



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

## 1.16 Cryo-Electron Microscopy and Tomography of Virus Particles

RH Rochat and W Chiu, Baylor College of Medicine, Houston, TX, USA

© 2012 Elsevier B.V. All rights reserved.

<b>1.16.1</b>	<b>Introduction</b>	312
<b>1.16.2</b>	<b>Single Particle Reconstruction</b>	313
<b>1.16.3</b>	<b>Tomography</b>	313
<b>1.16.4</b>	<b>Data Collection</b>	314
<b>1.16.5</b>	<b>Data Storage</b>	318
<b>1.16.6</b>	<b>Data Preprocessing</b>	319
1.16.6.1	Screening (Single Particle Reconstruction)	319
1.16.6.2	Boxing (Single Particle Reconstruction)	320
1.16.6.3	Contrast Transfer Function Determination (Single Particle Reconstruction)	321
<b>1.16.7</b>	<b>Single Particle Reconstruction</b>	323
1.16.7.1	Particle Alignment and Refinement	324
1.16.7.1.1	Symmetry-imposed alignment	325
1.16.7.1.2	Asymmetric alignment	327
1.16.7.2	Single Particle Data Subset Selection	327
1.16.7.3	Three-Dimensional Reconstruction	329
<b>1.16.8</b>	<b>Tomographic Reconstruction</b>	329
1.16.8.1	Tilt-Series Alignment	329
1.16.8.2	Alignment and the 'Missing Wedge'	329
<b>1.16.9</b>	<b>Data Interpretation</b>	331
1.16.9.1	Resolution Determination	331
1.16.9.2	Segmentation	331
1.16.9.3	Model Fitting	331
1.16.9.4	Visualization	333
<b>1.16.10</b>	<b>Biological Examples</b>	335
1.16.10.1	Icosahedral Reconstructions	335
1.16.10.2	Reconstructions without Imposing Symmetry	335
1.16.10.3	Tomographic Reconstruction	337
<b>1.16.11</b>	<b>Future Prospects</b>	337
<b>Acknowledgments</b>		337
<b>References</b>		338

### Abbreviations

3-D	three-dimensional	FFT	fast Fourier transform
CCD	charge-coupled device	MP	megapixel
cryo-EM	cryo-electron microscopy	MPSA	Multi-Path Simulated Annealing
cryo-ET	cryo-electron tomography	NCMI	National Center for Macromolecular Imaging
CTF	contrast transfer function	PDB	Protein Data Bank
		SNR	signal-to-noise ratio

### Glossary

**Asymmetric reconstruction** Reconstruction without icosahedral or any other symmetry enforcement.

**Charge-coupled device (CCD)** The CCD used in a transmission electron microscope is composed of a scintillator on top of an array of photon-sensitive pixels. Electrons that hit the scintillator produce a cascade of photons that are recorded as charge on each of the pixels. This pattern of charge (effectively the image) is then read from the array and digitally stored.

**Common lines** This approach for particle alignment relies on the fact that the Fourier transform of a real space

projection of a 3-D object is effectively the same as a 2-D slice in Fourier space passing through the origin of the 3-D Fourier transform of the object. As such, any two real-space projections will intersect along a line (the common line) in Fourier space. Every icosahedral particle has 37 self-common lines and these can be used to determine the icosahedral orientations of a set of particles on a per particle basis. As self-common lines are computed for projections of the same particle, this process is highly sensitive to the preserved icosahedral symmetry of a particle, and noise present in each single particle image. Similar to self-common lines are cross common lines which are described

for a pair of particles as opposed to just a single particle, making it possible to define a maximum of 60 cross common lines for any two icosahedral particles. Cross common line-based particle orientation refinement is advantageous because the magnitude of possible cross common lines reduces the detrimental effect noise has in the self-common lines approach.

**Conformational heterogeneity** Structural non-uniformity of a specimen due to biological or biochemical attributes.

**Contrast transfer function (CTF)** The CTF is a function that characterizes the electron optics used to form an image in a transmission electron microscope. It distorts the image due to defocus, astigmatism, and spherical aberration of the objective lens.

**Cryo-electron microscope** An electron microscope equipped with a specimen holder specifically designed to keep the specimen in the column at temperatures below that of liquid nitrogen.

**Cryo-electron Tomography (Cryo-ET)** The process whereby a frozen, hydrated sample is imaged at a series of tilt angles, and the images are then used to reconstruct a 3-D volume.

**Cryo-electron Microscopy (Cryo-EM)** Refers to the method of imaging a frozen, hydrated specimen kept below liquid nitrogen temperatures in an electron microscope and reconstructing the 3-D density map from multiple single images through computer processing.

**Cryo-ET** Refers to recording images of the same area of a frozen, hydrated specimen at multiple angles of tilt and reconstructing the 3-D density map from these images (a tilt series).

**Eucentric height** The position/height of the sample in an electron microscope column at which little or no shift is noted in the collected images while tilting the sample for tomography.

**Fiducial marker** Any object that is used to relate the relative position of one image to the next (often used to facilitate alignment and reconstruction of a tomogram).

**Icosahedral particle (icosahedron)** A polyhedron with 60 asymmetric units that are related by five-, three-, and twofold axes of symmetry.

**Inelastically scattered** An interaction between the sample and the electrons used in imaging that results in the loss of energy from the electrons.

**Nyquist** Refers to the maximum resolution retrievable in digital imaging that is twice the sampling interval. For example, if the sampling interval were 2 Å per pixel for a given image, Nyquist would be 4 Å.

**Objective lens** The lens system in an electron microscope that sits immediately below the sample in the column. This is the lens responsible for forming the first image of the sample.

**Particle orientation** The individual viruses imaged via cryo-EM and cryo-ET can be related to each other through their orientations. A three-coordinate system of the Euler angles ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) is typically used to define particle orientations. Once particle orientations are determined (typically during alignment or refinement), it is possible to reconstruct the 3-D density map.

**Power spectrum** A Fourier intensity representation of a function that can be 1-D, 2-D, or 3-D.

**Protein Data Bank (PDB)** A public repository for atomic models of proteins and nucleic acids determined experimentally by X-ray crystallography, nuclear magnetic resonance, and/or cryo-EM.

**Reconstruction** The production of a 3-D density map from a set of images of particles in different orientations (either cryo-EM or cryo-ET).

**Resolution** A metric referring to the smallest discernable separation between objects seen in an image or reconstruction. The method of assessing resolution in single particle cryo-EM is based on a numerical threshold in the Fourier shell correlation of two reconstructions computed from two independent sets of raw particle images.

**Secondary structure elements** Refers to the  $\alpha$  helices and  $\beta$  sheets in a protein.

**Signal-to-noise ratio** A measure of the level of noise in a measurement. This value is computed as the ratio of the signal intensity to the noise intensity.

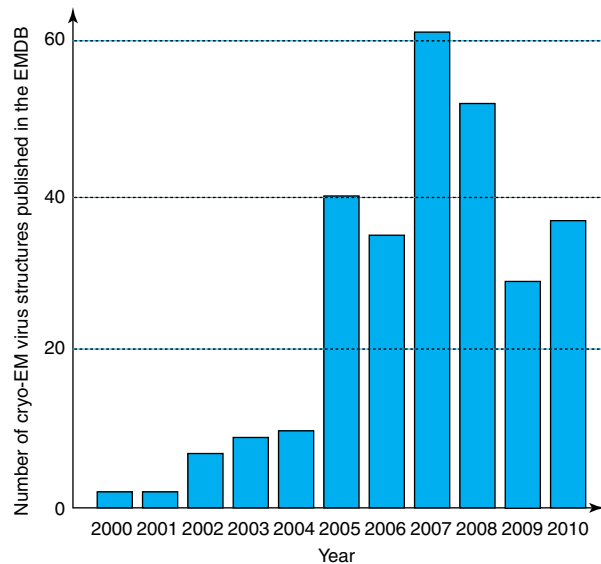
**Subtomogram** A subvolume of data computationally extracted from a complete tomogram.

### 1.16.1 Introduction

Human infectious disease is classified into five etiologies: bacterial, viral, parasitic, fungal, and prion. Viral infections are unique in that they recruit human cellular machinery to replicate themselves and spread infection. The number of viruses causing human disease is vast, and viruses can be broadly categorized by their structures. Many viruses, such as influenza, appear to be amorphous particles, whereas others, such as herpes simplex virus, rhinovirus, dengue virus, and adenovirus, have roughly symmetric structural components. Icosahedral viruses have been a target of electron microscopists for years, and they were some of the first objects to be reconstructed three-dimensionally from electron micrographs.<sup>1,2</sup> The ease with which highly

purified and conformationally uniform virus samples can be produced makes them an ideal target for structural studies. Apart from their biological significance, these virus samples have played a pivotal role in the development of new methodologies in the field of molecular biology as well as in cryo-electron microscopy (cryo-EM) and cryo-electron tomography (cryo-ET).

Since the turn of the century, there has been a tremendous increase in both the number of published cryo-EM virus structures and the quality thereof (Figure 1). Where sub-nanometer resolution icosahedral reconstructions were once the product of months of work, 6- to 10-Å reconstructions are quickly becoming a routine procedure that can be completed in days to weeks. Furthermore, near-atomic resolution reconstructions, although not as routine, are starting to make their



**Figure 1** Bar chart of the number of cryo-EM structures published in the Electron Microscopy Data Bank (EMDB: <http://www.ebi.ac.uk/pdbe/emdb>) from 2000–2010.

way into the literature.<sup>3–11</sup> Apart from icosahedral reconstructions, a growing trend in the field of cryo-EM is the reconstruction of viruses with no imposed symmetry, allowing for the identification and interpretation of non-icosahedral components of the virus (e.g., the tail and the portal complex).<sup>6,12–17</sup> Although asymmetric reconstruction is the ideal approach of any cryo-EM study, the resolution of these reconstructions has yet to reach the point where assignment of secondary structural features is definitive.

Virus reconstruction typically relies on one of two primary imaging modalities – single particle and tomography – the choice of which depends on the structural heterogeneity of the virus either *in vitro* or in the presence of a host cell. The major difference between the two techniques is the manner in which the data are recorded. In single particle data collection, individual virus particles on the grid are imaged only once, and it is the collection of thousands of these images that is used to reconstruct the virus (Figure 2). In cryo-ET, a single area of a grid containing many viral particles is imaged multiple times, at a variety of tilt angles, and it is the combination of these images that is used to generate a tomogram and subsequently an averaged density map (Figure 3). Although single particle data collection is often used for subnanometer resolution structural studies, in the case in which either structural or conformational non-uniformity exists (e.g., viral infection of a cell), tomography proves a more fruitful endeavor. Regardless of the technique used, implicit in each of these techniques is the need to process the raw data, and although the means of doing so are quite unique, the basic principle of aligning and averaging the data remains the same.

This chapter outlines common techniques used in the field of cryo-EM and cryo-ET for virus reconstruction, from data collection to processing and interpretation. Although many of these techniques have been well reviewed in the past,<sup>18–21</sup> because cryo-EM is a rapidly evolving field, advancements in both data collection procedures and data

processing algorithms are a frequent occurrence. The methodologies discussed in this chapter are commonly used at the National Center for Macromolecular Imaging (NCMI) for the reconstruction of viruses by cryo-EM and cryo-ET. This chapter is not intended to be a comprehensive compilation of all tools currently being used in the field but, rather, only a summary of the ones with which the authors have the most experience.

### 1.16.2 Single Particle Reconstruction

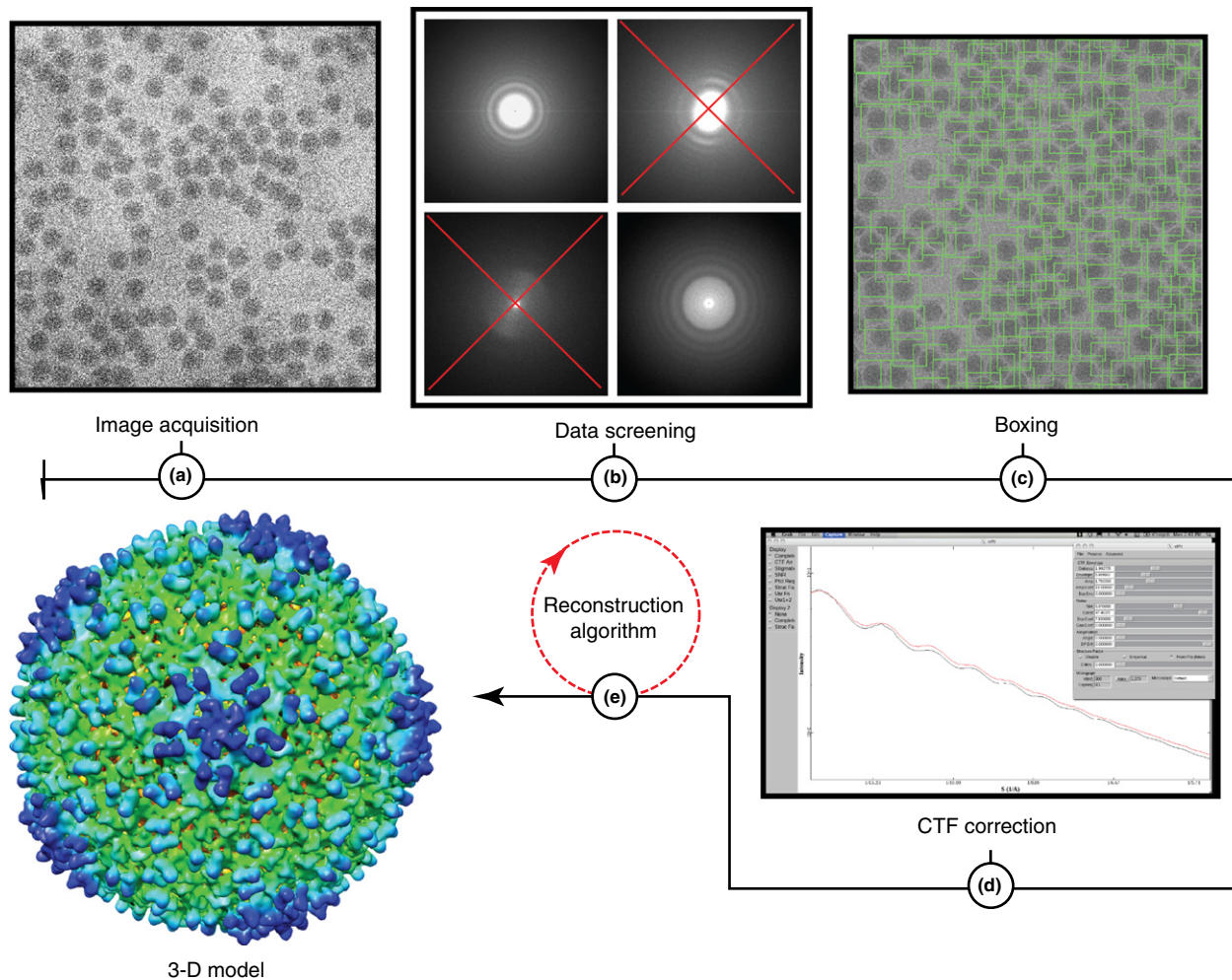
As a result of the conformational uniformity with which virus particles are produced, and the ease with which they can be purified, single particle reconstruction has long been used to solve the structure of icosahedral viruses.<sup>1,2</sup> Single particle reconstruction relies on imaging hundreds to hundreds of thousands of viral particles, computationally isolating each particle, and then combining the individual particles into a single 3-dimensional (3-D) density map (Figure 2). Because cryo-EM provides a projection of the virus onto the recording medium, the lack of an even distribution in the 3-D orientation of the particles can introduce bias when reconstructing the data. Accordingly, a single particle imaging session should result in the collection of a series of 2-D images that taken together offer a well-sampled view of the many different orientations of the particles. The process of single particle reconstruction involves isolating each viral particle from a set of images, determining its orientation and center, and then using this information to stitch the data back together into a single 3-D representation of the virus.

