Scipion Tutorial Series

National Center for Biotechnology

Biocomputing Unit

# Initial model Tutorial

Scipion Team, October 2017

## Intended audience

This tutorial provides an introduction to initial volume estimation using Scipion. Very limited knowledge about 3D-EM image processing and Scipion is required, however basic computer skills are necessary. You might want to go through Introductory tutorial first to get to know your way around Scipion.

## 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 error you find, as well as your suggestions for future editions by writing to scipion@cnb.csic.es.

# Contents

# 1 General Introduction

## 1.1 Software installation

The first step before start working on this tutorial is to download and install Scipion and EMAN2 package (Tang et al., 2007). In Scipion wiki there is all the information needed to download and install the packages.

## 1.2 Initial Volume estimation

Single Particle Analysis (SPA) techniques can obtain three-dimensional (3D) maps of biological complexes at near atomic resolution by combining tens of thousands of projection images obtained in a Transmission Electron Microscope (TEM). In general, the reconstruction process leading to the final 3D map requires the use of an approximate low resolution initial model to be refined in further steps. In this tutorial we present three different methods to obtain such initial model using Scipion: Random Conical Tilt, 3D-RANSAC and Eman2.

# 2 Random Conical Tilt

One of the oldest 3D reconstruction techniques in TEM field is the Random Conical Tilt (RCT) (see Radermacher et al. (1987)). In RCT, we assume that identical protein particles are randomly oriented within the sample, so when we take a low-dose TEM image, we're getting multiple views of the same structure. The specimen stage of the microscope is first tilted and a low-dose TEM image is taken (figure 1A). A second image of the same specimen area is taken at 0 degrees (figure 1B). These two images represent a tilt pair. For each tilt pair of images any processing program will need to relate each particle to its tilted version: estimate the tilt axis position and tilt angle. After several tilt pairs are collected, single particles in untilted images can be aligned (and classified) and this alignment will determine spatial location of corresponding tilted particle in each pair (figure 1C). Aligned untilted images within single class represent the same view (figure 1D). Now a unique orientation of tilted

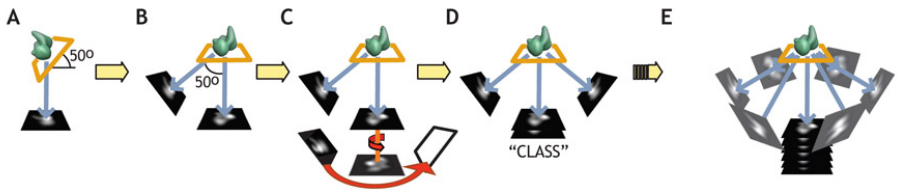


Figure 1: Random conical tilt reconstruction method (fig. from Leschziner's lab website).

particles is determined and a 3D reconstruction can be produced for each class (figure 1E).

## 2.1 Getting started

The test data you will work on can be downloaded using the following command:

```
scipion testdata --download rct
```

All the files will be downloaded in `$SCIPION_HOME/data/tests/rct`.

After the download, you must launch the main GUI by typing: `scipion` Then, create a new project by clicking Create project button, type a *Project name* and click OK. The main project window will appear. On the top of the left panel, there is a **View** list with separate categories of protocols: SPA, RCT, MDA etc. To follow the tutorial you should select the **Random Conical Tilt** category in this list.

## 2.2 Import tilted pairs of micrographs

The first step is to import the tilted pairs of micrographs into your Scipion project. To do this, press on **Import Micrograph Pairs** protocol. In *Pattern untilted* and *Pattern tilted* you must indicate where your untilted and tilted micrograph files are stored (click on Browse button). The complete patterns are:

```
$SCIPION_HOME/data/tests/rct/micrographs/F_rct_u_*.tif
```

```
$SCIPION_HOME/data/tests/rct/micrographs/F_rct_t_*.tif
```

In this first version of the protocol pairs assignment is done by micrograph order but in next versions a wizard will be provided.

Modify the parameters of the **Import Micrograph Pairs** protocol according to the ones shown in figure 2.

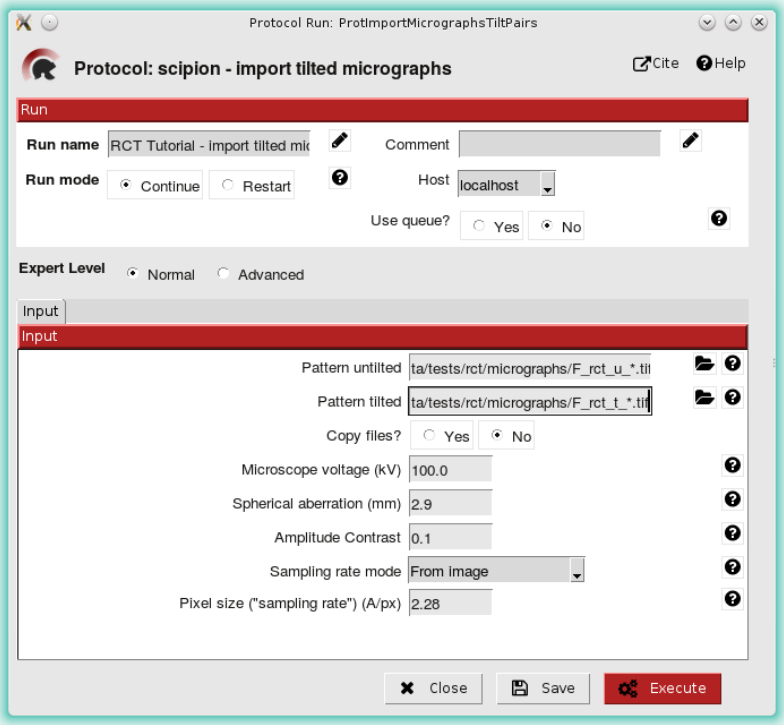


Figure 2: Import Micrograph Pairs protocol GUI.

When you have completed the form, click on the Execute button.

Once the protocol has finished you can press the Analyze results button and a new window will open where you can see the imported micrographs pairs as shown in figure 3.

