Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images

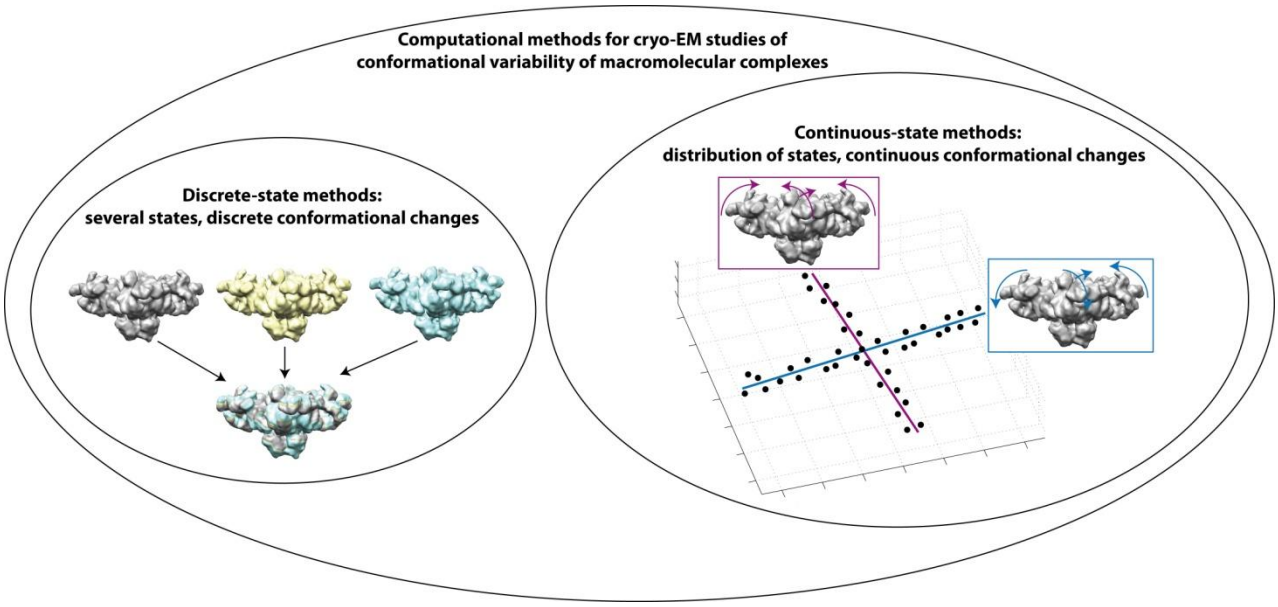
Slavica Jonic

This is the author's version of an article that has been published in *Current Opinion in Structural Biology*. Changes were made to this version by the publisher prior to publication. The final version is available at http://dx.doi.org/ 10.1016/j.sbi.2016.12.011 since 11 January 2017.

Highlights

- Recent cryo-EM technology allows near-atomic structural resolution of many complexes
- Conformational changes, functionally relevant, hinder higher resolution of complexes
- Most image processing procedures do not consider continuous conformational changes
- New methods are needed to observe details of the change cycle directly from images
- This review will help the future computational developments in cryo-EM

Graphical abstract



This is the author's version of an article that has been published in *Current Opinion in Structural Biology*. Changes were made to this version by the publisher prior to publication. The final version is available at http://dx.doi.org/ 10.1016/j.sbi.2016.12.011 since 11 January 2017.

Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images

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).

Short title: Cryo-EM conformational variability analysis methods

Keywords: Structure; dynamics; macromolecular complexes; cryo-electron microscopy; conformational changes; single-particle analysis

Abstract: Thanks to latest technical advances in cryo-electron microscopy (cryo-EM), structures of macromolecular complexes (viruses, ribosomes, etc.) are now often obtained at near-atomic resolution. Also, studies of conformational changes of complexes, in connection with their function, are gaining ground. Conformational variability analysis is usually done by classifying images in a number of discrete classes supposedly representing all conformational states present in the specimen. However, discrete classes cannot be meaningfully defined when the conformational change is continuous (the specimen contains a continuum of states instead of a few discrete states). For such cases, first image analysis methods that explicitly consider continuous conformational changes were recently developed. The latest developments in cryo-EM image analysis methods for conformational variability analysis are the focus of this review.

INTRODUCTION

Cryo-electron microscopy (cryo-EM) image collection and analysis technique referred to as single-particle analysis is based on the principle of reconstructing a 3D average structure from images (2D parallel-beam projections) of many copies of the same macromolecular complex (particle) at unknown random orientations and positions in a thin layer of vitreous ice untilted in the microscope [1]. The 3D reconstruction requires extracting particles into individual images (often referred to as single-particle images) and determining the particle orientation (three Euler angles) and position (in-plane coordinates of the particle centre) for each single-particle image. For a high resolution of the reconstruction, images should be sorted in the way to obtain a 3D average structure by combining only images of identical particle structures.