Although the first single particle cryo-EM reconstructions were published approximately 40 years ago,<sup>1,2,22</sup> advancements are still being made in the field that have enabled the resolution of these reconstructions to push well beyond the subnanometer threshold.<sup>4–6,8–10</sup> The driving force behind these advancements is the ability to identify not only the secondary structural elements but also the C $\alpha$  backbone and side chains of the individual subunits of a virus.

### 1.16.3 Tomography

Although purified samples with high structural uniformity are ideal for single particle reconstruction, it is not always possible to work within a system that lends itself to the single particle approach. In the presence of structurally diverse samples, a technique that can capitalize on the information content of just a few particles of similar or identical conformation is needed. The theory behind tomography is that by collecting a series of images from a sample, at a variety of tilt angles, it is possible to reconstruct a 3-D volume density from the 2-D projections (Figure 3(a)).<sup>23,24</sup> This approach allows for a great deal of low-resolution 3-D information to be extracted from the handful of particles in the tomogram. Under standard tomographic imaging conditions, a series of 71 images are taken from  $-70^\circ$  to  $70^\circ$  at a step size of  $2^\circ$  ( $60^\circ$  and  $80^\circ$  tilt holders are available as well). These images can then be aligned to each other using fiducial markers present in the images (Figure 3(b)), from which a volume (a tomogram) can be extracted (Figure 3(c)).<sup>25</sup> From this volume, it is possible to





**Figure 2** Single particle image processing schematic. (a) Image processing begins with the compilation of a series of micrographs. (b) Subtle aberrations in the micrographs can be visualized in the FFT of the raw data. Images with obvious aberrations, such as drift, vibration, and astigmatism, should not be considered for further processing. (c) Once the data have been screened, the particles in the image can be boxed. Depending on the program used, boxing can be done either manually or automatically; caution must be observed when autoboxing because the algorithms are not 100% specific. (d) After boxing, the CTF of the data in the boxed particles must be CTF corrected. Accurate CTF correction is necessary for high-resolution reconstruction. (e) The final, and most computationally intensive, step in image processing is reconstruction. In this step, the orientations for each of the particles are approximated and then refined. Finally, all of the data are combined to produce a 3-D model.

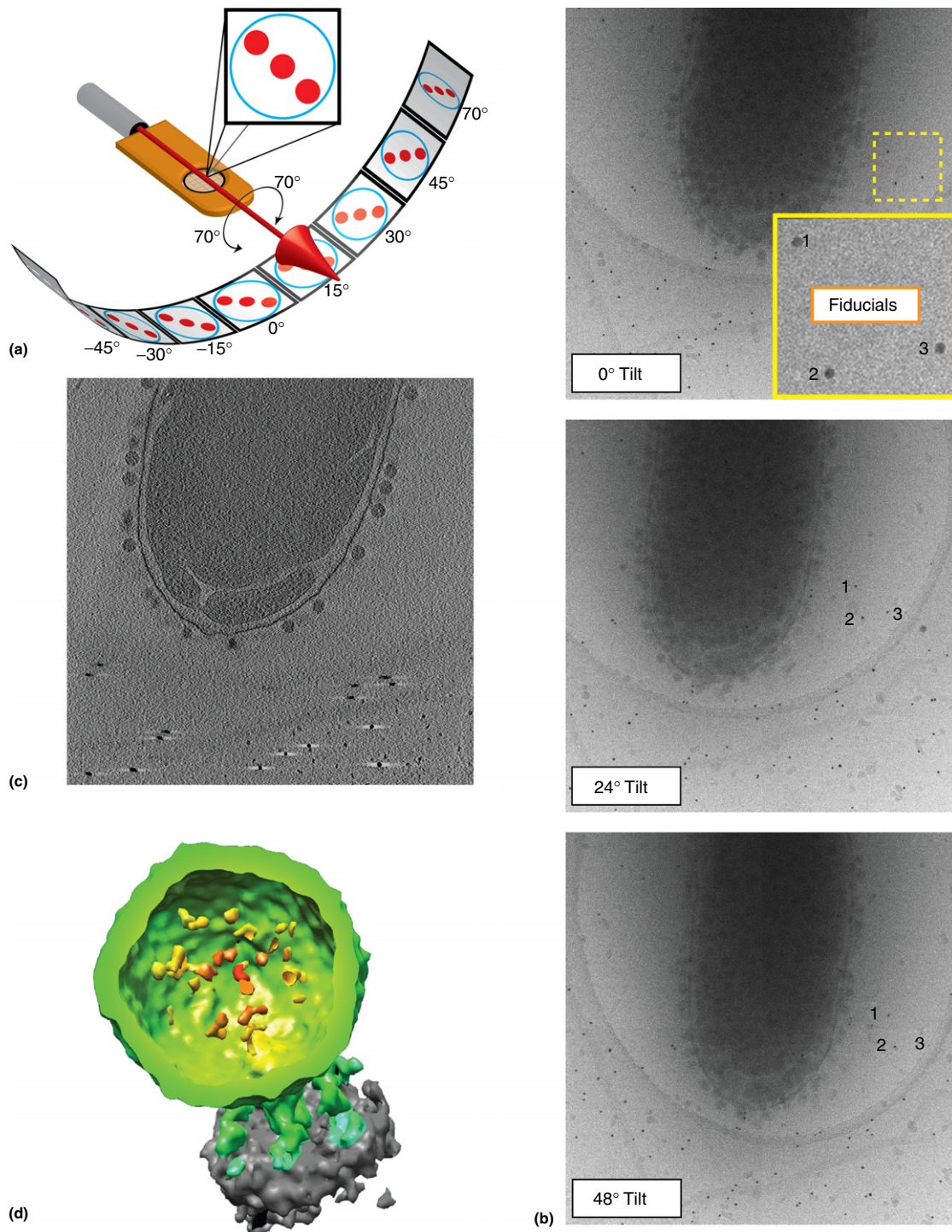
extract individual subtomograms corresponding to each of the particles. These subtomograms can then be classified, aligned, and averaged to obtain a single 3-D model.<sup>26–28</sup>

One consideration for tomography is the high dose that the sample receives during the tilt-series imaging. In single particle data collection, the sample is often subjected to between 20 and 25 electrons per  $\text{\AA}^2$  per micrograph.<sup>20,29</sup> In tomography, however, because a single area of the grid is imaged approximately 70 times, the sample is subjected to far more radiation, making damage a consideration. In general, during tomography, the dose per image is set to a point at which the total dose per tomogram (71 images) is between 80 and 100 electrons per  $\text{\AA}^2$ .<sup>27</sup> Even at this dose level, radiation damage begins to become an issue and will limit the maximum achievable resolution through this technique.<sup>30,31</sup> Therefore, in order to limit the total dose delivered in a single tomogram, on average,

the dose per image must be reduced to nearly 1/20th of what is commonly used for a single particle micrograph. Lowering the total dose per image reduces the signal-to-noise ratio (SNR) in the data and effectively lowers the maximum attainable resolution from the data, the gain from ‘dose fractionation’ notwithstanding.<sup>32</sup> Due to the manner in which a single tomographic series is recorded, the maximum resolution achieved so far with these techniques only approaches the nanometer threshold,<sup>33</sup> which is far below the current standard set by the single particle approach.<sup>4–11</sup>

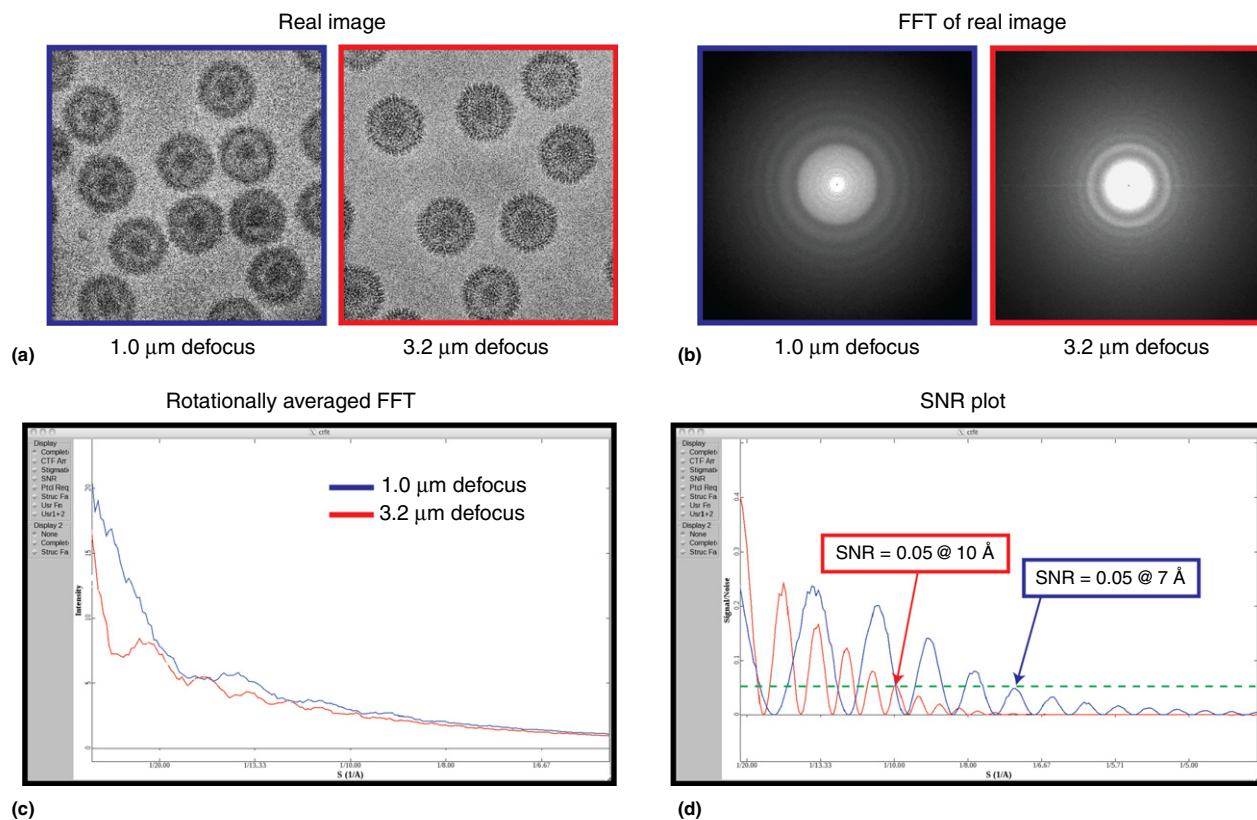
#### 1.16.4 Data Collection

Both cryo-specimen preparation and low-dose imaging for single particle and tomography have been well described and



**Figure 3** Schematic of tomographic data collection. (a) During tomography, the grid holder is rotated across its maximum tilt range ( $-70^\circ$  to  $70^\circ$  shown), and a single low-dose image is typically taken every two degrees (the step size), resulting in a total of 71 images (for a  $70^\circ$  holder). (b) Gold nanoparticles (enlarged view shown in yellow) are typically used as fiducial markers. Gold is an ideal marker because even in low-contrast images, the gold appears as an extremely electron-dense object, making it easy to identify. Tracking fiducial markers through a tilt series allows each of the individual images that comprise the tomogram to be aligned with respect to each other, facilitating tomographic reconstruction (c) and subvolume extraction. Although the object of interest in this tomogram (the cell with the virus attached) appears to move as the sample is tilted, this is an artifact of sample rotation and is corrected for by aligning the gold fiducials across the series of images. (c) A slice through a reconstructed tomogram showing virus particles attached to a cell as well as the fiducials (dark spots with white halos) that were used to reconstruct the volume. (d) Subvolumes containing individual particles can be extracted from the tomogram and averaged to yield the 3-D model of a virus (sliced and radially colored) in the process of infecting a cell (outer membrane shown in gray) [EMID-5219]. Adapted from Chang, J. T.; Schmid, M. F.; Haase-Pettingell, C.; Weigele, P. R.; King, J. A.; Chiu, W. Visualizing the structural changes of bacteriophage epsilon15 and its *Salmonella* host during infection. *J. Mol. Biol.* **2010**, *402*, 731–740.





