

Chapter 19

Single Particle Cryo-electron Microscopy and 3-D Reconstruction of Viruses

Fei Guo and Wen Jiang

Abstract

With fast progresses in instrumentation, image processing algorithms, and computational resources, single particle electron cryo-microscopy (cryo-EM) 3-D reconstruction of icosahedral viruses has now reached near-atomic resolutions (3–4 Å). With comparable resolutions and more predictable outcomes, cryo-EM is now considered a preferred method over X-ray crystallography for determination of atomic structure of icosahedral viruses. At near-atomic resolutions, all-atom models or backbone models can be reliably built that allow residue level understanding of viral assembly and conformational changes among different stages of viral life cycle. With the developments of asymmetric reconstruction, it is now possible to visualize the complete structure of a complex virus with not only its icosahedral shell but also its multiple non-icosahedral structural features. In this chapter, we will describe single particle cryo-EM experimental and computational procedures for both near-atomic resolution reconstruction of icosahedral viruses and asymmetric reconstruction of viruses with both icosahedral and non-icosahedral structure components. Procedures for rigorous validation of the reconstructions and resolution evaluations using truly independent de novo initial models and refinements are also introduced.

Key words Cryo-EM, Image processing, Icosahedral reconstruction, Asymmetric reconstruction, Near-atomic resolution, Virus

1 Introduction

Electron cryo-microscopy (cryo-EM), X-ray crystallography, and NMR are the three major methods for the determination of 3-D structures of biological samples. Although high-resolution 3-D structures of biological macromolecules are still mostly determined by X-ray crystallography, the rapid progress of both instrumental and computational technique developments has established cryo-EM as an indispensable method for studies of structure and dynamics of large macromolecular complexes and viruses [1–5]. Among the cryo-EM subfields, single particle 3-D reconstruction of icosahedral viruses has been leading the progresses towards near-atomic

resolutions (3–4 Å) and is now considered a preferred approach over X-ray crystallography [6].

Icosahedral virus particles were among the first biological specimens for which 3-D molecular structures have been solved using electron microscopy, image processing, and 3-D reconstruction [7]. Because of their large size, high symmetry, and availability in large quantities, icosahedral virus particles frequently have been studied structurally by single particle cryo-EM [8]. The improvements from subnanometer resolutions (6–10 Å) [9–11] to near-atomic resolution (3–4 Å) [1, 12–16] for icosahedral virus particles in recent years are remarkable. Not only the structures of a wide range of icosahedral viruses have now been determined in this resolution range by several groups [1, 12–16], but also there is a clear trend that near-atomic resolution reconstructions of icosahedral viruses are rapidly becoming routinely achievable. At these resolutions, full atomic models of protein subunits can be built for detailed structural analysis of virus assembly and maturation. Further improvements to reach beyond 3 Å resolution is also expected to occur soon. In addition to high-resolution icosahedral reconstructions, single particle asymmetric reconstruction of complex virus structures with both icosahedral and non-icosahedral features has become not only feasible but also contributed many novel structure observations in dsDNA viruses [17–23].

These progresses were made possible by advances in all aspects of cryo-EM study: instrumentation, image acquisition, image processing/3-D reconstruction algorithm and software, and computing resource. The following sections will describe each of these steps and discuss the protocols for both icosahedral and asymmetric single particle 3-D reconstruction in the order that parallels the pipeline for a typical cryo-EM study.

2 Materials

2.1 Specimen Preparation

1. General tools for cryo-EM such as tweezers, grid boxes, filter papers, and liquid nitrogen dewars and tanks.
2. 100 mM Tris or phosphate buffer (typically pH 7.0–8.0 with 50–100 mM salt).
3. 0.1 % (w/v) polylysine solution (e.g., Sigma P8920).
4. Pure organic solvent such as acetone, ethyl acetate, or chloroform.
5. Ion-exchange chromatography (e.g., Adenovirus Purification ViraKit™ from VIRAPUR).
6. Density gradient centrifugation (e.g., BECKMAN COULTER Optima™ series ultracentrifuge).

2.2 Cryo-EM Data Collection

1. 200–300 kV transmission electron cryo-microscope with field emission gun and low-dose kit. FEI or JEOL microscopes are most commonly used.
2. TEM grids. There are many types of grids available from various vendors. Quantifoil® grids from Quantifoil Micro Tools GmbH and C-flat™ grid from Protochips are two popular brands.
3. Glow discharge system for grid treatments (e.g., Quorum Technologies SC7620 or Q150T).
4. Plunge-freezing device (e.g., automatic FEI Vitrobot™, semiautomatic Gatan Cryoplunge™ 3, or homemade manual plunger).
5. Cryo-holder and cryo-transfer station (e.g., Gatan or Oxford Instruments).
6. Dry pumping station (e.g., Gatan Instruments).
7. CCD camera for digital recording (4K×4K or larger cameras).
8. Photographic films (e.g., Kodak Electron Image Film SO-163) and dark room.
9. Film scanner (e.g., Nikon Super CoolScan 9000 ED).

2.3 Image Processing and 3-D Reconstruction

1. Graphical workstation for particle selection and map visualization (e.g., Linux desktop).
2. Computer clusters for image processing (e.g., 64 bit Linux cluster).
3. Image processing and 3-D reconstruction software (e.g., jspr, EMAN2, and EMAN).

3 Methods

3.1 Sample Preparation

A homogeneous sample of sufficient concentration and quantity is important for high-resolution single particle 3-D reconstruction of viruses. Multiple biophysical methods including ion-exchange chromatography (e.g., Adenovirus Purification ViraKit™ from VIRAPUR), filter, and density gradient centrifugation are often used to separate particles by charge, size, and density. Dialysis is often used as the last step to exchange samples into a buffer that is more suitable for cryo-EM imaging but might not be necessarily suitable for long-term storage of the sample. To verify homogeneity, negative staining check could be used. Sample concentration is another important factor to consider. A good starting point is at approximately 10^{12} particle/ml or 1.0 mg/ml level concentration. One should pay special attention to the pH, salt concentration, and the additives of the buffer solutions. Typically, 100 mM Tris or phosphate buffer is sufficient for virus particles. Sucrose/glycerol, detergent, and high salt concentration should be avoided in the buffer.

3.2 TEM Grid Selection and Treatment

TEM grids are used to hold a thin layer of virus particles for TEM imaging. A few parameters should be considered when choosing grids.

3.2.1 Support Film

Both grids with continuous carbon support film and holey carbon film can be used for high-resolution data collection. Due to less background noise, holey carbon grids are preferred for most samples. However, due to particle surface properties and charge, many virus particles or even different states of the same virus tend to stay on the carbon film or along the edge of the holes. The easiest way to overcome this issue is to use a homemade continuous carbon grid or the Ted Pella ultrathin continuous carbon film (<3 nm) on holey carbon support film grid (300 mesh or 400 mesh). There are many ways to modify the carbon film surface properties to facilitate adsorption of different samples. By aging the grid, the carbon surface becomes hydrophobic. Glow discharging the grid using air will make the carbon surface more hydrophilic and negatively charged. A simple way to make the carbon surface hydrophilic and positively charged is treating the carbon surface with 0.1 % (w/v) polylysine solution after glow-discharging treatment. Simply apply a drop of polylysine solution on a grid, blot off the majority to leave a thin layer of solution on the grid, and then air-dry. This extra layer of polylysine will slightly increase the background noise, but is very effective to attract negatively charged particles onto carbon surface.

3.2.2 Grid Hole Size and Mesh Size

400 mesh grids are often the preferred choice. When holey carbon grids are used, the hole size should be just slightly larger than the recording medium (film or CCD) at the working magnification to improve mechanical stability and to minimize charging during exposure.

3.2.3 Homemade Holey Carbon Grid vs. Commercial Holey Carbon Grid

One big advantage for the commercial grids with patterned holes [24] is that automated data collection is possible due to their consistent hole size and pattern. However, homemade grids are cheaper with variable sized holes [25], which makes these grids a good choice for sample and freezing condition screening purposes.

3.3 Sample Freezing

To preserve native structure and to enhance tolerance for radiation damage, the virus samples should be embedded in vitreous ice first using fast plunge freezing and then imaged at liquid N₂ temperature [26, 27]. Sample freezing can be done in three ways: manual, semiautomatic, and automatic. In either way, the plunge freezing for high-resolution data collection needs to minimize the vitreous ice thickness (slightly thicker than the particle size), preventing crystalline ice formation, and to reduce ice contamination. A big advantage for semiautomatic and automatic cryo-plunger is the use of an environmental chamber to control humidity [28]. The FEI Vitrobot™ can also set blotting force as a parameter. Safety regula-

tions should be abided for infectious viruses according to their biosafety classes [29]. For example, biosafety level (BSL) 2 viruses should be plunge-frozen in a biosafety hood and BSL-3 viruses can only be frozen in a certified BSL-3 lab. The typical sample freezing procedure is shown as below.

1. Clean the grid using an organic solvent such as acetone, ethyl acetate, or chloroform to remove residual plastics used during grid production.
2. Glow discharge grids. Place grids on a clean glass slide with carbon side facing up. To increase hydrophilicity, glow discharge in air for 15–30 s at $2\text{--}3 \times 10^{-1}$ mbar with a current of 20–30 mA. Glow discharging with N_2 gas will make grid surface hydrophobic.
3. Optional grid treatment such as polylysine coating to modify carbon surface property.
4. Fill up liquid ethane and liquid nitrogen in the coolant container of the cryo-plunger. The proper temperature for liquid ethane should be around -173 to -178 °C. It is important not to overfill the coolant to avoid ethane and ice contamination.
5. Set up all parameters and humidity on the cryo-plunger. The parameters are different for different samples. Generally, one can start from one blotting on two sides, 3–5 s blotting time, and 100 % humidity.
6. Use the designated tweezers to carefully pick up the grid at the very edge and mount it on the cryo-plunger. Mount the filter papers on the blotting pads.
7. Retract the tweezers into the environmental chamber and wait until the humidity reaches 100 % (or desired value).
8. Apply 3–5 μl of sample on the grid from side window, then blot the excess solution, and plunge the grid into the liquid ethane (*see Note 1*).
9. Carefully remove the tweezers from the mount and quickly transfer the grid into liquid nitrogen surrounding the liquid ethane cup. If the grid will be used immediately, use a small piece of filter paper to suck off the extra liquid ethane on grid surface before transferring it into a grid box. Once the grid is transferred into the grid box, it can be kept in liquid nitrogen for long-term storage.

3.4 Imaging Conditions

Due to radiation damage, biological samples must be imaged under cryo-conditions using low doses. Depending on target resolution, particle size, and particle distribution, the combination of dose, magnification, exposure time, gun lens, aperture size, spot size, beam spread area, and defocus range should be planned at the

beginning of the session. To achieve high-resolution data collection, one should pay attention to the following conditions.

3.4.1 Temperature

Temperature should be monitored at all stages of data collection from sample loading to the end of the image data collection (except for the side entry-type cryo-holders because the mechanical stability of the sample will be affected by the cable connections). Before loading frozen grids (*see Note 2*), the side entry cryo-holder tip should be cooled down to $-180\text{ }^{\circ}\text{C}$. During data collection, the temperature should never rise above $-165\text{ }^{\circ}\text{C}$. Increased sample temperature risks phase transition from vitreous ice to crystalline ice.

3.4.2 Electron Dose

The typical dose for cryo-samples is $\sim 20\text{ e}/\text{\AA}^2$. However, at higher magnifications ($>60,000\times$), this dose is not enough for proper film exposure. In this case, $\sim 25\text{ e}/\text{\AA}^2$ is a balanced trade-off for increasing signals for most samples.

3.4.3 Magnification

Sufficiently large magnification should be used to ensure that the final image sampling is at least $3\times$ finer than the targeted resolution. For example, to reach 3 \AA resolution, the image sampling should be around $1\text{ \AA}/\text{pixel}$ or even finer.

3.4.4 Beam Size

The beam intensity, gun lens, and spot size should be adjusted accordingly to ensure a coherent (and ideally parallel if possible) beam illumination with a slightly larger illumination area ($10\text{--}20\%$) than the CCD or film.

3.4.5 Exposure Time

The typical exposure time is about 1.0 s (*see Note 3*).

3.4.6 Defocus

The images should be taken at underfoci (i.e., defocused) to increase the contrast. However, excessively large defocus will blur the images and high-resolution details. For high-resolution purpose, the defoci should be in the range of $0.5\text{--}3.0\text{ }\mu\text{m}$.

3.5 Microscope Alignment

To achieve optimal optic quality, the microscope should be carefully aligned at the beginning of data collection and monitored throughout the entire imaging session. Special attention should be paid to the following several critical steps.

3.5.1 Eucentric Height and Eucentric (Standard) Focus

The purpose of eucentricity (eucentric height at eucentric focus) adjustment is to ensure that the specimen is at the position that the optics (objective lens) are optimized and that there is minimal image shift while tilting the specimen. After a sample grid is loaded, the objective lens current is set to the default value (i.e., push “eucentric focus” or equivalent button) and then the image defocus is set to zero by mechanically adjusting the sample stage along the Z-axis to minimize the image contrast. This can be done at relatively low magnification ($\sim 3,000\times$) in search mode or at the

working magnification in exposure mode for better accuracy. The eucentric height (specimen z -shift) can also be set by wobbling the α angle (i.e., sample tilt angle) between $+30^\circ$ and -30° and adjusting the sample stage along the Z -axis to minimize the image movement.

