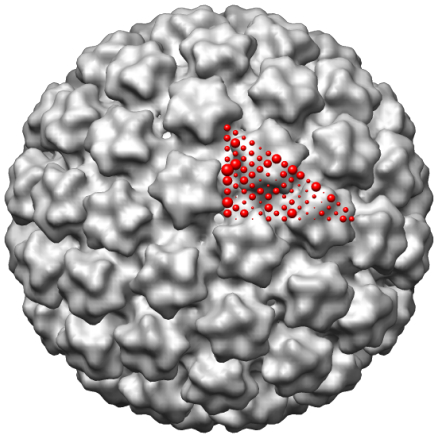
Scipion Tutorial Series

National Center for Biotechnology

Biocomputing Unit

Introductory Tutorial



Scipion Team, April 2018

Intended audience

This tutorial provides a general introduction to Scipion, an image processing framework to obtain 3D models of macromolecular complexes using Electron Microscopy (EM). It is designed for people without any prior knowledge of Scipion, however some basic knowledge about EM image processing is assumed and basic computer skills are required.

We'd like to hear from you

We have tested and verified the different steps described in this demo to the best of our knowledge, but since our programs are in continuous development you may find inaccuracies and errors in this text. Please let us know about any errors, as well as your suggestions for future editions, by writing to scipion@cnb.csic.es.

Contents

1 Getting started

1.1 Software Installation

To follow this tutorial you will need to have SCIPION properly installed in your system. To do so, you can execute the following commands:

```
git clone https://github.com/I2PC/scipion.git
cd scipion
./scipion config
./scipion install -j 5
```

You will also need to install additional EM software packages: CTFFind (any version) (Mindell and Grigorieff, 2003) Eman2.12 (Tang et al., 2007). Additionally, chimera package can be installed for volume visualization. For the full documentation please refer to the Scipion installation page.

1.2 Workflow summary

The main goal of this tutorial is to provide a quick introduction to EM image processing in Scipion. It will focus on Single Particles Analysis, illustrating the combination of different software packages. Following is a brief summary of the steps that will be done:

1. **Import micrographs**: register the set of micrographs into the project and provide important acquisition parameters (such as accelerating voltage, spherical aberration, pixel size etc).

2. **Downsample micrographs**: reduce the micrographs size to speed-up further computations.

3. **Estimate CTF**: Estimate the Contrast Transfer Function with CTFFind, select good quality micrographs. CTF correction itself is performed in the next steps.

4. **Particle picking**: Select particle coordinates using Xmipp3 manual/supervised picking (or Eman2 boxer).

5. **Extract particles**: extract the particles using selected coordinates. Some preprocessing operations are also applied to extracted particles.

6. **Import volume**: register a 3D volume into the project. This volume will be used as an initial reference for 3D refinement step.

7. **3D refinement**: Use Xmipp Projection Matching to iteratively refine the initial model using extracted particles.

2 Reconstruction of a Viral Capsid

In this demo, we will use the *single particle analysis (SPA)* approach to obtain a 3D reconstruction of a *Bovine Papillomavirus*. The EM images have been kindly provided by Dr. Grigorieff's Lab. Micrographs were collected at 300 kV and a calibrated magnification of 56,588, giving a pixel size of 1.237 Å (Wolf et al., 2010).

To start working on this tutorial type:

```
scipion tutorial intro
```

This command will download (if this was not done earlier) the test data in directory

`$SCIPION_HOME/data/tests/xmipp_tutorial`

Note that you will need write permissions for this folder. After downloading, a new project will be created. The project window will open as shown in Figure 1. The command we used has created a new project and imported an existing workflow template. This way, we have a basic template that will make it easier to follow the processing pipeline.

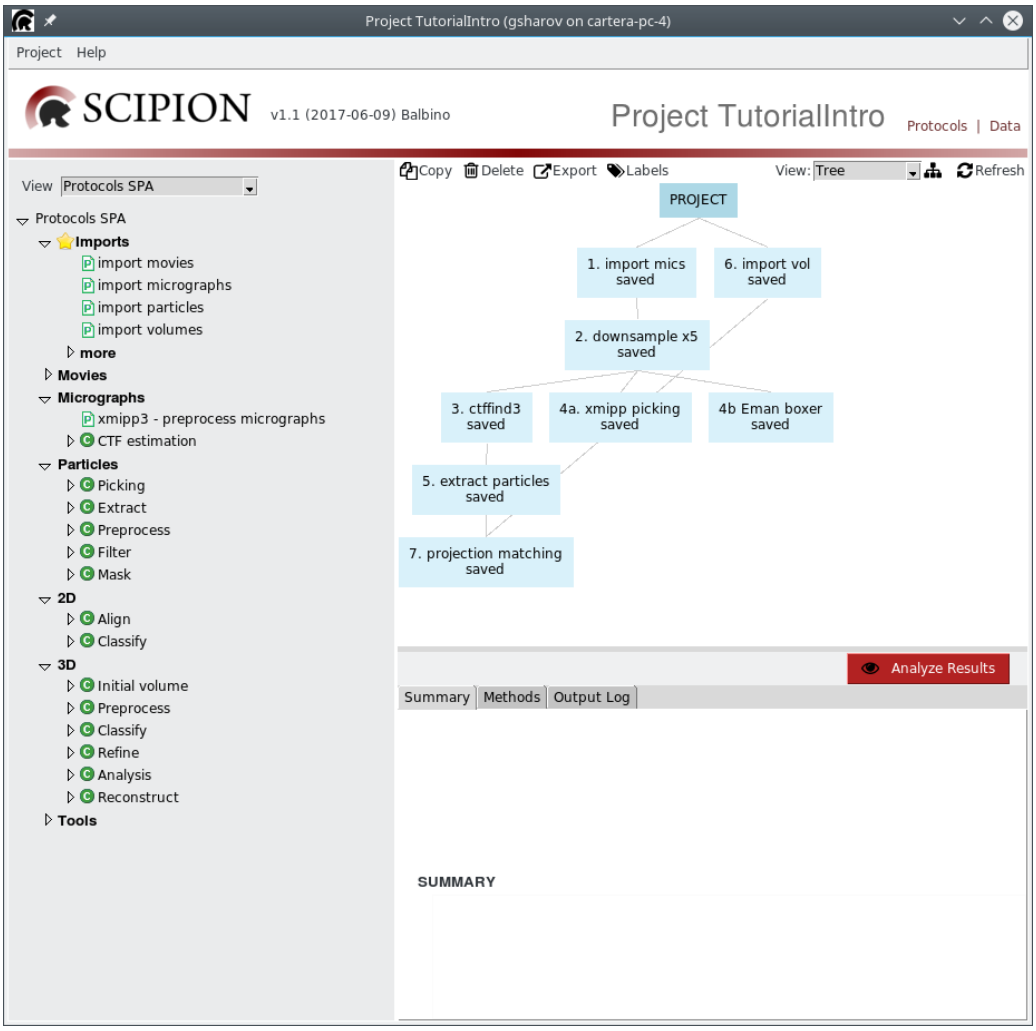


Figure 1: Project window with a pre-loaded workflow template.

