Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Advances in image processing for single-particle analysis by electron cryomicroscopy and challenges ahead

JL Vilas[2], N Tabassum[1], J Mota[2], D Maluenda[2], A Jiménez-Moreno[2], T Majtner[2], JM Carazo[2], ST Acton[1] and COS Sorzano[2,3]

Electron cryomicroscopy (cryoEM) is essential for the study and functional understanding of non-crystalline macromolecules such as proteins. These molecules cannot be imaged using X-ray crystallography or other popular methods. CryoEM has been successfully used to visualize macromolecular complexes such as ribosomes, viruses, and ion channels. Determination of structural models of these at various conformational states leads to insight on how these molecules function. Recent advances in imaging technology have given cryoEM a scientific rebirth. As a result of these technological advances image processing and analysis have yielded molecular structures at atomic resolution. Nevertheless there continue to be challenges in image processing, and in this article we will touch on the most essential in order to derive an accurate three-dimensional model from noisy projection images. Traditional approaches, such as k-means clustering for class averaging, will be provided as background. We will then highlight new approaches for each image processing subproblem, including a 3D reconstruction method for asymmetric molecules using just two projection images and deep learning algorithms for automated particle picking.

Addresses
[1] Virginia Image and Video Analysis, Univ. of Virginia, P.O. Box 400743, Charlottesville, VA 22904, USA
[2] Biocomputing Unit, Centro Nacional de Biotecnología (CNB-CSIC), Darwin, 3, Campus Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain
[3] Bioengineering Lab, Escuela Politécnica Superior, Universidad San Pablo CEU, Campus Urb. Montepríncipe s/n, 28668, Boadilla del Monte, Madrid, Spain

Corresponding author: Sorzano, C.O.S. (coss@cnb.csic.es)

Current Opinion in Structural Biology 2018, **52**:127–145

This review comes from a themed issue on **Biophysical and computational methods**

Edited by **Gregory A Voth** and **Mark Yeager**

For a complete overview see the Issue and the Editorial

https://doi.org/10.1016/j.sbi.2018.11.004

Introduction

Cryo-electron microscopy (cryoEM) of single particles has been established as a key technique for the elucidation of the three-dimensional structure of biological macromolecules. The *Nature Methods* Method of the Year (2015) and the Nobel Prize in Chemistry (2017) endorse this view. CryoEM is currently capable of achieving quasi-atomic resolution (1.8 Å) in some specimens, and visualizing specimens with molecular weights below 100 kDa with a resolution better than 4 Å [1•]. Beside that, CryoEM can yield key insight into the dynamics of macromolecules [2–4], and it provides a solid base for structure-based drug design, although some technical problems in this arena remain open [5].

The main advances in the last five years have come from multiple sources: (1) more sensitive and faster direct electron detectors, (2) faster and more robust image processing algorithms, and (3) more reproducible sample preparation techniques.

In this review we address the developments in image processing algorithms over the last five years. To begin, we summarize advances in other aspects of EM (not covered in this review) that also affect the image quality:

- *Image formation process*. Much attention has been placed on a better understanding of the physicochemical processes leading to radiation damage [6–8], beam induced movement [9,10] characterizing camera noise (modeling the noise produced by sensors capturing EM images) [11,12], modelling and correcting optical aberrations [13,14•,15], especially the defocus gradient along the specimen [16,17,18•], the effects of charging [19,20], the design and use of phase plates as a way to increase contrast [21•,22,23], and single band imaging as a way to address the defocus gradient [24,25,26•].
- *Better detectors*. Direct electron detectors (DEDs) have caused a quantum leap in EM. The current trends include thinner back-ends as a way to reduce the actual size of the point spread function, increased quantum efficiency of the detector in order to increase its sensitivity, and faster readouts as a way to better correct for the beam-induced movement [27,28••].

- *Better sample preparation*. Research in sample preparation has focused on increasing the sample stability [29] and reducing the amount of sample required for vitrification, as a way to increase freezing speed and reproducibility [30–33].

Our presentation is organized as follows: Section 2 reviews the advances during the last five years in image processing algorithms for single-particle analysis. Section 3 presents current challenges in the field from the algorithmic point of view, followed by conclusions. Figure 1 displays a graphical summary of the main topics discussed. The blue arrows between 2D Processing and 3D Analysis depict the cyclical nature of different stages, in which the order of steps may vary from method to method.

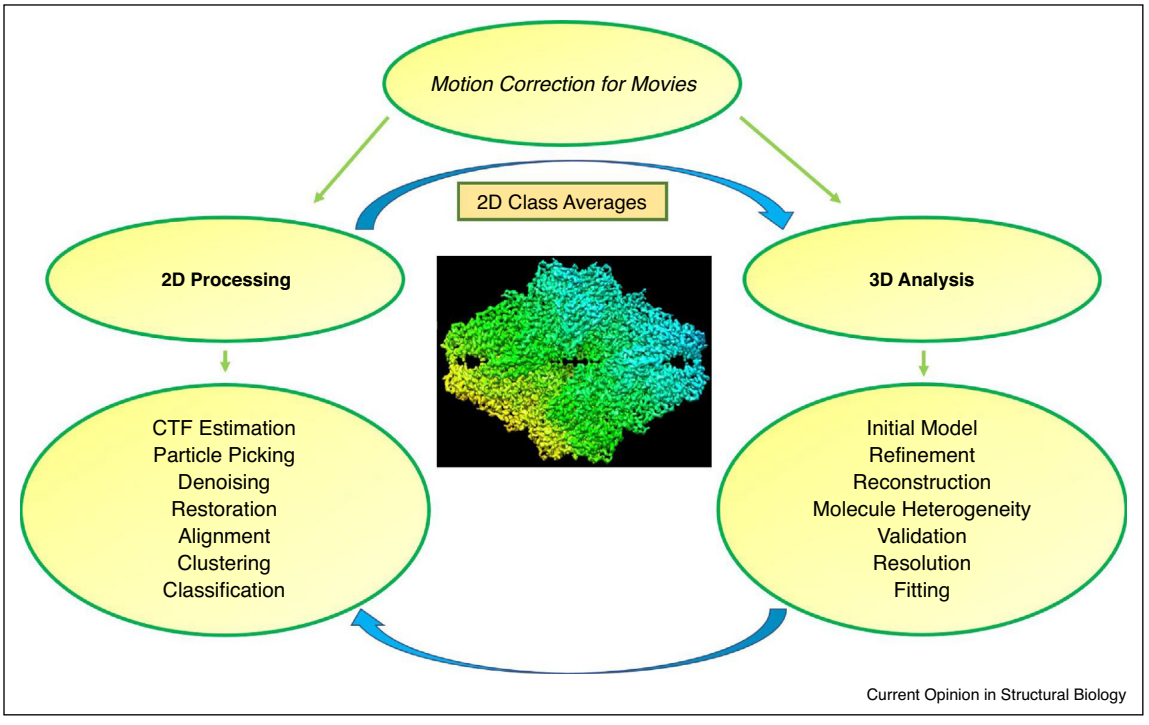
Recent advances in image processing algorithms for single-particle analysis

Software packages tend to be very inclusive, covering the whole pipeline from image acquisition to the final 3D reconstruction (Relion [34•], Eman2 [35], Frealign and Cistem [36], Xmipp [37•], Spider [38], Sparx [39], Bsoft [40]). These packages even include small tools from other software providers solving specific image processing problems. Two large integrative platforms have appeared in the domain: Scipion [41] and Appion [42]. In these platforms, the user may easily call different algorithms from different providers, and the system automatically performs the necessary conversions. In recent years, many engineering groups are contributing software that solve very specific problems along the image processing pipeline. These tools tend to be incorporated in the integrative platforms.

Movies and micrographs

The contrast between the sample and its background is one of the factors that determine the final quality of an image. Grant and Grigorieff [43•] demonstrated a method of using optimal exposure values to filter movie frames, yielding images with improved contrast that resulted in higher-resolution 3D reconstructions. They examined how quickly a large virus-like particle is damaged under the electron beam. These experiments identified an optimum range of exposure to electrons that provides the highest image contrast at any given level of detail. Their findings were used to design an exposure filter that can be applied to the movie frames. With higher contrast, greater levels of structural information can be obtained. However, this increase in contrast requires the use of longer exposure to the electron beam. To overcome this issue, instead of recording a single image, it is possible to

Figure 1



Summary of the main topics discussed in this review. Pictured is a 3D reconstruction of β-galactosidase (isosurface representation of Xmipp highres reconstruction).

record movies in which the movement of the sample under the electron beam can be tracked. The correction of beam-induced specimen movement was solved by a number of algorithms. Ripstein et al. [44] explained and compared several of the most popular existing algorithms for computationally correcting specimen movement including *Motioncorr* [45], *alignframes_lmbfgs* and *alignparts_lmbfgs* [46], *Unblur* [43•], and others, while summarizing all the advantages of each technique.

While conceptually simple, the algorithms used to perform motion correction vary widely, because each alignment routine uses different criteria to guide and smooth the alignment. By understanding the different approaches, we may achieve insights to better design the next generation of alignment software. McLeod et al. [47] presented a software package *Zorro*, which provides robust drift correction for dose fractionation by use of an intensity-normalized cross-correlation and logistic noise model to weight each cross-correlation in the multi-reference model and filter each cross-correlation optimally. Frames are reliably registered with low dose and defocus. The package utilizes minimal heuristics that minimizes the number of arbitrary input parameters required by the user. The most critical input parameters, weighting of peak significance and B-filter strength, are performed automatically.

Recently, a novel software tool *MotionCor2* [48] used anisotropic correction of beam-induced motion. The algorithm is based on an experimentally validated model that describes the sample motion as a local deformation that varies smoothly throughout the exposure. It combines the correction of both uniform whole-frame motion and anisotropic local motion, and it streamlines all the necessary preprocessing steps including bad pixel detection and correction before proceeding with the normal image processing pipeline.

One challenge in analyzing movies relates to their acquisition using a DED, in which there may be non-negligible differences between the gain of different sensor areas. Therefore, approaches to estimate the DED camera gain at the pixel level were developed. Afanasyev et al. [49] assimilate the gain of the camera to the standard deviation of each pixel over a large number of movies and prove that this is a successful way of identifying dead pixels. However, Sorzano et al. [50] showed that this approach does not provide a consistent gain estimation; therefore, they introduced a different approach to estimate the DED camera gain at each pixel from the movies. Their algorithm iteratively refines the gain image using local smoothness of the histograms of image rows and columns. A monitor of the gain estimate can be set to warn the user if the residual acquisition gain goes beyond certain limits (defined by the user as thresholds on its standard deviation and other percentile based parameters).

2D processing

CTF estimation

An electron microscope, as with any other imaging device, has a number of physical aberrations that distort the ideal projections, by modulating amplitudes and phases of the recorded electrons. To achieve the highest resolution, it is necessary to correct these distortions by estimating and correcting the contrast transfer function (CTF). The fitting procedure consists in an iterative adjustment minimizing the discrepancy between simulated and experimental power spectral densities (PSD) using a non-linear optimization that depends on an initial estimation of the model parameters, particularly the defocus. Several improvements of the CTF estimation have been done in recent years trying to improve the computation time and the accuracy, due to the large number of micrographs to analyze. A novel parameter-free approach has been presented [51] that uses a fast way to recover the defocus and astigmatism of the CTF without the need for non-linear optimization procedures and an initial estimation of the defocus. This method is available in Xmipp 3.0 [37•]. Other software has been developed for the CTF estimation such as CTFFIND4, which provides an improved version of CTFFIND3 that is faster and more suitable for images collected using modern technologies such as dose fractionationand and the use of a phase plate [52]. Gctf accelerates the CTF estimation using graphics processing units (GPU). This approach maximizes the cross-correlation of a simulated CTF with the logarithmic amplitude spectra of observed micrographs after background subtraction. In addition, an approach for local CTF refinement of each particle in a micrograph or frames in a movie is provided to improve the accuracy of CTF determination [53]. With the different programs available, it is becoming more difficult to compare their results across several runs and to select the best parameters to measure the CTF quality. To address this difficulty, a new parameter, the so-called CTF resolution, has been proposed in which they measure the correlation falloff of the calculated CTF oscillations against the normalized oscillating signal of the data. It is a robust metric to select the best parameters for each micrograph.

