

# FREALIGN: High-resolution refinement of single particle structures

Nikolaus Grigorieff \*

*Howard Hughes Medical Institute, Department of Biochemistry, Rosenstiel Basic Medical Sciences Research Center,  
Brandeis University, 415 South Street, Waltham, MA 02454, USA*

Received 4 April 2006; received in revised form 14 May 2006; accepted 16 May 2006  
Available online 2 June 2006

## Abstract

The refinement of three-dimensional reconstructions and correction for the contrast transfer function of the microscope are important steps in the determination of macromolecular structures by single particle electron microscopy. The algorithms implemented in the computer program FREALIGN are optimized to perform these tasks efficiently. A general overview and details on how to use FREALIGN are provided. The program is free and available for download on the author's web page.

© 2006 Elsevier Inc. All rights reserved.

*Keywords:* Image processing; Electron microscopy; Protein structure; FREALIGN

## 1. Introduction

The determination of protein structures by electron microscopy of isolated molecules and complexes (single particles) has become one of the main techniques used to visualize large cellular structures, such as the ribosome (e.g. Klaholz et al., 2004; Rawat et al., 2006) or spliceosome (Boehringer et al., 2004; Jurica et al., 2004), or smaller complexes, such as the transferrin receptor (Cheng et al., 2004). In many cases, this technique now yields structures at sub-nanometer resolution where secondary structure can be recognized (e.g. Cheng et al., 2004; Fotin et al., 2004; Halic et al., 2005; Ludtke et al., 2004). The structure determination from images of single particles usually begins with the selection of particles, a process that can now be automated (Zhu et al., 2004). Several steps then follow in which the particle images are aligned and classified. An initial three-dimensional (3D) reconstruction is obtained from these class averages using, for example, the random conical tilt (Radermacher, 1988) or angular reconstitution procedure (van Heel, 1987). At this stage,

the five alignment parameters for each particle ( $x$ ,  $y$  translation and three Euler angles) are roughly known. For an improved structure, these parameters need to be refined. The most commonly used refinement procedures are based on an iterative algorithm in which an existing 3D structure is used as a reference to determine the initial alignment parameters with higher accuracy. Iterations of alternating 3D reconstruction, using the improved alignment parameters, and realignment of the particles based on the updated 3D structure, are performed until the alignment parameters and 3D structure remain constant. This type of refinement may not converge on the correct or true set of the parameters but, instead, on an incorrect local optimum. Global optima can be found using a systematic grid search. The computer program FREALIGN (Fourier REconstruction and ALIGNment) has been designed to carry out efficient refinement of single particle structures, starting from an initial structure and approximately known single particle alignment parameters (Grigorieff, 1998). FREALIGN can also be used to determine initial alignment parameters, or to find global optima, in a systematic grid search using a reference 3D structure.

This article is primarily intended as an overview of FREALIGN features with important information on how to use them. Some of the theoretical background

\* Corresponding author. Fax: +1 781 736 2419.  
E-mail address: [niko@brandeis.edu](mailto:niko@brandeis.edu)

and details on the algorithms implemented in FREALIGN are described elsewhere (Grigorieff, 1998; Rosenthal and Henderson, 2003; Stewart and Grigorieff, 2004; Wolf et al., 2006).

## 2. Overview and general philosophy

FREALIGN differs from other well-established image processing packages, such as SPIDER (Frank et al., 1996), IMAGIC (van Heel et al., 1996) and EMAN (Ludtke et al., 1999). While these other packages are designed to perform all necessary procedures to obtain a refined 3D structure from single particle images, the purpose of FREALIGN has always been focused on refinement and 3D reconstruction. The algorithms in FREALIGN were first implemented in 1996 to provide an efficient way to calculate reconstructions fully corrected for the contrast transfer function (CTF) of the electron microscope. This development coincided with similar developments for SPIDER (Zhu et al., 1997). The new algorithms in FREALIGN also introduced an efficient procedure for refining 3D structures by working entirely in Fourier space. This considerably accelerated the CTF correction and calculation of projections to match the images of single particles. CTF correction and high-resolution refinement are now also routinely included in other image processing packages. Nevertheless, FREALIGN implements one of the fastest refinement and reconstruction algorithms, and it has produced some of the highest resolution single particle structures (Cheng et al., 2004; Fotin et al., 2004; Furst et al., 2003; Rosenthal and Henderson, 2003). It is entirely written in Fortran 77 and currently can read image files in the MRC/CCP4 and SPIDER format. It is usually run as part of a shell script that provides the necessary control parameters and input/output files names.

