Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Likelihood-based structural analysis of electron microscopy images

Pilar Cossio[1,2] and Gerhard Hummer[2,3]

Likelihood-based analysis of single-particle electron microscopy images has contributed much to the recent improvements in resolution. By treating particle orientations and classes probabilistically, uncertainties in the reconstruction process are explicitly accounted for, and the risk of bias towards the initial model is diminished. As a result, the quality and reliability of the reconstructions have greatly improved at manageable computational cost. Likelihood-based analysis of electron microscopy images also offers a route to direct coordinate refinement for dynamic systems, as an alternative to 3D density reconstruction. Here, we review recent developments in the algorithms used for reconstructions of high-resolution maps, and in the integrative framework of combining likelihood methods with simulations to address conformational variability in cryo-electron microscopy.

Addresses
[1] Biophysics of Tropical Diseases, Max Planck Tandem Group, University of Antioquia, Medellín, Colombia
[2] Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany
[3] Institute of Biophysics, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

Corresponding author: Cossio, Pilar (pilar.cossio@biophys.mpg.de)

Current Opinion in Structural Biology 2018, **49**:162–168

This review comes from a themed issue on **Theory and simulation**

Edited by **Kresten Lindorff-Larsen** and **Robert Best**

https://doi.org/10.1016/j.sbi.2018.03.004

Introduction

Likelihood-based methods play a central role in the analysis of images from electron microscopy (EM) experiments. Cryo-EM produces 2D projection images of individual particles frozen at near-native conditions. 3D density maps can be reconstructed if the 2D projections can be appropriately classified, and the coverage of orientation space is dense. However, the low signal-to-noise ratio of EM images makes it challenging to unambiguously determine the orientation and conformation of each individual particle. These uncertainties are taken into account using a likelihood function that assigns a probability to the orientations and conformational classes of each particle. This probabilistic approach has substantially reduced the risk of bias towards the starting model.

The foundations for likelihood-based techniques in cryo-EM were set by Sigworth in 1998 [1] with a method to align synthetic images for 2D class averaging. The underlying average is iteratively computed as a weighted sum over the possible in-plane rotations and translations of the individual images. The likelihood function quantifies the probability that an image arises by chance given a particular class-average. A key ingredient of any likelihood function is a model of the errors. For cryo-EM images, already the simplest model, white noise, helps avoid misalignment. Sigworth's method [1] was extended to multiple 2D references [2], and in 2005, it was optimized over a real cryo-EM dataset [3].

In a major step forward, likelihood formulations were then introduced into 3D reconstruction methods. Likelihood functions in 3D reconstruction quantify the degree of consistency between a fixed number of 3D maps and the particle images [4,5]. The maps are iteratively optimized to maximize a likelihood function that has been marginalized with respect to certain model parameters by integrating them out [6] (see Eq. [1]). However, the convergence of these methods is affected by the image quality. For noisy images, the methods can get trapped in a local optimum [7], leading to globally suboptimal reconstructions.

The recent dramatic advances in image quality have given a major boost to likelihood-based reconstruction methods. Direct electron detection cameras [8,9] record low-dose/low-defocus images. A fast frame readout rate makes it possible to correct for beam-induced motion and radiation damage [10–12]. The superior images analyzed with advanced algorithms, using fast likelihood-based formulations, have made it possible to reconstruct 3D maps at unprecedented resolution with reasonable computational costs [13,14].

However, sharper EM signals show that for many biomolecular systems the particle images do not represent a small set of discrete states but rather a continuous ensemble of conformations [15]. Reconstructing 3D maps of flexible systems is challenging because standard methods require a small number of conformational classes so that there are sufficient particle orientations to cover the 3D orientation space of each class. Extensive conformational variability constitutes a major challenge in cryo-EM.

The purpose of the review is twofold: first, to discuss the advances and optimizations in the likelihood-based algorithms that generate 3D reconstructions with unprecedented levels of speed and accuracy; and, second, to highlight alternative methods that address the challenge of conformational variability in the cryo-EM data by directly refining the conformational ensemble from simulations using a likelihood-based probability.

Likelihood and Bayesian analysis in cryo-EM

The likelihood function in cryo-EM reconstruction quantifies the probability that an experimental particle arises from a particular 3D model. Accuracy and computational speed depend on how the likelihood is formulated. The likelihood function arises from an assumption about the distribution of noise. Standard likelihood functions are based on Gaussian [4,5] or Poisson [16,17••] noise models. Assuming a Gaussian likelihood is equivalent to using the cross-correlation coefficient as a goodness-of-fit measure between the experimental image and a 2D image calculated from the 3D model.

