (*prefix*_0m._ls).
Matrix	*.p4p	This file contains unit cell information. When the integration is finished, a file called *prefix*_0m.p4p is created which contains the updated cell information. There is also a file written, *prefix*_0u.p4p which contains the unconstrained cell constants. This file can be manually created in PROTEUM by selecting "export>p4p" file from the "Sample" menu in the upper right corner. The p4p file also contains the table for the detector spatial correction. If you're creating a new database entry to work with old data, be sure to read in a p4p file before continuing after opening the entry by selecting "Import>p4p" from the "Sample" menu.
Active Mask	*.sfrm	This is an image file which contains the mask for the beamstop shadow. The filename contains the frame prefix, run number and frame number (0001). For example, *prefix*_am_01_0001.sfrm. You can view this file in PROTEUM as you would any image file to verify that SAINT is properly masking out the shadow.
Charting	*.cht	This file contains all the charts that were displayed in PROTEUM during the integration. The file can be re-opened in PROTUEM by clicking on the "Integrate Images" plugin and selecting "Open Chart File" from the Chart menu in the upper right corner of the GUI.

SADABS
Data scaling

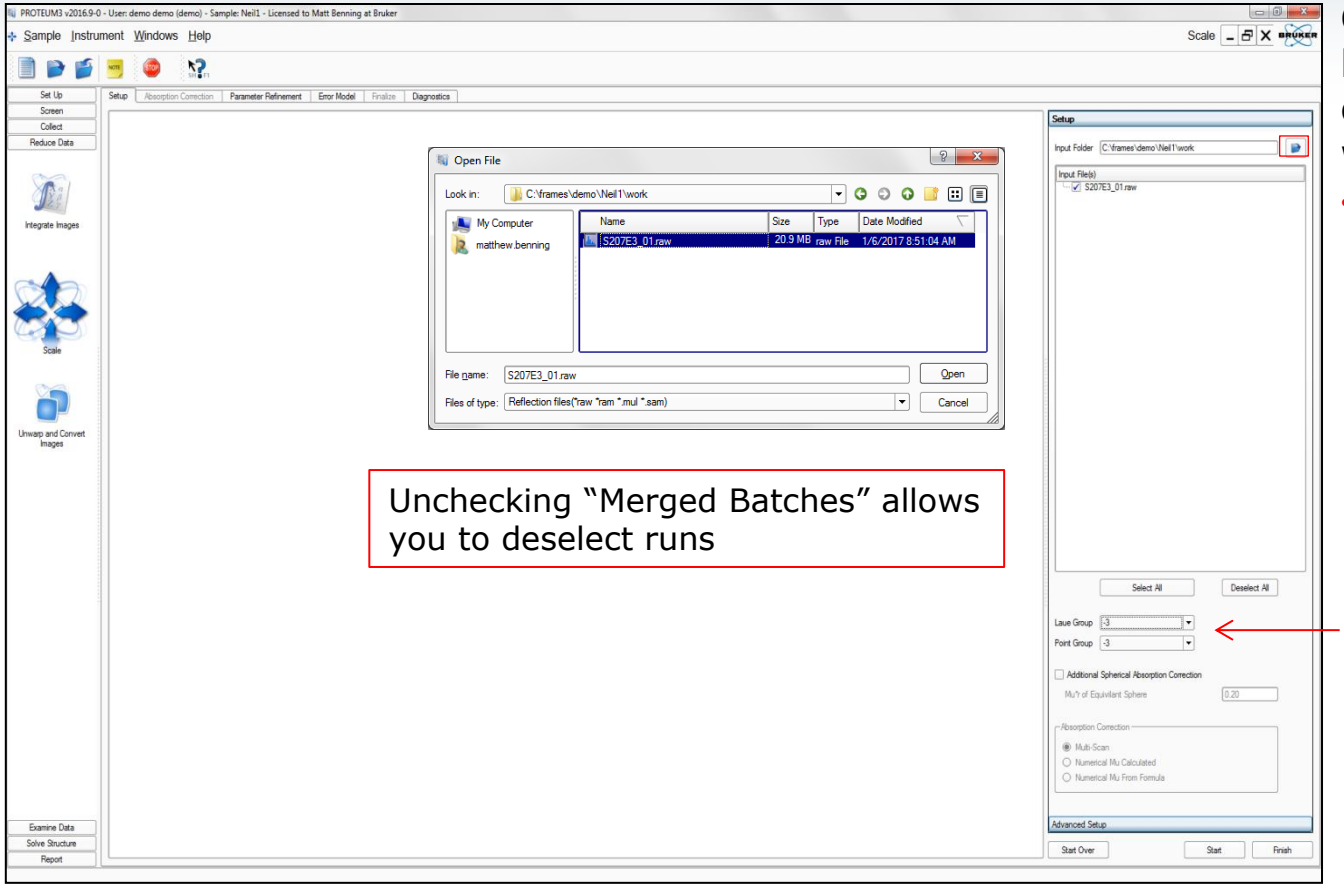
Steps during scaling:

- Scaling: determination of scaling and absorption parameters that assure the data is internally consistent
- Error model: the standard deviations of the intensities are modelled so that they are consistent with the deviation of the individual intensities from the mean intensity of group of equivalents.

Systematic errors:

- Absorption of the primary beam by the crystal (and support)
- Crystal decomposition
- Intensity variation of the primary beam (e.g. synchrotron)
- Changes in the effective volume irradiated.
- Beam inhomogeneity.

SADABS
Inputting Raw files



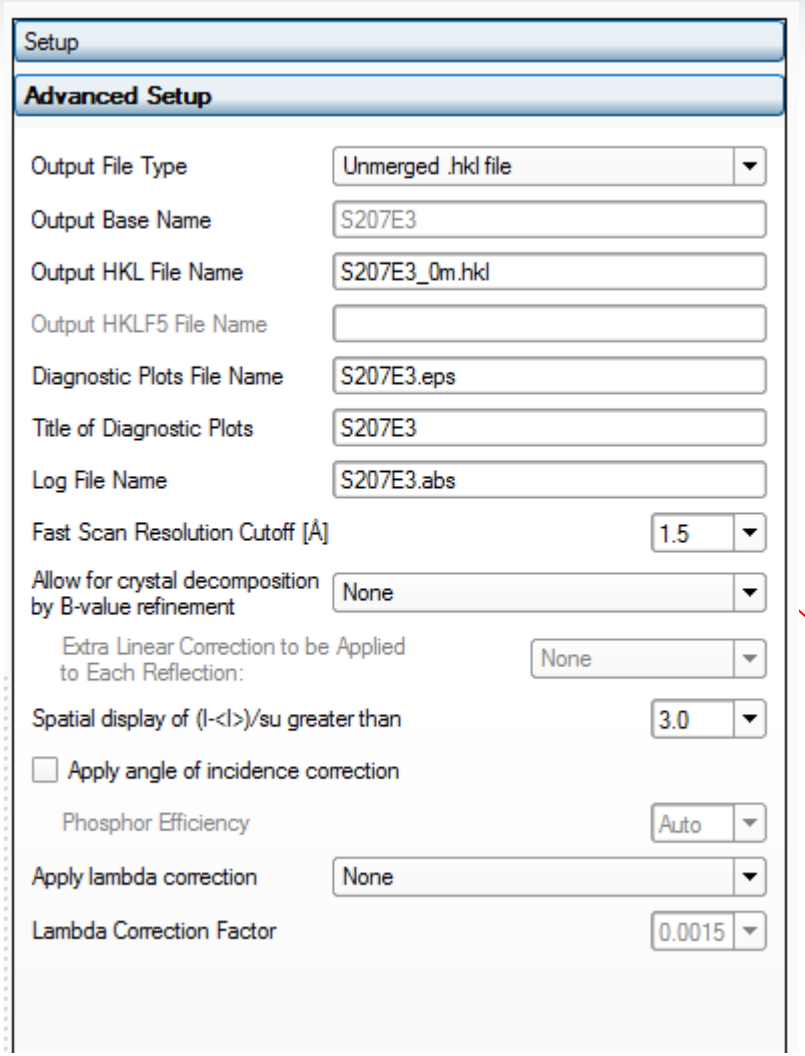
Clicking the browser button for the base name opens the selection window.