Thanks to recent instrumental and methodological developments for cryo-EM, near-atomic resolution of EM density maps is now more frequent [2-6] and studying conformational dynamics of macromolecular complexes is allowed [7-10]. The term "multiparticle reconstruction" is sometimes used to refer to the reconstruction of multiple different structures coexisting in a sample. The multiparticle reconstruction involves a strategy of

sorting the mixed population of images into groups (classes) such that each class ideally contains images of only one structure. The majority of available computational methods simplify the problem of structural heterogeneity by assuming that the number of different structures in the sample is small (typically, smaller than 10), and rely on reducing this number by optimized biochemical specimen preparation protocols [11-17]. However, as macromolecular complexes are not rigid but flexible entities, the obtained classes may not be perfectly homogeneous (particles assigned to the same class will rarely, if ever, have perfectly identical structures). This flexibility is usually referred to as conformational flexibility.

The class heterogeneity can also be partly explained by small errors made in determining the unknown parameters (orientation, position, and conformational state of each particle) due to a low signal-to-noise ratio (SNR) of cryo-EM images. Experimental techniques based on tilting the specimen have been proposed to reduce the number of unknown parameters [18-20], but a low quality of tilted data often makes the particle sorting as difficult as when using only untilted images. Thus, this review will be focused on computational methods for analyzing conformational variability of particles without tilting the specimen.

For a long time, specimen heterogeneity had a completely negative connotation and was only seen as a factor limiting the resolution of 3D reconstruction. However, it has recently been shown that the identification of conformational transitions from heterogeneous samples can be useful for studying molecular mechanism of action of complexes [7-10,21,22]. These findings have changed our regard about specimen heterogeneity. In particular, they opened doors to studying continuous conformational changes (a continuum of conformational states rather than a few discrete states). The development of methods that explicitly address continuous conformational changes is currently in progress [7,8].

Standard, class-based methods assuming a few discrete states will here be referred to as discrete-state methods while those exploring continuous conformational changes will be referred to as continuous-state methods. Among discrete-state methods, those performing the conformational variability analysis in 2D use either no prior structural information (ab initio methods) [11,12,23,24] or a 3D preliminary model to orient images [13,25-27]. Ab initio methods have a low robustness for low SNR data and have only been demonstrated in the context of separating images in two classes (e.g., small and large particles, ligand bound and unbound complexes, open and closed states). The main problem with methods using 3D preliminary models is that labeling of clusters in each projection direction, for their combination in multiple 3D reconstructions, may be relatively easy in the special, two-state case of presence and absence of an additional mass (e.g., bound and unbound ligand), but it is generally a difficult task. For these reasons, discrete-state methods performing conformational variability in 3D are currently more popular and will be reviewed here together with the latest developments for studying continuous conformational changes.

DISCRETE-STATE METHODS BASED ON CONFORMATIONAL VARIABILITY ANALYSIS IN 3D

Discrete-state methods that analyze conformational variability in 3D can be classified in the following three groups: multireference classification, maximum likelihood (ML) classification, and classification based on statistical analysis (**Fig. 1**).

Multireference classification

Supervised multireference classification approaches require good prior knowledge of all conformational states coexisting in the sample. They use 3D references with expected features (also referred to as templates) to assign, to each particle image, the conformational state of the most similar 2D projection of 3D templates [28,29]. The main problem with these approaches is a bias towards the 3D templates as they are usually chosen subjectively.

Approaches that iteratively refine alignment, reconstruction, and classification help reduce supervision and bias. They belong to the class of K-means clustering algorithms, with a first approach of this type (developed for 2D classification) dating back to 1985 [30]. In each iteration of 3D iterative multireference classification approaches, each image is assigned the conformational state, orientation, and translation parameters of the most similar 2D projection of 3D references and a 3D reconstruction is computed from each class of images (determined by the assigned conformational state) to update the 3D references for the next iteration [31,32]. Iterations are performed until obtaining stable 3D references. A good initial guess and a small number of necessary 3D references are required for a good performance of these methods. An approach adding 3D references incrementally has also been proposed [33], which makes the competition of 3D references for particles stronger and reduces the initial reference bias. This bias can alternatively be reduced by combining multireference classification with ML methods (discussed next), as shown in [34].

