Basic image processing course, ReNaFoBis Oléron school, 7.6.2015, B. Klaholz

Useful hints:
To get a screenshot from your displayed images: Maj cmd 4 → selection; Maj cmd 3 → full screen
Capture screen shots under Mac: Control-Shift-4 → select region of interest with mouse
Control-Shift-3 → full screen shot
Display in imagic: use " * " or "quit" to leave, NOT CTRL C;
Use CTRL Z and bg / fg to create batch job or bring it back

To start Imagic:
i **[shortcut for starting IMAGIC program]**

To start a display:
disp [under linux or within IMAGIC]
Input image file, loc#s [checker_8] : **file_name**
Size of the display window [600,600] : **[hit return for default]**
Type of cursor:
CROSS SQUARE CIRCLE
Please specify option [CROSS] : **[hit return for default]**
Parameters to be changed:
NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO] : **[hit return for default]**
To adjust scale of display:
scale
4

Other useful options in command window of display:
grey [to adjust the dynamic range of the image]
interactive
0,0 [full range] or for example -10,10 [limited range]
file [read in another file]
filename
dev [device, size of display window]
600,1200
erase [removes displayed image, to display freshly another one]
profile [to make profile]
Use cursor to position profile: NO
Starting point (IMAGE coordinates X,Y): 1,1
End point (IMAGE coordinates X,Y): 65,65 [center of a 128,128 image, i.e. center of powerspectrum]

To start a second display:
Ctrl Z
bg [background, batch job]
To quit the display:
*** [or] quit very import!!! (otherwise display problems)**

reactivate a background job: fg [foreground], then stop it with * or quit

When interpreting the results, consider that the absolute scales on the y-axis can be different!

II. Pre-processing:

a) Display a digitized micrograph / negative or CMOS camera image of single particles imaged by cryo-EM

In your team directory:

cp ../micrograph/* .
boxer &

read in one of the files called **1.mrc**:
File→ read Micrograph **do you see anything?**
Process→ Median Filter 5x5 (makes a block convolution) **now better?**
adjust grey values/contrast: **middle mouse button**
change scale to **0.4**
change box size to 80 or 128 [adjust box size to the particle size: should be ~2/3 of the image size; will be smaller than 128 if you use the **1.mrc** image from the CMOS camera; ideally, values of the type of 64, 96, 128, 256, 512, 1024, 2048, 4096 etc. should be used for fast Fourier transform calculations)

select ~6-10 particles
Boxes→ Autobox, adjust parameters for a reasonable selection and let it select automatically; afterwards, deselect some bad images manually

b) Calculate a power-spectrum:

ctfit
→ Open particle set
File name: 1_ptcl
Adjust grey values to see the power-spectrum better, adjust parameters to make the predicted spectrum fit with experimental spectrum, adjust defocus value to make the high-resolution peaks fit (not the first peak and first zero which contain information from the particle itself, e.g. secondary structure elements)
Compare with power-spectra from other defocus values: file names: 7_ptcl 9_ptcl 10_ptcl

III. Processing of real experimental data

The basic steps of a structure determination of single particles:

a) pre-process the data: bandpass-filter
b) centering / alignment
c) multivariate statistical analysis (MSA) and classification
d) angle assignment
e) 3D reconstruction

In your team directory:
cp ../data/* .

1) display particle set:

disp [under linux or within IMAGIC]
Input image file, loc#s [] : **CMOS_ctf-append_500**
Size of the display window [600,600] : **[hit return for default]**
Type of cursor:
CROSS SQUARE CIRCLE
Please specify option [CROSS] : **[hit return for default]**
Parameters to be changed:
NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO] : **[hit return for default]**

To start a second display:
Ctrl Z

2) bandpass-filter
i **[shortcut for starting IMAGIC program]**
IMAGIC-COMMAND : inc-pre

** INCPREP (vs. 21-May-2007) welcomes you **
Use MPI parallelisation [NO] : **NO [hit return for default]**
Input file, image loc#s [ctf-append_1000] : CMOS_ctf-append_500
Output file, image loc#s [ctf-append_1000-bp] : CMOS_ctf-append_500-bp