- Select any filename to input a single raw file

Point Group

- The point group will be set based on the assignment during indexing but you can change it by clicking the arrow
- To keep the Friedel mates separate uncheck the "Use only centrosymmetry point groups" box. All possible point groups will then be available.

SADABS
Advanced setup



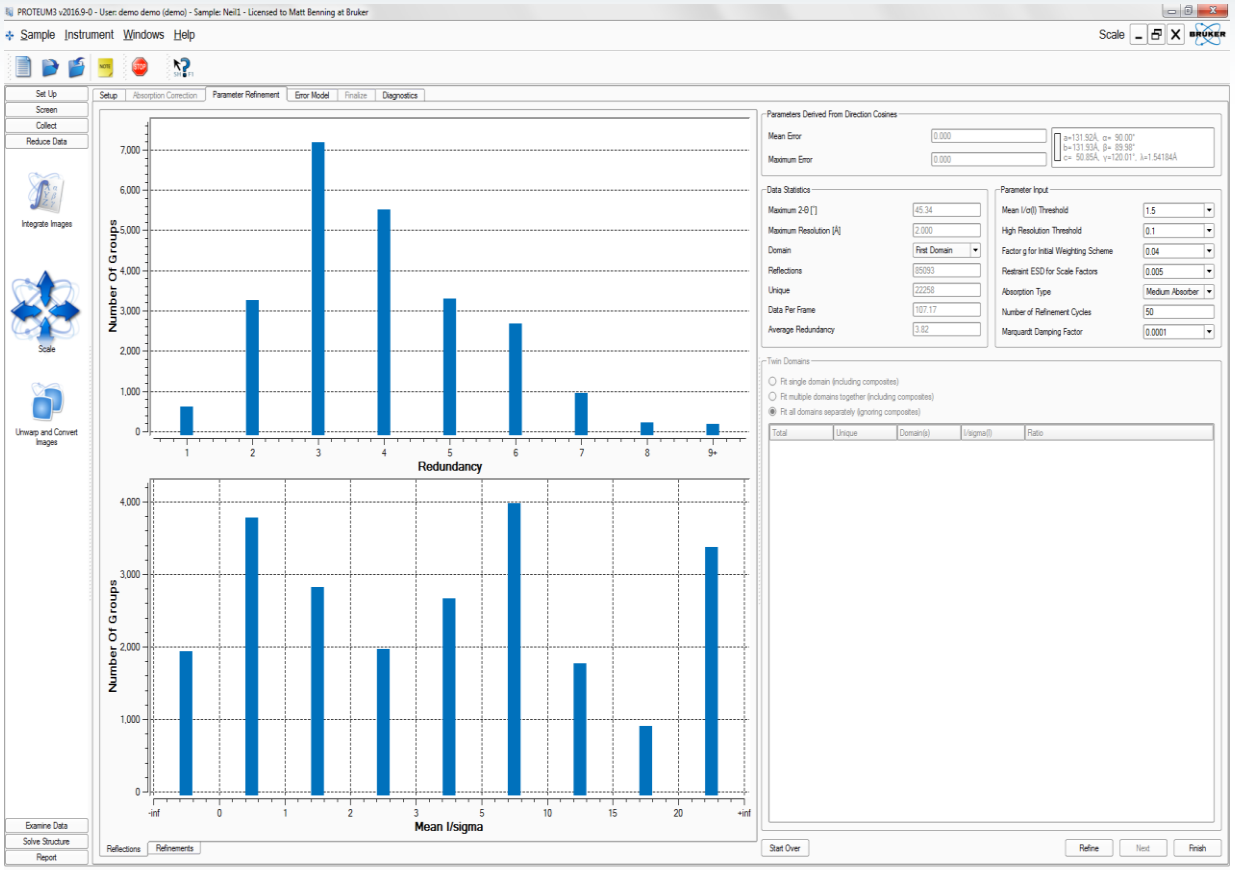
Output filenames are suggested based on the entry name. These can be changed by editing the box.

Zero-dose correction

Compare the same reflection collected as a function of time to model radiation decay

- Linear
- Quadratic

SADABS
Scale factors



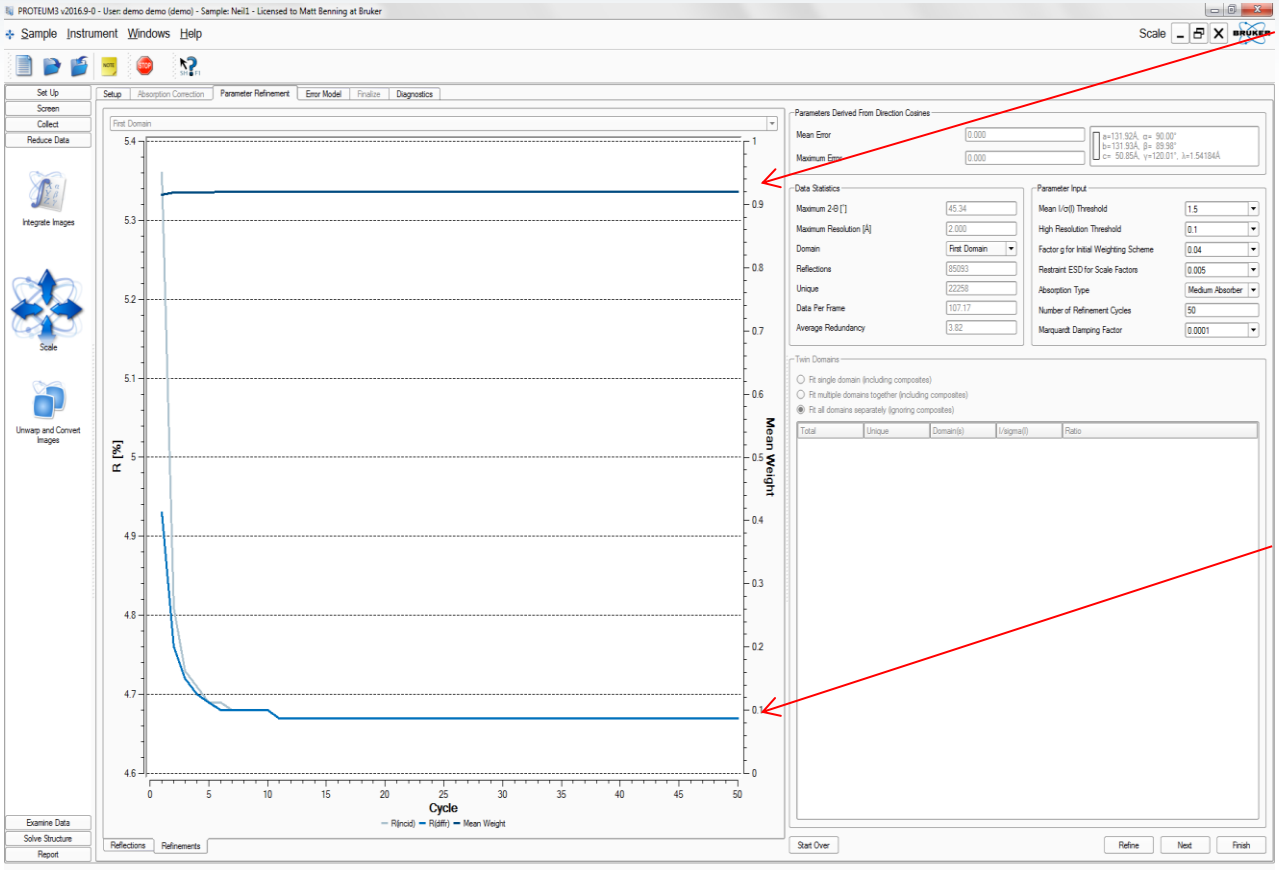
Check function

- Unconstrained cell constants and instrument error. Mean error should be >0.005.