ML classification

ML methods estimate a set of density maps that best describe the given heterogeneous set of images while assuming independence among data pixels and Gaussian white noise model in real space or reciprocal space [14,15,35-37]. In a simplified form, they can be seen as multireference classification methods where the discrete assignments of image orientation, translation, and class membership (to one of a fixed number of classes) are replaced by probability-weighted integrations over all possible assignments. Each image contributes to all orientations and translations, and to all classes with some weight. These methods are less susceptible to initial model bias and overrefinement, especially with low SNR data. Several variants of the ML approach are available in EM software such as XMIPP [38], RELION [14], and FREALIGN [15]. Among them, implementations in RELION and FREALIGN are currently most popular [10,21,22,39,40]. While the FREALIGN approach assumes independence of real-space pixels (allowing to easily mask noise around the particle), the RELION approach assumes independence of reciprocal-space pixels (allowing to easily take into account the contrast transfer function of the microscope). Finally, a zero-mean Gaussian prior on reciprocal-space pixels, in RELION, makes a smoothing effect that mitigates accumulation of noise in the 3D reconstruction during the refinement process.

ML methods are local multireference refinement methods, meaning that their results depend on initial 3D references. Incorrect initial references may lead to incorrect results. The initial references can be obtained by performing an iteration of ML optimization for randomly drawn subsets of data [35,37], by calculating reconstructions from random subsets of data as in the bootstrap approach (discussed next) [41], or by using known templates as in supervised multireference classification [28]. A low-pass filtering of initial references to 40-80 Å is normally recommended to reduce bias. Also, as the convergence of ML methods is not guaranteed, several rounds of them are usually required, following a hierarchical scheme

using heuristic criteria. The requirement for an initial setting (guessing) of the number of classes and the assumption of discrete conformational changes are the main limitations of ML methods. Different numbers of classes should be tested to find the largest number that produces classes with new features [42]. When the underlying changes are continuous, discrete classes cannot be meaningfully defined, even if the number of classes is increased at the price of higher computational cost, due to random fluctuations in the sampled density of states [43].

Classification based on statistical analysis

Random resampling of a set of images allows estimating a global 3D variability from the variability of subsets. The global 3D voxel-by-voxel variance can be calculated based on a posteriori analysis of a large series of 3D reconstructions from sets of images obtained by random, bootstrap resampling of the given set of images [41]. The new sets, of the same size as the given set, are generated using a combination of duplication and removal of images randomly selected from the given set of images. Another approach to compute the 3D variance is by 3D reconstruction from 2D variance images calculated from particle images assigned to the same projection directions, where the projection-direction assignment is done by image alignment with a low-resolution preliminary model [9]. Penczek and collaborators have showed that 3D variance can be used to localize variable areas in the initial 3D reconstruction (obtained by combining all images) and sort particles in two classes corresponding to ligand binding and unbinding, which is referred to as focused classification [16]. Data classification based on eigenvector analysis (principal component analysis (PCA) or multivariate statistical analysis (MSA)) of a series of 3D reconstructions from randomly selected subsets of images has also been proposed [17,44].

Resampling-based methods do not require a prior knowledge of conformers, but rely on a preliminary model (usually obtained by combining all images) to assign orientation and translation parameters to each image. While multireference and ML classification methods require that the number of classes is set before the data analysis (ideally, the number of classes should be known a priori), resampling-based methods allow setting this number at a later step (the number of classes can be determined after the 3D variance analysis, by making a balance between intraclass and interclass variances, which is less arbitrary than the setting of this number in the case of multireference and ML classification methods).

Interestingly enough, methods based on eigenvector analysis allow mapping of the entire set of images onto a common low-dimensional space determined by the most important eigenvectors [17,45]. Such map of images can reveal whether the underlying conformational change is discrete or continuous, if the image mapping space is determined using eigenvectors of the actual covariance matrix. However, an efficient estimation of the covariance matrix is currently limited to images of small sizes [45] or image regions [46]. Alternatively, eigenvectors of this matrix can be estimated without the matrix estimation [17,47]. While the method proposed in [17] estimates the eigenvectors using a series of 3D reconstructions from random subsets of images, the method proposed in [47] estimates the eigenvectors directly from images using a likelihood-based approach to PCA. The covariance matrix can be assembled from the estimated eigenvectors, but its accuracy is not guaranteed due to the estimation errors.

This is the author's version of an article that has been published in *Current Opinion in Structural Biology*. Changes were made to this version by the publisher prior to publication. The final version is available at http://dx.doi.org/ 10.1016/j.sbi.2016.12.011 since 11 January 2017.

