

Accepted Manuscript

Cryo-electron microscopy analysis of structurally heterogeneous macromolecular complexes

Slavica Jonić

PII: S2001-0370(16)30055-1  
DOI: doi:[10.1016/j.csbj.2016.10.002](https://doi.org/10.1016/j.csbj.2016.10.002)  
Reference: CSBJ 146

To appear in: *Computational and Structural Biotechnology Journal*

Received date: 9 August 2016  
Revised date: 4 October 2016  
Accepted date: 11 October 2016



Please cite this article as: Jonić Slavica, Cryo-electron microscopy analysis of structurally heterogeneous macromolecular complexes, *Computational and Structural Biotechnology Journal* (2016), doi:[10.1016/j.csbj.2016.10.002](https://doi.org/10.1016/j.csbj.2016.10.002)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Cryo-electron microscopy analysis of structurally heterogeneous macromolecular complexes

Slavica Jonić

IMPMC, Sorbonne Universités - CNRS UMR 7590, UPMC Univ Paris 6, MNHN, IRD UMR 206,  
75005 Paris, France.

Corresponding author (Slavica Jonic, IMPMC-UMR 7590, Université Pierre & Marie Curie, Case  
courrier 115, 4 Place Jussieu, 75005 Paris, France, Phone : +33 1 44 27 72 05, Fax : +33 1 44 27 37  
85, E-mail : Slavica.Jonic@impmc.upmc.fr).

**Keywords:** Cryo-electron microscopy; single particle analysis; macromolecular complexes;  
heterogeneity; structure; flexibility; conformational changes; dynamics

**Abstract:** Cryo-electron microscopy (cryo-EM) has for a long time been a technique of choice for determining structure of large and flexible macromolecular complexes that were difficult to study by other experimental techniques such as X-ray crystallography or nuclear magnetic resonance. However, a fast development of instruments and software for cryo-EM in the last decade has allowed that a large range of complexes can be studied by cryo-EM, and that their structures can be obtained at near-atomic resolution, including the structures of small complexes (e.g., membrane proteins) whose size was earlier an obstacle to cryo-EM. Image analysis to identify multiple coexisting structures in the same specimen (multiconformation reconstruction) is now routinely done both to solve structures at near-atomic resolution and to study conformational dynamics. Methods for multiconformation reconstruction and latest examples of their applications are the focus of this review.

## INTRODUCTION

Recent instrumental and methodological developments for cryo-electron microscopy (cryo-EM) [1-19] made that the structures of macromolecular complexes are now often determined at subnanometer and near-atomic resolutions [20-41]. The most exciting results in terms of resolution and size of solved structures are currently being obtained with the latest-generation cryo-electron microscopes equipped with direct electron detectors (registering electrons directly rather than via a scintillator and recording movies allowing for correction of the specimen motion occurring during beam exposure) and software for automated collection of images, in combination with the use of advanced image analysis methods and high performance computing platforms [42-44].

First structures at near-atomic resolution were obtained for large complexes with high symmetry such as icosahedral-symmetry viruses [20,21]. However, several works have recently shown that cryo-EM can be used for near-atomic resolution of structures of small complexes (170-500 kDa) with low symmetry [22,27,41] or no symmetry [26,32], where the best resolution (1.8 Å) was obtained for 334 kDa glutamate dehydrogenase [40]. Bartesaghi and collaborators have pointed out that, rather than imaging technologies or image-processing methods, the major bottleneck to a routine cryo-EM determination of structures at resolutions close to 2 Å is currently the preparation of specimens of adequate quality that takes into account intrinsic protein flexibility [27]. Regarding larger complexes, subnanometer resolution is currently often achieved [24,25,28,30,36] and near-atomic resolution is becoming more and more frequent [23,29,31,33-35,37-39].

Three-dimensional (3D) reconstruction from heterogeneous sets of images normally results in low-resolution density maps. Thus, data heterogeneity analysis to isolate images of complexes of similar molecular compositions and conformations is a usual prerequisite to structural determination at high resolution. Biochemical procedures can usually be optimized so that the majority of complexes in the specimen, if not all of them, have the same molecular composition. However, the same composition rarely means the same conformation, due to the flexibility of complexes. Thus, conformational heterogeneity of specimens is usually analyzed by image analysis and classification methods. The reconstruction of different coexisting structures from the same sample will here be referred to as multiconformation reconstruction. It involves a classification strategy that assigns the particles having similar structures (similar molecular compositions and similar conformations) to the same class of particles. Multiconformation reconstruction is used to obtain high-resolution structures and provides insights into conformational dynamics of macromolecular complexes. Multiconformation reconstruction methods will be reviewed here together with the latest examples of their applications.

Image classification in multiconformation reconstruction can be supervised or unsupervised. Supervised classification aims at sorting images into classes based on information on expected conformational states (prior knowledge about the distribution of states), and its use is limited to studying systems where this prior information is available. The majority of methods reviewed here belong to the group of unsupervised classification methods whose aim is to find actual conformational states without a prior knowledge about the distribution of states. Due to computational limitations, the majority of available multiconformation reconstruction methods assume specimens with relatively few different conformations of complexes (usually, less than 10) and restrained compositional heterogeneity. They also take

into account that biochemical preparation of specimens is usually optimized to reduce the number of different structures coexisting in the same specimen. These methods are sometimes referred to as discrete conformational heterogeneity methods. They differ in the number of required initial 3D models (0, 1, or several) (**Fig. 1**), but a more important difference among these methods is whether they analyze heterogeneity at image level (in a 2D space) or at volume level (in a 3D space). Thus, these methods will here be grouped in two groups (2D and 3D heterogeneity analysis methods) and reviewed in two separate sections.