NOTE The ability to export/import workflows in SCIPION is a great way to reproduce previous processing steps. It is particularly useful to repeat steps on similar samples or to share knowledge between users.

In the project window, the left panel shows a list with processing tasks (protocols) that can be used. The type of protocol list can be changed in the *View* menu (SPA is the default one). You can also search protocols by Ctrl+F. The right panel displays the tree of protocols executed (runs) by user and their state: saved, running, finished

or aborted. Users can visualize the runs in a list or tree view (by clicking on a View menu at the top right). Finally, the bottom right panel displays information for the selected protocol run, such as inputs and outputs, execution logs or documentation. The special Analyze Results button is used to visualize outputs and plot results.

2.1 Preprocessing

2.1.1 Import Micrographs

The first step is to import the micrographs into your Scipion project. To do this, double-click the 1. import mics box (or use left menu Imports » import micrographs), modify input as explained below and click on Execute to run the protocol. Micrographs are located in the folder:

```
$SCIPION_HOME/data/tests/xmipp_tutorial/micrographs/*.mrc
```

Acquisition info that you have to provide is shown on Figure 2. After execution, the project window will display new information as shown in Figure 3. If we press the Analyze Results button, imported micrographs will be loaded.

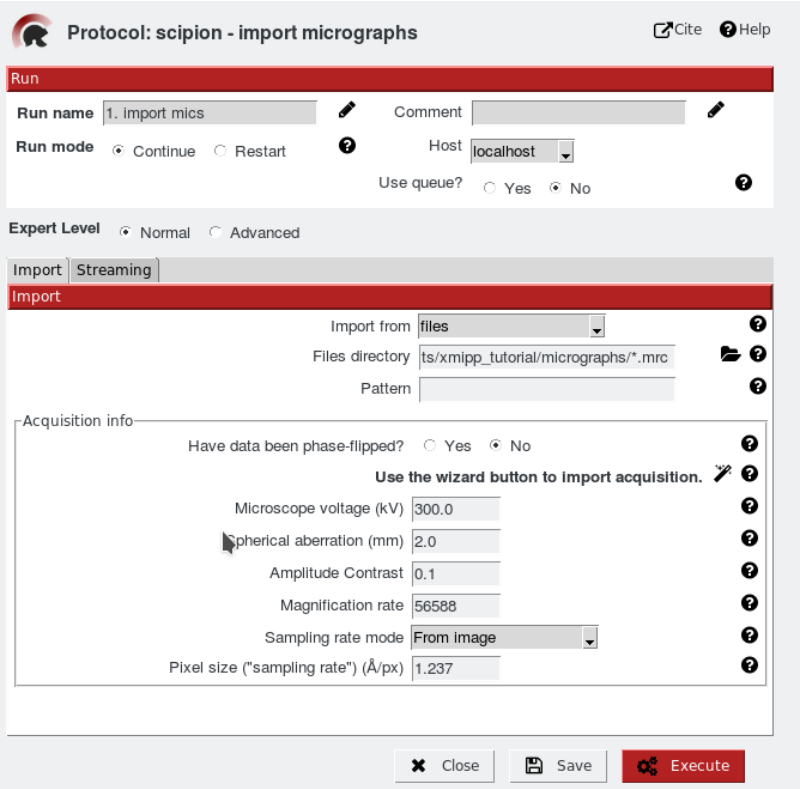


Figure 2: **Import Micrographs** Protocol GUI.

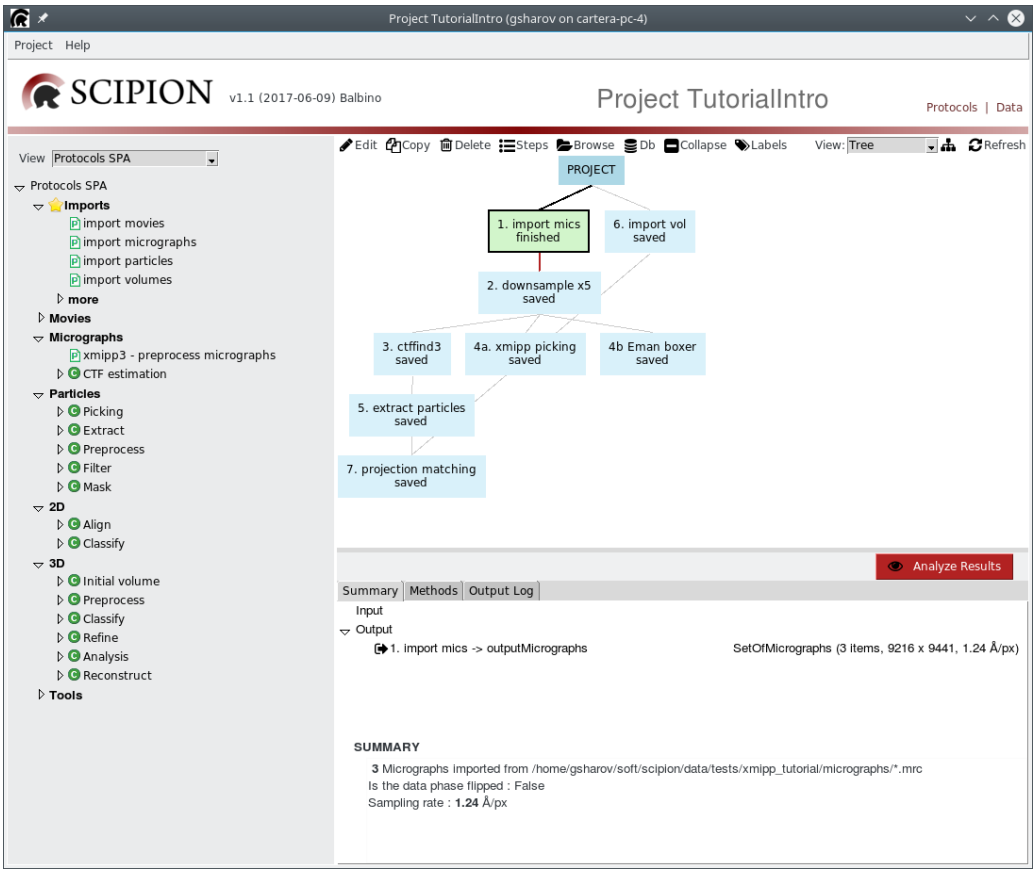


Figure 3: Project window after **Import Micrographs** protocol has finished.

