IMAGIC and Frealign

HANDS-ON

In the following you will find a work flow describing how IMAGIC images can be processed in Frealign

Version 25-Aug-2014

Installing
Frealign

1. Installing Frealign

Frealign is distributed as an archive that can be downloaded here:

grigoriefflab.janelia.org/frealign

The archive contains source code that can be compiled on Linux and Mac OSX platforms, as well as precompiled versions for Linux and Mac OS. After downloading, unpack the archive in a location where you keep installed programs using

```
tar -xvzf frealign.tar.gz
```

Change directory into the created Frealign installation directory and type

```
./INSTALL
```

This should produce a line of code that needs to be added to the .bashrc or .cshrc file in your home directory, depending on which shell you run. After that, log out and back in to activate the Frealign installation. To check that Frealign is installed correctly, type

```
frealign_help
```

This should produce a list of available Frealign commands with a short description.

Note for Max OSX users: In addition to the steps above, you need to move the content of the "bin_OSX" folder into the "bin" folder inside the Frealign installation directory.

IMAGIC
to
Frealign

2. IMAGIC

This chapter describes, which image files have to be processed in **IMAGIC** and how to create all files, which are needed to refine your results with **Frealign**.

2.1 IMAGIC images needed

The following image are needed:

(a) The original images picked from your micrographs. These images should be normalised but not aligned, not CTF corrected, not band-pass filtered etc.

(b) The processed images (CTF corrected, aligned and Euler angles assigned – either using angular reconstitution or multi-reference alignment = projection matching)

(c) The untreated micrographs and the PLT coordinate file (output from command PICK-PARTICLES) in case the particles had been picked from a pre-treated micrograph (as is the case in the Brazil School hands-on).

NOTE 1:

All "processed" images (b) MUST have their untreated starting image in the "original" image file (a) or untreated micrograph images (c)

NOTE 2:

Note in the following this manual refers to or suggests file names written in blue. These are file names used in the Brazil School hands-on and are not necessarily the ones you used in your analysis.

2.2 Last Multi-Reference Alignment

The "processed images" (c) are the latest images, which were multi-reference aligned (in the Brazil School hands-on the file name is like: particles_mra_best).

NOTE:

Each multi-reference alignment in **IMAGIC** is a PROJECTION-MATCHING if the references are forward projections of a 3-D volume.

2.3 Extract your Particles from the Untreated Patches

If you picked the particles (particles_mra_best) from raw micrographs (micrograph_raw or micrograph_raw_coarse) you can continue with chapter 1.4.

If you used the Brazil School Hands-On you picked from processed micrographs (micrograph_flip). Therefore you have to re-do the particle extraction using the non-filtered, non-CTF-corrected micrographs (which you can find on the directory "Dataset_Wormhemoglobin/Micrographs_297_raw" on the Brazil School network drive). You need the PLT coordinates file, which was the output of the command PICK- PARTICLES. In the Brazil School hands-on the file name was coord_ccf.plt.

1. Before you extract these images check if the micrograph numbers are available in the headers of micrograph images because **Frealign** needs to know these numbers. Call HEADER:

```
IMAGIC-COMMAND: header

Specify option                        : read
Read options available                : micrograph_id
Source of header values               : file
Input (header) file, image loc#s      : micrograph_raw_coarse
```

2. If no micrograph numbers are specified in the headers try to find if you once created a text file with these numbers (filenumbers.txt).

   If such a file does not exist you have to create this text file (filenumbers.txt) with your text editor. Write the micrograph numbers, one number per line and image.

   If the micrograph numbers are unknown you can simply write the location numbers into the headers. If you do not want to type in all numbers you can copy the printout of the above HEADER command and globally remove "not specified").

