# Cryo-EM Data Processing of β-Galactosidase in cryoSPARC 4.7 Tutorial

Workflow Overview:

Import Micrographs

CTF Refinement

Blob Picking

Inspecting particles

Extracting particles

2D Classification

Select 2D classes

Template picking

Inspecting particles

Extracting particles

2D Classification

Select 2D classes

Ab-initio model

Refinement

This workshop introduces the use of CryoSPARC for **exploratory data processing**, a strategy aimed at rapidly identifying, reconstructing, and refining molecular species within heterogeneous samples. While the examples focus on CryoSPARC, the broader concepts and workflows are equally relevant to single-particle analysis with other software tools.

## Introduction: Dashboard, Projects, Workspaces and Jobs

Connection to Cryosparc instance from SV-cluster (*note: only available during the course*)

1. Connect on the EPFL network (VPN)
2. Go to this address: http://izar3.hpc.epfl.ch:39060/
   Log in : stdepfl@gmail.com ||| PASS: stdepfl123

The Dashboard provides at-a-glance information on your Projects, Workspaces and status of Jobs. It also shows the change log for new versions of cryoSPARC. The header and footer contain links to Projects view, Workspaces and the identity of the current user.

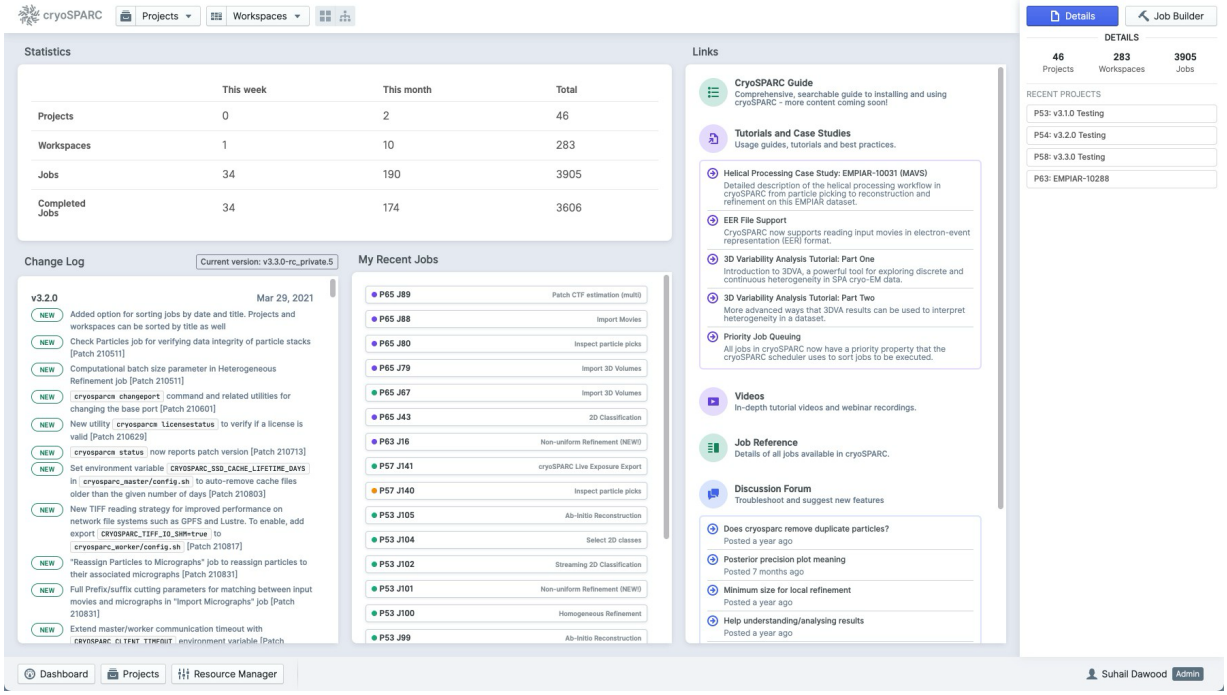
CryoSPARC organizes your workflow by Project, e.g, P1, P2, etc. Projects contain one or more Workspaces, which in turn house Jobs.

Projects are strict divisions. Files and jobs from different projects are stored in dedicated project directories and jobs cannot be connected from one project to another.

For this course, refer to **project P16 BIO-643-2025.**

Workspaces enable the logical separation of jobs and workflows, allowing them to be more easily managed within a large project. Jobs may be connected across workspaces and each job may belong to more than one Workspace.

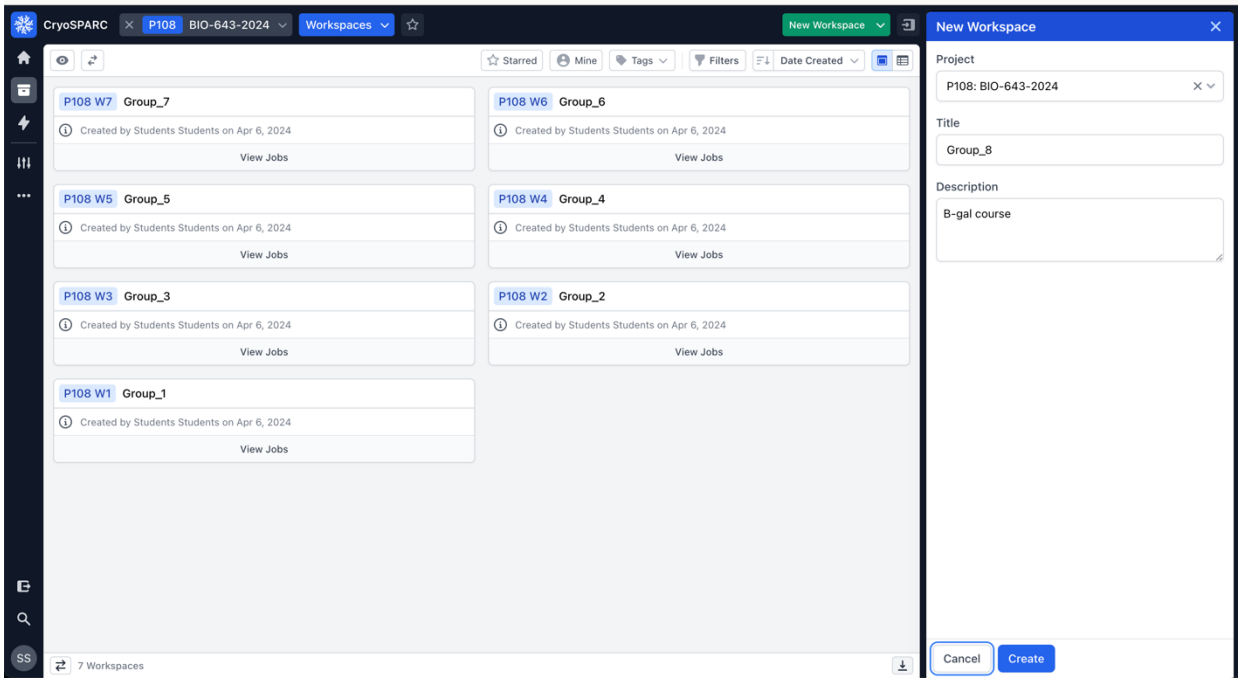
This allows us to properly manage datasets, methods of processing and also for sharing.