Parameters refinement

- Scale factor restraint prevents overfitting data. Can loosen a bit, 0.01
- Absorption type, medium works well for most but if there are heavy atoms and enough data can try strong absorber

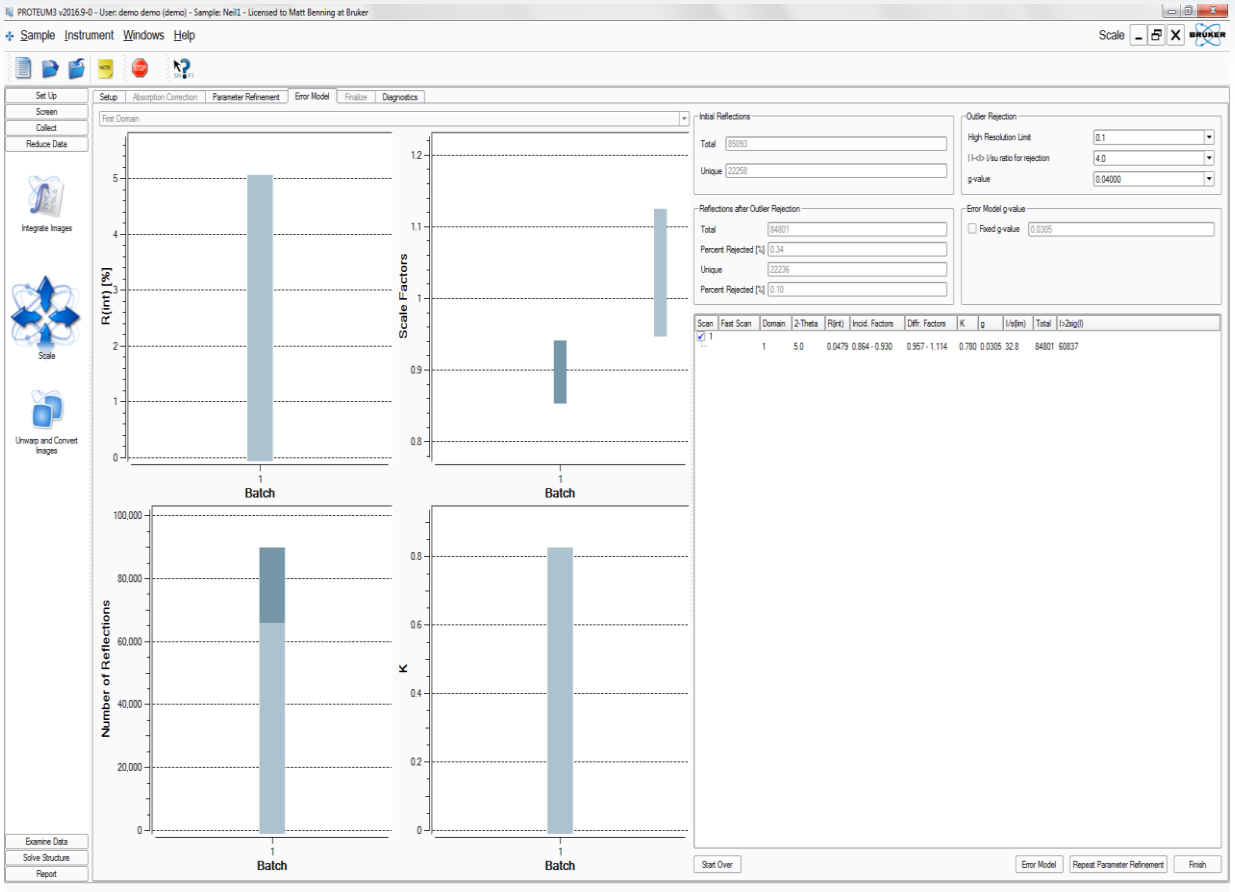
SADABS
Scale factor



- Blue line shows the mean weight of the observations for all the reflections. As the observations get farther from the mean, they are down weighted. If the Mean Weight falls below 0.75, the data agreement is not good.

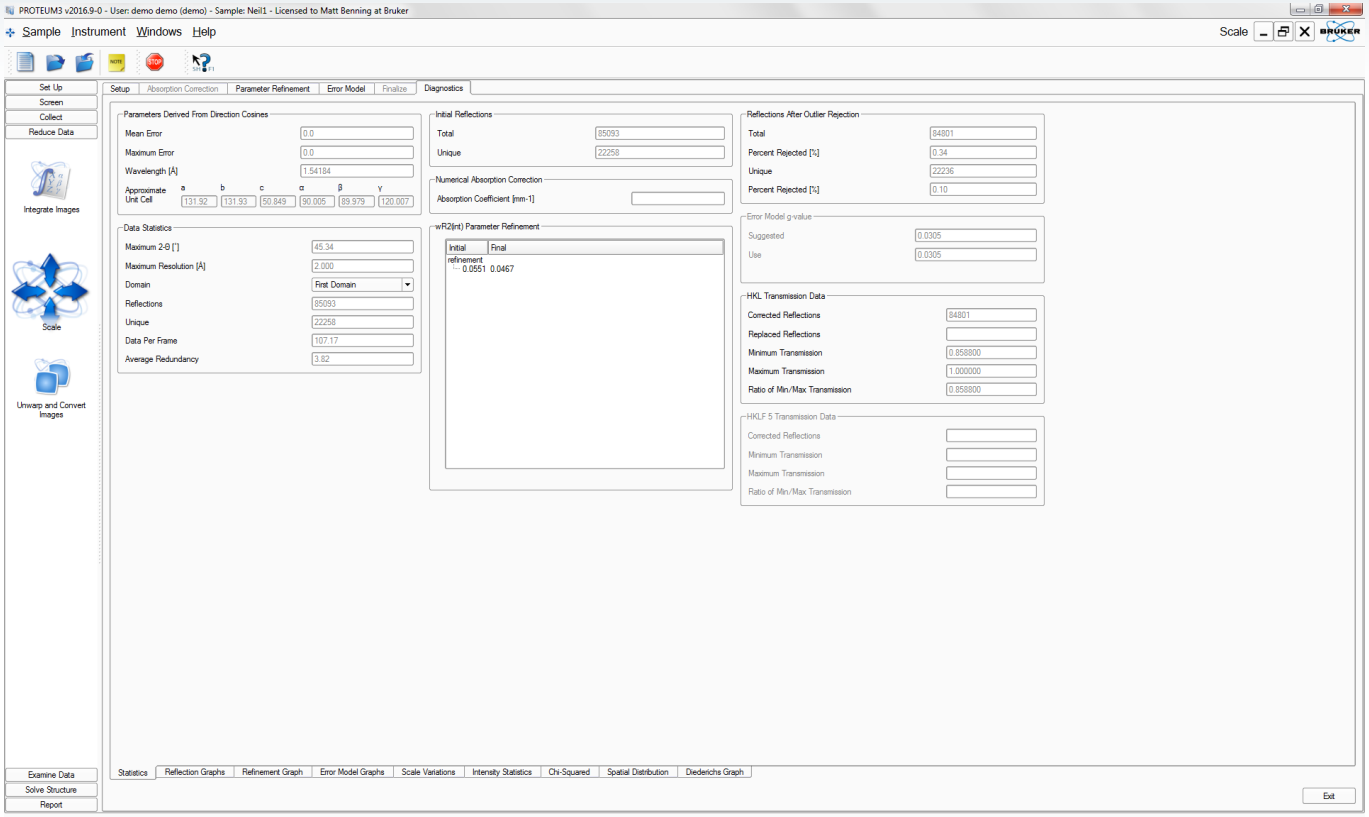
- Light blue line represents the Rfactor with scale factors only, the dark blue line is the Rfactor adjusted for adsorption. Most of the time they will converge but when there is a significant absorption affect, the blue line may exhibit a lower Rfactor.

