EMBO Practical Course on
Image Processing for Cryo-EM

5 - 15 September 2017

Practical 2: Using SCIPION for movies alignment, CTF estimation, particles picking and extraction.

Contents

Getting Started **1**

Day 1: Movies alignment, CTF estimation and Particles picking **2**

0.1 Creating a New Project and Importing a Workflow 2
0.2 Importing Movies 4
0.3 Movies Alignment 6
0.4 Estimating CTF 7
0.5 Analyzing Results and CTF recalculation 8
0.6 Particle Picking 10
0.7 Extracting Particles 12

Getting Started

Software Requirements

To follow this tutorial you will need to have SCIPION (de la Rosa-Trevín et al., 2016) properly installed on your system. This should be the case for this course. If you are following this practical on your own and SCIPION is not installed, the following page will provide detailed information about the installation:

`https://github.com/I2PC/scipion/wiki/How-to-Install`

We are going to use some Xmipp3 (de la Rosa-Trevín et al., 2013) programs, but they come installed with Scipion. Aditionally, we will need to install other EM software packages such as: motioncor2 (Zheng et al., 2017), CTFFind4 (Rohou and Grigorieff, 2015), Relion2 (Kimanius et al., 2016) and Eman2 (Tang et al., 2007).

Practical Summary

The main goal of this practical is to illustrate the combination of different EM software packages in the same project. Here is a brief summary of the steps that will be done:

Test Data

If you are following this tutorial in the EMBO 2017 course, the data folder should be located in: `/PATH/TO/DATA/FOLDER`

If not, you can still download the test data from: `download.data.uk` It should be a single tar file of about 6Gb. You need to uncompress it and you will have the data folder.

Through this tutorial we will refer to the data folder as `$DATA/`

Day 1: Movies alignment, CTF estimation and Particles picking

0.1 Creating a New Project and Importing a Workflow

We need to start by creating a new SCIPION project before any processing can be done. For that, you should open SCIPION from a terminal:

```
> scipion
```

After that, the window with all projects will be shown (it can have an empty list if there is not previous project for this user). You can then click on Create Project button and enter the project name, for example: **embo2017_day** (as shown in Fig. 1). After clicking on Create, a new window will appear showing your new empty project. Don't worry, in the next step we will import a workflow template that will guide you through the processing pipeline of this practical.

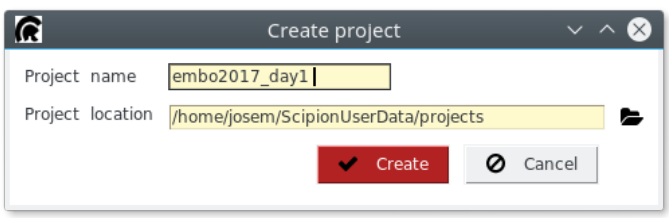

Figure 1: Create Project dialog.

TIP

SCIPION projects are located by default `$HOME/ScipionUserData/projects`, but this can be changed in a configuration file. Moreover, each project can be located in a different location and a link is created in `$HOME/ScipionUserData/projects`. Projects are self-contained folders that can be moved from one computer to another.