*The cryoSPARC user interface*

**Step 1: Create a Workspace** (here it has been done: Group_1to7)
Use Workspaces to organise or separate portions of the cryo-EM workflow for convenience or experimentation.

Alternatively, select the Projects drop-down in the header. This opens a searchable list of all Projects associated with your user account. Select an entry to open the associated project.



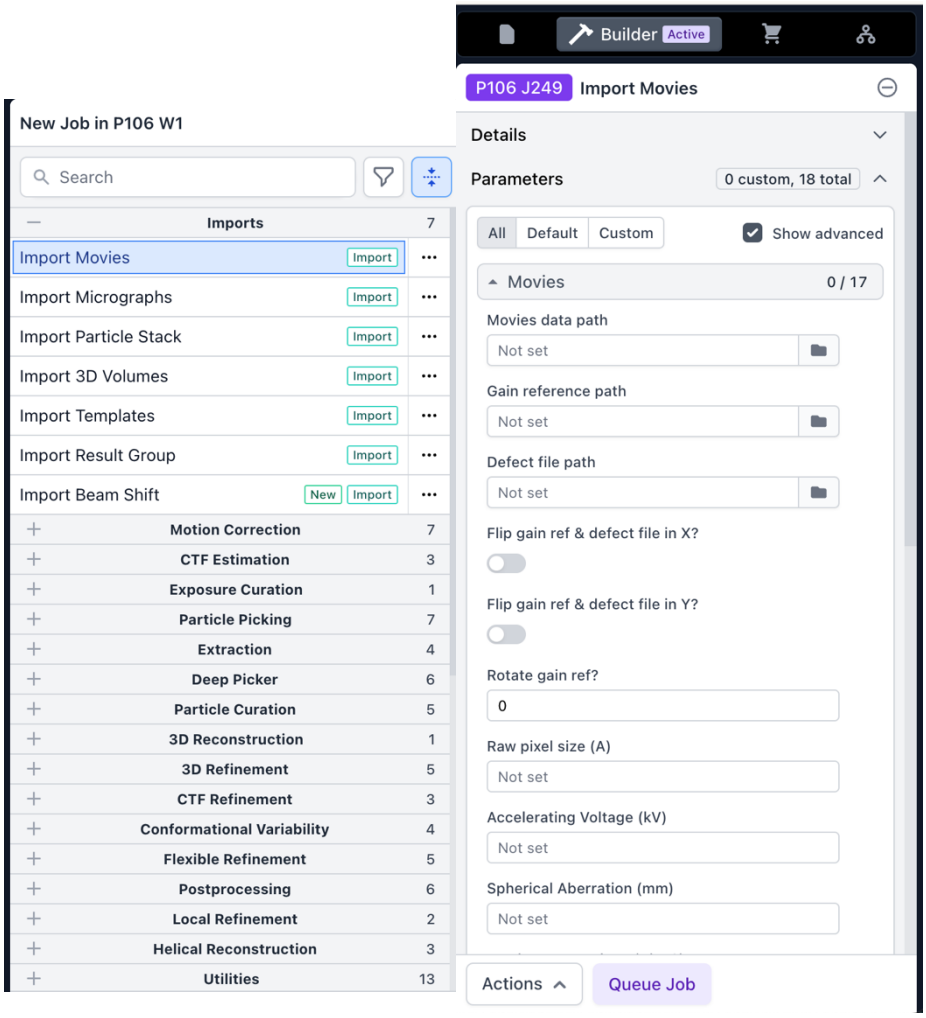

Create a New Workspace with the "+ Add" button in the header or press N on your keyboard. Alternatively, select New Workspace from the Project Details panel on the right side of the screen. Set a Title (may be changed later) and optionally a description.

Click "Create". This will create the new Workspace.

You will be by group of 2-3 people. You can edit one on the group_# and rename it if you wish with your name.
When clicking on the workspace there is a details tile tab on the right of your screen contain workspace information. Change the Title here.

**Step 2: Import Micrographs.**

In cryoSPARC, navigate to the new Workspace. To do so, navigate to the project to see a list of workspaces, and then into the workspace.
Select the Job builder in the right sidebar (🔨). The Job Builder displays all available job types by category (e.g., workflows, imports, motion correction, etc.).




Select the import micrographs job type in the Job builder. This creates a new job within the current Workspace, displayed as a card. By default, new jobs are set to Building status, indicated on the job card in purple. To change parameters, select the Building job and toggle between active or inactive building states with B on your keyboard, or click the "Building" badge on the job card.

Select the data path: Click the file browse icon and select the image files .mrc. To select multiple files, use a wildcard, e.g., *.mrc. This selects all files that match the wildcard expression. The file browser displays the list of selected files along with the number of matches at the bottom. For this tutorial dataset, navigate to the directory where the test data was downloaded.

Movie data path : /work/ptpsp-cryo/yoan/dci-collection/Glacios_PTPSP_Bgal_Data_20220303/100_best_v2/DW/*doseweighted.mrc
Edit Job parameters from the Builder; enter the following parameters:

Pixel size (Å): 0.926
Accelerating voltage (kV): 200
Spherical abberation (mm): 2.7
Total exposure dose (e/Å^2): 20

After changing a parameter, the blue D icon changes to a green S. This indicates the
The parameter is different from its default value.