SADABS
Error model

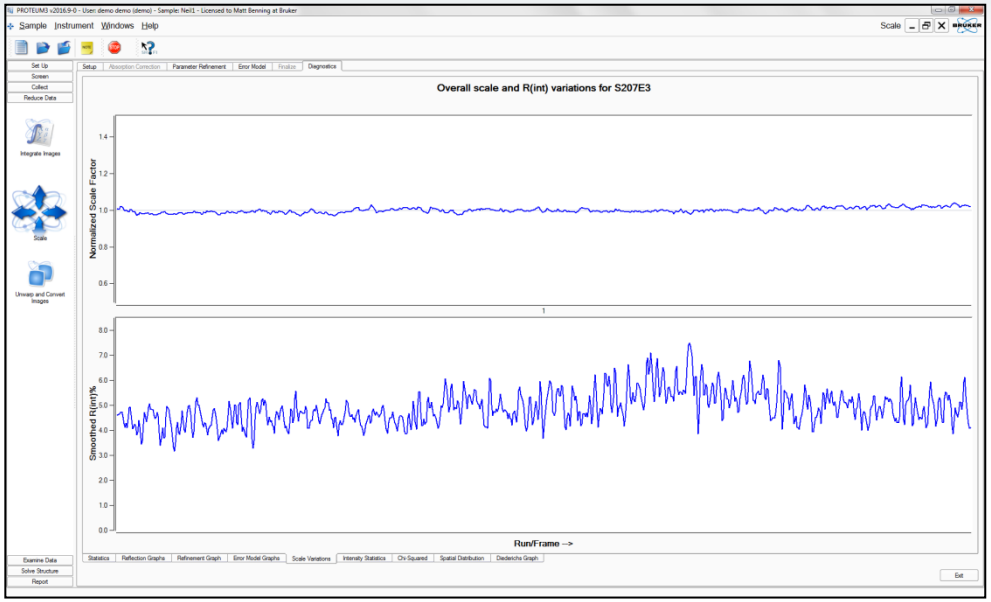


- Determination of an error model for errors that cause equivalent reflections to disagree.
- It deletes a small number of reflections that are completely incompatible with their equivalents, for example reflections blocked by the beam stop etc.
- Then determines an error model for the remaining reflections by fitting χ^2 to unity to put σ(I) onto an absolute scale.

SADABS
Diagnostics

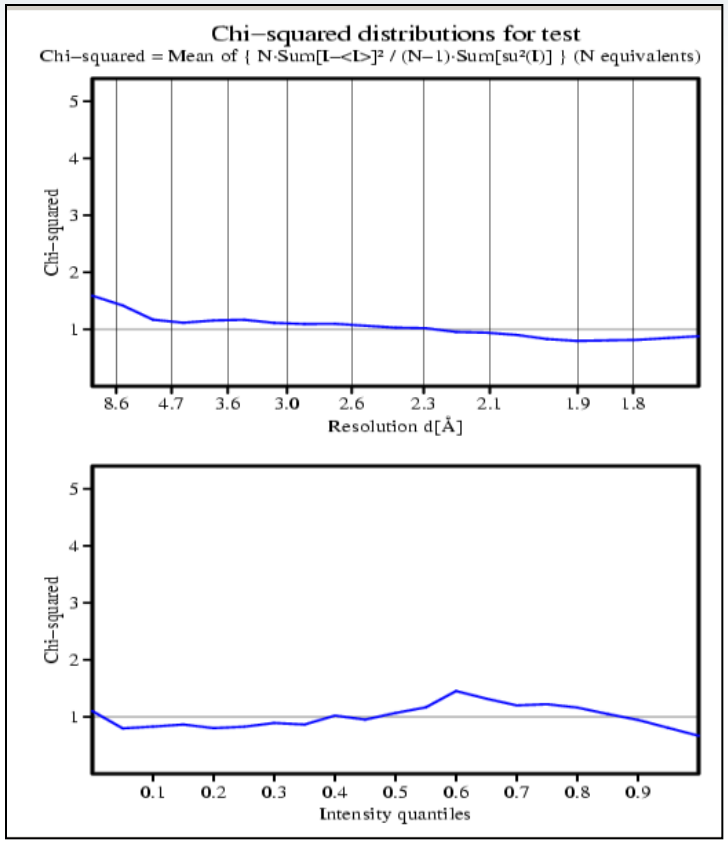


SADABS
Plots

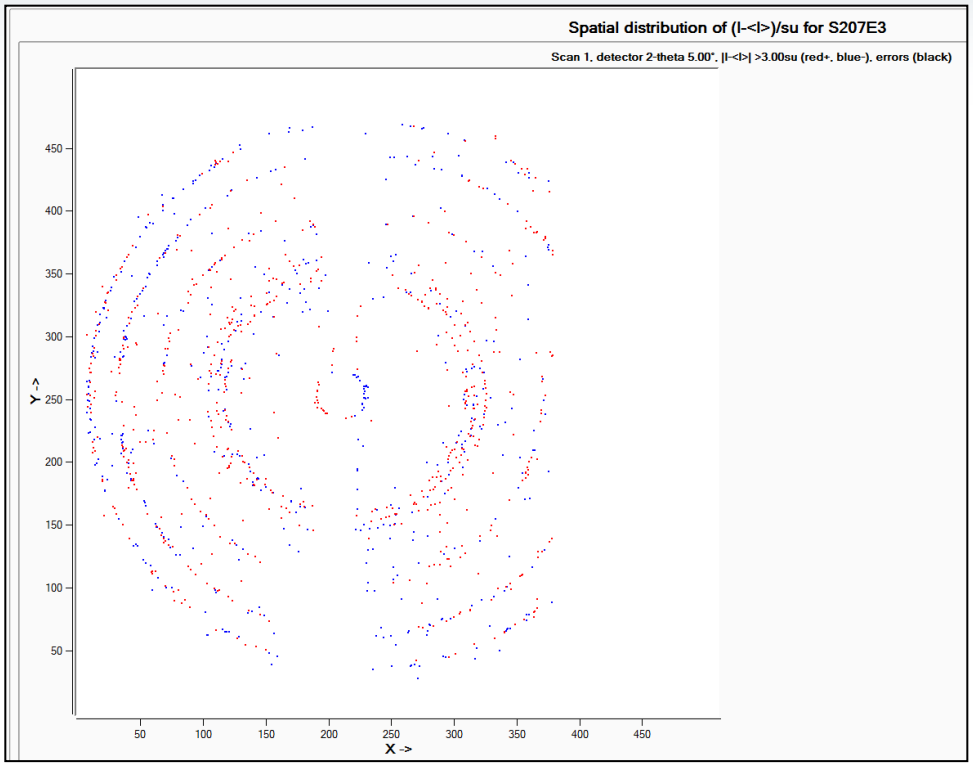


- Upper graph: scale factors *versus* frames and runs. Big variation are due to different illuminated volume.
- Bottom graph: R_{int} *versus* frames and run.

SADABS
Plots



χ^2 *versus* resolution and intensity. It should be closer to 1.



Outliers relative to detector area for each different 2θ angle. Show bad pixels, shadows, ice rings...

