Quantitative Methods Workshop 2022 | Cryo-EM 3D reconstruction

We will be using the software suite cryoSPARC (https://cryosparc.com/) to illustrate the single particle reconstruction technique commonly used to determine the 3D structure of protein complexes. Today's sample is the 20S proteasome from a thermophilic yeast. This particle contains the key active sites of the proteasome that carry out much of the protein degradation in eukaryotic cells. For the aficionados, note that this preparation lacks the regulatory particle known as the 19S. This dataset was described here: https://elifesciences.org/articles/06380, and this tutorial is based on a similar tutorial found here: https://guide.cryosparc.com/processing-data/cryo-em-data-processing-in-cryosparc-introductory-tutorial

Each group (i.e. table) will work through tutorial as a team. Some of the tasks are computationally intensive and may take as long as 10 minutes to finish. Please take this time to consider the discussion questions described in this document.

My hope is that you can all log into the workstation so you can interact with the web interface, but please only create and submit one version of each job as a group. This will likely work best if you designate one submitter, or if you rotate this responsibility around the group.

Finally, have fun!!

Login and setup

1. Go to XXX
2. Log in with your assigned login/password XXX;XXX
3. Go to the project named T20S – this stands for the 20S proteasome from Thermoplasma
4. Click on the first job, which imported the movies

The movies consist of 38 frames, and cryoSPARC has displayed the sum of those frames. Note a few observations about the images displayed. Think about (and note if you'd like) how these observations may impact our ability to learn the 3D structure of this protein complex.

Motion correction [~5 minutes]

1. Create a new **Path motion correction (multi)** job, and drag the micrographs from Job 1 into the "movies" input form.
2. Click "queue"

While this job is running, discuss why we need to perform motion correction. Additionally, discuss the advantages and disadvantages of treating the micrograph as one rigid body, as a series of adjacent patches, or, in the limit, as individual pixels.

Watch the output as the job is running. What do you notice about the motion corrected micrographs? What do you notice about the motion of adjacent patches (see the red lines that depict the motion of each patch)? What might explain the motion you see in each patch.

CTF estimation [~2 minutes]

1. Create a **CTF Estimation (CTFFIND4)** job. Connect the patch motion output to the exposures input on this patch CTF estimation job.
2. Click "queue"

While this job is running, take a moment to discuss the origin of the contrast transfer function – if this is completely new to everyone in your group, please don't hesitate to ask one of the instructors for guidance. You can learn more about the Contrast Transfer Function here: https://cryoem101.org/chapter-5/#part3, and, after the workshop, you can learn about the origins of phase contrast here: https://www.youtube.com/watch?v=HTJsYTGAhYg&ab_channel=caltech

Particle picking I [~5 mintues]

1. Start a **Manual Picker** job
2. Click the interactive tab at the top left of the job.
3. Set the box size to 340 px
4. Try changing the "Lowpass filter" slider.

Discuss what you see as you move the slider. Discuss the different shapes you see – why might we be observing these different shapes?

5. Set lowpass filter to ~10 angstroms
6. Pick ~10-20 particles from ~10 micrographs by clicking on the center of the particle. You'll want to work quickly here and don't worry about getting it perfect. In total we want ~100-200 particles, with a roughly equal proportion of "rings" and "barrels".

Having now picked a series of ~150 particles, as a group hypothesize the overall shape for this molecule.

7. Click "Done picking – extract particles"

2D classification I [~2 mintues]

1. Start a **2D classification job**
2. Connect the output of your Manual picking job to the input of the 2D classification job
3. Set the number of classes to 2
4. Click queue

Watch the job progress. What do you notice with each iteration? Compare the final 2D class averages to your original particle images. What do you notice? Hypothesize why a circular mask (the area around the edges that is all gray, and is not considered during the 2D classification procedure) is applied during the 2D classification?

We had set our number of classes to 2. Why might we have chosen such a small number of classes, and why might that be problematic?