Launch the job on the MASTER queue node.

**Step 3: CTF Estimation**

Select Patch CTF Estimation (multi) in the Job Builder to create a new job. This job type requires micrographs as the input.

On the main workflow, click on your last Import Micrographs job, then tap your Space Bar. Navigate to Outputs. Drag and drop the imported_micrographs output into the Micrograph placeholder in the Job Builder.

Note in Cryosparc it is very intuitive. Inputs are commonly called the same, and outputs have the same name.

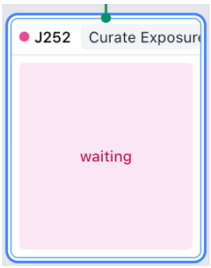
(100 micrographs)

For the queue, Launch the job on the lane called: `izar3-ptpsp-cryo-SSDExclusive_1h`

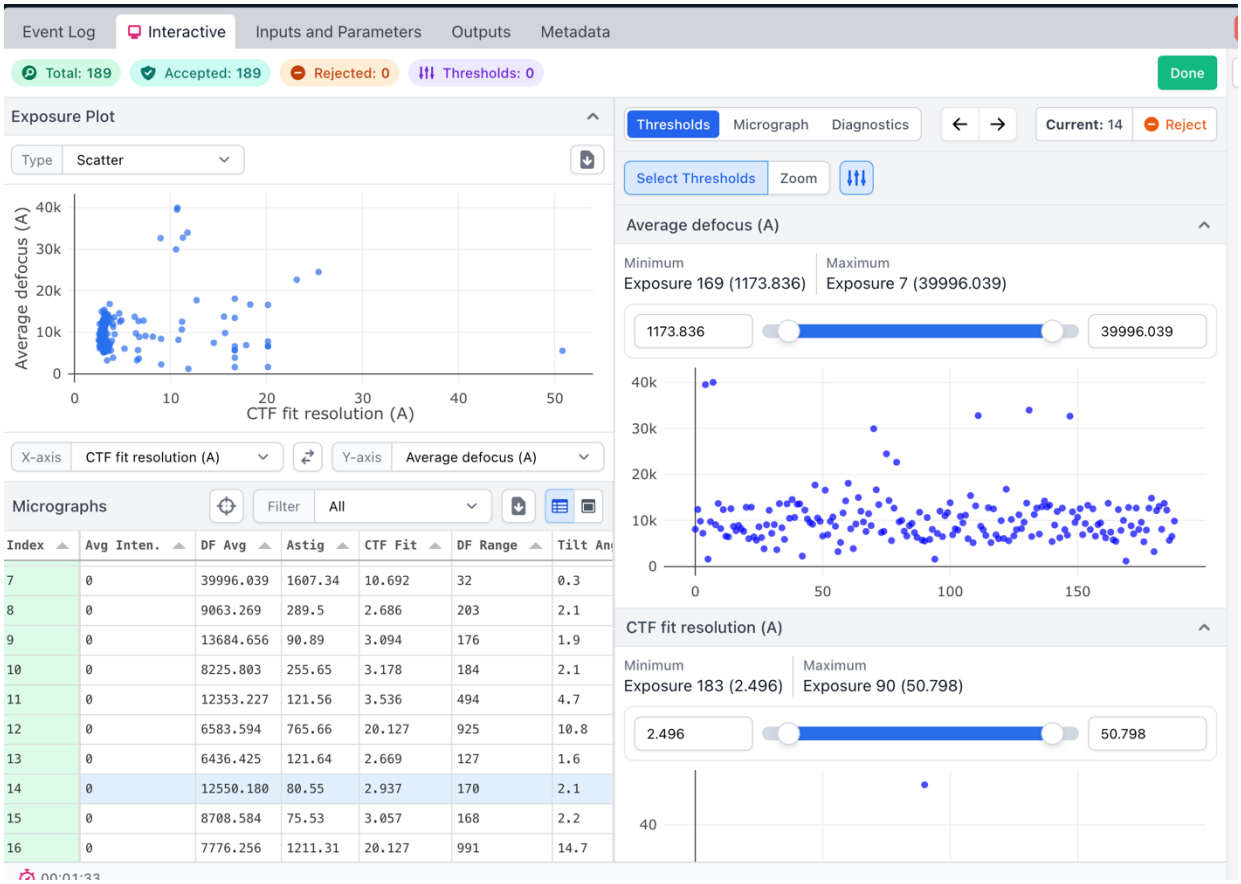
It will take a few minutes to run. During this time, it is good to do some research on β-Galactosidase protein structure.

> Find a PDB model
> Do you have any cryoEM structure known on this protein ?
> Gather information you might find interesting to solve its cryoEM structure. Shape, size, symmetry, Oligomers…

**Step 3a: Curate Exposures**




This is a job that allows us to select the micrographs that we wish to keep based on certain parameters such as defocus, ice thickness etc. The idea is to access the quality of the entire set to discard very bad images. You can do it by selecting the corresponding job under Exposure Curation and the manually Curate Exposure tool. Select the job from the list and drag the input micrographs. Then press queue, and once the job is ready, the tile becomes pink with "waiting ", then press "space bar" to open the GUI interface.




Look at the diagnostics interface. If you press on one of the blue dots you will see the corresponding images. You can screen it according to the maximum resolution of the image, the defocus, ice thickness. Look around and see if there are images you may want to remove from the list. If yes, press the red Reject button. Once you've done, press the green button "done".

*Note: For this tutorial, most images have been selected to be of good quality. But you might explore some diagnostic plots and maybe find some bad one.*

**Step 4: Blob Picking**

We do not have a template for our "unknown sample" so we will have to tell the computer to pick "blobs". There are many solutions out there, but commonly this is a good start to avoid bias.

Select the blob-picking job from the menu and complete the size parameter range with the values you found from PDB structures.

If you wish or you feel it may make sense based on your β-galactosidase structure interpretation, you can change the picking parameters from blob to elliptical or ring and see what's happening. It's fast so play around.

