



National Center for CryoEM Access and Training

cryoSPARC tutorial

Created: 04/13/2020, Updated: 07/15/2020

Software requirements

Required : Web browser (Chrome is recommended by cryoSPARC developers)

Recommended : UCSF Chimera for visualization (<https://www.cgl.ucsf.edu/chimera/>)

Tutorial Overview

Background: cryoSPARC is a software package for processing single particle cryo-EM data that is accessed through a web browser. More information, including installation instructions for your own workstation/cluster, advanced tutorials, and discussion boards can be found at www.cryosparc.com

Aim: This tutorial will take you through processing a small test dataset that easily reaches high resolution in order to introduce you to the basics of processing single particle cryo-EM data.

This tutorial is broken up into different sections. Each will begin with a small amount of background information followed by a step-by-step guide.

- **Step #:** this symbol will indicate the individual steps of the tutorial. These are the action items you will follow.
 - This symbol will contain specific details about that Step, explaining why specific actions were taken, for example the meaning behind different parameters, and when/if to change them from the values we used.

AWS Instance Information

Instance Type: p2.8xlarge

Number of GPUs: 8

RAM: 488

Filesystem: efs

Path to projects: /efs/projects

Path to test dataset: /test_data/Movies

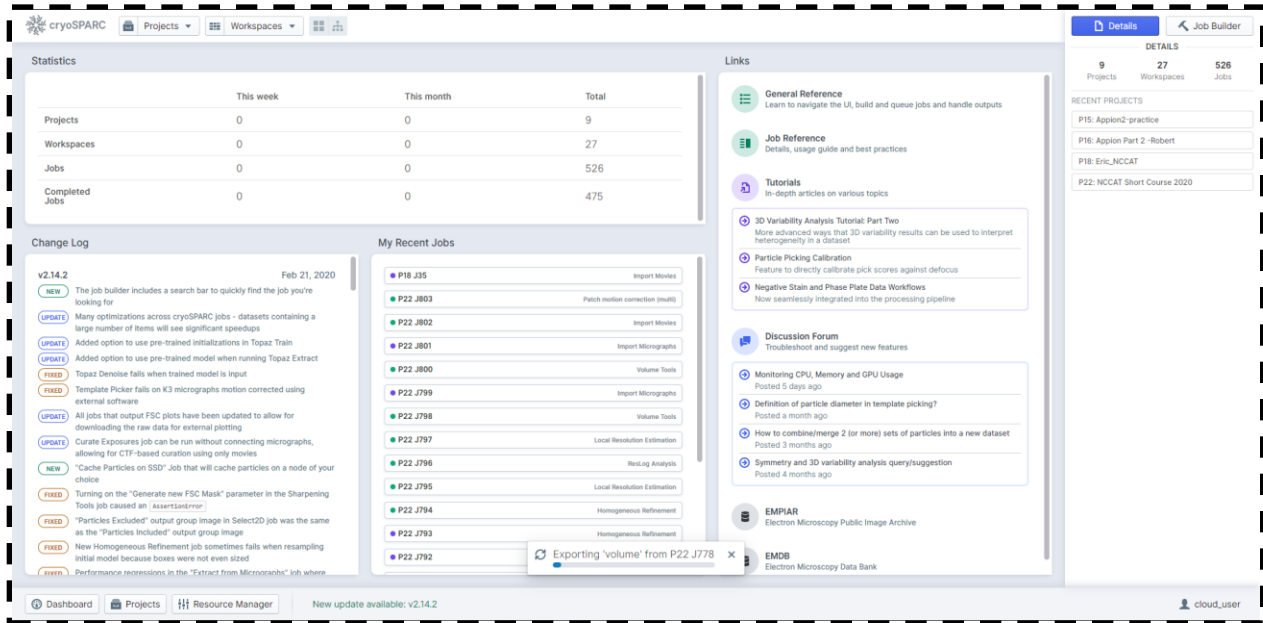
Accessing cryoSPARC

- **Step 1:** In your web browser, enter the cryoSPARC URL from your Stion dashboard
- **Step 2:** Sign in with the following information:
 - Email: Provided on the dashboard
 - Password: Provided on the dashboard

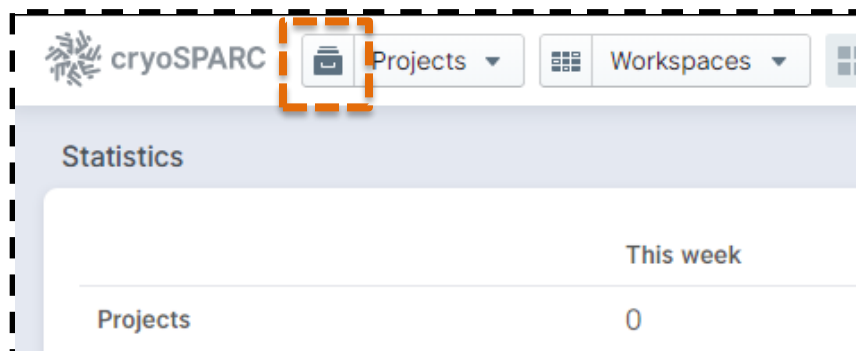
Creating a Project and Workspace

Background: The general processing hierarchy in cryoSPARC is Project -> Workspace -> Job. Projects are often used for specific samples or biological systems, with Workspaces usually being separate datasets. Jobs are the basic processing operations of cryoSPARC which exist within different Workspaces. Jobs can be moved or linked to different Workspaces as the user wishes.