**Figure 4** Defocus contrast enhancement. Low-frequency contrast is extremely useful for particle alignment, but it comes at the cost of high-resolution information, which is lost as a function of the exponentially decaying envelope function. (a) Contrast enhancement due to defocus can be seen in the real images collected on the electron microscope. Although the ‘real image’ data taken at a high defocus (3.2 μm) appear to be sharper, this is a reflection of the low-resolution contrast enhancement obtained by defocusing the image and does not necessarily make it more useful for the purposes of reconstruction. (b) The difference in defocus is also visualized in Fourier space, where the image taken closer to focus has broader rings that extend further outward in Fourier space. (c) At 1.0 μm defocus (blue line), the second CTF zero in the rotationally averaged FFT is pushed out to a spatial frequency that is nearly as far as the signal extends for the 3.2-μm defocus micrograph (red line), illustrating how low defocus images contain more high-resolution (high spatial frequency) data. To ensure that one has a complete data set that can be reconstructed to a subnanometer resolution, it is important to have data collected across a broad range of defoci – high defocus for initial model generation and low defocus for high-resolution reconstructions. (d) A further complication of high-defocus images is the rapidly decaying envelope that limits the information content (signal-to-noise ratio (SNR)) of these images. The point at which the SNR of the data collected at these two defocus values falls below 0.05 is a reflection of the diminished information content of high-defocus images.

are practiced routinely.<sup>24,34,35</sup> One of the first decisions the electron microscopist must make concerns what microscope to use – specifically, what voltage should be used. For the purpose of virus reconstruction, the two most commonly used voltages are 200 and 300 kV. The major differences between these two voltages concern contrast and depth of field. Although the 200-kV microscope has slightly better contrast, it has a smaller depth of field. Contrast is less of an issue for viruses, but the differences in depth of field between the two voltages become a consideration for large viruses where the top and bottom of the virus particles are no longer at the same defocus.<sup>21</sup> For these reasons, 300-kV electron microscopes are the instrument of choice for high-resolution structural studies.

Electron images are convoluted by the contrast transfer function (CTF) and the envelope function of the electron microscope.<sup>35</sup> Due to the oscillating nature of the CTF in frequency space, there are zeros of the function at different

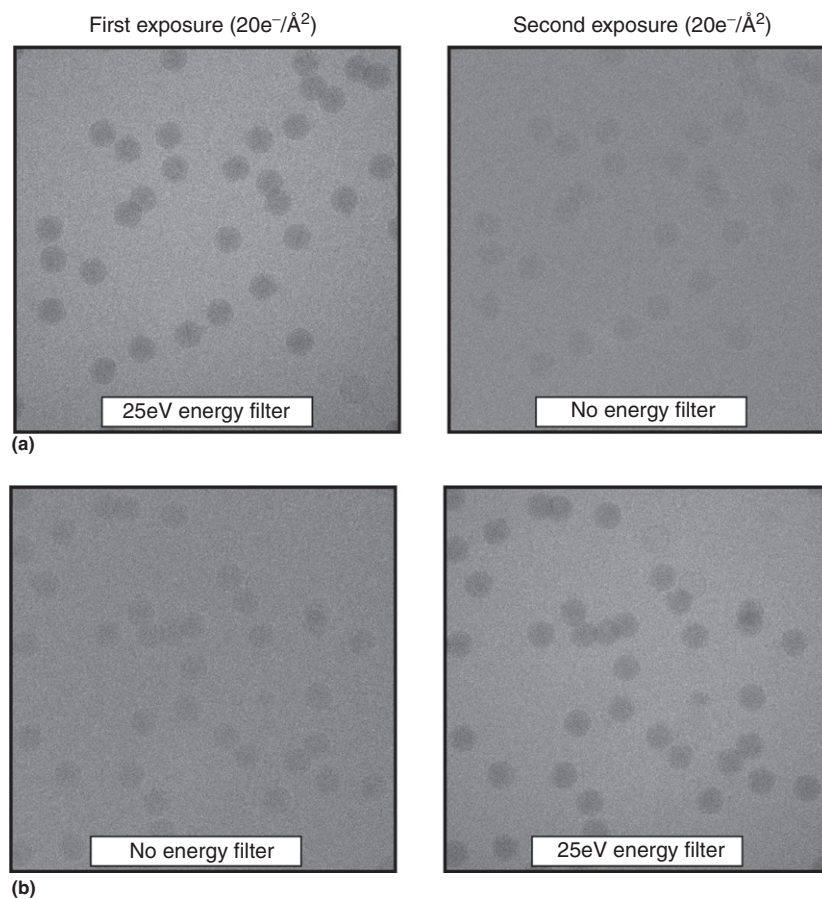
frequencies for a given defocus (see Section 1.16.6.3). To reconstruct an unknown virus from low to high resolution, it is necessary to collect the data at varying defocus (Figure 4) so that the merged images will have no information gap up to the targeted resolution of the reconstruction. Because the apparent image contrast is higher at higher defocus, high-defocus/high-contrast images are ideal for building an initial model for a reconstruction because the data contain low-resolution details. However, the usefulness of these high-defocus images is limited by the fact that for some instruments, at high defocus, the SNR rapidly decreases with increasing spatial frequency (Figure 4(d)) due to the spatial coherence of the electron gun. As a result, the same images used to determine an initial model are generally not ideal for high-resolution reconstructions because they have less signal at high spatial frequencies. Traditionally, viruses were imaged at a large defocus (>3 μm), optimized for good low-resolution contrast, which made it

possible for image processing software of the time to determine the orientations of the particles for the purpose of reconstruction.<sup>36,37</sup> High-resolution virus studies relied on focal pair series that coupled two images from the same area of a grid, one at high defocus to determine the approximate icosahedral orientations and another close to focus to maximize high-resolution information content.<sup>29,38</sup> However, because the microscope's spatial coherence and image processing software have improved, it is no longer necessary to image far from focus, and on average, for data collected on a 300-kV electron microscope, the defocus range used today for a typical high-resolution reconstruction varies from 0.5 to 2.5  $\mu\text{m}$ .

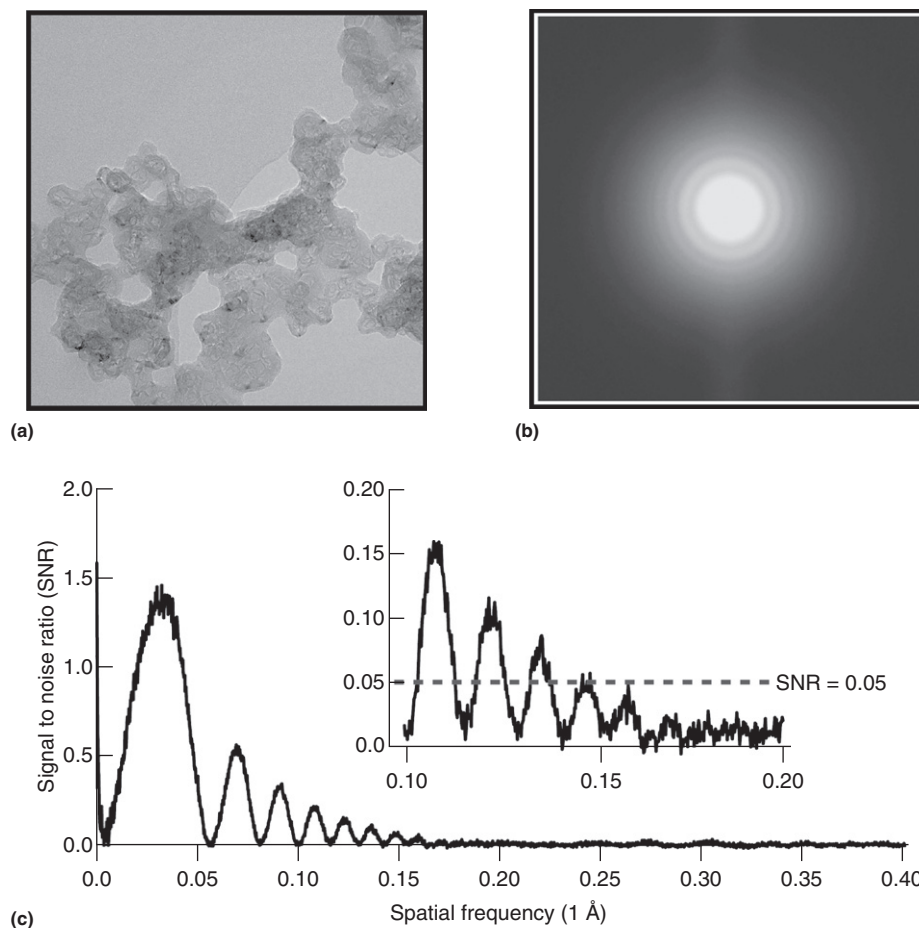
Because defocus-based contrast enhancement has the limitation that it reduces the SNR at high spatial frequencies, it is less useful for collecting high-resolution data. However, to address the issues of poor contrast in the transmitted electron beam, while preserving high-resolution features, energy filtering has been implemented in the workflow of data collection.<sup>39,40</sup> By filtering the transmitted beam, it is possible to exclude electrons that were inelastically scattered by the sample, thus enhancing the contrast of the resulting image, especially at low defocus (Figure 5). Energy filters are becoming

more commonplace in cryo-EM, and they have been used to solve structures at resolutions of approximately 4  $\text{\AA}$ .<sup>5,10,41</sup>

The final step in data collection is choosing which medium to use to record data. Not long ago, photographic film was the only medium for data collection, but advances in fabrication techniques have resulted in the development of charge-coupled devices (CCDs) capable of recording images from electron microscopes. Due to improvements in the scintillators used on the new generation of CCD cameras, CCDs have started to replace film as the preferred medium for data collection, even for near-atomic resolution reconstructions.<sup>9,11,42,43</sup> Inherently, CCDs have many advantages compared to photographic film; some of the most notable advantages are high contrast at low frequency, instantaneous assessment of data quality, and elimination of the labor-intensive process of film development and digitization. Like any recording medium, CCD cameras have their own modulation transfer function, and it is possible to empirically determine the retrievable resolution of these devices by estimating the frequency at which the SNR in images of amorphous carbon falls below 5% (Figure 6). Previous work has shown that depending on the CCD used,



**Figure 5** In-column energy filtering. Energy filters have been implemented in electron microscopes to reduce the number of inelastically scattered electrons that reach the detector. Electrons that are inelastically scattered by the sample have lower energy and can be filtered out before reaching the detector. Energy filtering results in image contrast enhancement in addition to reduction in background noise. These effects can be seen in micrographs taken of the same area at 1.1- $\mu\text{m}$  defocus, both with and without an energy filter, in a 300-kV JEM3200FSC microscope. The exposure series shown in panels a and b demonstrate that total exposure, resulting from imaging the same area, does not account for the differences observed when using a 25-eV energy filter.



**Figure 6** (a) A raw image of graphitized carbon imaged at a detector magnification of  $\times 70\,600$ . (b) Fourier transform of the carbon showing the Thon rings characteristic of the CTF of the data. (c) The spectral SNR of the rotationally averaged Fourier transform. The spatial frequency at which the SNR falls below 0.05 indicates that for this image, the recording medium recovered data up to approximately  $2/5$  Nyquist.

the achievable resolution for these devices ranges between  $2/5$  and  $2/3$  Nyquist frequency.<sup>42,44</sup> Determining the achievable resolution limit of a CCD is necessary because it dictates at what magnification data should be collected when targeting a specific resolution for a reconstruction.

### 1.16.5 Data Storage

Regardless of the medium used to collect data, individual images must be digitized before they can be processed for reconstruction. Although a digitized film negative requires approximately 100 MB to store (assuming a  $9\text{-}\mu\text{m}$  per pixel resolution), the raw data (the negative) can be stored indefinitely in a low-humidity, light-free environment. Alternatively, data from a CCD exist solely as a digital file and must be maintained as such. The data storage requirements of a single micrograph collected on CCD depend directly on the size of the CCD. The first CCDs used for biological applications were 1-megapixel (MP) cameras that required 4 MB of storage per frame.<sup>45,46</sup> These were followed by 4-MP,<sup>47</sup> 16-MP,<sup>44,48</sup> and 100-MP cameras requiring 16, 64, and 480 MB of storage per

frame, respectively. Although the cost of data storage is relatively low (currently less than \$0.10 per GB), a single session on the microscope can result in the collection of between 26 and 192 GB, depending on the CCD camera used and the number of images collected. One effect of this growing trend toward higher megapixel CCDs is the nearly exponential growth in stored data. Fortunately, data storage capabilities have grown exponentially during approximately the past decade, making it possible to store all of these data. Nevertheless, as higher resolution reconstructions are sought, the amount of data storage required for these reconstructions may exceed local resources. Although the cost of digital collection and short-term storage is far less than that of film development, the need to maintain the data electronically with a high level of fidelity for an indefinite amount of time reinforces the need for a local centralized electronic database for electron micrographs.