A novel phase contrast technique called the volta phase plate (VPP) [21•] has been developed in recent years to increase the contrast in electron micrographs. The phase shift introduced by a physical phase plate inserted in the microscope column maximizes contrast in low frequencies, thus producing a better contrast between particles and their background. The main problem of this method is that the image acquisition is in-focus and it is not possible to estimate the CTF, precluding corrections for physical aberrations. Danev et al. [55] proposed the use of the VPP with a bit of defocus. The advantage of this proposal is that the defocus can now be readily identified

through the oscillations of the Thon rings; however a drawback is that the small defocus causes damping of some high frequencies. The CTF correction for the Volta Phase Plate data is available in the three software packages mentioned previously.

Particle picking

Because of the strong background noise, low contrast images, and sample heterogeneity, it is necessary to record a large number of single-particle images in order to determine a reliable high-resolution 3D reconstruction. Methods for particle picking from micrographs can be divided into two main categories. The first is manual particle picking, which is obviously laborious and time-consuming. A large amount of human effort is required to obtain a sufficient number of high quality particles. Moreover, manual picking is considered subjective and can introduce bias and inconsistency.

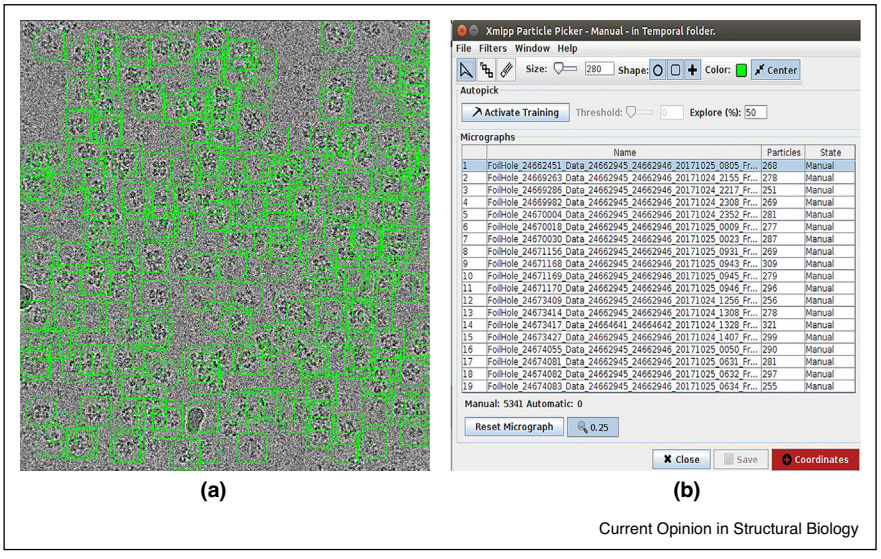
Therefore, the second category consisting of semi-automated and automated methods is more popular. This category includes generative approaches, which measure the similarity to a certain reference image. A typical generative approach is a template-matching technique, which is employed in RELION [56,57] or in highly parallel GPU-accelerated gEMpicker [58]. The input consists of a micrograph and images containing 2D templates for matching. The idea behind template-matching is that the cross-correlation between a template image and a micrograph is larger in the presence of the template. Template images may be chosen as a disk with a radius corresponding to the particle size with its edges softened by application of a Gaussian kernel [59]. Another alternative is *Gautomatch* developed by Zhang [60], which is a GPU accelerated program for flexible and fully automated particle picking from micrographs with or without templates. The automatic particle picker can learn from the particles picked by the user [61]. This method is available in Xmipp 3.0 software [37•], and an example is shown in Figure 2.

Since automatic and semi-automatic particle pickers are selecting a non-negligible number of incorrect particles, particle quality assessment and a sorting method based on multivariate statistical analysis of a particle set could be used to separate erroneously picked particles from correct ones [62]. The problem of discriminating between particles on carbon and particles in ice is solved by detecting carbon supports using the EMHP package [63].

In recent years, deep learning methods have been employed for particle picking in regular micrographs (i.e., not tilted pairs). *DeepPicker* [64] consists of two modules, where a labeled set of positive and negative samples are used to train a convolutional neural network (CNN) model, while in the particle picking module, the trained CNN classifier is then used to select particle images from the input micrographs. DeepEM is another recent method using deep CNN [65].

In cases when an initial model is not available, a low-to-medium resolution model can be obtained from negatively stained samples by the random conical tilt (RCT)

Figure 2



Use of the Xmipp particle picker with user input to select single particles. (a) Particles are detected and highlighted in the recorded micrograph. (b) Xmipp particle picker interface with a list of all micrographs showing the number of particles found in each micrograph.

[66] or orthogonal tilt reconstruction (OTR) [67] procedures. The basis for these two methods is in collecting two images of the same sample at different tilt angles, and identifying and boxing particles in both images. An accurate solution to finding both the particle correspondence and the tilt-axis estimation was proposed in [68] using *MaverickTilt* software to determine the tilt pairs from independent particle coordinates from images [69]. Vilas et al. introduced a method to automatically find correspondences of particles in the untilted and tilted micrographs [70], and the method is available in Scipion [41].

Denoising and image restoration

During the acquisition process, images are usually degraded by blur and noise. Most imaging devices, like CMOS and CCD cameras, are photon counting devices where the resulting noise is non-additive and signal-dependent, which can be modelled by a mixed Poisson–Gaussian (PG) distribution, often encountered in astronomy [71,72], biology [73] and medicine [74]. Image restoration methods (CTF correction and denoising) are based on estimating the original images from these blurred and noisy observations. In the first step, restoration methods are separated into two groups, non-blinded and blinded, depending on whether the point spread function (PSF) is known or not.

In addition, the non-blinded image restoration techniques are broadly categorized into two kinds of approaches [75••]. The first is an approach known as *phase flipping*, which involves flipping the sign of the Fourier coefficients at frequencies for whose CTF amplitude is negative, ignoring the effect of the CTF on the Fourier amplitudes. Phase flipping is easy to implement and preserves the noise statistics. The second commonly used approach is Wiener filtering (WF), which takes into account both the phases and amplitudes of the Fourier coefficients. However, to calculate the Wiener filter a prior estimation of the spectral signal to noise ratio (SSNR) of the signa is required, which by itself is a challenging problem.

Bhamre et al. [76] presented a new approach for non-blinded image restoration of cryoEM images based on a modified Wiener filtering. They named it the covariance Wiener filter (CWF) because the main algorithmic step is the estimation of the covariance. CWF performs phase and amplitude CTF correction, as well as denoising, thus improving the SNR of the resulting images. In particular, CWF applies Wiener filtering in the data-dependent basis of principal components (*eigenimages*), while traditional Wiener filtering is applied in the data-independent Fourier basis.

The first step of CWF is estimation of the covariance matrix of the underlying clean images, whereas the second step is solving a deconvolution problem to recover the underlying clean images using the estimated covariance.

In this statistical model, the Fourier transformed clean images are assumed to be independent, identically distributed (i.i.d.) samples. Since the clean images are two-dimensional (2D) projections of a certain 3D structure in different orientations, the covariance matrix represents the overall image variability due to the 3D structure, the distribution of orientations, and the varying contrast due to changes in ice thickness and structural variability, which are all of course unknown at this stage. While these model assumptions do not necessarily hold in reality [77,78], they simplify the analysis and lead to excellent denoising.

The method is thought to deal with images that have an additive white noise, which has equal intensity at different frequencies. However, for a more realistic colored noise process, with different power spectra, the images are processed in order to *whiten* the noise. The noise power spectrum is estimated using the pixels in the corners of the experimental images. One can define a new *effective* CTF including the whiten filter to estimate the new covariance matrix. However, this case is ill-conditioned, and it takes a large number of iterations for the conjugate gradient to converge to the desired solution. Instead, a well conditioned linear system is sought similar to one in the case of white noise.

The second step of the CWF is to use the estimated covariance to solve the associated deconvolution problem using Wiener filtering. The result is a denoised and CTF-corrected image for each experimental image.

On the other hand, in many situations it is difficult to accurately estimate the PSF (or the CTF) and blind methods may be preferable. Bajic et al. [79] presented a novel restoration method for images degraded with PG noise which jointly estimates the original image and the PSF from the observed data. Although the method was not designed to process cryoEM images, they illustrate its applicability in this field.

To simultaneously recover the original image and the PSF, the method minimizes an objective function. That function firstly contains a term that depends on the targets (clean image and PSF), driving the solution towards the observed data. Secondly, a regularization term which only depends on the clean image provides a noise suppression, whereas a parameter controls the trade-off of the two terms. The role of the regularization term is to provide numerical stability and it may be designed based on the desired characteristics of the unknown image, such as wavelet-based sparsity, smoothness, small total variation, etc.

During the clean image estimation, minimization of the objective function is seen as a constrained optimization problem that can be optimized by means of an iterative gradient-based method.

2D alignment, clustering, and classification

One of the main drawbacks of the cryoEM single-particle analysis is dealing with images with poor SNR, which requires the acquisition of a large number of experimental images. Therefore, averaging all similar and aligned images can substantially enhance the SNR. The averaged images are normally referred to as 2D averages, and they can be used to produce a reliable 3D starting model [80–82]. Most approaches used to simultaneously 2D align and cluster (SAC) are based on multi-reference alignment (MRA) following a k-means strategy. This strategy involves some randomized initial cluster centers followed with an iterative local-search-based cluster assignment and in-plane rotation [83]. It is possible to employ a previous step of principal component analysis (PCA), so that the clustering is actually performed using a low dimensional representation of the particles, accelerating the process.

The results from MRA using k-means strongly depends on the cluster initialization and the number of classes [84], compromising the reproducibility and robustness of the method. Reboul et al. [85] presented a stochastic hill climbing (SHC) method based on random walks, where the correlation maximizing step of k-means is replaced with the relaxed requirement of identifying *the first in-plane rotation and cluster assignment that improves the previous correlation, given random sequences of in-plane rotation and cluster assignments*. Thus, the references are randomly ordered and the rotation scan is also performed randomly. As soon as a configuration is improving the previous best correlation, the random walk ends and the next particle is processed. Since the cluster centers are not updated until all particles are done, the random walk is performed on all particles independently. The result is faster and less-dependent on the initialization in comparison to previous approaches.

Besides improving the SNR, 2D classification can be useful to remove contaminants and aggregated or denatured particles. Usually the input dataset is too heterogeneous. The degree of heterogeneity in a cluster can be analyzed using a great variety of procedures, for example via PCA of each cluster, obviously after removing the variability caused by image misalignment. Outliers can be identified through their Mahalanobis distance to the centroid [86,87] of the PCA subspace composed by the first few components. The Mahalanobis distance measures how many standard deviations away a point is from the mean of a distribution. Images close to the cluster centroid as measured by the Mahalanobis distance form the class core [86].

If our 2D clustering is hierarchical [88], the class core can be further refined by considering the subset of images that are basically classified together in the whole hierarchical process. Usually, outliers swap between several classes whereas the true projections tend to remain together in a stable behavior. This refined subset is called the stable core. To be more flexible, the implementation can relax this condition. In this way, the stable core is a subset of these particles that have remained together for all classification levels (within a certain level) of tolerance).

The previous methods are devoted to discrete classification; however, these approaches may not be well suited for macromolecules that exhibit continuous molecular motions. In this situation, several low-resolution maps showing different states of the molecule can guide the alignment and 2D classification of cryoEM images, for example [89].

3D analysis

The 3D reconstruction process can be seen as an optimization problem in which we need to move through a solution landscape where every point represents a 3D model. Each model has an associated energy that depends on the error between that model and the 2D experimental images. The aim of this process is to reach the optimal 3D model considering the information carried by the 2D images. This task is a major challenge in the field and significant effort has been applied by several researchers to develop algorithms to solve the problem.

