Chapter 28

Structural Analysis of Protein Complexes by Cryo Electron Microscopy

Tiago R.D. Costa, Athanasios Ignatiou, and Elena V. Orlova

Abstract

Structural studies of biocomplexes using single-particle cryo-electron microscopy (cryo-EM) is now a well-established technique in structural biology and has become competitive with X-ray crystallography. The latest advances in EM enable us to determine structures of protein complexes at 3–5 Å resolution for an extremely broad range of sizes from ~200 kDa up to hundreds of megadaltons (Bartesaghi et al., Science 348(6239):1147–1151, 2051; Bai et al., Nature 525(7568):212–217, 2015; Vinothkumar et al., Nature 515(7525):80–84, 2014; Grigorieff and Harrison, Curr Opin Struct Biol 21(2):265–273, 2011). The majority of biocomplexes comprise a number of different components and are not amenable to crystallisation. Secretion systems are typical examples of such multi-protein complexes, and structural studies of them are extremely challenging. The only feasible approach to revealing their spatial organisation and functional modification is cryo-EM. The development of systems for digital registration of images and algorithms for the fast and efficient processing of recorded images and subsequent analysis facilitated the determination of structures at near-atomic resolution. In this review we will describe sample preparation for cryo-EM, how data are collected by new detectors, and the logistics of image analysis through the basic steps required for reconstructions of both small and large biological complexes and their refinement to nearly atomic resolution. The processing workflow is illustrated using examples of EM analysis of a Type IV Secretion System.

Key words Cryo-electron microscopy, Sample preparation, Single particle analysis, Image processing, Type IV secretion system

1 EM Advances in Studies of Macro-Complexes (Type IV Secretion Systems)

The complexity of experimental and computational procedures used for studies of the structure–function relationships of biological complexes is growing significantly. One has to use different approaches to uncover conformational changes linked to the functional activity of the complexes. X-ray, nuclear magnetic resonance (NMR), and electron microscopy (EM), combined with biochemical and biophysical methods, allow for a deeper understanding of the mechanisms which underlie such macromolecular complex functions. This was clearly demonstrated by studies of a

Laure Journet and Eric Cascales (eds.), *Bacterial Protein Secretion Systems: Methods and Protocols*, Methods in Molecular Biology, vol. 1615, DOI 10.1007/978-1-4939-7033-9_28, © Springer Science+Business Media LLC 2017

ribosome [1]. Recent advances in EM, such as the invention of direct electron detection cameras, systems for automated data collection, and the development of new powerful image processing algorithms, have dramatically expanded the range of biological macromolecules suitable for study by this technique. The main advantages of EM are that it does not require crystallisation of samples and it is able to work with biocomplexes within a large range of sizes: from ~150 kDa to several hundreds of megadaltons [2–5]. Additionally, image processing software packages have been significantly improved to make it possible to analyse the quality of images and distortions caused by the microscope that prevent obtaining high-resolution structures. New approaches were developed to reveal more consistently the sample quality: to assess its homogeneity, to separate different conformations within the sample, and evaluate distributions of particles between different conformations [6–8]. A number of different packages with sophisticated image processing algorithms are used for the analysis of macromolecular complexes with different symmetries or asymmetry. Another positive aspect of modern achievements is that computing power is steadily increasing, making it possible to analyse many hundreds of thousands of particle images from heterogeneous samples. Nonetheless, the basic workflow of sample imaging and image processing remains the same (Fig. 1) [9].

This tremendous success in improving the resolution of EM structures during the last decade would not have been possible without cryo sample preparation and what we now call cryo-EM. This approach for sample preparation and the combination of methods of X-ray crystallography with EM have made it possible to achieve near-atomic-resolution details for some of the Gram-negative bacterial secretion systems. These structures have given us unprecedented understanding of the mechanistic details of how bacteria assemble these highly specialised nano-machines to secrete proteins and DNA to the bacterial extracellular space and to the eukaryotic or bacterial target cell. Among the Gram-negative bacterial secretion systems, Type IV Secretion Systems (T4SS) possess the unique ability to secrete proteins, DNA, or protein–DNA complexes in an adenosine triphosphate–dependent process. Given the competency for T4SSs to secrete such a variety of substrates involved in the pathogenesis and spreading of conjugative plasmids encoding antibiotic resistance genes, this secretion system become an important target for structural biology studies [10].

Among all Gram-negative T4SSs, those encoded by the pTi plasmid of *Agrobacterium tumefaciens*, together with the conjugative pKM101 and pR388 plasmids from *E. coli*, are the best characterized. This macromolecular structure is composed of 12 proteins: VirB1–VirB11 and VirD4 [11]. The first main advance in the understanding of the general architecture of a T4SS took place when the cryo-EM structure of the so-called core outer-membrane

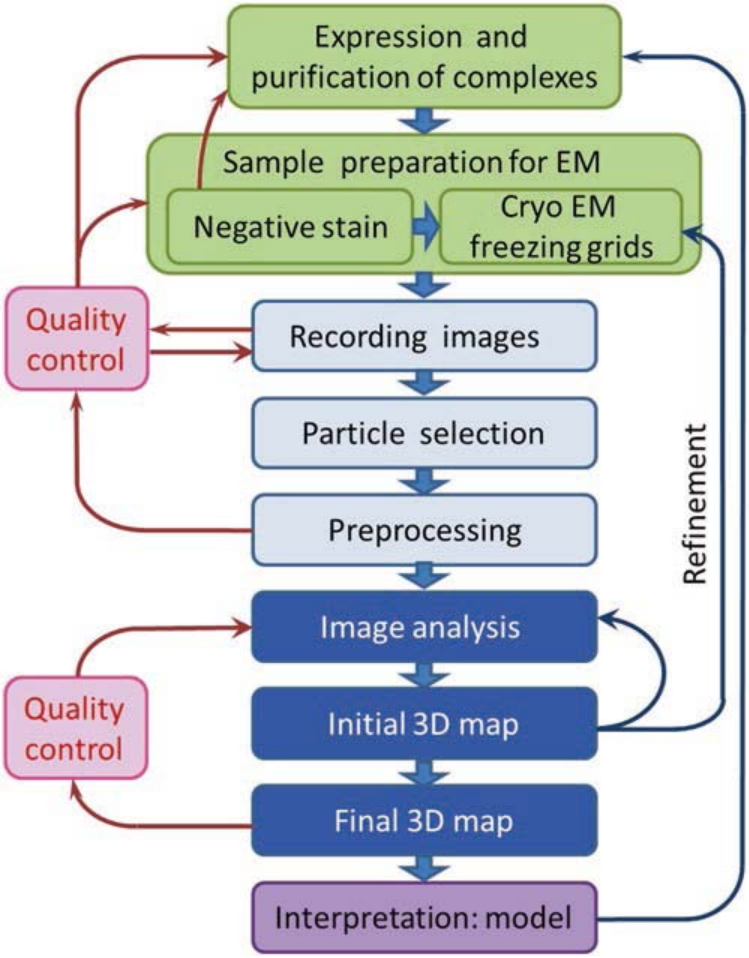

Fig. 1 Workflow of EM structural analysis. In *green* is the experimental part of structural analysis. The computational part is shown in *light* and *dark blue*; the initial steps of processing are shown in *light blue*. They include image frame alignment, CTF correction, normalisation, and filtering. The subsequent steps—alignment, statistical analysis, determination of particle orientations, and initial three-dimensional reconstruction (3D)—are shown in *dark blue*. The final step (*light purple*) is the interpretation of the maps obtained

complex (OMC, encoded by the conjugative pKM101 plasmid) was solved with a resolution of 15 Å. This 1.1 MDa structure, which spans both the outer and inner membrane, is made of 14 copies of VirB7, VirB9, and VirB10 proteins (Fig. 2a) [12]. Further, the resolution of the same core OMC was improved to 12.4 Å, which provided further details on the structural organisation of the proteins that form this complex (Fig. 2b) [13]. Recently, the almost complete full structure of the T4SS (VirB3–VirB10) encoded by the conjugative R388 plasmid was solved by negative staining (NS). This remarkable structure provided the first view of both the outer and the bipartite inner-membrane complex (IMC) and how these are linked by a structure called stalk (Fig. 2c) [14].

This review might not be complete from the point of view of a specialist and it might not provide sufficient mathematical background for the reader. However, we will try to give a general

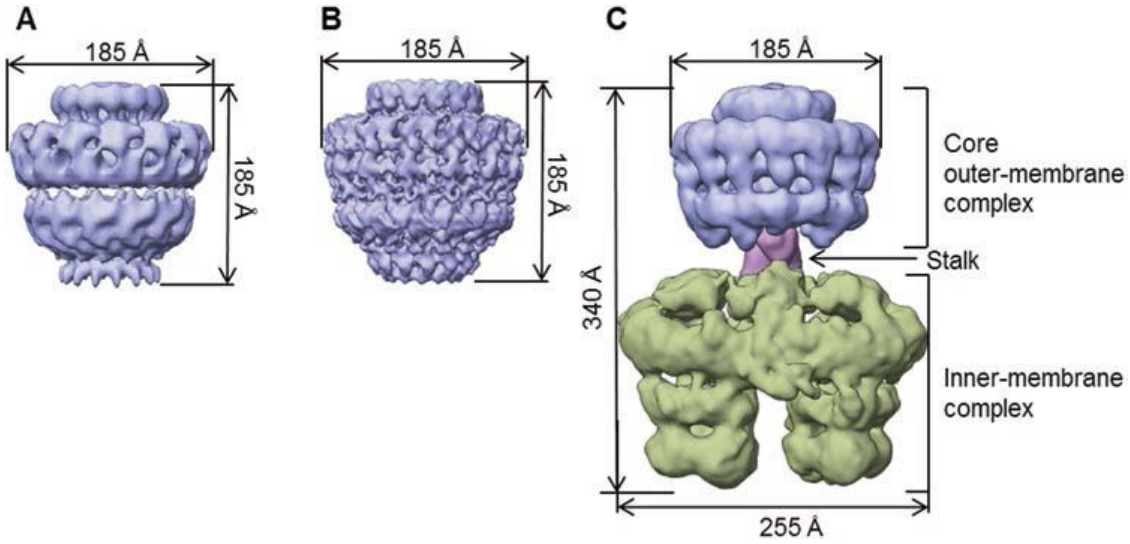

Fig. 2 EM maps of different T4SS structures. (**a**) Cryo-EM structure of core outer-membrane complex at 15 Å resolution. (**b**) Cryo-EM structure of the core outer-membrane complex at 12.4 Å resolution. (**c**) Structure of almost complete the T4SS complex (in negative stain) at overall resolution of ~20 Å

overview of imaging using an electron microscope and the current basic steps of structural analysis. This will include the outline of sample cryo preparation, the effects of radiation damage, and advances in the procedure of data collection. We will describe steps considered to be pre-processing determination of orientations of the particle images, and methods used to obtain structures and how they can be evaluated. Since this is a rather short review, we will not describe here how an image is obtained in an electron microscope. This information can be found in other reviews and books [8, 15]; readers interested in more details on the topics described here may consult the references provided at the end of the chapter.

2 Sample Preparation in Cryo-EM

Although EM provides much better resolution than light microscopy, it has the disadvantage that samples must be imaged in a vacuum. This is due to the fact that images are created by a beam of electrons in a transmission electron microscope. Without a vacuum, electrons become very quickly absorbed by air since they collide with air molecules and lose their energy and direction of scattering, therefore to obtain a high-quality image of a sample, it is necessary to keep the electron path free of air molecules (under a vacuum) to allow electrons to move directly to the sample. In their native conditions biological objects (macromolecules and cells) are immersed in water solutions. To be visualised in an electron microscope under vacuum the bio complexes have to be made rigid and stable to avoid drying out or undergo structural changes during the exposure time at data collection.

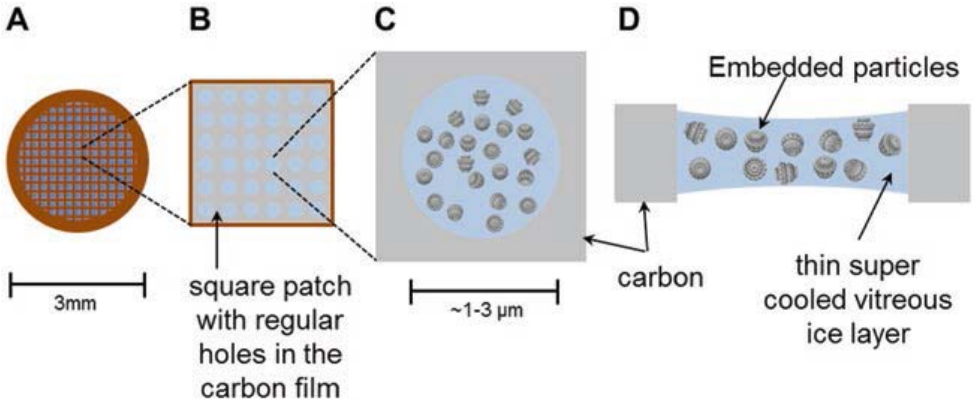

Fig. 3 Cryo-EM sample preparation. (**a**) 3 mm copper mesh grid covered with a film of holey carbon. (**b**) Magnified image of square patch showing microscopic holes in carbon. (**c**) Enlarged image of a single hole containing a layer of vitrified ice with protein molecules. (**d**) Cross section of a hole with particles embedded in ice

Cryo-EM methods of sample preparation allow to preserve the structural integrity of biocomplexes, keeping them in a nearly native hydrated state in the vacuum system of the microscope. The method proposed by Dubochet et al. [16, 17] is now a well-established, standard technique for freezing aqueous solutions of samples on cryo-EM grids. The EM grid is a metallic round plate (~3 mm in diameter and usually made of copper) with a fine mesh. The size of the mesh is typically chosen depending on the experiment, but the most commonly used type has 400 squares per inch (Fig. 3). Depending on the sample, a continuous thin layer of carbon film or one perforated with irregular (lacey grids) or regular holes should be put on the top of the metal grid. One can use manufactured grids with regular holes in carbon films. The grids should be chosen according to the size and shapes of holes and distance between holes, which are most suitable for the given sample (e.g. Quantifoil grids, Quantifoil Micro Tools GmbH; C-flat grids, Protochips, Inc.) with regularly arranged holes for automated and manual data collection (www.protochips.com; http://www.agarscientific.com).