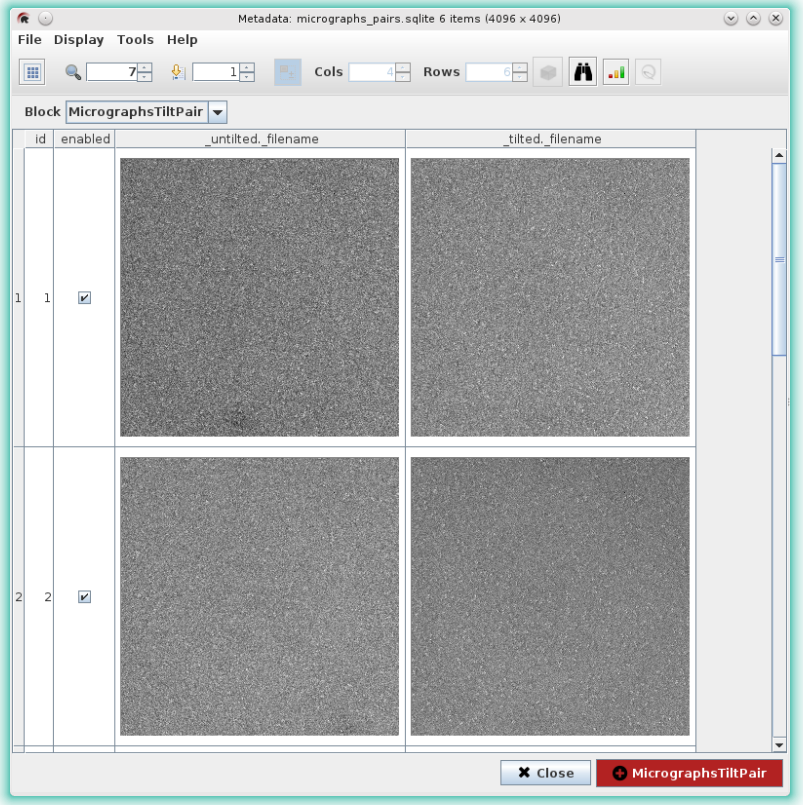


Figure 3: Tilt pair micrographs have been imported.

## 2.3 Particle Picking

Once the tilted pair micrographs have been imported you need to pick both tilted and untilted particles. There are two ways for performing this task, a manual picking and an automatic picking which will be shown in section 2.5. In this section the manual picking procedure is introduced. Press on **Picking micrograph pairs** protocol and the form shown in figure 4 will appear. Select the Micrographs Tilt Pair object produced by the import protocol as input and click on the Execute button. The Xmipp particle picking GUI will open.

If you find some particles that are difficult to see due to noise and low contrast of the images, you can use different filters offered by the Xmipp GUI in the Filters

menu. For instance you might select the Brightness/Contrast filter and press the Auto button that will set the brightness and contrast to the optimal values for the selected micrograph. Then switch to another micrograph and do the same. The resulting interface will be much easier to start picking.

The picking procedure includes the following steps:

1. Find the first particle in both untilted and tilted micrograph. First, pick the untilted one and then its tilted pair. Set particle size to 120 px.

2. Pick three more particle pairs like the first one, paying attention to the correct position of the tilted ones, so the tilt estimation can be properly done. During the picking of the first few pairs, the manual error correction will help the algorithm to learn and get a better tilt parameters estimation. To carefully locate a corresponding tilted particles, it is better to zoom in into the micrograph. Note that zooming in or out in the untilted micrograph will also affect the tilted one, but not the other way around. To zoom in, press ⇧+scroll up, to zoom out, press ⇧+scroll down. To see more options, go to Help > Tips or address to the documentation.

3. Finally, continue picking of only untilted particles (tilted ones will automatically be picked), but keep an eye on the tilted ones in order to prevent its position deviation due to accumulated error. The amount of particles to pick for a good reconstruction varies depending on the tilt angle, the quality of the micrographs and many other factors, but a good approach to start could be 2000 particle pairs.

However, if you don't feel like picking manually thousands of particles you can import the already picked particle pairs by clicking on File > Import coordinates and specifying the following path:

```
$SCIPION_HOME/data/tests/rct/positions
```

Take into account that the protocol status will not change to **Finished** but will remain as **Interactive**. This allows you to execute it again if you wish to pick more particles.

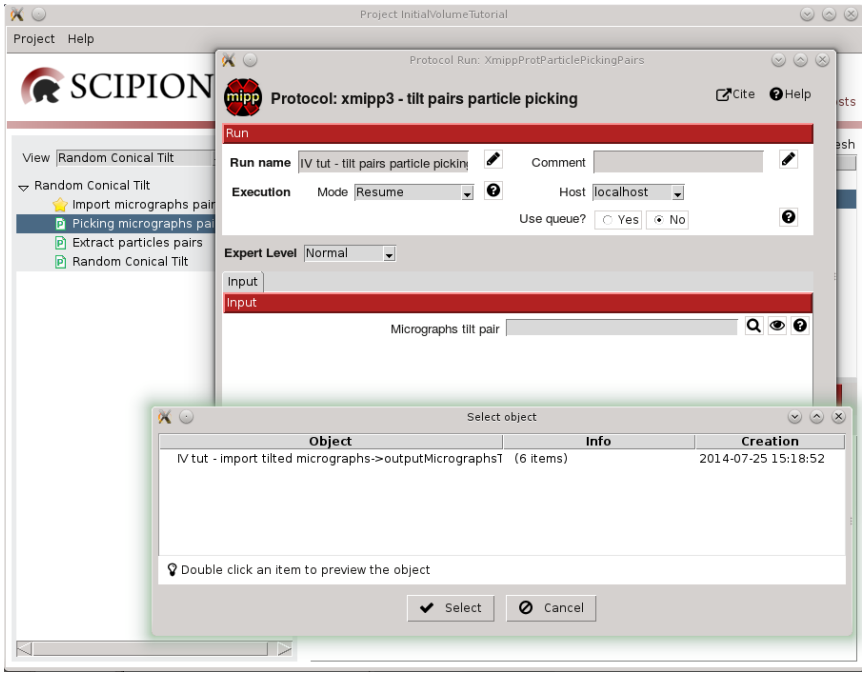


Figure 4: Picking micrograph pairs protocol GUI.