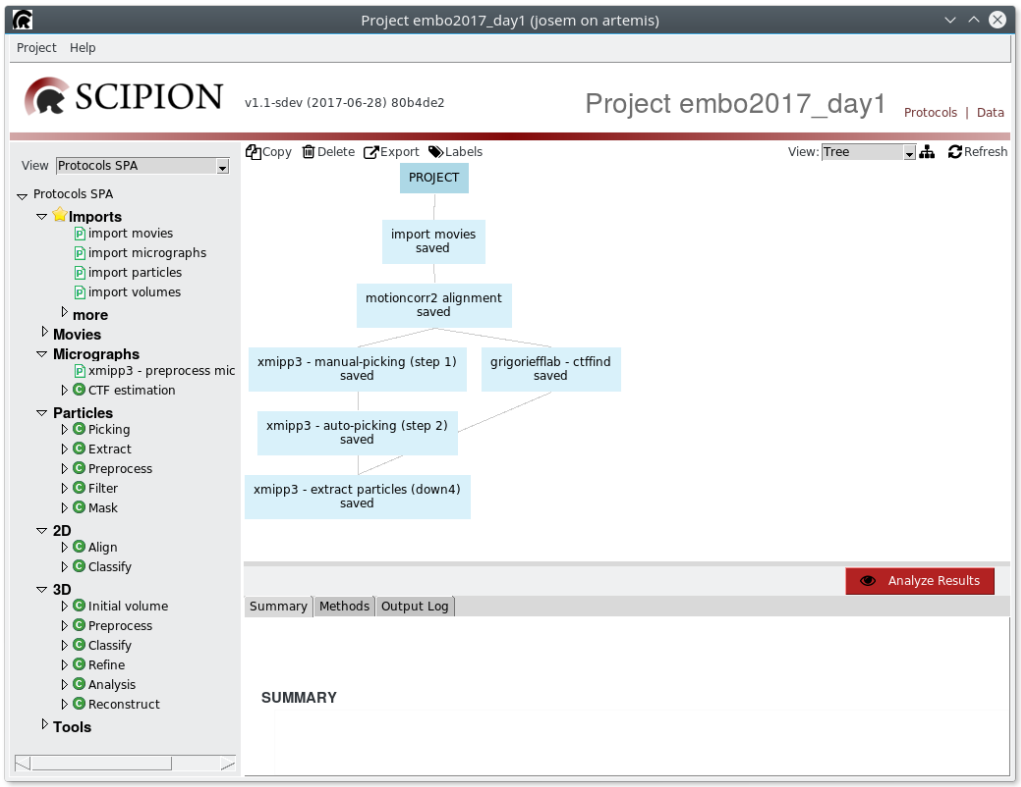

Figure 2: Project main GUI. Left panel displays a tree with processing tasks (protocols) that can be used. The top right panel displays the sequence of protocols executed (runs) by user and their state: saved, running, finished or aborted. Bottom right panel displays information for the selected run, such as inputs and outputs, execution logs or documentation.

From the project GUI we can import a workflow template that will load all the steps and parameters that will guide us through this practical. For doing that, you must go to the main menu: Project > Import workflow and use the browser dialog to search for the workflow file. In our case, it should be `$DATA/workflow1.json`. After selecting the file, click on the Import button and your project should have the workflow loaded, as shown in Figure 2.

TIP

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.

0.2 Importing Movies

In SCIPION the *import* is almost the only place where the user needs to deal directly with files. In our model each protocol has very well defined inputs and outputs, which are data objects. These objects (SetOfMovies, SetOfParticles, Volume, CTFModel, etc.) encapsulate the underlying files and formats. When importing data like SetOfMovies, SetOfMicrographs or SetOfParticles, the user provides critical information (such as the Pixel Size). This information will not be requested any more and should be properly propagated.

To import movie files, double-click the import movies box in the workflow. The protocol form with parameters will open as shown in Figure 3. In this case, the proper values for all the acquisition parameters have been loaded from the template. We only need to provide the path to the directory where the movies are. You can either use the Browse icon to select the path or type it directly in the entry field. For this tutorial it should be `$DATA/Movies`.

