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