

EMAN2.2 step-by-step tutorial

21~23 March 2018 - cryo-EM Workshop

ISGC 2018 - Academia Sinica, Taipei, Taiwan

Wei-Cheng Huang, Shih-Chi Luo, Piotr Draczknowski*¹, Shang-Te Danny Hsu*²

*¹e-mail: piotrd@gate.sinica.edu.tw *²e-mail: sthsu@gate.sinica.edu.tw

The logic of the tutorial and color coding of the information:

- The text you see in *italics* will generally refer to:
 - labels in the GUI, such as buttons to press
 - or commands to write in terminal
- **Items coloring in red refer to the most common pitfalls or mistakes** which may lead to problems with the data processing pipeline.
- **Crucial parameters that have to be changed/specified or selected** in EMAN GUI in order to successfully complete the tutorial run. **Other parameters can be left as default!**
- **Output file/files of the job**
- Additional information or explanations of some optional parameters which do not have to be changed during the tutorial.

Main features and functionalities of EMAN GUI

- EMAN2 will work best with a **3-button scroll mouse**, though there are alternatives using keyboard modifiers for one button mice or trackpads on Macs.
 - In most EMAN2 windows (2D images, 3D volumes, plots, etc.), the middle mouse button will open a control panel for the widget, which is different for each widget type
 - The right mouse button is used for panning in 2D or 3D image windows, and can be used to zoom (by shift+dragging), and to reset the zoom (clicking) in plot windows.
 - The scroll-wheel will generally act as a zoom. control-panel for more precise control
 - If you have a one-button mouse, holding down the alt/option key combined with a mouse click will serve the same role as a middle-click.

References:

- G. Tang, L. Peng, P.R. Baldwin, D.S. Mann, W. Jiang, I. Rees & S.J. Ludtke. (2007) EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol. 157, 38-46. PMID: 16859925
- J.M. Bell, M. Chen, P.R. Baldwin & S.J. Ludtke. (2016) High Resolution Single Particle Refinement in EMAN2.1. Methods. 100, 25-34. PMC4848122

Tutorial DATA set (full dataset and reference):

EMPIAR 10025 20S proteasome

<https://www.ebi.ac.uk/pdbe/emdb/empiar/entry/10025/>

EMDataBank Validation Challenges



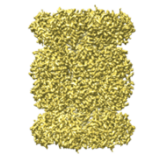
<http://challenges.emdatabank.org/?q=map-challenge-targets>

Online EMAN2.2 tutorial at EMAN Wiki and discussion:

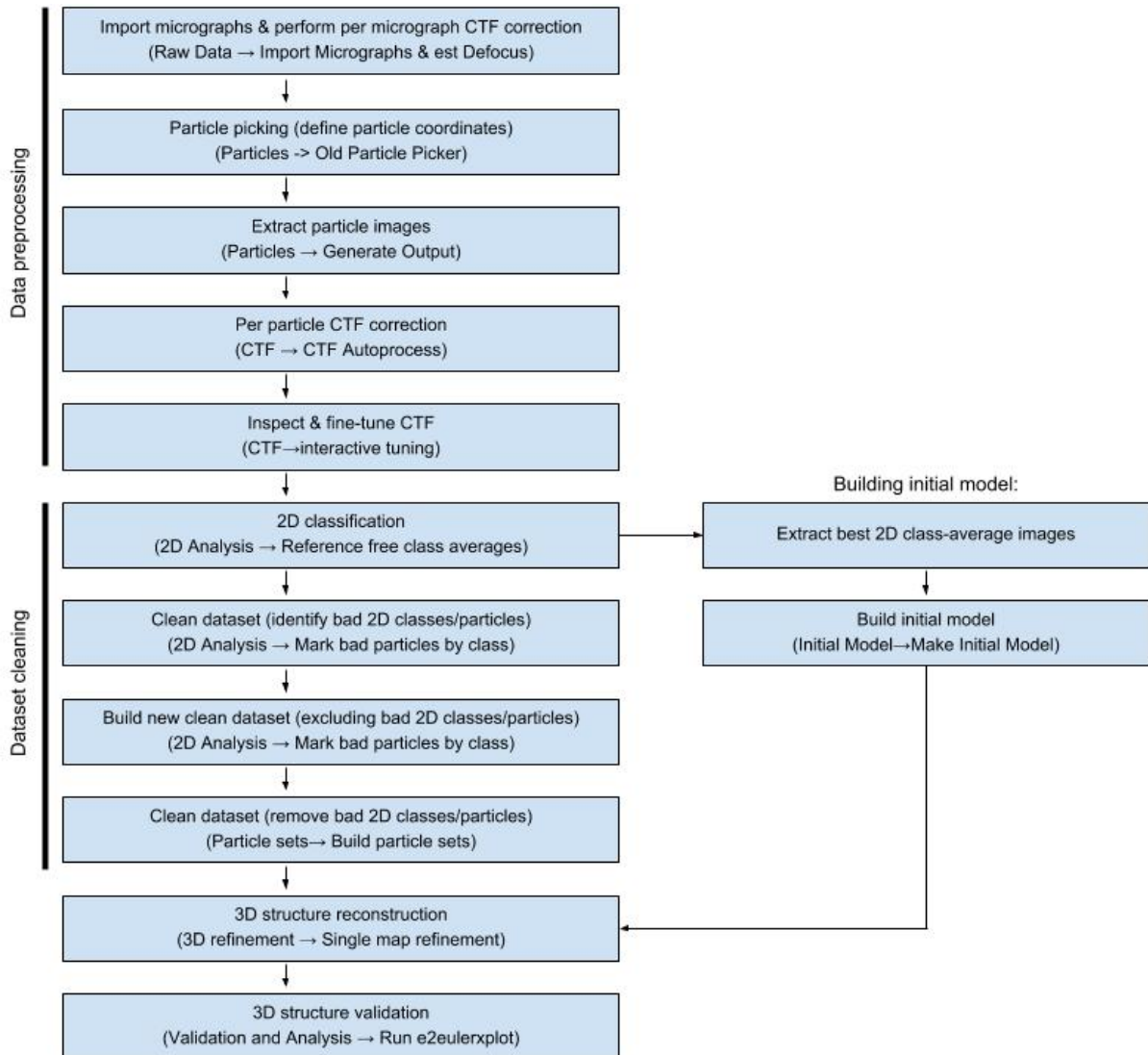
<http://blake.bcm.edu/emanwiki/EMAN2/Tutorials>

<https://groups.google.com/forum/#!forum/eman2>

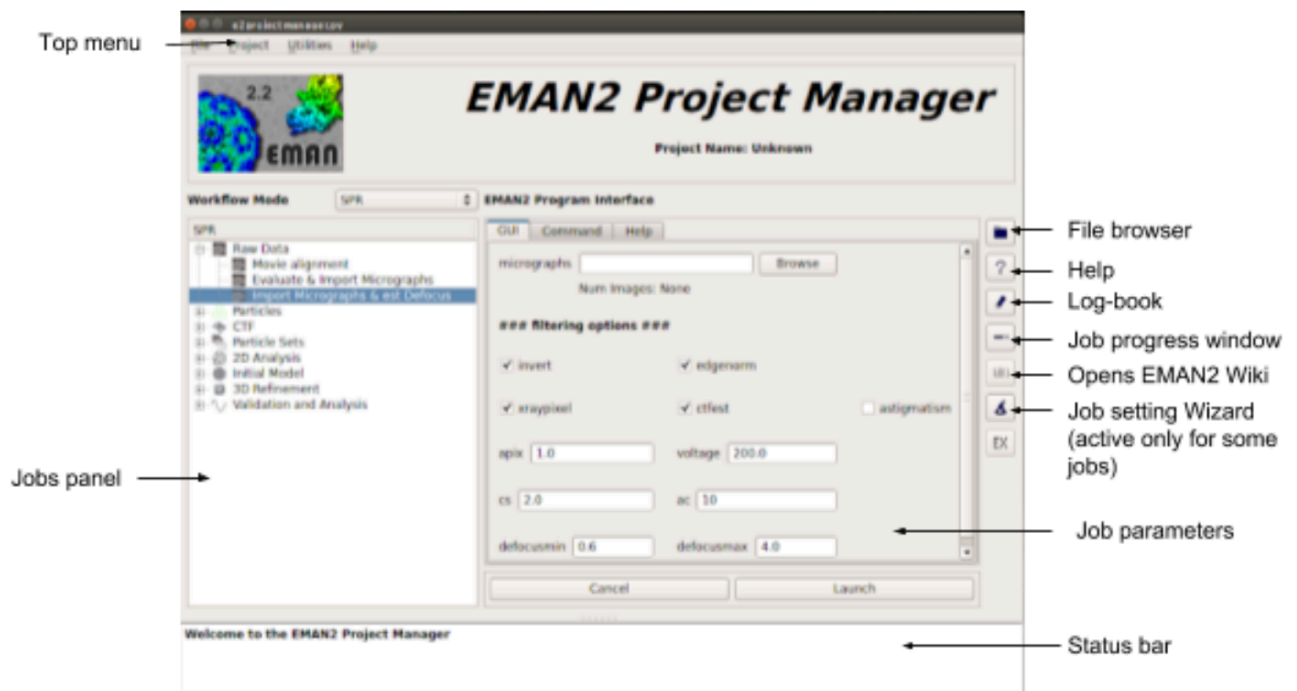
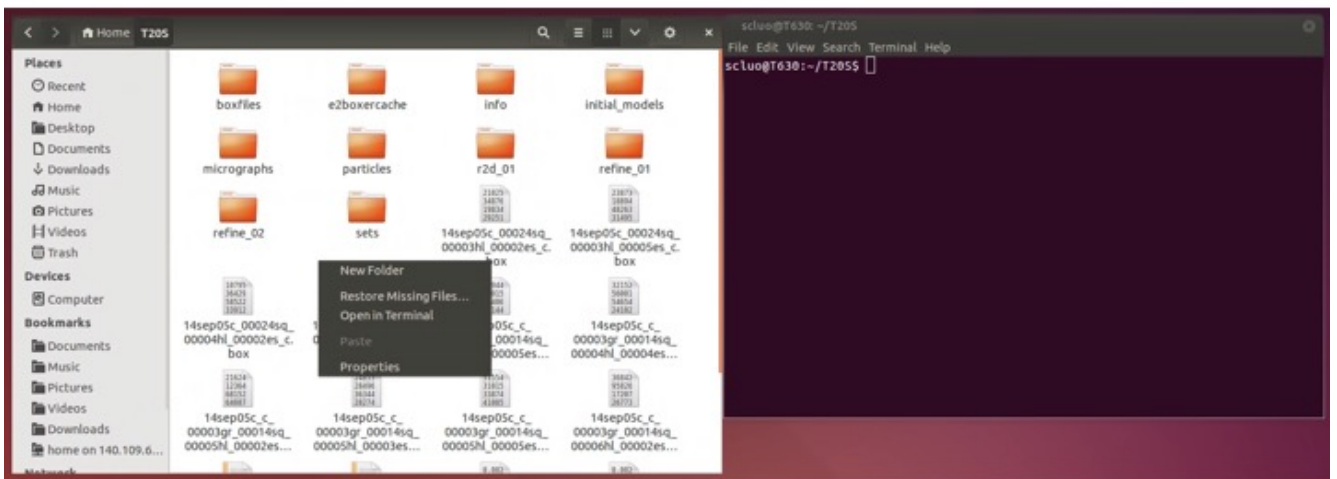
EMPIAR-10025