A drop of a sample (~3 µL) is applied to a glow-discharged (to make the surface more hydrophilic) grid; the sample is kept for a short time on the grid (0.5–2 min, depending on the sample) and then the gird is maintained on a plunger. Excess sample is blotted to make a thin layer of sample solution and then the grid immediately plunged into liquid ethane (or propane) that has a temperature of −182 °C. The ethane must be cooled prior sample freezing by liquid nitrogen (Fig. 4). Plunge freezing in liquid ethane takes place in ~10^{-5} s, trapping the biological molecules in their native, hydrated state embedded in amorphous ice that is like solid water. Cooling by plunging into liquid ethane is much faster than plunging directly into liquid nitrogen because the liquid ethane is used

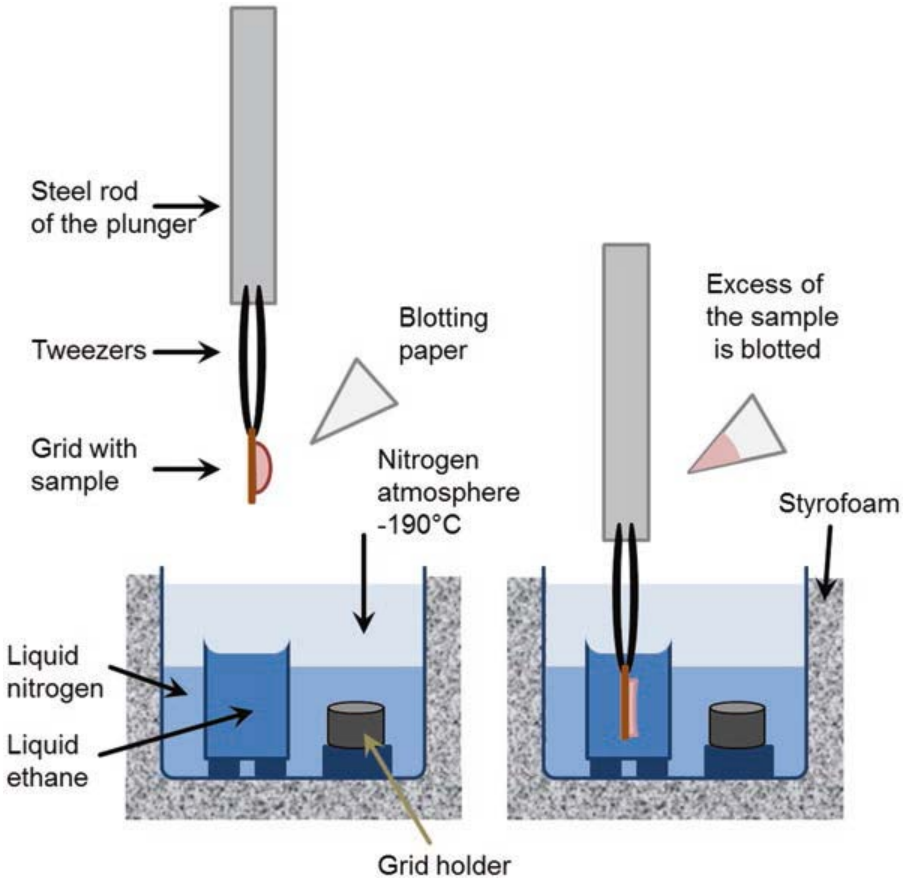

Fig. 4 Sample vitrification. *Left panel*: grid with sample applied being held in tweezers; *right panel* : after blotting excess of a sample, grid is plunged into container filled with liquid ethane. The top level of the container should be immersed in the nitrogen atmosphere that has a temperature slightly above the temperature of liquid nitrogen. The grid is then transferred into the grid holder. The transfer must be done without taking the grid out of the nitrogen atmosphere

near its freezing point so it does not evaporate and produce an insulating gas layer. This fast freezing prevents the formation of ice crystals and keeps samples in a nearly native hydrated state [18, 19]. More detailed information on the vitrification of samples can be found in papers by M. Samsó and R.A. Grassucci [20, 21].

The grids must be kept all the time at temperatures not higher as –170 °C (in a storage, during transfer to the microscope using a cryo-transfer holder, and at imaging in the electron microscope); otherwise, ice will change its conformation and start to make crystals that will destroy the sample and contaminate the grids. Another important advantage of cryo-EM is that liquid nitrogen temperatures reduce the radiation damage induced by electron beam when it passes through samples [22, 23]. Nowadays automated and controlled devices (Vitrobots) have been developed, thereby allowing higher reproducibility in grid preparation [24, 25]. However, it is recommended that the first estimation of the sample quality should

be done using the NS technique, which is fast, robust, and reliable [26]. It allows a quick assessment of sample quality, concentration, and suitability for the subsequent cryo preparations.

3 Image Acquisition by Digital Detectors

3.1 CCD Cameras

EM, like all modern photographic systems, is related to the acquisition of images and now uses digital cameras. All digital cameras convert analogue optical signals into digital format so the steps in developing and scanning film have become redundant. The first digital cameras used charge-coupled device (CCD) sensors, which had been invented in 1969 by W.S. Boyle and G.E. Smith at Bell Telephone Laboratories (Nobel Prize, 2009) [27]. The concept behind this device was based on the transformation of an analogue signal, such as photon energy, into an electrical charge using specially designed photo sensors. The magnitude of the charge registered by the sensor is proportional to the energy of photons absorbed by the sensor. CCD chips consist of an array of photosensitive elements. In the readout mechanism, charges are successively transferred to a reading register, amplified, and converted into a digital signal. The number of reading registers determines the speed of the CCD image recording. Since there are only a few of them, the readout of these cameras is not very high.

However, in electron microscopes, the process of recording electrons is more complicated than with photosensors, which are not able to register electrons. Therefore, the sensors were modified in such a way that a mono- or polycrystalline scintillator that converts electron energy into photons was placed on top of the photosensor, and only then were photons converted into an electrical signal (Fig. 5a) [28, 29]. Unfortunately, a CCD camera's sensitivity decreases with increased voltage of an electron microscope, so thicker scintillator layers are needed to improve the electron detection efficiency. Thick layers of scintillators affect the image quality, which is degraded because the higher-energy electrons are scattered through several adjacent sensors, reducing the image resolution. Nonetheless, these cameras provided experience and understanding for the development of automated data collection in EM.

3.2 Direct Electron Detectors

In the last decade, new digital detectors have been designed enabling the detection of electrons without the intermediate step of transforming electrons into photons and then into an electrical signal (Fig. 5b). Direct detector devices (DDDs) use an array of radiation-hardened active pixel sensors (a pixel circuit) which are integrated into a silicon complementary metal-oxide semiconductor (CMOS) chip [30, 31]. In this case the electron energy is transformed directly into an electrical signal. Another advance in this

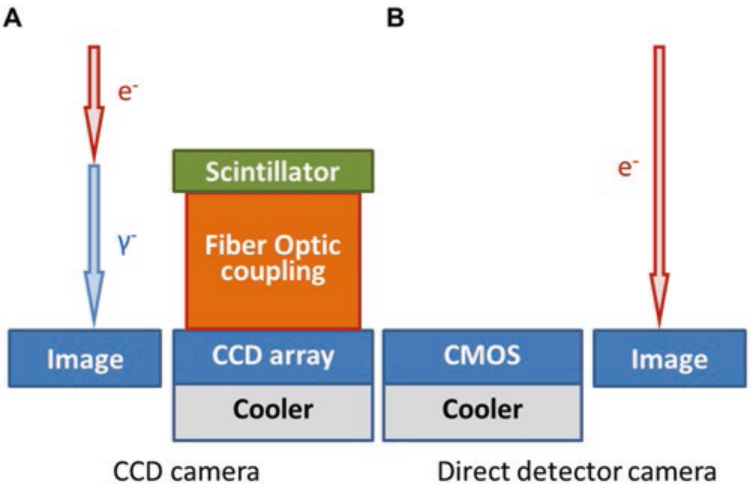

Fig. 5 Digital cameras. (**a**) In CCD-based cameras, the electrons hit a scintillator, generating light, which is partially captured by fibre optics, and directed onto the cooled CCD chip. (**b**) In direct detectors (two *right panels*), active pixel sensors, mostly based on CMOS technology, are capable of capturing and directly detecting incident electrons

technology is that an amplifier is built into each pixel and allows fast signal readout from each individual sensor (or pixel) nearly simultaneously. This makes it possible to separate a single exposure into a set of smaller subexposures. In cryo-EM, this provides a valuable option for electron dose fractionation that is important in studies of radiation-sensitive biological samples. The resultant image subframes can then be used for specimen drift correction, which is not possible with conventional CCD cameras.

The photon-electron conversion step is removed by the CMOS semiconductor technology, where the fibre optics became unnecessary allowing to improve the signal-to-noise ratio (SNR) in images registered by DDD compared to those from CCD. The quality of DD detectors is best described in terms of the detective quantum efficiency (DQE) [32, 33]. The DQE is a measure of the efficiency of signal transfer by the camera and defined as a ratio of the SNR in the output image registered by the camera sensors to the SNR at the input image:

$$DQE = \left(SNR_{out}\right)^2 / \left(SNR_{in}\right)^2$$

The ratio depends on the spatial frequency (sizes of the details) of the image. A perfect detector would not distort the input signal, so in an ideal system the output should be the same as the input. Therefore, the DQE of an ideal system would be equal to one for all frequencies. In reality, cameras distort the fine details in images, and this is reflected by a significant decline in the DQE at high frequencies [34, 35].

Direct detectors make it possible to register electrons within a high range of energy and are now used in 300 keV microscopes.

The high sensitivity of such a system has made it possible to reduce the size of sensors, and additional improvements in software have offered a new mode of image exposure, such as counting mode, where the system records single electrons, as implemented in Gatan's K2 cameras [33, 34, 36, 37].

3.3 Micrograph Subframe Alignment

Cryo-EM images using DDD cameras can record from 7 subframes on Falcon (FEI) to 40–50 subframes using direct electron (DE) or K2 (Gatan) cameras. Thus, the data recorded with DDD cameras represent sets of image frames (movies) that can be motion-corrected. Such high rates of image recording can reveal distortions of images induced by drift of the grid (sample) within the EM. Typically, frame alignment starts from frame N−1 that is aligned to the last frame. Then these two images are summed, and frame N−2 will be aligned to this sum. Then frames N−1 and N−2 will be summed, and frame N−3 is aligned to this new sum. The process is repeated in the same way towards the first frame. There are variations in algorithms when summation is done not for two frames for the following alignment but for four or five frames or all frames. Alignment is refined iteratively when on the next round of alignment the total sum obtained during the previous round is used as a reference. The entire procedure improves, firstly, SNR of the reference and then the quality of the alignment. Nowadays a number of software packages can be used for the frame alignment [36, 38–41]. The image shown in Fig. 6a represents the sum of the original frames without any correction. The trajectory of the image shift during these several exposures indicates that the movement of initial shifts of the sample is large initially but then slows down (Fig. 6b). A power spectrum from the sum of the frames without motion correction demonstrates that the Thon rings are not very sharp: they fade fast owing to the small shifts in different directions shown in Fig. 6b (red dots). When the movie frames are aligned (the motion correction), the Thon rings become symmetrical, going up to 3 Å (Fig. 6c). This indicates the presence of high-resolution details in the images. Summation of the motion-corrected subframes generates the final sharper image (Fig. 6d).