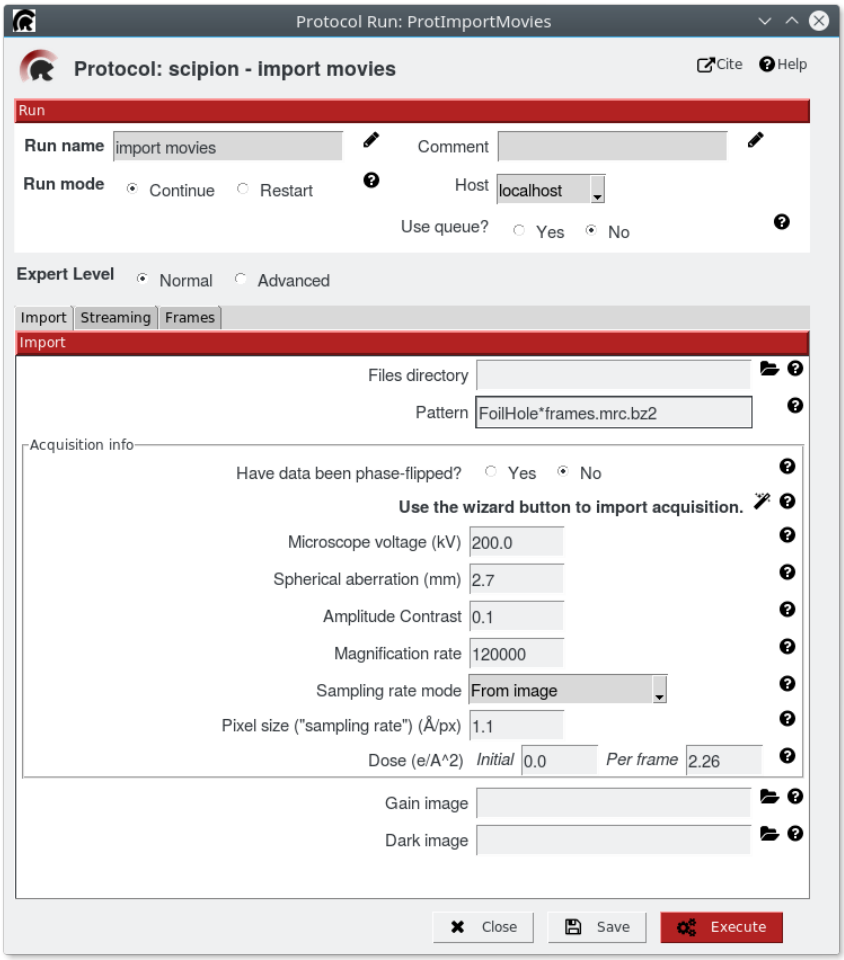

Figure 3: Import Movies protocol. In this case, only the path needs to be filled.

After selecting the path, we can press the Execute button and the box for this run should become yellow (running state) and then green (finished). Then, the summary tab should display some information such as the number of movies imported and the number of frames.

TIP

Since important information is provided during the *import* step, it is recommended to take your time to check that all input parameters are correct. When importing, the binary files are not copied to the project to avoid data duplication. Instead, soft links are created pointing to the files location. If you move your project to another computer these links may be broken. There is an advanced option where you can set *Copy files?* to Yes. In this case, the project is more self-contained at the price of using more disk space.

0.3 Movies Alignment

Aligning the individual frames of movies is necessary to correct BIM-induced image blurring and restore important high resolution information. In this practical we will use `motioncor2` (Zheng et al., 2017) for movie alignment. For that, we just need to open the motioncor2 alignment box and executed it, since all the parameters have been preloaded.

In general, one of the most important parameters at this step is to choose the frame range to use for aligning the frames and producing the final average micrograph. Additionally, the `motioncor2` protocol allows to apply a dose weight to each frame taking into account the accumulated dose. Take you time in the protocol form and click on the Help button of other parameters to know their usage.

After launching the movie alignment protocol, we can go further and start the CTF estimation. Due to the stream processing capability in Scipion, we don't need to wait until the first protocol finish to start the second one in this case.

ALTERNATIVES

- xmipp - Optical Flow alignment
- grigorieff lab - Unblur + Summovie

0.4 Estimating CTF

The next step is to estimate the CTFs (Contrast Transfer Functions) of the micrographs using *CTFFind4* (Rohou and Grigorieff, 2015). This protocol estimate the PSD (Power Spectral Density) of the micrographs and the parameters of the CTF (defocus U, defocus V, defocus angle, etc.).