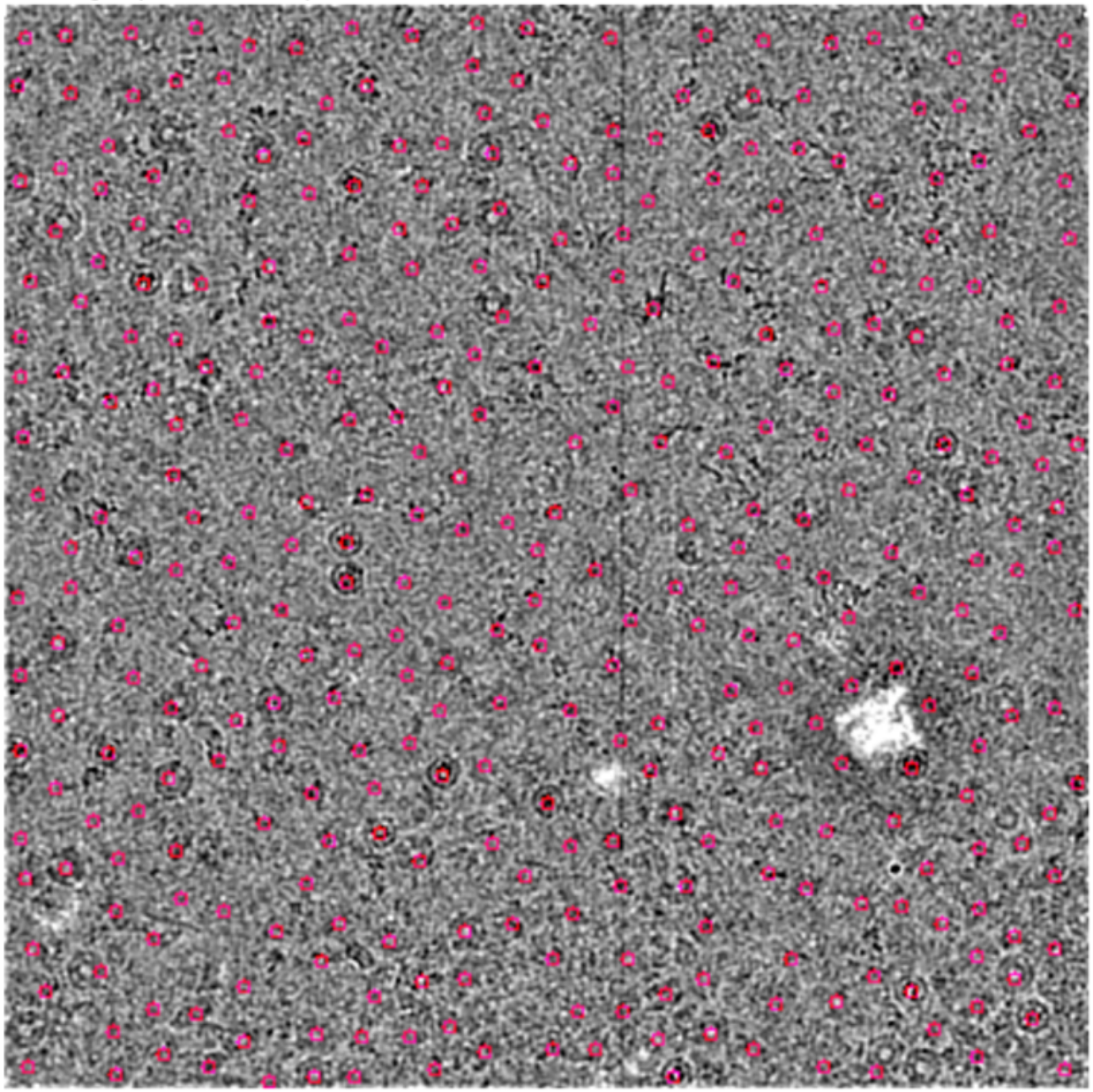
You can change those 2 additional parameters:

Number of micrographs to plot: 50

Maximum number of local maxima:1000

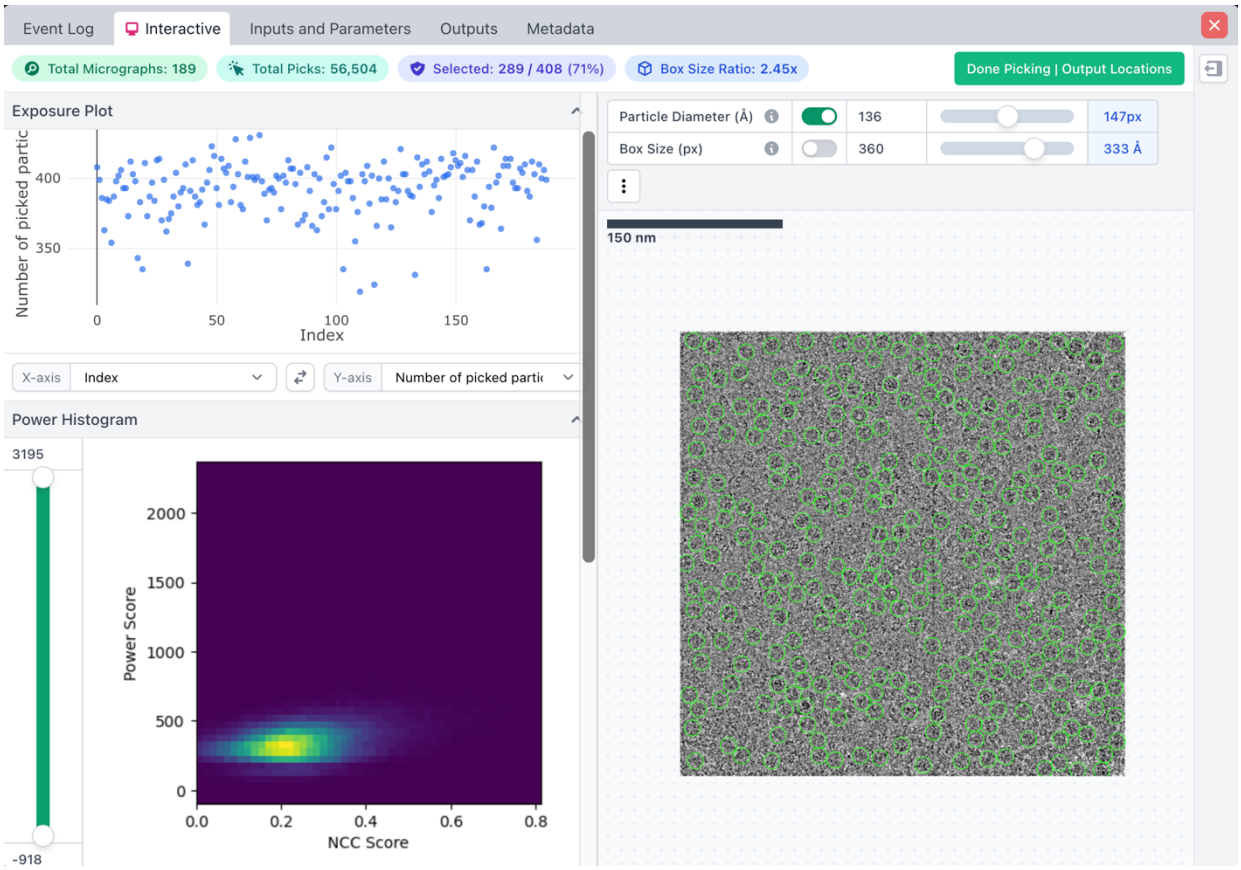
Queue the job on `izar3-ptpsp-cryo-SSDExclusive_1h`

