Latest advances in image processing for single particle analysis by electron cryomicroscopy and challenges ahead

J.L. Vilas[b], N. Tabassum[a], J. Mota[b], D. Maluenda[b], A. Jiménez-Moreno[b], T. Majtner[b], J.M. Carazo[b], S.T. Acton[a], C.O.S. Sorzano[b,c*]

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

Abstract

Electron cryomicroscopy (cryo-EM) 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. Cryo-EM has been successfully used to visualize molecules such as ribosomes, viruses, and ion channels, for example. Obtaining structural models of these at various conformational states leads to insight on how these molecules function. Recent advances in imaging technology have given cryo-EM a scientific rebirth. Because of imaging improvements, image processing and analysis of the resultant images have increased the resolution such that molecular structures can be resolved at the atomic level. Cryo-EM is ripe with stimulating image processing challenges. In this article, we will touch on the most essential in order to build an accurate structural three-dimensional model from noisy projection images. Traditional approaches, such as k-means clustering for class averaging, will be provided as background. With this review, however, we will highlight fresh approaches from new and varied angles for each image processing sub-problem, including a 3D reconstruction method for asymmetric molecules using just two projection

*Corresponding author
Email address: coss@cnb.csic.es (C.O.S. Sorzano[b,c])

images and deep learning algorithms for automated particle picking.

Keywords: Cryo-electron microscopy, Single Particle Analysis, Image processing algorithms

---

1. Introduction

Cryo-Electron microscopy (cryo-EM) 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. Cryo-EM is currently capable of achieving quasi-atomic resolution (1.8Å) in some specimens, and visualizing specimens with molecular weights below 100 kDa with a resolution better than 4Å [1]. Beside that, Cryo-EM can yield key insight into the dynamics of macromolecules [2, 3, 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 detectors at the microscope, 2) faster and more robust image processing algorithms, and 3) more reproducible sample preparation techniques.

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

- Image formation process. Much attention has been placed on better understanding of the physicochemical processes leading to radiation damage [6, 7, 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 charging effect [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 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 its freezing speed and reproducibility [30, 31, 32, 33].

This paper is organized as follows: in Section 2 we review the advances during the last five years in image processing algorithms for Single Particle Analysis. In Section 3 we expose the current open problems in the field from the algorithmic point of view, and present conclusions. A graphical summary of the main topics discussed is shown in Figure 1. The blue arrows between 2D Processing and 3D Analysis depict the cyclical nature of different stages - the order of steps may vary from method to method.

2. Recent Advances in Image Processing Algorithms for Single Particle Analysis

In terms of software, large 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.

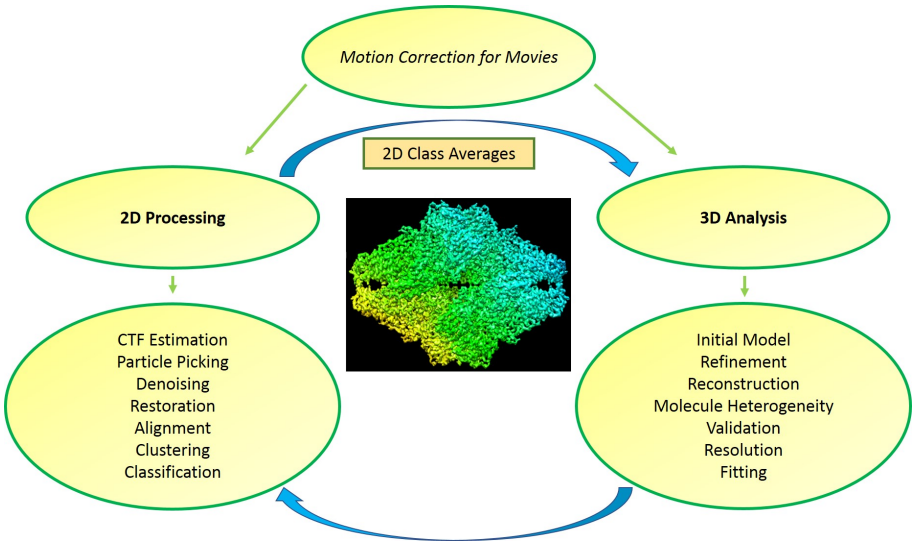


Figure 1: Summary of the main topics discussed in this review. Pictured is a 3D reconstruction of β-galactosidase (isosurface representation of Xmipp highres reconstruction.)

2.1. 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 lead to higher resolution reconstructions. They were studying 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 record movies in which the movement of the sample under the electron beam can be tracked. The correction of 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. Through understanding the different options, 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 here with low dose and defocus. The package utilizes minimal heuristics that minimizes the number of arbitrary input parameters required of 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] for anisotropic correction of beam-induced motion was introduced. 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 the normal cryo-EM processing pipeline.

Another problem with movies is related to their acquisition using Direct Electron Detector (DED), where non-negligible differences between the gain of different sensor areas could be introduced. 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 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.)

2.2. 2D Processing

2.2.1. 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 reach the best 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 during the last years trying to improve the computation time and the accuracy, due to the large amount of micrographs to analyze. A novel parameter-free approach has been presented in [51] in which a fast way to recover the defocus and astigmatism of the CTF without the need of non-linear optimization procedures and an initial defocus estimation is proposed. 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 fractionation and phase plate [52]. Gctf accelerates the CTF estimation using GPU. The main target of this is to maximize the cross-correlation of a simulated CTF with the logarithmic amplitude spectra of observed micrographs after background subtraction. Also, 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

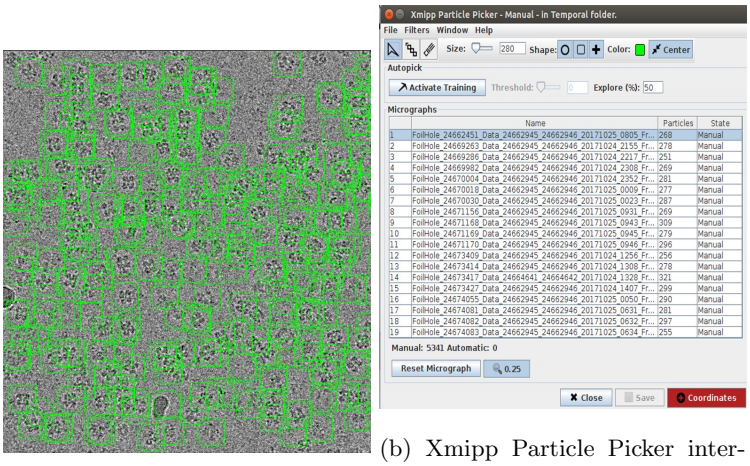
is becoming more difficult to compare their results across several runs and to select the best parameters to measure the CTF quality. To address this difficulty, a new parameter has been proposed in [54]. They introduce for this purpose the so-called CTF resolution, where they measure the correlation falloff of the calculated CTF oscillations against the normalized oscillating signal of 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 during the last years trying to get more contrast in the electron micrographs. The phase shift brought in by a physical phase plate introduced in the microscope column allows for the maximum 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, so it is not possible to correct physical aberrations. Danev et al. [55] proposed using the VPP with a bit of defocus. The advantage of this proposal is that the defocus can now be readily be identified through the oscillations of the Thon rings, and its drawback is that the small defocus causes some high frequencies to be damped. The CTF correction for Volta Phase Plate data is available in the three software implementations mentioned earlier.

2.2.2. Particle Picking

Because of the strong background noise, low contrast images, and sample heterogeneity, typically a large number of single-particle images is required for reliable 3D reconstruction. Methods for particle picking from micrographs can be divided into two main categories. The first one is a manual picking process, which is usually a laborious and time-consuming task. It requires a large amount of human effort to obtain a sufficient number of particles that also must be of high quality for high-resolution 3D reconstruction. Moreover, manual picking is considered subjective and can introduce bias and inconsistency.

Therefore, currently more popular is the second category consisting of semi-automated and automated methods. This category includes generative approaches, which measure the similarity to a certain reference image. A typical



(a) Detected particles

(b) Xmipp Particle Picker interface

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.

representative of generative methods is a template-matching technique, which is employed in RELION [56, 57] or in highly parallel GPU-accelerated gEMpicker [58]. The input here consists of a micrograph and images containing 2D templates to match. 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 could 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 Kai Zhang [60], which is a GPU accelerated program for flexible and fully automatic particle picking from cryo-EM micrographs with or without templates. The automatic particle picker can learn also from the user the particles of interest by way of the method given in [61]. This method is available in Xmipp 3.0 software [37], and an example of use 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 sort-

ing method based on multivariate statistical analysis of a particle set could be used to separate most 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 *EMHP* package [63].

