Protocol for:
cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination

Single particle electron cryomicroscopy (cryo-EM) is a powerful method for determining the structures of biological macromolecules. With automated microscopes, cryo-EM data can often be obtained in a few days. However, processing cryo-EM image data to reveal heterogeneity in the protein structure and to refine 3-D maps to high resolution frequently becomes a severe bottleneck, requiring expert intervention, prior structural knowledge, and weeks of calculations on expensive computer clusters. This protocol describes the steps required to use the cryoSPARC (cryo-EM single particle *ab initio* reconstruction and classification) software package to reconstruct and refine high resolution cryo-EM maps without prior structural knowledge in hours or minutes on an inexpensive desktop computer.

Introduction

CryoSPARC (cryo-EM Single Particle Ab-initio Reconstruction and Classification) is a software package for processing cryo-EM single particle data. CryoSPARC includes GPU accelerated implementations of 1) stochastic gradient descent (SGD) for ab-initio 3-D structure determination of potentially heterogeneous samples and 2) branch and bound (BnB) refinement for extremely rapid high-resolution refinement of 3-D maps to near-atomic resolutions. CryoSPARC also includes a powerful but simple web-browser based graphical user interface that allows non-expert users with minimal training to manage increasingly large collections of cryo-EM image datasets, run, schedule and queue computational jobs, view computational results including figures as they are computed in real time by the algorithms, and allows multiple users to simultaneously use the same computational hardware remotely.

Computational tasks in cryoSPARC have associated parameters, each with default values that work well on a wide range of input datasets. Parameters can be modified, and the final values used in an experiment are recorded by the software for simple reproducibility and inspection.

CryoSPARC is developed jointly by Structura Biotechnology Inc. and the University of Toronto, Canada. Both academic licenses and commercial licenses are available through Structura Inc. at www.cryosparc.com.

The following protocol describes how to use cryoSPARC to perform ab-initio heterogeneous reconstruction and high-resolution refinement in minutes or hours on a single computer with a single GPU.

Equipment

Computer Hardware
A desktop workstation or rackmount server with:
    + Quad-core or better Intel CPU
    + 32GB+ CPU RAM
    + NVIDIA GPU with Compute Capability 3.5+ and 8 GB+ GPU RAM
    + 256GB+ SSD drive
    + Linux OS (we recommend Ubuntu 16.04)
    + NVIDIA Drivers and CUDA 6.5+

Data
    + One or more raw cryo-EM particle stacks in .mrc or .mrcs format
    + One or more meta-data files in .star, .par, .txt or .emx formats, including CTF values for each particle (defocus, astigmatism, phase shift)
    + Imaging parameters for the dataset, including accelerating voltage, amplitude contrast, spherical aberration, pixel size

Software
CryoSPARC version 0.2.36 or newer
(Note that installing cryoSPARC will also automatically install all dependencies)

Procedure

Step 1) Start cryoSPARC on the GPU computer.

After installation and configuration, the `cryosparc` command should be available in your PATH. Simply type the following at the command line (Figure 1)
$ cryosparc start
This will begin the cryosparc host processes and provide you with a web address where cryoSPARC can be used. Typically it will be http://localhost:38000

Figure 1. cryoSPARC start command.

Step 2) On the GPU computer or any other computer connected to the same local network as the GPU computer, use the Google Chrome browser to navigate to the web address from step 1).

You will be presented with the cryoSPARC main dashboard (Figure 2).
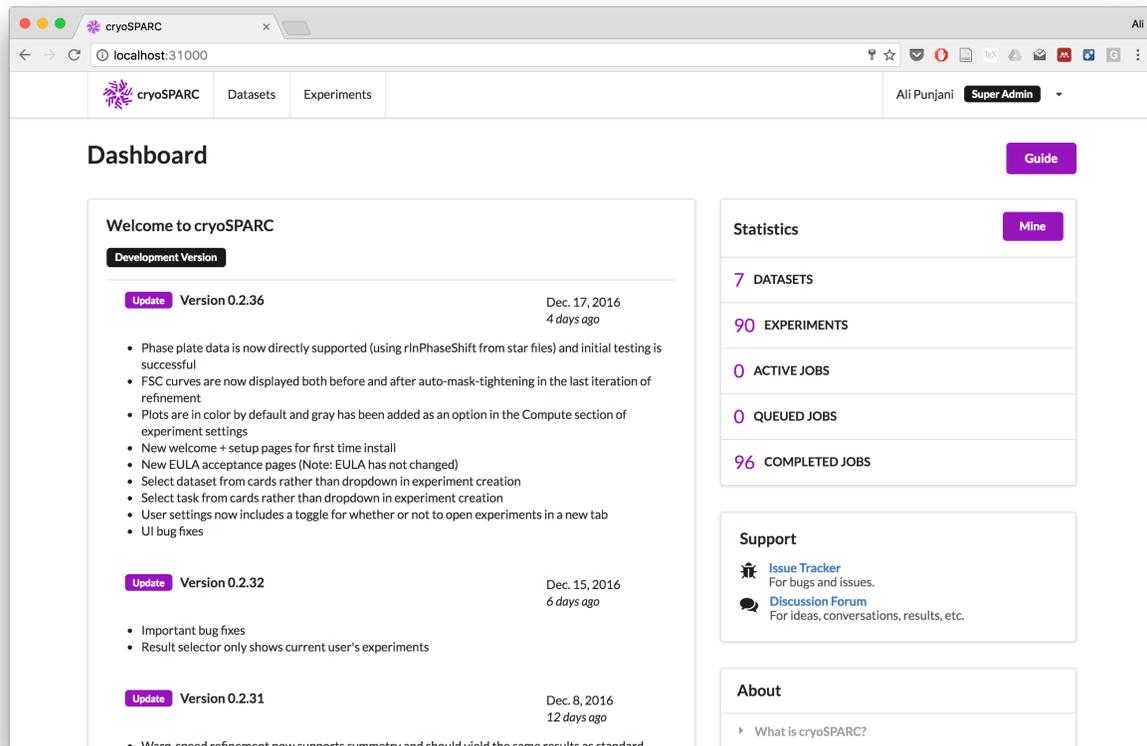
Figure 2. cryoSPARC main dashboard. The dashboard indicates version number, changelog, update availability, current job processing status, utilization, and help and support links.

