EMBO Practical Course on Image Processing for Cryo EM
1 –11 September 2015

Practical 5: First steps in the analysis of heterogeneity

Introduction

In addition to their arbitrary positions and orientations in the plane of the image projection, the particles may have different out-of-plane orientations, which will give rise to different projections. To sort the images into groups with common orientations, statistical analysis and classification are essential tools in "alignment by classification". Initial class averages selected from a first round of classification can serve as references to bring similar images to the same in-plane position and orientation and to separate different out-of-plane views. A few iterations of these alignment and classification steps provide good averages representing the characteristic views in the data set.
The progress of an alignment can be evaluated by examining the average and variance images. The average of an aligned set of similar images should improve the contrast and visible detail during refinement, and the variance should decrease. In addition, the cross-correlation (CC; maximum value of the normalized CCF) between references and raw images should increase during refinement. The use of statistical analysis and classification of images is important for discriminating variations from any source, differences in defocus, different particle orientations that reflect different 2D projections of a 3D structure, structural variations within an orientation group, and eventually conformational changes of the complexes.

Heterogeneity imposes significant limitations to the achievable resolution, since information from different molecular states in a heterogeneous ensemble will be combined into one reconstruction and hinder the quality of both the reconstruction and the interpretation of the result. In some cases, particular conformations can be trapped biochemically before EM imaging; however, this is not always experimentally possible. So we need to find a method to sort different complexes by computational means. Here, we will focus on the *a priori* method based on double multivariate statistical analysis (MSA) of features in the 2D images to detect image variations that reflect changes both in orientation and conformation. This 2D statistical analysis has been used to separate particles of different size (White et al., 2004) and also with different ligand occupancy (Elad et al., 2007, 2008).

To calculate the 3D map from a set of projection views, the relative orientations of the 2D projections must be determined. We will use in this practical determination of angles using an approach based on the search of common lines in the real space (van Heel, 1987, van Heel et al., 2000). It is based on the theorem that: for any set of 2D projections of a given 3D structure each pair of 2D projections has at least one 1D (line) projection in common. A common lines approach in real space for arbitrary symmetry was developed by van Heel and colleagues and implemented in IMAGIC (van Heel et al., 1996). For each 2D image, a set of 1D projections is calculated and presented as an image (sinogram) whose lines are formed of the series of 1D projections from 0° to 360°. It is important to note that centring of images is essential for angle assignment by common lines, because shifting the 2D image shifts the 1D projections.

When the Euler angles have been assigned, a new 3D map can be calculated and the procedure iterated with the new set of reprojections. It is advisable to make several trials to get an initial 3D reconstruction by angular reconstitution and to check the consistency of results, especially with asymmetric structures. Once a consistent initial 3D map has been obtained, the

structure can then be refined by further cycles of alignment, classification, and common line searching.

The aim of this session is to align your data to specific references, get new classes, to analyse eigen vectors and their meaning, how they can be used to sort out images of different complexes. You will separate images according to their features. Then you will find out the orientation of the classes (characteristic views) to build the first 3D model (in IMAGIC). Finally you will use this model to create references for a new cycle of alignment. You will learn some possible way how to separate images of different complexes.

Good luck!

To read:

1. Baker, T.S., and Cheng, R.H. (1996) A model-based approach for determining orientations of biological macromolecules images by cryoelectron microscopy. *J. Struct. Biol.* 116, 120-130
2. Crowther, .A., DeRosier, D.J., and Klug,A. (1970) The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proc. R.Soc. Lond*, 317, 319-340
3. *Elad*, N., Clare, D.K., Saibil, H.R., and *Orlova*, E.V. (2008) Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. J. Struct. Biol. 162, 108–120.
4. Fuller SD, Butcher SJ, Cheng RH, Baker TS (1996) Three-dimensional reconstruction of icosahedral particles--the uncommon line. *J Struct Biol*. 116, 48-55.
5. Gabor T Herman, Joachim Frank, (2014) Computational Methods for Three-Dimensional Microscopy Reconstruction
6. Mastronarde D., (1997) Dual-Axis Tomography: An Approach with Alignment Methods That Preserve Resolution J Struct Biol. Dec;120:343-352
7. Orlova EV, Saibil HR. (2010) Methods for three-dimensional reconstruction of heterogeneous assemblies. Methods Enzymol.;482:321-41
8. Orlova EV, Saibil HR. (2011) Structural analysis of macromolecular assemblies by electron microscopy. Chem Rev., 111(12):7710-48 (Free article)
9. Penczek, P., Grassuci, R.A., and Frank, J. (1994) The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryoelectron microscopy of biological particles. *Ultramicroscopy,* 53, 251-270
10. Radermacher, M. (1988) The three-dimensional reconstruction of single particles form random and non-random tilt series. *J. Electron Microsc. Tech.* 9, 359-394
11. Serysheva, I., Orlova, E.V., Sherman, M., Chiu, W., Hamilton, S., and van Heel, M., (1995) The skeletal muscle calcium-release channel in its closed state visualized by electron microscopy and angular reconstitution. *Nature Struct. Biol.* , 2, 18-14
12. van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A. (2000) Single-particle electron cryo-microscopy: towards atomic resolution. Q Rev Biophys. Nov;33(4):307-69.
13. van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M. (1996) A new generation of the IMAGIC image processing system. J Struct Biol. Jan-Feb;116(1):17-24
14. Van Heel, M. (1987) Angular reconstitution: a posteriori assignment of projection directions for 3D reconstructions. *Ultramicroscopy,* 21, 114-126
15. White, H. E., Saibil, H. R., Ignatiou, A., and Orlova, E. V. (2004). Recognition and separation of single particles with size variation by statistical analysis of their images. J. Mol. Biol. 336, 453–460.

Login to the server using the instructions provided previously. Open a terminal window and in your home directory type

> cd PRAC-5-6

You will find a set of files with the extension **job*.b :** these are scripts that will be used in the practical.

Files:
E1_data_or.hed and **E1_data_or**.img -> *model data for the analysis*
E1_data_a3.hed and **E1_data_a3**.img -> *aligned model data for the analysis*
msa-mask.hed and **msa-mask**.img -> *mask for the statistical analysis of 2D images*
model_st.hed and **model_st**.img -> *starting 3D model at low resolution*
ref_set_ini_a1.hed and **ref_set_ini_a1**.img -> *set of references for the M-R-A*
anch*.hed and **anch***.img -> *different anchor sets for determination of angle orientations*
and some other files necessary for running jobs smoothly.

You can find **DISPLAY** options in general Introduction **(GI**), pages 6-8.

Centering of the data set:

You will start processing of the data set E1_data_or. You can see what is it using DISPLAY command. To start processing please use the script :

job1_ali_center.b (see **mo-ac** options in **GI** pages 14-15)

What is in the script? See below, comments that are actually not in the script are indicated by the arrow and written in italic. “echo” is a command to display a line of text or a variable value on the screen.

echo "! "
echo "! IMAGIC program: summer ----------"
echo "! "
/s/emib/s/imagic/150710/incore/summer.e <<EOF
TOTAL_SUM *-> summation of all images in the file **E1_data_or,** the file has 3000 images*
E1_data_or *-> input file that contains all your images that you are going to process*
sum_1 *-> output file = sum of all images in the file*
none
EOF
echo "! "
echo "! IMAGIC program: rotatrim ----------"
echo "! "
/s/emib/s/imagic/150710/stand/rotatrim.e <<EOF
sum_1 *-> input file, the result of the previous step*
IMAGE *-> type of the information*
sum_1_aver *-> output file, rotationally averaged total sum of images*
EOF
echo "! "
echo "! IMAGIC program: alidir ----------"
echo "! "
/s/emib/s/imagic/150710/openmpi/bin/mpirun -np 5 -x IMAGIC_BATCH *(the next line is the continuation)* **/s/emib/s/imagic/150710/align/alidir.e_mpi <<EOF**
IMAGES *-> type of the information for processing*