2.1.2 Downsample Micrographs

After importing the micrographs we can perform the next step: **Downsample Micrographs**. Just double click on 2. downsample x5 box (or use left menu Micrographs > xmipp3 - preprocess micrographs) . This protocol combines multiple Xmipp programs that perform different operations on the micrographs (Figure 4). We want to reduce the micrographs size by a factor of 5 (this operation is usually called downsampling or binning).

The *invert contrast* flag must be activated, so that particles will become white over a dark background. Now press the Execute button and wait till the protocol finishes.

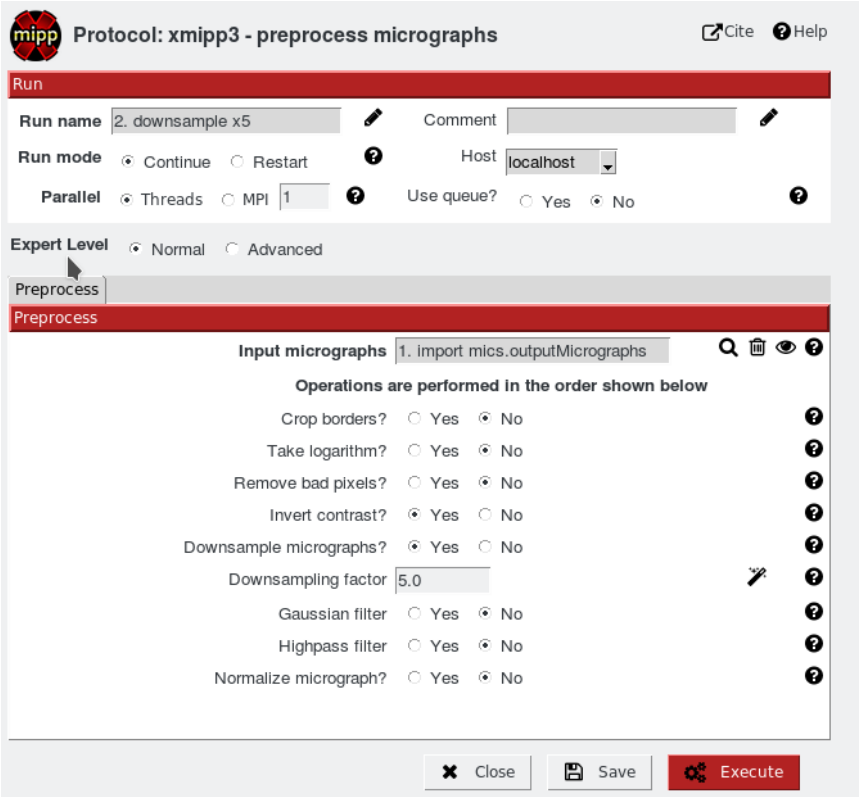


Figure 4: **Preprocess Micrographs** Protocol GUI.

2.1.3 CTF Estimation

The next step is to estimate the CTF (Contrast Transfer Function) of the micrographs either using *CTFFind* (Mindell and Grigorieff, 2003) or *Xmipp CTF estimation*. These protocols estimate the PSD (Power Spectral Density) of the micrographs and the parameters of the CTF (defocus U, defocus V, defocus angle, etc). They cut the micrographs into several smaller patches with a desired window size. After that, they compute the Fourier Transform of each patch and calculate an average.

To estimate the CTF you will need to provide parameters describing the frequency region to be analyzed. The parameters shown in Figure 5 are the proper ones for this example and are preloaded for you in the 3. ctffind3 box, so just double click on it (or use left menu Micrographs > CTF estimation > grigoriefflab - ctffind). If you

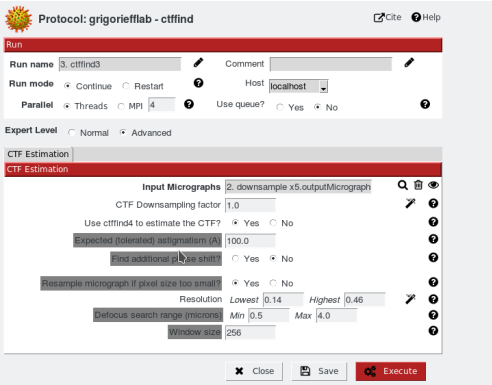


Figure 5: CTFFind protocol GUI.

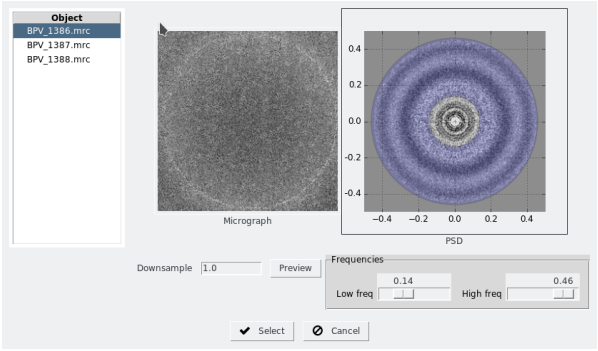


Figure 6: Wizard for choosing the frequencies.

want to try Xmipp3 for CTF estimation, double click on Micrographs > CTF estimation > xmipp3 - ctf estimation - this protocol is similar to CTFFind protocol.

The limiting frequencies must be such that all zeros of the PSD are contained within those frequencies. There is a wizard, shown in Figure 6, that helps you to do that. To see all available options, choose the `Expert Level` : **Advanced**. Please note that the range of defocus values can differ from the one used in this example, and it must be chosen according to your data set. After verifying all input parameters, press the Execute button and wait till the protocol finishes.

The CTFs of good micrographs typically show multiple concentric rings, as in Figure 7 left, extending from the image center towards its edges. Bad micrographs may lack rings or have very few rings that hardly extend from the image center. Another reason to discard micrographs may be a presence of strongly asymmetric rings (astigmatism, Figure 7 center) or rings that fade in a particular direction (drift, Figure 7 right).

When the protocol (either CTFFind or Xmipp CTF estimation) is finished you may click on the Analyze Results button (Figure 8). To discard micrographs with bad CTFs you may click on them with the mouse right button and press **Disable**. Once you finish the selection, press on the Micrographs button to create your subset of micrographs containing only enabled items.