   Update the headers with command HEADER:

```
IMAGIC-COMMAND: header

Specify option                        : write
Options available                     : micrograph_id
Source of header values               : file
Plt input file with header values     : filenumbers.txt
Input (header) file, image loc#s      : micrograph_raw_coarse
```

3. If the micrograph numbers are available in the headers you van extract the wanted particle images. As usual call the command CUT-IMAGE:

```
IMAGIC-COMMAND: cut-im

Mode of operation                     : aperiodic
Input file, image loc#s               : micrograph_raw_coarse
Output file, image loc#s              : particles_ccf
Output image dimensions X,Y           : 180,180   Use large-image sizes
                                                  Note: the images are
                                                  not yet CTF corrected
                                                  (point spread)
Coordinates (plt) file                : coord_ccf
```

2.4 Extract the "active" Images

During image processing you usually have excluded some "bad" images so the number of images in the untreated image file (particles_ccf) and in the last multi-reference aligned file (particles_mra_best) is not the same any more.

There are two cases now:
(a) If you only want to use these "good" images in **Frealign** follow this chapter.
(b) You do not have to necessarily exclude images and you can use all original images in **Frealign**. In this case continue with chapter 1.5.

Continuing here you want to use only those original images in **Frealign**, which are available in the last multi-reference aligned image file (particles_mra_best). Call the command EXTRACT-IMAGE:

```
IMAGIC-COMMAND: extract-image

What should be copied                    : 2d
Mode of copy operation                   : EXTRACT
Input file, NO loc#s                     : particles_ccf
Output file, image loc#s                 : particles_frealign
Source of image locations                : original_images
Header file of processed images          : particles_mra_best
Reset "original location" header value: yes
```

All important header values were taken over into the headers of the output file (particles_frealign).

2.5 Sort Images by Micrograph Numbers

It is a good idea to sort the images by micrographs numbers, although this sorting is not necessarily needed in **Frealign**.

There are two cases now:
(a) If you only want to sort images by micrographs numbers follow this chapter.
(b) If you do not want to perform this sorting continue with chapter 1.6.

First check if the particle images are already sorted by their micrograph identification numbers.

```
IMAGIC-COMMAND: header

Specify option                      : look
Options available                   : micrograph_id
Input (header) file, image loc#s    : particles_frealign
```

If necessary, sort with the command SORT-IMAGES:

```
IMAGIC-COMMAND: copy-image

Input file, image loc#s             : particles_frealign
Output file, image loc#s            : particles_copy

IMAGIC-COMMAND: sort-image

What should be copied               : 2d
Mode of copy operation              : SORT
Input file, NO loc#s                : particles_copy
Output file, image loc#s            : particles_frealign
Source of SORT values               : header
Criteria for SORT                   : index
How to specify the index            : label
Label wanted                        : mident
```

```
Sort up or down                          : up
How many images wanted                   : 0                    all images
```

2.6 Export to Frealign

The command IMAGIC-TO-FREALIGN converts the un-treated images into a **Frealign** MRC file and creates a **Frealign** parameter file.

```
IMAGIC-COMMAND: imagic-to-frealign

Option used for current command    : CONVERT
Input file with ORIGINAL images    : particles_frealign
Header file of PROCESSED images    : particles_mra_best
Invert contrast/densities          : zero_float_after_invert
Source of Euler angles             : mra_header
Source of magnification values     : interactive_constant
Image magnification                : 59000          you have to know
Source of micrograph numbers       : header
Source of defocus values           : header
```

Now all files are created, which you need to work with **Frealign** (particles_frealign.mrc and particles_frealign.par).

Refining
with
Frealign

3. Refining with Frealign

You should choose a name for your project (for example "hem" for worm hemoglobin) and change the created **Frealign** files appropriately (for example, hem_stack.mrc and hem_1_r1.par).

The **Frealign** parameter file always consists of the project name (here "hem", a cycle number (here "1") and class number (here "r1").

NOTE:

A **Frealign** class number and a MSA classification number in **IMAGIC** are not the same.

3.1 Preparing `mparameters` File

Frealign runs are controlled by parameters in the `mparameters` file that must be placed in the working directory from which **Frealign** will be launched. To obtain a template for this file, type

```
frealign_template
```

This template file needs to be edited to fit your computing environment (here 8 CPU cores), microscope parameters and sample specifications.