3.5.2 Objective Lens Astigmatism Correction

For high-resolution imaging, objective lens astigmatism correction is the most important adjustment, which should be done as the last alignment step at the working magnification or at a higher magnification for better accuracy in exposure mode. The key here is to fine-tune the power spectra (FFT of the image) as circular as possible by monitoring the live FFT of CCD images. This should be done on a continuous carbon area with ice cleaned by the electron beam. Because the shape of the power spectra is more sensitive at small defocus, the adjustment should be done at small underfocus with only one or two Thon rings appeared on the live FFT image. Image binning should be avoided because the high-resolution information is lost after binning. If the Thon rings appear square, the objective lens aperture is likely to be contaminated and needs to be cleaned.

3.5.3 Coma-Free Alignment

The purpose for coma-free alignment is to further minimize the residual beam tilt. During coma-free alignment, the beam will be switched alternatively between positive and negative values of the same absolute tilt along an axis (X or Y). By monitoring slightly underfocused live FFT images of an electron-burned continuous carbon area, any residual beam tilt is minimized by fine-tuning the beam tilt to make both power spectra converge to similar shape. Repeat this alignment for both X - and Y -axis. Check objective lens astigmatism again after coma-free alignment. If necessary, repeat astigmatism and coma-free alignments iteratively.

3.6 Image Acquisition

Low-dose imaging mode should be used to minimize exposure of samples to electrons (*see* Fig. 1). First, the whole grid should be surveyed at low magnification (200 – $5,000\times$) in search mode with a low-intensity beam and large defocus to find areas with thin ice and good particle distribution. Mark these positions using a corresponding function on the microscope or computer. Once this is done, go to one marked position, and set up the magnification, illumination area, off-axis focusing distance, and direction angle for focusing mode. Then, change to exposure mode to set up the low-dose imaging conditions including spot size, illumination area, defocus value, and exposure time. Make sure that the dose is measured using an open grid area (e.g., any broken area or holes without ice). To avoid the lens hysteresis problem, one should cycle through the sequence of “search \rightarrow focus \rightarrow exposure” a few times before actual imaging and keep this order during the entire imaging session (e.g., do not

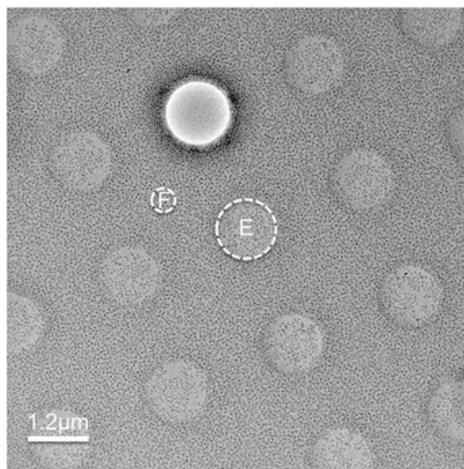


Fig. 1 Low-dose imaging. Shown is a low-mag search mode image of ice-embedded evenly distributed virus particles. The grid carbon bar region (*circle with “F” label*) will be used to determine focus at high magnification in focus mode. The final image of a hole with particles will be taken in exposure mode (*circle with “E”*). The grid is a C-flat 1.2/1.3 grid with a hole diameter of 1.2 and 1.3 μm hole spacing

go back to focus mode after exposure). During imaging, if several marked squares are close enough (e.g., 3×3 squares), there is no need to adjust eucentricity or alignment. Once the new imaging area is far from the original alignment area, one should recheck the eucentricity and astigmatism to make sure that they are not too much off. Once the imaging conditions are set up, the repetitive imaging through different holes can be done manually or automatically using appropriate software such as Legikon [30].

3.7 Data Quantity

The amount of data for different projects is different depending on sample properties (size, symmetry, surface feature, etc.) and targeted reconstruction resolution. For low-resolution (15 \AA and lower) 3-D reconstruction, a few hundred icosahedral virus particles are sufficient. Higher resolution reconstruction will need more particles with subnanometer resolution needing a few thousand particles. For near-atomic resolution 3-D reconstructions, about 50,000–100,000 icosahedral virus particles should be imaged. This amount of data is equivalent to about 500–1,000 films (or 2,000–4,000 4K \times 4K CCD images) for icosahedral virus of ~ 50 nm in diameter.

3.8 Image Digitization

For image processing and 3-D reconstruction, the image pixel values should be linear to the density projection of target structures. It is important that the digital images are consistent with this requirement. From electron optics and image formation theory of TEM instruments, the electron wave intensities at the imaging plane are proportional to the structural densities [31] and thus the

pixel values for digital camera (i.e., CCD and DDD) images and the optical densities (O.D.) of photographic films (i.e., film darkness) are also linear to the structural densities. The digital camera images can be used as is. However, extra attention should be paid to the photographic films as they must be further digitized using film scanners.

Many different scanners including both commercial products and specially designed scanners have been used in the cryo-EM field [32]. Currently, the most popular scanner is the Nikon Super CoolScan 9000 ED (*see Note 4*) which scans at 4,000 dpi (i.e., 6.35 $\mu\text{m}/\text{pixel}$). The scanner records the light intensities transmitted through the film and the saved image pixel values are thus the transmittance instead of the optical density values (*see Note 5*). The scanned images should be transformed so that the pixel values become linear to optical densities. This conversion can be done using the *log* transform of transmittance.

The convention for image contrast (i.e., brighter or darker pixels for particles relative to background pixels) is different for different image processing software. We adopted the positive contrast convention in which the particles should be brighter in graphic display (i.e., larger pixel values) than the background. The contrast can be inverted by multiplying each pixel value by -1 . For CCD images, the contrast needs to be inverted while the scanned image by a Nikon scanner is already in suitable contrast.

The following command will perform image format and O.D. conversion of scanned images (*see Note 6*):

```
nikontiff2mrc.py <input.tif> <output.mrc> --ODconversion=<0|1>
--invert=<0|1>
```

The following command will invert the contrast of CCD images:

```
e2proc2d.py <input.dm3> <output.mrc> --mult=-1
```

3.9 Image Processing and 3-D Reconstruction

Single particle 3-D reconstructions of icosahedral viruses (*see Notes 7 and 8*) were among the earliest EM reconstructions of biological structures. Many general software packages or specialized programs have been developed by different groups to perform all or a subset of the image processing and 3-D reconstruction tasks (*see http://en.wikibooks.org/wiki/Software_Tools_For_Molecular_Microscopy*). Comprehensive coverage of all the software packages is beyond the scope of this protocol. Instead, we will focus on image processing strategies and a set of programs that we routinely use in our projects. We use EMAN [33] and EMAN2 [34] packages as the general image processing software and have developed many additional programs for icosahedral and asymmetric reconstruction of viruses. If not specially explained, in this protocol, the programs starting with “*e2*” are EMAN2 programs while programs *jspr.py*, *jalign*, *j3dr*, *images2lst.py*, *nikontiff2mrc.py*, *filterMicrograph.py*,

goodImageSizes.py, *batchboxer.py*, *fitctf2.py*, *flipHand.py*, *common-Images.py*, and *symreduce.py* are developed in the Jiang lab.

As the entire image processing and 3-D reconstruction project consists of multiple tasks in different stages, the protocol is divided into a series of sections in an order parallel to that of image processing tasks carried out in an actual project. Each of the following sections will focus on one of the tasks and the overall image processing strategy will be summarized in the final section.

3.10 Particle Selection

For single particle cryo-EM, the particles in a micrograph must be individually selected and then saved for subsequent processing. While it is possible that particle selection is performed directly from the micrograph of original sampling, we prefer a three-step process that includes prefiltering of micrographs to enhance contrast, particle selection using the prefiltered micrographs, and particle output from original images.

3.10.1 Prefiltering of Micrographs

As cryo-EM images have low contrast due to low-dose imaging and small defocuses, the particles can often be difficult to detect. To enhance the contrast, multiple filtering steps are performed:

- *Binning*. This will remove noises in the high-resolution ranges. Binning also significantly reduces image size and speeds up subsequent selection process. The amount of binning depends on the initial sampling and we typically bin the micrographs by 4–6× so that the virus particle diameters are around 100 or fewer pixels in the binned images.
- *Low-pass filtering*. This further removes noise and enhances the particle contrast. Though the signals at higher resolutions are removed by binning and low-pass filtering, it does not affect particle selection as visibility and localization of particles are based mostly on low-resolution signals.
- *Gradient removal*. This will remove the overall density gradient in the micrograph caused by ice thickness variation, make the contrast of entire micrograph uniform, and eliminate the need of adjusting image display brightness/contrast parameters for different regions of the micrograph.
- *Removal of pixels with extreme values* from X-ray pixels in CCD images or dust on film during scanning. It will help image display with proper brightness/contrast and also improve the robustness of automated particle selection.

The following command will perform these filtering processes with some of the filtering parameters automatically determined using the specified particle diameter and starting image sampling:

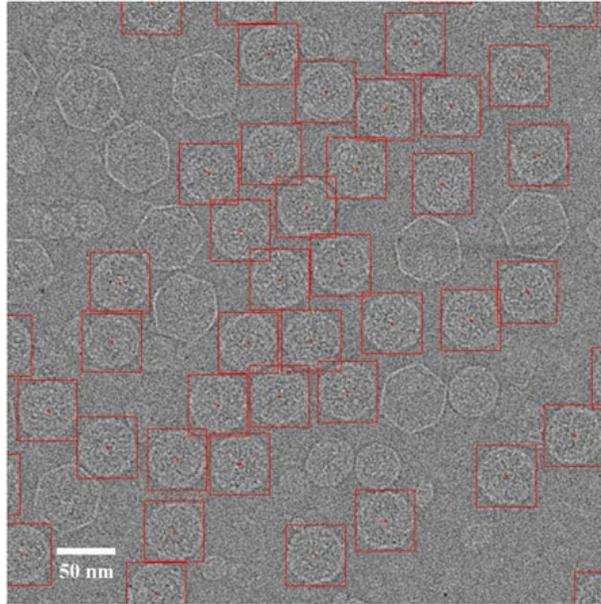


Fig. 2 Particle selection. A typical low-dose image of ice-embedded bacteriophage T7 capsid I particles imaged at $3.0\ \mu\text{m}$ underfocus using a FEI Titan Krios microscope at 300 kV with dose setting at $\sim 25\ \text{e}/\text{\AA}^2$. The *red boxes* indicate selected capsid I particles. The particles are more packed than optimal distribution but can still be effectively refined with appropriate methods (*see Note 9*). Contaminant larger capsid II particles and smaller vesicles are not selected since they can be easily distinguished from capsid I particles

```
filterMicrograph.py <micrographs> --diameter=<Angstrom> --apix=<
  Angstrom/pixel> --shrink=<n>
```

3.10.2 Particle Selection

The prefiltered micrograph images are then used for particle selection. The task is to locate good particles that are free of contamination and isolated from neighboring particles (*see Note 9*). While there are many automated particle selection methods [35], no automated method so far can be fully trusted. In practice, a hybrid approach should be adopted in which manual screening is performed after automated selection.

The selected particles should be well centered to benefit subsequent 2-D alignment steps that determine the particle orientation and center parameters. While automated selection methods in general result in well-centered particles, manual selection often results in particles with significantly more spread of centering. To improve the centering for manual selection, an effective visual guide is to use particle box (or circle if the program allows) about the same size as the virus particle diameter so that the edges of well-centered particles are tangent to the boxes. Off-centered particles will protrude out of the box and can be easily detected by the user

(*see* Fig. 2). Use of this visual guide is facilitated by the separation of particle selection step from output of selected particles. When the two steps are combined in one task, boxes significantly larger than the particles likely lead to poorer centering for manually selected particles. Once the particles in the micrograph are automatically selected and then screened, only the particle locations (i.e., the x, y coordinates for the particle center in the micrograph) need to be saved.

The following command will launch the graphic interface for particle selection that supports both automated and manual selection:

```
e2boxer.py <input filtered micrographs>
```

3.10.3 Particle Output

The saved particle center coordinates, after being adjusted for the amount of binning in prefiltering, can then be used to “cut” the particles from original micrographs. In this step, the box size can be changed to allow sufficient padding (typically 25–50 %) around the particles. The box size should use a “good” number (*see* **Note 10**). Square boxes should be used unless the software explicitly states support for rectangular images. The individual particle images should be normalized (i.e., set mean to 0 and variance to 1) to make particles from micrographs of different ice thickness have similar pixel value ranges.

The following command will “cut” individual particles with particles from the same micrograph saved in same file in HDF format while a separate image file for each micrograph (*see* **Note 6**). The particle location and source micrograph are also recorded as metadata in the HDF file.

```
batchboxer.py <input particle location files> --scale=<n> --removeS  
peckle=1 --normalize=1
```

3.11 CTF Determination

TEM images are modulated by the contrast transfer function (CTF) of the objective lens in TEM optical system [31, 36, 37]. Since the CTF functions cannot be precisely preset and can vary significantly among different images, it is critical to accurately determine the CTF parameters for each micrograph for proper corrections and to reach high-resolution 3-D reconstructions.