The accuracy of the resulting reconstructions depends on how well the models and experimental uncertainties are represented. Detailed descriptions account for the model orientation, electron density projection, particle center translations, intensity uncertainties, and blurring effects caused by the intentional setting of the microscope out of focus (which are described using a contrast transfer function (CTF) [18]). However, sophisticated likelihood functions require the optimization of many variables (so-called 'nuisance parameters'), and thus entail higher computational costs. Most [19–21,22••] 3D reconstruction methods (Table 1) optimize only the orientations and translations but assume a constant microscope defocus for each individual particle. BioEM [23], an ensemble refinement method, treats the CTF parameters, intensity normalization and offset as additional nuisance variables.

Calculating the likelihood function in Fourier space brings several advantages. The projection slice theorem makes it possible to obtain a 2D projection from a 3D Fourier-transformed model without having to rotate or project it. In reciprocal space, the Gaussian error model can include colored noise by giving variable weights to different spatial frequencies. Moreover, in Fourier space it is straightforward to separate high and low frequency modes. This has been recently exploited in a branch-and-bound algorithm [22••] to discard poor orientations, as identified by using only the low frequency modes in a reduced likelihood function. Conversely, in a real-space treatment the costs are reduced by masking the particle and discarding the regions of only noise from the calculation [21].

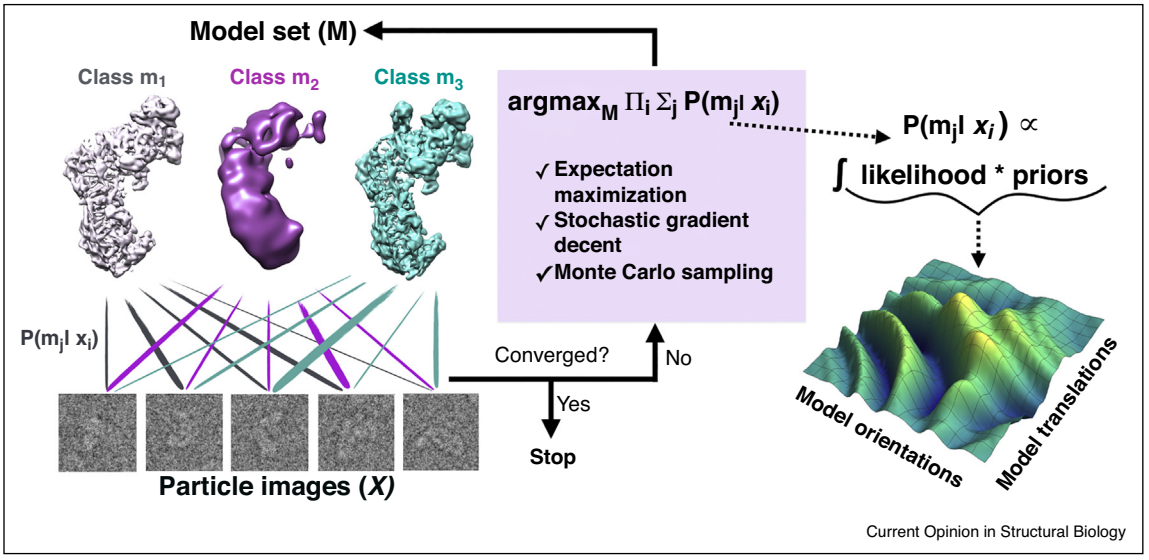
Prior knowledge about models and their parameters reduces the uncertainties in the cryo-EM data analysis. For white noise, likelihood maximization is equivalent to minimizing the squared difference of calculated and observed intensities in a least-squares fit. However, for low signal-to-noise ratios, these calculations can lead to erroneous particle classification [24]. Therefore, to reduce the errors, prior probabilities often modulate the likelihood function (creating a joint likelihood or Bayesian posterior; see Figure 1 right). Possible prior functions include Gaussian distributions of the particle centers

Table 1

Recent likelihood-based methods for 3D reconstruction and ensemble refinement from cryo-EM particle images. We report the name and reference of each method, the type of analysis (maximum or marginalized likelihood or Bayesian) and the reconstruction or ensemble refinement methods employed. We specify if the method performs Fourier-based 3D density map reconstructions from particle classification or direct model coordinate refinement

