BIROn - Birkbeck Institutional Research Online

Pyle, Euan and Zanetti, Giulia (2021) Current data processing strategies for cryo-electron tomography and subtomogram averaging. Biochemical Journal 478 (10), pp. 1827-1845. ISSN 0264-6021.

Downloaded from: https://eprints.bbk.ac.uk/id/eprint/44466/

Usage Guidelines:
Please refer to usage guidelines at https://eprints.bbk.ac.uk/policies.html or alternatively contact lib-eprints@bbk.ac.uk.

Biochemical Journal (2021) **478** 1827–1845
https://doi.org/10.1042/BCJ20200715

Review Article

Current data processing strategies for cryo-electron tomography and subtomogram averaging

Euan Pyle and Giulia Zanetti

Institute of Structural and Molecular Biology, Birkbeck College, Malet St., London WC1E 7HX, U.K.

Correspondence: Giulia Zanetti (g.zanetti@mail.cryst.bbk.ac.uk)

Cryo-electron tomography (cryo-ET) can be used to reconstruct three-dimensional (3D) volumes, or tomograms, from a series of tilted two-dimensional images of biological objects in their near-native states *in situ* or *in vitro*. 3D subvolumes, or subtomograms, containing particles of interest can be extracted from tomograms, aligned, and averaged in a process called subtomogram averaging (STA). STA overcomes the low signal to noise ratio within the individual subtomograms to generate structures of the particle(s) of interest. In recent years, cryo-ET with STA has increasingly been capable of reaching sub-nanometer resolution due to improvements in microscope hardware and data processing strategies. There has also been an increase in the number and quality of software packages available to process cryo-ET data with STA. In this review, we describe and assess the data processing strategies available for cryo-ET data and highlight the recent software developments which have enabled the extraction of high-resolution information from cryo-ET datasets.

Introduction

Recent developments in cryo-electron microscopy (EM) have enabled the determination of structures of macromolecules at atomic resolution using single-particle analysis (SPA) [1,2]. Obtaining atomic and near-atomic resolution enhances our understanding of the precise mechanisms of protein function and could facilitate structure-based drug design, if sufficient resolution can be reached on clinically relevant drug targets. The ability of EM to image molecules previously intractable by other structural biology techniques, combined with improvements in sample preparation [3,4], data collection [1,5], and processing software [6–8], has dramatically improved our understanding of many classes of proteins such as TRP channels, class-D GPCRs, and bacterial secretion systems [9–12].

Currently, SPA is the favoured cryo-EM method to obtain high-resolution protein structures. However, SPA requires that the protein(s) of interest be purified to a high compositional and conformational homogeneity, that the sample lies in a range of different orientations on the grid, and that individual particles do not overlap with one another. Many proteins do not meet these requirements and thus are not suitable for SPA. Consequently, various studies instead utilise cryo-electron tomography (ET). Cryo-ET utilises a series of tilted two-dimensional (2D) projection images to reconstruct a three-dimensional (3D) representation (or cryo-tomogram) of the sample of interest. This allows the separation of molecules that are stacked or overlapping along the projection view. A cryo-tomogram represents a unique object which can be directly interpreted subject to high level of noise. However, for protein structure determination, individual particles (or subtomograms) can be extracted from the tomogram, aligned to a common reference, and averaged together to generate a structure in a process known as subtomogram averaging (STA). STA improves the low signal to noise ratio (SNR) and has been increasingly used to generate subnanometre resolution structures [13–15]. One distinct advantage of Cryo-ET over SPA is that it can be used to image proteins *in situ* and can, therefore, be used to answer biological questions regarding a protein's native structure and interactions. Cryo-ET and SPA can give insights into structures of proteins embedded in their native cellular environment, also in

*Giulia Zanetti was the recipient of the Biochemical Society's 2021 Colworth Medal.

Received: 26 February 2021
Revised: 19 April 2021
Accepted: 26 April 2021

Version of Record published: 18 May 2021

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Biochemical Journal (2021) 478 1827–1845
https://doi.org/10.1042/BCJ20200715

combination with techniques such as focussed ion beam (FIB) milling for thick samples [16–18]. In addition to obtaining more physiologically relevant structures, the co-ordinates of particles can be annotated into a tomogram so that their interactions with the environment and their geometrical relationship to one another can be visualised.

A brief overview of the steps in a typical cryo-ET and STA data processing pipeline is described in Figure 1; however, it should be noted that this workflow can vary depending on the protein of interest. Additionally, several of these steps are iterative and may be also be repeated and refined at later points in the pipeline. A list of software packages available for each step in this workflow is presented in Table 1 and Table 2.

Cryo-ET with STA has increasingly been able to derive subnanometer resolution structures of a range of proteins and complexes such as the COPI, COPII, and clathrin coats [19–22], viruses including HIV and SARS-CoV-2 [23–25], and protein conducting channels [26]. However, cryo-ET with STA does not routinely reach the same resolution as SPA due to several factors. Firstly, the SNR of each image is much lower in cryo-ET than SPA. This is due to the need to image the same particle multiple times during the acquisition of a tilt-series which also requires a reduction the electron dose for each tilt image in comparison with SPA. Additionally, the contrast of the image is reduced at high tilt angles due to increased sample thickness. The low SNR can cause resolution-limiting issues at several points within the STA pipeline such as difficulties in CTF estimation. Secondly, the measured position and the tilt angle of the grid during the acquisition of the tilt-series may be inaccurate due to mechanical instability in the microscope. Tilt-series alignment (TSA) is carried out to detect and correct for inaccuracies in grid position and tilt angle. The TSA must be sufficiently accurate to preserve high-resolution information after tomogram reconstruction. Any inaccuracies will cause the loss of high-resolution information during the back-projection of the individual images to form the tomogram. Finally, it is typically only possible to tilt the grid at a maximum of ±60° before the sample becomes too thick to image effectively due to reduced contrast or to other mechanical impediments, such as the grid support bars. Therefore, during the reconstruction of the tomogram, there are a range of missing views causing resolution anisotropy and the smearing of the tomogram along the direction perpendicular to the tilt axis. This is known as the 'missing-wedge' effect. While the anisotropy can be overcome by averaging a sufficient number of particles in each tomogram in a range of different orientations, the missing-wedge can cause inaccuracies in the alignments which ultimately limit the achievable resolution.

Some of the issues limiting the resolution of cryo-ET and STA can be partially solved by advances in data collection protocols. For example, radiation damage over the tilt-series can be distributed in a way which maximises the SNR within the tomogram by using a dose-symmetric tilt scheme [27]. However, some of these problems, such as the missing-wedge effect, cannot currently be resolved by improvements to data collection protocols and are instead being addressed with data processing protocols. This review describes and assesses the various options available for processing cryo-ET data with STA. We also analyse the recent developments in software packages for cryo-ET with STA. We describe how these software packages maximise the quality of the tomograms, and/or structure(s) at each stage along the cryo-ET with STA pipeline. We also discuss the issues that continue to negatively impact the resolution of the structures generated by the cryo-ET with STA workflow.

Reference-free motion correction

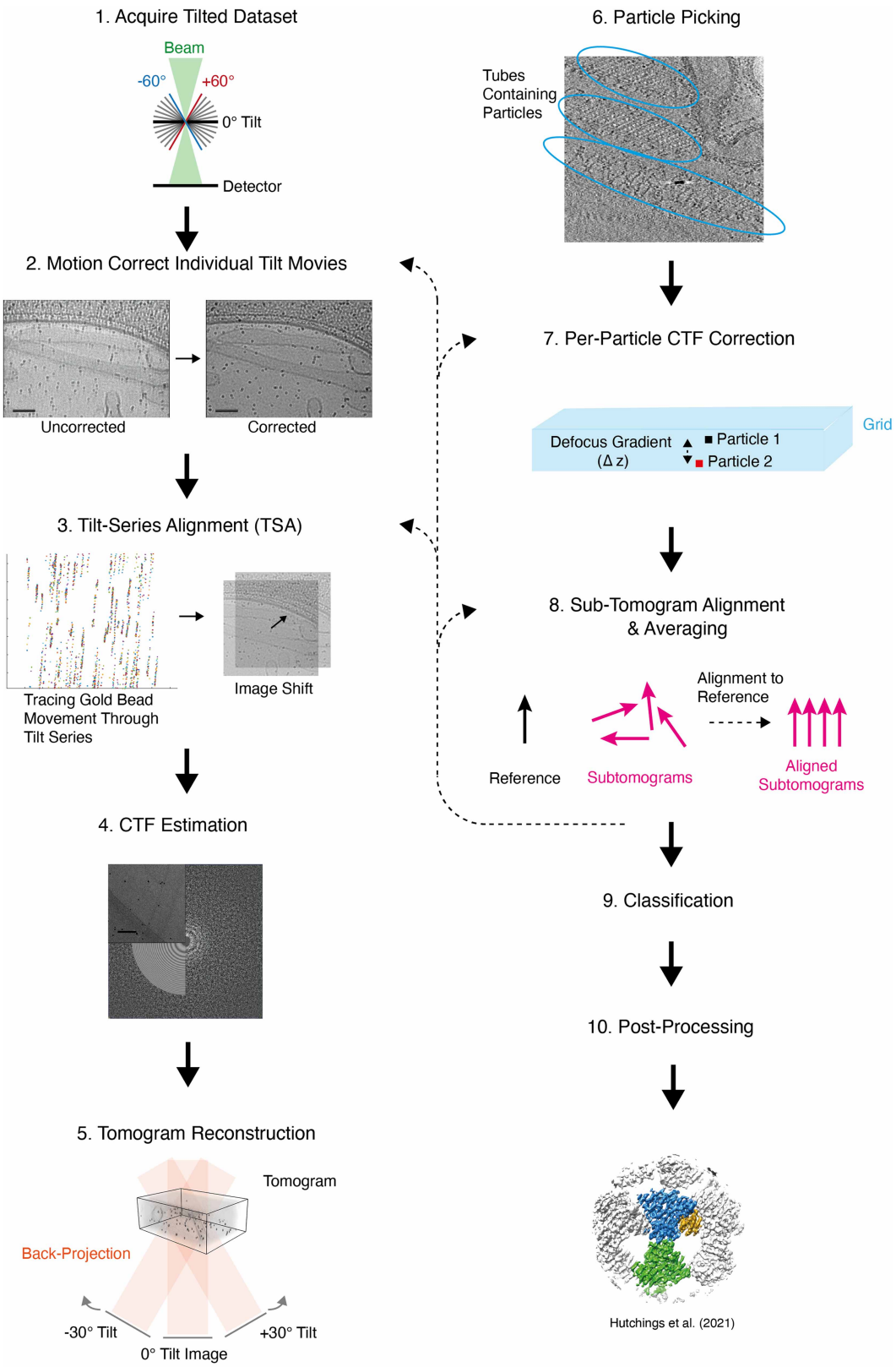
Direct electron detector cameras have enabled the acquisition of images recorded as a stack of dose-fractionated frames or 'movies'. Over the course of each movie, the images are blurred by the motion of particles within the grid causing the loss of high-resolution information. Several factors influence the change in the positions and orientations of particles during the recording of a movie including mechanical stage drift, Brownian motion, molecular vibration, and beam-induced motion (BIM) such as bending of the support foil and ice [4,28,29]. Mechanical stage drift causes uniform motion of the sample and can be corrected by software which tracks global translational shift between frames of the movie. In contrast, BIM can cause anisotropic, global, and locally correlated movement of particles. For each movie and corresponding 2D image within a tilt-series, much of the particle motion can be detected and corrected for by reference-free motion correction software.