3.4 Radiation Damage

Images in an electron microscope are generated by the electron beam that illuminates the sample and then the image is formed by electromagnetic lenses in the plane of the camera. While a short wavelength of the electron beam improves dramatically the resolution of images of biological molecules, it was proved that biosamples are very sensitive to the high-energy electron irradiation that takes place during imaging. Changes in biological complexes depend on the time of the overall exposure (cumulative) dose and were estimated using spot fading diffraction experiments on two-dimensional (2D) crystals [42–44]. Therefore, a 3D structure derived from experiments when a sample was overexposed can

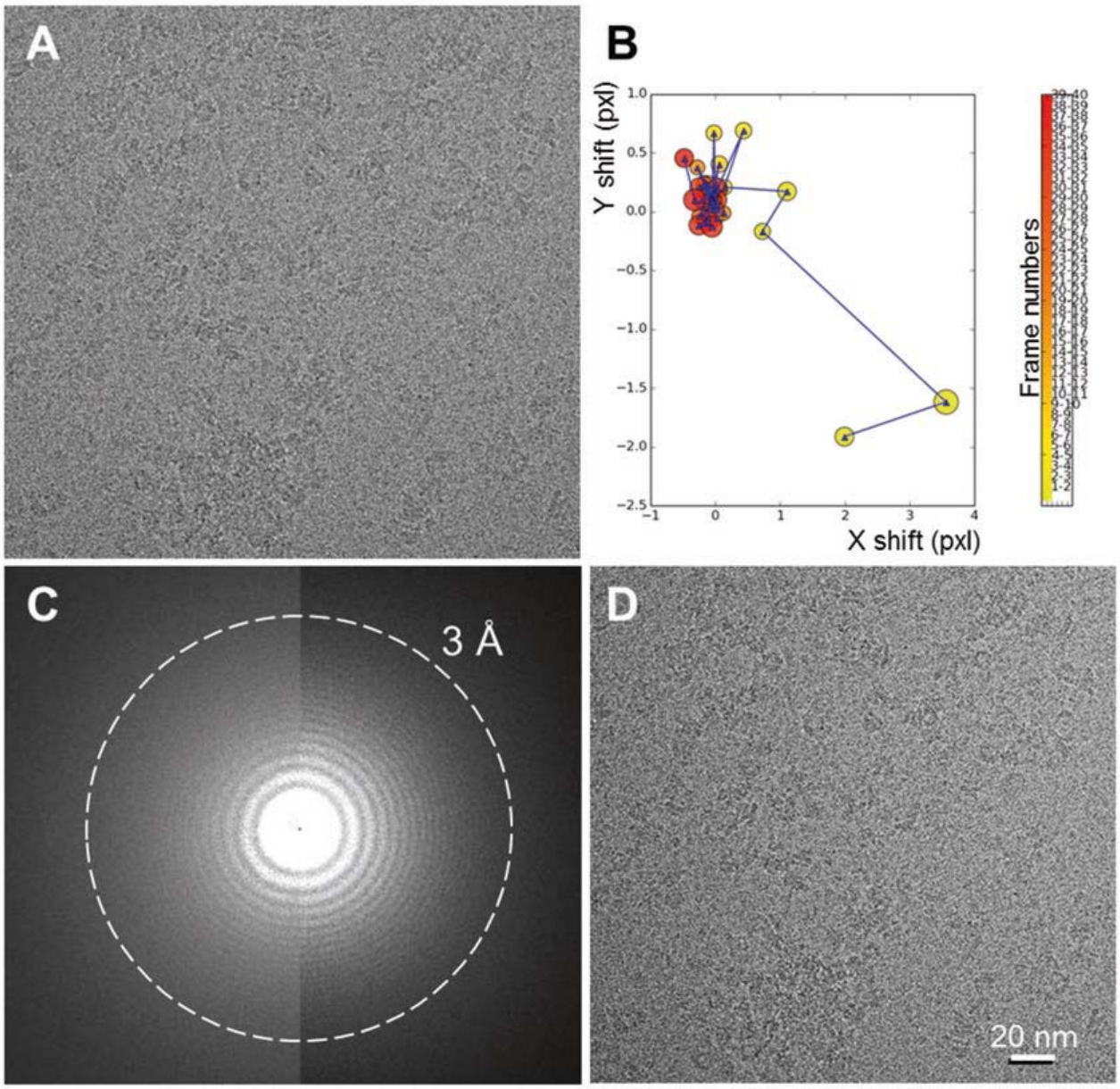

Fig. 6 Cryo-EM images with motion correction. (**a**) Representative cryo-EM image of vitrified T4SS particles. (**b**) Trace of motion in *X* and *Y* directions of frames. (**c**) *Left* : power spectrum from the sum of raw movie frames without motion correction. *Right* : power spectrum from the sum of movie frames after motion correction. (**d**) Sum of movie frames that were shifted according to determined shifts shown in (**b**). Protein is *black* in these images

differ remarkably from the structure of the native molecule. High-energy electrons of the electron beam in EM may induce displacements, bond breakage, and mass loss of low-atomic-number elements such as carbon, nitrogen, and oxygen [45]. It has been shown by crystallography that exposure of crystals to X-rays induces decarboxylation of glutamate and aspartate residues, the breakage of disulphide bonds, and the loss of hydroxyl groups from tyrosine and the methylthio group of methionine [46].

Cryo-EM imaging carries the major benefit of reducing radiation damage as samples are kept at cryogenic temperatures during imaging. Vitrified samples preserve their native structure and are

imaged well at liquid nitrogen temperatures [22, 47]. Such low temperatures increase tolerance to ionising radiation damage [44, 48] since the free radicals generated from inelastic scattering events are unable to diffuse through the sample and cause secondary damage [49]. In addition, the freezing also constrains the movement and degrees of freedom of the atoms of a molecule after a bond is broken, thereby limiting the structural rearrangement produced during irradiation [44]. As a result, keeping and imaging samples at liquid nitrogen temperature improves radiation tolerance two- to sixfold over room temperature imaging [44, 48].

Another important approach has been used for a number of years in EM which is the usage of the low-dose mode during data collection. Low-dose imaging is based on reducing the amount of time a sample is exposed to electrons by focusing on an adjacent area that is sufficiently close to the area of interest but does not overlap it. All modern electron microscopes that are used for biological studies come with pre-installed low-dose software allowing for efficient exchange between imaging modes. The search mode is a low-magnification overview image used to identify areas of interest while the imaging (or photo) mode is used for actual data collection at high magnification. The focus mode is typically set at a higher magnification than imaging mode, but the beam is shifted to an adjacent area. Such an interchange between these modes is implemented in systems for automated data collection and allows significant reduction of radiation damage.

The next and now very fast evolving method is the usage of dose fractionation, which is provided by current direct detector technology. DDDs have a very high speed of frame readouts. Depending on the detector type (FEI, Gatan, or DE) and the available software, it is possible to record from 7 to 60 subframes per exposure. Typically, the images in the first two or three frames demonstrate large sample shifts, while later on the movement slows down. However, the last frames indicate often that the sample has been damaged by the beam (Fig. 6b). Bartesaghi and co-authors compared the density maps reconstructed from different fractions of the total exposure (10, 20, or 30 $e^{-}/Å^{2}$). Analysis of these high-resolution cryo-EM structures show that densities for residues with positively charged and neutral side chains are well resolved, while the residues with negatively charged side chains, having weaker densities, were less resolved [50]. The negatively charged glutamate and aspartate show on average 30% less density than the similarly sized neutral glutamine and asparagines [50], which is consistent with observations in X-ray analysis [46]. Therefore, the user can use all frames for the quality assessment of images and samples and then only the first half or first two thirds of subframes (depending on the type of the DDD used in experiments) for the reconstruction of the native complex [42, 43, 48, 50, 51].

4 Image Analysis of Micrographs

4.1 Contrast Transfer Function

The aim of single-particle reconstruction is to obtain an accurate representation of the 3D structure of a molecule using a set of 2D projection image data. The macromolecular complexes are considered as thin objects so their images can be described as linear projections of the Coulomb potential of the molecular complex [44]. This is a primary condition necessary for the subsequent reconstruction procedure. However, images produced by electron microscopes do not directly represent projections of the molecules under study. Deviations from the real densities of projections are induced by aberrations in the optical system of the microscope [52].

The function that describes the actual representation of every single point in the registered image of a theoretically correct projection is called the contrast transfer function (CTF) of the microscope [44, 52, 53]. The CTF is defined by the acceleration voltage (electron wavelength), the type of electron source (beam coherence), and aberrations of the objective lens (C_s, C_c, and astigmatism). The major factors affecting the CTF are the degree of spherical aberration (C_s) of the objective lens and level of defocus (Δf). As a result, the CTF modulates the amplitudes and phases of the electron diffraction pattern formed in the back focal plane of the objective lens. For any given defocus setting the features of a specimen are modulated through positive and negative contrast. The CTF limits the amount of information which can be obtained from electron images. At zero crossings of the CTF no information is transmitted, and specimen features corresponding to such spatial frequencies will not be visible in the final image.

A transmission electron microscopy (TEM) image can be represented as a power spectrum (Fourier space), which demonstrates the magnitude of the various frequency components contained within the image (Figs. 7 and 8). The effect of the CTF on an image is that the power spectrum looks like it is oscillating and appears as concentric rings or Thon rings [54], which indicate the location of the minima and maxima in frequency space. Dark regions show the positions of all zero crossings of the CTF, and bright regions correspond to areas where the CTF has either positive or negative contrast (Figs. 7 and 8).

The major effect on images of biological samples from the spherical aberration of the objective lens is to cause phase changes, and therefore the representation of densities in image is altered significantly. On the other hand, biological samples viewed in ice under close to focus conditions demonstrate very little amplitude contrast since the difference in their densities and water density is very small [8, 44]. Thus, the images are typically taken far from focus (in underfocus mode) to increase the weight of low frequencies and thereby improve the visibility of particles [8, 44]. Here it

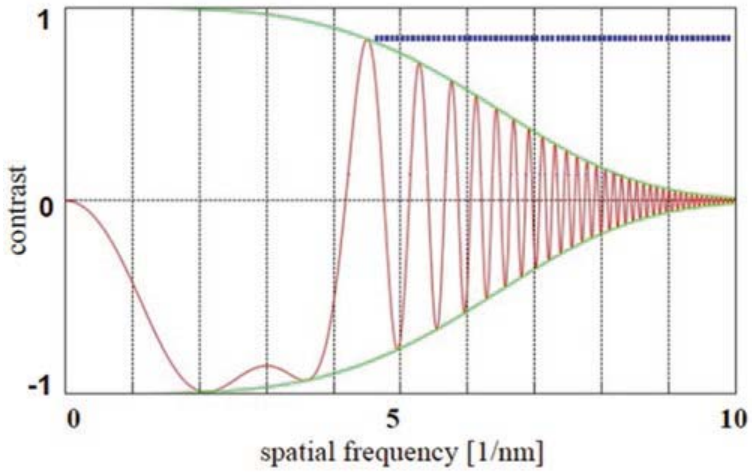

Fig. 7 CTF with envelope function. *Dotted blue line* : amplitude of all frequencies in perfect microscope; *green line*: effect of envelope function on CTF (*red*) resulting in suppression of high spatial frequencies

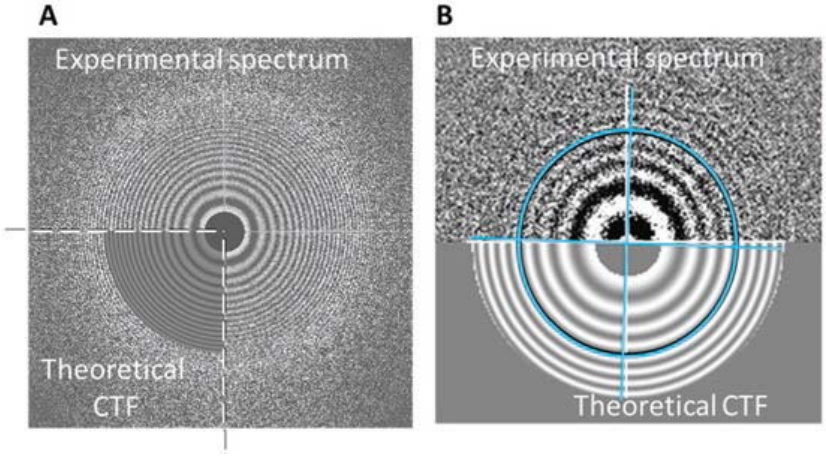

Fig. 8 Assessment of CTF parameters. (**a**) Comparison of theoretically calculated CTF (*left bottom quadrant*) with CTF seen in experimental spectrum. For an accurate CTF determination the Thon rings from both image parts should match accurately. (**b**) Identification of axes of astigmatism which are superimposed over Thon rings of an actual observed power spectrum and compared with the theoretical spectrum. The spectrum of a micrograph shown here indicates that there is a small astigmatism, ~2%, and the axes of ellipse are slightly tilted, shown in *light blue*

should be mentioned that low frequencies are responsible for the overall shape and appearance of particles in images. However, high defocusing induces changes in the distribution of density information related to fine details that could be lost owing to the attenuation of amplitudes at high frequencies. The level of defocus used for imaging depends on the size of the biocomplex. The images of small particles (~100–300 kDa) are taken with a large defocus, sometimes up to 6–7 μm, while viruses with diameters of at least 50 nm can be imaged at 0.5–1.0 μm.

The CTF for biological samples can be described by the formula

$$Phase\ CTF = -2\sin\left[\pi\left(\Delta f \lambda q^2 - C_s\lambda^3 q^4 / 2\right)\right],$$

Phase CTF C_s = spherical aberration constant; Δf = defocus; q = spatial frequency; λ = electron wavelength. The spherical aberration coefficient and the electron wavelength are the only constants, and these values remain fixed for each electron microscope [52].

4.2 Defocus Determination and Correction of CTF

To correct an image for CTF effects and obtain an image that corresponds to the projection, it is necessary to determine its defocus and to check it for astigmatism and drift. The nominal value of defocus set on the microscope does not usually represent the actual defocus obtained in the final digital image or micrograph. This occurs because, although the acceleration voltage and spherical aberration remain constant, the common deviations in sample thickness and the position of the supporting film will cause local variations in defocus. As a result, the CTF related to the defocus should be determined for each image frame. Finding the exact level of defocus and astigmatism in cryo-EM images is of absolutely crucial importance when working on the production of a high-resolution structure.

CTF determination is performed by calculating the sum of power spectra (or amplitudes) of small patches (256 × 256 or slightly larger) from the sum of all subframes. This spectrum is correlated with a number of CTFs theoretically calculated in a range of possible defocus values. A maximum correlation between the observed and a theoretical CTF would indicate the actual defocus of the values and will define frequencies where the phases must be flipped (Fig. 8). Different options for (semi-) automated defocus determination are available in a number software packages, such as EMAN2, CTFIND, and IMAGIC5 [55–57].