SADABS
Output

Output files in the *work* subdirectory

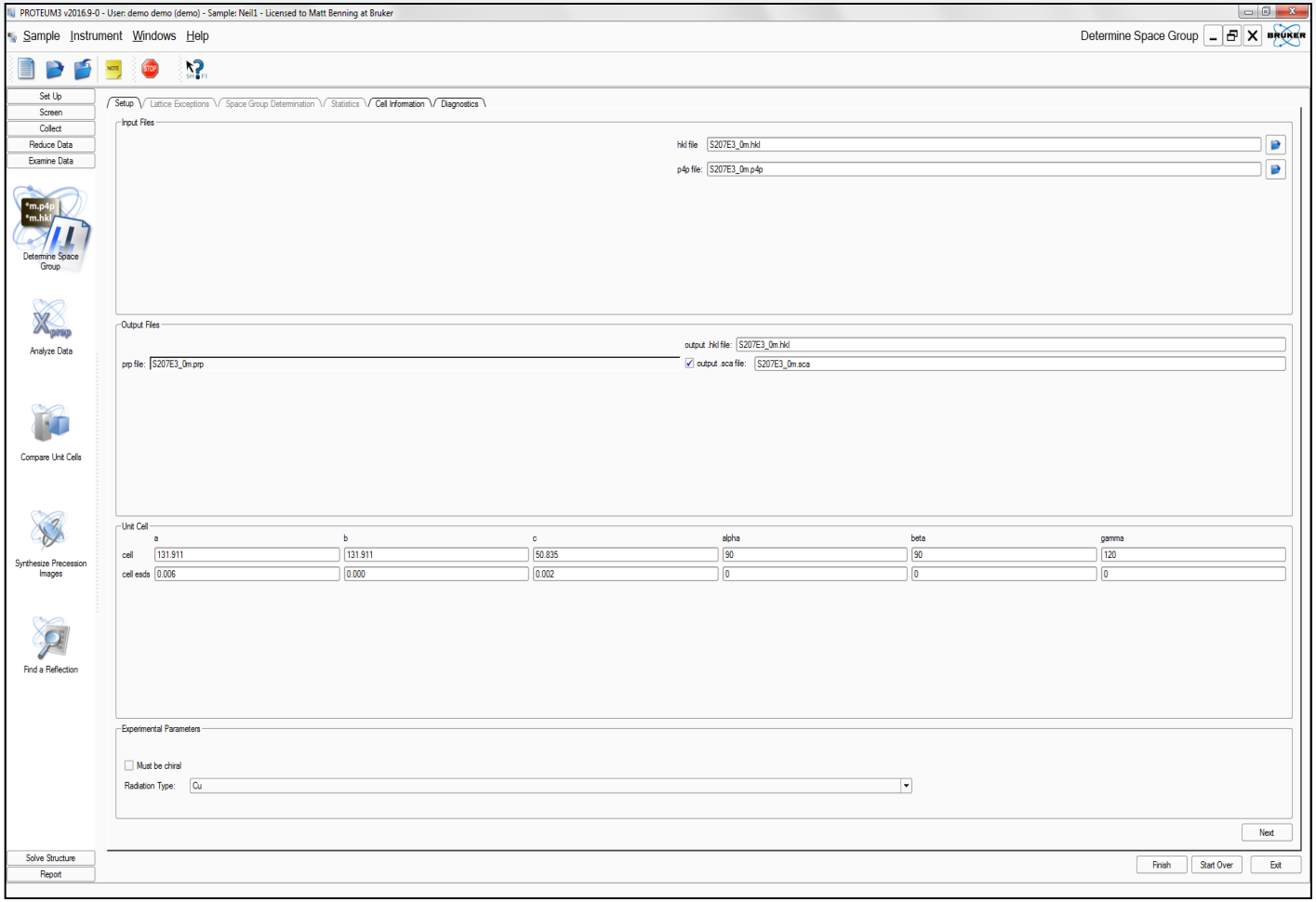
Output Files	Extension	Description
Scaled Intensities	*.hkl	File contains the scaled, unmerged intensities in SHELX HKLF4 format
Log	*.abs	Log file from SADABS

XPREP

- Space group determination and data statistics are carried out with the software XPREP.
- Steps during space group determination:
  - Determine metric symmetry and lattice group
  - Determine Laue symmetry (R_{int})
  - Find systematic absences

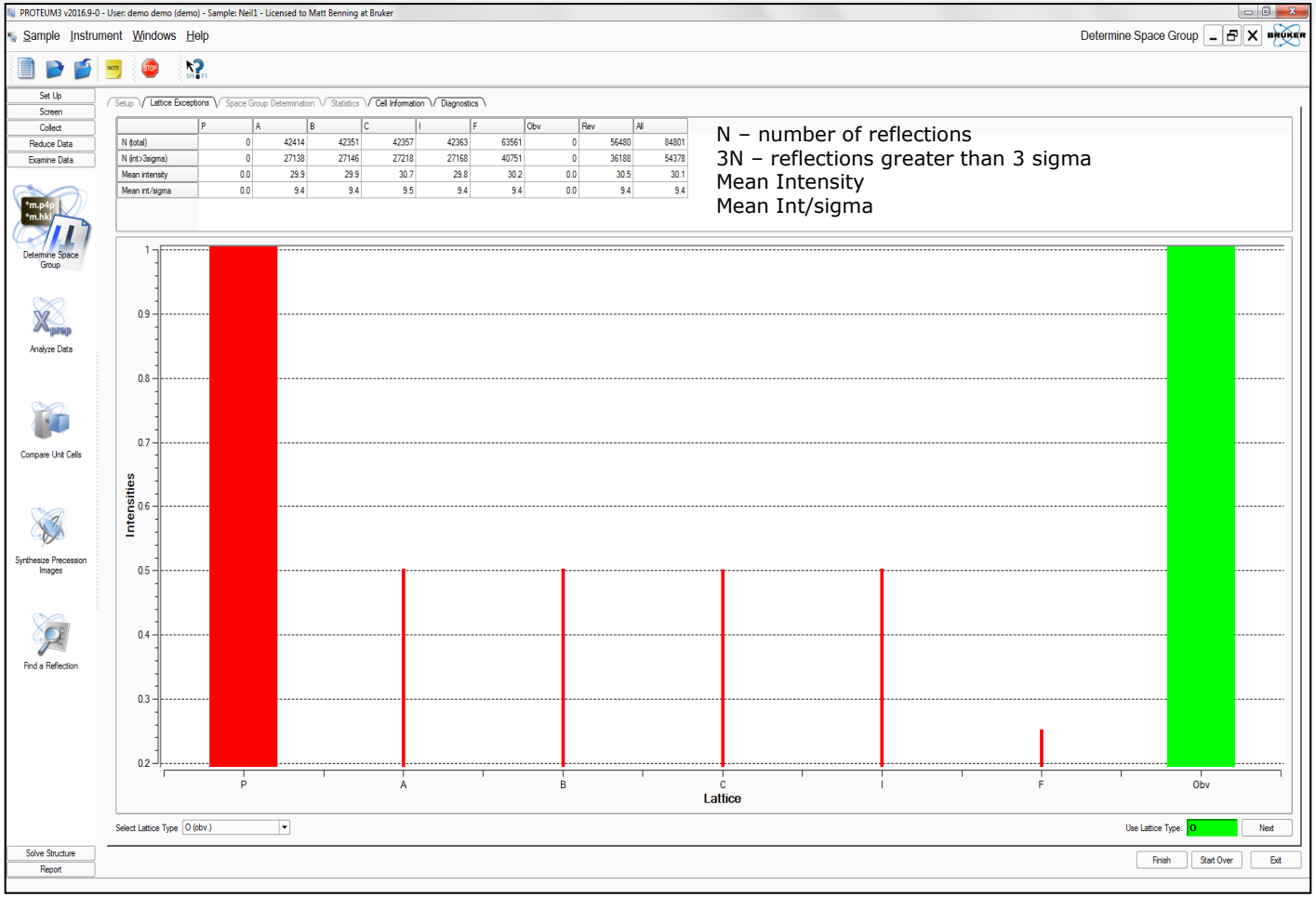
- XPREP can also be used to calculate statistics, calculate anomalous signal, merged data, prepare files for ShelxD...