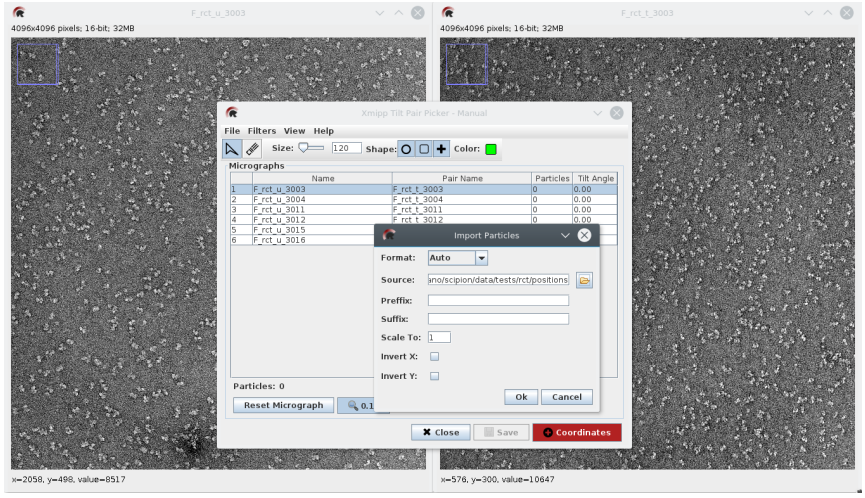


Figure 5: Xmipp GUI for picking particle pairs.

## 2.4 Extract particle pairs

Now that coordinates of the particles are known they have to be extracted as particle pairs. To do so, select the Extract particles pairs protocol and fill in the parameters as shown in figure 6. Although particles were picked with a box size of 120 pixels, we will downsample them by a factor of two to speed up next steps. On the Preprocess tab select **Invert contrast: No**

Once the protocol has finished, you can review the particles by clicking on the Analyze results button as shown in figure 7. If you don't like some of the particles, you can disable them and create a subset of the good ones. Just click on the ParticlesTiltPair red button and give the new subset a name.

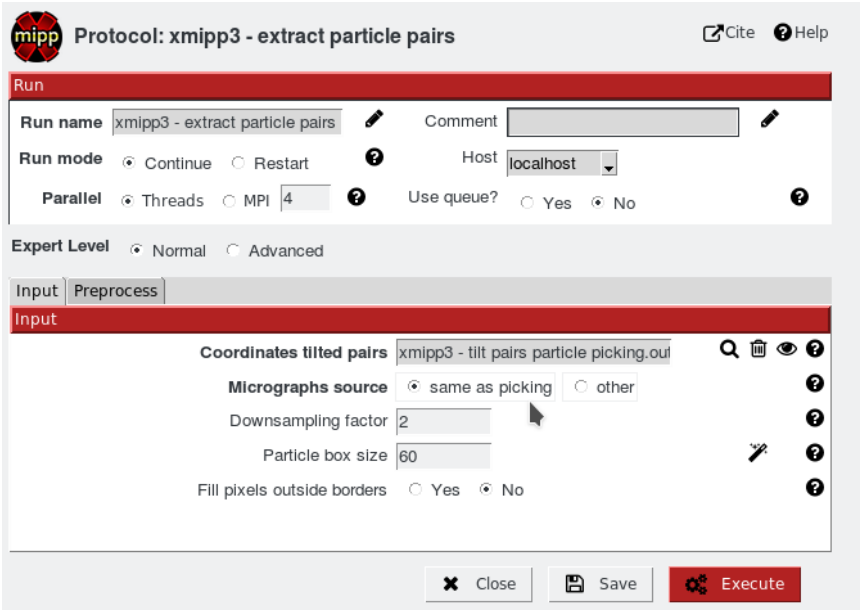


Figure 6: Extract particle pairs protocol GUI.

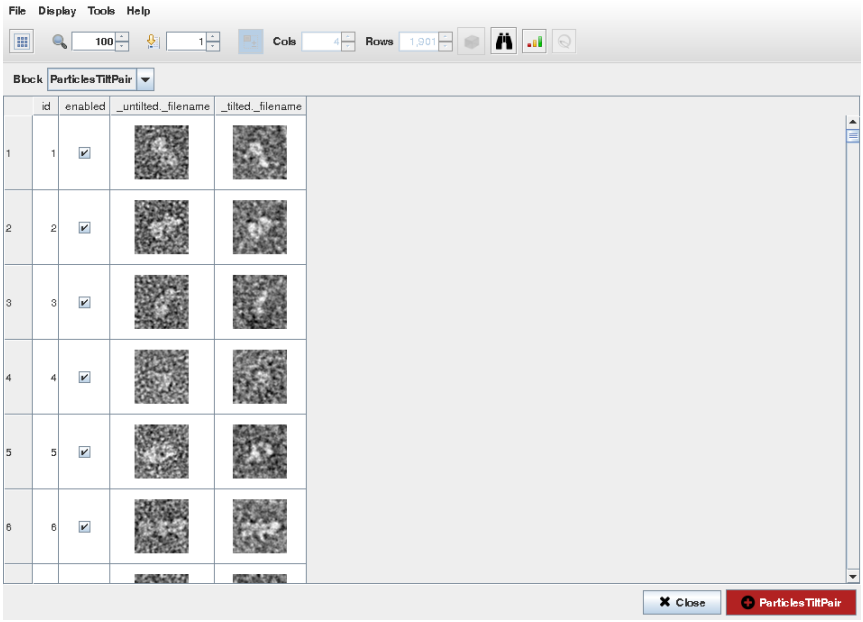


Figure 7: The results of Extract particle pairs protocol.

## 2.5 Automatic picking of tilt pairs

Picking particle pairs can be tedious due to the fact that is a manual process. In addition, as it was mentioned in the previous section, picking particles may be also difficult because of the low contrast and the low signal to noise ratio of the micrographs. For this reason, an automatic tilt pairs assignment has been implemented in Scipion via **Xmipp3 - Assignment tiltpair** protocol. It allows to find tilt parameters between the coordinates of particles in untilted and tilted micrograph, that were picked separately.

The workflow for automatic assignment of tilt pairs consists of five steps, see figure 8:

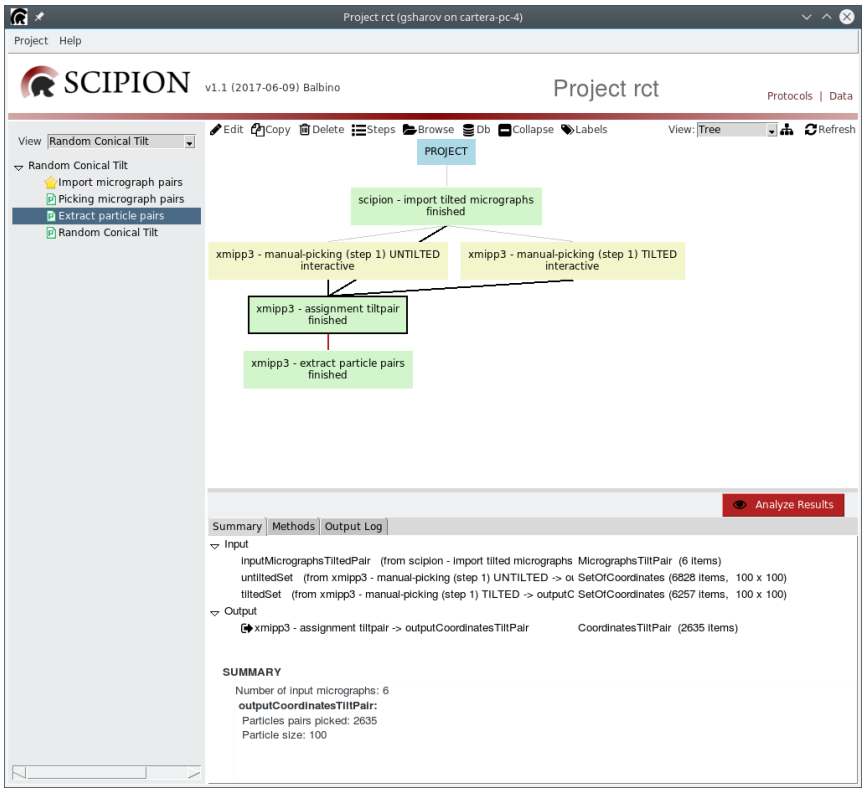


Figure 8: Workflow for automatic assignment of tilt pairs.

1. Import tilt pair micrographs, as described in 2.2.

2. Pick particles in untilted micrographs using any available picking protocol from SPA menu: **xmipp - manual-picking(step 1)** and **xmipp - auto-picking(step 2)**, **igbmc - auto-picking** etc.

3. Pick particles in tilted micrographs.

4. Run **Xmipp3 - assignment tilt pair** protocol: fill the protocol parameters as shown in figure 9. This protocol is also located in **SPA** view.

5. Extract particle pairs.

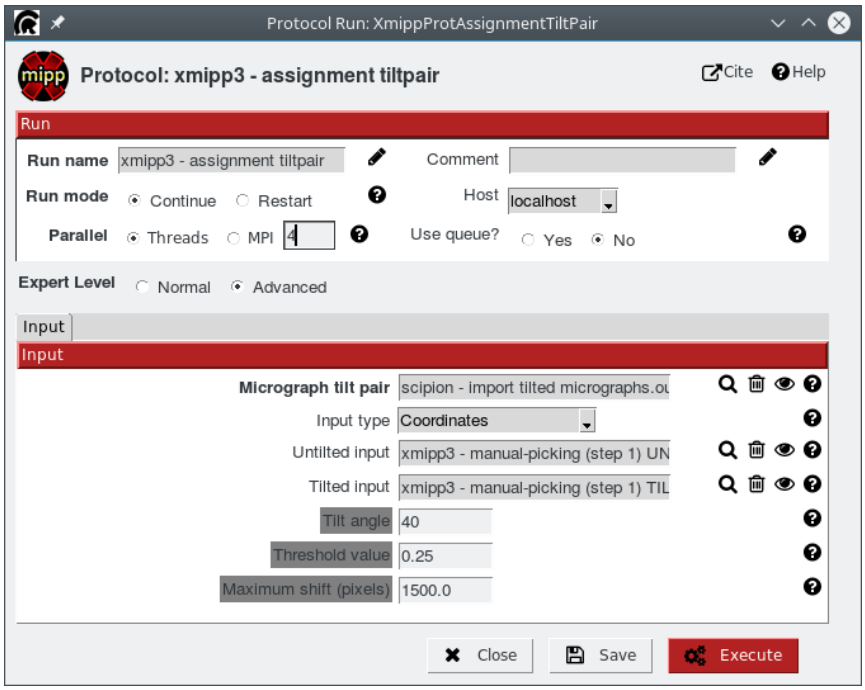


Figure 9: Xmipp3 - Assignment tiltpair protocol GUI.

**Xmipp3 - Assignment tiltpair** protocol makes use of six input parameters, three of them are mandatory and the other ones are optional and for advanced users:

**Required parameters:**

- *Micrograph tilt pair*: This field must contain a set of micrograph tilt pairs, untilted and tilted.
  The other two required parameters can be a set of coordinates or a set of particles, the user can choose between them.

- *Untilted input*: This field requires the untilted coordinates/particles which were previously picked.

- *Tilted input*: The same as before, but for the tilted micrographs.

**Optional parameters:**

- *Tilt angle*: If the approximate tilt angle is known, this value can be provided here. The protocol will search for a tilt angle in an interval $[estimation - 15^{o}, estimation + 15^{o}]$. If the tilt angle is unknown, then the value is $-1$ and the protocol search the full range of angles. By default, $tiltangle = -1$.

- *Threshold value*: The matching of a tilt pair will be required when the transformation between untilt and tilt points exhibits a distance equal or less than threshold*particle radius. Threshold is 0.25 by default.

- *Maximum Shift (pixels)*: Due to imprecise alignment of the tilt axis of the microscope stage, usually there is a shift between images of the tilted and untilted micrographs. Thereby, this option allow to control the maximum shift (in pixels) between both micrographs. 1500 pixels by default.

Finally, once you have run the protocol, if you click Analyze results, the particle pairs will be displayed on the micrographs, see figure 10. Afterwards, you can extract particle pairs as described in 2.4.

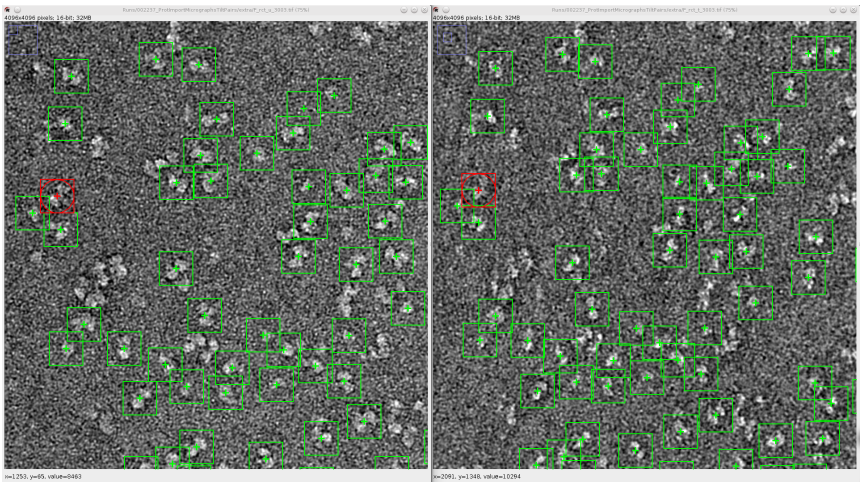

Figure 10: Assignment tiltpair protocol results. Particle tilt pairs are shown in the micrographs.