Method	Likelihood-based methods		Reconstructions		Dynamics and/or ensemble refinement			
			3D density maps refinement from particle averaging and classification			Direct model coordinate refinement		
	Maximum or marginalized likelihood	Bayesian	Expectation maximization	Stochastic gradient descent	Covariance or principal components	Monte Carlo sampling	Hybrid/ integrative simulations	Maximum entropy or minimal ensemble
XMIPP [19]	X		X					
RELION [20]	X		X					
FREALIGN [21]	X		X					
cryoSPARC [22••]	X		X	X				
sMAP-EM [16]	X	X	X					
MLV [32]	X		X		X			
Tagare *et al.* [34•]	X		X		X			
Joubert and Habeck [17••]		X				X		
BioEM [23]		X					X	X
EMageFit [53•]	X					X	X	X
Mosaics — EM [52•]						X	X	X

Figure 1



3D reconstruction and classification in cryo-EM. (Left) Likelihood-based reconstruction methods in cryo-EM aim to build a small set of model classes, $M = \{m_1, m_2, \ldots, m_j, \ldots\}$, (e.g. from [26]) that maximize the product over all particle images, $X = \{x_1, x_2, \ldots, x_i, \ldots\}$, of the sum of their posterior probabilities. Expectation-maximization, stochastic gradient descent or Monte Carlo sampling methods are used to find the optimal model classes. (Right) The posterior of each model class is the integral over the nuisance parameters (model orientations, translations etc.) of the likelihood function times the prior probabilities.

[19–21], of the reciprocal coordinates of the model [20], or of the microscope defocus (Cossio *et al.*, unpublished data). These priors regularize the problem and they can depend on adjustable 'hyperparameters'. However, if the priors are too strong, they shift the maxima; and if they are too weak, one effectively falls back to a least-squares calculation.

Marginalized likelihood and Bayesian methods do not search for optimal model parameters, but treat these as hidden variables by integrating them out of the joint likelihood. The marginalized likelihood of a 3D model m given particle images X, is the joint likelihood integrated over the nuisance parameters φ,

$$P(X|m) \propto \int L(X|\varphi, m)p(\varphi|m)d\varphi, \quad (1)$$

where $L(X|\varphi,m)$ is the likelihood function and $p(\varphi|m)$ is the probability distribution of the parameters φ conditional to model m. The Bayesian posterior is $P(m|X) \propto P(X|m)p(m)$ where $p(m)$ is the model prior. Thus, an essential feature of these approaches, as opposed to maximum likelihood, is that parameter probability distributions are not extremalized with respect to model parameters but integrated over.

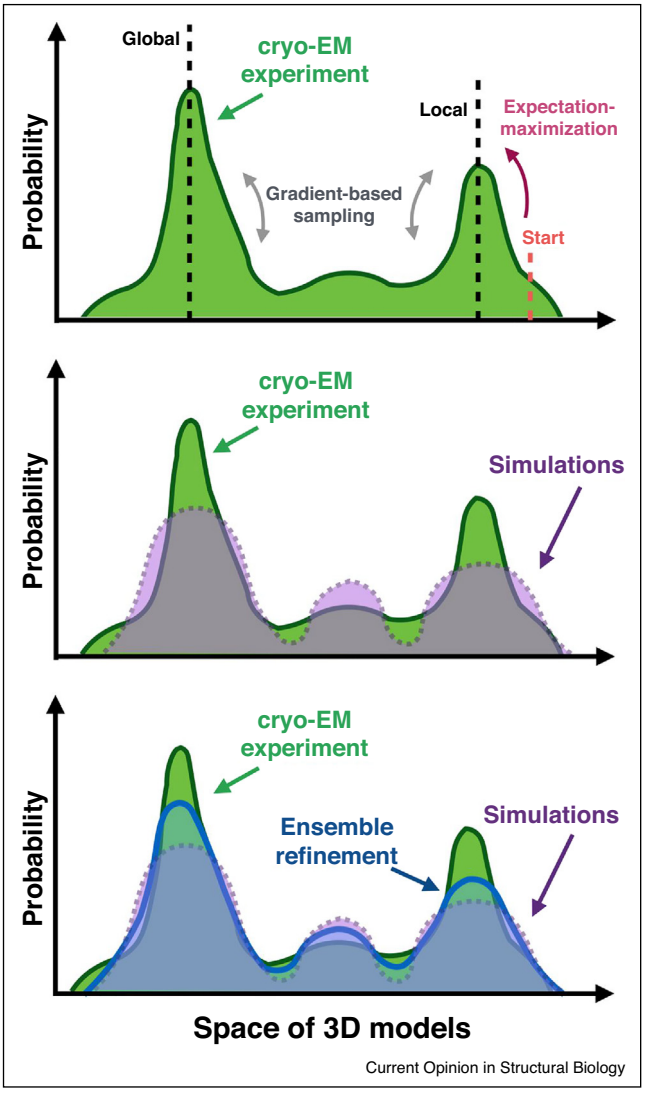
However, calculating the marginalized likelihood is computationally expensive. Whereas some integrals in Eq. [1] can be solved analytically [23], most have to be solved numerically, requiring an exhaustive parameter sampling. The numerical integration of the posterior can be sped up by iterative optimization of the integration grid or by a careful choice of the prior probabilities (e.g. for the orientations [20,25]).