Step 3) Click on the 'Datasets' button in the top navigation bar to arrive at the datasets page (Figure 3). From here, click on 'New Dataset' to create a new particle dataset in cryoSPARC (Figure 4). Click on the Configure tab on the left, and click the 'Add New File Set' button to begin adding a new set of particle stack and meta data files to cryoSPARC.
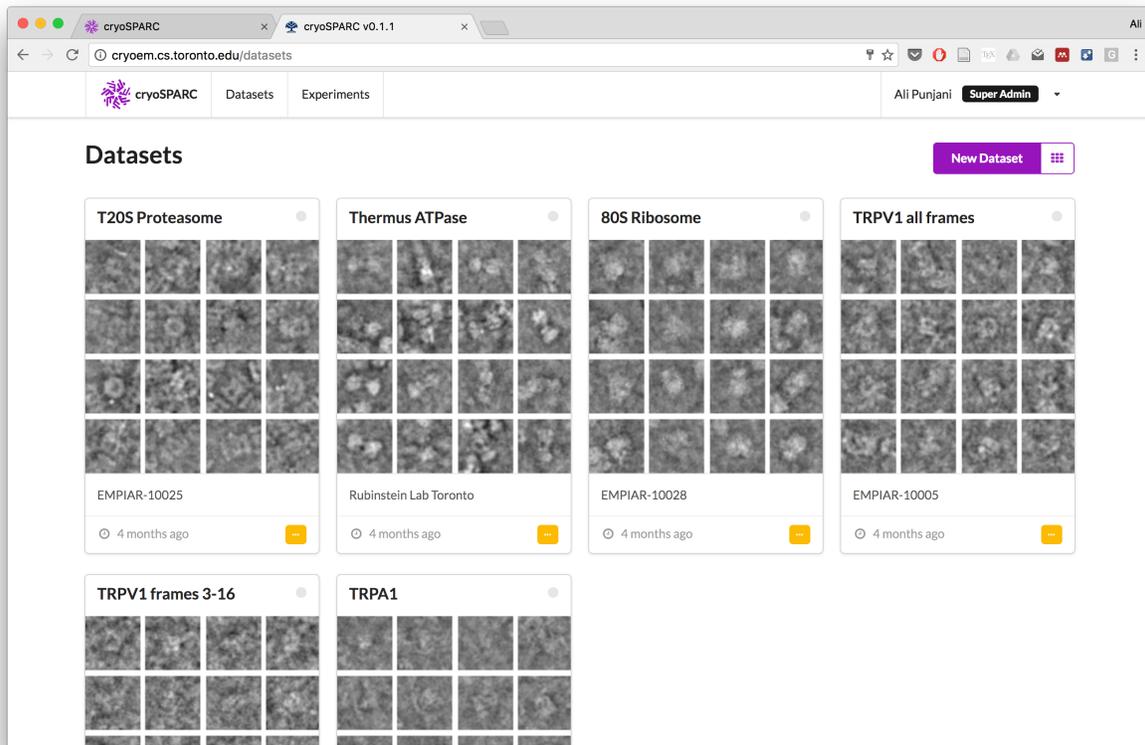
Figure 3. Datasets management page. All datasets are shown along with a random selection of particle images from each. Click on a dataset to view it's details.
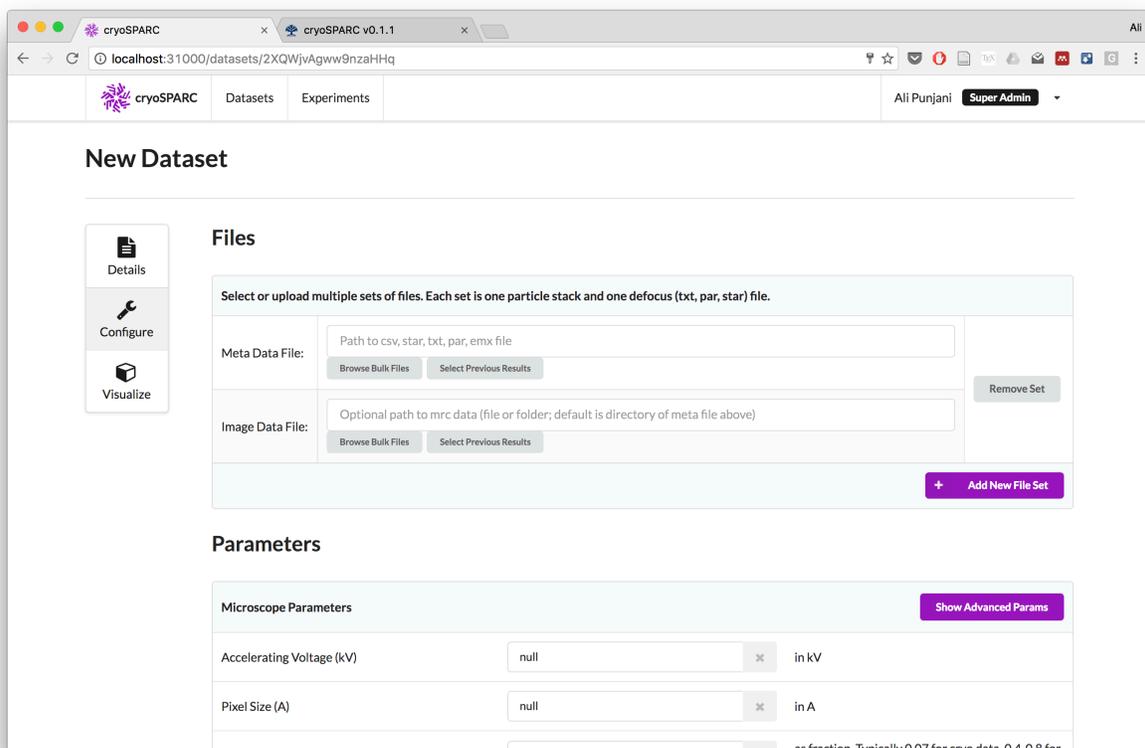
Figure 4. New dataset page in cryoSPARC. File sets including particle stack and meta data file can be indicated, along with imaging microscope parameters.