## 3. Basic algorithm

The main part of the algorithm implemented in FREALIGN has been described previously (Grigorieff, 1998) and performs three tasks: CTF correction of the image data, 3D reconstruction, and refinement of the reconstruction. The CTF is calculated including astigmatism and both amplitude and phase contrast, with defocus parameters determined previously (e.g. using CTFFIND3, Mindell and Grigorieff, 2003). The Fourier transform of the image of a single particle is then multiplied by the CTF and the image phases are shifted to move the particle to the image center, according to the particle's  $x, y$  translation determined in a previous alignment step. After an inverse Fourier transform a circular mask with radius RO (see below) and a cosine edge is applied to mask out noisy background. After Fourier-transforming a third time the image phases are shifted to move the particle to the origin of the image. The two-dimensional (2D) Fourier transform of the image is then oriented in 3D according to the angles determined in a previous alignment cycle and added to the 3D Fourier transform

of the new reconstruction. Since, in general, the sample points of the discrete 2D Fourier transform of the image will not coincide with sample points of the Fourier transform of the reconstruction, an interpolating function has to be used. Such a function is derived from the Fourier transform of a box with dimensions equal to the dimensions of the reconstructed volume. Thus, in the interpolation step, a box transformation is centered on each sample of the 2D Fourier transform of the image and evaluated at the nearest neighbor of the 3D Fourier transform of the reconstruction. In the process of the reconstruction from images of all particles, sums are accumulated at each sample point of the 3D Fourier transform:

$$R_i = \frac{\sum_j w_j^2 b^2 c_j P_{ij}}{f + \sum_j (w_j b c_j)^2} \quad (1)$$

Here,  $R_i$  represents sample  $i$  in the 3D Fourier transform of the reconstruction,  $P_{ij}$  is a sample from the Fourier transform of particle image  $j$  (before CTF correction) contributing to sample  $R_i$ ,  $c_j$  is the CTF for image  $j$  corresponding to that point in the image,  $b$  is the box transformation, and  $w_j$  is a weighting factor describing the quality of the image.  $f$  is a constant, similar to a Wiener filter constant, which prevents over-amplification of terms when the rest of the denominator is small. A Wiener filter constant represents the noise level in the data which varies for each sample  $R_i$ . However, the noise level is not explicitly determined here, and  $f$  is set to a fixed value keeping a balance between noise amplification and signal reproduction at each point. Since data are normally collected at different defocuses the term  $f$  has its largest effect at low resolution where the sum in the denominator assumes its smallest values.  $f$  is set to be 10% of the average value (averaged over all samples in the 3D Fourier transform) of the sum in the denominator. Since the image of a particle recorded in the microscope has the CTF already applied once in the microscope, one can write

$$P_{ij} = c_j O_{ij} \quad (2)$$

with  $O_{ij}$  the sample of the Fourier transform of the projection of the original particle.  $O_{ij}$  is never actually observed in the microscope. However, when combining Eqs. (1) and (2) it becomes clear that Eq. (1) represents a least-squares fit of sample  $R_i$  to all samples  $O_{ij}$  with sample weights  $w_j b c_j$ . A similar least-squares fit was used in the merging of images obtained from viruses by electron cryo-microscopy (Bottcher and Crowther, 1996; Bottcher et al., 1997). The weighting factor  $w_j$  allows weighting of contributions from individual images according to their correlation coefficients with the reference. The weighting factor is given as

$$w_j = \exp(-ap_j g^2) \quad (3)$$

where  $g$  is the spatial frequency in  $\text{\AA}^{-1}$ ,  $p_j$  is the arc cosine of the correlation coefficient between image  $j$  and the reference and  $a$  is a constant converting the phase residual into an appropriate temperature factor (see Eq. (5)).

The reconstruction described by Eq. (1) is corrected for the CTF and, therefore, accomplishes the first two tasks listed above. For refinement, the Fourier transform of the image is oriented in 3D according to the previously determined particle angles and the image phases are shifted to move the particle to the image origin. A weighted correlation coefficient (Stewart and Grigorieff, 2004) is calculated between the image Fourier transform and a 3D Fourier transform of the reference. As before, the Fourier transform of a box is used to interpolate between the two Fourier transforms. To refine the particle angles and  $x$ ,  $y$  translation, a Powell optimization algorithm (Harwell Numerical Library 1979, Harwell, UK) is used to maximize the correlation coefficient for each image.

#### 4. Running FREALIGN

Fig. 1 shows a flow chart with input and output data. Input consists of control parameters needed to run

FREALIGN, image data (particle image stack, 3D reference reconstruction) and particle parameters (Euler angles,  $x$ ,  $y$  translation, magnification, defocus and astigmatism, arc cosine of the correlation coefficient). A list of control parameters is provided in Table 1. Output includes image data (refined 3D reconstruction—this overwrites the input 3D reference, matching projections, a number of 3D diagnostic files), refined particle parameters and changes in these parameters compared with the input. FREALIGN performs one cycle of refinement. To carry out several cycles, the refined 3D reconstruction and new particle parameters have to be used as new inputs for another FREALIGN cycle. This iteration is usually performed as a loop in a script, for example a shell script used to run FREALIGN. The script also provides the control parameters and names of the input and output files such as image files and particle parameter files. Image files and particle parameters are stored in separate text files. The flow chart in Fig. 2