Database prototypes have been developed that take advantage of the rich amount of metadata implicit in the collection of just a single CCD frame.<sup>49,50</sup> In addition to providing a central repository for data storage and archive, these databases can archive all the metadata related to each



The screenshot displays the EMEN web interface. At the top, the NCM I logo and navigation menu are visible. A breadcrumb trail shows the path: EMEN2 → Equipment → All groups → NCM I → Project HSV (Herpes Simplex Virus) → JEOL3200FSC → Subproject HSV1 wildtype B capsids → JEOL3200FSC → Imaging 2009/04/14 → ccd\_hsv1-b-1510. Below the trail, a metadata table for frame ccd\_hsv1-b-1510 is shown, detailing parameters such as magnification, dose rate, exposure time, and camera settings. To the right, a thumbnail of a cryo-EM micrograph shows several dark, circular virus particles against a noisy background. The interface includes various toolbars for editing, permissions, and image manipulation.

Metadata for Frame: ccd_hsv1-b-1510	
Frame ID :	ccd_hsv1-b-1510
Set Magnification:	80.0
Dose rate:	23.0 e/A2/sec
Exposure time:	1.0 s
A/pixel:	1.33 A/pix
Energy filter :	15eV Omega
<b>Camera</b>	
Camera:	Gatan 4kx4k US
Frame type:	Gain Normalized
Frame width:	4096 pixels
<b>Stage</b>	
Stage pos, X:	0.0 degree
B-factor in drift direction:	0.0 degree
<b>Assessment</b>	
Ice type:	vitreous
Ice cond:	great ice
Ice thick:	Perfect
Image quality:	5
<b>Corrected CTF Parameters</b>	
Meas defocus:	1.87733995914 um
Calibrated mag:	80.0 K
B-factor:	91.4024963379 A^2
SNR Max:	
Amp contrast:	10.0 %

**Figure 7** Electron Microscopy Electronic Notebook (EMEN). EMEN is an in-house electronic database developed by the NCM I for the storage and dissemination of electron microscopy data.<sup>50</sup> EMEN is available through the Internet (user restricted) and allows collaborators worldwide to connect and view and/or upload data relevant to the approximately 200 projects currently contained within the database. (a) Every entry in the database is hierarchically linked through a system of parent records, allowing the user to trace back every aspect of an imaging session, from microscope used to sample preparation and freezing. (b) As data are uploaded to the EMEN, metadata from the header of the image files are parsed and stored within the database. The EMEN platform allows for the inclusion of other subjective metadata such as image quality, and it provides an excellent resource for data mining. (c) A thumbnail of every image in EMEN is accessible via the web interface and provides an excellent resource for collaborators to visually browse entire data sets.

CCD frame in terms of the microscope lens parameters, cryo-specimen preparation parameters, and also biochemical purification procedures.<sup>50</sup> Furthermore, these databases should prove to be a useful tool for data organization, retrieval, and dissemination (Figure 7). Although not a requirement for the cryo-electron microscopist, these local databases provide an excellent means for investigators to store their data securely and visualize their data via the Internet.

### 1.16.6 Data Preprocessing

The next step in the reconstruction workflow is preprocessing the electron micrographs collected from the microscope. For single particle reconstructions, preprocessing typically follows the same three steps: screening, boxing, and CTF

determination. Each of these steps is essential to the reconstruction process, and although they can be as time-consuming as the data collection process, efforts have been made to automate them.<sup>51,52</sup> Alternatively, preprocessing tomogram data entails aligning the series of tilt images to each other. Due to the nature in which tomogram data are collected, the data are screened during data collection to ensure that if there is an aberrant image in the tilt series, another can be collected at the same tilt angle to prevent a loss of data for that angle.

#### 1.16.6.1 Screening (Single Particle Reconstruction)

The first step in any model reconstruction and refinement is gathering all of the data frames and deciding which of them

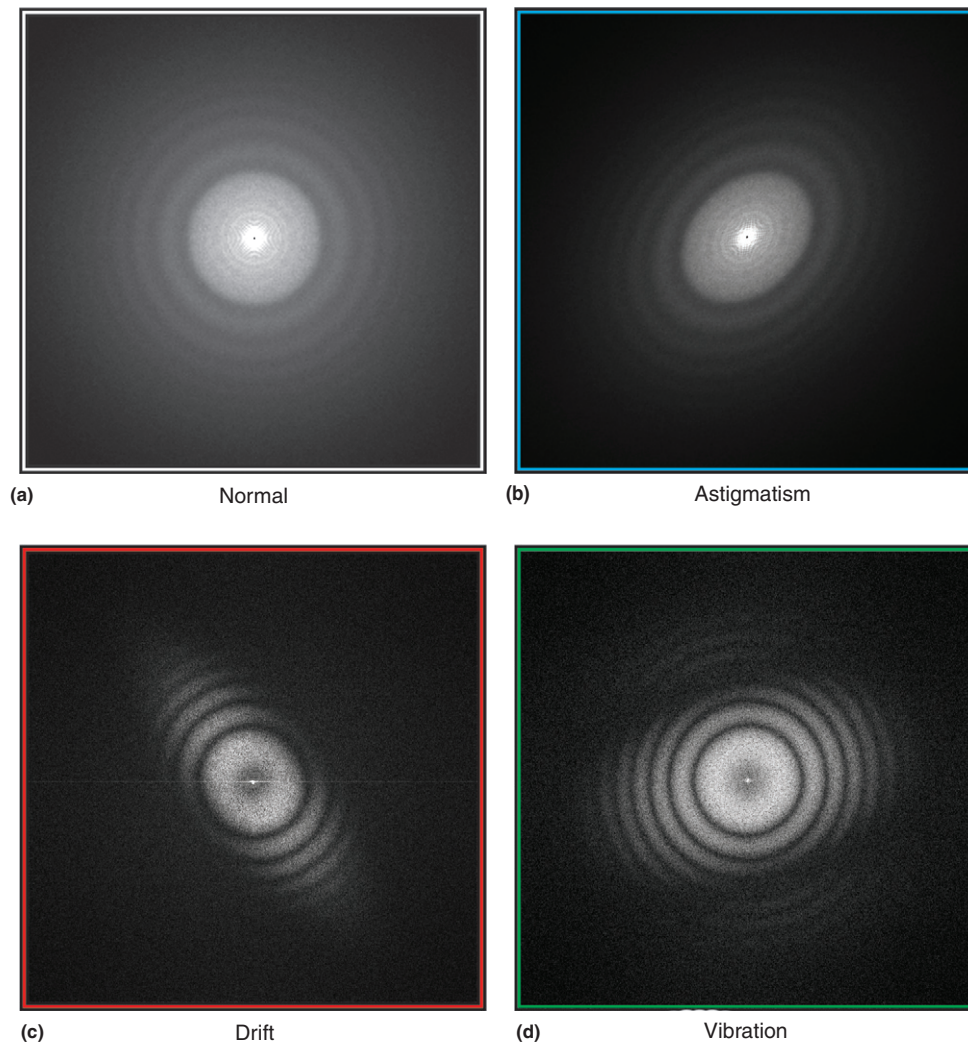


are ideal for further consideration. Because human eyes are the most rapid screening device, when collecting data on a CCD camera, it is possible to screen the images as they are being collected. However, because it is possible that there are subtle aberrations in the image undetectable to the human eye (e.g., astigmatism, vibration, and drift), the power spectrum of each image should be inspected as well (Figure 8). By excluding obviously bad images from the reconstruction, one can reduce the number of particles that could otherwise yield confounding or perhaps erroneous results. If the images are not screened during imaging, they should at least be processed before or during boxing by visual inspection of both the raw

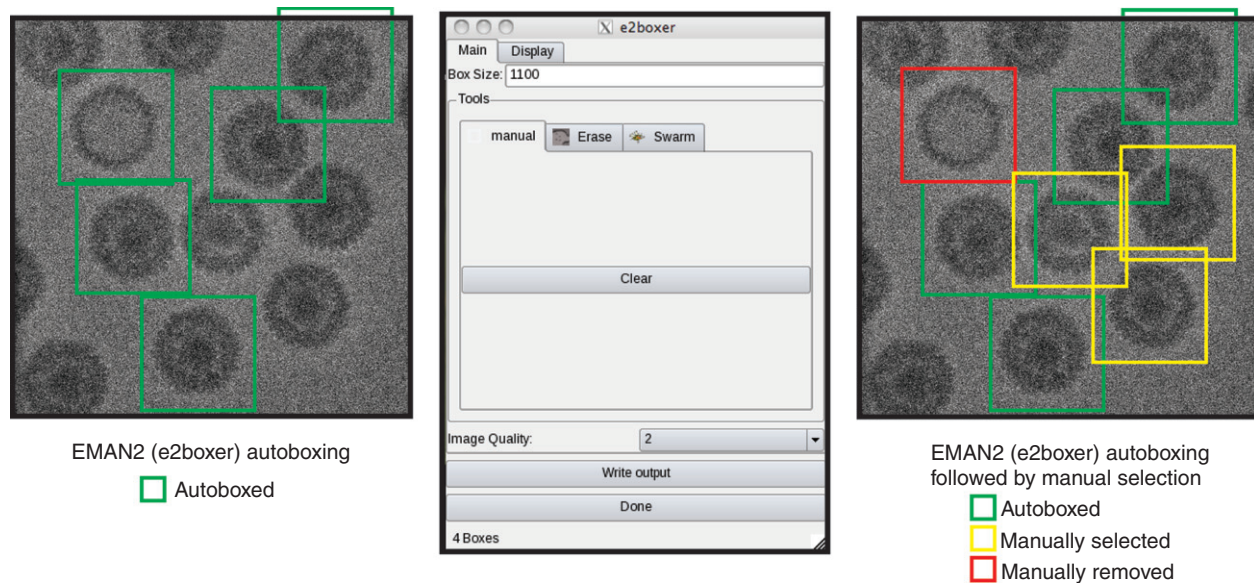
data and the power spectrum. Images that contain features as shown in Figure 8 should be excluded from the process.

#### 1.16.6.2 Boxing (Single Particle Reconstruction)

The next step in the process of single particle analysis is selecting individual virus particles in the micrographs, also known as boxing. The only parameters under control during the boxing process are the box size with which the individual particles are boxed and the position of these boxes with respect to particles they 'box.' It is important to note that an adequate box size in addition to reasonable box centering



**Figure 8** Micrograph aberrations visible in the power spectrum. (a) When an image is free of aberrations, the signal will appear as a series of concentric rings extending outward from the center of the FFT. (b) Astigmatism appears as elliptical rings in the FFT of an image and has the tendency to flatten out a rotationally averaged power spectrum, making it difficult to fit the CTF for astigmatic data. Astigmatism is an aberration of the illumination source, and because subtle differences in the sample can affect astigmatism, it is important to correct this aberration in the objective lens system while the sample is in place. (c) Drift occurs when the sample is moving during the time of exposure. Drift is often the cause of thermal instability, or not allowing the stage to come to rest after moving it. The characteristic feature of drift is symmetric degradation of the FFT across all spatial frequencies. (d) Vibration is the most insidious aberration because it can be of a physical or an acoustic etiology. Although at first glance vibration appears similar to drift in the FFT, it differs from drift in that it tends to occur at specific spatial frequencies, effectively degrading the signal beyond that resolution. The presence of a strong ring at low spatial frequencies and disconnected rings at high spatial frequencies is a sign of vibration.



**Figure 9** Autoboxing specificity and sensitivity. (Left) When *e2boxer* is used to automatically box particles in an image, it does a good job of selecting and centering around particles in the micrograph. Although the automatic selection algorithm does not select every particle in the micrograph, the sensitivity can be turned up to catch these particles. However, caution must be used to avoid making the algorithm so sensitive that it begins to falsely select particles. (Right) After running the autoboxing algorithm, it is possible to manually edit the particles selected by *e2boxer*. In the case of herpes simplex virus type 1, when different capsid types are present in the micrograph, it is important to manually select those missed by the autoboxer (yellow box) and remove those of a different type (red boxes). From Tang, G.; Peng, L.; Baldwin, P.; Mann, D.; Jiang, W.; Rees, I.; Ludtke, S. EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **2007**, *157*, 38–46.

about the particle is required for reconstruction. Furthermore, care must be taken in choosing the desired box size because a box that is just a single pixel smaller or larger than optimal can have a dramatic effect on the speed with which the data are processed. In general, the box size is set to be approximately 1.5 to 1.75 times the size of the particle so as to ensure that there is adequate background in the box to generate a noise profile for each particle for accurate CTF correction, alignment, and refinement. Programs are available (e.g., EMAN and EMAN2) that simplify this process by allowing the user to view individual micrographs, select particles, and export the single particle images.<sup>51,52</sup> However, caution must be exercised because the process is neither 100% specific or sensitive in any software. Furthermore, if a sample is not 100% homogeneous, the automatic boxing programs may inappropriately include other particles of the same shape and size. Therefore, it is recommended that automatic selection be followed by an additional round of manual inspection so as to exclude any erroneously selected particles and include those particle missed by the autoboxing algorithm (Figure 9). The advantage of using an autoboxing program is that all of the boxes are automatically centered on the selected particles, a process that is very labor-intensive if performed manually. Because it is easier to delete boxes generated by the autoboxing programs than it is to manually select and center boxes, it is recommended that the threshold for autoboxing be set low enough to box all features in a micrograph. The same discretion that is used in image screening should also be used when boxing or inspecting boxed particles. Features such as charging<sup>53</sup> that may not be readily visible in the power spectrum of an entire micrograph may become apparent when viewing the fast

Fourier transform (FFT) of an individual particle, in which case the particle should either not be selected or be deleted.