Among many of the parameters (defocus, B-factor, noises, etc.), the defocus value (*see* **Note 11**) is the most critical parameter to be determined. The signature of CTF is its oscillatory nature [38] which can be seen easily as the Thon rings [39] in the image power spectra (*see* Fig. 3). The Thon rings oscillate more frequently at larger defocuses and less at smaller defocuses (*see* Fig. 4). This relationship between defocus and Thon rings is the basis of both manual and automated CTF fitting methods.

The image power spectra can be computed from either selected particles [40] or directly from the micrograph using grid boxing

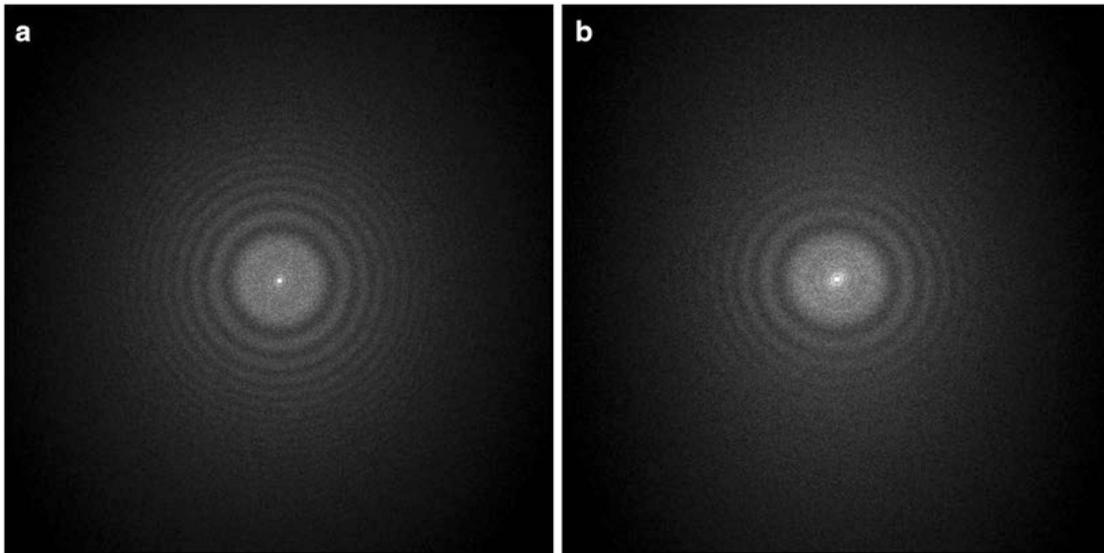


Fig. 3 Image power spectra. (a) Isotropic Thon rings with minimal astigmatism; (b) elongated Thon rings indicating significant astigmatism. The defocus is smallest along the most elongated direction

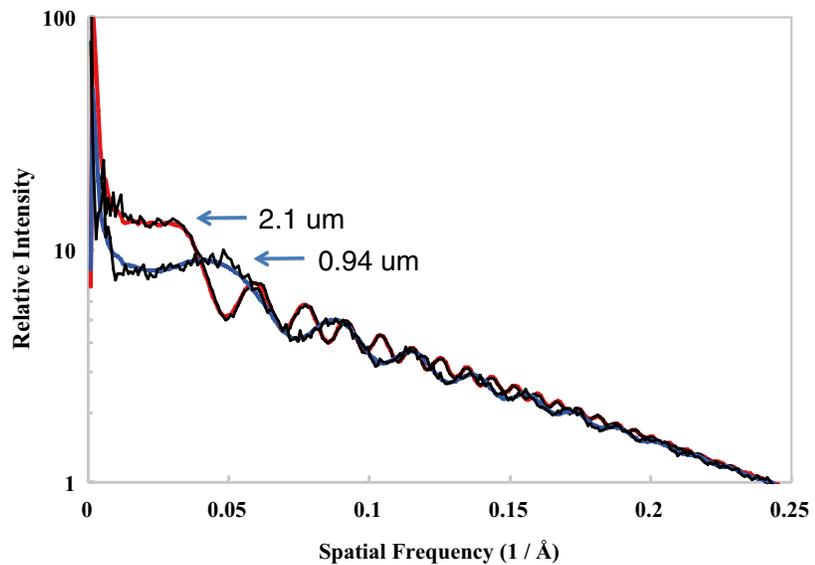


Fig. 4 CTF fitting. Shown are well-fitted CTF curves of two micrographs with a defocus of 2.1 μm and 0.94 μm , respectively

[41]. We suggest the former approach as it will enhance the Thon rings by avoiding the micrograph background areas empty of particles. Since the background areas with vitreous ice do not have strong Thon rings, their inclusion will weaken the average power spectra and contribute negatively to the reliability of CTF fitting.

To efficiently process a large number of images needed for high-resolution 3-D reconstructions, we suggest a two-step CTF fitting process that starts with automated fitting [41–46] and then

completes with interactive graphic screening/verification. The reason for the need of final manual screening in practice is due to the inevitable occasional failures of automated fitting methods [42] and lack of reliable detection of these failures by the automated methods. In our recent automated fitting method, we have integrated four different automated fitting algorithms which can cross-validate the fitting results and only prompt users with the inconsistent results among the methods for further manual screening [42]. Once CTF fitting is completed, the determined CTF parameters should be saved to the image headers that are needed for subsequent CTF correction and further refinement.

The following commands will perform each of the three tasks:

- Automated fitting (*see Note 12*)


```
fitctf2.py <input particle images> --cs=<mm> --voltage=<kV>
--apix=<Angstrom/pixel> --oversample=<n>
```
- Interactive screening


```
fitctf2.py<input particle images>--screenCtf=1
```
- Setting CTF parameters to image header


```
fitctf2.py<input particle images>--setParm=1
```

In these commands, we assume that the images are free of astigmatism, which is a reasonable assumption for a well-aligned microscope by experienced microscopists. Though many fitting methods including *fitctf2.py* are able to determine astigmatism, we typically leave it to later high-resolution refinement steps (*see* Subheading 3.15). Due to this postponed handling of astigmatism and further refinement of the focus values at the level of individual particles in later high-resolution refinement steps, we typically only record the CTF parameters fitted from power spectra as initial coarse values without CTF phase correction. Instead, CTF correction is performed as part of 2-D alignment and 3-D reconstruction tasks.

3.12 Image Quality Evaluation

Despite careful sample preparation, microscope alignment, and data collection, it is almost always true that the qualities of a subset of images are poor and these images should be discarded. Depending on the type of quality issues, poor quality images can be identified at different steps. For example, during film digitization step, images of very thick ice, large amount of ice contamination, very few particles, or severe charging/drifts do not need to be scanned. However, stringent image quality evaluation is generally evaluated using the image power spectra and often integrated as part of the manual screening/verification of fitted CTF parameters.

A high-quality image should have isotropic Thon rings in its 2-D power spectra and the Thon rings should extend to high resolutions (i.e., closer to the edge of 2-D power spectra). Elongated

Thon rings indicate a significant level of astigmatism (*see* Fig. 3). Astigmatic images should be discarded if the image processing and 3-D reconstruction software does not support determination and correction of astigmatism. In the image processing strategy described here, astigmatic images can be effectively utilized. When the Thon rings are significantly weaker along some direction than its perpendicular direction, the specimen has undergone significant drift or charging during exposure. These images should be discarded. Sometimes, a sharp peak can be seen at 3.7 Å in the 1-D power spectra, which is indicative of a significant level of ice contamination. Occasionally, we have also seen sharp peaks in the power spectra which were ultimately traced back to a malfunctioning scanner.

The Thon rings gradually become weaker at higher resolutions and the rate of weakening can be used to quantify the image quality. The quality factor is termed as B-factor (*see* **Note 13**) of which the value can be obtained during CTF fitting [40]. Smaller B-factors correspond to better quality images and B-factors of low-dose cryo-EM images of biological samples taken with modern TEMs with a FEG gun can reach values around 200 Å² or better. Studies of B-factor distributions have found that B-factors become slightly larger at large defocuses [40], so it is beneficial to collect images at smaller defocuses. In consideration of decreased image contrast at a smaller defocus, a defocus range of 1–2 µm should be a good compromise of the different factors for near-atomic resolution 3-D reconstruction of many viruses.

It is often observed that the image qualities can vary significantly even among images taken in same session. It is common practice to only select the best fraction of the images for inclusion in further image processing and 3-D reconstruction. The selection can be done using a graphic CTF fitting program such as *ctffit* in EMAN to examine the resolution of the last visible CTF peaks in the 1-D power spectra plot (*see* Fig. 4 and **Note 14**). Depending on the goal of the project, the resolution limit should be adjusted accordingly. For example, only images with CTF peaks beyond 6 Å will be included for projects aiming at near-atomic resolution (3–4 Å) reconstructions. However, it must be pointed out that the visibility of CTF peaks at high resolution is affected not only by imaging quality but also by other factors and the resolution cutoff should not be overly aggressive, especially for near-atomic resolution projects. For example, a small number of particles in a single micrograph might not have sufficient scattering power for clearly visible CTF peaks at high resolutions. Another common reason is the focus variations due to either residual astigmatism, small local tilt of the sample, or simply the particle *Z*-position variations in thick ice. Averaging of the power spectra of different focuses will effectively accelerate the apparent decay rate of CTF peak heights and even introduce zero CTF peak heights at a certain resolution beyond

which weak CTF peaks become visible again [47]. These second groups of CTF peaks are out of sync with the first group of CTF peaks at lower resolutions, and their peak positions cannot be simultaneously fitted with a single defocus. These focus variations can be effectively determined and corrected with proper 2-D alignment methods (*see* Subheading 3.15). From these analyses and experiences with many datasets, we suggest that cryo-EM image qualities are often better than what the highest resolution CTF peaks or the fitted B-factors indicate. It is now a common observation that 3-D reconstructions can reach a resolution significantly beyond the apparent limit of the images.

The particle images that pass the quality screening are deemed good and will be included for subsequent image processing and 3-D reconstruction. All these particle images can be pooled into a single large dataset using the following command:

```
images2lst.py<input image files><output data set.lst>
```

3.13 Initial Icosahedral Model

As single particle cryo-EM images are 2-D projections of the to-be-determined 3-D structure at random views, the inverse problem is to determine the 3-D structure from these 2-D images using computational image processing methods. Current image processing methods rely on iterative processes in which the 3-D reconstruction is iteratively improved. It is critical that the initial 3-D model is correctly constructed before proceeding to full refinement. If an initial model is already known from earlier studies of the same virus or its homolog, that initial model can be used (*see* **Note 15**) and the user can skip the step on building the initial model in this section. For a project on a new virus structure, several methods have been developed in the field to build a *de novo* initial model:

- *Self-common line method.* This is the classic method that takes advantage of the icosahedral symmetry of many virus structures. Based on Fourier central section theorem [48], the 2-D Fourier transform of a projection image is equivalent to a central section of 3-D Fourier transform of the original 3-D structure and the central sections of different views intersect at a common line through the Fourier origin [49–51]. Due to icosahedral symmetry, the symmetry-related views of the same particle can have up to 37 pairs of common lines, which are often termed as self-common line (*see* **Note 16**). The unique property of this method is that the information from a single image is self-sufficient to determine the particle orientation without need of a reference. The weakness of this method is that the common lines are clustered for views near symmetry axes that introduce biases [52]. This method also only works well for images with large defocus values and good

contrast, which was the underlying need for a focal pair imaging strategy, often used in the past [51]. We now seldom use this method.

- *Symmetry view method.* This EMAN method intentionally searches for particle images with best five-, three-, and twofold symmetry characteristics and uses these particles to construct the first crude 3-D model that will be further refined. This method is available in the EMAN program *starticos*.
- *Synthetic model method.* An icosahedral shape geometry can be computationally synthesized that approximates the size, angularity, and shell thickness of the target virus based on visual information from the particle images. We implemented this method via the processes *mask.icos* and *mask.dodecahedron* that can be executed via *e2proc3d.py* program.
- *Random de novo model method.* In this method, random orientation parameters (i.e., Euler angles; see **Note 17**) are assigned to the particles for reconstruction of the initial 3-D density map model. Since the view parameters are random, the initial model is essentially a spherical average without any meaningful surface features. However, the model does reflect the average size and shell thickness of the viral structure which are often sufficient for subsequent iterative refinements to converge to correct structure (see Fig. 5) [53, 54]. This is our preferred method for building initial model.