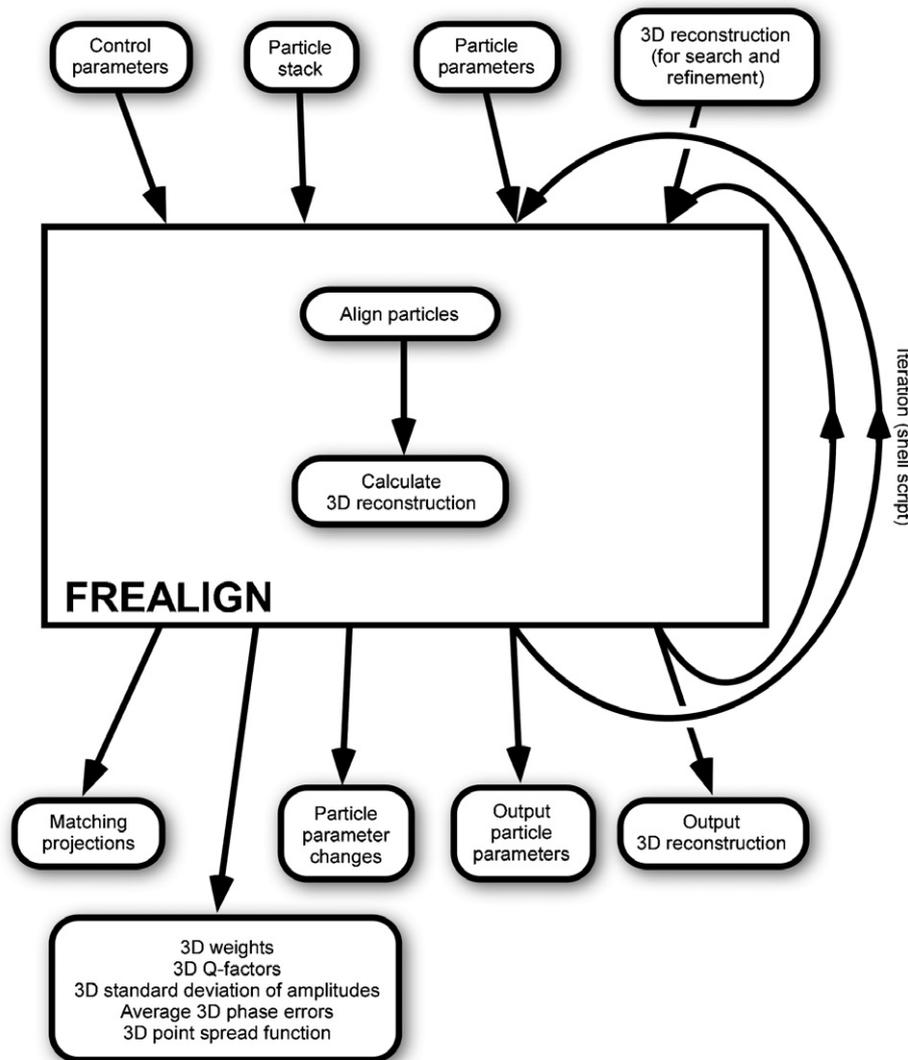


Fig. 1. Overview of input and output parameters and files for FREALIGN. Using the data provided on input, FREALIGN performs one cycle of particle refinement and reconstruction. For repeated cycles, FREALIGN needs to be called from a script, for example a shell script, and the refined 3D reconstruction and output particle parameters replace the inputs from the previous cycle.