T20S Proteasome at 2.8 Å Resolution		Contains:
Publication:	2.8 Angstrom resolution reconstruction of the Thermoplasma acidophilum 20S proteasome using cryo-electron microscopy. Campbell MG, Veesler D, Cheng A, Potter CS, Carragher B 	 micrographs
Elife	4 e06380-e06380 (2015)	
PMID:	25760083	
DOI:	10.7554/eLife.06380	
EMD-6287		
Related EMDB entry:		
Deposited:	2015-02-26	
Released:	2015-03-06	
Last modified:	2015-03-06	
Dataset size:	2.0 TB	
Dataset DOI:	10.6019/EMPIAR-10025	
Version history:	1 2015-07-01 Added Relion star file with coordinates: data/14sep05c_raw_196 /run1_shiny_mp007_data_dotstar.txt	

Workflow overview:



1. Right click on desktop & open terminal
2. Go to a clean project directory and load your raw data
Type: `cd T20S_proteasome_EMAN`
This folder from now on will be called “project directory”
3. Type: `e2projectmanager.py`
This will open EMAN2.2 GUI window.



4. From the top menu of the EMAN GUI window select:

Project → **Edit Project**

and set overall properties of the project:

Project name = **T20S** (can be anything whatever you want)

Particle Mass = 700 (molecular mass of the imaged particles in kDa)

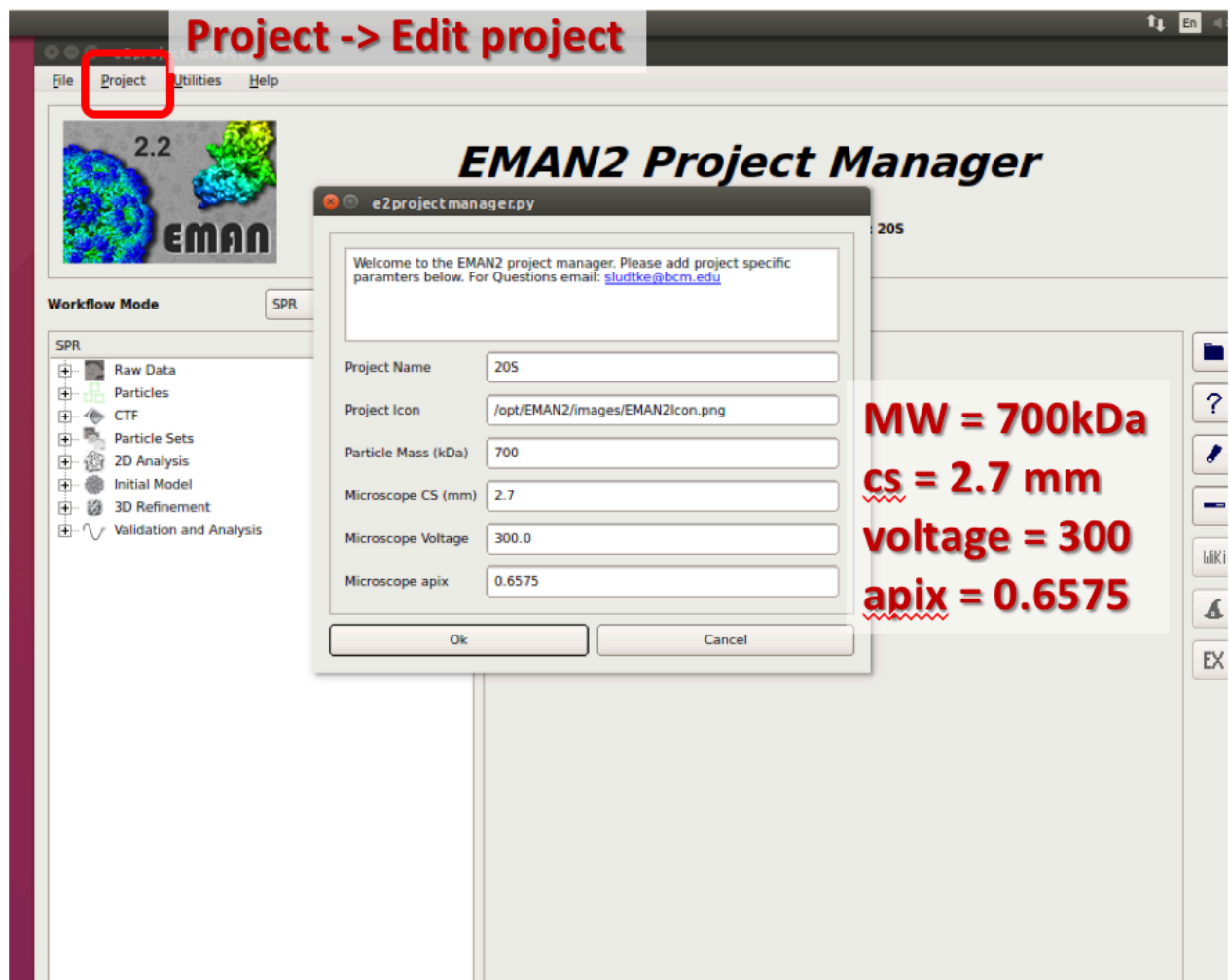
Microscope Cs = 2.7 (depends on your microscope)

Microscope Voltage = 300 (depends on your microscope stings)

Microscope apix = 0.6575 (depends on magnification used, specifies Å/pixel ratio)

Confirm by **OK**

Output file: info/project.json



Due to time limitation of this hands-on tutorial section, we only use **10 micrographs** in this demonstration instead of **196 micrographs** dataset in the publication.

From the left-side panel (Jobs panel) of the EMAN GUI select:

5. **Raw Data** → **Import Micrographs & est Defocus.**

a. Press the *Browse* button, then highlight all of the *.mrc files in that “./Micrographs” folder. Confirm the selection by *OK*.

b. In the main GUI window select checkboxes: **edgenorm**, **xraypixel**, **invert** and **ctffest**

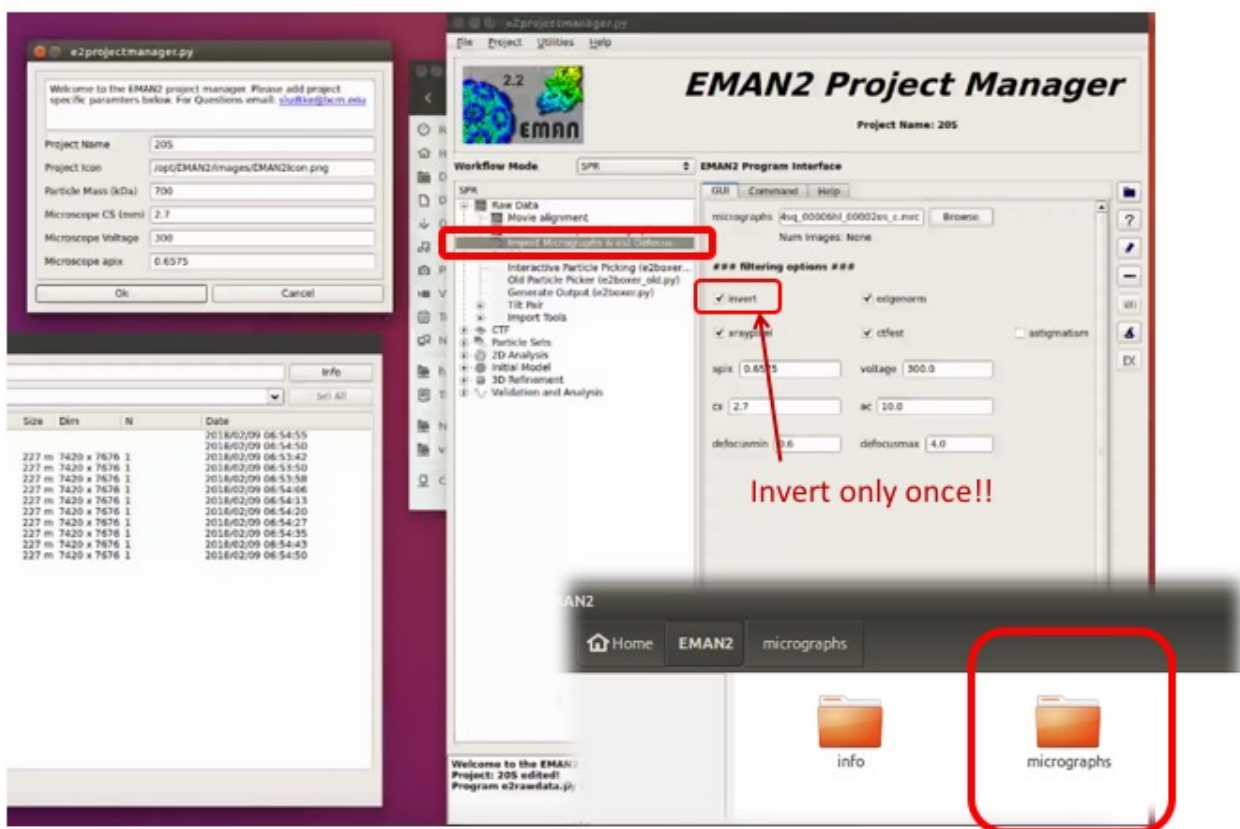
The **astigmatism** parameter will try to cure CTF deformations (elongated, non-circular, oval Thon rings instead of nice, even and circular CTF pattern) caused by imperfections of your microscope alignment. Not necessary in case of the tutorial dataset.