Sometimes the CTF estimation algorithm may fail to find the rings even if they can be seen by eye. If this is the case, you may help the algorithm to find the rings

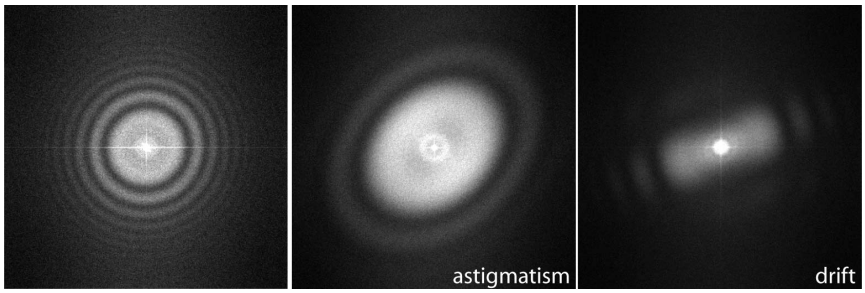


Figure 7: CTFs of a good, astigmatic and drifting micrograph, respectively.

by clicking on the image with the mouse right-button and choosing **Recalculate CTF** in the menu. A graphical interface will help you to correctly identify the rings. You must provide the first CTF zero and the search range, and then press OK. When you finish, press the Recalculate CTFs button.

2.1.4 Particle Picking

Now you are ready to pick particles from micrographs. You can pick particles with either *Xmipp3 manual picking* or *Eman boxer* tool (You can find more picking protocols in Particles » Picking). We illustrate here particle picking with both Eman and Xmipp using a box size of 110 pixels.

Xmipp particle picking

Double click on 4a. xmipp picking box (or use left menu Particles » Picking » xmipp3 - manual picking) and press the Execute button.

This Xmipp picking protocol has manual and semi-automatic picking mode. Last one can be launched via Activate training button. Particles can still be added or removed manually after auto-picking to correct the mistakes. When you are done picking particles you can click on Coordinates to register the output. For more information, you can address to picker documentation.

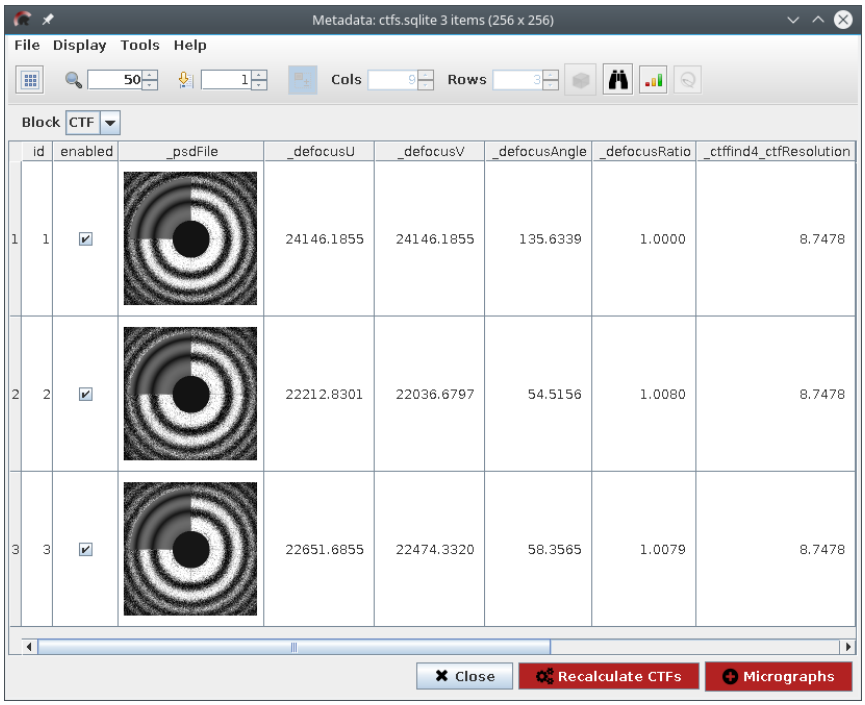


Figure 8: Output of CTFFind protocol showing estimated CTFs for all micrographs.

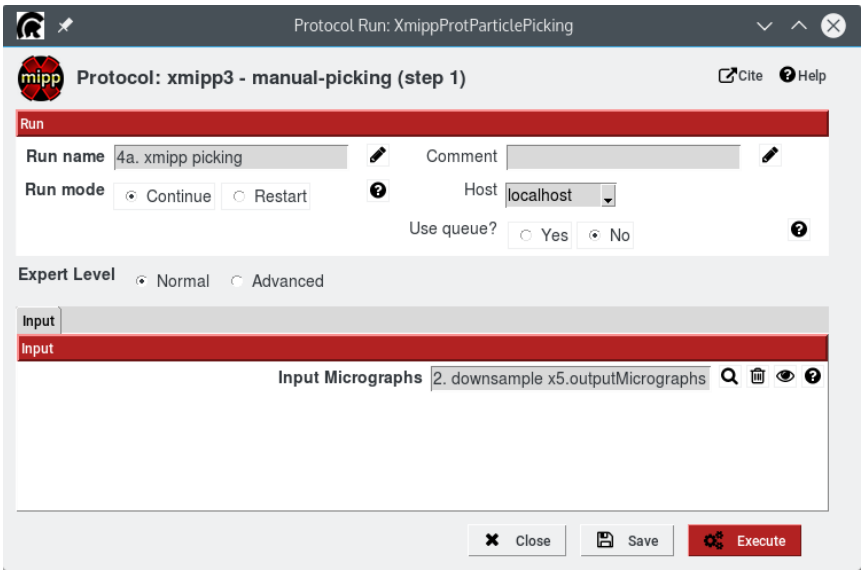


Figure 9: Xmipp manual/supervised picking protocol GUI.

Following is a summary of the control actions that you will need during picking:

- Use ⇧+mouse wheel in the overview window to zoom in and out.
- Mark particles with the mouse left button. You may adjust its position by clicking the left mouse button on the selected particle and dragging it to the new position.
- Use ⇧+left mouse on a selected particle in order to remove it.
- You can apply filters to the micrographs, so that you may see the particles better. Select in the menu Filters as many filters as you like.

EMAN2 boxer

Double click on 4b. Eman boxer box (or use left menu Particles » Picking » eman2 - boxer) and press the Execute button - this will start the EMAN2 picking interface. Please enter 110 for both box size and particle diameter for better results.

Eman2 boxer provides manual picking and several modes of automatic picking. Please refer to its webpage for further information. When you are done picking

particles, press Done. You will be asked if you want to register output. If you choose Yes, output coordinates are generated.