Despite these advances, finding optimal 3D models for dynamic systems remains a major challenge in cryo-EM. A complete optimization of the posterior requires a thorough sampling not only of the parameters but, more importantly, of all possible conformations (i.e. the model space; Figure 2 top). In the following, we describe recent algorithms that iteratively optimize models in a computationally efficient manner. The majority of methods use Fourier-based reconstruction techniques to determine 3D density maps from particle alignment, averaging and classification. However, we also discuss integrative modeling methods that directly optimize the coordinates of 3D structural models, instead of reconstructing maps from particle classification.

3D reconstruction and classification

Cryo-EM likelihood-based reconstruction methods aim to obtain a set of heterogeneous 3D model classes that maximize the product over all particles of the sum of their posterior probabilities (exemplified with complex I [26] in Figure 1). In most cases, this is accomplished using the expectation-maximization algorithm [27]. Initially, this algorithm calculates the joint likelihood as a probability,

Figure 2



Ensemble refinement of continuous space of biomolecular conformations in cryo-EM. (Top) A schematic representation of the posterior probability as a function of the space of all possible 3D models. The expectation-maximization algorithm converges to the nearest local maximum (red arrow), therefore, the starting model (orange line) is important. Gradient-based methods (grey arrows) can iteratively build models with lower or higher probabilities, enabling the escape from local maxima and a wide search of the conformational space. (Middle) A representation of the conformational ensemble obtained from simulations, shown in purple. (Bottom) Ensemble refinement methods directly bias the model coordinates or reweigh the distribution from the simulations (blue), resulting in an ensemble that is in accordance with the posterior probability from the cryo-EM images and the potential energy from the simulations.

or expectation, of the parameters and model classes for each particle conditional on the current 3D maps. In a subsequent step, improved 3D maps are reconstructed as the average over all particles and all orientations weighted by the joint likelihood from the previous step. This process is repeated, and the algorithm guarantees reconstructing maps with equal or higher probability at each step. The algorithm is halted when the 3D maps stop improving. XMIPP [19] and RELION [20] use the expectation-maximization algorithm to optimize the classes and orientations, while FREALIGN [21] employs a mixed approach, obtaining optimal parameters from least-squares fits, and classes from expectation maximization.

A drawback of the expectation-maximization algorithm is that it guarantees convergence only to the nearest maximum but not to the global one. The starting models are important because they determine the local maxima to which the algorithm converges (orange and red arrows in Figure 2 top). Initial high-resolution models can strongly bias the weights of the particles to favor those that best resemble the starting model, as shown in [7]. Unfortunately, it is difficult to determine if the converged models correspond to the true global optimum. Therefore, it is advantageous to start with maps generated randomly or statistically from the particle images (e.g. [28,29]). Alternatively, Joubert and Habeck [17••] use a Bayesian framework with Monte Carlo sampling to build low-resolution reconstructions, avoiding initial model bias. These authors use Gaussian mixture models to represent the 3D density and reduce the conformational degrees of freedom.