Step 4) Click on the 'Browse Bulk Files' button to browse for a meta data (.star, .par, .txt, .emx) file for your dataset. Repeat for the particle stack if necessary (.star files contain paths to corresponding .mrcs files). Enter microscope parameters if necessary (.star files contain microscope parameters).

Click on the Visualize tab on the left, and click Launch to have cryoSPARC load your data and display it to verify the integrity of the provided files and the correctness of microscope parameters (Figure 5).
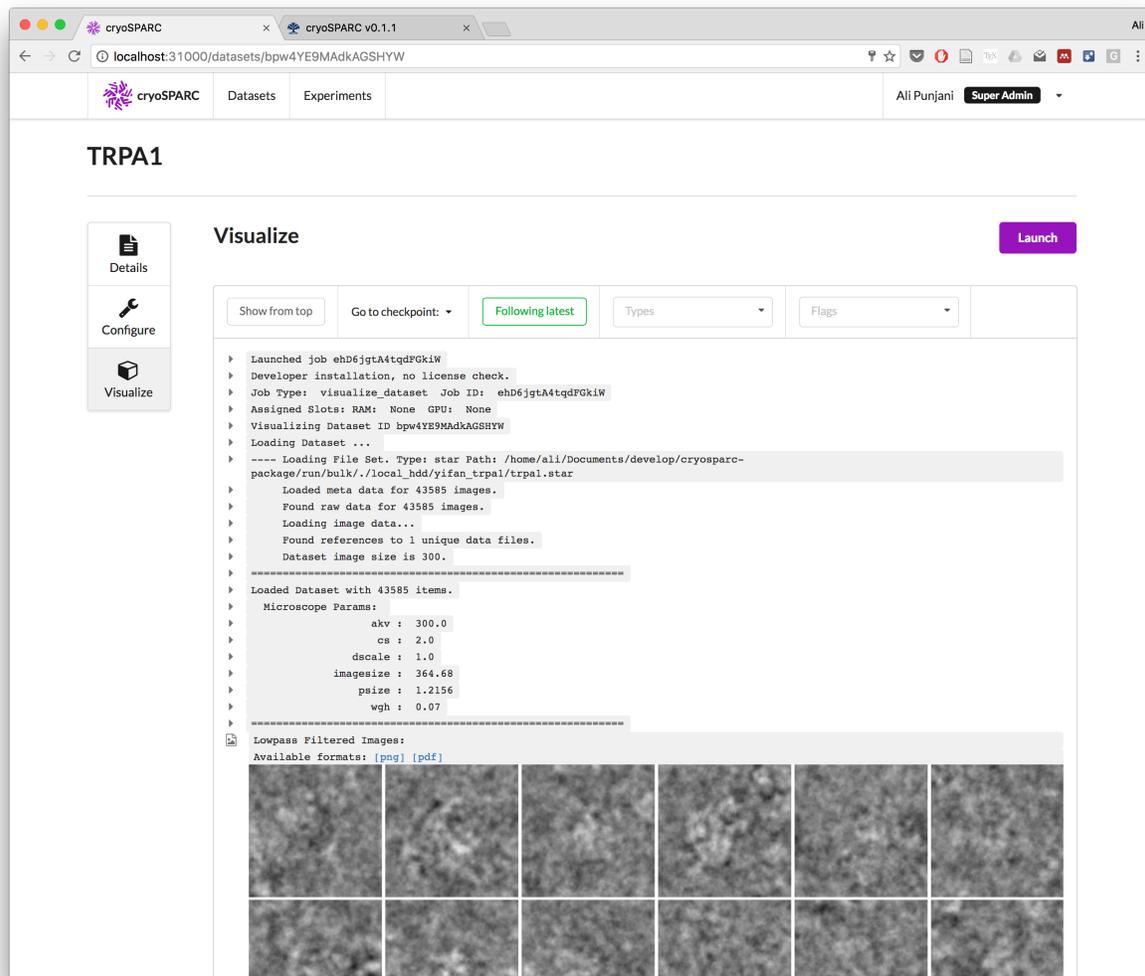
Figure 5. Dataset visualization. Read the microscope parameters in the output to check that they are correct, and inspect the visualized particle images.