The image will be band-pass filtered.
Please specify:
Low frequency cut off [] : **0.025** **[roughly particle size, pixel size: 3Å, Nyquist 6Å]**
Remaining low-freq. transmission [] : **0.1** **[leave 10% of low frequencies]**
High frequency cut off [] : **0.5** **[high frequency cut off]**
The image will be masked by a circle. Please specify
the mask radius (pixels or fraction of inner radius)
If you specify a drop-off it will be a soft mask.
Mask radius, drop-off [] : **0.999** **[keep maximum to the edge of a circular area]**
Desired new sigma [] : **3** **[normalise the variance to 3 sigma]**
Invert the image densities [NO] : **[hit return for default]**

Display the filtered version of the particles for comparison

3) calculate the total sum of the particle images which will serve as a reference for particle centering

IMAGIC-COMMAND : **inc-sum**

** SUMMER (vs. 14-June-2007) welcomes you **

Mode of summing:
CONDITIONAL_SUM SOME_SUM TOTAL_SUM
Please specify option [TOTAL_SUM] : **[hit return for default]**
Input file, NO loc#s [] : **CMOS_ctf-append_500-bp**
Output file, image loc#s [] : **CMOS_ctf-append_500-bp_sum**

Display the file **CMOS_ctf-append_500-bp_sum**

4) particle centering:

IMAGIC-COMMAND : **ali-dir**

** ALIDIR (vs. 19-July-2007) welcomes you **
Alignment modes available:
TRANSLATIONAL ROTATIONAL HORIZONTAL VERTICAL
ALL
Please specify option [] : **TRANSLATIONAL**
Correlation functions available:
CCF MCF
Please specify option [] : **CCF**
Input file, image loc#s [] : **CMOS_ctf-append_500-bp**
Output file, image loc#s []: **CMOS_ctf-append_500-bp_cent1**
Reference file, image loc []: **CMOS_ctf-append_500-bp_sum**
Give this reference a number (1,2,...) [0] : **[hit return for default]**
Options to filter the reference(s):
NO_FILTER LOWPASS
Please specify option []: **LOWPASS** **[filtering the reference; try also NO_FILTER]**
Halfwidth value for low-pass filter [] : **0.1** **[e.g. 10% of the Nyquist frequency]**
Max shift (pixels/fraction of radius) [] **0.3** **[e.g. 30% of the image size]**
Full output? [] **NO**
Maximum allowable (radial) shift is … pixels.
…

IMAGE	#-ITER	ANGLE	XSHIFT	YSHIFT	CCC	
1	1	0.00	-1.14	3.81	0.1768	
2	1	0.00	-2.32	0.08	0.1810	etc.

display the files:
CMOS_ctf-append_500-bp **CMOS_ctf-append_500-bp_cent1** (CMOS_ctf-append_500-bp_cent2)
to check the success of the centering

6) create a mask for the area to be considered during multivariate statistical analysis (MSA)
IMAGIC-COMMAND : **test-im**

** TESTIM (vs. 11-July-2007) welcomes you **

Output filename, image loc#s []: **msamask**
Image dimensions X,Y [96, 96] : **84,84** **[hit return for default]**
IMAGIC data formats you can choose:
PACK INTG REAL COMP RECO
Please specify option [REAL] : **[hit return for default]**
Currently, you can choose:
…
Please specify option [] : **DISC**
Disc radius (pixel or fraction of inner radius) [] : **0.75** **[to be adjusted to particle size]**

7) multivariate statistical analysis (MSA):
IMAGIC-COMMAND : **msa-run**
Use MPI parallelisation [NO] : **NO** **[hit return for defaults]**

** MSA (vs. 3-Sep-2011) welcomes you **
Choose mode of operation:
FRESH_MSA REFINE
Please specify option [FRESH_MSA] : **[hit return for default]**

MSA distances:
EUCLIDIAN CHISQUARE MODULATION
Please specify option [MODULATION] : **[hit return for default]**

Input (= output) file (aligned "images") [] : **CMOS_ctf-append_500-bp_cent1**
Input MSA mask file [msamask] : **[hit return for default]**
Eigenimages output file [: **eigenim**
Pixel coordinates output file [] : **pixcoos**
Eigenpixel vectors output file [] : **eigenpix**
Number of iterations (<65) [] : **25**
Number of eigenimages (< 70) [] : **40**
Overcorrection factor (0 < ocf < 0.9) [0.8] : **0.8** **[hit return for default]**
Rootname for results file, NO ext. [msa] : **[hit return for default]**

Display the file **eigenim**

8) hierarchical ascendant classification:

IMAGIC-COMMAND : **msa-class**

** CLASSIFY (vs. Sept. 2006) welcomes you **

Input to be classified:

IMAGES PIXEL-VECTORS SEQUENCES
Please specify option [IMAGES] : **[hit return for default]**
Input (=output) file (treated by MSA)[] : **CMOS_ctf-append_500-bp_cent1**
Percentage of images to be ignored [0] : **[hit return for default]**
Active eigenimages for classification [] : **30**
Use default classification options [YES] : **[hit return for default]**
What number of classes do you wish [] : **50** **[total particle number divided by number of members per class (usually 10-20, or 3-5 with high-contrast images)]**

Name of output results files []: **classes0_50**

9) form class averages:

IMAGIC-COMMAND : msa-sum

** CLASSUM (vs. 26-Feb-2007) welcomes you **

Input images to be summed [] : **CMOS_ctf-append_500-bp_cent1**
Rootname of input classification files [] : **classes0_50**
Output class averages [] : **classums0_50**
Downweight small classes [NO] : **[hit return for default]**

Mode of summing statistics:
NONE VARIANCE S-IMAGE I-IMAGE FT
Please specify option [NONE] : **[hit return for default]**
Fraction of worst class members to ignore [0] : **[hit return for default]**

Display the file classums0_150 (and keep it displayed, use Ctrl Z)

10) band-pass filter the class averages:
IMAGIC-COMMAND : band-pass
** INCBAND (vs. Feb. 2007) welcomes you **
Input file, loc#s []: **classums0_50**
Output file, image loc#s [] : **classums0_50-bp**
Filter options available:
BAND-PASS HIGH-PASS LOW-PASS
INVERSE_BAND-PASS
Please specify option [BAND-PASS] : **BAND-PASS**
The image will be band-pass filtered.
Please specify
Low frequency cut off [] : **0.04** **[remember about Nyquist frequency…]**
Remaining low freq. transmission [0.005]: **0.005** **[hit return for default]**
High frequency cut off [] : **0.5**
ASQ filter the images too [NO]: **[hit return for default]**

Display the file classums0_150-bp (and keep it displayed, use Ctrl Z, bg to put the job into background)

11) Assigning angles without a reference, based on common lines

IMAGIC-COMMAND : **ang-rec**
** EULER (vs.) welcomes you **
Pointgroup symmetry:

C1	1	C2	2
C3	3	C4	4
C5	5	C6	6
C7	7	C8	8
C9	9	C10	10
C11	11	C12	12
C13	13	C14	14
C15	15	C16	16
C17	17	C18	18
C19	19	C20	20
C21	21	C22	22
CN	N	D2	222
D3	32	D4	422
D5	52	D6	622
D7	72	D8	822
D9	92	TETRAGONAL	23
O (CUBIC)	432	ICOSAHEDRAL	532
NONE			

Please specify option [] : **C1** **[C1 Point-group symmetry for an asymmetric object]**

Option for angular reconstitution:

NEW	ANCHOR_SET
C1_STARTUP	SELF_SEARCH
SINOGRAM	SINE_CORRELATION
PREDICT_SINE	CORR_PEAKS

Please specify option [] : **C1_STARTUP**
Input (classum) images, NO loc#s [] : **classums0_50-bp**
Loc# of THREE (classum) images to be used [] : **1; 13; 25** **[choose 3 different views; separate location numbers by ";"]**
Output (ordered) image file [my_ordered] : **[hit return for default; selected class averages will be put into a new file called my_ordered]**
Output sinograms, NO loc#s [my_sino] : **[YES, hit return for default; sinogram file]**
ASQ filter the sinogram lines [YES] : **[hit return for default; amplitude square-root filtering]**
Linear mask radius for sinograms [] : **0.7** **[depends on particle size]**
Output sinecorr file, NO loc#s [my_sine] : **[hit return for default; sinogram correlation file]**
Wanted angular increment in search [5.0] : **[hit return for default]**
Minimum inter-euler stay away angle [30.] : **[hit return for default]**
Full output of the results [NO] : **YES**

Are the relative angles clearly bigger than ~40°? If not, select another set of 3 views and start again; Careful to not select mirror images (which have opposite viewing angles and are therefore similar)

