CHAPTER SIX

Processing of Structurally Heterogeneous Cryo-EM Data in RELION

S.H.W. Scheres[1]
MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, United Kingdom
[1]Corresponding author: e-mail address: scheres@mrc-lmb.cam.ac.uk

Contents

1. Introduction 126
2. New Algorithmic Concepts 128
   2.1 Regularization: The Empirical Bayesian Approach 128
   2.2 Prevention of Overfitting: The Gold-Standard Approach to Refinement 129
   2.3 Getting Clean, High-Resolution Maps: The Postprocessing Approach 131
   2.4 Beam-Induced Motion Correction and Radiation-Damage Weighting: The Movie-Processing Approach 132
3. A Typical High-Resolution Structure Determination Procedure 134
4. Dealing with Structural Heterogeneity 142
   4.1 3D Classification with Exhaustive Angular Searches 142
   4.2 Detection of Remaining Structural Heterogeneity 143
   4.3 3D Classification with Finer, Local Angular Searches 144
   4.4 Masked 3D Auto-Refinement 144
   4.5 Masked 3D Classification 146
   4.6 Masked 3D Refinement/Classification with Partial Signal Subtraction 147
   4.7 Dealing with Pseudo-Symmetry 150
   4.8 Multibody Refinement 151
   4.9 A More Elaborate Example 152
5. Outlook 154
Acknowledgments 154
References 154

Abstract

This chapter describes algorithmic advances in the RELION software, and how these are used in high-resolution cryo-electron microscopy (cryo-EM) structure determination. Since the presence of projections of different three-dimensional structures in the dataset probably represents the biggest challenge in cryo-EM data processing, special

Methods in Enzymology, Volume 579
ISSN 0076-6879
http://dx.doi.org/10.1016/bs.mie.2016.04.012

emphasis is placed on how to deal with structurally heterogeneous datasets. As such, this chapter aims to be of practical help to those who wish to use RELION in their cryo-EM structure determination efforts.

1. INTRODUCTION

Over the past two decades, statistical methods have become increasingly popular for the processing of cryo-electron microscopy (cryo-EM) images. In 2010, we wrote a review in this same series of *Methods in Enzymology* about a class of methods that are based on the optimization of a likelihood function (Sigworth, Doerschuk, Carazo, & Scheres, 2010). In that review, we explained how the averaging of two-dimensional (2D) projection images, or similarly their reconstruction into a three-dimensional (3D) map, can be considered as an incomplete data problem. The incompleteness of the data lies in the fact that the relative orientations of the 2D projection images are unknown, since the individual macromolecular complexes adopt random orientations on the experimental support. The way in which these unknown, or hidden, parameters are treated constitutes the main difference between the maximum-likelihood approach and alternative, least-squares methods that had dominated the field until then. In both refinement approaches one iteratively compares images from one or more 2D or 3D references in many different orientations with each experimental particle image, and references for the next iteration are calculated from averages over the experimental particles. In the least-squares approach one calculates a single, best orientation and class for each particle based on the squared difference between the reference and the particle, or the closely related cross-correlation coefficient. Each experimental particle then contributes only in its best orientation and class to the average that will form the reference for the next iteration. In the maximum-likelihood approach one employs a statistical noise model to calculate posterior probabilities for all possible orientations and classes, and each particle contributes to all references and in all orientations, but these are weighted according to the posterior probabilities. This treatment of the hidden parameters is called marginalization, and it results in a smearing out of each particle over multiple orientations and classes if the noise in the data is too strong to uniquely identify their correct assignment.

Because the statistical noise models are typically based on Gaussian distributions, the maximum-likelihood approach is closely related to the

least-squares approach. In fact, in the absence of any noise in the data it becomes straightforward to identify the best orientation and class, and the two approaches become identical (Sigworth, 1998). However, because of the high noise levels in cryo-EM images, the two approaches typically behave rather differently on experimental data. These differences are most important during the initial iterations of the optimization, when low-resolution references contain too little information to unambiguously assign orientations and classes. Upon convergence, when high-resolution references provide much more information, the posterior probability distributions often approach delta functions, corroborating the selection of only the single most likely assignment for each particle.

By 2010, maximum-likelihood approaches had been used in a number of experimental studies, mostly for 2D and 3D classification tasks (Scheres, 2010). Since then, the use of likelihood-based approaches in cryo-EM has increased steeply. Existing implementations of 2D and 3D maximum-likelihood classification in the XMIPP package (Scheres, Nunez-Ramirez, Sorzano, Carazo, & Marabini, 2008) have remained in use, while Niko Grigorieff also implemented a new classification approach in FREALIGN that marginalizes over the class assignments but still treats the orientational assignments in a least-squares manner (Lyumkis, Brilot, Theobald, & Grigorieff, 2013). Probably the steepest increase in the use of likelihood-based methods has been due to the introduction of a new, empirical Bayesian approach (Scheres, 2012a) that was implemented in the RELION program (Scheres, 2012b). This approach differs from previously available likelihood optimization approaches in the introduction of a regularization term to the likelihood function. The resulting regularized likelihood optimization algorithm has proven useful for both high-resolution reconstruction as well as 2D or 3D classification in a wide range of experimental studies (Bai, McMullan, & Scheres, 2015).

This chapter reviews the algorithmic concepts that were implemented in RELION and that are new compared to the approaches existing in 2010. It also describes how this software may be used in high-resolution cryo-EM structure determination. One of the major challenges in many projects is the presence of multiple, different 3D structures in the data. Many macromolecules adopt a range of different conformations as an intrinsic part of their functioning, and many samples are not purified to absolute homogeneity. If left untreated, the presence of structural heterogeneity in the data will lead to blurred regions in the maps and possibly to incorrect interpretations. Because many different approaches to deal with structural

heterogeneity exist, this probably represents the most challenging aspect of image processing in RELION. Therefore, this chapter is most detailed in its description of how to deal with structural heterogeneity. As such, it is also aimed as a useful resource for those who wish to use RELION in their cryo-EM structure determination efforts.

2. NEW ALGORITHMIC CONCEPTS

2.1 Regularization: The Empirical Bayesian Approach

The main algorithmic advance in RELION is the introduction of a regularization term to the likelihood target function. Previously available approaches optimized an unregularized likelihood function, which expresses the agreement between the experimental data and the reconstruction, while marginalizing over the hidden parameters. The problem with the unregularized target is that cryo-EM 3D reconstruction is severely ill-posed, ie, there are many possible noisy reconstructions that fit the data equally well. In order to define a unique solution one needs to introduce additional information, which is known as regularization. The key question then becomes, what information is available about the reconstruction in the absence of experimental data? A commonly employed approach to similar questions in machine learning is known as Tikhonov regularization, where one restrains the square of the Euclidean (or L2) norm of the solution (Tikhonov, 1943).

By describing both the signal and the noise components of the data using Gaussian distributions, Tikhonov regularization can also be understood from a Bayesian perspective. In this framework, the belief that the L2-norm of the reconstruction is limited is expressed as a prior on the solution. Optimization of the posterior distribution, which is proportional to the multiplication of the prior with the original likelihood function, is called maximum-a-posteriori (MAP) optimization and is mathematically equivalent to Tikhonov regularization. Whereas in standard Bayesian methods the prior is fixed before any data are observed, inside RELION parameters of the prior are estimated from the data themselves. This type of algorithm is referred to as an empirical Bayesian approach.

Both the likelihood and the prior are expressed in the Fourier domain, where all signal and noise components are assumed to be independent. In the likelihood function, the assumption of independent Gaussian noise in the Fourier domain is identical to a previously introduced, unregularized likelihood approach in the Fourier domain (Scheres, Nunez-Ramirez, et al., 2007).

Modeling the power of the noise as a function of spatial frequency in the Fourier domain allows for the description of nonwhite experimental noise and effectively results in a χ^2-weighting of the different spatial frequencies in alignment. In addition, the Fourier domain formulation permits a convenient description of the effects of microscope optics and defocusing (by the so-called contrast transfer function, or CTF). In the prior, the power of the Fourier components of the reconstruction is also restrained as a function of spatial frequency. Thereby, the prior basically acts as a Fourier filter that imposes smoothness on the reconstruction in real space.

In fact, optimization of the regularized likelihood target by the standard expectation-maximization algorithm (Dempster, Laird, & Rubin, 1977) results in an update formula for the reconstruction that shows strong similarities with previously introduced Wiener filters (Penczek, 2010a). Yet, while in 2010 it was still far from obvious how the Wiener filter for 3D reconstruction could be implemented to have its intended meaning (Penczek, 2010b), the empirical Bayesian perspective allowed an elegant and straightforward derivation of the correct implementation of the Wiener filter for 3D reconstruction. This filter depends on estimates for the power of both the signal and the noise as a function of spatial frequency. By constantly re-estimating these powers from the data themselves, RELION effectively calculates the best possible filter, in the sense that it yields the least noisy reconstruction, at every iteration of the optimization process.

Because optimal weights for the alignment and reconstruction are estimated without user intervention, the regularized likelihood optimization algorithm is intrinsically easy-to-use in the sense that is does not rely on an expert user to tune many critical parameters. This makes RELION suitable for automation and has likely contributed to its rapid uptake in the field.

2.2 Prevention of Overfitting: The Gold-Standard Approach to Refinement

Although the constantly updated Wiener filter explicitly dampens high-frequency components in the reconstruction, overfitting of the data is not entirely obviated. The reason for this lies in the empirical nature of the Bayesian approach. At every iteration, the widths of the Gaussians in the prior are estimated from the power of the reconstruction itself. Therefore, once one over-estimates the power of the true signal due to an inadvertent build-up of a small amount of noise in the reconstruction, one will allow even more noise in the next iteration. This can lead to overfitting, where noise in the model iteratively builds up.