There are several software packages which carry out motion correction of global movements including MotionCorr and UnBlur [30,31]. MotionCorr measures image shifts within the frames of the movie and generates a least-squares estimate of the movements between each movie frame. Whilst this deals well with global translations, it is harder to correct for locally correlated motions of particles induced by BIM. Several software packages, such as MotionCor2, additionally corrects for local non-uniform motions by dividing the movie frames up into

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Biochemical Journal (2021) **478** 1827–1845
https://doi.org/10.1042/BCJ20200715



Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Biochemical Journal (2021) **478** 1827–1845
https://doi.org/10.1042/BCJ20200715

Figure 1. Overview of a typical cryo-ET with STA data processing workflow. Part 2 of 2

Note that several steps can be iteratively refined and/or returned to at a later point in the workflow. The steps that are most commonly iterated are indicated by dotted arrows. The order in which these steps are carried out may vary. (1) During data collection, 2D movies of the sample are captured at fixed points over a range of tilt angles (typically ±60°). (2) Each individual movie is motion-corrected, dose-weighted, and averaged to form a series of 2D images hereafter referred to as a tilt-series. (3) The position of the grid in each image is refined by tracing the position of fiducial markers (usually gold beads) throughout the tilt-series (left) and transforming individual images (right) to ensure a smooth fiducial marker trajectory. This process is known as tilt-series alignment (TSA). (4) Contrast transfer function (CTF) estimation is then carried out for each 2D image in the tilt-series by fitting a CTF model (bottom left) to the Fourier transform (right) of the real image (top left). (5) The 2D images in the tilt-series are backprojected to form a 3D volume, otherwise known as a tomogram. (6) Particles of the protein of interest are picked within the tomogram. Small 3D volumes, or subtomograms, containing one or more particle(s) are then extracted. (7) CTF correction can be carried out at this point or elsewhere in the workflow. (8) The subtomograms are aligned to a reference and averaged. This average is used as a new reference until the alignment reaches convergence and a structure with maximal resolution is generated. The resolution of the resulting structure can be assessed by 'gold-standard' Fourier shell correlation (FSC), spectral SNR methods, or local resolution estimations. (9) Conformational and compositional heterogeneity can be assessed by classification of subtomograms after or during STA. This may lead to the generation of multiple structures with different compositions or conformations. (10) Optionally, a range of post-processing techniques can then be used to improve the quality of the structure such as sharpening. The data presented here was generated by Hutchings et al. [19,20].

tiles and tracking the correlated movements within these subdivisions. Other motion correction software that can account for local motion caused by BIM include alignparts_lmbfgs, Xmipp, and Zorro [32–35].

Major advances in the motion correction implementation have occurred in the years after the introduction of direct electron detectors. More recently, some software packages have integrated incremental improvements in their motion correction algorithms for better accuracy or speed. For example, MotionCor2 v1.4.0 has been developed to improve local motion correction at higher tilt angles [36]. Additionally, Warp is a software package released by Tegunov et al. [37] which carries out all the pre-processing steps in the tomography pipeline, including motion correction. Warp utilises a similar motion correction strategy to MotionCor2 [32,37].

Reference-free motion correction software is often sufficiently accurate at correcting local particle motion in single-particle datasets [32]. However, in tomography, the accuracy and precision of reference-free motion correction is sometimes limited by the low SNR of the images. This issue is more pronounced at higher tilt angles. Furthermore, reference-free motion correction software cannot account for the grid deformations that build up with the acquisition of each subsequent movie within a tilt-series.

Reference-based motion correction

Reference-based motion correction methodologies have been developed to supplement reference-free motion correction for both SPA and cryo-ET/STA [38–40]. In reference-based motion correction, a 3D reference of the particle of interest is used to generate high SNR 2D projections. In STA, the reference used is typically the subtomogram average generated by the first round of alignment and averaging (see subtomogram averaging). The 2D reference projections are then used to track and correct for the movement of individual particles within low SNR tilt-series frames. As reference-based motion correction conceptually overlaps with the refinement of particle positions on a per-tilt basis it will be discussed below (see per-particle tilt refinement).

Tilt-series alignment: manual or automated?

TSA is necessary prior to tomogram reconstruction due to inaccuracies in image tracking during data collection. TSA is carried out by detecting and tracking the movement either of high contrast fiducial markers such as gold beads (fiducial-based alignment) or of features in the images such as organelles or membranes (patch tracking or fiducial-less alignment) between tilt images. This data is then used to identify and correct for any inaccuracies in the calculated tilt angle, magnification, or in-plane rotation. Fiducial-based alignment is more accurate and is, therefore, more commonly used when aiming to generate high-resolution structures. The position of the gold beads within each tilt image can be input manually by the user, which represents a time-consuming step sometimes taking multiple weeks for large datasets. However, it is also possible to detect the position of gold beads using automated fiducial marker detection algorithms, which shorten the time taken for this step but may compromise TSA accuracy.

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Table 1. Partial list of software used prior to tomogram reconstruction in cryo-ET with STA

	Motion correction	Tilt-series alignment	CTF estimation	Tomogram reconstruction
IMOD [40–42]		✓	✓	✓
Warp [36]	✓		✓	
EMAN2 [44]		✓	✓	✓
Dynamo [46]		✓		
RELION [76]			✓	
emClarity [53]			✓	✓
MotionCorr [30]	✓			
MotionCor2 [32]	✓			
UnBlur [31]	✓			
Alignparts_lmbfgs [33]	✓			
Zorro [34]	✓			
Xmipp [35]	✓			
RAPTOR [43]		✓		
AuTom [45]		✓		
UCSF tomo [47]		✓		
Protomo [48]		✓		
TomoAlign [51] [52]		✓		
CTFFIND4 [70]			✓	
Gctf [71]			✓	

A commonly used tool for TSA is the IMOD software package [41,42]. IMOD allows both manual input of fiducial marker positions and automated marker tracking [43]. Other software can also be integrated with IMOD to carry out TSA using automated marker tracking such as RAPTOR [44]. Several software packages, including Warp and emClarity, integrate IMOD into their workflows for TSA. EMAN2, AuTom, Dynamo, Protomo, and UCSF tomo can be used as alternatives to IMOD for automated TSA using fiducial markers [45–48] or marker-free patch-based alignment [45,46,49].

Many recent improvements have focused on improving the accuracy of automatic detection and tracking of markers so automated TSA can reach the same level of accuracy as time-consuming manual TSA. Automated marker tracking in IMOD was recently improved by implementing several adjustments including the application of a Sobel filter to reduce alignment error, introducing robust fitting to disregard any outliers from the alignment, and by using only the highest-quality markers when an excess of markers is available [43]. Despite these developments, automated TSA sometimes fails [43]. In this case, it is possible to resolve the problem by switching back to manually aligning the tilt images for individual tomograms. Additionally, it is sometimes necessary to remove 'bad images' from tilt-series and the appropriate parameters for marker detection may require optimisation thus making the process not fully automatic [43]. AuTom can also perform automated TSA based on the principles presented by Han et al. [50]. AuTom is reported to have a lower failure rate during TSA compared with IMOD and thus require less manual intervention. The EMAN2 software package

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Biochemical Journal (2021) **478** 1827–1845
https://doi.org/10.1042/BCJ20200715

Table 2. Partial list of software used after tomogram reconstruction in cryo-ET with STA

	Particle picking	CTF correction	Alignment and averaging	Per-particle refinement	Classification
IMOD [40–42]	✓				
Warp [36]	✓	✓			
EMAN2 [44]	✓	✓	✓	✓	✓
Dynamo [46]	✓		✓		✓
RELION [76]		✓	✓		✓
emClarity [53]	✓	✓	✓	✓	✓
M [54]				✓	
pyTOM [60]	✓		✓		
Chimera (Pick Particle) [62,63]	✓				
DeepFinder [64]	✓				
Xmipp Deep Consensus [65]	✓				
SPHIRE-crYOLO [66]	✓				
NovaCTF [74]		✓			
PEET [83]			✓		
AV3 [85]			✓		
Jsubtomo [86]			✓		

also includes recent developments in automated patch tracking TSA [45]. EMAN2 uses an iterative approach to TSA which starts with highly binned images. Here, the software selects identifiable 'landmarks' from within the tilt-series and refines the TSA. It then generates a binned tomogram with relatively large alignment errors. 2D projections of the 3D landmarks within the tomogram are then used to align the next iteration of TSA, upon which the binning of the tilt images is reduced. As this process is iterated, EMAN2 refines the alignment further and removes outliers. The focus of the TSA in EMAN2 seems to prioritise minimising manual intervention and the increase in data throughput. Chen et al. [45] claim EMAN2 was able to automatically align tomograms where IMOD had previously failed.

Manual input of fiducial marker position remains the most common TSA method for high-resolution STA. Fully automated TSA has not yet become standard practice perhaps due to concerns of resolution-limiting inaccuracies in the detection and tracking of the fiducial markers. However, Chen et al. [45] do not consider the current accuracy of the automated TSA within EMAN2 to be resolution limiting. Additionally, automated TSA has been used in a recent 4.6 Å resolution structure of the COPII inner coat and in a 3.4 Å resolution structure of HIV-1 CA-SP1 hexamer [20,51]. For the COPII inner coat structure, automated TSA routines implemented within the Dynamo software package were coupled with some in-house scripts to minimise the residual error in each tomogram. Furthermore, the accuracy of the initial TSA may be less essential to obtaining high-resolution structures due to recent advances in TSA refinement. Therefore, it is likely we will see increased usage of automated TSA protocols for high-resolution structures in the near future, especially when used in conjunction with TSA refinement tools. Additionally, manual TSA will become a less viable option as data collection for cryo-ET becomes higher throughput and users need to analyse higher numbers of tomograms.