The whole 3D reconstruction process is commonly managed by starting with an initial model, which can be seen as an estimation of the starting point in the solution landscape. The subsequent rounds of refinement steps move along the whole landscape, improving the reconstructed model in every step. The refinement algorithms can easily get stuck in local minima of the solution landscape [90]. Therefore, good design of the initial volume estimation and refinement algorithms are key to determining a final, accurate 3D structure.

Initial model

The goal of the initial model procedure is to create a low-resolution molecular density of the underlying structure, that can be further refined into a high-resolution map. This process is especially important for molecules with unknown structure, as using an incorrect initial model can lead to bias in the final map, or slow convergence of the refinement algorithm.

In recent years a plethora of initial model algorithms have appeared. Currently, there are a sufficiently high number of methods such that at least one of them will produce a suitable initial volume.

A family of these new algorithms is based on the Central Slice Theorem [91] which states that the Fourier transform of a 2D image belonging to a certain projection direction, corresponds to a slice of the 3D Fourier transform of the volume in the perpendicular direction. So, every pair of the 2D images coming from different projection directions will intersect at a common line in Fourier space Such common line methods [80,92–95] are based on this theorem. Wang et al. [92] described an algorithm based on synchronization to determine the direction of all the 2D images at once. Combining the common lines outcomes for pairs of images, yields a global assignment of orientations that maximizes the number of satisfied pairwise relations. The idea of synchronization was further studied in [94] where a graph-partitioning algorithm was suggested to consistently assign orientations, giving a confidence value to each one. A typical problem with these methods is that they are prone to detect false common lines. Wang et al. [93] proposed a method for dealing with this problem, in which the orientations were estimated by minimization of the sum of unsquared residuals, adding a spectral normalization term to avoid artificial clustering that appears with overlapping slices in the Fourier space. The algorithm proposed in [95] presented a way to model the errors in the estimated common lines by giving them a probability value. However, the main drawback of the common lines approaches has still not been overcome, as they still tend to easily fail when the detection rate of common lines is too low due to the low SNR in typical cryoEM 2D images [96].

Another approach to the generation of a starting model is to follow a statistical framework, for example, [82,97–99], in which the alignment parameters can be found by optimizing some related quantity. Elmlund et al. [97] presented a probabilistic initial 3D volume generation method called (PRIME), where each image is assigned to a range of orientations with the highest correlations. Then, the initial 3D model is generated by giving a weight to every image in every specific orientation proportional to the obtained correlation. The method in [82] is based on dimensional reduction of 2D class averages with the aim of obtaining representative sets of class images with the main structural information. Then, with the 2D representative image sets several initial models are generated. The best initial model can be determined using random sample consensus (RANSAC).

Joubert and Habeck [98] used was based on Bayesian inference approach in which a pseudo-atomic model is used to represent the 3D structure, whilst the estimation of the unknown 3D structure and image orientations is carried out with a maximum *a posteriori* optimization. However, it must be noted that a low number of pseudo-atoms in the pseudo-atomic model could generate inaccurate structural representations. The algorithm presented by Sorzano et al. [99] followed a maximum likelihood approach where the projection parameters are treated as hidden random variables and the goal is to find the volume that maximizes the likelihood of observing the experimental images (although normally this algorithm is applied to 2D class averages). The method results in a weighted least squares problem, in which the weights are given by both the experimental image and the projection direction. This method introduced an important idea in the field: not only can the experimental images vote during the construction of a model, by assigning a weight to each projection direction, but the projection directions can also vote and help in the decision of the weights of the experimental images.

The main drawbacks of statistical approaches are the following: the computational complexity is usually high due to the iterative framework, and, as they need some first estimation to iterate until deriving the definitive initial model, there is a tendency to easily finish in local minima. This is the problem with a solution landscape containing plenty of local minima — algorithms may get trapped in these less optimal solutions.

In 2018, a new approach to *ab initio* modeling was presented that does not require estimation of the viewing directions of projections. Assuming that the projection orientations are uniformly distributed across the sphere, Levin et al. [100] showed that a low-resolution estimate is achievable by using just two denoised projections. The authors use Kam's autocorrelation method and solve for the missing orthogonal matrices by using projection matching. There are a few limitations to this method, one being the assumption that viewing directions are distributed uniformly, as some molecules have preferred orientations. Nevertheless, this method may be a fresh, promising direction in model initialization research.

Finally, [101] a particle swarm optimization method was introduced that collects different initial volume proposals from other algorithms and considers them to be individuals of a population of initial volumes. Particle swarm optimization refers to allowing candidate solutions, called "particles", to traverse, or "swarm", the search space of solutions and approach the optimal solutions. This population is evolved using an algorithm combining stochastic gradient descent and particle swarm optimization. Ordinarily, the whole population converges to a single structure, which is usually a correct initial volume.

In many cases, it is not possible to build an initial model following the common lines approach applied to cryoEM images. In this situation, one may to use negatively stained samples and the random conical tilt (RCT) [66] or orthogonal tilt reconstruction (OTR) [67] procedures to obtain a low-to-medium resolution starting-model.

Although there is a wide range of possibilities to tackle the initial volume estimation, this is still an open problem,

but to a much lesser extent than it was about five years ago. Even so, more robust algorithms are still in need, since there are situations in which the existing algorithms fail to produce a satisfactory result.

Refinement and reconstruction

A key step in the cryoEM image processing pipeline is determination of a 3D reconstruction that is compatible with the 2D projection images and has a sufficiently high resolution to interpret molecular details in the structure. This is the problem that refinement and reconstruction methods try to solve, and the main challenge is that 2D projection images are contaminated by a huge amount of noise. Fortunately, this can often be obviated by simple collecting a very large number (sometimes millions) of images, and the averaging of many images by SPA can greatly reduce the noise level. The remaining challenges to achieving a high-resolution reconstruction are mainly limited by incomplete coverage of the viewing directions, limiting effects of the CTF, and execution time. We can find plenty of reconstruction methods, mainly organized in two families: direct Fourier inversion and iterative algorithms.

Direct Fourier inversion methods are based on the Central Slice Theorem using the method of common lines [91]. These approaches are well suited to handle a large number of projections, with a reasonable computational burden and high-accuracy when the angular coverage of the set of projections fully fills the 3D Fourier space and the SNR is sufficient to accurately determine the common lines. However, when we do not have a good angular coverage the outcomes generated by these methods may not yield optimal solutions. Abrishami et al. [102] dealt with the angular coverage problem by introducing a gridding-based direct Fourier method that used a weighting technique to compute a uniform sampled Fourier transform. This proposal followed the general idea of [34•] and added a weighting scheme for each projection that was estimated in an iterative way — evaluating a function similar to a kernel interpolator.

Another line of research has sought to incorporate *a priori* information in the 3D reconstruction process. Some iterative procedures have exploited sparse representation of the reconstructed volume. For instance, Moriya et al. [103] assumed a median root prior which favored locally monotonic reconstructions. Xu et al. [104] used an improved L^2 gradient flow method (L2GF) in which an energy functional consisting of a fidelity term and a regularization term was employed. For a review of iterative algorithms, the interested reader is referred to [105••]. The use of different reconstruction algorithms also depends on the user, because they might result in slightly but non-negligible differences, exemplified by the two reconstruction methods shown in Figure 3.

The main drawback of existing refinement and reconstruction methods is the difficulty of managing the projection images. There are a limited number of projection images available, which impedes the ability to correctly pose the inverse problem. Another drawback is the high computational cost, even when using highly optimized implementations on GPUs.

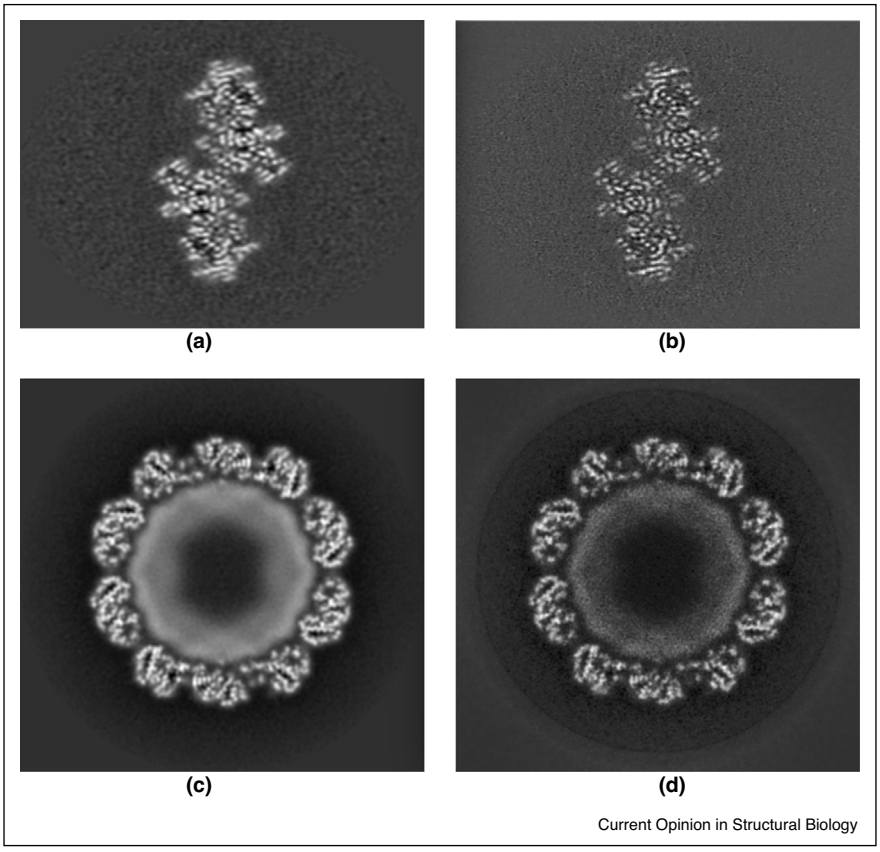
More general statistical methods are also gaining popularity. Dvornek et al. [106] proposed a novel speedup of the expectation—maximization algorithm. The idea behind the approach was to represent the 2D experimental images and the model projections in two low-dimensional subspaces. The matching between experimental and projections images was performed in the subspace bases. Because the number of basis elements is much smaller than the number of images and projections, substantial speedup was achieved. The main difference between this algorithm and that proposed in [34•] is that the latter is implemented in the Fourier domain whilst the subspace in [106] can be applied in Fourier or spatial domains. In [107] the stochastic gradient descent (SGD) and Bayesian marginalization algorithms were used to recover multiple 3D states of the molecule. The algorithm started with an arbitrary computer-generated random initialization that was incrementally refined with random selection of 2D images. The potential pitfall of this algorithm is that it relies on an arbitrary initial map reference and there may be bias towards the initial map. However, the SGD approach is supposed to help in this regard.

Molecular heterogeneity

Macromolecules can undergo conformational changes due to their functional states and interactions with other macromolecules and their environment. CryoEM can be performed with samples maintained in a physiologic environment so it is possible to visualize different molecular conformations, which poses a great challenge in the development of processing algorithms to analyze the molecular structures. In addition, samples may contain different oligometric states, disordered species, aggregates, molecular contaminants and areas with crystalline ice, which may confound image processing. Identifying and accounting for heterogeneity is an active field of research in cryoEM. Just focusing on different conformational states, we divide the approaches into four main families: physical, statistical, covariance analysis, and projection subtraction methods.