Step 5) Click on the Experiment button in the main navigation bar. This will take you to the experiment management page (Figure 6). Click 'New Experiment' to start a new experiment. In the New experiment page, select your dataset, then select the task you wish to perform (in this case Ab-Initio Heterogeneous Reconstruction). Then click on the parameters page to see default parameters of the experiment (Figure 7). In this case, you can set the number of classes you wish to reconstruct but for this tutorial we will leave it as 1 class.
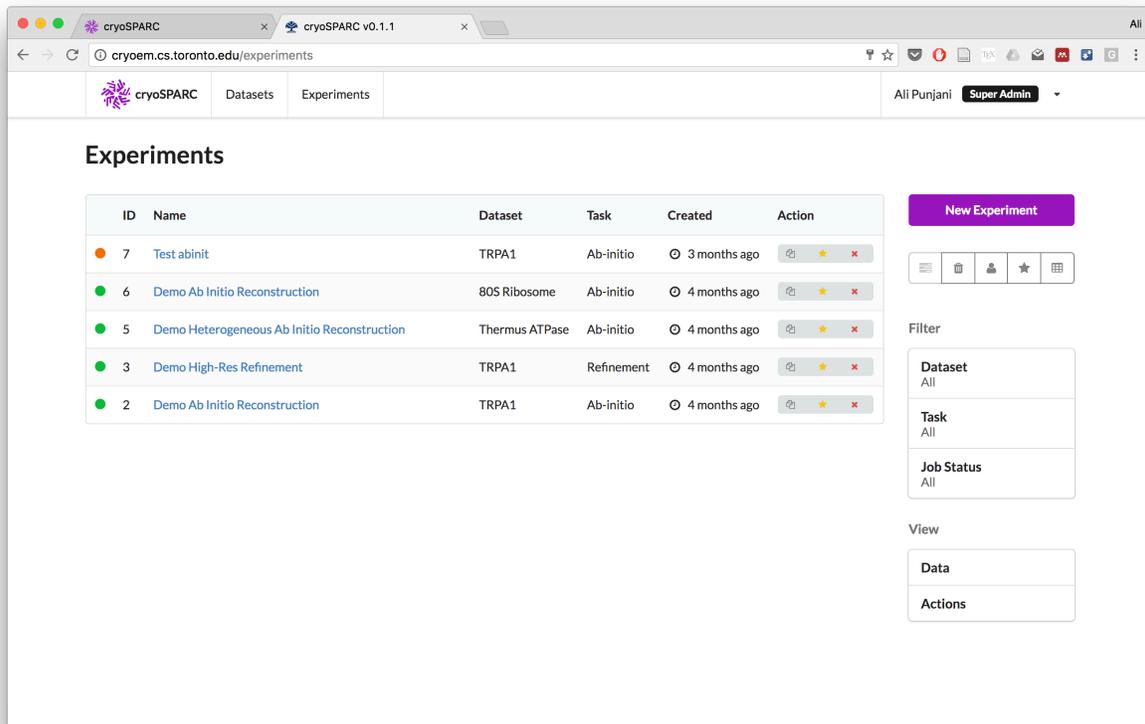
Figure 6. Experiment management page. View, list, filter, sort experiments that you have created, and view their status.
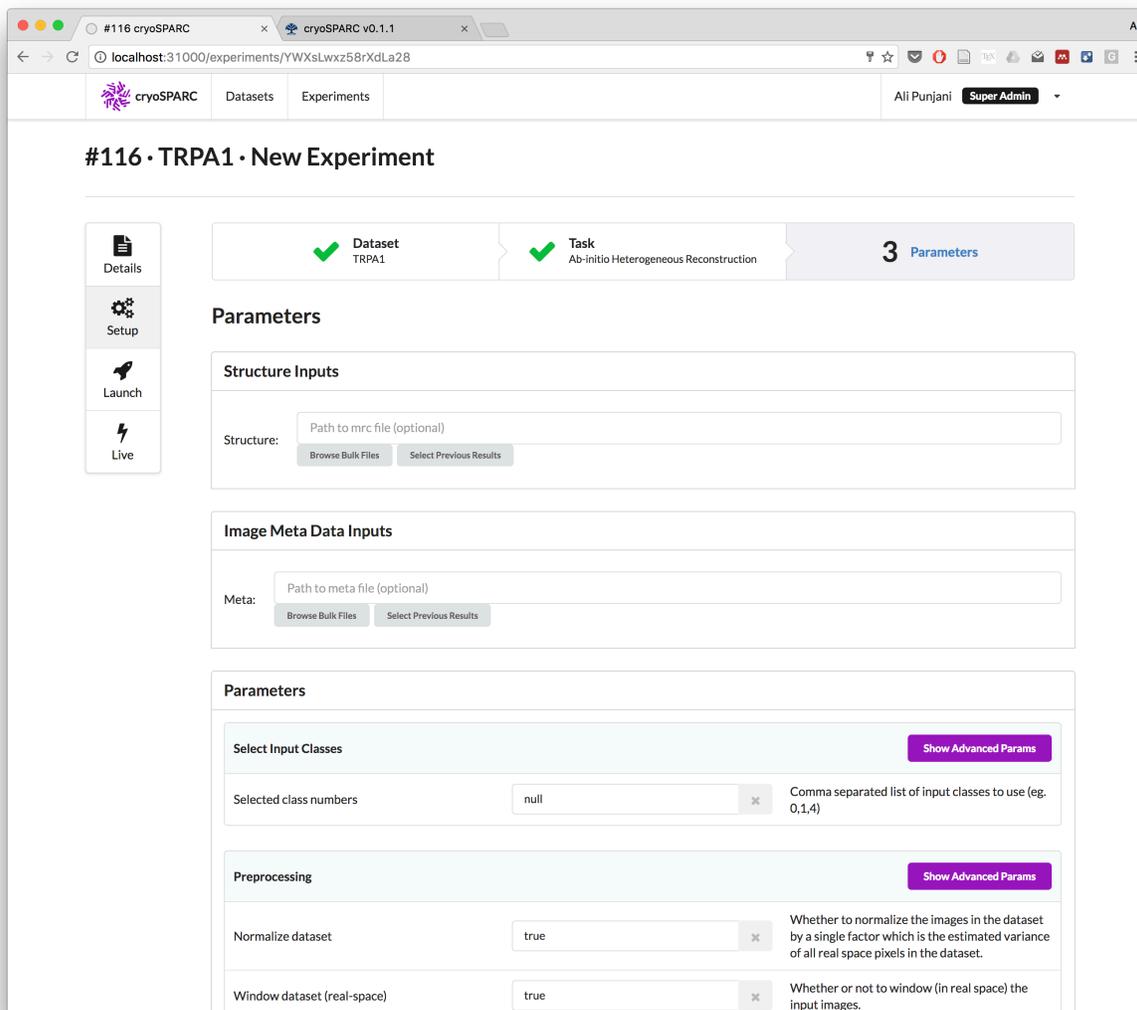
Figure 7. Parameters set-up page for a new experiment. Each parameter has a default value and no modifications need to be made to achieve the reported results. Modified parameters are indicated in green, and parameters are grouped by section into basic and advanced parameters.