Astigmatic images have power spectra which are not rotationally symmetric, and this can complicate and reduce the accuracy of CTF determination. Generally cryo-EM images which have greater than 5% astigmatism are not used for further processing except in special cases where strong astigmatism can be used to recover information in areas where the CTF crosses zeroes. The level of astigmatism can be calculated as follows:

$$Astigmatism = \left(Defocus_{\max} - Defocus_{\min} / Defocus_{avg}\right)$$

An EM projection image is only considered a faithful representation of the observed object of interest if it has been corrected for the CTF modulation effects of the microscope. This can only be done after CTF determination. Phase correction is carried out in reciprocal space by multiplying the alternating rings of the CTF by

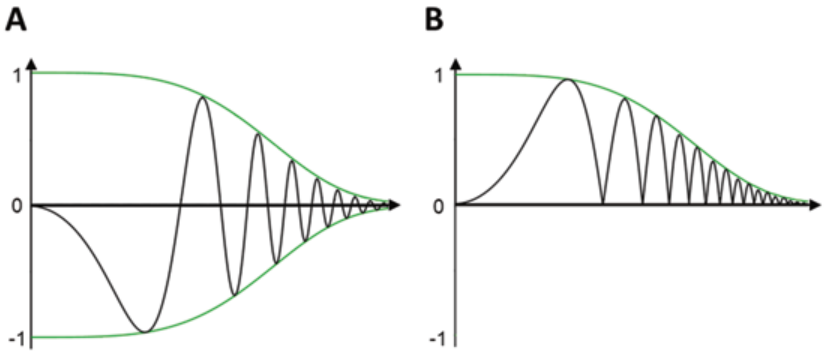

Fig. 9 CTF correction. (**a**) CTF oscillates changing contrast from negative to positive depending on frequencies. Information is lost only where CTF crosses zero line. (**b**) Negative lobes of uncorrected CTF are flipped over to positive (correction of CTF by phase flipping). The missing information can be recovered by collecting images at different defocus levels which fill these zero regions with information

−1 at positions where the contrast transfer is negative and +1 where the contrast transfer is positive. This has the effect of reversing or "flipping" the negative lobes of the CTF into positive contrast, thereby restoring the correct image phases (Fig. 9).

The missing information where the CTF crosses zeroes are restored by combining images at different defocuses, so that where some images lack spatial information at a particular frequency others will provide the complementary missing information. High spatial frequencies are suppressed by envelope decay, so amplitude correction is also important for maximising high-resolution details. This operation usually involves applying a Wiener filter [58] to remove noise from the CTF prior to amplitude amplification.

4.3 Particle Selection

The structural analysis process in EM begins with selecting images of individual particles from micrographs. This involves recording their unique locations (x,y) within the image field and saving these coordinates in a data file that will be used in the next steps of processing. This can be done interactively using packages like Xmipp [59], EMAN2 [55], Ximdisp [60], RELION (semi-automated selection of cryo-EM particles in RELION-2 [61], and others. The simplest way to do this is to select a single particle image by clicking on the image with a mouse. The coordinates of these points will be stored and then used to extract individual particles within a square box of designated dimension, for example, 500 × 500 pixels. The cut-out area must be large enough to retain all the image data around the object with as little background as possible. Particles can also be selected automatically with particle identification/selection programs, for example, Autopicker [62], BShow [63], and FindEM [64]. These programs use the assessment of local correlation to measure the degree of similarity between reference images and then a small area of the raw micrograph. Areas which show maximum correlation to the references are boxed out.

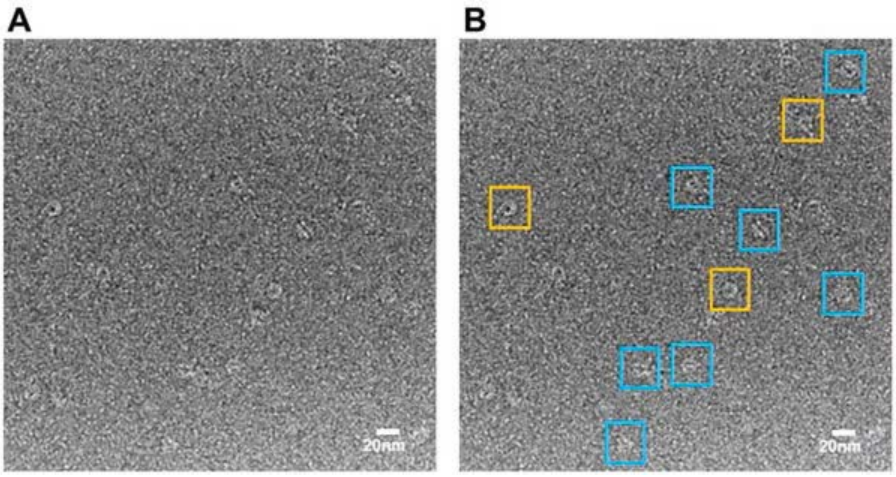

Fig. 10 Particle picking. (**a**) Cryo-EM micrograph of T4SS core outer-membrane complex. (**b**) Particles outlined by *yellow squares* represent end views; side views are outlined by *squares* in *cyan*

The poor contrast in ice images and the presence of artefacts which could resemble target particles can generally increase the inaccuracy of automatic selection.

The images of selected particles must not overlap with other particles and particles in images should not be distorted. In Fig. 10 we show an example of a micrograph of a vitrified sample of the T4SS core OMC. Images were recorded on a 4096 × 4096 Gatan CCD camera with a low electron dose on a Tecnai F20 FEG microscope operating at a voltage of 200 kV, a magnification of 68,100, and a defocus range of 1250–3500 nm.

4.4 Normalisation of Data

The normalisation of all images is an essential pre-processing step. The contrast and intensity can vary from image to image during data acquisition, even when all the EM settings are the same. This effect arises because of a number of factors, including differences in the thickness of the carbon support film or ice, particle orientations, uneven staining, or the merging of images from different data collection sessions. Normalisation standardises the densities of images by setting the mean pixel grey value of each image to the same level, commonly zero, and also rescaling the standard deviation to an equal value for every particle image. Without normalisation the density variations, such as very bright or very dark regions within an image, could bias the cross-correlation procedures which are later used for alignment and calculation of particle classes. Typically, normalisation is based on the following formula:

$$\rho_{i,j}^{norm} = \left(\left(\rho_{i,j} - \rho_{avg}\right) / \sigma_{old}\right)\sigma_{new},$$

σ_{old} and σ_{new} are the standard deviations of the original and target images respectively, and $\rho_{i,j}$ is the density of a pixel in the image array coordinates.

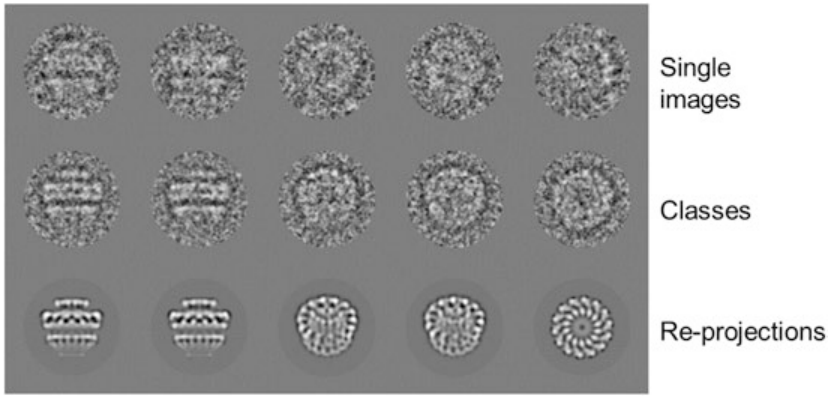

Fig. 11 Alignment and classification of T4SS core outer-membrane complex images. *Upper panel* : representative images of core outer-membrane complex; *middle panel* : class averages of images at nearly the same orientations; *bottom panel* : corresponding reprojections from final 3D model

4.5 Alignment of Particle Images

One of the most important steps in image processing is the reduction of noise and enhancement of the signal. Different factors, such as insufficient coherence of the electron beam, quality of amorphous ice (due to uneven distribution of salts and some other effects in buffers), supporting films, and the noise of registering cameras, contribute to decreasing the SNR in particle images. Since these types of noise are not related to a signal from the sample, the averaging of particle images improves the SNR. However, to retrieve reliable information using averaging, images must represent the same particles in the same orientations [44]. Therefore, images of particles in the same orientations should be identified, and before averaging, the images should be aligned rotationally and translationally with respect to each other. Alignment compares all images to a reference image and shifts them so that they are in the same position as the reference. Usually, normalised particle images should be centred or aligned to the reference that represents a typical view of the complex.

One of the possible options for starting analysis is to align all images of a data set to the rotationally averaged total sum of all images (with no alignment). The centred images are then subjected to multivariate statistical analysis (MSA) (*see* following discussion) to obtain a number of averages of images which are aligned only translationally and grouped according to common features. The best characteristic views (averages of groups of images with the lowest variations among them) are used for multi-reference alignment. The most reliable classes are centred and used as new references for the next round of alignment (Fig. 11). The procedure can be repeated several times alternating with MSA [65]. In other cases one can use alignment in Fourier space [66] or so-called reference-free alignment implemented in EMAN2 and SPIDER [55, 67].

4.6 Statistical Analysis and Classification

To improve the SNR, aligned images which have high similarity between each other should be grouped together, and for this purpose, statistical analysis and classification are used. Different approaches have been developed to reduce a large number of variables to a limited number of important parameters [68].

4.6.1 Principal Component Analysis

Principal component analysis (PCA) reduces the number of variables to find the most significant variations in the measurements [44, 65]. The essence of the procedure is a transformation of a set of observations of possibly correlated variables (in our case images) into a set of values of uncorrelated variables called principal components. In the complete representation, a number of the principal components are equal to the number of original variables. However, since images contain a high level of noise, the number of the meaningful components becomes much smaller. The principal components are described by the eigenvectors of the data matrix. PCA is the simplest of the true eigenvector-based multivariate analyses [65, 68].

4.6.2 Factor Analysis

Factor analysis is designed to identify variations in a number of original variables using predefined "significant" factors that are often defined by a researcher [44]. That requires specific assumptions about the underlying features of the object under study, such as average density or the perimeter and size of specific domains.

4.6.3 Maximum Likelihood Estimation

Maximum likelihood estimation (ML) is a method that assesses parameters that would correspond to a statistical model. When applied to a data set (such as our image data set) and given a statistical model (the initial 3D model), ML provides estimates of how our new reconstruction would correspond to the proposed model and what sort of deviations could be observed. This concept can be stated in different words: once a model is specified with parameters (to a certain extent) and data have been collected (our EM images), one is able to evaluate how well the model fits the observed data. The quality of this correlation is assessed by finding parameter values of a model that best fit the data—a procedure called *parameter estimation*. In the EM case, that would be a 3D model that corresponds in the best way to the data set; otherwise, the model must be modified. ML has many properties which should be taken into account during estimation: sufficiency (complete information about parameters reflecting features of interests), consistency (numbers of images related to this or some other 3D model), efficiency (lowest possible variance of parameter estimates), and perhaps other practical parameters [69, 70]. This method is successfully used in the analysis of 3D reconstructions and implemented in RELION [71].

4.6.4 Classification

Classification is done once the principal components or important factors of the data are defined. Cluster analysis is a tool to identify groups of similar objects. This kind of analysis is used for grouping

similar images (particles in the same orientations), and in EM, two implementations are used. One is K-means (used in SPIDER, EMAN, and XMIPP [55, 59, 67]), where the user defines a number of classes (K, typically not bigger than 10) that should be obtained and the algorithm randomly assigns each image to one of the classes [72]. These starting (random) points are called centroids or seeds. The centroids should be placed as far away from each other as possible. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. Averages are calculated for each class, and the distance between each image and the obtained averages are calculated (Fig. 12a). A new class will be formed by the images that were closest to one of the averages. Then the class averages are recalculated. This process is done iteratively until the images stop moving between classes. The K-means method is reasonably fast and works better in low-dimensional space since dimensionality increases time and a local minima problem may occur.

Another method of classification is hierarchical ascendant classification (HAC) (implemented in IMAGIC, SPIDER, and EMAN2). There are two main streams: agglomerative, which is a "bottom-up" approach where each observation starts in its own cluster and pairs of clusters are merged as one moves up the hierarchy, and divisive, a "top-down" approach where all observations start in one cluster and splits are performed recursively as one moves down the hierarchy. HAC is based on the Ward criterion [73], which minimises the intra-class variance while maximising the inter-class variance. In IMAGIC a version of the agglomerative approach is used. This criterion is used in the pair-wise merging of the classes that are supposed to be obtained to form a HAC tree (Fig. 12b). The user chooses a number of classes that would be needed for subsequent processing, and the HAC tree is cut at that level. Sometimes it is hard to achieve the lowest intra-class variance since this algorithm does not allow one to move images from one class to another. However, this can be achieved by weighting parameters during reclassification.

