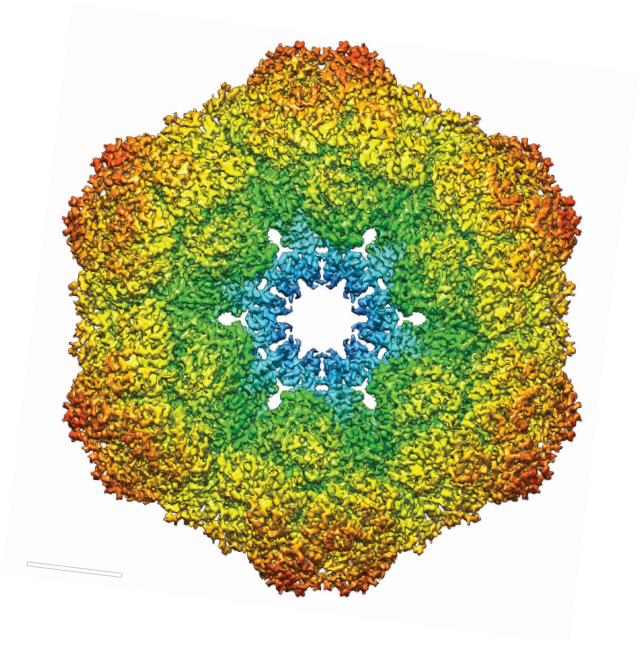
SINGLE PARTICLE CRYO-EM

HANDS-ON

Wormhemoglobin Analysis

(GISP)



Version 1-Nov-2023

www.ImageScience.de

1. Remember: Some IMAGIC Basics

This chapter is on how to work with the **IMAGIC** software.

An "**IMAGIC** image file" consists of a header file (".hed") and the image density file (".img"):

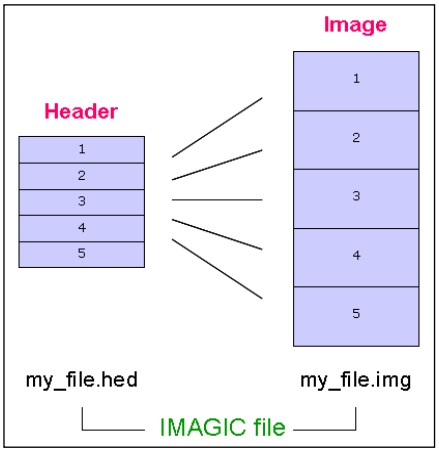


Fig. 1: IMAGIC file

The image file contains the actual image density values, while the header file contains information about the images ("meta data") as a set of records that can be accessed through different labels. For example:

IMN	image location number (1,2,3,...)
IXLP	number of lines per image
IYLP	number of pixels per line
IZLP	number of sections if input is a 3-D volume
REF	multi-reference number
CLASSNO	class number
ALPHA	Euler alpha angle
BETA	Euler beta angle
GAMMA	Euler gamma angle
etc…	

An additional PLT text file can be associated to an **IMAGIC** file to store further meta-data like:

coordinates of particles
contour of masks
image numbers
Euler angles
graphics (curves)
etc…

The PLT file may contain a maximum of five numerical values per line, separated by blanks or by commas.

A few other **IMAGIC** text (ASCII) files may be generated during processing:

CLS files are classification files containing classes and their members

LIS files contain information printed during execution of a program

LOG files contain output of programs when running as batch job (script)

DAT files contain data for various purposes

DFF (deFault Files) are used to store your last answers

The **IMAGIC** coordinate system is a right-handed system with its (1,1) origin in the top-left corner of the image. The length of the lines (number of rows/columns) is **NY** and the number of lines is **NX**:

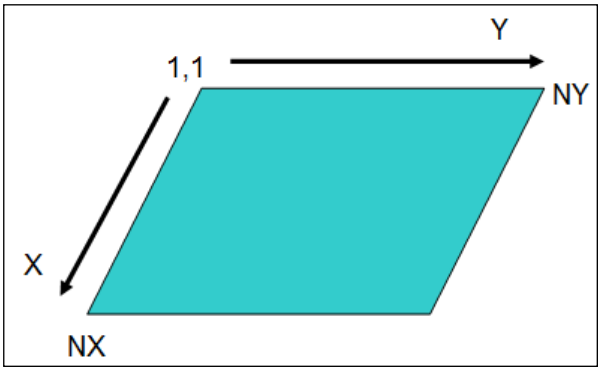


Fig. 2: IMAGIC 2-D coordinate system

The **IMAGIC** coordinates for a 3-D volume are the following:

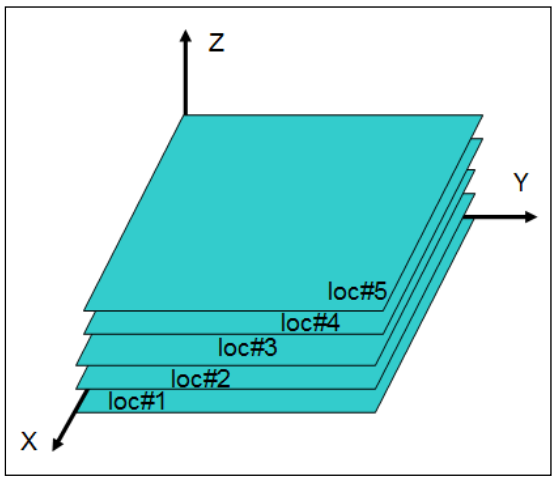


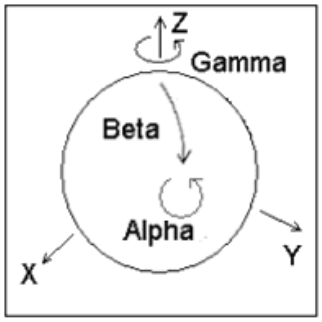
Fig. 3: IMAGIC 3-D coordinate system

Note that **Z** = **X** x **Y** as required for a right-handed co-ordinate system.

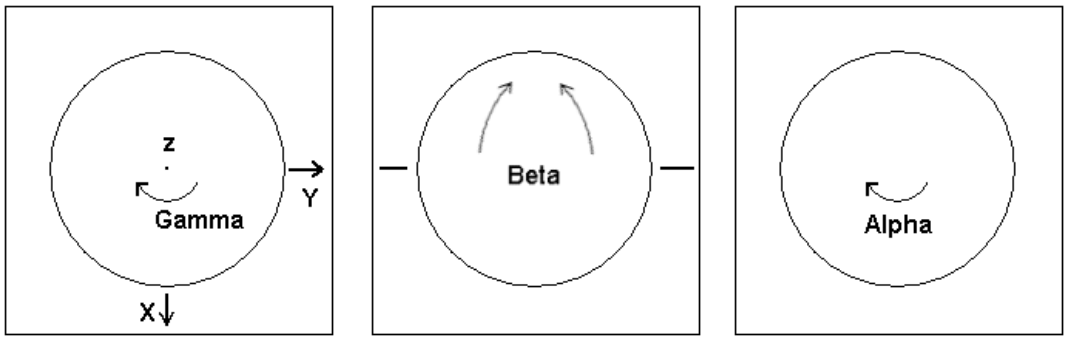
In **IMAGIC** 3-D orientations are defined by three Euler angles Alpha, Beta and Gamma.

From the perspective of an external viewer (like every IMAGIC image used/created in commands ANGULAR-RECONSTITUTION, THREED-SURFACE, THREED-FORWARD, etc.) the Euler angles are defined as follows:

The first rotation is a rotation around the Z-axis by GAMMA, followed by a rotation BETA around the new Y-axis and a rotation ALPHA around the new Z-axis.



But normally a user does not think in this way but tries to imagine how the particle would look like "in his hands":



a) Look at the particle along the Z-axis ("north pole")
b) Rotate the particle clockwise by Gamma
c) Rotate the particle into the plane clockwise by Beta
d) Rotate the particle clockwise by Alpha

PLEASE NOTE:

The important angles to define a 3-D orientation are Beta and Gamma. Alpha is only the final in-plane rotation.

The parameters and file names in the **GISP** pages will often have default values which appear in the text/value boxes.

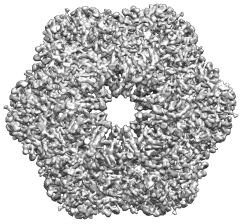
GISP/IMAGIC remembers the last values you have entered for a specific command. These values (stored in the DFF files) become the default values the next time the command is started in that working directory. In general, if you do not know how to answer a question, the default values serve as an intelligent first guess.

When following the **GISP** single pafrtilce workflow all output file names will be created automatically. When running additional **IMAGIC** commands using the "Commands" button you are free to choose whatever names you wish. However, bear in mind you will have to remember what you've chosen for the next commands.

MPI refers to parallel processing. If your notebook computer (or any other type of computer) has multiple cores commands which are using parallel processing will ask you if you want to run the command in parallel or not. In the beginning always first run on a single core (non-parallel mode) which will give you more feedback (answer NO). Later when using many images your answer may be YES. Note that the number of processors to be used should be at least the number of nodes PLUS 1:

Throughout this hands-on, words that appear in GREEN refer to **IMAGIC** commands. Words in red are required/suggested input values. Suggested file names are in blue.

YOUR NOTES:



2. The Worm Hemoglobin Data Set

Hemoglobin (Hb) is the iron-containing oxygen-transport metalloprotein present in the red blood cells of vertebrates. In earth worms (Lumbricus terrestris), the hemoglobin (or "haemoglobin" and also known as "erythrocruorin") is extra-cellular, freely dissolved in the blood. It is a 3.6 MDa dodecameric assembly with D6 (622) point-group symmetry consisting of 12 protomers or 1/12th subunit.

Data collection:

Micrographs were collected as 7-frame movies on an FEI Titan KRIOS with a Cs corrector and a X-FEG operated at 300 kV.

Spherical aberration:	0.02 mm
Focal distance:	3.4 mm
Objective aperture:	120 μm
Pixel Size:	1.11 Å (coarse 2: 2.22 Å; coarse 4: 4.44 Å)
Size of a single micrograph:	4096 x 4096

YOUR NOTES:

On the hands-on directory, you will find two main directories:

- On the data directory **whgb_data** you will find a number of files containing micrographs and intermediate results which we are going to use during the subsequent image analysis.
- Additional data is stored in the **whgb_data_additional** directory, which contains the following sub-directories:

  Directory **01_whgb_micrographs_mrc** containing:

  15 micrographs aligned movie sums in **MRC** format

  Directory **02_whgb_micrographs_imagic** containing:

  70 raw micrographs (10 movies of 7 frames each) in a single IMAGIC file with 70 locations (full 4096x4096 pixels)

  Directory **03_whgb_micrographs_preprocessing** containing:

  3500 pre-processed (including camera correction; anisotropic magnification correction) and 4-times coarsened ("C4") micrographs (500 movies)

  Directory **04_whgb_micrographs_moviealigned** containing:

  500 camera corrected, anisotropic magnification corrected, aligned 4-times coarsened micrograph movie-sums

  Directory **05_whgb_micrographs_ctf_correction** containing:

  Files associated with the automatic CTF correction

  Directory **06_whgb_particle_picking** containing:

  Files associated with particle picking

  Directory **07_whgb_particle_classification** containing:

  Files associated with particle picking

  Directory **08_whgb_first_3d_reconstruction** containing:

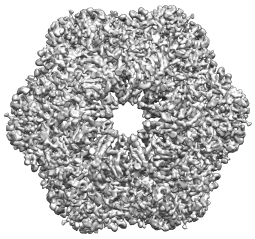
  Files associated with the first 3-D reconstruction

  Directory **09_whgb_movie_spectra** containing:

  Files associated with movie spectra

Copy the files in **whgb_data** to your own computer:

- Create a subdirectory **whgb_data** in your working directory.
- Copy all files in the data directory **whgb_data** of the hands-on directory into your own **whgb_data** folder.



3. Image Analysis: Short Overview

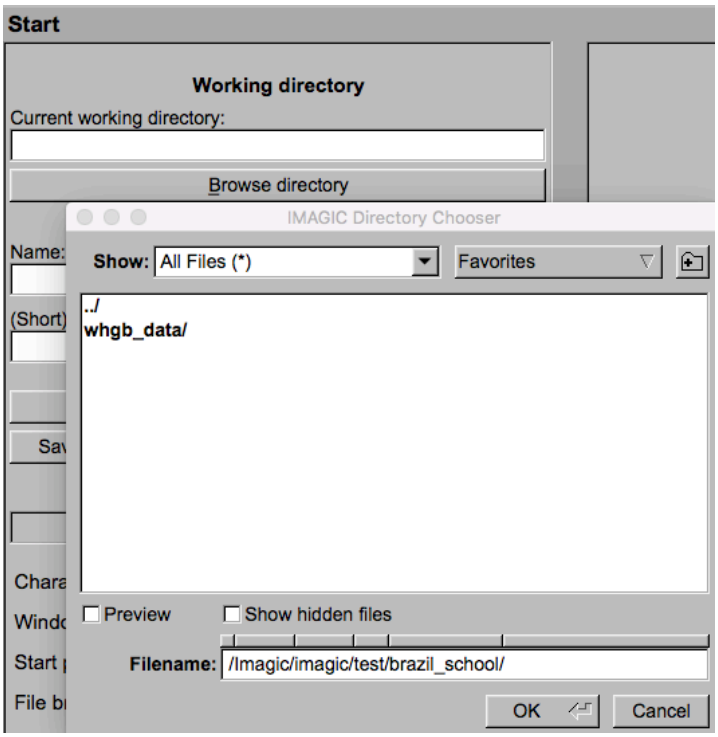
The hands-on processing of the worm hemoglobin dataset:

- Import Micrographs to IMAGIC
- Camera Correction
- Movie Alignment
- Playing with the CTF
- CTF correct the Micrographs
- Particle Picking
- Extract/box and pre-treat the single particle Images
- MSA-Classification
- Angular Reconstitution and 3-D Reconstruction Start-Up
- 3-D Reconstruction Refinement
- Iterative Refinements
- Multi-Reference Alignment (“old fashioned”)
- Fourier Shell Correlation
- More…

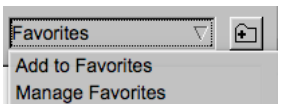
4. Start

Before doing any calculations, you have to define some project parameters:

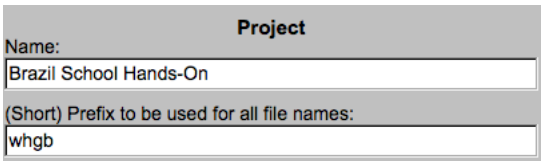
1. Specify your working directory. You can type the name into the text box or use the "Browse directory" button.



NOTE: You can store your directory in "Favorites".



2. Type in the name of your project and, if wanted, some information. Note that is only for your information.

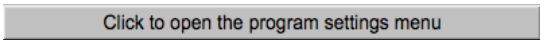


3. Also specify a SHORT prefix. Output file names on the **GISP** workflow pages will be created automatically using this prefix. When using "Commands" the output file names have to be specified.

4. Save the information given by clicking the 

button.

5. You can also change some **GISP** program settings:



May be, your computer window/monitor is too small and you want to reduce the **GISP** windows:

In this case you normally also have to adjust the font size:

Save the settings and note that the **GISP** will re-start.

6. Click the 

button to start working.

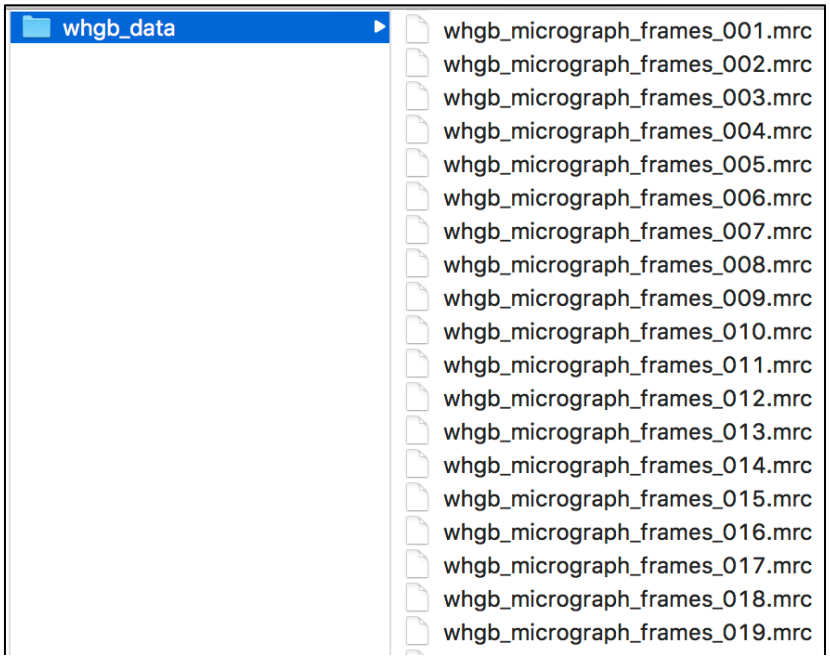
YOUR NOTES:

5. Import Micrographs

The raw data will first be organized as a "stack" of images, which will be treated as your initial, raw data. To achieve this, the micrographs will be appended together into one single file.

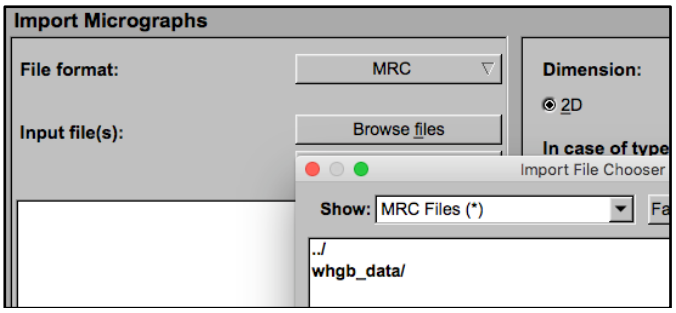
Before you begin appending the micrographs you should look at the original micrographs to get an idea of how to convert your micrographs to an **IMAGIC** stack file.

On the hands-on directory **whgb_data** you will find 70 MRC-formatted images (whgb_micrograph_frames_001.mrc, whgb_micrograph_frames_002.mrc ... whgb_micrograph_frames_070.mrc). The micrographs are movie frames of size 1024x1024 pixels.

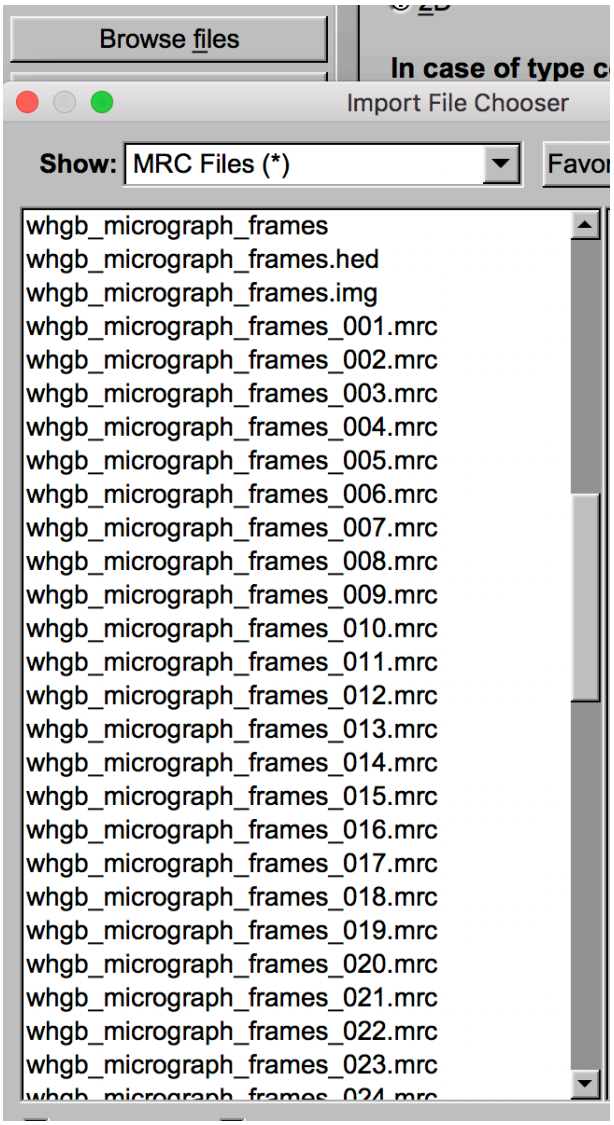


7. To be able to process the input micrographs in **IMAGIC** you need to convert them into **IMAGIC** format. Open the page **Import Micrographs**.
8. Choose MRC as format.

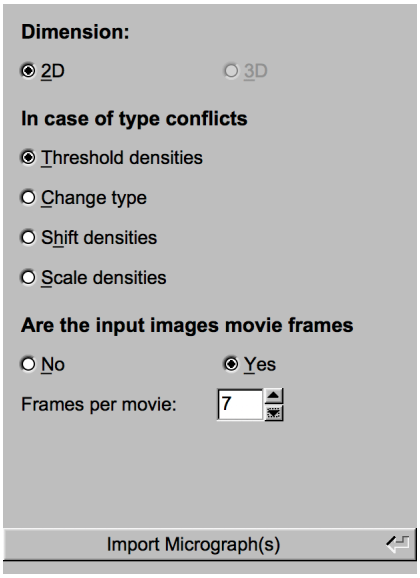
9. Click the Browse file button and choose directory **whgb_data**



and select the MRC movie frame files whgb_micrograph_frames_001.mrc to whgb_micrograph_frames_070.mrc.



10. There can be a type conflict between **MRC** and **IMAGIC** (move the mouse over the text in the program window to get help). Usually we would choose Change type but the files are created in EPU so that Threshold densities is suggested to store disk space.
11. Also specify that the input **MRC** files are movies with 7 frames per movie.



12. Click the "Import Micrograph(s)" button to create the **IMAGIC** file whgb_micrographs containing the stack of all movie frames.
13. Check the micrograph movie frames in the display on the right-hand side.

NOTE:

Each image is a movie frame. So, it is difficult to detect particles in the displayed images. In Contrast use the option Invert so that the particles look white.

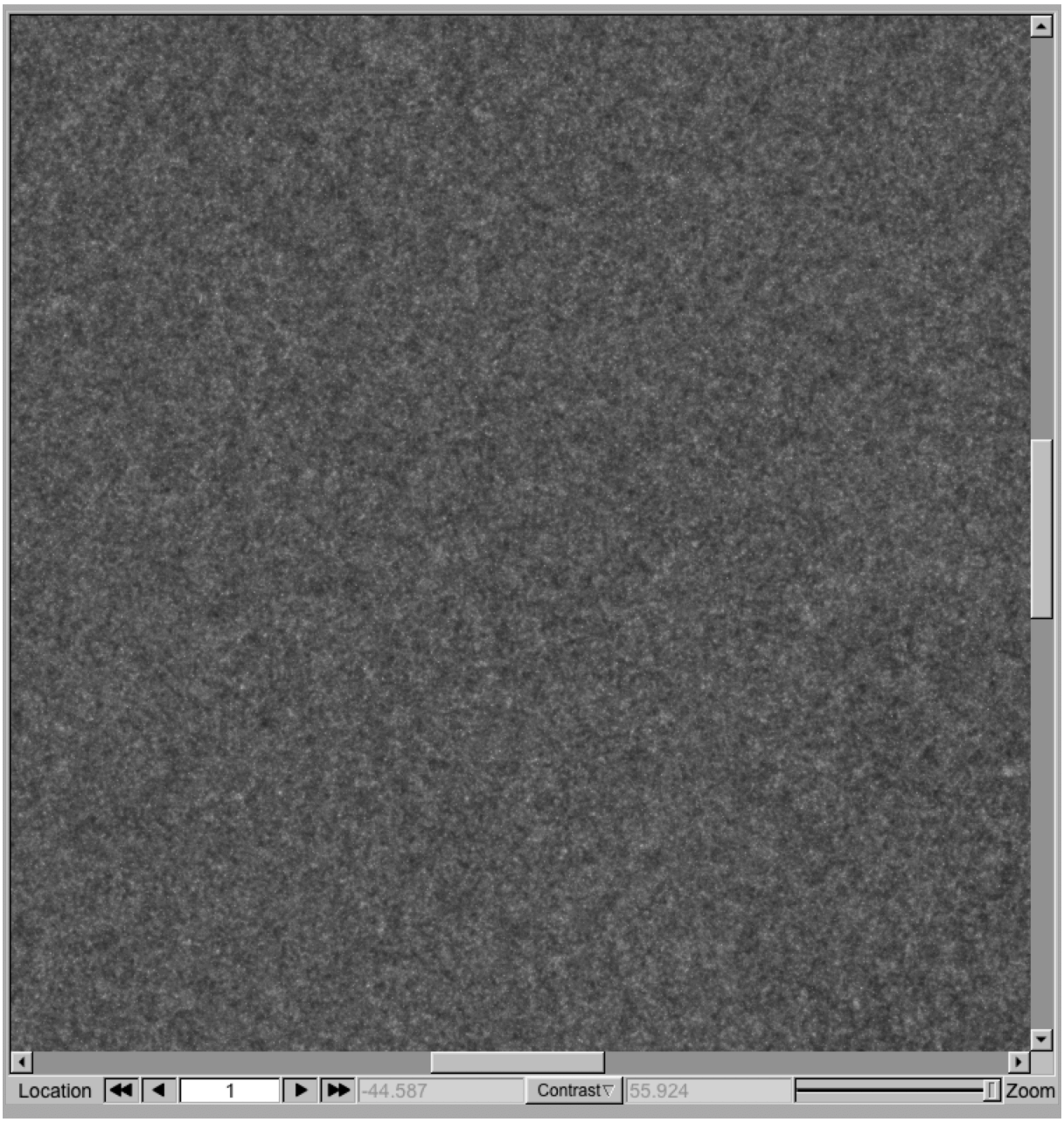


Fig. 5: One of the displayed movie frames

NOTE:

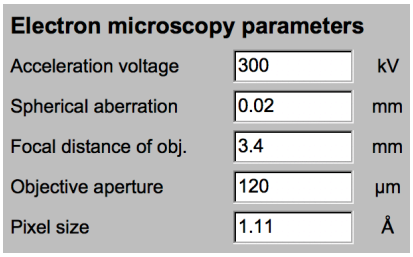
All **IMAGIC** pages provide context related detailed help. Move the cursor to any item wanted and wait until the help window pops up.