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Biochemical Journal (2021) 478 1827–1845
https://doi.org/10.1042/BCJ20200715

Tilt-series alignment refinement

Improving the accuracy of the TSA can increase the resolution of maps generated by cryo-ET with STA. TSA refinement can be carried out using either data from later points in the STA workflow or by integrating further steps immediately after the initial TSA. After the initial TSA, TomoAlign builds a model of the deformation of the sample throughout the tilt-series due to BIM. Here, fiducial markers are used to track the deformation of the grid throughout the tilt-series and this information is used to refine the TSA parameters [52,53]. Several software packages, such as emClarity and M, provide the user with the opportunity to refine the TSA at the level of individual tilts based on structure(s) generated by STA [54,55]. Due to conceptual overlap we will further discuss TSA refinement in the per-particle tilt refinement section.

Tomogram reconstruction

After TSA, the individual tilt images are used to produce a 3D reconstruction of the sample. Several methods of tomogram reconstruction can be used. For data being processed by STA, tomogram reconstruction is typically carried out by weighted backprojections (WBP) [56]. Here, each image in a tilt-series is weighted to balance spatial frequencies and then backprojected into a real space volume. Alternatively, the simultaneous iterative reconstruction technique (SIRT) and the similar Simultaneous Algebraic Reconstruction Technique (SART) [57,58] can be used for tomogram reconstruction. In both SIRT and SART, an algorithm is used to optimise the predicted projections of a tomogram and its corresponding tilt images. These methods are aimed at increasing the SNR for direct interpretation of the tomogram compared with WBP. The increased contrast assists processes such as subtomogram alignment and particle picking. However, both techniques lose high frequencies that are determined to have a low SNR. Consequently, high-resolution information is lost in SIRT and SART reconstructions. Therefore, WBP, which preserves high-resolution information, is preferred for generating high-resolution structures by STA.

During tomogram reconstruction, filters can be applied to generate better contrast and lower noise within the tomogram and CTF correction can be performed (see CTF correction). Noisy high tilt images can be low-pass filtered to limit the noise these images contribute to the tomogram. Images can also be filtered according to their accumulated dose over the tilt-series [31].

Particle picking

In STA, multiple copies of the protein of interest are present within each tomogram. After tomogram reconstruction, it is necessary to locate the positions of the particles within the 3D volume. After the location of the particles has been identified, small 3D subvolumes of each individual particle (or subtomograms) within the tomogram are computationally extracted in preparation for subtomogram alignment and averaging. To generate the highest possible resolution structure from the available data, it is essential to locate as many 'real' particles of the protein of interest as possible and to minimise the selection of noise or other molecules as false-positive particles. Alignment of insufficient numbers of 'real' particles and excessive numbers of false-positive particles limits the reliability of the structure generated by STA. Cryo-ET data presents additional difficulties in particle picking compared with SPA. For example, *in situ* tomograms are compositionally heterogenous making the protein of interest more difficult to locate. Furthermore, low SNR combined with the missing-wedge effect makes small particles difficult to accurately identify. There are several widely used methods for particle picking in the cryo-ET workflow including manual picking, template matching, geometric surface picking, and machine-learning-based picking.

Several software packages allow manual particle picking including Dynamo, IMOD, and EMAN2 [41,45,47,59]. This involves the user visually inspecting the entire tomogram and selecting positions which they identify as their protein of interest. This requires some *a priori* information on the general shape or location of the protein they wish to investigate. In Dynamo, once multiple particles have been selected, it is possible for the user to manually align the particles which can be useful in generating an initial reference for subtomogram alignment and averaging [47]. However, manual particle picking can be challenging due to noise and to the difficulty of recognising particles in the 3D tomogram on the computer screen. Binning, denoising, and filtering the tomograms can make particles more easily identifiable for manual particle picking. The obvious and major disadvantage of manual particle picking is that it is extremely time-consuming. As a result, manual picking is usually only used to generate an initial reference for template matching and/or subtomogram alignment. However, for some samples other particle picking techniques may be ineffective, leaving manual picking as the only suitable option.

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Biochemical Journal (2021) **478** 1827–1845
https://doi.org/10.1042/BCJ20200715

Template matching is an automated method of particle picking which is incorporated into multiple software packages including Dynamo, Warp, pyTom, emClarity, and EMAN2 [37,45,47,54,59–61]. Initially, a template volume must be generated. To manually generate a template volume, a limited number of particles should be manually picked, extracted into subtomograms, aligned, and averaged to produce a low-resolution reference. It is also possible to use a pre-existing structure as a template volume, but care must be taken to avoid reference bias, for example by heavily low-pass filtering the reference [62]. Once a template is available it is used to find particles by cross-correlation. Template matching is a very common method of particle picking as it is a mostly automatic process and it is, therefore, faster than manual particle picking. However, anisotropic resolution due to the missing-wedge effect can limit the identification of particles in some orientations. Furthermore, for complex and compositionally heterogenous tomograms, such as *in situ* data, template matching may produce excessive false positives.

Geometry-based particle picking is a semi-automated method included in Dynamo and the Pick Particle Chimera Plugin [60,63,64]. Here, particle picking is assisted using distinctive geometric features within the tomogram and *a priori* information about the protein of interest. For example, membranes are generally easy to identify within a tomogram and if the protein of interest is known to be membrane associated it is possible to limit particle picking to the surface of the membrane. This has been applied to several biological structures such as virions [23] and coated membrane tubules [19]. Dynamo can be used to annotate a range of geometric shapes including cylinders and spheres within a tomogram [60]. Additionally, surfaces can be manually traced and segmented. The position of the particle relative to the geometric surfaces can also be used to help the initial alignment of the particles. Geometry-based particles picking is particularly useful in complex data such as *in situ* tomograms. It limits the number of false positives by incorporating *a priori* information about particle location. However, similar to template matching, this approach may introduce reference bias if the user is not careful. Moreover, this process is currently only semi-automated as the user has to manually identify the position of the geometric shapes within each tomogram.

Recently, multiple software packages have incorporated machine-learning and deep-learning capabilities into the particle picking process including DeepFinder, Xmipp Deep Consensus, SPHIRE-crYOLO, EMAN2, and Warp [37,65–67]. Most of these software packages have been tailored to SPA datasets; however, some are beginning to incorporate compatibility for cryo-tomograms and *in situ* feature recognition. EMAN2 allows multiple methods for particle picking, including one using neural networks to identify and annotate *in situ* features such as vesicles and microtubules [45,68]. This process does require the user to train the neural network. Here, the user must outline and annotate the feature of interest, such as the endoplasmic reticulum membrane, as well as regions not containing the feature of interest. The neural network can then find and annotate the feature throughout multiple tilt-series and tomograms. Subtomograms can then be extracted from the identified regions. The neural network picking in EMAN2 was able to identify isolated ribosome particles within *in situ* tomograms. These particles could be extracted as subtomograms and were used to generated low-resolution structures via STA [68]. As this process is near-fully automated, it can be carried out with greater speed than the semi-automated geometry-based particle picking. This approach is especially advantageous for *in situ* data where it can take an expert 1 week to fully annotate a tomogram [68]. In contrast, the neural network-based approach takes 2.5 h to annotate a feature on a $2000 \times 2000 \times 500$ voxel tomogram on the workstation used by Chen et al. [68]. Further binning can reduce processing time further. The reported particle picking accuracy of this technique for individual ribosomes varied but was reported to reach as high as 91% [68]. Chen et al. [68] do note some current limitations of the software. For example, gold fiducial markers, carbon edges and other high contrast objects that weren't included in the neural network training data can cause issues with feature identification [68].

DeepFinder is another deep-learning-based particle picking software for cryo-ET datasets which is included in the Scipion package [65]. Similar to EMAN2, DeepFinder requires a neural network to be trained before finding features across multiple tomograms. During benchmarking, DeepFinder outperformed other particle picking approaches [65]. Other software packages are also beginning to incorporate machine-learning particle picking approaches for tomography datasets. For example, the latest release of the machine-learning-based crYOLO (v1.8) has extended support to include cryo-ET data [67]. This feature is currently in the beta phase of testing.

CTF correction

CTF estimation and correction is carried out in SPA by modelling phase and amplitude oscillations in the Fourier transform of the images. This model can be used to correct the CTF and restore the signal. Accurate CTF correction is essential to recover high-resolution information and generate interpretable structures using cryo-ET with STA [69,70].

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

The CTF is more challenging to detect and estimate in cryo-ET than in SPA due to the low SNR of individual images, especially at high tilts and high sample thickness, as the Thon rings are less prominent and therefore harder to accurately model. Furthermore, the CTF is modulated by defocus and the tilting of the sample causes a defocus gradient across images. Each image contains an average of all the CTFs across this defocus gradient which dampens the CTF, especially at higher frequencies, which makes accurate detection and estimation more challenging. Inaccurate CTF estimation can lead to the loss of high-resolution information. Several software packages can be used for initial CTF detection and estimation including CTFFIND4, CTFPlotter in IMOD, and gctf [41,71,72]. To improve CTF estimation in cryo-ET, the sensitivity of CTF estimation software must be increased to overcome the low SNR of the individual tilt images.

CTF estimation in cryo-ET is complicated by the defocus gradient across each tilted image. For thin samples, a simple tile-based method is often used to account for the defocus gradient caused by tilting. Here, the defocus can be more accurately estimated by splitting each image in the tilt-series into strips and assigning a different defocus value to each strip [73]. However, for thicker samples, the low SNR in small tiles can cause resolution-limiting errors. Here, the average CTF from each image is estimated and assumed to represent the defocus at the tilt axis.

In both SPA and cryo-ET, particles lie at different heights within the ice. This affects the defocus, and therefore the CTF, of each particle within each image. One advantage of cryo-ET is that it is possible to determine the height, and therefore the defocus, of each particle due to the construction of a 3D volume of the sample. This allows the performance of CTF correction adjusted for particle height, which can be done over the whole tomogram before STA, or at the level of extracted subtomograms during STA.