In this protocol, we will focus on the random de novo model method. In general, only a small dataset at coarse sampling (via binning of the particle images) is needed for building the initial model. We typically bin the images 4× so that the sampling of the binned images is around 4–6 Å. The binning not only reduces the image size to speed up computation but also effectively enhances the image contrast for more robust initial model building. The following command will bin the images:

```
e2proc2d.py<input particle file><output particle.hdf>--meanshrink=<n>
```

A small number of particles (100–200) randomly selected from the entire dataset are typically used for initial model building:

```
images2lst.py<whole dataset.lst><subset.lst>--randomSample=<n>
```

Particles of homogenous conformation and good contrast will enhance the success rate of the random de novo model method. If the quality or contrast of the randomly selected particles consistently cause initial model building to fail, the user can manually screen the particle images and select particles at larger defocuses and with higher contrast. This can be done by running

```
e2display<whole dataset.lst>
```

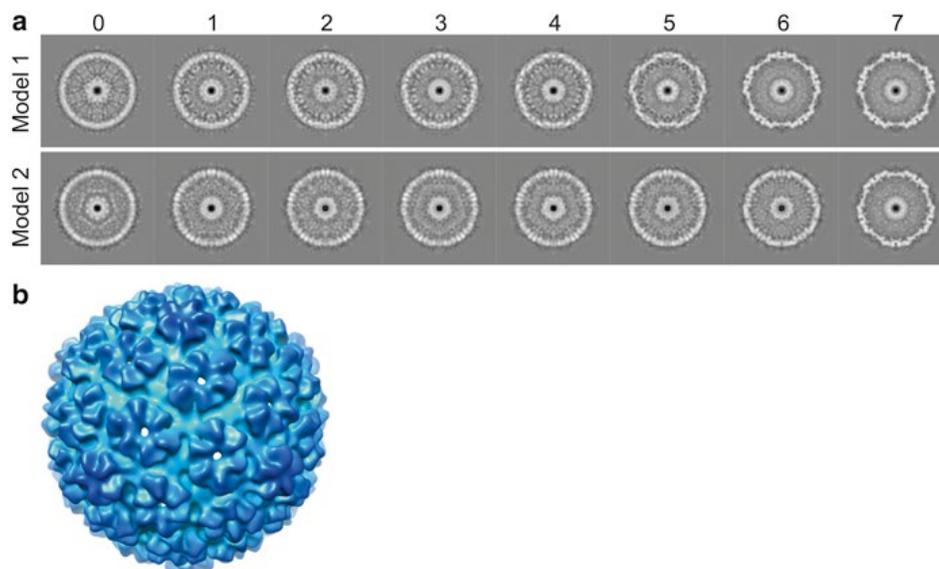


Fig. 5 De novo icosahedral reconstruction initial models. **(a)** Shown are the convergence histories for two random initial model refinements for bacteriophage T7 capsid I. The numbers at the *top* are the refinement iterations. The initial icosahedral model (iteration 0) was built using randomly assigned particle orientations. The two de novo models with opposite handedness illustrate that single particle cryo-EM 3-D reconstruction cannot uniquely determine the absolute handedness. **(b)** The threefold surface view of model-1 that is radially colored

to display the particles and then using the “Sets” tab function in its control widget.

The random de novo model construction and refinement can then be performed using the following command:

```
jspr.py<particles.lst> --nRepeat=<n> --sym=icos --diameter=<Angstrom>
--apix=<A/pixel> --iters 8 --cpus=<n>
```

As the random de novo model method starts from random parameters and some of the starting points might be too distant for the iterative refinements to converge to the correct solution, it might be necessary to repeat the random de novo model method multiple times (with different random initial views assigned to the particles) to obtain a correct initial model. The parameter `--nRepeat=<n>` can be used to specify the number of repeats. The multiple independent de novo models are also useful for removing bad particles in subsequent full dataset refinement (*see* Subheading 3.14).

In single particle cryo-EM, the image processing programs can always produce 3-D density maps, though currently no method can either guarantee that these density maps are correct models or offer a score to accurately quantify the reliability of the models. However, for a new project with a completely unknown target structure, a critical decision on whether the model is correct must be made. Here, we list a few criteria that can help make the decision:

- Consistent models (ignoring potential difference in handedness; *see* Fig. 5 and **Note 18**) from multiple random de novo model processes.
- Matching of model projections to experimental particle images using the determined view parameters.
- Appropriate level of structural details. For small datasets expecting only low-resolution reconstructions, correct models tend to have a small number of smooth large contiguous features while incorrect models tend to have too many discontinuous “details.”
- Presence of apparent holes or weaker densities along rotational symmetry axes, especially for high symmetries, such as icosahedral fivefold axis. This is from the physical principle that no two atoms can occupy the same location. Strong densities at symmetry axes would imply overlapping structures from the symmetry-related subunits.

3.14 Initial Icosahedral Refinement

Once the correct initial icosahedral model is built using a small subset of particle images, the initial model can be used to start the iterative refinement in which the 3-D model is improved and the orientation and center parameters of all particles are determined. We typically divide the iterative refinement into two phases, the initial coarse level refinement and the high-resolution refinement.

The initial refinement phase is to use global 2-D alignments to determine the approximate orientation and center parameters. It will also help decide which subset of particles should be considered as “bad” particles to be excluded from further high-resolution refinements. The initial refinement can be done using binned images and similar parameters as those used for initial model building. We typically use projection matching methods as implemented in program *jalign* for 2-D alignment and direct Fourier inversion method as implemented in *j3dr* for 3-D reconstruction. For virus particles with a diameter in range of 400–600 Å, a 2 or 3° angular step size can be used for projection matching. For 2-D alignment, two algorithms, one based on autocorrelation function (ACF) [33] and one based on polar Fourier transform (PFT) [55], can often be used. The ACF method is faster but the PFT method is preferred for more packed particles. A more detailed command with these choices is:

```
jspr.py <particles.lst> --sym=icos --diameter=<Angstrom> --apix=<A/pixel> --cpus=<n> --iters=13 --iter0 10 --initModels=<initial model map> --aligner=jalign --alignerOption "--superaligner gridSearch OrientationCenter:maskradius=<n> --orientgeneman:delta=2:inc_mirror=0:perturb=1 --aligner pft:flip=1:rmax=<n> --preprocess filter.ctf:type=1 --projector standard --cmp ccc --reconstructor j3dr
```

```
--reconstructorOption "--reconstructor fourier:mode=gauss_2:ctfweight=1 --preprocess filter.ctf:type=1 --preprocess xform.centerbyxform"
```

In this command, *filter.ctf:type=1* will perform CTF phase correction during both 2-D alignment and 3-D reconstruction, while *ctfweight=1* will perform amplitude-weighted CTF correction during 3-D reconstruction.

Similar to the need of a decision on whether the initial model is correct, a decision is also needed for the orientation and center parameters for each of the particles in the dataset after the initial refinement. The 2-D alignment task, regardless of the detailed algorithm being used, essentially just performs a ranking of different putative orientation/center combinations and returns one or a few best-ranked combinations. However, the best-ranked solution is by no means guaranteed to be correct. In general, the results for large defocus and high contrast images are more reliable than low contrast small defocus images. The alignment scores associated with the results are usually not robust enough to reliably discriminate good particles with correct solution from “bad particles” (see **Note 19**). In addition, the alignment scores not only depend on the orientation/center parameters but also are sensitive to defocusses, noises, contaminants, and neighbor particles which make it difficult to have a rational choice of score cutoff for correct alignment solutions. Our practical approach is to perform multiple independent alignments of particles using independently constructed de novo models and only keep those particles with consistent solutions among the multiple sets of independent solutions. The rationale for this approach is that:

- Alignment of good particles will be dominated by the consistent structural features in these models and independent alignment solutions should result in consistent solutions.
- Neither structural features nor noises in the models are consistent with the “bad particles” and the alignment solution is an essential random assignment. Since the noises in the independently built de novo models are not correlated, the random-like assignments by the multiple sets of independent alignments most likely will be uncorrelated and will be significantly different.

A potential concern of this approach is the increased computation for the repeat of refinements. This concern can be effectively alleviated: First, binned (4–6×) images are sufficient for initial alignment. The reduced image sizes will greatly speed up both data IO and computation time; second, a small number (e.g., 3) of refinement iterations are sufficient; and third, large angular step

size ($2\text{--}3^\circ$) for relatively low resolution ($20\text{--}30\text{ \AA}$) is sufficient which leads to greatly reduced computational need.

The following command will find the subset of particles with consistent orientation/center parameters among multiple sets of alignment results:

```
commonImages.py <set 1.lst> ... <set n.lst> <common subset.lst> --sym=  
icos --maxAngDiff=<degrees> --maxCenDiff=<pixels>
```

3.15 High-Resolution Icosahedral Refinement

The second phase of iterative refinements will further improve the 3-D reconstruction resolution to the limit allowed by the dataset. The high-resolution refinements will only use the subset of good particles found in initial refinement phase. Using the particle orientation and center parameters determined in the initial refinement phase as a starting point, the refinement will perform local gridless optimization to improve the particle parameters to higher accuracy (sub-degree for orientations and sub-pixel for centers). Furthermore, additional parameters such as astigmatism, defocus, and magnification will also be refined. Therefore, high-resolution refinements will go beyond orientation/center refinement with five parameters (3 for Euler angles and 2 for center) and include additional parameters (2 for astigmatism, 1 for defocus, and 1 for magnification) for the particles.

3.15.1 Refinement of Orientation and Center

The orientation accuracy requirement is dependent on the particle diameter and the target resolution $\left(\sim \frac{\text{resolution}}{\text{diameter}} \times \frac{180}{\pi}\right)$. For a virus of 700 \AA in diameter, an angular accuracy of 0.25° or higher will be needed to reach near-atomic resolutions ($3\text{--}4\text{ \AA}$). Such accuracy can be achieved by performing local refinement using numerical nonlinear optimization methods such as Nelder-Mead Simplex [56]. For each putative orientation, the corresponding projection can be efficiently computed by extracting the central section of the Fourier transform of the current 3-D density map, which explicitly uses the Fourier central section theorem [48]. The interpolation errors can be minimized by using the gridding method [57]. The following command shows how to perform the local refinement:

```
jspr.py <particles.lst> --sym=icos --diameter=<Angstrom> --apix=<A/  
pixel> --cpus=<n> --iters=23 --iter0 20 --initModels=<initial  
model map> --aligner=jalign --alignerOption "--aligner refine  
OrientationCenter:maskradius=<pixels>:steprot=<degrees>:ste  
pxy=<pixels> --projector fourierGriddingRSTM --cmp ccc  
--preprocess filter.ctf:type=1" --reconstructor j3dr --reconstructo  
rOption "--reconstructor fourier:mode=gauss_2:ctfweight=1 --pre  
process filter.ctf:type=1 --preprocess xform.centerbyxform"
```

in which the new options different from the initial refinement command are highlighted in bold.

The local refinement can be first performed using the binned images and then switched to the original particle images. The refinement parameters (center positions) and particle image file names can be converted using the following command:

```
images2lst.py <binned particles.lst> <original particles.lst>
  --multCenter=<n> --newdir=<directory of original sampling
  particle images>
```

3.15.2 Refinement of Astigmatism

As discussed in Subheading 3.11, cryo-EM images can still be mildly astigmatic, even for well-aligned microscopes. Minor astigmatism can often be ignored for low- to intermediate-resolution targets. However, for near-atomic resolution reconstructions (3–4 Å), the residual astigmatism should be accurately determined and then corrected. Here, we are concerned with only twofold astigmatism while higher order astigmatism will continue to be ignored. The amount of astigmatism (i.e., the defocus difference between the most elongated and the shortest directions of the Thon rings) and the direction (i.e., the most elongated Thon ring direction) can be determined by a 2-D search in which the particle image is multiplied by 2-D CTF function with putative astigmatism and then compared to the model projection. The astigmatism parameters refined this way should be more accurate than those determined based on elongated Thon rings in the power spectra. The following command shows how to perform the astigmatism search during 2-D alignment and correction during 3-D reconstruction (*see Note 20*):

```
jspr.py <particles.lst> --sym=icos --diameter=<Angstrom> --apix=<A/
  pixel> --cpus=<n> --iters=32 --iter0 30 --initModels=<initial
  model map> --aligner=jalign --alignerOption "---aligner refin
  eAstigmatism:batchsize=<n>:stepdfdiff=<um>:dfdiffrange=
  <um>:maskradius=<pixels>" --reconstructor j3dr --reconstruc
  tor Option "--reconstructor fourier:mode=gauss_2:ctfweight=1
  --preprocess filter.ctf:type=1 --preprocess xform.centerbyxform"
```

3.15.3 Refinement of Defocus

Thon ring-based CTF fitting methods typically assign the same defocus values to all particles in a micrograph with optional local tilt approximation considered by some methods [41, 58]. However, more stochastic variations of particle focuses, for example, from different Z-positions in thick ice, cannot be determined using Thon ring fitting. Another common source of small focus error is related to sample grid preparation using a thin continuous carbon film for particle support. Thon ring-based fitting will result in the average defocus of particles and the carbon film. If the entire micrograph instead of only selected particles is used to generate power spectra, the focus can even be skewed almost entirely to that of the carbon film. The particle focus can be further refined at individual particle level to eliminate these potential inaccuracies. The

focus refinement can be performed with a 1-D search in which the particle image is multiplied by the CTF function with putative defocus and current astigmatism parameters and then compared to the model projection. The defocus parameter refined this way should be more accurate than that based on Thon ring fitting. The following command shows how to perform the defocus search during 2-D alignment and correction during 3-D reconstruction (*see Note 20*):

```
jspr.py <particles.lst> --sym=icos --diameter=<Angstrom> --apix=<A/
pixel> --cpus=<n> --iters=32 --iter0 30 --initModels=<initial model
map> --aligner=jalign --alignerOption "---aligner refineMicrog
raphDefocus:batchsize=-1:stepdefocus=<um>:defocusrange=<u
m>:maskradius=<pixels>" --reconstructor j3dr --reconstructo
rOption "--reconstructor fourier:mode=gauss_2:ctfweight=1 --pre
process filter.ctf:type=1 --preprocess xform.centerbyxform"
```