TRANS *-> type of alignment: only translational*
CCF *-> what will be used: cross correlation (not the mutual correlation, that is used typically for images taken in negative stain)*
0.2 *-> allowing shift within 20% of a half size of the box*
E1_data_or *-> input file with the images before alignment*
E1_data_a1 *-> output file with aligned images*
INP *-> source of the reference, it will come from a file indicated on the next line*
sum_1_aver,1,1 *-> input file and location of the image that will be used as a reference*
1 *-> # of the reference*
NO *-> modification of the reference during iterative procedure of alignment*
LOW *-> filtering of the reference: low pass*
0.3 *-> high frequency cut-off*
YES *-> if iterative refinement of the reference is necessary*
3 *-> number of iteration*
TOTAL_AVERAGE *-> summing option for the reference*
0.8 *-> over-correction factor*
0.01 *-> threshold to stop reference refinement*
NO *-> store a new reference*
YES *-> full output of alignment results that you can check later in the log file*
YES *-> use or not MPI processors*
5 *-> # of MPI processors to be used*
EOF

Run the job as it has been described in the general introduction:

./job1_ali_center.b > & job1_ali_center.log &

Takes ~ 1 minute

The file **job1_ali_center.log** is the output file with the information on the job and results. To check if the job is running please type the following command:

tail –200 job1_ali_center.log

you will see on the screen the last 200 lines of the information of the last processing steps written in the log file. If the script has crashed, the error message will be written and seen on the screen. If this happens you have to correct the error in the script and usually it will be necessary to delete this log file (the command: **rm job.log**) or use a new name for the log file. If you want to see more details use the command;

nedit job1_ali_center.log

After centering of images we need to find characteristic views of our sample. For that you have to first run the multivariate statistical analysis. The outputfilesfrom this jobare usedas input files in the next stps of MSA

Please use the script **:**

./job2_msa_r1.b > & job2_msa_r1.log &

takes ~ 4 minutes

```
echo "! "
echo "! IMAGIC program: msa ----------------------------------------------"
```

```
echo "! "
/s/emib/s/imagic/150710/openmpi/bin/mpirun -np 5 -x IMAGIC_BATCH (the line continues)
/s/emib/s/imagic/150710/msa/msa.e_mpi <<EOF
YES                 -> use or not MPI processors
5                   -> # of  MPI processors to be used
NO                  -> NO_LOCAL_FILES will be used
FRESH_MSA
                    MSA distances:
                    EUCLIDIAN,  CHISQUARE,  MODULATION
MODULATION
E1_data_a1          -> input file with images that were centered in the previous job
msa-mask            -> input file, was loaded with the data set. It is just a circle (see general
                       introduction)
eig_r1              -> output file with eigen images, that we have to look at later
NO                  -> we are not using the default parameters
50                  -> # of MSA iterations
30                  -> # of eigenvectors to be used
3                   -> random number that will be used to start calculations of eigen vectors
0.8                 -> overcorrection factor (if it is small then the accuracy of the calculation
                       will be low, if it will be close to 1 (but should be less as 1.00) then the
                       accuracy will be slightly better but the calculation will be much longer)
msa_r1              -> output file with information on statistical analysis
EOF
echo "! "
echo "! IMAGIC program: classify ----------------------------------------------"
echo "! "
/s/emib/s/imagic/150710/msa/classify.e <<EOF
IMAGES
E1_data_a1          -> input file
0                   -> percentage of bad images that could be ignored
30                  -> # of eigenvectors that will be used in classification
YES                 -> to use default options (the # eigenvectors indicated above)
150                 -> # of classes you want
msa_r1_cl_150       -> a set of files with information on the classes and is used for
                              classification
EOF
echo "! "
echo "! IMAGIC program: classum ----------------------------------------------"
echo "! "
/s/emib/s/imagic/150710/msa/classum.e <<EOF
E1_data_a1          -> input aligned file
msa_r1_cl_150       -> input file form the previous step of classification
clsum_r1_150        -> output file with class sums
NO                  -> down weight small classes
0                   -> # percentage of bad images that could be ignored
NONE                -> summing statistic
EOF
```

It would be useful if you will look at the eigen images using the command **DISPLAY.** You will find some differences in characteristic views (classes).