Additionally, it is important that the handedness of the system be determined during CTF correction. Most CTF correction software packages make the default assumption that, upon rotating images so that the tilt axis is vertical, their right side is more defocussed for positive tilt angles. If this is not the case, the handedness of the system must be inverted otherwise CTF correction will be inaccurate (Figure 2). This can be done either through a flag in the CTF correction software or by manually rotating tilt images before reconstruction [41,45,74].

NovaCTF is a software package which accounts for sample thickness and different particle heights in CTF correction during tomogram reconstruction [75]. This is commonly known as 3D-CTF correction (Figure 2) [76]. NovaCTF carries out CTF correction for the whole tomogram. 3D-CTF correction can also be done on a per-particle basis where each subtomogram has its CTF corrected individually. RELION carries out per-particle CTF correction where each subtomogram is CTF corrected rather than the whole tomogram [77]. RELION uses the particle co-ordinates to generate a tilt-series of 2D images for each individual particle. The CTF can then be corrected based on the previous defocus calculations, the particle height in the tomogram, and the tilt-series geometry. To avoid interpolation issues, emClarity uses an alternative per-particle 3D-CTF correction approach where the entire tomogram is corrected for the phase of the CTF before tomogram reconstruction but the amplitudes are corrected for each subtomogram after tomogram reconstruction. The CTF correction parameters can be updated after TSA refinement in emClarity (see per-particle tilt refinement) [54].

Warp can also be used for CTF correction of tomography data. Warp models the CTF in image patches which take into account image tilt and local geometry [37]. M can be used in conjunction with Warp to refine CTF parameters [55]. Furthermore, M automatically adjusts the box size of the particles to a sufficiently large size to allow aliasing-free CTF correction which is particularly useful for preserving high-resolution information in particles which have a high defocus. Additionally, M detects and corrects for optical aberrations within the data such as astigmatism, anisotropic pixel size, and higher-order aberrations. The improvement in resolution from correcting these aberrations has previously been established in SPA pipelines [78–80]. The combination of CTF-related features in the Warp/M packages show impressive improvements in obtainable resolution [37,55].

A new software package, CARYON, utilises the comparative information in individual movie frames to obtain more accurate CTF parameters, with the potential to significantly impact our ability to extract high-resolution information from cryo-tomograms [81].

Subtomogram averaging

Alignment and averaging of subtomograms produces structures of the complex of interest which have improved SNR compared with the individual particles in the tomogram, and more isotropic resolution (Figure 3). The resolution of the structure generated by STA is affected by the precision and accuracy of the alignment of the subtomograms. The subtomogram alignment and averaging process begins with the alignment of the

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Biochemical Journal (2021) **478** 1827–1845
https://doi.org/10.1042/BCJ20200715

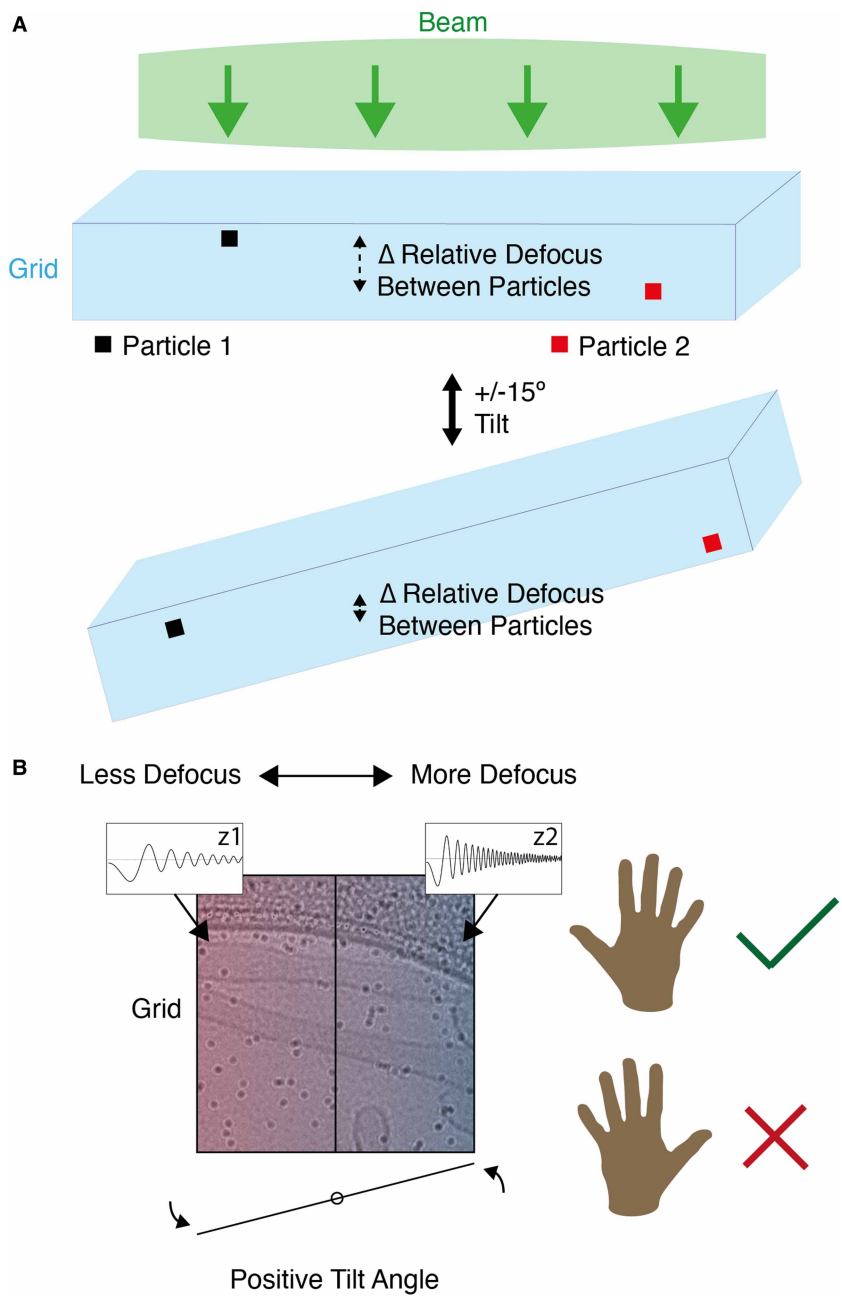


Figure 2. The contrast transfer function in cryo-tomography.

(**a**) The defocus of individual particles depends on Z-height in the ice. The Z-height of individual particles is used as part of 3D-CTF correction. Tilting the grid changes the relative defocus of particles at different heights in the ice. (**b**) The defocus gradient across the grid can be used to assess whether the correct handedness is being used. At positive tilt angles, the right side of the grid should be more defocused if the handedness is correct. Z1 shows the defocus of a particle on one side of the grid and its corresponding CTF. Z2 shows the defocus of a particle on the other side of the grid and its corresponding CTF.

subtomograms to a common reference (Figure 3a). Similar to template-based methods in particle picking (see particle picking), this reference can be manually derived by hand-picking particles, assigning an approximate orientation, aligning the particles accordingly, and calculating their average. Alternatively, the reference can be generated by using a homologous structure to the protein of interest which is low-pass filtered to avoid

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Biochemical Journal (2021) **478** 1827–1845
https://doi.org/10.1042/BCJ20200715

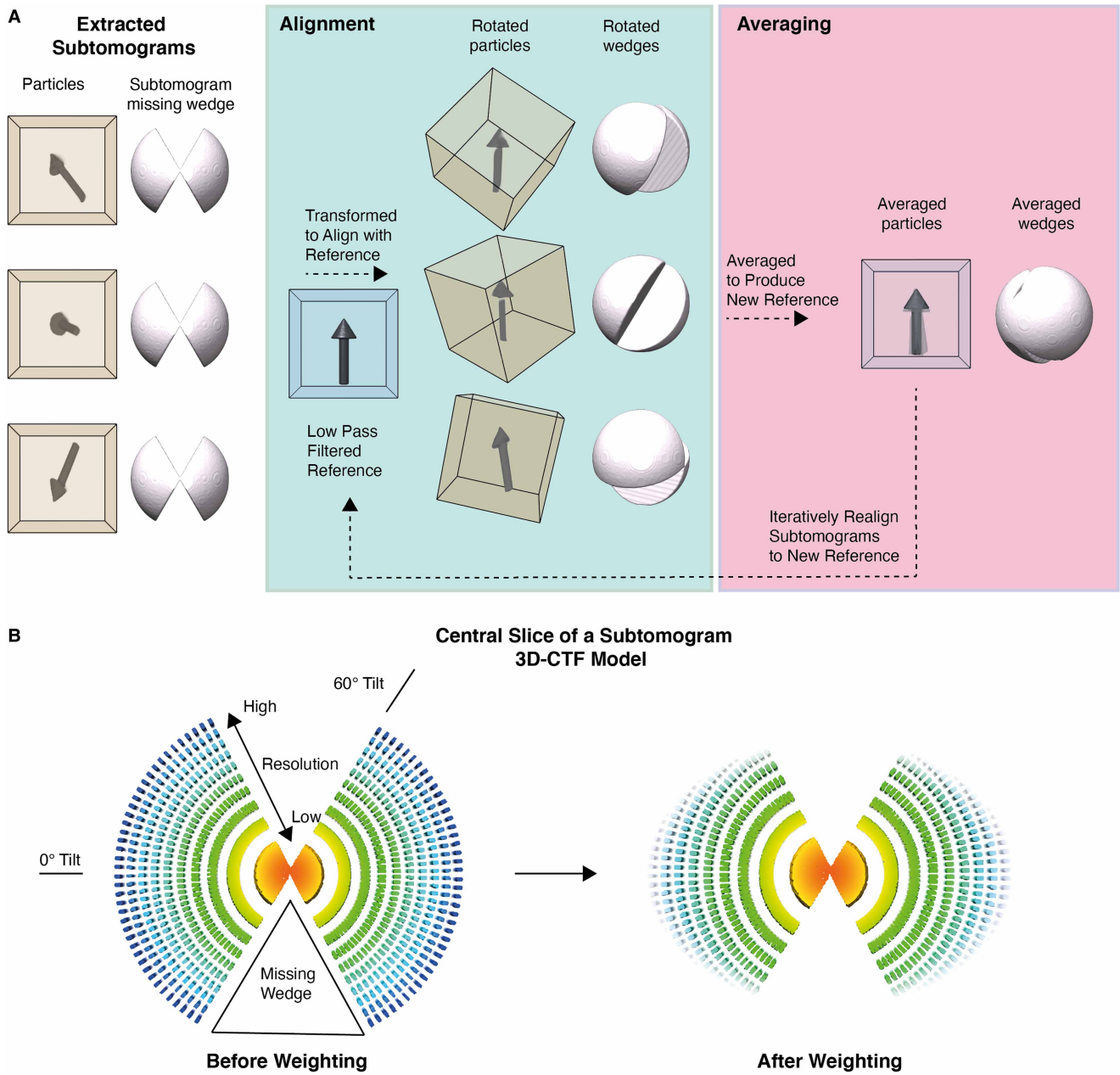


Figure 3. The general process behind subtomogram alignment and averaging.
(**a**) The cubes represent randomly orientated extracted subtomograms and the arrows show the orientation of the particles within the subtomograms relative to one another. Each particle has its corresponding missing-wedge representation to the right. (**b**) Visual representation of the 3D-CTF model of an individual subtomogram before (left) and after the high-resolution information loss at high tilts is taken into account by weighting (right). Note the blank areas corresponding to the missing-wedge and the loss of signal at high resolution at higher tilts.