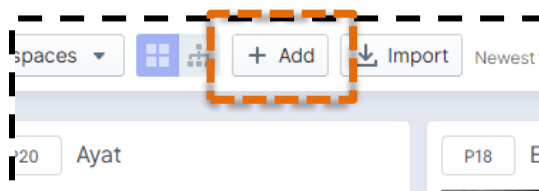
After logging in, you will be greeted by the cryoSPARC homepage. It contains information about changes in recent versions of the program, job history, and helpful links.



- **Step 3:** Click the button next to the word Projects in the top left corner
 - Clicking the word Projects itself will bring up a dropdown menu of projects you have created already. The same is true for the Workspaces button to the right.

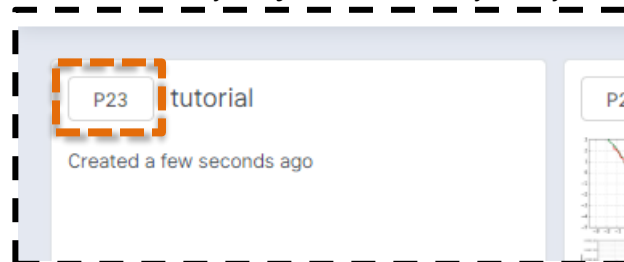


- **Step 4:** Click the "+ Add" button



- **Step 5:** Title your project
 - This is for your use only. cryoSPARC will assign this project a number, e.g. P1, and use that as the name for the project directory
- **Step 6:** Enter “/efs/projects” into the middle field. Optionally, click the folder button to the right and navigate to the directory using the GUI, then click “Create”
 - This is the directory where cryoSPARC will output results. It should be readable and writable, and contain enough space to comfortably store all processing results (several TB at least).

- **Step 7:** Click the “P##” button, located to the left of your project title to open the Project
 - The ## will depend on how many Projects are already in cryoSPARC-



- **Step 8:** From inside the Project you just created, again click the “+ Add” button and title your Workspace. After you click “Create”, click on the “P##-W##” button to open your Workspace.
 - Workspaces can be whatever you want - your individual data collections of a specific protein, imports of the same dataset from different preprocessing programs, etc.



Movie Preprocessing

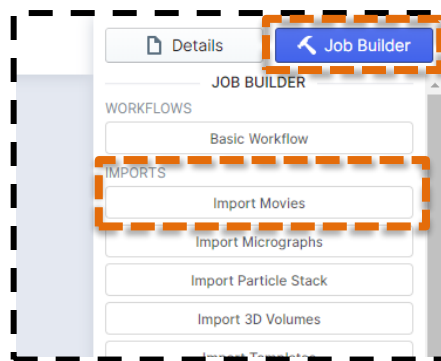
Background : This section will cover preprocessing of the data, focusing on three cryoSPARC jobs: "Import Movies", "Patch motion correction", and "Patch CTF estimation".

"Import Movies" will bring our raw movie frames into cryoSPARC while allowing us to specify certain data collection parameters while performing gain correction.

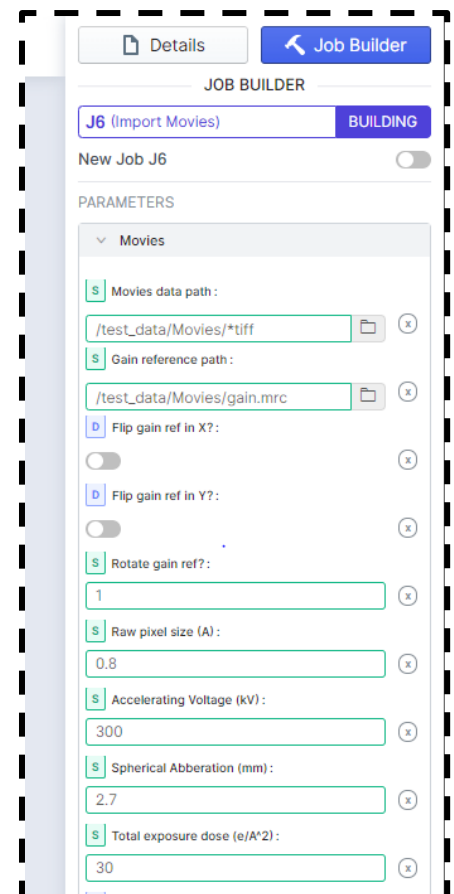
"Patch motion correction" will then take those raw movie frames and align them to account for sample and stage movement during the exposure before producing an aligned average called a "micrograph."