CONTINUOUS-STATE METHODS

While discrete-state methods assume a few distinct conformational states coexisting in the sample, continuous-state methods assume a continuum of conformational states i.e. continuous conformational changes and are under development. Currently, only two approaches consider continuous conformational changes explicitly. They are usually referred to as HEMNMA (Hybrid Electron Microscopy Normal Mode Analysis) [7,48] and manifold embedding [8,43]. These two approaches are presented in **Figure 2** and discussed here. They represent images in a low-dimensional space specifically suited to the continuous nature of conformational changes and allow a 3D visualization of conformational changes along trajectories in this space.

Manifold embedding

The manifold embedding approach [8] analyzes conformational variability in 2D space determined by one projection direction (using images with virtually the same orientations) and, then, combines the conformational variability information obtained in different projection directions to visualize it in 3D space. In each projection direction, it maps images onto a low-dimensional space using the diffusion map technique, which is a nonlinear dimensionality reduction technique based on computing eigenvectors and eigenvalues of a diffusion operator on the data. The manifold embedding approach requires that the image orientation is determined first. It assumes that effects of orientational changes dominate those of conformational changes, meaning that heterogeneity is not excessive so that each image can be associated with the correct projection direction using standard methods that do not take into account conformational heterogeneity (e.g., projection matching with a density map reconstructed from the entire, heterogeneous set of images).

HEMNMA

HEMNMA [7] analyzes and visualizes conformational variability directly in 3D space. It first iteratively computes the parameters of orientation, translation, and conformation for each image. More precisely, a reference (pseudo)atomic structure (an input atomic structure or a pseudoatomic structure computed from an input EM density map [49]) is elastically deformed using normal modes, the deformed structure is converted into a density map, and 2D projections of that map are compared with each image until the best matching 2D projection (the best matching density map) is found for this image, which determines its optimal Euler angles, in-plane translations, and conformational parameters (M displacement amplitudes along M normal modes). HEMNMA minimizes the risk of initial model bias by using large populations of different test density maps for different images (the density maps generated by deforming the reference (pseudo)atomic structure). The conformational parameters determined for the ensemble of images are then analyzed by PCA or other linear dimensionality reduction method to map images onto a low-dimensional space. Trajectories of conformational changes may be identified by exploring the most densely populated regions in the low-dimensional space and the reference (pseudo)atomic structure can then be displaced along the trajectories to visualize the changes in 3D. Also, density maps can be reconstructed along the trajectories from images with similar conformations. A user-friendly HEMNMA graphical interface is available in Xmipp and Scipion [48]. The graphical interface additionally allows low-dimensional image mapping using nonlinear dimensionality reduction methods, but the trajectory animation is unavailable when using nonlinear methods (the conversion into the original displacement space cannot be done in such cases) [48].

This is the author's version of an article that has been published in *Current Opinion in Structural Biology*. Changes were made to this version by the publisher prior to publication. The final version is available at http://dx.doi.org/ 10.1016/j.sbi.2016.12.011 since 11 January 2017.

Combining continuous-state methods with discrete-state methods

The 3D density maps reconstructed along a trajectory obtained with a continuous-state method can be regarded as discrete samples of a continuous trajectory. Extrapolation of continuous trajectories from their discrete, unordered samples obtained by discrete-state methods has also been addressed. In this context, the method referred to as StructMap projects a set of 3D density maps obtained by discrete-state methods onto a low-dimensional space of their conformational distances [50]. StructMap models each density map by elastic deformations (using normal modes) of other density maps and performs a statistical analysis of their differences to project the maps onto a common low-dimensional space. Distances among density maps in the low-dimensional space can be analyzed to select maps that would later be used to finely extrapolate continuous trajectories by analyzing images with HEMNMA or with manifold embedding [50]. The selected maps would be used to compute the reference pseudoatomic structures for HEMNMA or to orient images for manifold embedding via projection matching [50]. An illustration of combining discrete-state and continuous-state methods is shown in **Figure 3.**

DISCUSSION AND CONCLUSION

More often than to study conformational transitions, discrete-state classification methods are used to obtain high-resolution reconstruction of single conformations. Within the high-resolution reconstruction strategies, classification methods are used to select a subset of images containing particles with most consistent views and conformations (i.e. only those assigned to the classes of highest resolution), and the 3D reconstruction is then computed using only that image subset. In this context, many particle images are thrown away without fully understanding reasons for their inconsistence with other particles. For instance, a 3D reconstruction at near-atomic resolution is sometimes obtained by collecting ~1 million particle images, but retaining only one half after a 2D classification and less than one fourth for the final 3D reconstruction [51].