overfitting. The reference structure is then rotated in space and aligned to each subtomogram by cross-correlation or maximum-likelihood (ML) methods. The relative orientation of each subtomogram to the reference is stored as metadata, and used to rotate the subtomograms, which are then averaged to generate a new reference (Figure 3a). If alignment was successful, the new reference should have a better SNR than the previous reference which allows more accurate subtomogram alignment. This reference is then used in the next iteration. This alignment and the averaging process is repeated until the reference generated by the averaging step does not significantly improve on the reference used in the previous subtomogram alignment step, otherwise

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Biochemical Journal (2021) 478 1827–1845
https://doi.org/10.1042/BCJ20200715

known as convergence. Throughout iterations, parameters such as binning and angular sampling are refined to optimise alignments and processing times.

To avoid overfitting, gold-standard procedures must be used during subtomogram alignment [82]. Prior to initial alignment, the data is split into two halves. The Fourier shell correlation (FSC) between the references generated by each half dataset can be used to estimate the resolution of the references. The estimated resolution of each reference can be used to set an appropriate low-pass filter during alignment, and to assess convergence. Each subtomogram contains a missing-wedge of information which must be taken into account during the alignment process. The effects of the missing-wedge can be partially compensated for by applying a wedge-shaped mask to the reference in Fourier space so the reference has a missing-wedge in a similar position to the subtomogram [83].

A range of software packages can be used to carry out subtomogram alignment and averaging including PEET, EMAN2, RELION, Dynamo, Jsubtomo, AV3/PyTOM, and emClarity [45,47,54,61,84–87]. Recent advances in subtomogram alignment software have focused on improving the missing-wedge compensation in the reference. Most software packages currently use the standard binary wedge-mask method; however, EMAN2, RELION, and emClarity all utilise more advanced methods during missing-wedge compensation which account for the CTF and the SNR in each image over a tilt-series [45,54,85].

In RELION, a 3D model of the CTF of each subtomogram is generated by estimating the 2D CTFs of each image in their tilt-series and then combining this information with the known tilt angles of each image and the Z-height of the subtomogram (Figure 3b). The 3D CTF model describes the missing-wedge in 3D Fourier space (Figure 3b). RELION additionally applies weighting to the 3D CTF model which accounts for the decrease in the SNR of high tilts and the cumulative radiation damage during data collection (Figure 3b). This decreases the contribution of high-resolution Fourier components at high tilts to avoid using noise in subtomogram alignment [85]. Likewise, emClarity iteratively generates a '3D sampling function' for each subtomogram which is similar to the weighted 3D CTF model in RELION and accounts for directional anisotropy. To avoid interpolation between the positive and negative signal at low frequencies, emClarity corrects for the phase of the CTF during tomogram reconstruction and corrects for the amplitude of the CTF during subtomogram alignment.

Warp exports subtomograms alongside 3D CTF volumes indicating the missing-wedge and tilt-dependent weighting information [37]. Warp and M do not carry out STA; however, their workflow has been designed so that RELION can easily be integrated to carry out this step [88]. M allows the user to re-extract and further improve the alignment of particles after its refinement processes. Tools have recently been developed to integrate Warp and M with Dynamo [47,51].

Per-particle tilt refinement

Once subtomogram alignment and averaging has produced a good-quality map of the protein of interest, it is possible to use this structure to generate high SNR 2D reference projections to refine the position and orientation of particles within the data on a per-tilt basis. This information can be used to refine and optimise several steps previously carried out in the workflow such as motion correction and TSA. The refinement of the parameters within these steps can enable better subtomogram alignment and an improved quality map. Several software packages have recently incorporated per-particle tilt refinement within their STA workflow including M, emClarity, and EMAN2.

The emClarity software package includes a process called tomogram constrained-particle refinement (tomoCPR) which can be used to refine the TSA on a per-tilt basis. Whilst initially emClarity utilises IMOD to generate the TSA parameters and tomograms, tomoCPR is carried out after the initial subtomogram alignment. TomoCPR uses 2D reference projections to identify particles within the individual tilt frames and registers the movements of spatially proximal particles which are artificially clustered together in patches containing a fixed number of particles. Individual clusters of particles within these patches should experience the effects of BIM similarly and emClarity uses this information to smoothen the TSA for groups of neighbouring particles. TomoCPR is also an iterative process which can be refined after each iteration of STA. The improved TSA from tomoCPR allows emClarity to further improve and refine 3D-CTF correction and subtomogram alignment. TomoCPR is demonstrated by Himes et al. [54] as one of the most effective adjustments made within the emClarity workflow. Himes et al. [54] note that tomoCPR is especially effective in tomograms where there are limited numbers of fiducial markers or where the fiducial markers are spatially separated from the particles of interest. It is, therefore, possible that the improvement in TSA is derived from the increased coverage of the

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Biochemical Journal (2021) **478** 1827–1845
https://doi.org/10.1042/BCJ20200715

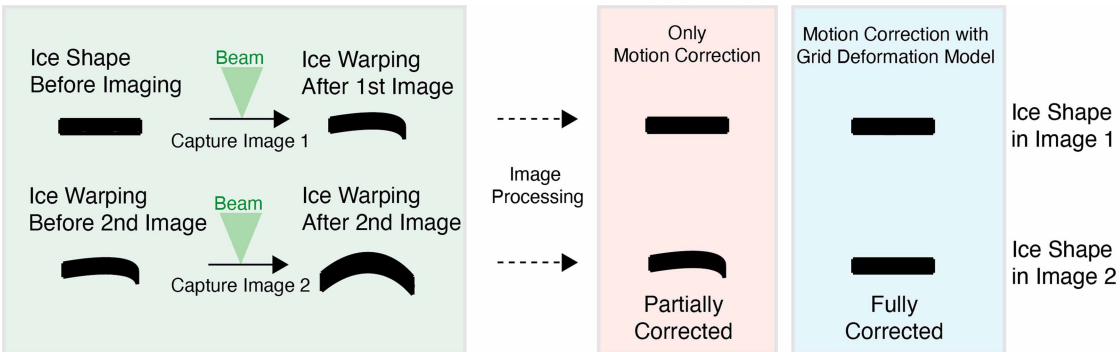


Figure 4. Comparison between correcting for ice deformation during the acquisition of a tilt-series by motion correction or motion correction with a grid deformation model.

Note that ice deformation and the accuracy/precision of the software is exaggerated for illustrative purposes.

grid in fiducial markers which track local changes better, rather than being due to the alignment of previously existing fiducial markers being optimised.

M is a tool which refines multiple steps within the STA workflow [55]. M is packaged with and utilised alongside Warp. M is used at the end of the STA pipeline to refine several steps across the workflow [55]. It identifies particles in each tilt image through cross-correlation with 2D projections of the reference(s). M accurately tracks the movement of particles and builds a model of the deformation of the grid caused by factors such as ice bending by treating particles as a physically connected group rather than individuals (Figure 4). This allows M to track and correct for the motion of particles caused by the deformation of the grid throughout the entire tilt-series rather than only correcting the motion in individual movies like reference-free motion correction software (Figure 4). Correcting for grid deformation also results in the improvement of the TSA. One of the key advantages of M is that it can identify and track the movement of particles using multiple references. This is particularly beneficial when imaging complex samples such as heterogenous and *in situ* data. Whilst the improvements in obtainable resolution of large ribosome and apoferritin particles demonstrated by Tegunov et al. [55] are promising, future studies will have to demonstrate the applicability of M to smaller particles. Furthermore, the applicability of M to particles which are grouped together in solely in one small region of a tomogram has not yet been established.

EMAN2 also carries out per-particle tilt refinement using the reference generated by the initial STA. Reference projections are used to refine the orientation parameters over a subtilt series generated for each particle and the quality of each tilt is assessed and weighted correspondingly. This corrects for misalignments of particles within the tilt-series, effectively refining the TSA for each particle. Chen et al. [45] report that this subtilt refinement protocol improved the resolution of their ribosome structures from 16 and 13 Å to 8.5 and 9.3 Å, respectively.

It is clear that recent improvements in the TSA refinement can have a major impact on the quality of the reconstructed tomograms. The improvement in tomogram quality that results from an accurate TSA can help refine other steps within the STA workflow, such as subtomogram alignment and 3D-CTF correction. Iterating TSA, 3D-CTF correction, and subtomogram alignment is clearly of strong benefit as refinements of each step in the workflow yield mutual, incremental improvements.

Classification

Proteins often exhibit both conformational and compositional heterogeneity. Particle heterogeneity can cause the loss of high-resolution information during subtomogram alignment and averaging, particularly within local regions containing structural variability within the data. However, heterogeneity within the data can also be used to elucidate the different states of the protein of interest. Consequently, classification techniques are often applied during or after subtomogram alignment to evaluate the differences between individual subtomograms. This ensures that particles in different conformational or compositional states are not aligned to one another and averaged to a single structure and can be used to separate and derive multiple structures of the same protein in different states. The classification process can also be used to discard subtomograms which either do

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Biochemical Journal (2021) 478 1827–1845
https://doi.org/10.1042/BCJ20200715

not contain the particle of interest or are otherwise compromised. Classification can be challenging in cryo-ET due to the effects of the missing-wedge making the differences between individual subtomograms difficult to detect and due to the limited number of particles in some datasets. There are multiple software packages which can classify aligned particles including Dynamo, RELION, emClarity, and EMAN2 [45,47,54,85].