Multi Reference Alignment – MRA and MSA

Usually MRA has to be performed after centring of images using translational alignment.

For the multi-reference alignment we have to prepare references from class-averages that represent characteristic **DIFFERENT** views of the molecule. Display the file **"clsum_r1_150"** using the command **DISPLAY** (see the general introduction)

Takes ~ 10 min

You should go to the line "NEXT STEP" on the next page since the following job is an optional task, but it will give you an idea how references can be prepared:

With the command **display** have a look at the classes and with the option **select** make a list of **5-6 images** that can be used for the next round of **multi-reference-alignment.**

Display the file with classes **clsum_r1_150 using DISPLAY and do the following:**

```
Parameters to be changed:
NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO]       : select

 Chosen OPTION is             : SELECT

Output (PLT) file for loc#s (NO ext.)
[clsum_r1_150]                                     : ref_set_in
m**NOTE: Please select image locations by clicking into the
        wanted image on the screen

        Red box  : selected
        Black box: removed

        Continue selecting images with CONTROL click
        Quit selecting images with SHIFT-CONTROL click
        Move the cursor to the point of interest.
        Press the right mouse button for zoom and release to select the
point
```

or, use the middle mouse button to select the point without zooming. In the bottom of the window with images you will find two small boxes: a green with "next" and a red with "stop" , that make procedure od selection much easier.

Numbers of images selected for the first round of **M-R-A and MSA** will be written in the plt file (you have to give a file name before selection images e.g. **"ref_set_in.plt**") All references should be in one new file:

```
IMAGIC-COMMAND : exc-copy (EXCLUSIVE-COPY)
Please specify option                    2D_IMAGES/SECTIONS
Mode of copy operation
Please specify option                    EXTRACT
Input file, NO loc#s                     clsum_r1_150
Output file, image loc#s                 ref_set
Get image loc#s from
Please specify option                    plt
PLt file name                            ref_set_in.plt
Numbers wanted                           ALL
```

Then references have to be prepared: as they have to be aligned to have similar orientations. It is recommended to orientate images so that the rotational symmetry will be along the X-axis, which is the vertical one in Imagic. Typically you have to find angles yourself by rotating the image and displaying the rotated image. You can do it using the command **ROT-IM**. For all

references the centre of the image mass (intensities) has to be in the centre of the image frame. Use the command **ali-mass (ALIGNMENT-MASS-OF-IMAGES)** for that**.** Multi-reference alignment runs in a more stable way if all references are normalized to the same SIGMA (**σ** = standard deviation of densities, **NORM-VAR**). We recommend applying a circular mask with **a soft edge** to mask out the noisy background that surrounds the characteristic views.

That may take up to 30-45 minutes with many questions.

"NEXT STEP"

To make your life easier please use the file **ref_set_ini_a1** for your first round of **M-R-A** that contains pre-prepared references. You can look at them using **Display** command.