In recent years, deep learning methods start to be employed for particle picking in regular micrographs (not tilted pairs.) *DeepPicker* [64] consists of two modules, where in model training, labeled 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 input micrographs. Another recent model also derived from a deep CNN is *DeepEM* [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) [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, 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] along with *MaverickTilt* software determining tilt pairs from independent particle coordinates from images [69]. Vilas *et al.* introduced a method of automatically finding correspondences of particles in the untilted and tilted micrographs [70]. The method is available in Scipion [41].

2.2.3. 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 and it can be modelled by a mixed Poisson-Gaussian (PG) distribution, often encountered also in astronomy [71, 72], biology [73] and medicine [74]. Image restoration methods (CTF correction and denoising) are based on estimating original images from these blurred and noisy observations. In one first step, restoration methods

can be separated in two big groups non-blind and blind, depending on whether the Point Spread Function (PSF) is known or not.

In addition, the non-blind image restoration techniques can also be 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 but 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.

T. Bhamre et al. [76] presented a new approach for non-blind image restoration of cryo-EM images based on a modified Wiener filtering. They name 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 projections of a certain three-dimensional molecule in different orientations, the covariance matrix represents the overall image variability due to the three-dimensional 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. B. 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 cryo-EM 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 which 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.

2.2.4. 2D Alignment, Clustering, and Classification

One of the main drawbacks of the cryo-EM single particle analysis is to deal with images with very poor SNR. However, a large number of experimental images is usually acquired. 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, 81, 82]. The most used methods to simultaneously 2D align and cluster (SAC) are based on the 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. C. 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. Usually the input dataset is too heterogeneous. The degree of heterogeneity in a cluster can be analyzed using a great variety of procedures, e.g. 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 stable core. To be more flexible, the implementation can relax this condition. In this way, the stable core is a subset of these particles which have been together for all classification levels (with a certain number of tolerance).

The previous methods are devoted to discrete classification; however, this kind of approaches could not be well suited with macromolecules 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 cryo-EM images, e.g. [89].

2.3. 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 collected. The aim of this process is to reach the optimal 3D model considering the information carried by the 2D cryo-EM images. This task is a main 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 starting with an initial model estimation, which can be seen as an estimation of the start-

ing point in the solution landscape, followed by a refinement to move along the whole landscape, improving the reconstructed model in every step. The refinement algorithms easily get stuck in local minima of the solution landscape [90]. Therefore, a good design of the initial volume estimation and refinement algorithms is key in the accuracy of the final 3D model generated.

2.3.1. 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 whose structure is unknown, as using an incorrect initial model can lead to bias in the final map, or slow convergence of the refinement algorithm.

In the recent years a plethora of initial model algorithms have appeared and, if 5 years ago the initial volume was an important problem, 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 are based on the Central Slice Theorem [91] that 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 line in the Fourier space, named the common line. The methods [80, 92, 93, 94, 95] are based on this theorem. [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, a global assignment of orientations that maximizes the number of satisfied pairwise relations can be derived. The idea of synchronization was further studied in [94] where a graph-partitioning algorithm is suggested to consistently assign orientations, giving a confidence value to each one. One typical problem with these methods is that they are prone to detect false common lines. In [93] a method dealing with this problem is proposed, in which the orientations were estimated by minimization of the sum of unsquared residuals,

adding a spectral norm term to avoid the 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 giving them a probability value. However, the main drawback of the common lines approaches has not been overcome yet, as they still tend to easily fail when the detection rate of common lines is too low due to the low SNR in typical cryo-EM 2D images. [96]

Another usual approach to the initial model problem is to follow a statistical framework, e.g., [97, 82, 98, 99], in which the alignment parameters can be found optimizing some related quantity. [97] presented a probabilistic initial 3D volume generation (PRIME) where each image is assigned to a range of orientations with the highest correlations. Then, the 3D initial model is generated 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 class average 2D images 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).

[98] was based on Bayesian inference. 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 taken into account that a low number of pseudo-atoms in the pseudo-atomic model could generate inaccurate structural representations. The algorithm presented in [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 ends up in a weighted least squares problem, in which the weights are given by both the experimental image and the projection direction. Actually, this method introduced an important idea in the field: not only experimental images can vote during the construction of a model by assigning a weight to

each projection direction, but 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 getting the definitive initial model, tend 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] show 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. However, the methods shown in this paper may lead model initialization research in a fresh, promising direction.

Finally, [101] a particle swarm optimization method is 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, is not possible to build an initial model following the common cryo-EM pipeline. In this situation, it is possible to use negatively stained samples and the Random Conical Tilt (RCT) [66] or Orthogonal Tilt Reconstruction (OTR) [67] procedures, obtaining a low-to-medium resolution 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 five years ago. More robust algorithms are still in need, since there are situations in which the existing algorithms are not able to produce a satisfactory result.

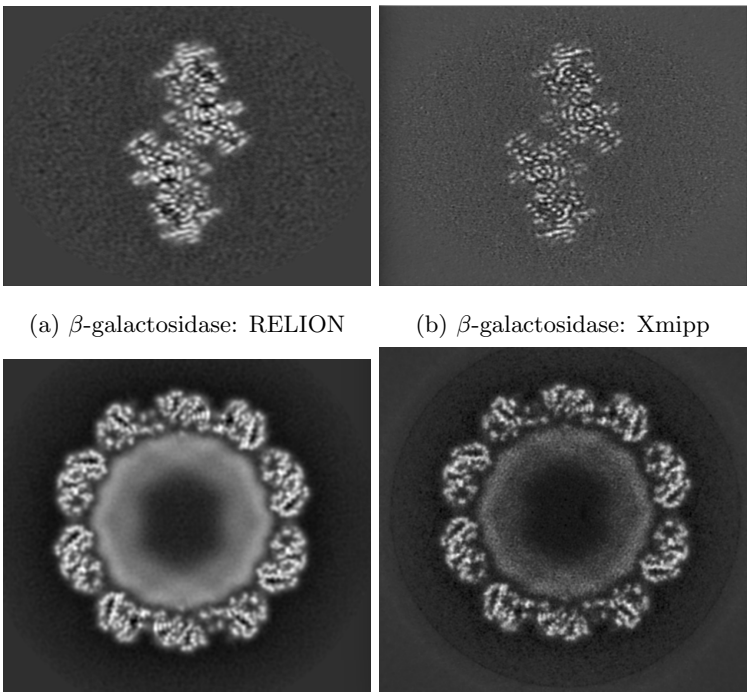
2.3.2. Refinement and Reconstruction

One key step in the cryo-EM image processing pipeline is the 3D reconstruction of a model compatible with the available 2D images coming from projections of the molecule under study, achieving a resolution sufficiently to interpret details in the macromolecular structures. This is the problem that refinement and reconstruction methods try to solve.

Despite the fact that 2D projection images are contaminated by a huge amount of noise, thanks to the large number of available images in SPA, the averaging of many images coming from the same direction is able to greatly reduce the noise level, making the reconstruction process 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 [91]. They are well suited to handle a large number of projections, which is common in SPA, with a reasonable computational burden and high accuracy when the angular coverage of the set of projections fully fills the 3D Fourier space. However, when we do not have a good angular coverage the outcomes generated by these methods cannot be 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 in which every projection direction with weights is estimated in an iterative way - evaluating a function similar to a kernel interpolator.

Another research line 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]



(a) β-galactosidase: RELION (b) β-galactosidase: Xmipp

(c) Virus: RELION (d) Virus: Xmipp

Figure 3: Examples of two reconstructed strutures 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*)

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 algorithm depends on the user, because they might cast slightly but non-negligible results, an example showing two reconstruction methods is 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 graphic processing units (GPUs).

More general statistical methods are gaining popularity recently. [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 main problem of this algorithm, since it essentially relied on an arbitrary initial map, was the sensitivity to be biased towards the initial map, although the SGD is supposed to help in this regard.