c. Press **Launch**

This will import the micrographs and estimate per micrograph defocus (CTF correction)

Output file: micrographs/*.hdf

If you expect some degree of bad micrographs and your dataset is not too big (or you have time for visually inspecting the whole dataset) it is possible to exclude the low-quality micrographs already at the level of data import. This is done by choosing *Raw Data* → *Evaluate & Import Micrographs* instead of the procedure described above. This will run program called e2evalimage.py. For details refer to EMAN2.2 tutorial at EMAN Wiki.



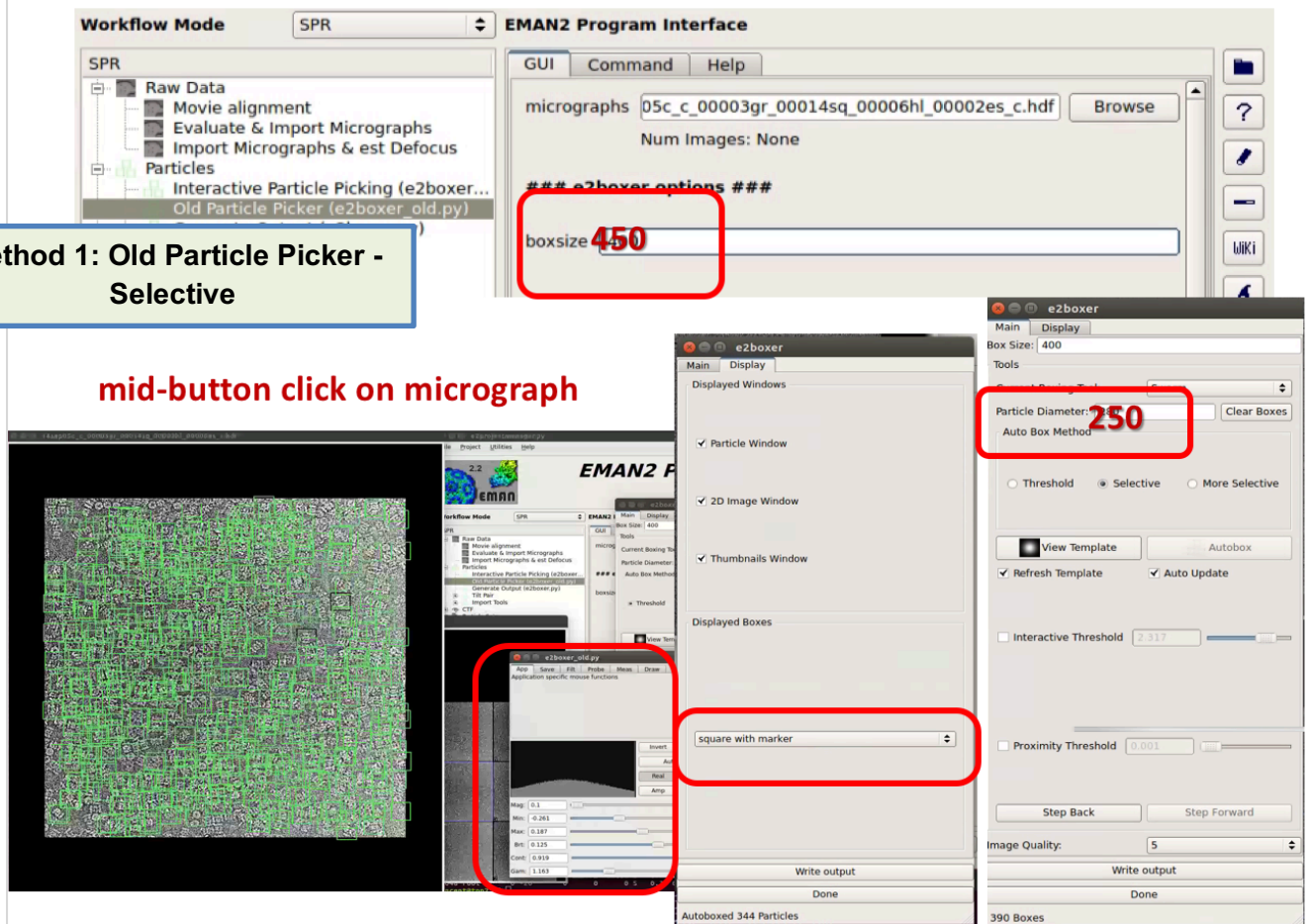
6. *Particles -> Old Particle Picker*

- ◆ Select all micrographs with *Browse*
- ◆ Boxsize = 450, *Launch*
- ◆ Particle Diameter = 250
- ◆ Arrange windows so micrograph fills most of display
- ◆ In *Selective* mode left click to select a few 'good' particles, and the autoboxing will be started spontaneously.
- ◆ If you don't like the autoboxing result, press *Clear Boxes*.
- ◆ When finished with an image, do NOT *write output*, just select the next image from the 'Thumbnails' window
- ◆ When all the images are done the particle picking, press *Done*

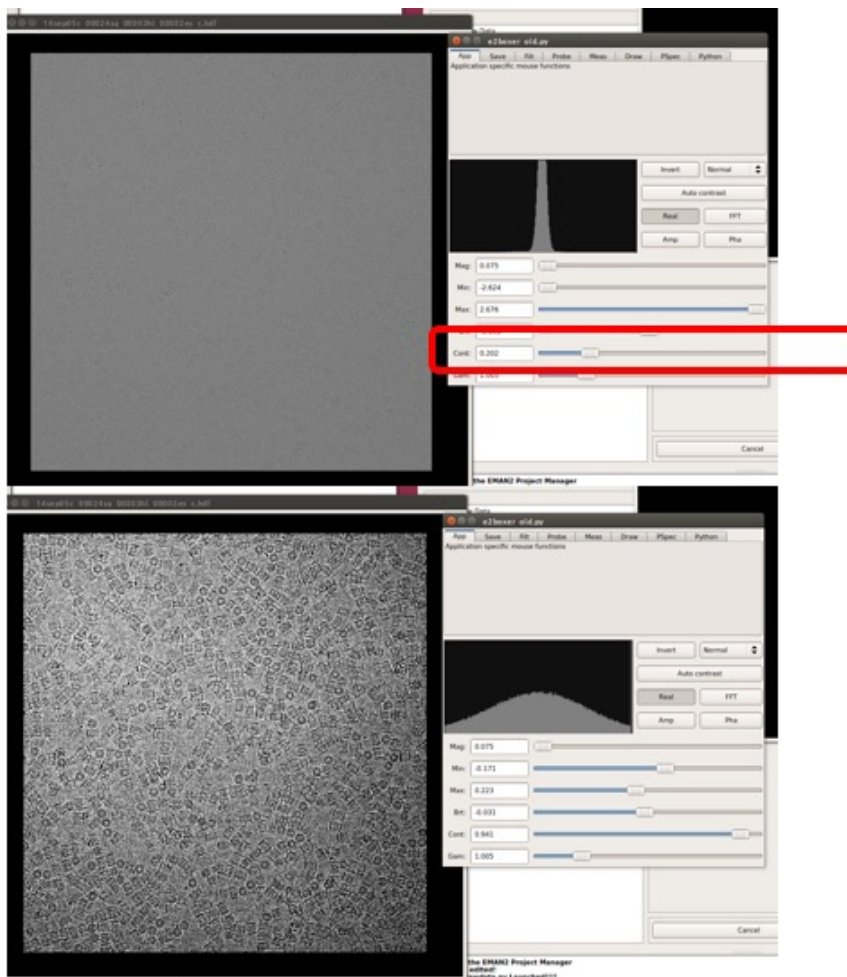
Output file: boxfiles/*.box files with particle coordinates

Method 1: Old Particle Picker - Selective

mid-button click on micrograph

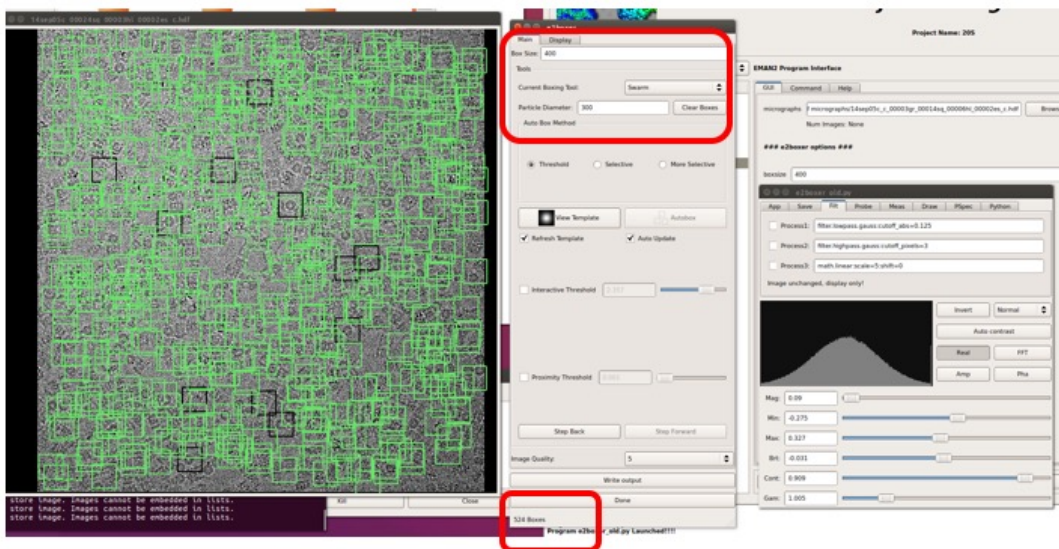


Change contrast of selected micrograph by dragging the SIGMA bar.