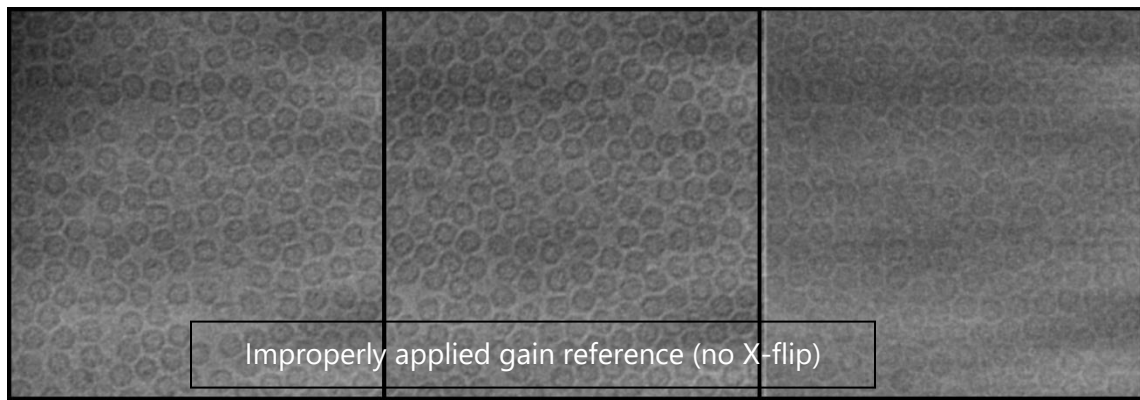
"Patch CTF estimation" will use the output from Patch motion correction and attempt to measure additional parameters that vary from one micrograph to another; astigmatism, defocus, estimated resolution, etc.

- **Step 9:** Click the "Job Builder" button, then click the "Import Movies" button, found under the Imports section."

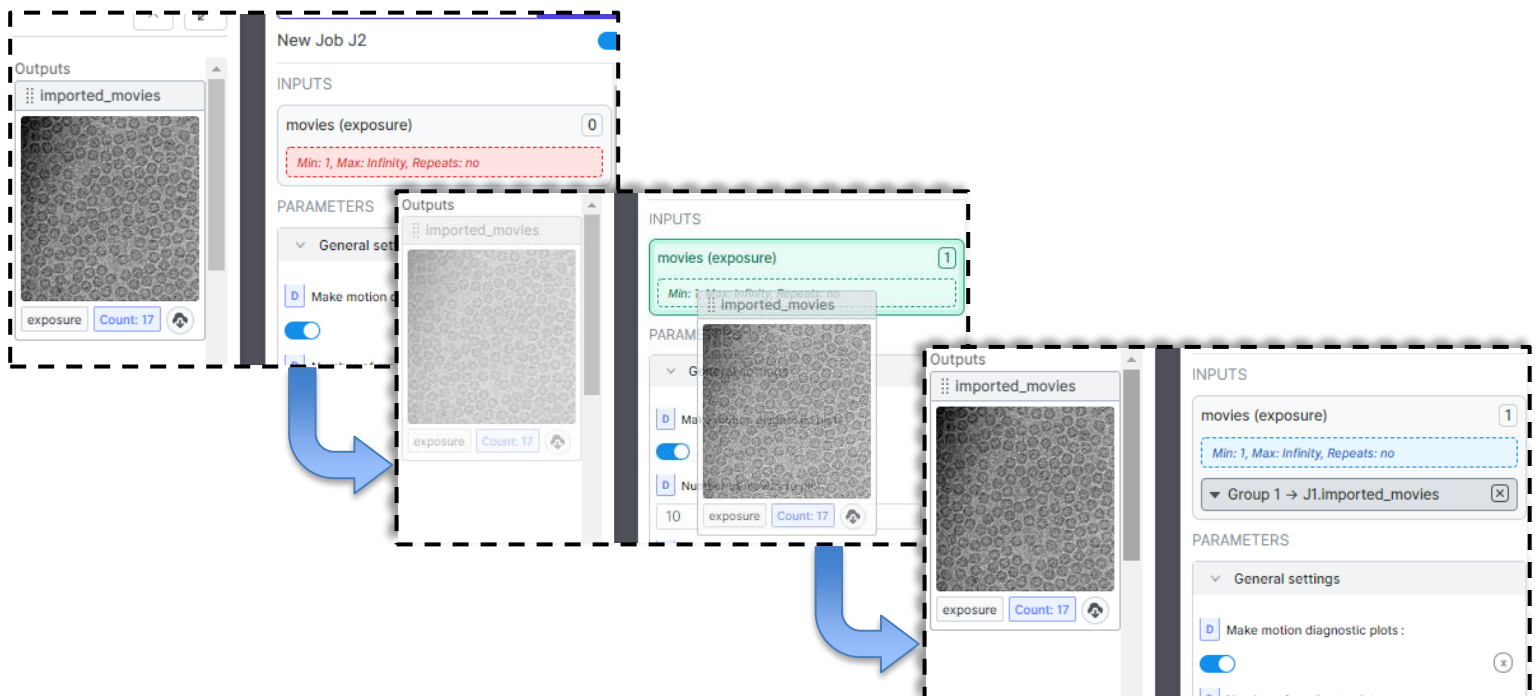
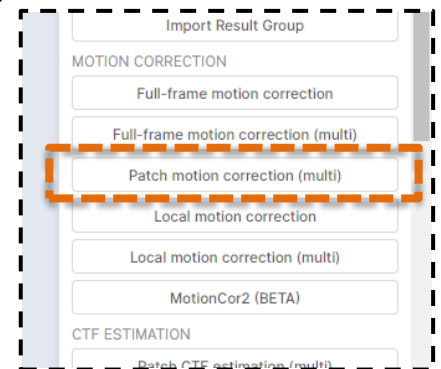


- **Step 10:** Fill in the parameters as shown on the right
 - Movies data path: This is the path to the raw movies you wish to import. The field accepts wildcards.
 - Gain reference path: The path to the gain reference.
 - Flip gain ref in X/Y?: Whether to flip the gain reference along the X-or Y-axis. Necessary for some microscopes but not others. Consistent artifacts like the ones pictured below are indicative of an improperly flipped gain.
 - Rotate gain ref?: Like flipping the gain, some setups require that the gain be rotated. The number entered here correspond to the number of 90° turns.
 - Raw pixel size: Parameter set during imaging.
 - Accelerating voltage: Determined by your microscope.
 - Spherical Abberation: Determined by your microscope.
 - Total exposure dose: Parameter set during imaging.