### 1.16.6.3 Contrast Transfer Function Determination (Single Particle Reconstruction)

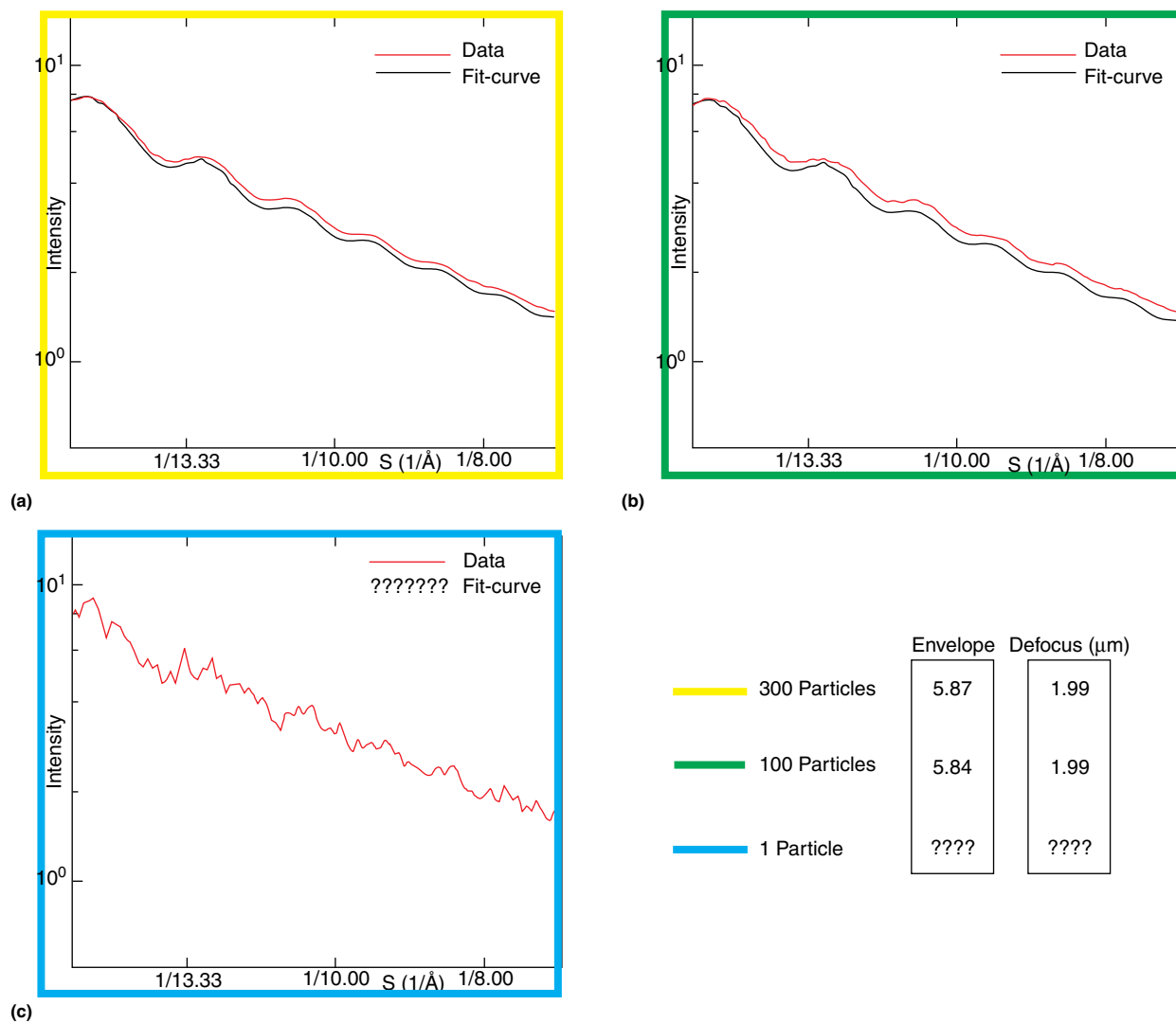
A characteristic feature of any image taken on an electron microscope is the presence of a CTF, which appears as a sinusoidal pattern with decaying periodicity in the power spectrum.<sup>54,55</sup> The CTF (eqn [1]), a function of spatial frequency ( $S$ ), describes how much contrast is present in a single micrograph and is dependent on both the microscope and the imaging conditions.<sup>54</sup>

$$\text{CTF}(S) = - \left[ \sin(\gamma(S)) * \sqrt{(1 - Q)^2 + \cos(\gamma(S)) * Q} \right]$$

$$\gamma(S) = \pi * [(-C_s * \lambda^3 * S^4) / 2 + \Delta z * \lambda * S^2] \quad [1]$$

Whereas two of these variables – spherical aberration of the objective lens ( $C_s$ ) and wavelength of the accelerated electrons ( $\lambda$ ) – are a function of the microscope, the others – defocus ( $\Delta z$ ) and amplitude contrast ( $Q$ ) – describe features characteristic of electron microscope conditions. In this formulation, the micrograph is assumed to be astigmatism free at least up to the resolution of the reconstruction.

One of the most laborious, and important, tasks in any cryo-EM experiment is fitting the CTF of one's data. Because the CTF is a function of parameters specific to an individual micrograph, it is necessary to fit the CTF for every micrograph. Furthermore, because the CTF is also a function of defocus, it



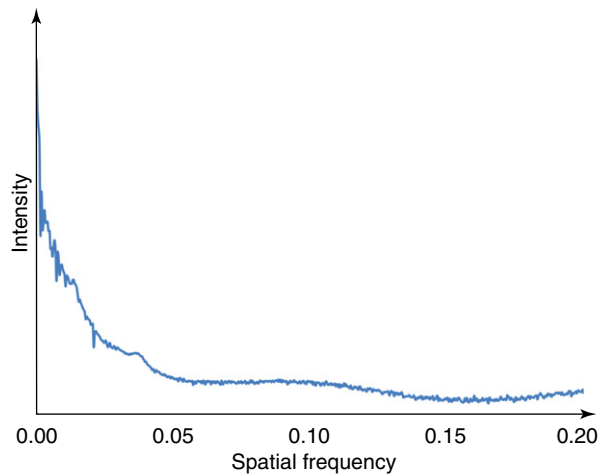
**Figure 10** Signal as a function of number of boxed particles. In each window (a–c), the fit CTF (black) is shifted below the raw data (red) to show features of the two curves; in a true fit, the two curves should be nearly superimposed. (a) In a micrograph with 300 particles, when all 300 are boxed, it is relatively easy to fit the CTF for the data. Furthermore, high-resolution data are visible as ripples in the raw data at high spatial frequency in the averaged FFT for all 300 particles. (b) When looking at a subset of 100 of the total 300 particles in the micrograph, it is still possible to fit the raw data out to nearly the same resolution (envelope). (c) When only one particle is boxed, it becomes very difficult to fit the raw data accurately, and there is no guarantee that a fit of the data made by eye is correct. As a result of the error inherent in fitting single particle CTF data, care must be taken to verify the accuracy of single particle CTF measurements.

may be advantageous to extend this to every particle in a single image frame because variations in ice thickness can allow for multiple focal heights. For any given image, taking a rotational average of the FFT of all the particles in that image produces the CTF. However, depending on the sample and the imaging conditions, SNR becomes an issue when trying to accurately fit a small number of particles, especially when trying to determine the CTF for a small number of particles (Figure 10). In addition, in order to accurately fit the CTF of the data, it is recommended that a structure factor be used during fitting. Structure factors can be computed from X-ray solution scattering<sup>56</sup> or through manual fitting of a subset of the single particle data (Figure 11).

The CTF can be fit manually using programs such as *ctfit* in EMAN, in which the parameters discussed previously are

modulated to fit the experimental curve (Figure 12). During CTF correction, the parameter that is by far the most sensitive to error is the defocus determined for the individual images. Because phase can be flipped and lost at different frequencies during the imaging process, fitting the defocus parameter in the CTF is a way to attempt to recover this information. The standard method for defocus determination is to correctly identify the ‘zeros’ in the CTF and fit them accordingly. Because the CTF zeros are less sensitive to defocus at low spatial frequencies, it is important to fit the defocus parameter at high spatial frequencies (as far out as one still has signal) so as to ensure that one does not flip the phase of the experimental data in relation to the fit (Figure 12). For this reason, the best approach to fitting the CTF is to correctly determine the noise profile for a micrograph and then adjust the defocus

as a last step. In addition to the parameters discussed previously, Fourier amplitude fall-off (known as the envelope function) is another parameter that must be fit for each CTF



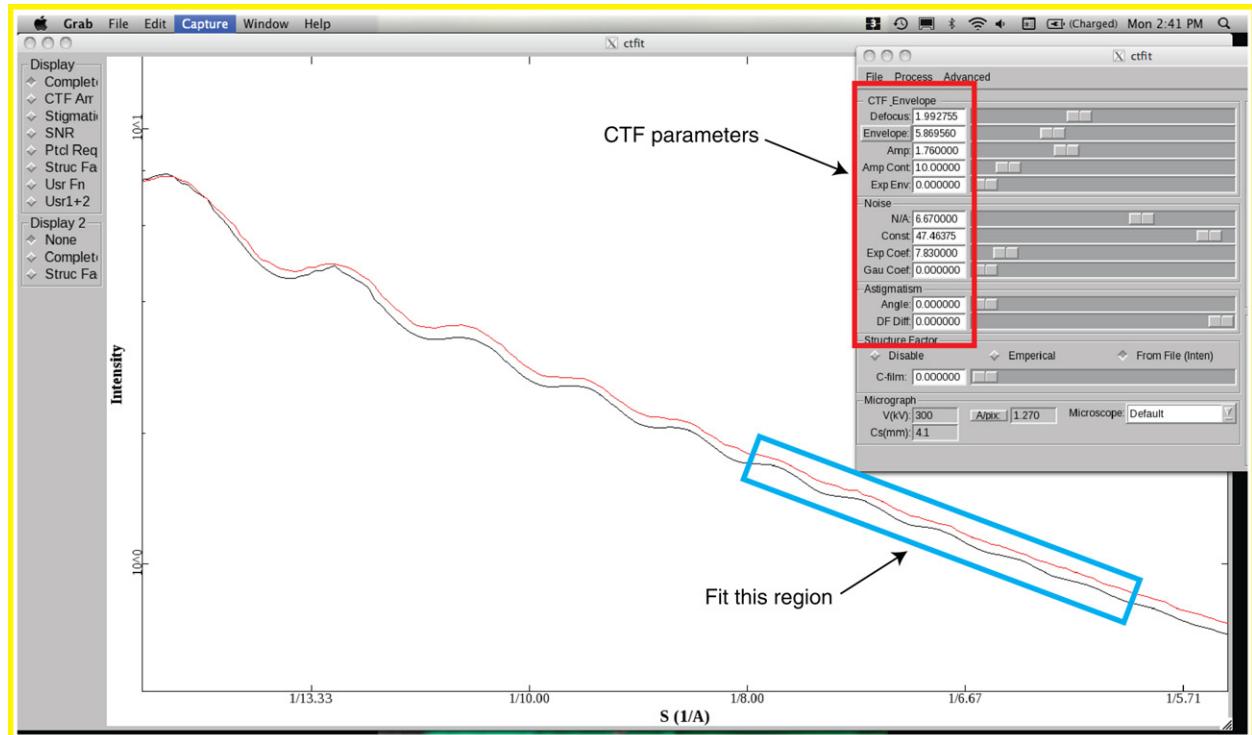
**Figure 11** X-ray solution scattering data for herpes simplex virus type I B-capsid. A solution scattering, or structure factor, curve is necessary for accurately fitting the parameters of a CTF, and it is further used during reconstruction to radially scale the densities in the generated map. Although it is more difficult to ascertain a solution scattering curve because it is the result of X-ray scattering experiments, it is possible to approximate this curve by generating a structure factor curve from the raw cryo-EM data (<http://blake.bcm.edu/emanwiki/EMAN1/FAQ>).

curve and contributes significantly to the overall signal, especially at high spatial frequencies. Because this fall-off is a function of the instrument, specimen conditions, and the recording medium, it is not expected as a component of the theoretical CTF shown in eqn [1]. As such, for the purposes of CTF fitting, the effective envelope function can be assumed to decay as a Gaussian or an exponential function with respect to spatial frequency. A further complication in fitting the CTF of raw data is the presence of astigmatism, and although for most practical applications astigmatism is assumed to be insignificant, it is possible to determine and correct for its presence.<sup>57</sup>

### 1.16.7 Single Particle Reconstruction

The 3-D reconstruction of a virus from multiple single particle images is the last step in the computationally intensive process of particle alignment.<sup>1</sup> The general principle behind the theory is that given a series of single particle images, once you have determined their icosahedral (or asymmetric) orientations, you can combine the data to form one cohesive 3-D model. Although this process can be applied to a small number of particles, unless a large number of particles are used, many features – such as preferred orientation, limited defocus range, and noise – keep a reconstruction from reaching high resolution. To circumvent these issues, it has been common practice to use an ever-increasing number of particles.<sup>58</sup>

The majority of the subnanometer resolution virus maps that have been published to date assume icosahedral



**Figure 12** Contrast transfer function (CTF) fitting. Programs such as EMAN<sup>51</sup> allow the user to fit the CTF for a series of single particle images by modulating a variety of variables. Accurate CTF correction is necessary for high-resolution reconstructions. To maximize the information recovered from the CTF, it is important to verify the fit at high spatial frequencies (cyan box). Failure to check the fit at both low and high spatial frequencies can introduce phase flipping in the data, obscuring the retrieval of high-resolution information from the data. From Ludtke, S.; Baldwin, P.; Chiu, W. EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **1999**, *128*, 82–97.