12) 3D reconstruction
IMAGIC-COMMAND : **true**

** TRUE3D (vs. Jan. 2007) welcomes you **

MPI parallelisation:
ONLY_3D ALL NO
Please specify option [] : **NO**
Please specify option [] : **ALL in one**

Pointgroup symmetry to be used:
C1 1 C2 2
…
Please specify option [] : **C1**
Use default 3D reconstruction options [YES] : **[hit return for default]**
Input 2D (classum) images, loc#s [] : **my_ordered**

Source of Euler angles:
ANGREC_HEADER_VALUES PLT_FILE
MRA_HEADER_VALUES
Please specify option [ANGREC_HEADER_VALUES] : **[hit return for default]**
Output 3D rec. filename, loc#s []: **3d_0-1** **[file which will contain the 3D reconstruction, sections by sections after weighted back-projection]**
Output file for reprojections, NO loc#s [] : **3d_0-1-reproj** **[reprojections according to the same Euler angles as the input images]**
Output file for error projections, NO loc# [] : **3d_0-1-err** **[difference between reprojection and input image, i.e. reflects amount of error]**
Mask the reconstruction [] : **YES**
Radius of the mask [] : **0.85**
Hamming window factor [] : **0.6**
Object size as fraction of image size []: **0.8**

Now display the files my_ordered and 3d_0-1-reproj for comparison, do they look correct?

13) Add more views to the angular reconstitution to improve the structure:
IMAGIC-COMMAND : **ang-rec**

** EULER (vs. 27-Sep-2006) welcomes you **

Pointgroup symmetry:
C1 1 C2 2
…
Please specify option []: **C1**
Option for angular reconstitution:
NEW ANCHOR_SET

C1_STARTUP SELF_SEARCH
SINOGRAM SINE_CORRELATION
PREDICT_SINECORR_PEAKS
Please specify option [C1_STARTUP] : **NEW**
Option of NEW:
FRESH ADD REMOVE_PROJ
Please specify option [ADD] : **ADD**
Input (classum) images, NO loc#s []: **classums0_50-bp**
Location number(s) wanted [] : **1;23;31;45 [select one image or a series of images]**
Output (ordered) image file [my_ordered] : **[hit return for default]**
Output sinograms, NO loc#s [my_sino] : **[hit return for default]**
ASQ filter the sinogram lines [YES] : **[hit return for default]**
Linear mask radius for sinograms [] : **0.7 [as before]**
Output sinecorr file, NO loc#s [my_sine] : **[hit return for default]**
Wanted angular increment in search [5.0] : **[hit return for default]**
Full output of the results [YES] : **NO**

14) 3D reconstruction with more views
IMAGIC-COMMAND : true

** TRUE3D (vs. Jan. 2011) welcomes you **
MPI parallelisation:
ONLY_3D ALL NO
Please specify option [NO] : **[as before]**
Pointgroup symmetry to be used: …
Please specify option [C1] : **[as before]**
Use default 3D reconstruction options [YES] : **[as before]**
Input 2D (classum) images, loc#s [my_ordered] : **[as before]**

Source of Euler angles:
ANGREC_HEADER_VALUES PLT_FILE
MRA_HEADER_VALUES
Please specify option [ANGREC_HEADER_VALUES] : **[as before]**
Output 3D rec. filename, loc#s [3d_0-1] : **3d_0-2**
Output file for reprojections, NO loc#s [3d_0-1-reproj] : **3d_0-2-reproj**
Output file for error projections, NO loc# [3d_0-1-err]: **3d_0-2-err**
Mask the reconstruction [YES] : **[as before]**
Radius of the mask [0.85] : **[as before]**
Hamming window factor [0.6] : **[as before]**
Object size as fraction of image size [0.8] : **[as before]**

Compare again new files my_ordered and 3d_0-2-reproj

15) Make forward projections
(could be used as references for a multiple-reference alignement, here only for comparing forward projections of 3d_0-1 and 3d_0-2)

IMAGIC-COMMAND : **threed-for**

** FORWARD (vs. Jan. 2007) welcomes you **
Input 3D image file [] : **3d_0-1**
Output file for forward projections []: **3d_0-1-24**
Threshold 3D density value [-99999] : **[hit return for default]**
Option used for current IMAGIC command: FORWARD
Mode of interpolation for projecting:
NEAREST_NEIGHBOUR BILINEAR SPLINE
SINC NARROWING WIDENING
OBLIQUE_SAMPLING HEADERS_ONLY
Please specify option [WIDENING] : **[hit return for default]**
Choose projection option:
FILE ANOTHER INTERACTIVE ORTHOGONAL
SPIRAL TETRAHEDRON TOMOGRAPHY STEREO
UNIFORM ICOSAHEDRON ASYM_TRIANGLE RANDOM
Please specify option [ASYM_TRIANGLE] : **[hit return for default]**