Below is the original template file (left) and edited file for worm hemoglobin (right):

```
Control parameter file to run Frealign
======================================

This file must me kept in the project work

Note: Please make sure that project and sc

# Computer-specific setting
cluster_type          none        ! Set to '
nprocessor_ref         8          ! Number c
nprocessor_rec         4          ! Number c
mem_per_cpu         2000          ! Memory a

# Refinement-specific parameters
MODE                   1          ! 1, 2, 3
start_process          2          ! First cy
end_process            2          ! Last cyc
res_high_refinement    6.0        ! High-res
res_high_class         8.0        ! High-res
thresh_reconst         0.0        ! Particle
nclasses               1          ! Number c

# Search-specific parameters
res_search            30.0        ! High-res
thresh_refine         50.0        ! Mode 4:
DANG                 200.0        ! Mode 3 a
ITMAX                200          ! Mode 2 a
Bsearch             2000.0        ! B-factor

# Dataset-specific parameters
data_input            particle    ! Root nam
raw_images            /path/partice_stack
image_contrast         N          ! N or P.
outer_radius         170.0        ! Outer ra
inner_radius           0.0        ! Inner ra
mol_mass            2500.0        ! Molecula
Symmetry              C1          ! Symmetry
pix_size               1.237      ! Pixel si
dstep                  5.0        ! Pixel si
Aberration             2.0        ! Sherical
Voltage              300.0        ! Beam acc
Amp_contrast           0.07       ! Amplituc
```

```
Control parameter file to run Frealign
======================================

This file must me kept in the project work

Note: Please make sure that project and sc

# Computer-specific setting
cluster_type          none        ! Set to '
nprocessor_ref         8          ! Number c
nprocessor_rec         8          ! Number c
mem_per_cpu         2000          ! Memory a

# Refinement-specific parameters
MODE                   1          ! 1, 2, 3
start_process          2          ! First cy
end_process            4          ! Last cyc
res_high_refinement   20.0        ! High-res
res_high_class         8.0        ! High-res
thresh_reconst         0.0        ! Particle
nclasses               1          ! Number c

# Search-specific parameters
res_search            30.0        ! High-res
thresh_refine         50.0        ! Mode 4:
DANG                 200.0        ! Mode 3 a
ITMAX                200          ! Mode 2 a
Bsearch             2000.0        ! B-factor

# Dataset-specific parameters
data_input             hem        ! Root nam
raw_images             hem_stack
image_contrast         N          ! N or P.
outer_radius         150.0        ! Outer ra
inner_radius           0.0        ! Inner ra
mol_mass            3600.0        ! Molecula
Symmetry              D6          ! Symmetry
pix_size               5.44       ! Pixel si
dstep                 56.0        ! Pixel si
Aberration             2.7        ! Sherical
Voltage              300.0        ! Beam acc
Amp_contrast           0.07       ! Amplituc
```

3.2 Running Frealign

After the mparameters file has been modified, **Frealign** can be launched by typing

```
frealign_run_refine
```

Frealign will create a local scratch directory that contains all temporary files. Progress can be monitored by inspecting the file `frealign.log` in the work directory. This file will also contain error messages. It may be useful to issue to following command in the working directory for continuous monitoring:

```
tail -f frealign.log
```

If, for any reason, a **Frealign** run has to be aborted before in finishes normally, type

```
frealign_kill
```

Other **Frealign** commands available and you are encouraged to issue

```
frealign_help
```

Note: parameters set in the `mparameters` file can be changed "on the fly", i.e. while **Frealign** is running. Therefore, if after inspection of intermediate results you decide to change a parameter, for example the refinement resolution, you do not need to stop **Frealign**. Simply change the parameter in the `mparameters` file and when the next refinement cycle starts it will use the new parameter. It is also possible to terminate **Frealign** in this way. For example, if cycle 3 is running and you decide you want to terminate refinement after cycle 3, you can change the value for "end_process" in the `mparameters` file to "3". When cycle 3 finishes **Frealign** will stop running.