Micrograph J276/imported/014834060009830452735_00023732264751766912



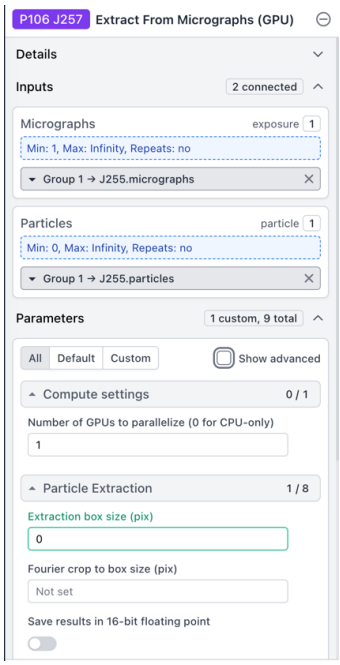
**Step 5: Inspect Picks**
You can inspect your peaking using “inspect particles picks” job.

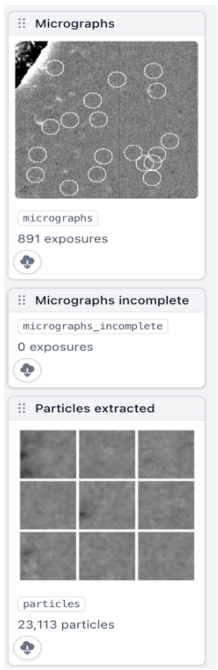
Same that for the previous inspection job, once the tile is pink you can open the GUI interface with space bar and start looking around. This your picking good ? Over or under-picked? You can maybe adjust the quality by playing with the power spectrum in the bottom left panel.
You need to provide some sort of parameters, including NCC score, local power as bellow.
You can change the low pass filter and the pixel box size to view better.
The goal is to remove particles that are bad and keep ones that are good
Press done when finish.



**Step 6: Extract from Micrographs**

This job extracts particles from the respective micrographs.
Select Extract from Micrographs (GPUs).
Open the recently completed Inspect Picks job. Drag and drop both the micrographs and particle outputs into the corresponding inputs on the Job Builder.




In the Job Builder, look under the Particle Extraction section and change the Extraction box size (pix) to 2x the size of the particles you are working with. Attention: you have to convert in pixels ! Look at the values of your input data ☺
Refer to this: https://blake.bcm.edu/emanwiki/EMAN2/BoxSize to set a proper box (makes calculation way faster).

Change the Fourier Crop to box by half of the value you put on the extraction box size. This will bin by 2 the data to make calculation faster also.

Once the job is completed, you'll notice the number of resulting particles could be less than the input; this is due to the fact that the particle extraction process excludes picks that are too close together, or near the edges of the micrographs

**Step 7: 2D Classification**

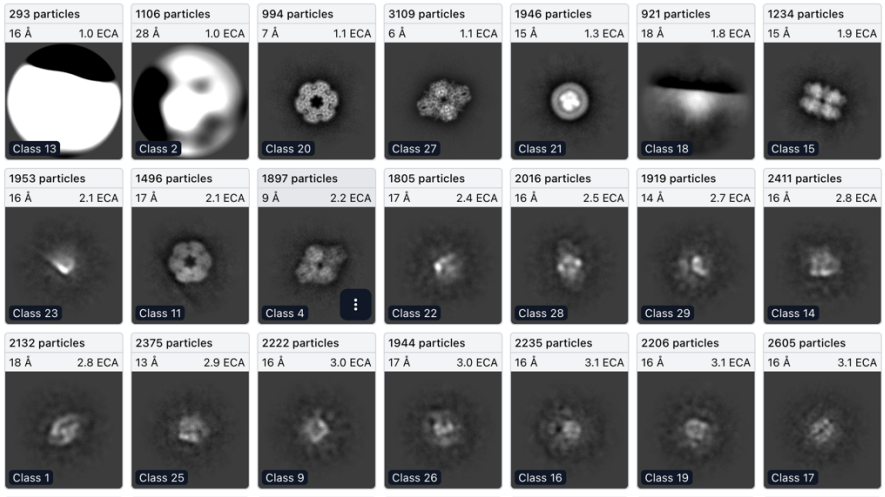
Select 2D Classification from the Job Builder.
Drag and drop the particle output from the previously completed Extract from Micrographs, into the input and queue the job.

The only thing you need to modify is (Number of 2D classes :25) you can see this option by clicking on the Advanced toggle on the top right below BUILDING)

**Step 8: Select 2D Classes**

Select Select 2D Classes from the Job Builder.
Drag and drop both the particles and class_averages outputs from the most recently completed 2D Classification job

Interactive Select 2D job depicting good quality classes selected.




You can see that not all the classes are good. Pick the best classes you think are good. Be careful, you can bias what you will obtain later!

