Things you will need to know before launching the GUI:
[where a name is given in bold_italics this denotes a simple keyword that could be used to set the parameter when using the command line implementation of simple]

1. Things about your local computer system:
   a. Where are your data located? [e.g. /home/user/name/project1 ?]
   b. Are you going to run your jobs locally or on a cluster using a queueing system? [e.g. slurm?]
   c. What is the architecture of the computers you will be running the jobs on? [e.g. how many CPU cores with shared memory (nthr), how many processors are available for you to run your job over allowing for hyperthreading (nparts)]. Here the important thing is how many CPU sockets you will be using (nparts) and how many CPU threads should you launch on that socket (nthr). Nthr should generally be identical to the number of physical CPU cores on the socket but on systems with hyperthreading you may want to increase nthr. The thread affinity of OpenMP is set to be close within a socket to reduce latency.
   Example setting:
   CPU cluster with 10 nodes, each with 2 x 16 core processors nthr=16 nparts=20

2. Things about your microscope and data collection:
   a. Microscope voltage [e.g. 300kV for Titan Krios, 200kV Talos Arctica]
   b. Microscope cs value [e.g. 2.7 for Titan Krios or Talos Artica]
   c. Fractional amplitude contrast [e.g. ~0.1 for cryo data from standard protein, 0.2-0.3 for negative stain data]
   d. Size of pixels in your recorded movies/micrographs – this is dependent on what detector was used, exactly where it is positioned c.f. the sample in the microscope and the magnification at which the data were collected. In SIMPLE this value is referred to as the “Sampling Distance” or smpd and has units of Angstroms/pixel. If you have taken your data on most modern microscopes you will be able to read this value from the microscope control software. It may also be correctly written in the header of the movies/micrographs (inspect this on the command line using the command “simple_exec prg=iminfo fname=name_of_your_movie.mrc” if you haven’t noted this down at the time of data collection).
   e. If you wish to perform dose weighting you will also need to know the dose rate (in electrons per second, dose_rate) and exposure time (seconds, exp_time) used to collect the data.
   f. Defocus range set during collection – this may be needed to fit ctf parameters (particularly if the range used is unusual) and can also be useful in selecting problem movies/micrographs which can sometimes be identified as those that end up with estimated defocus values significantly outside the range set during collection.

3. Things about your sample:
   a. Molecular weight of the object you are imaging (you will use this when you wish to apply a molecular mask during 3D volume refinement).
   b. Anything you know or suspect about any internal symmetry – if there is none, then the symmetry of the object (referred to as pgrp denoting point group symmetry) is c1.
   c. Longest dimension of the object – this can be measured by inspecting your micrographs/motion corrected movies in the “Manual Picking” module of the GUI. This information is needed to
calculate a suitable radius (in pixels, referred to as \textit{msk}) for a soft spherical mask that is applied during processing centered particles [e.g. longest dimension of object is 120\(\text{A}\), \textit{smpd} is 1.2 \(\text{A/pix}\) therefore \textit{msk}=\frac{120}{(1.2\times2)}=50 \text{ pixels}]

Information for Samples Used In These Examples

1. Urease
   This is a 1.4MDa object with tetrahedral (\textit{pgrp}=t) symmetry
   a. Negative stain data
      i. Microscope voltage \textit{kv}=200
      ii. Microscope \textit{cs}=2.7
      iii. Amplitude Contrast \textit{fraca}=0.25
      iv. Defocus range -0.1 to -1 micron
      v. Sampling Distance \textit{smpd}=1.43 \text{Angstroms/pixel}
      vi. Object Radius in pixels \textit{msk}=80 [and a box of 240 pixels for extraction]
   b. CryoEM data
      i. Microscope voltage \textit{kv}=200
      ii. Microscope \textit{cs}=2.7
      iii. Amplitude Contrast \textit{fraca}=0.1
      iv. Defocus range -1 to -3 microns
      v. Sampling Distance \textit{smpd}=1.26 \text{Angstroms/pixel}
      vi. Object Radius in pixels \textit{msk}=

2. Beta Galactosidase
   This is a small number of particles from the EMDB data set deposited by XXXX
   i. Sampling Distance \textit{smpd}=1.272
   ii. Object Radius in pixels \textit{msk}=88
TASK ONE - Make the first project in the SIMPLE GUI

This task will show you how to start the SIMPLE GUI and create a project that points to a directory you have already created on the local file system. You should have already unpacked the test data sets into three directories with the data needed. The first project we create will be to work with micrographs from a negative stain urease sample.