## 2.6 Alignment and classification with CL2D

Once the particles are extracted, the untilted ones have to be aligned and classified in order to distinguish different states of the particle or sort out bad ones. In this tutorial you will perform a CL2D classification using Xmipp3 package.

In general, 2D classification works better on "as good as possible" images. That means we might apply a band pass filter (see in SPA view, Particles » Filter » xmipp3 - filter particles) to allow the algorithm to concentrate on the important image features and not to get distracted by high-frequency noise or background gradient.

CL2D protocol can be found also in SPA view, in 2D » Classify group. Select the set of *untilted* particles produced in the previous step or the subset of untilted good ones if you have removed some particles before. Fill in the rest of the parameters as shown in figure 11 and press Execute.


Figure 11: CL2D protocol GUI.

This protocol will take some time to run (depending on your computer power and number of MPI processes chosen). Once it has finished you can visualize the obtained 2D classes by clicking on Analyze results button. A new window will open: see figure 12. Here it is necessary to select what kind of results are going to be analyzed. There are three groups of classes: classes, cores, and stables cores. In our case, we will use stable cores. These are composed by those particles that have stayed together throughout all iterations.

Each class should contain between 100 and 200 particles, more or less. Less than

100 will result in reconstructions of low quality, so the initial model estimation will be not done properly.

Some extra assessment can be done using the Principal Component Analysis (PCA) of the class images. First, right click on the desired class and select **Open images**, Images within the chosen class will show up. Now select in the viewer menu Tools > PCA - this will calculate 4 representative eigenimages for a set of your particles. If they are very noisy it means that there are not enough particles per class, so in principle you could use PCA tool for several classes.

Apart from that, there is also a subjective assessment, and it comes from what the user is looking for. Each class shows you a mean image that can give you an idea of what you will get if you do a reconstruction with those images. The user must select those classes that are more promising. In our case, using this criteria we can discard classes that seem to show particles in a top view (or just one part of the particle). Then, as you did with extracted particles, you should create a subset of classes to be used as input for next step, as shown on figure 13

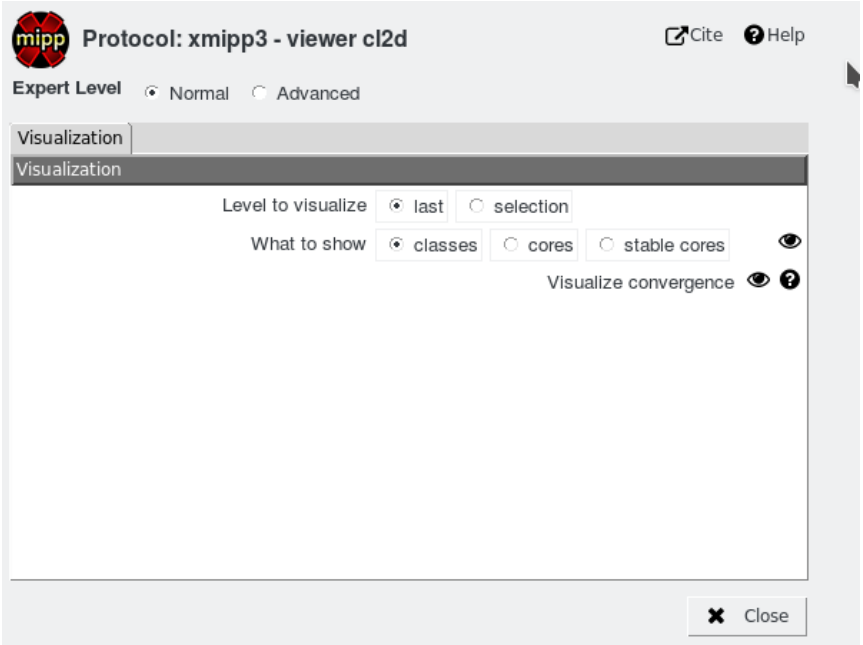


Figure 12: Results viewer of the CL2D protocol. A type of classes must be selected for display.

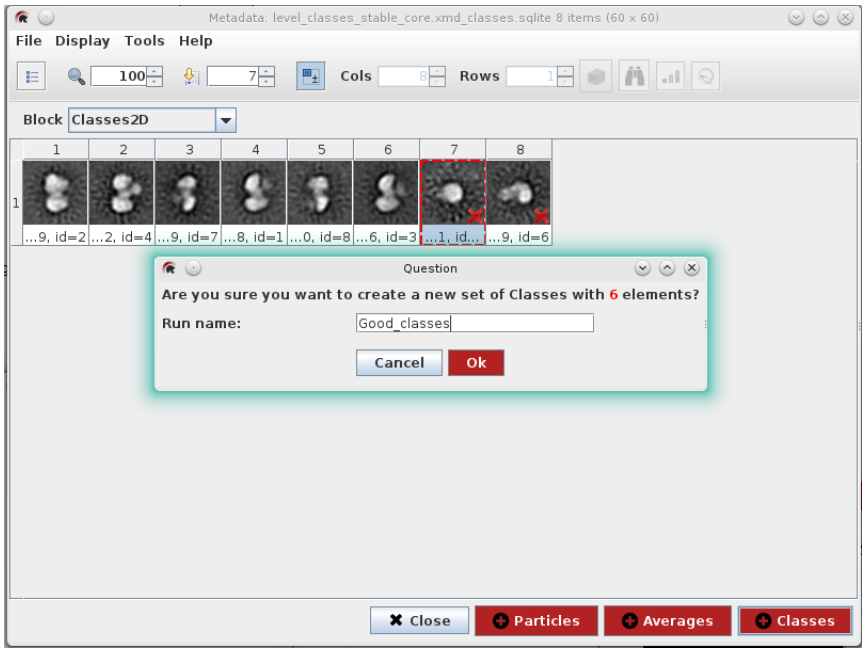


Figure 13: Two classes were discarded. By clicking on the red Classes button, a subset with good classes is created.

## 2.7 Initial volume generation with Random Conical Tilt

Once you have the desired classes, the only step left is RCT reconstruction. Select **Random Conical Tilt** in the View menu, then open the protocol named **Random Conical Tilt** and fill in the form as shown on figure 14. You have to provide the particle tilt pairs object produced in the extraction step (or a subset) and the particles or classes you want to reconstruct. Whichever particles you choose, they should contain alignment information.