XPREP
Space Groups and Statistics



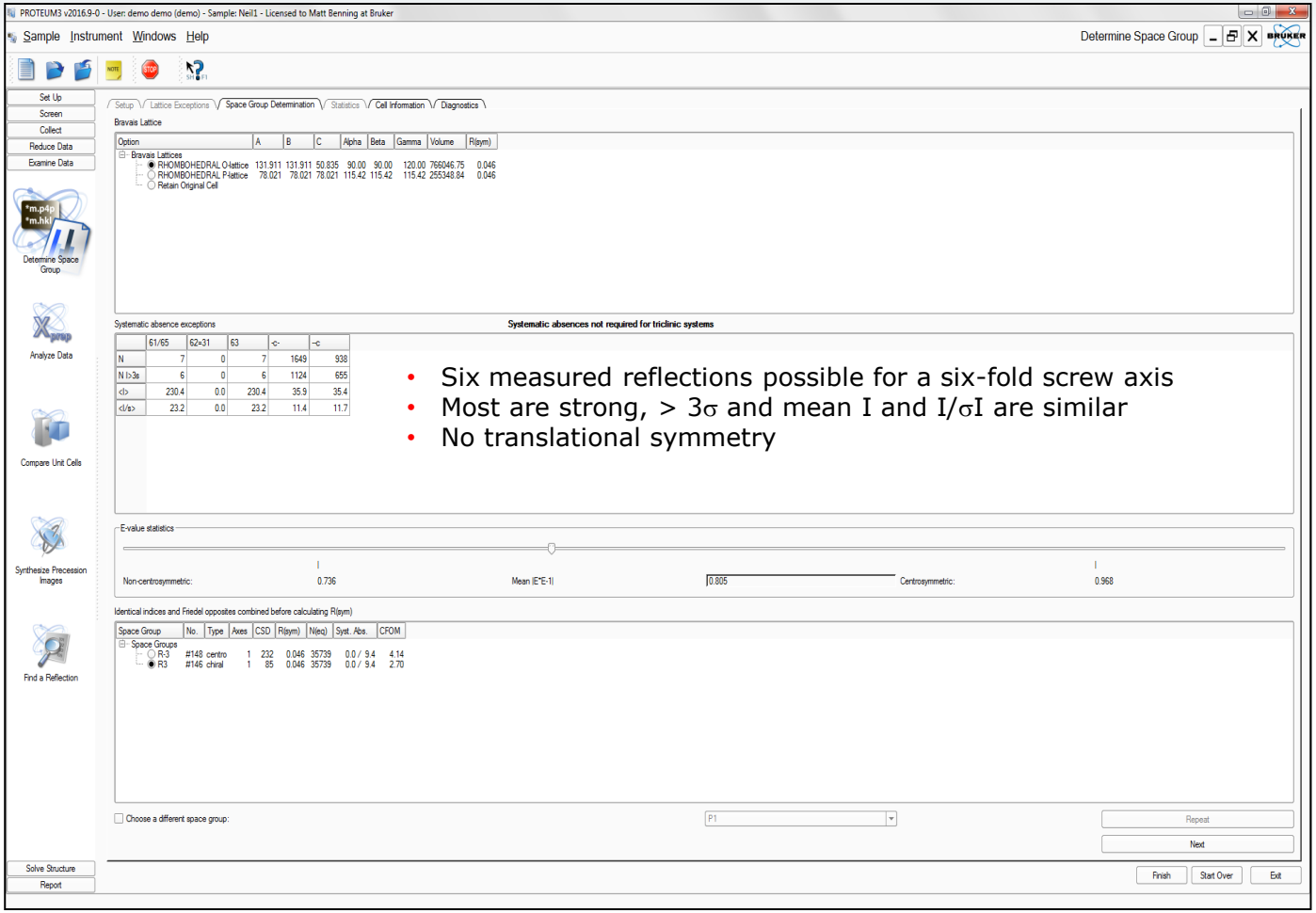
- Most of the information comes from the Database
- Can output a scalepack in addition to SHELX HKL

XPREP
Space Groups and Statistics



- Find the correct metric symmetry (correct lattice type) by checking systematic absences

XPREP
Space Groups and Statistics

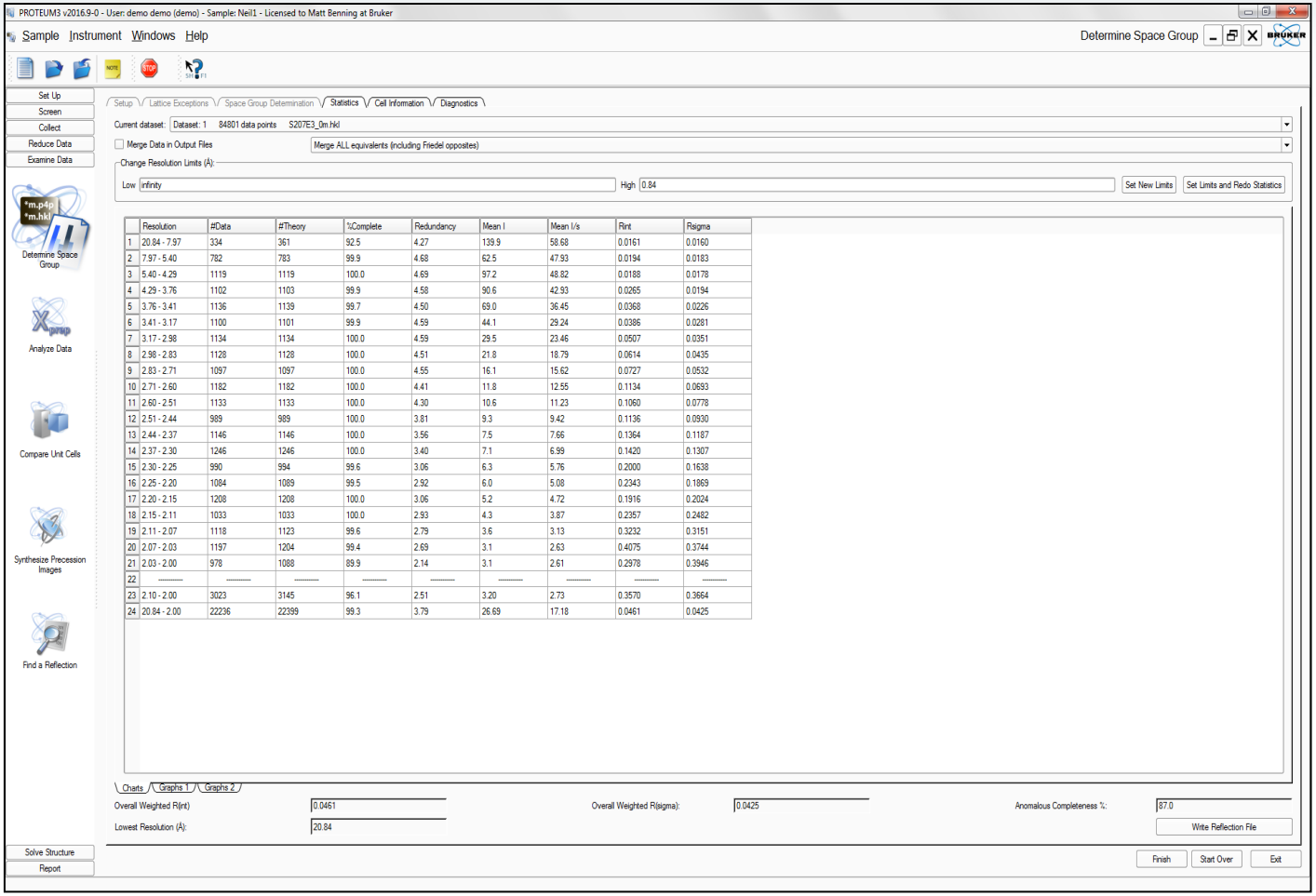


- Six measured reflections possible for a six-fold screw axis
- Most are strong, > 3σ and mean I and I/σI are similar
- No translational symmetry

- Find translational symmetry by looking at the potential systematic absences
- Will only have Screw axis for protein crystals