Development of methods for analyzing quasicontinuums of conformational states is an active field of research that will here be only briefly discussed (Outlook section). These methods will be fully reviewed in a separate publication.

## **2D HETEROGENEITY ANALYSIS METHODS**

In this section are reviewed methods that perform 3D reconstruction of different structures identified by analyzing structural heterogeneity at the level of images. Some of these methods use 3D starting models to determine the orientation of images while the other methods, referred to as *ab initio* methods, use no prior structural information.

### **Orientation determination without a starting 3D model**

The orientation of images can be determined based on the central section theorem [45]. This theorem states that the Fourier transform of a 2D projection is a plane intersecting the origin of the 3D object's Fourier transform and that this plane is parallel to the projection plane [45,46]. Any two non-parallel 2D projections of the same 3D object will therefore share a common line in Fourier space. Thus, the orientation of images can be determined by

determining the relative orientation of common lines between the 2D Fourier transforms of images [47,48]. The 3D model of the object obtained using images and the determined orientation is referred to as *ab initio* 3D model.

If the given set of images is heterogeneous, the images have to be sorted into structurally homogeneous subsets (image sorting) and 3D geometrical relationships among the images have to be determined (image orienting). When using no prior 3D model, image sorting and orienting can be performed in two separate steps or simultaneously. In the two-step approach proposed in [49], image orienting is preceded by a classification of images in classes of similar orientations (orientation classes) and a classification of each orientation class in classes of similar structures (image sorting), and both classifications are based on 2D multivariate statistical analysis (MSA) [50,51]. This approach, here referred to as nonsimultaneous sorting and orienting, has been particularly efficient in separating small and large particle images or images of ligand bound and unbound complexes [49,52,53]. In the approach for simultaneous sorting and orienting proposed in [54], all 6 parameters (3 Euler angles, 2 shifts, and structure assignment) are considered simultaneously for all images by solving a multidimensional optimization problem and common line correlations in Fourier space [54]. The larger the expected number of different structures, the more complex is the optimization problem to solve. So far, this approach was only used to separate two conformational states, such as open/closed and ligand bound/unbound states [54,55].

The main problem with the methods in this group is their low robustness to noise. They are thus usually used with 2D average images that have a higher signal-to-noise ratio (SNR) than individual images [53,55]. Also, their applications in studies with more than two conformational states have not yet been demonstrated.

### **Orientation determination using a starting 3D model**

Methods in this group aim at facilitating recognition of structural variability by minimizing orientational variability. They assume that dissimilarities between images corresponding to different structures are larger than those between images corresponding to the same structure but having slightly different angular directions.

The orientational variability is minimized by determining the orientation of images with respect to a preliminary 3D model that is usually obtained by combining images from the entire heterogeneous data set. Images assigned to the same projection direction are then sorted in clusters by analyzing discrepancies between common lines [56-58] or between entire images or their regions [59-62]. Clusters in each projection direction are labeled (different structures are assigned to different clusters) and those with the same label in different projection directions are combined in the same 3D reconstruction. Cluster labeling is a difficult task and the labeling approaches are usually not trivial. For instance, in [57], distinct cluster averages corresponding to a selected view (the view selected visually as the view showing the highest variability) and presumably representing different conformers are used as conformational references for the conformational assignment of cluster averages in all other orientations based on the highest cross-correlation of common lines between the cluster averages and the conformational references. On the contrary, the approach proposed in [58] considers all cluster averages simultaneously instead of selecting a single representative view and defining conformational references, by computing all pairwise similarities between the cluster averages based on cross-correlation of common lines.

The preliminary 3D model should have good quality and a potential model bias should be considered. In **Figure 1**, these methods are referred to as 2D variance analysis methods to be distinguished from the 3D variance analysis methods that also use an initial 3D model to orient images but analyze heterogeneity at volume level (classification based on 3D variance analysis that is described below).

The 2D heterogeneity analysis methods have also been used with globally homogenous data sets to select the most self-consistent subset of particles for a high-resolution 3D reconstruction. For instance, a procedure involving MSA-based classification of images, ab initio reconstruction, and iterative refinement has recently resulted in the first subnanometer-resolution structures of the complete portal-phage tail interface that mimic the states before and after DNA release during phage infection [63].

### **3D HETEROGENEITY ANALYSIS METHODS**

In this section are reviewed methods that perform 3D reconstruction of different structures identified by analyzing heterogeneity at the level of 3D reconstructions. These methods require one or several initial 3D models and can be classified in the following three groups: multireference classification, classification based on 3D variance analysis, and classification based on maximum likelihood.

#### **Multireference classification**

Methods that sort images by evaluating their similarity with a set of structures manually selected to contain expected features (3D references) [64,65] belong to the class of supervised classification methods and will here be referred to as supervised multireference classification



methods. In these methods, all given 3D references are first projected in directions uniformly distributed over a 3D sphere. Then, each particle image is compared with all obtained 2D projections (via projection matching that involves rotational and translational alignment of images with projections and computing their cross-correlation coefficients) to find the most similar projection and to assign the particle image to the corresponding 3D reference (the 3D reference from which the most similar 2D projection comes from). Each obtained class thus contains images that are the most similar to one of the given 3D references. Finally, a 3D reconstruction is computed from each obtained class of images using the rotational and translational parameters determined via the alignment with the best-matching reference. The required important prior knowledge (already available 3D template structures for all conformational states potentially present in the sample) makes supervised multireference classification methods useless when this prior information is not available, which is often the case.