4.7 Determination of Particle Orientation

To obtain the 3D structure of a biocomplex from EM images, the orientation of each of the individual particle images must be determined. The location of an individual molecule can be identified by the X, Y, and Z coordinates, and the shifts of different particles with respect to each other can be described in the same way as shifts in X, Y, and Z. The particles can also be rotated by α, β, and γ angles, called *Euler angles*. This means that a molecule has six degrees of freedom in space. In the microscope images correspond to the projections along the Z-axis of the translational system of coordinates, so the shift in the Z direction is not significant (we assume that the electron beam is parallel), but the shifts in the X and Y directions should be determined. During translational

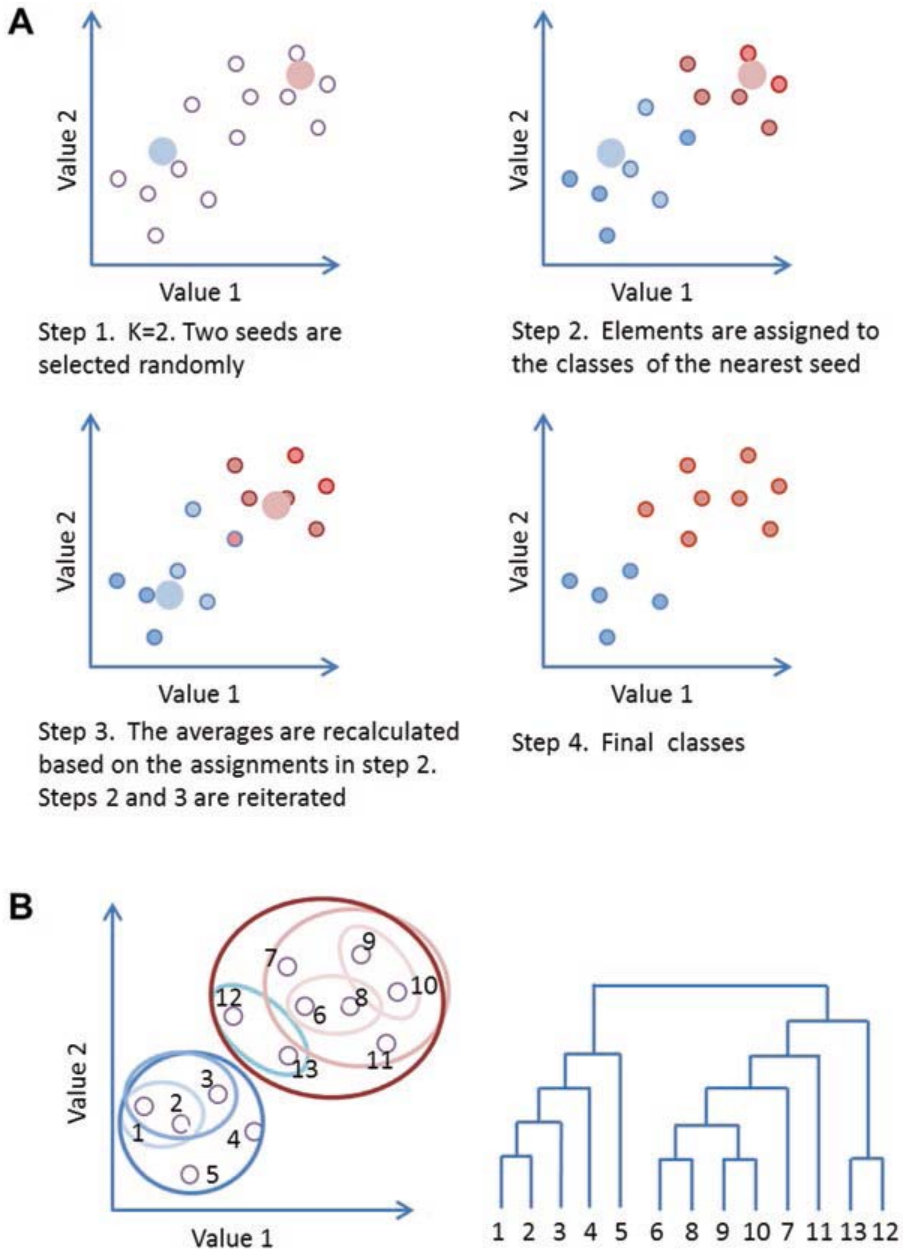

Fig. 12 Principles of image classification. (**a**) K-means classification. Experimental data represented by *empty circles* with their characteristic parameters v1 and v2. The initial randomly selected seeds are shown in *light blue* and *pink* . Results of classification steps are shown in *circles* coloured correspondingly. (**b**) Hierarchical classification. The *left panel* shows position of points in plane and their successive forming of classes; the *right panel* shows the classification tree

alignment, the centre of the molecule is set to $X = 0$ and $Y = 0$. To calculate the 3D map from individual images or class sums, it is necessary to determine the orientations of the characteristic views (classes) relative to each other.

4.7.1 Random Conical Tilt

The random conical tilt (RCT) technique is used to obtain an initial model of a new complex about which little is known. The method is based on a typical property of samples having a preferential orientation on a grid when negatively stained. The RCT

approach is a reliable method for generating an unbiased initial 3D model obtained by experimental measurements. Two electron micrographs of the same part of the grid (at magnification ~30–40K) are required; the first one is typically at a high tilt (45–60°, where a goniometer is used), and the second image of the same area is taken without tilt [74]. The tilt respectively to the *Z* axis is known (the same as that used for the tilt of the first image), and rotation around the *Z* axis is obtained from images of the same particles but in the untilted micrograph. As soon the tilted images are centred, angles are assigned to them (Euler angles). With the relative orientation in space determined, a 3D reconstruction of the object can be produced [44]. Although such a model could be far from prefect, it will be a good start for subsequent refinement.

4.7.2 Projection Matching

Projection matching requires an initial model and is based on a simple principle of comparison of images with projections of the model [44]. As a 3D template (an initial map), one can use the low-resolution negative-stain EM 3D reconstruction, the low-pass filtered X-ray model, or the EM map of a homolog. This template is projected in all possible directions covering an entire Euler sphere with a certain angular increment. Then the images or class averages of the data set are compared with these references, and the angles corresponding to the reference with the best cross-correlation will be assigned to the image [67]. Projection matching helps to determine out-of-plane rotations of the object. During the angular determination refinement, the angular increment between projections is reduced, or additional projections with a small increment can be calculated around the initial angle. This method is easy to use; however, it is extremely time consuming due to a long computation during which it is necessary to try all possible in-plane alignments and to compare each image to a set of references. Nonetheless, multi-processor computers can be used to speed up the process. Once the Euler angles are assigned to all images or class averages, a new 3D reconstruction will be calculated and a new set of refined higher resolution model projections computed for the next round of projection matching. Several different software programs, such as IMAGIC [57], EMAN2 [55], and SPIDER [67], offer the option of projection matching.

4.7.3 Angular Reconstitution

EM images represent projections of embedded molecules in random orientations. If there is no initial model, the angular reconstitution technique is a method of choice to determine the orientation of images relative to the other images. The common line projection theorem postulates that every pair of 2D projections of the same 3D object has at least one mutual 1D line projection called a *common line projection* [8, 75]. Thus, by matching 1D line projections for different images, we can identify the relationship between the 2D projections and determine the angles between the common lines and, therefore, determine the relative Euler angles of the images.

A set of 1D projections of an image is obtained when the image is rotated 360° at 1° intervals. The set of 1D projections forms a sinogram (because a trajectory of projections of one point at successive rotations from 1° to 360° corresponds to a sine function). During angular reconstitution [75] each 1D projection of the first image is compared with each 1D projection calculated from the second 2D image (Fig. 13). The line that has the highest correlation (in theory it should be equal to 1) is considered as a common line of these two projections. Sinograms of different images are compared pair-wise line by line, checking the correlation to find common 1D projections (the most similar 1D projections) between two selected 2D images. At least three images are required to determine an initial orientation with respect to the object. Sinograms are generated for all classes, and the search for common lines is performed for all images. The angles between common 1D projections are used to assign an orientation to the class average. Then other classes are added to the initial set of three [75, 76]. It is also possible to determine orientations using common lines in Fourier space [77, 78]. The central section theorem states that the 2D Fourier transform of a 2D projection represents a 2D central section through the 3D Fourier transform of the 3D density. In Fourier space, a common line corresponds to the cross section of Fourier transforms of images. This means that two Fourier transforms from two different 2D projections of the same 3D object have one common central line [77, 78]. Here a comparison of the radial lines of the Fourier transform of one image with all possible radial lines of the Fourier transform of the other image is performed. Again, as in real space, the angle between common lines of the two images with respect to the third one gives the angle between these two views. A combination of EM and angular reconstitution has become an important method for analysing 3D structures of non-crystallised molecules.

4.8 3D Reconstruction

EM images are considered as 2D projections of a 3D object [44]. This is due to the large depth of focus in TEM images. The depth of defocus is related to the acceleration voltage: the higher the voltage, the greater the depth of focus, which can be up to 200 nm. Therefore, the image should represent a projection (the total sum of electron densities along the beam rays) in the image plane produced. However, for the images to be considered as real projections, they must be corrected for CTF effects (*see* earlier discussion). Once the CTF correction is done and the Euler angles have been assigned to each projection image or class average, the 3D electron density for the particle can be determined.

Several approaches are used to calculate the 3D densities of the molecules from their projections. Since the current trend is towards complete automation of the image processing, two approaches are commonly used in EM owing to the efficiency of their

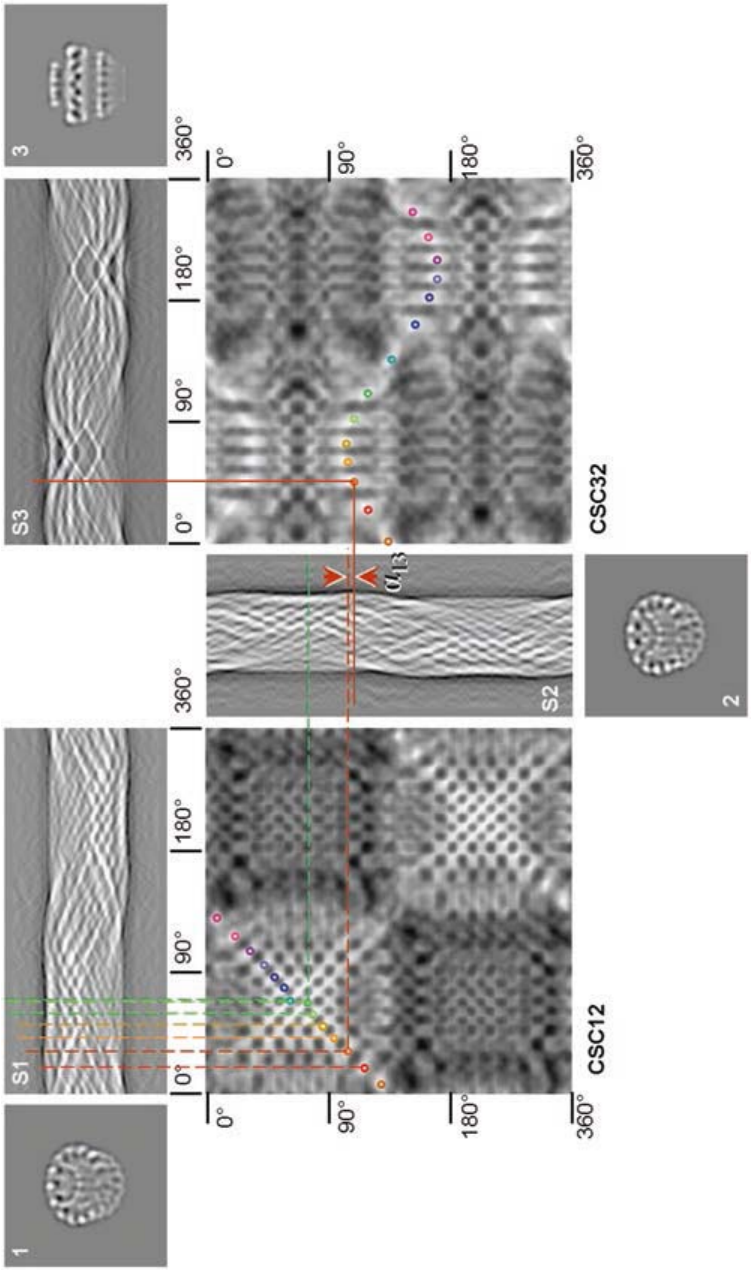

Fig. 13 Sinograms and sinocorrelation functions for three projections. A sample object is a model of the T4SS core outer-membrane complex, which has 14-fold symmetry. Images corresponding to projections are numbered 1–3. S1, S2, and S3 (sinograms) are sets of 1D projections of corresponding projections 1–3. CSC12 and CSC32 correspond to cross-sinogram correlation (CSC) functions between projections 1 and 2 and projections 3 and 2, respectively. Each point of the CSC function represents the correlation coefficient of a pair of lines from the two sinograms. There are 14 common lines between each pair of images since the object has 14-fold symmetry. The highest correlations (*rainbow circles*) point to the position of common lines. Some of them are shown by *dashed lines* between projections 2 and 1, and the *solid red line* shows one common line between projections 2 and 3 (the other are not shown). Each CSC function has all peaks doubled because projections from 180° to 360° mirror those from 0° to 180°. The angular distance between the common lines (*red solid* and *dashed*) shows the angle α_{13} between projections 1 and 3

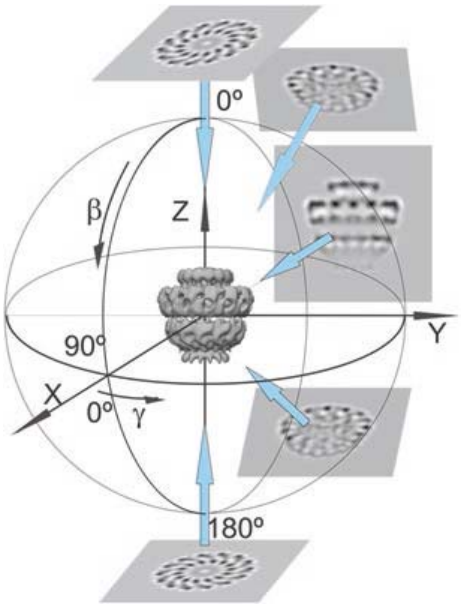

Fig. 14 3D reconstruction in real space. Different class averages of the core outer-membrane complex are shown and situated around the Euler sphere. Each class average is back-projected along its assigned Euler angle in real space. Densities of each 2D class average are stretched as rays through 3D space and the crossover points of the intersecting rays will sum together defining the 3D electron density of the entire object

implementations. In the first one, the reconstruction is calculated in real space and based on filtered back-projection algorithms; in the other one, reconstructions are performed using Fourier space [79–81].