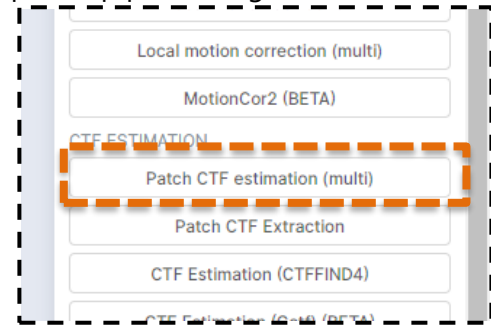




- **Step 11:** Click **Queue** then select the appropriate lane, followed by **Create**
 - The “appropriate” lane (aka node) depends on the computational environment. Your instructor will inform you of which you should use. Select it for every job you run from here on out.
- **Step 12:** Close the job window with the space bar. It can also be closed by clicking the X button in the top right corner.
- **Step 13:** Click “Job Builder” then select “Patch motion correction”
 - The Job Builder window will now change to the input page for the Patch motion correction job.
- **Step 14:** Click on the Import Movies job in the main window and press the space bar.
 - This will open the Import Movies job. Note the Outputs section on the right side of the window, and the small box titled “Imported_movies”
- **Step 15:** Drag-and-drop the “Imported_movies” output into the “movies (exposure)” input in the Patch motion correction Job Builder window with the left mouse button.
 - This drag-and-drop approach is how all outputs from one job will be used as inputs for another job. Inputs with a red box and Min value of 1 are required. Others have a blue box and a Min value of 0. These are optional inputs.



- **Step 16:** Launch the Patch motion correction job by clicking Queue, selecting the appropriate lane, then clicking Create.
 - For most datasets, the default parameters are fine. At the bottom of the Job Builder window, you can specify the number of GPUs to parallelize in order to speed up processing.
- **Step 17:** When Patch motion correction job finishes (the circle next to the job number will change from flashing blue to green), create a Patch CTF estimation job
- **Step 18:** Open the Patch motion correction job with the space bar, then drag-and-drop the “micrographs” output into the “exposures (exposure)” input box for the Patch CTF estimation job
- **Step 19:** Launch the Patch CTF estimation job.
 - Again, for most datasets the default parameters are acceptable. There are a few that may be changed on a somewhat regular basis:
 - Amplitude contrast: Generally 0.1 for cryo-EM samples and 0.25 for negative stain
 - Minimum/Maximum search defocus: The defocus limits the program will consider. By default, it is 0.1 – 4.0 μm which is usually okay but may need to be adjusted.



Particle picking and 2D Classification

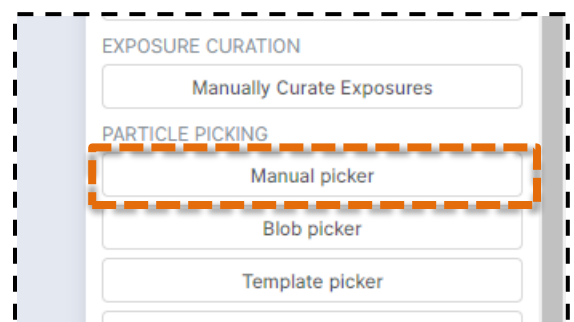
Background: After image preprocessing, the next step is selecting all of the particles in each micrograph. There are many ways to pick particles, with some working better than others for specific datasets. This test dataset is fairly easy to pick (the ice is thin and the particle is large and highly symmetric), but certain automated techniques (i.e. “blob picking”) struggle with it due to the lack of density in the center. Thus, we will use a combination of two techniques.

“Manual picking” will be used to select a small number of particles in order to create an initial average.