In the physical approaches, we can find a family of algorithms based on anisotropic network model (ANM), which is a direct application of the normal mode analysis, and molecular dynamics (MD) to predict the collective motions of structures and to describe full atomic molecular motions, respectively. Gur et al. [108] combined both with a Monte Carlo/metropolis scheme to randomly select the modes to deform the structure with the aim of generating trajectories between two conformational states. Costa et al. [109] used ANM and MD to

Figure 3



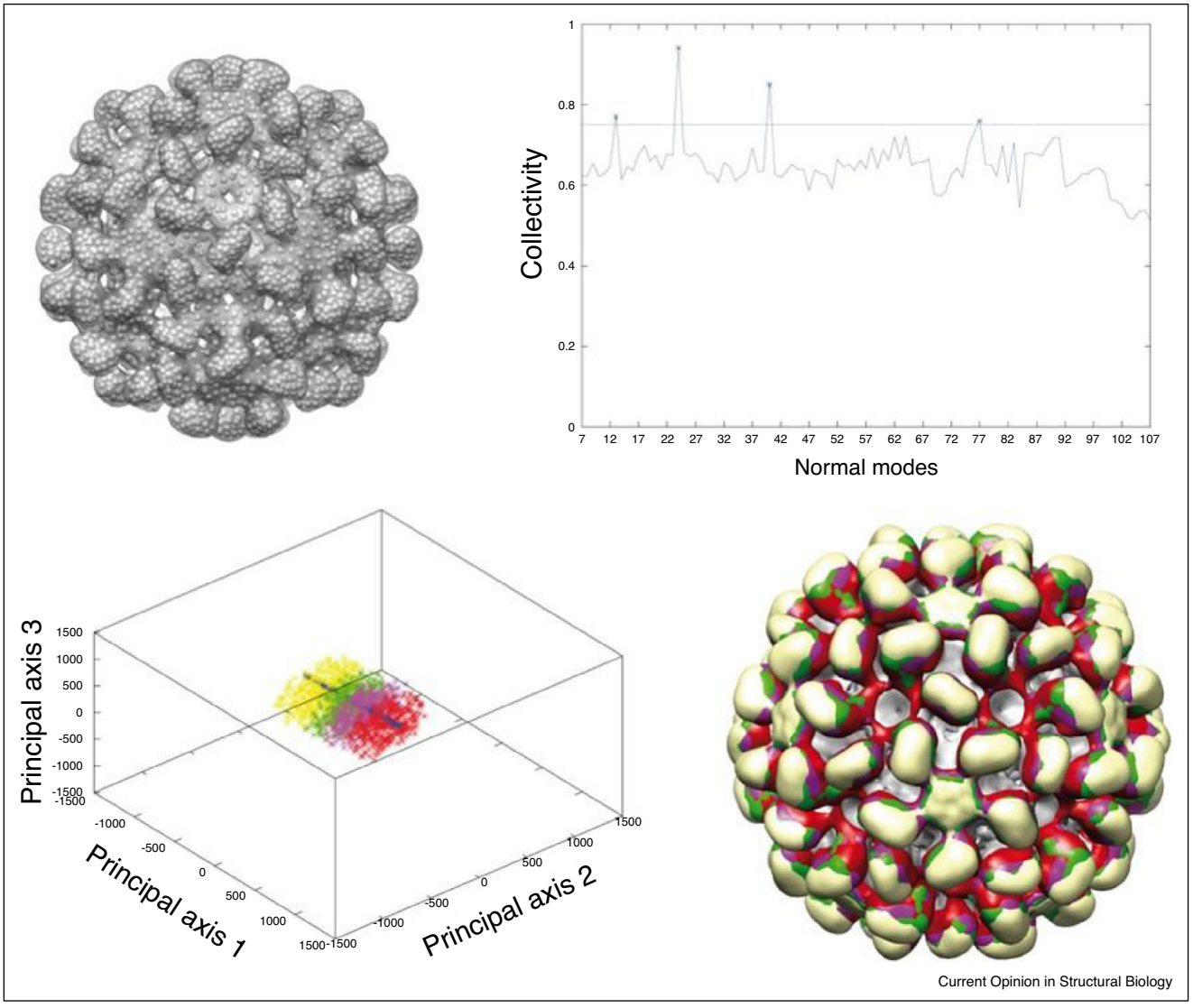
Examples of two reconstructed structures using RELION autorefine (*left*) and Xmipp highres (*right*). Despite the input date were the same, both algorithms cast different degree of detail keeping the same structure. The representative slices from 3D reconstruction of β-galactosidase (EMDB entry 10013) (*top*) and Brome mosaic virus (EMDB entry 10010) (*bottom*).

couple local and global motions efficiently. The method performed a large number of MD simulations, each corresponding to the excitation of a randomly determined linear combination of selected normal modes. Similarly, Kurkcuoglu et al. [110] used combinations of ANMs to calculate the conformational space for a molecule, and a clustering procedure was applied to construct representative substrates.

Among the statistical approaches is a method for sorting structural states found [111], which was based on bootstrapping of 3D sub-ensembles and 3D multivariate statistical analysis followed by 3D classification. Sorzano et al. [112] presented a method to analyze distances among elastically aligned pairs of EM models. Each experimental 3D model was transformed by elastic deformation and compared with other models in terms of structural and conformational differences. Punjani et al. [107], that was described in the previous section, was also developed to refine multiple high-resolution 3D models directly from single-particle images using SGD and Bayesian marginalization algorithms. Haselbach et al. [113] studied conformational variability by combining an iterative 3D classification approach with 3D principal component analysis (PCA). 3D classification gave hundreds of 3D structures, which were ordered according to their conformational similarities by applying PCA. Thus, this method identified motion patterns of flexible components in a conformational landscape. An example is shown in Figure 4.

A different approach to discover heterogeneity in cryoEM images consists of estimating the covariance of the reconstructed model. Katsevich et al. [114] proposed a new estimator in Fourier space that converges to the population covariance matrix as the number of images grows, but this method involves the inversion of a high-dimensional linear operator. Instead of inverting the original linear operator And'en et al. [115], proposed to use the conjugate gradient, achieving a lower computational complexity and the possibility of including the CTF correction. Liao et al. [116] estimated the whole covariance matrix, instead of only its main eigenvectors. Hence, this approach avoided the resampling problem and enabled the analysis of covariance in localized regions.

Figure 4



Top, left: 3D EM density map of the Tomato Bushy Stunt Virus and its pseudoatomic representation. *Top, right:* collectivity of the normal modes of the pseudoatomic representation. *Bottom, left:* projection of the deformation parameters estimated for experimental images onto a 3D principal component (PCA) space. Clustering of these projections into four classes. *Bottom, right:* The corresponding reconstructions of the four identified classes in the PCA space are shown; their isosurface representation is superposed using the same colors than the identified classes, exhibiting a conformational change.

The work described in [117] used fluctuation–dissipation theory for estimating a spring-and-mass mechanical model. Thus, this approach was able to transform the covariance matrix into a generative mechanical model of the complex.

The last family of methods to deal with structural heterogeneity is based on focusing the refinement process on the region where the motion is mostly taking place, masking out the fixed parts of the images. This procedure is usually named projection subtraction, and during 3D refinement it only accounts for those parts of the images where the structural variability can be found. Bai et al. [118] proposed to subtract projections of the fixed part of the molecule from every experimental image. In this way, the modified experimental image only contains the moving parts of the macromolecule. This procedure required knowledge of the relative orientation of each particle, which was obtained from a consensus refinement of the entire data set against a single, unmasked reference. A similar idea was published in [119], where a first 3D estimated model was separated into different modules according to prior knowledge. For every module, the orientation parameters were calculated by maximizing

the cross-correlation coefficient. However, this method assumed that the resolution of the initial 3D model was high enough to discriminate different modules. One of the main drawbacks of the projection subtraction approaches is that the moving element needs to be rigidly moving and of sufficient size so that the subtracted projections can be correctly aligned.

Despite all the progress in the analysis of heterogeneity, several difficulties remain. First, the 3D models need to be reconstructed from 2D images, making it difficult to connect the models reconstructed from thousands of 2D experimental images with the actual conformational state associated with a corresponding projection. Moreover, the noise problem must be highlighted, as 2D experimental images have a SNR well below 1 (which means that there is much more noise power than signal power). This problem poses a limit on the resolution that can be achieved in the 3D models reconstructed with SPA, making some conformational states indistinguishable.

Validation of results

The reconstruction workflow involves many steps in which the user decisions might determine the quality or even the validity of the cryoEM density map. The low SNR of cryoEM images complicates the reconstruction process. In particular, it can induce problems at critical steps, especially in the angular assignment of particles. The result may be maps of low quality, or worse, maps that are erroneous. One approach for map validations is to use corroborating data from other techniques such as X-ray scattering or NMR spectroscopy, or alternatively by using the experimental images that must be in agreement with the volume. A set of methods addressed to validate the map have been proposed.

1 *Overfitting detection*: Overfitting phenomena occurs particularly at high resolution. A reconstructed volume using noisy particles should stand out in the resolution of the map. One approach for validation is to generate a reconstruction derived from a data set in which a certain number of experimental particles have been replaced with noisy particles [120]. The goal is to analyze the resolution of the reconstructed volume before and after noise substitution. If both resolutions are consistent, then an there is an alignment problem.
2 *Tilt pair validation*: This was the first validation method developed by Henderson and Rosenthal [121–123] and requires a recording sample images at two different tilt angles. The geometry constraint introduced by the tilt angle and direction must be conserved when the particle's tilt pairs are aligned with the obtained volume, that is the angular relation between the untilted and tilted particle. The results of the angular alignment are simply plotted in a polar plot, in which radial measure represents the tilt angle and the angle shows the tilt direction. When the volume is in agreement with the angular alignment, the plot will exhibit a cluster. However, a high level of noise may introduce alignment errors which are manifested as scattered points in the polar plot. The existence of such clusters can be analyzed statistically. [124].
3 *Alignability validation*: These methods measure the alignability of the images used for reconstruction [125,126]. Leaving out issues of particle symmetry, each particle will be a map projection under one direction and it is expected that the most probable orientations for each particle form a cluster in the projection sphere. Additionally, if we make a *de novo* angular assignment, it is expected that the new angular assignment is consistent with the angular assignment used for the reconstruction. In contrast, pure noise images are expected to behave in the opposite way: the most probable directions are not clustered, and the *de novo* angular assignment does not coincide with the assigned angles.
4 *Atomic model validation*: Many structures elucidated by cryoEM were previously obtained by other techniques such as X-ray crystallography or NMR spectroscopy. In these cases, the atomic model is known. Then, the cryoEM density map must follow the atomic model at least at low-intermediate resolution.

Resolution

Once the macromolecular structure has been obtained and validated, it is necessary to report a quality measurement of the map, given by the resolution. There is no consensus about a universal definition of resolution, the most widespread being the size of the smallest reliable detail in the map. However, from an optical point of view, resolution has a clear definition as the ability of an imaging system to distinguish two separated points in an acquired image. The Rayleigh criterion can be considered as the standard in optics [127]. It should be highlighted that this definition implies that resolution is a property of the imaging system, instead of a property of the acquired image (i.e., the map in cryoEM). Nevertheless, when the imaging system is omitted and only the image is analyzed, other criteria are used, for example, Johnson criteria [128].

In cryoEM, the resolution has been traditionally analyzed in a global sense, that is, reporting a single parameter that summaries the quality of the map. For a comprehensive review of these resolution measures, the reader is referred to [129••]. The most used global resolution method is the Fourier shell correlation (FSC) in which correlation of two band-pass filtered independent reconstructions is measured. The resolution is defined as the central frequency of the band-pass filter at which the correlation drops below a given threshold. The problem with this measure is that it is a self-consistency measure of the

reconstruction process, rather than a quality measure of the reconstructed volume, e.g. it rewards systematic errors during the reconstruction process. To do that, the *Gold Standard* procedure is carried out. It consists in splitting the set of particles in two sets, and then performing two independent reconstructions [130,131]. This is a self-consistency measurement because both reconstructions should cast similar maps. If one of the reconstructions exhibits overfitting, it will not correlate with the other. Despite the gold standard, there is still some overfitting. In this regard, the phase-randomization method can be used to calculate the true FSC-resolution by noise substitution of particle phases beyond a certain frequency [132]. CryoEM images present low SNR and even particles of noise can be aligned, that is features of noise correlate with the reference [120,133,134], in particular at high frequencies. When many particles of noise are aligned, those poor features are reinforced and a model bias is introduced. This problem is called the *phantom in the noise* or *Einstein from noise*.