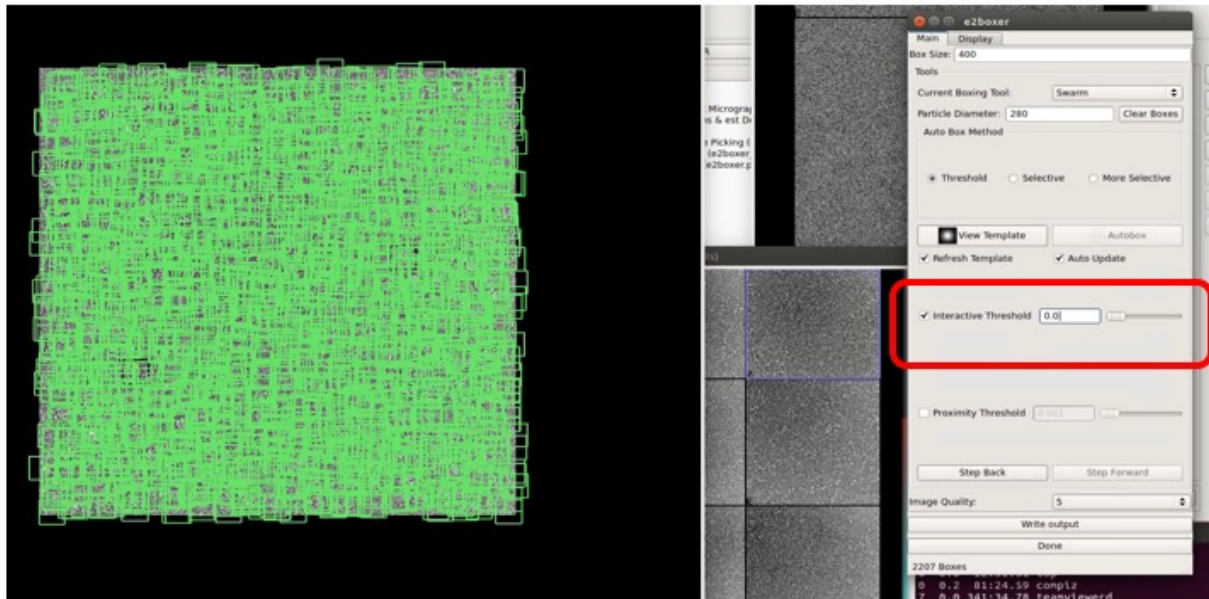


Method 2: Old Particle Picker - Threshold

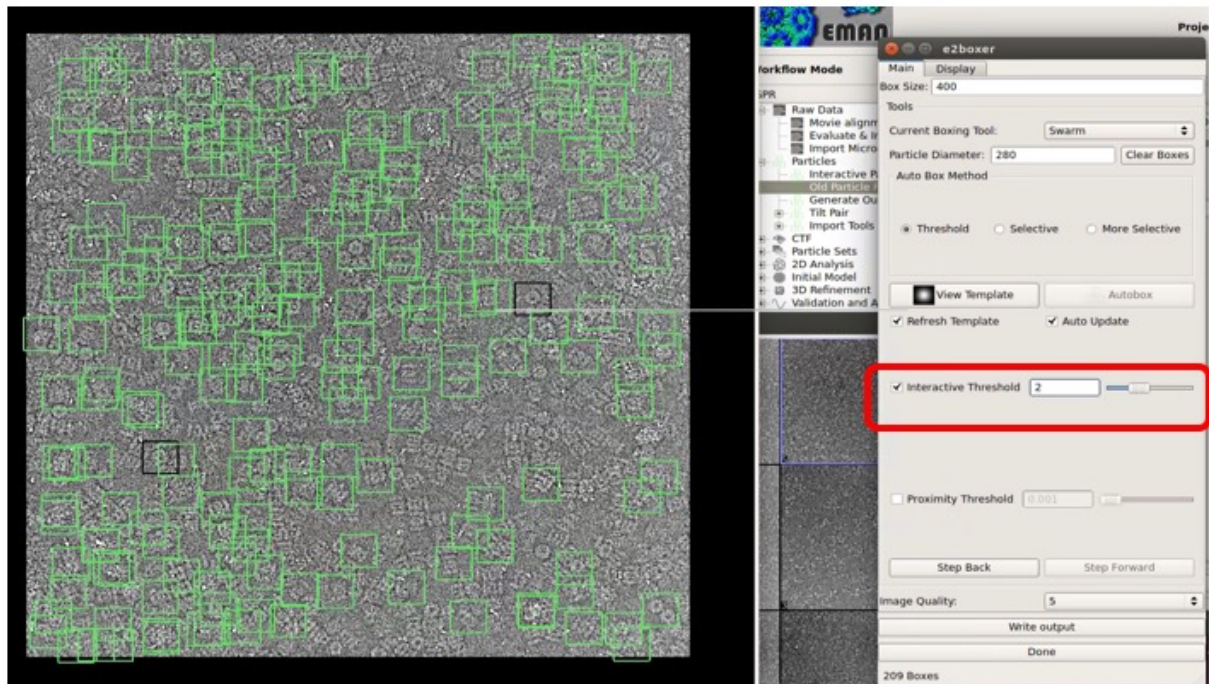
Select few more 'model' partials



Threshold = 0 (too low!)



Threshold = 2

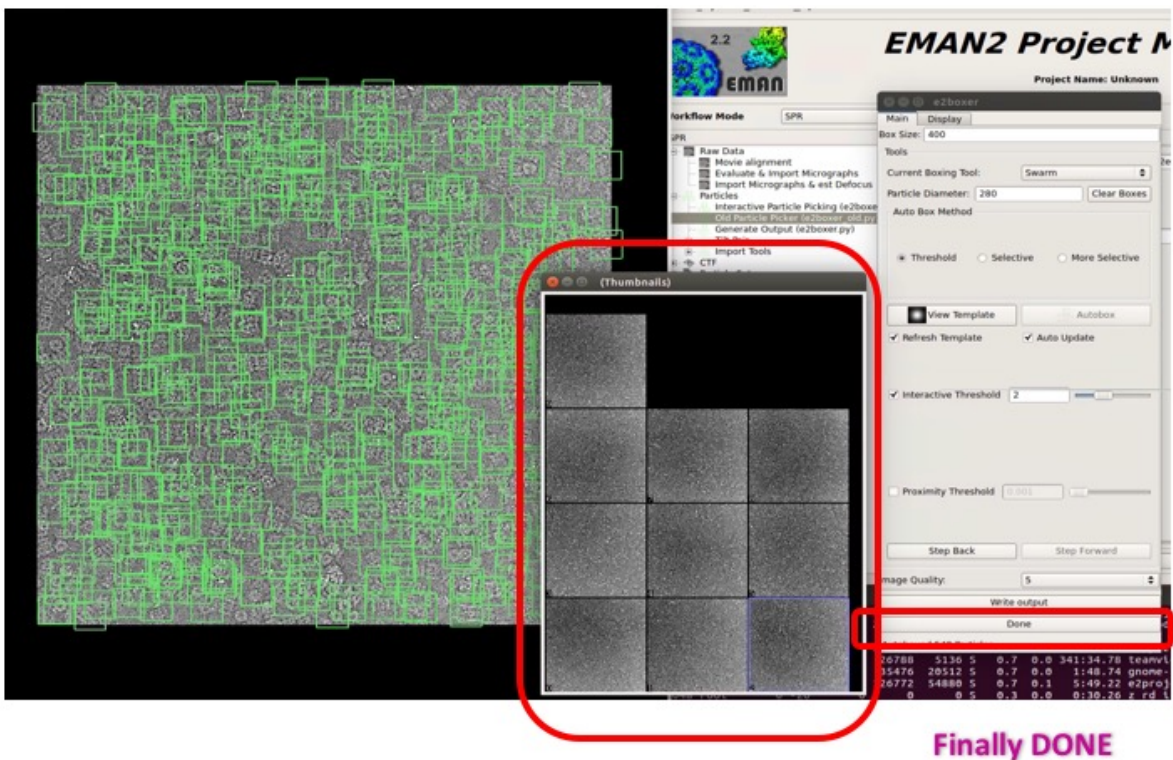


Pick 4 more particles -> more good models

Threshold = 2



Go through the other micrographs



7. Particles → Generate Output

- a. Select checkboxes:

write_ptcls and **allmicrographs**

By checking all micrographs box, you don't need define which micrographs to process.