1. Start a terminal and type “simple” followed by return key.
2. Open firefox.
3. Type the address “localhost:8088” followed by return in browser address bar.

4. You should now have the SIMPLE GUI launched. In the top right corner find the “Select Project” selector.

5. Left click on selector and left click “Manage Projects” to open the Project Manager page.
6. On the project manager page left click on the SIMPLE logo next to “Project Manager” on the left hand side of the page to bring up the menu.

![Project Manager](image)

7. Left click in “New Project” box to bring up the “New Project” window.

![New Project](image)

8. Input the data as shown below using a left click to locate in each box. Due to the complexity of the paths on these computers we will just enter the path directly in the “Project Folder” box. The path will be of the form “/data/labXX/simple_workshop/urease” where labXX is the name of the computer you are sitting at (found on a sticker at the top left of the screen).

![New Project](image)

When using your own install you will be able to left click on the folder icon and then browse to the project folder you want to use. Note that you can directly type the path into the “Project Folder” box on this first screen but you cannot type the path into the box at the top of the file browser on the window generated by clicking on the folder icon.

9. You have now created the first project.
TASK TWO - Determine the Contrast Transfer Function (CTF) parameters for the urease micrographs

In this task we will use the module “CTF estimation” to determine the CTF parameters for a set of negative stain urease micrographs. Unless otherwise specified “select” means “left click on item”.

1. Use the project selector to select the “urease” project you created in Task One.

2. Now create a new job by selecting the “menu” icon to the LHS of the “SIMPLE” logo highlighted above.

3. Selecting the “New Job” item will bring up the set of modules that can be run in the central panel.

4. Select the “CTF Estimation” module
5. Fill in the “Job Parameters” page as shown below and click the “Next” button - this is common to all modules and will not be described in later tasks.

6. For all modules this next page contains the set of parameters which are required to run the module. The grey text in the boxes are short descriptors of what needs to be entered in the box with any default value shown in [].

For the “CTF Estimation” module the inputs need to be:
(i) a set of micrographs - which can be given as list in the upper box (this list might have been generated by another module) or by indicating the folder containing the micrographs, as we will do by clicking on the folder icon to the right of the “Micrographs Directory” box and selecting the “micrographs” folder that we have created within the “urease” folder.
7. The next page contains a set of optional parameters. In this case we need to change the default value for the “Minimum Expected Defocus” to 0.1 microns (see below). All other boxes can be left unfilled and will default to suitable values.

8. The next page defines how the job will be run. The settings shown below are those used for the small desktop PCs used in Oxford for the workshop. Enter the parameters shown below. This step is common to all modules and will not be described in later tasks. If you will be submitting your job to queuing system then indicate this in the box and further pages will appear prompting you for details of the queuing system.

9. The job should take ~2min to complete on a i9 core desktop. Jobs can be inspected from the project “History” page. Selecting the “cog” icon will bring up a menu with a variety of job related tasks including re-running the job, deleting the job, viewing the job log file and viewing any outputs from the job.
10. View the results by selecting the "View Output" option. For this module you will see a panel of micrographs and power spectra. The LHS of the power spectrum is the experimental one and the RHS the predicted one using the CTF parameters determined.

The CTF parameters can be viewed by selecting the orange "CTF" box within the micrograph frame. Close this pop-up by clicking on the "X".

The data can be sorted in a variety of ways. Click on the "Sort" logo and choose sort to see the options.

Choose how to sort the images and then close the pop-up. You may need to scroll up and down to refresh the browser window.
The display contrast is altered by right clicking anywhere over the images to bring up the “View Controls” pop-up. Alter slider bars and hit update to change.

Left clicking a micrograph/power spectrum pair will de-select it (it will be greyed out) and the selection could be saved from the menu accessed via the “S” logo. Left clicking on the greyed out item will re-select it. We do not need to deselect any of the micrographs in this dataset.

The data can be sorted in a variety of ways. Click on the “S” logo and choose sort to see the options.

11. When you have finished inspecting the results close the viewer window by clicking on the “x” to return to the project history menu.

The principles of selecting, sorting, saving selections, opening and closing pop-ups and altering view parameters are consistent between different modules and will not repeated in further tasks.
TASK THREE - Manual pick of particles in urease negative stain micrographs

In this task we will use the module “Manual picking” to demonstrate how the SIMPLE GUI can be used to manually pick a small number of micrographs to give a sense of size and generate particles to be used in a 2D clustering and then used to make references to automatically pick micrographs. Unless otherwise specified “select” means “left click on item”.