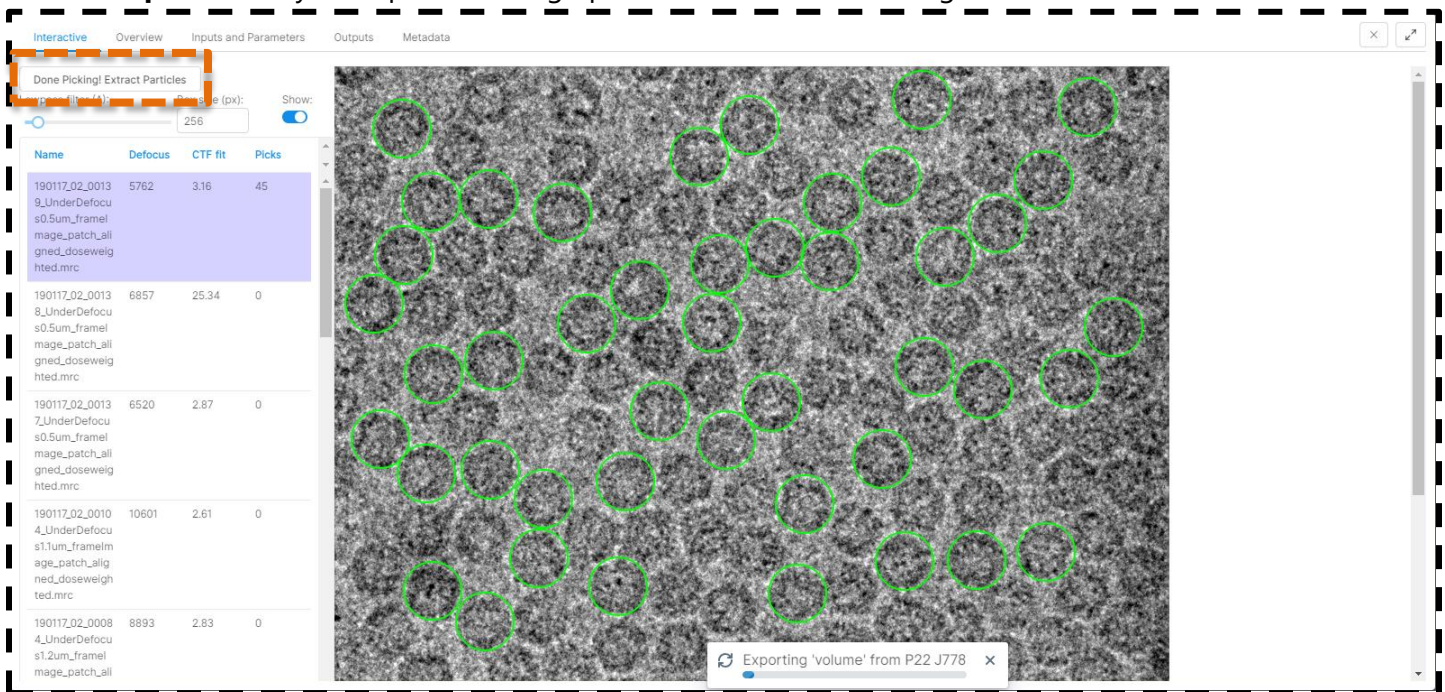
“Template picking” will take that average and use it to automatically find particles in the images.

The averages obtained after particle picking will be calculated using 2D Classification, which will do two things at once; classify each particle into distinct “classes”, and create an average of the particles in each of those classes. This will be used both as a diagnostic to see if our particle picking was accurate, and as a way to remove incorrect particle picks.

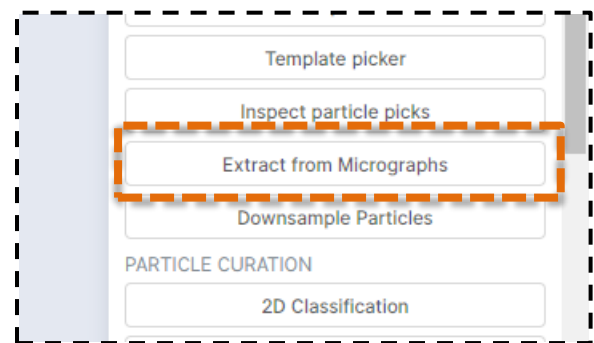
- **Step 20:** Once Patch CTF estimation has completed, create a Manual picker job.
- **Step 21:** Drag-and-drop the “exposures” output from Patch CTF estimation into the “micrographs (exposure)” input box in the Job Builder window and launch the job.
 - Remember, job windows are opened and closed with the space bar. *The input window for a job being built can be opened and closed by pressing B on the keyboard.*



- **Step 22:** Click the “Interactive” button in the top left corner of the window.
 - This is the main Manual picking UI. Micrographs are selected on the left. Lowpass filters, contrast intensity, and box size can be adjusted at the top. Note that this box size is purely for display.
- **Step 23:** Manually select ~50 particles. Particles are selected using the left mouse button, and can be removed with the right mouse button.
 - For this dataset, far fewer than 50 particles would be fine. For other datasets, 50 may be far from enough.
 - For non-symmetric particles, it is important to try and pick as many different orientations of the particle as possible. It is also generally advised to pick particles from a variety of different micrographs (different defocus values, different ice thicknesses, etc.)
- **Step 24:** Once you’ve picked enough particles, click “Done Picking! Extract Particles”



- **Step 25:** Now the particles must be extracted from the micrographs. Create an Extract from Micrographs job. Drag-and-drop the “micrographs” output from Manual Picking into the “micrographs” input box, and the “particles” output into the “particles” input box.
- **Step 26:** Set the “Extraction box pize (pix)” parameter to 416. Set the “Fourier fropp to box size (pix)” parameter to 208.



- The details behind finding the right box size are beyond this tutorial, but a good starting point is given by the following equation (for 300 kV, at 200 kV the multiplier is 20):

$$Box[\text{\AA}] = \frac{MaxDefocus[\text{\AA}]}{25 * BestPossibleResolution[\text{\AA}]}$$

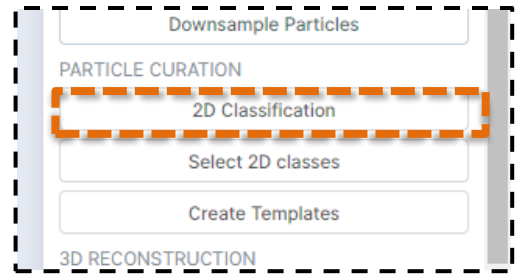
This recommendation may be too small if your particle is very large.