However, as the pioneers of the local resolution showed, one number does not fit all [135]. It has been shown that resolution is actually a tensor (it depends on the location within the volume and the direction) [129••], and the global resolution summarizes this rich information into a single number. The local quality differences have their origin in the reconstruction process. The SPA workflow considers that all particles (projections of the macromolecular complex) are identical and uniformly distributed on the projection sphere. Unfortunately, reality differs from this assumption because of heterogeneity and angular orientation. The heterogeneity has been identified as one of the main problems in cryoEM [136], and contradicts the assumption of SPA that all particles are identical copies of the same complex. Thus, we distinguish heterogeneity due to (1) the macromolecular complexes not being rigid and presenting a certain degree of flexibility, i.e. conformational heterogeneity; (2) despite the purification efforts some proteins present slight, but not negligible, structural heterogeneity. Radiation damage can also be responsible for this kind of heterogeneity. In any case the heterogeneous region of the macromolecule will be blurred. The angular assignment of particles is the second main source that induces local variations in the cryoEM density map. If the sample presents preferred directions or even lack of information in others, the distribution of angular assignments will be non-uniform, and will result in different regions of the map having different resolutions [137]. To overcome this problem of angular coverage, Tab et al. [138] showed that by tilting the sample the overall resolution can be increased and the quality map improves.

Blocres was the first method for estimating local resolution in cryoEM maps [135]. It extends the FSC measurement in a local sense. Thus, by means of two half maps and a moving window centered in the voxel of interest a local FSC can be calculated. The critical point is to set the window size. Logically, this is a self-consistency measurement, and it preserves all FSC properties. Interestingly, *Blocres* introduced the possibility of computing the locally filtered map at the local resolution values.

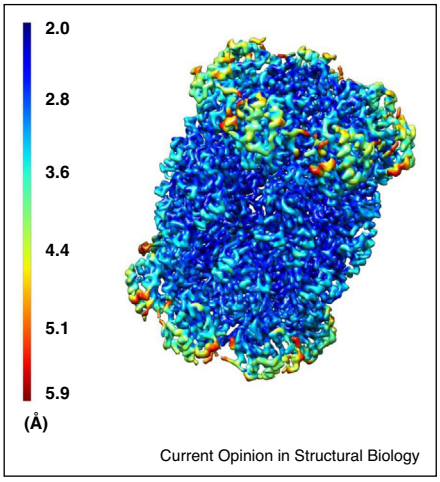
Nowadays, ResMap is a commonly used method for an estimate of local resolution [139]. Its rationale is the local detection of a sinusoidal signal above the noise level in a statistical sense. This task is carried out by use of a steerable function basis that models sinusoidal signals as linear combinations. Moreover, this method overcomes the drawback of using two half maps by computing local resolution maps using just a single volume or two half maps. In addition, it considers the spatial correlation in terms of resolution between closest voxels and computes a False Discovery Rate, that is in an hypothesis the expected value of the number of resolutions wrongly assigned over the total number of resolutions assigned correctly.

Recently, a new method called *MonoRes* for estimating local resolution has been described [140]. The idea of this method is to measure the local energy of the macromolecule and the energy distribution of the noise. The discrimination between noise and particle is provided by a mask. Thus, a frequency sweep performs hypothesis tests to determine if the energy of each voxel in the filtered map is significantly higher than the energy of noise at that frequency. This new method has the advantage of being fully automatic without user intervention, computationally faster than other approaches, and invariant under *b*-factor correction, and any other isotropic frequency correction. In addition, it also provides a locally filtered map at the local resolution values, as shown in Figure 5.

Fitting an atomic model

Thus far, we have discussed methods for building and refining a 3D reconstruction of the macromolecule being imaged. This reconstruction is in reality a map of the Coulomb potential. The ultimate interest in the research community is focused on an atomic level structural model of the macromolecule. Initially, fitting can be performed for secondary structure elements (SSEs) such as α-helices and β-sheets. Initial methods from the early 2000s focused on one particular SSE for search, but in more recent years, with SSELearner (2012) and the like, different SSE types can be resolved using just one method [141]. There are different approaches to fitting multiple SSEs. SSELearner uses a local structure tensor to characterize shape at density voxels. A support vector machine is trained with discriminatory tensors and known SSEs. This learning approach uses previously solved structures to solve similar unsolved molecular structures [142].

Figure 5



Local resolution map of the thermoplasma acidophilum 20S proteasome using the MonoRes method [140].

When fitting to 3D density maps, both rigid body fitting and flexible fitting mechanisms can be used. Rigid fitting is often used as a precursor to flexible fitting, which then makes allowances for conformational changes. These changes occur especially during interaction of the protein with other proteins. Another precursor to flexible fitting can be coarse graining. Coarse graining combines multiple atoms based on neighborhood arrangement into psuedoatoms that can be arranged into a low resolution model. (The reader is referred to the article on coarse-graining in this series by Pak and Voth.) This can save computational energy when modeling large molecules [143]. The coarse grained model can then be refined, like rigid fitting, with flexible fitting, which requires search of the solution space of possible conformations. Many methods use simulated annealing to find the best fit [144].

Best fit can be determined using a variety of metrics, the oldest being cross-correlation between the estimated structure and the density reconstruction. Different metrics have been proposed over the years, including surface area agreement with the density model, stereochemistry metrics considering atomic bonding and van der Waals forces, and others. Recent work has shown that a combined metric of local mutual information and amount of overlap with the density reconstruction performs better than cross-correlation alone [145]. It seems that along with validation methods for 3D reconstructions, evaluation of atomic models is a promising direction for cryoEM research.

Atomic model refinement is also a popular topic of current research which goes hand in hand with model evaluation. Current work improves fitting of amino acid sidechains by using multiple local optimization results instead of one global optimization result [146]. For model refinement, researchers have also analyzed physical properties that should be taken into consideration, such as partial charges on atoms [147].

Building an accurate atomic model is possible even without a reliable 3D density map. As noted in previous discussions, we know that molecules have certain preferred orientations within a grid. If the set of orientations only includes a few possible rotations, then 3D reconstruction through traditional methods is intractable. Traditionally in these situations, 2D class averages are compared to candidate models, which are represented by a graph of SSE components and amino acid side-chains [148]. Comparisons are performed based on similar metrics as when fitting to density maps. In 2015, electron atomic scattering factors (EASF) have been used to generate 3D EM volumes from atomic models. The EASF for each element represents the shape of atoms as seen by electrons in the electron beam, and is related to the elastic scattering of electrons. These EASF functions can be sampled to create an atomic model of a macromolecule, that can then be used with any of a number of popular software tools to generate a density map of the molecule [78].

Another exciting new direction for atomic model fitting is to find the pathway of conformational change. Matsumoto et al. generate various atomic models with different conformations, which are then deconstructed into their hypothetical prior 2D projections. The projections are compared to actual projection images, building a distribution of conformations from the best matches. From this distribution, the path of conformational changes that a protein undergoes can be estimated, which is important for understanding functional relationships [149].

Conclusions — current image processing challenges

Despite the recent impressive successes of cryoEM and SPA, this modality is still a very active research area, with advances in sample preparation methods [7], camera detection efficiency [7,136,150••], specimen stabilization under the beam [150••], better electron optics (energy filters, aberration corrections) [151••,152,153••], in-focus phase contrast [7], computational means to validate structures [7,136,154•], wider access to high-end microscopes [7,150••], and better training [7]. From the data analysis point of view, we would like to complement this list with the following considerations:

1 *Better BIM correction*: Specimen movement under the electron beam is a serious issue. There has been steady progress in this area, with proposals at the level of sample preparation [155,156], computational frame alignment [157•] and dose weighting [43•,158].

However, the best way to combine all these approaches is still unclear, and even some BIM effects, such as out-of-plane rocking along the beam direction, are not yet addressed by any method.

2 *Finer aberration corrections*: Microscope aberrations that have not been corrected by hardware must be estimated and corrected by software. Many attempts have been made to correct for spherical aberrations [159], magnification anisotropy [160], or local defocus changes [161], but their use is not widespread, probably indicating that improvements in the image processing workflow are required. Even such a basic task as focus determination is far from trivial and reliable for high resolution [162]. In addition, the weak-phase approximation is violated for large specimens, and at high resolution the Central Slice Theorem does not hold as an image formation model [14•,16,151••]. This implies that beyond a given resolution, reconstruction algorithms are not correctly handling frequency coordinates. Finally, the much anticipated introduction of phase plates as a way to avoid defocusing [163] poses additional challenges, since focus determination in these conditions is especially difficult.

3 *Handling homogeneity/heterogeneity and flexibility*: Particle flexibility and heterogeneity is at the same time a blessing and a curse of EM. On one side, flexibility reveals the dynamics of the macromolecule under study. On the other side, only homogeneous sets of particles can be reconstructed at atomic resolution. The compromise between a data set being as large as possible and as homogeneous as possible is still an open problem, particularly due to the low contrast and SNR of the acquired images. Significant advances in this regard have been made in recent years [57,164]. However, the issue is far from settled, particularly in those cases in which conformational changes correspond to a continuous distribution of states. This issue has been explored in some works [89,165], but this problem still needs further investigation. A particularly challenging situation occurs when studying a macromolecule of unknown structure. Indeed, most image classification algorithms are designed as local optimizers that start from a reasonably good initial map. If this map is not available, algorithms may easily derive nonsensical structures. There are specific initial volume algorithms to handle this issue [166]. However, currently, no algorithm has been designed specifically with flexibility/heterogeneity in mind.

4 *Integration with other information sources*: With very few exceptions [167], current reconstruction processes do not consider any source of information other than the projection images produced by the microscope. After a 3D map is obtained, modeling — especially the modeling of large macromolecular complexes — certainly benefits from other sources of information, such as cross-linking and mass spectroscopy [168] or protein–protein interaction data [169]. However, the explicit algorithmic incorporation of *a priori* information about the type of signals (macromolecular maps) being handled is missing in the field.

5 *Validation*: For the good and for the bad, data analysis always produces a model of the macromolecular structure. Unfortunately, due to the high level of noise and the high dimensionality of the optimization process, the chances of getting trapped in a local minimum are not negligible. There are two possible manifestations of a local minimum: (1) the overall shape of the structure is incorrect (despite the fact that its projections are compatible, to a certain degree, with the experimental images); (2) small details of the structure are incorrect (the algorithm has overfitted noise). The first problem can be alleviated if similar maps are obtained when starting from several initial models. However, automatic algorithms capable of detecting this situation are still needed [120,122,125,126]. The second case can be alleviated by independently processing two halves of the data [170]. But the field needs better data processing strategies that do not imply using only a half of the dataset at hand.

6 *Standardization*: Thanks to the success of cryoEM as an imaging technique, many engineering groups are getting involved in the global research effort and adding software for solving specific problems. In addition, we have the traditional software packages that cover the whole image processing pipeline (Relion [171], Eman [172], Xmipp [37•,173], Spider [38], Imagic [174], Frealign [175], etc.) and systems that integrate algorithms from multiple sources (Appion [42] and Scipion [41,176]). This ecosystem of software lacks a common standard for interchanging information. Although some attempts have been proposed at the level of metadata [177] and geometry [178••], they have not been widely adopted. In addition, the field lacks a mechanism to report the image processing steps carried out from the acquired movies to the final 3D reconstruction.

7 *Data management*: The number of solved structures is growing rapidly year by year. Thus, the structural biology community and in particular the EM-community is appreciating the importance of sharing this information. To achieve that, web services such as the EMDataBank (http://www.emdatabank.org) and Worldwide Protein Data Bank (wwPDB; http://wwpdb.org) have been especially valuable. Other databases such as EMPIAR (http://www.ebi.ac.uk/pdbe/emdb/empiar/) pursue raw data availability. For a good review on data management and databases in structural biology see [179].

Conflicts of interest statement

Nothing declared.