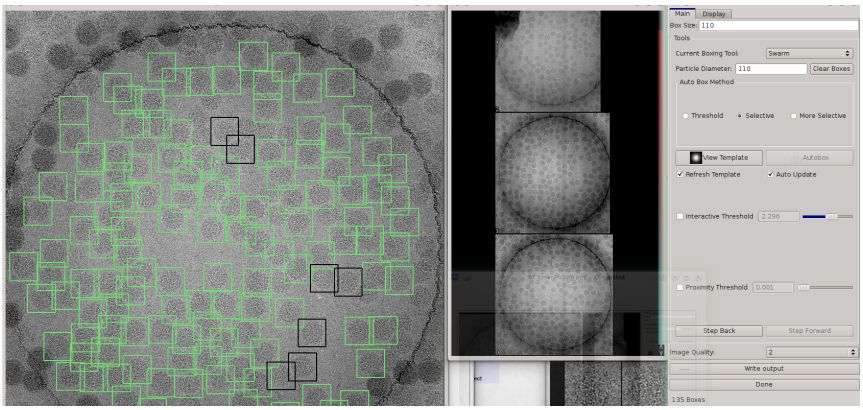


Figure 10: EMAN2 boxer, an interface for particle picking.

2.2 Extract Particles

Double click on 5. extract particles box (or use Particles » Extract » xmipp3 - extract particles). This will launch our next protocol that will allow you to extract, normalize and correct the CTF phases of your picked particles, among other things. Modify the default parameters according to the Figures 11 and 12. The following parameters are available:

- The *coordinates* of the particles in the micrographs, which are taken from the results of the previous step. Also, in the same tab, set the *particle box size* in pixels (in our case 110 px).
- The *invert contrast* flag. If activated, bright regions become dark regions and the other way around. This flag should not be set for that tutorial since we have already inverted the contrast of the micrographs.
- The *phase flipping* flag. If activated, the protocol corrects the CTF phases of your particles. We need to activate this flag since we plan to use projection matching in Xmipp.

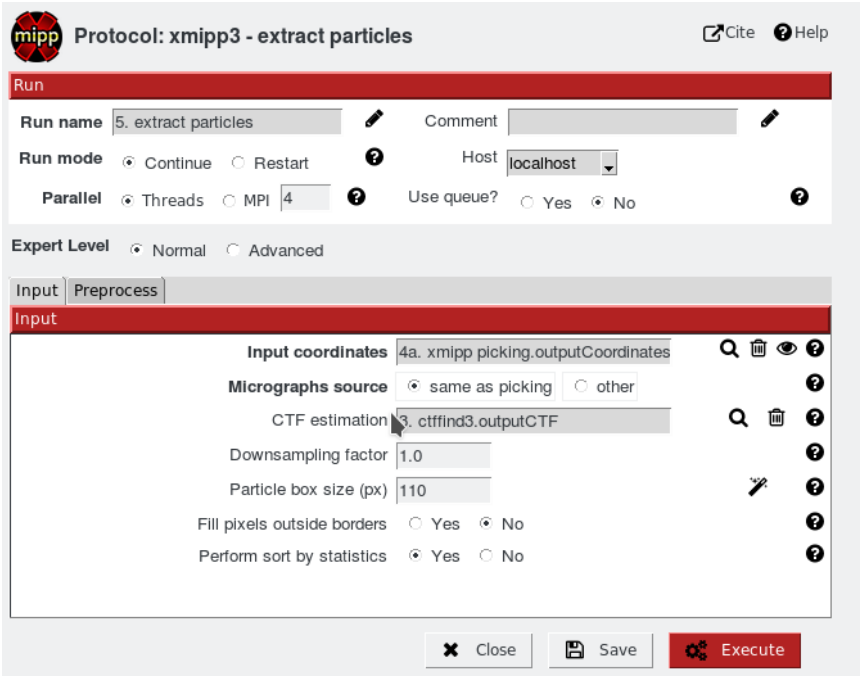


Figure 11: Extract particles protocol GUI (`Input` tab)

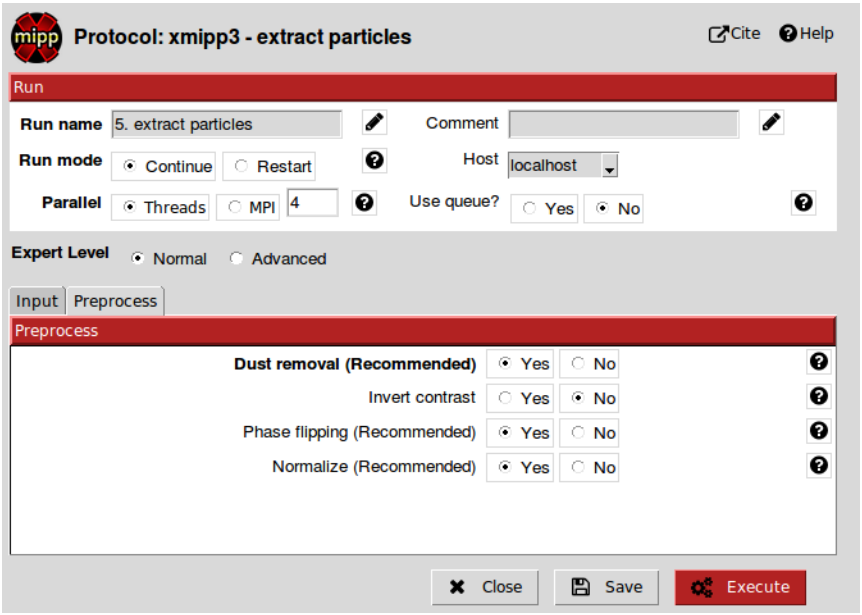


Figure 12: Extract particles protocol GUI (`Preprocess` tab)

- The *normalize* flag. If activated (recommended), the particles are normalized to have zero mean and a standard deviation of one for the background pixels.

The protocol also can sort the particles based on general statistics assigning to each particle a Z-score value. Particles with low Z-score are reliable and the ones with large Z-score are normally outliers. After the protocol has finished, press the Analyze Results button in the main window to check the extracted and normalized images.

2.2.1 Particles Selection

By default, particles displayed in gallery mode are sorted by Z-score. To see the score associated with each particle you can switch to the table view by pressing top left button. Together with particles window the Z-score plot will pop up. You can sort the particles by Z-score and choose the right cutoff using this plot. In this case a good value is in range between 2 and 3. To remove bad particles, you can right click on them and choose "Disable". For more information on the viewer, look at the documentation. To create a new set of particles (only enabled ones) you can click on Particles button (see figure 13).