XPREP

Space Groups and Statistics



	Resolution	#Data	#Theory	%Complete	Redundancy	Mean I	Mean I/s	Rint	Rsigma
1	20.84 - 7.97	334	361	92.5	4.27	139.9	58.68	0.0161	0.0160
2	7.97 - 5.40	782	783	99.9	4.68	62.5	47.93	0.0194	0.0183
3	5.40 - 4.29	1119	1119	100.0	4.69	97.2	48.82	0.0188	0.0178
4	4.29 - 3.76	1102	1103	99.9	4.58	90.6	42.93	0.0265	0.0194
5	3.76 - 3.41	1136	1139	99.7	4.50	69.0	36.45	0.0368	0.0226
6	3.41 - 3.17	1100	1101	99.9	4.59	44.1	29.24	0.0386	0.0281
7	3.17 - 2.98	1134	1134	100.0	4.59	29.5	23.46	0.0507	0.0351
8	2.98 - 2.83	1128	1128	100.0	4.51	21.8	18.79	0.0614	0.0435
9	2.83 - 2.71	1097	1097	100.0	4.55	16.1	15.62	0.0727	0.0532
10	2.71 - 2.60	1182	1182	100.0	4.41	11.8	12.55	0.1134	0.0693
11	2.60 - 2.51	1133	1133	100.0	4.30	10.6	11.23	0.1060	0.0778
12	2.51 - 2.44	989	989	100.0	3.81	9.3	9.42	0.1136	0.0930
13	2.44 - 2.37	1146	1146	100.0	3.56	7.5	7.66	0.1364	0.1187
14	2.37 - 2.30	1246	1246	100.0	3.40	7.1	6.99	0.1420	0.1307
15	2.30 - 2.25	990	994	99.6	3.06	6.3	5.76	0.2000	0.1638
16	2.25 - 2.20	1084	1089	99.5	2.92	6.0	5.08	0.2343	0.1869
17	2.20 - 2.15	1208	1208	100.0	3.06	5.2	4.72	0.1916	0.2024
18	2.15 - 2.11	1033	1033	100.0	2.93	4.3	3.87	0.2357	0.2482
19	2.11 - 2.07	1118	1123	99.6	2.79	3.6	3.13	0.3232	0.3151
20	2.07 - 2.03	1197	1204	99.4	2.69	3.1	2.63	0.4075	0.3744
21	2.03 - 2.00	978	1088	89.9	2.14	3.1	2.61	0.2978	0.3946
22	----------	----------	----------	----------	----------	----------	----------	----------	----------
23	2.10 - 2.00	3023	3145	96.1	2.51	3.20	2.73	0.3570	0.3664
24	20.84 - 2.00	22236	22399	99.3	3.79	26.69	17.18	0.0461	0.0425

XPREP
Output

Output files in the *work* subdirectory

Output Files	Extension	Description
Log	*.prp	The file is actively updated as you navigate through XPREP or "Space Groups and Statistics" (PROTEUM's GUI interface for XPREP).
Different file formats		The intensity file output from SADABS (*.hkl) can be converted to other file formats using XPREP. Using the "W" option from the "Read, modify or merge DATASETS" ([D]) menu, you can output the intensities in Scalepack, CNS or X-PLOR formats. You can also output a Scalepack HKL file from "Space Groups and Statistics" by checking the "output .sca file" box.

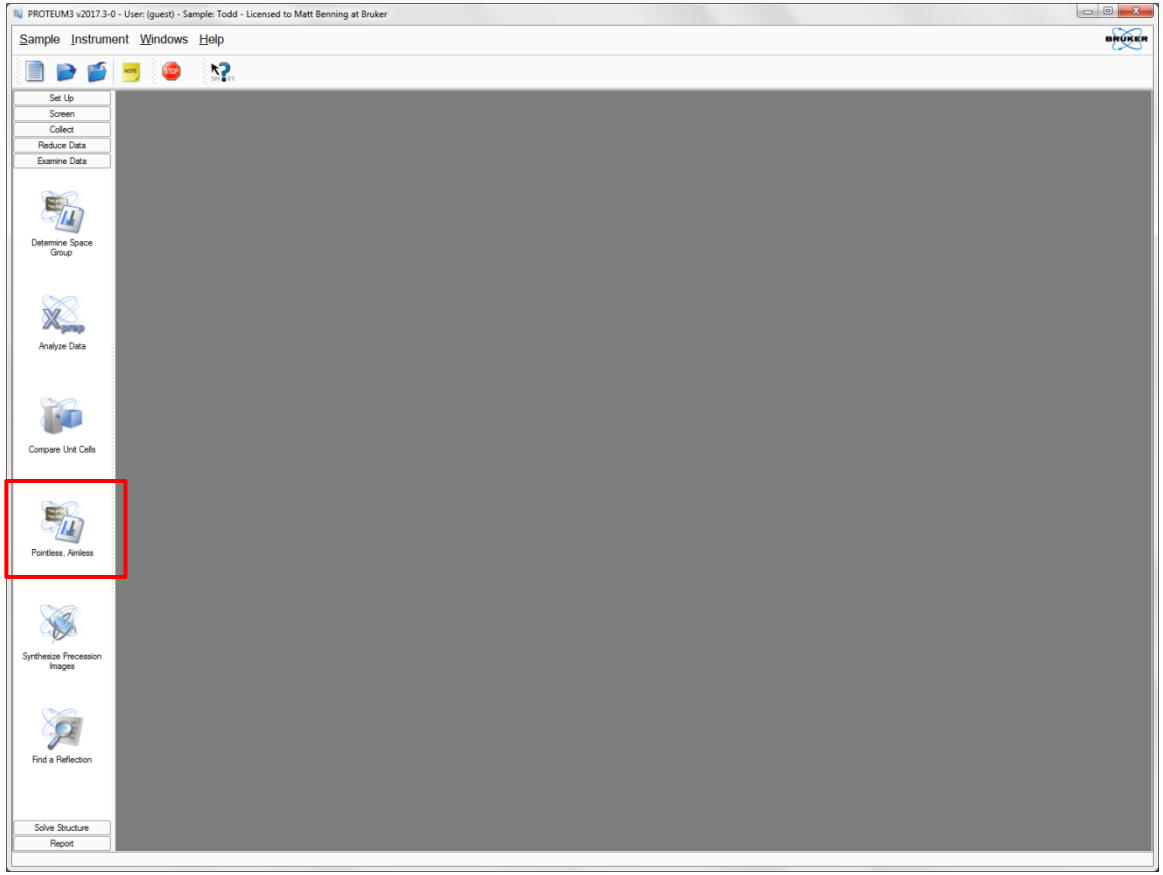
Pointless, Aimless

If you have CCP4 installed, add the following 3 lines to the end of the bn-config.py file