To estimate the CTF you will need to select the frequency region to be analyzed (Figure 4). The limiting frequencies must be such that all zeros of the CTF are contained within those frequencies. There is a wizard, shown in Figure 5, that helps in choosing those frequencies. To see all available options, choose the *Advanced* expert level and click on the Help button for any specific parameter.

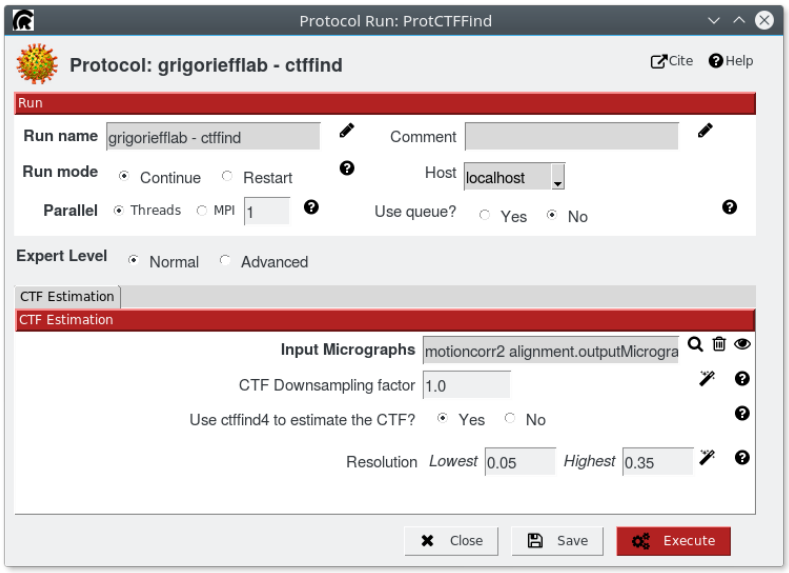

Figure 4: CTFFind protocol.

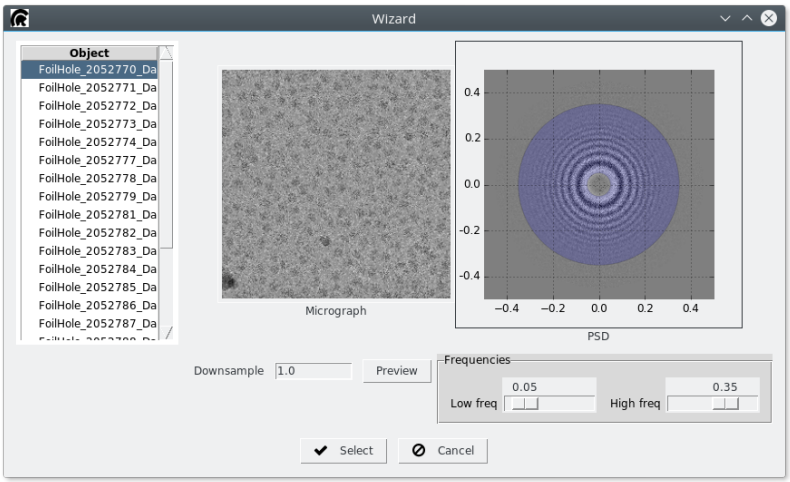

Figure 5: Wizard to help selecting the frequency region.

ALTERNATIVES

- gCTF
- xmipp - CTF estimation

0.5 Analyzing Results and CTF recalculation

The CTFs of good micrographs typically have multiple concentric rings, shown in Figure 6 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. A reason to discard micrographs may be the presence of strongly asymmetric rings (astigmatism, Figure 6 center) or rings that fade in a particular direction (drift, Figure 6 right).