By 2010, overfitting and over-estimation of resolution were still common problems in many cryo-EM structure determination projects, and this was recognized by a community-driven task-force for the validation of cryo-EM structures (Henderson et al., 2012). One of their recommendations was to prevent overfitting by a so-called gold-standard approach to refinement. In this approach one splits the data into two halves and one refines independent reconstructions against each half-set. The Fourier shell correlation (FSC) between the two independent reconstructions then yields a reliable resolution estimate, so that the iterative build-up of noise can be prevented. The original approach to estimate resolution from FSC curves was always intended to be used with two independently refined reconstructions (Harauz & van Heel, 1986). Nevertheless, common practices in the field had evolved toward the refinement of a single reconstruction, where the resulting angles would be used to make two (no longer independent) reconstructions from random half-sets at every iteration. The apparent rationale for this was that a reconstruction from all of the data would provide a better reference for alignment than the noisier reconstructions from only half of the data. However, evidence that this reasoning does not necessarily hold for cryo-EM refinements was obtained from tilt-pair experiments (Henderson et al., 2011). These experiments showed that alignments are dominated by the lower spatial frequencies, which are almost indistinguishable between reconstructions from all or half of the data. Further evidence came from a comparison of the two refinement approaches on simulated data and on experimental datasets for which a high-resolution crystal structure was available. Although the gold-standard approach gave nominally lower-resolution estimates, the resulting maps were actually better than when only a single reconstruction was refined (Scheres & Chen, 2012). Moreover, the previously introduced FSC = 0.143 threshold at which to interpret the resolution of the reconstruction (Rosenthal & Henderson, 2003) had been deemed as too optimistic when using the suboptimal refinement approach, and a more conservative FSC = 0.5 threshold had become the norm. The new experiments illustrated that the problem lies in the inflated FSC curves from the suboptimal refinement approach, and that the FSC = 0.143 threshold performs well when two independent reconstructions are used (Scheres & Chen, 2012).

These results prompted the implementation of a gold-standard refinement approach in RELION. In this implementation, the iterative build-up of noise is avoided by estimating the widths of the Gaussian priors from the FSC curve between two independently refined reconstructions.

Combined with a new approach to estimate the accuracy of alignment from the estimated signal-to-noise ratios in the data, this led to the implementation of "3D auto-refinement." In this approach, the estimated angular accuracies are used to automatically increase angular sampling rates during the refinement up to the point where the noise in the data prevents one from distinguishing smaller angular differences. This procedure allows the calculation of high-resolution reconstructions from structurally homogeneous datasets without intermediate user intervention (Scheres, 2012b).

2.3 Getting Clean, High-Resolution Maps: The Postprocessing Approach

In order to strictly prevent overfitting, the two independently refined reconstructions are not masked when calculating the FSC curves inside 3D auto-refinement. This leads to an underestimation of the true resolution of the reconstructed signal during refinement because the signal is restricted to a central region of the map and the surrounding solvent region merely contributes noise. Because orientational and class assignments are predominantly driven by the lower frequency content of the images, they are usually not noticeably affected by this underestimation of resolution. However, upon convergence the highest possible amount of information needs to be extracted from the reconstruction.

The noise in the solvent region may be removed by masking. Masking is a multiplication operation with a 3D map (the mask) that has values in the range of zero outside a region of interest to one inside the region. By masking out the solvent region from the two half-reconstructions, the noise gets reduced and the FSCs will increase. However, besides this beneficial effect of solvent removal, FSC values may also increase due to an undesirable convolution effect of masking. Multiplication with a mask in real-space corresponds to a convolution with the Fourier transform of the mask in the Fourier domain. A mixing of the stronger and better-correlating, low-frequency Fourier-components with the weaker and less-correlating, high-frequency components will cause an increase in FSCs that does not reflect the true signal-to-noise ratio in the maps. The more detailed features the mask has, the stronger this convolution effect is. Also this problem was recognized by the EM validation task-force (Henderson et al., 2012).

Adaptation of a method that was originally devised to estimate the amount of overfitting in refinement provided a solution to correct FSC curves for the convolution effects of masking (Chen et al., 2013). In this approach, the phases of Fourier components of the two half-reconstructions

with spatial frequencies higher than a given cutoff are randomized. The cutoff for phase-randomization is chosen to be considerably lower than the resolution of the reconstructions without masking away the solvent. Then, a solvent mask is applied to the two phase-randomized half-reconstructions. In the absence of convolution effects, the resulting FSC curves should be zero beyond the phase-randomization cutoff. Therefore, any nonzero correlations are due to convolution effects and one can correct the FSC curve between the two masked half-reconstructions without phase randomization. The corrected FSC curve reflects the increase in resolution caused by removing the solvent noise, without being affected by mask convolution effects.

The phase-randomization approach to correct FSC curves was implemented in the postprocessing approach of RELION, where it was combined with a previously proposed method to sharpen the reconstruction (Rosenthal & Henderson, 2003). This is needed because high-frequency components in the reconstruction are dampened in the image formation process itself, in the detection process and in the image processing. This falloff is often modeled by a Gaussian using a B-factor, in analogy to the temperature factor in X-ray crystallography. Application of a negative B-factor leads to sharpening of the map. The B-factor value may be estimated automatically for reconstructions that extend beyond 10 Å, or may be set by the user for lower-resolution reconstructions. Multiplication of the sharpened map with the corrected FSC curve typically leads to clean reconstructions with excellent high-resolution details.

2.4 Beam-Induced Motion Correction and Radiation-Damage Weighting: The Movie-Processing Approach

With the advent of direct-electron detectors in 2012/13, the possibility arose to collect movies during the exposure of the sample to the electron beam. This allowed for two improvements that were previously inaccessible: beam-induced motion correction and radiation-damage weighting. When the electrons hit the sample, inelastically scattered electrons deposit energy and chemical bonds in the sample are broken. The exact mechanisms of what happens are unknown, but one observation is that the sample starts to move upon exposure to the electron beam (see chapter "Specimen Behavior in the Electron Beam" by Glaeser). This beam-induced motion causes a blurring in the images that can be corrected by movie-processing, since each of the movie frames contains a sharper snapshot of the moving objects. Grigorieff and coworkers were the first

to illustrate the potential of beam-induced motion correction for large rotavirus particles (Brilot et al., 2012; Campbell et al., 2012).

Correcting for beam-induced motions of smaller complexes is more difficult, since the movie frames contain only a fraction of the total electron dose each and are thus extremely noisy (see chapter "Processing of Cryo-EM Movie Data" by Rubinstein). Still, implementation of a movie-refinement approach, where one marginalizes over the orientations of running averages of multiple movie frames led to much improved maps. Using this approach, reconstruction of only thirty thousand ribosome particles led to a map in which side chain densities were clearly visible (Bai, Fernandez, McMullan, & Scheres, 2013). For even smaller particles, ie, with molecular weights less than 1 MDa, the running averages of multiple movie frames still contain too much noise to reliably follow beam-induced motions for individual particles. Therefore, the movie-processing approach was adapted based on the observation that neighboring particles often move in similar directions (despite the fact that particles further away on the same micrograph may move in very different directions). By fitting straight lines through the most likely translations from the original movie-processing approach, and by considering groups of neighboring particles in these fits, the high noise levels in the estimated movement tracks could be sufficiently reduced to allow beam-induced motion correction also for smaller complexes (Scheres, 2014).

The approach to simultaneously fit beam-induced motions for groups of neighboring particles was combined with a novel way of handling radiation-damage weighting. Radiation damage starts with the breakage of chemical bonds in the sample, which will rapidly destroy the high-resolution content of the images. Subsequently, secondary structure elements and protein domains are unfolded, and eventually the entire macromolecular complex will be destroyed. Consequently, low-resolution information in the images will persist longer than the high-resolution information. To model the dose- and resolution-dependent effects of radiation damage on the images, the calculation of a B-factor was proposed for each movie frame. In the resulting approach, the movie-frames of each particle are aligned according to the fitted motion tracks, and single-frame reconstructions are used to estimate a B-factor for each movie frame. Then, the aligned movie frames of each particle are averaged with frequency-dependent weights according to their relative B-factors. By weighting the different spatial frequencies in each movie frame differently, the useful information from each movie frame is retained. For example, whereas later movie frames may hardly contain any high-resolution information, they may still contribute constructively

to the lower-resolution information content. Because the lower-resolution information may still be useful in particle alignment, this approach is more attractive than the alternative of throwing away later movie frames for high-resolution reconstruction. Because it results in particle images with improved signal-to-noise ratios this approach was called "particle polishing," a sideways reference to a quirk in the field whereby the crispest images are referred to as "shiny" (which was first used by Xuekui Yu at the 2007 Gordon research conference for 3D-EM).

3. A TYPICAL HIGH-RESOLUTION STRUCTURE DETERMINATION PROCEDURE

Fig. 1 shows a flowchart of a typical high-resolution structure determination project from movies acquired on a direct-electron detector. Micrograph-wide beam-induced motions are first corrected by aligning the frames of the recorded micrograph movies, and an average micrograph is calculated for every movie. This step is often performed using the MOTIONCORR program (Li et al., 2013). Secondly, one uses another third-party program to estimate the parameters of the CTF for each average micrograph. Implementation of a wrapper in the RELION GUI makes it convenient to use CTFFIND (Mindell & Grigorieff, 2003; Rohou & Grigorieff, 2015) or Gctf (Zhang, 2016) for this task.