Pointgroup symmetry to be used:
C1 1 C2 2 **....**
Please specify option [C1] : **[hit return for default]**
Option to chose Euler angles:
EQUIDIST RANDOM
Please specify option [EQUIDIST] : **[hit return for default]**
Option for Euler angle alpha:
ZERO ROTATE
Please specify option [ZERO] : **[hit return for default]**
Wanted angular increment in search [] : **45**
Please specify option [] : NO

Do the same for 3d_0-2;
Compare files 3d_0-1-24 and 3d_0-2-24 , did the quality of the reconstruction improve?

If time allows: refine the structure by using forward projections as references:
run m-r-a (multi-reference-alignment), then msa and classification and 3D reconstruction (beginning of the iterative structure refinement procedure)

Optional (not part of practicals at Oleron school, but useful to see Fourier effects / reciprocity of dimensions in Fourier space, and border effects of masks:

I. Illustration of the Fourier transformation

cp ../FT-effects/* . **[copy over the images to work on]**

Files names are:	description:
checker_8	**checkerboard array**
checker_32	**checkerboard array**
disc	**sharp disc**
disc-smooth	**smooth disc**
square	**square**

Calculate Fourier transformation of these images:
i **[shortcut for starting IMAGIC program]**
IMAGIC-COMMAND : **fft**

** INCFFT2D (vs. Aug. 2005) welcomes you **

Input file, image loc#s [] : **checker_8**
Output file, image loc#s [] : **checker_8-fft** **[give output file name]**
Mode of operation:
FORWARD_FFT REVERSE_FFT AUTO_CORRELATION
SELF_CORRELATION POWER_SPECTRUM AMPLITUDE_SPECTRUM
Please specify option [] : **FORWARD**
[hit return]

[do the same for the other images: produce files:
checker_8-fft, checker_32-fft, disc-fft, disc-smooth-fft, square-fft
quit **[or Ctrl C] when finished**
display these files in IMAGIC:
disp **[under linux or within IMAGIC]**
Input image file, loc#s [checker_8] : **checker_8**
Size of the display window [600,600] : **[hit return for default]**
Type of cursor:
CROSS SQUARE CIRCLE
Please specify option [CROSS] : **[hit return for default]**
Parameters to be changed:
NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO] : **[hit return for default]**
To adjust scale of display:
scale
4

disp
Input image file, loc#s [checker_8] : **checker_32** [next file name]
etc.

Switch between display windows to compare the images (do not move displays around such that they remain aligned with respect to each other)

Then display the corresponding FT's, files:
checker_8-fft
checker_32-fft
Switch between these display windows to compare the images

With the same procedure, compare sharp and smooth discs and the square:

disc	disc-fft
disc-smooth	disc-smooth-fft
square	square-fft

When displaying the files **disc-fft** and **disc-smooth-fft** you can draw a profile of the spectrum:
In the display command window:
Parameters to be changed:
NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO] : **profile**
Use cursor to position profile [NO] : **[NO; hit return for default]**
Starting point (IMAGE coordinates X,Y) [1,1] : **1,1**
End point (IMAGE coordinates X,Y) [128,128] : **65,65**
Parameters to be changed:
NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO] : **[hit return for default]**
Output device (X_WINDOWS, PS, FILE) [X_WINDOWS] : **[hit return for default]**

Display settings:
device
600, 1200
scale
4
file
filename
grey
-10,10
profile
1,1
65,65

Basic Steps in Single particle image processing and 3D reconstruction

I. Pre-processing

- Digitization of micrographs (negatives); not needed if CCD/CMOS images

- particle selection, « boxing »

- correction of the contrast transfer function

- band-pass filtering and normalisation of particle images

II. Structure determination

- particle centering / alignments

- MSA (multivariate statistical analysis) + classification

- angle assignment

- angular reconstitution

- projection matching

- 3D reconstruction

- structure refinement

- resolution assessment

- map interpretation; fitting of known structures, atomic model building…

Some basic concepts of cryo-EM & 3D reconstruction

Correct terms are important (be precise and rigorous in science :-)