Particle picking II [~10 minutes]

1. Start a **Blob picker job**.
2. Connect the output of your CTF estimation job to the micrographs input of this picker job.
3. Set the minimum particle diameter to 100 angstroms

4. Set the maximum particle diameter to 200 angstroms
5. Click queue

Blob pickers are a quick and unbiased way to select particles in a micrograph. Here, three templates are created, each composed of a Gaussian blob with different radii. These blobs are then used as "templates" to identify particles in the micrograph by rasterizing them along the micrograph and identifying regions that strongly correlate with the blob. Particles that are too big or too small will be missed, as well all of the particle-free regions of the micrograph.

Discuss the types of particles for which this sort of approach would fail.

6. Start a **Inspect particle picks** job
7. Connect the blob picker particle outputs and micrograph outputs to this job.
8. Click queue
9. Click the interactive tab at the top left of the job.
10. Change the boxsize to 20 and the lowpass filter to 10 A to better highlight the center of the picks. You may need to click a new micrograph for the green boxsize circles to update
11. Try adjusting the power threshold, which sets the range of contrast allowed. Start by raising the minimum and watch as particle picks disappear. Once you no longer see picks in the ice-only regions of the micrograph, try adjusting the right slider to eliminate high contrast artifacts.

    Discuss what types of particles are lost as you narrow the range of allowed contrast.
12. Try adjusting the NCC threshold, which sets a minimum normalized cross correlation value.

    Discuss what types of particles are lost as you increase the minimum required cross correlation.

13. Pick settings you are happy with – we recommend an NCC of ~0.20, and power range of 575 – 1050)
14. Click done picking, output pick locations.
15. Start an **Extract from Micrograph** job, connecting the micrographs and particles outputs from your inspect picks job.
16. Set extraction box size to 440 pix and Fourier crop to box size 256 pix
17. Click queue

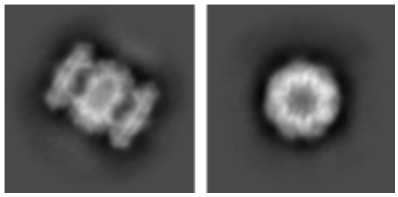
While you are waiting for this job to complete, discuss in the group the advantages of disadvantages of the two particle picking strategies we've tried. Brainstorm alternative strategies one might deploy…thinks CNNs! If you have time, think about why we've chosen to Fourier crop our images and what this does to the maximum resolution we could achieve.

2D classification II [~10 minutes]

1. Start a **2D classification job**
2. Connect your blob picker output – we expect ~10,000 particles
3. Set the number of 2D classes to 20

Watch the job closely as it proceeds. What do you notice about the classes emerging? Do you see the particles shifting in the XY plane? How might that be happening?

4. Start a **Select 2D classes** job, connecting the outputs of the 2D classification above (particles and templates)
5. Select two classes – one with the "ring" view, and one with the "barrel" view.
    a. There will likely be multiple classes of each view – choose the ones that are best centered and showing the highest resolution features. Example below.



Particle picking III [~5 minutes]

1. Start a **Template picker** job
2. Connect your micrographs from your CTF estimation job and the templates you selected above
3. Set the particle diameter at 190 angstroms

Watch the picks coming from this template-based picker – are they better or worse than the template picker? What are the potential downsides of such a template-based approach?