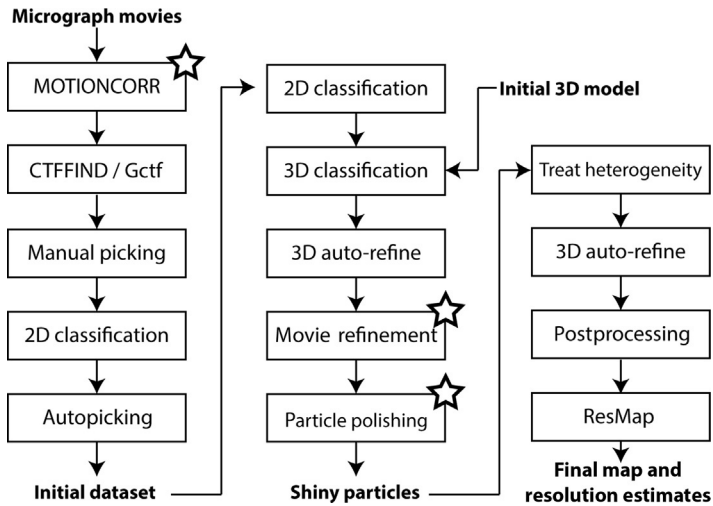

Fig. 1 Workflow of a typical structure determination project in RELION. Processes indicated with a star are skipped if movies are unavailable.

Next, one needs to identify the positions of individual particles in all micrographs. For this task the user first manually selects suitable particles from a few micrographs: typically around a thousand particles is enough. These particles are extracted into small square images, called boxes, where the size of the box is typically set to around two times the largest dimension of the complex of interest (see Fig. 2). Sometimes smaller boxes are used for highly elongated particles. Upon particle extraction, the individual particle images are normalized to have zero-mean and unity-variance in the background noise area, which is defined as the area outside a circular mask (see also Fig. 2). The diameter of this mask is set slightly larger than the largest dimension of the complex. A ramping background density may be modeled at this stage by fitting a 2D plane through the background pixels and subtracting this plane from the particle.

After extraction, the particles are subjected to a first round of reference-free 2D class averaging, which is called 2D classification in RELION. This task is performed using the same regularized likelihood optimization algorithm as described above, although in this case the references are 2D images and one marginalizes only over in-plane orientations. For cryo-EM data the number of classes for this run is typically set such that at least 50–100 experimental particles contribute on average to each class, so often one uses around 20 classes at this stage. From the resulting 2D class averages, one selects a few (usually not more than 5–10) representative views, which are used as templates for automated particle picking of all micrographs (Scheres, 2015). It is important to check that these templates are centered inside the particle box to prevent bias in the autopicked coordinates. In case the templates are not centered, they can be moved to their center-of-mass (assuming the signal is positive, ie, white) using the `relion_image_handler` program. Besides the templates, the autopicking algorithm has two additional parameters that are important: a threshold that expresses how restrictive the particle picking is (with higher threshold values being more restrictive, ie, picking fewer particles) and a minimum inter-particle distance. The minimum inter-particle distance is often set in the range of 50–70% of the longest dimension of the macromolecular complex of interest. Useful thresholds are often in the range of 0.1–0.5, where noisier micrographs generally require lower autopicking thresholds. For very noisy data, useful values as low as 0.05 have been observed, but in these cases one should be extra vigilant of template bias. Template bias leads to a dangerous situation where averages of experimental particle images reproduce the reference image used to pick the particles, even if the picked particles only

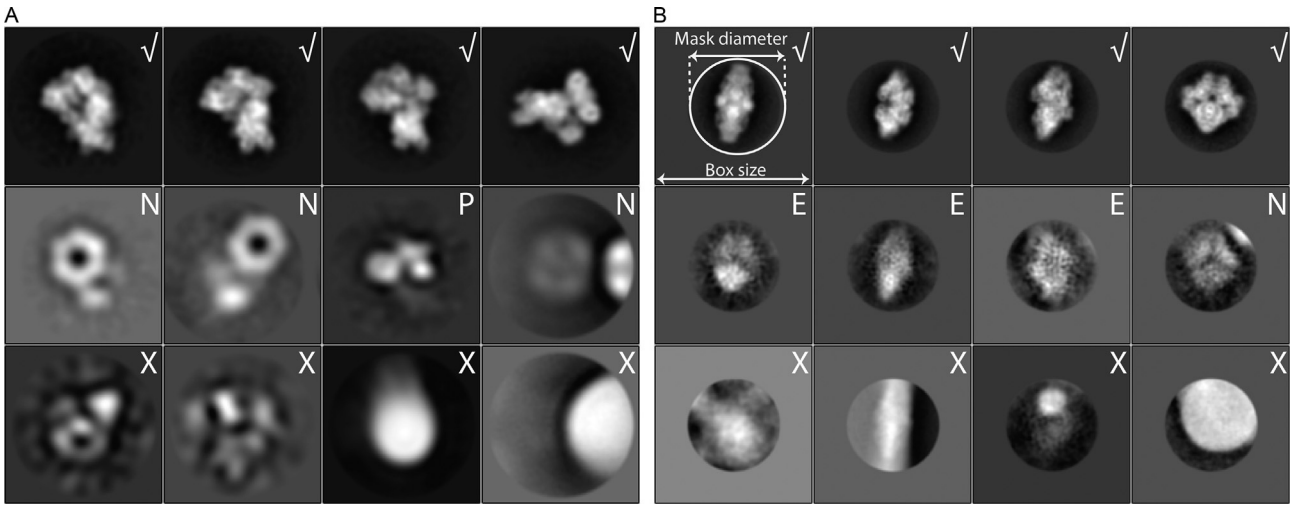

Fig. 2 Examples of class averages from a 2D classification of cryo-EM particles of an *E. coli* DNA replication complex (A) and *E. coli* β-galactosidase (B). Suitable class averages are indicated with a "✓." Class averages corresponding to large Einstein-from-noise classes are indicated with an "E," class averages with interfering neighboring particles are indicated with an "N," a class of a partially formed complex is indicated with a "P," and classes corresponding to otherwise unsuitable particles are indicated with an "X." The particle box size and the mask diameter used for particle extraction and normalization are indicated in the first class average in (B).

contain background noise. This is often referred to as "Einstein-from-noise," as a popular example to illustrate this effect uses a template picture from Einstein (Henderson, 2013; van Heel, 2013). To aid in the recognition of template bias and to prevent high-resolution bias, the template images are typically low-pass filtered to around 20–30 Å resolution. A more detailed description of how to recognize template bias is given below.

After extraction of the autopicked particles, one needs to identify those particles that are suitable for high-resolution structure determination. There are multiple reasons why particles may be unsuitable. The autopicking algorithm itself has been observed to be prone to selecting false positives in the form of high-contrast artifacts (Scheres, 2015), while noise-only images may result from low autopicking thresholds. In addition, particles may be partially unfolded during cryo-EM sample preparation, impurities in the sample may be mistaken for the particle of interest, or too closely neighboring particles may prevent their adequate alignment. A first, computationally cheap approach to identify false positives from the autopicking is called "particle sorting." This approach subtracts the corresponding templates from each experimental particle, and calculates a range of features on the difference images. Sorting all particles on the sums of their Z-scores over all features (Scheres, 2010) then may be more convenient to identify outliers in large datasets.

One of the most effective ways of selecting suitable particles is 2D classification. Typically, for large datasets one uses approximately 200 classes, while for smaller datasets, one restricts the number of classes such that at least 50–100 particles contribute to each class on average. Manual inspection of the resulting 2D class averages, often ordered on the number of particles contributing to each class, is then used to select suitable particles. Provided full CTF correction is performed, suitable particles will result in 2D class averages with strong, white density for the macromolecular complex and a black and featureless background. The macromolecular density should ideally not consist of white blobs for each protein domain, but contain clearly visible features of projected secondary structure elements throughout the complex. Overfitting of noise sometimes results in radially extending "hairy" artifacts in the solvent regions of the class averages. These undesired artifacts are indicative of bad alignments and typically only appear with very noisy data. Prevention of these artifacts may sometimes be achieved by limiting the resolution of the Fourier components that are used in the expectation step of the 2D classification. Often a resolution limit in the range of 15–20 Å is useful in initial runs, and this limit may be relaxed in later runs.

The selection of classes that are suitable for high-resolution reconstruction requires user supervision. High-contrast false positives from the autopicking and relatively small, low-resolution class averages are relatively easy to distinguish from the suitable classes. For the novice user, Einstein-from-noise classes are often harder to identify, yet their removal from the dataset is important. Sometimes, the class averages of Einstein-from-noise classes remain limited in resolution, and they can be readily identified by their blob-like aspect. In other cases, and especially for smaller macromolecular complexes, Einstein-from-noise classes may also give rise to relatively high-resolution averages. These class averages are often more grayish in their appearance than the suitable white-on-black classes, and high-resolution features often occur just as strongly in the surrounding background as in the apparent density of the complex. Fig. 2 shows examples of suitable and unsuitable 2D class averages of a 250 kDa DNA replication complex from *E. coli* (Fernandez-Leiro, Conrad, Scheres, & Lamers, 2015) and of the 450 kDa *E. coli* β-galactosidase complex (Scheres, 2015). Selection of the suitable classes may be performed conveniently through the GUI, and the selected particles from one 2D classification may be fed into an ensuing one. Sometimes, one also performs a subsequent 2D classification on particles that were selected from classes that did not look optimal in an attempt to rescue suitable particles that would otherwise be discarded. Most often, not more than three subsequent 2D classifications are performed. Strongly depending on the quality of the original micrographs and the threshold used in the autopicking, 20–80% of the autopicked particles may be discarded at this stage.