3.3 Interpreting Results

When a **Frealign** refinement cycle finishes, new `.par` and map files are generated with updated cycle numbers. The end of each `.par` file contains a Fourier Shell Correlation (FSC) table, as well as other tables:

C	NO.	RESOL	RING RAD	FSPR	FSC	Part_FSC	sqrt Part_SSNR	sqrt Rec_SSNR	CC	EXP. C	SIG C	ERFC	TOTVOX
C	2	522.24	0.0104	0.04	0.999	1.000	7.6605	150.08	0.0000	0.0000	0.0000	0.00	17
C	3	261.12	0.0208	0.02	0.999	1.000	12.2895	197.61	0.0000	0.0000	0.0000	0.00	41
C	4	174.08	0.0313	0.36	1.000	1.000	8.7815	242.54	0.0000	0.0000	0.0000	0.00	89
C	5	130.56	0.0417	0.26	0.999	1.000	6.1829	188.86	0.0000	0.0000	0.0000	0.00	129
C	6	104.45	0.0521	0.95	0.999	1.000	3.4398	136.96	0.0000	0.0000	0.0000	0.00	225
C	7	87.04	0.0625	0.97	0.997	1.000	1.6125	84.48	0.0000	0.0000	0.0000	0.00	253
C	8	74.61	0.0729	1.03	0.997	1.000	1.5789	87.23	0.0000	0.0000	0.0000	0.00	393
C	9	65.28	0.0833	1.45	0.994	1.000	1.1361	63.40	0.0000	0.0000	0.0000	0.00	461
C	10	58.03	0.0938	1.00	0.997	1.000	1.2171	86.83	0.0000	0.0000	0.0000	0.00	613
C	11	52.22	0.1042	1.46	0.997	1.000	1.1105	88.06	0.0000	0.0000	0.0000	0.00	713
C	12	47.48	0.1146	1.79	0.996	1.000	1.0575	80.51	0.0000	0.0000	0.0000	0.00	845
C	13	43.52	0.1250	3.03	0.991	0.999	0.6389	51.71	0.0000	0.0000	0.0000	0.00	1025
C	14	40.17	0.1354	3.54	0.988	0.999	0.5329	44.30	0.0000	0.0000	0.0000	0.00	1229
C	15	37.30	0.1458	4.60	0.981	0.998	0.4701	35.75	0.0000	0.0000	0.0000	0.00	1313
C	16	34.82	0.1563	5.31	0.964	0.997	0.4164	25.76	0.0000	0.0000	0.0000	0.00	1585
C	17	32.64	0.1667	10.65	0.920	0.993	0.3470	16.82	0.0000	0.0000	0.0000	0.00	1701
C	18	30.72	0.1771	15.30	0.860	0.987	0.4210	12.30	0.0000	0.0000	0.0000	0.00	2021
C	19	29.01	0.1875	29.71	0.651	0.958	0.5845	6.77	0.0000	0.0000	0.0000	0.00	2181
C	20	27.49	0.1979	62.73	0.276	0.824	0.1820	3.06	0.0000	0.0000	0.0000	0.00	2473
C	21	26.11	0.2083	33.88	0.615	0.952	0.1787	6.27	0.0000	0.0000	0.0000	0.00	2697
C	22	24.87	0.2188	26.15	0.702	0.967	0.1624	7.61	0.0000	0.0000	0.0000	0.00	2949