2.3.3. Molecule Heterogeneity

Macromolecules can undergo conformational changes due to their functional needs and the interaction with other molecules and the environment. For this reason, in the 2D cryo-EM images it is possible to visualize different molecule conformations, which poses a great challenge in the development of processing algorithms to analyze the molecular structures. Heterogeneity is currently an active field of research in cryo-EM as to get the highest resolution in the 3D model reconstruction is essential to discover the presence of different conformations. In this review, 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. [108] combined both with Monte Carlo/Metropolis scheme to randomly select the modes to deform the structure with the aim of generating trajectories between two conformational states. In [109] ANM and MD were also used to couple local and global motions efficiently. The method performed a large number of MD simulations, each of them corresponding to the excitation of a randomly determined linear combination of selected normal modes. Similarly, in [110] combinations of ANMs were used to calculate the conformational space for a molecule, and a clustering procedure was applied to construct representative substates.

Among the statistical approaches is a method for sorting structural states found in [111]. It was based on bootstrapping of 3D sub-ensembles and 3D multivariate statistical analysis followed by 3D classification. In [112] a method to analyze distances among elastically aligned pairs of EM models was presented. 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. [113] studied the conformational variability 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 is able to identify motion patterns of flexible components in a conformational landscape. An example is shown in Figure 4.

A different approach to discover heterogeneity in cryo-EM data consists of estimating the covariance of the reconstructed model. [114] proposed a new estimator in the Fourier space that converges to the population covariance matrix as the number of images grows, but this method involves the inversion of a

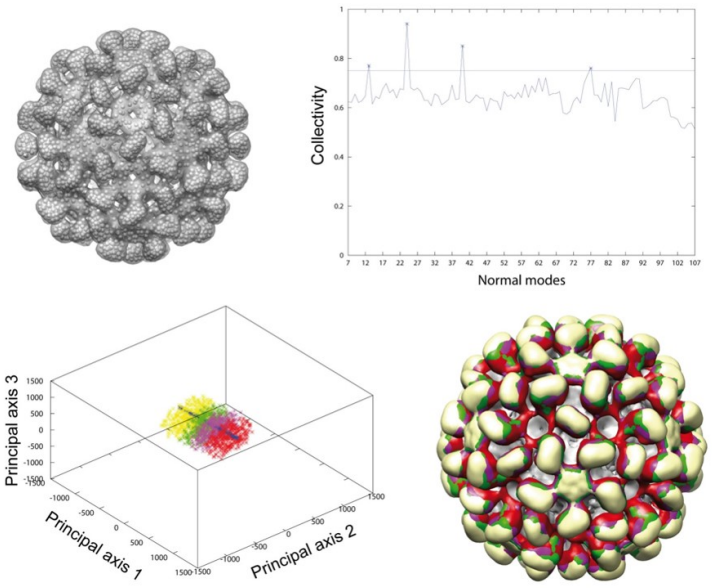


Figure 4: *Top, left:* 3D Electron density map of the Tomato Bushy Stunt Virus and its pseudoatomic representation. *Top, right:* collectivity of the normal modes of the pseudoatomic representation. *Bottom, left:* projection of the deformation parameters estimated for experimental images onto a 3D Principal Component (PCA) Space. Clustering of these projections into 4 classes. *Bottom, right:* The corresponding reconstructions of the 4 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.

high-dimensional linear operator. In [115], instead of inverting the original linear operator, it was proposed to use the conjugate gradient, achieving a lower computational complexity and the possibility of including the CTF correction. [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.

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 it is able to take into account during 3D refinement only those parts of the images where the structural variability can be found. [118] proposed to subtract projections of the fixed part of the molecule from every experimental image. This way, the modified experimental image only contains the moving part of the molecule. 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 enough size so that the subtracted projections can be correctly aligned.

Despite all the research in heterogeneity, the main 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 to a 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.

2.3.4. Validation of Results

The reconstruction workflow involves many steps in which the user decisions might determine the quality or even the validity of the electron density map. The low SNR of cryo-EM images complicates the reconstruction process. In particular, it can induce problems in critical steps, especially in the angular assignment of particles. Thus, low quality maps can be obtained or, in the worst case, a wrong map can be elucidated. The map validation can be carried by means of external techniques as X-rays or NMR, 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. By substituting a certain number of experimental particles by noisy particles and reconstructing, a validation can be carried out [120]. The goal will be to analyze the resolution of the reconstructed volume before and after noise substitution. If both resolutions are consistent, then an aligning problem is detected.
2. *Tilt Pairs Validation*: This was the first validation method [121, 122, 123] and requires a measurement of the sample 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, i.e. the angular relation between the untilted and tilted particle. The results of the angular alignment are simply plotted in a polar plot, in which the the radial measure represents 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. The high level of noise might introduce non-negligible alignment errors which are shown as scattered points in the polar plot; to analyze the existence of clusters an statistical approach is required [124].

3. *Alignability validation*: These methods aim at measuring the alignability of the set of images used for reconstruction [125, 126]. Leaving out symmetrical issues, 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 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 cryo-EM were previously obtained by other techniques such as X-ray crystallography or NMR. In these cases, the atomic model is known. Then, the electron density map must follow the atomic model at least at medium-low resolution.

2.4. Resolution

Once the macromolecular structure has been obtained and validated, it is necessary to report a quality measurement of its electron density map. The resolution tries to answer this regard. 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 capability of an imaging system of distinguishing 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 (map in cryo-EM). Nevertheless, when the imaging system is omitted and only the image is analyzed, other criteria are used, e.g., Johnson

criteria [128].

In cryo-EM, the resolution has been traditionally analyzed in a global sense, that is, reporting a single parameter called global resolution 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) where the 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 consist in sppliting 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]. Cryo-EM images present low SNR and even particles of noise can be aligned i.e. features of noise correlate with the reference [133, 134, 120], 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 orien-

tation. The heterogeneity has been identified as one of the main problems in cryo-EM [136], and contradicts the SPA hypothesis 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 electron 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 cast better solved directions than others [137]. To overcome this problem of angular coverage, [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 maps in cryo-EM [135]. It extends the FSC measurement in a local sense. Thus, by means of two half maps and a moving window centered in the interest voxel a local FSC can be calculated. The critical point is to set the window size. Logically, this is a self-consistency measurement, as the FSC itself, and it preserves all FSC properties. Interestingly, *Blocres* introduced the possibility of computing the locally filtered map at the local resolution values.

Nowadays, the most spread method in local resolution measurements is *ResMap* [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 means of a steerable function basis that allows for modeling of sinusoidal signals by means of 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 i.e. in an hypothesis the expected value of the number of resolutions wrong assigned over the total number of resolution assigned.

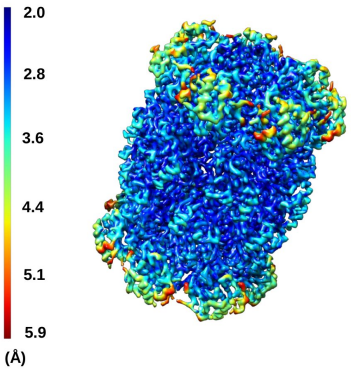


Figure 5: Local resolution map of the Thermoplasma acidophilum 20S proteasome using the MonoRes method [140].

Recently, a new method called *MonoRes* for estimating local resolution has been published [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 is carried out performing 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 local filtered map at the local resolution values, shown in Figure 5.

2.4.1. Fitting an Atomic Model

Thus far, we have discussed methods for building and refining a 3D reconstruction of the molecule being imaged. This reconstruction is in reality just a density map. The ultimate interest in the research community is focused on an atomic level structural model of the macromolecule. Initially, a 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]

When fitting to 3D density maps, both rigid 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. This can save computational energy when modeling large molecules. [143] The coarse grained model can then be refined, like rigid fitting, with flexible fitting - flexible fitting 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 cryo-EM 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. More recently, 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]

3. Conclusions - Current Image Processing Challenges

Despite the recent successes of cryo-EM, this modality is still a very active research area, and experimental advances are still in development including sample preparation [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 [154, 7, 136], 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. The steady progress in this area is clear and positive, 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 us out-of-plane rocking along 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 still a better match into the processing workflow is required. Even such a basic task as focus determination is far from trivial and reliable for high resolution [162]. Additionally, 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 [151, 14, 16]. 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 helps to reveal the dynamics of the macromolecule under study. On the other side, only homogeneous sets of particles can be reconstructed to 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 im-

ages. 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 [165, 89], 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 find nonsensical structures. There are specific initial volume algorithms to handle this issue [166]. However, currently, there is no algorithm specifically designed with flexibility/heterogeneity in mind.