- Fourier cropping the box size to 208 bins the particles by 2x to speed up processing time.
 - This doubles our Nyquist frequency, but for this small dataset that will not impact the resolution we get.

➤ **Step 27:** Create a 2D Classification job and drag-and-drop the “particles” output from the Extract from Micrographs job into the “particles” input box.

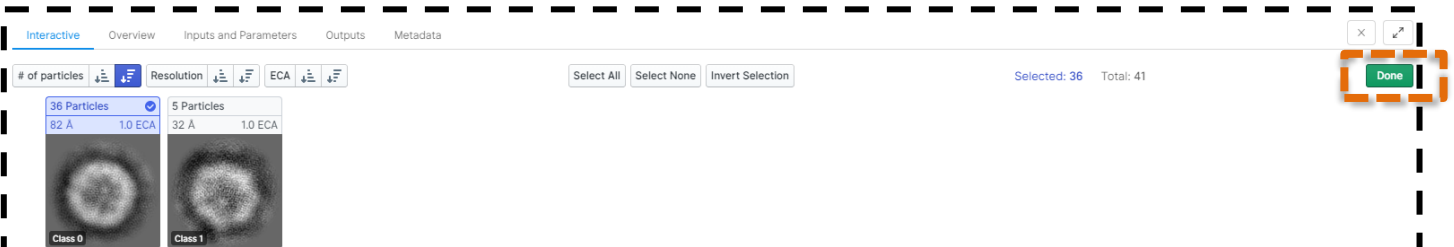
➤ **Step 28:** Set the “Number of 2D classes” parameter to 2 and run the job.

- For most projects, the default 2D Classification parameters will work well. Some, like the number of iterations, the batch size, and whether to force maximization over poses/shifts can be changed to produce better results for smaller/lower SNR particles.
- In this case, we only need 2 classes, and likely all/most of the particles will end up in a single class. The usual rule of thumb is ~100 particles/class, with no more than ~200. Processing time scales with this number.
- Increasing the number of GPUs can significantly increase processing time.



➤ **Step 29:** Create a Select 2D Classes job and use the outputs from 2D Classification as inputs by dragging-and-dropping (the “class_averages” output will go into the “templates” input). Run the job.

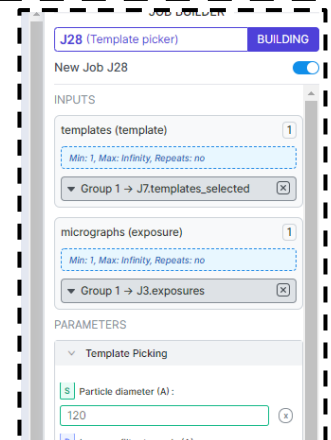
➤ **Step 30:** Click the Interactive button. Select the one class that looks the best. Click Done.



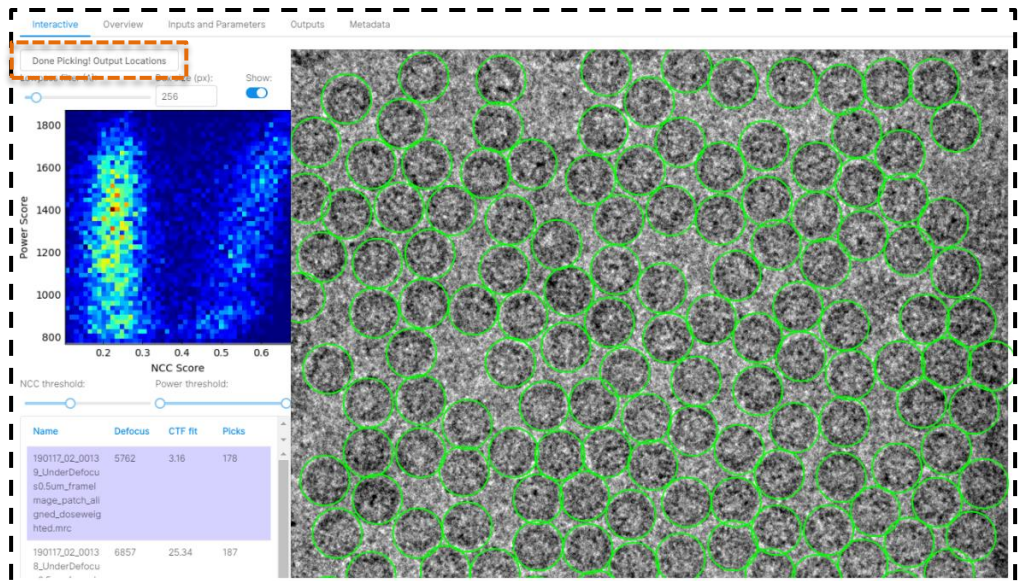
➤ **Step 31:** Create a Template picker job. Open the Select 2D Classes job and drag-and-drop the “templates_selected” output into the “templates” input. Close this job with the space bar and open the Patch CTF estimation job. Use the exposures from that job as the “micrographs” input for template picking. Set the Particle Diameter (Å) to 120. Run the job.

➤ **Step 32:** Create an Inspect Particle Picks job, using the output from the Template Picker. Use default parameters.