In the protocol GUI, the **Alignment parameters** tab contains the **Thin Object** option, that must be set to "Yes" if the particle is not more or less spherical. When the object is going to change its width significantly between the untilted and the tilted image, we can considered this as a "thin object".

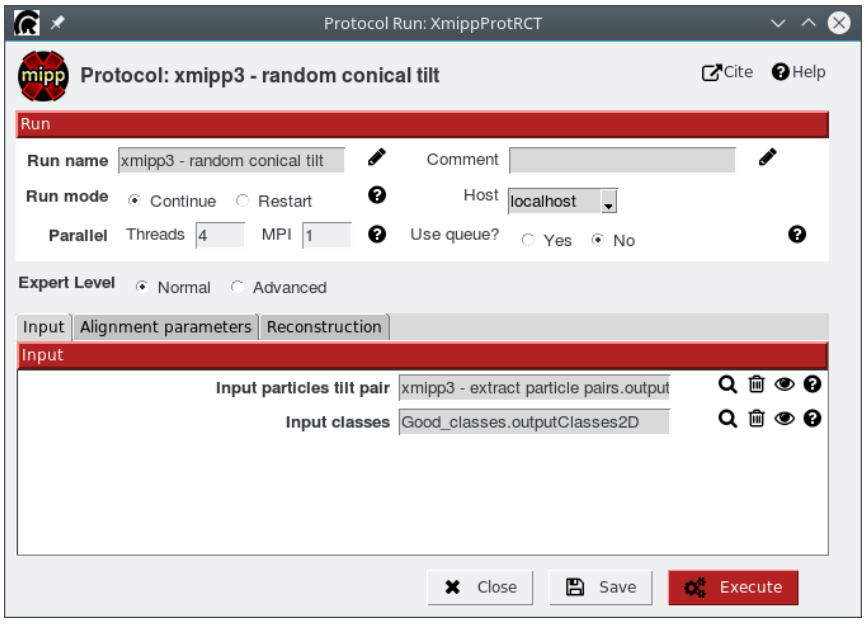


Figure 14: RCT protocol GUI.

RCT protocol can also low-pass filter the volumes after the reconstruction is done, producing both filtered and unfiltered volumes as output. In order to visualize the volumes click on Analyze results after the protocol has finished, and a window as the one in figure 15 will show up. The volume can be visualized in slices by double click on the image.

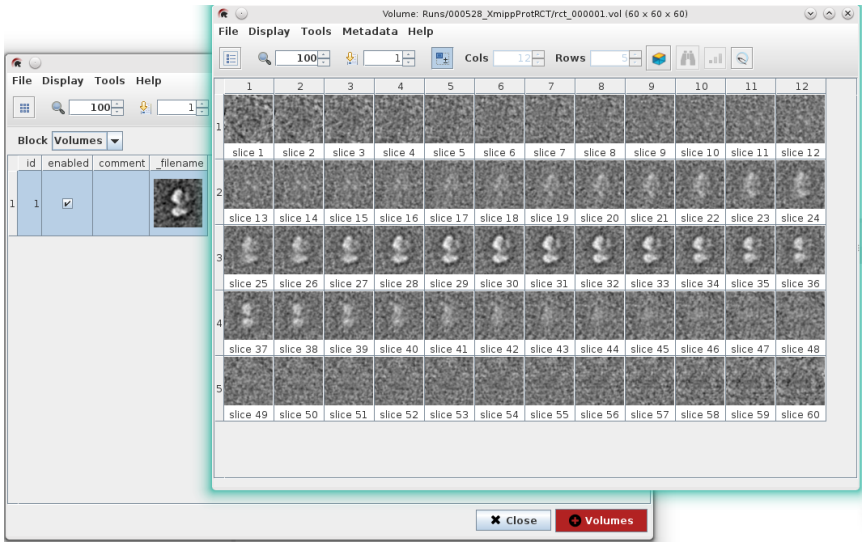


Figure 15: RCT protocol output: a 3D volume.

In addition, the volume can be opened with Chimera software, simply by clicking on the Chimera icon, when the volume is analyzed slice by slice, see figure 15

This is enough to have an initial 3D volume, but to have a better and more defined volume, we can also re-filter it by running the **xmipp - filter volumes** protocol found in 3D » Preprocess group of the protocols in **SPA** view. Protocol parameters are shown in figure 16

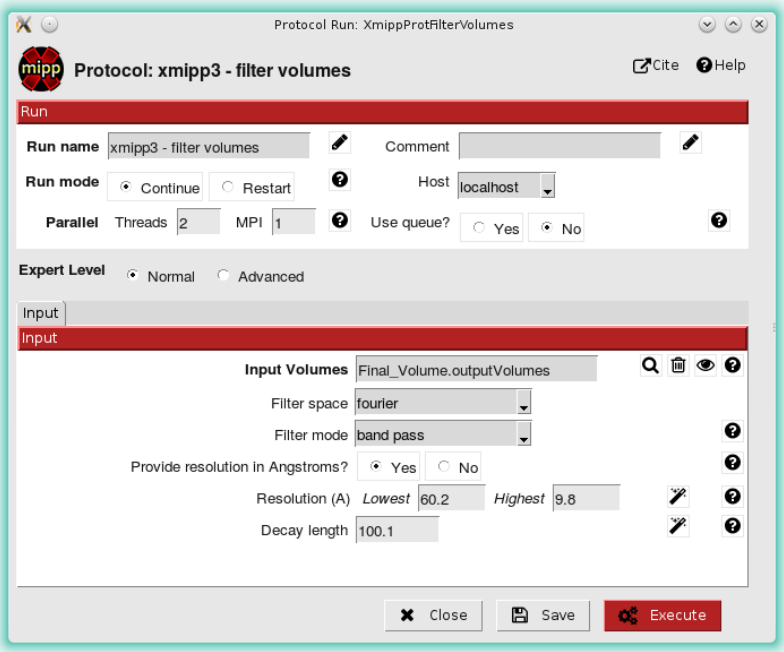


Figure 16: Filter volume protocol GUI.

# 3 Xmipp RANSAC

We will see now how to use the Xmipp RANSAC method, which can obtain a reliable low-resolution initial volume from a set of projection images without a priori information (Vargas et al., 2014).