4. Complement 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 in need [122, 120, 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 cryo-EM as an imaging technique, many engineering groups are getting involved in the global research effort and adding new small pieces of software solving specific problems. In addition, we have the traditional software packages that cover the whole image processing pipeline (Relion [171], Eman [172], Xmipp [173, 37], Spider [38], Imagic [174], Frealign [175], ...) and systems that integrate algorithms from multiple sources (Appion [42] and Scipion [41, 176]). This ecosystem of software lacks a common standard of interchanging information. Although some attempts have been proposed at the level of metadata [177] and geometry [178], they have not been widely adopted. Additionally, the field is lacking 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 year after year. Thus, the structural biology community and in particular EM-community is getting awareness about sharing this information. To achieve that, there are some web services as they are: The EMDataBank (http://www.emdatabank.org), Worldwide Protein Data Bank (wwPDB; http://wwpdb.org). Other databases such as EMPIAR (http://www.ebi.ac.uk/pdbe/emdb/empiar/) pursues raw data availability. For a good review on data management and databases in structural biology see [179].

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

[1] *A. Merk, A. Bartesaghi, S. Banerjee, V. Falconieri, P. Rao, M. I. Davis, R. Pragani, M. B. Boxer, L. A. Earl, J. L. S. Milne, S. Subramaniam, Breaking cryo-EM resolution barriers to facilitate drug discovery, Cell 165 (2016) 1698–1707,
Breakthrough experimental work showing the capabilities of electron microscopy to visualize small molecules and quasi-atomic resolution.

[2] M. Bonomi, G. T. Heller, C. Camilloni, M. Vendruscolo, Principles of protein structural ensemble determination., Current Opinion in Structural Biology 42 (2017) 106–116.

[3] J. Frank, Time-resolved cryo-electron microscopy: Recent progress., Journal of Structural Biology 200 (3) (2017) 303–306.

[4] S. Jonić, Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images, Current Opinion in Structural Biology 43 (2017) 114 – 121.

[5] S. Rawson, M. McPhillie, R. Johnson, C. Fishwick, S. Muench, The potential use of single-particle electron microscopy as a tool for structure-based inhibitor design, Acta Crystallographica. Section D, Structural Biol 73 (6).

[6] R. F. Egerton, Control of radiation damage in the TEM., Ultramicroscopy 127 (2013) 100–108.

[7] R. M. Glaeser, How good can cryo-EM become?, Nature Methods 13 (1) (2016) 28–32.

[8] M. Mishyna, O. Volokh, Y. Danilova, N. Gerasimova, E. Pechnikova, O. S. Sokolova, Effects of radiation damage in studies of protein-DNA complexes by cryo-EM., Micron 96 (2017) 57–64.

[9] G. McMullan, K. R. Vinothkumar, R. Henderson, Thon rings from amorphous ice and implications of beam-induced brownian motion in single particle electron cryo-microscopy., Ultramicroscopy 158 (2015) 26–32.

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

[11] X. Li, S. Q. Zheng, K. Egami, D. A. Agard, Y. Cheng, Influence of electron dose rate on electron counting images recorded with the K2 camera., Journal of Structural Biology 184 (2013) 251–260.

[12] H. Shigematsu, F. J. Sigworth, Noise models and cryo-EM drift correction with a direct-electron camera., Ultramicroscopy 131 (2013) 61–69.

[13] M. Vulović, R. B. G. Ravelli, L. J. van Vliet, A. J. Koster, I. Lazić, U. Lücken, H. Rullgård, O. Öktem, B. Rieger, Image formation modeling in cryo-electron microscopy., Journal of Structural Biology 183 (1) (2013) 19–32.

[14] *M. Vulović, L. M. Voortman, L. J. van Vliet, B. Rieger, When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-EM., Ultramicroscopy 136C (2013) 61–66,
Limitations of the weak-phase object assumption in EM.

[15] P. W. Hawkes, The correction of electron lens aberrations., Ultramicroscopy 156 (2015) A1–A64.

[16] P. J. B. Koeck, A. Karshikoff, Limitations of the linear and the projection approximations in three-dimensional transmission electron microscopy of fully hydrated proteins., Journal of Microscopy 259 (3) (2015) 197–209.

[17] I. Lobato, D. Van Dyck, MULTEM: A new multislice program to perform accurate and fast electron diffraction and imaging simulations using graphics processing units with CUDA., Ultramicroscopy 156 (2015) 9–17.

[18] *K. H. Downing, R. M. Glaeser, Estimating the effect of finite depth of field in single-particle cryo-EM., Ultramicroscopy 184 (2017) 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] C. J. Russo, R. Henderson, Microscopic charge fluctuations cause minimal contrast loss in cryoEM., Ultramicroscopy 187 (2018) 56–63.

[20] C. J. Russo, R. Henderson, Charge accumulation in electron cryomicroscopy., Ultramicroscopy 187 (2018) 43–49.

[21] *R. Danev, B. Buijsse, M. Khoshouei, J. M. Plitzko, W. Baumeister, Volta potential phase plate for in-focus phase contrast transmission electron microscopy., Proc. Natl. Acad. Sci. USA 111 (44) (2014) 15635–15640,
The use of Volta Phase Plates have revolutionzed the contrast in the electron microscope.

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

[23] P. J. B. Koeck, Design of an electrostatic phase shifting device for biological transmission electron microscopy., Ultramicroscopy 187 (2018) 107–112.

[24] P. J. B. Koeck, Annular dark field transmission electron microscopy for protein structure determination., Ultramicroscopy 161 (2016) 98–104.

[25] P. J. B. Koeck, An aperture design for single side band imaging in the transmission electron microscope., Ultramicroscopy 182 (2017) 81–84.

[26] *C. J. Russo, R. Henderson, Ewald sphere correction using a single side-band image processing algorithm., Ultramicroscopy 187 (2018) 26–33,
How to overcome the limits of the central slice theorem for achieving high resolution.

[27] M. Kuijper, G. van Hoften, B. Janssen, R. Geurink, S. De Carlo, M. Vos, G. van Duinen, B. van Haeringen, M. Storms, FEI's direct electron detector developments: Embarking on a revolution in cryo-TEM., Journal of Structural Biology 192 (2) (2015) 179–187.

[28] **A. R. Faruqi, G. McMullan, Direct imaging detectors for electron microscopy, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 878 (2018) 180–190,
Excellent review on direct detectors, the device that has revolutionzed the quality of electron microscopy images.

[29] A. Chari, D. Haselbach, J.-M. Kirves, J. Ohmer, E. Paknia, N. Fischer, O. Ganichkin, V. Möller, J. J. Frye, G. Petzold, M. Jarvis, M. Tietzel, C. Grimm, J.-M. Peters, B. A. Schulman, K. Tittmann, J. Markl, U. Fischer, H. Stark, Proteoplex: stability optimization of macromolecular complexes by sparse-matrix screening of chemical space., Nature Methods 12 (2015) 859–865.

[30] I. Razinkov, V. P. Dandey, H. Wei, Z. Zhang, D. Melnekoff, W. J. Rice, C. Wigge, C. S. Potter, B. Carragher, A new method for vitrifying samples for cryoEM., Journal of Structural Biology 195 (2016) 190–198.

[31] S. A. Arnold, S. Albiez, A. Bieri, A. Syntychaki, R. Adaixo, R. A. McLeod, K. N. Goldie, H. Stahlberg, T. Braun, Blotting-free and lossless cryo-electron microscopy grid preparation from nanoliter-sized protein samples and single-cell extracts., Journal of Structural Biology 197 (2017) 220–226.

[32] X. Feng, Z. Fu, S. Kaledhonkar, Y. Jia, B. Shah, A. Jin, Z. Liu, M. Sun, B. Chen, R. A. Grassucci, Y. Ren, H. Jiang, J. Frank, Q. Lin, A fast and

effective microfluidic spraying-plunging method for high-resolution single-particle cryo-EM., Structure 25 (2017) 663–670.e3.