14. Click the Next button to continue.

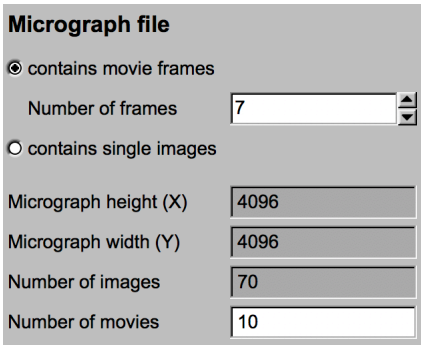
6. The Project Page

On this page you have to give some project information before the image analysis can start.

1. Usually, the imported micrographs do not contain the information how they were imaged. So, you have to give this information here:



2. You also have to specify if the input micrographs are movie frames or not:



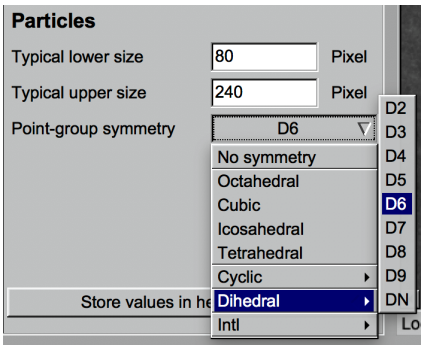
3. Some procedures (like boxing particles, filter etc.) need information about the size of the particles.

   If you can see particles in the displayed micrograph movie sums you can give this information here. Else simply store the zero values.

   To specify a size double click into the related value box. Now move the cursor into the displayed image and move the cursor to the centre of a particle. Press the left mouse button und keep it pressed until the red circle has the wanted size. The diameter of the circle is the size specified.

   Of course, you can also specify the sizes by typing the values.

4. In addition, we need information about the symmetry of the particles, (which is D6 or 622 for worm hemoglobin):



5. Do not forget to store all given parameters:

REMEMBER:

All **IMAGIC** pages provide context related detailed help. Move the cursor to any item wanted and wait until the help windows pops up.

This is no more mentioned in the following.

6. Click the  button to continue.

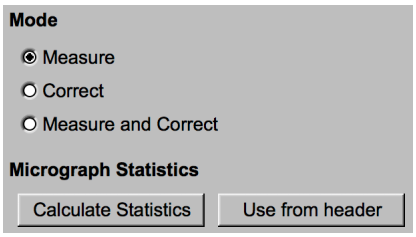
7. Camera Correction

The *a posteriori* camera correction procedure allows improving the quality of the micrographs images based on the statistics of the full dataset. In many cases, it also helps to improve the movie alignments. We perform the camera correction on each movie frame image using the total average image and the corresponding sigma image of all frames in the dataset.

7.1. Measure the Camera Statistics

The stack of input micrographs only contains 70 movies. This number of micrographs is too small to get a reasonable statistic of the camera behaviour.

1. Anyway, you can use the Measure option to get a feeling how one can calculate the camera statistics:

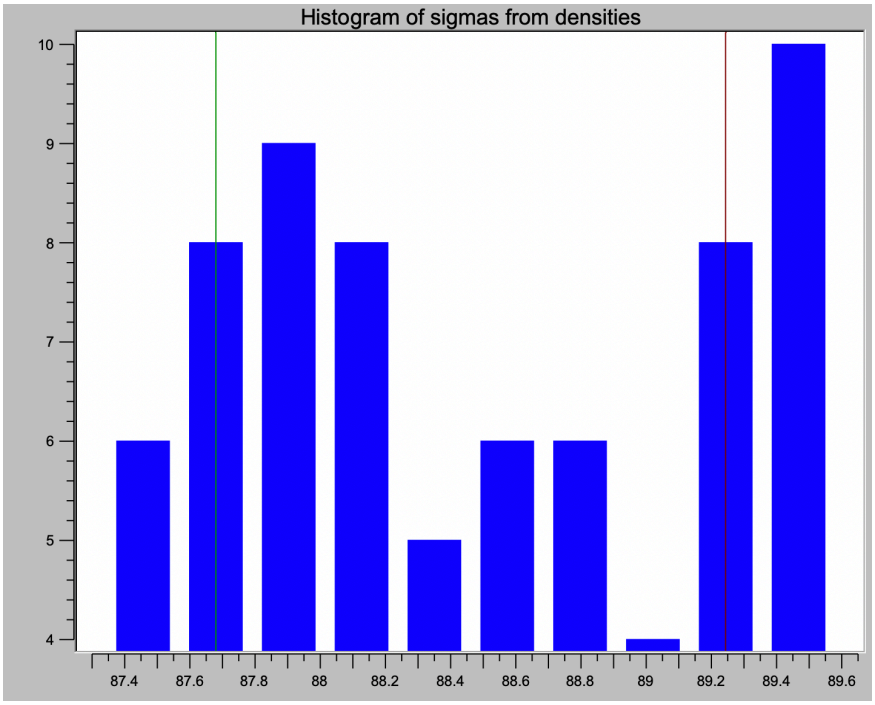


2. First, calculate the (overall image) statistics of all input micrographs (whgb_micrographs).

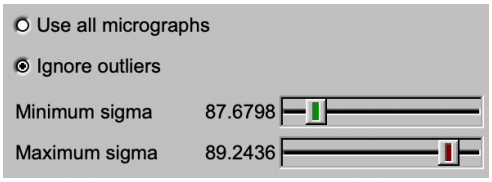
   Click the 

button.

3. The histogram of sigma values is shown on the right-hand side. Usually this histogram has a Gaussian like shape (which is not the case for our too small test data-set).



4. If the shape looks correct you can ignore outlier micrographs while calculating the camera statistics. Use the sliders define a minimum and a maximum value for sigma. The vertical lines in the histogram will help you to check the chosen values.



NOTE:

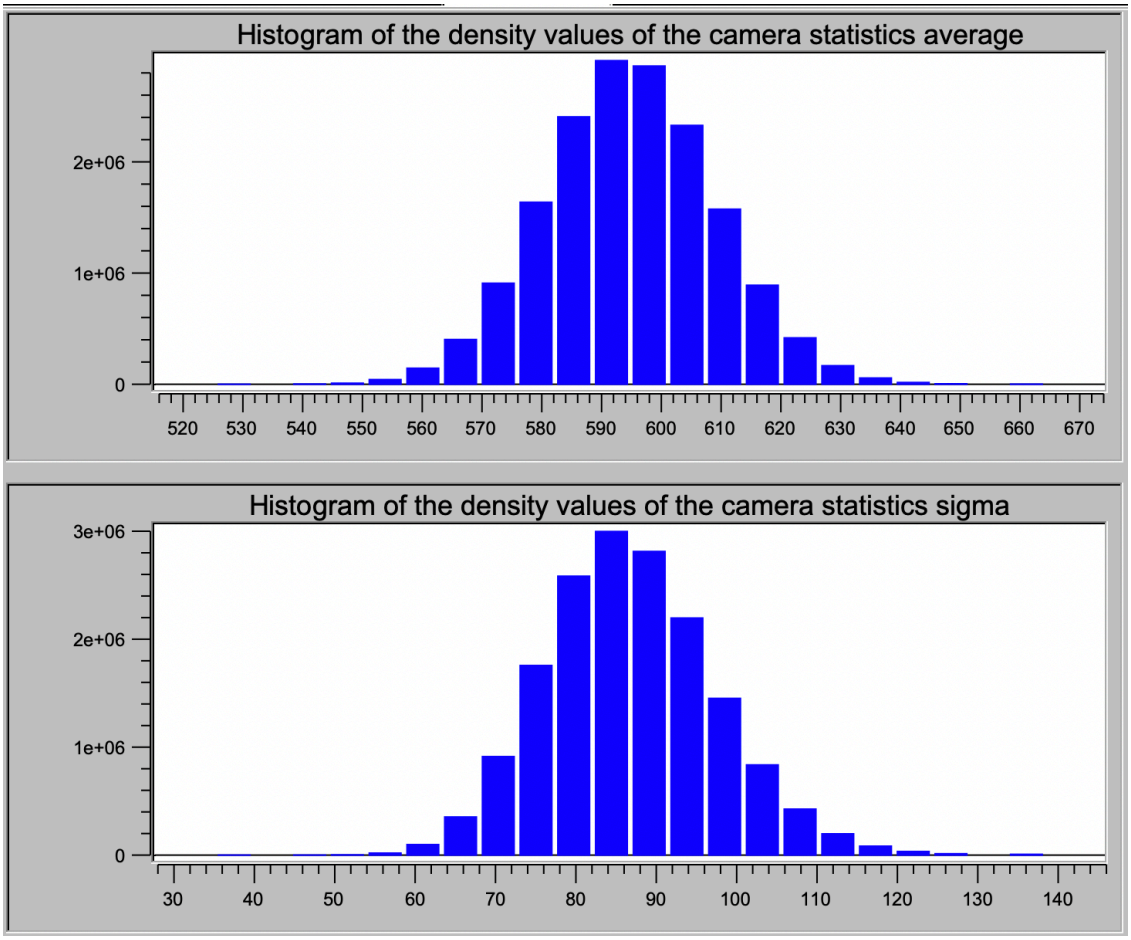
It is always important to exclude micrograph images containing too high contrast and features like grid bars, junk, ice crystals, etc.

Those micrograph images often correspond to the extreme values of the histogram of sigma and can be excluded automatically based on the histograms. Only micrographs corresponding to the histogram part between the red lines will be used for the camera correction.

5. The camera statistics of the selected micrographs will be calculated after clicking the "Measure" button.

6. Ignore the warning: 

7. The camera statistics is shown in two histograms (average and sigma).



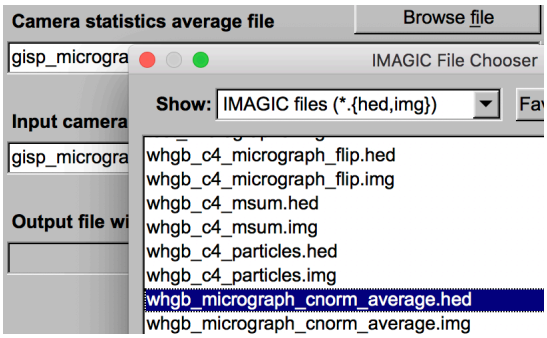
The histograms usually have a Gaussian like shape. In case of camera errors (blind or dark pixels, for example) these can easily be seen as vertical lines.

8. Also have a look at the displayed average and sigma images and check for camera errors. Always use a Zoom factor of 1!

7.2. *A posteriori* Camera Correction

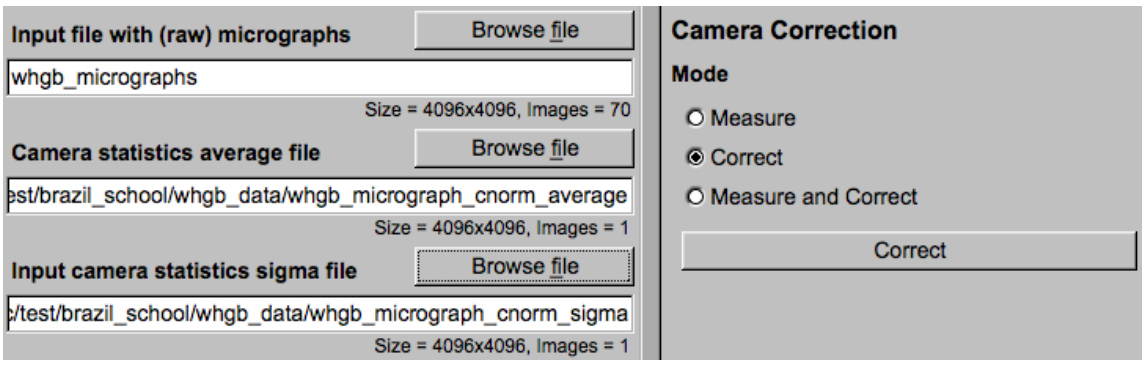
We have already done a camera measure with a sufficant large number of micrographs. You can find the results (whgb_micrograph_cnorm_average and whgb_micrograph_cnorm_sigma) on the hands-on directory **whgb_data**.

1. As usual, click the "Browse" button and "IMAGIC files" and go to directory **whgb_data**. Choose the camera statistics average file.



2. In the same way, choose the camera statistics sigma file.

3. Now you can camera correct the input micrographs (whgb_micrographs) using these "correct" camera statistics files.



4. Check the camera corrected micrographs (whgb_micrographs_cnorm) displayed in one of the tabs on the right-hand side.

5. Also have a look at the "correct" camera statistics files (average and sigma)

NOTE:

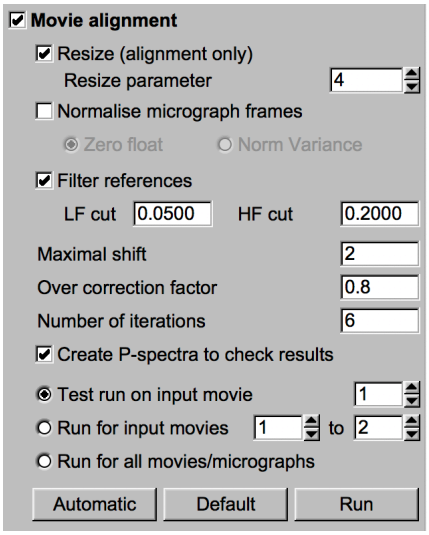
Always use a zoom factor of 1 when checking camera statistics (Click Gallery: Off and Zoom = 1.0).

8. Movie Alignment

Movie alignment is performed onto the camera corrected micrographs (whgb_micrographs_cnorm).

Movie alignment are three main calculations:

- Resize the micrographs to be used in movie alignment to speed-up the calculations
- Low-pass filter the micrographs
- Movie alignment



NOTE:
Play around with the parameters running movie alignment on a single or a small range of micrographs.

1. As usual, click the 

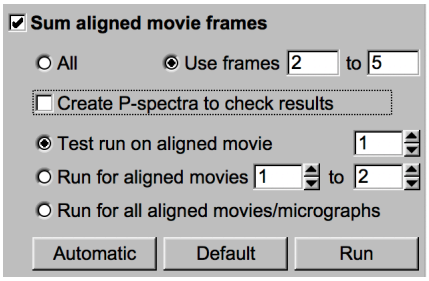
button to start the calculations.

2. Compare the input and the aligned movie frames.

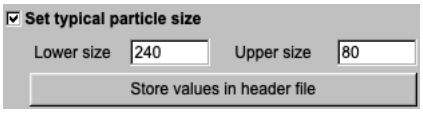
3. Also compare the related P-spectra.

4. If the movie alignment of the test micrographs is okay run the movie alignment for all micrographs.

If you did not yet specify the size of the particles you can do it here using the movie sums of the aligned micrographs



and

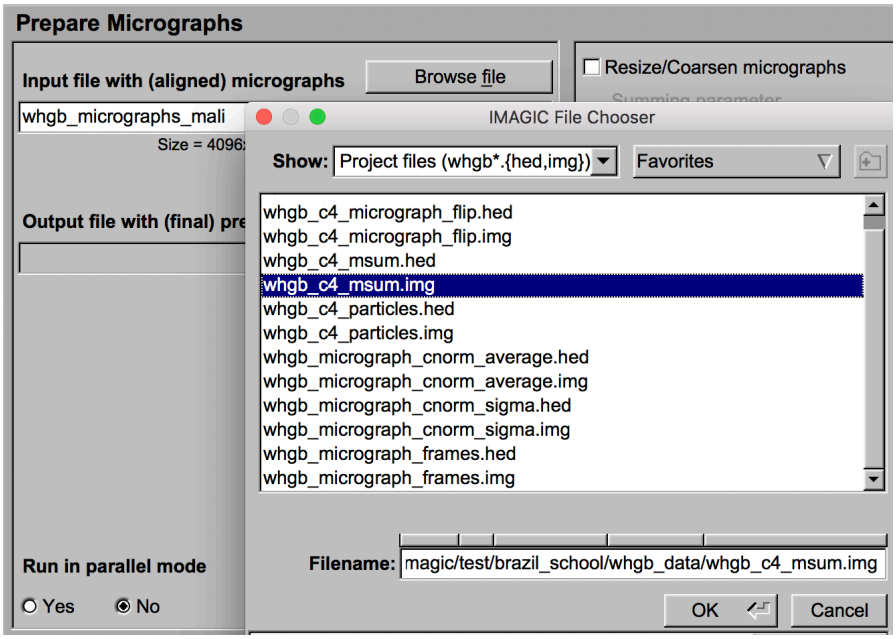


Please refer to chapter 7 (Project Page).

YOUR NOTES:

9. Prepare Micrographs

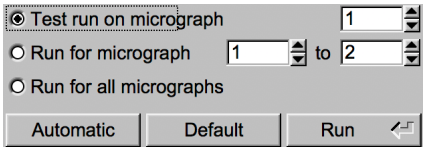
Now continue the practical with the 500 aligned and movie summed micrographs provided in the hands-on directory **whgb_data**. The **IMAGIC** file name is whgb_c4_msum.



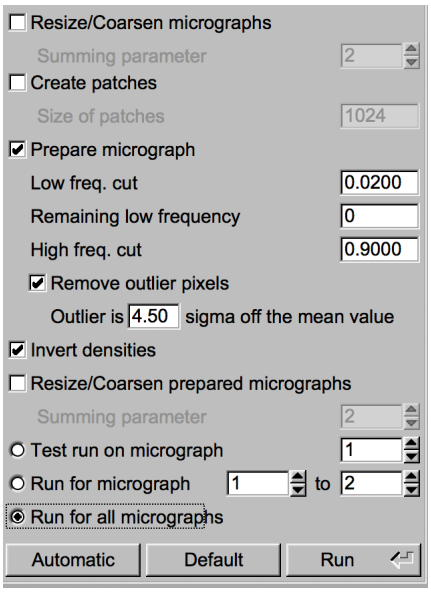
1. Suppress extreme low frequencies by applying a band-pass filter. One can also resize and/or normalize the micrograph images as well as invert the image densities (particles must be WHITE in **IMAGIC**).

   NOTE:
   The micrographs in the data directory are already binned so you don't have to resize them anymore.

2. First choose a single micrograph and play around with various band-pass filter parameters to see the effect of a band-pass filter applied to the micrographs.



3. Finally use the following parameter for all micrograph images.



4. As usual, continue with the Next button.

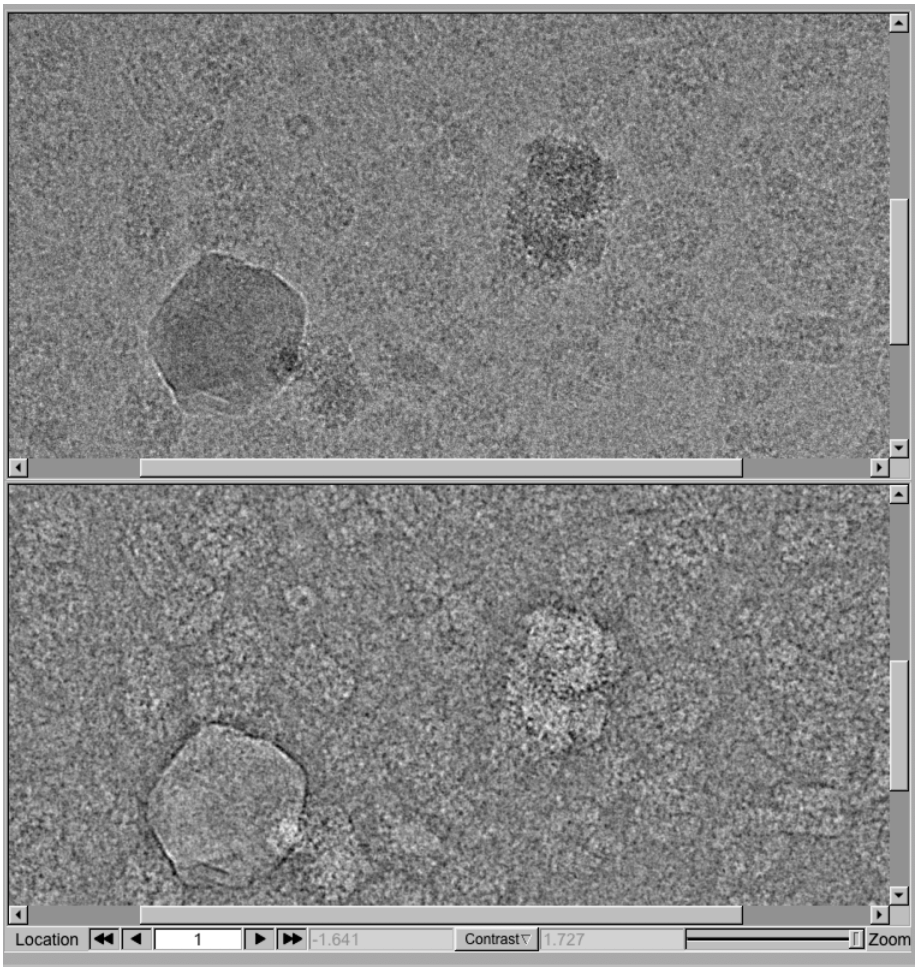
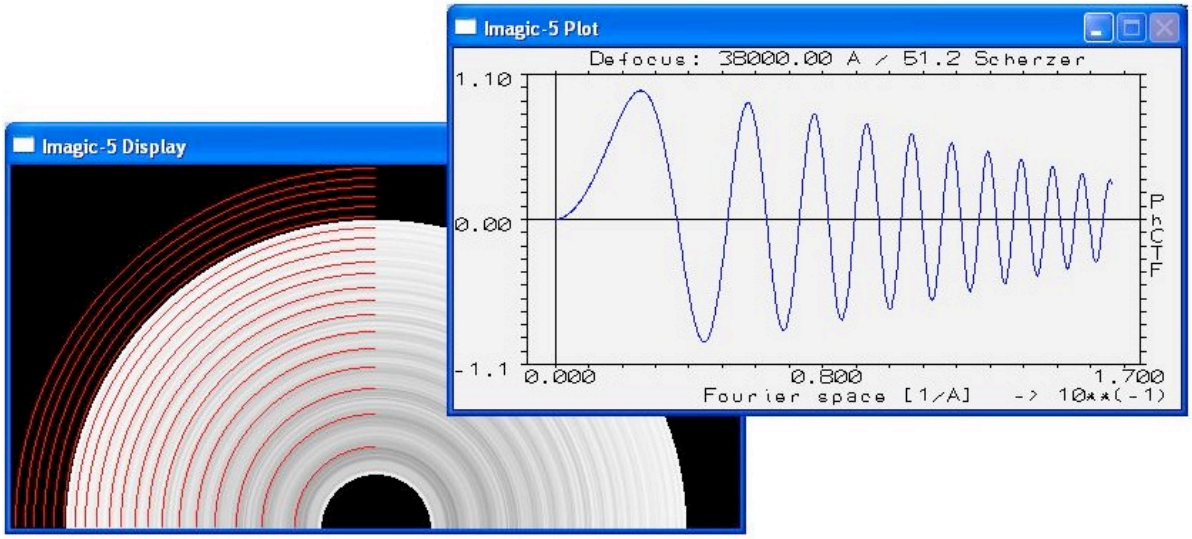


Fig. 8: A micrograph image before and after the preparation

10. Contrast Transfer Function (CTF)



As discussed in the lectures an electron microscope unfortunately does not image all frequencies equally. This exercise is meant to play around with the command TRANSFER, which is an interactive program to calculate the (rotationally symmetric) CTF according to chosen microscope parameters.

10.1. Start the IMAGIC Command TRANSFER

TRANSFER is a terminal program in a line-by-line modus. To run it you have to do the following:

- Click the Command button.
- Choose or type in the command TRANSFER-FUNCTION-1D and press "Enter".

A terminal/command window will open and the TRANSFER command will start.

Note that using this old-style (non-GUI) **IMAGIC** commands means that you have to type in all answers requested and options wanted.

10.2. Playing around with EM Parameters and their Influence on the CTF

1. Command TRANSFER is an interactive command with many parameters. You can use the keywords written in capitals to change the related parameter. TRANSFER allows you to change the various parameter settings until you type CR/ENTER, which means NO CHANGES, i.e. go ahead and display the CTF curve:

   TRANSFER displays the settings:

```
Current TRANSFER settings:
==========================

Desired TRANSFER function           : Phase CTF
Acceleration VOLTAGE                : 200          kV
Relativistic WAVE length in Angstrom: 0.025045     Angstrom
CHROMATIC aberration                : None
SPHERICAL aberration constant       : 2.2          mm
FOCAL length of objective           : 1.6          mm
APERTURE of objective lens          : 50.0         micro m
Coherent illumination SOURCE/ANGLE  : 0.0
DEFOCUS value                       : 890.7438965  Angstrom
GENERAL defocus values              : 1.2          Scherzer
OBJECT size defocus envelope        : Off
LENGTH of transfer function         : 640          pixel
PIXEL size in curve                 : 1.0          Angstrom
-----------------------------------------------------------
MODE of operation                   : Calculation of CTF
Output DESTINATION for plot(s) is   : IMAGIC plot

Change options (VOLT,DESTIN.,MODE, etc. ...) [NO] :
```

2. First give CR/ENTER to display the contrast transfer function (CTF) related to the given parameters in the TRANSFER settings.