Three-dimensional reconstructionin real space. These methods calculate the 3D distribution of densities in the space of objects. In EM, the exact filtered back-projection method [82] is used more often. In this method, each image of the data set (or classes) is stretched along a direction defined by the found orientations of images. Electron densities in three dimensions are obtained by the summation of rays from stretched projections. The electron density generated by these summations produces densities for each voxel. As more projections are included in the 3D reconstruction, the voxels become better defined (Fig. 14). The angular distribution of different images should evenly cover the Euler sphere or the asymmetric triangle (for particles with symmetry). This is essential to achieve an even representation of all details in the structure, or at least the set of images chosen for the reconstruction should have a distribution of angles that covers a great circle of the Euler sphere [83]. A patchy distribution of angles leads to the appearance of stripes in the 3D electron density map. To avoid an additional low-frequency background induced by the projection-stretching procedure, images are filtered in advance (although in some packages the filter is applied on the resulting 3D

reconstruction). A high-pass filtering of the input 2D projections corrects the overweighting of the low-frequency components, thereby restoring amplitude balance and, thus, minimisation of blurring. The exact filter algorithm used in IMAGIC computes a specific filter unique to each 2D projection [82, 84].

Three-dimensional reconstructionin Fourier space is based on a theorem which states that the Fourier transform of a 2D projection of a 3D object constitutes a central section of the 3D Fourier transform of the object [85]. This means that the Fourier transform of projections from different angular views can be merged to fill up the Fourier space with different 2D sections calculated from the images (or classes). Recovery of the 3D structure of an object in real space is done by reverse transformation of its 3D Fourier transform (Fig. 15) [85–87].

The size of trustworthy resolved details can be assessed using a formula derived by A. Cowther [86] assuming that projections are evenly distributed:

$$R = D / N$$

where N is the number of views, D the particle diameter, and R the target resolution. N can be significantly decreased for the same resolution if the complex has a high order of point group symmetry.

In Fourier space the large number of central sections used causes overlapping of the central sections near and at the point of origin. This leads to overweighting of the low-frequency components in the Fourier transform and therefore to an effect similar to a simple back-projection in real space, such as blurring of the reconstruction. Thus, currently used algorithms based on Fourier methods employ down-weighting of low frequencies or a high-pass map filter.

4.9 Structure Refinement

All single-particle EM packages use nearly the same procedure for the refinement of structures after the first 3D model is obtained (Fig. 1). It is done by a realignment of single-particle images with new references obtained from the new model. It is often combined with the determination of angles: projection matching with a smaller angular increment or a local refinement of the angles [88]. Reprojections of the new model can be used as a new anchor set in angular reconstitution to refine the orientation of the classes. The anchor set is a set of projections calculated from the first 3D model which are used to determine orientations for new classes or realigned images. The angular increment between projections used as the anchor set is usually chosen so that 100–150 projections (which is much less compared to projection matching) will be calculated and used for the refinement of angles. New Euler angles are found and assigned to the new class averages. Sorting classes

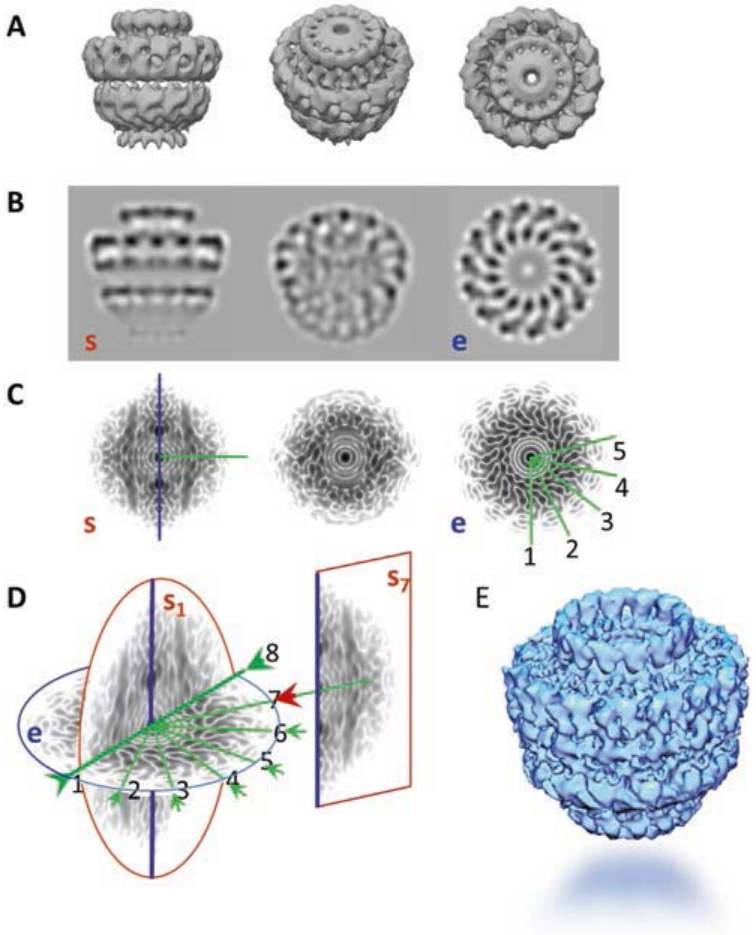

Fig. 15 Reconstruction using Fourier transform (FT) of classes (images). (**a**) T4SS core outer-membrane complex (14-fold symmetry) is observed in different orientations on the supporting film. (**b**) Projections of these particles that are equivalent to EM images in the direction of the electron beam. (**c**) FT from projections shown in **b**. *s* : FT of side view; *e* : FT of end view. The corresponding images are shown in **b**. (**d**) A pair of 2D transforms that share at least one common line in reciprocal space. Common lines between the side view (*s*) projection and the end view (*e*) projection are indicated by *green lines* ; *purple line*: common line between side views (shown in s_1 and s_7). The angles between pairs of common lines determine the relative Euler angle orientations. (**e**) An inverse Fourier transformation of the combined 2D transforms generates an improved real-space structure seen as electron density

according to the errors between classes and reprojections helps to facilitate the refinement procedure. While all implementations share the same principles of refinement, the details of the algorithms and the degree to which the user can control the process vary significantly [9].

4.10 Evaluation of Quality Structure

When the map of a new structure is obtained, the molecular mass and its oligomeric state should be estimated. The map should be inspected at the 1σ threshold of the density level, which typically corresponds to the molecular weight of the complex in the study. If the complex consists of several interacting proteins, the map

should not have disconnected fragments of densities; they should be continuous at densities well above the background noise. The concept of resolution is based on an assessment of the minimal distance between two points in an image at which they can still be distinguished from one another. That criterion was formulated as the Rayleigh criterion: when the centre of a peak of one point image falls exactly on the first zero of the image of the second point. In electron crystallography, the signal-related Fourier components of the image are linked to the reflections on a regular lattice, the *reciprocal lattice*, and a resolution is defined by frequencies of the reflections that are above the background noise and therefore available for Fourier synthesis [89]. This crystallographic resolution, R_c, and Raleigh's point-to-point resolution distance, d, are related by $d = 0.61/R_c$. How can this be done in an objective way in single-particle analysis? For several decades researchers have used a concept of Fourier shell correlation (FSC) (*see* following discussion). However, in recent years, owing to the tremendous achievements of single-particle cryo-EM, several modified approaches have been proposed and methods of resolution assessment have moved closer to criteria used in X-ray crystallography.

4.10.1 Fourier Shell Correlation and Gold Standard Approach

In single-particle analysis there is no well-defined periodical pattern in Fourier spectra. Common practice today is to look for data consistency by splitting the data set randomly in half and compare the two resulting averages (3D reconstructions). The resolution of a 3D map (the size of reliable details) could be assessed by FSC. Two Fourier transforms are calculated, and the corresponding spherical shells are compared using normalised cross-correlation as a function of spatial frequency (R) (radius in the Fourier space). The value of the cross-correlation is used to assess the frequencies (or size of details) at which these two maps start to differ. If the correlation falls below the 0.5 threshold, then the details are considered to be different. Currently, several criteria are used to determine the threshold used in FSC, and a threshold of 0.1432 has become rather popular [90, 91]. It should be noted that FSC resolution depends on how the data were split and which threshold of FSC is used for the assessment. Scheres and Chen proposed a "gold standard" method for the evaluation of structural quality [92]. According to this approach, the initial data set is divided into two halves from the very beginning, and two models are refined independently. As soon as the structures are obtained, the FSC curves can be determined as usual. The FSC between two independent reconstructions shows that when the gold-standard procedure is used, the resolution of the final results depends on how well the structures are aligned and what sort of mask was used to remove surrounding noise (Fig. 16). This separation of the data set into two equal subsets and the independent refinement helps to avoid bias towards the same model used on initial steps for the alignment and determination of angles.

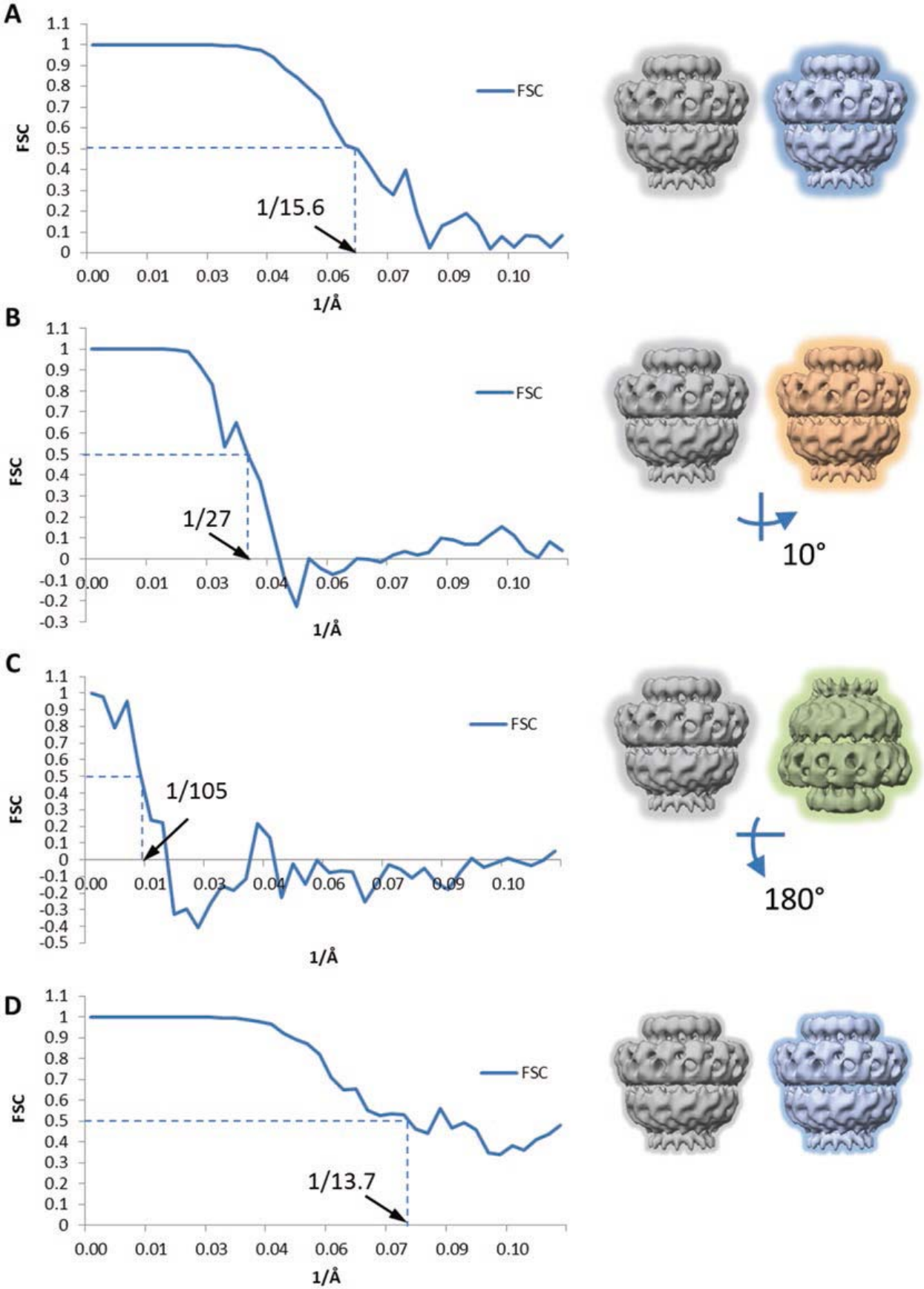

Fig. 16 Examples of Fourier shell correlation (FSC) and resolution assessments. (**a**) FSC curvature of two independent 3D structures that have been well aligned and masked with a loose soft mask shown as a wide halo around the particle images. A resolution assessed at the 0.5 level is 15.6 Å. (**b**) Here the second structure is

The disadvantage of the FSC is that for the resolution assessment one needs to split the data into two halves, but this reduces the resolution of the final 3D model since the entire image data set was not used for the same 3D reconstruction. The other approach to the evaluation of detail reliability reconstructed in 3D is the randomisation of phases (or one can use both randomisation of phases and amplitudes) above the frequencies critical for detail assessment. The main concept behind the approach is to modify an original data set of particle images in such a way that the amplitudes and phases beyond a certain chosen frequency are substituted by random values. This randomisation of phases (and sometimes amplitudes) at high frequencies is equivalent to replacing high-frequency structural details by noise. The modified data set is subsequently subjected to the same image processing procedure used for the original experimental data. The FSC between these two structures usually demonstrates a sharp drop at the same frequency where substitution of phases and amplitudes was carried out [93]. Any non-zero FSC values beyond the resolution where the noise was introduced reflects a level of bias during image processing. It is important to note that a data set with high-frequency noise contains little information about the real structure compared to the original data set, so the particle orientations may be less accurately defined and may affect the value of FSC at frequencies close to the threshold selected for the phase substitution. The behaviour of the FSC curves at high frequencies may also be affected by 3D masking. The FSC can be improved if the featureless regions are masked out. However, a tight mask with a very sharp boundary can produce strange artefacts in FSC such as rising up to the Nyquist frequencies, indicating that there are non-reliable details in the structure (Fig. 16).