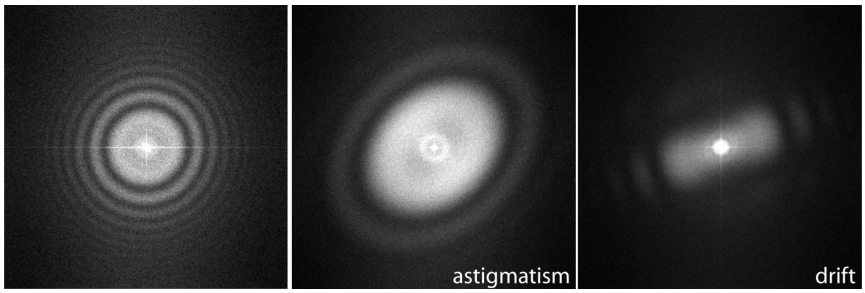

Figure 6: CTFs of good, astigmatic and drift micrographs, respectively.

The output from Ctffind4 (or any other CTF estimation protocol) is shown when clicking on the Analyze Results button (Figure 7). To discard micrographs with bad CTFs you may click with the mouse right button and choose **Disable**. Once you finish the selection, press on the red Micrographs button to create a new subset of micrographs with only the enabled ones.

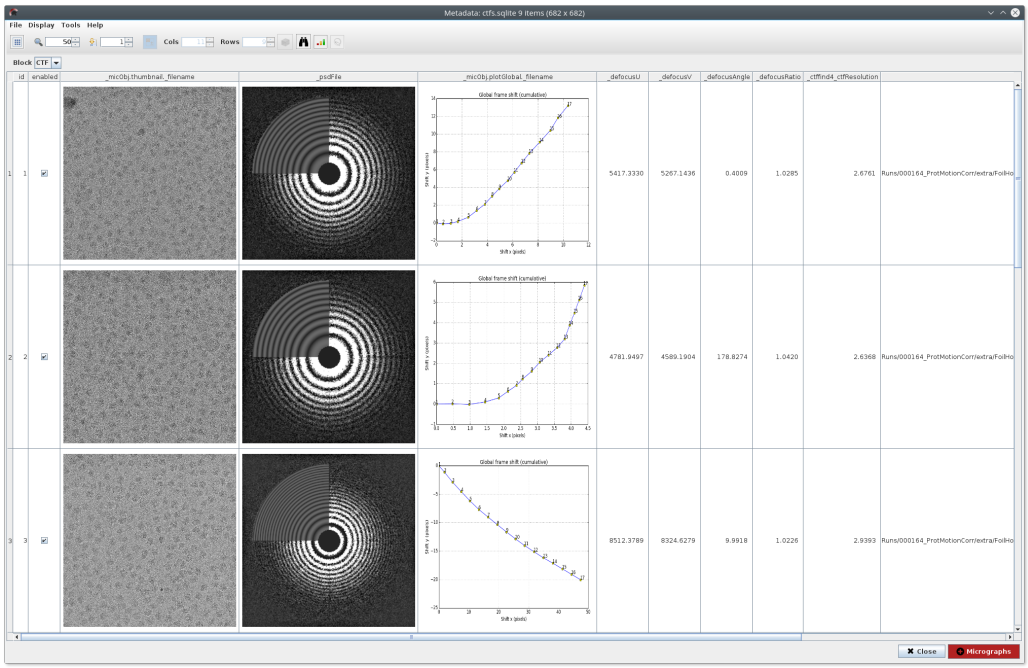

Figure 7: Results visualization after CTF estimation.

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

0.6 Particle Picking

Picking is an important step to select your particles from the micrograph images. Manual picking can be very tedious and many picking tools have been developed that can be more or less convenient depending on the sample and personal preferences. In SCIPION there are integrated several picking tools, allowing users to select the one which better fits their needs. It this tutorial we are going to use the Xmipp picking.

Xmipp particle picking is divided in two steps: (1) manual/supervised picking and (2) completely automatic picking. For the manual/supervised picking, we open the xmipp3 - manual picking box, select the input micrographs and execute it. This box will become light yellow, meaning that this is an interactive job that we can relaunched at any time.

The Xmipp picking GUI (Figure 8) contains a control panel with the list of micrographs and some other parameters. The micrograph that we are picking is displayed in a separate window and we can apply to it a number of filters/enhancements (like Gaussian blurring, Invert contrast, adjust histogram etc.) just to improve the visualization of particles. Go to to the menu Help > Tips to check the picking controls.