[33] V. P. Dandey, H. Wei, Z. Zhang, Y. Z. Tan, P. Acharya, E. T. Eng, W. J. Rice, P. A. Kahn, C. S. Potter, B. Carragher, Spotiton: New features and applications., Journal of Structural Biology 202 (2) (2018) 161–169.

[34] *S. H. W. Scheres, RELION: implementation of a bayesian approach to cryo-EM structure determination., Journal of Structural Biology 180 (3) (2012) 519–530,
Current reference algorithm for 3D classification and reconstruction.

[35] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, S. J. Ludtke, Eman2: an extensible image processing suite for electron microscopy., Journal of Structural Biology 157 (2007) 38–46.

[36] T. Grant, R. Alexis, N. Grigorieff, cisTEM, user-friendly software for single-particle image processing., eLife 7 (e35383).

[37] *J. M. de la Rosa-Trevín, J. Otón, R. Marabini, A. Zaldívar, J. Vargas, J. M. Carazo, C. O. S. Sorzano, Xmipp 3.0: an improved software suite for image processing in electron microscopy., Journal of Structural Biology 184 (2) (2013) 321–328,
An integrative software platform with over 200 image processing protocols from over 20 different software packages.

[38] T. R. Shaikh, H. Gao, W. Baxter, F. J. Asturias, N. Boisset, A. Leith, J. Frank, Spider image processing for single-particle reconstruction of biological macromolecules from electron micrographs, Nature Protocols 3 (2008) 1941–1974.

[39] M. Hohn, G. Tang, G. Goodyear, P. R. Baldwin, Z. Huang, P. A. Penczek, C. Yang, R. M. Glaeser, P. D. Adams, S. J. Ludtke, SPARX, a new environment for cryo-EM image processing., Journal of Structural Biology 157 (1) (2007) 47–55.

[40] B. Heymann, D. Belnap, Bsoft: Image processing and molecular modeling for electron microscopy, Journal of Structural Biology 157 (1) (2007) 3–18.

[41] J. M. de la Rosa-Trevín, A. Quintana, L. Del Cano, A. Zaldívar, I. Foche, J. Gutiérrez, J. Gómez-Blanco, J. Burguet-Castell, J. Cuenca-Alba, V. Abrishami, J. Vargas, J. Otón, G. Sharov, J. L. Vilas, J. Navas, P. Conesa, M. Kazemi, R. Marabini, C. O. S. Sorzano, J. M. Carazo, Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy., Journal of Structural Biology 195 (1) (2016) 93–99.

[42] G. C. Lander, S. M. Stagg, N. R. Voss, A. Cheng, D. Fellmann, J. Pulokas, C. Yoshioka, C. Irving, A. Mulder, P. W. Lau, D. Lyumkis, C. S. Potter, B. Carragher, Appion: an integrated, database-driven pipeline to facilitate em image processing., Journal of Structural Biology 166 (2009) 95–102.

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

[44] Z. A. Ripstein, J. L. Rubinstein, Processing of cryo-EM movie data., Methods in Enzymology 579 (2016) 103–124.

[45] X. Li, P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard, Y. Cheng, Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM., Nature Methods 10 (6) (2013) 584–590.

[46] J. L. Rubinstein, M. A. Brubaker, Alignment of cryo-EM movies of individual particles by optimization of image translations., Journal of Structural Biology 192 (2) (2015) 188–195.

[47] R. A. McLeod, J. Kowal, P. Ringler, H. Stahlberg, Robust image align-

ment for cryogenic transmission electron microscopy., Journal of Structural Biology 197 (2017) 279–293.

[48] S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, Motioncor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy., Nature Methods 14 (2017) 331–332.

[49] P. Afanasyev, R. B. G. Ravelli, R. Matadeen, S. De Carlo, G. van Duinen, B. Alewijnse, P. J. Peters, J.-P. Abrahams, R. V. Portugal, M. Schatz, M. van Heel, A posteriori correction of camera characteristics from large image data sets., Scientific Reports 5 (2015) 10317.

[50] C. O. S. Sorzano, E. Fernández-Giménez, V. Peredo-Robinson, J. Vargas, T. Majtner, G. Caffarena, J. Otón, J. L. Vilas, J. M. de la Rosa-Trevín, R. Melero, J. Gómez-Blanco, J. Cuenca, L. del Cano, P. Conesa, R. Marabini, J. M. Carazo, Blind estimation of ded camera gain in electron microscopy, Journal of Structural Biology 203 (2) (2018) 90–93.

[51] J. Vargas, J. Otón, R. Marabini, S. Jonic, J. M. de la Rosa-Trevín, J. M. Carazo, C. O. S. Sorzano, FASTDEF: Fast defocus and astigmatism estimation for high-throughput transmission electron microscopy., Journal of Structural Biology 181 (2) (2013) 136–148.

[52] A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs., Journal of Structural Biology 192 (2) (2015) 216–221.

[53] K. Zhang, Gctf: real-time ctf determination and correction., Journal of Structural Biology 193 (2016) 1–12.

[54] L. K. Sheth, A. L. Piotrowski, N. R. Voss, Visualization and quality assessment of the contrast transfer function estimation., Journal of Structural Biology 192 (2) (2015) 222–234.

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

[56] S. H. W. Scheres, Semi-automated selection of cryo-EM particles in RELION-1.3., Journal of Structural Biology 189 (2).

[57] D. Kimanius, B. O. Forsberg, S. H. Scheres, E. Lindahl, Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2, eLife 5 (2016) e18722.

[58] T. V. Hoang, X. Cavin, P. Schultz, D. W. Ritchie, gempicker: A highly parallel gpu-accelerated particle picking tool for cryo-electron microscopy, BMC Structural Biology 13 (1) (2013) 25.

[59] R. Langlois, J. Pallesen, J. T. Ash, D. Nam Ho, J. L. Rubinstein, J. Frank, Automated particle picking for low-contrast macromolecules in cryo-electron microscopy., Journal of Structural Biology 186 (1) (2014) 1–7.

[60] J. Kai Zhang – Gautomatch, `http://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/`, accessed: 2018-05-16.

[61] V. Abrishami, A. Zaldívar-Peraza, J. M. de la Rosa-Trevín, J. Vargas, J. Otón, R. Marabini, Y. Shkolnisky, J. M. Carazo, C. O. S. Sorzano, A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs., Bioinformatics 29 (19) (2013) 2460–2468.

[62] J. Vargas, V. Abrishami, R. Marabini, J. M. de la Rosa-Trevín, A. Zaldivar, J. M. Carazo, C. O. S. Sorzano, Particle quality assessment and sorting for automatic and semiautomatic particle-picking techniques., Journal of Structural Biology 183 (3) (2013) 342–353.

[63] Z. Berndsen, C. Bowman, H. Jang, A. B. Ward, EMHP: an accurate automated hole masking algorithm for single-particle cryo-EM image processing, Bioinformatics 33 (23) (2017) 3824–3826.

[64] F. Wang, H. Gong, G. Liu, M. Li, C. Yan, T. Xia, X. Li, J. Zeng, Deeppicker: A deep learning approach for fully automated particle picking in cryo-EM., Journal of Structural Biology 195 (3) (2016) 325–336.

[65] Y. Zhu, Q. Ouyang, Y. Mao, A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy, BMC Bioinformatics 18 (1).

[66] M. Radermacher, W. Hoppe, 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] A. Leschziner, Chapter nine - the orthogonal tilt reconstruction method, in: G. J. Jensen (Ed.), Cryo-EM, Part B: 3-D Reconstruction, Vol. 482 of Methods in Enzymology, Academic Press, 2010, pp. 237 – 262. doi:https://doi.org/10.1016/S0076-6879(10)82010-5.
URL http://www.sciencedirect.com/science/article/pii/S0076687910820105

[68] M. Shatsky, P. Arbelaez, B. G. Han, D. Typke, S. E. Brenner, J. Malik, R. M. Glaeser, Automated particle correspondence and accurate tilt-axis detection in tilted-image pairs., Journal of Structural Biology 187 (1) (2014) 66–75.

[69] F. Hauer, C. Gerle, J.-M. Kirves, H. Stark, Automated correlation of single particle tilt pairs for random conical tilt and orthogonal tilt reconstructions., Journal of Structural Biology 181 (2) (2013) 149–154.