3. Then play around with different pixel sizes:

```
Change options (VOLT,DESTIN.,MODE, etc. ...) [NO] : pixel

Pixel size measured in Angstrom                    : 2.22
```

4. Always give CR/ENTER to display the CTF curve.

5. Next change the DEFOCUS value to 10000 Angstrom:

```
Change options (VOLT,DESTIN.,MODE, etc. ...) [NO] : defocus
Defocus value                                    : 10000
```

6. Now give CR/ENTER to display the CTF curve.

7. Now you can play around with other defocus values (200, 500, 3000, 30000...) and notice their influence on the CTF.

NOTE:

In Scherzer focus (GENERAL defocus value = 1.0 Scherzer) you have good image contrast over a large range of frequencies but, unfortunately, you have very little image contrast in the low frequencies and, as a result, you cannot recognize your particles. When using large defocus values, lower frequencies are transferred better, improving the visibility of the particles. But, unfortunately, you now get more frequencies, which are not imaged at all (the "zeroes") and even worse, some frequencies are imaged with reversed contrast.

8. Also play around with other parameters (VOLTAGE etc.) and examine the related CTF curves.

9. For a low VOLTAGE parameter also define a CHROMATIC aberration. Examine the related CTF curves.

NOTE:

A large amount of chromatic aberration creates an envelope function, which is imposed onto the CTF so that the very high frequencies are not transferred any more. Even CTF correction cannot restore these higher frequencies.

10.3. Interactive CTF Correction

Before using the automatic CTF estimation/correction procedures it is a good idea to interactively try to correct the CTF for a few micrographs.

Copy the files test_micrographs.hed and test_micrographs.img that you can find on hands-on directory **whgb_data** to your working directory. The files contain two micrographs (coarsened by a factor of 2 - the pixel size is 2.22 Å).

Use command TRANSFER.

1. First specify the important EM parameters. In your case:

```
Change options                          : voltage
Acceleration voltage in kV              : 300
...
Change options                          : spherical
Spherical aberration in mm              : 0.02
...
Change options                          : focal
Focal distance                          : 3.4
...
Change options                          : pixel
Pixel size measured in Angstroms        : 2.22
...
Change options                          : aperture
Aperture of objective                   : 120
...
```

2. To estimate the CTF call option FIND_CTF:

```
Change options                    : mode
Dimension of the data set         : 2d
Choose mode of operation          : find
Input file, image loc#s           : test_micrographs,one
                                          location number
Default filter parameters         : yes
```

TRANSFER displays: the CTF curve;

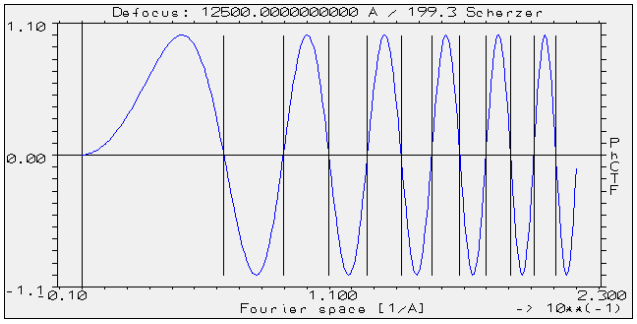


Fig. 6a: CTF plot in TRANSFER

the profile of the rotational power spectrum (RPS);

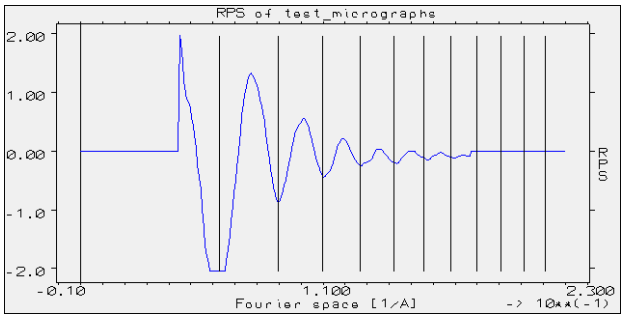


Fig. 6: RPS plot in TRANSFER

as well as the rotational power image.

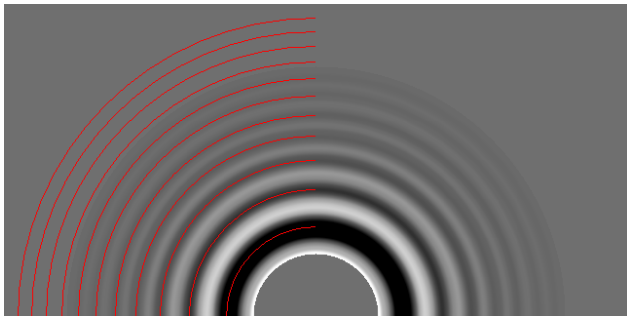


Fig. 7: RPS image in TRANSFER

In the CTF curve vertical lines mark the zeroes. The positions of these zeroes are also shown in the rotational power spectrum profile (vertical lines) and the rotational power spectrum image (red lines).

3. Play around with various defocus values until the zeroes in the CTF curve (lower curve) and the zeroes/Thon rings in the micrograph images (images above) are the same.

Defocus values found:

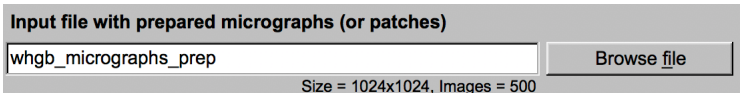
Give a "*" to exit the **IMAGIC** command and close the terminal/command window.

Knowing how one can correct for the CTF you will now use the (automatic) CTF correction in the **GISP** program. Go to the "**Prepare CTF Correction**" page.

YOUR NOTES:

11. Prepare CTF Correction

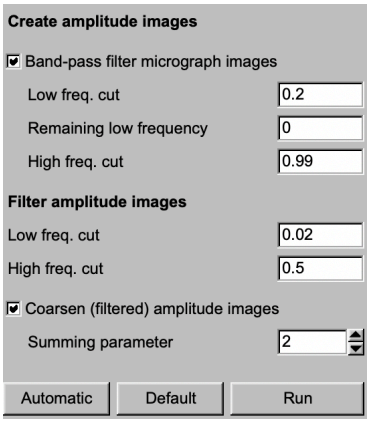
Input are the prepared micrograph images (whgb_micrographs_prep).



11.1. Calculate and Check Amplitude Images

1. First, the amplitude images (whgb_ampl) of the micrographs (whgb_micrographs_prep) are calculated.

   Before the amplitudes are calculated the micrographs will be masked and once more band-pass filtered. These are the first filter parameters, which you can specify. The amplitude images itself will also be masked and band-pass filtered (especially the background has to be removed by reducing the low frequencies). Having applied this filter the Thon rings should be better visible. Output will be the pre-treated amplitudes (whgb_ampl).



2. Now it is necessary to check if the filter parameters were chosen correctly. Therefore, all pre-treated amplitudes (whgb_ampl) will be averaged and displayed on the right-hand side. The red lines indicate the chosen MSA mask radii (see chapters 11.2 / 11.3).

Also, a profile along the central line is shown.

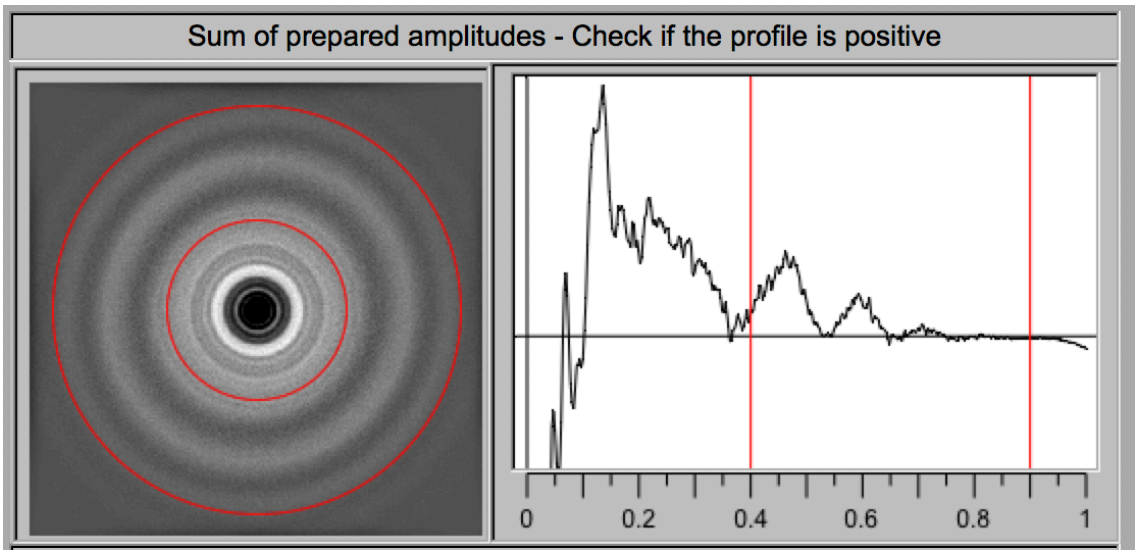


Fig. 9a: Averaged prepared amplitudes and profile along a central line (half part starting in the centre)

The second display row shows the created amplitude images. If the filter parameters are chosen correctly you should see Thon rings.

3. Play around with the filter parameters and see how they influence the amplitude images.

   If the CTF curve does not converge to zero, the low frequencies are not yet reduced enough and the filter parameters should be enhanced:

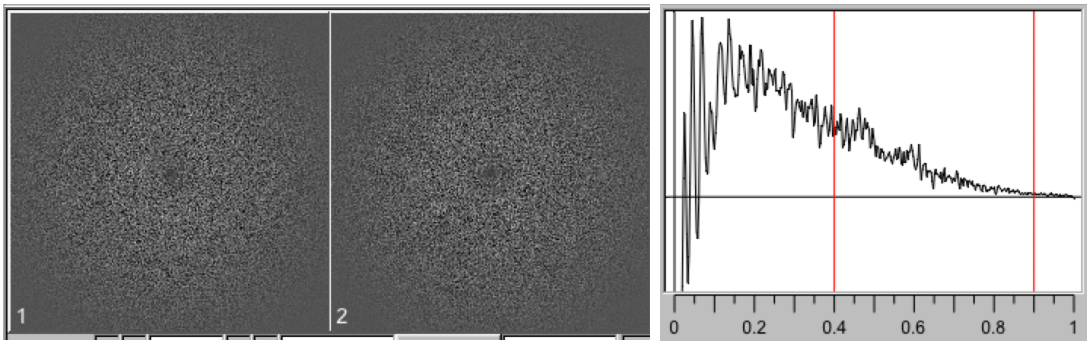


Fig. 9b: Amplitude images and related profile - Low frequency cut = 0.1

If the CTF curve approaches zero for high frequencies the band-pass parameters were chosen correctly:

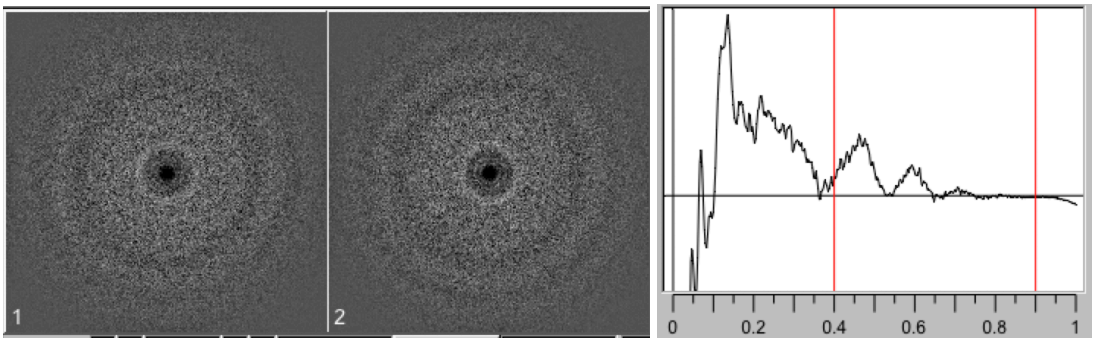
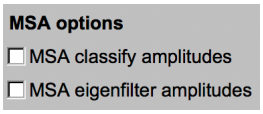


Fig. 9c: Amplitude images and related profile - Low frequency cut = 0.02

11.2. Options on how to estimate the CTF

There are three options on how to estimate the CTF:

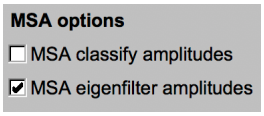
- If the Thon rings are clearly visible in each of the amplitude images simply use these prepared images.

  In this case you click "Next" and continue with the page "**CTF Correction**". But this is not suggested here.

  NOTE:
  For a huge stack of input micrographs search this option can be very time consuming

- MSA filter the amplitude images: The amplitude images are treated by MSA. Only the MSA eigenimages clearly showing Thon rings are used to MSA eigenfilter the amplitude images. These eigenfiltered amplitude images are expected to show Thon rings.

  NOTE:
  For a huge stack of input micrographs this option can also be very time consuming

- MSA classify amplitude images: The amplitude images are treated by MSA and only the MSA eigenimages clearly showing Thon rings are used to MSA classify the amplitude images. The related class averages should clearly show Ton rings and will be used to estimate the CTF.

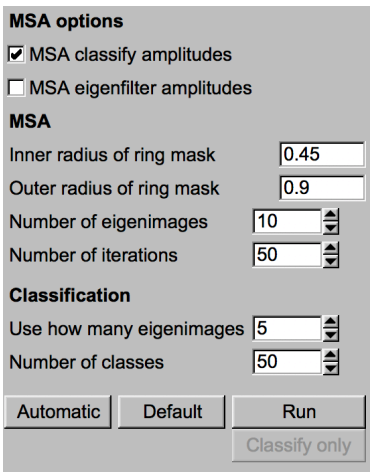
  Because the number of class averages is smaller than the number of amplitude images the CTF search using this option is much faster than the two other options described above.

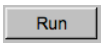
11.3. Estimate CTF using MSA and Classification

In contrast to the individual amplitude images the class averages will show the Thon rings much better which are needed to find the defocus values.

1. Certain areas are not of interest and should not be taken into consideration for MSA and classification. Check if the ring mask correctly masks out the unwanted inner and outer parts. Use the red rings in the amplitude average image and the red horizontal line in the related profile for help (see chapter 11.1).

   Start with the automatically given MSA and classification parameters. You can later adjust the parameters when checking the first results (eigenimages, class averages etc.).



   As usual click  to start MSA and classification.

NOTE:

MSA-Classification is a CPU intensive command when using large data sets. Depending on the size of the data it may be a good idea in "Real Science" to run it in "Parallel Mode".

But note that in "Parallel Mode" usually the program print-out is less detailed than in single processor mode.

2. Check the MSA eigenimages and find out the range of MSA eigenimages showing Thon rings.

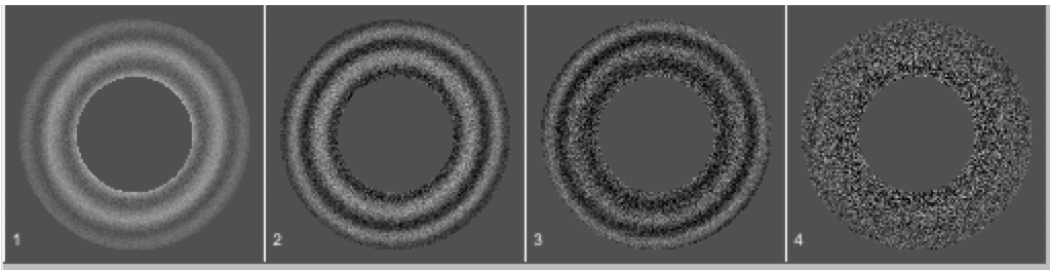
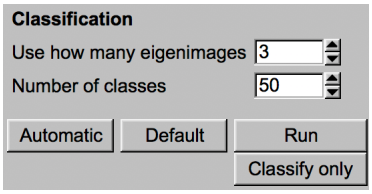


Fig. 11: Some eigenimages of the MSA treated micrograph amplitude images

3. For classification only use the eigenimages showing Thon rings.

   Note:
   You don't have to re-calculate the full MSA and classification ("Run") when only changing the number of eigenimages to be used for classification or the number of classes. You can click the "Classify only" button.



4. The finally created class averages of the amplitude images should clearly show Thon rings.

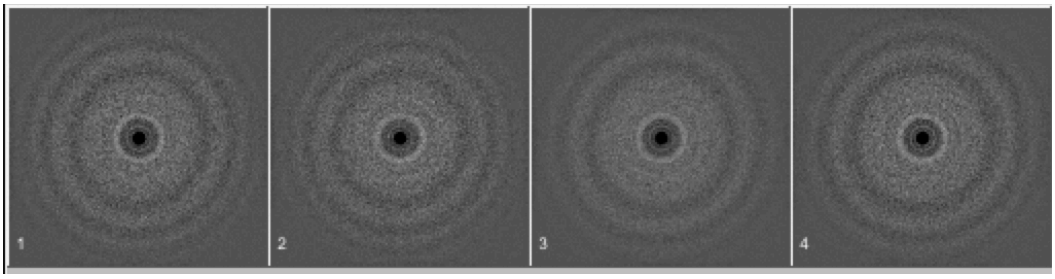


Fig. 12: Some MSA class averages of the micrograph amplitude images

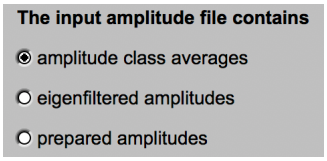
5. As usual, click the 

button to continue.

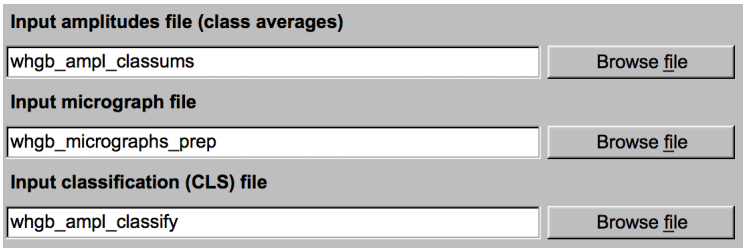
12. CTF Correction

There are three options depending of the type of input amplitudes images to be used as input (as already described in chapter 11.1).

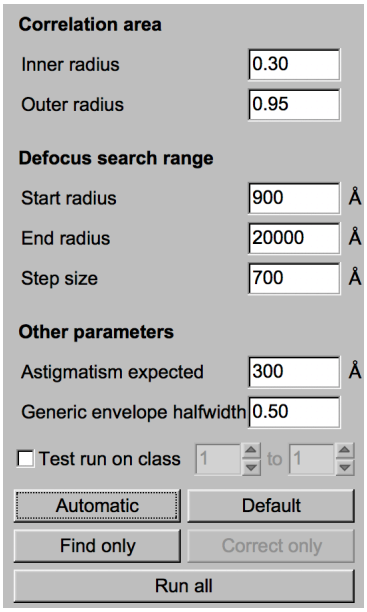
1. Since you created MSA class averages of the amplitudes images you should choose the (suggested) option "The input amplitudes file contains amplitude class averages".



2. Input files are the prepared micrograph files (whgb_micrograph_prep) and the related MSA class averages (whgb_ampl_classums) – both are suggested automatically. CTF correction also needs the related classification results file (whgb_ampl_classify) - also suggested automatically.



3. First adjust the CTF search parameters



Important parameters are the following:

- Inner and outer correlation radius: A normalized cross correlation is used to compare the filtered experimental amplitude image to the theoretical CTFs. The centre and periphery of the amplitude image do not contain rings and are not important in the estimation. Therefore, the cross-correlation is only computed over a ring area specified by two radii. You can play with these parameters to obtain the best estimation - or simply try the suggested values.
- Defocus range and step size: Here you can set the parameters for the initial brute force search. The first parameter is the start of the search, the second is the end of the search and the third is the step size over which the search is conducted. You can play with these parameters to obtain the best estimation - or simply try the suggested values.

Note:
You can test the parameters by using a single or a range of micrograph(s) as test only:

4. Click the  button to estimate the CTF.

5. Check the defocus values found in the program print-out on the left-hand side.

6. Also have a look at the half-half images (whgb_ampl_half_half) in the display on the right-hand side. Each image will contain a) in the left half the input amplitude image, and b) in the right half the "estimated" CTF. These "half-half" images should be used to check the accuracy of the CTF estimation. The Thon rings of both halves should fit.

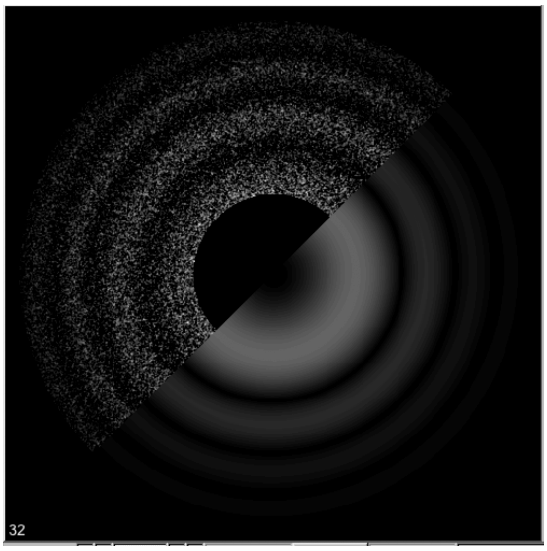


Fig. 13a: A good CTF half-half image

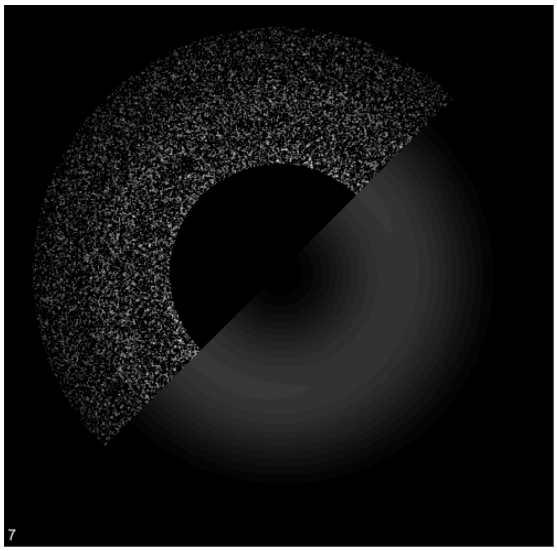


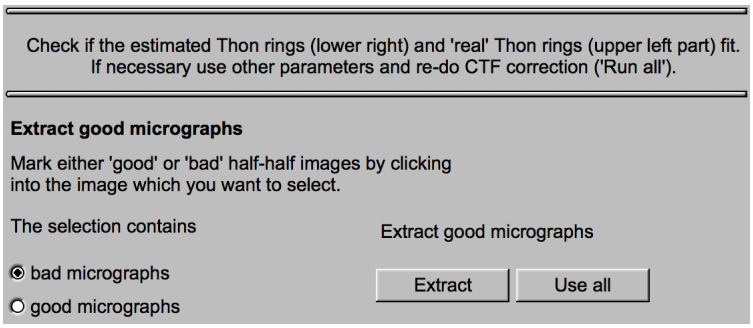
Fig. 13b: A bad CTF half-half image

7. Having found the correct CTF parameters you can now CTF correct the micrographs (flip the phases). The CTF corrected micrographs are whgb_micrograph_flip_all.

   Click the [Correct only] button

8. If wanted you can select good or bad micrograph by clicking into the displayed CTF half-half images.

   Use either the "Extract" button to extract the good micrographs or click the "Use all" button if you would like to use all micrographs. The final CTF corrected micrographs are stored in whgb_micrograph_flip.



6. As usual, click the [Next] button to continue.

NOTE:

You can alternatively estimate the defocus parameters with CTFFIND3 or CTFFIND4 (Linux or Mac OS X).

CTFFIND3 / **CTFFIND4** are programs provided by the Grigorieff lab (http://grigoriefflab.janelia.org/ctf) and is licensed under the terms of the GNU Public License version 3 (GPLv3). The programs are not part of **IMAGIC** but **CTFFIND3** can be used within the **IMAGIC** environment with command CTFFIND3, if the **CTFFIND3** program is installed in the FREALIGN directory of **IMAGIC**.

