



Frequently asked questions

Q: Can Frealign be used to generate an initial reconstruction?

A: This will generally not be possible. To use Frealign, an initial reconstruction will have to be supplied from another source, or it can be calculated using Frealign if particle alignment parameters are known.

Q: How good does the initial reconstruction have to be for Frealign to be able to perform successful refinement?

A: This will depend on the details of the particle and the degree of heterogeneity. Frealign's ability to converge to the correct structure from a rough starting structure is not as good as that of some of the other available single particle software. Frealign was designed primarily for high-resolution refinement and will work best if the resolution of the starting reconstruction is correct at a resolution of 10 to 15 Å.

Q: How do I get started if I have an initial reconstruction?

A: If an initial reconstruction is available, together with a particle stack (normalized, same pixel dimensions as the reconstruction), Frealign can determine initial particle alignments by running a global parameter search with Mode 3 (keyword MODE). Initial reconstruction and stack need to follow Frealign's naming convention. To run a global search with Mode 3, one also has to choose an angular step size for the search (keyword DANG). For a thorough search, a step size of 5 degrees is recommended. The user also has to supply a text file that lists for every particle four numbers: micrograph number (this could simply be a serial number for the micrographs), defocus value 1 (in Å), defocus value 2 (in Å) astigmatic angle (in degrees). This file also has to be named according to Frealign's naming conventions.

Q: How do I generate a list of micrograph numbers and defocus values for each particle?

A: This depends on how the defocus values were determined and the particles were picked/boxed. Users will have to write their own scripts to generate the list from the CTF results and particle picking files. If the CTF parameters were determined with CTFFIND4 and particles were boxed with EMAN's boxer program, an initial Frealign parameter file can be generated using [this script](#). The script assumes that CTF result files and boxer files share a root name, followed by a serial number and file extensions identifying the CTF and boxer files. It is essential that the order of the lines correspond to the order of the particles in the image stack.

Q: Which files are needed to start a Frealign run?

A: At a minimum, the particle image stack (normalized), a particle alignment parameter file and the control parameter file (mparameters) are needed. A 3D reconstruction is also required if initializing a global search (for example, using Mode 3).

Q: Can Frealign automatically perform pixel binning to speed up processing?

A: No, there is no automatic feature. However, Frealign comes with a tool, `resample.exe` (or `resample_mp.exe` for multiple CPUs) that can perform the binning. Both the particle image stack and 3D reconstructions have to be binned to maintain compatible pixel sizes. Also, in the mparameters file, the value for the pixel size (keyword `pix_size`) and effective detector pixel size (keyword `DSTEP`) have to be updated. For example, if the original pixel size is 1 Å and detector pixel size is 5 µm and 2 x 2 binning is used, the new pixel size will be 2 Å and the new detector pixel size will be 10 µm. The shift alignment parameters calculated by Frealign are listed in Å and therefore should not be affected by binning. If the user wants to switch back to unbinned data, the pixel sizes have to be changed back to the original values and the last reconstruction should be recalculated to generate a map with the smaller pixel size.

Q: How many CPUs should be used to run Frealign?

A: Frealign's computation is quite efficient and refinement (Mode 1) can be run on a workstation with multiple CPUs (for example, 16 CPUs). However, for large datasets (several 100,000 particles) and large particle sizes (larger than 256 pixels), or when a global search is run (Mode 3) it is recommended to run Frealign on a computer cluster using as many CPUs as are available (for example, 500). The number of CPUs can be set in the mparameters file using keyword `nprocessor_ref`. Frealign can also calculate 3D reconstructions using multiple CPUs (keyword `nprocessor_rec`). However, due to extensive disk I/O, it is recommended not to use more than about 50 CPUs (this will depend on the network speed).

Q: What is the difference between Mode 1, 2, 3 and 4?

A: Mode 1 is the standard mode used for local refinement. It assumes that the particle alignment parameters are already fairly close to their true values. Mode 2 is used to search N randomly picked orientations (N is set by keyword `ITMAX`). This may be useful to test if the current particle alignments are already close to their final values. If they are not, the randomized search will find improved orientations for a significant number of particles. The number of significantly changed orientations can be monitored by the user by inspecting the `.shft` files in Frealign's scratch directory. If the alignment parameter changes listed in the `.shft` files are frequent and large, this indicated that refinement has not yet converged. Mode 3 is used to perform a systematic angular grid search (the angular step size is set by keyword `DANG`). This is useful when initial particle alignment parameters have to be determined. Mode 4 is a combination of Mode 2 and 3. It performs N multiple grid searches with randomly picked starting orientations.

Q: What is the difference between search (Mode 2, 3, and 4) and refinement (Mode 1)?

A: In a search, many different orientations are tested, followed by local refinement at each test orientation. A search can therefore find better particle orientations, even if these are very different from the current best orientations. However, a search is computationally expensive. In refinement, only orientational parameters close to the current best parameters are searched. This type of refinement is therefore unlikely to determine the correct orientation parameters if these are very different from the current parameters. A refinement runs much faster than a search.

Q: What is the difference between FSC and Part_FSC in the table at the end of the particle parameter file?

A: The FSC table list the Fourier Shell Correlation curve for the two reconstructions calculated from even and odd-numbered particles. For this calculation, the reconstructions are masked with a spherical mask with a radius set by keyword `outer_radius`. This FSC includes some noise that is inside the spherical mask and not part of the particle density. To get a more accurate resolution estimate, one can apply a tighter mask to the two half reconstructions. This is done in some software packages. In Frealign, no tight mask is applied. Instead, the FSC that would have been obtained with a tight mask is estimated from the volume fraction that is occupied by the particle. This estimated FSC is tabulated under `Part_FSC` and should indicate the resolution of the reconstruction more accurately than the table listed under FSC.