Another group of multireference classification methods “select 3D references” automatically, via an iterative procedure in which 3D references compete for particles and are updated in each iteration [66-70]. Such methods belong to the class of unsupervised classification methods and they will here be referred to as unsupervised multireference classification methods. As supervised multireference classification methods, unsupervised multireference classification methods first assign each image to the best-matching reference from the set of given 3D references (by projection matching) and then compute 3D reconstruction from images assigned to the same 3D reference. However, contrary to supervised multireference classification methods, these methods use the 3D reconstructions obtained in the first iteration to update the 3D references for the next iteration. The iterations consisting of alignment, classification, and reconstruction steps are repeated until obtaining stable 3D references,

which allows obtaining new structural features (on the 3D references) with new iterations. An iterative procedure consisting of alignment-classification rounds was initially used in 2D work (using 2D class averaging instead of 3D reconstruction from image classes), where it was referred to as multi-reference alignment [71]. This approach is a version of K-means clustering algorithm, which estimates the unknown cluster centers based on the data and assigns the data to the nearest cluster (the nearest cluster center). In the EM context, the cluster centers are the reference structures (at a current iteration) and the usual measure of distance between images and the centers of clusters is the correlation between the images and the projections of the reference structures. Another version of this approach adds new 3D references progressively [72]. This approach, referred to as incremental K-means-like approach [72], starts by aligning the entire set of images using only one initial reference. After stabilizing the particle orientation parameters, it adds a new reference; if the two references are not too similar, the second reference will attract the particles that do not fit well the first reference and these particles will be used to update this second reference. After a refinement of the two-reference alignment (using a decreasing angular step size and a decreasing search range) and a stabilization of the orientation parameters, a new reference may be added and the process repeated until the number of references starts to exceed the number of intrinsic divisions within the dataset, which can be observed as a poor reconstruction (usually from a very small subpopulation of particles) after adding a new reference [72]. The particle images assigned to corrupt reconstructions are assumed to contain degraded particles and are removed [72]. The incremental K-means-like approach was recently used for isolating rotated from unrotated 80S ribosome images (using a low-resolution 80S ribosome from rabbit as reference) before their further splitting guided by 3D variability analysis (see the next subsection), and this combined procedure resulted in solving 11 states with conformational and compositional differences along the elongation circle of human 80S ribosome [73].

The main problem with multireference classification methods is a potential bias with the provided initial 3D references, which is usually observed as absence of new structural details on the reconstructed density maps. The initial references could be obtained by angular reconstitution, but they should be correct enough to avoid incorrect classifications. Also, a higher number of required references results in a lower performance of these methods.

These methods can also be used to separate particles to extract the most self-consistent data subset for a high-resolution 3D reconstruction. For instance, a near-atomic resolution of Porcine circovirus PCV2 virus-like particle was recently obtained using a strategy based on multireference alignment, classification, and refinement [39].

### **Classification based on 3D variance analysis**

Methods in this group are based on the principle of estimating a global 3D variability from the variability of data subsets. They require a prior knowledge of 3 Euler angles and 2 in-plane shifts for each particle image. However, they do not require that the number of different structures (classes) is known a priori. Instead, the number of classes can be determined a posteriori (after the 3D variability analysis) so that the classes show high inter-class and low intra-class variances.

One approach is to compute the 3D variance volume using a series of 3D reconstructions from bootstrap random sampled subsets of a given set of images. Bootstrap resampling was introduced to EM by Penczek and collaborators who proposed to generate new sets of images in which some images randomly selected from the given data set are duplicated and some are omitted so that the new data sets have the same size as the given data set [74]. The global 3D

variance is then computed using a series of 3D reconstructions obtained from the new data sets [74]. In [75], this 3D variance was used for localizing variable areas in a preliminary 3D model (obtained by combining the entire heterogeneous data set) and sorting particles in two classes corresponding to ligand binding and unbinding. More precisely, a 3D spherical mask placed in the region of highest 3D variance was projected in different directions to create 2D binary masks, and images with higher average densities in the regions outlined by respective 2D masks were labeled as ligand-bound states while those with lower average densities were labeled as ligand-unbound states [75]. Classification in a localized (masked) region is known as focused classification [75]. In another approach, the 3D variance was obtained by backprojection of 2D variance images calculated from images with the same projection directions and it was used in a focused classification approach to separate ligand-bound states from ligand-unbound states along the elongation circle of human 80S ribosome [73]. Also, classification approaches based on eigen analysis (multivariate statistical analysis or principal component analysis) of a series of 3D reconstructions from random subsets of images have been proposed [68,76,77] as well as approaches to estimate the actual covariance matrix or its eigenvectors directly from images [78-81].

### **Classification based on maximum likelihood**

Maximum-likelihood (ML) estimation was introduced to EM by Provencher and Vogel in the context of 3D reconstruction of viruses and ribosomes [82,83]. Later, an improved ML-based virus reconstruction approach was developed by Doerschuk and Johnson [84] and a 2D ML approach to general single-particle structure determination was proposed by Sigworth [85]. An extension of this latter approach to 3D, involving a membership of each experimental image to one of a fixed number of classes, was proposed by Scheres and colleagues, and it is referred to as ML3D [86]. ML3D considers that image orientation and class membership are

random variables and computes 3D models by expectation-maximization (E-M) recursive optimization. E-M algorithm is equivalent to a multireference projection-matching algorithm, where the discrete assignments of image orientation, translation, and class membership are replaced with probability-weighted integrations over all possible assignments and each image contributes to all orientations, translations, and classes with some weight. In practice, integrations are replaced with summations over a set of discrete projection directions, in-plane rotations, in-plane translations, and classes. 3D ML-based cryo-EM image classification was first demonstrated using ML3D, which resulted in solving structures of 70S ribosomes with bound and unbound elongation factor G that were consistent with those obtained by supervised multireference classification [86].