## 3.1 Initial Volume Generation with RANSAC

Switch to the **SPA** view if you are not already there and select the 3D > Initial volume > xmipp3 - ransac protocol. As input we will use the classes obtained in the Random Conical Tilt section, so we can compare the volumes generated by different techniques. Fill in the form parameters as shown in figure 17 and explained below.


Figure 17: Xmipp RANSAC protocol GUI.

- There is no symmetry, so leave "c1" as the *symmetry group*.
- The *angular sampling* (in degrees) defines how fine the projection gallery of the volume is explored.
- The *number of RANSAC iterations* is the number of random volumes generated, in our case 100.
- We choose not to perform *dimensionality reduction* because there are not enough input classes. Instead we instruct the protocol to take 6 random images

and construct a volume with those. If the number of classes is large enough, you may choose to perform the dimensionality reduction and then the random images are selected from a low-dimensional space on which the input classes have been projected.

- The *inlier threshold* is a correlation-based cutoff to determine if a class is an inlier of the generated random volume or not. Note that in case where classes have very low noise, this value must be increased (to 0.8-0.95) and, on the other hand, in case with high noise, this value must be decreased.

- The *Number of best volumes to refine* is the number of volumes, among all the generated ones (100 in this case), that are going to be further refined using projection matching approach for the selected *Number of iterations to refine the volumes*.

- Finally, *Max frequency of the initial volume* is the maximum frequency of the initial volume in Angstroms.

Press the Execute button and wait until the protocol has finished. Then the output volumes can be visualized by clicking on Analyze Results as seen in figure 18.

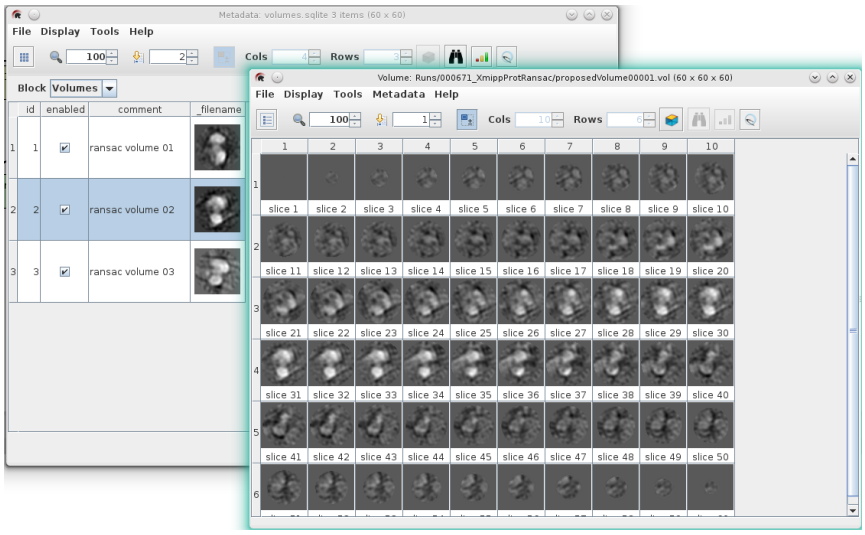


Figure 18: Xmipp RANSAC output volumes.

# 4 EMAN2 Initial Model

In this section we will show how to obtain an initial model in Scipion using EMAN2 initial model protocol.

## 4.1 Initial Volume Generation with EMAN

Change to the **SPA** view if you are not already there and select the 3D > Initial volume > eman2 - initial model protocol. As we did before, use the classes obtained in the Random Conical Tilt section as input, so we can compare the generated volumes. Also, fill in the other form parameters as shown in figure 19.

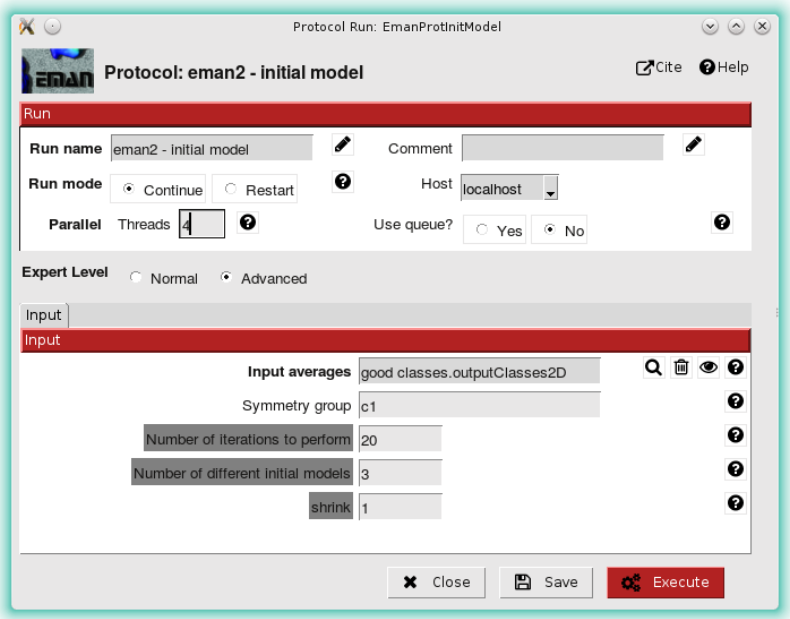


Figure 19: EMAN Initial Model protocol GUI.

Once the protocol has finished, display the generated output volumes by clicking on Analyze results button. Double-click on any volume to get a window as the one in figure 20.