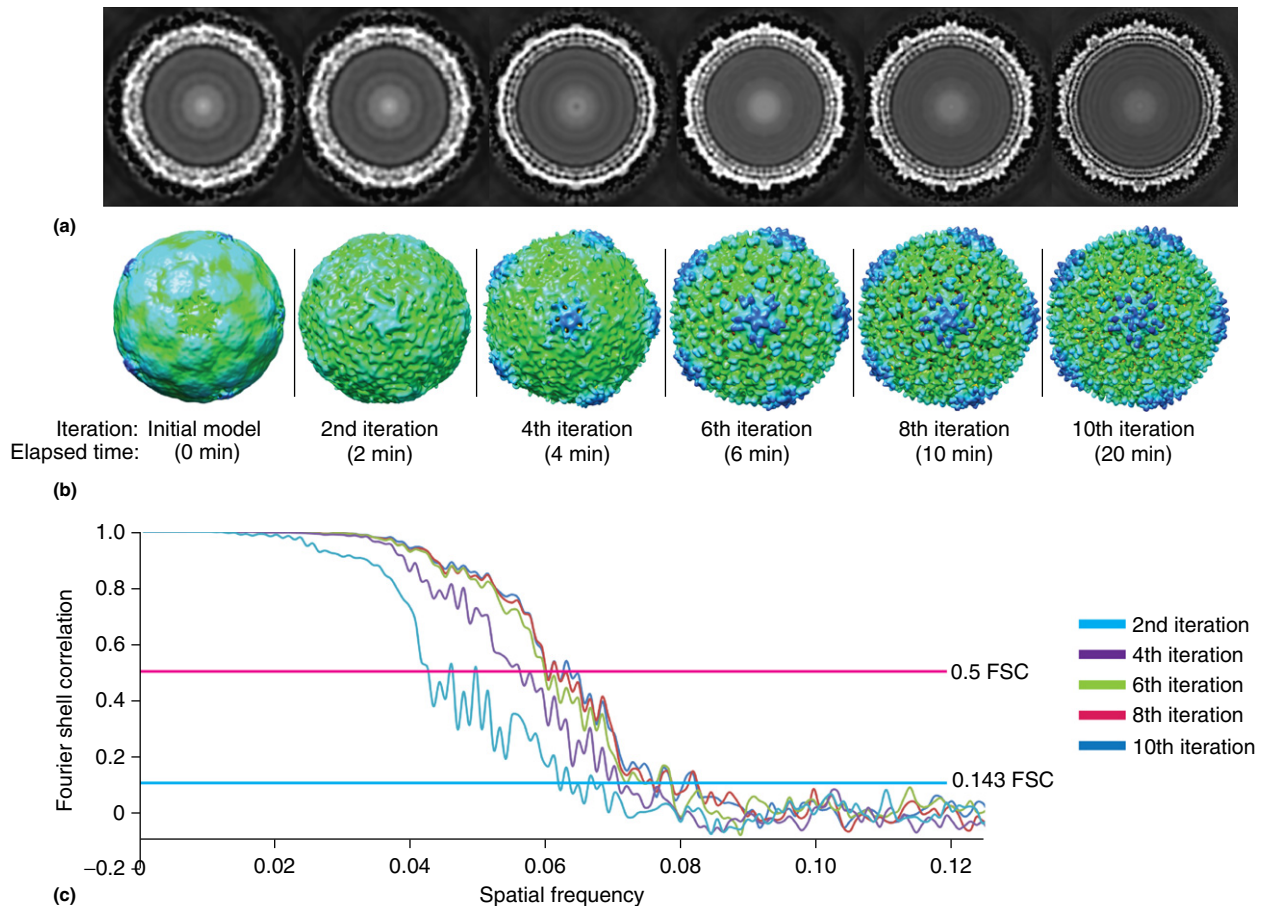
symmetry, but in the past few years, a handful of asymmetric virus structures have been published.<sup>6,12–16,59,60</sup> Although non-icosahedral reconstructions represent the ideal target for the microscopist, they are increasingly elusive because the asymmetric alignment procedure is far more complicated<sup>6</sup> and they require nearly 60 times as much data for an equivalent icosahedral reconstruction resolution.

### 1.16.7.1 Particle Alignment and Refinement

Three-dimensional reconstruction has been the source of a great deal of research in the field and has led to the development of many programs, including EMAN,<sup>51</sup> EMAN2,<sup>52</sup> MPSA,<sup>58</sup> Spider,<sup>61</sup> XMIPP,<sup>62</sup> IMAGIC,<sup>63</sup> FREALIGN,<sup>64</sup> AUTO3DEM,<sup>65</sup> SPARX,<sup>66</sup> and IMIRS.<sup>67</sup> Although some of these programs are more efficient than others, the choice of which to use is usually a matter of personal preference.

Depending on which program one chooses, the process of 3-D reconstruction can be further broken down into a variety of subprocesses. Programs such as EMAN and EMAN2 have

multiple alignment algorithms, from the traditional common lines approach to principal component analysis and projection matching. Although the projection matching algorithm is an approach that has been used to obtain near-atomic resolution structures,<sup>43</sup> it uses an exhaustive search to refine all five alignment variables (the three Euler angles ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) and the particle's center ( $x$ ,  $y$ )) simultaneously and as such is computationally expensive. On the other hand, the Multi-Path Simulated Annealing (MPSA) algorithm uses an algorithm based on cross common lines in Fourier space and handles model refinement separately from the initial orientation determination.<sup>58</sup> MPSA has been shown to be an effective means for icosahedral virus reconstruction,<sup>6,16,58</sup> and because it is approximately 1000 times faster than the exhaustive search described previously, it is extremely efficient for large data sets (a requirement for high-resolution viral reconstruction). Another advantage of MPSA is that unlike other similar reconstruction algorithms, it does not require an accurate initial model, making it possible to reconstruct uniquely shaped viruses *de novo* (i.e., free of model bias) in minutes



**Figure 13** *De novo* model generation using MPSA. Often, an accurate initial model for a virus is not available, making it necessary to perform a *de novo* reconstruction of the data. To initialize this process, a random spherical ‘initial model’ (a and b), with dimensions approximate to that of the data, is used as an initial model. The reported times are based on processing with 100 CPUs operating at 2.66 GHz. (a) and (b) Using MPSA, it is possible to resolve very coarse features of the bacteriophage  $\epsilon 15$  in as little as 2 min (after two iterations), and a near subnanometer resolution initial model can be generated in less than 1 h (orthoslice (a) and 3-D reconstruction (b)). (c) An FSC plot of these reconstructions shows that by the eighth iteration (after 10 min), the model has converged and the resolution is high enough to use as an initial model for refinement to higher resolutions.

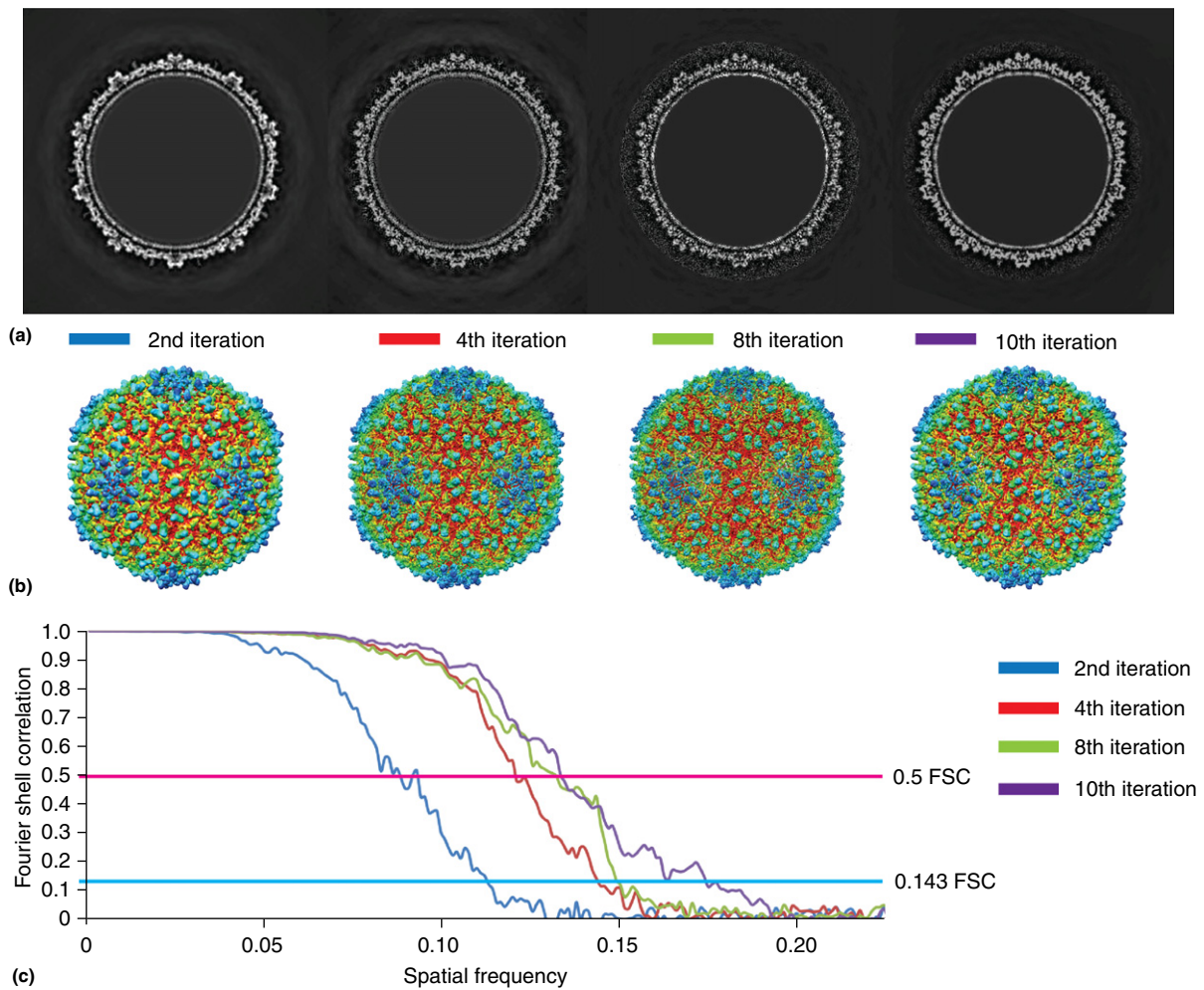
(Figure 13). This technique can be used to build accurate initial models for use in refinement of the data to sub-nanometer resolutions. Although any of the programs listed previously can be used and have been shown to work for virus reconstructions, the following section focuses on the use of MPSA as a means to do so.

#### 1.16.7.1.1 Symmetry-imposed alignment

The first method for virus particle alignment and reconstruction was described in 1970.<sup>1</sup> The method was broken into two separate processes: particle center determination using cross-correlation with a reference image and orientation determination through the Fourier space self common lines approach.<sup>1</sup> This process performed an exhaustive search over the entire orientation space in an asymmetric unit, searching for the orientation with the highest self common line correlation. These alignments were further refined by finding the orientations with the highest cross common line correlation

between raw particle images and reference projections from template models. The disadvantage of this approach is that the method decouples the center and orientation search, a problem that becomes an issue in the later steps of refinement. Furthermore, the self common lines approach is highly dependent on low spatial frequency contrast, making it difficult to properly identify the icosahedral orientation of particles imaged with a defocus less than 2  $\mu\text{m}$ .<sup>58</sup> Because close-to-focus image reconstruction was not possible with this algorithm, it was difficult to obtain high-resolution 3-D reconstructions from a single set of images.

As an alternative to the self common lines approach for initial particle alignment, cross common line correlation was proposed as a method that could potentially circumvent the problems of the self common lines approach, namely accurate particle center determination and low-frequency contrast dependence.<sup>37,68</sup> Cross common lines were first implemented as a search over the entire parameter space for a large set of



**Figure 14** Model refinement of  $\epsilon 15$  phage using the MPSA algorithm. Once an adequately resolved model for the data has been generated, it is possible to begin refining the individual particle orientations to produce a better resolved model. Refinement is an iterative process that incrementally improves model resolution by fine-tuning particle orientations before reconstruction. Refinement is best performed once the particle's approximate orientations are known as it becomes more computationally efficient to search a smaller orientation space. The successive improvements in the model can be seen in both the reconstructed data (orthoslice (a) and 3-D model (b)) and the FSC computed for the data (c).

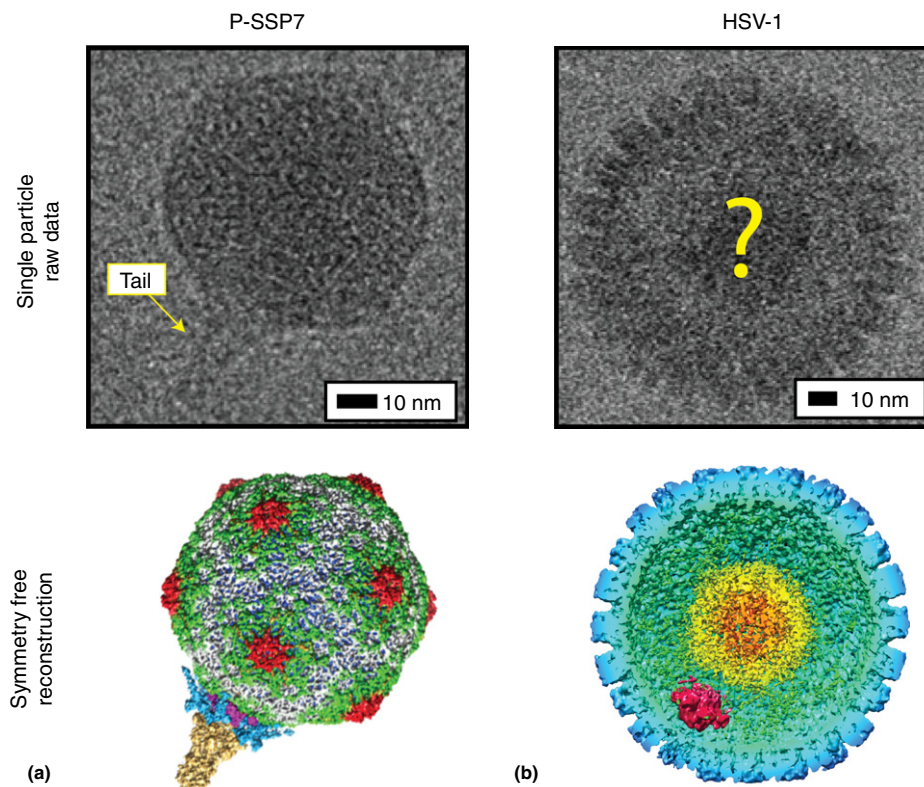


image data,<sup>37</sup> a method that was limited by the fact that it scales on the order of approximately  $10^7$ , making it a computationally expensive method for searching the entire five-parameter orientation space ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $x$ , and  $y$ ).