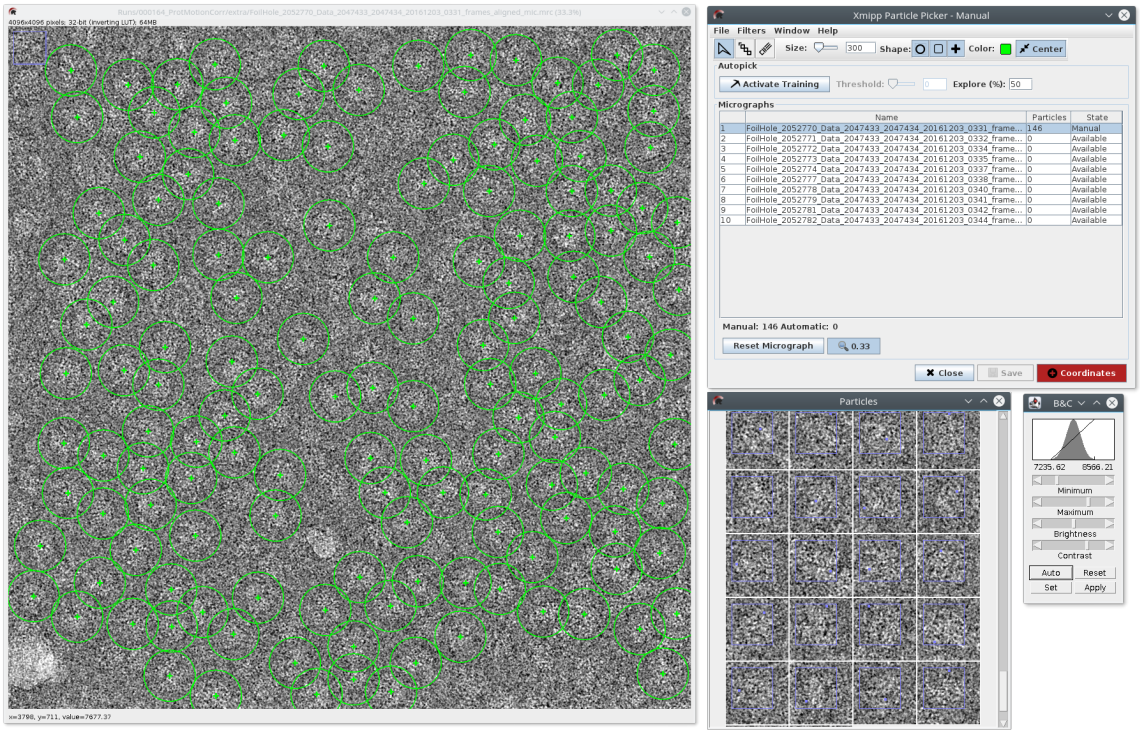

Figure 8: Xmipp picking interface.

In the manual/supervised step, we should start picking manually a few micrographs and then click the Activate Training button. At this point, the program will train a classifier based on machine learning and will propose some coordinates automatically. You can "correct" the classifier by adding missing particles or removing wrongly picked ones. After training with a few more micrographs, we can register the output coordinates by clicking on the Coordinates red button.

Then we can close the GUI and open the xmipp3 - automatic box and select both the previous execution of manual/supervised and all micrographs as input (choose *Micrographs to pick: other*). When executing, this will pick the rest of micrographs automaticly. At the end, we can review the picking coordinates and we still have a chance to add/remove particles.

TIP

Once we have trained the classifier during the manual/supervised step, you can pick another set of micrographs of newly collected data in an automated fashion.