However, the classes of lower resolution could come from a mixture of particles with different (possibly unique) conformations that are difficult to represent with the given, usually small number of classes. Discrete-state methods may provide a limited view of data heterogeneity due to the use of a small number of classes. Also, different discrete-state methods may produce different density maps using the same heterogeneous set of images, as in the case of a publicly available set of images of different states of EF-G binding to the 70S *E. coli* ribosome (a table showing the results obtained with this data set and different discrete-state methods is provided in [24]). Continuous-state methods can provide a broader view of data heterogeneity than discrete-state methods, which was shown in [7] using the mentioned set of images of mixed states of EF-G binding to the 70S *E. coli* ribosome and other test data sets. Also, they can help identify trajectories of conformational changes. Furthermore, the 3D reconstruction at high resolution is possible from subsets of images along the trajectories, but requires enough large and homogeneous image subsets. Continuous-state methods are under active development and are expected to have high impact on studies of dynamics of macromolecular complexes.

ACKNOWLEDGMENTS

This is the author's version of an article that has been published in *Current Opinion in Structural Biology*. Changes were made to this version by the publisher prior to publication. The final version is available at http://dx.doi.org/ 10.1016/j.sbi.2016.12.011 since 11 January 2017.

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

REFERENCES

1. Frank J: *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. New York: Oxford University Press; 2006.

2. 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.

3. Yu X, Jin L, Zhou ZH: **3.88 A structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy**. *Nature* 2008, **453**:415-419.

4. Liao M, Cao E, Julius D, Cheng Y: **Structure of the TRPV1 ion channel determined by electron cryo-microscopy**. *Nature* 2013, **504**:107-112.

5. 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.

6. Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, Subramaniam S: **2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor**. *Science* 2015, **348**:1147-1151.

•• 7. 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. HEMNMA approach to analyzing continuous conformational changes was introduced in this article. The article also shows performances of the method using several synthetic and experimental data sets. HEMNMA analyzes EM images using normal modes of a starting 3D model obtained by EM or X-ray crystallography. Conformational parameters obtained by image analysis are then used to map images onto a low-dimensional space where conformational changes can be explored (using animations of the starting model or 3D reconstructions along trajectories identified in most densely populated regions in this space).

•• 8. 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. Manifold embedding approach to analyzing continuous conformational changes was introduced in this article. The article also shows performances of the method using yeast 80S ribosome data. The method assumes that heterogeneity is not excessive so that each image can be associated with a correct projection direction using standard image-orienting methods that do not take into account conformational heterogeneity. It maps images with similar projection directions onto a low-dimensional space using diffusion maps (nonlinear dimensionality reduction) and combines the trajectory information from different projection directions into a 3D movie.

9. 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.

10. Banerjee S, Bartesaghi A, Merk A, Rao P, Bulfer SL, Yan Y, Green N, Mroczkowski B, Neitz RJ, Wipf P, et al.: **2.3 A resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition**. *Science* 2016, **351**:871-875.

11. 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.

12. 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.

13. 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.

• 14. Scheres SH: **RELION: implementation of a Bayesian approach to cryo-EM structure determination**. *J Struct Biol* 2012, **180**:519-530. This article introduces RELION software and shows how empirical Bayes method can be used for a high-resolution density map refinement.

• 15. Lyumkis D, Brilot AF, Theobald DL, Grigorieff N: **Likelihood-based classification of cryo-EM images using FREALIGN**. *J Struct Biol* 2013, **183**:377-388. This article describes the likelihood-based method combining 3D multiconformation alignment, classification, and refinement that is available in FREALIGN. The authors point out that both variants of maximum-likelihood approach, the one of FREALIGN and the one of RELION, are extended likelihood or empirical Bayes methods as both of them estimate hyperparameters of hierarchical prior distributions from the data using variants of maximum-likelihood methodology. However, the article shows that slightly better results may be obtained with RELION in some data cases.

16. 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.

• 17. Penczek PA, Kimmel M, Spahn CM: **Identifying conformational states of macromolecules by eigen-analysis of resampled cryo-EM images**. *Structure* 2011, **19**:1582-1590. This article describes a statistical analysis approach to 3D reconstruction of multiple conformational states from a heterogeneous set of images that is based on estimating eigenvectors of the covariance matrix without estimating this matrix. The eigenvectors are computed by principal component analysis of a series of 3D reconstructions from randomly selected subsets of images, and the images are classified after their projection onto a low-dimensional space determined by the most important eigenvectors, which allows a posterior determination of the number of classes.

18. 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.

19. Leschziner AE, Nogales E: **The orthogonal tilt reconstruction method: an approach to generating single-class volumes with no missing cone for ab initio reconstruction of asymmetric particles**. *J Struct Biol* 2006, **153**:284-299.