Table 1  
List of control parameters used to run FREALIGN

Card 1	CFORM,IFLAG,FMAG,FDEF,FASTIG,FFLIP,IEWALD,FMATCH,FHIST,FBEAUT,FCREF CFORM—Format of particle images and 3D reference structure (M or S for MRC or Spider format, respectively) IFLAG—Mode of operation (can be -4, -3, 0, 1, 2, 3 or 4). See main text for more detail. FMAG—Magnification refinement (F or T to disable/enable magnification refinement) FDEF—Defocus refinement (F or T to disable/enable defocus parameter refinement) FASTIG—Astigmatism refinement (F or T to disable/enable astigmatic angle refinement) FFLIP—Flip second Euler angle during a search with IFLAG >1 or IFLAG <0 (F or T to disable/enable, this option doubles the number of search angles tried in a search) IEWALD—Ewald sphere correction (0 to disable, 1 or 2 to enable using a simple or reference-based correction, -1 or -2 to enable correction with reversed handedness of the reconstruction) FMATCH—Matching projections (F or T to disable/enable output of matching projections) FHIST—History of ITMAX randomized trials in a search (F or T to disable/enable output) FBEAUT—Real-space symmetrization (F or T to disable/enable symmetrization) FCREF—Figure-of-merit filter (F or T to disable/enable application of the filter to the final reconstruction)
Card 2	RO,RI,PSIZE,WGH,XSTD,PBC,BOFF,DANG,ITMAX,IPMAX RO, RI—Outer and inner radius of the mask in Angstrom that is applied to the 3D reconstruction before output. A mask with radius RO is also applied to the input particle images PSIZE—Pixel size in Angstrom of the input and output 3D reconstruction WGH—Amplitude contrast used in the contrast transfer function correction XSTD—Threshold to for masking the input 3D structure (XSTD > 0, this is similar to solvent flattening) or the input particle images (XSTD < 0). In both cases, tight and shape-dependent masks are derived from a low-pass filtered input 3D structure. XSTD = 0 disables this function PBC—Resolution-dependent weighting of particle images for 3D reconstruction. A small value, for example 4, results in stronger discrimination between particles with different phase residuals while a large value, for example 100, results in approximately equal weighting of all particles BOFF—Approximate value of the average phase residual of all particles used for reconstruction DANG—Angular step size in degrees to generate equally spaced search angles, for example 10 ITMAX—Number of randomized search trials, for example 10 IPMAX—Number of potential matches in a search to be refined further to find the best match
Card 3	MASK—[0/1] parameters to include in refinement (usually 1 1 1 1 1)
Card 4	IFIRST,ILAST—First and last particles to include
Card 5	ASYM—Symmetry required Cn, Dn, T, O, I, I1, I2 or N (can be zero)
Card 5a	SYMOP—Symmetry matrix, only if N and N≠0, ((SYMOP(J,K,I), J = 1,3), K = 1,3), I = 1,N
Cards 6–12 describe	each dataset, terminating with RELMAG=0 (or -100 to disable calculation of a 3D reconstruction)
Card 6	RELMAG,DSTEP,TARGET,THRESH,CS,AKV,TX,TY RELMAG—Relative magnification to apply (1.0) DSTEP—Densitometer step size in microns for this data set (7.0) TARGET—Target phase residual for search/refine (15.0) THRESH—Worst phase residual for inclusion (90.0) CS,AKV—Spherical aberration coefficient, acceleration voltage (2.0, 120.0) TX,TY—Beam tilt in X, Y direction (0.0, 0.0)
Card 7	RREC,RMAX1,RMAX2,RBFACT—Resolution of reconstr., refinement low/high, B-factor
Card 8	FINPAT1—Particle image stack file name
Card 9	FINPAT2—Matching projections file name (if FMATCH=T)
Card 10	At least one of the following is required:
(a)	FINPAR—Input particle parameter file, required if IFLAG=0,1,2,3,4
(b)	NIN,ABSMAGPIN,IFILMIN,DFMID1IN,DFMID2IN,ANGASTIN,MORE—if IFLAG <0 NIN,ABSMAGPIN,IFILMIN—Number of particles, magnification, film number DFMID1IN,DFMID2IN,ANGASTIN—Defocus and astigmatism MORE—If set to 1 then more cards, if set to 0 then terminate card 10 For option (b), the information on these cards is used to create a new parameter file
Card 11	FOUTPAR—Output particle parameter file
Card 12	FOUTSH—Output particle parameter shift file (parameter changes compared with input)
Card 13	F3D—3D reference structure for input and refined structure for output (overwrites input)
Card 14	FWEIGH—3D weights file for output (sum of weights in the denominator of Eq. (1))
Card 15	FQF—3D Q-factor file for output (Q-factors for each voxel in Fourier space)
Card 16	FAMP—3D amplitude standard deviation file for output (rms amplitudes in Fourier space)
Card 17	FPHA—3D phase residual file for output (average phase diff. between images and reference)
Card 18	FPOI—3D point spread function for output (indicating anisotropies in resolution in real space)

Each card contains one line of parameters.



Table 2  
List of Fortran parameters to set the size of some arrays used in FREALIGN

NN1	Maximum pixel dimensions of particles (for example, 128)
NNBIG	Maximum pixel dimensions for padded reference structure generated inside FREALIGN (for example, 512). The padding accelerates some interpolation steps (see main text). NNBIG has to be at least as big as NN1 and can be as big as four times NN1 to achieve best performance
NNPART	Maximum number of particles from all input data sets that can be processed (for example, 150000)
NNSTAT	Flag to switch the calculation of output statistics (output files FQF, FAMP, FPHA and FPOI) on (NNSTAT = 1) or off (NNSTAT = 0). Switching output statistics off lowers the memory requirements to run FREALIGN