3.15.4 Refinement of Magnification

It is well known that the true magnification of TEM instruments can have significant differences from the nominal magnifications provided by the microscope vendors. It is a common task to calibrate the magnifications of each instrument separately using standard specimens with accurate known feature sizes. However, the reproducibility of magnification by the same instrument is generally considered excellent. The magnification for images obtained from the same microscope at the same nominal magnification is typically considered constant for all images in a dataset. When such assumption is invalid, the relative magnifications of all images need to be refined. When datasets obtained from different microscopes are combined, their relative magnifications also need to be accurately determined before they can be properly merged. The TEM magnification can be affected by many factors, for example, lens current, lens hysteresis, and illumination conditions [59]. The magnifications can change after major services to the electron optical systems or sometimes from causes difficult to pinpoint. Since even a relative small magnification difference can lead to significant errors, for example, 3.5 Å from 1 % magnification change for a virus of 700 Å in diameter, the relative magnifications of all images need to be refined to high accuracy (<1 %) for the 3-D reconstructions of large viruses aiming at near-atomic resolution. The magnification refinement can be performed with a 1-D search in which the model projections are computed at putative relative scales and then compared to the particle images. The following command shows how to perform the magnification search during 2-D alignment and correction during 3-D reconstruction (*see Note 20*):

```
jspr.py <particles.lst> --sym=icos --diameter=<Angstrom> --apix=<A/
pixel> --cpus=<n> --iters=32 --iter0 30 --initModels=<initial
model map> --aligner=jalign --alignerOption "---aligner
refineScale:batchsize=-1:maskradius=<pixels>: stepscale=0.001
```

```
:scalerange=0.02 --preprocess filter.ctf:type=1" --reconstructor
j3dr --reconstructorOption "--reconstructor fourier:mode=gauss_
2:ctfweight=1 --preprocess filter.ctf:type=1 --preprocess xform.
centerbyxform"
```

3.15.5 Other Factors

In addition to the above particle parameters that should be refined for high-resolution 3-D reconstructions, several other factors should also be considered (*see Note 21*):

- *Quality of reference map.* The reconstructed map might not strictly follow icosahedral symmetry due to either reconstruction algorithm limitations or uneven data distributions. It is beneficial to reapply icosahedral symmetry in real space to “beautify” the 3-D map.
- *Quality of reference projections for 2-D alignment.* Since all the single particle cryo-EM iterative refinements rely on comparison of experimental particle images with projections of a 3-D model, improved quality of the reference 3-D map will be beneficial. The map generated by the reconstruction program (*j3dr* in this protocol) includes not only the icosahedral shell of the virus but also the internal genome and external background. These internal and external densities do not share same icosahedral symmetry as the virus shell and only contribute more noises to 2-D alignments. These densities should be masked out in the 3-D map before projections are computed. Note that the masking should always use soft masks instead of a sharp step cutoff to avoid masking artifacts.

The following command shows how to improve the reference 3-D maps for 2-D alignments of next iteration of refinement:

```
e2proc3d.py <input map> <output map> --process xform.
symmetrize:sym=icos --process mask.icos:radius3f=<pixels>:maskso
ft=<pixels>:imask=1:omask=0:curvature=<val> --process mask.
auto3d:nshells=<pixels>:nshellsgauss=<pixels>: radius=<pixels>:th
reshold=<val>
```

The same *e2proc3d.py* options can also be supplied to *jspr.py* as the *--post3dOption* parameter for inclusion in the automated iterative refinements.

3.16 Asymmetric Reconstruction

Icosahedral symmetry has been assumed so far for the virus particles and all the refinements and 3-D reconstructions have enforced icosahedral symmetry. In addition to icosahedral shell, many viruses also have structural components without icosahedral symmetry, for example, the portal vertex in tailed dsDNA bacteriophages. These structural components will be completely smeared in icosahedral reconstructions. In recent years, asymmetric reconstruction methods have been developed to determine the complete structure with

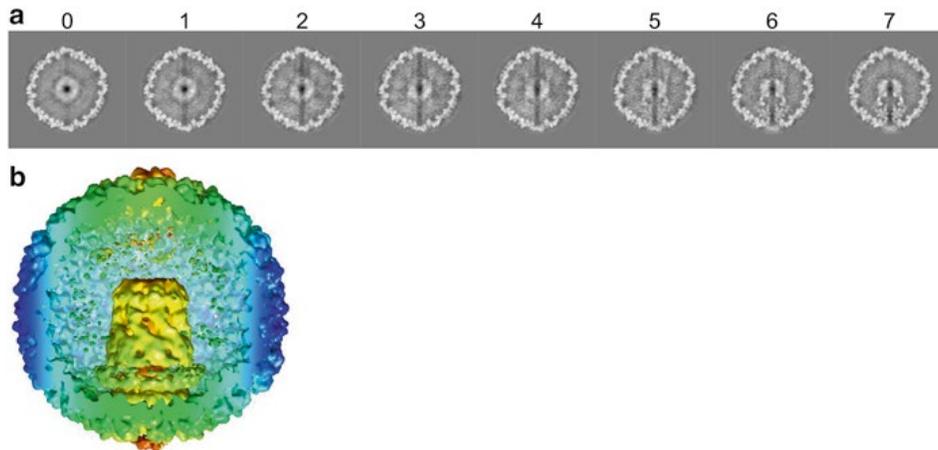


Fig. 6 De novo asymmetric reconstruction model. **(a)** Shown are the convergence histories for an asymmetric reconstruction refinement for bacteriophage T7 capsid I by relaxing the symmetry from icosahedral symmetry to C1 (i.e., no symmetry). The numbers at the top are the refinement iterations. The initial asymmetric model (iteration 0) was built from starting icosahedral orientation but randomly reassigned to one of the 60 views related by icosahedral symmetry. **(b)** The twofold surface view of the model that is cylindrically colored around the vertical axis. The icosahedral shell was cut open to reveal the internal core stack

both icosahedral shell and the non-icosahedral components simultaneously resolved [17–23]. Though the asymmetric reconstructions can in theory be performed directly using standard single particle reconstruction assuming no symmetry, the icosahedral features are often overwhelmingly dominant which makes it difficult to reliably determine the view of non-asymmetric features in this direct approach. A more successful asymmetric reconstruction strategy adopts a symmetry relaxation approach [17]. The virus structure is first determined as an icosahedral structure using methods as discussed in above sections. The icosahedral symmetry assumption is subsequently relaxed and the virus structure is reconstructed without enforcing any symmetry.

The symmetry relaxation step starts with the icosahedral orientations (i.e., Euler angles limited to one of the 60 asymmetric units) and aims to reassign each of the orientation to the best one of the 60 orientations related by icosahedral symmetry to maximize the matching of asymmetric reconstruction with experimental particle images. Similar to the random initial model for icosahedral reconstruction, the particle orientations can be randomized to one of the 60 icosahedral symmetry-related orientations. A 3-D model is then constructed without imposing any symmetry and used as initial asymmetric model for subsequent iterative refinements for asymmetric reconstruction. Alternatively, an initial asymmetric model can be synthesized by adding a featureless density blob to the expected location of asymmetric structural components to the icosahedral reconstruction. In the subsequent iterative refine-

ments, the 2-D alignment is limited to improving the reassignment of particle orientations to the best one of the 60 orientations related by icosahedral symmetry and the reconstruction is performed without imposing any symmetry. During iterative refinements, the icosahedral structure features will preserve their icosahedral symmetry while the asymmetric structural features will gradually emerge and become resolved (*see* Fig. 6 and **Note 22**).

The following command will perform random reassignment of particle orientations for initial random model generation:

```
symreduce.py <input eulers.lst> <output random eulers.lst> --sym=icos
--random
```

The following command will perform iterative refinements for asymmetric reconstructions:

```
jspr.py <particles.lst> --sym=c1 --diameter=<Angstrom> --apix=<A/
pixel> --cpus=<n> --iters=48 --iter0 40 --aligner=jalign --aligner-
Option "--superaligner symmetryRelax:startsymb=icos:endsym=
m=c1 --aligner dummy:maskradius=<pixels>" --reconstructor
j3dr --reconstructorOption "--reconstructor fourier:
mode=gauss_2: ctfweight=1 --preprocess filter.ctf:type=1 --prepro-
cess xform.centerbyxform"
```

3.17 Resolution Evaluation

In single particle cryo-EM, the resolution is evaluated by splitting the entire dataset into two halves (e.g., even and odd subsets), generating a 3-D reconstruction from each of the half datasets, calculating the Fourier shell correlation (FSC) curve between these two maps [60, 61], and then reporting the resolution at which the FSC curve becomes worse than a threshold (*see* **Note 23**). While this overall approach is a consensus in the field, several key details are still being hotly debated and the reported resolution number can vary significantly depending on how these details are handled. Here, we list a few of these details and our recommendations:

- *When to split data.* Almost all reported cryo-EM structures up to now have used a common 3-D model to iteratively refine the entire dataset until convergence. The dataset is only split after the final iteration of 2-D alignment. We will refer reconstructions generated from this late data split as *semi-independent reconstructions* instead of independent reconstructions as often stated in the literature. This approach has been shown repeatedly prone to model bias, overfitting to noises, and exaggerated high-resolution FSC curves [62–64]. We suggest that the data should be split after all preprocessing steps have been completed but before any initial model has been built. The two half datasets should never be mixed for 2-D alignment after the split. De novo models will be built independently for each half dataset. The subsequent iterative refinement of the half datasets should be confined within the respective half dataset and a

3-D model built from one half dataset should never be used to refine the other half dataset. The FSC will be calculated from the final 3-D reconstructions from each of the completely independent refinements. This complete separation ensures that no model bias is introduced in the refinement. Overfitting to noises might still occur within each of the half datasets but the overfitting should be independent and will not introduce artificial correlations between the 3-D reconstructions. We will refer the reconstructions generated from this strategy as *truly independent reconstructions* to distinguish them from those semi-independent reconstructions that were often inaccurately quoted as independent reconstructions in the literature. The FSC curves obtained from truly independent reconstructions have shown excellent resistance to artificial correlations from model bias and overfitting. The resolution number obtained from such FSC curve will more accurately represent the authentic resolution and quality of the map [62].

- *Threshold criterion.* Many threshold criteria such as 0.5, 0.143, and 3σ have been used in the literature [61, 65, 66]. We will use the 0.143 criterion as the SNR at this value for the map generated from the full dataset would be equal to one [66]. However, this point is valid only when the FSC is computed between truly independent 3-D reconstructions from half datasets split before refinement has started. It should be noted that it is inappropriate to adopt the 0.143 criterion for semi-independent reconstructions in which the refinements have used a single model to iteratively refine the entire dataset and the dataset is only split after many iterations of 2-D alignments. It might be tempting to use the 0.143 criterion even for semi-independent reconstructions so that a high-resolution number can be claimed. Unfortunately, many recent near-atomic 3-D reconstructions have used the 0.143 criterion even though they were based on semi-independent reconstructions [1]. In practice, it might be more appropriate to use the 0.5 criterion for FSC curves from semi-independent reconstructions to avoid overclaiming the resolution [12, 22].
- *Masking.* For large viruses, the icosahedral shell densities often only occupy a small fraction of the total volume while the majority of the volume is either the background (outside of the shell) or the encapsulated genome (inside the shell). Before the FSC curve is computed, the background and genome densities are often masked out so that only the icosahedral shell densities contribute to the FSC curves. However, it is important to avoid overaggressive masking which might inflate the FSC curves to a level that is dominated by the correlation between the masks instead of the structure features. In general, sharp masks should be avoided. Instead, masks with soft edges,

such as Gaussian or raised cosine, should be used. Adaptive masks that follow the shape of the structural features will also help minimize masking-induced artificial correlations. A well-behaved FSC curve should remain high (~ 1) at low resolutions, gradually decrease to zero, and then remain oscillating around zero across the entire spatial frequency to the edge (i.e., Nyquist resolution at $\frac{1}{2 \times \text{sampling}}$) [61]. If the FSC curve remains high at lower resolutions and then sharply drops to zero, it is a sign of strong model bias and inclusion of resolution range up to the sharp-drop resolution in the scoring function used for 2-D alignment. If the FSC curve does not decrease to zero or the FSC curve drops to zero but then increases above zero again, the density maps are most likely over-masked. When the FSC curve is plotted for publication or deposition, the entire range (from Fourier origin to Nyquist) of spatial frequencies should be included to faithfully report the resolution data [67]. Truncation before Nyquist might lead to doubt of the credibility of the FSC curve and the claimed map resolution.

Using *jspr.py*, the FSC curves will be computed when option `--eotest=1` is provided. When a single dataset is used for refinement, the FSC curves are computed from semi-independent reconstructions by splitting the dataset after 2-D alignment of last iteration refinement. When two or more independent datasets (e.g., the even and odd half datasets split before initial models are built) are used, the FSC curves are computed from the truly independent reconstructions from these independent datasets.