1. Use the project selector to select the “urease” project you created in Task One.

2. Now create a new job by selecting the “menu” icon to the LHS of the “SIMPLE” logo highlighted above.

3. Selecting the “New Job” item will bring up the set of modules that can be run in the central pane.

4. Select the “Manual Picking” module
5. Move through the Job Parameters page as in previous task and now fill in required parameters using the upper box to select the "ctffit_out.simple" file which has been written in the previous job folder (and contains a list of micrographs) and entering the sampling distance which is 1.43Å.

6. This opens the picking window and selecting a micrograph in the list will lead to its display in the centre (it will take a few seconds to load the image).
7. You can select to show both an outline of the box that will be used to extract the particle and a circle with the radius you define as appropriate for the particle by using the check boxes below the scale bar. Left clicking on a particle will place a box at that position, these can be removed by right clicking the box. The box should be big enough to safely encompass the particle with background around it (240 pixels is appropriate in this case). Play with setting the particle radius to different values (80 pixels is appropriate for these data) and rescaling the image - note this particle radius as you will use this as a mask radius for further operations. Two different zooms are shown below.

A particle count for that micrograph is shown at the top of the micrograph. Boxes are automatically written and do not need to be saved in an additional step. If you were doing this in real life, you would probably work through sufficient micrographs to select a few hundred to a thousand particles. But we will continue in the next task with boxes we have previous picked using the "Automated Picking module".
TASK FOUR - Extract particles in urease negative stain micrographs

In this task, we will use the module “Particle Extraction” to extract particles and create a parameters file with the CTF parameters from TASK TWO 1 from the urease micrographs for further processing. Unless otherwise specified, “select” means “left click on item”.

1. Use the project selector to select the “urease” project you created in Task One.

2. Now create a new job by selecting the “menu” icon to the LHS of the “SIMPLE” logo highlighted above.

3. Selecting the “New Job” item will bring up the set of modules that can be run in the central pane.

4. Select the “Particle Extraction” module
5. Move through the Job Parameters page as in previous task and now fill in required parameters using the upper box to select the “ctffit_out simple” file which has been written in the previous job folder (and contains a list of micrographs and CTF parameters), enter the sampling distance (1.43Å) and point towards the “boxes” folder that you will find in the “urease” area.

6. On the “Optional Parameters” page you need to indicate that these data are from negative stain micrographs on the selector (the default is to assume cryo images).

7. On the next page set the number of CPU threads to 4 and run the job - this will take less than 1 minute.
8. As in TASK TWO, you can use the “View Output” and “View Log” options from the project “History” page to view the output from the job. For the “View Output” you select a micrograph in the central selector and will then see the particles from that micrograph. A total particle count from that dataset is shown at the top. Display contrast etc are altered as in TASK TWO.
TASK FIVE - Cluster similarly oriented urease particles in 2D

In this task we will use the module “2D clustering” to define groups of urease particles that are the same view and average them to generate 2D clusters with a higher signal to noise. This process will also allow us to detect damaged or inappropriately picked particles.

1. Use the project selector to select the “urease” project you created in Task One.

2. Now create a new job by selecting the “menu” icon to the LHS of the “SIMPLE” logo highlighted above.

3. Selecting the “New Job” item will bring up the set of modules that can be run in the central pane.

4. Select the “2D clustering” module.
5. Move through the Job Parameters page as in previous task and now fill in required parameters using the upper box to select the "extract_out.simple" file written in the previous job, enter the sampling distance (1.43Å), give a suitable mask radius (80 pixels) and select CTF correction as shown below and ask for 100 2D clusters.

![Image of Prime 2D Clustering](image)

6. You do not need to select any of the optional parameters, so just click through that window and set up the job parameters as in TASK TWO. Then hit “Run”. This is a bigger computational job and will take ~6 minutes to run. As in TASK TWO, you can monitor progress by inspecting the logfile “View Log” and, as the cycles of 2D refinement run, you will be able to see the 2D clusters in the “View Output” screen. This module automatically down-samples the data at two levels during the process, before finally generating the clusters at the full sampling of the data. This, combined with the gradually improving allocation of like particles to the same cluster means more detail will become visible in the clusters as the iterations complete.

7. When the job completes, use “View Output” select the final (highest numbered cycle) and sort the clusters on the parameter “res” (see TASK TWO for instructions on how to access this menu).

![Image of View Output](image)

8. You can view the particles in each cluster and the parameters associated with each cluster by selecting the yellow INFO and PTCLS buttons in each one.
9. Deselect any fuzzy clusters - there are few here as negative stain images are of high contrast, but they are likely to be towards the bottom of the page if you have sorted on "res" and "ascending".