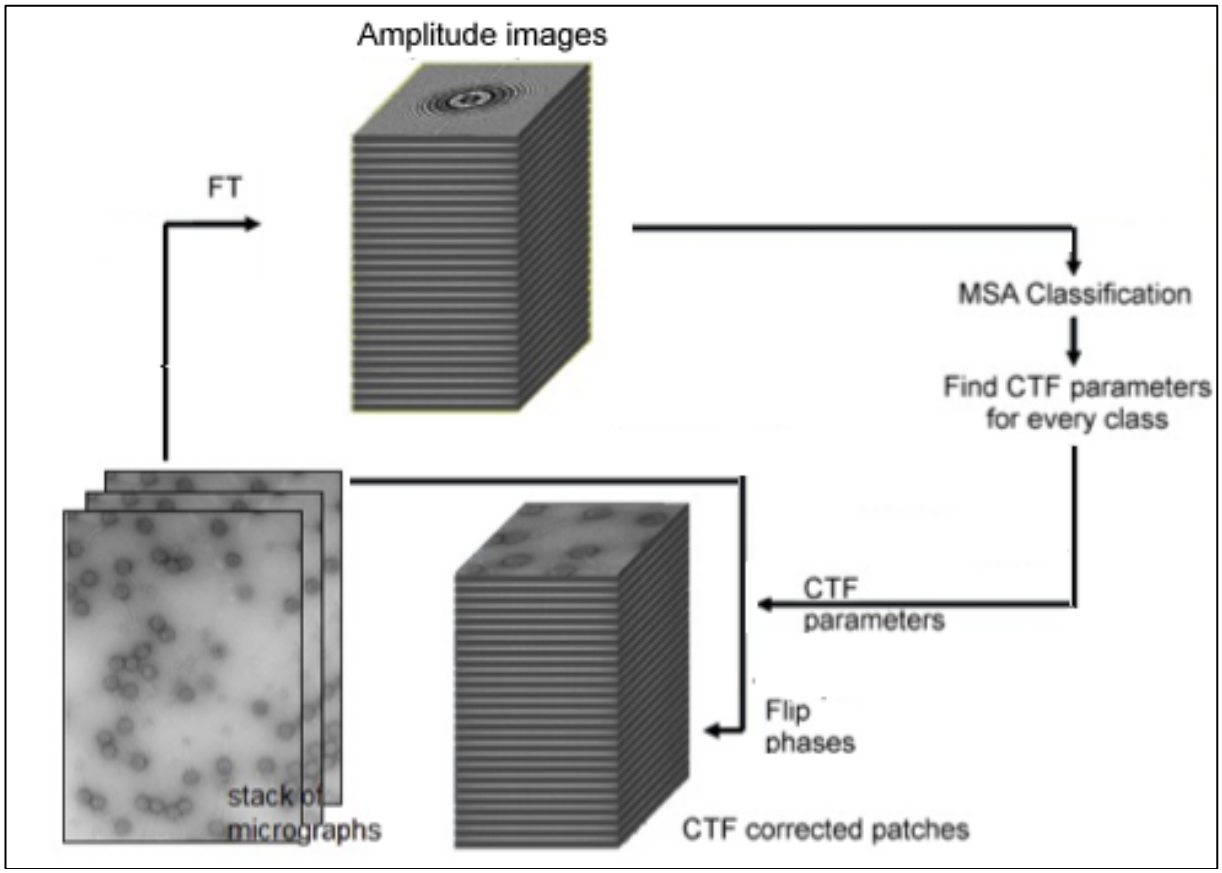


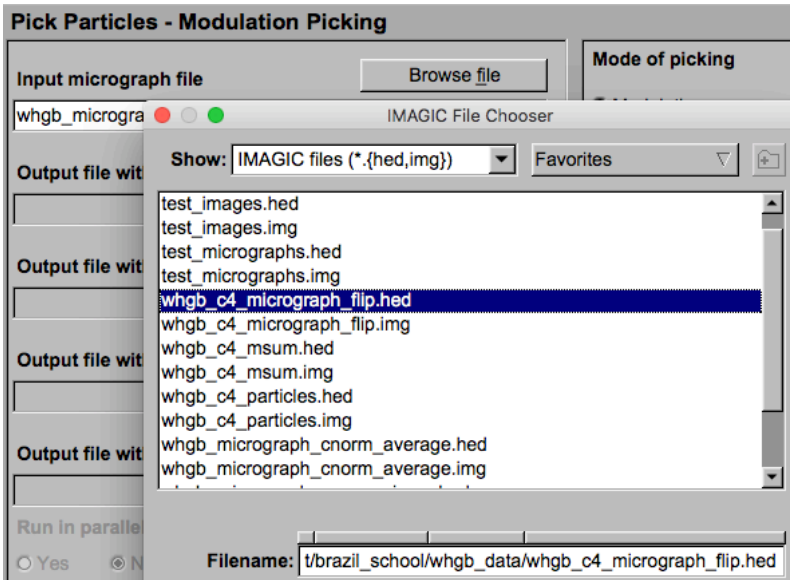
Fig. 14: Automatic CTF correction using MSA and Classification

YOUR NOTES:

13. Particle Picking

After CTF correction you are ready to select the particles from the stack of the CTF-corrected micrographs.

You will find the **IMAGIC** image file whgb_c4_micrograph_flip with 500 CTF corrected micrographs (movie sums) on the hands-on directory **whgb_data**, which you should use as input for all particle picking procedures.

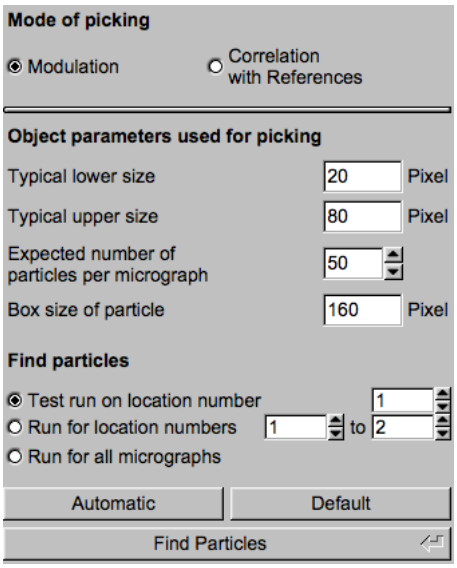


13.1. Pick Particles – Modulation Picking

Particle picking is an essential step in image processing. When working with low-contrast images of small proteins, and/or images taken close-to-focus, in which the particles are not clearly seen, it is important to avoid any bias in the data. Therefore, it is strongly recommended to do initial particle selection using the reference-free modulation picking (as was explained in the lectures).

Input are the CTF corrected movie sum micrographs (whgb_c4_micrograph_flip) provided on the hands-on directory **whgb_data**.

1. First search particles in a single or in a small number of micrographs.



2. Move the cursor over all input boxes to get help. Play around with the search parameters and check the particles found in the displayed micrographs.

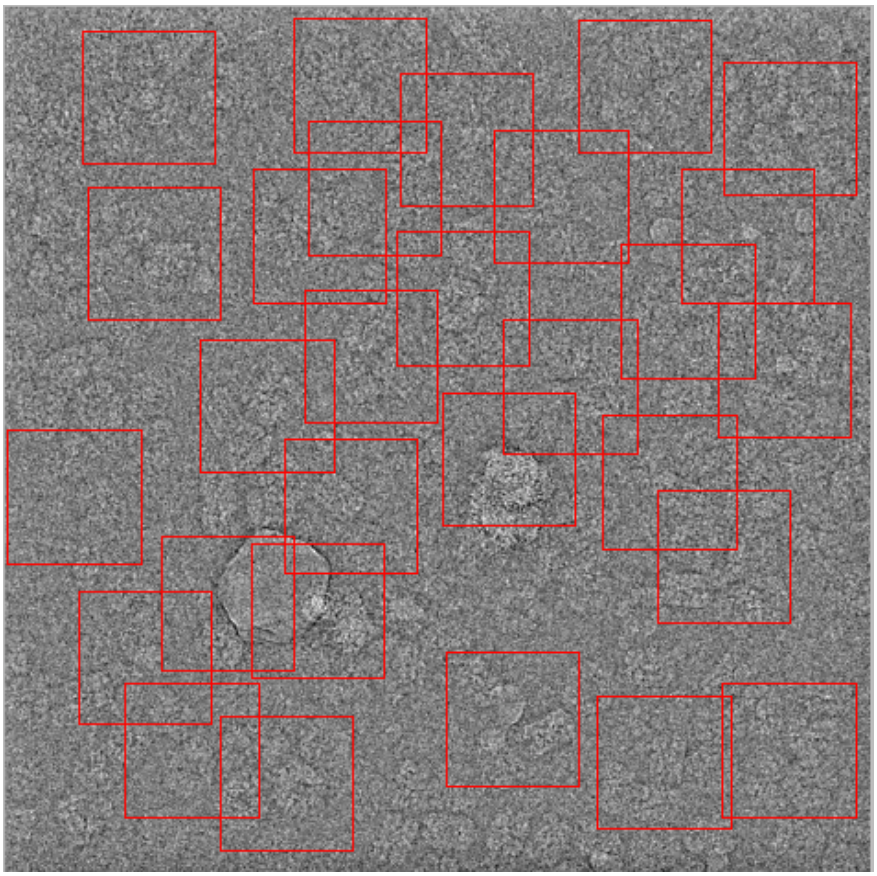


Fig. 15a: Modulation search: a micrograph

3. Also have a look at the modulation images.

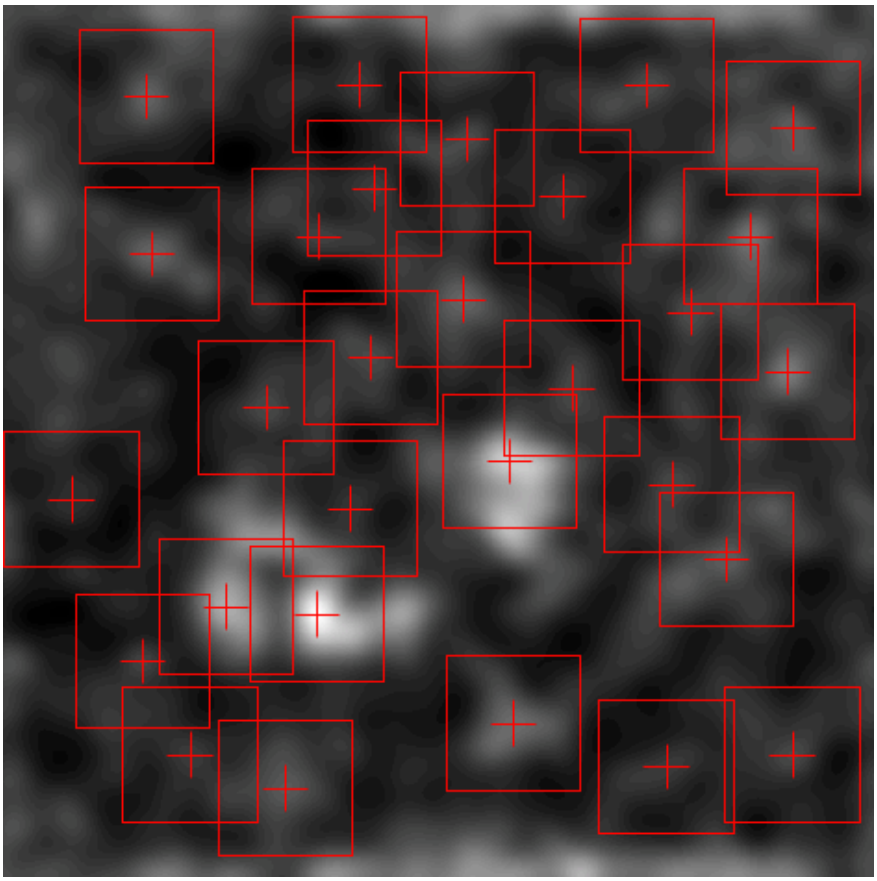
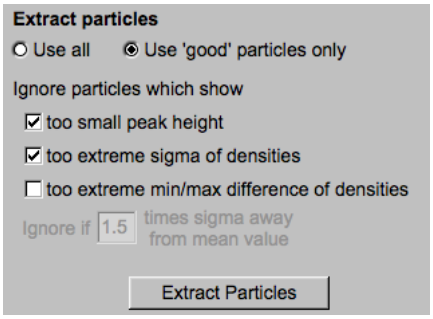


Fig. 15b: Modulation search: the related modulation image

4. Finally, search particles in all micrographs.

5. Now the particles can be extracted. Particles picked using this approach usually contain a lot of junk (ice, carbon foil, clumped particles). In a first approach, they can be sorted out by looking at the statistics of the picked particles.



6. Check the extracted particles displayed on the right-hand side.

You can see that the first displayed images show ice blobs etc. whereas the last images usually contain noise. The particle images are in between.

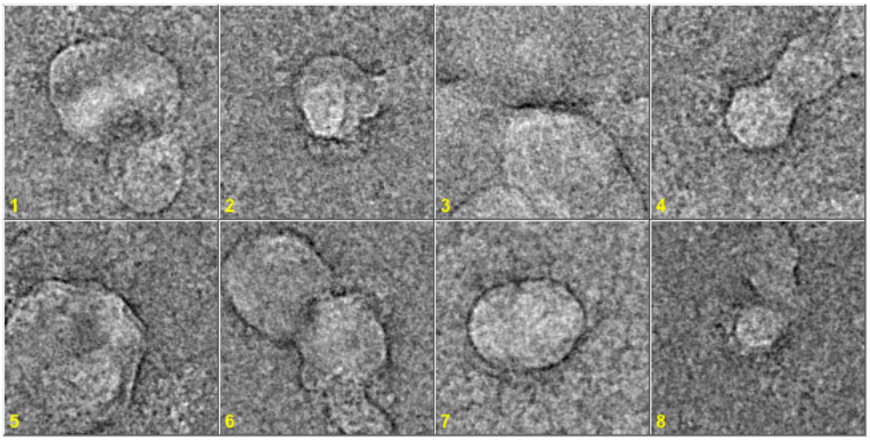


Fig. 16a: Some of the bad modulation picked particles at the beginning

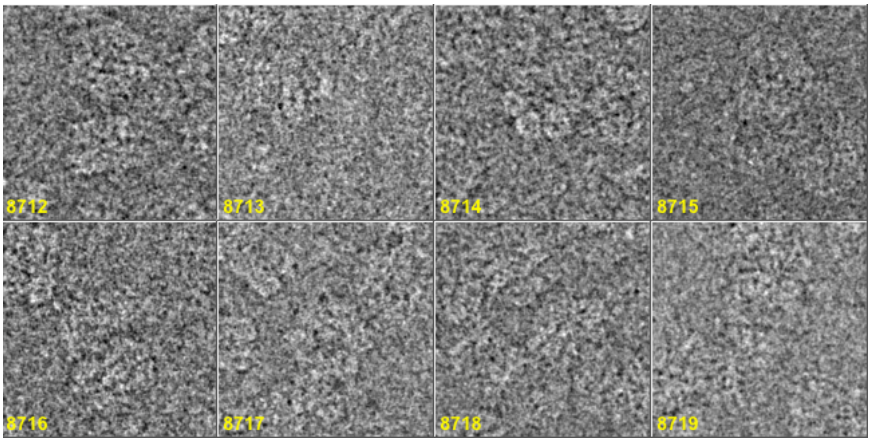


Fig. 16b: Some of the good modulation picked particles

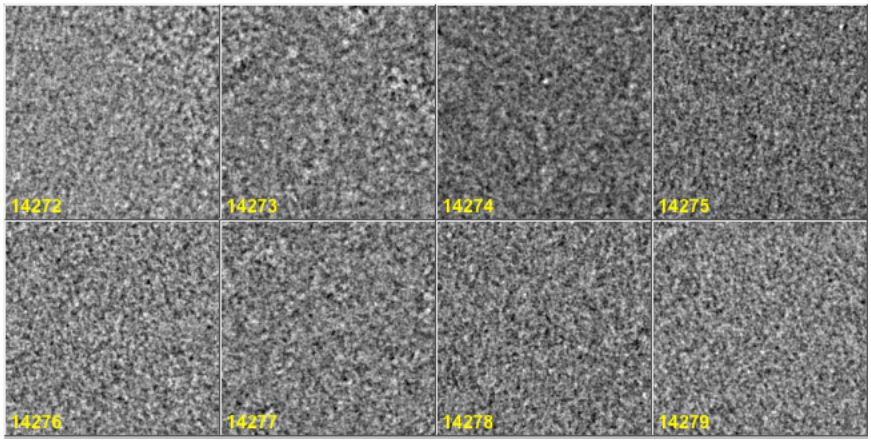


Fig. 16c: Some of the bad modulation picked particles at the end

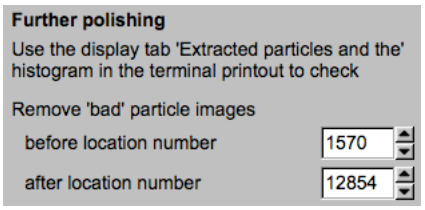
7. Check the extracted particles and write down the location numbers where the good particle images start and where they end.

   Also have a look at the histogram printed in the terminal window to get an idea where to find this range of good particle images:

```
Histogram of CCCs from whgb_mod_particles_pick

14279     0     0   3.43E-02 |
14277     2     2   5.01E-02 |
14258    21    19   5.81E-02 |
14200    79    58   6.60E-02 |*
14081   198   119   7.39E-02 |***
13887   392   194   8.18E-02 |*****
13596   683   291   8.98E-02 |*******
13089  1190   507   9.77E-02 |************
12326  1953   763   1.06E-01 |******************
11215  3064  1111   1.14E-01 |**************************
 9777  4502  1438   1.21E-01 |**********************************
 8112  6167  1665   1.29E-01 |****************************************
 6502  7777  1610   1.37E-01 |**************************************
 5136  9143  1366   1.45E-01 |********************************
 3984 10295  1152   1.53E-01 |***************************
 3117 11162   867   1.61E-01 |********************
 2460 11819   657   1.69E-01 |***************
 1998 12281   462   1.77E-01 |***********
 1640 12639   358   1.85E-01 |********
 1360 12919   280   1.93E-01 |*******
 1137 13142   223   2.01E-01 |*****
  965 13314   172   2.09E-01 |****
  842 13437   123   2.17E-01 |***
  706 13573   136   2.25E-01 |***
  602 13677   104   2.32E-01 |**
  518 13761    84   2.40E-01 |**
  449 13830    69   2.48E-01 |**
  375 13904    74   2.56E-01 |**
  316 13963    59   2.64E-01 |*
  271 14008    45   2.72E-01 |*
  218 14061    53   2.80E-01 |*
  188 14091    30   2.88E-01 |*
  147 14132    41   2.96E-01 |*
  118 14161    29   3.04E-01 |*
  102 14177    16   3.12E-01 |
    0 14279   102   3.28E-01 |**

Low values binned at  lower edge : 0
High values binned at higher edge: 97

Meaning of columns: # remaining, # accumulated, # in this bin, bin value
```

Fig. 17: Histogram of modulation search peak heights

8. Finally use the Remove Particles button to get the final modulation picked particles (whgb_mod_particles).

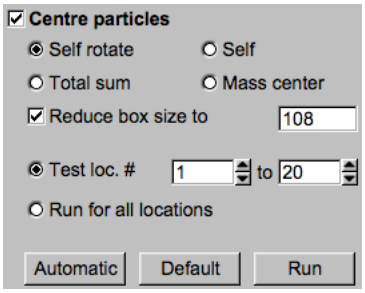


Note: the location numbers are YOUR choice.

13.2. Pick Particles – Select References from Modulation picked Particles

Modulation picked particles are usually not well centred. So, centring is the first processing we have to perform. The images of the centred particles will subsequently be MSA classified. Finally, you will extract good particle images which will be used as references for a subsequent correlation picking.

1. Play around with the various centring options:

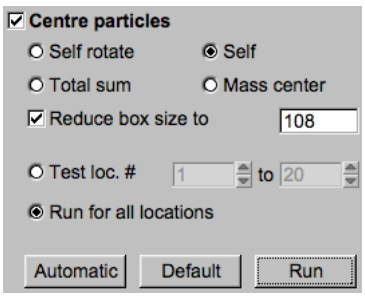


   Again, you can test the parameters for a few images before centring all images.

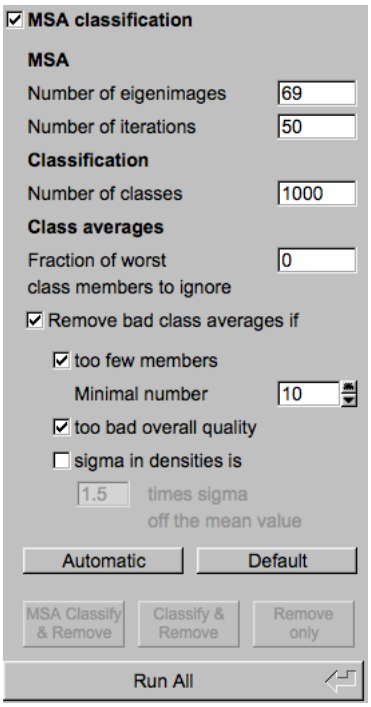
2. Look at both, the input and the centred particle images to check if the centring operation was okay.

   May be, you have to use more than one option by using centred images as new input files.

3. Finally run centring for all images.



5. The centred particle images will be MSA classified:



6. Have a look at the MSA eigenimages and the class averages.

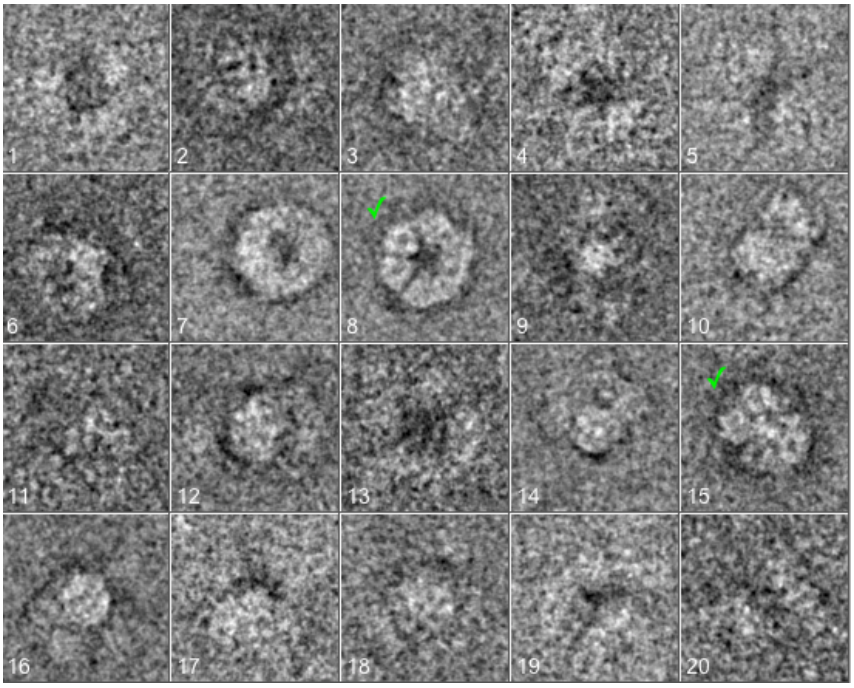
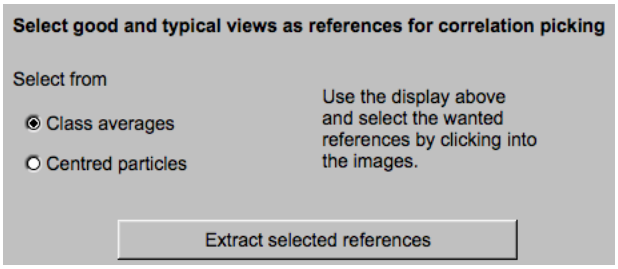


Fig. 17: Some class averages of the modulation picked particles. The green markers indicate selected images

NOTE:

After the first MSA and Classification you can re-run "Classify" and/or "Remove" with other options and parameters without re-calculating the more time-consuming MSA.

7. Check the displayed class averages and select good and typical particle views by clicking into the related image (a second click will de-select the image). Finally extract the selected references.

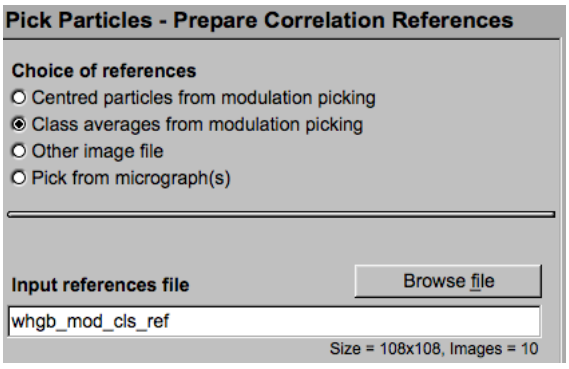


Note:
If the class averages do not show good and typical particle views you can alternatively select from the centred particle images (not suggested here).

7. In real science you would click the 

button to continue.

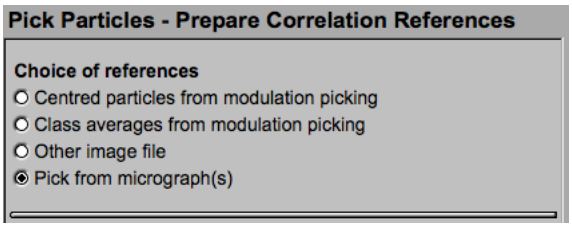
The suggested option on the next page (chapter 13.4) would be



Input are the created references (whgb_mod_cls_ref).

But here you should continue with another option to learn about an alternative option on how to get references for the correlation particle search.

Click the Next button and use the "Pick from Micrograph(s)" option on the next page:



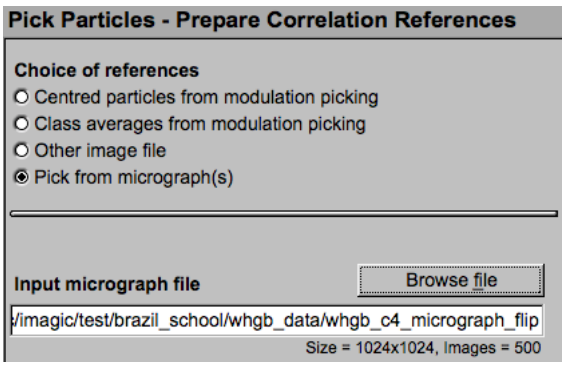
Input are the CTF corrected micrographs (whgb_c4_micrograph_flip) from your data directory **whgb_data**.

13.3. Initial Interactive Picking

Instead of creating picking references using modulation picking (chapters 13.1 and 13.2), you can also interactively pick reference particles. To avoid bias introduced by this selection the references will later be low-pass filtered and rotationally symmetrized.

1. If you are not on the related page click the Open Menu button.
2. Select the "**Prepare Correlation References**" page:

3. Choose option "Pick from micrograph(s)"



Input are the CTF corrected micrographs (whgb_c4_micrograph_flip) from the hands-on directory **whgb_data**.