20. Penczek PA, Grassucci RA, Frank J: **The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles**. *Ultramicroscopy* 1994, **53**:251-270.

21. 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**.

22. 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.

23. Orlova EV, Saibil HR: **Methods for three-dimensional reconstruction of heterogeneous assemblies**. *Methods Enzymol* 2010, **482**:321-341.

24. Elmlund D, Elmlund H: **Cryogenic electron microscopy and single-particle analysis**. *Annu Rev Biochem* 2015, **84**:499-517.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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.

30. van Heel M, Stoffler-Meilicke M: **Characteristic views of E. coli and B. stearothermophilus 30S ribosomal subunits in the electron microscope**. *Embo J* 1985, **4**:2389-2395.

31. 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.

32. Lukoyanova N, Kondos SC, Farabella I, Law RH, Reboul CF, Caradoc-Davies TT, Spicer BA, Kleifeld O, Traore DA, Ekkel SM, et al.: **Conformational changes during pore formation by the perforin-related protein pleurotolysin**. *PLoS Biol* 2015, **13**:e1002049.

33. Loerke J, Giesebrecht J, Spahn CM: **Multiparticle cryo-EM of ribosomes**. *Methods Enzymol* 2010, **483**:161-177.

34. 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 A resolution by Cs-corrected cryo-EM**. *Nature* 2015, **520**:567-570.

35. 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.

36. 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.

• 37. Scheres SH: **A Bayesian view on cryo-EM structure determination**. *J Mol Biol* 2012, **415**:406-418. This article describes the empirical Bayes method combining 3D multiconformation alignment, classification, and refinement that is available in RELION.

38. 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.

39. 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**.

40. 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 A resolution**. *Nature* 2016, **530**:298-302.

41. 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.

42. Scheres SH: **Classification of structural heterogeneity by maximum-likelihood methods**. *Methods Enzymol* 2010, **482**:295-320.

• 43. Frank J, Ourmazd A: **Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM**. *Methods* 2016, **100**:61-67. The conceptual outline of the manifold embedding approach to analyzing continuous conformational changes was here provided in greater detail.

• 44. 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. In this article, a method for data classification based on multivariate statistical analysis of a

series of 3D reconstructions from randomly selected subsets of images, used already in 2008, was described and demonstrated in more details.

• 45. Katsevich E, Katsevich A, Singer A: **Covariance Matrix Estimation for the Cryo-EM Heterogeneity Problem**. *SIAM J Imaging Sci* 2015, **8**:126-185. This article describes a methodology for estimating the entire actual covariance matrix from a heterogeneous set of particle images.

• 46. 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. Rather than estimating the entire covariance matrix, from a heterogeneous set of particle images, this article proposes the estimation in regions of interest to make it more efficient.

• 47. 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. This article describes a likelihood-based approach to principal component analysis of a heterogeneous set of particle images. This approach computes eigenvectors of the covariance matrix directly from images without estimating this matrix.

• 48. 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. This article describes a user-friendly graphical interface and a detailed protocol for analyzing continuous conformational changes with HEMNMA, which was out of scope of the article where HEMNMA was introduced.

49. Jonic S, Sorzano COS: **Coarse-graining of volumes for modeling of structure and dynamics in electron microscopy: Algorithm to automatically control accuracy of approximation**. *IEEE Journal of Selected Topics in Signal Processing* 2016, **10**:161 - 173.

• 50. 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. This article describes the StructMap method that projects a given set of EM density maps, obtained by a discrete-state method, onto a low-dimensional space of structural differences (distance space) and that can be combined with HEMNMA. Also, the article addresses extrapolation of continuous trajectories from their discrete, unordered samples obtained by discrete-state methods.

51. Walls AC, Tortorici MA, Bosch BJ, Frenz B, Rottier PJ, DiMaio F, Rey FA, Veesler D: **Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer**. *Nature* 2016, **531**:114-117.

FIGURES

This is the author's version of an article that has been published in *Current Opinion in Structural Biology*. Changes were made to this version by the publisher prior to publication. The final version is available at http://dx.doi.org/ 10.1016/j.sbi.2016.12.011 since 11 January 2017.