- b. Set:

boxsize = 450

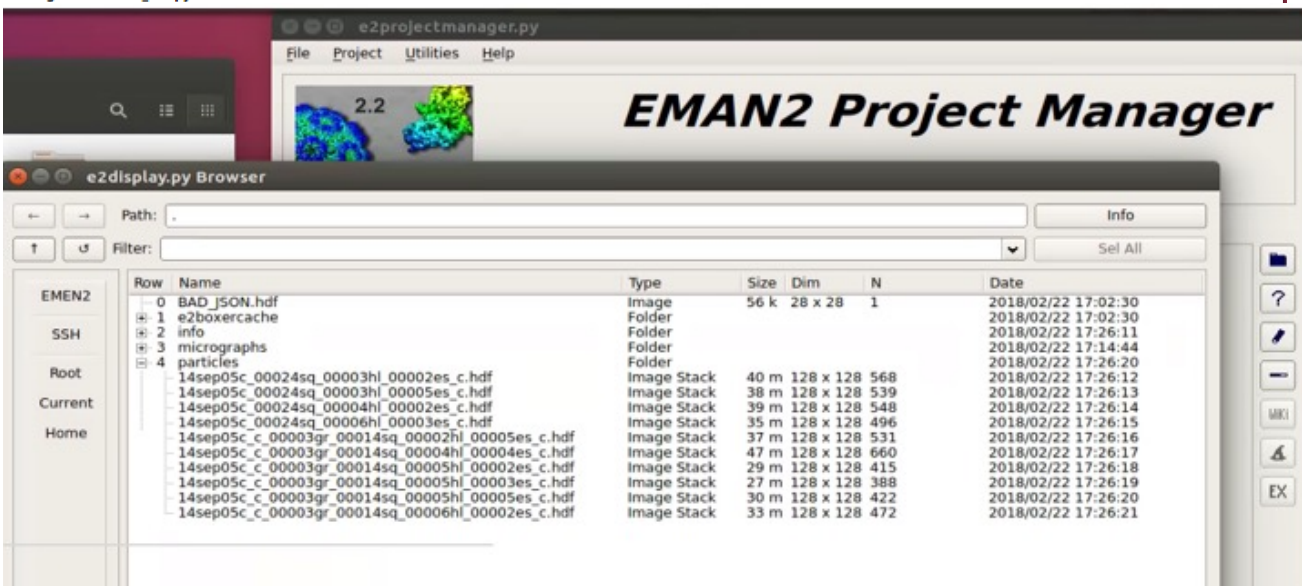
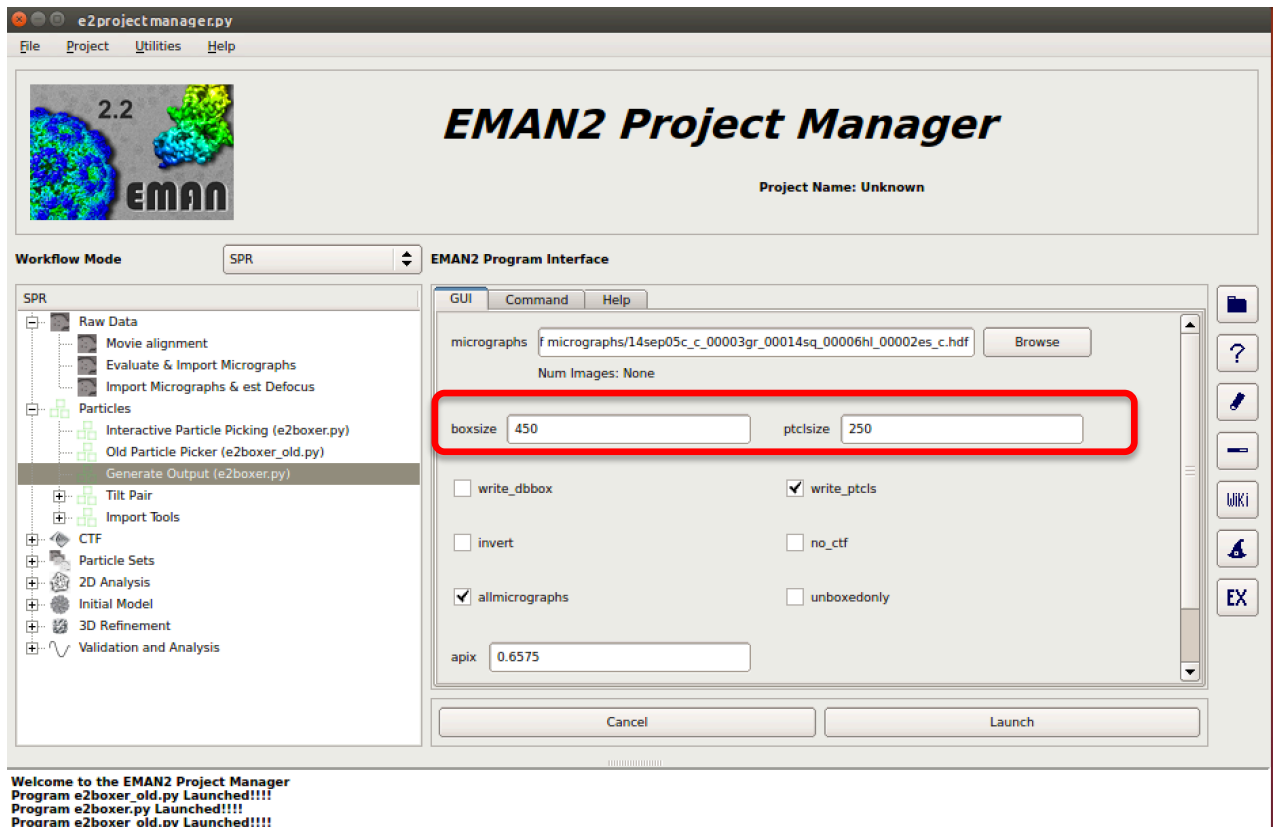
ptclsize = 250

apix = 0.6575

- c. Press **Launch**.

This will extract individual particle images from the micrographs and save them.

Output file: particles/*.hdf



8. **CTF** → **CTF Autoprocess**

a. Select checkbox:

hires, highdensity

For low-resolution data e.g. negative stain it may be better to choose midres or lowres box. This parameter defines the resolution of the downsampled data sets saved together with the full resolution particles.

highdensity parameter is used because of high density of 20S particles in the micrographs.

b. Set:

Threads = 20

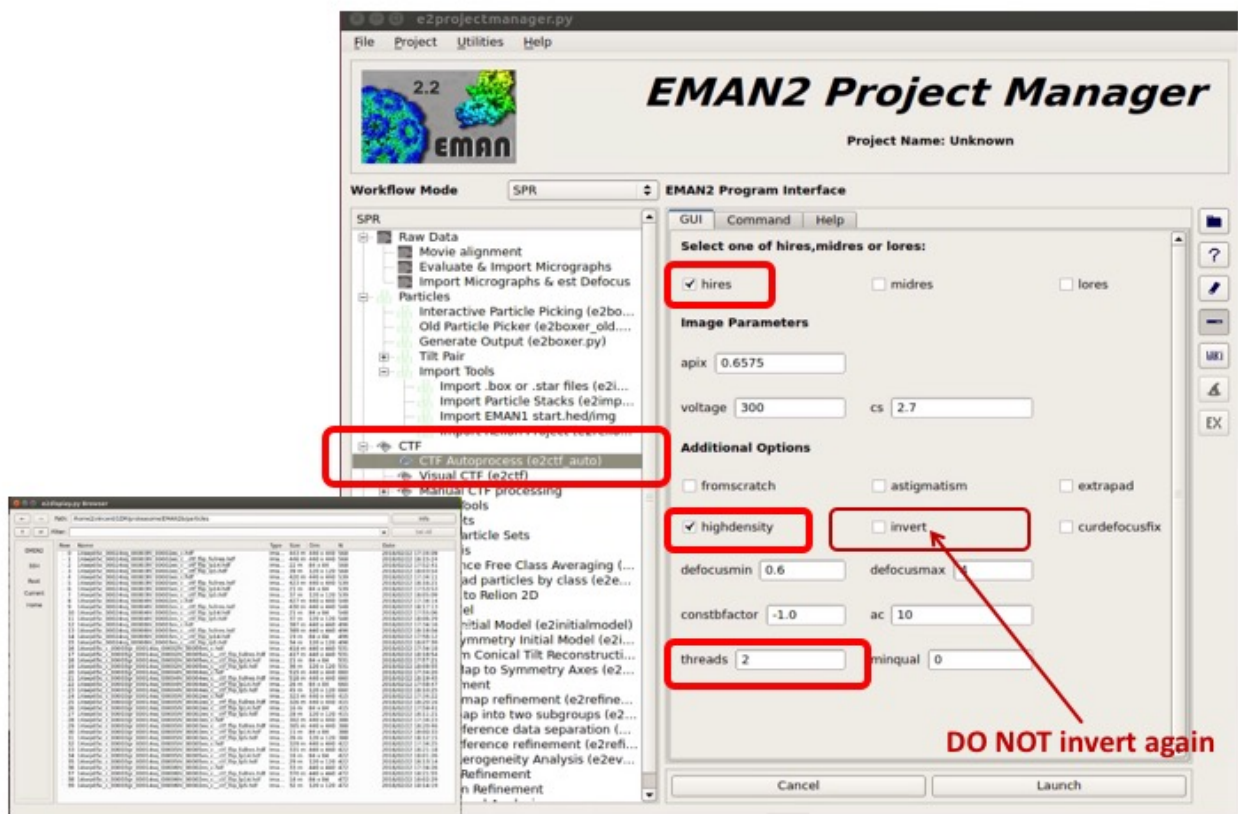
This basically tells the program how many processors to use on the local computer. You may increase it if you have free computational resources (CPUs) which will speed up calculations.