Step 6) Click the Launch tab on the left and select 'Enqueue'. This will queue your experiment for running in the cryoSPARC scheduler. When the job begins to run, you will see results streaming in real time on the Launch page. This includes plots and result files that are generated (Figure 8). You can directly download result maps and plot data from the interface. In this case, the single class ab-initio reconstruction will yield a medium-resolution structure of the TRPA1 ion channel in approximately one hour. When the job is finished, you may view the final result.
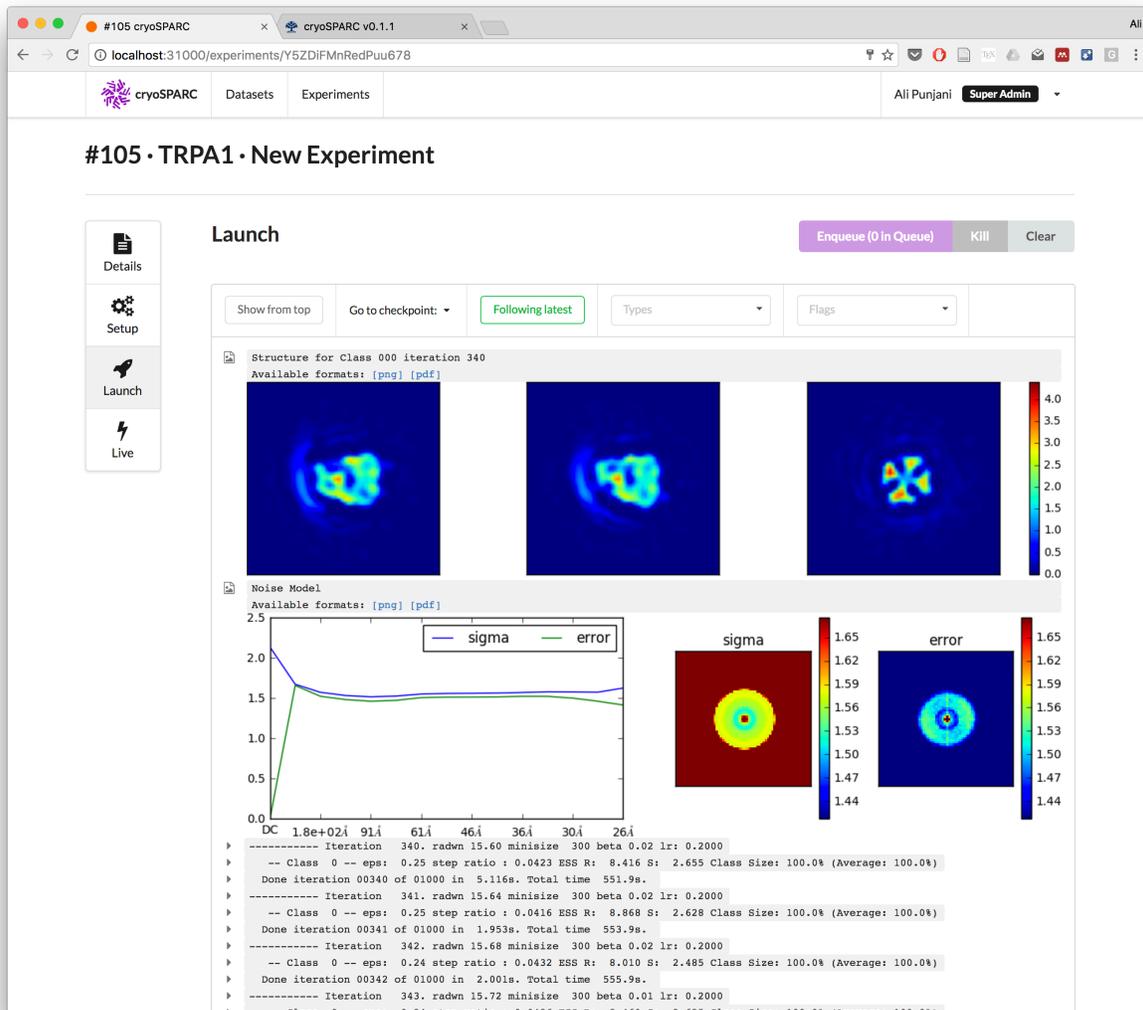
Figure 8. Real time streaming results from cryoSPARC processing an ab-initio structure reconstruction job. Plots, text, and result files are all visible in the interface.