[70] J. L. Vilas, J. Navas, J. Gómez-Blanco, J. M. de la Rosa-Trevín, R. Melero, I. Peschiera, I. Ferlenghi, J. Cuenca, R. Marabini, J. M. Carazo, J. Vargas, C. O. S. Sorzano, Fast and automatic identification of particle tilt pairs based on delaunay triangulation., Journal of Structural Biology 196 (2016) 525–533.

[71] F. Benvenuto, A. L. Camera, C. Theys, A. Ferrari, H. Lantéri, M. Bertero, The study of an iterative method for the reconstruction of images corrupted by poisson and gaussian noise, Inverse Problems 24 (3) (2008) 035016.

[72] D. L. Snyder, A. M. Hammoud, R. L. White, Image recovery from data acquired with a charge-coupled-device camera, Journal of the Optical Society of America A 10 (5) (1993) 1014–1023.

[73] T. E. Nichols, J. Qi, E. Asma, R. M. Leahy, Spatiotemporal reconstruction of list-mode PET data, IEEE Transactions on Medical Imaging 21 (4) (2002) 396–404.

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

[75] **P. A. Penczek, Image restoration in cryo-electron microscopy., Methods in Enzymology 482 (2010) 35–72,
Excellent review of image restoration in electron microscopy.

[76] T. Bhamre, T. Zhang, A. Singer, Denoising and covariance estimation of single particle cryo-EM images, Journal of Structural Biology 195 (1) (2016) 72–81.

[77] C. O. S. Sorzano, S. Jonic, R. Núñez Ramírez, N. Boisset, J. M. Carazo, Fast, robust and accurate determination of transmission electron microscopy contrast transfer function, Journal of Structural Biology 160 (2007) 249–262.

[78] C. O. S. Sorzano, J. Vargas, J. Otón, V. Abrishami, J. M. de la Rosa Trevín, S. del Riego, A. Fernández-Alderete, C. Martínez-Rey, R. Marabini, J. M. Carazo, Fast and accurate conversion of atomic models into electron density maps, AIMS Biophysics 2 (2015) 8–20.

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

[80] M. Van Heel, Angular reconstitution: A posteriori assignment of projection directions for 3d reconstruction, Ultramicroscopy 21 (1987) 111–123.

[81] N. Jaitly, M. A. Brubaker, R. H. Rubinstein, J. L. snf Lilien, A bayesian method for 3-D macromolecular structure inference using class average images from single particle electron microscopy, Bioinformatics 26 (19) (2010) 2406–2415.

[82] J. Vargas, A. L. Álvarez-Cabrera, R. Marabini, J. M. Carazo, C. O. S. Sorzano, Efficient initial volume determination from electron microscopy images of single particles, Bioinformatics 30 (2014) 2891–2898.

[83] J. Frank, M. Radermacher, P. Penczek, J. Zhu, Y. Li, M. Ladjadj, A. Leith, Spider and web: Processing and visualization of images in 3d electron microscopy and related fields, Journal of Structural Biology 116 (1996) 190–199.

[84] Z. Yang, J. Fang, J. Chittuluru, F. J. Asturias, P. A. Penczek, Iterative stable alignment and clustering of 2d transmission electron microscope images., Structure 20 (2) (2012) 237–247.

[85] C. F. Reboul, F. Bonnet, D. Elmlund, H. Elmlund, A stochastic hill climbing approach for simultaneous 2D alignment and clustering of cryogenic electron microscopy images., Structure 24 (2016) 988–996.

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

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

[88] C. O. S. Sorzano, J. R. Bilbao-Castro, Y. Shkolnisky, M. Alcorlo, R. Melero, G. Caffarena-Fernández, M. Li, G. Xu, R. Marabini, J. M. Carazo, A clustering approach to multireference alignment of single-particle projections in electron microscopy., Journal of Structural Biology 171 (2) (2010) 197–206.

[89] Q. Jin, C. O. S. Sorzano, J. M. de la Rosa-Trevín, J. R. Bilbao-Castro, R. Núñez Ramírez, O. Llorca, F. Tama, S. Jonić, Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes., Structure 22 (3) (2014) 496–506.

[90] C. O. S. Sorzano, R. Marabini, A. Pascual-Montano, S. H. W. Scheres, J. M. Carazo, Optimization problems in electron microscopy of single particles, Annals of Operations Research 148 (2006) 133–165.

[91] S. Zhao, H. Halling, A new fourier method for fan beam reconstruction, in: 1995 IEEE Nuclear Science Symposium and Medical Imaging Conference Record, IEEE, 1995, pp. 1287–1291.

[92] Y. Shkolnisky, A. Singer, Viewing direction estimation in cryo-EM using synchronization, SIAM Journal on Imaging Sciences 5 (3) (2012) 1088–1110.

[93] L. Wang, A. Singer, Z. Wen, Orientation determination of cryo-EM images using least unsquared deviations, SIAM Journal on Imaging Sciences 6 (4) (2013) 2450–2483.

[94] G. Pragier, I. Greenberg, X. Cheng, Y. Shkolnisky, A graph partitioning approach to simultaneous angular reconstitution., IEEE Transactions on Computational Imaging 2 (2016) 323–334.

[95] I. Greenberg, Y. Shkolnisky, Common lines modeling for reference free ab-initio reconstruction in cryo-em, Journal of Structural Biology 200 (2) (2017) 106–117.

[96] C. Kervrann, C. O. S. Sorzano, S. T. Acton, J. C. Olivo-Marin, M. Unser, A guided tour of selected image processing and analysis methods for fluorescence and electron microscopy, IEEE Journal of Selected Topics in Signal Processing 10 (2016) 6–30.

[97] H. Elmlund, D. Elmlund, S. Bengio, PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy., Structure 21 (8) (2013) 1299–1306.

[98] P. Joubert, M. Habeck, Bayesian inference of initial models in cryo-electron microscopy using pseudo-atoms., Biophysical Journal 108 (5) (2015) 1165–1175.

[99] C. O. S. Sorzano, J. Vargas, J. M. de la Rosa-Trevín, J. Otón, A. L. Álvarez-Cabrera, V. Abrishami, E. Sesmero, R. Marabini, J. M. Carazo, A statistical approach to the initial volume problem in single particle analysis by electron microscopy., Journal of Structural Biology 189 (3) (2015) 213–219.

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

[101] C. Sorzano, J. Vargas, J. Vilas, A. Jiménez-Moreno, J. Mota, T. Majtner, D. Maluenda, M. Martínez, R. Sánchez-García, J. Segura, J. Otón, R. Melero, L. del Cano, P. Conesa, J. Gómez-Blanco, Y. Rancel, R. Marabini, J. Carazo, Swarm optimization as a consensus technique for electron microscopy initial volume, Applied Analysis and Optimization 2 (2018) 299–313.

[102] V. Abrishami, J. R. Bilbao-Castro, J. Vargas, R. Marabini, J. M. Carazo, C. O. S. Sorzano, A fast iterative convolution weighting approach for gridding-based direct fourier three-dimensional reconstruction with correction for the contrast transfer function., Ultramicroscopy 157 (2015) 79–87.

[103] T. Moriya, E. Acar, R. H. Cheng, U. Ruotsalainen, A bayesian approach for suppression of limited angular sampling artifacts in single particle 3D reconstruction., Journal of Structural Biology 191 (3) (2015) 318–331.

[104] G. Xu, M. Li, C. Chen, A multi-scale geometric flow method for molecular structure reconstruction, Computational Science and Discovery 8 (2015) 014002.

[105] **C. O. S. Sorzano, J. Vargas, J. Otón, J. L. Vilas, M. Kazemi, R. Melero, L. del Caño, J. Cuenca, P. Conesa, J. Gómez-Blanco, R. Marabini, J. M. Carazo, A survey of the use of iterative reconstruction algorithms in electron microscopy, BioMed Research International 2017 (2017) 6482567, Excellent review of iterative reconstruction algorithms in electron microscopy.

[106] N. C. Dvornek, F. J. Sigworth, H. D. Tagare, Subspaceem: A fast maximum-a-posteriori algorithm for cryo-EM single particle reconstruction., Journal of Structural Biology 190 (2) (2015) 200–214.

[107] A. Punjani, J. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination., Nature Methods 14 (2017) 290–296.