In an attempt to capitalize on the advantages of the cross common lines approach and minimize the computational requirements of an exhaustive search, an alternative method was developed to treat particle alignment and refinement as an optimization problem.<sup>58</sup> This program implements an MPSA algorithm to search the parameter landscape for the lowest energy condition, corresponding to the orientation with the highest cross common lines correlation. The novelty of this program is that it uses a series of optimization paths that can communicate with each other, effectively allowing each individual path to jump to a point of lower energy identified by another path. This approach has been shown to reduce the computational load of the traditional expensive cross common lines search approximately 1000-fold. Furthermore, this process allows for local refinement of the particle orientations

once they have converged, providing a means by which high-resolution cryo-EM maps can be generated.

In the framework of the MPSA algorithm, this refinement is implemented as an iterative search over parameter space, confined to a region of the space that corresponds to the previous iterations results. Under the standard conditions of initial alignment, the MPSA algorithm searches a much wider space than during refinement, and by limiting the range of the search during refinement, the algorithm can finely probe the parameter space around orientations that are presumed to be relatively close to the true orientation. As such, it is necessary for the model to converge to near subnanometer resolutions before attempting to refine the data any further. Once a reasonable high-resolution model of the virus has been generated, it is possible to begin refining the orientations to generate a higher resolution final model. Because refinement is a computationally intensive process, it is typically reserved for finely tuning orientations that are already near their true orientation (Figure 14).



**Figure 15** Identifying non-icosahedral features during reconstruction. For virus reconstructions, the portal and/or portal complex are often used as the non-icosahedral feature from which the particles are aligned. (a) P-SSP7 phage has a very large portal complex with a protruding tail that extends well beyond the outer capsid surface [EMID-1715]. The portal in P-SSP7 is easily visible in the raw data, and as such it is far easier to reconstruct without any imposed symmetry. From Liu, X.; Zhang, Q.; Murata, K.; Baker, M. L.; Sullivan, M. B.; Fu, C.; Dougherty, M. T.; Schmid, M. F.; Osburne, M. S.; Chisholm, S. W.; Chiu, W. Structural changes in a marine podovirus associated with release of its genome into *Prochlorococcus*. *Nat. Struct. Mol. Biol.* **2010**, *17*, 830–836. Copyright by Nature. (b) The portal in the herpes simplex type I capsid resides on the inner capsid surface beneath a fivefold vertex. [EMID-5255] From Rochat, R. H.; Liu, X.; Murata, K.; Nagayama, K.; Rixon, F. J.; Chiu, W. Seeing the portal in herpes simplex virus type 1 B capsids. *J. Virol.* **2011**, *85*, 1871–1874. Because this feature is located internally and has a similar size as the penton, it is difficult, if not impossible, to visually identify the portal in raw micrographs, which is an issue with regard to asymmetric particle alignment.

The publication of several ultra-high-resolution virus structures ( $<4.0 \text{ \AA}$ ) has resulted in the introduction of many new procedures for model refinement.<sup>3,5,10,69</sup> Issues ranging from local differences in the specimen imaging area to subtle aberrations in the electron beam can have a profound effect on processing single particle data for a high-resolution reconstruction. When targeting a resolution less than  $4.0 \text{ \AA}$ , it is necessary to consider correcting these variations on a per micrograph (astigmatism and beam effects) and per particle (magnification, defocus, and tilt) basis. Because the local environment in which each image is collected can produce variations in magnification and astigmatism, these aberrations should be iteratively refined in the final stages of producing a high-resolution reconstruction.<sup>5,10,69</sup> In addition, because variation in ice thickness and specimen tilt can position individual particles at different focal distances, applying a global CTF correction for defocus can potentially ablate high-resolution features of the data.<sup>3,69</sup> To compensate for these differences, local refinement of the defocus for each particle in a micrograph can minimize phase flipping at high spatial frequencies and preserve the high-resolution features in the reconstruction (Figure 12). Because the resolution of single particle cryo-EM reconstructions are pushing nearer to atomic resolution, it is likely that further corrections will be necessary to compensate for irregularities in the beam and sample.

#### 1.16.7.1.2 Asymmetric alignment

The majority of virus reconstructions published in the literature are of specimens for which there is some assumed symmetry to the map. For many viruses, the assumed symmetry is icosahedral; however, due to the nature in which these viruses are built and mature, there are important molecular components in their overall structure that are not symmetric. Accordingly, there has been an increased interest in the determination of the structure of these viruses without imposed symmetry because it will help shed light on the process of viral assembly and maturation.<sup>6,12-16,70</sup> The assumption of symmetry is often made because it simplifies the process of orientation determination. Furthermore, by icosahedrally averaging a map, the information content of a data set is enhanced, making it possible to achieve high-resolution reconstructions with nearly 60 times less data than for a reconstruction in which no symmetry is assumed.<sup>68</sup> The process of orientation determination for an asymmetric reconstruction differs from the traditional symmetric orientation search in that one must be able to identify the non-icosahedral components in the virus particle.<sup>6</sup> In the case of a phage such as P-SSP7, which has a large tail and portal structure, it is conceptually easy to understand how one would determine the asymmetric orientation of the particles because the non-icosahedrally symmetric feature is visible in the raw micrographs (Figure 15(a)). However, in the case of a virus such as HSV-1 in which the structure does not have such a protruding feature that is readily identifiable in raw micrographs, it is more difficult to find the true asymmetric orientation of the particle (Figure 15(b)). Once the asymmetric orientation of every particle has been determined, the process of reconstructing the virus asymmetrically is identical to any other reconstruction in which symmetry is assumed.

#### 1.16.7.2 Single Particle Data Subset Selection

Traditionally, the resolution of a map is improved as increasingly more particles are added to the reconstruction. However, this must be weighed against the observation that some data are of better quality than others, and inclusion of data that contain aberrations may in fact do more harm than good in enhancing the resolution of a reconstruction. The quality of a single particle can be thought of as a metric of conformational heterogeneity, icosahedral symmetry, and imaging conditions.<sup>56</sup> The better all three of these factors are, the more likely the particle will contribute to producing a high-resolution reconstruction.

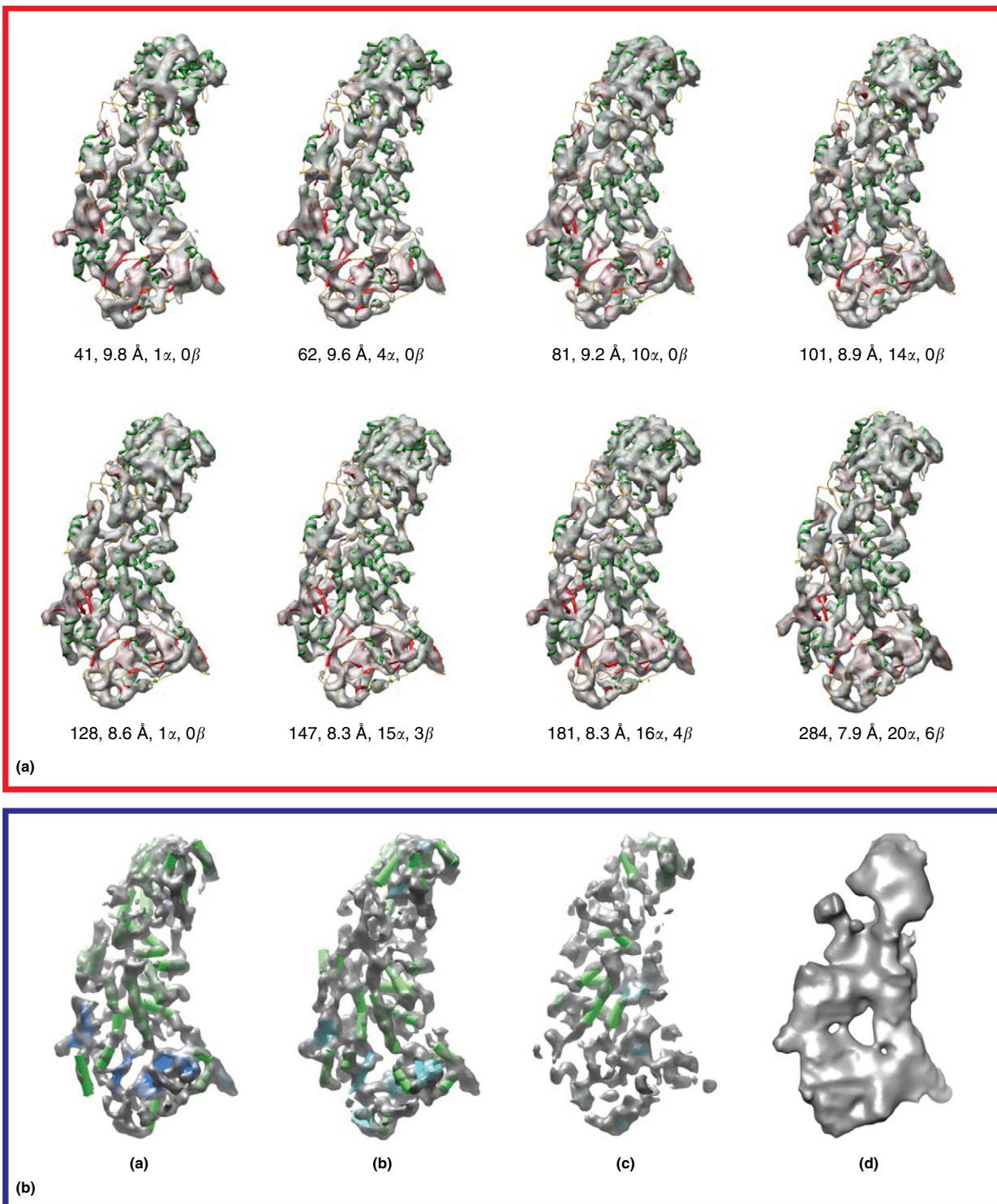
The number of particles required to achieve a specific resolution reconstruction depends on assumptions made about the data, both without<sup>1</sup> and with noise.<sup>71,72</sup> In the noise-free and even particle orientation sampling assumption, the predicted value of  $n = \pi D/d$ , where  $n$  is number of particles,  $D$  is the particle diameter, and  $d$  is the resolution.<sup>1</sup> By taking into account the presence of noise, the more rigorous estimate for resolution ( $d$ ) as a function of particle number for a given B-factor in the data is

$$(1/N_{\text{asym}})[(\langle S \rangle^2 / \langle N \rangle^2) \cdot 30\pi / N_e \sigma_e d] e^{B/2d^2}$$

where  $N_{\text{asym}}$  is the number of asymmetric units in the data (60 for icosahedral viruses),  $\langle S \rangle$  is the signal computed from the data,  $\langle N \rangle$  is the noise from the data,  $N_e$  is the dose at which the data were collected,  $\sigma_e$  is the elastic cross section for carbon, and  $B$  is the Fourier decay B-factor for the data. This equation has been simplified in the case of single particle cryo-EM as a plot of  $\log(N_{\text{asym}}d)$  versus  $1/d$ , where  $N_{\text{asym}}$  is the number of particles in the data set and  $d$  is the target resolution,<sup>72</sup> a formulation that in the presence of noise has been confirmed experimentally.<sup>58</sup> Consequently, as the B-factor for the data increases, so does the quantity of data needed for an equivalent resolution. It is important to note that this effective B-factor depends not only on microscope envelope function and experimental conditions but also on error in orientation and center alignment introduced during data processing.

In the event that there is good conformational heterogeneity in a sample, selecting a subset of 'good particles' is an effective means by which map resolution can be enhanced with the exclusion, not inclusion, of data.<sup>58</sup> It has been shown that with as few as 50 good particles, with an arbitrary initial model, it is possible to achieve subnanometer resolution icosahedral reconstructions, and these maps are accurate enough to begin to identify some long  $\alpha$  helices (Figure 16).<sup>58</sup> Furthermore, if multiple conformational states are present in a single micrograph, the different states must be isolated from each other before reconstruction.

Often, the 'best' particles are those that have been determined, through the process of iterative alignment and refinement, to consistently score higher than other particles.<sup>58</sup> Although there is no concrete method to determine *a priori* which particles are the best, there are many steps that can be taken to ensure that the quality of data you put into your reconstruction is of the highest quality possible. First, in the event that an accurate initial model for your sample exists,



**Figure 16** Resolvable features within the P3A subunit of RDV while minimizing the number of particles used. This data set is composed of 4865 raw particle images, of which 4200 were determined to be ‘good’,<sup>58</sup> and it demonstrates the effect of bad, good, and best particles on map resolvability. (a) From the top left to the bottom right, as the number of ‘best particles’ included in the reconstruction increases, so does the overall resolution of the map. As the resolution increases, so do the number of resolvable features within the map as predicted by SSEhunter and Helixhunter. (b) The particles used in a reconstruction have a drastic effect on map resolution and feature resolvability. The left-most reconstruction is composed of the best 284 particles (7.8 Å); as one randomly selects 284 of the good particles, the resolution decreases to 9.2 Å, and then it decreases to 10.8 Å as one randomly selects 284 particles from the entire data set. The final map on the right is a reconstruction of the 665 particles (52.4 Å) that were identified by the program as being ‘bad.’ Adapted from Liu, X.; Jiang, W.; Jakana, J.; Chiu, W. 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.* **2007**, *160*, 11–27.



there is no need to collect data at a defocus higher than 2  $\mu\text{m}$ . In addition, by carefully screening your data on the front end, it is possible to eliminate entire micrographs of poor-quality data. A second level of screening can be performed at the level of boxing, where it is possible to identify and remove particles that may be experiencing the effects of local charging. Although there is no way to directly control for conformational heterogeneity within your sample, if there are visible differences in the raw particle data, it is possible to ameliorate this problem by selecting subsets of the data (Figure 9). Nevertheless, even if 'bad' particles make it through these steps of screening, in some cases, the process of data refinement will eliminate particles that are not consistent with the existing pool of data.