Next, one typically also uses an initial 3D multireference refinement run (called 3D classification) to continue the selection of suitable particles for high-resolution structure determination. The initial 3D reference that is required for 3D classification or 3D auto-refinement cannot be generated inside RELION itself. Popular programs to calculate such initial models from the particle images themselves are EMAN2 (Ludtke, Baldwin, & Chiu, 1999; Tang et al., 2007; see also chapter "Single Particle Refinement and Variability Analysis in EMAN2.1" by Ludtke) and SIMPLE (Elmlund, Davis, & Elmlund, 2010). Alternatively, one may use maps of similar complexes from the EMDB, maps generated from related PDB entries, or maps obtained by random conical tilt reconstruction (Radermacher, Wagenknecht, Verschoor, & Frank, 1987) or subtomogram averaging (Walz et al., 1997). In some cases, in particular when the complex of interest has a known point-group symmetry, refinements may even be started from a

spherical blob. The number of classes to be used in the initial 3D classification run depends on the size of the dataset and the available computer power. For smaller datasets one uses on average at least 5000 particle images per class, while for larger datasets one rarely uses more than 8-10 classes to reduce computing costs. The initial 3D classification is run with exhaustive angular searches, and typically without imposing symmetry. As in the case of 2D classification, the resolution of the Fourier components that is taken into account in the expectation step can again be limited to prevent overfitting. More details about the 3D classification are given in the next section. Again, suitable classes will show strong protein-like features, for example, rod-like densities for α-helices, while unsuitable classes will be of lower resolution or may not resemble the complex of interest. At this stage, reconstructions that correspond to the complex of interest in different conformational states are usually still kept together for subsequent movie-refinement.

One problem that may arise at this point is that there are not enough different views for 3D reconstruction because the particles adopted a strongly preferred orientation on the experimental support. This problem, which is often already detectable from a shortage of different 2D class averages, may manifest itself in streaky reconstructions, where densities are smeared out in the direction of the predominant view. A related problem may be that different classes become streaky in different directions, which is an indication that the classification converged to separate different views rather than different structural states. Such classifications are typically not useful. As long as some of the minority views are available in the dataset, throwing away particles from the predominant view may help to balance the orientational distribution and thereby get better reconstructions and classifications. This was, for example, crucial in the structure determination of the dynactin complex (Urnavicius et al., 2015). In many cases it is probably more efficient to tweak the cryo-EM sample preparation procedure, for example, by adding small amounts of detergents or by changing the type of experimental support, in order to get more different views.

The selected particles after the initial 3D classification are then used for movie-processing. For this, one first runs a 3D auto-refinement of the selected particles. Some degree of structural heterogeneity in the data at this point is acceptable. The objective of this refinement is to determine the orientations of all particles with respect to a single consensus reference. If at this point the consensus map does not extend well beyond 10 Å resolution, then movie-processing will probably not be effective and could be skipped. Otherwise, the consensus refinement is continued for a single iteration, where

running averages of the movie frames of each particle are aligned to reference projections of the consensus reconstruction. The user has to define the width of the running averages, which depends on the dose per movie frame and on the size of the complex of interest. For large complexes (eg, larger than 1 MDa) an accumulated dose of 3–6 electrons (e^-) per $Å^2$ in the running averages is usually sufficient. Smaller complexes may benefit from wider running averages, eg, with an accumulated dose in the range of 7–10 $e^-/Å^2$. In the subsequent particle polishing step, the translations from the movie refinement are used to fit straight tracks that describe the beam-induced motion for each particle. How many neighbors contribute to the fitting of each particle depends on a user-controlled standard deviation of a Gaussian weighting function. Larger standard deviations are needed for smaller particles, as their individual tracks will be noisier and more neighbors are needed to get reliable fits. Useful values for the standard deviation therefore also depend on how many particles there are on each micrograph. For particles with molecular weights less than 0.5 MDa useful values are often in the range of 500–1500 Å, while values below 500 Å are often used for larger particles. Reconstructions with individual movie frames are then used to determine the power of the signal at every spatial frequency for radiation-damage weighting. This often represents a critical step, and inspection of plots of both the B-factors and the linear scale factors is highly recommended. In the weighting process, the absolute values of the scale and B-factors of the movie frames are not important, only the differences between them matter. The B-factors are typically relatively high negative numbers in the first few frames (eg, with an accumulated dose of 1–3 $e^-/Å^2$) due to rapid initial beam-induced motions; they become smaller for intermediate frames (4–10); and then increase again for later frames due to radiation damage (eg, from an accumulated dose of 10–15 $e^-/Å^2$ onward). The linear scale factors typically decrease in a fairly linear manner throughout the movie, possibly with the exception of the first one or two frames, which may be somewhat lower. As a typical example the resulting plots and frequency-dependent weights for each movie frame are shown for a dataset on bovine complex-I in Fig. 3. If these plots do not look as expected, then the user may change the widths of the running averages, or the standard deviation of the Gaussian that expresses how many neighboring particles influence the fitted motion tracks. Also changing the resolution limits for the B-factor calculations or the tightness of the mask may help to get better plots. Visual inspection of the fitted beam-induced motion tracks and the Guinier plots that are used to estimate the B-factors may be

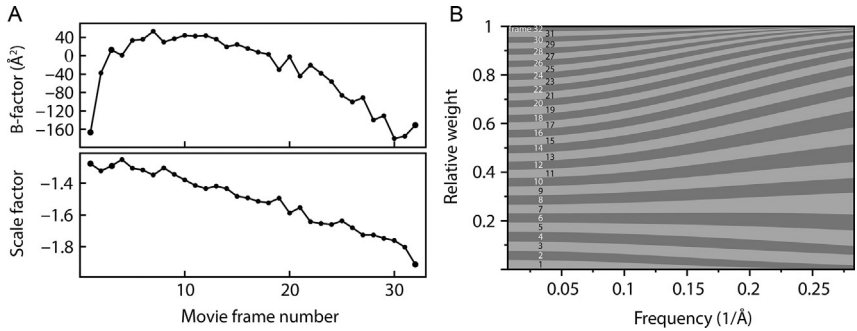

Fig. 3 Radiation-damage weighting. (A) B-factors and linear scale factors as estimated for all 32 movie frames of a cryo-EM dataset on bovine complex-I. (B) The resulting frequency-dependent weights for each movie frame. The width of each band indicates the relative weight at each frequency in each frame. At each frequency the sum over frames of the weights is unity. Note that the early and the later frames hardly contribute to the highest frequencies, whereas they do contribute to the lower frequencies. *This figure was adapted from Scheres, S. H. W. (2014). Beam-induced motion correction for sub-megadalton cryo-EM particles.* eLife 3*, e03665.*

helpful in identifying how these parameters may be changed. In addition, instead of performing reconstructions from individual movie frames to estimate the B-factors, one may perform reconstructions from running averages of multiple movie frames. The drawback of this approach is that large differences in the B-factors of sequential movie frames, as for example, observed in the first few frames in Fig. 3, cannot be modeled accurately.

After particle polishing, the resulting "shiny" particles may be used for further treatment of structural heterogeneity as outlined in more detail in the next section. Once a structurally homogeneous subset has been identified, a final 3D auto-refinement is performed, which is followed by post-processing to sharpen the map and to estimate its final resolution. The solvent mask for the postprocessing procedure can be calculated automatically from the reconstructed density, but the user should check that the resulting mask does not cut off part of the complex and that the edges of the mask are sufficiently wide. Too sharp edges on the mask will lead to strong convolution effects on the FSC curves, and unreliable (typically too low) resolution estimates. Therefore, it is important that the user inspects the FSC curves from the phase-randomization procedure. In particular, FSC values between the masked phase-randomized maps should be close to zero at the estimated resolution of the map. RELION uses cosine-shaped edges on masks, and useful widths of the mask edges are often in the range of 3–10 pixels. Masks around specific parts of the complex may be used to

estimate variations in local resolution. This may be useful to calculate better-resolved parts of the map to higher resolution, and more flexible parts to lower resolution. The overall reported resolution should however be calculated around the entire complex, and not merely reflect the resolution of the best ordered region. Finer variations in local resolution are often estimated using a wrapper from the RELION GUI to the RESMAP program (Kucukelbir, Sigworth, & Tagare, 2014). The same mask that encompasses the entire complex from the postprocessing approach should be provided to RESMAP in order to obtain reliable resolution estimates.

4. DEALING WITH STRUCTURAL HETEROGENEITY

4.1 3D Classification with Exhaustive Angular Searches

The main approach to distinguish projections from different 3D structures in RELION is 3D classification. Inside this approach, multiple reconstructions are refined simultaneously and one marginalizes over both the orientational and class assignments of the particle images. As demonstrated by the previously introduced unregularized likelihood approach to 3D classification in real-space (ML3D) (Scheres, Gao, et al., 2007), marginalization over both orientations and classes allows classification without the need for prior knowledge of the differences between the structures present in the data. This unsupervised classification is achieved by initializing multireference refinements from a single, low-resolution consensus model and assigning a random class to each particle in the first iteration. In this way, apart from the consensus model, the user only provides the number of desired classes. The actual number of different structures in the data is typically unknown, but in practice different runs with varying numbers of classes generate useful results. Because the user does not provide different references, against which to match the structural heterogeneity in the data, in machine-learning terminology this is called clustering instead of classification.