ML3D is available in XMIPP [10,87,88] and it is usually limited to intermediate resolution analysis because refinements with angular sampling steps finer than  $10^\circ$  are difficult to perform in practice [89]. The main reason for this is that ML3D keeps reference projections in computer memory to speed up computations, which makes that memory use becomes too large for angular sampling steps finer than  $10^\circ$  [86,89]. The two most recent and currently most popular ML-based methods overcome this problem by computing reference projections on-the-fly instead of precomputing and storing them in computer memory, thanks to the use of central section theorem for the projection operation. One of these two methods is available in RELION [6,90] while the other is available in FREALIGN [12,91]. Their performance is similar, though the implementation in RELION may give slightly better results in some data cases [12]. Both of them estimate parameters of prior distributions from data using variants of ML methodology and are thus referred to as extended likelihood or empirical Bayes [12].

To allow efficient optimization, ML-based methods simplify mathematical expressions by assuming independence of pixels and Gaussian white noise in either real space or reciprocal space. The approach assuming independence of real-space pixels cannot allow for correlations among real-space pixels that are due to the contrast transfer function of the electron microscope, while those assuming independence of reciprocal-space pixels cannot allow for real-space masking of noise around the particle to improve the SNR as real-space masking introduces correlations among reciprocal-space pixels [12,90]. ML3D and the FREALIGN approach assume independence of real-space pixels while the RELION approach assumes independence of reciprocal-space pixels. The zero-mean Gaussian prior on the reciprocal-space pixels in RELION makes a smoothing effect that reduces noise in the reconstruction. One can also note that the ML-based approach of RELION is an improved version of the Fourier-space version of ML3D that is referred to as MLF3D [92] and available in XMIPP (MLF3D performs many operations in the same way as ML3D and the computational costs of MLF3D and ML3D were observed to be comparable [92]). FREALIGN performs ML classification using E-M algorithm with fixed alignment parameters found by a non-ML parameter refinement. Rounds of non-ML parameter refinement are typically alternated with rounds of the E-M algorithm, which makes the method more efficient than implementations that are entirely based on ML [12].

ML-based methods can be used with heterogeneous data in order to determine multiple co-existing conformational states, but they can also be used with globally homogenous data, in which case the most self-consistent subset of particles could be isolated to obtain a single-state reconstruction at high resolution. A few examples of these two types of applications of RELION and FREALIGN ML-based methods will be cited here. It should however be noted that the initial models used by these approaches are usually obtained by other approaches

(e.g., initial model generation methods available in EMAN2 and SIMPLE software packages [3,18], random conical tilt technique [93], or selecting models from EMDB), which was also the case in the applications that are cited next.

In the context of identifying the most self-consistent subsets of particles for reconstructing single states at high resolution, the RELION approach resulted in near-atomic resolution of different structures, from small complexes, such as a 170 kDa asymmetric gamma-secretase [26] and a 300 kDa tetrameric TRPV1 ion channel [22], to large complexes, such as a 4.3 MDa asymmetric 80S ribosome [31] and a 5.5 MDa tetrameric ryanodine receptor RyR1 [35]. In the same context, the FREALIGN approach resulted in near-atomic resolution of structures of complexes such as a 250 kDa asymmetric L-protein of vesicular stomatitis virus [32], a 465 kDa tetrameric beta-galactosidase [27], and a 540 kDa hexameric AAA+ ATPase p97 [37].

In the context of classifying heterogeneous data sets to identify and structurally characterize different conformational states that coexist in the specimen, the RELION approach was combined with the use of masks and a modification of experimental images to improve consistency of their comparison with model projections. This approach separates experimental images based on their differences in a region of interest, which involves subtracting a residual signal (the signal outside the region of interest, but inside the complex) from the experimental images. The modification of experimental images improves consistency of their comparison with model projections in the region of interest, but it requires a preliminary determination of image orientation and translation parameters and a creation of two 3D masks, one for the region of interest and the other for the entire complex. This strategy produced results such as three distinct conformations of human gamma-secretase [94] and different local-region conformations of the yeast spliceosomal U4/U6.U5 tri-snRNP [95]. Regarding

characterization of coexisting conformational states with FREALIGN, several examples of results have been recently obtained, such as three distinct coexisting functional states of human p97 with occupancies of zero, one, or two molecules of ATP $\gamma$ S per protomer [37], seven distinct states of bovine mitochondrial ATP synthase with different modes of bending and twisting [96], and five 80S ribosome structures formed with the Taura syndrome virus internal ribosome entry site (IRES) and eEF2-GTP-sordarin suggesting a trajectory of IRES translocation required for translation initiation [97].

The main problem with ML-based classification methods is that the number of classes should be known from the beginning. To avoid using arbitrary numbers of classes, it is recommended to test different numbers in order to find the largest number that still produces classes with different features [89]. Also, there is a risk of bias to initial models, and to reduce it, the initial models are usually low-pass filtered to 40-80 Å.

## OUTLOOK

With specimens containing a small number of different conformations of the same complex, methods reviewed in this article can lead to near-atomic resolution of the different conformations. However, it would be difficult to use such methods with specimens containing a quasicontinuum of conformational states (a large number of discrete samples from a continuous trajectory of conformational change), as a meaningful definition of a few discrete classes would be impossible in this case. Several methods have been recently developed that could help analyzing continuous conformational changes [79,81,98-100]. However, only two of them explicitly consider continuous conformational changes while aiming at visualizing the entire conformational distribution, animating trajectories of conformational changes, and reconstructing 3D structures along these trajectories. One of them is based on normal mode



analysis [99,101] while the other is based on manifold embedding [98,102]. The problem of quasicontinuous conformational states is challenging and it is not yet fully solved. Development of methods for their study is the first step towards full studies of dynamics of macromolecular complexes. Instead of biochemically isolating a few conformational states, the concept of studying entire macromolecular dynamics is based on the principle of allowing as many as possible different conformational states in the specimen. Because of the expected high impact, methodological developments for such studies are currently in progress and will be reviewed in a separate publication.