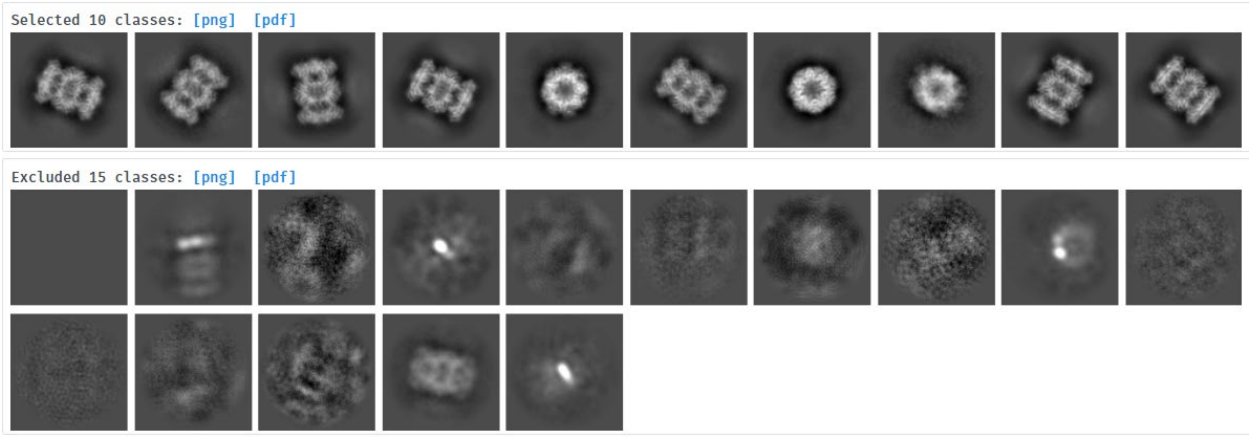
4. Start a **Inspect particle picks** job
5. Connect the blob picker particle outputs and micrograph outputs to this job.
6. Click queue
7. Click the interactive tab at the top left of the job.
8. As above, pick settings or the NCC threshold and power range you are happy with – we recommend an NCC of ~0.35, and power range of 930 – 1990)
9. Click done picking, output pick locations.
10. Start an **Extract from Micrograph** job, connecting the micrographs and particles outputs from your inspect picks job.
11. Set extraction box size to 440 pix and Fourier crop to box size 256 pix
12. Click queue

2D classification III [~10 minutes]

1. Start a **2D classification job**
2. Connect your template picker output – we expect ~10,000 particles
3. Set the number of 2D classes to 20

Watch the job closely as it proceeds. What is different about this round of 2D classification vs. what you observed with the blob picker? Once this has finished, what do you observe about the resulting classes? Hypothesize about the observed symmetry in the molecule, noting that this molecule is composed of only two subunit types (alpha and beta). What can this tell us about the overall stoichiometry/composition of the particle? The T20S is a potent protease, capable of degrading many cellular proteins. This activity must be carefully regulated such that it does not erroneously degrade proteins in the cell. Given this information and your classes, hypothesize where the proteolytic sites might be, and how the T20S might regulate access to those sites.

6. Start a **Select 2D classes** job, connecting the outputs of the 2D classification above (particles and templates)
4. Select all classes that show high resolution features – see below for example classes to include/exclude



3D reconstruction [Ab-initio – 5 minutes; Refinement – 5 minutes]

1. Start an **Ab-initio reconstruction** job
2. Connect the particles from your selected 2D classes and queue this job.
   This job is now generating a lower resolution 3D volume using our provided particles. In doing so, it also coarsely assigns projection angles to each of the particles. Once iteration 100 has completed, inspect the resulting figures and discus amongst the group what these figures likely represent. Please feel free to ask Joey, Laurel, or Barrett for additional guidance.

   Once the ab-initio job has finished, we can "refine" the structure. Here, we will use the initial model generated above to iteratively improve the angular assignments to each particle.

3. Start a **Homogeneous refinement** job
4. Connect the particles and ab initio volume from above.
5. **Set the symmetry to D7 and queue this job**
   While this job is running consider the equation below, which provides insight into how this algorithm works:

$$V^{MAP} = \arg\max_{V} \sum_{1}^{N} log \int p(X_i|\emptyset, V)p(\emptyset)\, d\emptyset + log\, p(V)$$

Where V is the 3D volume, V^{MAP} is the maximum a posteriori volume, X is a given 2D image drawn from a set of N images, and $\emptyset$ is the set of projection angles assigned to those N images.

Here, the intuition is that given the estimated Volume^{K} at iteration k, the N images are first aligned to this volume (E-step), then with the updated alignments, a new Volume^{K+1} is calculated (M-step). Discuss why this approach is highly sensitive to the initial model provided as Volume^{0}?

Would this approach work if we had imaged a mixture of different types of particles (e.g. ribosomes and proteasomes)? If not, why not? How could one trivially alter the equation above to allow one to simultaneously determine multiple structures?

6. Once this job has completed, click the download button under the Output volume, choose map_sharp. If you have chimera installed on your computer, you can view that volume by dragging it into the ChimeraX window. If nobody at your table has chimera installed, asked one of the instructors to help you view it on their laptop.

   Look at the structure closely – can you see protein secondary structure? What about individual side chains? Think back to the initial movies you imported and how little signal appeared to be in those images…amazing, right?