The simultaneous refinement of multiple references precludes the application of the gold-standard refinement approach, as the randomly seeded classification convergence may be very different for the two halves of the data. Therefore, inside the 3D (and also the 2D) classification approach in RELION, the resolution of each of the models is estimated directly from their power spectrum. This could in principle lead to overfitting, and the user has control over this by adjustment of the regularization parameter T. Lower values of T will prevent higher-resolution features to appear in the reconstructions and thus keep overfitting at bay. In practice, one does not change T very much. Most 2D classifications are performed with

$T=2$, while 3D classifications typically use values around $T=4$. One important exception to this is masked classification, which is described in more detail below.

If one starts without any knowledge of the relative orientations of the particles, then one needs to marginalize over all orientations in the likelihood function, or in other words one needs to perform exhaustive angular searches. To limit the associated computational costs, exhaustive angular searches in 3D classification are typically performed using a relatively coarse angular sampling. Because RELION uses the Healpix algorithm (Gorski et al., 2005) to sample the first two Euler angles on the sphere, only certain angular sampling rates are allowed. A sampling rate of 7.5 degrees is most often used for exhaustive searches. For highly symmetrical structures, eg, with icosahedral or octahedral symmetry, a 3.7 degree sampling may be better, although 3D classifications with exhaustive searches are typically performed without imposing symmetry, even if the complex of interest is known to have symmetry. In these asymmetric classifications one aims to remove junk or otherwise unsuitable particles from the dataset, and imposing symmetry on junk particles might make them look more like the structures of interest. Because each junk particle is probably different from all other junk particles, the concept of a "junk class" does not necessarily exist. Nevertheless, suitable particles tend to group together in one or more classes that yield reconstructions with better protein features than classes with junk particles. The latter tend to yield reconstructions with lower resolutions and without apparent symmetry.

If large conformational or compositional differences exist between the particles in the dataset, then 3D classification with exhaustive angular searches may also separate these out. Sometimes, this initial stage of separating out junk particles and very large conformational differences is repeated several times, where the selected subset from a previous run is used as input for the next.

4.2 Detection of Remaining Structural Heterogeneity

Once a subset of non-junk particles has been selected, all relative orientations of the selected particles may be determined with improved accuracy in a 3D auto-refinement run. If multiple different structures were encountered in the first 3D classification round, then separate 3D auto-refinements may be performed for each of the corresponding subsets. After 3D auto-refinement, postprocessing may then be used to sharpen the refined maps and to calculate their true resolution after solvent masking. Indications

whether remaining structural heterogeneity is still present may then be obtained from visual inspection of the resulting reconstructions (both before and after postprocessing) for blurry parts of density. Often, looking at the maps in 2D slices, for example, in RELION itself, is highly complementary to looking at 3D rendered maps in a program like UCSF Chimera (Goddard, Huang, & Ferrin, 2007; Pettersen et al., 2004). In addition, the calculation of a map with local resolution variations by the wrapper to the ResMap program (Kucukelbir et al., 2014) is useful, since large variations in local resolution are a strong indicator of remaining structural heterogeneity.

4.3 3D Classification with Finer, Local Angular Searches

Small conformational or compositional differences may not be separated well using the relatively coarse angular samplings that are typically used in 3D classification with exhaustive angular searches. One option to separate structures with smaller differences is to use the refined orientations from a previous (consensus) 3D auto-refinement as input for a 3D classification with finer angular samplings. In order to reduce the computational costs of this classification, one performs only local searches of the angles around the angles from the consensus refinement. This approach implicitly sets the prior probabilities of angles that deviate much from the input ones to zero, ie, one assumes that the angles from the auto-refinement are close to the true angles. This assumption probably holds reasonably well when conformational changes are relatively small, and thereby this approach is highly complementary to 3D classification with exhaustive searches, which is better suited for large conformational differences.

An extreme, and computationally even cheaper, alternative to 3D classification with local angular searches is to keep the orientations from the consensus refinement completely fixed. In this type of 3D classification, one only marginalizes over the class assignment and assumes that the orientations from the consensus refinement are the correct ones. Because in this calculation one only needs to compare each particle image with a single projection for each of the references, these calculations can be performed very rapidly. Most often, these 3D classifications without angular searches are performed within masks, as explained below.

4.4 Masked 3D Auto-Refinement

In many cases, the structural heterogeneity of interest is of a continuous nature, for example, when a particular subunit rotates relative to the rest

of the complex. In other cases, multiple subunits all move independently from each other. In such cases, masked 3D auto-refinement may be a suitable tool to deal with the multitude of different 3D structures in the data.

In this approach, one applies a 3D mask to the reference at every iteration. By masking out part of the complex, the experimental particle images are aligned only with respect to that part of the structure that lies within the mask. Thereby, the refinement becomes effectively insensitive to what happens in the regions outside the mask. Masks may be derived from fitted PDB models, or generated using the `relion_mask_create` program. Interactive editing of masks may be performed using the "Volume eraser" tool in UCSF Chimera (Pettersen et al., 2004). In order to prevent artifacts in the Fourier domain due to the convolution effects of masking, one again needs to use soft edges on the masks. Cosine-shaped soft edges may be added to masks using the `relion_mask_create` program. Often, useful widths of the soft edges are in the range of 3–10 pixels.

An insightful example of how masked 3D auto-refinement may be useful to deal with continuous motions was performed for the cytoplasmic ribosome of the *Plasmodium falciparum* parasite (Wong et al., 2014). As often observed for cryo-EM datasets of ribosomes, the ribosomes in this sample adopted multiple different ratcheted states, where the small subunit rotates relative to the large subunit. Because many intermediate rotated states exist, the structural heterogeneity can be interpreted as an almost continuous inter-subunit rotation. The resulting reconstruction from all particles showed relatively well-defined density for the large subunit (which dominates the alignment), while density for the small subunit was more blurred. Separation of this rotation in a discrete number of classes, like one would do in 3D classification, would be difficult. One would need to use a very large number of classes to describe the near-continuous motion, and one would end up with very few particles per class.

However, by applying a mask around the large subunit and aligning all particles to the mass within the mask, one can keep the entire dataset together. Conversely, one can do the same for the small subunit. For the *P. falciparum* dataset, this approach resulted in much improved reconstructions for both subunits. In addition, as one obtains two sets of angles for all particles, the differences between these sets carry information about the ratcheted state of each particle, although this information was not used in the *P. falciparum* study. One problem with these masked refinements is that one ends up with two separate (masked) reconstructions for the two subunits, in which information about the interface between the subunits

may not be well defined. Therefore, apart from the two separately masked reconstructions, the map from the unmasked refinement of the entire ribosome remained useful to inspect the subunit interface. The reconstruction from the unmasked refinement was also used to report on a single resolution estimate.

4.5 Masked 3D Classification

Masked 3D classification is the multireference equivalent of masked 3D auto-refinement. By masking out a region of interest from all references at every iteration, the classification can be focused on a specific region of interest, while structural variability in other regions is ignored. This is useful for many types of structural heterogeneity. For example, it may be used to separate complexes from which parts are missing in the case of non-stoichiometric complex formation, or one may separate conformational differences in one part of the structure while ignoring differences in other parts. Masked 3D classification is also useful to describe continuous motions in parts of the structure by dividing the data into subsets. Whereas 3D auto-refinement could in principle describe continuous motions without subdivision into a discrete number of subsets, auto-refinement cannot be applied to cases where the flexible region is very small. In such cases, finding the orientations of the particles with respect to that small region will be difficult. This is because projections of the masked references will only contain a small amount of mass, and the corresponding signal in the particle images will be drowned in the experimental noise and the projected mass of the rest of the complex. As a rule of thumb, in order to be able to apply masked 3D auto-refinement, one needs to have at least as much molecular mass inside the mask as one would need for the structure of an isolated complex, say at least 150–200 kDa. Many macromolecular complexes have flexible parts that are much smaller than this, and for these cases masked 3D classification provides a useful alternative.

In the masked 3D classification approach one typically provides the angles as determined in a consensus refinement as input and one does not perform any orientational searches. Apart from strongly reduced computational costs, this approach has the added benefit that particles cannot rotate or translate their region of interest outside of the mask, which would require even more classes. The result of these masked 3D classifications without angular searches is that the potentially continuous motion of the region of interest is divided into a discrete number of subsets. Each of these subsets

will naturally contain fewer particles than the original dataset, but the density of the subset reconstruction may be significantly improved, in particular when certain conformational states are more recurrent than others. If the size of the mask is a borderline case between masked auto-refinement or masked classification, then one may also employ local angular searches around the angles found in the consensus refinement. Because the 3D classification approach estimates the resolution from the power spectrum of each of the reconstructions, and because these reconstructions are now masked to select a potentially very small region, one typically uses higher values of the regularization parameter T in masked classifications. Whereas in unmasked 3D classifications T is typically kept around 4, values of 10–40 have been used in masked classification. Smaller regions require larger values of T. Interpretation of the maps by the user is required to find a useful value of T that balances the build-up of noise with the ability to visualize high-resolution features and separate small structural differences.