3.18 Map Sharpening

The density maps generated by reconstruction programs are rarely filtered optimally due to many reasons, for example, CTF correction and alignment errors [66]. The maps in general need to be further “sharpened” for subsequent structural analysis. The sharpening is generally performed in Fourier space by applying one or multiple Fourier filters to boost or suppress the Fourier amplitudes at different spatial frequencies (*see Note 24*). The sharpening process can be divided into two logical steps:

- *Boost signals at high-resolution frequencies* to a level approaching that of an ideal structure. An inverse Gaussian low-pass filter is generally used and the “B-factor” in this inverse Gaussian filter can be estimated from the decay slope of the calculated structural factors of the density map [66]. The corresponding slope for an ideal structure is approximately flat. If the structural factors are known from experimental measurements (e.g., X-ray solution scattering) [40] or computational fitting of multiple micrographs (<http://blake.bcm.edu/emanwiki/EMAN1/FAQ/>

StructureFactor), the structural factors can be used as a reference to boost the density map Fourier amplitudes.

- *Suppress noises at high-resolution frequencies.* A density map reconstructed from experimental data is necessarily worse than an ideal structure and it is not justified to boost the high-resolution Fourier amplitudes to the same level of an ideal structure [66]. It is thus necessary to apply another low-pass filter to dampen the high-resolution Fourier amplitudes to a level appropriate for the resolution of the map. This can be done with a standard Gaussian low-pass filter with its B-factor related to the resolution ($B=4 \times \text{resolution}^2$) or a low-pass filter derived from the FSC curve (i.e., the c_{ref} curve, $c_{\text{ref}} = \sqrt{\frac{2 \times \text{FSC}}{\text{FSC} + 1}}$) [66].

These two logical steps can be realized using a single inverse Gaussian filter if Gaussian filter is used in both steps,

```
e2proc3d.py <input map> <output map> --process filter.lowpass.
autob:bfactor=<Å²>
```

or using two consecutive Fourier filters if both structural factor and FSC curves are used in the two steps, respectively:

```
e2proc3d.py <input map> <output map> --process filter.setstrucfac:
sffile=<filename>:smooth=<n> --process filter.file:file=<fsc file>:cref=1
```

The density map voxel value range is often arbitrary and can vary dramatically depending on what filters and software are used. Though not strictly necessary to rescale the voxel ranges, it will make rational choice of threshold for visualization more convenient if the map voxel values are rescaled relative to some meaningful references. We typically renormalize the final sharpened density map by setting the background mean to zero and background variance to one. With this normalization, the voxel values will become equivalent to Z-score, i.e., the number of sigmas above noise level. This normalization can be performed using this following command:

```
e2proc3d.py <input map> <output map> --process normalize.mask.
circlemean: radius=<pixels>:ringwidth=<pixels>
```

3.19 Validation

As discussed in Subheading 3.13 on initial models, single particle cryo-EM image processing and 3-D reconstruction procedures will always produce some density maps though these reconstructions are not guaranteed to be valid. It has occurred in the field that drastically different reconstructions were reported for the same structure by multiple groups [67]. In this section, we will divide the validation task into three different levels according to the resolutions. Note the resolutions used for this division are somewhat arbitrary and mostly for the convenience of discussion.

3.19.1 Low Resolution (10 Å and Lower)

Validation at low resolution is basically the same problem as the construction of the correct initial model which is discussed in Subheading 3.13. For most icosahedral viruses, it is in general straightforward to obtain the correct initial model at low resolutions. Consistent structure among multiple de novo models using random model approaches is a reliable test if the structure is correct. Another useful validation method is the tilt pair method [66, 68]. In this method, a small dataset of tilt pair images is obtained and particle pairs are selected from the untilted and tilted images. The orientations of the tilted and untilted particles are compared and the orientation differences are tested for their consistency with the known tilt angle around the tilt axis. Since the tilt angle and axis information are not included in model construction and particle orientation determination, these information can be used as reliable validation references. The tilt pair validation method can be performed using the online server (<http://cryoem.nimr.mrc.ac.uk/software>) or *e2tiltvalidate.py*. For icosahedral viruses, another useful validation is the T number detected from the structure which should be consistent with that of viruses in the same family of known structures and the copy number estimated from biochemical analysis of the particle composition.

3.19.2 Intermediate Resolution (5–10 Å)

Once the reconstructions have been validated at low resolution, the overall architecture of the structure is then considered correct and validation at higher resolutions concerns mostly the structural details and if the observed structural details are consistent with the expected level of details at stated resolutions. At intermediate resolutions (also referred to as subnanometer resolutions), the intrinsic structural features, such as protein secondary structure elements, can serve as reliable validation references [69, 70]. For example, the alpha-helices should be clearly visible as rodlike densities and neighboring helices should be clearly resolved. Beta-sheets should also be visible as smooth curved thin platelike densities. Lack of such clearly resolved secondary structure elements strongly suggests that the resolution is lower than nanometer resolution. If a subnanometer resolution is suggested by the FSC curve, the reconstruction and FSC curve likely suffer from severe model bias and overfitting to noises.

3.19.3 High Resolution (5 Å and Higher)

As the resolution improves, finer structural details are expected [1, 12–16]. At 5 Å or higher resolution, often referred to as near-atomic resolutions, the alpha-helices should obtain obvious helicity with periodic surface bumps instead of just smooth rods. The strands in the beta-sheet should begin to be resolved. Densities from a few bulky side chains will become visible. The protein backbone should become visually traceable for most of the regions except for a small number of regions where the densities are less resolved due to intimate interactions or local flexibility of the struc-

ture. The C-alpha model can be constructed but accurate assignment of sequence identity remains challenging. When the resolution is higher than 4 Å, the increasing number of significant side-chain densities will allow more accurate assignment of the amino acids. The all-atom model can now be constructed though the rotamer conformations of most side chains are still uncertain until the resolution becomes closer to 3 Å at which many of the side-chain densities become sufficiently resolved to include atomic model of entire side chain. For a few best-resolved regions, small bump densities corresponding to the carbonyl oxygen atoms and the small shape differences among Tyr, Phe, and Trp can be seen [71].

It is worthwhile to point out that cautions should be taken for the apparent “side-chain densities” because the small bumps on the main chain densities might originate from noises [62]. As discussed in Subheading 3.16, the iterative refinements are prone to overfitting to noises, which will lead to some of the small density bumps difficult to distinguish from authentic structural features. To validate these side-chain-like densities, the corresponding regions of other unique copies of the same protein in the asymmetric unit should be examined if the T number of the virus structure is larger than 1. The truly independent reconstructions can also be used to cross-validate the authenticity of the “side-chain” densities.

3.20 Computing

Image processing and 3-D reconstruction of large viruses to high resolutions are both data and computational intensive tasks that require significant computational resources. Currently, the most popular platform is 64-bit Linux systems using Intel or AMD CPUs. The preprocessing tasks (particle selection and CTF fitting) are typically performed on a Linux workstation with multiple CPU cores, multi-gigabytes of memory, and sufficient storage space (up to hundreds of gigabytes to terabytes for large datasets). For construction of initial models using small datasets of binned images, a Linux workstation or a small Linux cluster should be sufficient. However, high-resolution refinements using the entire data (10^3 – 10^4 particle images of sizes from 400^2 to $1,000^2$) will require much more computing resources, such as large Linux clusters with hundreds of CPU cores.

Flexible usages of different computational resources are supported by *jspr.py*. These resources include single computer with one CPU, single computer with multiple CPUs or cores, collection of multiple computers sharing a common file system, dedicated computer cluster sharing a common file system, and HTCCondor system with geographically separated computers without shared file system. Many of the underlying algorithms were parallelized using OpenMP for shared memory or MPI for distributed memory systems (*see Note 25*). Depending on resources available to the user, one resource or combinations of different resources can be used by *jspr.py* with appropriate options. The following com-

mand shows an example that submits the computational tasks to a dedicated cluster using PBS queue for job management:

```
jspr.py <input dataset and other options> --cpus=<n> --pbs=1
--pbsNum Nodes=<n> --pbsppn=<n> --pbsQueue=<name>
--pbsWall Time=<hours>
```

3.21 Overall Image Processing Strategy

By integrating the above sections on the individual image processing tasks and the need of validations, a suggested overall image processing strategy is shown in Fig. 7 and briefly explained as following:

- Micrographs are digitized and filtered/binning for particle selection.
- Particles are selected first using an automated method and then manually screened.
- CTF parameters are determined first using automated fitting and then manually verified. CTF parameters are stored in the particle image header. Poor quality images are discarded.
- Split whole dataset into even and odd halves, and the two halves are independently processed throughout the rest of image processing.
- For each of the halve datasets, multiple consistent de novo initial models are built using the random model method. Each of the initial models is built from a different small subset of binned particle images randomly selected from the half dataset.
- For each of the halve datasets, the multiple de novo initial models are used to independently determine the initial orientation and center parameters. Particles with consistent parameters are identified for inclusion in further refinements.
- The orientation/center parameters are further refined first using binned images and then using original images.
- Further refine the scale, astigmatism, and defocus parameters of the particles.
- Compute the FSC curve using the two final truly independent reconstructions, one from each of the half datasets. The resolution of the final reconstruction will be read from this curve using the FSC=0.143 criterion.
- Merge the two half datasets and reconstruct the 3-D density map from the combined dataset. This map is the final reconstruction of the project for structural analysis.

The workflow can be carried out using command line programs presented in the above individual sections. The processing of multiple independent models and datasets can be executed simultaneously in the same *jspr.py* run. The workflow is also available as the *spr-icos* mode that we implemented using the extensible *e2project manager.py* graphic user interface (see Fig. 8 and Note 26).

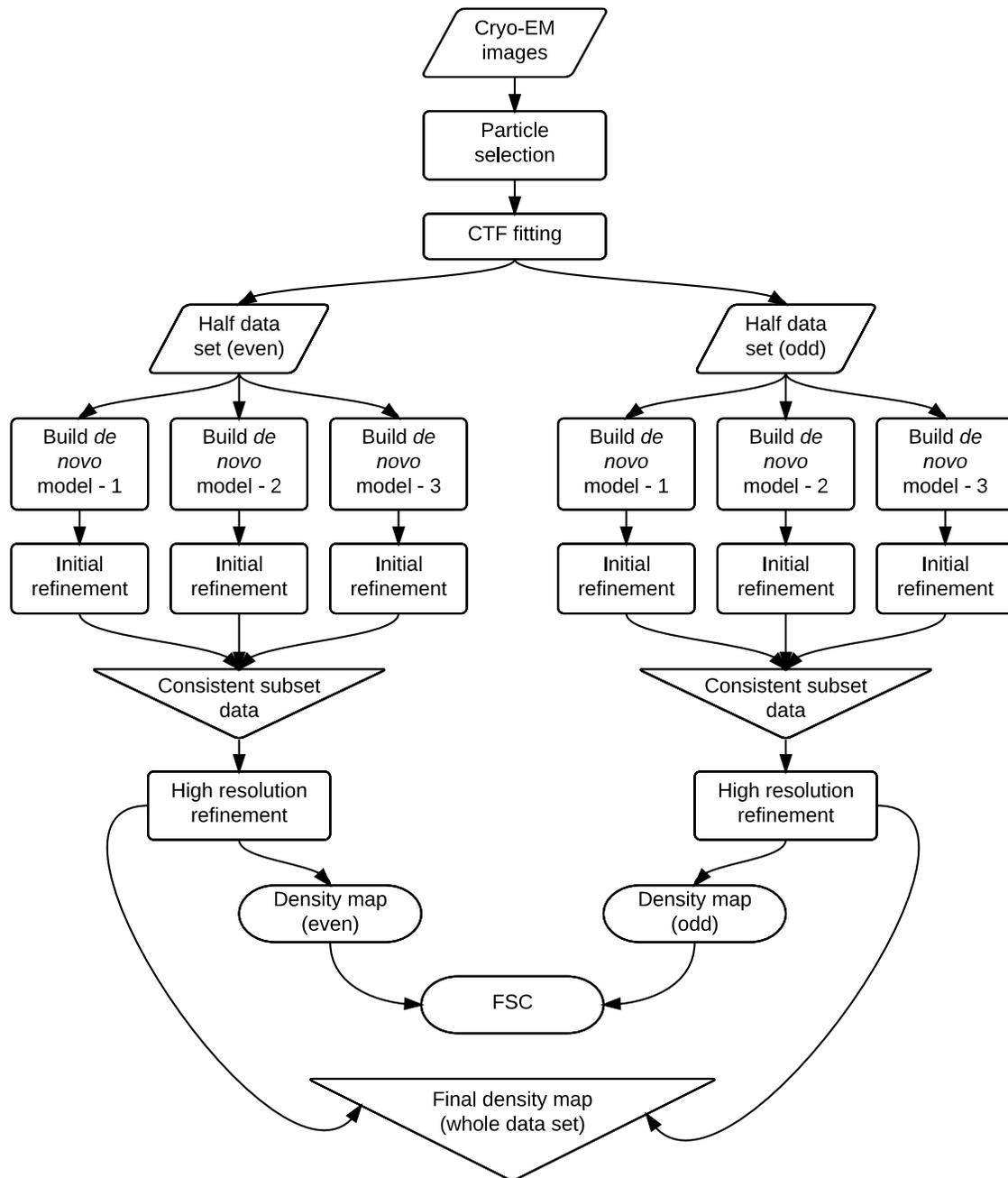


Fig. 7 Image processing strategy. Detailed discussions about each of the steps in this strategy are in Subheadings [3.9–3.21](#)

4 Notes

1. The blotting is a critical step that will ultimately decide if the resulting grid is usable. Insufficient blotting will result in too thick ice and over-blotting will leave a bare grid. Since the blotting task is to remove >99.9 % (3–5 μ l) of the solution and leave <0.1 % (~1 nl) of the sample on the grid, it is extremely sensitive to the



Fig. 8 Graphic user interface. The GUI for the workflow shown in Fig. 7 is implemented as the “SPR-Icos” mode in the EMAN2 *e2projectmanager.py* program. The image processing programs are also available as an online computing tool on DiaGrid

sample viscosity, grid surface hydrophobicity, environmental humidity, filter paper wetness, contact between the filter paper and grid, and the duration of the blotting. One should be prepared for the potential difficulties in this step, especially for sample buffers including glycerol/sucrose, detergent, or high concentration of salt (1 M). Currently, there is no method that can guarantee the success of every blotting/freezing. Practice and experience through systematic trial and error is the only solution to get around the difficulties.