Recently, Punjabi *et al.* developed cryoSPARC [22••], a method that finds the global optimum with high computational efficiency starting from any conformation. CryoSPARC first uses a stochastic gradient descent (SGD) algorithm to quickly build models from random subsets of particles. The models are built by using the gradient of the posterior probability [30] and can have lower probabilities than those from the previous iteration. This makes it possible to escape from local maxima and explore the conformation space (grey arrows Figure 2 top). However, so far SGD only provides intermediate resolution maps. CryoSPARC then uses these as starting models for the expectation-maximization algorithm, benefiting from the observation that, even close to the maximum, high-frequency information can be discarded using the branch-and-bound algorithm. These improvements reduce the time needed to reconstruct high-resolution maps from days to hours, as revealed with AAA+ unfoldase reconstructions [31].

Continuous conformations: cryo-EM ensemble refinement

Cryo-EM has the advantage that it does not measure averages but truly single-molecule data: each particle is a snapshot of the system frozen in a, possibly different, conformation. For stable biomolecules, with a small number of states, it is possible to visualize conformational variability using the expectation-maximization algorithm by computing the variance [32,33] or principal

components [34•] of the reconstructions. However, for biomolecules that have a continuous conformational spectrum (e.g. because of flexible linkers) reconstructing 3D maps is challenging [35].

For systems that have dynamic regions, some methods [36–39] perform a masked 2D or 3D refinement where parts of the complex (e.g. the flexible parts) are masked out. RELION subtracts the signal of these masked parts from the particle images [38••,39], and the modified particles are realigned and used to reconstruct the structures within the mask. These methods were used to characterize distinct conformations of the human γ-secretase [40]. A limitation is that the entire 3D maps and their orientations are assumed to be sufficiently good to mask and subtract the signal from the experimental particles, which for intrinsically flexible systems could be problematic.

To address continuous conformational changes, some techniques have focused on 2D particle classification [15], instead of using likelihood-based methods. Bootstrapping [41] and multivariate statistical analysis [42] pinpoint and classify regions of high variance in the images. By mapping the particles onto a low-dimensional projection of the conformation space using manifold embedding [43,44] or normal mode analysis [45], it is possible to identify continuous structural changes. Examples include the work cycle of the ribosome [43], the inflammasome assembly [44], or the motions in Pol α-B [45]. A challenge for these methods is to discriminate between changes in conformation and differences in orientation.

Marginalized likelihood and Bayesian methods can also address conformational dynamics (Table 1) by modifying the structural ensemble from molecular simulations (Monte Carlo or molecular dynamics) to account for experimental information [46•]. Some hybrid-methods use the 3D density map to add to the molecular force-field a biasing force that arises from the cross-correlation [47] or likelihood [48] between the simulated structure and 3D map. However, for highly dynamic systems, a low-resolution 3D map represents uncertainty in the model, and not necessarily an underlying dynamics.

For flexible biomolecules, one can adapt concepts from integrative structural modeling [49] and ensemble refinement [50] (Figure 2). The posterior probability, $P(m|X)$, of each model for given cryo-EM particles can be used to define a conformational ensemble. In the context of simulations, one could add the negative logarithm of the posterior probability to the force-field potential energy [50] to directly bias the model coordinates. Alternatively, one can conduct an *a posteriori* reweighing of the ensemble of conformations obtained from the simulations using either maximum entropy or minimal ensemble techniques [50,51•]. The posterior probabilities of the individual cryo-EM particles can then be used to extract the minimal number of models [23] required to account for the cryo-EM images.

Less computationally demanding are hybrid methods that combine 2D class-averages and Monte Carlo simulations. A multiscale natural moves algorithm simultaneously refines the model orientations and conformations using an *ad hoc* energy that quantifies the cross-correlation to the 2D class-averages [52•]. An integrative approach [53•] combines class-average maximum-likelihood scores with additional experimental information (e.g. crosslinking restraints) to assemble macromolecular complexes using simulated annealing. With these methods, conformational states of the Mm-cpn chaperone [52•] and of the TfR–Tf complex [53•] were resolved.

Conclusions

Likelihood-based methods constitute a major advance in cryo-EM image analysis and classification. Algorithm optimizations that include prior probabilities for incompletely resolved parameters, the evaluation of the likelihood in reciprocal space with a branch-and-bound algorithm, and novel reconstruction techniques, facilitate the rapid reconstruction of high-resolution maps of stable biomolecules. The bias towards the initial model can be greatly reduced if the maps are built by using the gradient of the posterior probability [22••].