c. Press **Launch**

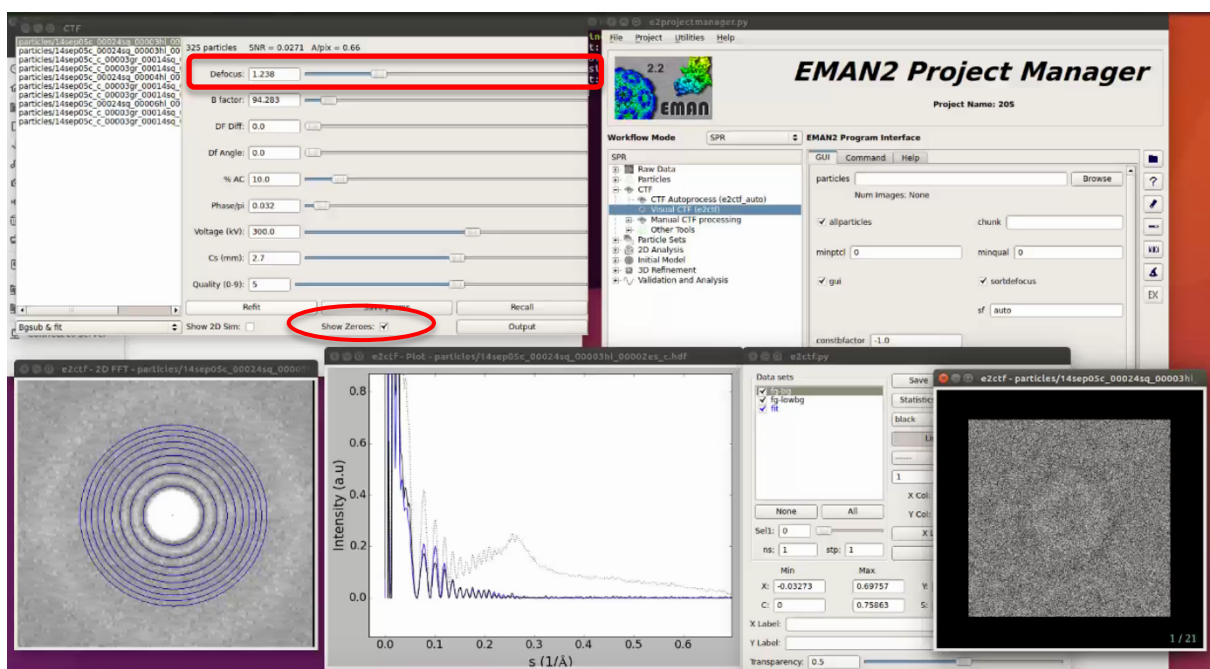
Output file: **sets/all_ctf_flip_fullres.lst**

all_ctf_flip_lp14.lst

all_ctf_flip_lp5.lst



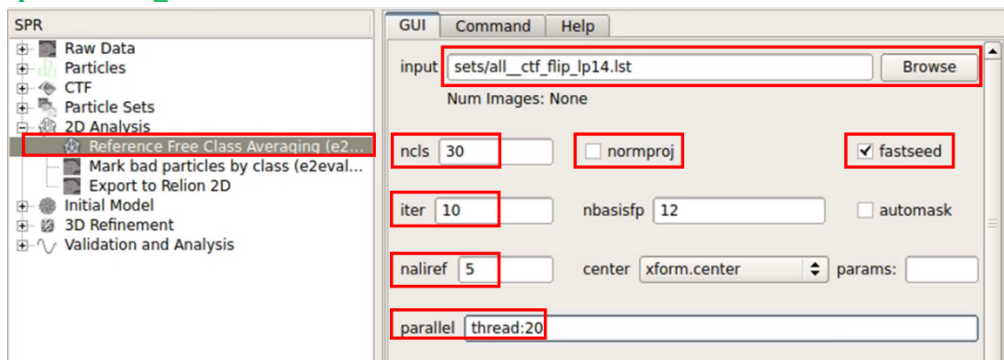
9. **CTF** → **Visual CTF** (not crucial for the 20S tutorial dataset)
- Select checkboxes:
allparticles, **sortdefocus** and **gui**
You can leave the *particles* field empty
 - Press **Launch**
You will get 4 windows
 - “CTF” window allows you to select which image to work on, and adjust the various parameters for that image interactively
 - “Particles” window shows the average of the first 20 particles (without alignment) as well as the first 20 particles individually (use up and down arrow in that window)
 - “2-D FFT” window shows the average 2-D power spectrum for the particles in the current image
 - Plot window with several curves. In "Bgsub & Fit" mode,
 - The black curve and dotted black curve are averaged power spectrums for the experimental data provided (they differ only in a background subtraction scheme)
 - The blue curve is the model fit based on the parameters shown in the control panel.
 - In the “CTF” window check the **Show Zeroes** box.
 - Using left side panel in the “CTF” window (with the list of micrographs) switch between micrograms and evaluate correctness of the calculated CTF oscillations. Check the first dozen closest to focus and last dozen far from focus images (top and bottom of the list). For each selected micrograph:
 - Oscillations in blue and black curves shown in Plot window should match. At this point we can expect significant deviations at low resolution, this it is acceptable. There is no need for a perfect fit in amplitudes as long as the zeroes are in the right places.
 - The zeroes are also depicted as Blue rings in “2-D FFT” window. They should match the black rings visible of your CTF pattern.
 - If you find an image with a significantly incorrect defocus (i.e. blue line oscillations “CTF” window do not match black lines and Blue rings in “2-D FFT” window do not match the black rings visible of your CTF pattern), then adjust it (using the *Defocus* slider) so it’s approximately correct, and hit the *Refit* button (you can use it several times if necessary). **## If you correct any defocus values, before going to next step, rerun the automatic process in step 8.**



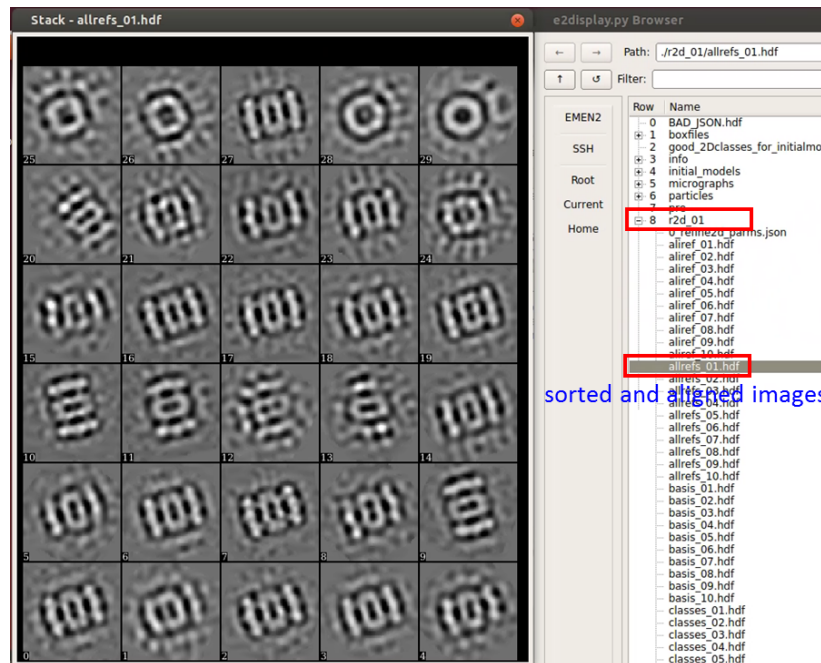
10. **2D Analysis** → **Reference free class averages**

- a. Select checkbox:
fastseed (speed up calculations)
- b. Uncheck:
normproj (unchecking *normproj* will tend to do a better job in separating junk from real particles, whereas keeping *normproj* will tend to produce more different views (of lower quality).
- c. Set:
input = sets/all__ctf_flip_lp14.lst (we will use downsampled data to speed up calculations)
iter = 10
ncls = 30 (number of classes)
naliref = 5 (number of references that will be used to classify particles)
parallel = thread:20 (or however many cores you have)
- d. Press **Launch**

Output file: **r2d_01/...**



When checking the results, it is often better to look at the allrefs_X.hdf file instead of classes_X.hdf (where X is the number of iteration). Both contain the same images, but **allrefs contains sorted and aligned images** to make it easier to compare the different averages visually.



11. **2D Analysis** → **Mark bad particles by class**

- a. There are no options, just press **Launch**
- b. From the window *e2evalparticles* that will appear **select r2d_01/allrefs_10.hdf** (i.e. aligned images class averages from last iteration of 2D classification).