## SUMMARY

In this article, I reviewed methods for analyzing structural heterogeneity of macromolecular complexes in cryo-EM images and examples of recent applications of these methods. These methods assume specimens containing a small number of different coexisting structures (typically, less than 10) and differ in the way of analyzing the heterogeneity (2D or 3D analysis) and in the number of required initial 3D models (0, 1, or several). When specimens contain complexes with a small number of different structures, 3D reconstructions at near-atomic resolution may result from the classes of images obtained with these methods, as shown in several works cited in this review. However, their use is limited in the case of continuous conformational changes because of potentially large numbers of different conformational states that could be collected in this case (the large number of states would be difficult to classify in a relatively small number of classes used by these methods).

Except for supervised multireference classification methods, all methods reviewed here belong to the class of unsupervised classification methods, which aim at finding actual conformational states without a prior knowledge about the distribution of states. Methods that

do not use available 3D models (ab initio methods) have advantage over other methods regarding the model bias, but they base the 3D reconstruction of different conformations on conformational variability analysis and sorting in 2D space. Methods that attack the heterogeneity problem directly in 3D space are more straightforward, but as they use available 3D models (at least for initial image orientation), the potential model bias should be carefully considered when using these methods. Currently, the most popular 3D methods are maximum-likelihood-based classification methods available in RELION [6,90] and FREALIGN [12,91] software packages.

This article is a quick start overview of cryo-EM image analysis methods that deal with discrete conformational heterogeneity of macromolecular complexes. In this context, the most relevant and recent articles were discussed. A collection of articles covering this aspect of 3D EM can also be found at [https://biocomp.cnb.csic.es/3DEM-Methods/index.php/Main\\_Page](https://biocomp.cnb.csic.es/3DEM-Methods/index.php/Main_Page).

## ACKNOWLEDGMENTS

The author thanks GENCI-CINES/IDRIS for HPC resources [x2014072174, x2015072174, x2016072174].

*Conflict of interest statement:* None declared.

## REFERENCES

1. Chen DH, Jakana J, Liu X, Schmid MF, Chiu W: **Achievable resolution from images of biological specimens acquired from a 4k x 4k CCD camera in a 300-kV electron cryomicroscope.** *J Struct Biol* 2008, **163**:45-52.

2. Henderson R, Chen S, Chen JZ, Grigorieff N, Passmore LA, Ciccarelli L, Rubinstein JL, Crowther RA, Stewart PL, Rosenthal PB: **Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy.** *J Mol Biol* 2011, **413**:1028-1046.
3. Elmlund D, Elmlund H: **SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles.** *J Struct Biol* 2012, **180**:420-427.
4. Leith A, Baxter W, Frank J: **Use of SPIDER and SPIRE in image reconstruction.** In *International Tables for Crystallography*. Edited by: John Wiley & Sons, Inc; 2012:620-623. [Arnold E, Himmel DM, Rossmann MG (Series Editor), vol F: Crystallography of Biological Macromolecules.]
5. Scheres SH, Chen S: **Prevention of overfitting in cryo-EM structure determination.** *Nat Methods* 2012, **9**:853-854.
6. Scheres SH: **RELION: implementation of a Bayesian approach to cryo-EM structure determination.** *J Struct Biol* 2012, **180**:519-530.
7. van Heel M, Portugal R, Rohou A, Linnemayr C, Bebeacqua C, Schmidt R, Grant T, Schatz M: **Four-dimensional cryo-electron microscopy at quasi-atomic resolution: IMAGIC 4D.** In *International Tables for Crystallography*. Edited by: John Wiley & Sons, Inc; 2012:624-628. [Arnold E, Himmel DM, Rossmann MG (Series Editor), vol F: Crystallography of Biological Macromolecules.]
8. Bai XC, Fernandez IS, McMullan G, Scheres SH: **Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles.** *Elife* 2013, **2**:e00461.
9. Chen S, McMullan G, Faruqi AR, Murshudov GN, Short JM, Scheres SH, Henderson R: **High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy.** *Ultramicroscopy* 2013, **135**:24-35.
10. de la Rosa-Trevin JM, Oton J, Marabini R, Zaldivar A, Vargas J, Carazo JM, Sorzano CO: **Xmipp 3.0: an improved software suite for image processing in electron microscopy.** *J Struct Biol* 2013, **184**:321-328.
11. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA, Cheng Y: **Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM.** *Nat Methods* 2013, **10**:584-590.
12. Lyumkis D, Brilot AF, Theobald DL, Grigorieff N: **Likelihood-based classification of cryo-EM images using FREALIGN.** *J Struct Biol* 2013, **183**:377-388.
13. Scheres SH: **Beam-induced motion correction for sub-megadalton cryo-EM particles.** *Elife* 2014, **3**:e03665.
14. Chiu PL, Li X, Li Z, Beckett B, Brilot AF, Grigorieff N, Agard DA, Cheng Y, Walz T: **Evaluation of super-resolution performance of the K2 electron-counting camera using 2D crystals of aquaporin-0.** *J Struct Biol* 2015, **192**:163-173.
15. Grant T, Grigorieff N: **Automatic estimation and correction of anisotropic magnification distortion in electron microscopes.** *J Struct Biol* 2015, **192**:204-208.