4.10.2 Spectral Signal-to-Noise Ratio

The concept of resolution estimation using an assessment of the SNR in spectra of the reconstruction (SSNR) where an entire data set was used has been suggested by Unser and collaborators [94–97] and is similar to the *Q*-factor [98, 99]. The basis of the method is to measure the consistency between the input data for the reconstructed 3D map and a corresponding set of reprojections computed. The method estimates the relative energy

Fig. 16 (continued) rotated by 10° around the rotational axis. While the overall shape still coincides well between both structures small details are not in the register. It is reflected in the FSC curve which declines much faster and the resolution at the 0.5 threshold indicates only 27 Å, corresponding to a size of major domains in the structure. (**c**) The structures are not aligned. FSC falls down even earlier at lower frequencies indicating only consistency in overall sizes. (**d**) FSC between two 3D structures when the same tight mask was applied. The increase at high frequencies indicates the correlation between masks imposed on the structures, and here the resolution is overestimated

contribution of the reconstructed signal and noise components by calculating two independent reconstructions [96]. SSNR characterises the quality of the reconstruction as a function of the radial frequency. The bottom line is that one will only trust those signal frequency components whose energy is above what would have been obtained had the algorithm been applied to noise only.

4.10.3 Local Estimation of Resolution

The third approach whose popularity has risen in recent years analyses the 3D density maps obtained from a whole data set by computing the cross-correlation between neighbouring voxels in the Fourier domain, the Fourier Neighbor Correlation (FNC). A 3D mask is applied on a 3D structure in such a manner that values outside the mask are altered for zeroes and pixels remain unchanged only in part of the cube. Any density inside the structure mask counts as signal plus noise, whereas any density outside this mask counts as noise. This operation can be represented in Fourier space as a convolution of the Fourier transform of the mask with the Fourier transform of the noise. The convolution provides an assessment of correlations between the Fourier terms. The method has been implemented in a computer program called RMEASURE [100]. It is used for 3D reconstructions only.

The ResMap algorithm [101] is based on initialising a local sinusoid model at $r = 2d$, where d is the voxel spacing in angstroms (Å). Likelihood ratio tests are conducted at all voxels in the volume. At a fixed wavelength equal to d, the standard likelihood ratio test can detect whether a local sinusoid is a meaningful part of the model approximation. The test requires an estimate of the noise variance, which can be evaluated from the region surrounding the structure. The smallest r at which the likelihood ratio test passes at a given p-value defines the resolution. The p-value is the measure of whether the outcome of the attempt is due to an actual effect or a mere random chance. Voxels that pass the test are assigned a resolution r, while those that fail the test will be examined at a larger r. The algorithm produces a local resolution map with a number assigned to every voxel in the density map.

5 Interpretation and Fitting of Atomic Models

The final verification of the quality of the obtained EM map is done by an analysis of fitting or "docking" of known atomic structures, homologous atomic models, or results obtained by de novo tracing of the polypeptide chain into subunits or protein component domains (Fig. 17). In recent decades the resolution (the smallest reliable details in a structure) of most EM maps produced by single-particle EM analysis was between 20 and 30 Å. At such a low resolution large domains can be recognised according to their overall shape. But this level of detail does not provide sufficient information on the interaction between proteins and possible

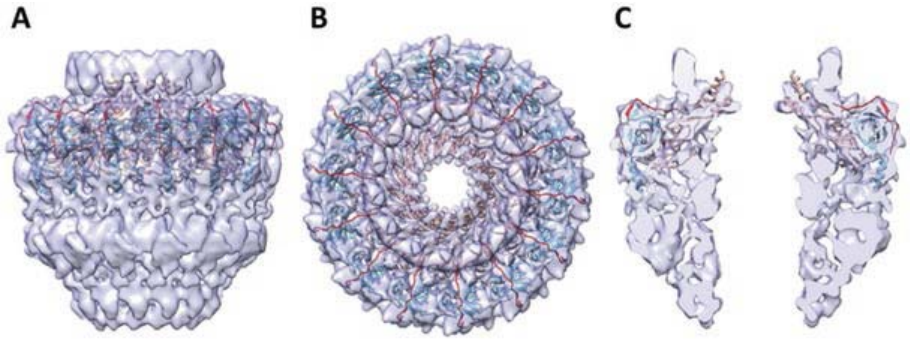

Fig. 17 Fitting of atomic model (PDB: 3JQO) to cryo-EM map (EMD-2232) of core outer-membrane complex. (**a**) Side view. (**b**) Top view. (**c**) Central cross section

active sites. The use of antibodies and different methods of labelling specific domains allow to localise domain positions and construct reasonable pseudo-atomic models based on EM maps; nonetheless, these results require substantial additional biochemical research to verify the interpretation. In subnanometer-resolution maps (6–9 Å range), densities corresponding to α-helices reveal characteristic cylindrical densities with a twist that makes more accurate fittings and determination of structural handiness. At the 4.5 Å level one can see a separation of strands in the β-layers. A resolution of around 4 Å reveals densities corresponding to large amino acid side chains [50, 102, 103], and at resolutions of around 3.7 Å one can do de novo tracing of a polypeptide chain using methods developed in X-ray crystallography [104].

An efficient approach is when pseudo-atomic models obtained by homology modelling are fitted into cryo-EM density maps to build atomic models of individual proteins. The basic principle behind the fitting procedure is the assessment of the correlation between a density map and the model. The maximum cross-correlation value between the EM map and modelled densities of the atomic structure indicates the best fit of the model. Such fitting is performed for maps at a resolution greater than 3.7 Å. Depending on the software used, a search can be carried out in reciprocal or real space. The initial stage of the fitting procedure is carried out either manually or automatically as a so-called rigid fit if a homologous atomic model exists. If there is no such a model, it can be built up with online homology servers like Phyre2 [105] or I-Tasser [106]. These models can be initially fitted into the EM density map using Chimera [107]; and then locations of flexible domains can be identified in Coot [108] or in a more automated way with FlexEM [109] or IMODFIT [110]. The flexible fitted structure can be optimised and checked for clashes using PHENIX [111]. This last step makes it possible to fix the positions of secondary elements simultaneously while conforming to geometrical restraints.

Recently, a new method, electron microscopy–iterative modular optimisation (EM–IMO) was published [112]. It endeavours to build, modify, and refine the local structures of protein models

using cryo-EM maps as constraints. A multi-parameter refinement strategy which combines EM–IMO and molecular dynamics with fine-tuning parameters allows one to build backbone models for the different conformations of proteins at near-atomic resolution in cryo-EM maps. The use of EM–IMO demonstrates that homology modelling and a multi-parametric refinement protocol offer a practical strategy for building atomic models based on medium- to high-resolution cryo-EM density maps [113]. Recent developments in single-particle cryo-EM now allow structures to be resolved at resolutions close to 3 Å. To facilitate the interpretation of EM reconstructions, X-ray packages such as REFMAC and PHENIX were modified for optimal fitting of atomic models to EM maps because external structural information can enhance the reliability of the derived atomic models, stabilise refinement, and reduce overfitting [104].

An atomic model obtained as a result of flexible fitting should be evaluated for its correctness and consistency with requirements of interactions between atoms. A Ramachandran plot [114] is routinely used to visualise the distribution of dihedral torsion angles. These angles of a polypeptide backbone are the most dominant local structural factors which dictate protein folding. Geometrical constraints and steric clashes between atoms of the main chain and side chains of each residue generate sometimes incorrect orientations of neighbouring amino acids. The angles are grouped into favoured and disallowed regions within the plot, which indicates the overall structural quality. Also, for each type of secondary structure, i.e. α-helix or β-sheet, there will be a characteristic range of allowed torsion angles, and these are shown on the plot.

6 Conclusions

Although impressive improvements have been made recently in the field of EM, many challenges remain in structural studies like large multi-protein complexes with low or no symmetry. Large complexes are typically flexible or can be unstable, so better approaches to dealing with sample heterogeneity are needed, which means more computer power will be required. Flexible, multi-domain protein structures could be uncovered by an iterative approach where their spatial organisation is determined by solving structures of domains step by step. Here, the largest domain could be tackled first, then “subtracted” computationally from the experimental image to get at the next largest domain; this process could then be repeated sequentially until the complete structure and overall architecture have been determined. Cryo-EM and methods of image analysis have become important and powerful tools in the analysis of biocomplexes. Recent advances in EM have greatly contributed to unravelling the so far elusive structural and mechanistic

details of Gram-negative bacterial secretion systems. Such a level of detailed information has not only led to a better understanding of the mechanistic traits underlying the process of substrate secretion through the T4SS apparatus but also given us the unique opportunity to visualise how this bacterial nano-machine is structurally organised within both outer and inner membranes. This valuable structural and mechanistic knowledge can now be used to design and develop new antibacterial compounds which target critical bacterial processes and could help us to curb the spread of pathogenicity and antibiotic resistance.

Acknowledgments

The authors thank Dr. H. White for reading the manuscript and useful discussions that led to improvements in the manuscript. This work was funded by MRC Grant MR/K012401/1 to E.V.O. The authors apologise for not covering all methods fully owing to space constraints.

References

1. Steitz TA (2008) A structural understanding of the dynamic ribosome machine. Nat Rev Mol Cell Biol 9(3):242–253
2. Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, Subramaniam S (2015) 2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. Science 348(6239):1147–1151
3. Bai XC, Yan C, Yang G, Lu P, Ma D, Sun L, Zhou R, Scheres SH, Shi Y (2015) An atomic structure of human gamma-secretase. Nature 525(7568):212–217
4. Vinothkumar KR, Zhu J, Hirst J (2014) Architecture of mammalian respiratory complex I. Nature 515(7525):80–84
5. Grigorieff N, Harrison SC (2011) Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. Curr Opin Struct Biol 21(2):265–273
6. Grant T, Grigorieff N (2015) Automatic estimation and correction of anisotropic magnification distortion in electron microscopes. J Struct Biol 192(2):204–208
7. Scheres SH (2010) Classification of structural heterogeneity by maximum-likelihood methods. Methods Enzymol 482:295–320
8. Orlova EV, Saibil HR (2010) Methods for three-dimensional reconstruction of heterogeneous assemblies. Methods Enzymol 482:321–341
9. Cheng Y, Grigorieff N, Penczek PA, Walz T (2015) A primer to single-particle cryo-electron microscopy. Cell 161(3):438–449
10. Costa TR, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M, Waksman G (2015) Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nat Rev Microbiol 13(6):343–359
11. Ilangovan A, Connery S, Waksman G (2015) Structural biology of the Gram-negative bacterial conjugation systems. Trends Microbiol 23(5):301–310
12. Fronzes R, Schafer E, Wang L, Saibil HR, Orlova EV, Waksman G (2009) Structure of a type IV secretion system core complex. Science 323(5911):266–268
13. Rivera-Calzada A, Fronzes R, Savva CG, Chandran V, Lian PW, Laeremans T, Pardon E, Steyaert J, Remaut H, Waksman G, Orlova EV (2013) Structure of a bacterial type IV secretion core complex at subnanometre resolution. EMBO J 32(8):1195–1204
14. Low HH, Gubellini F, Rivera-Calzada A, Braun N, Connery S, Dujeancourt A, Lu F, Redzej A, Fronzes R, Orlova EV, Waksman G (2014) Structure of a type IV secretion system. Nature 508(7497):550–553
15. Spence JCH (2003) High resolution microscopy, 3rd edn. Oxford University Press, New York