The simplest form of classification is cross-correlation thresholding. This consists of assessing the cross-correlation of each subtomogram to the reference and discarding particles below a specific cross-correlation threshold. This is a fast and easy way to remove 'bad particles' from the subtomogram alignment and averaging steps. Care should be taken with the value of the cross-correlation threshold so as not to remove particles which could contribute useful information [19]. Visual inspection of some particles can assist with setting this threshold.

One of the most common methods used to classify particles is principal component analysis (PCA) with k-means or hierarchical clustering [47,54,86]. PCA can be described as a voxel-by-voxel comparison between aligned volumes. K-means or hierarchical clustering are methods of grouping similar images into the same class, effectively separating subtomograms into different groups. Ideally, each group should contain particles of either different conformation, composition, or image quality. PCA is not always suitable for large datasets as it is computationally expensive, in situations where one homogeneous conformation dominates the dataset, or for data where the initial subtomogram alignment fails due to excessive heterogeneity. In multiple reference alignment (MRA) classification, multiple references are generated and the subtomograms are aligned to each reference [47]. The cross-correlation of each subtomogram is calculated for each reference and subtomograms are paired with the reference with which they have the highest cross-correlation. One major shortcoming of MRA is that is requires *a priori* knowledge about the different particle conformations or compositions in the data. ML methods have also been developed to classify subtomograms [89]. The ML statistical framework has been shown to perform better than cross-correlation based methods in situations of poor SNR, avoiding the classification procedure falling into false local minima [90]. ML classification of subtomograms is implemented in RELION [77,85].

EmClarity implements considerable developments in the classification of subtomograms [54]. It utilises its 3D-CTF model (otherwise known as the 3D-sampling-function) to more accurately estimate the effect of the missing-wedge and the SNR on individual subtomograms at higher resolutions. The generation of the 3D-sampling-function was previously discussed in the missing-wedge compensation steps of subtomogram alignment (see subtomogram averaging). emClarity applies the 3D-sampling-function during the classification of each subtomogram which allows higher-resolution information to be utilised in the classification process. This allows the separation of classes of subtomograms with smaller differences than was previously possible. Himes et al. [54] demonstrated the ability of emClarity to distinguish between populations of the 80S ribosomes with subtle conformational differences.

Currently, the classification capabilities of cryo-ET and STA software packages are not as sophisticated as their SPA counterparts. Whilst there have been some recent advances in subtomogram classification, particularly in emClarity, more work is required to improve the classification capabilities of STA software. This is especially critical for *in situ* data in which crowded cellular environments, limited numbers of the particle of interest, noise, and high heterogeneity in the protein(s) of interest can limit the ability to effectively classify subtomograms. This can cause considerable limitations in the achievable resolution.

Post-processing

After one or more structures have been obtained by subtomogram alignment, averaging, and classification, it is possible to further optimise information derived from the data by using post-processing techniques. For structures generated by STA at a resolution lower than ∼5 Å, it is common for sharpening to be the only post-processing step. During sharpening, the structure factors of the Fourier transformation of the structure are reweighted according to the B-factor of the data [8,88,91].

For higher-resolution STA structures, most post-processing methods are directly adapted from SPA and X-ray crystallography workflows. For example, density modification for cryo-EM has recently adapted from X-ray crystallography by Terwilliger et al. [92,93]. Density modification is used to improve the map according to *a priori* knowledge of the expected structure factors of proteins and solvent. Some STA studies have begun to use density modification to improve the resolution of their maps [20]. Generally, density modification will be most successful where the solvent and particle are clearly separated which may be challenging for *in situ* data. Other post-processing methods which can be used for higher-resolution STA data include LocScale and DeepEMhancer which enhance local sharpening [94,95].

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Another recent development in structure refinement has been integrated into the M software package [55]. Prior to reconstructing the final map, M carries out a denoising procedure based on deep learning. Denoising accounts for the local resolution in the generated structure and filters Fourier components to avoid amplifying noise. To estimate the local resolution in the structure, M trains a neural network to recognise and effectively filter noise. Tegunov et al. [55] noted an improvement in resolution isotropy after using their denoising model.

Discussion

There are several approaches one can take to choose the ideal software and workflow to process cryo-ET data by STA. As different software packages often outperform one another in different areas of the STA workflow, this can introduce challenges in adapting parameters and reformatting the data and metadata to the requirements of different software packages at various points in the pipeline. Recently, more tools have been released to facilitate converting data and metadata to the required format to enable simple interchange between different software packages such as the recent work by Burt et al. [51] which integrates the Warp/M workflow with Dynamo. Reaching the highest possible resolution also often comes at the expense of time as it can require greater computational resources. Consequently, the user should carefully choose their preferred methodology based on a balance of resolution and time. Direct comparison between software packages is not straightforward, as their usage is highly dependent on the type and quantity of data and on the resolution necessary to answer a given biological question. However, it is possible to assess the areas in the STA workflow in which there has been considerable progress and which areas require improvement.

One area in which much progress has been made in recent years is correcting for BIM over a tilt-series, which cause resolution-limiting inaccuracies during TSA and reconstruction. Several software packages have introduced features which correct for this. Specifically, the deformation model in M and tomoCPR in emClarity have both been shown to generate impressive improvements in resolution [54,55]. The continued improvements to automated TSA processes have also reduced the reliance on manual TSA for high-resolution structures which previously represented a time-consuming bottleneck. Use of neural networks in particle picking is another area which shows great potential in both improving the resolution of structures by avoiding picking false-positive particles and by speeding up what was previously another time-consuming bottleneck in the processing workflow. The integration of many steps in the cryo-ET and STA workflow into software packages with GUIs such as EMAN2 and Warp/M has made data processing more accessible to inexperienced users. However, several outstanding issues still remain for processing cryo-tomography data with STA. One of the main issues is data processing speed which can prevents high-throughput data analysis. This will become more critical as advances in data collection, such as fast-incremental single-exposure tilt schemes, increase the speed of acquiring cryo-ET datasets [96]. Whilst several software packages have utilised GPU integration to increase speed [97], processes such as subtomogram alignment can sometimes take considerable computing time and resources. Improving the speed of the STA pipeline will take a combination of increased automation, adapting software to use the most time-efficient methodologies, and improving the ease of use for new users.

There has recently been considerable progress in the field of *in situ* cryo-ET [98]. Analysing the structure of proteins in their native state *in situ* is one of the key goals in structural biology. However, *in situ* data poses specific challenges in areas such as particle picking and classification. Some progress has been made in the area of *in situ* tomogram feature annotation and particle picking by neural networks [68]; however, classification remains an issue particularly for small particles or particles in low abundance. More work will have to be done to extract the full amount of information possible out of *in situ* data. Nevertheless, if the data processing techniques for cryo-ET data with STA consistently improve, it is possible that *in situ* cryo-ET could be used to address previously intractable biological questions.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This work was supported by grants from BBSRC (BB/T002670/1) and ERC (StG 852915 — CRYTOCOP).

Author Contributions

E.P. wrote the original draft of the manuscript and prepared the figures. E.P. and G.Z. reviewed and edited the manuscript.

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Biochemical Journal (2021) **478** 1827–1845
https://doi.org/10.1042/BCJ20200715

Acknowledgement

We thank Sander Van der Verren for helpful comments on the manuscript.

Abbreviations

BIM, beam-induced motion; Cryo-ET, cryo-electron tomography; CTF, contrast transfer function; EM, electron microscopy; FIB, focussed ion beam; FSC, Fourier shell correlation; ML, maximum-likelihood; MRA, multiple reference alignment; PCA, principal component analysis; SART, simultaneous algebraic reconstruction technique; SIRT, simultaneous iterative reconstruction technique; SNR, signal-to-noise ratio; SPA, single-particle analysis; STA, subtomogram averaging; TomoCPR, tomogram constrained-particle refinement; TSA, tilt-series alignment; WBP, weighted backprojections.