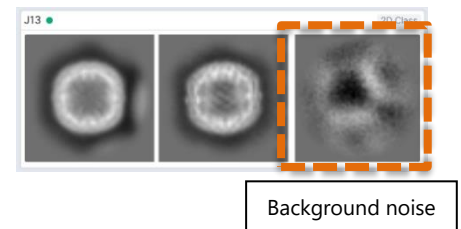
- It is likely that Template Picker picked a large number of “particles” that are actually background noise. This job lets us remove them before extraction.



- **Step 33:** Click the Interactive button in the top left corner. This will display a number of micrographs and the picks on them. Note that many are centered around background noise. Use the NCC threshold and Power threshold sliders to limit the picks only to real particles. Click "Done Picking! Output Locations"



- Appropriate NCC and power threshold numbers will vary by dataset. For this dataset, leave the power threshold as is, and increase the NCC threshold to ~0.34, as seen above.
- **Step 34:** Create another Extract from Micrographs job as before, with an Extraction box size of 416 pix, Fourier cropped to 208 pix.
 - You should end up with approximately 2,500 particles.
- **Step 35:** Repeat 2D Classification, this time with 2-4 classes.
 - It is possible that even after the Inspect Particle Picks job, some picks over background noise were included. Depending on how many were, you may need more or less classes. Usually 3 works well.
- **Step 36:** IF Step 35 revealed a class that looks like background noise (see image on the right), create another Select 2D Classification job and select the classes that look good.

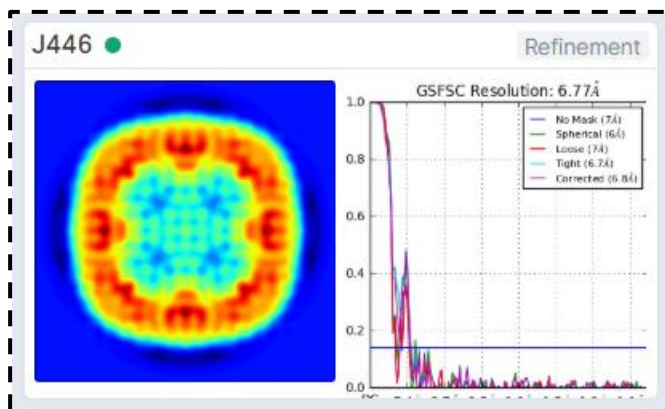


3D Reconstruction

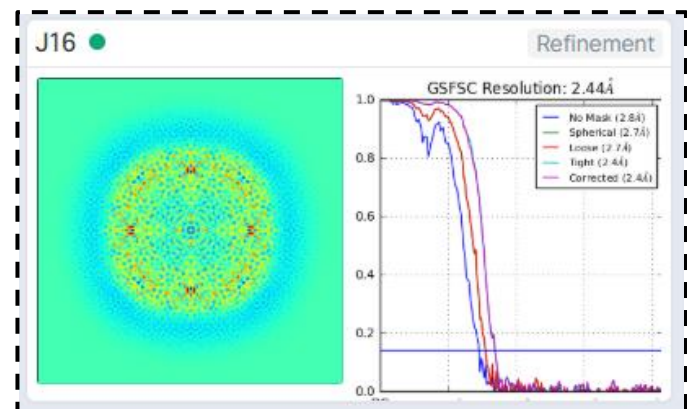
Background: Now that we have a clean stack of particles, as confirmed by 2D Classification, we move on to the final phase; obtaining a 3D reconstruction. This will happen in two stages; *ab initio* modeling where an initial low-resolution map is obtained, followed by homogeneous refinement where that map is refined to high resolution. Much more can be said about 3D reconstruction, but given the high level of homogeneity in this sample, it will easily reach high resolution with only two fairly quick jobs.

- **Step 37:** Create and run an *Ab-Initio* Reconstruction job using either the "particles" output from Step 34 (if your 2D classes all looked like clean particles) or the "particles_selected" output from Step 35 (if there was one or more classes that contained background noise) as inputs.

- **Step 38:** Change the *Ab-Initio* Reconstruction job parameters as follows, then run the job
 - Maximum resolution (Angstroms): 4
 - Initial resolution (Angstroms): 6
 - Number of initial iterations: 0
 - Number of final iterations: 200
 - Initial minibatch size: 1000
 - Final minibatch size: 1000
- While the default parameters will produce a suitable low-resolution map, these parameters produce one that is much higher resolution and will help avoid a potential problem in the next step. However, it will increase the amount of time this job takes. *NOTE:* Not all samples should have ab initio models constructed with these parameters. Usually the default parameters are best.
- **Step 39:** Create a Homogeneous Refinement job (If you have the “NEW!” refinement, use that) using the particles and volume output from *Ab-Initio* modeling as inputs.
 - Default parameters (save for symmetry) are fine. For other datasets, tweaking options related to higher-order aberration correction (e.g. optimizing per-particle defocus or per-group CTF params) may lead to better results. Experiment with them to see what yields the best results for your personal datasets.
- **Step 40:** Set the Symmetry option to O (for octahedral) and run.
 - Unless you are absolutely certain of your symmetry, it is best to leave it at C1. Run it, then verify the symmetry in Chimera (<https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/midas/measure.html#symmetry>), and if it is indeed symmetric, re-run refinement.
 - **If you see strange results from Homogeneous Refinement that resemble the image below,** it is because cryoSPARC has mis-identified the symmetry axis. This is less likely to happen if you used the suggested parameters in Step 38. If it does happen, simply re-run the refinement and hope it works this time.