4.6 Masked 3D Refinement/Classification with Partial Signal Subtraction

The masked 3D auto-refinement and classification approaches suffer from a fundamental inconsistency. Inside these calculations, experimental projections of the entire complex are compared with projections of masked references, which contain only part of the complex. Therefore, the density in the experimental particle image that comes from the part of the complex that lies outside the mask will effectively act as noise in the comparisons. This noise will deteriorate the orientational and or class assignments. In the *P. falciparum* ribosome case described above, the part of the experimental particle images that corresponded to the large subunit caused larger inconsistencies in the masked refinement of the small subunit than vice versa. Consequently, the resolution in the masked refinement of the small subunit was considerably lower than for the large subunit.

The inconsistency in comparing experimental projections with projections from masked references can be reduced by subtraction of part of the signal from the experimental images (Fig. 4). In this approach, one first performs a consensus 3D auto-refinement of the entire dataset. Then, one designs two masks. The first mask encapsulates the region of interest on which one will perform a masked 3D classification (or a masked 3D auto-refinement). The second mask comprises the entire complex except for the region that lies inside the first mask. This second mask is applied to the reconstruction from the consensus refinement, and projections of this

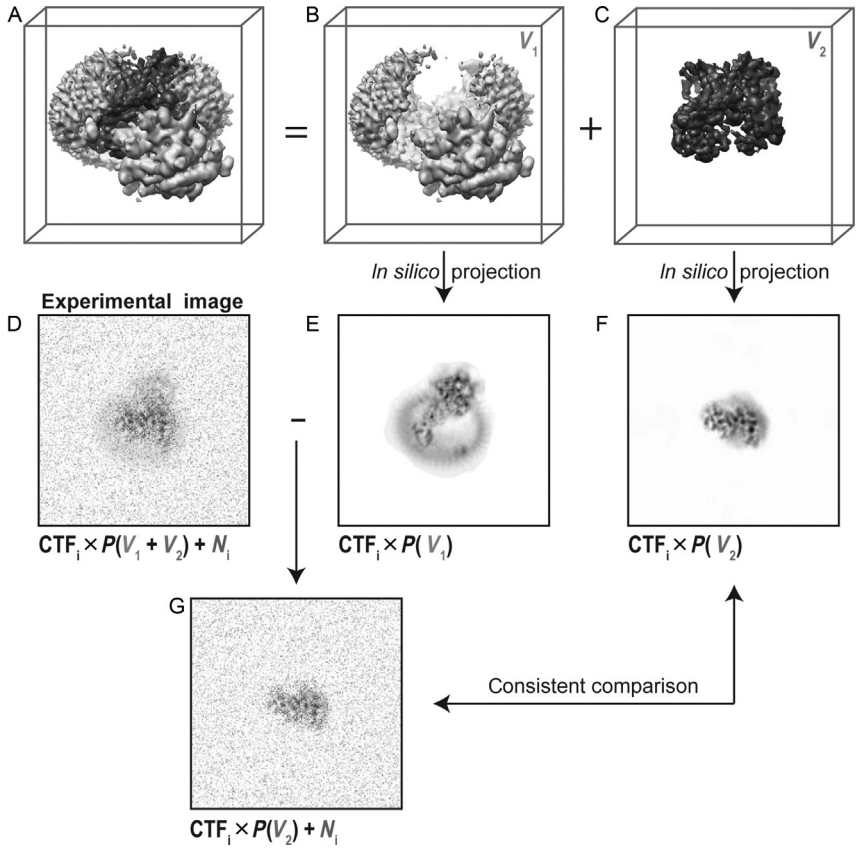

Fig. 4 Partial signal subtraction. (A) A 3D model of a complex of interest. (B) The part of the complex one would like to ignore (V_1). (C) The part of the complex one would like to focus classification on in a masked classification or 3D auto-refine (V_2). (D) An experimental particle image is assumed to be a noisy, CTF-affected 2D projection of the entire complex in (A). (E) A CTF-affected 2D projection of V_1. (F) A CTF-affected 2D projection of V_2. (G) Partial signal subtraction consists of subtracting the CTF-affected 2D projection of V_1 (E) from the experimental particle (D). This results in a modified experimental particle image (G) that only contains a noisy and CTF-affected projection of V_2. *This figure was adapted from Bai, X., Rajendra, E., Yang, G., Shi, Y., Scheres, S.H. (2015b). Sampling the conformational space of the catalytic subunit of human γ-secretase.* eLife 4. *doi:10.7554/eLife.11182.*

masked reconstruction are subtracted from all experimental particle images in the dataset. For this subtraction, one takes the effects of the CTF for each particle into account, and one uses the orientations as determined in the consensus refinement. This creates a new dataset of experimental images from which part of the signal was subtracted. The `relion_reconstruct` program

may be used to calculate a 3D reconstruction of the subtracted particles in order to verify the subtraction. The subtracted dataset can then be used for masked 3D classification or auto-refinement. Again, multiple scenarios are possible. Orientational searches can be performed exhaustively, restricted to local searches, or kept fixed at the orientations from the consensus refinement. If the remaining part of the complex is small, masked 3D classifications may be used to divide the data into a discrete number of subsets, while continuous motions in larger parts of the complex may be described using masked 3D auto-refinements.

Early applications of the partial signal subtraction approach were used by Michael Rossmann and colleagues to study symmetry mismatches in bacteriophage ϕ29 (Morais et al., 2003) and flaviviruses (Zhang, Kostyuchenko, & Rossmann, 2007). Because their reconstruction algorithm did not maintain the correct absolute grayscale of the experimental projections, it was necessary to optimize a scale factor between the reconstruction and the experimental particles. This is not necessary inside RELION, where the original grayscale is maintained in the reconstruction. Moreover, whereas the early work on viruses was done with CTF phase-flipped experimental images, inside RELION both the phases and amplitudes of the CTF of each particle are taken into account in the subtraction process.

Implementation of the subtraction approach inside RELION (Bai, Rajendra, Yang, Shi, & Scheres, 2015) recently allowed separation of three different conformations of the human gamma-secretase complex. In this case, a masked 3D classification was performed on the catalytic subunit, while projections of the other three subunits were subtracted from the experimental particle images. Because the region with the mask of the catalytic subunit only comprised approximately 30 kDa, orientations were kept fixed at those obtained in a consensus refinement of the entire (170 kDa) complex. After identification of three distinct conformations of the catalytic subunit, unmasked 3D auto-refinement of the original experimental particle images led to three reconstructions of the entire complex to near-atomic resolution. The observation that these structures differed only in the orientation of a few alpha-helices illustrates the exciting potential of masked classification with partial signal subtraction in separating cryo-EM images which differ only at the secondary structure level. The same approach was also used by Ilca and colleagues to subtract viral capsid densities in order to visualize an RNA polymerase bound inside double-stranded RNA bacteriophage φ6 (Ilca et al., 2015), while Zhou and colleagues used their own modified version of RELION to subtract projections of NSF rings

to analyze structural variability in the SNAP–SNARE complex (Zhou, Huang, et al., 2015).

4.7 Dealing with Pseudo-Symmetry

Many macromolecular complexes adopt some form of symmetry, but often this symmetry is only true for part of the complex. A generally applicable approach to deal with pseudo-symmetric complexes was recently used to solve the structure of a human apoptosome complex (Zhou, Li, et al., 2015). In this case, seven copies of the central protein form a ring-like hub with C7 symmetry. However, this symmetry is broken due to the dynamic nature of seven protruding spokes. Because each of the seven spokes seems to flex with respect to the central hub in an independent manner from the other six spokes, none of the complexes is truly symmetric. Consequently, a consensus refinement with imposed C7 symmetry led to a map with an overall resolution of 3.8 Å, where the central hub was well-resolved but the resolution for the spokes was considerably lower.

In order to deal with the pseudo-symmetry, the dataset was artificially expanded according to the pseudo-symmetric C7 point group. In this case, the dataset was enlarged sevenfold, by replicating each particle and adding $(n/7)*360°$ with $n=1,\ldots,7$ to the first Euler angle of every particle from the C7 consensus refinement. In this way, each of the seven spokes of every particle is oriented onto every position on the C7-symmetric ring. Masked classification within a single spoke and without angular searches then resulted in the identification of a subset of all spokes that were in a comparatively highly populated conformation relative to the symmetrical ring. Subsequent masked 3D auto-refinement in C1, where the mask included the entire central hub plus the single protruding spoke, yielded a reconstruction in which the spoke was determined to a resolution of 5 Å, which allowed docking of available crystal structures with confidence. Note that this C1 refinement was done with only local angular searches around the expanded set of orientations in order to prevent copies of the same particle from contributing to the reconstruction more than once in the same orientation. Because symmetry expansion is not as self-evident for noncyclic space groups, a stand-alone command line program was implemented to expand a dataset for any given symmetry point group. This program, which is called `relion_particle_symmetry_expand`, is already available upon request and will be part of RELION-2.0.

It should be noted that this approach shows similarities with an earlier approach to deal with pseudo-symmetric structures that was described by

Briggs et al. (2005). In their study of Kelp fly virus capsids, they identified two different types of fivefold vertices in the pseudo-icosahedral capsids. They used the orientations from an icosahedral consensus refinement to extract subimages of all vertices of each capsid, and used the known orientations from the icosahedral refinement to select similarly oriented vertices for multivariate statistical analysis and classification. Refinements of the identified classes were then performed exploiting the known orientations of the vertices with respect to the entire capsid.

Although not used for the apoptosome structure, the symmetry-expansion approach can also be combined with the partial signal subtraction approach. One could, for example, expand the set of orientations according to the broken point-group symmetry, and then subtract the consensus density for the symmetric part of the structure plus all-but-one of the asymmetric features from the experimental particle images. Then, one would perform a masked 3D classification or auto-refinement on the remaining asymmetrical feature, possibly followed by a masked 3D auto-refinement that includes both the symmetrical part of the complex and the single asymmetrical feature.