4. Check the micrographs displayed on the right-hand side and select good particles representing typical views.

5. To select a reference, click into the centre of the wanted particle (another click will de-select it).

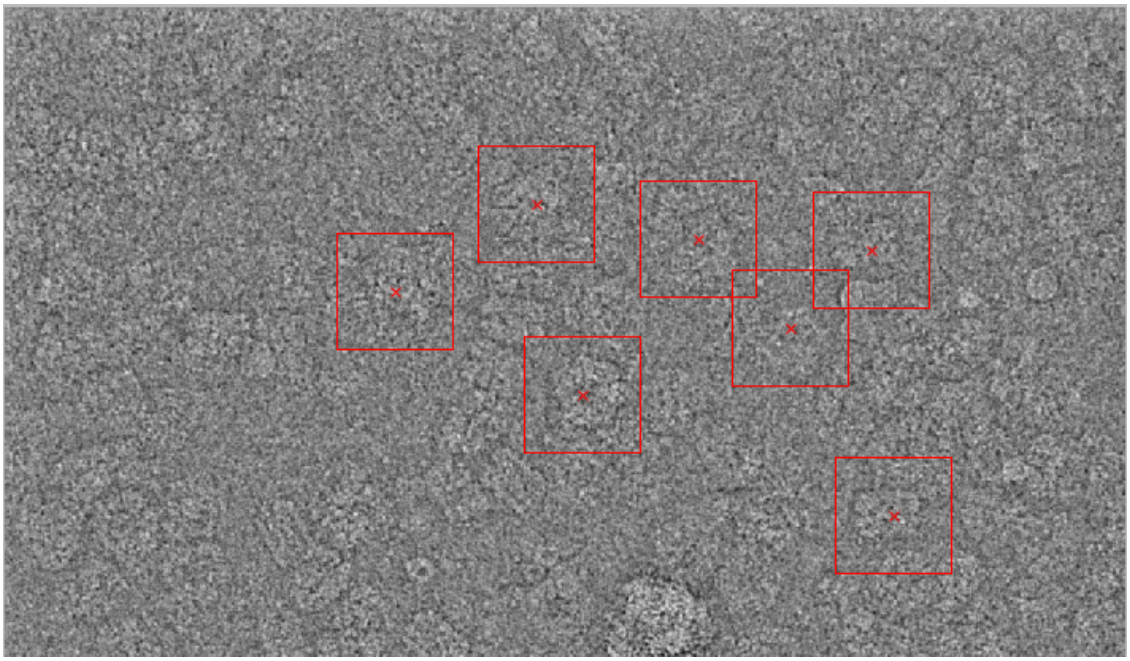
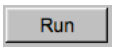


Fig. 18: Part of a micrograph with some selected reference particles

6. If all references are selected (3 to 6 particles) click the 

button.

7. The selected particles will be extracted and you can prepare them.

13.4. Pick Particles – Prepare Correlation References

Before correlation particles search you have to prepare the references.

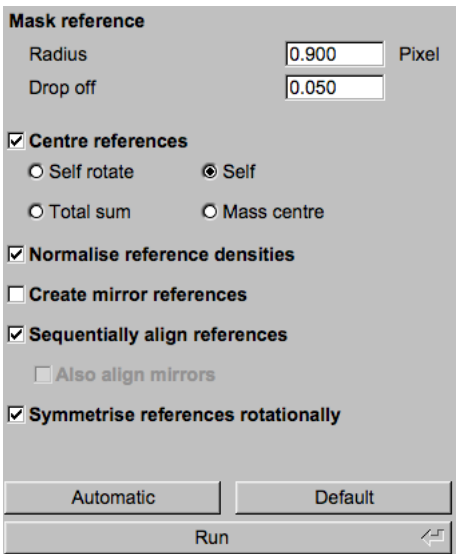
1. At the edges of the images there are features which are not related to the particles. This information is "removed" by applying a circular mask.

2. The reference particles have to be well centred. Play around with the various options to get good results.

3. You should also normalize the reference images.

4. The number of references is small, so you can sequentially align them.

5. The particles on the micrographs show all rotational orientations. To find all these rotated versions you should rotationally average the references.



6. Check the results: Are the references well aligned? Are they correctly aligned (check the side views)? Have a look at the final references.

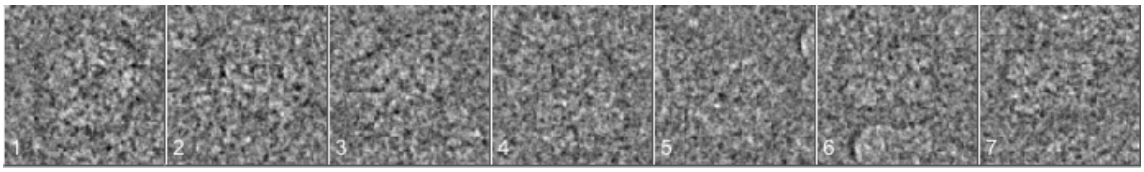


Fig. 19a: Some of the chosen references (whgb_refs_picked)

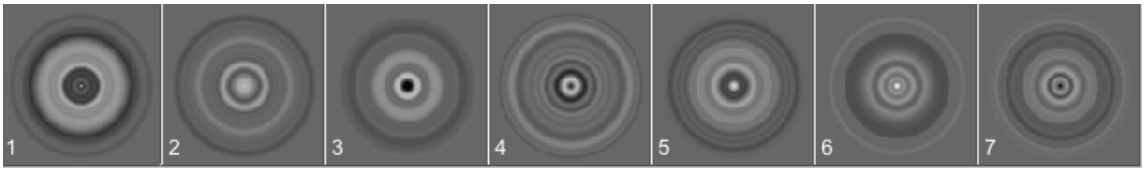
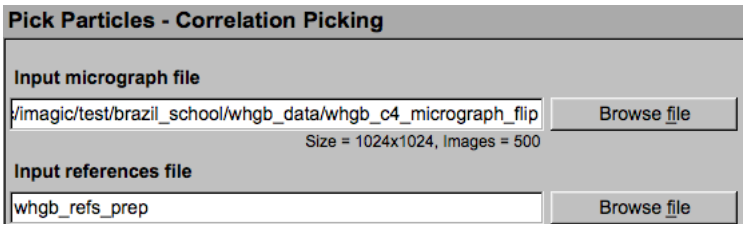


Fig. 19b: The related final prepared references (whgb_refs_prep)

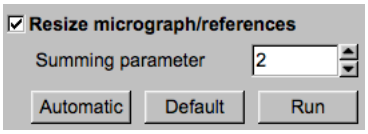
13.5. Pick Particles – Correlation Picking

Now you can pick particles using correlation picking.

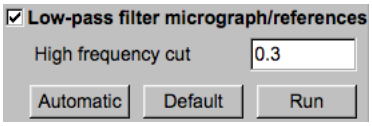
Input are the CTF corrected micrographs (whgb_c4_micrograph_flip) from the hands-on directory **whgb_data** and the prepared references (whgb_refs_prep). As suggested.



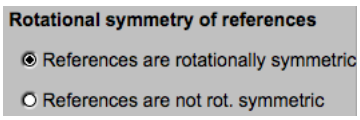
1. To speed up the calculations the micrographs can be resized (only used during the correlation search):



2. It is very important to strongly low-pass filter the references to avoid overfitting (correlation of noise):

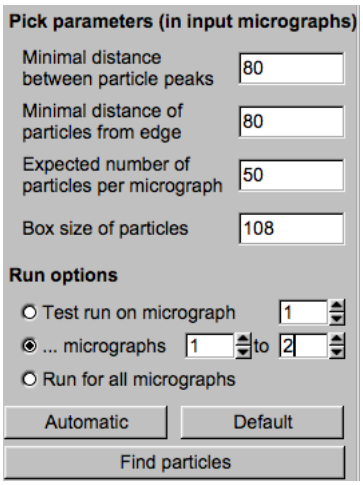


3. Your references (whgb_refs_prep) are rotationally averaged. Click the related button.



If the references are not rotationally averaged rotated versions of the references will be created during particle search. When re-searching particles at a later stage of a "real science analysis" with better references you would go for this option. This option is very time-consuming. So, do not use this option here.

4. Play around with the pick parameters. Run the search for only a limited number of micrographs to test how the given parameters influence the search.



NOTE: The particles are not yet extracted.

5. Check the particles found in the displayed micrographs.

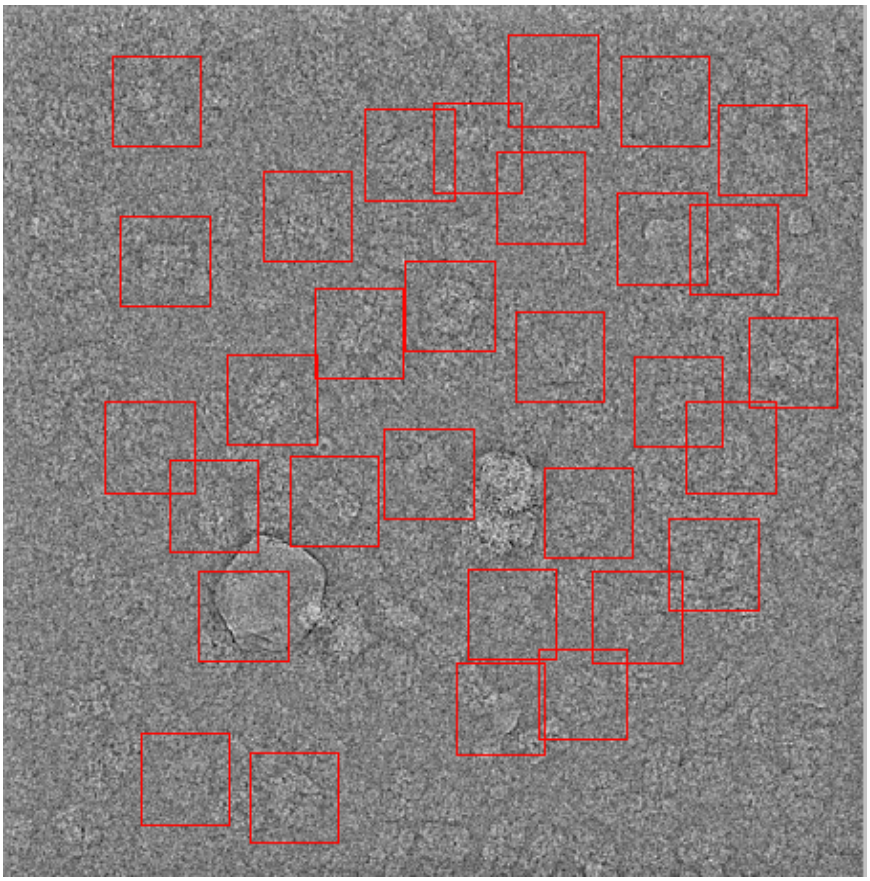


Fig. 20a: Correlation search of particles: a micrograph with particles found

6. Also have a look at the correlation images.

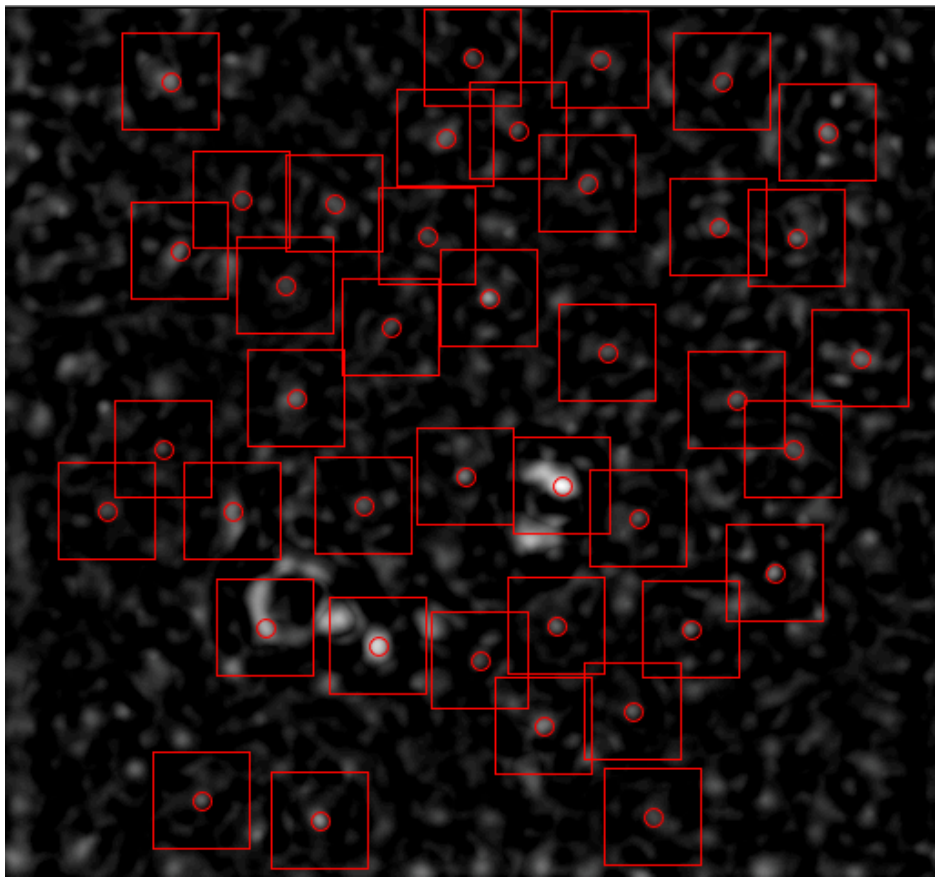
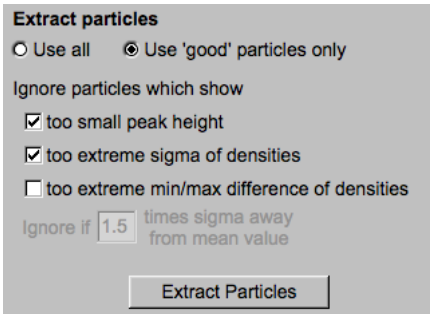


Fig. 20b: Correlation search of particles: the related correlation image

7. Finally, search particles in all micrographs

8. Now the particles can be extracted. Like in Modulation Picking, the images contain a lot of picked junk (ice, carbon foil, clumped particles). In a first approach, they can be sorted out by looking at the statistics of the picked particles.



9. Check the extracted particles displayed on the right-hand side.

You can see that the first displayed images contain ice blobs etc. whereas the last images usually contain noise. The particle images are in between.

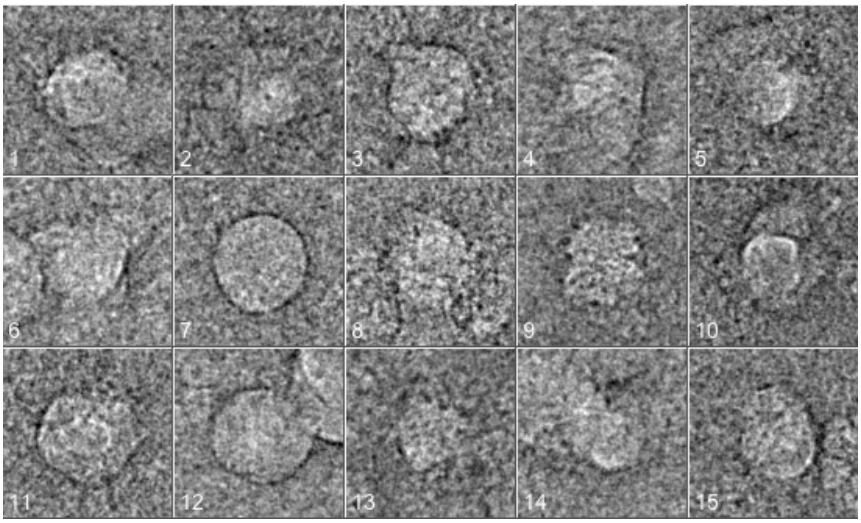


Fig. 21a: Some of the bad correlation picked particles at the beginning

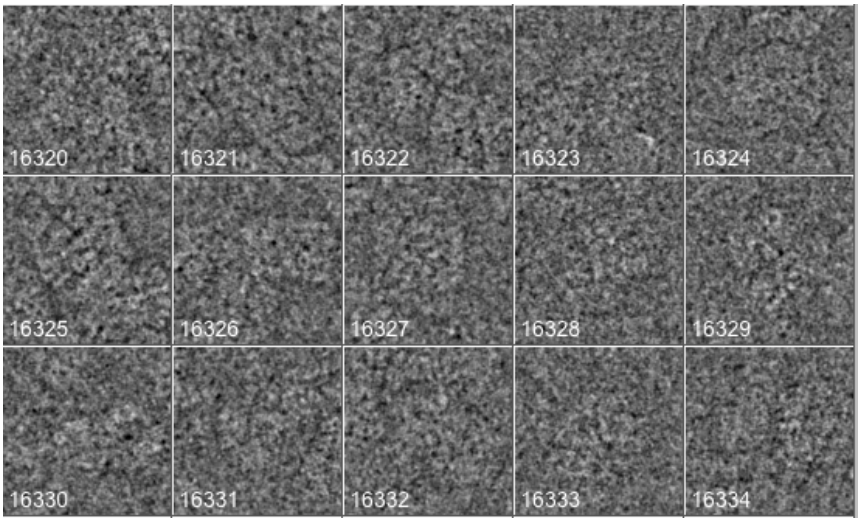


Fig. 21b: Some of the good correlation picked particles

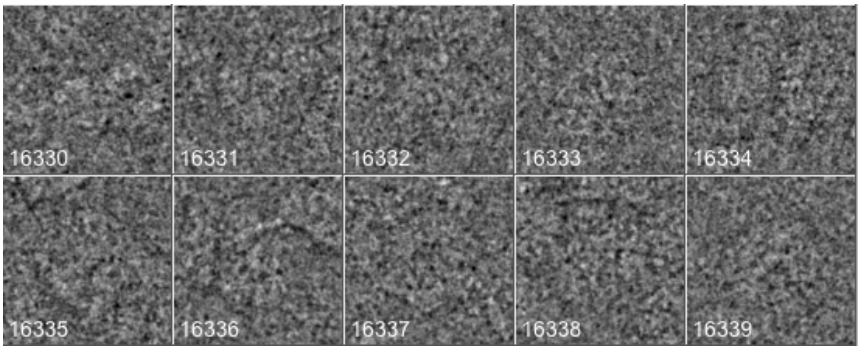


Fig. 21c: Some of the bad correlation picked particles at the end

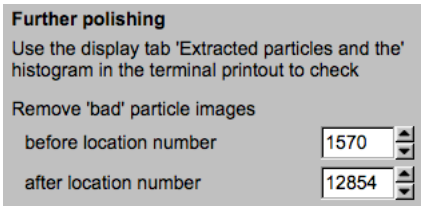
10. Check the extracted particles and write down the location numbers where the good particle images start and where they end.

    Also have a look at the histogram printed in the terminal window to get an idea where to find this range of good particle images:

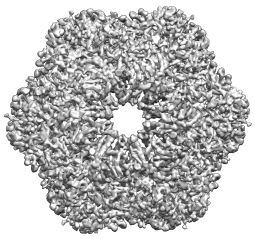
```
Histogram of CCCs from gisp_ccf_particles_pick_good

16341     0     0   1.15E+00 |
15627   714   714   1.27E+00 |********************
14703  1638   924   1.34E+00 |************************
13563  2778  1140   1.40E+00 |*******************************
12337  4004  1226   1.46E+00 |*********************************
11085  5256  1252   1.52E+00 |**********************************
 9696  6645  1389   1.58E+00 |****************************************
 8322  8019  1374   1.64E+00 |***************************************
 7109  9232  1213   1.71E+00 |********************************
 5999 10342  1110   1.77E+00 |******************************
 4991 11350  1008   1.83E+00 |***************************
 4138 12203   853   1.89E+00 |***********************
 3415 12926   723   1.95E+00 |*******************
 2832 13509   583   2.01E+00 |****************
 2336 14005   496   2.07E+00 |**************
 1954 14387   382   2.14E+00 |**********
 1604 14737   350   2.20E+00 |*********
 1348 14993   256   2.26E+00 |*******
 1126 15215   222   2.32E+00 |******
  934 15407   192   2.38E+00 |*****
  804 15537   130   2.44E+00 |****
  701 15640   103   2.51E+00 |***
  590 15751   111   2.57E+00 |***
  509 15832    81   2.63E+00 |**
  428 15913    81   2.69E+00 |**
  380 15961    48   2.75E+00 |*
  335 16006    45   2.81E+00 |*
  295 16046    40   2.87E+00 |*
  265 16076    30   2.94E+00 |*
  241 16100    24   3.00E+00 |*
  214 16127    27   3.06E+00 |*
  188 16153    26   3.12E+00 |*
  162 16179    26   3.18E+00 |*
  149 16192    13   3.24E+00 |
  130 16211    19   3.31E+00 |*
    0 16341   130   3.43E+00 |****

Low values binned at  lower edge : 0
High values binned at higher edge: 118

Meaning of columns: # remaining, # accumulated, # in this bin, bin value
```

Fig. 22: Histogram of the correlation coefficients

11. Finally use the Remove Particles button to get the final (correlation) picked particles (whgb_ccf_particles).



Note: The location numbers are YOUR choice.



14. Alignment-By-Classification (2-D ABC)

Alignment by classification is a method by which no reference is used. Neither external references nor references generated from 3-D volumes are used at this stage.

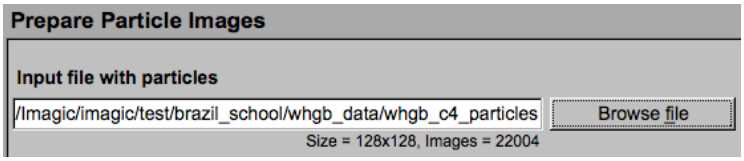
The (2-D) ABC procedure consists of the following steps:

- Prepare the particle images to concentrate on the most relevant information frequency range (chapter 13). Especially remove low frequencies (background ramps etc.) and the very high frequencies (noise).
- Centre the particle images (chapter 13). Usually this step is not necessary for particles picked with generic rotationally symmetric picking references. Nevertheless, something to keep in mind.
- MSA (eigenimage-eigenvalue) data compression classification (chapter 16) Study the eigenimages and make sure they contain details relevant to the particles you study.
- Unsupervised automatic classification followed by a carefully study (and sorting) of the best class averages (chapter 16).

15. Prepare Particle Images

At this stage of the analysis you have a number of options to prepare the picked/boxed particle images (whgb_ccf_particles).

To "synchronise" the data of all course participants we do not continue with the created best particles file (whgb_ccf_particles). Input will be the boxed particle images (whgb_c4_particles) from the hands-on directory **whgb_data**. The file contains picked/boxed particles, which we have prepared for you.



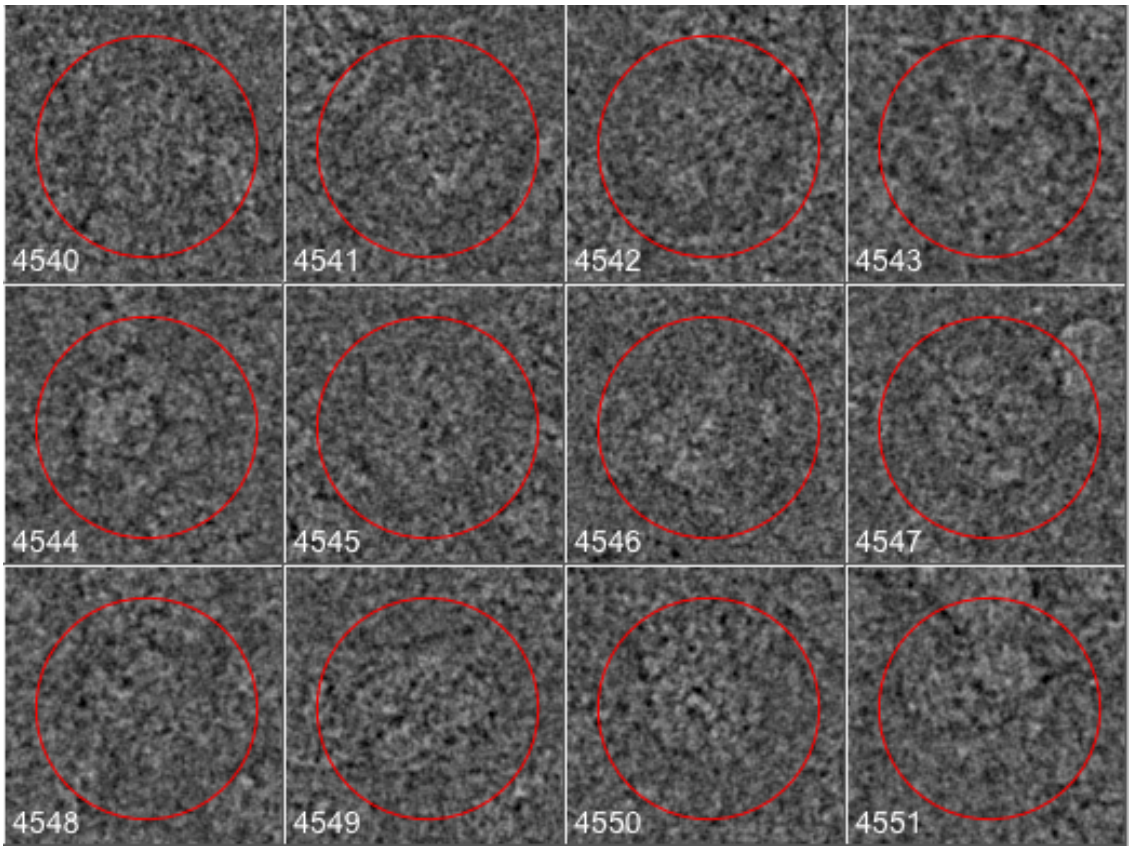


Fig. 23: Some of the picked/boxed particles (the red circle indicates the chosen MSA mask).

1. It usually is a good idea to specifically band-pass filter the particle images for further image analysis. Remember: a band-pass filter reduces the influence of unimportant or disturbing spatial frequencies. The very low spatial frequencies are suppressed because they only describe unwanted low-resolution background. The very high spatial frequencies are also strongly suppressed because they describe noise.