The mode of operation is controlled by an integer called IFLAG that can assume 0 (reconstruction only), 1 (local parameter refinement followed by reconstruction), 2 (local refinement with ITMAX random starting angles), 3 (systematic parameter search), 4 (systematic search with ITMAX random starting angles and subsequent local refinement), -4 (similar to 4, but with the creation of a new particle parameter file) and -3 (similar to 3, but with the creation of a new particle parameter file). Local refinement refers here to optimization using a Powell conjugate gradient algorithm that starts from a previous estimate of the parameters while systematic search refers to a grid search over the entire parameter space. Modes 0 and 1 are the standard modes for reconstruction and refinement. Mode 2 is useful for testing, after refinement, whether the final particle parameters represent a global optimum or whether other, better alignment parameters can be found. In this mode, many different random starting angles are tested. The number of starting angles to be tested is given by the control parameter ITMAX that could, for example, be set to 100. Ideally, for a refined structure, this randomized search should not find different parameters for most of the particles. Mode 3 is used to determine initial parameters for particles when a 3D reference structure is available. The angular step size for the systematic search is given by DANG, and angles are calculated according to a formula given by Penczek et al. (1994). Option -3 is useful, in this respect, because it also creates a parameter file that would otherwise have to be created by the user from scratch with randomly assigned parameters. Option 4 is a combination of mode 2 and 3. In this mode, repeated systematic searches are carried out as for mode 3, but each systematic search starts with random starting angles. This allows two different search strategies. In the first strategy, the number of repeated systematic searches (parameter ITMAX) is small (e.g. 5) and the angular step size (DANG) is also small (e.g. 5 degrees), resulting in a tight, but time-consuming search. In the second strategy, the number of repeated searches is large (e.g. 100) and the angular step size is also large (e.g. 100 degrees). In some studies it was found that the second orientation is more successful in identifying the correct orientation of particles (Rosenthal and Henderson, 2003).

Other control parameters can be set to enable refinement of CTF parameters (FDEF, FASTIG) and magnification (FMAG), to write out matching projections (FMATCH), to correct for the Ewald sphere curvature (IEWALD, Wolf

et al., 2006), to impose real-space symmetry on a reconstruction (FBEAUT) or to apply a filter based on a figure-of-merit to the reconstruction (FCREF, Rosenthal and Henderson, 2003). If a symmetry operator is supplied on input (ASYM), the symmetry is always imposed on the reconstruction by inserting multiple copies of the Fourier transform of the image into the Fourier transform of the structure that is being reconstructed, using all possible symmetry-related orientations. The option FBEAUT only controls an additional real space symmetrization of the structure at the end of the reconstruction process. This is sometimes useful in removing small asymmetries in the reconstruction that arise from interpolation errors in reciprocal space. Finally, the exclusion of any of the five particle parameters (three Euler angles and  $x$ ,  $y$  translation) from the search or refinement can be controlled by the MASK parameter that is usually set to 1 1 1 1 1 (all parameters are active). Setting any of the MASK entries to 0 will deactivate the corresponding parameter (Euler angles  $\psi$ ,  $\theta$ ,  $\phi$ ; translations  $x$ ,  $y$ ), which can be useful in certain procedures where only some parameters should be refined.

#### 4.2. Parameters in the Fortran code (Table 2)

There are a few parameters that determine the dimensions or arrays in FREALIGN. They are all set in the Fortran file `frealign_vX.f` where X is the program version and currently equal to 7. The parameter determining the maximum linear dimensions of input particle images and the 3D reference structure is NN1. It is set, for example, to 128. A second parameter that determines the amount of padding of the 3D reference structure (see below) is NNBIG. It has to be greater or equal to NN1 and, if system resources allow it, should be set to four times the linear dimensions of the input 3D reference structure. The parameter NNPART determines the maximum number of particles that can be used (see below) and is set, for example, to 150000. When one of these parameters is changed the program needs to be recompiled. To avoid the need to recompile for every new project, the dimensions should be set to the maximum values possible given the available system memory. The system memory requirements for running FREALIGN are determined mainly by the parameters NN1 and NNBIG. When NN1 = NNBIG = 256, about 900 MB of memory are needed. When these two parameters are set to 512, about 5.8 GB are needed. An approximate formula for the needed system memory M (in bytes) is

$$M = 40 * NN1^3 + 4 * NNBIG^3 + 200 * 10^6. \quad (4)$$

It should be noted that the pixel dimensions of particle images and 3D structures is limited to even numbers.

The evaluation of some statistical output arrays (FQF, FAMP, FPFA and FPOI, see Fig. 1) can be disabled by setting the Fortran parameter NNSTAT to 0 (it is normally set to 1). Disabling the statistical arrays reduces the memory requirements of FREALIGN substantially, to about

$$M = 24 * NN1^3 + 4 * NNBIG^3 + 200 * 10^6. \quad (5)$$

#### 4.3. Multiple data sets

It is possible to supply multiple data sets on input. These data sets can differ in their native magnification, microscope parameters (spherical aberration coefficient  $C_s$ , acceleration voltage, beam tilt) and the resolution interval that should be included in the refinement and reconstruction (control parameters RMAX1 and RMAX2, see Table 1). Thus, data sets collected on different days and on different microscopes can be merged. Each data set requires the input of a separate particle image stack and a parameter file. Currently, up to 20 separate data sets can be used, and up to 150000 particles can be included in total. The limit of the number of particles can be increased, however, by changing the dimension variable NNPART in the Fortran source code (see above).