Once clicking 'Done', the job will generate outputs for each group of classes and particles, one for the selected set and another for the excluded set:

*Note: Usually, it is good to resubmit the good and bad particles to a second round of classification, to make sure during alignment and re-orientation steps nothing has been missed.*

We will use the good classes for Template picking to pick. This is also so-called reference-based particle picking. It could lead to bias (or Einstein from white noise) when used as the first and over picked signal from a random background.

We will repeat the same steps as above (From Steps 4 to 8, but instead of Blob picking in Step 4 we will use Template picking building option)

**Step 4b: Template picker**

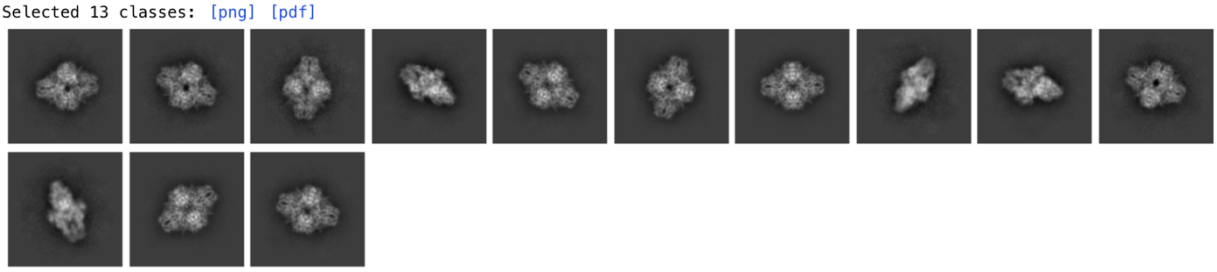
Using your newly selected 2D classes, drop them into a new Template picker job as templates. For micrographs, we will start from the beginning so use the micrographs from the output of your Patch CTF job.
Define the diameters you want for the picking. Then keep the other value as default.

Continue as **Steps 5b, 6b, 7b, 8b** as previously, but with each job's outputs. You can try

to change parameters if you wish.

At this point we will have Blob picked, 2D classified, and selected the best initial 2D classes. From these 2D classes we have improved particle picking by using them as templates for the template picker.

With our now final set of particles selected based upon their 2D classes we will perform an ab-initio 3D reconstruction.



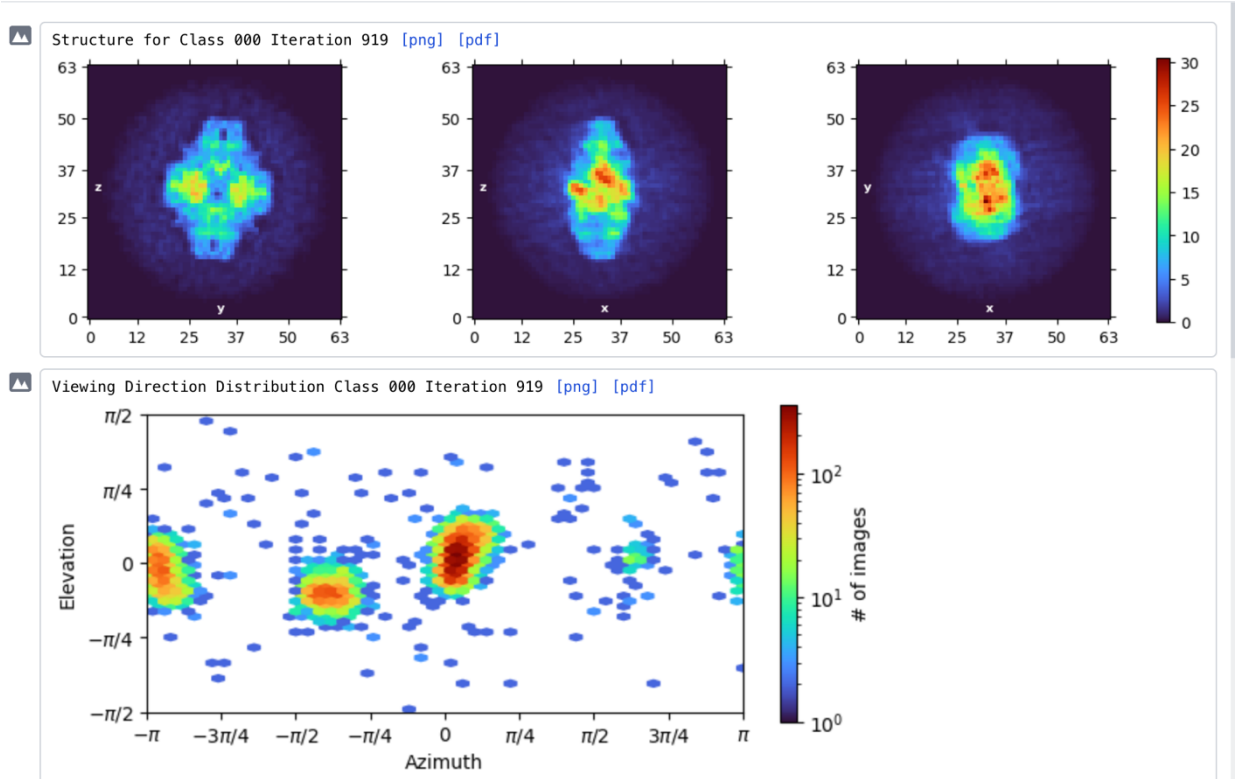

**Step 9: Ab-initio Reconstruction**

Select Ab-initio Reconstruction from the Job Builder.

Drag and drop the particles_selected output from the most recently completed Select 2D classes job (classes selected from the result of the template picker) into the Particle stacks input in the Job Builder.