4.8 Multibody Refinement

At an early stage of the structure determination process of the spliceosomal U4/U6.U5 tri-snRNP particle from the Nagai lab (see also below) an automated approach to iterative masked refinements with partial signal subtraction from multiple regions was explored (Nguyen et al., 2015). The spliceosome has been an example of a notoriously difficult cryo-EM sample for many years due to its extremely dynamic nature (Lührmann & Stark, 2009). The 1.4 MDa Y-shaped tri-snRNP complex may be considered to consist of four more-or-less independently moving regions: a central "body"; a more flexible "foot" and "head"; and an extremely flexible "arm." In the automated approach, four regions were refined in parallel using masked 3D auto-refinement with partial signal subtraction. At every iteration, a mask around each of the four regions was applied in turn, and the signal of the other three regions was subtracted from every particle. Masked alignment of the four independent regions at every iteration led to four sets of optimal orientations for every particle, one orientation for each region. These optimal orientations were then used to perform a better signal subtraction in the next iteration. This still experimental implementation, which was called multibody refinement, holds the promise to automatically refine multiple independently moving regions, or bodies, almost without

user-intervention. However, the current implementation still suffers from strong artifacts at the boundaries between the regions, and thus requires further development to make it generally applicable.

4.9 A More Elaborate Example

A recent near-atomic resolution structure of the spliceosomal U4/U6.U5 tri-snRNP complex from the Nagai lab (Nguyen et al., 2016) represents an insightful example of how several of the approaches described above may be combined to describe molecular motions in a highly dynamic complex. Despite the extensive amount of flexibility in this complex, maps in which the majority of the complex could be resolved to near-atomic resolution could be obtained using the procedures that are outlined in Fig. 5.

After following the standard procedures as outlined in Fig. 1, a refinement of 140 thousand shiny particles led to a consensus map with an overall resolution of 3.7 Å. However, due to apparent flexibility in the complex, the reconstructed density in large portions of the map was not of sufficient quality to allow building of an atomic model. Separate masked 3D auto-refinements on the body, the foot and the head, with subtraction of the signal corresponding to the other regions, resulted in improved reconstructions of these three regions to local resolutions of 3.6, 3.7, and 4.2 Å, respectively. The improvement in density was largest for the head region, presumably because of its very dynamic attachment to the body. Whereas large parts of the head domain were only resolved to 8–10 Å resolution in the consensus map, the 4.2 Å reconstruction from the masked refinement with partial signal subtraction allowed the building of an atomic model for most of the protein and RNA in this region.

Because the arm region is even more flexible than the head, and because of its relatively small size, masked 3D auto-refinement of this region did not work well. Instead, masked 3D classification with subtraction of the signal from the other three regions was performed on this region. In an initial classification, a mask around only the arm region was used to divide the data according to three different relative orientations of the arm with respect to the body. Subsequent subtraction of only the head and the foot region, combined with masked 3D auto-refinement of the body and the arm regions together yielded three overall resolutions of 4.5–4.6 Å. Local resolution estimates indicated that the arm regions in these maps were resolved to resolutions in the range of 6.2–7.5 Å.

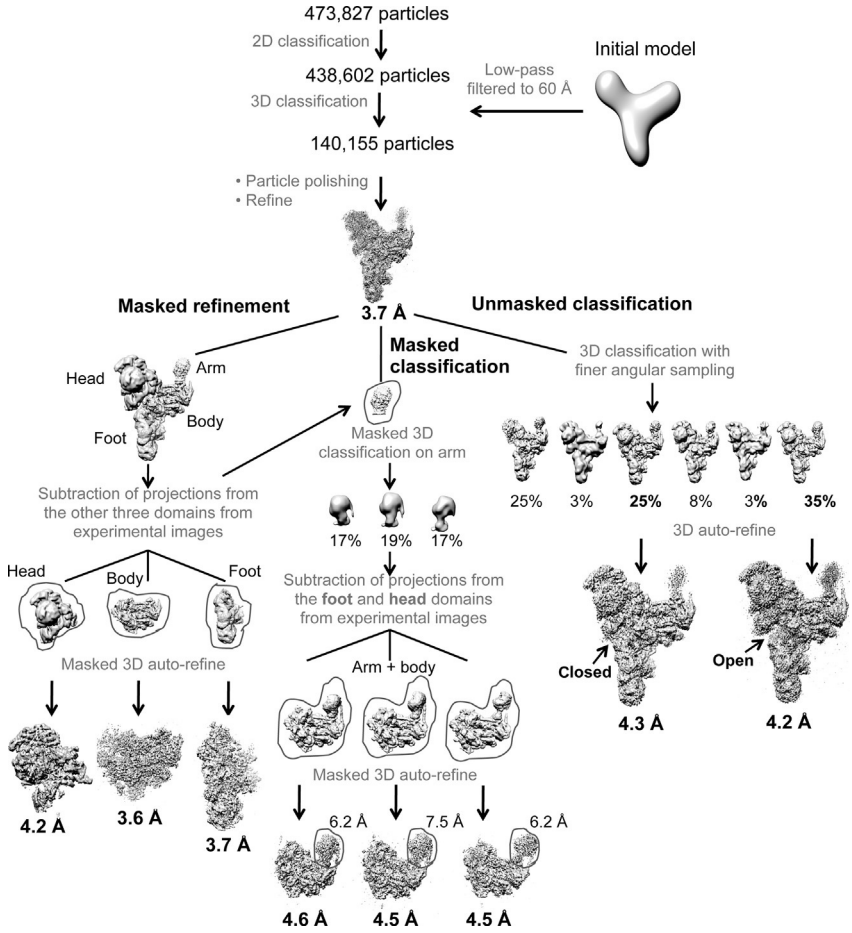

Fig. 5 An example of how to deal with structural heterogeneity is shown in the form of a flowchart for the processing of cryo-EM data on the spliceosomal U4/U6.U5 tri-snRNP complex. The various steps are described in more detail in the main text. *This figure was adapted from Nguyen, T.H.D., Galej, W.P., Bai, X.-C., Oubridge, C., Newman, A.J., Scheres, S.H.W., et al. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution.* Nature, 530, *298–302. doi:10.1038/nature16940.*

In addition to the masked 3D classifications and auto-refinements, an unmasked 3D classification with 1.8 degree local angular searches centered around the orientations from the consensus refinement of the entire complex was also useful to identify an open and a closed conformation of the body and the head regions.

5. OUTLOOK

RELION is still under active development. Its high computational costs (typically ~100–200 thousand CPU hours to process a single dataset) are currently an important bottleneck, especially for those labs that do not have access to large computer clusters. Ongoing efforts in vectorization of the code and its implementation for graphical processing units (GPUs) will probably lead to significant reductions in these costs in the near future. This will allow on-the-fly processing of micrographs during data acquisition, and procedures to automatically execute a predefined workflow of individual tasks are currently being developed. In addition, new algorithmic developments to handle helical assemblies; new approaches for electron tomography (Bharat, Russo, Löwe, Passmore, & Scheres, 2015); and further developments of multibody refinement and other classification tools for flexible complexes are all active areas of research. Hopefully, these developments will continue to push the boundaries of cryo-EM structure determination.

ACKNOWLEDGMENTS

I thank Drs. Xiao-chen Bai, Anthony Fitzpatrick, and Titia Sixma for critical comments on the chapter. This work was funded by the UK Medical Research Council (MC_UP_A025_1013).

REFERENCES

Bai, X.-C., Fernandez, I. S., McMullan, G., & Scheres, S. H. (2013). Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife*, *2*, e00461. http://dx.doi.org/10.7554/eLife.00461.

Bai, X., McMullan, G., & Scheres, S. H. W. (2015). How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences*, *40*, 49–57. http://dx.doi.org/10.1016/j.tibs.2014.10.005.

Bai, X., Rajendra, E., Yang, G., Shi, Y., & Scheres, S. H. (2015). Sampling the conformational space of the catalytic subunit of human γ-secretase. *eLife*. *4*. http://dx.doi.org/10.7554/eLife.11182.

Bharat, T. A. M., Russo, C. J., Löwe, J., Passmore, L. A., & Scheres, S. H. W. (2015). Advances in single-particle electron cryomicroscopy structure determination applied to sub-tomogram averaging. *Structure*, *23*, 1743–1753. http://dx.doi.org/10.1016/j.str.2015.06.026.

Briggs, J. A. G., Huiskonen, J. T., Fernando, K. V., Gilbert, R. J. C., Scotti, P., Butcher, S. J., et al. (2005). Classification and three-dimensional reconstruction of unevenly distributed or symmetry mismatched features of icosahedral particles. *Journal of Structural Biology*, *150*, 332–339. http://dx.doi.org/10.1016/j.jsb.2005.03.009.

Brilot, A. F., Chen, J. Z., Cheng, A., Pan, J., Harrison, S. C., Potter, C. S., et al. (2012). Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology*, *177*, 630–637. http://dx.doi.org/10.1016/j.jsb.2012.02.003.

Campbell, M. G., Cheng, A., Brilot, A. F., Moeller, A., Lyumkis, D., Veesler, D., et al. (2012). Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure*, *20*, 1823–1828. http://dx.doi.org/10.1016/j.str.2012.08.026.