By cryo-EM, we obtain:

- a "3D reconstruction" (initial or refined)
- a "cryo-EM map" or "density map"
- a "structure"

technically:

- back-projection
- angular reconstitution
- random conical tilt
- tilt series / tomogram

NOT:

- an "envelope" (would be SAXS or neg. stain. EM)
- a "volume", units would be Å^3 (e.g. volume of a pocket, volume x density = mol. mass)
- a "surface", units would be Å^2 (e.g. interaction surface between 2 proteins)
- a "model", would be a molecular model *fitted to* the map (crystallography/cryo-EM)
  or a model *compatible with* SAXS data or NMR restraints;
  other "models": "homology model", "hypothetical model", "working model"

IGBMC

Some basic concepts of cryo electron microscopy

Correct terms are important:

A classification is based on a statistical analysis:

- multivariate statistical analysis (MSA) provides information on variance (variability) which serves to merge similar images into class averages (classes); is *independent* of a reference

- classes *are NOT*: the sum of images that correlate best with a reference (through a multi-reference alignment)

Some basic concepts of cryo electron microscopy

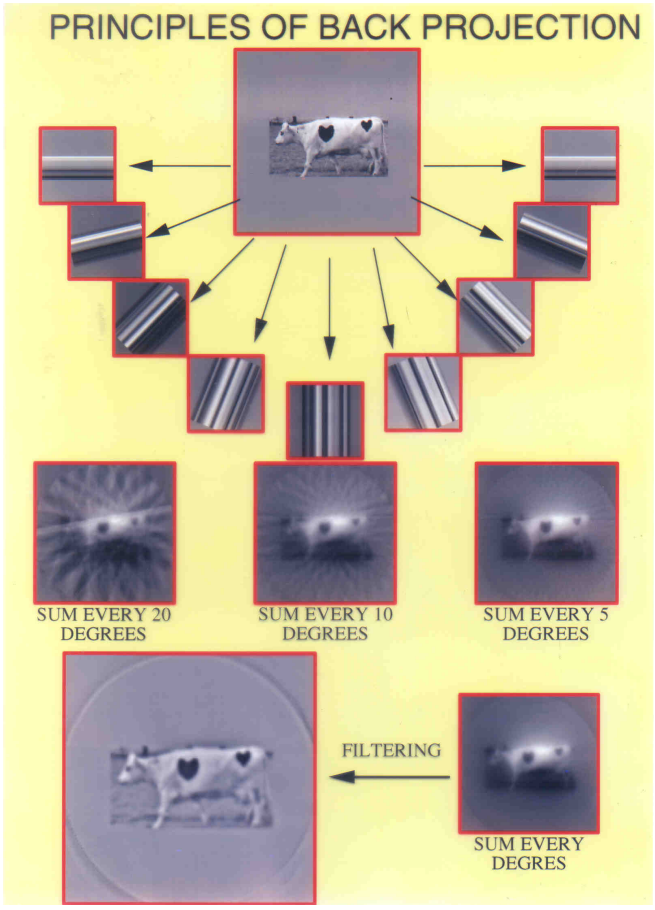
Basic aspects:

- **"resolution" corresponds to "frequency" in image processing (1/ Å)**
- **Nyquist frequency is = 2 x pixel size, e.g. 1 Å / pixel → Nyquist = 2 Å**
- **interpolations during 2D image alignment and 3D reconstruction limit the possible resolution to about 2/3 of the Nyquist frequency, i.e. here ~ 3 Å** *(exception: super-reso)*

pixels in 3D: "voxel"

Consider:

- **any correlation calculation (e.g. alignment) is biased by the reference used**
- **resolution estimation, criteria used:**
  - **0.5, arbitrary, historically from the virus field, tends to underestimate resolution**
  - **0.143 (Henderson) and ½ bit (van Heel)**
  - 3 σ, not used anymore (over-estimation)
  - **features in the map: can we see dsRNA helices (~10-12 Å resolution), α-helices (~8 Å), ß-sheets (~5 Å) or side chains (4-2.5 Å depending on size)?**


PRINCIPLES OF BACK PROJECTION
SUM EVERY 20 DEGREES
SUM EVERY 10 DEGREES
SUM EVERY 5 DEGREES
FILTERING
SUM EVERY DEGRES