2. The cryo-transfer process should be practiced very carefully and be performed rapidly and smoothly. It is essential to ensure minimal exposure to the room moisture and to avoid ice contamination and excessive temperature increase of the grid.

3. When film or CCD is used to record images, around one second continuous exposure is typically used. It is known that beam-induced motions during this “long” exposure can blur the images and degrade the image resolution. With next-generation direct electron detectors capable of high-speed exposures (e.g., 40 frames/s for DDD camera), it will be possible to acquire a series of short exposures that “freeze” the motion in each frame. Improved image resolution can then be achieved from computational averaging of the frames to eliminate the motion blur [72].
4. The Nikon scanner has been used in several near-atomic resolution virus reconstructions and should be regarded as sufficient for most cryo-EM projects. However, it should also be recognized that the scanner introduces additional artifacts to the images [32, 73]. The film holder needs to be modified to include glass plates that can hold the film flat and level during scanning for uniform quality across entire film [74]. Its strong anisotropic MTF causes the power spectra to decay significantly faster in the X -direction than in Y -direction. Other performance concerns including positional accuracy and orthogonality have also been reported [32]. A bigger concern is the uncertainty in its commercial availability as its production has already been discontinued. Replacement scanner models need to be carefully evaluated [32, 73] first to ensure sufficient scan quality for high-resolution cryo-EM projects. A safer long-term plan is to complete the transition to all digital recording using CCD or next-generation direct electron detectors such as the DDD from Direct Electron, K2 from Gatan, or Falcon from FEI.
5. An O.D. step tablet can be used to test if the scanner saves transmittances instead of O.D. values (see <http://imagej.nih.gov/ij/docs/examples/calibration>). Simply scan the tablet with linear gradient of O.D. values and plot the pixel values along the gradient direction. A curved instead of linear plot will indicate that the pixel values are recorded as transmittance and need conversion to O.D. to make the plot linear.
6. There is no community standard on image formats in the cryo-EM field. The Gatan CCD images are saved in proprietary DM3 format while Nikon scanner saves images in 16-bit TIFF format. We typically convert the micrograph level images to MRC format. The selected particles are saved in HDF format for its support for storing user-defined attributes without the need to modify the format definition as required by other image formats of fixed headers. Particle sets are collected in LST format that simply points to the actual images in binary formats. The LST format also supports arbitrary *key=value* pairs to record important image processing results such as

Euler angles, center positions, and refined defocus values. The benefits of LST format include easy access as a text format, disk space savings, and flexible particle set manipulation (merge, split, common, diff, etc.). Another important but often overlooked benefit is to keep particle files in binary format unmodified during image processing and to preserve their original pristine state for reproducible research. The 3-D reconstructions are typically stored in MRC format for its compatibility with CCP4 density map format in X-ray crystallography. EMAN and EMAN2 I/O functionalities natively support nearly all image formats used in the cryo-EM field and can be used to interconvert images in any of the formats.

7. Some viruses, for example, T4 bacteriophage, do not have icosahedral symmetry. Instead, their structure can be considered as elongated icosahedron with D5 symmetry, a subset of icosahedral symmetry. Their structures can, in general, be solved using most single particle cryo-EM software by simply changing the symmetry parameter from icosahedral to D5. For example, `--sym=d5` instead of `--sym=icos` should be used for *jspr.py*.
8. While icosahedral symmetry is often associated with viruses, it should be noted that icosahedral symmetry also exists for other biological structures, for example, pyruvate dehydrogenase complex (PDC) [66]. The symmetry of PDC can also be referred to as dodecahedral that shares exactly the same sets of symmetries as icosahedral and is processed as icosahedral symmetry during image processing and 3-D reconstruction. Icosahedron and dodecahedron are the dual of each other (see <http://en.wikipedia.org/wiki/Icosahedron>) and can be interconverted by truncating the fivefold vertex (icosahedron to dodecahedron) or threefold vertex (dodecahedron to icosahedron).
9. In single particle cryo-EM, a fundamental assumption is that the images of different particles are all projections of identical 3-D structure though the views of the different particles can be different. When neighbor particles or contaminants are very close to the particle, this fundamental assumption about the data is violated. In addition, the oscillatory CTF modulation of EM images effectively spreads the particle information beyond the apparent particle boundary. When two particles are very close, the information from each particle will contaminate the other particle. Neighbor particles also present additional challenges in 2-D alignments especially for alignment methods that determine the particle in-plane rotational angle before the center (e.g., EMAN/EMAN2 autocorrelation-based 2-D alignment) [33, 34]. Thus, isolated particles are preferred in general. However, this preference should not be overly enforced to discard all micrographs with closely packed particles as those micrographs might be the only data available due

to difficulties in sample freezing. With the CTF pre-corrected to suppress signal spreading, a suitable alignment method (e.g., PFT method which explicitly searches the center and determines the best in-plane rotation for each trial center position) [55, 75], and appropriate mask size, even particles in closely packed array-like images (*see* Fig. 2) can be effectively used for near-atomic resolution 3-D reconstructions.

10. “Good” image size needs to satisfy a few requirements or preferences. First, it should be an even number; second, its largest prime number factor should be as small as possible and ideally 2 (i.e., 2^n). Both are derived from the fast Fourier transform (FFT), which is frequently used in cryo-EM image processing algorithms. FFT is only supported for 2^n numbers in many early software but is now supported for any number in modern FFT libraries (e.g., FFTW). However, the speed of FFT is still faster for the “good” numbers specified above. Third, it is preferable that the image size can be divided by a large 2^n number but at least 4 due to potential implementation details of some image processing algorithms. Fourth, the particle image size of the original sampling should also allow the image size after binning to satisfy the first three preferences. The program “goodImageSizes.py” will print a list of numbers that satisfy these preferences and the user can choose a number from the list that is suitable for the project.
11. Underfocuses (up to a few micrometer) are typically used in cryo-EM to improve the image contrast. The commonly used term “defocus” refers to underfocus (i.e., the sample is slightly below the Z -position in the column that would have been imaged at focus). Cryo-EM software do not have a consistent convention in the sign for defocus values. A negative sign is used in EMAN version 1 while a positive sign is used in EMAN version 2 and in this protocol.
12. To adequately sample the Thon ring oscillations in Fourier space, the particle image size should be sufficiently large since the number of samplings in FFT is half of the image size. For large viruses, the image size in general should be sufficiently large. However, for more frequent Thon ring oscillations due to larger defocuses used for small particles, the user should ensure that Fourier sampling is sufficiently fine by using the `--oversampling=<n>` option. When $n>1$, the particle images are padded n times before power spectra are computed. With the same reason, appropriate oversampling factors should also be used for CTF correction in 2-D alignment and 3-D reconstruction tasks using the `filter.ctf:type=1:oversample=<n>` processor.
13. The B-factor defined in EMAN version 2 is consistent with the convention used in X-ray crystallography. However, the

corresponding B-factor in EMAN version 1 will be 4 times smaller ($B_{EMAN2} = 4 \times B_{EMAN1}$).

14. Due to the large dynamic range of image power spectra from low to high resolutions, the power spectra (both 2-D and 1-D) are often displayed in *log* mode (i.e., the power spectra pixel values are first transformed using *log* function before being mapped to screen). When displayed in *log* mode, both the power spectra background and the CTF peak heights should decrease nearly linearly.
15. There is no consensus view convention for icosahedral structures [76]. We adopt the convention in which fivefold is along *Z*-axis and a twofold is along *Y*-axis. The *e2proc3d.py* program can be used to interconvert among these conventions. For example, to convert a virus map in another popular view convention with all XYZ axes along twofold axis into this view convention, the following command can be used:


```
e2proc3d.py <input map> <output map> --icos2fto5f
```
16. There is a common misconception that common lines only exist for icosahedral structures. It should be clarified that cross common lines between different particles exist for structures of any symmetry or no symmetry and self-common lines exist for any symmetric structures. In practice, common line methods are typically used for icosahedral structures due to the large number of common lines from the 60-fold symmetries. The significantly smaller number of common lines from lower symmetry structures and the high level of noises in cryo-EM make common line methods less robust and rarely used for non-icosahedral structures.
17. The conventions for Euler angles are often different for different software. Proper conversion of the angles should be performed when multiple software are used for the same dataset. EMAN and EMAN2 support most of the conventions and their interconversion [77].
18. It should be noted that single particle cryo-EM 3-D reconstructions cannot determine the absolute handedness of the structure. Either one of two hands is equally consistent with the images due to the fact that TEM images are projections of the structure and structures of opposite handedness can produce identical projections. It is common to obtain models of opposite handedness from the same dataset from different runs of the random de novo model method (see Fig. 4). The absolute handedness can only be determined using additional information, for example, from tilted images, from comparison to homologous structures of known handedness, or from protein chemistry if solved at near-atomic resolutions. However, it is straightforward to flip the handedness at either the 3-D or 2-D level:

- At the 3-D level by mirroring along an axis:

```
e2proc3d.py <input density map> <output density map.mrc> --process xform.mirror:axis=<x|y|z>
```
 - At the 2-D level by adding 180° to the in-plane rotation angle of each particle and then recomputing the 3-D map:

```
flipHand.py <input images file> <output images.lst>
```
19. “Bad particles” here refer to any particles that cannot pass the criteria for inclusion in subsequent high-resolution refinement. These particles can be contaminants that are selected by mistake; authentic particles but with low contrast, neighbor particles, or noises, etc.; and authentic particles in different conformations. When the particles exist in a small number of discrete conformations, the structure of each of the conformations can often be successfully determined using multiple-model-based refinement that is also supported by *jspr.py*.
 20. For astigmatism, defocus, and magnification refinement, the *jspr.py* program will assign all particles from the same micrograph to a single *jalign* task. As a result, a single micrograph is assumed by default to have uniform astigmatism, defocus, and magnification. However, it is also possible to refine astigmatism, defocus, and magnification using a subset or even a single particle by specifying an appropriate *batchsize=<n>* parameter: *n* = -1 will use all particles, *n* = 1 will use a single particle, and *n* > 1 will use more than one particle. It often slightly improves the reconstructions when a subset or a single particle instead of all particles in the micrograph is refined as a unit. Such improvements suggest local variations in the micrographs [41, 58, 78].
 21. Two additional factors, beam tilt and depth of field, should also be considered in theory. For the Titan Krios TEM, the three condenser system makes it easy to achieve parallel beam illumination. The coma-free alignment ensures that the beam is parallel to the optical axis with minimal tilt. With this tilt-free parallel illumination, beam tilt-induced phase shifts of the experimental images are significantly minimized. The depth of field problem is more relevant to large viruses where the differences of defocus for different *Z*-heights (i.e., from top to bottom) of a single particle cannot be ignored. The depth of field problem is equivalent to the Ewald sphere problem in which the Ewald sphere can no longer be approximated as planes. The depth of field/Ewald sphere problem is important only at high resolutions and should in theory be corrected at near-atomic resolutions. Several correction methods [47, 79–82] have been proposed and shown effective for simulated data. However, effective corrections for experimental cryo-EM images remain elusive.

22. The asymmetric reconstruction procedure implicitly assumes that all particles used in reconstruction have identical structure including how the different structural components, icosahedral and non-icosahedral, are organized in the particles. This is the same assumption by any single particle cryo-EM 3-D reconstructions. For particles with defined component structures but varying organizations of the components in different particles (i.e., same set of LEGO pieces assembled differently), this assumption will lead to asymmetric reconstructions with only the dominant components resolved while the less dominant components will remain smeared, for example, the resolved tail spikes and portal ring but smeared core densities in the asymmetric reconstruction of bacteriophage $\epsilon 15$ [17, 83].
23. The resolution reported based on the FSC at a threshold is an estimate of the average quality of the map. It is common that different regions of a structure have different qualities due to conformational variability. Some of the better-resolved regions should be at higher resolution than what the average resolution indicates [84, 85].
24. Fourier phases are generally not modified by the sharpening process. Fourier phase changes will move features in real space while Fourier amplitude changes mostly affect the level of smearing of features in real space.
25. The two major computational tasks in the iterative refinements are 2-D alignment performed by the *jalign* program and 3-D reconstruction performed by *j3dr*. The 2-D alignment tasks belong to the class of “embarrassingly parallel” tasks and can be trivially parallelized by dividing the entire dataset into many small chunks with each chunk assigned to a separate subtask. The 3-D reconstruction task needs to merge all particles into a single 3-D model, which makes it necessary to use MPI for parallelization as implemented in *j3dr*.
26. The *e2projectmanager.py* GUI implementation of our workflow has also been implemented as an online computational tool on DiaGrid (<http://diagrid.org/tools/cryoem>) which aims to provide users not only the preinstalled software but also large-scale computational resources needed by image processing and 3-D reconstruction of viruses.