16. Dubochet J, Adrian M, Chang JJ, Homo JC, Lepault J, McDowall AW, Schultz P (1988) Cryo-electron microscopy of vitrified specimens. Q Rev Biophys 21(2):129–228
17. Jaffe JS, Glaeser RM (1987) Difference Fourier analysis of "surface features" of bacteriorhodopsin using glucose-embedded and frozen-hydrated purple membrane. Ultramicroscopy 23(1):17–28
18. Dubochet J, Lepault J, Freeman R, Berriman A, Homo JC (1982) Electron microscopy of frozen water and aqueous solutions. J Microsc 124(3):219–237
19. Lepault J, Dubochet J (1986) Electron microscopy of frozen hydrated specimens: preparation and characteristics. Methods Enzymol 127:719–730
20. Cabra V, Samso M (2015) Do's and don'ts of cryo-electron microscopy: a primer on sample preparation and high quality data collection for macromolecular 3D reconstruction. J Vis Exp 95:52311
21. Grassucci RA, Taylor DJ, Frank J (2007) Preparation of macromolecular complexes for cryo-electron microscopy. Nat Protoc 2(12):3239–3246
22. Adrian M, Dubochet J, Lepault J, McDowall AW (1984) Cryo-electron microscopy of viruses. Nature 308(5954):32–36
23. Baker LA, Rubinstein JL (2010) Radiation damage in electron cryomicroscopy. Methods Enzymol 481:371–388
24. Tivol WF, Briegel A, Jensen GJ (2008) An improved cryogen for plunge freezing. Microsc Microanal 14(5):375–379
25. Vos MR, Bomans PH, Frederik PM, Sommerdijk NA (2008) The development of a glove-box/Vitrobot combination: air-water interface events visualized by cryo-TEM. Ultramicroscopy 108(11):1478–1483
26. Jensen GJ (2010) Cryo-EM. Part A: sample preparation and data collection. Preface. Methods Enzymol 481:xv–xvi
27. Boyle WS, Smith GE (1970) Charge coupled semiconductor devices. J Bell Syst Tech 49(4):587–593
28. McMullan G, Cattermole DM, Chen S, Henderson R, Llopart X, Summerfield C, Tlustos L, Faruqi AR (2007) Electron imaging with Medipix2 hybrid pixel detector. Ultramicroscopy 107(4–5):401–413
29. Faruqi AR, Henderson R (2007) Electronic detectors for electron microscopy. Curr Opin Struct Biol 17(5):549–555
30. Ramachandra R, Bouwer JC, Mackey MR, Bushong E, Peltier ST, Xuong NH, Ellisman MH (2014) Improving signal to noise in labeled biological specimens using energy-filtered TEM of sections with a drift correction strategy and a direct detection device. Microsc Microanal 20(3):706–714
31. Veesler D, Campbell MG, Cheng A, Fu CY, Murez Z, Johnson JE, Potter CS, Carragher B (2013) Maximizing the potential of electron cryomicroscopy data collected using direct detectors. J Struct Biol 184(2):193–202
32. Cunningham IA (1999) Practical digital imaging and PACS. Advanced Medical Publishing for American Association of Physicists in Medicine, USA
33. McMullan G, Chen S, Henderson R, Faruqi AR (2009) Detective quantum efficiency of electron area detectors in electron microscopy. Ultramicroscopy 109(9):1126–1143
34. McMullan G, Faruqi AR, Clare D, Henderson R (2014) Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. Ultramicroscopy 147:156–163
35. Ruskin RS, Yu Z, Grigorieff N (2013) Quantitative characterization of electron detectors for transmission electron microscopy. J Struct Biol 184(3):385–393
36. Bammes BE, Rochat RH, Jakana J, Chen DH, Chiu W (2012) Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency. J Struct Biol 177(3):589–601
37. Milazzo AC, Moldovan G, Lanman J, Jin L, Bouwer JC, Klienfelder S, Peltier ST, Ellisman MH, Kirkland AI, Xuong NH (2010) Characterization of a direct detection device imaging camera for transmission electron microscopy. Ultramicroscopy 110(7):744–747
38. Campbell MG, Cheng A, Brilot AF, Moeller A, Lyumkis D, Veesler D, Pan J, Harrison SC, Potter CS, Carragher B, Grigorieff N (2012) Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. Structure 20(11):1823–1828
39. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA, Cheng Y (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat Methods 10(6):584–590
40. Abrishami V, Vargas J, Li X, Cheng Y, Marabini R, Sorzano CO, Carazo JM (2015) Alignment of direct detection device micrographs using a robust optical flow approach. J Struct Biol 189(3):163–176
41. Afanasyev P, Ravelli RB, Matadeen R, De Carlo S, van Duinen G, Alewijnse B, Peters PJ, Abrahams JP, Portugal RV, Schatz M, van Heel M (2015) A posteriori correction of camera characteristics from large image data sets. Sci Rep 5:10317
42. Glaeser RM (1971) Limitations to significant information in biological electron microscopy as a result of radiation damage. J Ultrastruct Res 36(3):466–482

43. Chiu W, Jeng TW (1982) Electron radiation sensitivity of protein crystals. Ultramicroscopy 10(1–2):63–69
44. Frank J (2006) Three dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state, 2nd edn. Oxford University Press, New York
45. Egerton RF, Li P, Malac M (2004) Radiation damage in the TEM and SEM. Micron 35(6):399–409
46. Burmeister WP (2000) Structural changes in a cryo-cooled protein crystal owing to radiation damage. Acta Crystallogr D Biol Crystallogr 56(Pt 3):328–341
47. Taylor KA, Glaeser RM (1976) Electron microscopy of frozen hydrated biological specimens. J Ultrastruct Res 55(3):448–456
48. Chiu W (1986) Electron microscopy of frozen, hydrated biological specimens. Annu Rev Biophys Biophys Chem 15:237–257
49. Knapek E, Dubochet J (1980) Beam damage to organic material is considerably reduced in cryo-electron microscopy. J Mol Biol 141(2): 147–161
50. Bartesaghi A, Matthies D, Banerjee S, Merk A, Subramaniam S (2014) Structure of beta-galactosidase at 3.2-a resolution obtained by cryo-electron microscopy. Proc Natl Acad Sci U S A 111(32):11709–11714
51. Carlson DB, Evans JE (2012) Low-dose imaging techniques for transmission electron microscopy. The transmission electron microscope. InTech, China
52. Erickson HP, Klug A (1971) Measurement and compensation of defocusing and aberrations by Fourier processing of electron micrographs. Philos Trans R Soc B 261(837):105–118
53. Wade RH (1992) A brief look at imaging and contrast transfer. Ultramicroscopy 46:145–156
54. Thon F (1966) Zur Defokussierungsabh ä ngigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung. Naturforschg 21a:476–478
55. Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ (2007) EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol 157(1):38–46
56. Rohou A, Grigorieff N (2015) CTFFIND4: fast and accurate defocus estimation from electron micrographs. J Struct Biol 192(2): 216–221
57. van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M (1996) A new generation of the IMAGIC image processing system. J Struct Biol 116(1):17–24
58. Wiener N (1964) Extrapolation, interpolation, and smoothing of stationary time series. Wiley, New York
59. de la Rosa-Trevin JM, Oton J, Marabini R, Zaldivar A, Vargas J, Carazo JM, Sorzano CO (2013) Xmipp 3.0: an improved software suite for image processing in electron microscopy. J Struct Biol 184(2):321–328
60. Smith JM (1999) Ximdisp–a visualization tool to aid structure determination from electron microscope images. J Struct Biol 125(2–3):223–228
61. Scheres SH (2015) Semi-automated selection of cryo-EM particles in RELION-1.3. J Struct Biol 189(2):114–122
62. Langlois R, Pallesen J, Ash JT, Nam Ho D, Rubinstein JL, Frank J (2014) Automated particle picking for low-contrast macromolecules in cryo-electron microscopy. J Struct Biol 186(1):1–7
63. Heymann JB, Belnap DM (2007) Bsoft: image processing and molecular modeling for electron microscopy. J Struct Biol 157(1):3–18
64. Roseman AM (2004) FindEM--a fast, efficient program for automatic selection of particles from electron micrographs. J Struct Biol 145(1–2):91–99
65. Van Heel M, Portugal RV, Schatz M (2009) Multivariate statistical analysis in single particle (Cryo) electron microscopy. In: Verkley A, Orlova E (eds) An electronic textbook: electron microscopy in life science. 3D-EM Network of Excellence, London
66. Grigorieff N (2007) FREALIGN: high-resolution refinement of single particle structures. J Struct Biol 157(1):117–125
67. Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol 116(1):190–199
68. Bartholomew DJ, Steele F, Galbraith J, Moustaki I (2008) Analysis of multivariate social science data. Statistics in the social and behavioral sciences series, 2nd edn. Taylor & Francis, USA
69. Myung IJ (2003) Tutorial on maximum likelihood estimation. J Math Psyc 47(1):90–100
70. Sigworth FJ (1998) A maximum-likelihood approach to single-particle image refinement. J Struct Biol 122(3):328–339
71. Scheres SH (2012) A Bayesian view on cryo-EM structure determination. J Mol Biol 415(2):406–418
72. Macqueen J (1967) Some methods for classification and analysis of multivariate observations. Proc Fifth Berkeley Symp Math Stat Prob 1:281–297
73. Ward JHJ (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58(301):236–244

74. Guan W, Lockwood A, Inkson BJ, Mobus G (2011) A piezoelectric goniometer inside a transmission electron microscope goniometer. Microsc Microanal 17(5):827–833
75. Van Heel M (1987) Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. Ultramicroscopy 21(2):111–123
76. van Heel M, Orlova EV, Harauz G, Stark H, Dube P, Zemlin F, Schatz M (1997) Angular reconstitution in three-dimentional electron microscopy: historical and theoretical aspects. Scanning Microsc 11:195–210
77. Crowther RA (1971) Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. Philos Trans R Soc Lond Ser B Biol Sci 261(837):221–230
78. Fuller SD (1987) The T=4 envelope of Sindbis virus is organized by interactions with a complementary T=3 capsid. Cell 48(6): 923–934
79. Herman GT (1980) Image reconstruction from projections: the fundamentals of computerized tomography. Academic, New York
80. Penczek PA (2010) Fundamentals of three-dimensional reconstruction from projections, vol 482. Methods in enzymology: Cryo-EM, part B, 3-D reconstruction. Academic, Elsevier, San Diego, CA
81. Orlova EV, Saibil HR (2011) Structural analysis of macromolecular assemblies by electron microscopy. Chem Rev 111(12):7710–7748
82. Harauz G, van Heel M (1986a) Exact filters for general geometry three-dimensional reconstruction. Optik 73:146–156
83. Orlov SS (1976) Theory of three dimensional reconstruction—conditions of a complete set of projections. Sov Phys Crystallogr 20:312–314
84. Radermacher M (1988) Three-dimensional reconstruction of single particles from random and nonrandom tilt series. J Electron Microsc Tech 9(4):359–394
85. De Rosier DJ, Klug A (1968) Reconstruction of three dimensional structures from electron micrographs. Nature 217(5124):130–134
86. Crowther RA, DeRosier DJ, Klug A (1970) The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. Proc R Soc A 317 (1530)
87. DeRosier DJ, Moore PB (1970) Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry. J Mol Biol 52(2):355–369
88. Penczek PA (2008) Single particle reconstruction. In: Shmueli U (ed) International tables for crystallography. Springer, New York, pp 375–388
89. Glaeser RM, Downing KH, DeRosier DJ, Chiu W, Frank J (2007) Electron crystallography of biological macromolecules. Oxford University Press, New York
90. van Heel M, Schatz M (2005) Fourier shell correlation threshold criteria. J Struct Biol 151(3):250–262
91. Rosenthal PB, Henderson R (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J Mol Biol 333(4): 721–745
92. Scheres SH, Chen S (2012) Prevention of overfitting in cryo-EM structure determination. Nat Methods 9(9):853–854
93. Chen S, McMullan G, Faruqi AR, Murshudov GN, Short JM, Scheres SH, Henderson R (2013) High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. Ultramicroscopy 135:24–35
94. Unser M, Trus BL, Steven AC (1987) A new resolution criterion based on spectral signal-to-noise ratios. Ultramicroscopy 23(1):39–51
95. Unser M, Trus BL, Frank J, Steven AC (1989) The spectral signal-to-noise ratio resolution criterion: computational efficiency and statistical precision. Ultramicroscopy 30(3):429–433
96. Unser M, Sorzano CO, Thevenaz P, Jonic S, El-Bez C, De Carlo S, Conway JF, Trus BL (2005) Spectral signal-to-noise ratio and resolution assessment of 3D reconstructions. J Struct Biol 149(3):243–255
97. Penczek PA (2002) Three-dimensional spectral signal-to-noise ratio for a class of reconstruction algorithms. J Struct Biol 138(1–2): 34–46
98. Kessel M, Radermacher M, Frank J (1985) The structure of the stalk surface layer of a brine pond microorganism: correlation averaging applied to a double layered lattice structure. J Microsc 139(Pt 1):63–74
99. van Heel M, Hollenberg J (1980) The stretching of distorted images of two-dimensional crystals. Electron microscopy at molecular dimensions. Springer, Berlin
100. Sousa D, Grigorieff N (2007) Ab initio resolution measurement for single particle structures. J Struct Biol 157(1):201–210
101. Kucukelbir A, Sigworth FJ, Tagare HD (2014) Quantifying the local resolution of cryo-EM density maps. Nat Methods 11(1):63–65
102. Zhang R, Alushin GM, Brown A, Nogales E (2015) Mechanistic origin of microtubule dynamic instability and its modulation by EB proteins. Cell 162(4):849–859

103. Clare DK, Orlova EV (2010) 4.6A cryo-EM reconstruction of tobacco mosaic virus from images recorded at 300 keV on a 4k x 4k CCD camera. J Struct Biol 171(3): 303–308
104. Brown A, Long F, Nicholls RA, Toots J, Emsley P, Murshudov G (2015) Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. Acta Crystallogr D Biol Crystallogr 71(Pt 1):136–153
105. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10(6):845–858
106. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER suite: protein structure and function prediction. Nat Methods 12(1):7–8
107. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera—a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612
108. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60(Pt 12 Pt 1):2126–2132
109. Topf M, Lasker K, Webb B, Wolfson H, Chiu W, Sali A (2008) Protein structure fitting and refinement guided by cryo-EM density. Structure 16(2):295–307
110. Lopez-Blanco JR, Chacon P (2013) iMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. J Struct Biol 184(2):261–270
111. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66(Pt 2):213–221
112. Zhu J, Cheng L, Fang Q, Zhou ZH, Honig B (2010) Building and refining protein models within cryo-electron microscopy density maps based on homology modeling and multiscale structure refinement. J Mol Biol 397(3):835–851
113. Lindert S, Alexander N, Wotzel N, Karakas M, Stewart PL, Meiler J (2012) EM-fold: de novo atomic-detail protein structure determination from medium-resolution density maps. Structure 20(3):464–478
114. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7:95–99