Acknowledgements

The authors would like to acknowledge support from: Spanish Ministry of Economy and Competitiveness through Grants BIO2013-44647-R,

BIO2016-76400-R (AEI/FEDER, UE),Comunidad Autonoma de Madrid through Grant: S2017/BMD-3817, Instituto de Salud Carlos III, PT13/0001/0009, PT17/0009/0010, and European Union (EU) and Horizon 2020 through grant West-Life (EINFRA-2015-1, Proposal: 675858), CORBEL (INFRADEV-1-2014-1, Proposal: 654248), Elixir - EXCELERATE (INFRADEV-3-2015, Proposal: 676559), iNEXT (INFRAIA-1-2014-2015, Proposal: 653706), EOSCpilot (INFRADEV-04-2016, Proposal: 739563). The authors acknowledge the support and the use of resources of Instruct, a Landmark ESFRI project.

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. • Merk A, Bartesaghi A, Banerjee S, Falconieri V, Rao P, Davis MI, Pragani R, Boxer MB, Earl LA, Milne JLS *et al.*: **Breaking cryo-EM resolution barriers to facilitate drug discovery**. *Cell* 2016, **165**:1698-1707.
Breakthrough experimental work showing the capabilities of electron microscopy to visualize small molecules and quasi-atomic resolution.

2. Bonomi M, Heller GT, Camilloni C, Vendruscolo M: **Principles of protein structural ensemble determination**. *Curr Opin Struct Biol* 2017, **42**:106-116.

3. Frank J: **Time-resolved cryo-electron microscopy: recent progress**. *J Struct Biol* 2017, **200**:303-306.

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

5. Rawson S, McPhillie M, Johnson R, Fishwick C, Muench S: **The potential use of single-particle electron microscopy as a tool for structure-based inhibitor design**. *Acta Crystrallogr D* 2017, **73**.

6. Egerton RF: **Control of radiation damage in the TEM**. *Ultramicroscopy* 2013, **127**:100-108.

7. Glaeser RM: **How good can cryo-EM become?** *Nat Methods* 2016, **13**:28-32.

8. Mishyna M, Volokh O, Danilova Y, Gerasimova N, Pechnikova E, Sokolova OS: **Effects of radiation damage in studies of protein-DNA complexes by cryo-EM**. *Micron* 2017, **96**:57-64.

9. McMullan G, Vinothkumar KR, Henderson R: **Thon rings from amorphous ice and implications of beam-induced Brownian motion in single particle electron cryo-microscopy**. *Ultramicroscopy* 2015, **158**:26-32.

10. Hettler S, Kano E, Dries M, Gerthsen D, Pfaffmann L, Bruns M, Beleggia M, Malac M: **Charging of carbon thin films in scanning and phase-plate transmission electron microscopy**. *Ultramicroscopy* 2018, **184**:252-266.

11. Li X, Zheng SQ, Egami K, Agard DA, Cheng Y: **Influence of electron dose rate on electron counting images recorded with the K2 camera**. *J Struct Biol* 2013, **184**:251-260.

12. Shigematsu H, Sigworth FJ: **Noise models and cryo-EM drift correction with a direct-electron camera**. *Ultramicroscopy* 2013, **131**:61-69.

13. Vulović M, Ravelli RBG, van Vliet LJ, Koster AJ, Lazić I, Lücken U, Rullgård H, Öktem O, Rieger B: **Image formation modeling in cryo-electron microscopy**. *J Struct Biol* 2013, **183**:19-32.

14. • Vulović M, Voortman LM, van Vliet LJ, Rieger B: **When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-EM**. *Ultramicroscopy* 2013, **136C**:61-66.
Limitations of the weak-phase object assumption in EM.

15. Hawkes PW: **The correction of electron lens aberrations**. *Ultramicroscopy* 2015, **156**:A1-A64.

16. Koeck PJB, Karshikoff A: **Limitations of the linear and the projection approximations in three-dimensional transmission electron microscopy of fully hydrated proteins**. *J Microsc* 2015, **259**:197-209.

17. Lobato I, Van Dyck D: **MULTEM: a new multislice program to perform accurate and fast electron diffraction and imaging simulations using graphics processing units with CUDA**. *Ultramicroscopy* 2015, **156**:9-17.

18. • Downing KH, Glaeser RM: **Estimating the effect of finite depth of field in single-particle cryo-EM**. *Ultramicroscopy* 2017, **184**:94-99.
The finite depth of field of the electron microscopy does not have a significant impact on standard analyses, but it may have on large objects.

19. Russo CJ, Henderson R: **Microscopic charge fluctuations cause minimal contrast loss in cryoEM**. *Ultramicroscopy* 2018, **187**:56-63.

20. Russo CJ, Henderson R: **Charge accumulation in electron cryomicroscopy**. *Ultramicroscopy* 2018, **187**:43-49.

21. • Danev R, Buijsse B, Khoshouei M, Plitzko JM, Baumeister W: **Volta potential phase plate for in-focus phase contrast transmission electron microscopy**. *Proc Natl Acad Sci USA* 2014, **111**:15635-15640.
The use of Volta Phase Plates have revolutionzed the contrast in the electron microscope.

22. Fan X, Zhao L, Liu C, Zhang J-C, Fan K, Yan X, Peng H-L, Lei J, Wang H-W: **Near-atomic resolution structure determination in over-focus with volta phase plate by Cs-corrected cryo-EM**. *Structure* 2017, **25**:1623-1630.

23. Koeck PJB: **Design of an electrostatic phase shifting device for biological transmission electron microscopy**. *Ultramicroscopy* 2018, **187**:107-112.

24. Koeck PJB: **Annular dark field transmission electron microscopy for protein structure determination**. *Ultramicroscopy* 2016, **161**:98-104.

25. Koeck PJB: **An aperture design for single side band imaging in the transmission electron microscope**. *Ultramicroscopy* 2017, **182**:81-84.

26. • Russo CJ, Henderson R: **Ewald sphere correction using a single side-band image processing algorithm**. *Ultramicroscopy* 2018, **187**:26-33.
How to overcome the limits of the central slice theorem for achieving high resolution.

27. Kuijper M, van Hoften G, Janssen B, Geurink R, De Carlo S, Vos M, van Duinen G, van Haeringen B, Storms M: **FEI's direct electron detector developments: embarking on a revolution in cryo-TEM**. *J Struct Biol* 2015, **192**:179-187.

28. •• Faruqi AR, McMullan G: **Direct imaging detectors for electron microscopy**. *Nucl Instrum Methods Phys Res A* 2018, **878**:180-190.
Excellent review on direct detectors, the device that has revolutionzed the quality of electron microscopy images.

29. Chari A, Haselbach D, Kirves J-M, Ohmer J, Paknia E, Fischer N, Ganichkin O, Möller V, Frye JJ, Petzold G *et al.*: **Proteoplex: stability optimization of macromolecular complexes by sparse-matrix screening of chemical space**. *Nat Methods* 2015, **12**:859-865.

30. Razinkov I, Dandey VP, Wei H, Zhang Z, Melnekoff D, Rice WJ, Wigge C, Potter CS, Carragher B: **A new method for vitrifying samples for cryoEM**. *J Struct Biol* 2016, **195**:190-198.

31. Arnold SA, Albiez S, Bieri A, Syntychaki A, Adaixo R, McLeod RA, Goldie KN, Stahlberg H, Braun T: **Blotting-free and lossless cryo-electron microscopy grid preparation from nanoliter-sized protein samples and single-cell extracts**. *J Struct Biol* 2017, **197**:220-226.

32. Feng X, Fu Z, Kaledhonkar S, Jia Y, Shah B, Jin A, Liu Z, Sun M, Chen B, Grassucci RA *et al.*: **A fast and effective microfluidic spraying-plunging method for high-resolution single-particle cryo-EM**. *Structure* 2017, **25**:663-670 e3.

33. Dandey VP, Wei H, Zhang Z, Tan YZ, Acharya P, Eng ET, Rice WJ, Kahn PA, Potter CS, Carragher B: **Spotiton: new features and applications**. *J Struct Biol* 2018, **202**:161-169.

34. • Scheres SHW: **RELION: implementation of a Bayesian approach to cryo-EM structure determination**. *J Struct Biol* 2012, **180**:519-530.
Current reference algorithm for 3D classification and reconstruction.

35. Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ: **Eman2: an extensible image processing suite for electron microscopy**. *J Struct Biol* 2007, **157**:38-46.

36. Grant T, Alexis R, Grigorieff N: **cisTEM, user-friendly software for single-particle image processing**. *eLife* 2018, **7**.

37. • de la Rosa-Trevín JM, Otón J, Marabini R, Zaldívar A, Vargas J, Carazo JM, Sorzano COS: **Xmipp 3.0: an improved software suite for image processing in electron microscopy**. *J Struct Biol* 2013, **184**:321-328.
An integrative software platform with over 200 image processing protocols from over 20 different software packages.

38. Shaikh TR, Gao H, Baxter W, Asturias FJ, Boisset N, Leith A, Frank J: **Spider image processing for single-particle reconstruction of biological macromolecules from electron micrographs**. *Nat Protoc* 2008, **3**:1941-1974.

39. Hohn M, Tang G, Goodyear G, Baldwin PR, Huang Z, Penczek PA, Yang C, Glaeser RM, Adams PD, Ludtke SJ: **SPARX, a new environment for cryo-EM image processing**. *J Struct Biol* 2007, **157**:47-55.

40. Heymann B, Belnap D: **Bsoft: image processing and molecular modeling for electron microscopy**. *J Struct Biol* 2007, **157**:3-18.

41. de la Rosa-Trevín JM, Quintana A, Del Cano L, Zaldívar A, Foche I, Gutiérrez J, Gómez-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.

42. Lander GC, Stagg SM, Voss NR, Cheng A, Fellmann D, Pulokas J, Yoshioka C, Irving C, Mulder A, Lau PW *et al.*: **Appion: an integrated, database-driven pipeline to facilitate EM image processing**. *J Struct Biol* 2009, **166**:95-102.

43. • Grant T, Grigorieff N: **Measuring the optimal exposure for single particle cryo-EM using a 2.6Å reconstruction of rotavirus VP6**. *elife* 2015, **4**.
Dose weighting of the movie frames is crucial to preserve high quality information in the micrographs.

44. Ripstein ZA, Rubinstein JL: **Processing of cryo-EM movie data**. *Methods Enzymol* 2016, **579**:103-124.

45. 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.

46. Rubinstein JL, Brubaker MA: **Alignment of cryo-EM movies of individual particles by optimization of image translations**. *J Struct Biol* 2015, **192**:188-195.

47. McLeod RA, Kowal J, Ringler P, Stahlberg H: **Robust image alignment for cryogenic transmission electron microscopy**. *J Struct Biol* 2017, **197**:279-293.

48. Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y, Agard DA: **Motioncor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy**. *Nat Methods* 2017, **14**:331-332.

49. Afanasyev P, Ravelli RBG, Matadeen R, De Carlo S, van Duinen G, Alewijnse B, Peters PJ, Abrahams J-P, Portugal RV, Schatz M, van Heel M: **A posteriori correction of camera characteristics from large image data sets**. *Sci Rep* 2015, **5**:10317.

50. Sorzano COS, Fernández-Giménez E, Peredo-Robinson V, Vargas J, Majtner T, Caffarena G, Otón J, Vilas JL, de la Rosa-Trevín JM, Melero R *et al.*: **Blind estimation of DED camera gain in electron microscopy**. *J Struct Biol* 2018, **203**:90-93.

51. Vargas J, Otón J, Marabini R, Jonic S, de la Rosa-Trevín JM, Carazo JM, Sorzano COS: **FASTDEF: fast defocus and astigmatism estimation for high-throughput transmission electron microscopy**. *J Struct Biol* 2013, **181**:136-148.

52. Rohou A, Grigorieff N: **CTFFIND4: fast and accurate defocus estimation from electron micrographs**. *J Struct Biol* 2015, **192**:216-221.

53. Zhang K: **Gctf: real-time CTF determination and correction**. *J Struct Biol* 2016, **193**:1-12.

55. Danev R, Tegunov D, Baumeister W: **Using the volta phase plate with defocus for cryo-EM single particle analysis**. *eLife* 2017, **6**.