#### 4.4. Particle parameters

Euler angles, translation, magnification and defocus are defined for each particle image in a stack. These parameters are stored in separate text (ASCII) files for each data set and the file names are provided on input. The conventions used for the Euler angles are the same as those used in SPIDER (Frank et al., 1996). First, the object is rotated clockwise around the  $z$ -axis (angle phi), followed by a clockwise rotation around the  $y$ -axis (theta) and a clockwise rotation around the new  $z$ -axis (psi). The translations are given in pixels. It is important that the order of particles in the parameter file corresponds with the order of particles in the stack. Furthermore, in the parameter file, the micrograph from which each particle was chosen is noted. It is important that particles from the same micrograph are kept together in the parameter file to facilitate the input of all parameters from one micrograph when the magnification or CTF parameters are to be refined. The refinement of these parameters is done for entire groups of particles from single micrographs to boost the signal and accuracy of the refinement. However, if data are included in the refinement at a resolution higher than 10 Å, as specified by the control parameter RMAX2, the refinement of CTF parameters is done individually for each particle.

In a refinement, the parameters from a previous cycle are read in, refined, and written out into a new parameter file. Any changes in the parameters are recorded in a

parameter shift file that can be used to track the convergence of the refinement. The image stack is not altered, nor is a new stack created with images in their refined orientation and translation. Therefore, the parameter file contains all the information needed to reconstruct a structure from the original, windowed particles. This strategy saves storage space because it does not duplicate the image stack, which is often quite large. It also avoids the accumulation of interpolation errors that may arise from repeated rotation and translation of the images. The output parameter file also records the weighted correlation coefficient between particle and reference (see below) and its change from the previous cycle. For historical reasons, the correlation coefficient is converted, using the arc cosine function, to produce a number that corresponds, to some degree, with a phase residual. Therefore, in a successful refinement, the recorded 'phase residuals' (PRES) in the parameter file will become smaller.

#### 4.5. Data flow

After reading control and particle parameters, search angles are generated if IFLAG is greater than 1 (or negative). For the purpose of this discussion, the case of local refinement is considered (IFLAG = 1). In this case, the 3D reference structure is read in (see Fig. 2). Depending on the available system memory, this structure is padded to twice or four times the dimensions of the original structure before Fourier transformation. The padding leads to an over-sampling of the Fourier transform and improved interpolation when calculating projections (central sections through the Fourier transform). The amount of padding is determined in the program by the Fortran dimension variable NNBIG (see above). For example, for fourfold padding, NNBIG has to be at least four times the linear pixel dimension of the input 3D reference structure. When fourfold padding is possible, the calculation of central sections is significantly accelerated because the interpolation between the 3D array of the reference structure and the 2D array of the central section is reduced to include only the nearest neighbor.

An input 3D structure is required for any type of refinement or search, when matching projections are to be generated, or when the 3D reference structure is used to derive masks (XSTD not equal to 0). The type of mask used depends on the sign of XSTD. For a positive value, the 3D reference is masked with a soft-edged mask derived by thresholding a low-pass filtered version of the 3D reference. The value of XSTD is taken as the threshold in multiples of standard deviations above the mean of the low-pass filtered reference. This process is analogous to solvent flattening in X-ray crystallography. If XSTD is negative, its absolute value is also used to threshold a low-pass filtered version of the 3D reference. However, the resulting binary mask is not used to mask the 3D reference, but to calculate projections corresponding with the particle images. These projections resemble the shape of the particles

and, after applying a soft edge, are used as tighter “customized” masks for the input particle images.

FREALIGN is then ready to process particles, one at a time. A particle image is read from the input image stack, Fourier transformed, multiplied with the CTF, inverse-Fourier transformed, masked (if XSTD is greater or equal to 0, this is just a soft-edged circular mask), and Fourier transformed again. The conventions for defocus values and astigmatic angle are the same as in the MRC image processing package (Crowther et al., 1996) and the programs CTFFIND3 and CTFTILT (Mindell and Grigorieff, 2003). The CTF multiplication before the masking avoids some of the loss of high-resolution information, which is contained in the fringes surrounding each particle, due to the image defocus (Rosenthal and Henderson, 2003). Euler angles and translation are refined by maximizing a weighted correlation coefficient between particle image and projection (central section) of the 3D reference structure (Grigorieff, 1998; Stewart and Grigorieff, 2004, see above). The refinement starts off with parameters from a previous refinement cycle and uses a Powell optimization algorithm. Therefore, this refinement includes only a local search. For a global search, one of the other modes of operation must be used (IFLAG greater than 1). The search and refinement can be influenced by amplification or attenuation of high-resolution terms using the control parameter RBFACT (Rosenthal and Henderson, 2003). Fourier terms in the particle image are multiplied by