2.3 3D Reconstruction: Projection Matching

Having the right angles for each projection image is crucial for making a 3D reconstruction. However, in a single particle experiment you don't know a priori the angles and you have to estimate them. The most popular way of estimating them is by comparing the projections of a volume that is similar to the volume to be reconstructed (initial model) with the images obtained from the microscope.

A possible approach is to generate equidistant projections from the initial model. The particles of the experimental data set are then compared (for example, by cross-correlation) to each reference projection. A "similarity" coefficient (for example, cross-correlation coefficient) is calculated between each experimental particle and reference projection. Each individual experimental particle is matched to the reference projection that gave the highest "similarity" coefficient. Therefore, it is assumed

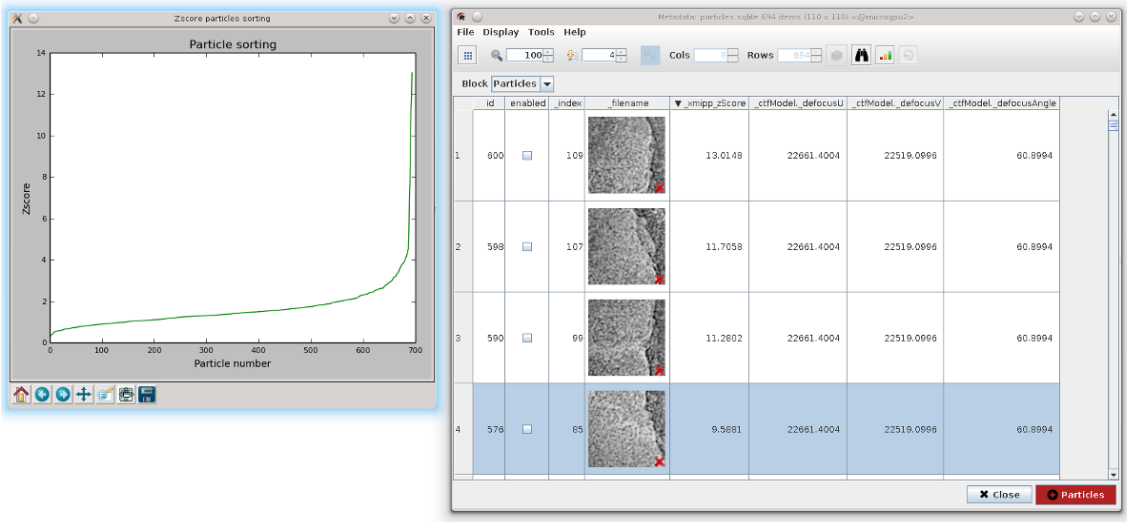


Figure 13: Analyze Results of the **Extract Particles** protocol: Z-score plot and table view of extracted articles. sorted by Z-score. Red crosses mark disabled particles.

that this experimental particle was projected with the same Euler angles as the reference projection.

Since an initial map is necessary to start the protocol, we should first import it. Double click on the 6. import vol box (or use Imports » import volumes) and import from folder

`$SCIPION_HOME/data/tests/xmipp_tutorial/volumes/`

the volume `BPV_scale_filtered_windowed_110.vol` with *sampling rate* 6.185.

To run the projection matching protocol double click on the 7. projection matching box or go to 3D » Refine » xmipp3 - projection matching. Input parameters are shown in Figures 14, 15, 16 and 17. We have pre-selected them in this tutorial workflow for your convenience and below we explain only some of them. For more details for each parameter please use the respective **Help** button.

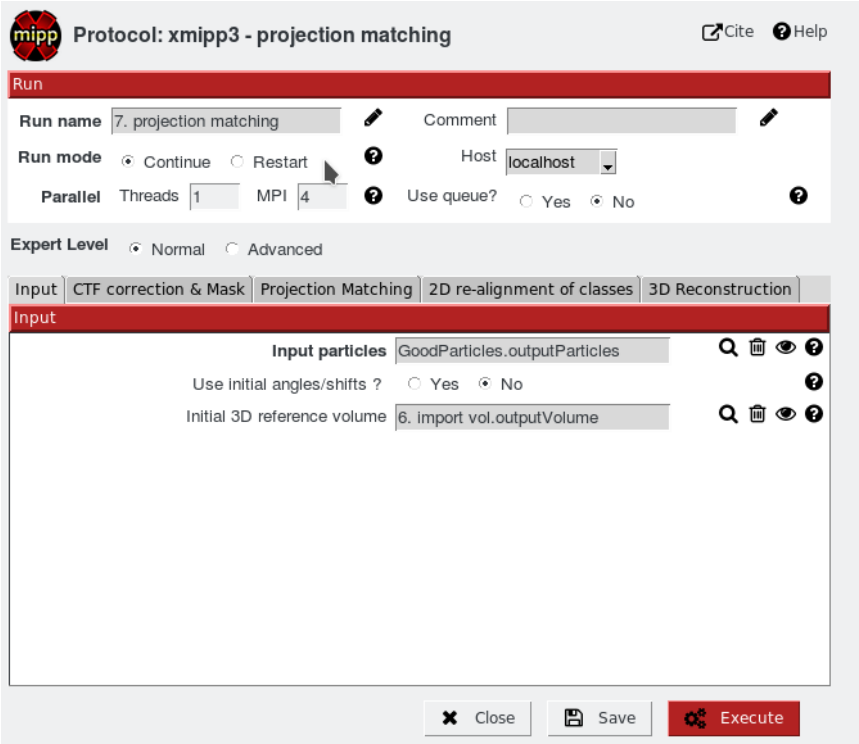


Figure 14: **Projection Matching** protocol GUI: **Input** tab.