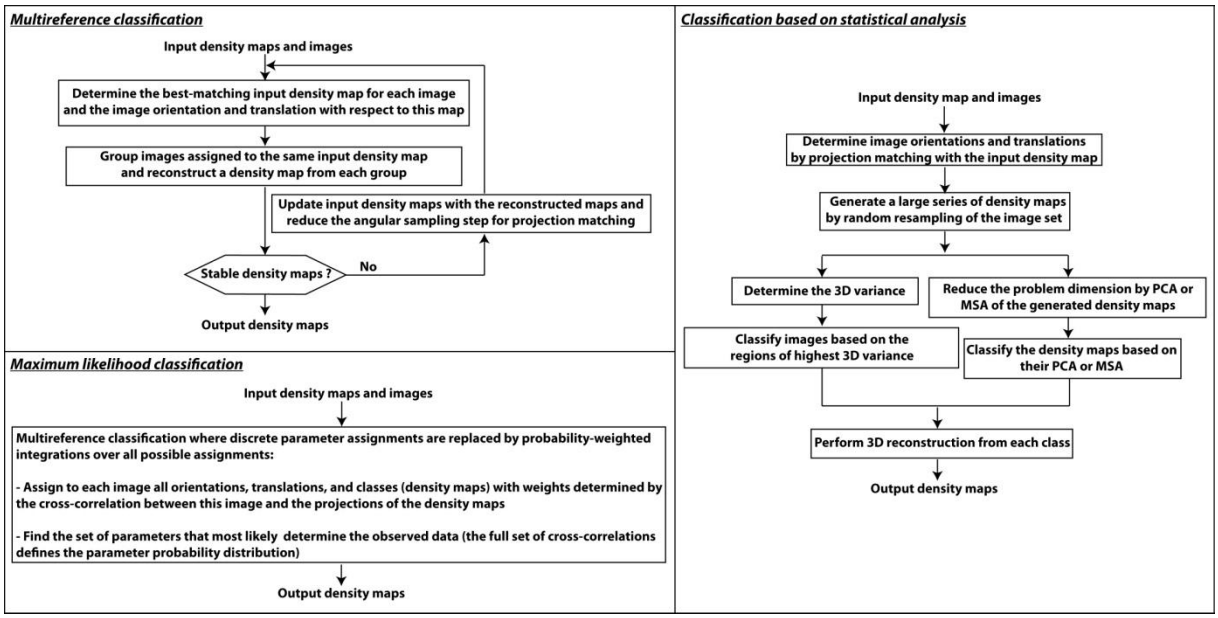


Figure 1: Discrete conformational analysis methods reviewed here, referred to as discrete-state methods.

This is the author's version of an article that has been published in *Current Opinion in Structural Biology*. Changes were made to this version by the publisher prior to publication. The final version is available at http://dx.doi.org/ 10.1016/j.sbi.2016.12.011 since 11 January 2017.

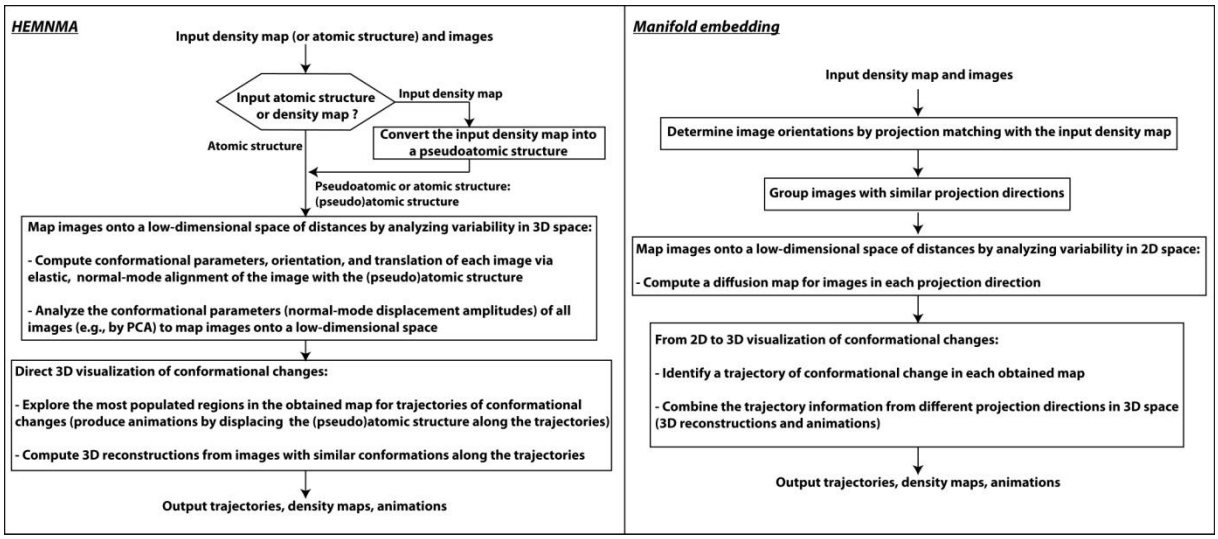


Figure 2: Continuous conformational analysis methods reviewed here, referred to as continuous-state methods.