C	23	23.74	0.2292	20.20	0.790	0.979	0.1641	9.61	0.0000	0.0000	0.0000	0.00	3237
C	24	22.71	0.2396	15.81	0.852	0.986	0.1905	11.90	0.0000	0.0000	0.0000	0.00	3577
C	25	21.76	0.2500	16.81	0.827	0.983	0.2113	10.85	0.0000	0.0000	0.0000	0.00	3761
C	26	20.89	0.2604	22.61	0.768	0.976	0.2840	9.02	0.0000	0.0000	0.0000	0.00	4249
C	27	20.09	0.2708	42.02	0.505	0.926	0.3624	5.01	0.0000	0.0000	0.0000	0.00	4465
C	28	19.34	0.2813	72.24	0.179	0.728	0.1112	2.31	0.0000	0.0000	0.0000	0.00	4937
C	29	18.65	0.2917	43.36	0.511	0.928	0.1180	5.07	0.0000	0.0000	0.0000	0.00	5013
C	30	18.01	0.3021	30.57	0.649	0.958	0.1235	6.74	0.0000	0.0000	0.0000	0.00	5677
C	31	17.41	0.3125	31.91	0.619	0.952	0.1288	6.33	0.0000	0.0000	0.0000	0.00	5773
C	32	16.85	0.3229	42.82	0.504	0.926	0.1427	5.00	0.0000	0.0000	0.0000	0.00	6489
C	33	16.32	0.3333	66.27	0.241	0.797	0.1788	2.80	0.0000	0.0000	0.0000	0.00	6653
C	34	15.83	0.3438	78.11	0.118	0.622	0.0637	1.81	0.0000	0.0000	0.0000	0.00	7229
C	35	15.36	0.3542	69.04	0.213	0.769	0.0560	2.58	0.0000	0.0000	0.0000	0.00	7561
C	36	14.92	0.3646	61.78	0.288	0.833	0.0650	3.16	0.0000	0.0000	0.0000	0.00	8033
C	37	14.51	0.3750	51.77	0.402	0.892	0.0995	4.07	0.0000	0.0000	0.0000	0.00	8469
C	38	14.11	0.3854	67.95	0.227	0.783	0.1200	2.69	0.0000	0.0000	0.0000	0.00	8889
C	39	13.74	0.3958	86.32	0.042	0.349	0.0415	1.04	0.0000	0.0000	0.0000	0.00	9441
C	40	13.39	0.4063	76.43	0.150	0.685	0.0428	2.09	0.0000	0.0000	0.0000	0.00	9965
C	41	13.06	0.4167	68.29	0.231	0.787	0.0652	2.72	0.0000	0.0000	0.0000	0.00	10309
C	42	12.74	0.4271	66.15	0.249	0.803	0.0988	2.86	0.0000	0.0000	0.0000	0.00	11033
C	43	12.43	0.4375	85.57	0.046	0.371	0.0559	1.09	0.0000	0.0000	0.0000	0.00	11433
C	44	12.15	0.4479	91.73	-0.018	-0.178	0.0177	0.66	0.0000	0.0000	0.0000	0.00	12169
C	45	11.87	0.4583	84.26	0.064	0.457	0.0368	1.30	0.0000	0.0000	0.0000	0.00	12561
C	46	11.61	0.4688	84.05	0.062	0.449	0.0718	1.28	0.0000	0.0000	0.0000	0.00	13157
C	Average			60.27	0.323	0.694	0.3745	15.63	0.0000	0.0000	0.0000	0.00	