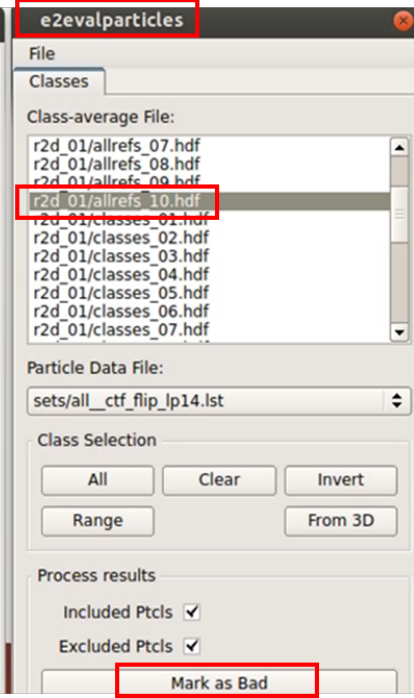
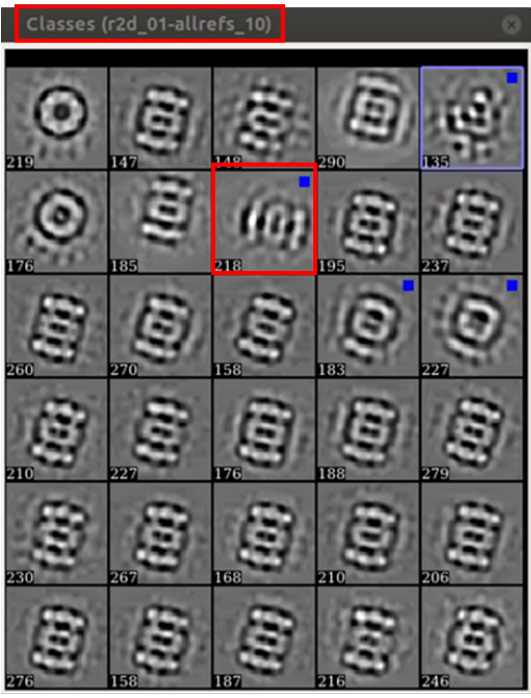
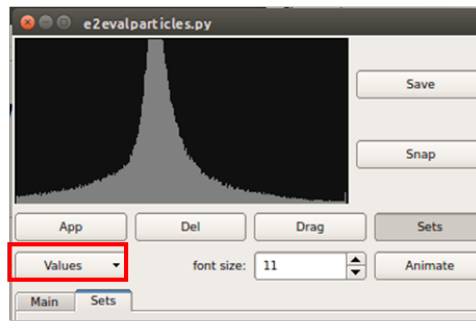
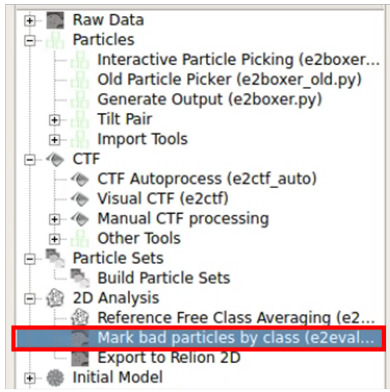
This will cause 3 additional windows to open:

- Included Particles
 - Excluded Particles
 - third will show the class-averages from the selected file (can be covered by the other two windows)
- c. Double-click on the class-averages that look obviously bad and do not resemble your target complex (i.e. incomplete complexes, ice contamination etc.). This will put a blue square mark in the upper right corner of the class-average image.

It is expected that 3-6 classes should look obviously wrong and therefore, be marked as bad. There is a random element to the 2-D class-averaging process, so everyone will see a somewhat different set of class-averages.

Single click on a class-average will display particles associated with the class (in Included and Excluded Particles windows). This may help to decide whether mark this class as a “bad” and exclude all the particles in this class from further calculations or not. The goal is to eliminate 80-90% of the bad particles without removing too many good ones. The 2Dclassification & Mark bad particles data-set cleaning procedure can be repeated so it is not necessary to try to remove all the bad particles after a single round of 2D classification if some of the class-averages are doubtful. Number of particles associated with each class-average can be directly displayed by middle-clicking on the class-averages window and then selecting *ptcl_repr* from the *Values* drop-down menu)

- d. **After marking all the “bad” classes press the *Mark as Bad* button** in the *e2evalparticles* window and confirm that you want to do this in the dialog that appears.
- e. Exit the program (close the *e2evalparticles* window) when you’re done.

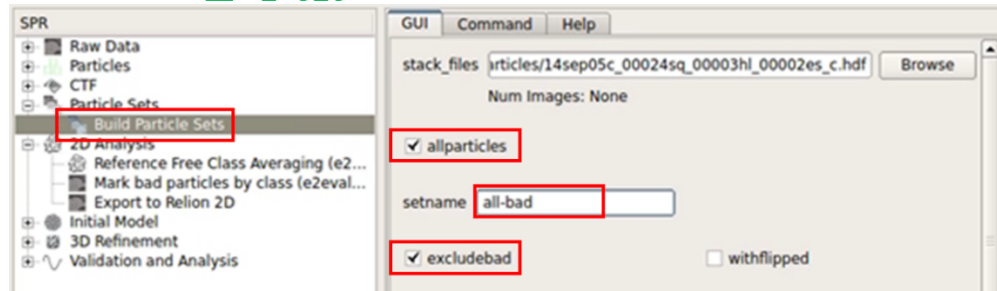


12. **Particle sets** → **Build particle sets**

- a. Select checkboxes:
allparticles and **excludebad**
- b. You can leave the *stack_files* field empty or select all the files in */particles* directory.
- c. **Setname = all-bad**
- d. Press **Launch**


This will build data-sets (full resolution and low-pass filtered particle lists) excluding all the particles associated with class-averages marked as “bad” in the previous step

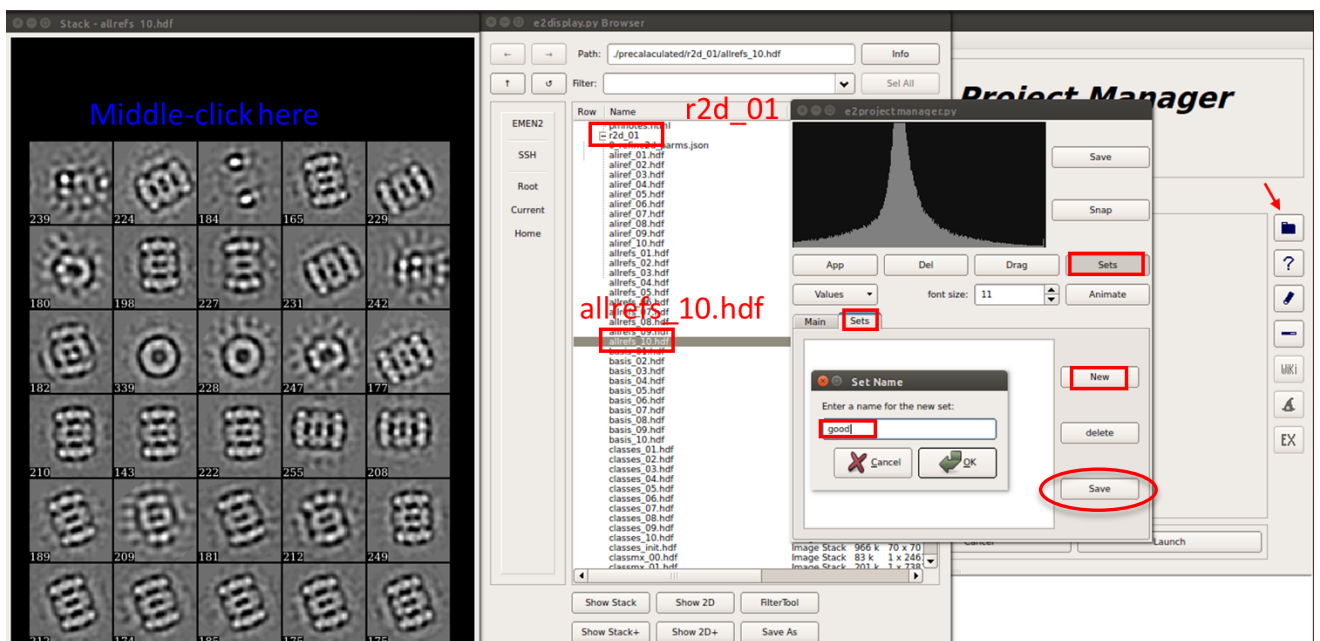
Output file: sets/all-bad_ctf_flip_fullres.lst
all-bad_ctf_flip_lp14.lst
all-bad_ctf_flip_lp5.lst



13. Select good class-average images to build initial model.


The selected class-average will be saved as particle images and then, in the following step, used to estimate the initial structure of the complex (the images of individual particles associated with the selected class-averages will not be used which will minimize computational costs).

- Open the **file browser** ( top icon on the right side of the project manager window)
- Browse to *r2d_01*, and double click on *allrefs_10.hdf*
- Middle-click on the new window showing the class-averages to bring up the control panel.
- In the control panel press the **Sets** button under the histogram plot, then also select the **Sets** tab (next to *Main*).
- Press the **New** button and type **good** in the box that appears. Click on the colored word **good** to make it active.
- Now single-click on class-averages you think they are good. You will see a colored mark appear. Click a second time if you change your mind and wish to exclude the average.
- Mark 10-15 good averages representing different views of the complex.
- With the set **good** highlighted, please press the **Save** button in the lower right corner of the control panel (**not the Save button in the upper right**). This will save the class-average images to a specified file. Name the file **good_classes.hdf**
- You can now close the browser.



14. **Initial Model**→**Make Initial Model**

- a. Set:
 - Input = good_classes.hdf**
 - iter = 24**
 - tries = 10**
 - sym = d**
 - Symmetry Number = 7**
 - random orientation = **randorient** checked
- b. Press **Launch**
- c. Evaluate results and identify best initial model

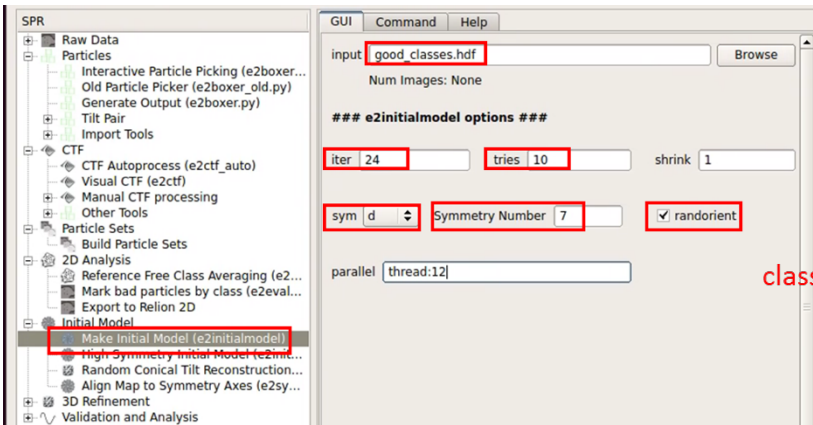
- using **file browser**  go to **/initial_models** folder
- check **model_XX_YY.hdf** (3D density) for each of the obtained models. Single click on the file and then click **Show 3D+** at the bottom of the file browser window. **Double-clicking the file will open the 3D map along with previously displayed one!** Middle-click in the 3D viewer window will show a histogram window. The density display threshold can be changed by sliding the *Threshold* slider at the bottom of the window.
- Good model should exhibit barrel shaped 3D density at low iso-surface display threshold with four 3-4 rings/layers visible at high threshold setting.
- Check the **model_XX_YY_aptcl.hdf** file for each of the models. Selecting the file and clicking **Show Stack+** button will display used class-average images with it respective model projection.
- For a good initial model, the class-average images and their model projection should agree with each other very well (aside from noise).