To run the job type:
./ job3_ref_ali_msa2.b > & job3_ref_ali_msa2.b &

-> takes ~ 15 minutes

Script **job3_ref_ali_msa2.b**

echo "! "
echo "! "
echo "! IMAGIC program: mralign ----------"
echo "! "
/s/emib/s/imagic/150710/openmpi/bin/mpirun -np 5 -x IMAGIC_BATCH *(the line continues)*
/s/emib/s/imagic/150710/align/mralign.e_mpi <<EOF
YES *-> MPI processors to use/not*
5 *-> # of MPI processors to be used*
FRESH
ALL_REFERENCES
ALIGNMENT
BOTH (ROT AND TRANS)
TRANS
CCF *-> MCF (Mutual cross correlation) option is recommended for images taken in negative stain, because they have a significant amount of low frequencies (information of the molecular shape). The option **CCF** (cross-correlation) is recommended for images such as those in ice.*
E1_data_a1 *-> input file of the previous round of alignment*
E1_data_a2 *-> output file of the current round of alignment*
E1_data_or *-> input file, the original without any alignment. The original band-pass filtered file is required to perform the final rotation with only one interpolation of the original image. Otherwise multiple interpolations during comparison with different references will lead to the loss of high-resolution information.*
ref_set_ini_a1 *-> input file with pre-prepared references*
NO *-> input file*
0.15 *-> maximum shift with respect to the original images. The restriction should be applied to avoid large displacements of images that sometimes take place because of noise.*
0.07 *-> maximum shift during the current alignment, the next step of alignment should not be very big compared to original shifts. This allows us to refine the position of the particle image.*
-180,180 *-> range of angles with respect to the original images*

-180,180 *-> range of angles during current alignment (on the following steps it can be limited to a smaller range*
HIGH *-> this parameter defines a step size to search the best rotational alignment in polar coordinates. The smaller step corresponds to the highest precision*
0.01,0.7 *-> This option is related with presentation of images in polar coordinates. It is very rare in polar coordinates, that densities near to the origin of the coordinate system reflect features of the structure, therefore it is recommended to slightly step away from the centre. The outer radius depends on the size of the object within the image frame. Here the main features will be restricted by ~ 2/3 of the radius of the frame. If the object (molecule) has bigger size within the frame, this parameter has to be increased.*
12 *-> number of iterations, strongly recommended a number > 10.*
YES *-> full output of all parameters, you can see them in the log file.*
EOF
echo "! "
echo "! IMAGIC program: msa -----" *(see comments on pages 4 and 5: first MSA after centring)*
echo "! "
/s/emib/s/imagic/150710/openmpi/bin/mpirun -np 5 -x IMAGIC_BATCH *(the line continues)*
/s/emib/s/imagic/150710/msa/msa.e_mpi <<EOF
YES
5
NO
FRESH_MSA
MODULATION
E1_data_a2 **->** *File obtained after a round of MRA*
msa_mask **->** *as it was loaded and previously used in the first MSA round*
eig_r2 **->** *output file with eigen images*
NO
50
40
5
0.8
msa_r2
EOF
echo "! "
echo "! IMAGIC program: classify ----------"
echo "! "
/s/emib/s/imagic/150710/msa/classify.e <<EOF
IMAGES
E1_data_a2
0
40
YES
100
msa_r2_cl_100 **->** *a set of files with information on the classes*
EOF
echo "! "
echo "! IMAGIC program: classum ----------"
echo "! "
/s/emib/s/imagic/150710/msa/classum.e <<EOF
E1_data_a2
msa_r2_cl_100

clsum_r2_100 *-> output*
NO
0
NONE
EOF

Display the file containing the class averages "**clsum_r2_100"** using the command **DISPLAY**. Please compare your visual estimation of class quality with the evaluation given in the file **msa_r2_cl_100.lis**. In this file information on the quality of the classes is given in tables at the very end of the file. These tables provide the number of images per class, the average distance between the class members, the distances between classes, and the overall quality of the classes. To look at the LIS file you have to get out of IMAGIC by typing a symbol "*****" and then type the UNIX command **nedit msa_r2_cl_100.lis &**. The file will be open in a new window.
Moreover, you will find by the visual inspection that some characteristic views in **clsum_r2_100** have an extra blob and some other not.
After that please check images in **eig_r2,** you will see that there is strong variations in densities on the top of the main shape of the particles indicating on not complete occupancy of the extra domain on the complex.