16. Grant T, Grigorieff N: **Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6.** *Elife* 2015, **4**:e06980.
17. Rohou A, Grigorieff N: **CTFFIND4: Fast and accurate defocus estimation from electron micrographs.** *J Struct Biol* 2015, **192**:216-221.
18. Bell JM, Chen M, Baldwin PR, Ludtke SJ: **High resolution single particle refinement in EMAN2.1.** *Methods* 2016, **100**:25-34.
19. de la Rosa-Trevin JM, Quintana A, Del Cano L, Zaldivar A, Foche I, Gutierrez J, Gomez-Blanco J, Burguet-Castell J, Cuenca-Alba J, Abrishami V, et al.: **Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy.** *J Struct Biol* 2016, **195**:93-99.
20. Yu X, Jin L, Zhou ZH: **3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy.** *Nature* 2008, **453**:415-419.
21. Zhang X, Settembre E, Xu C, Dormitzer PR, Bellamy R, Harrison SC, Grigorieff N: **Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction.** *Proc Natl Acad Sci U S A* 2008, **105**:1867-1872.
22. Liao M, Cao E, Julius D, Cheng Y: **Structure of the TRPV1 ion channel determined by electron cryo-microscopy.** *Nature* 2013, **504**:107-112.
23. Allegretti M, Mills DJ, McMullan G, Kuhlbrandt W, Vonck J: **Atomic model of the F420-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector.** *Elife* 2014, **3**:e01963.
24. Vinothkumar KR, Zhu J, Hirst J: **Architecture of mammalian respiratory complex I.** *Nature* 2014, **515**:80-84.
25. Allegretti M, Klusch N, Mills DJ, Vonck J, Kuhlbrandt W, Davies KM: **Horizontal membrane-intrinsic alpha-helices in the stator a-subunit of an F-type ATP synthase.** *Nature* 2015, **521**:237-240.
26. Bai XC, Yan C, Yang G, Lu P, Ma D, Sun L, Zhou R, Scheres SH, Shi Y: **An atomic structure of human gamma-secretase.** *Nature* 2015, **525**:212-217.
27. Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, Subramaniam S: **2.2 Å resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor.** *Science* 2015, **348**:1147-1151.
28. Fan G, Baker ML, Wang Z, Baker MR, Sinyagovskiy PA, Chiu W, Ludtke SJ, Serysheva, II: **Gating machinery of InsP3R channels revealed by electron cryomicroscopy.** *Nature* 2015, **527**:336-341.
29. Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV, Stark H: **Structure of the E. coli ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM.** *Nature* 2015, **520**:567-570.

30. Fischer M, Rhinow D, Zhu Z, Mills DJ, Zhao ZK, Vonck J, Grininger M: **Cryo-EM structure of fatty acid synthase (FAS) from *Rhodosporidium toruloides* provides insights into the evolutionary development of fungal FAS.** *Protein Sci* 2015, **24**:987-995.
31. Khatter H, Myasnikov AG, Natchiar SK, Klaholz BP: **Structure of the human 80S ribosome.** *Nature* 2015, **520**:640-645.
32. Liang B, Li Z, Jenni S, Rahmeh AA, Morin BM, Grant T, Grigorieff N, Harrison SC, Whelan SP: **Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy.** *Cell* 2015, **162**:314-327.
33. Pang Y, Bai XC, Yan C, Hao Q, Chen Z, Wang JW, Scheres SH, Shi Y: **Structure of the apoptosome: mechanistic insights into activation of an initiator caspase from *Drosophila*.** *Genes Dev* 2015, **29**:277-287.
34. von der Ecken J, Muller M, Lehman W, Manstein DJ, Penczek PA, Raunser S: **Structure of the F-actin-tropomyosin complex.** *Nature* 2015, **519**:114-117.
35. Yan Z, Bai XC, Yan C, Wu J, Li Z, Xie T, Peng W, Yin CC, Li X, Scheres SH, et al.: **Structure of the rabbit ryanodine receptor RyR1 at near-atomic resolution.** *Nature* 2015, **517**:50-55.
36. Aylett CH, Sauer E, Imseng S, Boehringer D, Hall MN, Ban N, Maier T: **Architecture of human mTOR complex 1.** *Science* 2016, **351**:48-52.
37. Banerjee S, Bartesaghi A, Merk A, Rao P, Bulfer SL, Yan Y, Green N, Mroczkowski B, Neitz RJ, Wipf P, et al.: **2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition.** *Science* 2016, **351**:871-875.
38. Hospenthal MK, Redzej A, Dodson K, Ukleja M, Frenz B, Rodrigues C, Hultgren SJ, DiMaio F, Egelman EH, Waksman G: **Structure of a Chaperone-Usher Pilus Reveals the Molecular Basis of Rod Uncoiling.** *Cell* 2016, **164**:269-278.
39. Liu Z, Guo F, Wang F, Li TC, Jiang W: **2.9 Å Resolution Cryo-EM 3D Reconstruction of Close-Packed Virus Particles.** *Structure* 2016, **24**:319-328.
40. Merk A, Bartesaghi A, Banerjee S, Falconieri V, Rao P, Davis MI, Pragani R, Boxer MB, Earl LA, Milne JL, et al.: **Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery.** *Cell* 2016, **165**:1698-1707.
41. Walls AC, Tortorici MA, Bosch BJ, Frenz B, Rottier PJ, DiMaio F, Rey FA, Veerler D: **Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer.** *Nature* 2016, **531**:114-117.
42. Kuhlbrandt W: **Biochemistry. The resolution revolution.** *Science* 2014, **343**:1443-1444.
43. Cheng Y, Grigorieff N, Penczek PA, Walz T: **A Primer to Single-Particle Cryo-Electron Microscopy.** *Cell* 2015, **161**:438-449.
44. Nogales E: **The development of cryo-EM into a mainstream structural biology technique.** *Nat Methods* 2016, **13**:24-27.
45. Bracewell RN: **Strip integration in radio astronomy.** *Austr J Phys* 1956, **9**:198-217.