Output file: initial_models/model_XX_YY.hdf

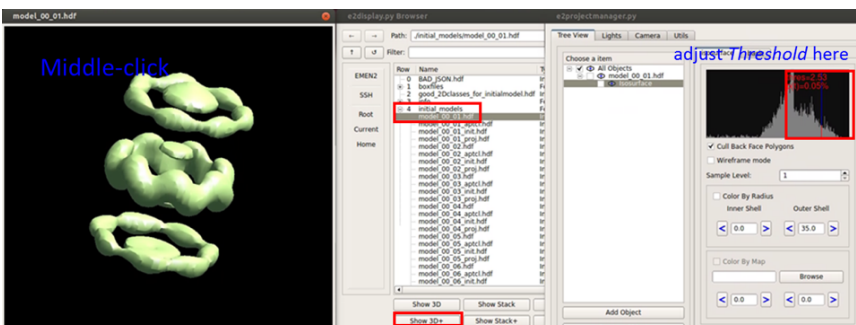
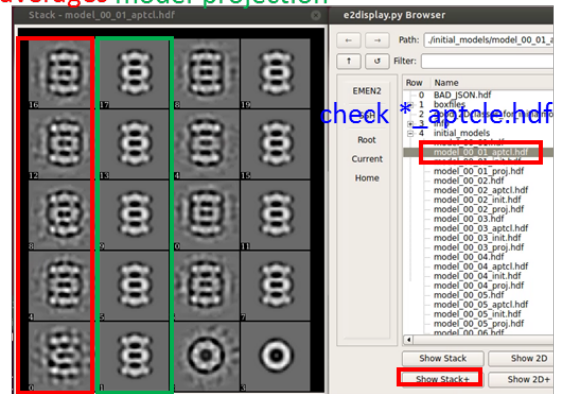
model_XX_YY_aptcl.hdf
model_XX_YY_init.hdf
model_XX_YY_proj.hdf

XX will be 00 when you first run the initial model job.

YY is the number of the initial model generated in this run of the job.



class-averages model projection



15. **3D refinement** → **Single map refinement (e2refine_easy)**

a. Set:

Input: `sets/all-bad_ctf_flip_lp5.lst` (for a new data-set/refinement it is better to use first low-pass filtered data or even a subset of the data to speed-up calculations and quickly assess the quality of the data and identify potential problems)

model: `initial_models/model_00_01.hdf` (or the model that you identified as the best)

targetres = 5 (target resolution 3D reconstruction)

sym = d7

iter = 6

parallel = thread:20 (or how ever many cores you have)

threads: 20

--threads option - In addition to --parallel, some commands have a --threads option. There are a few commands which cannot be run using the generic multi-computer parallelism provided by --parallel. These commands may still be able to take advantage of multiple cores on a single machine. --threads is the number of available processors on a single computer. It should be specified in addition to --parallel when both are available.

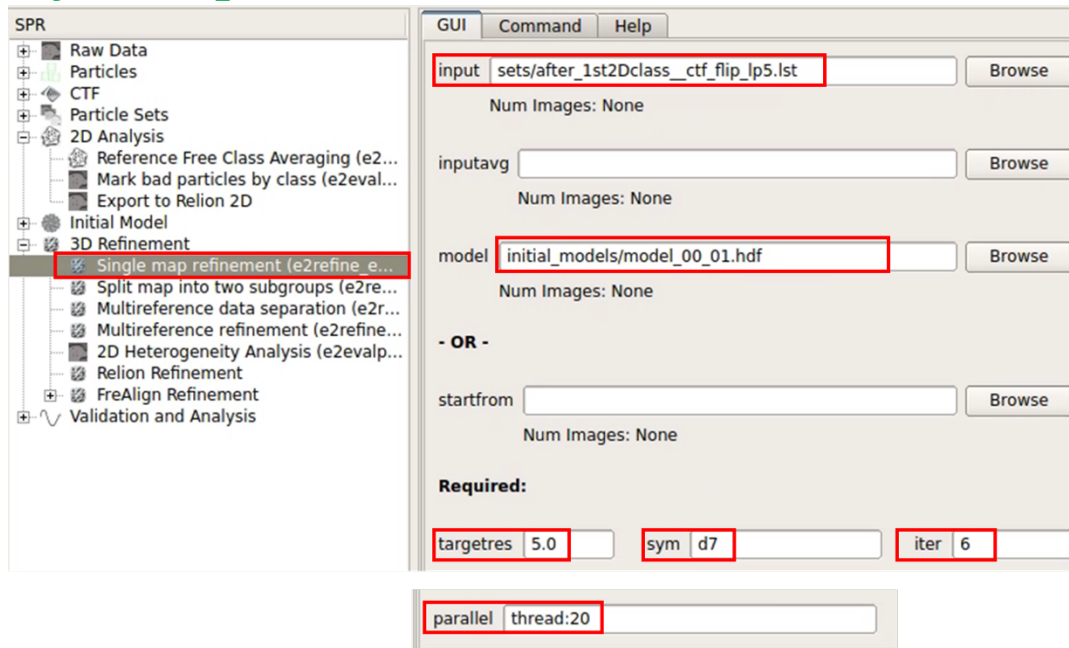
b. Press **Launch**

c. The progress of the refinement and results from each interaction can be viewed while the job is running by opening the `refine_01/report/index.html` file.

d. Quickly evaluate final result by looking at the 3D map from the last iteration (`/refine_01/thread_06.hdf`)

e. Comparing experimental data (`/refine_01/classes_06_odd.hdf` or `...even.hdf`) with projections of the final 3D map (`/refine_01/projections_06_odd.hdf` or `...even.hdf`) can be another good way to evaluate how good the calculated density describes observed particle images and also identify some of the missing views of the complex in your data-set.

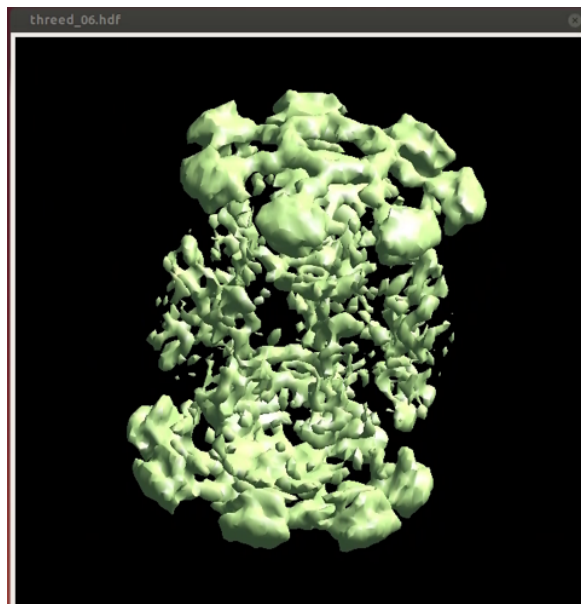
Output file: `refine_01/...`



	Row	Name	Type	Size	Dim	N	D
EMEN2	0	BAD_JSON.hdf	Image	2 m	834 x 832	1	20
SSH	1	boxfiles	Folder				20
	2	good_2Dclasses_for_initialmodel.hdf	Image Stack	411 k	70 x 70	12	20
Root	3	info	Folder				20
	4	initial_models	Folder				20
Current	5	micrographs	Folder				20
	6	particles	Folder				20
Home	7	pre	Folder				20
	8	r2d_01	Folder				20
	9	raw	Folder				20
	10	refine_01	Folder				20
		0_refine_parms.json	JSON	2 k	-	-	20

Current		projections_up_odd.nrr	Image Stack	4 m	120 x 120	77	20
Home		report	Folder				20
		converge.pdf	-	21 k	-	-	20
		converge.png	Image	96 k	640 x 480	1	20
		index.html	HTML	41 k	229 ln	-	20
		resolution.pdf	-	21 k	-	-	20
	resolution.png	Image	69 k	640 x 480	1	20	

		threed_05.nrr	Image	6 m	120 x 120 x 120	1	20
		threed_05_even.hdf	Image	6 m	120 x 120 x 120	1	20
		threed_05_odd.hdf	Image	6 m	120 x 120 x 120	1	20
		threed_06.hdf	Image	6 m	120 x 120 x 120	1	20
		threed_06_even.hdf	Image	6 m	120 x 120 x 120	1	20
		threed_06_odd.hdf	Image	6 m	120 x 120 x 120	1	20
		threed_even_unmasked.hdf	Image	6 m	120 x 120 x 120	1	20
		threed_odd_unmasked.hdf	Image	6 m	120 x 120 x 120	1	20
	11	sets	Folder				20



16. *Validation and Analysis* → *Run e2eulerxplor*

- a. Press **Launch**
- b. Middle-click somewhere on the plot background
- c. Select result of your last refinement interaction (i.e. `/refine_01/threed_06.hdf`)
- d. Good reconstruction should have all the views of the complex covered i.e. your data-set contains particle images representing views of the complex from all of the possible angles providing unambiguous 3D reconstruction of the studied structure.
- e. You can check which views of the complex are most- or less abundant in your data-set by double-clicking on the tallest and shortest bars respectively.
- f. For C1 symmetry the plot has a sphere-like shape. The plot observed for 20S is a result of used D7 symmetry.