Acknowledgements

We thank Dr. Philips Sewer for providing the T7 capsid I sample used as examples in the figures. We also thank Dr. Agustin Avila-Sakar and Mr. Frank Vago for suggestions to the manuscript. The research has been supported by NIH.

References

1. Grigorieff N, Harrison SC (2011) Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. *Curr Opin Struct Biol* 21:265–273
2. Crowther RA (2010) From envelopes to atoms: the remarkable progress of biological electron microscopy. In: Ludtke SJ, Prasad BVV (eds) *Advances in protein chemistry and structural biology: recent advances in electron cryomicroscopy*, Pt A, vol 81. Academic Press, San Diego, pp 1–32
3. Jensen G (ed) (2010) *Cryo-EM Part A: sample preparation and data collection*. Academic, San Diego
4. Jensen G (ed) (2010) *Cryo-EM Part B: 3-D reconstruction*. Academic, San Diego
5. Frank J (2006) Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state. Oxford University Press, New York
6. Harrison SC (2010) Virology. Looking inside adenovirus. *Science* 329:1026–1027
7. Crowther RA (2008) The Leeuwenhoek lecture 2006. Microscopy goes cold: frozen viruses reveal their structural secrets. *Philos Trans R Soc Lond B Biol Sci* 363:2441–2451
8. Prasad BV, Schmid MF (2012) Principles of virus structural organization. *Adv Exp Med Biol* 726:17–47
9. Jiang W, Chiu W (2007) Cryoelectron microscopy of icosahedral virus particles. *Methods Mol Biol* 369:345–363
10. Jiang W, Ludtke SJ (2005) Electron cryomicroscopy of single particles at subnanometer resolution. *Curr Opin Struct Biol* 15:571–577
11. Chiu W, Baker ML, Jiang W et al (2005) Electron cryomicroscopy of biological machines at subnanometer resolution. *Structure* 13:363–372
12. Jiang W, Baker ML, Jakana J et al (2008) Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. *Nature* 451:1130–1134
13. Liu H, Jin L, Koh SB et al (2010) Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. *Science* 329:1038–1043
14. Cheng L, Sun J, Zhang K et al (2011) Atomic model of a cypovirus built from cryo-EM structure provides insight into the mechanism of mRNA capping. *Proc Natl Acad Sci U S A* 108:1373–1378
15. Yu X, Ge P, Jiang J et al (2011) Atomic model of CPV reveals the mechanism used by this single-shelled virus to economically carry out functions conserved in multishelled reoviruses. *Structure* 19:652–661
16. Zhang X, Sun S, Xiang Y et al (2012) Structure of Sputnik, a virophage, at 3.5-Å resolution. *Proc Natl Acad Sci U S A* 109(45): 18431–18436
17. Jiang W, Chang J, Jakana J et al (2006) Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature* 439:612–616
18. Chang J, Weigele P, King J et al (2006) Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. *Structure* 14:1073–1082
19. Morais MC, Tao Y, Olson NH et al (2001) Cryoelectron-microscopy image reconstruction of symmetry mismatches in bacteriophage phi29. *J Struct Biol* 135:38–46
20. Tang J, Olson N, Jardine PJ et al (2008) DNA poised for release in bacteriophage phi29. *Structure* 16:935–943
21. Xiang Y, Morais MC, Battisti AJ et al (2006) Structural changes of bacteriophage phi29 upon DNA packaging and release. *EMBO J* 25:5229–5239
22. Liu X, Zhang Q, Murata K et al (2010) Structural changes in a marine podovirus associated with release of its genome into *Prochlorococcus*. *Nat Struct Mol Biol* 17:830–836
23. Lander GC, Tang L, Casjens SR et al (2006) The structure of an infectious P22 virion shows the signal for headful DNA packaging. *Science* 312:1791–1795
24. Ermantraut E, Wohlfart K, Tichelaar W (1998) Perforated support foils with pre-defined hole size, shape and arrangement. *Ultramicroscopy* 74:75–81
25. Fukami A, Adachi K (1965) A new method of preparation of a self-perforated micro plastic grid and its application. *J Electron Microsc* 14:112–118
26. Dubochet J, Adrian M, Chang JJ et al (1988) Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* 21:129–228
27. Adrian M, Dubochet J, Lepault J et al (1984) Cryo-electron microscopy of viruses. *Nature* 308:32–36
28. Frederik PM, Hubert DH (2005) Cryoelectron microscopy of liposomes. *Methods Enzymol* 391:431–448
29. Jeng TW, Talmon Y, Chiu W (1988) Containment system for the preparation of vitrified-hydrated virus specimens. *J Electron Microsc Tech* 8:343–348
30. Suloway C, Pulokas J, Fellmann D et al (2005) Automated molecular microscopy: the new Leginon system. *J Struct Biol* 151:41–60

31. Hanszen KJ (ed) (1971) The optical transfer theory of the electron microscope: fundamental principles and applications. Academic, New York
32. Henderson R, Cattermole D, McMullan G et al (2007) Digitisation of electron microscope films: six useful tests applied to three film scanners. *Ultramicroscopy* 107:73–80
33. Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128:82–97
34. Tang G, Peng L, Baldwin PR et al (2007) EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* 157: 38–46
35. Zhu Y, Carragher B, Glaeser RM et al (2004) Automatic particle selection: results of a comparative study. *J Struct Biol* 145:3–14
36. Erickson HP, Klug A (1971) Measurement and compensation of defocusing and aberrations by Fourier processing of electron micrographs. *Phil Trans Roy Soc Lond B* 261:105–118
37. Erickson HP, Klug A (1970) The Fourier transform of an electron micrograph: effects of defocussing and aberrations, and implications for the use of underfocus contrast enhancement. *Ber Bunsen* 74:1129–1137
38. Jiang W, Chiu W (2001) Web-based simulation for contrast transfer function and envelope functions. *Microsc Microanal* 7:329–334
39. Thon F (1971) Phase contrast electron microscopy. Academic, New York
40. Saad A, Ludtke SJ, Jakana J et al (2001) Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination. *J Struct Biol* 133:32–42
41. Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142: 334–347
42. Jiang W, Guo F, Liu Z (2012) A graph theory method for determination of cryo-EM image focuses. *J Struct Biol* 180:343–351
43. Yang C, Jiang W, Chen DH et al (2009) Estimating contrast transfer function and associated parameters by constrained non-linear optimization. *J Microsc* 233:391–403
44. Sorzano CO, Jonic S, Nunez-Ramirez R et al (2007) Fast, robust, and accurate determination of transmission electron microscopy contrast transfer function. *J Struct Biol* 160:249–262
45. Huang Z, Baldwin PR, Mullapudi S et al (2003) Automated determination of parameters describing power spectra of micrograph images in electron microscopy. *J Struct Biol* 144:79–94
46. Mallick SP, Carragher B, Potter CS et al (2005) ACE: automated CTF estimation. *Ultramicroscopy* 104:8–29
47. DeRosier DJ (2000) Correction of high-resolution data for curvature of the Ewald sphere. *Ultramicroscopy* 81:83–98
48. Mersereau RM, Oppenheim AV (1974) Digital reconstruction of multidimensional signals from their projections. *Proceedings of the IEEE*, vol 62, pp 1319–1338. <http://dx.doi.org/10.1109/PROC.1974.9625>
49. Crowther RA, Amos LA, Finch JT et al (1970) Three dimensional reconstructions of spherical viruses by Fourier synthesis from electron micrographs. *Nature* 226:421–425
50. Crowther RA (1971) Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. *Philos Trans R Soc Lond B Biol Sci* 261:221–230
51. Thuman-Commike PA, Chiu W (2000) Reconstruction principles of icosahedral virus structure determination using electron cryomicroscopy. *Micron* 31:687–711
52. Fuller SD, Butcher SJ, Cheng RH et al (1996) Three-dimensional reconstruction of icosahedral particles—the uncommon line. *J Struct Biol* 116:48–55
53. Liu XG, Jiang W, Jakana J et al (2007) Averaging tens to hundreds of icosahedral particle images to resolve protein secondary structure elements using a Multi-path Simulated Annealing optimization algorithm. *J Struct Biol* 160:11–27
54. Yan XD, Dryden KA, Tang JH et al (2007) Ab initio random model method facilitates 3D reconstruction of icosahedral particles. *J Struct Biol* 157:211–225
55. Baker TS, Cheng RH (1996) A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *J Struct Biol* 116:120–130
56. Nelder JA, Mead R (1965) A simplex-method for function minimization. *Comput J* 7:308–313
57. Yang Z, Penczek PA (2008) Cryo-EM image alignment based on nonuniform fast Fourier transform. *Ultramicroscopy* 108:959–969
58. van Heel M, Gowen B, Matadeen R et al (2000) Single-particle electron cryo-microscopy: towards atomic resolution. *Q Rev Biophys* 33: 307–369
59. van Duinen G, van Heel M, Patwardhan A (2005) Magnification variations due to illumination curvature and object defocus in transmission electron microscopy. *Opt Express* 13: 9085–9093
60. Saxton WO, Baumeister W (1982) The correlation averaging of a regularly arranged bacterial cell envelope protein. *J Microsc* 127:127–138
61. Penczek PA (2010) Resolution measures in molecular electron microscopy. *Methods Enzymol* 482:73–100

62. Scheres SH, Chen S (2012) Prevention of overfitting in cryo-EM structure determination. *Nat Methods* 9:853–854
63. Stewart A, Grigorieff N (2004) Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy* 102:67–84
64. Sigworth FJ (1998) A maximum-likelihood approach to single-particle image refinement. *J Struct Biol* 122:328–339
65. van Heel M, Schatz M (2005) Fourier shell correlation threshold criteria. *J Struct Biol* 151: 250–262
66. Rosenthal PB, Henderson R (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* 333:721–745
67. Henderson R, Sali A, Baker ML et al (2012) Outcome of the first electron microscopy validation task force meeting. *Structure* 20: 205–214
68. Henderson R, Chen S, Chen JZ et al (2011) Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. *J Mol Biol* 413:1028–1046
69. Jiang W, Baker ML, Ludtke SJ et al (2001) Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J Mol Biol* 308:1033–1044
70. Baker ML, Ju T, Chiu W (2007) Identification of secondary structure elements in intermediate-resolution density maps. *Structure* 15:7–19
71. Zhang X, Jin L, Fang Q et al (2010) 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell* 141:472–482
72. Campbell MG, Cheng A, Brilot AF (2012) Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure* 20:1823–1828
73. Roseman AM, Neumann K (2003) Objective evaluation of the relative modulation transfer function of densitometers for digitisation of electron micrographs. *Ultramicroscopy* 96:207–218
74. Typke D, Nordmeyer RA, Jones A et al (2005) High-throughput film-densitometry: an efficient approach to generate large data sets. *J Struct Biol* 149:17–29
75. Joyeux L, Penczek PA (2002) Efficiency of 2D alignment methods. *Ultramicroscopy* 92:33–46
76. Heymann JB, Chagoyen M, Belnap DM (2005) Common conventions for interchange and archiving of three-dimensional electron microscopy information in structural biology. *J Struct Biol* 151:196–207
77. Baldwin PR, Penczek PA (2007) The transform class in SPARX and EMAN2. *J Struct Biol* 157:250–261
78. Booy FP, Pawley JB (1993) Cryo-crianking - what happens to carbon-films on copper grids at low-temperature. *Ultramicroscopy* 48: 273–280
79. Wolf M, DeRosier DJ, Grigorieff N (2006) Ewald sphere correction for single-particle electron microscopy. *Ultramicroscopy* 106: 376–382
80. Leong PA, Yu X, Zhou ZH et al (2010) Correcting for the ewald sphere in high-resolution single-particle reconstructions. *Methods Enzymol* 482:369–380
81. Zhang X, Zhou ZH (2011) Limiting factors in atomic resolution cryo electron microscopy: no simple tricks. *J Struct Biol* 175:253–263
82. Wan Y, Chiu W, Zhou ZH (2004) Full contrast transfer function correction in 3D Cryo-EM reconstruction. 2004 International conference on communication, circuits, and systems, vol 2. Chengdu, China, pp 960–964. <http://dx.doi.org/10.1109/ICCCAS.2004.1346339>
83. Guo F, Liu Z, Vago F et al (2013) Visualization of uncorrelated, tandem symmetry mismatches in the internal genome packaging apparatus of bacteriophage T7. *Proceedings of the National Academy of Sciences* 110:6811–6816
84. Yu G, Zhang D, Guo F et al (2013) Cryo-EM Structure of a Novel Calicivirus, Tulane Virus. *PLoS ONE* 8:e59817
85. Cardone G, Heymann JB, Steven AC (2013) One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. *J Struct Biol*