[108] M. Gur, J. D. Madura, I. Bahar, Global transitions of proteins explored by a multiscale hybrid methodology: application to adenylate kinase., Biophysical Journal 105 (2013) 1643–1652.

[109] M. G. S. Costa, P. R. Batista, P. M. Bisch, D. Perahia, Exploring free energy landscapes of large conformational changes: Molecular dynamics

with excited normal modes., Journal of Chemical Theory and Computation 11 (6) (2015) 2755–2767.

[110] Z. Kurkcuoglu, I. Bahar, P. Doruker, Clustenm: Enm-based sampling of essential conformational space at full atomic resolution., Journal of Chemical Theory and Computation 12 (2016) 4549–4562.

[111] B. P. Klaholz, Structure sorting of multiple macromolecular states in heterogeneous cryo-EM samples by 3D multivariate statistical analysis, Open Journal of Statistics 5 (07) (2015) 820–836.

[112] C. O. S. Sorzano, A. L. Alvarez-Cabrera, M. Kazemi, J. M. Carazo, S. Jonić, Structmap: Elastic distance analysis of electron microscopy maps for studying conformational changes., Biophysical Journal 110 (8) (2016) 1753–1765.

[113] D. Haselbach, I. Komarov, D. E. Agafonov, K. Hartmuth, B. Graf, O. Dybkov, H. Urlaub, B. Kastner, R. Luhrmann, H. Stark, Structure and conformational dynamics of the human spliceosomal bact complex., Cell 172 (2018) 454–464.e11.

[114] E. Katsevich, A. Katsevich, A. Singer, Covariance matrix estimation for the cryo-EM heterogeneity problem., SIAM Journal on Imaging Sciences 8 (2015) 126–185.

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

[116] H. Y. Liao, Y. Hashem, J. Frank, Efficient estimation of three-dimensional covariance and its application in the analysis of heterogeneous samples in cryo-electron microscopy., Structure 23 (6) (2015) 1129–1137.

[117] Y. Gong, P. C. Doerschuk, 3-d understanding of electron microscopy images of nano bio objects by computing generative mechanical models, in:

2016 IEEE International Conference on Image Processing (ICIP), IEEE, 2016, pp. 3161–3165.

[118] X. C. Bai, E. Rajendra, G. Yang, Y. Shi, S. H. Scheres, Sampling the conformational space of the catalytic subunit of human γ-secretase., eLife 4.

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

[120] B. Heymann, Validation of 3DEM reconstructions: The phantom in the noise, AIMS Biophysics 2 (2015) 21–35.

[121] P. B. Rosenthal, R. Henderson, Optimal determination of particle orientation, absolute hand, and contrast loss in single particle electron-cryomicroscopy, Journal of Molecular Biology 333 (2003) 721–745.

[122] R. Henderson, S. Chen, J. Z. Chen, N. Grigorieff, L. A. Passmore, L. Ciccarelli, J. L. Rubinstein, R. A. Crowther, P. L. Stewart, P. B. Rosenthal, Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy., Journal of Molecular Biology 413 (5) (2011) 1028–1046.

[123] S. Wasilewski, P. B. Rosenthal, Web server for tilt-pair validation of single particle maps from electron cryomicroscopy., Journal of Structural Biology 186 (1) (2014) 122–131.

[124] C. J. Russo, L. A. Passmore, Robust evaluation of 3d electron cryomicroscopy data using tilt-pairs., Journal of Structural Biology 187 (2) (2014) 112–118.

[125] J. Vargas, J. Otón, R. Marabini, J. M. Carazo, C. O. S. Sorzano, Particle alignment reliability in single particle electron cryomicroscopy: a general approach., Scientific Reports 6 (2016) 21626.

[126] J. Vargas, R. Melero, J. Gómez-Blanco, J. M. Carazo, C. O. S. Sorzano, Quantitative analysis of 3D alignment quality: its impact on soft-validation, particle pruning and homogeneity analysis., Scientific Reports 7 (2017) 6307.

[127] M. Born, E. Wolf, Principles of Optics, 7th Edition, Cambridge Univ. Press, 1999.

[128] J. Johnson, Analysis of image forming systems, Image Intensifier Symposium (1958) 244–273.

[129] **C. O. S. Sorzano, J. Vargas, J. Otón, V. Abrishami, J. M. de la Rosa-Trevín, J. Gómez-Blanco, J. L. Vilas, R. Marabini, J. M. Carazo, A review of resolution measures and related aspects in 3D electron microscopy., Progress in Biophysics and Molecular Biology 124 (2017) 1–30,
Excellent review on resolution measures for electron microscopy and signal to noise ratio.

[130] N. Grigorieff, Resolution measurement in structures derived from single particles, Acta Crystallographica Section D 56 (2000) 1270–1277.

[131] S. H. W. Scheres, A Bayesian view on cryo-EM structure determination., Journal of Molecular Biology 415 (2) (2012) 406–418.

[132] S. Chen, G. McMullan, A. R. Faruqi, G. N. Murshudov, J. M. Short, S. H. W. Scheres, R. Henderson, High-resolution noise substitution to measure overfitting and validate resolution in 3d structure determination by single particle electron cryomicroscopy., Ultramicroscopy 135 (2013) 24–35.

[133] R. Henderson, Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise., Proc. Natl. Acad. Sci. USA 110 (45) (2013) 18037–18041.

[134] M. van Heel, Finding trimeric HIV-1 envelope glycoproteins in random noise., Proc. Natl. Acad. Sci. USA 110 (45) (2013) E4175–E4177.

[135] G. Cardone, J. B. Heymann, A. C. Steven, One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions., Journal of Structural Biology 184 (2) (2013) 226–236.

[136] E. Nogales, The development of cryo-EM into a mainstream structural biology technique, Nature Methods 13 (1) (2016) 24–27.

[137] K. Naydenova, C. J. Russo, Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy., Nature Communications 8 (2017) 629.

[138] Y. Z. Tan, P. R. Baldwin, J. H. Davis, J. R. Williamson, C. S. Potter, B. Carragher, D. Lyumkis, Addressing preferred specimen orientation in single-particle cryo-EM through tilting, Nature Methods 14 (8) (2017) 793–796.

[139] A. Kucukelbir, F. J. Sigworth, H. D. Tagare, Quantifying the local resolution of cryo-EM density maps., Nature Methods 11 (2014) 63–65.

[140] J. L. Vilas, J. Gómez-Blanco, P. Conesa, R. Melero, J. M. de la Rosa Trevín, J. Otón, J. Cuenca, R. Marabini, J. M. Carazo, J. Vargas, C. O. S. Sorzano, MonoRes: automatic and accurate estimation of local resolution for electron microscopy maps, Structure 26 (2018) 337–344.

[141] J. Esquivel-Rodríguez, D. Kihara, Computational methods for constructing protein structure models from 3D electron microscopy maps, Journal of Structural Biology 184 (2013) 93–102.

[142] D. Si, S. Ji, K. A. Nasr, J. He, A machine learning approach for the identification of protein secondary structure elements from electron cryomicroscopy density maps, Biopolymers 97 (2012) 698–708.

[143] Y. L. Chen, M. Habeck, Data-driven coarse graining of large biomolecular structures, PLoS One 12 (2017) 1–17.

[144] M. S. Chapman, A. Trzynka, B. K. Chapman, Atomic modeling of *cryo*-electron microscopy reconstructions - joint refinement of model and imaging parameters, Journal of Structural Biology 182 (2013) 10–21.

[145] A. P. Joseph, I. Lagerstedt, A. Patwardhan, M. Topf, M. Winn, Improved metrics for comparing structures of macromolecular assemblies determined by 3D electron-microscopy, Journal of Structural Biology 199 (2017) 12–26.

[146] R. Y. Wang, Y. Song, B. A. Barad, Y. Cheng, J. S. Fraser, F. DiMaio, Automated structure refinement of macromolecular assemblies from cryo-EM maps using rosetta, eLife 5 (2016) 1–22.

[147] J. Wang, On contribution of known atomic partial charges of protein backbone in electrostatic potential density maps, Protein Science 26 (2017) 1098–1104.

[148] J. Velázquez-Muriel, K. Lasker, D. Russel, J. Phillips, B. M. Webb, D. Schneidman-Duhovny, A. Sali, Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images, PNAS 109 (2012) 18821–18826.