Step 7) To refine a structure that has been reconstructed by ab-initio reconstruction, return to the experiments page, click 'New Experiment' and select your dataset. This time, select Heterogeneous Refinement as the task. In the job parameters, use the 'Browse previous results' button to select the output from the previous ab-initio experiment as the starting reference for high-speed refinement (Figure 9). In refinement experiments, every parameter has default values and none need to be changed, except if you wish to enforce symmetry during refinement. In this case we can enforce C4 symmetry.
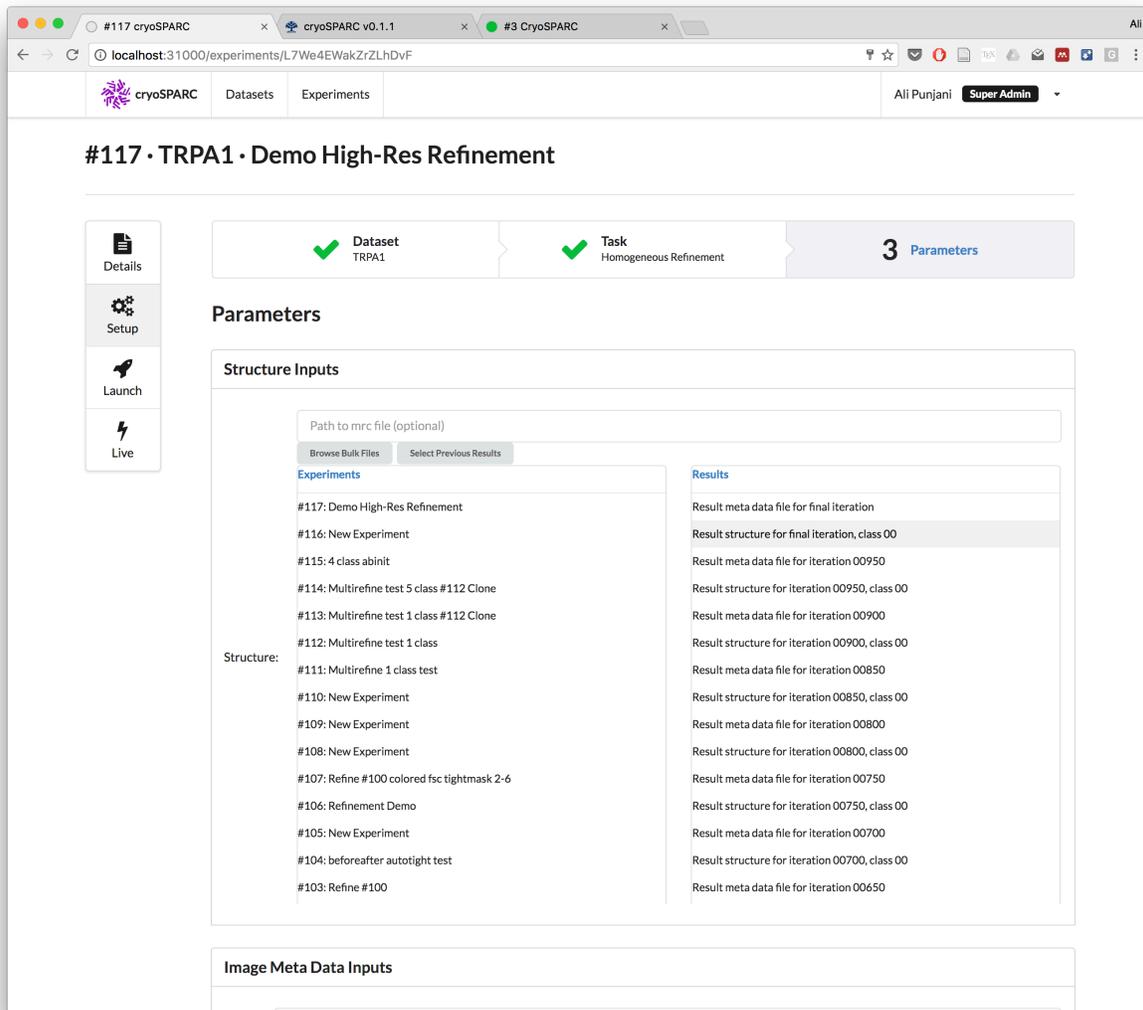
Figure 9. Browse previous results to select the starting map for a high-resolution refinement experiment.