Semi-flexible systems can be characterized using the variance [32] and principal components [34•] of the 3D reconstructions, or using masked 3D refinement with particle signal subtraction [38••]. For highly flexible systems, the posterior probability can directly bias the structural ensemble from simulations to be consistent with the particle images. However, cryo-EM refinement of dynamic biomolecules remains computationally expensive because of the vast model space that has to be explored. Software benefits from modern hardware like GPUs [22••,54•,55]. Nevertheless, there remains room for improvement in the methods. A key emerging challenge is the development of fast algorithms that use the gradient of the posterior from individual images (e.g. from [54•,55]) as a biasing force in molecular dynamics simulations.

Conformational variability is also a challenge for 3D map validation, where the gold-standard approach is not easily applicable with multiple reference reconstructions. Measures that validate the consistency of the particle alignment with the models [56] or use the posterior probability [23] can aid model validation. But robust and standardized protocols to determine the accuracy of the maps and ensembles of intrinsically flexible biomolecules are lacking in cryo-EM. Thus, despite their success likelihood-based methods in single-particle refinement, remain

under continued development to address both recognized and emerging challenges in model reconstruction and refinement, in structural variability, and in model validation.

Acknowledgements

The authors thank Dr. Edoardo D'Imprima and Dr. Matteo Allegretti for helpful feedback and discussions. P.C. and G.H. were supported by the Max Planck Society. P.C. was also supported by Colciencias and the University of Antioquia, Colombia.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Sigworth FJ: **A maximum-likelihood approach to single-particle image refinement**. *J Struct Biol* 1998, **122**:328-339.

2. Pascual-Montano A, Donate LE, Valle M, Bárcena M, Pascual-Marqui RD, Carazo JM: **A novel neural network technique for analysis and classification of EM single-particle images**. *J Struct Biol* 2001, **133**:233-245.

3. Scheres SHW, Valle M, Núñez R, Sorzano COS, Marabini R, Herman GT, Carazo JM: **Maximum-likelihood multi-reference refinement for electron microscopy images**. *J Mol Biol* 2005, **348**:139-149.

4. 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.

5. Scheres SHW, Núñez-Ramírez R, Gómez-Llorente Y, San Martín C, Eggermont PPB, Carazo JM: **Modeling experimental image formation for likelihood-based classification of electron microscopy data**. *Structure* 2007, **15**:1167-1177.

6. Sigworth FJ, Doerschuk PC, Carazo JM, Scheres SHW: **Maximum-likelihood methods in cryo-EM. Part I: Theoretical basis and overview of existing approaches**. *Methods Enzym* 2010, **482**:263-294.

7. Henderson R: **Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise**. *Proc Natl Acad Sci USA* 2013, **110**:18037-18041.

8. McMullan G, Chen S, Henderson R, Faruqi AR: **Detective quantum efficiency of electron area detectors in electron microscopy**. *Ultramicroscopy* 2009, **109**:1126-1143.

9. McMullan G, Clark AT, Turchetta R, Faruqi AR: **Enhanced imaging in low dose electron microscopy using electron counting**. *Ultramicroscopy* 2009, **109**:1411-1416.

10. Brilot AF, Chen JZ, Cheng A, Pan J, Harrison SC, Potter CS, Carragher B, Henderson R, Grigorieff N: **Beam-induced motion of vitrified specimen on holey carbon film**. *J Struct Biol* 2012, **177**:630-637.

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. Scheres SHW: **Beam-induced motion correction for sub-megadalton cryo-EM particles**. *Elife* 2014, **3**:e03665.

13. Kühlbrandt W: **The resolution revolution**. *Science* 2014, **343**:1443-1444.

14. Cheng Y: **Single-particle Cryo-EM at crystallographic resolution**. *Cell* 2015, **161**:450-457.

15. Jonić S: **Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images**. *Curr Opin Struct Biol* 2017, **43**:114-121.

16. Moriya T, Acar E, Cheng RH, Ruotsalainen U: **A Bayesian approach for suppression of limited angular sampling artifacts in single particle 3D reconstruction**. *J Struct Biol* 2015, **191**:318-331.

17. •• Joubert P, Habeck M: **Bayesian inference of initial models in cryo-electron microscopy using pseudo-atoms**. *Biophys J* 2015, **108**:1-6.
Bayesian analysis of cryo-EM images and Monte Carlo sampling to build low resolution models, avoiding initial model bias. Pseudo-atoms are used to represent the 3D model.