$$w_B = \exp[-RBFACT/4 * g^2]. \quad (6)$$

As before,  $g$  is the special frequency in  $\text{\AA}^{-1}$ . Some important notes concerning the use of RBFACT should be made here. A non-zero value for RBFACT changes the weighting of the Fourier terms in the particle image and can be used to tune search and refinement operations. However, a negative value for RBFACT that results in the amplification of terms at higher resolution can cause over-fitting and noise bias (Stewart and Grigorieff, 2004). Over-fitting leads to artificially high values in the commonly used Fourier Shell Correlation (FSC) curve to estimate the resolution of the reconstruction (Harauz and van Heel, 1986). It also reduces the actual resolution of the reconstruction. FREALIGN uses an automated weighting scheme designed to optimize the refinement (Stewart and Grigorieff, 2004). Setting RBFACT to a negative value (e.g. -1000) during a limited number of refinement cycles (e.g. 3 cycles) may be useful in “shaking up” the particle parameters to lead the refinement out of local optima. However, refinement cycles with non-zero RBFACT values should always be followed by a few refinement cycles (e.g. 5 cycles) with RBFACT set to zero to ensure optimal weighting of particle image data and a more reliable FSC curve.

Upon convergence of the refinement, the particle image is inserted into the 3D Fourier transform of the reconstruction, and the refined parameters are written into a new parameter file. A resolution-dependent weighting factor

$w_j$  (see Eq. (3)) is used to reflect the quality of the particle (Grigorieff, 1998)

$$w_j = \exp[-(PRES - BOFF)/PBC * g^2]. \quad (7)$$

PRES is the arc cosine of the weighted correlation coefficient obtained in the refinement, BOFF is the average of PRES values of all particles, and PBC is a constant set by the user. BOFF can either be provided by the user and set to approximately the average of all PRES values, or it can be set to 0 on input and will then be calculated internally by using PRES values listed in the input particle parameter file. The value PBC determines how much weighting should be done: a large value (e.g. 100) will make the weighting function relatively insensitive to differences in PRES between particles, while a small value (e.g. 4) will discriminate more between particles with different PRES values.

Internally, two 3D Fourier transforms are updated, one with even-numbered particles and the other with odd-numbered particles. The two reconstructions are used at the end of the reconstruction process to evaluate the FSC curve for estimating the resolution. Any changes in the parameters are written into a separate file (see above). If matching projections are requested, the projection matching the particle is written into a stack in a montage with the original particle image, for easy comparison. The matching projections for particle images with pixel dimensions greater than or equal to 128 are demagnified to enable a more efficient visual inspection of the matching projections. A visual inspection of matching projections is recommended to ensure proper alignment of particles, especially when their initial orientations are determined in a search.

When all particles have been processed, the sum of the two reconstructions is inverse-Fourier transformed, masked with a spherical mask and written out with other files that provide information about the distribution of the data in the reconstruction, errors, and a point spread function (Grigorieff, 1998). A table indicating the FSC curve (Harauz and van Heel, 1986), differential phase residual (DPR, Frank et al., 1981; Penczek et al., 1994), Q-factor (QFACT, Kessel et al., 1985; van Heel and Hollenberg, 1980) and spectral signal-to-noise ratio (SSNR, Unser et al., 1987, 1989) is also written out.

#### 4.6. Distributed computing

The refinement of  $N$  particles can easily be distributed across many nodes of a computer cluster by running several instances of FREALIGN in parallel using IFLAG = 1. Each instance processes a subset of particles that can be specified with the control parameters (IFIRST, ILAST). By setting the control parameter RELMAG in the line terminating the list of input data sets to -100.0, the calculation of 3D reconstructions is omitted, speeding up the refinement. RELMAG is used to set the nominal magnification of a data set and is usually set to 1.0, except in the line terminating the list of input data sets. When all

parameters have been refined, the output parameter files from all instances of FREALIGN can be combined and used in a final run of FREALIGN to calculate a new 3D reconstruction (IFLAG = 0). Shell scripts controlling the distribution of FREALIGN jobs on a Linux or UNIX cluster are included with the FREALIGN release. The scripts allow specification of the number of processors that should be used and they perform the partitioning of the data. For this distribution method no special environment is required, such as MPI or other parallel computing environments. The only requirement is that all computers running instances of FREALIGN have access through a network to the same data disk to retrieve and store parameters.

### Acknowledgments

The author thanks Richard Henderson for contributing to the code for refinement, and implementing symmetry operations, Matthias Wolf for contributing to code necessary to accommodate the Ewald sphere curvature, and Christine Villeneuve for restructuring the code to make it more modular. These people also contributed with discussions and ideas, as did the following people: Peter Rosenthal, Tony Crowther, Alex Stewart, David DeRosier and Alex Fotin. Extensive testing of FREALIGN was performed by Johannes Fürst, Thomas Walz and Yifan Cheng. The author gratefully acknowledges financial support from the German Research Foundation (research fellowship), Medical Research Council, UK, the Humboldt Foundation (research fellowship), and the National Institutes of Health, Grant 1 P01 GM-62580.