References

1 Yip, K.M., Fischer, N., Paknia, E., Chari, A. and Stark, H. (2020) Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–161 https://doi.org/10.1038/s41586-020-2833-4
2 Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P.M.G.E., et al. (2020) Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 https://doi.org/10.1038/s41586-020-2829-0
3 Jain, T., Sheehan, P., Crum, J., Carragher, B. and Potter, C.S. (2012) Spotiton: a prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. *J. Struct. Biol.* **179**, 68–75 https://doi.org/10.1016/j.jsb.2012.04.020
4 Naydenova, K., Jia, P. and Russo, C.J. (2020) Cryo-EM with sub-1 Å specimen movement. *Science (New York, N.Y.)* **370**, 223–226 https://doi.org/10.1126/science.abb7927
5 McMullan, G., Faruqi, A.R., Clare, D. and Henderson, R. (2014) Comparison of optimal performance at 300 keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy* **147**, 156–163 https://doi.org/10.1016/j.ultramic.2014.08.002
6 Punjani, A., Rubinstein, J.L., Fleet, D.J. and Brubaker, M.A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 https://doi.org/10.1038/nmeth.4169
7 Punjani, A., Zhang, H. and Fleet, D.J. (2020) Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 https://doi.org/10.1038/s41592-020-00990-8
8 Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E. et al. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, 163 https://doi.org/10.7554/eLife.42166
9 Liao, M., Cao, E., Julius, D. and Cheng, Y. (2013) Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107–112 https://doi.org/10.1038/nature12822
10 Madej, M.G. and Ziegler, C.M. (2018) Dawning of a new era in TRP channel structural biology by cryo-electron microscopy. *Pflugers Arch.* **470**, 213–225 https://doi.org/10.1007/s00424-018-2107-2
11 Velazhahan, V., Ma, N., Pándy-Szekeres, G., Kooistra, A.J., Lee, Y., Gloriam, D.E. et al. (2020) Structure of the class D GPCR Ste2 dimer coupled to two G proteins. *Nature* **7**, 195 https://doi.org/10.1038/s41586-020-2994-1
12 Rapisarda, C., Tassinari, M., Gubellini, F. and Fronzes, R. (2018) Using Cryo-EM to investigate bacterial secretion systems. *Annu. Rev. Microbiol.* **72**, 231–254 https://doi.org/10.1146/annurev-micro-090817-062702
13 Schur, F.K. (2019) Toward high-resolution in situ structural biology with cryo-electron tomography and subtomogram averaging. *Curr. Opin. Struct. Biol.* **58**, 1–9 https://doi.org/10.1016/j.sbi.2019.03.018
14 Zhang, P. (2019) Advances in cryo-electron tomography and subtomogram averaging and classification. *Curr. Opin. Struct. Biol.* **58**, 249–258 https://doi.org/10.1016/j.sbi.2019.05.021
15 Hutchings, J. and Zanetti, G. (2018) Fine details in complex environments: the power of cryo-electron tomography. *Biochm. Soc. Trans.* **46**, 807–816 https://doi.org/10.1042/BST20170351
16 Marko, M., Hsieh, C., Schalek, R., Frank, J. and Mannella, C. (2007) Focused-ion-beam thinning of frozen-hydrated biological specimens for cryo-electron microscopy. *Nat. Methods* **4**, 215–217 https://doi.org/10.1038/nmeth1014
17 Mahamid, J., Schampers, R., Persoon, H., Hyman, A.A., Baumeister, W. and Plitzko, J.M. (2015) A focused ion beam milling and lift-out approach for site-specific preparation of frozen-hydrated lamellas from multicellular organisms. *J. Struct. Biol.* **192**, 262–269 https://doi.org/10.1016/j.jsb.2015.07.012
18 Kelley, K., Jaroenlak, P., Raczkowski, A.M., Eng, E.T., Bhabha, G., Potter, C.S. et al. (2020) Waffle method: A general and flexible approach for FIB-milling small and anisotropically oriented samples. *bioRxiv* **17**, 2020.10.28.359372 https://doi.org/10.1101/2020.10.28.359372
19 Hutchings, J., Stancheva, V., Miller, E.A. and Zanetti, G. (2018) Subtomogram averaging of COPII assemblies reveals how coat organization dictates membrane shape. *Nat. Commun.* **9**, 4154–4158 https://doi.org/10.1038/s41467-018-06577-4
20 Hutchings, J., Stancheva, V.G., Brown, N.R., Cheung, A.C.M., Miller, E.A. and Zanetti, G. (2021) Structure of the complete, membrane-assembled COPII coat reveals a complex interaction network. *Nat. Commun.* **12**, 2034 https://doi.org/10.1038/s41467-021-22110-6
21 Kovtun, O., Dickson, V.K., Kelly, B.T., Owen, D.J. and Briggs, J.A.G. (2020) Architecture of the AP2/clathrin coat on the membranes of clathrin-coated vesicles. *Sci. Adv.* **6**, eaba8381 https://doi.org/10.1126/sciadv.aba8381
22 Dodonova, S.O., Aderhold, P., Kopp, J., Ganeva, I., Röhling, S., Hagen, W.J.H. et al. (2017) 9 å structure of the COPI coat reveals that the Arf1 GTPase occupies two contrasting molecular environments. *eLife* **6**, 352 https://doi.org/10.7554/eLife.26691
23 Ke, Z., Oton, J., Qu, K., Cortese, M., Zila, V., McKeane, L., et al. (2020) Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **588**, 498–502 https://doi.org/10.1038/s41586-020-2665-2
24 Mattei, S., Glass, B., Hagen, W.J.H., Kräusslich, H.-G. and Briggs, J.A.G. (2016) The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science (New York, N.Y.)* **354**, 1434–1437 https://doi.org/10.1126/science.aah4972
25 Wan, W., Kolesnikova, L., Clarke, M., Koehler, A., Noda, T., Becker, S. et al. (2017) Structure and assembly of the Ebola virus nucleocapsid. *Nature* **551**, 394–397 https://doi.org/10.1038/nature24490

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

26 Pfeffer, S., Burbaum, L., Unverdorben, P., Pech, M., Chen, Y., Zimmermann, R. et al. (2015) Structure of the native Sec61 protein-conducting channel. *Nat. Commun.* **6**, 8403 https://doi.org/10.1038/ncomms9403
27 Hagen, W.J.H., Wan, W. and Briggs, J.A.G. (2017) Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 https://doi.org/10.1016/j.jsb.2016.06.007
28 McMullan, G., Vinothkumar, K.R. and Henderson, R. (2015) Thon rings from amorphous ice and implications of beam-induced Brownian motion in single particle electron cryo-microscopy. *Ultramicroscopy* **158**, 26–32 https://doi.org/10.1016/j.ultramic.2015.05.017
29 Brilot, A.F., Chen, J.Z., Cheng, A., Pan, J., Harrison, S.C., Potter, C.S. et al. (2012) Beam-induced motion of vitrified specimen on holey carbon film. *J. Struct. Biol.* **177**, 630–637 https://doi.org/10.1016/j.jsb.2012.02.003
30 Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S. et al. (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 https://doi.org/10.1038/nmeth.2472
31 Grant, T. and Grigorieff, N. (2015) Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 https://doi.org/10.7554/eLife.06980
32 Zheng, S.Q., Palovcak, E., Armache, J.-P., Verba, K.A., Cheng, Y. and Agard, D.A. (2017) Motioncor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 https://doi.org/10.1038/nmeth.4193
33 Rubinstein, J.L. and Brubaker, M.A. (2015) Alignment of cryo-EM movies of individual particles by optimization of image translations. *J. Struct. Biol.* **192**, 188–195 https://doi.org/10.1016/j.jsb.2015.08.007
34 McLeod, R.A., Kowal, J., Ringler, P. and Stahlberg, H. (2017) Robust image alignment for cryogenic transmission electron microscopy. *J. Struct. Biol.* **197**, 279–293 https://doi.org/10.1016/j.jsb.2016.12.006
35 Abrishami, V., Vargas, J., Li, X., Cheng, Y., Marabini, R., Sorzano, C.ÓS. et al. (2015) Alignment of direct detection device micrographs using a robust optical flow approach. *J. Struct. Biol.* **189**, 163–176 https://doi.org/10.1016/j.jsb.2015.02.001
36 Zheng, S. JISCMail - CCPEM Archives. https://www.jiscmail.ac.uk/cgi-bin/wa-jisc.exe?A2=ind2011&L=CCPEM&O=D&P=6421
37 Tegunov, D. and Cramer, P. (2019) Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 https://doi.org/10.1038/s41592-019-0580-y
38 Bartesaghi, A., Lecumberry, F., Sapiro, G. and Subramaniam, S. (2012) Protein secondary structure determination by constrained single-particle cryo-electron tomography. *Structure (London, England: 1993)* **20**, 2003–2013 https://doi.org/10.1016/j.str.2012.10.016
39 Scheres, S.H. (2014) Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3**, e03665 https://doi.org/10.7554/eLife.03665
40 Zivanov, J., Nakane, T. and Scheres, S.H.W. (2019) A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ.* **6**, 5–17 https://doi.org/10.1107/S205225251801463X
41 Kremer, J.R., Mastronarde, D.N. and McIntosh, J.R. (1996) Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 https://doi.org/10.1006/jsbi.1996.0013
42 Mastronarde, D.N. (1997) Dual-Axis tomography: an approach with alignment methods that preserve resolution. *J. Struct. Biol.* **120**, 343–352 https://doi.org/10.1006/jsbi.1997.3919
43 Mastronarde, D.N. and Held, S.R. (2017) Automated tilt series alignment and tomographic reconstruction in IMOD. *J. Struct. Biol.* **197**, 102–113 https://doi.org/10.1016/j.jsb.2016.07.011
44 Amat, F., Moussavi, F., Comolli, L.R., Elidan, G., Downing, K.H. and Horowitz, M. (2008) Markov random field based automatic image alignment for electron tomography. *J. Struct. Biol.* **161**, 260–275 https://doi.org/10.1016/j.jsb.2007.07.007
45 Chen, M., Bell, J.M., Shi, X., Sun, S.Y., Wang, Z. and Ludtke, S.J. (2019) A complete data processing workflow for cryo-ET and subtomogram averaging. *Nat. Methods* **16**, 1161–1168 https://doi.org/10.1038/s41592-019-0591-8
46 Han, R., Wan, X., Wang, Z., Hao, Y., Zhang, J., Chen, Y., et al. (2017) Autom: A novel automatic platform for electron tomography reconstruction. *J. Struct. Biol.* **199**, 196–208 https://doi.org/10.1016/j.jsb.2017.07.008
47 Castaño-Díez, D., Kudryashev, M., Arheit, M. and Stahlberg, H. (2012) Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J. Struct. Biol.* **178**, 139–151 https://doi.org/10.1016/j.jsb.2011.12.017
48 Zheng, S.Q., Keszthelyi, B., Branlund, E., Lyle, J.M., Braunfeld, M.B., Sedat, J.W. et al. (2007) UCSF tomography: an integrated software suite for real-time electron microscopic tomographic data collection, alignment, and reconstruction. *J. Struct. Biol.* **157**, 138–147 https://doi.org/10.1016/j.jsb.2006.06.005
49 Noble, A.J. and Stagg, S.M. (2015) Automated batch fiducial-less tilt-series alignment in Appion using Protomo. *J. Struct. Biol.* **192**, 270–278 https://doi.org/10.1016/j.jsb.2015.10.003
50 Han, R., Wang, L., Liu, Z., Sun, F. and Zhang, F. (2015) A novel fully automatic scheme for fiducial marker-based alignment in electron tomography. *J. Struct. Biol.* **192**, 403–417 https://doi.org/10.1016/j.jsb.2015.09.022
51 Burt, A., Gaifas, L., Dendooven, T. and Gutsche, I. (2021) Tools enabling flexible approaches to high-resolution subtomogram averaging. *bioRxiv* 2021.01.31.428990 https://doi.org/10.1101/2021.01.31.428990
52 Fernandez, J.-J., Li, S., Bharat, T.A.M. and Agard, D.A. (2018) Cryo-tomography tilt-series alignment with consideration of the beam-induced sample motion. *J. Struct. Biol.* **202**, 200–209 https://doi.org/10.1016/j.jsb.2018.02.001
53 Fernandez, J.-J., Li, S. and Agard, D.A. (2019) Consideration of sample motion in cryo-tomography based on alignment residual interpolation. *J. Struct. Biol.* **205**, 1–6 https://doi.org/10.1016/j.jsb.2019.01.005
54 Himes, B.A. and Zhang, P. (2018) Emclarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nat. Methods* **15**, 955–961 https://doi.org/10.1038/s41592-018-0167-z
55 Tegunov, D., Xue, L., Dienemann, C., Cramer, P. and Mahamid, J. (2021) Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nat. Methods* **18**, 186–193 https://doi.org/10.1038/s41592-020-01054-7
56 Radermacher, M. (2006) Weighted Back-projection methods. In *Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell* (Frank, J., ed.), pp. 245–273, New York, NY: Springer New York
57 Gilbert, P. (1972) Iterative methods for the three-dimensional reconstruction of an object from projections. *J. Theor. Biol.* **36**, 105–117 https://doi.org/10.1016/0022-5193(72)90180-4
58 Andersen, A.H. and Kak, A.C. (1984) Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrason. Imaging* **6**, 81–94 https://doi.org/10.1177/016173468400600107