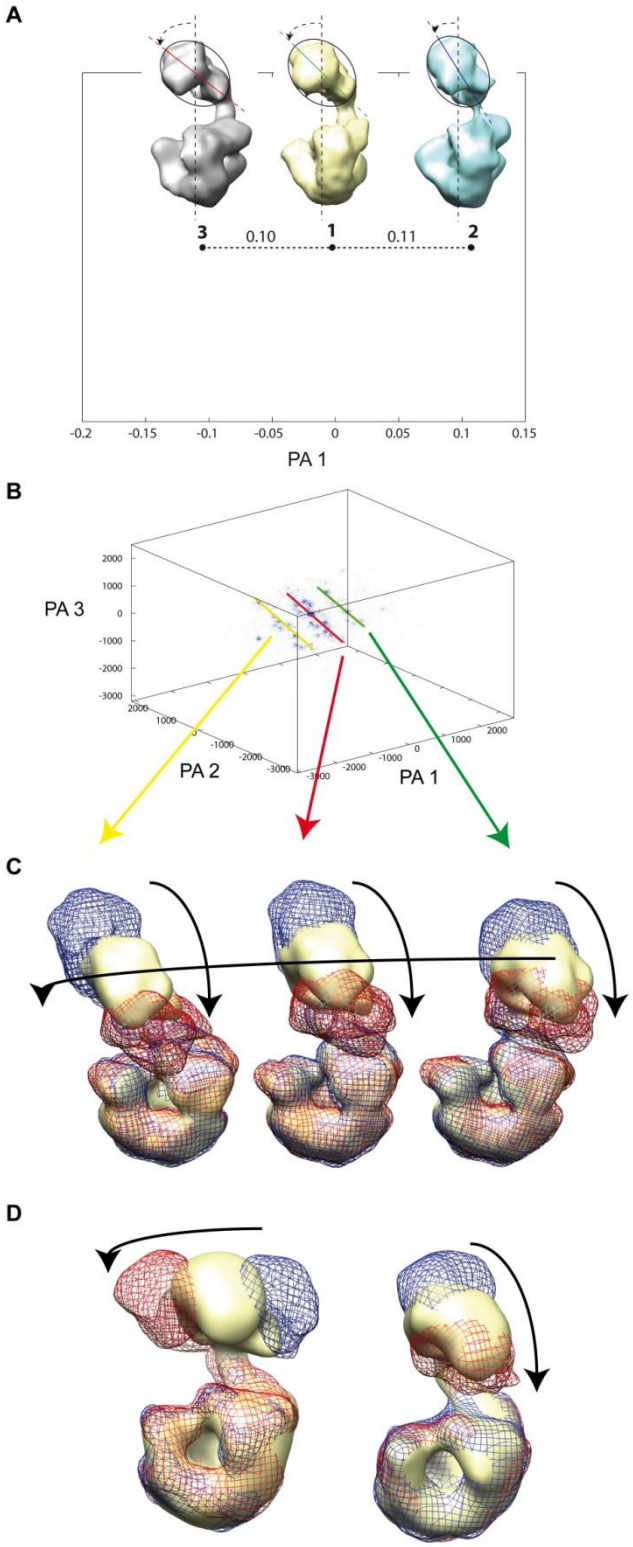


Figure 3: HEMNMA initialized with StructMap in the context of combining discrete-state and continuous-state methods. (A) StructMap projects a set of EM density maps obtained by a discrete-state method (here, ML classification method proposed in [35]) onto a low-dimensional (here, 1D) density-map distance space (in this space, each density map is marked

with an index and a point). (B) HEMNMA projects a set of images onto a low-dimensional (here, 3D) image distance space using normal modes of a pseudoatomic structure from the density-map distance space shown in (A) (here, a pseudoatomic structure of density map 3) to interpret the conformation in each image (in the image distance space, each image is marked with a point). (C) Animated displacement of the pseudoatomic structure along trajectories identified by exploring the most densely populated regions in the distance space shown in (B). (D) 3D reconstructions of density maps from subsets of images with similar conformations according to the distribution shown in (B). PA 1-3: principal axes 1-3. The workflow is illustrated using DNA polymerase Pol α–B subunit complex of the eukaryotic primosome. The distance spaces, the pseudoatomic-structure animations, and the reconstructions used to illustrate the workflow were reproduced with permission from [50] (panel (A)) and [7] (panels (B)-(D)).