10. Save your selection using the menu accessed from the "S" logo using the default name and location.

This completes TASK FIVE. The selected clusters will be used to generate an initial 3D model of urease in the next task.
TASK SIX - Generate initial ab-initio 3D model from 2D clusters

In this task we will use the module “Initial Model Generation” to derive an initial 3D model of our object. We will do this using the 2D clusters as input as their enhanced signal to noise will aid establishing the correct shape. In this task we will assume that we know the symmetry of the object (which is tetrahedral), as this will speed our progress. In real life you would likely first perform a reconstruction in c1 (no symmetry within the particle), inspect the 3D volume and then repeat the process imposing the symmetry you observe in the asymmetric reconstruction.

1. Use the project selector to select the “urease” project you created in Task One.

2. Now create a new job by selecting the “menu” icon to the LHS of the “SIMPLE” logo highlighted above.

3. Selecting the “New Job” item will bring up the set of modules that can be run in the central pane.

4. Select the “Initial Model Generation” module
5. Move through the Job Parameters page as in previous task and now fill in required parameters using the upper box to select the “prime2D_selected_solediame” file written in the previous job, enter the sampling distance (1.43 Å), give a suitable mask radius (80 pixels) and select “l” for tetrahedral point group symmetry.

6. In the optional parameters use the selector to indicate that the “Point Group Symmetry” is known so that the symmetry is used from the first iterations (again speeding the calculation).

This module performs several different operations at various downsamplings of the original data before mapping this back to produce a final 3D volume (as an xxx.mrc file). You can view the log file in the usual way and look at a comparison of reprojections of the final volume compared to the 2D clusters were the input for this module but, as we have not yet implemented a 3D viewer within the GUI (watch this space!), you will have to open a terminal window, navigate to the folder and launch “chimera” to view the output volume (which will be called recvol_state01_iterXX_pproc.mrc where XX is the number of final iteration before convergence and can be found in the log file). The task will take ~10 minutes to complete on a i9 core desktop.
Using the view output option will show a comparison of reprojections of the volume (the upper of each pair) and the 2D clusters (lower of each pair)- note that the options for redisplay on this page are somewhat differently laid out on this page! You are seeing a preview of how the output pages will look in the next version of the SIMPLE GUI.

This completes TASK SIX and our work with this negative stain data set we will now move to looking at cryo-em data.
TASK SEVEN - Generate and refine a 3D model for beta galactosidase starting from a set of particles and defocus parameters

In this task we will recapitulate many of the steps you performed with the negative stain urease data to generate and initial 3D model from a set of beta galactosidase cryo-erm particles and defocus parameters.

1. As you did in TASK ONE, create a new project for beta galactosidase pointing at the directory containing the appropriate data.

2. Now run a 2D clustering job as you did for the urease data in TASK FIVE. The appropriate parameters for this job are shown below.

3. This job will take ~13 minutes to complete.

4. Once completed, make a selection of detailed 2D classes, save the selection and use these to generate an initial 3D model as we did in TASK SIX. Suitable parameters are shown below.
5. This job will take ~8min to complete. Inspect the 3D model generated in chimera. If you are happy that all looks correct, we will then move on to refining this volume using the particle set, to generate a high resolution volume.

6. In the project menu, start a new “3D refinement job”

7. move through the job parameters pages as before and fill in the required parameters pointing to the output from the ini3D run. This output has merged the selection of particles from the 2D clusters and the 3D orientation parameters at the particle level and so allows the refinement of the volume at the particle level to now be run. On the optional parameters page, select 0.1 for the fractional update per iteration to speed the cycles. These should take ~5 min per iteration and will take 20-30 iterations to converge.
8. You can follow the progress of the refinement by inspection of the log file and by viewing the job output (see below) where the resolution (as determined by gold-standard FSC) of the volume is reported. You will also see the sections from the volume of the most recent iteration with the “View Output” option. The initial resolution should be better than 10 A and will rapidly improve. After ~15 cycles, the resolution should be ~4 A and can be further improved by calculating a mask and then performing further refinement using that mask (see TASK NINE, post-processing). Volumes can be inspected in chimera—the xxx.pproc.mrc are filtered at the FSC resolution and are easier to inspect. If you want to perform other forms of post-processing of the volumes see the next task for a description of how to apply B-factors, masks etc.

This completes TASK SEVEN.
TASK EIGHT - Postprocess 3D volume, make mask and perform further 3D refinement for beta galactosidase

In this task we use inspect the results of our 3D refinement of beta galactosidase, apply a B factor to that density, and calculate a mask for further rounds of 3D refinement.