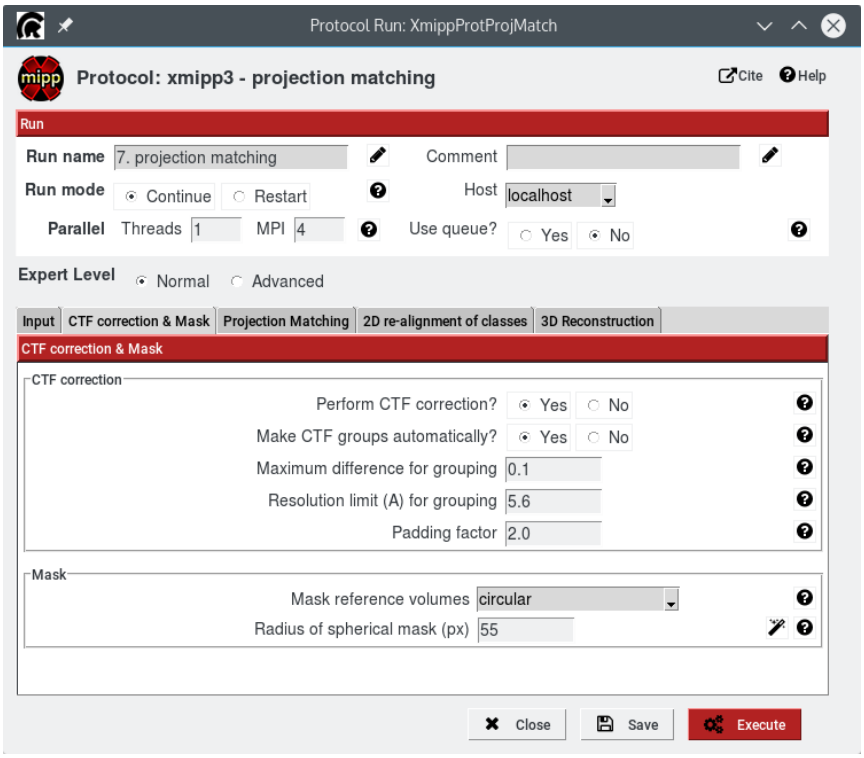


Figure 15: **Projection Matching** protocol GUI: **CTF Correction & Mask** tab.

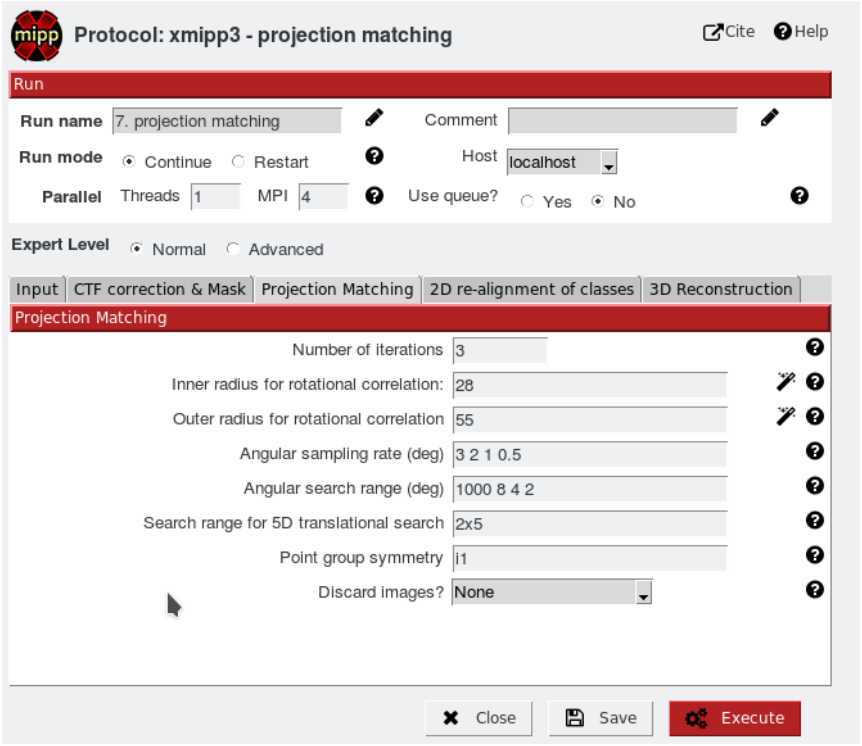


Figure 16: **Projection Matching** protocol GUI: **Projection matching** tab.

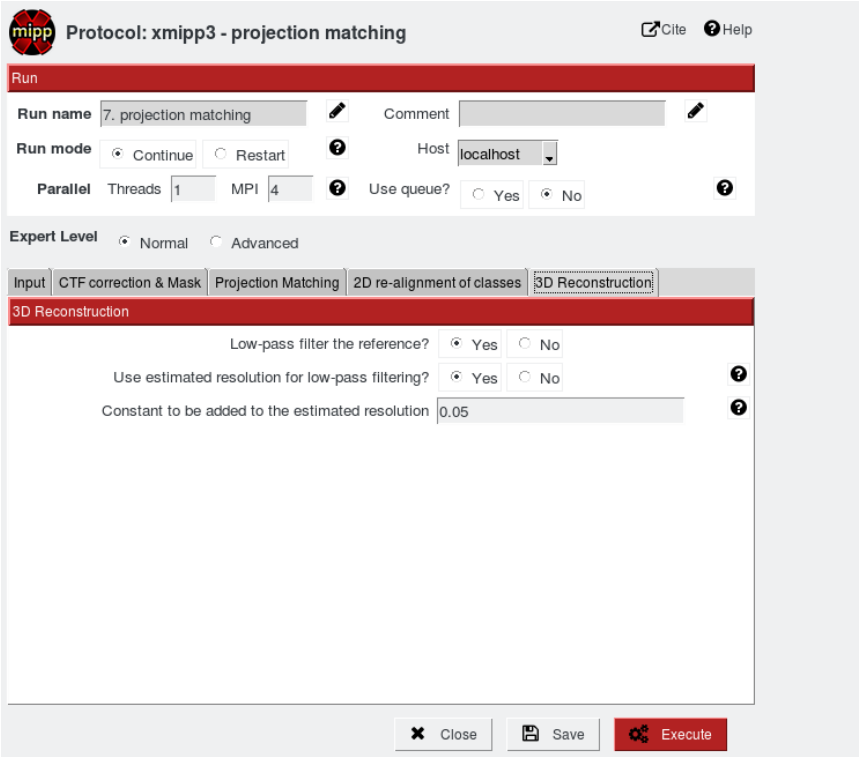


Figure 17: **Projection Matching** protocol GUI: **3D reconstruction** tab.

We only need to check the key parameters explained below and execute protocol.

- *Input particles:* Previously created set of good particles.
- *Initial 3D reference volumes:* Initial volume that has been imported in the previous step.
- *Mask reference volumes:* Circular. Masking the reference volume will increase the signal to noise ratio.
- *Radius of spherical mask(px):* 55. Since particles size is 110, radius of spherical mask should be 55.
- *Point group symmetry:* i1. This is the symmetry type of our virus.

When the protocol is finished, in order to visualize the reconstructed volume, click on Analyze Results button. Protocol viewer window will pop up (Figure 18).