[149] A. Matsumoto, N. Miyazaki, J. Takagi, K. Iwasaki, 2D hybrid analysis: approach for building three-dimensional atomic model by electron microscopy image matching, Scientific Reports 7 (2017) 1–12.

[150] **K. R. Vinothkumar, R. Henderson, Single particle electron cryomicroscopy: trends, issues and future perspective, Quarterly Reviews of Biophysics 49 (2016) e13: 1–25,
Excellent review on current experimental limitations of electron microscopy and challenges ahead.

[151] **X. Zhang, Z. H. Zhou, Limiting factors in atomic resolution cryo electron microscopy: no simple tricks., Journal of Structural Biology 175 (3) (2011)

253–263,
Excellent review of the theoretical limitations of electron microscopy.

[152] J. C. Yang, M. W. Small, R. V. Grieshaber, R. G. Nuzzo, Recent developments and applications of electron microscopy to heterogeneous catalysis., Chemical Society Reviews 41 (24) (2012) 8179–8194.

[153] **R. R. Schröder, Advances in electron microscopy: A qualitative view of instrumentation development for macromolecuar imaging and tomography, Archives of Biochemistry and Biophysics 581 (2015) 25–38,
Excellent review of the current state of the electron microscope as an optic device.

[154] *R. Henderson, A. Sali, M. L. Baker, B. Carragher, B. Devkota, K. H. Downing, E. H. Egelman, Z. Feng, J. Frank, N. Grigorieff, W. Jiang, S. J. Ludtke, O. Medalia, P. A. Penczek, P. B. Rosenthal, M. G. Rossmann, M. F. Schmid, G. F. Schröder, A. C. Steven, D. L. Stokes, J. D. Westbrook, W. Wriggers, H. Yang, J. Young, H. M. Berman, W. Chiu, G. J. Kleywegt, C. L. Lawson, Outcome of the first electron microscopy validation task force meeting., Structure 20 (2) (2012) 205–214,
Recommendations for structure validation.

[155] L. Passmore, C. J. Russo, Methods in Enzymology. The Resolution Revolution: Recent Advances In cryoEM, Academic Press, 2016, Ch. Specimen preparation for high-resolution cryo-EM, pp. 51–86.

[156] R. F. Thompson, M. Walker, C. A. Siebert, S. P. Muench, N. A. Ranson, An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology., Methods (San Diego, Calif.) 100 (2016) 3–15.

[157] *V. Abrishami, J. Vargas, X. Li, Y. Cheng, R. Marabini, C. O. S. Sorzano, J. M. Carazo, Alignment of direct detection device micrographs using a robust optical flow approach., Journal of Structural Biology 189 (2015)

163–176,
Review on sample preparation for electron microscopy.

[158] J. M. Spear, A. J. Noble, Q. Xie, D. R. Sousa, M. S. Chapman, S. M. Stagg, The influence of frame alignment with dose compensation on the quality of single particle reconstructions., Journal of Structural Biology 192 (2) (2015) 196–203.

[159] N. Fischer, P. Neumann, A. L. Konevega, L. V. Bock, R. Ficner, M. V. Rodnina, H. Stark, Structure of the *E. coli* ribosome-EF-Tu complex at 3a resolution by Cs-corrected cryo-EM, Nature.

[160] T. Grant, N. Grigorieff, Automatic estimation and correction of anisotropic magnification distortion in electron microscopes., Journal of Structural Biology 192 (2) (2015) 204–208.

[161] L. M. Voortman, E. M. Franken, L. J. van Vliet, B. Rieger, Fast, spatially varying CTF correction in TEM., Ultramicroscopy 118 (2012) 26–34.

[162] R. Marabini, B. Carragher, S. Chen, J. Chen, A. Cheng, K. H. Downing, J. Frank, R. A. Grassucci, J. Bernard Heymann, W. Jiang, S. Jonic, H. Y. Liao, S. J. Ludtke, S. Patwari, A. L. Piotrowski, A. Quintana, C. O. S. Sorzano, H. Stahlberg, J. Vargas, N. R. Voss, W. Chiu, J. M. Carazo, Ctf challenge: Result summary., Journal of Structural Biology 190 (3) (2015) 348–359.

[163] R. Danev, W. Baumeister, Cryo-EM single particle analysis with the volta phase plate., eLife 5 (2016) e13046.

[164] A. Punjani, M. A. Brubaker, D. J. Fleet, Building proteins in a day: Efficient 3d molecular structure estimation with electron cryomicroscopy., IEEE Transactions on Pattern Analysis and Machine Learning 39 (2017) 706–718.

[165] A. Dashti, P. Schwander, R. Langlois, R. Fung, W. Li, A. Hosseinizadeh, H. Y. Liao, J. Pallesen, G. Sharma, V. A. Stupina, A. E. Simon, J. D. Din-

man, J. Frank, A. Ourmazd, Trajectories of the ribosome as a brownian nanomachine., Proc. Natl. Acad. Sci. USA 111 (49) (2014) 17492–17497.

[166] C. O. S. Sorzano, J. Vargas, J. M. de la Rosa-Trevín, A. Jiménez-Moreno, R. Melero, M. Martínez, P. Conesa, J. L. Vilas, R. Marabini, J. M. Carazo, High-resolution reconstruction of single particles by electron microscopy, Journal of Structural Biology 204 (2) (2018) 329–337.

[167] J. Velázquez-Muriel, K. Lasker, D. Russel, J. Phillips, B. M. Webb, D. Schneidman-Duhovny, A. Sali, Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images., Proc. Natl. Acad. Sci. USA 109 (46) (2012) 18821–18826.

[168] A. Politis, F. Stengel, Z. Hall, H. Hernández, A. Leitner, T. Walzthoeni, C. V. Robinson, R. Aebersold, A mass spectrometry-based hybrid method for structural modeling of protein complexes., Nature Methods 11 (4) (2014) 403–406.

[169] J. Segura, R. Sanchez-Garcia, D. Tabas-Madrid, J. Cuenca-Alba, C. O. S. Sorzano, J. M. Carazo, 3DIANA: 3D domain interaction analysis: A toolbox for quaternary structure modeling., Biophysical Journal 110 (4) (2016) 766–775.

[170] S. H. W. Scheres, S. Chen, Prevention of overfitting in cryo-EM structure determination., Nature Methods 9 (9) (2012) 853–854.

[171] S. H. W. Scheres, Methods in Enzymology. The Resolution Revolution: Recent Advances In cryoEM, Academic Press, 2016, Ch. Processing of structurally heterogeneous cryo-EM data in RELION, pp. 125–157.

[172] S. J. Ludtke, 3-d structures of macromolecules using single-particle analysis in eman., Methods in Molecular Biology 673 (2010) 157–173.

[173] C. O. S. Sorzano, R. Marabini, J. Velázquez-Muriel, J. R. Bilbao-Castro, S. H. W. Scheres, J. M. Carazo, A. Pascual-Montano, XMIPP: A new

generation of an open-source image processing package for electron microscopy, Journal of Structural Biology 148 (2004) 194–204.

[174] M. van Heel, G. Harauz, E. V. Orlova, R. Schmidt, M. Schatz, A new generation of the IMAGIC image processing system, Journal of Structural Biology 116 (1996) 17–24.

[175] N. Grigorieff, Methods in Enzymology. The Resolution Revolution: Recent Advances In cryoEM, Academic Press, 2016, Ch. Frealign: an exploratory tool for single-particle cryo-EM, pp. 191–226.

[176] C. M. Pablo, G. José, Q. Adrián, de la Rosa Trevín José Miguel, Z. Airén, C. A. Jesús, K. Mohsen, V. Javier, del Cano Laura, S. Joan, S. C. O. S., C. J. María, Scipion web tools: Easy to use cryo-em image processing over the web, Protein Science 27 (1) (2018) 269–275.

[177] R. Marabini, S. J. Ludtke, S. C. Murray, W. Chiu, J. M. de la Rosa-Trevín, A. Patwardhan, J. B. Heymann, J. M. Carazo, The electron microscopy exchange (emx) initiative., Journal of Structural Biology 194 (2) (2016) 156–163.

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

[179] G. J. Kleywegt, S. Velankar, A. Patwardhan, Structural biology data archiving - where we are and what lies ahead, FEBS letters.