Step 8) Once again, click on the Launch tab on the left for the refinement experiment, then click Enqueue. The job will begin to run and report results in real time. This refinement will take approximately one hour to reach 4.2A resolution (Figure 10). Results maps both sharpened and unsharpened, along with half-maps, masks, and FSC curves can be downloaded directly from the interface. The resulting structure can be viewed locally using UCSF Chimera (Figure 11).
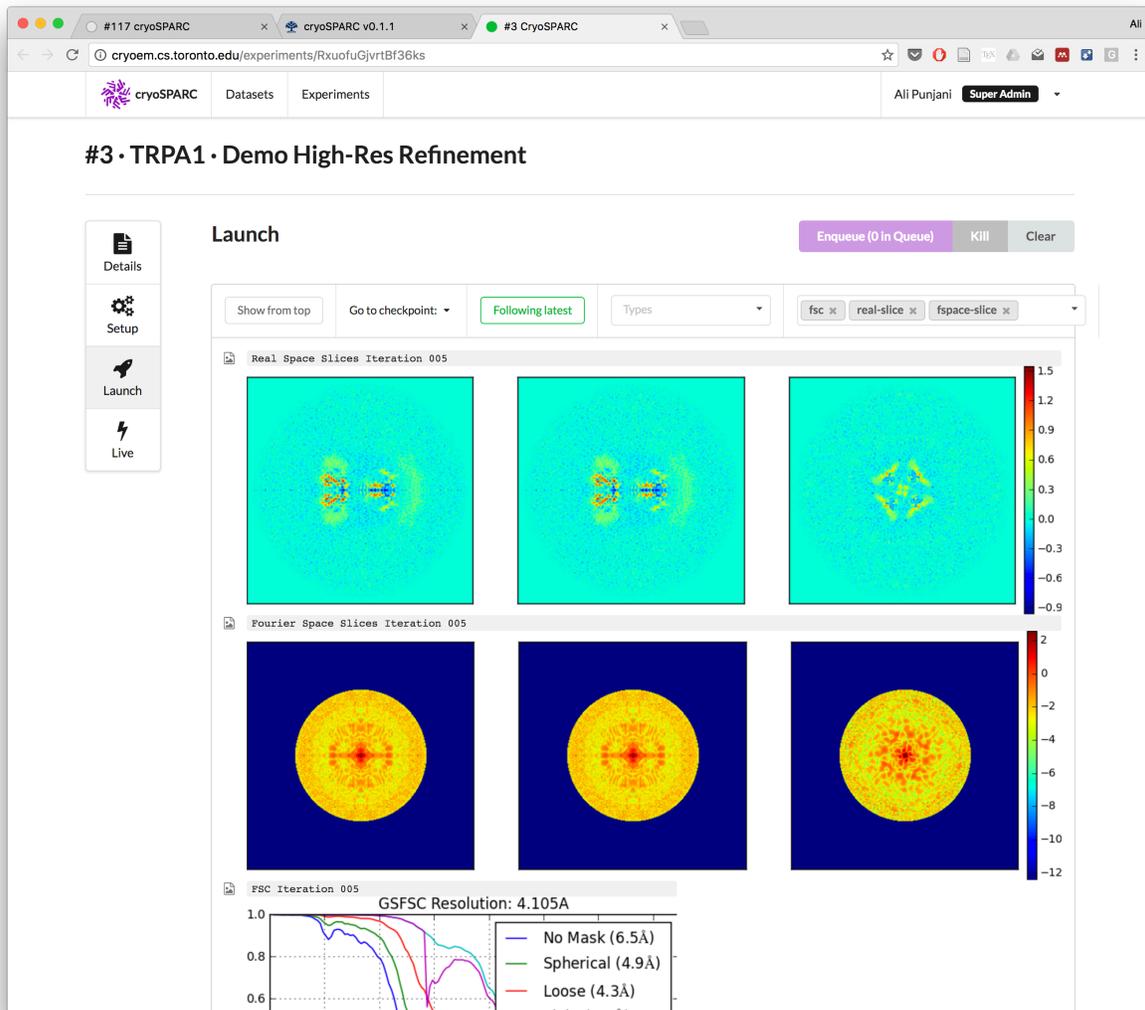
Figure 10. Results during a refinement experiment. Structures and FSC plots are generated in real time while computation progresses.
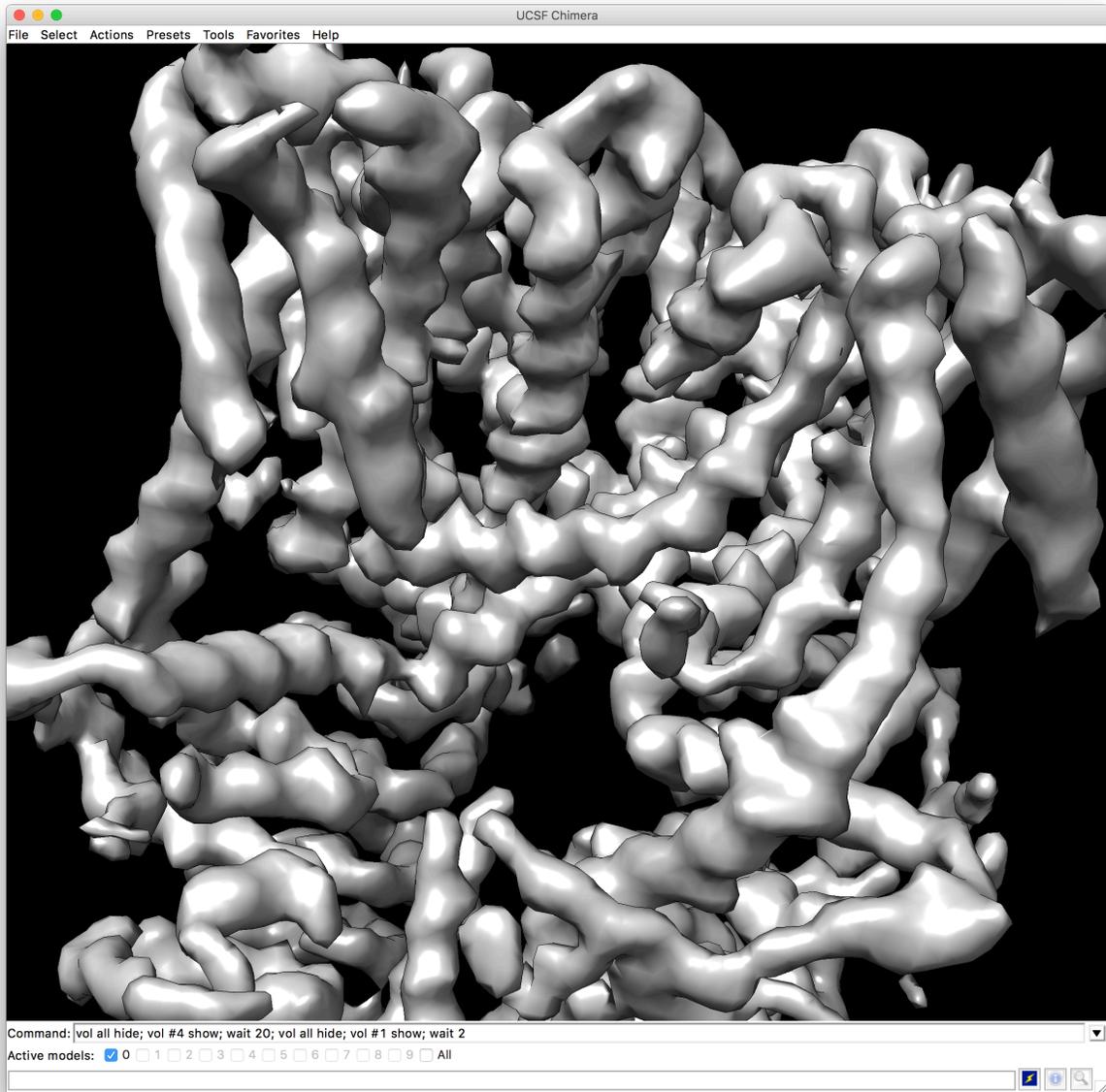
Figure 11. CryoSPARC result visualized in UCSF Chimera.