Q: Why does the FSC or Part_FSC show negative values?

A: Both curves include some noise and when the correlation is weak, this noise can lead to negative correlations.

Q: What the meaning of the OCC column in the particle parameter file?

A: This column lists the fraction (multiplied by 100) of the particle belonging to a particular class. If only one class is present, all particles should show a value of 100. Frealign uses these values as weights to calculate a reconstruction.

Q: What is the meaning of the LogP column in the particle parameter file?

A: This column lists the log values of the likelihoods calculated for each particle. These values are used for Maximum Likelihood classification by Frealign. During the course of refinement, the scores should increase.

Q: What is the meaning of the SIGMA column in the particle parameter file?

A: The SIGMA values indicate the standard deviation of the noise in each particle. These values are used for Maximum Likelihood classification by Frealign.

Q: What is the meaning of the SCORE column in the particle parameter file?

A: The particle scores are a combination of correlation coefficients (multiplied by 100) between particles and reference and Maximum-Likelihood like parameter restraints. During the course of refinement, the scores should increase.

Q: What is a good SCORE, what not?

A: SCORE values vary widely according to the signal-to-noise ratio present in the images. For large particles with good contrast, the values tend to be higher (for example, 60 or 70) while for small particles they are lower (for example, 20 to 30). SCORE values also depend on the high resolution limit (keyword `res_high_refinement`) used because the data at higher resolution tends to be noisier than at lower resolution. Therefore, increasing the resolution limit usually decreases the SCORE values.

Q: What is the meaning of the CHANGE column in the particle parameter file?

A: This column lists the changes in the SCORE values from the previous cycle. Convergence of the refinement is usually indicated by smaller values in the CHANGE column compared to earlier refinement cycles. For a successful refinement, the score changes are typically in the range of 0.1 or smaller.

Q: Where can I find the two reconstructions calculated from even and odd particles that are used for the Fourier Shell Correlation?

A: The two reconstructions can be found in Frealign's scratch directory. They contain the string "maps1" and "map2" in their file names.

Q: How is the resolution limit used in the search/refinement set?

A: The resolution limit is set in `mparameters`, keyword `res_high_refinement`. This is the resolution used to low-pass filter the reference during particle alignment. It is not adjusted automatically and should be set carefully by the user according to the current resolution of the reference reconstruction (see FSC tables). In a successful refinement, the resolution indicated by the FSC (or `Part_FSC`) should significantly exceed that set by `res_high_refinement`. If the FSC indicates a resolution close to the user-set limit, this may indicate overfitting. In this case, the FSC curve cannot be used for a reliable resolution estimate.

Q: What is the difference between `mparameter` keywords `res_high_refinement` and `res_high_class`?

A: Like `res_high_refinement`, `res_high_class` indicates a resolution limit. However, the `res_high_class` limit is used to evaluate the log likelihood values used for classification. This gives the user the option to evaluate differences between particles at a lower resolution than the resolution used for particle alignment. If `res_high_refinement` is set to a lower resolution (larger values) than `res_high_class`, `res_high_class` is automatically adjusted to match `res_high_refinement`.

Q: When should I start classification?

A: Classification is started simply by setting the value for keyword `nclasses` to a number larger than 1 (in `mparameters`). It is recommended to do this after running a number of refinement cycles with a single class until the refinement converges (score changes are small on average).

Q: How do I sharpen the final map (apply a negative B-factor)?

A: Frealign comes with a program called `bfactor.exe`. It can be used to apply a negative B-factor and a low-pass filter (to limit the resolution of the map to the resolution limit indicated by the FSC curve). When applying a negative B-factor and low-pass filter, the cosine edge filter should be used. A Gaussian filter will essentially undo the B-factor sharpening. The amount of sharpening (the value of the negative B-factor) can be estimated from a Guinier plot. `bfactor.exe` estimates the B-factor that should be applied from such a plot in a user-specified resolution interval. However, unless the reconstruction has a resolution significantly higher than 10 Å this estimate can have a significant error. In this case, the best way to estimate the correct B-factor for sharpening is by trial and error: too much sharpening will lead to significant noise features in the map while too little sharpening will produce a map that does not display details that should be visible at the nominal resolution of the map.

Q: How do I low-pass filter my maps?

A: A low-pass filter can be applied using the program `bfactor.exe`. Both Gaussian and cosine-edge filters can be applied. For the cosine-edge filter, a width of 5 pixels (in Fourier space) for the edge usually yields good results.

Q: Can Frealign work with Spider particle image stacks?

A: No, if Spider formatted files are used, they have to be simple 3D files, i.e. files that have a single header and a series of 2D sections.

Q: Where can I find more information about different Frealign options and parameters?

A: Frealign comes with a `README.txt` file that can be found in the `man Frealign` directory. It contains more detailed information about input parameters.

Q: How can I merge classes that were produced by 3D classification?

A: Classes can be combined into a single class using the program `merge_classes.exe`. It reads the particle image stack and alignment parameter files of the classes to merge and writes out a new stack containing only particle images from the selected classes, as well as a single parameter file describing the particle alignments.

Q: How can I select particles belonging to a particular class?

A: A single class or multiple classes can be selected and combined into a new single class using the program `select_classes.exe`. It reads the particle image stack and alignment parameter files of all classes and allows the user to specify the classes that should be selected for output into a new stack and parameter file.

Q: Some programs ask to specify the image format as M, S or I. What does this mean?

A: M, S and I stand for MRC, Spider and IMAGIC format.