46. Bracewell RN: *The Fourier Transform And Its Applications* edn 3rd: McGraw-Hill Science/Engineering/Math; 1999.
47. van Heel M: **Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction.** *Ultramicroscopy* 1987, **21**:111-123.
48. Penczek PA, Zhu J, Frank J: **A common-lines based method for determining orientations for  $N > 3$  particle projections simultaneously.** *Ultramicroscopy* 1996, **63**:205-218.
49. White HE, Saibil HR, Ignatiou A, Orlova EV: **Recognition and separation of single particles with size variation by statistical analysis of their images.** *J Mol Biol* 2004, **336**:453-460.
50. van Heel M, Frank J: **Use of multivariate statistics in analysing the images of biological macromolecules.** *Ultramicroscopy* 1981, **6**:187-194
51. van Heel M, Portugal R, Schatz M: **Multivariate Statistical Analysis of Large Datasets: Single Particle Electron Microscopy.** *Open Journal of Statistics* 2016, **6**:701-739.
52. Elad N, Clare DK, Saibil HR, Orlova EV: **Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections.** *J Struct Biol* 2008, **162**:108-120.
53. Orlova EV, Saibil HR: **Methods for three-dimensional reconstruction of heterogeneous assemblies.** *Methods Enzymol* 2010, **482**:321-341.
54. Elmlund D, Davis R, Elmlund H: **Ab initio structure determination from electron microscopic images of single molecules coexisting in different functional states.** *Structure* 2010, **18**:777-786.
55. Elmlund D, Elmlund H: **Cryogenic electron microscopy and single-particle analysis.** *Annu Rev Biochem* 2015, **84**:499-517.
56. Zhou ZH, McCarthy DB, O'Connor CM, Reed LJ, Stoops JK: **The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes.** *Proc Natl Acad Sci U S A* 2001, **98**:14802-14807.
57. Hall RJ, Siridechadilok B, Nogales E: **Cross-correlation of common lines: a novel approach for single-particle reconstruction of a structure containing a flexible domain.** *J Struct Biol* 2007, **159**:474-482.
58. Shatsky M, Hall RJ, Nogales E, Malik J, Brenner SE: **Automated multi-model reconstruction from single-particle electron microscopy data.** *J Struct Biol* 2010, **170**:98-108.
59. Klaholz BP, Myasnikov AG, Van Heel M: **Visualization of release factor 3 on the ribosome during termination of protein synthesis.** *Nature* 2004, **427**:862-865.
60. Briggs JA, Huiskonen JT, Fernando KV, Gilbert RJ, Scotti P, Butcher SJ, Fuller SD: **Classification and three-dimensional reconstruction of unevenly distributed or symmetry mismatched features of icosahedral particles.** *J Struct Biol* 2005, **150**:332-339.
61. Fu J, Gao H, Frank J: **Unsupervised classification of single particles by cluster tracking in multi-dimensional space.** *J Struct Biol* 2007, **157**:226-239.

62. Clare DK, Pechnikova EV, Skurat EV, Makarov VV, Sokolova OS, Solovyev AG, Orlova EV: **Novel Inter-Subunit Contacts in Barley Stripe Mosaic Virus Revealed by Cryo-Electron Microscopy.** *Structure* 2015, **23**:1815-1826.
63. Chaban Y, Lurz R, Brasiles S, Cornilleau C, Karreman M, Zinn-Justin S, Tavares P, Orlova EV: **Structural rearrangements in the phage head-to-tail interface during assembly and infection.** *Proc Natl Acad Sci U S A* 2015, **112**:7009-7014.
64. Valle M, Sengupta J, Swami NK, Grassucci RA, Burkhardt N, Nierhaus KH, Agrawal RK, Frank J: **Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process.** *Embo J* 2002, **21**:3557-3567.
65. Gao H, Valle M, Ehrenberg M, Frank J: **Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset.** *J Struct Biol* 2004, **147**:283-290.
66. Heymann JB, Cheng N, Newcomb WW, Trus BL, Brown JC, Steven AC: **Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy.** *Nat Struct Biol* 2003, **10**:334-341.
67. Brink J, Ludtke SJ, Kong Y, Wakil SJ, Ma J, Chiu W: **Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis.** *Structure* 2004, **12**:185-191.
68. Simonetti A, Marzi S, Myasnikov AG, Fabbretti A, Yusupov M, Gualerzi CO, Klaholz BP: **Structure of the 30S translation initiation complex.** *Nature* 2008, **455**:416-420.
69. Galkin VE, Orlova A, Vos MR, Schroder GF, Egelman EH: **Near-atomic resolution for one state of F-actin.** *Structure* 2014, **23**:173-182.
70. Lukoyanova N, Kondos SC, Farabella I, Law RH, Reboul CF, Caradoc-Davies TT, Spicer BA, Kleinfeld O, Traore DA, Ekkel SM, et al.: **Conformational changes during pore formation by the perforin-related protein pleurotolysin.** *PLoS Biol* 2015, **13**:e1002049.
71. van Heel M, Stofferl-Meilicke M: **Characteristic views of E. coli and B. stearothermophilus 30S ribosomal subunits in the electron microscope.** *Embo J* 1985, **4**:2389-2395.
72. Loerke J, Giesebrecht J, Spahn CM: **Multiparticle cryo-EM of ribosomes.** *Methods Enzymol* 2010, **483**:161-177.
73. Behrmann E, Loerke J, Budkevich TV, Yamamoto K, Schmidt A, Penczek PA, Vos MR, Burger J, Mielke T, Scheerer P, et al.: **Structural snapshots of actively translating human ribosomes.** *Cell* 2015, **161**:845-857.
74. Penczek PA, Yang C, Frank J, Spahn CM: **Estimation of variance in single-particle reconstruction using the bootstrap technique.** *J Struct Biol* 2006, **154**:168-183.
75. Penczek PA, Frank J, Spahn CM: **A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation.** *J Struct Biol* 2006, **154**:184-194.