Chen, S., McMullan, G., Faruqi, A. R., Murshudov, G. N., Short, J. M., Scheres, S. H. W., et al. (2013). High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy*, *135*, 24–35. http://dx.doi.org/10.1016/j.ultramic.2013.06.004.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum-likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Elmlund, D., Davis, R., & Elmlund, H. (2010). Ab initio structure determination from electron microscopic images of single molecules coexisting in different functional states. *Structure*, *18*, 777–786.

Fernandez-Leiro, R., Conrad, J., Scheres, S. H., & Lamers, M. H. (2015). cryo-EM structures of the E. coli replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and τ. *eLife*. *4*. http://dx.doi.org/10.7554/eLife.11134.

Goddard, T. D., Huang, C. C., & Ferrin, T. E. (2007). Visualizing density maps with UCSF Chimera. *Journal of Structural Biology*, *157*, 281–287. http://dx.doi.org/10.1016/j.jsb.2006.06.010.

Gorski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., et al. (2005). HEALPix—A framework for high resolution discretization, and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, *622*, 759–771. http://dx.doi.org/10.1086/427976. ArXivastro-Ph0409513.

Harauz, G., & van Heel, M. (1986). Exact filters for general geometry three dimensional reconstruction. *Optik*, *73*, 146–156.

Henderson, R. (2013). Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 18037–18041. http://dx.doi.org/10.1073/pnas.1314449110.

Henderson, R., Chen, S., Chen, J. Z., Grigorieff, N., Passmore, L. A., Ciccarelli, L., et al. (2011). Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. *Journal of Molecular Biology*, *413*, 1028–1046. http://dx.doi.org/10.1016/j.jmb.2011.09.008.

Henderson, R., Sali, A., Baker, M. L., Carragher, B., Devkota, B., Downing, K. H., et al. (2012). Outcome of the first electron microscopy validation task force meeting. *Structure*, *20*, 205–214. http://dx.doi.org/10.1016/j.str.2011.12.014.

Ilca, S. L., Kotecha, A., Sun, X., Poranen, M. M., Stuart, D. I., & Huiskonen, J. T. (2015). Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nature Communications*. *6*. http://dx.doi.org/10.1038/ncomms9843. article number 8843.

Kucukelbir, A., Sigworth, F. J., & Tagare, H. D. (2014). Quantifying the local resolution of cryo-EM density maps. *Nature Methods*, *11*, 63–65. http://dx.doi.org/10.1038/nmeth.2727.

Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., et al. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods*, *10*, 584–590. http://dx.doi.org/10.1038/nmeth.2472.

Ludtke, S. J., Baldwin, P. R., & Chiu, W. (1999). EMAN: Semiautomated software for high-resolution single-particle reconstructions. *Journal of Structural Biology*, *128*, 82–97. http://dx.doi.org/10.1006/jsbi.1999.4174.

Lührmann, R., & Stark, H. (2009). Structural mapping of spliceosomes by electron microscopy. *Current Opinion in Structural Biology*, *19*, 96–102. http://dx.doi.org/10.1016/j.sbi.2009.01.001.

Lyumkis, D., Brilot, A. F., Theobald, D. L., & Grigorieff, N. (2013). Likelihood-based classification of cryo-EM images using FREALIGN. *Journal of Structural Biology*, *183*, 377–388. http://dx.doi.org/10.1016/j.jsb.2013.07.005.

Mindell, J. A., & Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *Journal of Structural Biology*, *142*, 334–347. http://dx.doi.org/10.1016/S1047-8477(03)00069-8.

Morais, M. C., Kanamaru, S., Badasso, M. O., Koti, J. S., Owen, B. A. L., McMurray, C. T., et al. (2003). Bacteriophage φ29 scaffolding protein gp7 before and after prohead assembly. *Nature Structural & Molecular Biology*, *10*, 572–576. http://dx.doi.org/10.1038/nsb939.

Nguyen, T. H. D., Galej, W. P., Bai, X.-C., Oubridge, C., Newman, A. J., Scheres, S. H. W., et al. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature*, *530*, 298–302. http://dx.doi.org/10.1038/nature16940.

Nguyen, T. H. D., Galej, W. P., Bai, X., Savva, C. G., Newman, A. J., Scheres, S. H. W., et al. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature*, *523*, 47–52. http://dx.doi.org/10.1038/nature14548.

Penczek, P. A. (2010a). Image restoration in cryo-electron microscopy. In G. Jensen (Ed.), *Methods in enzymology. Cryo-EM, Part B: 3-D reconstruction* (pp. 35–72): Academic Press.

Penczek, P. A. (2010b). Fundamentals of three-dimensional reconstruction from projections. In G. Jensen (Ed.), *Methods in enzymology. Cryo-EM, Part B: 3-D reconstruction* (pp. 1–33): Academic Press.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*, 1605–1612. http://dx.doi.org/10.1002/jcc.20084.

Radermacher, M., Wagenknecht, T., Verschoor, A., & Frank, J. (1987). Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *Journal of Microscopy*, *146*, 113–136.

Rohou, A., & Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, *192*, 216–221. http://dx.doi.org/10.1016/j.jsb.2015.08.008.

Rosenthal, P. B., & Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology*, *333*, 721–745.

Scheres, S. H. W. (2010). Classification of structural heterogeneity by maximum-likelihood methods. In G. Jensen (Ed.), *Methods in enzymology. Cryo-EM, Part B: 3-D reconstruction* (pp. 295–320): Academic Press.

Scheres, S. H. W. (2012a). A Bayesian view on cryo-EM structure determination. *Journal of Molecular Biology*, *415*, 406–418. http://dx.doi.org/10.1016/j.jmb.2011.11.010.

Scheres, S. H. W. (2012b). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, *180*, 519–530. http://dx.doi.org/10.1016/j.jsb.2012.09.006.

Scheres, S. H. W. (2014). Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife*, *3*, e03665.

Scheres, S. H. W. (2015). Semi-automated selection of cryo-EM particles in RELION-1.3. *Journal of Structural Biology*, *189*, 114–122. http://dx.doi.org/10.1016/j.jsb.2014.11.010.

Scheres, S. H. W., & Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. *Nature Methods*, *9*, 853–854. http://dx.doi.org/10.1038/nmeth.2115.

Scheres, S. H. W., Gao, H., Valle, M., Herman, G. T., Eggermont, P. P. B., Frank, J., et al. (2007). Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature Methods*, *4*, 27–29. http://dx.doi.org/10.1038/nmeth992.

Scheres, S. H. W., Nunez-Ramirez, R., Gomez-Llorente, Y., San Martin, C., Eggermont, P. P. B., & Carazo, J. M. (2007). Modeling experimental image formation for likelihood-based classification of electron microscopy data. *Structure*, *15*, 1167–1177.

Scheres, S. H. W., Nunez-Ramirez, R., Sorzano, C. O. S., Carazo, J. M., & Marabini, R. (2008). Image processing for electron microscopy single-particle analysis using XMIPP. *Nature Protocols*, *3*, 977–990.

Sigworth, F. J. (1998). A maximum-likelihood approach to single-particle image refinement. *Journal of Structural Biology*, *122*, 328–339.

Sigworth, F. J., Doerschuk, P. C., Carazo, J.-M., & Scheres, S. H. W. (2010). An introduction to maximum-likelihood methods in Cryo-EM. In G. Jensen (Ed.), *Methods in enzymology. Cryo-EM, Part B: 3-D reconstruction* (pp. 263–294): Academic Press.

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., et al. (2007). EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, *157*, 38–46. http://dx.doi.org/10.1016/j.jsb.2006.05.009.

Tikhonov, A. N. (1943). On the stability of inverse problems. *Doklady Akademii Nauk SSSR*, *39*, 195–198.

Urnavicius, L., Zhang, K., Diamant, A. G., Motz, C., Schlager, M. A., Yu, M., et al. (2015). The structure of the dynactin complex and its interaction with dynein. *Science*, *347*, 1441–1446. http://dx.doi.org/10.1126/science.aaa4080.

van Heel, M. (2013). Finding trimeric HIV-1 envelope glycoproteins in random noise. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, E4175–E4177.

Walz, J., Typke, D., Nitsch, M., Koster, A. J., Hegerl, R., & Baumeister, W. (1997). Electron tomography of single ice-embedded macromolecules: Three-dimensional alignment and classification. *Journal of Structural Biology*, *120*, 387–395. http://dx.doi.org/10.1006/jsbi.1997.3934.

Wong, W., Bai, X.-C., Brown, A., Fernandez, I. S., Hanssen, E., Condron, M., et al. (2014). Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *eLife*, *3*, e03080. http://dx.doi.org/10.7554/eLife.03080.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *Journal of Structural Biology*, *193*, 1–12. http://dx.doi.org/10.1016/j.jsb.2015.11.003.

Zhang, Y., Kostyuchenko, V. A., & Rossmann, M. G. (2007). Structural analysis of viral nucleocapsids by subtraction of partial projections. *Journal of Structural Biology*, *157*, 356–364. http://dx.doi.org/10.1016/j.jsb.2006.09.002.

Zhou, Q., Huang, X., Sun, S., Li, X., Wang, H.-W., & Sui, S.-F. (2015). Cryo-EM structure of SNAP-SNARE assembly in 20S particle. *Cell Research*, *25*, 551–560. http://dx.doi.org/10.1038/cr.2015.47.

Zhou, M., Li, Y., Hu, Q., Bai, X.-C., Huang, W., Yan, C., et al. (2015). Atomic structure of the apoptosome: Mechanism of cytochrome c- and dATP-mediated activation of Apaf-1. *Genes & Development*, *29*, 2349–2361. http://dx.doi.org/10.1101/gad.272278.115.