1. The “3D Refinement Module” has written out a set of files for each iteration which include an unfiltered volume (called recvol_state01_iterXXX.mrc where XXX is the cycle number), a volume which has been filtered to limit the resolution to the high resolution limit as determined by the FSC (recvol_state01_iterXXX_pproc.mrc) - this is the one you generally inspect and a set of binary files with information for applying the FSC and anisotropic volume filters (fsc_state01.bin, aniso_vol_filt.mrc). We are now going to modify the final volume by applying a B factor correction and calculate a mask for further 3D refinement.

2. First we need to establish a threshold for contouring the volume for drawing a mask. To do this open the final cycle pproc.mrc in chimera.

4. Contour the volume to the point that noise becomes visible.

5. Now readjust the contours until the the volume is smooth as in the lower panel.
6. You now need to note down the contour level from within the Chimera “Volume Viewer” box. Note - you will need to click in this box and scroll to the right to see the level of precision required.

7. Once you have this number you are ready to launch a “Post Processing” job within the SIMPLE GUI. Move through the Job Params page and fill in as shown below selecting the unfiltered volume from the final cycle of the 3D refinement job.

8. We will now calculate a suitable B factor to apply to the volume by calculating a Guinier Plot. Fill in the values shown below and select Estimate. The calculated B factor will be output to the screen.

9. On the next page we will select to generate a mask file using the threshold value you determined in Chimera. Make sure to turn automasking on and to fill in the molecular weight of beta galactosidase in kDa.
10. The output volume will be filtered using the fsc and anisotropic filters calculated during the 3D refinement run, masked and the B factor will be applied to give a volume with substantially more detail than the unprocessed one.

11. Now start another “3D Refinement” job using the .simple file generated at the end of your earlier job but using masking with the automask.mrc file generated by this run (see below) and selecting the neighbourhood refinement mode.

12. This will run for a small number of iterations and will generally yield at least a 3.5Å volume.
TASK NINE - Motion correction and generation of initial references for automated particle picking within the simple pipeline

In this task we use the "Motion Correction" module to automatically motion correct and dose weight twelve urease cryo-em movies, then you can perform "Manual Picking" on the motion corrected micrographs, estimate the CTF parameters, extract the particles and perform 2D clustering. We will then setup the automated pipeline to perform motion correction, CTF estimation, auto-picking and extraction in a fully automated fashion as an example of how the pipeline processing can be run during data collection.

1. Create a "Motion Correction" job

![Motion Correction Module]

2. This module will (i) correct dead/hot detector pixels (ii) motion correct the movie and weight the frames according to the level of agreement between them and (iii) perform dose weighting of the frames via the method of Grant & Grigorieff, eLife 2015. Various experimental parameters must be known to perform this - see below for suitable values for these movies.
3. Then setup how the job should run on your computer:

4. Once this job has completed, use the `unblur.out.simple` file as the input for a CTF Estimation job (as you did in TASK TWO). Using the parameters as shown below (all the optional parameters can be left at default values on this occasion):

5. Now go and perform a manual picking as you did in TASK THREE, then extract the particles using those boxes as in TASK FOUR and run a 2D clustering as in TASK FIVE making sure that in all of these you use the correct sampling distance (1.26 Angstroms/pixel), mask radius (90 pixels) and a box size of 320 pixels should suffice. As you will have only picked a small number of particles in the manual picking stage - perform the 2D clustering in just a few classes (5 or so is probably sufficient for this very spherical object to generate references for further picking).
5. We are now going to use the “Manual Picking” module to establish good parameters for using the 2D class averages as references for automatically picking these micrographs in the SIMPLE pipeline. Re-open the “Manual Picking” module (as in TASK THREE). We are going to use the “Reference Based Picking” panel and use the 2D class averages as picking references and set up the required parameters.

6. First select the 2D class averages from the final iteration of your earlier 2D job in the top box and then hit autopick. This will generate the references (which will be written as pickrefs.mrc in the directory associated with this job) and perform a picking with default parameters. Play with the parameters boxed below until you are happy with the level of picking and then make a note of the values you like. You are now ready to run the autopicking module to fully pick the micrographs.
7. Select the “Automated Picking” module

8. Now setup the autopicking pointing to the list of selected micrographs from ctfhit, the 2D clusters that you used to generate the picking reference in your test picking:
9. In the optional parameters enter the values for Deviation (\textit{ndev}) and Box Overlap (\textit{thres}) that you determined in the “Manual Picking” module.

10. Once this has run you will be able to inspect the particles picking using the “View Output” option.

This completes this task and gives you an introduction to the main functions available through the SIMPLE GUI.