Resolution = 12.4 Å, according to the 0.143 threshold

It is important not to delete this table from the `.par` files because **Frealign** uses it to apply appropriate weighting to the data during refinement. There are two FSC tables, "FSC" and "Part_FSC". The "FSC" table is calculated from two reconstructions calculated internally by **Frealign**, each containing half the data. These reconstructions are masked with a sphere of radius "outer_radius", as specified in the `mparameters` file. This, these reconstructions usually contain background noise next to the particle density. "Part_FSC" provides a more accurate resolution estimate by excluding most of this background.

3.4 Updating Refinement Resolution

The resolution of the reconstruction should increase significantly after only a few cycles, often even after just one cycle. The resolution of the refinement as specified in the `mparameters` file ("res_high_refinement") should be adjusted correspondingly to include signal at higher resolution. It is important to limit the resolution to a value that is lower than the resolution of the current reconstruction as indicated by the "Part_FSC" table. For example, if a resolution of 10 Å is indicated, the refinement resolution should be limited to about 12 Å. When this is done, the FSC curves should not be affected by over-fitting and bias and should reflect the true resolution of the reconstruction (sometimes referred to as "gold-standard" FSC).

3.5 Sharpening Maps

Frealign does not apply additional sharpening to the reconstruction that is usually necessary to visualize high-resolution details. It also does not set Fourier terms beyond the resolution limit to zero. To sharpen a map and apply an appropriate low-pass filter, the program **Bfactor** can be used, which can be downloaded here:

grigoriefflab.janelia.org/bfactor

YOUR NOTES:

3.6 Multi-Reference Refinement/Classification

Starting classification of a dataset using **Frealign** is achieved by changing the number of classes from "1" to the higher number (for example, "3", see below). **Frealign** will then run a maximum-likelihood like algorithm to refine separate classes. Convergence may take many cycles (10s up to 100). In the example below, the starting and ending cycle numbers ("start_process" and "end_process") were also updated to continue from the initial refinement done using a single class (see section 7). Furthermore, the refinement resolution was updated to reflect the improved resolution of the reconstruction after four cycles of refinement (see FSC table, section 3.3).

```
Control parameter file to run Frealign
======================================

This file must me kept in the project work

Note: Please make sure that project and sc

# Computer-specific setting
cluster_type          none       ! Set to '
nprocessor_ref         8         ! Number c
nprocessor_rec         8         ! Number c
mem_per_cpu         2000         ! Memory a

# Refinement-specific parameters
MODE                   1         ! 1, 2, 3
start_process          2         ! First cy
end_process            4         ! Last cyc
res_high_refinement   20.0       ! High-res
res_high_class         8.0       ! High-res
thresh_reconst         0.0       ! Particle
nclasses               1         ! Number c

# Search-specific parameters
res_search            30.0       ! High-res
thresh_refine         50.0       ! Mode 4:
DANG                 200.0       ! Mode 3 a
ITMAX                200         ! Mode 2 a
Bsearch             2000.0       ! B-factor

# Dataset-specific parameters
data_input             hem       ! Root nar
raw_images             hem_stack
image_contrast          N        ! N or P.
outer_radius          150.0      ! Outer ra
inner_radius            0.0      ! Inner ra
mol_mass             3600.0      ! Molecula
Symmetry               D6        ! Symmetry
pix_size               5.44      ! Pixel si
dstep                  56.0      ! Pixel si
Aberration              2.7      ! Sherical
Voltage               300.0      ! Beam acc
Amp_contrast           0.07      ! Amplituc
```

```
Control parameter file to run Frealign
======================================

This file must me kept in the project work

Note: Please make sure that project and sc

# Computer-specific setting
cluster_type          none       ! Set to '
nprocessor_ref         8         ! Number c
nprocessor_rec         8         ! Number c
mem_per_cpu         2000         ! Memory a

# Refinement-specific parameters
MODE                   1         ! 1, 2, 3
start_process          5         ! First cy
end_process           20         ! Last cyc
res_high_refinement   13.0       ! High-res
res_high_class         8.0       ! High-res
thresh_reconst         0.0       ! Particle
nclasses               3         ! Number c

# Search-specific parameters
res_search            30.0       ! High-res
thresh_refine         50.0       ! Mode 4:
DANG                 200.0       ! Mode 3 a
ITMAX                200         ! Mode 2 a
Bsearch             2000.0       ! B-factor

# Dataset-specific parameters
data_input             hem       ! Root nar
raw_images             hem_stack
image_contrast          N        ! N or P.
outer_radius          150.0      ! Outer ra
inner_radius            0.0      ! Inner ra
mol_mass             3600.0      ! Molecula
Symmetry               D6        ! Symmetry
pix_size               5.44      ! Pixel si
dstep                  56.0      ! Pixel si
Aberration              2.7      ! Sherical
Voltage               300.0      ! Beam acc
Amp_contrast           0.07      ! Amplituc
```

3.7 Interpreting Classification Results

Classification will produce the `.par` and map files for each class. In our example, after the final has finished, the following files should be present in the working directory:

```
hem_20_r1.mrc
hem_20_r1.par
hem_20_r2.mrc
hem_20_r2.par
hem_20_r3.mrc
hem_20_r3.par
```

Files corresponding to earlier cycles will also be present. To display statistics (average class occupancy, score and likelihood), type

```
frealign_calc_statistics 20
```

The above command will show the statistics for cycle 20. The resolution ("FSC_Part") and the maps need to be inspected to determine the meaning of each class. The largest class is not always the "best" class.

Depending on the dataset, some classes accumulate "junk", i.e. particles that are damaged, dirt and sometimes also misaligned particles. Classification is therefore a convenient way to "clean" a dataset and remove non-matching particles that could reduce the quality and resolution of a reconstruction.

YOUR NOTES:

Content