### References

- Boehringer, D., Makarov, E.M., Sander, B., Makarova, O.V., Kastner, B., Luhrmann, R., Stark, H., 2004. Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nat. Struct. Mol. Biol.* 11, 463–468.
- Bottcher, B., Crowther, R.A., 1996. Difference imaging reveals ordered regions of RNA in turnip yellow mosaic virus. *Structure* 4, 387–394.
- Bottcher, B., Wynne, S.A., Crowther, R.A., 1997. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* 386, 88–91.
- Cheng, Y., Zak, O., Aisen, P., Harrison, S.C., Walz, T., 2004. Structure of the human transferrin receptor-transferrin complex. *Cell* 116, 565–576.
- Crowther, R.A., Henderson, R., Smith, J.M., 1996. MRC image processing programs. *J. Struct. Biol.* 116, 9–16.
- Fotin, A., Cheng, Y., Sliz, P., Grigorieff, N., Harrison, S.C., Kirchhausen, T., Walz, T., 2004. Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature* 432, 573–579.
- Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., Leith, A., 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* 116, 190–199.
- Frank, J., Verschoor, A., Boublik, M., 1981. Computer averaging of electron micrographs of 40S ribosomal subunits. *Science* 214, 1353–1355.
- Furst, J., Sutton, R.B., Chen, J., Brunger, A.T., Grigorieff, N., 2003. Electron cryomicroscopy structure of *N*-ethyl maleimide sensitive factor at 11 Å resolution. *EMBO J.* 22, 4365–4374.
- Grigorieff, N., 1998. Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 Å in ice. *J. Mol. Biol.* 277, 1033–1046.
- Halic, M., Becker, T., Frank, J., Spahn, C.M., Beckmann, R., 2005. Localization and dynamic behavior of ribosomal protein L30e. *Nat. Struct. Mol. Biol.* 12, 467–468.
- Harauz, G., van Heel, M., 1986. Exact filters for general geometry 3-dimensional reconstruction. *Optik* 73, 146–156.
- Jurica, M.S., Sousa, D., Moore, M.J., Grigorieff, N., 2004. Three-dimensional structure of C complex spliceosomes by electron microscopy. *Nat. Struct. Mol. Biol.* 11, 265–269.
- Kessel, M., Radermacher, M., Frank, J., 1985. The structure of the stalk surface layer of a brine pond microorganism: correlation averaging applied to a double layered lattice structure. *J. Microsc.* 139 (Pt 1), 63–74.
- Klaholz, B.P., Myasnikov, A.G., Van Heel, M., 2004. Visualization of release factor 3 on the ribosome during termination of protein synthesis. *Nature* 427, 862–865.
- Ludtke, S.J., Baldwin, P.R., Chiu, W., 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* 128, 82–97.
- Ludtke, S.J., Chen, D.H., Song, J.L., Chuang, D.T., Chiu, W., 2004. Seeing GroEL at 6 Å resolution by single particle electron cryomicroscopy. *Structure* 12, 1129–1136.
- Mindell, J.A., Grigorieff, N., 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* 142, 334–347.
- Penczek, P.A., Grassucci, R.A., Frank, J., 1994. The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* 53, 251–270.
- Radermacher, M., 1988. Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *J. Electron Microsc. Tech.* 9, 359–394.
- Rawat, U., Gao, H., Zavialov, A., Gursky, R., Ehrenberg, M., Frank, J., 2006. Interactions of the release factor RF1 with the ribosome as revealed by Cryo-EM. *J. Mol. Biol.* 357, 1144–1153.
- Rosenthal, P.B., Henderson, R., 2003. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* 333, 721–745.
- Stewart, A., Grigorieff, N., 2004. Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy* 102, 67–84.
- Unser, M., Trus, B.L., Frank, J., Steven, A.C., 1989. The spectral signal-to-noise ratio resolution criterion: computational efficiency and statistical precision. *Ultramicroscopy* 30, 429–433.
- Unser, M., Trus, B.L., Steven, A.C., 1987. A new resolution criterion based on spectral signal-to-noise ratios. *Ultramicroscopy* 23, 39–51.
- van Heel, M., 1987. Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* 21, 111–123.
- van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., Schatz, M., 1996. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* 116, 17–24.
- van Heel, M., Hollenberg, J., 1980. The stretching of distorted images of two-dimensional crystals. In: Baumeister, W. (Ed.), *Electron Microscopy at Molecular Dimensions*. Springer, Berlin, New York, pp. 256–260.
- Wolf, M., DeRosier, D.J., Grigorieff, N., 2006. Ewald sphere correction for single-particle electron microscopy. *Ultramicroscopy* 106, 376–382.
- Zhu, J., Penczek, P.A., Schroder, R., Frank, J., 1997. Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S *Escherichia coli* ribosome. *J. Struct. Biol.* 118, 197–219.
- Zhu, Y., Carragher, B., Glaeser, R.M., Fellmann, D., Bajaj, C., Bern, M., Mouche, F., de Haas, F., Hall, R.J., Kriegman, D.J., et al., 2004. Automatic particle selection: results of a comparative study. *J. Struct. Biol.* 145, 3–14.