- ccp4 = "C:/CCP4-7/7.0"
- ccp4_range = [22.0,1.85]
- ccp4_autoprocess = True

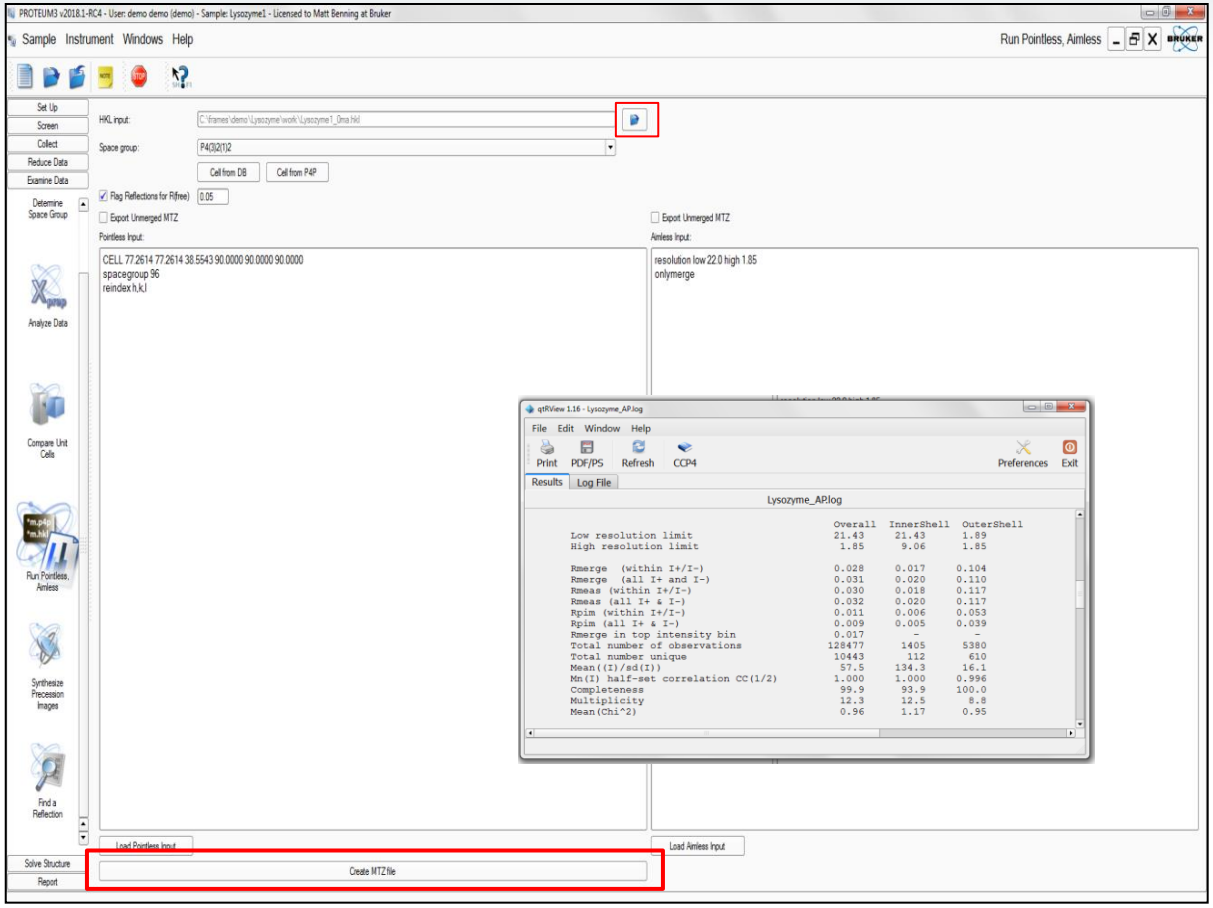
Pointless, Aimless

- Open the “Examine Data” menu
- Select the “Pointless, Aimless” icon



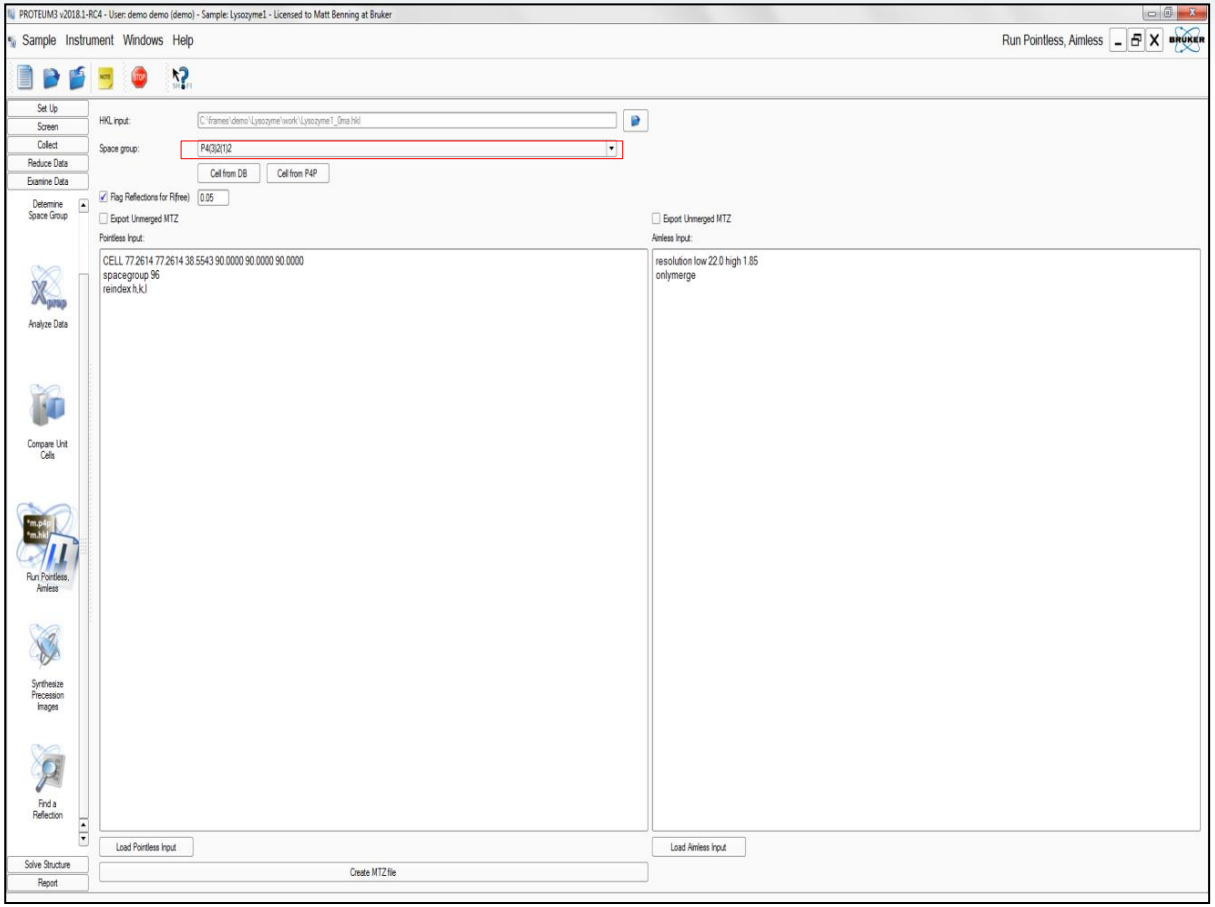
Pointless, Aimless

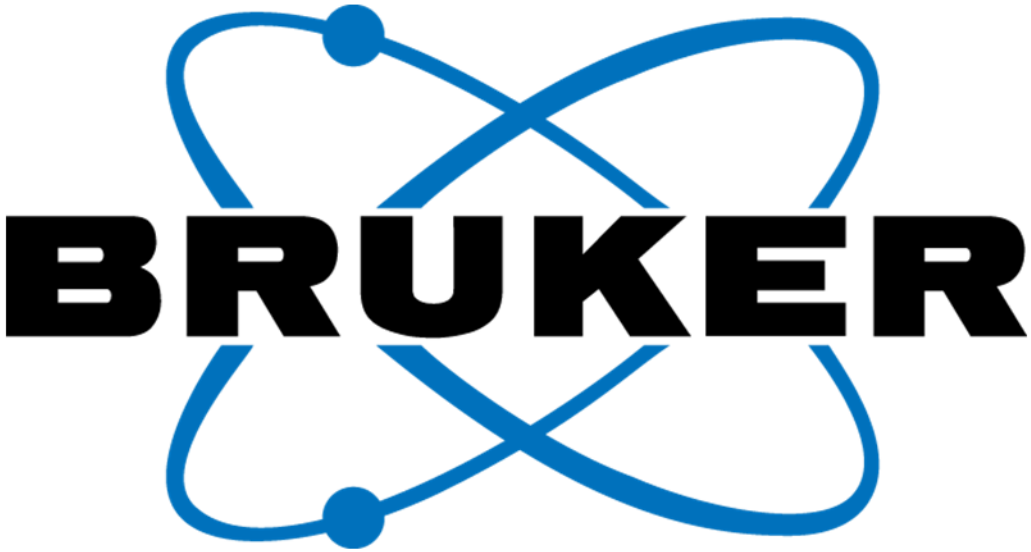
- If there is no MTZ file in the work folder, PROTEUM will automatically run Pointless and Aimless based on default values and display the aimless output.
- Default resolution 25 – 1.85 Å
- The pointless and aimless fields are editable so you can add keywords, change the defaults and click "create MTZ file" at the bottom left to rerun the programs. The new Aimless log will appear when both programs are finished.
- If the space group is not assigned (default), PROTEUM lets pointless perform a space group search.
- The plugin will search for the *HKL filename*_0m.hkl in the work directory but you can also search for a HKL file using the browser button.



Pointless, Aimless

- If you want to assign a space group, select the desired group in the box below the input HKL filename. This will fix the space group to the that group assigned.
- A merged MTZ file is written out by Aimless, if you want to write out a unmerged MTZ file as well, check the "Export Unmerged MTZ"
- Output files are written to the work folder.
  - *Entry prefix*_AP.log is the output logfile from Aimless
  - *HKL filename*_merged.mtz is the merged MTZ file output by Aimless
  - *HKL filename*_umerged.mtz is the corresponding unmerged MTZ





www.bruker.com