Takes ~ 15 min

Separation of images

To understand if the structures are different one approach albeit not the only one even in **IMAGIC** is to select images that form different types of characteristic views.
The characteristic views with different features were selected in advance using **DISPLAY** option **SELECT.** The selection has been done in advance and the list of picked images is saved in the files: **extra_den_a2.plt** and **no_den_a2.plt**
To run the job type:

./ job4_separ_sub_msa_a2.b > & job4_separ_sub_msa_a2.log &

-> takes ~ 4 minutes

See the script using **nedit job4_separ_sub_msa_a2.b**
This job includes:

1. using command **MSA-EXTR** extracting images that form classes with the extra density using **PLT** file **extra_den.plt**. Images will be extracted from the file **E1_data_a2** with the aligned images using results of classification written in the file **msa_r2_cl_100. cls**.
2. using command **MSA-EXTR** extracting images that form classes with NO extra density using **PLT** file **no_den.plt**. Images will be extracted from the file **E1_data_a2** with the aligned images using results of classification written in the file **msa_r2_cl_100. cls**.
3. As output you will get subdata sets: **set_extra_den_a2** (with the extra blob) and **set_no_den_a2** (no blob)
4. For each subset we will rerun the **MSA** analysis and will get new classes to check how well the data were separated. In this step we will increase the number of classes during analysis of each subset to 300. Outputs will be in the files **clsum_extra_den_a2_300** and **clsum_no_den_a2_300**
5. Information on classification will be written in files **cl_extra_den_a2_300.lis** and **cl_no_den_a2_300.lis**

Display the file containing class averages **clsum_extra_den_a2_300** using the command **DISPLAY**. Then check the file **clsum_no_den_a2_300.** You will find that the first data set is

more homogenous compared to the second one. However some number of classes with the extra blob still can be found in the second subset.

-> takes ~ 6 minutes

Angular reconstitution and three-dimentional (3D) reconstructions

The approach to search for the angular orientation is based on common lines between the views obtained during the previous step of the practical. Typically to start this process one has to start at least with three different or more different views such as an end view, a side view that may have orientations close to **90**˚ relative to Z axis (β angle) and another side view with beta in a range of 60-120 degrees and a significant difference in gamma. In this case one has use an option C1 start and provide the file where the location of these particular views will be found (as locations at the input). The other views should be sequentially added and their angles will be determined with respect to the previous set of images. However this procedure takes time and a bit of thinking. One has to see if the program has provided a reasonable solution. Some errors can arise due to noise in the images (if there are few images in that class), bad alignment and poor centring.

To save time we will use an initial model to start the structural analysis. In this case we will use a low resolution map **model_st,** from whicha set of projections **anch_set_0** has been calculated. Display this set of projections using **DISPLAY.**

Run the command:

./ job5_euler_3d_a2.b > & job5_euler_3d_a2.log &
-> takes ~ 30 minutes (you can have a coffee break)

See the script using **nedit job5_euler_3d_a2.b**
This batch job does the following:

1. using the command **THREE-FORW** to calculate projections from **model_st** to get **anch_set_0**
2. using the command **EULER** to calculate angular orientations for the data set **clsum_extra_den_a2_300** (classes from images with an extra blob)

```
Point-group symmetry:
  C1          1          C2          2
  C3          3          C4          4
 ……
 ……
Please specify option [ICOSAHEDRAL]              : C1-> asymmetrical
object
Option for angular reconstitution:
NEW_IMAGE, ANCHOR_SET, C1_STARTUP, SELF_SEARCH, REFINE_EULERS_SET,
SINOGRAM, SINE_CORRELATION, PREDICT_SINECORR_PEAKS, RANDOM_STARTUP
 Please specify option [NEW_IMAGE]               : anch_set (here we
                                                   identify the option
                                                   that will be used; we
                                                   are using a set of
                                                   different projections
                                                   from the low passed
                                                   model to start our
                                                   structural analysis
```

```
Option of ANCHOR_SET:
  FRESH  REFINE
Please specify option [FRESH]                    : FRESH

Anchor set options:
  SINGLE_ANCHOR_SET        OWN_ANCHOR_SET
  EACH_TO_BEST_ANCHOR_SET  ALL_TO_BEST_ANCHOR_SET
Please specify option [SINGLE_ANCHOR_SET]        : SINGLE_ANCHOR_SET

How are the input (classum) images available:
  IMAGES  SINOGRAMS
Please specify option [IMAGES]                   : IMAGES
Input(=output) (classum) images, NO loc#s
[my_classum]                                     :
clsum_extra_den_a2_300
Sinogram file, NO locations [my_sino]            : sino_extra
ASQ filter the sinogram lines [NO]               : yes
Linear mask radius for sinograms [.9]            : 0.75

How is the anchor set available:
  IMAGES  SINOGRAMS
Please specify option [IMAGES]                   :
Input anchor set IMAGES, NO loc#s [my_arset]     : anch_set_0
Output anchor set sinograms, NO loc#s [my_arsino]: anch_sino
Output sinecorr file, NO loc#s [my_sinecorr]     : sinecorr_exd
Delete output sinecorr file(s) [YES]             :
Wanted angular increment in search [5.0]         : 4.0

Choose criterion for peak search:
  FISHER_TRANSFORM  PEAK_OVER_STDV  NONE
Please specify option [FISHER_TRANSFORM]         : peak

          NONE           :  A conventional peak search is launched in
                            which the max sum of all (symmetry)
                            related peaks with respect to the anchor
                            set is the search criterion

        PEAK_OVER_STDV   :  The peak criterion is the same as in NONE
                            above, but now the max sum is divided by
                            standard deviation (STDV) among the
                            (symmetry) related peaks with respect to
                            the anchor set

        FISHER_TRANSFORM :  Correlation values (ranging from -1 to + 1)
                            are 'Fisher transformed'(see: Wikipedia)
                            to avoid the quenching of the variances
                            against the top value of "1".

Use MPI parallelisation [NO]                     : yes
Number of processors to be used [5]              : 5

Output of results:
  FINAL_OUTPUT  SHORT_OUTPUT
Please specify option [FINAL_OUTPUT]             : FINAL_OUTPUT
Printout results histograms [YES]                : YES
```

3. after this with **EULER** command it calculates angular orientations for the data set **clsum_no_den_a2_300** (classes from images without an extra blob)

4. using the command **TRUE-THREE** a 3D reconstruction from all 300 classes is calculated from **clsum_extra_den_a2_300**

```
  MPI parallelisation:
    ONLY_3D  BOTH  NO_MPI
  Please specify option [NO_MPI]                  : both
  Number of processors to be used [3]             : 5

  Mode of 4D operation:
    ALL_IN_ONE_3D, 3D_MEMBERSHIP_IN_HEADER, RANDOM_3D_MEMBERSHIP
    SEQUENTIAL_ASSIGNMENT, MULTIPLE_RAND_ASSIGN, HEADER_3D_MEMBERSHIP,
    SPLIT_3D_MEMBERSHIP, FOURIER_SHELL_CORRELATION

  Please specify option [3D_MEMBERSHIP]              : all

  Point-group symmetry to be used:
    C1           1           C2           2
    C3           3           C4           4
    C5           5           C6           6
    And many others
  Please specify option [C1]                         : C1
  Use default 3D reconstruction options [YES]        :
  Input 2D (class-sum) images, NO loc#s [my_classum] :
clsum_extra_den_a2_300

  Source of Euler angles:
    ANGREC_HEADER        MRA_HEADER
    RANDOM_EULER_ANGLES
  Please specify option [ANGREC_HEADER_VALUES]       :
  Update input headers (3D references) [YES]         :

  Output file for 3D reconstructions, NO loc#s
  [my_3d]                                            : 3d_extra_den_a2_300
  Output file for re-projection images, NO loc#s
  [my_repro]                                         : repr_exd
  Output file for error projection images, NO loc#
  [my_err]                                           : err

  Spherically mask the reconstruction [YES]          :
  Radius of the mask [0.8]                           : 0.75

  Also create a normalized 3D volume [NO]            :

  Hamming window factor [0.75]                       : 0.999
  Object size as fraction of image size [0.8]        : 0.7
```