56. Scheres SHW: **Semi-automated selection of cryo-EM particles in RELION-1.3**. *J Struct Biol* 2015, **189**.

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

58. Hoang TV, Cavin X, Schultz P, Ritchie DW: **gempicker: a highly parallel GPU-accelerated particle picking tool for cryo-electron microscopy**. *BMC Struct Biol* 2013, **13**:25.

59. Langlois R, Pallesen J, Ash JT, Nam Ho D, Rubinstein JL, Frank J: **Automated particle picking for low-contrast macromolecules in cryo-electron microscopy**. *J Struct Biol* 2014, **186**:1-7.

60. Zhang, JK. Gautomatch, http://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/, accessed: 2018-05-16.

61. Abrishami V, Zaldívar-Peraza A, de la Rosa-Trevín JM, Vargas J, Otón J, Marabini R, Shkolnisky Y, Carazo JM, Sorzano COS: **A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs**. *Bioinformatics* 2013, **29**:2460-2468.

62. Vargas J, Abrishami V, Marabini R, de la Rosa-Trevín JM, Zaldivar A, Carazo JM, Sorzano COS: **Particle quality assessment and sorting for automatic and semiautomatic particle-picking techniques**. *J Struct Biol* 2013, **183**:342-353.

63. Berndsen Z, Bowman C, Jang H, Ward AB: **EMHP: an accurate automated hole masking algorithm for single-particle cryo-EM image processing**. *Bioinformatics* 2017, **33**:3824-3826.

64. Wang F, Gong H, Liu G, Li M, Yan C, Xia T, Li X, Zeng J: **Deeppicker: a deep learning approach for fully automated particle picking in cryo-EM**. *J Struct Biol* 2016, **195**:325-336.

65. Zhu Y, Ouyang Q, Mao Y: **A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy**. *BMC Bioinform* 2017, **18**.

66. Radermacher M, Hoppe W: **Properties of 3-D reconstruction from projections by conical tilting compared to single axis tilting**. *Proc. Seventh European Congress on Electron Microscopy I* 1980:132-133. URL: https://ci.nii.ac.jp/naid/10017254264/en/.

67. Leschziner A: **The orthogonal tilt reconstruction method**. In *Cryo-EM, Part B: 3-D Reconstruction, Vol. 482 of Methods in Enzymology*. Edited by Jensen GJ. Academic Press; 2010:237-262 http://dx.doi.org/10.1016/S0076-6879(10)82010-5. Chapter nine, URL: http://www.sciencedirect.com/science/article/pii/S0076687910820105.

68. Shatsky M, Arbelaez P, Han BG, Typke D, Brenner SE, Malik J, Glaeser RM: **Automated particle correspondence and accurate tilt-axis detection in tilted-image pairs**. *J Struct Biol* 2014, **187**:66-75.

69. Hauer F, Gerle C, Kirves J-M, Stark H: **Automated correlation of single particle tilt pairs for random conical tilt and orthogonal tilt reconstructions**. *J Struct Biol* 2013, **181(2)**:149-154.

70. Vilas JL, Navas J, Gómez-Blanco J, de la Rosa-Trevín JM, Melero R, Peschiera I, Ferlenghi I, Cuenca J, Marabini R, Carazo JM *et al.*: **Fast and automatic identification of particle tilt pairs based on Delaunay triangulation**. *J Struct Biol* 2016, **196**:525-533.

71. Benvenuto F, Camera AL, Theys C, Ferrari A, Lantéri H, Bertero M: **The study of an iterative method for the reconstruction of images corrupted by Poisson and Gaussian noise**. *Inverse Probl* 2008, **24**:035016.

72. Snyder DL, Hammoud AM, White RL: **Image recovery from data acquired with a charge-coupled-device camera**. *J Opt Soc Am A* 1993, **10**:1014-1023.

73. Nichols TE, Qi J, Asma E, Leahy RM: **Spatiotemporal reconstruction of list-mode PET data**. *IEEE Trans Med Imaging* 2002, **21**:396-404.

74. Delpretti S, Luisier F, Ramani S, Blu T, Unser M: **Multiframe sure-let denoising of timelapse fluorescence microscopy images**. *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2008*; *IEEE: 2008:149-152*.

75. •• Penczek PA: **Image restoration in cryo-electron microscopy**. *Methods Enzymol* 2010, **482**:35-72.
Excellent review of image restoration in electron microscopy.

76. Bhamre T, Zhang T, Singer A: **Denoising and covariance estimation of single particle cryo-EM images**. *J Struct Biol* 2016, **195**:72-81.

77. Sorzano COS, Jonic S, Nú nez Ramírez R, Boisset N, Carazo JM: **Fast, robust and accurate determination of transmission electron microscopy contrast transfer function**. *J Struct Biol* 2007, **160**:249-262.

78. Sorzano COS, Vargas J, Otón J, Abrishami V, de la Rosa Trevín JM, del Riego S, Fernández-Alderete A, Martínez-Rey C, Marabini R, Carazo JM: **Fast and accurate conversion of atomic models into electron density maps**. *AIMS Biophys* 2015, **2**:8-20.

79. Bajić B, Lindblad J, Sladoje N: **Blind restoration of images degraded with mixed Poisson–Gaussian noise with application in transmission electron microscopy**. *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*; *IEEE: 2016:123-127*.

80. Van Heel M: **Angular reconstitution: A posteriori assignment of projection directions for 3d reconstruction**. *Ultramicroscopy* 1987, **21**:111-123.

81. Jaitly N, Brubaker MA, Rubinstein RH, snf Lilien JL: **A Bayesian method for 3-D macromolecular structure inference using class average images from single particle electron microscopy**. *Bioinformatics* 2010, **26**:2406-2415.

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

83. Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A: **Spider and web: processing and visualization of images in 3D electron microscopy and related fields**. *J Struct Biol* 1996, **116**:190-199.

84. Yang Z, Fang J, Chittuluru J, Asturias FJ, Penczek PA: **Iterative stable alignment and clustering of 2D transmission electron microscope images**. *Structure* 2012, **20**:237-247.

85. 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.

86. Sorzano COS, Vargas J, de la Rosa-Trevín JM, Zaldívar-Peraza A, Otón J, Abrishami V, Foche I, Marabini R, Caffarena G, Carazo JM: **Outlier detection for single particle analysis in electron microscopy**. *Proc. Intl. Work-Conference on Bioinformatics and Biomedical Engineering*; *IWBBIO: 2014:950*.

87. Bhamre T, Zhao Z, Singer A: **Mahalanobis distance for class averaging of cryo-EM images**. *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*; *IEEE: 2017:654-658*.

88. 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.

89. Jin Q, Sorzano COS, de la Rosa-Trevín JM, Bilbao-Castro JR, Nú nez Ramírez R, Llorca O, Tama F, Jonić S: **Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes**. *Structure* 2014, **22**:496-506.

90. Sorzano COS, Marabini R, Pascual-Montano A, Scheres SHW, Carazo JM: **Optimization problems in electron microscopy of single particles**. *Ann Oper Res* 2006, **148**:133-165.

91. Zhao S, Halling H: **A new Fourier method for fan beam reconstruction**. *1995 IEEE Nuclear Science Symposium and Medical Imaging Conference Record*; *IEEE: 1995:1287-1291*.

92. Shkolnisky Y, Singer A: **Viewing direction estimation in cryo-EM using synchronization**. *SIAM J Imag Sci* 2012, **5(3)**:1088-1110.

93. Wang L, Singer A, Wen Z: **Orientation determination of cryo-EM images using least unsquared deviations**. *SIAM J Imag Sci* 2013, **6**:2450-2483.

94. Pragier G, Greenberg I, Cheng X, Shkolnisky Y: **A graph partitioning approach to simultaneous angular reconstitution**. *IEEE Trans Comput Imag* 2016, **2**:323-334.

95. Greenberg I, Shkolnisky Y: **Common lines modeling for reference free ab-initio reconstruction in cryo-EM**. *J Struct Biol* 2017, **200**:106-117.

96. Kervrann C, Sorzano COS, Acton ST, Olivo-Marin JC, Unser M: **A guided tour of selected image processing and analysis methods for fluorescence and electron microscopy**. *IEEE J Select Top Signal Process* 2016, **10**:6-30.

97. Elmlund H, Elmlund D, Bengio S: **PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy**. *Structure* 2013, **21**:1299-1306.

98. Joubert P, Habeck M: **Bayesian inference of initial models in cryo-electron microscopy using pseudo-atoms**. *Biophys J* 2015, **108**:1165-1175.

99. Sorzano COS, Vargas J, de la Rosa-Trevín JM, Otón J, Álvarez-Cabrera AL, Abrishami V, Sesmero E, Marabini R, Carazo JM: **A statistical approach to the initial volume problem in single particle analysis by electron microscopy**. *J Struct Biol* 2015, **189**:213-219.

100. Levin E, Bendory T, Boumal N, Kileel J, Singer A: **3D *ab initio* modeling in cryo-EM by autocorrelation analysis**. *2018 IEEE 15th International Symposium on Biomedical Imaging* 2018:1569-1573.

101. Sorzano C, Vargas J, Vilas J, Jiménez-Moreno A, Mota J, Majtner T, Maluenda D, Martínez M, Sánchez-García R, Segura J, Otón J *et al.*: **Swarm optimization as a consensus technique for electron microscopy initial volume**. *Appl Anal Optim* 2018, **2**:299-313.

102. Abrishami V, Bilbao-Castro JR, Vargas J, Marabini R, Carazo JM, Sorzano COS: **A fast iterative convolution weighting approach for gridding-based direct Fourier three-dimensional reconstruction with correction for the contrast transfer function**. *Ultramicroscopy* 2015, **157**:79-87.

103. 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.

104. Xu G, Li M, Chen C: **A multi-scale geometric flow method for molecular structure reconstruction**. *Comput Sci Discov* 2015, **8**:014002.

105. •• Sorzano COS, Vargas J, Otón J, Vilas JL, Kazemi M, Melero R, del Ca no L, Cuenca J, Conesa P, Gómez-Blanco J, Marabini R, Carazo JM: **A survey of the use of iterative reconstruction algorithms in electron microscopy**. *Biomed Res Int* 2017, **2017**:6482567.
Excellent review of iterative reconstruction algorithms in electron microscopy.

106. Dvornek NC, Sigworth FJ, Tagare HD: **Subspaceem: A fast maximum-a-posteriori algorithm for cryo-EM single particle reconstruction**. *J Struct Biol* 2015, **190**:200-214.

107. Punjani A, Rubinstein J, Fleet DJ, Brubaker MA: **cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination**. *Nat Methods* 2017, **14**:290-296.

108. Gur M, Madura JD, Bahar I: **Global transitions of proteins explored by a multiscale hybrid methodology: application to adenylate kinase**. *Biophys J* 2013, **105**:1643-1652.

109. Costa MGS, Batista PR, Bisch PM, Perahia D: **Exploring free energy landscapes of large conformational changes:**

molecular dynamics with excited normal modes. *J Chem Theory Comput* 2015, **11**:2755-2767.

110. Kurkcuoglu Z, Bahar I, Doruker P: **Clustenm: ENM-based sampling of essential conformational space at full atomic resolution**. *J Chem Theory Comput* 2016, **12**:4549-4562.

111. Klaholz BP: **Structure sorting of multiple macromolecular states in heterogeneous cryo-EM samples by 3D multivariate statistical analysis**. *Open J Stat* 2015, **5**:820-836.

112. Sorzano COS, Alvarez-Cabrera AL, Kazemi M, Carazo JM, Jonić S: **Structmap: elastic distance analysis of electron microscopy maps for studying conformational changes**. *Biophys J* 2016, **110**:1753-1765.

113. Haselbach D, Komarov I, Agafonov DE, Hartmuth K, Graf B, Dybkov O, Urlaub H, Kastner B, Luhrmann R, Stark H: **Structure and conformational dynamics of the human spliceosomal bact complex**. *Cell* 2018, **172**:454-464 e11.

114. Katsevich E, Katsevich A, Singer A: **Covariance matrix estimation for the cryo-EM heterogeneity problem**. *SIAM J Imag Sci* 2015, **8**:126-185.