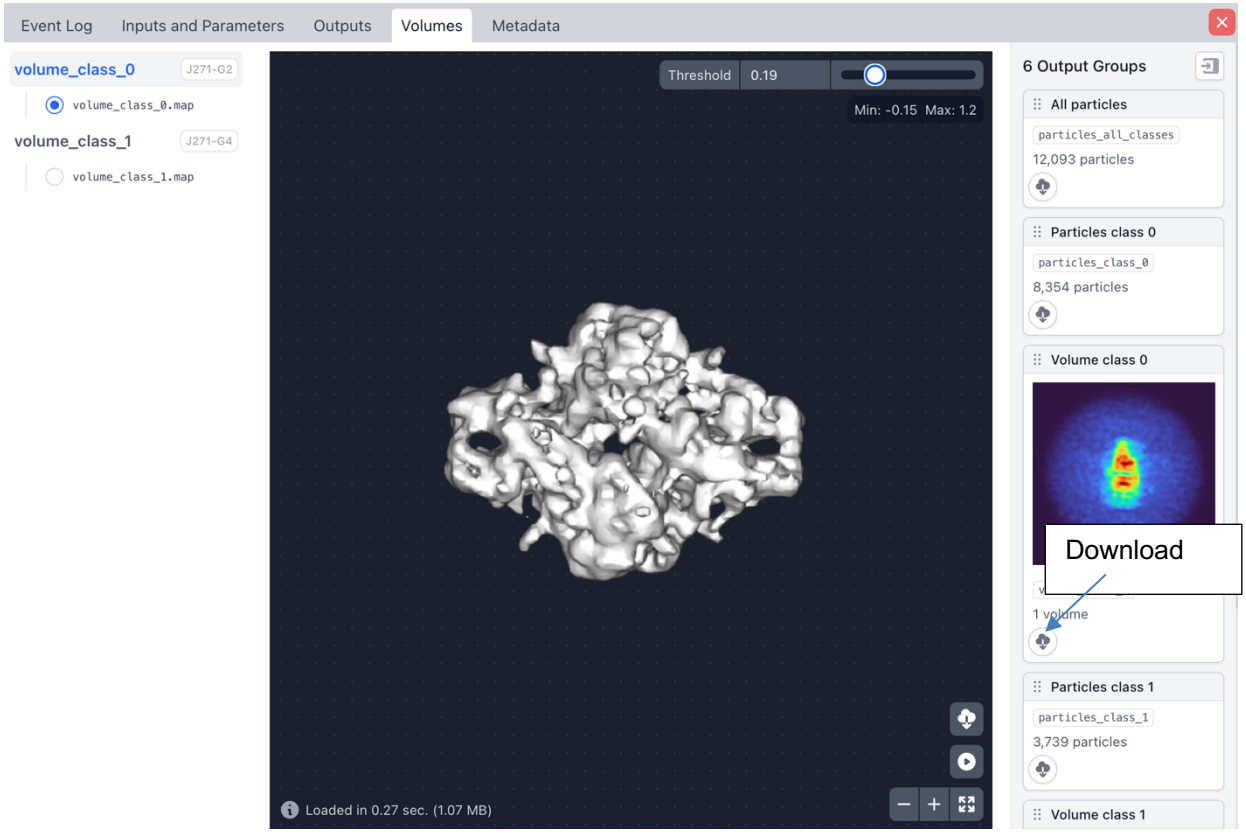
Set the number of classes by 2 if the total number of particles selected is above (5.000). That will help to clean up bad particles into a 3D volume that doesn't contain all the same features. That is a nice tip to know.

Do you know if B-gal as a symmetry or, do you see one?




Queue the job. Results appear in real-time in the stream log as iterations progress. Ab-initio reconstruction should resolve the Beta-Gal structure to a coarse resolution.

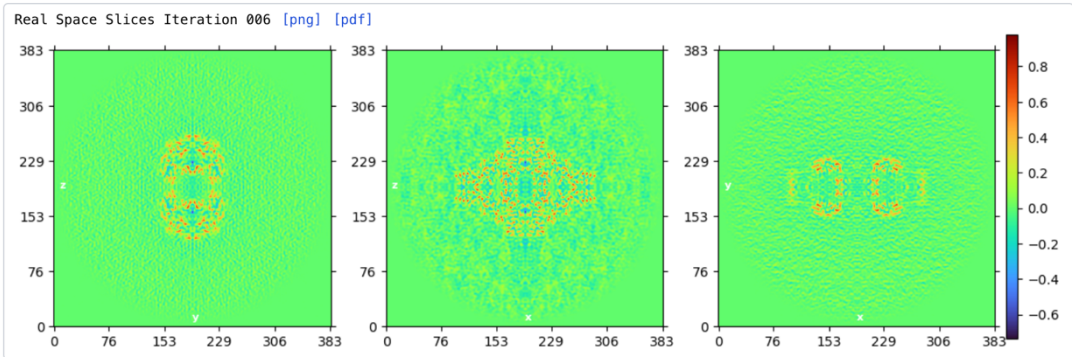
The Viewing direction distribution graph is showing you in a 2D plan all the 3D view you have of your sample. White means 0 view (and therefore 0 particles), blue from 1 to 10, then green then red >100. Do you think you have enough?
You can go to the volume viewer tab and loot at your volume (play with the Threshold button). Or click on the cloud to download the volume and open it by chimeraX for example.