2. A circular mask is imposed to remove the unwanted information at the edges not describing the particles.

3. The particle images will be normalised which means that the densities are zero-floated (the average is set to zero) und sigma is set to the specified (arbitrary) value.

4. The provided particles (whgb_c4_particles) are already well centred. So, do NOT use the centring option here.

> REMEMBER:
>
> Move the cursor over all input boxes and option buttons to get context sensitive help.

5. Play around with the parameters and optionable only using the first 20 particle images. Have a look at the prepared images to see the effects.

6. Use a very strong low-pass filter and find out what is described by the removed high frequencies:

Note that a low-pass filter reduces the high frequencies which describe the fine details (noise but also fine particle details):

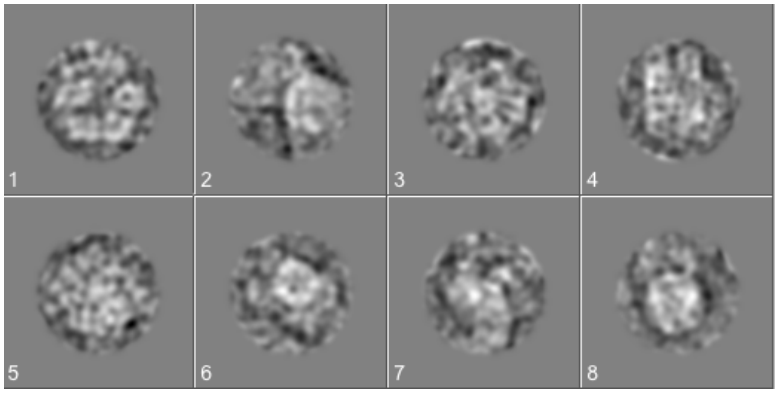


Fig. 24a: Some strongly low-pass filtered particle images

7. Now apply a very strong high-pass filter and find out what is described by the removed low frequencies:

Note that a high-pass filter reduces the low frequencies which describe the large details (background but also particle shapes etc.):

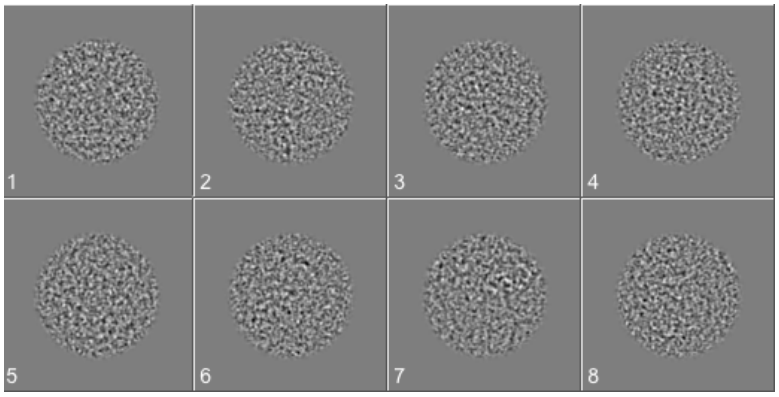


Fig. 24b: Some strongly high-pass filtered particle images

8. Play around with other filter parameters. Finally, use a reasonable band-pass filter like this:

Impose this band-pass filter into all particle images.

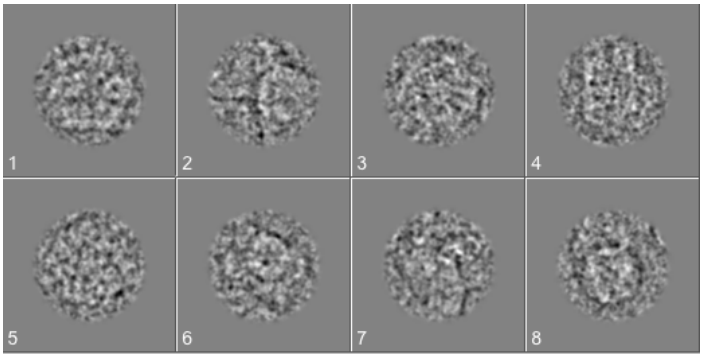


Fig. 24c: Some of the final band-pass filtered particle images

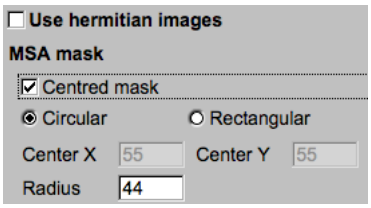
16. First Multivariate Statistical Analysis and Classification

The aim of Multivariate Statistical Analysis (MSA) and classification is to find similar images (views of the particle) so that we can average them to reduce the noise level (improve the signal-to-noise ratio "SNR") and to find the "typical" views, which we would like to use to calculate a 3-D reconstruction.

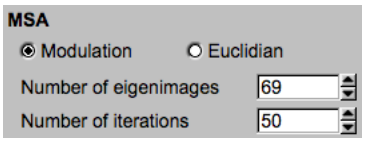
Input are the prepared particle images (whgb_prep).

MSA and classification consists of different steps:

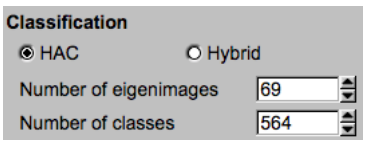
1. Specify a circular mask (use the mouse and the display to specify the mask). The mask defines which parts of images are to be analysed ("area of interest"). Only pixels falling within this mask are actually contributing to the analysis.



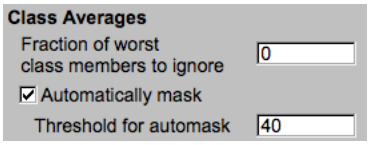
2. The MSA eigenimage-eigenvalue calculations: you can define the number of eigenimages and the number of iterations (usually large).



3. Classification of the MSA treated particle images: the number of classes you choose is related to the average number of images per class you would like. You can play with this value to see how the quality of the classes is affected. Ideally, you would have as few members per class as possible whilst still obtaining high contrast class averages.

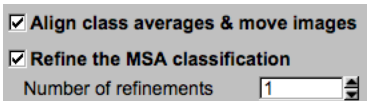


4. Create class averages: Average all the particles that belong to the same class.



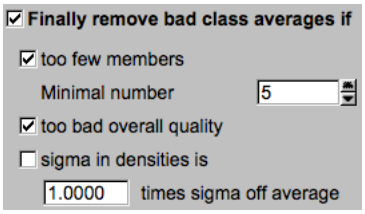
Always mask the class averages. Only the particles information is wanted during the subsequent image analysis.

5. The class averages can be sequentially aligned and the class images moved by their class average. A new refined MSA classification can be calculated now using the moved images as input.



This refinement is part of the (new) Alignment-By-Classification procedure and is not used in this practical.

6. Remove bad class averages: You can remove "bad" class averages which usually can be found by checking the number of class members, the overall quality in MSA and classification or the sigma of the class average densities.



7. After the first MSA run have a look at the eigenimages. The eigenimages of a centred dataset are a good way of examining the information content of a dataset.

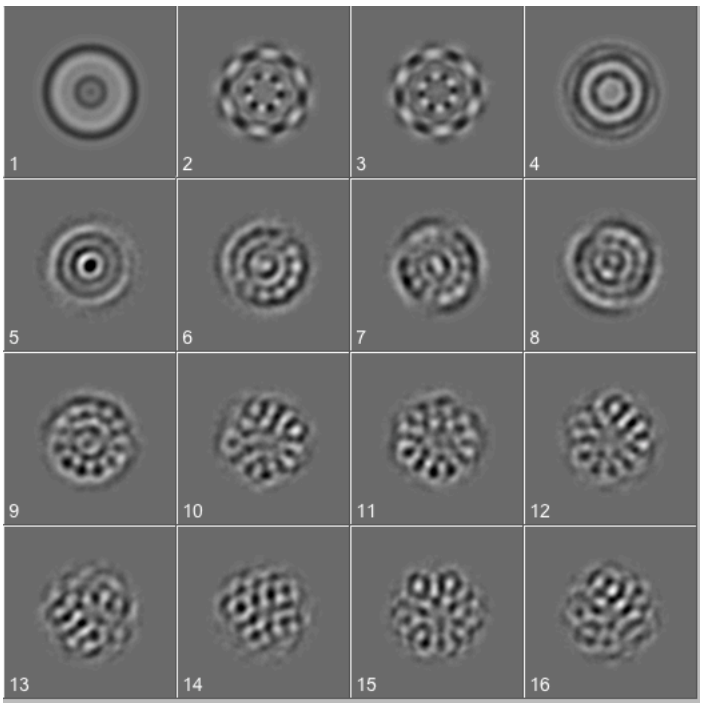


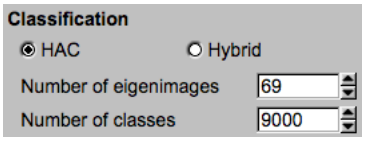
Fig. 26: The first MSA eigenimages of the (rotational) unaligned worm hemoglobin data-set. (Remember that worm hemoglobin has D6/622 symmetry.)

Note that the first eigenimage always shows (a sort of) average of all images.

The worm-hemoglobin has D6/622 symmetry and the images were not yet rotationally aligned, so you should find eigenimages that are rotated to each other (like a sine and cosine wave, as the 2nd and 3rd in figure 26) showing this 6-fold cyclical symmetry.

8. Now play around with the classification parameters and check the results. If you do not change the MSA parameters you do NOT have to re-calculate MSA. You can use the "Classify only" and /or the "Remove only" button.

First create a huge number of class averages.



Note that a too large number of class members does not give good averages. One needs a reasonable number of images per class to get good class average images:

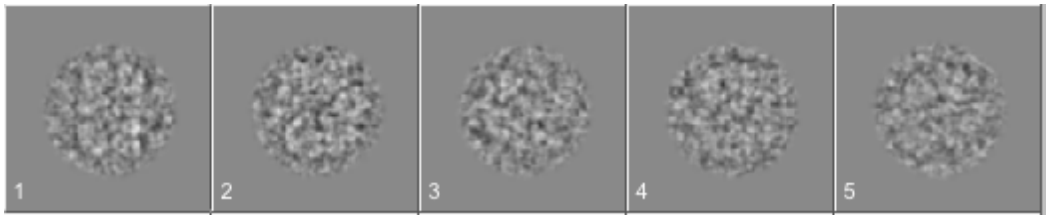
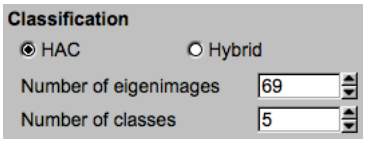


Fig. 27a: Some class averages when using a huge number of classes

Create a very small number of classes:



Note that a too small number of classes smears out all details.

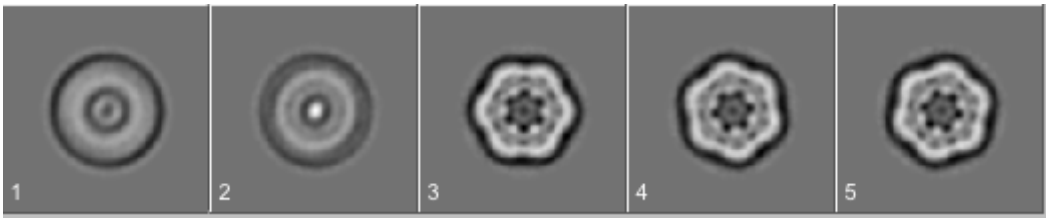
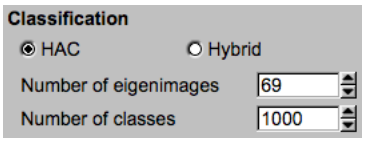


Fig. 24b: Some class averages when using a small number of classes

Finally create a reasonable number of class averages (10 to 20 members per class):



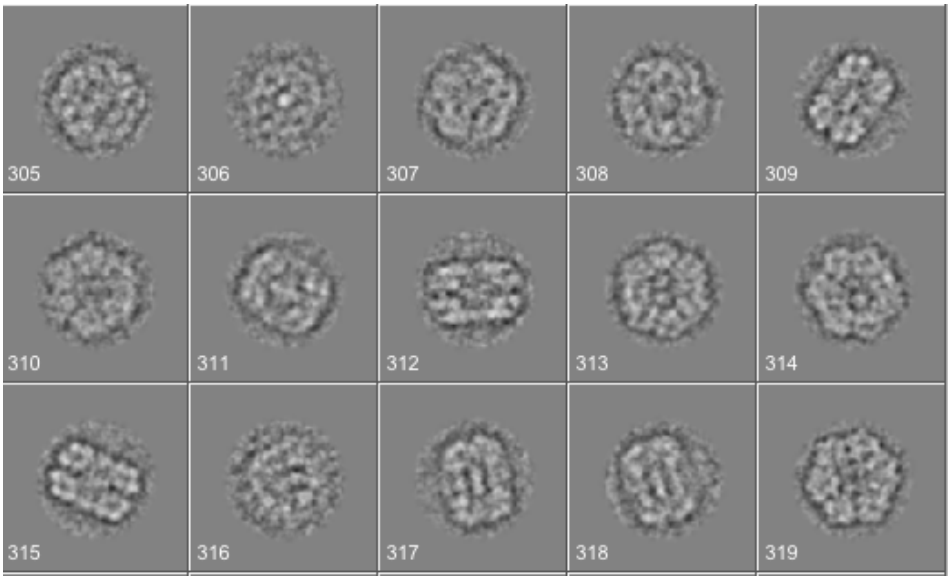


Fig. 27c: Some class averages when using a reasonable number of classes

9. But even with a reasonable number of classes you will see that there are "good" class averages showing particles views with high resolution but also a number of "bad" class averages showing classes with low resolution (weird ice clumps, for example). This is why removing bad class averages (6) is important.

10. If wanted the MSA and classification procedure can be refined: The class averages can be sequentially aligned to each other and the class member images moved by their aligned class averages. The moved particle images are MSA classified again. The idea is to align the single particles images without correlating any noisy images (which is a risk of over-fitting and reference bias).

    This part of the ABC work-flow is time consuming and should be skipped in this hands-on (your class averages are already aligned very well).

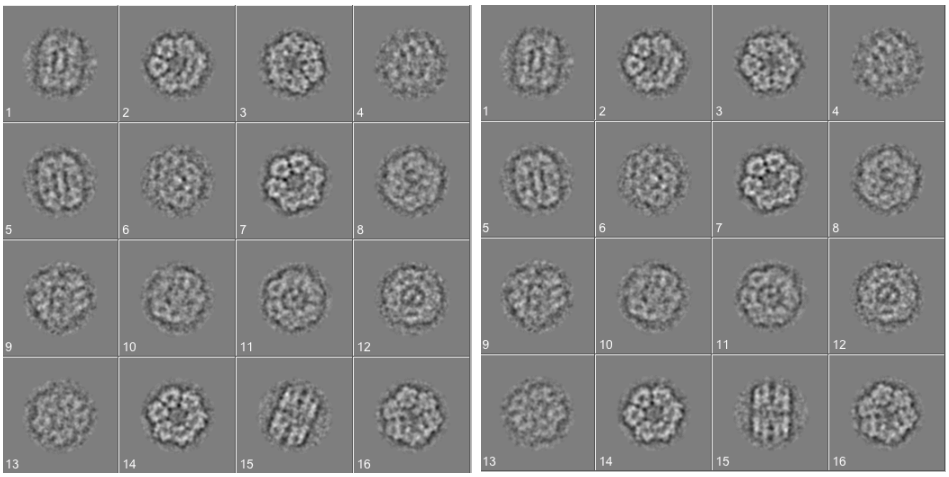
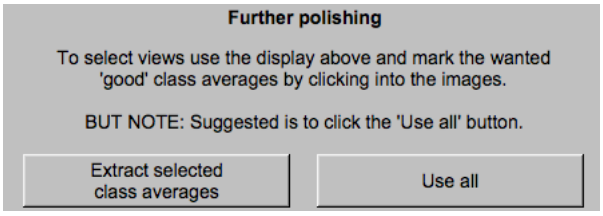
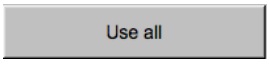


Fig. 27d: Some class averages before and after the sequential alignment

11. Finally, you can extract good / exclude bad micrographs.



But suggested is to click the  button

The final class averages are named whgb_classums_best

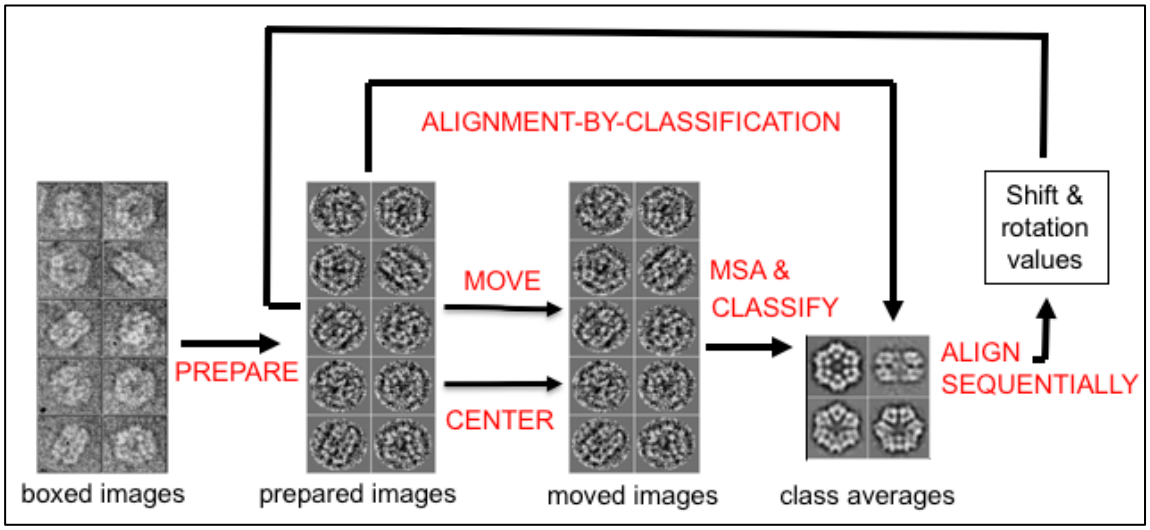


Fig. 28: Alignment-by-Classification without Multi-Reference Alignment

17. 3-D Reconstruction – Extract Class Averages

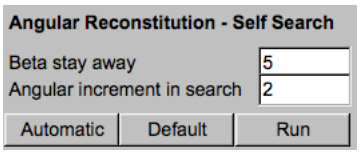
17.1. Angular Reconstitution – Self-Search

This option is not necessarily needed. But it can be helpful in case you want to extract class averages by range or interactively.

Angular reconstitution / self-search is a procedure to check how well each single class average conforms to the given point-group symmetry of the particles. This only works for high symmetric point-groups.

The idea of angular reconstitution / self-search is to sort the class-averages with the smallest residual for the given point-group symmetry to start up a 3-D reconstruction. Each class-average image is examined exclusively with respect to itself.

1. The worm hemoglobin particles have a D6/622 symmetry so you can run angular reconstitution / self-search here to get a feeling what are good class averages to start-up the 3-D reconstruction.



2. The class averages used are displayed and their beta angle is printed in yellow.

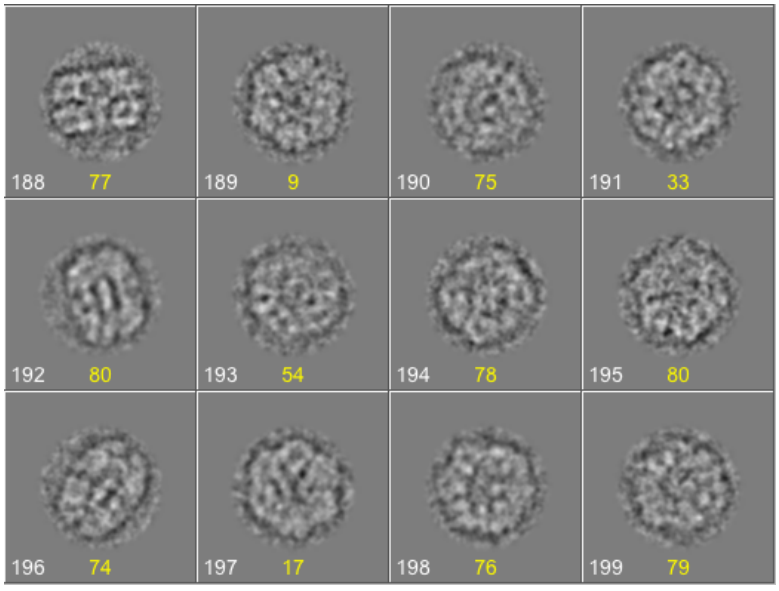


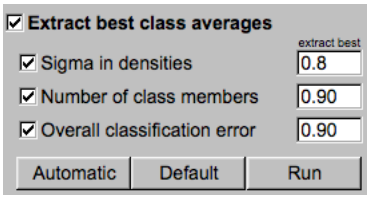
Fig. 29: Some class averages with self-search Euler angle Beta

Check if the particle views and the beta angles fit to each other. A close-to-top view is expected to have a beta angle around 0 to 5 degrees, a close-to-side view an angle around 70 to 90 degrees.

17.2. 3-D Reconstruction – Sort Class Averages

For the first random start-Up 3-D reconstruction you should only start with the best class averages. Although you already removed bad class averages during MSA and classification (chapter 16) you should extract only the very best class averages here.

The first option is to sort and the class averages by sigma in the densities, the number of class members and the over classification error:



Please specify the number of best class averages to be extracted. Note: "0.8" means extract the 80% best, "800" means extract the 800 best class averages.

17.3. Select Class Averages for Initial 3-D Reconstruction

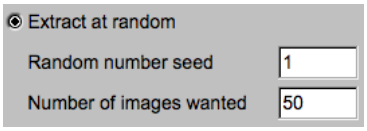
Now, there are a number of options on how to finally specify the (sorted) class averages to be used for the initial 3-D reconstruction:

- 

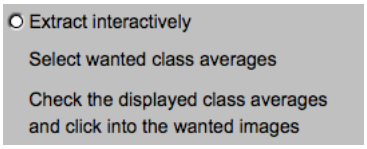
  Use all (sorted) class averages.

- 

  Only the range of (sorted) class averages specified will be used. If the class averages are sorted this can be a good option.

- 

  The specfied number of class averages is extracted at random (suggested).

- 

The class averages to be used are chosen interactively. Use this option if you would like to fully control the selection. But note that this can be a bias. This option can also be helpful in practicals to get an idea how automatic procedures work.

As usual check the displayed class averages (and their beta angles) and select good and typical particle views by clicking into the related image (a second click will de-select the image).

For the subsequent 3-D reconstruction you should select a few good class-averages with very different Euler angles Beta:

a) First select two intermediate view. Both views should look like intermediate views and should have "intermediate" Euler angles Beta (around 40°-60°).
b) Next select a (close-to-) side view, which are the ones that have rectangular like shape. Try to find such a view, whose Euler angle Beta is around 70-80°.
c) Finally select a (close-to-) top view (round shape) with a Beta angle around 10°.
d) Select 2-3 other intermediate or side views.

18. 3-D Reconstruction – Random Start-Up

Once you have good class averages you need to find their relative 3-D orientation (Euler angles).

Worm hemoglobin is a molecule with D6/622 point-group symmetry. This high degree of symmetry makes the initial angular assignment much easier than with lower degrees of symmetry where, for example the influence of preferred orientations is much higher.

There are two main options for the 3-D reconstruction start-up options:

- Random start-up using multiple 3-D volumes:

  The idea is to create multiple 3-D volumes. In each 3-D reconstruction refinement, the input class averages can move to another 3-D volumes and 3-D volumes with a too small number of members are removed. Usually, this procedure ends up in a 'good' initial single 3-D volume or in a few 'good" 3-D volumes (may be, already describing some heterogeneity in the data-set).
- Random start-up a single 3-D volume

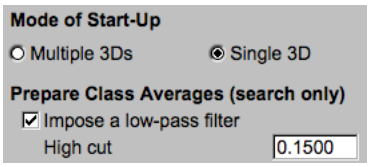
  A single 3-D start-up volume will be created which is refined in a number of iterations.

In this practical we first choose the option to create a single 3-D volume. The option creating multiple start-up 3-D volumes is explained in chapter 17.3.

18.1. 3-D Reconstruction – Random Start-Up – Single 3-D

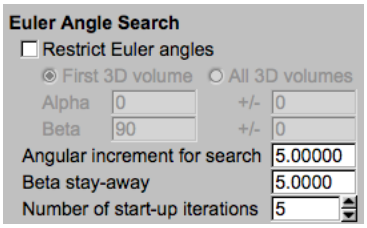
Input are the extracted class averages (whgb_classums_select).

1. First, you have to specify a number of parameters related to angular reconstitution (Euler angle search). As usual move the cursor over text and boxes to get help.



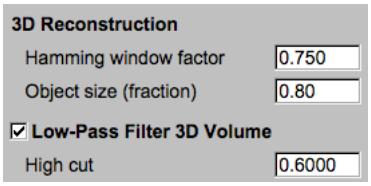
Always low-pass filter the class averages at this stage of the analysis. Of course, this filter is only used during the Euler angles search.

Specify and check the parameters for the search of Euler angles.

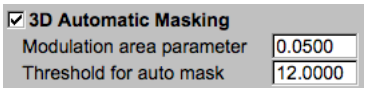


Do not use restricted Euler angles. This is only for elongated particles, tomographiv geometries etc.

2. There are also a number of 3-D reconstruction parameters you have to specify:



3. Very important is the masking of the 3-D volume. This is done automatically for which you have to specify the following parameters:



Note: You can later refine this mask using other parameters.

4. It is possible to refine the first 3-D volume. One can iterate the process of Euler angle search (angular reconstitution / anchor-set refinement) and 3-D reconstruction. See figure 30.

But this option is not suggested here in this practical.