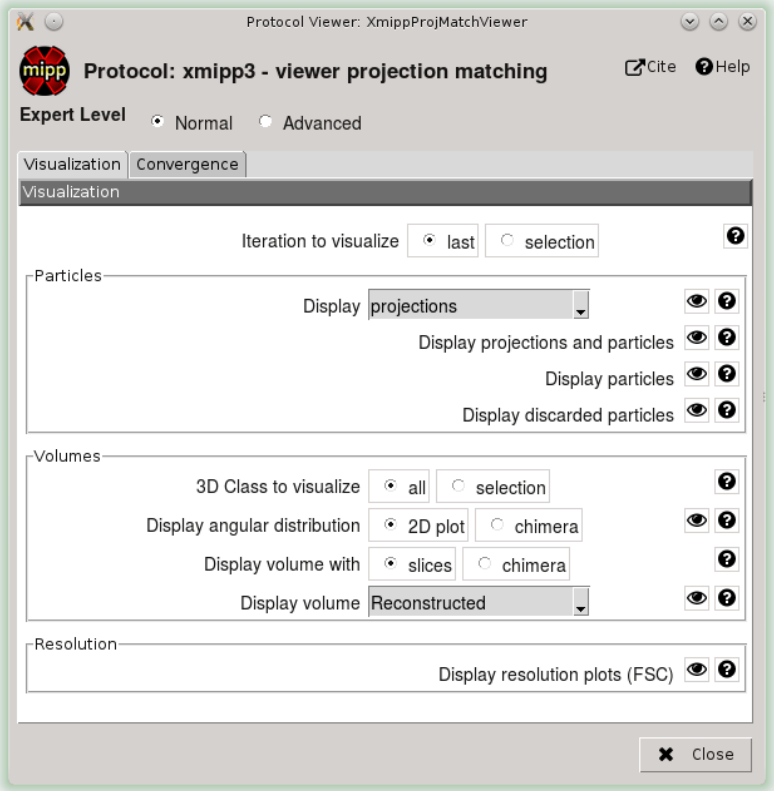


Figure 18: Projection matching protocol viewer.

In the field *Display volume: Reconstructed* press the **Eye** button and you will see the volume slice by slice (Figure 19). To visualize the volume with Chimera, now click on Chimera icon at the top right of the window (Figure 20).

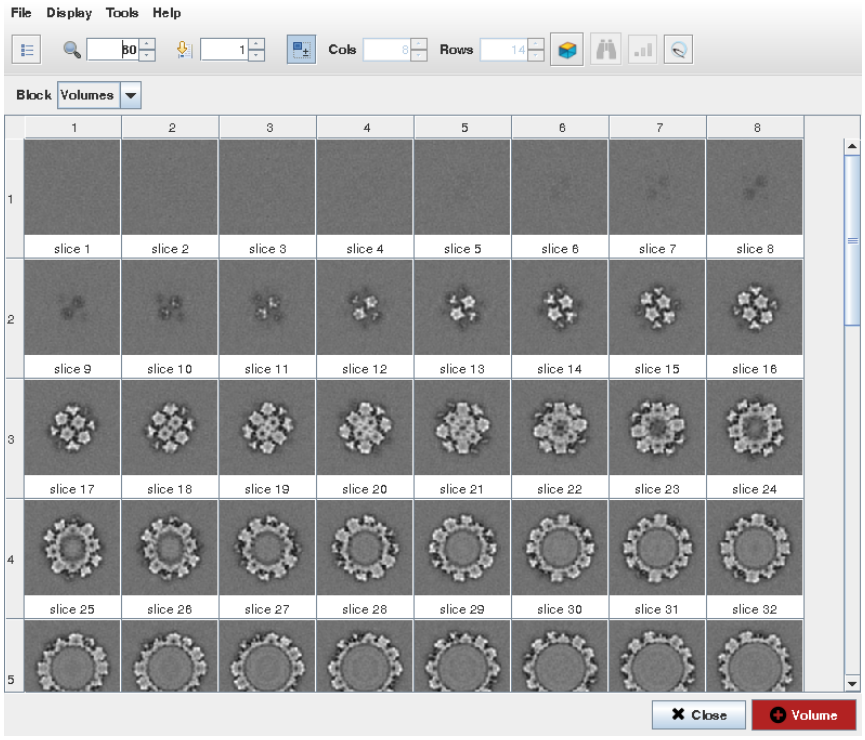


Figure 19: Reconstructed volume slices.

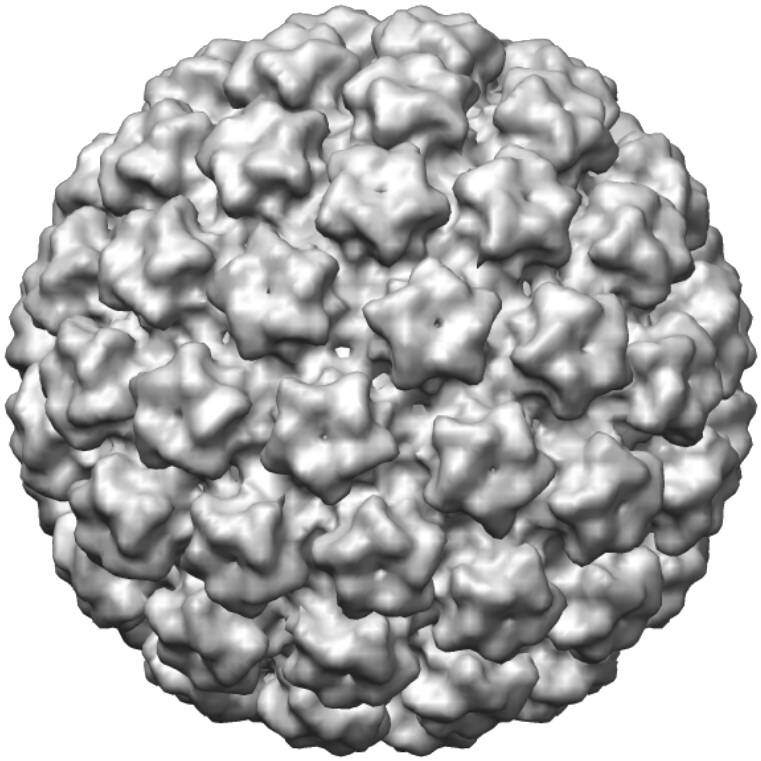


Figure 20: Reconstructed volume in Chimera.

References

Mindell, J. A. and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *JSB*, 142:334 – 347.

Tang, G., Peng, L., Baldwin, P., Mann, D., Jiang, W., Rees, I., and Ludtke, S. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J. Struc. Bio.*, 157:38–46.

Wolf, M., Garcea, R. L., Grigorieff, N., and Harrison, S. C. (2010). Subunit interactions in bovine papillomavirus. *Proc Natl Acad Sci U S A*, 107(14):6298–6303.