18. Penczek PA: **Image restoration in cryo-electron microscopy**. *Methods Enzym* 2010, **482**:35-72.

19. Scheres SHW, Núñez-Ramírez R, Sorzano COS, Carazo JM, Marabini R: **Image processing for electron microscopy single-particle analysis using XMIPP**. *Nat Protoc* 2008, **3**:977-990.

20. Scheres SHW: **A Bayesian view on cryo-EM structure determination**. *J Mol Biol* 2012, **415**:406-418.

21. Lyumkis D, Brilot AF, Theobald DL, Grigorieff N: **Likelihood-based classification of cryo-EM images using FREALIGN**. *J Struct Biol* 2013, **183**:377-388.

22. •• Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA: **cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination**. *Nat Methods* 2017, **14**:290-296.
3D reconstruction method that uses a stochastic gradient descent algorithm to rapidly build low-resolution models and avoids getting trapped in local maxima. The models are starting points for the expectation-maximization using a branch-and-bound algorithm to build high-resolution reconstructions (see also [30]).

23. Cossio P, Hummer G: **Bayesian analysis of individual electron microscopy images: towards structures of dynamic and heterogeneous biomolecular assemblies**. *J Struct Biol* 2013, **184**:427-437.

24. Sorzano COS, Bilbao-Castro JR, Shkolnisky Y, Alcorlo M, Melero R, Caffarena-Fernández G, Li M, Xu G, Marabini R, Carazo JM: **A clustering approach to multireference alignment of single-particle projections in electron microscopy**. *J Struct Biol* 2010, **171**:197-206.

25. Gorski KM, Hivon E, Banday AJ, Wandelt BD, Hansen FK, Reinecke M, Bartelman M: **HEALPix — a framework for high resolution discretization, and fast analysis of data distributed on the sphere**. *Astrophys J* 2004, **622**:759-771.

26. D'Imprima E, Mills DJ, Parey K, Brandt U, Kühlbrandt W, Zickermann V, Vonck J: **Cryo-EM structure of respiratory complex I reveals a link to mitochondrial sulfur metabolism**. *Biochim Biophys Acta* 2016, **1857**:1935-1942.

27. Dempster AP, Laird NM, Rubin DB: **Maximum likelihood from incomplete data via the EM algorithm**. *J R Stat Soc B* 1977, **39**:1-38.

28. Vargas J, Alvarez-Cabrera A-L, Marabini R, Carazo JM, Sorzano COS: **Efficient initial volume determination from electron microscopy images of single particles**. *Bioinformatics* 2014, **30**:2891-2898.

29. Reboul CF, Bonnet F, Elmlund D, Elmlund H: **A stochastic hill climbing approach for simultaneous 2D alignment and clustering of cryogenic electron microscopy images**. *Structure* 2016, **24**:988-996.

30. Punjanji A, Brubaker MA, Fleet DJ: **Building proteins in a day: efficient 3D molecular reconstruction**. *IEEE Trans Pattern Anal Mach Intell* 2017, **39**:706-708.

31. Ripstein ZA, Huang R, Augustyniak R, Kay LE, Rubinstein JL: **Structure of a AAA+ unfoldase in the process of unfolding substrate**. *Elife* 2017, **6**:1-14.

32. Wang Q, Matsui T, Domitrovic T, Zheng Y, Doerschuk PC, Johnson JE: **Dynamics in cryo EM reconstructions visualized with maximum-likelihood derived variance maps**. *J Struct Biol* 2013, **181**:195-206.

33. Fu C, Wang K, Prevelige PE, Catalano CE, Johnson JE, Doerschuk PC: **Detecting asymmetry in the presence of symmetry with maximum likelihood three-dimensional**

reconstructions of viruses from electron microscope images. *IET Image Process* 2016, **10**:624-629.

34. • 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.
A method that extends the Fourier-slice theorem to calculate the principal components in cryo-EM reconstructions using the expectation-maximization algorithm.

35. White HE, Ignatiou A, Clare DK, Orlova EV: **Structural study of heterogeneous biological samples by cryoelectron microscopy and image processing**. *Biomed Res Int* 2017, **2017**:1032432.

36. Briggs JAG, Huiskonen JT, Fernando KV, Gilbert RJC, 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.

37. Grigorieff N: **Frealign: an exploratory tool for single-particle Cryo-EM**. *Methods Enzymol* 2016, **579**:191-226.