5. using the command **EXC-COPY** class averages in **clsum_extra_den_a2_300** will be sorted according to errors with reprojections and the best 50 (with the lowest errors) are extracted into the file **clsum_extra_den_a2_300_50**
6. using the command **TRUE-THREE** the next 3D reconstruction from 50 classes is calculated using **clsum_extra_den__a2_300_50**
7. using the command **TRUE-THREE** a 3D reconstruction from all 300 classes is calculated from **clsum_no_den_a2_300**

8. using the command **EXC-COPY** class averages in **clsum_no_den_a2_300** will be sorted according to errors with reprojections and the best 100 (with the lowest errors) are extracted into the file **clsum_no_den_a2_300_100**
9. using the command **TRUE-THREE** the next 3D reconstruction from 100 classes is calculated using **clsum_no_den_a2_300_100**

The most interesting for us results will be in files: **3d_extra_den_a2_300_50** and **3d_no_den_a2_300_100.** You can look at them using the command **DISPLAY**. Compare sections
112, 160 in both reconstructions, in the structure 3d_extra_den_a2_300_100 you will see an extra densities locates in sections 140-156 (use the option **LOCATION** in **DISPLAY**)

You can visualize the distribution of the Euler angles by going out of IMAGIC and typing:
X_euler clsum_extra_den_a2_300.hed 3 3 (keep in mind that here the extension "hed" is important.

3D masking

To do masking of your 3d use the command **thr-aut-mask (THREED-AUTOMATIC-MASKING**)

```
IMAGIC-COMMAND : thr-auto-mask

What should be masked                                  : 3D_VOLUME

Auto-masking options:
  DO_IT_ALL  REFINE_THRESHOLD
Please specify option [DO_IT_ALL]                      : DO_IT_ALL
Input 3D volume file [3d_no_den_a2_300_50]             : 3d_extra_den_a2_300_50
Output file containing masked input 3D
[3d_no_den_a2_300_50m]                                 : 3d_extra_den_a2_300_50m
Output modulation/variance volume [my_mod_varian]      :
Output file with binary mask [3d_mask_extra_den_1]     : 3d_mask_extra_den_1
Masking based on local modulation [YES]                :

 To find the mask the 3D volume will be band-pass filtered.
Please specify the band-pass parameters

Low, high frequency cutoff [0.06,0.2]                  : 0.06,0.2 (it is
important                                                          to keep
these                                                              parameters.
For fun                                                            you can
change
      parameters but
            change the output file
                  names and have a look at
                                                              the results using
                                                                    DISPLAY)

 Specify the size of the local area over which the modulation/
 variance is calculated by the following low pass parameter

Low-pass parameter (pixels/fraction) [0.04]            : 0.04

Threshold options:
```

```
  MANUAL  AUTOMATIC
Please specify option [AUTOMATIC]                  : automatic
Auto-threshold percentage [5]                      : 5

 Follow-up operations include:

 THREED-AUTO-MASKING       : to refine auto masking parameters
 THREED-SURFACE            : to view the cleaned 3D reconstruction results
 EM2EM                     : to export results to other programs like
CHIMERA
 THREED-FORWARD-PROJECTION : to generate references for M-R-A
                             to generate an anchor-set for ANGULAR-RECONST
                             to generate masks to impose on 2D images
                             (input for THREED-RECONST)
```

Please check results of masking using **DISPLAY**. Please mask reconstruction **3d_no_den_a2_300_100** with the same parameters**.**

See the workflow chart of the practical on the next page.

Workflow of practical 5