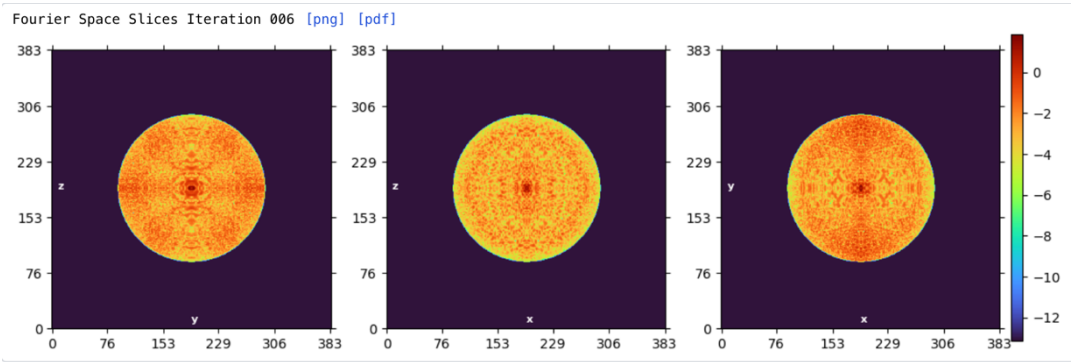


## Step 10: Homogeneous Refinement

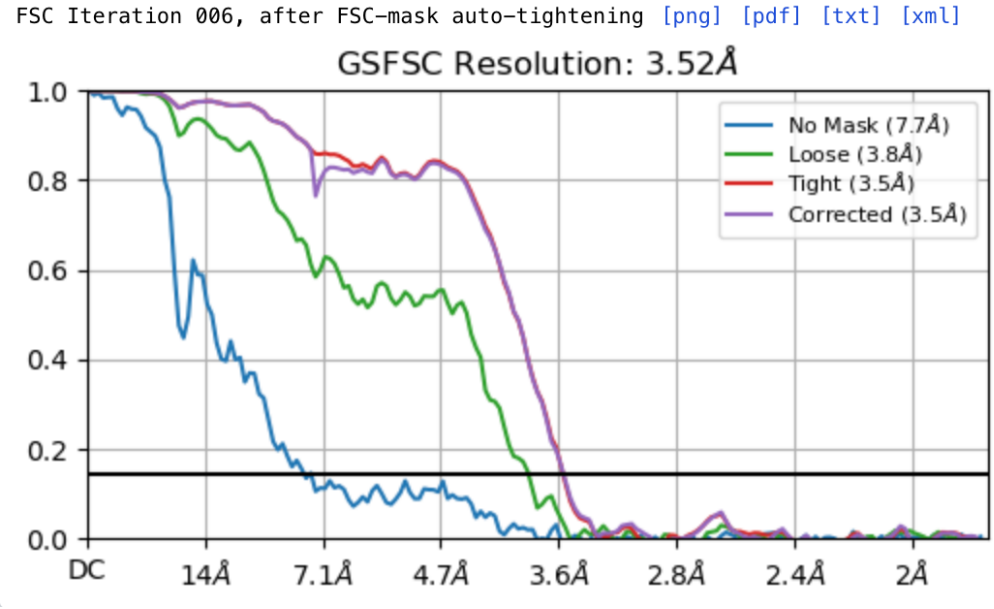
Select Homogeneous Refinement from the Job Builder.
Drag and drop both the particles_all_classes and volume from the recently completed Ab-initio Reconstruction into the Particle Stacks and Initial Volume inputs, respectively. Set the following parameter:
Symmetry: Find the B-gal symmetry or try the one you feel and see the differences.

Queue the job. Results appear in real time in the stream log. The refinement job performs a rapid gold-standard refinement using the branch-and-bound algorithm. The job displays the current the resolution and other diagnostic information for each iteration.



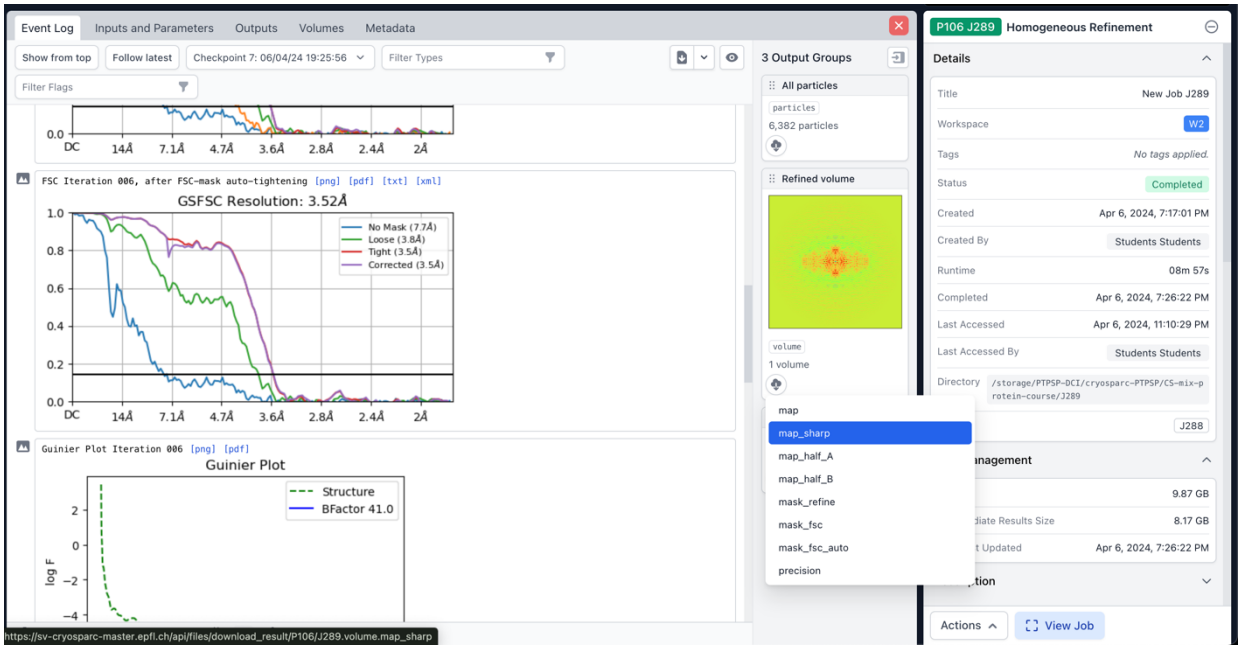
Event log of the completed Homogeneous Refinement job




FSC plot for the final refinement iteration
Once complete, download the volume and/or mask directly from the Outputs section on the right-hand side: Select the drop-down to choose the outputs you wish to download.

A refinement job outputs a map_sharp, the final refined volume with automatic B-factor sharpening applied and filtered to the estimated FSC resolution.
Your resolution will depend on many factors. In this case because we are using the same data, principally it will be due to the number of particles. If you are achieving higher resolution, try going back to your 2D classes and selecting a lower (or poorer) set of particles. If you are at lower resolution already, go back to your 2D classes and add in some classes you previously rejected.

Some helpful notes :
-Everyone will get different results from the same data.
-You results depend on the particles you use
-This then depends on how you selected, picked, and classified your particles




Within the dob details dialog, output groups listed have a download menu with various options.

What resolution can you archive? Can you confidently build a model into it ? And if yes, where do you start?

To go further:

You might want to optimize the data quality and polishing. Try reclassifying the 2D particles from your latest model on your own — you might identify some bad particles.

You could also try re-extracting after 3D alignments with no binning to see if the resolution improves. It might also help to test a smaller or larger box size.

Try to download the final map and open it with ChimeraX softaware. Download a PDB file of beta-gal and see if you can it the model into the Coloumb density.