5. The resulting 3-D volume will be visualized as a stack of 2-D sections (from bottom to top) and as a number of surface views. Specify a threshold value for the latter.

   Note: Very often the suggested automatic value is not correct. In such a case you can re-calculate the surface views later uisng nother threshold value.

6. Click Run all to start the 3-D reconstruction. Check the program print-out.

   Note that you can re-do certain operations wihout re-doing the full calculation:

   Filter Re-calculate 3-D filter, masked 3-D volume and surface views.

   Mask Re-calculate the masked 3-D volume and the surface views.

   Surface Re-calculate the surface views.

7. Try to understand the extensive output printed in the terminal window. Various indicators are available to see whether the Euler angle assignment is successful. After one or two iterations, the Euler angles assigned should stabilize (zero angles between current and previous Euler angle assignment).

8. Check the results:

   It is important to check "by eye" how well the re-projections match the input class averages (see display tabs on the right-hand side). By flicking between the two tabs compare how well these two match.

   If they do not match (especially if the rectangular shaped close-to-side views have roundish re-projections) the Euler angle assignment was not correct and you should re-do the procedure using other parameters, and/or another combination of your selected class averages.

9. Note:

   Very often the automatic threshold value for the surface views is not correct. In such a case re-calculate the surface views using another threshold value:

   Surface

NOTE:

You can also visualize the surface of your 3-D volume with the non-IMAGIC program **CHIMERA** which has an **IMAGIC** plugin. In the browse window select the **IMAGIC** file with the extension .hed or .img (both will work)

If you like to visualize the sections or the surface views in an endless loop (movie) press the Movie button. Input is either the file whgb_3d_startup_masked or the file whgb_3d_X, where X is the refinement number.

10. How to continue:

Results are okay. Continue with the next Page

The results are okay. Continue with the next page and calculate a refined 3D reconstruction. This is the same as clicking the "Next" button.

Redo 3D Random Start-Up (another random seed is used)

Redo the calculations on this page. Clicking this button will start a new "Run all". Another random number seed to set the initial Euler angles is used automatically.

Choose another Set of Class Averages

Return to the "**3D-Reconstruction – Extract Class Average**" page to choose another combination of class averages.

Instead of creating a 3D single 3D volume you can also create multiple 3-D random start-up volumes. Refer to the next chapter.

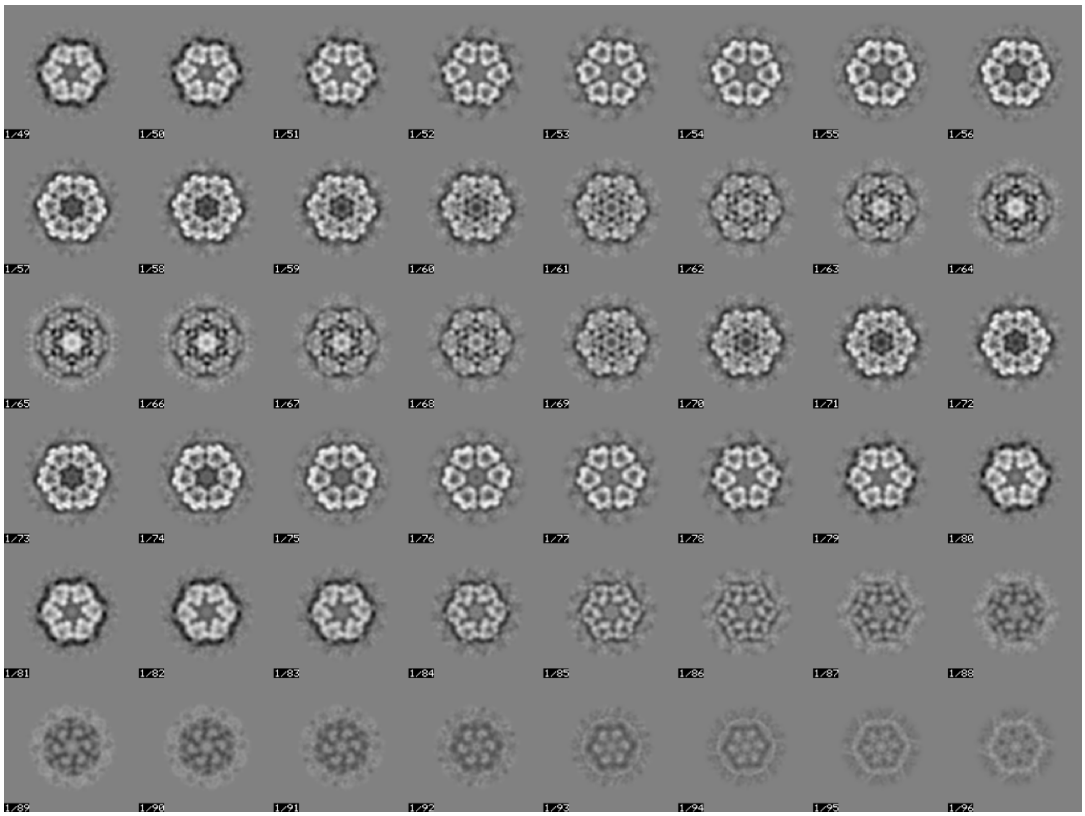


Fig. 30: A 3-D reconstruction of the worm hemoglobin particles shown section by section.

18.2. 3-D Reconstruction – Random Start-Up – Multiple 3-D

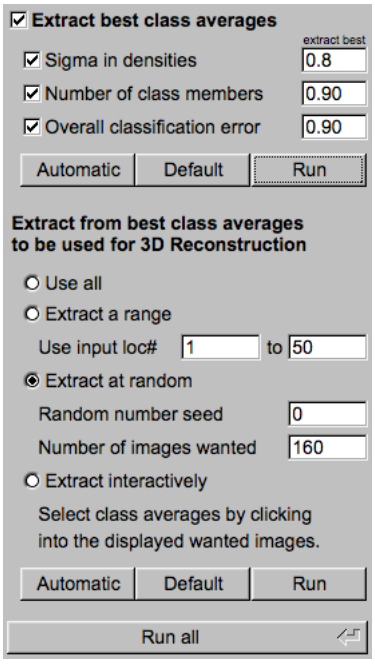
The idea is to create multiple 3-D volumes. In each new Euler angles search, the input class averages can move to another 3-D volumes and 3-D volumes with a too small number of members will be removed. Usually, this procedure ends up in a 'good' initial single 3-D volume or in a few 'good" 3-D volumes (may be, already describing some heterogeneity in the data-set).

This procedure is time consuming. In this practical only use a small number of class averages (whgb_classums_select) and only a few 3-D volumes to start with.

1. Go back to the page "**3D Reconstruction – Sort Class Averages**".

   As done for the single 3D reconstruction sort/extract and select the best class averages. If the sorted best class averages were already created for the single 3-D calculations you don't have to re-calculate them, of course.

   Note that you need a larger number of input class averages because are creating multiple 3D volumes here:

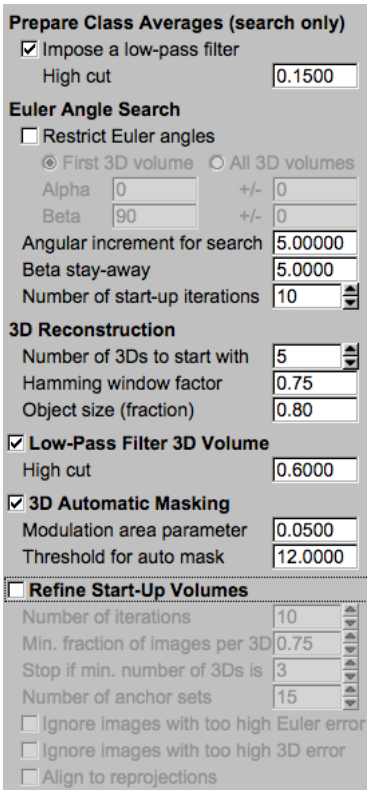


   As before click the "Next" button to continue.

2. Now you are ready to calculate a multiple 3-D random start-up reconstruction.

Input are all class averages (whgb_classums_best).

The parameters you can choose are more or less the same as in single 3-D mode (chapter 18.1.).



In contrast to the single 3-D calculations, you have to specify how many multiple 3-D volumes are to be calculated:

As done in single 3-D mode, click 

to start the 3-D reconstruction. Check the program print-out.

3. As before, you can re-do certain operations wihout re-doing the full calculation:

Re-calculate 3-D filter, masked 3-D volume and surface views

Re-calculate the masked 3-D volume and the surface views

Re-calculate the surface views

4. Check the results:

Do the re-projections match the input class averages (see display tabs on the right-hand side). By flicking between the two tabs compare how well these two match.

If they do not match (especially if the rectangular shaped close-to-side views have roundish re-projections) the Euler angle assignment was not correct and

you should re-do the procedure using other parameters, and/or another combination of your selected class averages.

5. If needed, re-calculate the surface views using another threshold value:

   Surface

6. Visualize the 3-D volumes in **Chimera**.

7. The options on how to continue did not change:

   Results are okay. Continue with the next Page

   The results are okay. Continue with the next page and calculate a refined 3D reconstruction. This is the same as clicking the "Next" button.

   Redo 3D Random Start-Up (another random seed is used)

   Redo the calculations on this page. Clicking this button will start a new "Run all". Another random number seed to set the initial Euler angles is used automatically.

   Choose another Set of Class Averages

   Return to the "**3D-Reconstruction – Extract Class Average**" page to choose another combination of class averages.

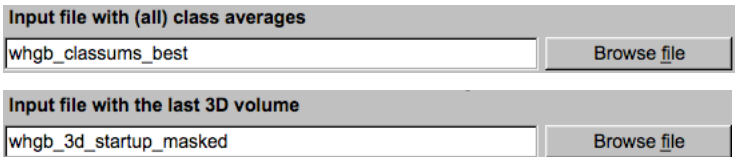
YOUR NOTES:

19. 3-D Reconstruction - Anchor-Set Refinement

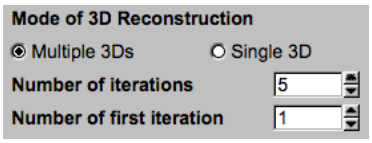
You have calculated first 3-D reconstructions, either using the single 3-D option (chapter 18.1.) or the multiple 3-D option (chapter 18.2.). Using angular reconstitution / anchor-set you can refine the previous image-processing steps.

First, one can use the 3-D volume to get all "typical" views if the 3-D volume is forward projected into all directions (if there is a point-group symmetry larger than C1 usually only into the so-called asymmetric triangle – refer to the lectures). The resulting "2-D forward projections" have well-defined Euler angles and can serve as references (a so called "anchor-set") to get (better) Euler angles of ALL class averages and a refined 3-D reconstruction. See figure 31.

1. Inputs are the last masked 3-D volume(s) (whgb_3d_startup_masked) and usually ALL class averages (whgb_classums).

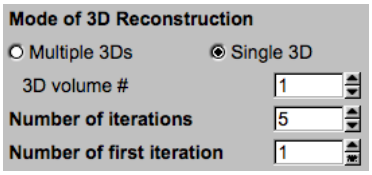


2. The 3-D reconstruction calculations on this page can be iterated using the created 3-D volume(s) as a new input 3-D volume (and the moved class averages as new input class averages). Please give the number of refinement iterations wanted and some related parameters:



(multiple 3-D mode)

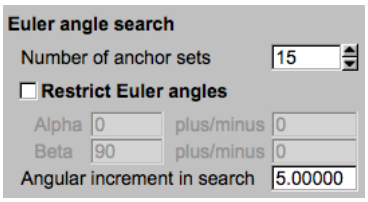
or:



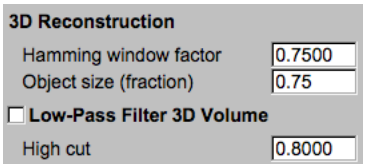
(single 3-D mode)

To keep track of the iterations you already did before, you can specify the number of the first new iteration. These numbers are used to automatically create output file names.

3. Specify some angular reconstitution parameters:

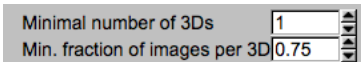


4. Also give some 3-D reconstruction values:



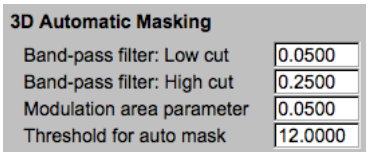
You may or may not, filter the 3-D volume.

Note that in multiple 3D mode you have to specfy a two important additional parameters:



An important procedure in multiple 3-D mode when iterating the Euler angles search and the 3-D reconstruction is that the class averages are allowed to "move" to another ("better") 3-D volume and that 3-D volumes with a too small number of members will be removed. You can specify the maximal number of iterations and the minimal number of 3-D volumes at which the iterations are to be stopped. But note that, nevertheless, in a certain iteration step only a single 3-D volume may "survive".

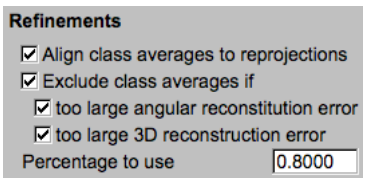
5. As before (chapter 18) masking the 3-D volume is very important and can be refined later (if needed).



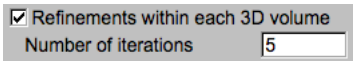
6. A possible refinement is to align the input class averages to their reprojections and re-calculate the 3-D volume with the aligned class averages.

Another refinement is to remove 'bad' class averages and to re-calculate the 3-D volume. During both, angular reconstitution (Euler angle search) and 3-D reconstruction, an error value is calculated. "Bad" class averages, wgich are the ones with a high angular reconstitution or a high 3-D reconstruction error are ignored in the refined 3-D reconstructions.

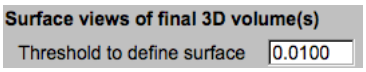
Note that all these refinements can take some computing time.



In multiple 3-D mode the competitive procedure of Euler angle search / 3-D reconstruction can be refined by a few refinement during which the class averages are no more allowed to "move" to another 3-D volume. The number of these refinements can be given here:



7. Like before (chapter 18) the resulting 3-D volume is shown section by section and as surface views for which you have to specify the threshold parameter:



8. Click  to start the 3-D reconstruction. Check the program print-out.

Like before (chapter 13) you can re-do a number of operations wihout re-doing the full calculation:

 Re-calculate 3-D filter, masked 3-D volume and surface views

 Re-calculate the masked 3-D volume and the surface views

 Re-calculate the surface views

9. As before (chapter 18) try to understand the extensive output printed in the terminal window. Various indicators are available to see whether the Euler angle assignment is successful. After one or two iterations, the Euler angles assigned should stabilize (zero angles between current and previous Euler angle assignment).

10. As before (chapter 18) check the results:

    Check "by eye" how well the re-projections match the input class averages (see display tabs on the right-hand side). By flicking between the two tabs compare how well these two match.

    Check the estimated Euler angles (printed in yellow).

    Note: Very often the suggested automatic value is not correct. In such a case you re-calculate the surface views.

    Also visualize the 3-D volumes in **Chimera**.

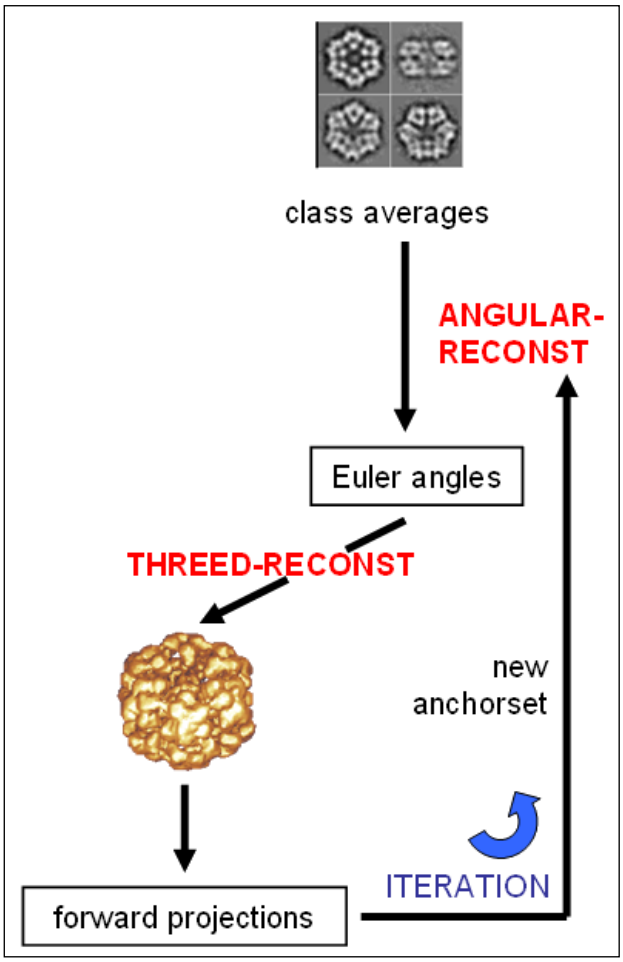


Fig. 31: Iteration of Angular Reconstitution and 3-D Reconstruction

NOTE:

If there is time you can continue with the next chapter (click the "Next" button) or calculate a Multi-Reference Alignment (chapter 21). But probably there is no time any more during the practical. So, enjoy your current 3-D volume(s) ☺ and find out the resolution with the Fourier-Shell-Correlation (chapter 23).

Nevertheless, it is worthwhile to read the refinement steps and iterations in the following chapters.

20. Align Class Averages / Move Particle Images and a new MSA and Classification

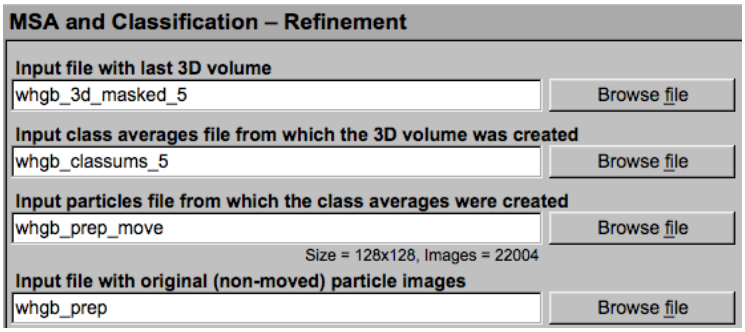
This is an important part of the 4-D Alignment-by-Classification work-flow. The noisy particle images are no more aligned to references (which always is a risk of bias and over-fitting even if you would use the forward projections from the last 3-D volume (like it is done in the "old-fashioned way" in chapter 20)).

Instead of aligning the noisy particle images the class averages are "3-D aligned" to their re-projections during the 3-D reconstruction process. The noisy particle images are subsequently MOVED using the rotation and shift values of the related class average (refer to the lectures).

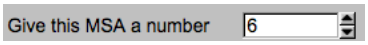
After this "move" a new "**MSA and Classification -Refinement**" is calculated followed by a new "**3-D Reconstruction - Anchorset Refinement**" (chapter 18).

1. Go to the "**MSA and Classification – Refinement**" page.

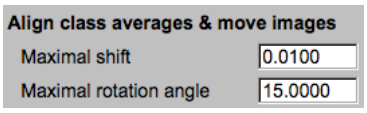
   Input is/are the last 3-D volume(s) created, the class averages from which the 3-D volume was created as well as the last moved and the "original" particles file.



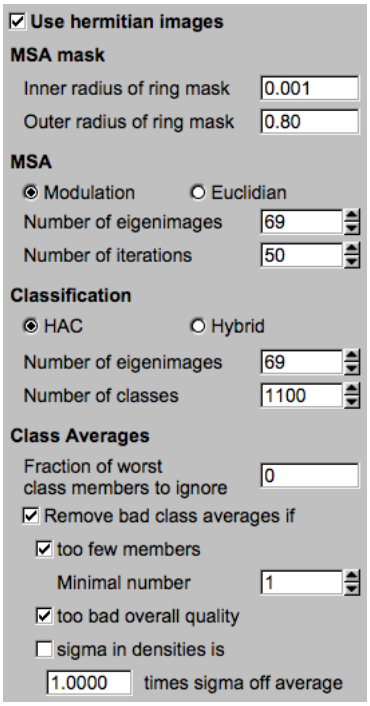
2. To keep track of your image processing give this MSA and classification a number. This will number will be used in the output file names. In the example shown here the new MSA number would be 6.



3. The input class averages used for 3-D reconstruction will be aligned to the related re-projections created during this 3-D reconstruction. The aligned input class averages are used to move the related class member particle images. If wanted, adjust the parameters:



4. A new MSA and classification will be applied to the moved particle images. The parameters are the ones you already know (chapter 16). If wanted you can now use "hermitain images" in the MSA calculations.

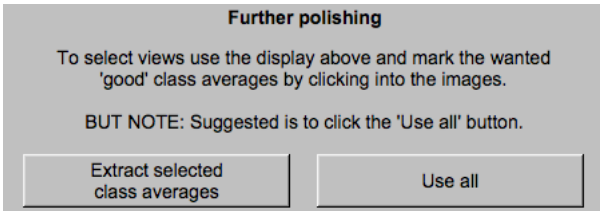


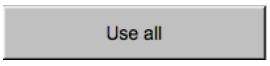
5. First compare the input and the moved particles in the display tabs on the right-hand side.

6. Also check the MSA eigenimages (refer to chapter 16).

7. Of course, also check the new class averages (again, refer to chapter 16).

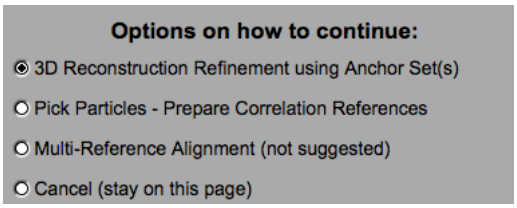
8. Like before (chapter 16) you can extract good / exclude bad micrographs:



Again, suggested is to click the 

button.

9. Clicking the Next ➡ button will give you some options on how to continue:



Usually, you will proceed with "3-D Reconstruction Refinement".

But may be, due to time restrictions you want to play around with alignments. In this case choose the "Multi-Reference Alignment" option – although this is an "old fashioned" appoach (chapter 21).

The option to continue with particles picking is another option used in later refinement circles (chapter 20).

YOUR NOTES:

21. Iteration Cycle(s)

1. The process of MSA and Classification, Euler angles search (angular reconstitution - anchor set refinement), 3-D reconstruction and align class averages / move images can be iterated:

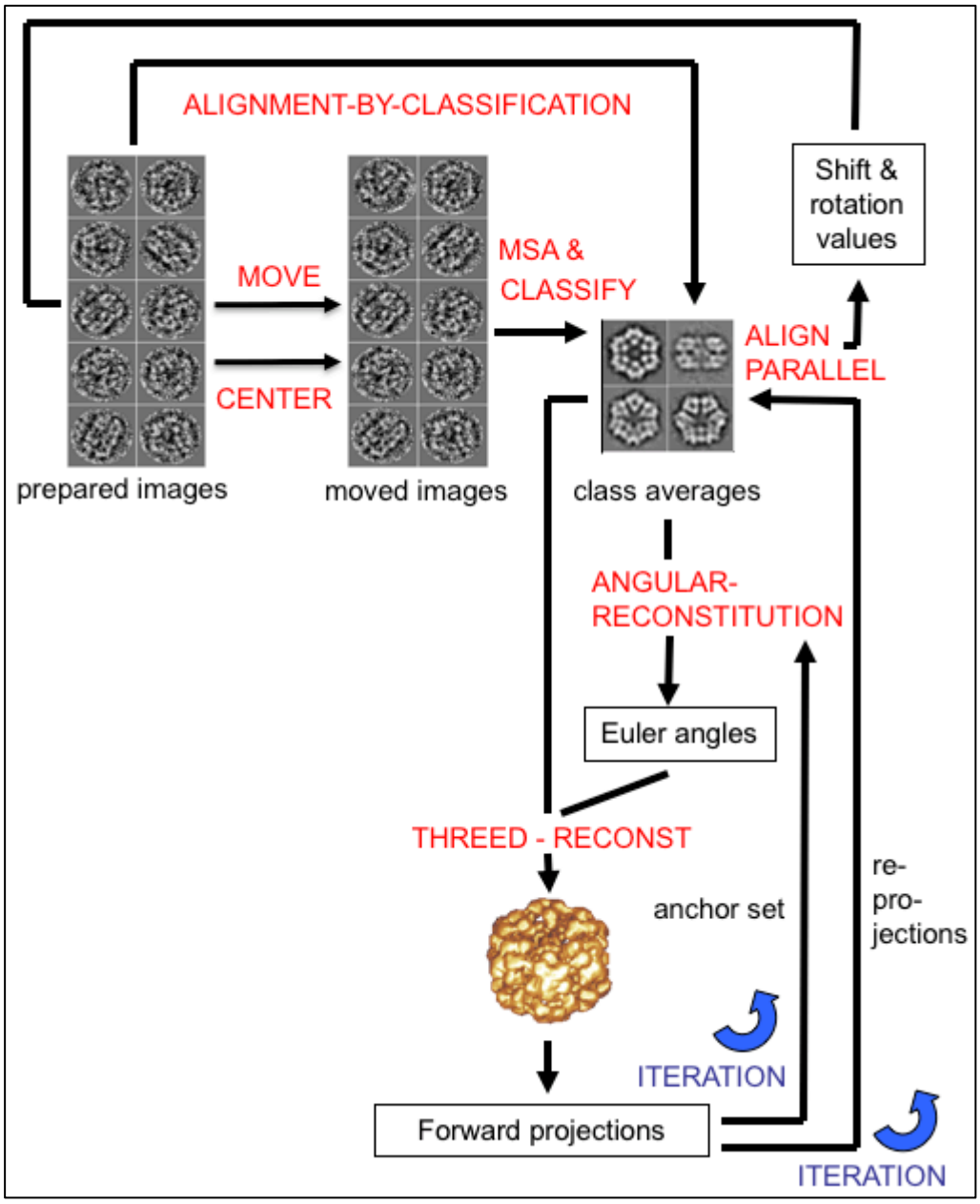


Fig. 32: 4-D Alignment-by-Classification

The refinement loops shown in figure 32 are repeated many times until you reach your desired resolution or convergence. This iterative refinement is the most time-consuming process. As the quality of your reconstruction increases you can use a finer angular increment for forward projecting and in the Euler angles search.