38. •• Scheres SHW: **Processing of structurally heterogeneous cryo-EM data in RELION**. *Methods Enzymol* 2016, **579**:125-157.
Review on the protocols for particle polishing and post-processing to reconstruct high-resolution maps with RELION. Emphasis is placed on how to deal with particles of heterogeneous systems using 3D masking and particle signal subtraction techniques.

39. Ilca SL, Kotecha A, Sun X, Poranen MM, Stuart DI, Huiskonen JT: **Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes**. *Nat Commun* 2015, **6**:8843.

40. Bai X-C, Rajendra E, Yang G, Shi Y, Scheres SH: **Sampling the conformational space of the catalytic subunit of human γ-secretase**. *Elife* 2015, **4**:e11182.

41. Penczek PA, Frank J, Spahn CMT: **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.

42. van Heel M, Portugal RV, Schatz M: **Multivariate statistical analysis of large datasets: single particle electron microscopy**. *Open J Stat* 2016, **6**:701-739.

43. 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.

44. Wu J, Ma Y, Congdon C, Brett B, Chen S, Xu Y, Ouyang Q, Mao Y: **Massively parallel unsupervised single-particle cryo-EM data clustering via statistical manifold learning**. *PLOS ONE* 2017, **12**:e0182130.

45. Jin Q, Sorzano COS, De La Rosa-Trevín JM, Bilbao-Castro JR, Núñez-Ramírez 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.

46. • Schröder GF: **Hybrid methods for macromolecular structure determination: experiment with expectations**. *Curr Opin Struct Biol* 2015, **31**:20-27.
Review on the strategies to combine molecular simulations with experimental data.

47. Goh BC, Hadden JA, Bernardi RC, Singharoy A, McGreevy R, Rudack T, Cassidy CK, Schulten K: **Computational methodologies for real-space structural refinement of large macromolecular complexes**. *Annu Rev Biophys* 2016, **45**:253-278.

48. Hanot S, Bonomi M, Greenberg CH, Sali A, Nilges M, Vendruscolo M, Pellarin R: **Multi-scale Bayesian modeling of cryo-electron microscopy density maps**. *bioRxiv* 2017 http://dx.doi.org/10.1101/113951.

49. Sali A, Berman HM, Schwede T, Trewhella J, Kleywegt G, Burley SK, Markley J, Nakamura H, Adams P, Bonvin AMJJ *et al.*: **Outcome of the first wwPDB hybrid/integrative methods task force workshop**. *Structure* 2015, **23**:1156-1167.

50. Hummer G, Köfinger J: **Bayesian ensemble refinement by replica simulations and reweighting**. *J Chem Phys* 2015, **143**:243150.

51. • Bonomi M, Heller GT, Camilloni C, Vendruscolo M: **Principles of protein structural ensemble determination**. *Curr Opin Struct Biol* 2017, **42**:106-116.
A review on the maximum entropy and maximum parsimony methods for protein ensemble determination when dealing with experimental measurements that are averages over the conformational ensemble.

52. • Zhang J, Minary P, Levitt M: **Multiscale natural moves refine macromolecules using single-particle electron microscopy projection images**. *Proc Natl Acad Sci U S A* 2012, **109**:9845-9850.
Multiscale natural moves Monte Carlo algorithm for the conformational refinement of macromolecules using the cross-correlation to 2D class-averages.

53. • Velazquez-Muriel J, Lasker K, Russel D, Phillips J, Webb BM, Schneidman-Duhovny D, Sali A: **Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images**. *Proc Natl Acad Sci U S A* 2012, **109**:18821-18826.
Integrative Monte Carlo method that combines experimental data from diverse techniques, and includes a likelihood-based score to 2D class-averages, for assembling macromolecular complexes.

54. • Cossio P, Rohr D, Baruffa F, Rampp M, Lindenstruth V, Hummer G: **BioEM: GPU-accelerated computing of Bayesian inference of electron microscopy images**. *Comput Phys Commun* 2017, **210**:163-171.
Ensemble refinement software that performs Bayesian inference of cryo-EM images, optimized for GPU parallelization. The posterior probability, is used to rank, classify and refine model ensembles.

55. Kimanius D, Forsberg BO, Scheres S, Lindahl E: **Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2**. *Elife* 2016, **5**:e18722.

56. Vargas J, Otón J, Marabini R, Carazo JM, Sorzano COS: **Particle alignment reliability in single particle electron cryomicroscopy: a general approach**. *Sci Rep* 2016, **6**:21626.