Wrong symmetry axis



Right symmetry axis

Visualization

Background: Now that we have our map, it is time to download it and view it using UCSF Chimera.

- **Step 41:** Download the “map_sharp” file by clicking on the download icon under “volume” in the output area of the Homogeneous Refinement job.

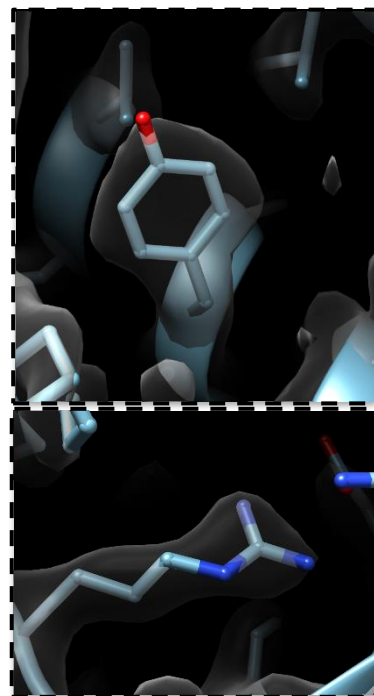
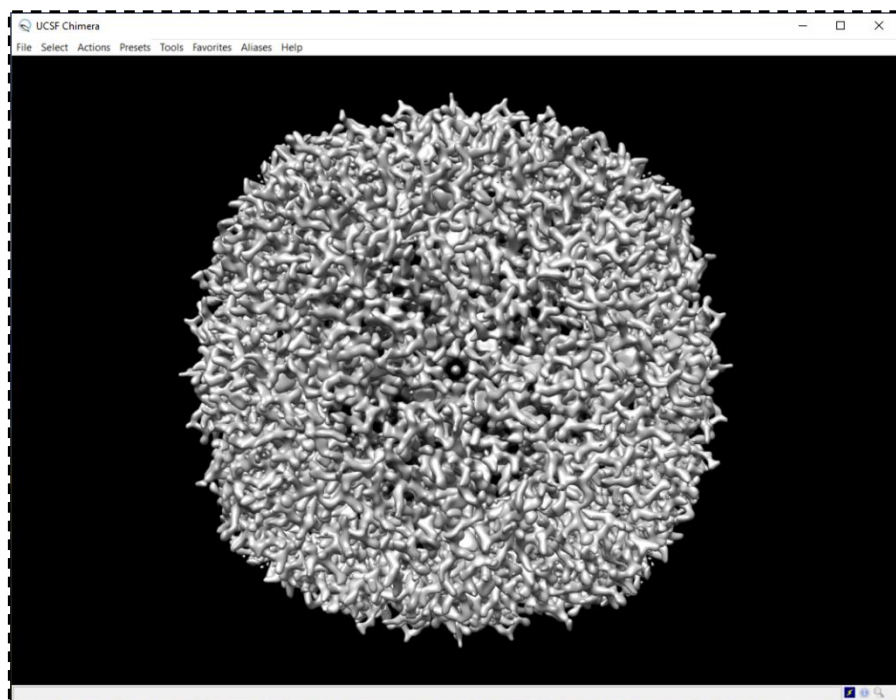
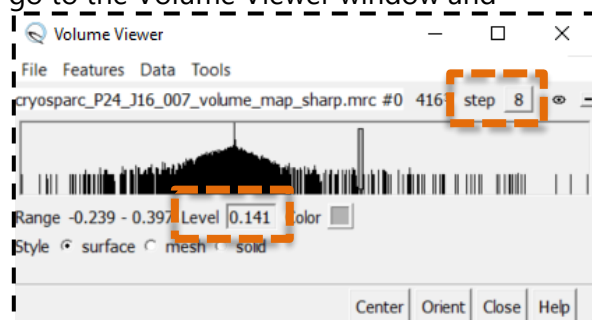
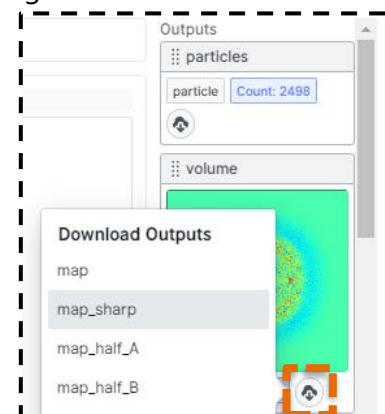
- This .mrc file will likely save to your Downloads folder, or whatever folder is your default download location for your internet browser.

- **Step 42:** Open UCSF Chimera. Click “File” then “Open” and click on the cryoSPARC map file.

- The file will be named
“cryosparc_P##_J##_###_volume_map_sharp.mrc”

- **Step 43:** With the map and Volume Viewer window open, go to the Volume Viewer window and change the step size from 8 (or maybe 4) to 1. Adjust the threshold to approximately 0.15.

- Check the pitch of the α -helices. It is possible that cryoSPARC got the wrong handedness, which can be corrected with the command “vop zflip #0”
- PDB 6V21 is a good atomic model to dock into this map. See Chimera documentation for more.



Acknowledgements: Created by Micah Rapp. Thank you to Ed Eng, Yong Zi Tan, and Misha Kopylov for processing advice. Data comes from EMPIAR-10248. Thank you to Swapnil Bhatkar for IT support and development of the AWS cryoSPARC instance.