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Biochemical Journal (2021) **478** 1827–1845
https://doi.org/10.1042/BCJ20200715

59 Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I. et al. (2007) EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 https://doi.org/10.1016/j.jsb.2006.05.009
60 Castaño-Díez, D., Kudryashev, M. and Stahlberg, H. (2017) Dynamo catalogue: geometrical tools and data management for particle picking in subtomogram averaging of cryo-electron tomograms. *J. Struct. Biol.* **197**, 135–144 https://doi.org/10.1016/j.jsb.2016.06.005
61 Hrabe, T., Chen, Y., Pfeffer, S., Cuellar, L.K., Mangold, A.-V. and Förster, F. (2012) Pytom: a python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *J. Struct. Biol.* **178**, 177–188 https://doi.org/10.1016/j.jsb.2011.12.003
62 Henderson, R. (2013) Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proc. Natl Acad. Sci. U.S.A.* **110**, 18037–18041 https://doi.org/10.1073/pnas.1314449110
63 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. et al. (2004) UCSF chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 https://doi.org/10.1002/jcc.20084
64 Qu, K., Glass, B., Doležal, M., Schur, F.K.M., Murciano, B., Rein, A. et al. (2018) Structure and architecture of immature and mature murine leukemia virus capsids. *Proc. Natl Acad. Sci. U.S.A.* **115**, E11751–E11760 https://doi.org/10.1073/pnas.1811580115
65 Moebel, E., Martinez-Sanchez, A., Larivière, D., Fourmentin, E., Ortiz, J., Baumeister, W. et al. (2020) Deep learning improves macromolecules localization and identification in 3D cellular cryo-electron tomograms. *bioRxiv* **89**, 2020.04.15.042747 https://doi.org/10.1101/2020.04.15.042747
66 Sánchez-García, R., Segura, J., Maluenda, D., Carazo, J.M. and Sorzano, C.O.S. (2018) Deep consensus, a deep learning-based approach for particle pruning in cryo-electron microscopy. *IUCrJ.* **5**, 854–865 https://doi.org/10.1107/S2052252518014392
67 Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., et al. (2019) SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 https://doi.org/10.1038/s42003-019-0437-z
68 Chen, M., Dai, W., Sun, S.Y., Jonasch, D., He, C.Y., Schmid, M.F. et al. (2017) Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *Nat. Methods* **14**, 983–985 https://doi.org/10.1038/nmeth.4405
69 Zanetti, G., Riches, J.D., Fuller, S.D. and Briggs, J.A.G. (2009) Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging. *J. Struct. Biol.* **168**, 305–312 https://doi.org/10.1016/j.jsb.2009.08.002
70 Kudryashev, M., Castaño-Díez, D. and Stahlberg, H. (2012) Limiting factors in single particle cryo electron tomography. *Comput. Struct. Biotechnol. J.* **1**, e201207002 https://doi.org/10.5936/csbj.201207002
71 Rohou, A. and Grigorieff, N. (2015) CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 https://doi.org/10.1016/j.jsb.2015.08.008
72 Zhang, K. (2016) Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 https://doi.org/10.1016/j.jsb.2015.11.003
73 Xiong, Q., Morphew, M.K., Schwartz, C.L., Hoenger, A.H. and Mastronarde, D.N. (2009) CTF determination and correction for low dose tomographic tilt series. *J. Struct. Biol.* **168**, 378–387 https://doi.org/10.1016/j.jsb.2009.08.016
74 Sanchez, R.M., Zhang, Y., Chen, W., Dietrich, L. and Kudryashev, M. (2020) Subnanometer-resolution structure determination in situ by hybrid subtomogram averaging - single particle cryo-EM. *Nat. Commun.* **11**, 3709 https://doi.org/10.1038/s41467-020-17466-0
75 Turoňová, B., Schur, F.K.M., Wan, W. and Briggs, J.A.G. (2017) Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *J. Struct. Biol.* **199**, 187–195 https://doi.org/10.1016/j.jsb.2017.07.007
76 Jensen, G.J. and Kornberg, R.D. (2000) Defocus-gradient corrected back-projection. *Ultramicroscopy* **84**, 57–64 https://doi.org/10.1016/S0304-3991(00)00005-X
77 Bharat, T.A.M. and Scheres, S.H.W. (2016) Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nat. Protoc.* **11**, 2054–2065 https://doi.org/10.1038/nprot.2016.124
78 Wu, M., Lander, G.C. and Herzik, Jr, M.A. (2020) Sub-2 angstrom resolution structure determination using single-particle cryo-EM at 200 keV. *J. Struct. Biol.* **4**, 100020 https://doi.org/10.1016/j.yjsbx.2020.100020
79 Zivanov, J., Nakane, T. and Scheres, S.H.W. (2020) Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ.* **7**, 253–267 https://doi.org/10.1107/S2052252520000081
80 Bromberg, R., Guo, Y., Borek, D. and Otwinowski, Z. (2020) High-resolution cryo-EM reconstructions in the presence of substantial aberrations. *IUCrJ.* **7**, 445–452 https://doi.org/10.1107/S2052252520002444
81 Ramlaul, K., Burt, A., de Martín Garrido, N., MacDonald, J.T., Palmer, C.M., Jakobi, A. et al. (2020) Direct information estimation from cryo-EM movies with CARYON. *bioRxiv* **11**, 2020.11.25.398891
82 Henderson, R., Sali, A., Baker, M.L., Carragher, B., Devkota, B., Downing, K.H., et al. (2012) Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205–214 https://doi.org/10.1016/j.str.2011.12.014
83 Bartesaghi, A., Sprechmann, P., Liu, J., Randall, G., Sapiro, G. and Subramaniam, S. (2008) Classification and 3D averaging with missing wedge correction in biological electron tomography. *J. Struct. Biol.* **162**, 436–450 https://doi.org/10.1016/j.jsb.2008.02.008
84 Heumann, J.M., Hoenger, A. and Mastronarde, D.N. (2011) Clustering and variance maps for cryo-electron tomography using wedge-masked differences. *J. Struct. Biol.* **175**, 288–299 https://doi.org/10.1016/j.jsb.2011.05.011
85 Bharat, T.A.M., Russo, C.J., Löwe, J., Passmore, L.A. and Scheres, S.H.W. (2015) Advances in single-Particle electron cryomicroscopy structure determination applied to Sub-tomogram averaging. *Structure (London, England: 1993)* **23**, 1743–1753 https://doi.org/10.1016/j.str.2015.06.026
86 Förster, F., Pruggnaller, S., Seybert, A. and Frangakis, A.S. (2008) Classification of cryo-electron sub-tomograms using constrained correlation. *J. Struct. Biol.* **161**, 276–286 https://doi.org/10.1016/j.jsb.2007.07.006
87 Huiskonen, J.T., Hepojoki, J., Laurinmäki, P., Vaheri, A., Lankinen, H., Butcher, S.J. et al. (2010) Electron cryotomography of Tula hantavirus suggests a unique assembly paradigm for enveloped viruses. *J. Virol.* **84**, 4889–4897 https://doi.org/10.1128/JVI.00057-10
88 Scheres, S.H.W. (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 https://doi.org/10.1016/j.jsb.2012.09.006
89 Stölken, M., Beck, F., Haller, T., Hegerl, R., Gutsche, I., Carazo, J.M. et al. (2011) Maximum likelihood based classification of electron tomographic data. *J. Struct. Biol.* **173**, 77–85 https://doi.org/10.1016/j.jsb.2010.08.005
90 Sigworth, F.J., Doerschuk, P.C., Carazo, J.M. and Scheres, S.H.W. (2010) An introduction to maximum-likelihood methods in cryo-EM. *Methods Enzymol.* **482**, 263–294 https://doi.org/10.1016/S0076-6879(10)82011-7

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

91 Rosenthal, P.B. and Henderson, R. (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 https://doi.org/10.1016/j.jmb.2003.07.013
92 Terwilliger, T.C., Sobolev, O.V., Afonine, P.V., Adams, P.D. and Read, R.J. (2020) Density modification of cryo-EM maps. *Acta Crystallogr. D Struct. Biol.* **76**, 912–925 https://doi.org/10.1107/S205979832001061X
93 Terwilliger, T.C., Ludtke, S.J., Read, R.J., Adams, P.D. and Afonine, P.V. (2020) Improvement of cryo-EM maps by density modification. *Nat. Methods* **17**, 923–927 https://doi.org/10.1038/s41592-020-0914-9
94 Sánchez-García, R., Gómez-Blanco, J., Cuervo, A., Carazo, J.M., Sorzano, C.O.S. and Vargas, J. (2020) DeepEMhacer: a deep learning solution for cryo-EM volume post-processing. *bioRxiv* **24**, 1181 https://doi.org/10.1101/2020.06.12.148296
95 Jakobi, A.J., Wilmanns, M. and Sachse, C. (2017) Model-based local density sharpening of cryo-EM maps. *eLife* **6**, 213 https://doi.org/10.7554/eLife.27131
96 Chreifi, G., Chen, S., Metskas, L.A., Kaplan, M. and Jensen, G.J. (2019) Rapid tilt-series acquisition for electron cryotomography. *J. Struct. Biol.* **205**, 163–169 https://doi.org/10.1016/j.jsb.2018.12.008
97 Castaño-Díez, D. (2017) The dynamo package for tomography and subtomogram averaging: components for MATLAB, GPU computing and EC2 Amazon web services. *Acta Crystallogr. D Struct. Biol.* **73**, 478–487 https://doi.org/10.1107/S2059798317003369
98 Böhning, J. and Bharat, T.A.M. (2020) Towards high-throughput in situ structural biology using electron cryotomography. *Prog. Biophys. Mol. Biol.* **23**, 3583 https://doi.org/10.1016/j.pbiomolbio.2020.05.010

Downloaded from http://portlandpress.com/biochemj/article-pdf/478/10/1827/912005/bcj-2020-0715c.pdf by Birkbeck College (University of London) user on 27 May 2021

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).