ALTERNATIVES

- EMAN2 - boxer (supervised)
- SPARX - gaussian
- Appion - dogpicker
- Relion - autopicking
- gEMpicker
- Gautomatch

0.7 Extracting Particles

Once we have a set of coordinates, we can proceed to particle extraction with Xmipp. We need to be careful with the options selected here, since this will affect later steps. The extract protocol (Figure 9) will allow us to extract, normalize and correct the CTF phases of your picked particles, among other things. The options are summarized below:

- The *coordinates* of the particles in the micrographs are taken from the results of the previous step. Also, in the same tab, select the *particle box size* in pixels (in this case 64 px).
- The *invert contrast* flag. If activated, bright regions become dark regions and the other way around.
- The *phase flipping* flag. If activated, the protocol corrects the CTF phases of your particles.
- The *normalize* flag. If activated (recommended), the particles are normalized to have zero mean and a standard deviation of one for the background pixels.

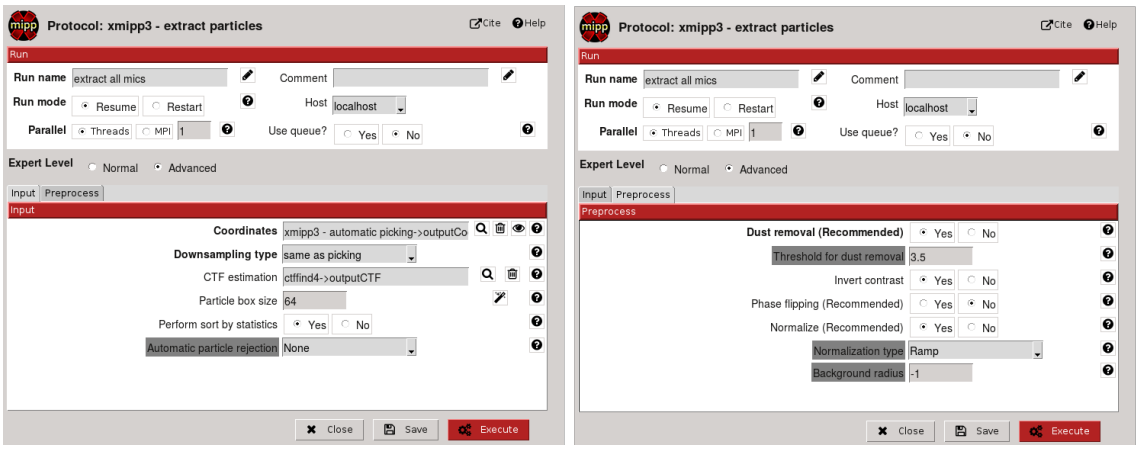

Figure 9: Extract particles protocol. Available options are shown.

In this case, we have choose to invert the contrast, since we will SPIDER and Xmipp later for 2D classification and it expects particles to be white over black. We have also chosen to do phase flipping, because these packages expect the input particles in this way. When using other packages (e.g, Relion, Frealign) it is not recommended to do the phase flipping, since they prefer to handle that correction internally.

We have also checked the *Sort by statistics* option, which tells the protocol to sort 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 usually outliers. Press the Analyze Results button in the main window to check the extracted and normalized images (Figure 10 and 11). If the *Sort by statistics* was checked, the particles will be sorted (ascending) by the z-score value and the z-score distribution over the particle set will also be plotted. If you want to remove some particles, e.g. because they are outliers, you can use right click > disable. To create a new set of particles you can click on Particles red button.