115. Andén J, Katsevich E, Singer A: **Covariance estimation using conjugate gradient for 3D classification in cryo-EM**. *IEEE 12th International Symposium on Biomedical Imaging*; *IEEE: 2015:200-204*.

116. 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.

117. Gong Y, Doerschuk PC: **3-D understanding of electron microscopy images of nano bio objects by computing generative mechanical models**. *2016 IEEE International Conference on Image Processing (ICIP)*; *IEEE: 2016:3161-3165*.

118. Bai XC, Rajendra E, Yang G, Shi Y, Scheres SH: **Sampling the conformational space of the catalytic subunit of human γ-secretase**. *eLife* 2015, **4**.

119. Shan H, Wang Z, Zhang F, Xiong Y, Yin C-C, Sun F: **A local-optimization refinement algorithm in single particle analysis for macromolecular complex with multiple rigid modules**. *Protein Cell* 2016, **7**:46-62.

120. Heymann B: **Validation of 3DEM reconstructions: the phantom in the noise**. *AIMS Biophys* 2015, **2**:21-35.

121. Rosenthal PB, Henderson R: **Optimal determination of particle orientation, absolute hand, and contrast loss in single particle electron-cryomicroscopy**. *J Mol Biol* 2003, **333**:721-745.

122. 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.

123. Wasilewski S, Rosenthal PB: **Web server for tilt-pair validation of single particle maps from electron cryomicroscopy**. *J Struct Biol* 2014, **186**:122-131.

124. Russo CJ, Passmore LA: **Robust evaluation of 3D electron cryomicroscopy data using tilt-pairs**. *J Struct Biol* 2014, **187**:112-118.

125. 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.

126. Vargas J, Melero R, Gómez-Blanco J, Carazo JM, Sorzano COS: **Quantitative analysis of 3D alignment quality: its impact on soft-validation, particle pruning and homogeneity analysis**. *Sci Rep* 2017, **7**:6307.

127. Born M, Wolf E: *Principles of Optics*. 7th edition. Cambridge Univ. Press; 1999.

128. Johnson J: **Analysis of image forming systems**. *Image Intens Symp* 1958:244-273.

129. •• Sorzano COS, Vargas J, Otón J, Abrishami V, de la Rosa-Trevín JM, Gómez-Blanco J, Vilas JL, Marabini R, Carazo JM: **A review of resolution measures and related aspects in 3D electron microscopy**. *Prog Biophys Mol Biol* 2017, **124**:1-30.
Excellent review on resolution measures for electron microscopy and signal to noise ratio.

130. Grigorieff N: **Resolution measurement in structures derived from single particles**. *Acta Crystrallogr D* 2000, **56**:1270-1277.

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

132. Chen S, McMullan G, Faruqi AR, Murshudov GN, Short JM, Scheres SHW, 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.

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

134. van Heel M: **Finding trimeric HIV-1 envelope glycoproteins in random noise**. *Proc Natl Acad Sci USA* 2013, **110**:E4175-E4177.

135. Cardone G, Heymann JB, Steven AC: **One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions**. *J Struct Biol* 2013, **184**:226-236.

136. Nogales E: **The development of cryo-EM into a mainstream structural biology technique**. *Nat Methods* 2016, **13**:24-27.

137. Naydenova K, Russo CJ: **Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy**. *Nat Commun* 2017, **8**:629.

138. Tan YZ, Baldwin PR, Davis JH, Williamson JR, Potter CS, Carragher B, Lyumkis D: **Addressing preferred specimen orientation in single-particle cryo-EM through tilting**. *Nat Methods* 2017, **14**:793-796.

139. Kucukelbir A, Sigworth FJ, Tagare HD: **Quantifying the local resolution of cryo-EM density maps**. *Nat Methods* 2014, **11**:63-65.

140. Vilas JL, Gómez-Blanco J, Conesa P, Melero R, de la Rosa Trevín JM, Otón J, Cuenca J, Marabini R, Carazo JM, Vargas J, Sorzano COS: **MonoRes: automatic and accurate estimation of local resolution for electron microscopy maps**. *Structure* 2018, **26**:337-344.

141. Esquivel-Rodríguez J, Kihara D: **Computational methods for constructing protein structure models from 3D electron microscopy maps**. *J Struct Biol* 2013, **184**:93-102.

142. Si D, Ji S, Nasr KA, He J: **A machine learning approach for the identification of protein secondary structure elements from electron cryo-microscopy density maps**. *Biopolymers* 2012, **97**:698-708.

143. Chen YL, Habeck M: **Data-driven coarse graining of large biomolecular structures**. *PLoS One* 2017, **12**:1-17.

144. Chapman MS, Trzynka A, Chapman BK: **Atomic modeling of *cryo*-electron microscopy reconstructions – joint refinement of model and imaging parameters**. *J Struct Biol* 2013, **182**:10-21.

145. Joseph AP, Lagerstedt I, Patwardhan A, Topf M, Winn M: **Improved metrics for comparing structures of macromolecular assemblies determined by 3D electron-microscopy**. *J Struct Biol* 2017, **199**:12-26.

146. Wang RY, Song Y, Barad BA, Cheng Y, Fraser JS, DiMaio F: **Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta**. *eLife* 2016, **5**:1-22.

147. Wang J: **On contribution of known atomic partial charges of protein backbone in electrostatic potential density maps**. *Protein Sci* 2017, **26**:1098-1104.

148. Velázquez-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**. *PNAS* 2012, **109**:18821-18826.

149. Matsumoto A, Miyazaki N, Takagi J, Iwasaki K: **2D hybrid analysis: approach for building three-dimensional atomic model by electron microscopy image matching**. *Sci Rep* 2017, **7**:1-12.

150. •• Vinothkumar KR, Henderson R: **Single particle electron cryomicroscopy: trends, issues and future perspective**. *Q Rev Biophys* 2016, **49**:e13 1-25.
Excellent review on current experimental limitations of electron microscopy and challenges ahead.

151. •• Zhang X, Zhou ZH: **Limiting factors in atomic resolution cryo electron microscopy: no simple tricks**. *J Struct Biol* 2011, **175**:253-263.
Excellent review of the theoretical limitations of electron microscopy.

152. Yang JC, Small MW, Grieshaber RV, Nuzzo RG: **Recent developments and applications of electron microscopy to heterogeneous catalysis**. *Chem Soc Rev* 2012, **41**:8179-8194.

153. •• Schröder RR: **Advances in electron microscopy: a qualitative view of instrumentation development for macromolecular imaging and tomography**. *Arch Biochem Biophys* 2015, **581**:25-38.
Excellent review of the current state of the electron microscope as an optic device.

154. • Henderson R, Sali A, Baker ML, Carragher B, Devkota B, Downing KH, Egelman EH, Feng Z, Frank J, Grigorieff N *et al.*: **Outcome of the first electron microscopy validation task force meeting**. *Structure* 2012, **20**:205-214.
Recommendations for structure validation.

155. Passmore L, Russo CJ: **Specimen preparation for high-resolution cryo-EM**. *Methods in Enzymology. The Resolution Revolution: Recent Advances in cryoEM*. Academic Press; 2016:51-86.

156. Thompson RF, Walker M, Siebert CA, Muench SP, Ranson NA: **An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology**. *Methods (San Diego, CA)* 2016, **100**:3-15.

157. • Abrishami V, Vargas J, Li X, Cheng Y, Marabini R, Sorzano COS, Carazo JM: **Alignment of direct detection device micrographs using a robust optical flow approach**. *J Struct Biol* 2015, **189**:163-176.
Review on sample preparation for electron microscopy.

158. Spear JM, Noble AJ, Xie Q, Sousa DR, Chapman MS, Stagg SM: **The influence of frame alignment with dose compensation on the quality of single particle reconstructions**. *J Struct Biol* 2015, **192(2)**:196-203.

159. Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV, Stark H: **Structure of the *E. coli* ribosome-EF-Tu complex at 3a resolution by Cs-corrected cryo-EM**. *Nature* 2015.

160. Grant T, Grigorieff N: **Automatic estimation and correction of anisotropic magnification distortion in electron microscopes**. *J Struct Biol* 2015, **192**:204-208.

161. Voortman LM, Franken EM, van Vliet LJ, Rieger B: **Fast, spatially varying CTF correction in TEM**. *Ultramicroscopy* 2012, **118**:26-34.

162. Marabini R, Carragher B, Chen S, Chen J, Cheng A, Downing KH, Frank J, Grassucci RA, Bernard Heymann J, Jiang W *et al.*: **CTF challenge: result summary**. *J Struct Biol* 2015, **190**:348-359.

163. Danev R, Baumeister W: **Cryo-EM single particle analysis with the volta phase plate**. *eLife* 2016, **5**:e13046.

164. Punjani A, Brubaker MA, Fleet DJ: **Building proteins in a day: efficient 3D molecular structure estimation with electron cryomicroscopy**. *IEEE Trans Pattern Anal Mach Learn* 2017, **39**:706-718.

165. 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 USA* 2014, **111**:17492-17497.

166. Sorzano COS, Vargas J, de la Rosa-Trevín JM, Jiménez-Moreno A, Melero R, Martínez M, Conesa P, Vilas JL, Marabini R, Carazo JM: **High-resolution reconstruction of single particles by electron microscopy**. *J Struct Biol* 2018, **204**:329-337.

167. Velázquez-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 USA* 2012, **109**:18821-18826.

168. Politis A, Stengel F, Hall Z, Hernández H, Leitner A, Walzthoeni T, Robinson CV, Aebersold R: **A mass spectrometry-based hybrid method for structural modeling of protein complexes**. *Nat Methods* 2014, **11**:403-406.

169. Segura J, Sanchez-Garcia R, Tabas-Madrid D, Cuenca-Alba J, Sorzano COS, Carazo JM: **3DIANA: 3D domain interaction analysis: a toolbox for quaternary structure modeling**. *Biophys J* 2016, **110**:766-775.

170. Scheres SHW, Chen S: **Prevention of overfitting in cryo-EM structure determination**. *Nat Methods* 2012, **9**:853-854.

171. Scheres SHW: **Processing of structurally heterogeneous cryo-EM data in RELION**. *Methods in Enzymology. The Resolution Revolution: Recent Advances in cryoEM*. Academic Press; 2016:125-157.

172. Ludtke SJ: **3-D structures of macromolecules using single-particle analysis in Eman**. *Methods Mol Biol* 2010, **673**:157-173.

173. Sorzano COS, Marabini R, Velázquez-Muriel J, Bilbao-Castro JR, Scheres SHW, 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.

174. van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M: **A new generation of the IMAGIC image processing system**. *J Struct Biol* 1996, **116**:17-24.

175. Grigorieff N: **Frealign: an exploratory tool for single-particle cryo-EM**. *Methods in Enzymology. The Resolution Revolution: Recent Advances in cryoEM*. Academic Press; 2016:191-226.

176. Pablo CM, José G, Adrián Q, de la Rosa Trevín José Miguel, Airén Z, Jesús CA, Mohsen K, Javier V, del Cano Laura S, Joan SCO, María SCJ: **Scipion web tools: easy to use cryo-EM image processing over the web**. *Protein Sci* 2018, **27**:269-275.

177. Marabini R, Ludtke SJ, Murray SC, Chiu W, de la Rosa-Trevín JM, Patwardhan A, Heymann JB, Carazo JM: **The electron microscopy exchange (EMX) initiative**. *J Struct Biol* 2016, **194**:156-163.

178. •• Sorzano COS, Marabini R, Vargas J, Otón J, Cuenca-Alba J, Quintana A, de la Rosa-Trevín JM, Carazo JM: **Interchanging geometry information in electron microscopy single particle analysis: mathematical context for the development of a standard**. *Computational Methods for Three-Dimensional Microscopy Reconstruction*. Springer; 2014:7-42.
Interchanging information between packages requires common formats and clear geometrical definitions. This article focuses on the latter.

179. Kleywegt GJ, Velankar S, Patwardhan A: **Structural biology data archiving – where we are and what lies ahead**. *FEBS Lett* 2018.