2. Also, one can go back to (correlation) particle picking now using the (better) references created from the latest best 3-D volume and repeat the MSA classification, Euler angles search (angular reconstitution / anchor set) and 3-D reconstruction iterations.

   But make sure that you (strongly) low-pass filter your references before picking to avoid reference bias/over-fitting.

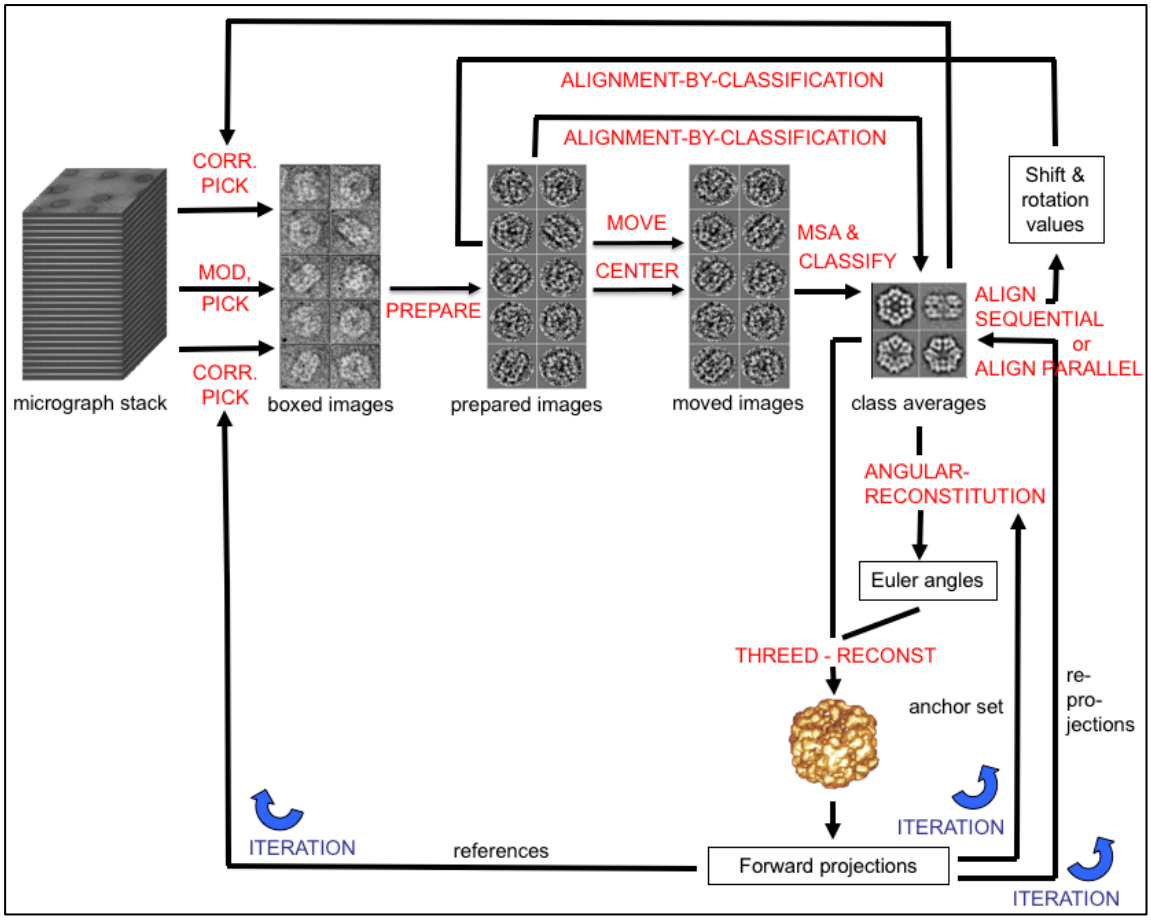


Fig. 33: Re-do Particle Picking / Iterations

3. You can also pick from the un-coarsened data set if the limit of the current sampling is reached and refine your results using these full resolution images.

Of course, due to time restrictions these iterations are again no more part of this hands-on.

22. Multi-Reference-Alignment (MRA) - "old fashioned"

Multi-reference Alignment is no more used in the 4-D ABC workflow because it uses explicit reference images for aligning raw (noisy) images which bears the risk of reference bias and over-fitting. MRA is the "old traditional" way of aligning an image data-set. It has largely been superseded by the 4-D ABC approach (refer to the lectures). But that does not mean that these classical procedures are no longer available.

To keep track of your iterations you can give this alignment a number. As usual, this number will be used to create output file names.

22.1. Get the References for the Multi-Reference-Alignment

To align the particle images you need references. In principle, there are two options to get these references:

- Select the references from class averages:

  Check the displayed class averages and interactively select "good" images with high contrast and showing all typical views. See (2).

  But note that this selection can be a strong bias.

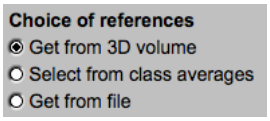
- Generate the references from your last 3-D volume:

  Create forward projections of your last 3-D volume into all directions (of the asymmetric triangle). See (1).

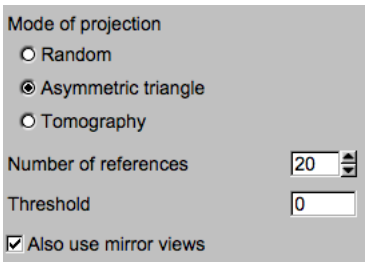
  In contrast to the interactive selection from class averages here you really get all typical views. Another advantage is that the references are perfectly (3-D) aligned to each other.

  So, when using this option do NOT centre or align these references.

1. You have already created (a) 3-D volume(s). Therefore you should choose the second option and create references from your last 3-D volume.



You can specify some parameters, especially the number of references wanted.

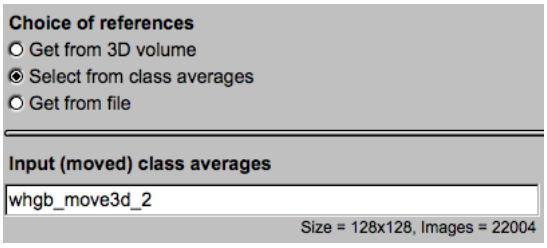


NOTE:

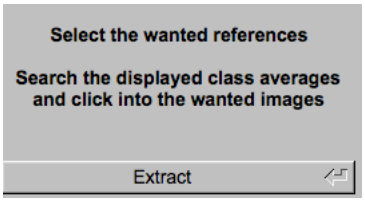
In contrast to the creation of anchor sets we here create a higher number of references and also use the mirror versions.

Click the Generate references button to generate the references.

2. As mentioned (and not suggested here) you can also select the refences from your class averages:

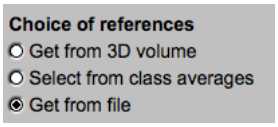


As usual select the references by clicking into the wanted images displayed on the right-hand side.



The references will be extracted when clicking Extract the button.

The "**Multi-Reference Alignment**" page will be re-loaded with the "Get from file" option. Input references are your selected references (whgb_mrarefs_X, where X is an iteration number).



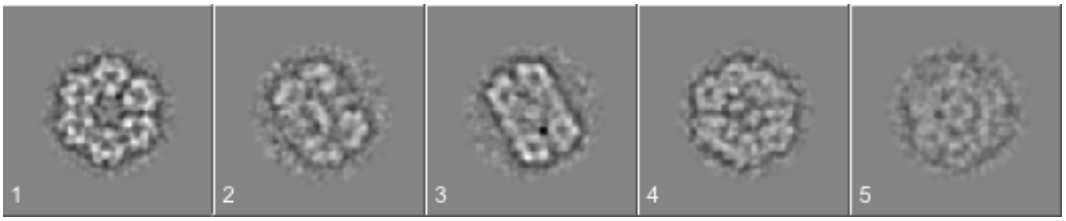


Fig. 34a: Some references chosen from class averages

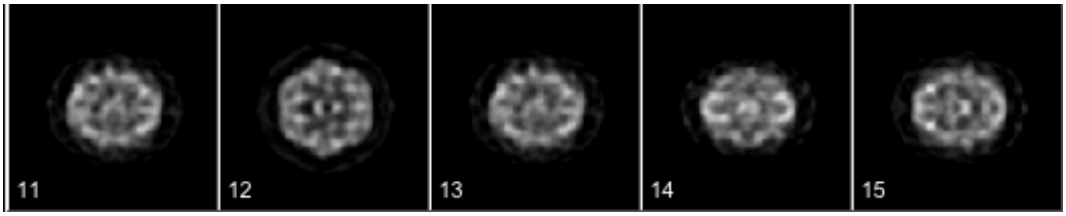


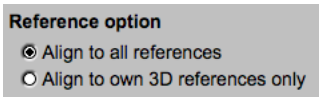
Fig. 34b: Some references created from an initial 3-D volume

22.2. Multi-Reference-Alignment

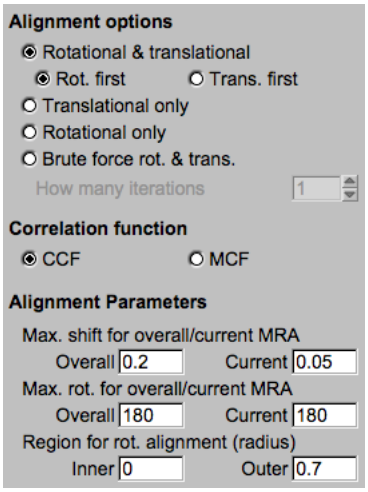
The input particle images will be correlated to all references and finally aligned to the reference to which it fits best.

1. Having extracted / generated references the alignment page will open up. The input reference file is the one with your references. The input particles file is expected to contain the last moved particles (whgb_move3d_X). The "original" non-moved particle images (whgb_prep) are also needed. As was explained in the lecture, in order to avoid interpolation artefacts, **IMAGIC** keeps the original filtered data and continually calculates the "equivalent move" necessary to reach the latest alignment.

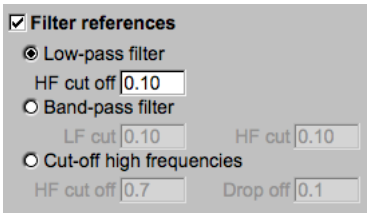
2. If your references were generated from multiple 3-D volumes you can align each particle image either to all references or to its own 3-D volume references only.

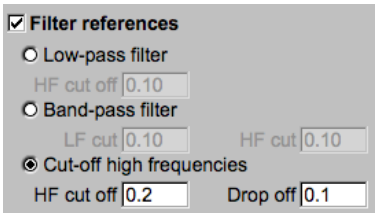


3. Some other parameters can be specified. Do not forget to get help by moving the cursor over the related boxes.



4. Do not forget to strongly suppress the high frequencies during alignments to avoid bias and over-fitting.

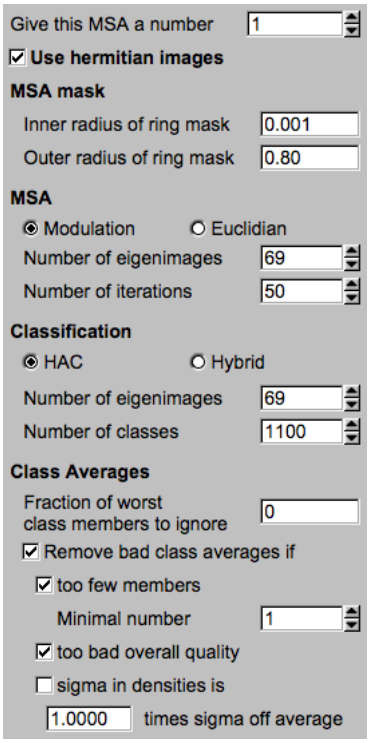




5. Run the alignment and check the printout in the terminal window as well as the aligned images in the display on the right-hand side.

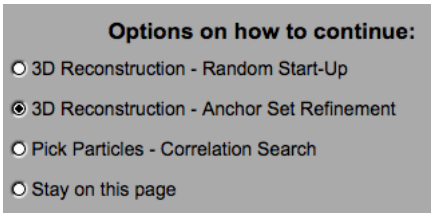
22.3. MSA and Classification of the Aligned Images

1. After the alignment has completed, you will run a MSA and classification on the aligned images (whgb_mra_X) to get new class averages (whgb_classums_X).



   Refer to chapter 16.

2. Clicking the Next button, you are asked on how to continue:



   Usually the next step is a new "3-D Reconstruction - Anchor-Set Refinement".

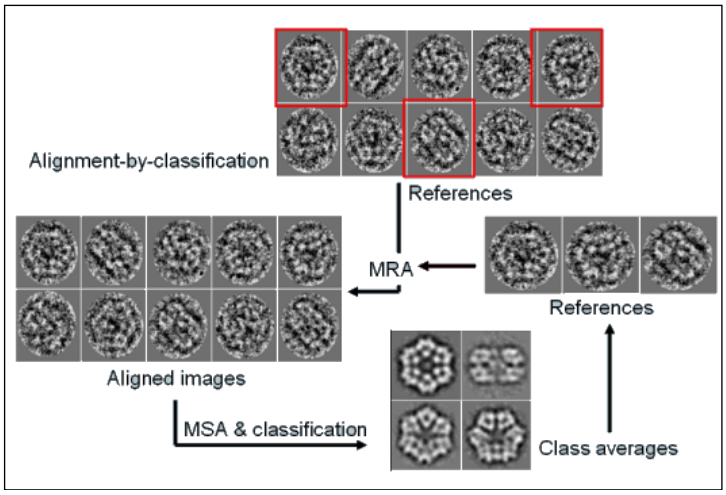


Fig. 35: (2-D) Alignment by Classification using Multi-Reference Alignment

22.4. Iteration Cycle(s)

You can iterate this multi-reference alignment / MSA classification cycle until you feel your class averages are of sufficient quality.

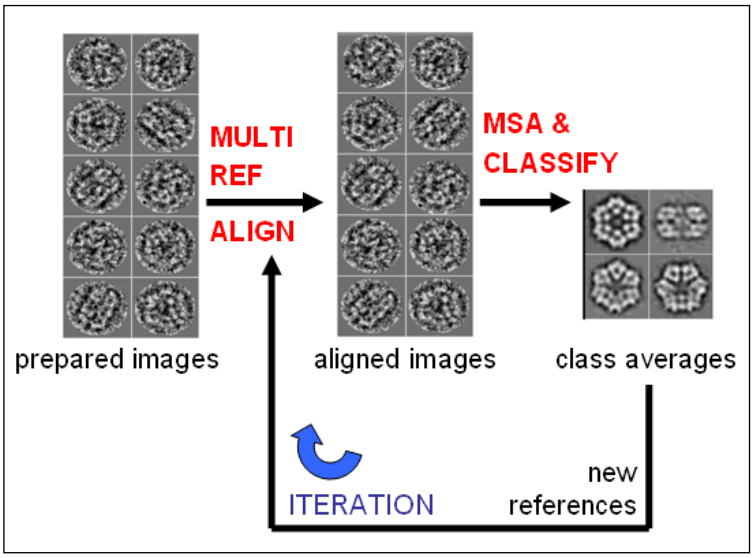


Fig. 36: Iteration of Multi-Reference Alignment and MSA Classification

Again: These iterations are not part of the practical.

You can also iterate multi-reference alignment, MSA classification and the Euler angle search / 3-D reconstruction:

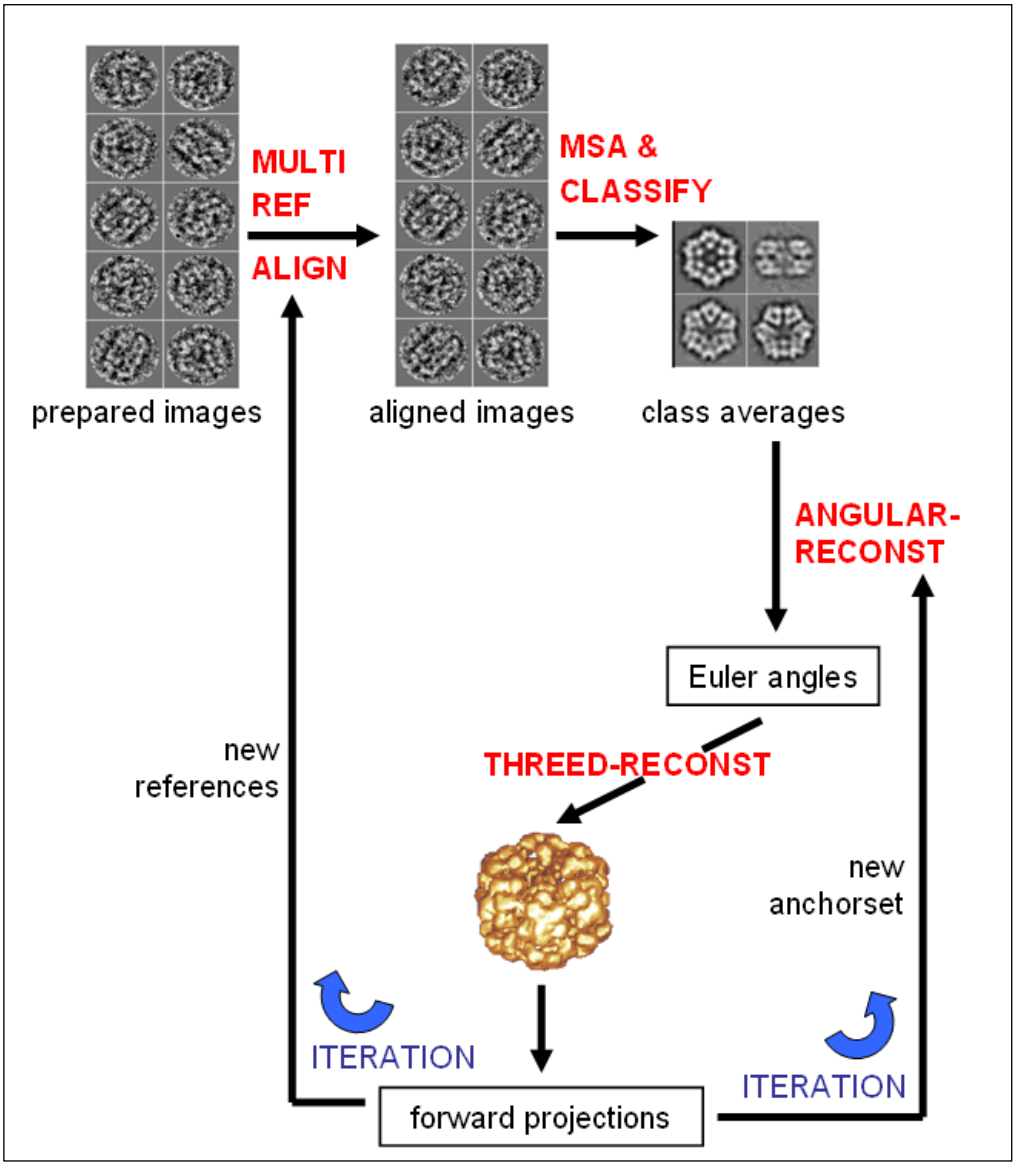


Fig. 37: Iterate Multi-Reference Alignment / MSA and Classification and Angular Reconstitution. / 3-D Reconstruction

This refinement loop is repeated many times until you reach your desired resolution or convergence. As the quality of your reconstruction increases you can use a finer angular increment for forward projecting your alignment references.

Again: These iterations are not part of the practical.

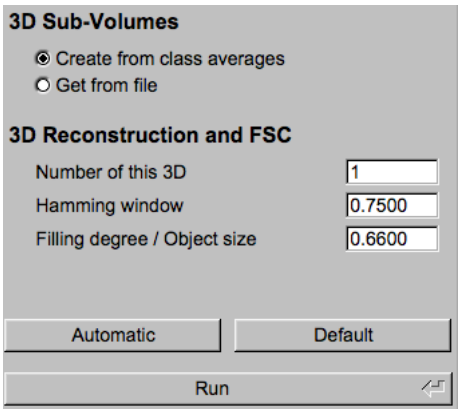
23. Fourier Shell Correlation (Estimate the Resolution)

In nearly all publications the Fourier shell correlation (FSC) is used to estimate the resolution of a 3-D reconstruction.

Remember, that the FSC is not really a resolution measure but a criterion to compare the similarity of two 3-D volumes. If it is used to estimate the resolution of a 3-D reconstruction one has to make sure that the two 3-D subsets do not contain artificial similarities or the same systematic errors. The best approach would be to calculate two 3-D volumes completely independently.

You cannot do this in this practical. To get an idea how the FSC can be calculated and interpreted you will calculate 3-D volumes from two-subsets, which we assume to be the "two independent data sets".

1. Input are the class averages from which you created your last 3-D volume.



2. These class averages will be used to create the two sub-sets needed to calculate the Fourier shell correlation.

IMPORTANT NOTE:

Do NOT use any mask!! Using the same masks in both sub-volumes adds a common feature which correlates very well and pretends a non-existent high resolution.

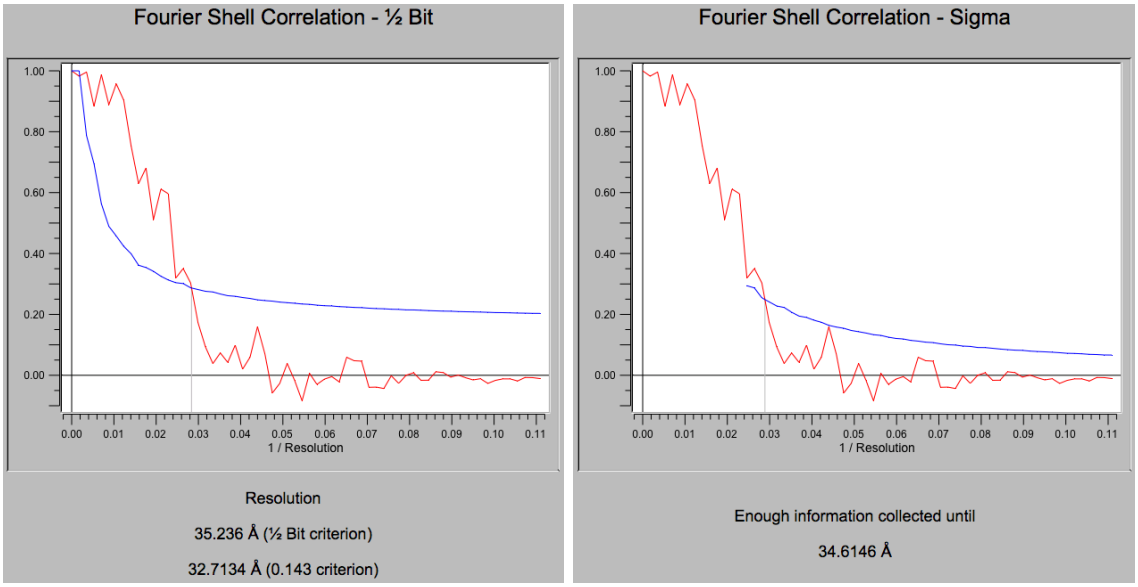


Fig. 38: A Fourier shell correlation plot (1/2-bit and sigma threshold)

The crossing of the Fourier Shell Correlation and the ½-bit information threshold curve (figure 38 - left) expresses where you have already collected a sufficient amount of data in the final 3-D reconstruction to allow a direct structural interpretation at that resolution level. The ½-bit curve is calibrated to approximately yield resolution values comparable to resolution values in use in X-ray crystallography (FOM).

The crossing of the Fourier Shell Correlation and the (modified) 3-sigma curve (figure 38 - right) indicates where the FSC systematically emerges above the expected random correlations of the background noise. This criterion indicates at which spatial frequency you are systematically gaining information significantly above the random noise level. When you continue collecting information by adding more data of the same quality to the data set you would certainly improve the data set up to - and maybe even somewhat beyond - this point.

NOTE:

The FSC is a measure to compare the similarity of two 3-D data sets. If it is used to estimate the resolution of a 3-D reconstruction you have to make sure that the two 3-D subsets do not contain artificial similarities (like masks, for example).

It is good practise not to interpret resolution curves, which are too close to the high end of the resolution curve (the right-hand side of the FSC curve). In other words: you should never claim any resolution level beyond 2/3rd of the Nyquist frequency.

If the sampling size is 4.4 Angstrom per pixel/voxel then the attainable resolution is about 13.2 Angstrom rather than the theoretical Nyquist frequency of 8.8 Angstrom. If the resolution is better than 3x the sampling size your data set is under-sampled and you should re-scan your micrographs with a higher resolution and re-do the image analysis.

Whilst the ½-bit curve provides a single figure for your resolution it is important to always take into account the curve as a whole when judging the quality of the reconstruction.

The best resolution measure is still the resolution of the biological details, which you can see in your 3-D reconstruction.

SOME WEBSITES:

www.ImageScience.de

www.single-particles.org

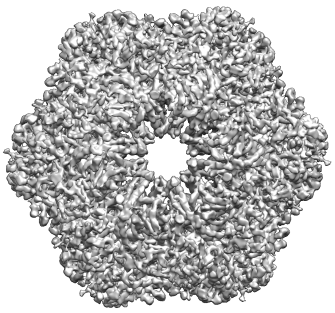
www.brazil-school.org

3dem.ucsd.edu

ERROR HINTS:

We tried to find and correct all errors and typos before, during and after the Brazil School. If you still find some mistakes please send your error hints to michael@ImageScience.de so that we can improve this tutorial. Thank you very much.

YOUR NOTES:



CONTENT