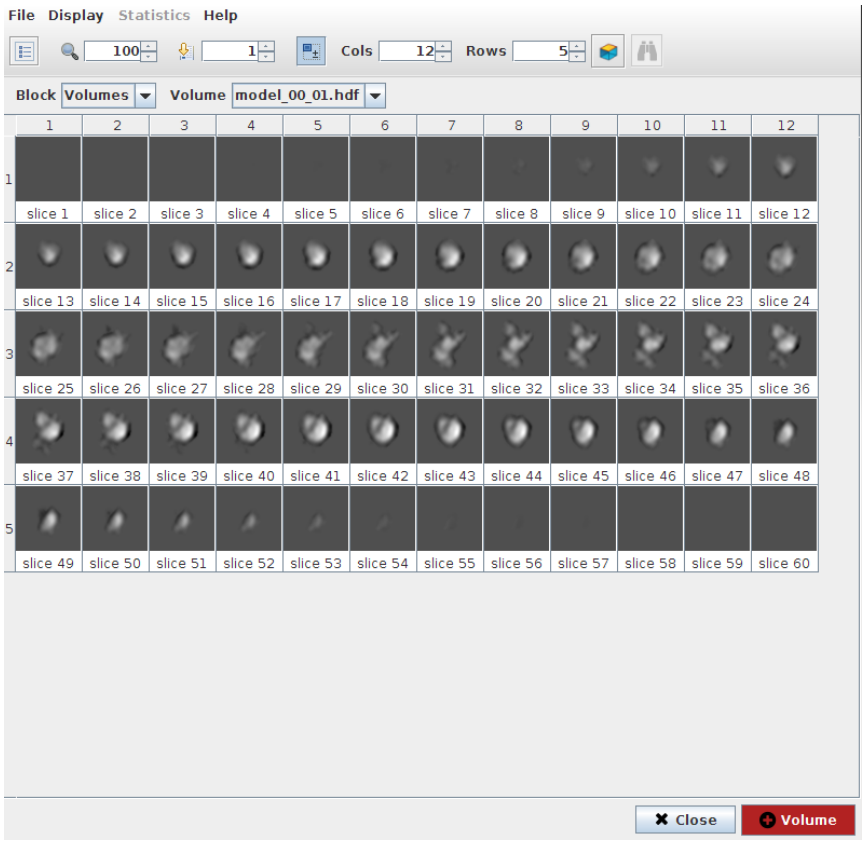


Figure 20: EMAN has generated output volumes.

# 5 Xmipp Reconstruct Significant

In this section we will show how to obtain an initial model in Scipion using Xmipp Reconstruct Significant protocol. This algorithm sets the initial volume problem in a *weighted least squares* framework. The weights are calculated by a statistical approach based on the cumulative density function of image similarities. The algorithm is not particularly fast since slow convergence is promoted as a way to avoid local minima. However, the reconstructed volume normally corresponds to a correct estimate of the underlying structure.

## 5.1 Initial Volume Generation with Reconstruct Significant

In the **SPA** view select the 3D > Initial volume > xmipp3 - reconstruct significant protocol. Fill in the form parameters as shown in figure 21 and 22.

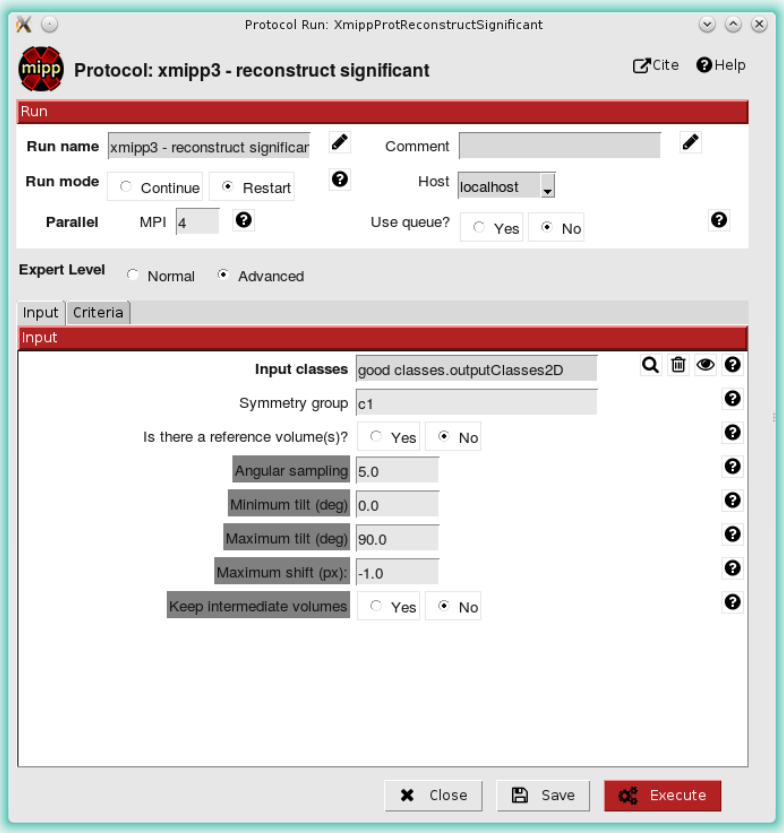


Figure 21: Xmipp Reconstruct Significant protocol GUI. Input parameters


Figure 22: Xmipp Reconstruct Significant protocol GUI. Criteria parameters.

Once the protocol has finished you can go and view the generated output volume by clicking on Analyze results, as you can see in figure 23.

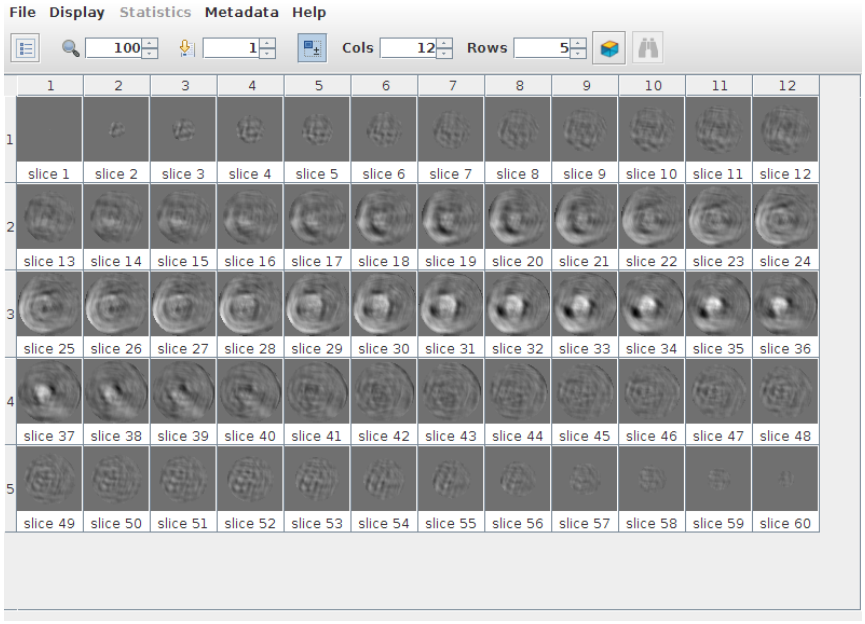


Figure 23: Xmipp Reconstruct Significant results.

# References

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.

Vargas, J., Álvarez-Cabrera, A.-L., Marabini, R., Carazo, J. M., and Sorzano, C. O. S. (2014). Efficient initial volume determination from electron microscopy images of single particles. *Bioinformatics*, 30(20):2891–2898.