76. Penczek PA, Kimmel M, Spahn CM: **Identifying conformational states of macromolecules by eigen-analysis of resampled cryo-EM images.** *Structure* 2011, **19**:1582-1590.
77. Klaholz BP: **Structure sorting of multiple macromolecular states in heterogeneous cryo-EM samples by 3D multivariate statistical analysis.** *Open Journal of Statistics* 2015, **5**:820-836.
78. Anden J, Katsevich E, Singer A: **Covariance estimation using conjugate gradient for 3D classification in CRYO-EM.** In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI) 16-19 April 2015*: 2015:200-204.
79. Katsevich E, Katsevich A, Singer A: **Covariance Matrix Estimation for the Cryo-EM Heterogeneity Problem.** *SIAM J Imaging Sci* 2015, **8**:126-185.
80. Liao HY, Hashem Y, Frank J: **Efficient estimation of three-dimensional covariance and its application in the analysis of heterogeneous samples in cryo-electron microscopy.** *Structure* 2015, **23**:1129-1137.
81. Tagare HD, Kucukelbir A, Sigworth FJ, Wang H, Rao M: **Directly reconstructing principal components of heterogeneous particles from cryo-EM images.** *J Struct Biol* 2015, **191**:245-262.
82. Provencher SW, Vogel RH: **Three-dimensional reconstruction from electron micrographs of disordered specimens. I. Method.** *Ultramicroscopy* 1988, **25**:209-221.
83. Vogel RH, Provencher SW: **Three-dimensional reconstruction from electron micrographs of disordered specimens. II. Implementation and results.** *Ultramicroscopy* 1988, **25**:223-239.
84. Doerschuk PC, Johnson JE: **Ab initio reconstruction and experimental design for cryo electron microscopy.** *IEEE Transactions on Information Theory* 2000, **46**:1714-1729.
85. Sigworth FJ: **A maximum-likelihood approach to single-particle image refinement.** *J Struct Biol* 1998, **122**:328-339.
86. Scheres SH, Gao H, Valle M, Herman GT, Eggermont PP, Frank J, Carazo JM: **Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization.** *Nat Methods* 2007, **4**:27-29.
87. Scheres SH, Nunez-Ramirez R, Sorzano CO, Carazo JM, Marabini R: **Image processing for electron microscopy single-particle analysis using XMIPP.** *Nat Protoc* 2008, **3**:977-990.
88. Sorzano CO, Marabini R, Velazquez-Muriel J, Bilbao-Castro JR, Scheres SH, Carazo JM, Pascual-Montano A: **XMIPP: a new generation of an open-source image processing package for electron microscopy.** *J Struct Biol* 2004, **148**:194-204.
89. Scheres SH: **Classification of structural heterogeneity by maximum-likelihood methods.** *Methods Enzymol* 2010, **482**:295-320.
90. Scheres SH: **A Bayesian view on cryo-EM structure determination.** *J Mol Biol* 2012, **415**:406-418.
91. Grigorieff N: **FREALIGN: High-resolution refinement of single particle structures.** *J Struct Biol* 2007, **157**:117-125.



92. Scheres SH, Nunez-Ramirez R, Gomez-Llorente Y, San Martin C, Eggermont PP, Carazo JM: **Modeling experimental image formation for likelihood-based classification of electron microscopy data.** *Structure* 2007, **15**:1167-1177.
93. Radermacher M, Wagenknecht T, Verschoor A, Frank J: **Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*.** *J Microsc* 1987, **146 ( Pt 2)**:113-136.
94. Bai XC, Rajendra E, Yang G, Shi Y, Scheres SH: **Sampling the conformational space of the catalytic subunit of human gamma-secretase.** *Elife* 2015, **4**.
95. Nguyen TH, Galej WP, Bai XC, Oubridge C, Newman AJ, Scheres SH, Nagai K: **Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution.** *Nature* 2016, **530**:298-302.
96. Zhou A, Rohou A, Schep DG, Bason JV, Montgomery MG, Walker JE, Grigorieff N, Rubinstein JL: **Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM.** *Elife* 2015, **4**:e10180.
97. Abeyrathne PD, Koh CS, Grant T, Grigorieff N, Korostelev AA: **Ensemble cryo-EM uncovers inchworm-like translocation of a viral IRES through the ribosome.** *Elife* 2016, **5**.
98. Dashti A, Schwander P, Langlois R, Fung R, Li W, Hosseinizadeh A, Liao HY, Pallesen J, Sharma G, Stupina VA, et al.: **Trajectories of the ribosome as a Brownian nanomachine.** *Proc Natl Acad Sci U S A* 2014, **111**:17492-17497.
99. Jin Q, Sorzano CO, de la Rosa-Trevin JM, Bilbao-Castro JR, Nunez-Ramirez R, Llorca O, Tama F, Jonic S: **Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes.** *Structure* 2014, **22**:496-506.
100. Sanchez Sorzano CO, Alvarez-Cabrera AL, Kazemi M, Carazo JM, Jonic S: **StructMap: Elastic Distance Analysis of Electron Microscopy Maps for Studying Conformational Changes.** *Biophys J* 2016, **110**:1753-1765.
101. Sorzano CO, de la Rosa-Trevin JM, Tama F, Jonic S: **Hybrid Electron Microscopy Normal Mode Analysis graphical interface and protocol.** *J Struct Biol* 2014, **188**:134-141.
102. Frank J, Ourmazd A: **Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM.** *Methods* 2016, **100**:61-67.

## FIGURE LEGENDS

**Figure 1:** Multiconformation reconstruction methods reviewed here.

No initial 3D model	Initial 3D model	Several initial 3D models
Nonsimultaneous sorting and orienting Simultaneous sorting and orienting	2D variance analysis 3D variance analysis	Multireference classification Maximum likelihood