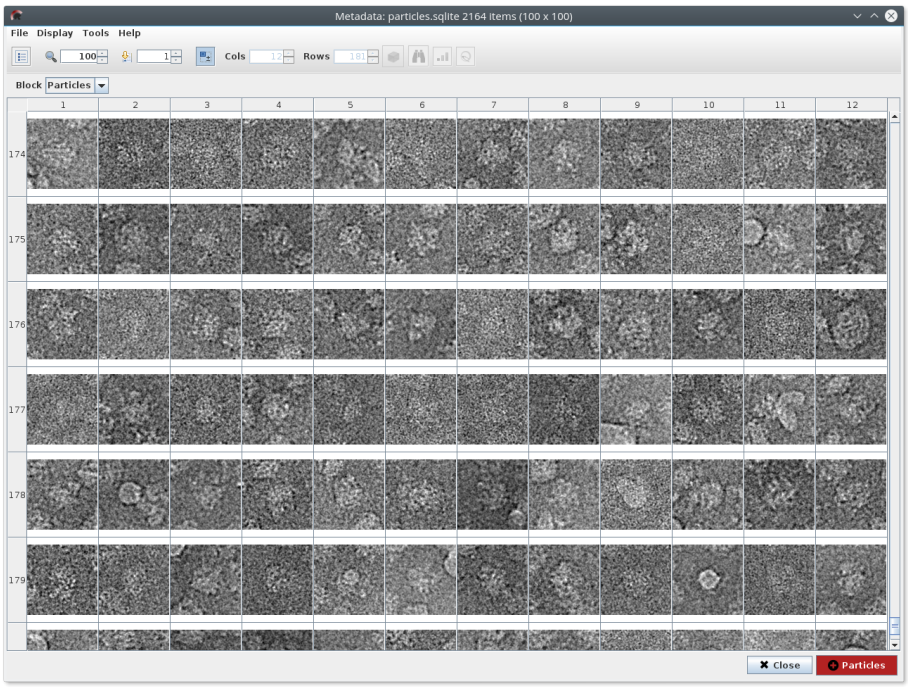

Figure 10: Particles sorted by z-score, highest values means worse particles.

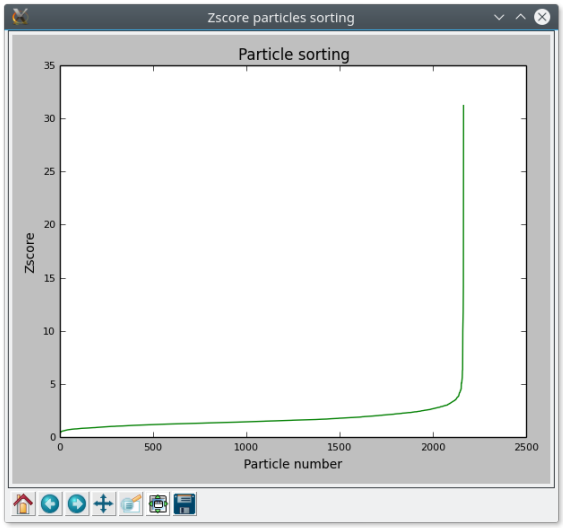

Figure 11: Plot of the z-score values for the particle set.

TIP

The GUI to visualize particles can be used with other protocols to create subsets based on any quality parameter. Particles can be easily ordered by any column and then selected based on that order to discard bad particles.

ALTERNATIVES

- Relion2 - particles extraction

References

de la Rosa-Trevín, J., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J., and Sorzano, C. (2013). Xmipp 3.0: An improved software suite for image processing in electron microscopy. *J. Struc. Biol.*, 184(2):321 – 328.

de la Rosa-Trevín, J., Quintana, A., del Cano, L., Zaldívar, A., Foche, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., Vargas, J., Otón, J., Sharov, G., Vilas, J., Navas, J., Conesa, P., Kazemi, M., Marabini, R., Sorzano, C., and Carazo, J. (2016). Scipion: A software framework toward integration, reproducibility and validation in 3d electron microscopy. *J. Struc. Biol.*, 195(1):93 – 99.

Kimanius, D., Forsberg, B., Scheres, S. H. W., and Lindahl, E. (2016). Accelerated cryo-em structure determination with parallelisation using gpus in RELION-2. *eLife*, 5:e18722.

Rohou, A. and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struc. Biol.*, 192(2):216–221.

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J. Struc. Biol.*, 157(1):38–46.

Zheng, S., Palovcak, E., Armache, J.-P., Verba, K., Cheng, Y., and Agard, D. (2017). Motioncor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature methods*, 14(4):331–332.