### 1.16.7.3 Three-Dimensional Reconstruction

Once the orientations of all of the particles in a data set have been determined, it is possible to reconstruct the 3-D model of the virus. This process works by merging all of the 2-D particle images, after CTF correction, into a single 3-D map. 3-D reconstruction programs typically have a method to generate these 3-D density maps, and to achieve this task EMAN and EMAN2 use the programs *make3d* and *e2make3d*, respectively. These two programs are based on a direct Fourier inversion method to generate the 3-D volumetric data, and they have a variety of command line parameters that allow the user to specify symmetries or data preprocessing steps. Other reconstruction programs use a similar algorithm, FREALIGN,<sup>73</sup> and although this approach is more computationally efficient than the Fourier-Bessel method, as used in IMIRS,<sup>67</sup> the memory requirements are far greater. If a map is being processed asymmetrically, it is important to note that there is no difference in the reconstruction algorithm, other than the fact that no symmetry is applied to the map.

## 1.16.8 Tomographic Reconstruction

As mentioned previously, cryo-ET is an alternative technique for visualizing and reconstructing viruses in or near their native environment. Like single particle cryo-EM, the process of averaging subtomograms relies on aligning particles of identical conformations together to enhance the model resolution. The process of tomogram reconstruction has been well reviewed elsewhere,<sup>74</sup> and the following section focuses on subtomogram averaging for virus structure determination.

### 1.16.8.1 Tilt-Series Alignment

Because a tomogram is a 3-D representation of a volume of data recorded from a series of 2-D images, it is necessary to accurately align all of the 2-D data to itself in order to generate a tomogram. Depending on how well the eucentric height was determined, there may be little or no translational shift in the data. Nevertheless, even if there is little shift across all tilt angles, it is necessary to track the movement of fiducial markers across the entire tilt range. Theoretically, one should need only a handful of fiducial markers in a single tilt series to

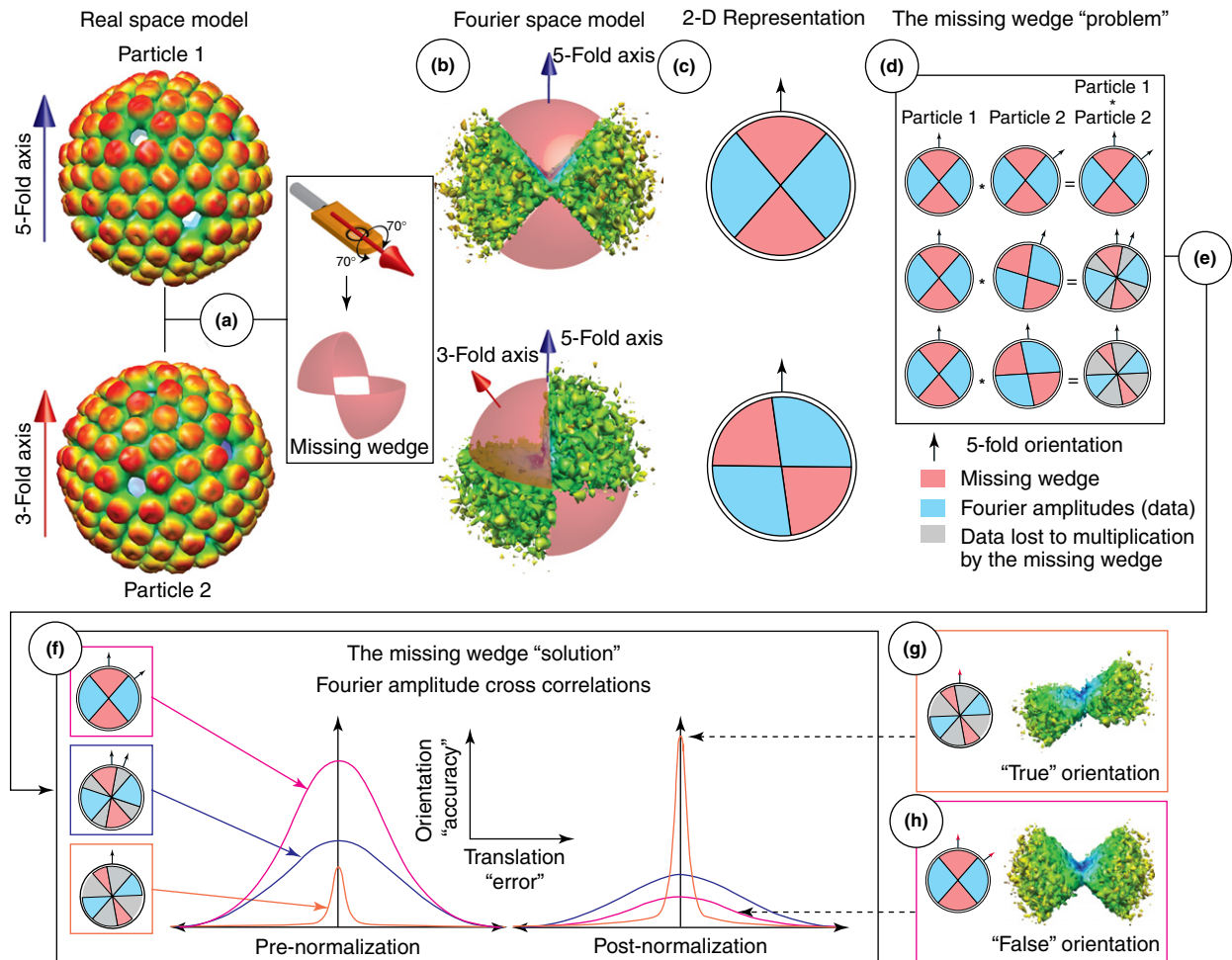
accurately align the tomogram; however, this is complicated by the fact that at some tilt angles, these markers may become obscured or unidentifiable by other features in the micrograph. For this reason, it is best to have many fiducial markers present in the micrograph so that even at high-degree tilt angles one can still align the tomogram (Figure 3(b)).

Image alignment is performed using a graphical user interface (GUI) computer program that allows the user to identify fiducial markers and track them through the entire stack of images. For alignment and reconstruction, one of the more popular tools is the publically available program IMOD.<sup>75</sup> By relating the fiducial markers to themselves across the entire stack of images in the tomographic tilt series, it is possible to align and extract a 3-D volume representative of the imaged area. Furthermore, because a single tomogram can occupy 10–100 GB of space, it is often trimmed prior to further processing. Eventually, the 3-D volumetric reconstruction can then be used to extract subtomograms for classification, alignment, and averaging.<sup>26–28</sup>

Whereas it is common to correct the CTF in single particle reconstructions, the total dose per micrograph in cryo-ET is so low that accurately correcting the CTF is not trivial. Because tomographic reconstructions are typically targeted at low resolution ( $<20 \text{ \AA}$ ), CTF correction has not normally been included as part of the data processing scheme. Although it is possible to ignore the CTF for most tomographic reconstructions, the manifestation of the limited tilt angle present in tomography introduces another confounding artifact known as the 'missing wedge' and requires the user to address the issues of alignment and the missing wedge during averaging.<sup>28</sup>

### 1.16.8.2 Alignment and the 'Missing Wedge'

Just as in single particle cryo-EM, there are a variety of schemes for determining the orientation and alignment of the particles in the extracted subtomograms.<sup>25,26,28,51,76</sup> Accordingly, a well-resolved tomographic average requires the individual subtomograms containing the extracted particles to be aligned with respect to each other – a process that is rarely as straightforward as in single particle cryo-EM. Because most cryo-holders have a maximum tilt angle of  $\pm 70^\circ$  (although some  $80^\circ$  holders are available), the total coverage for a tomogram is limited to  $140^\circ$  at best (Figure 17(a)). Ideally, a tomogram would contain data collected across a full  $180^\circ$ ; however, because this is not possible with the currently available cryo-holders, this lack of information from  $70^\circ$  to  $90^\circ$  is manifested as a missing wedge of data in Fourier space (Figure 17(b) and (c)), and if it is not corrected for, it can distort the final reconstruction. An additional complicating factor is that because the tilt angle for a group of particles in a single tomogram is identical, all the particles have the same missing wedge. However, because the particles are randomly oriented in the sample, each particle has different missing wedge data with respect to its orientation (Figure 17(b) and (c)). Because the missing wedge is manifest as a value of zero in the FFT of the data, calculation of the cross-correlation between two particles can result in the elimination of a great deal of information in Fourier space because cross-correlation involves multiplication by these zeros (Figure 17(c)). To



**Figure 17** The 'missing wedge' problem. (a) 3-D reconstruction of an icosahedral virus oriented along two different axes (fivefold (top) and threefold (bottom)). During tomography, the limited tilt range of the sample holder results in the creation of a missing wedge of data that is visible in the 3-D Fourier transform of the data. (b) The 3-D Fourier transform of data from two icosahedral particles shows that when the particles are oriented along the same axis (fivefold), their missing wedges are not necessarily oriented with respect to each other. (c) A 2-D simplification of the 3-D Fourier transforms shown in panel b; the Fourier amplitudes are shown in cyan and the missing wedge in red (both sets of data are still oriented along the fivefold axis). (d) To illustrate the missing wedge problem, the product of the Fourier amplitudes for the data in panels b and c is shown. Because the missing wedge is manifest as a zero value in Fourier space, any multiplication by the missing wedge will eliminate Fourier amplitudes in the data (light red wedges). When the total sum of Fourier amplitudes is used to choose the best orientation of the two particles with respect to each other, the data eliminated by the missing wedge will result in selection of the wrong orientation. In this example, the best orientation (at the bottom) is not selected because the majority of the Fourier amplitudes were eliminated when multiplied by the missing wedge. (e) To minimize the effect of this pitfall caused by the missing wedge, an alternative approach, for orienting two particles to each other, is needed that can circumvent this problem. (f) Computing the Fourier amplitude cross-correlations reveals that the highest correlation belongs to the orientation that maximized the Fourier amplitudes in panel d. Unfortunately orientations determined this way may not be entirely accurate as they tend to be biased by the missing wedge. However, if the cross-correlations are normalized with respect to each other (i.e., by setting the mean = 0, and standard deviation = 1), the sharp short peak in the pre-normalized data (orange) is transformed to the tallest peak in the post-normalized data. As a result of this normalization, the true orientation is chosen correctly (g), even though the overall Fourier amplitudes are less than those for other false orientations (h). This process is advantageous because it is less sensitive to the presence of the missing wedge which can obscure accurate orientation determination.

address the fact that the presence of missing data in the Fourier space can hinder proper alignment of the sub-volumes,<sup>27,28</sup> procedures have been developed to circumvent this problem by normalizing the cross-correlations calculated for two particles at different orientations (Figure 17(f)).<sup>28</sup> Although this approach has been met with success, there is no guarantee that the missing wedge problem has been solved, especially for high-resolution tomographic reconstructions.

Alignment of tomographic data requires the three Euler angles ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) for every computationally extracted particle to be determined through a series of orientation searches. This process is typically initialized through comparison with an approximate model of the virus. However, in the event that no adequate reference model is available, it is necessary to generate an initial model from the available data. The raw data can be used to generate an adequate initial model by

comparing subtomograms to each other in a process known as ‘all-vs-all’ comparison.<sup>76</sup> This initial model can be further improved through a series of iterative refinements in which the individual subtomograms are classified, aligned, and averaged to a single or multiple 3-D models. Furthermore, as in single particle cryo-EM, particles can be aligned with respect to their asymmetric features, making it possible to generate single particle cryo-ET maps without imposing symmetry in the map. Nevertheless, again as in single particle cryo-EM, the assumption of symmetry during reconstruction dramatically enhances map resolution.<sup>16</sup> To generate a 3-D model from cryo-ET data, once the orientation search is complete and the particles have been aligned, the data can be reconstructed into a 3-D model by averaging the individual subtomograms together while accounting for the missing wedge information from each particle.<sup>28</sup>

### 1.16.9 Data Interpretation

The final step in the structural determination of a virus is interpretation of the resulting 3-D density map, specifically resolution determination, segmentation, and model fitting. However, care must be taken not to overinterpret the map beyond its best approximated resolution: 20 Å for gross structure, 12 Å for individual protein domains, 9 Å for long and smooth  $\alpha$  helices and large  $\beta$  sheets, 4.7 Å for bumpy  $\alpha$  helices and possibly  $\beta$  strands, less than 4.5 Å to possibly determine a C $\alpha$  backbone trace and bulky side chain features, and less than 3.6 Å to resolve ambiguity in  $\beta$  strand and loop connectivity.<sup>77–79</sup>

#### 1.16.9.1 Resolution Determination

The most commonly accepted method for resolution determination is the one-half Fourier shell correlation (0.5 FSC) criterion;<sup>80,81</sup> however, other criteria of 0.143 or 0.3 have also been proposed for resolution definition.<sup>72,82,83</sup> Because there is no standard practice for reporting map resolution, care must be taken in the comparison of any two maps, and it may require that the reported resolutions be recalibrated prior to doing so.

An FSC curve is generated by calculating the correlations between two density maps generated from two halves of a single data set (eqn [2]), where  $s$  is the radius of the shells used to evaluate the FSC,  $\Delta s$  is the width of these shells.  $F_1$  is the complex structure factor of the first map, and  $F_2^*$  is the complex conjugate of the structure factor of the second map. For example, if one has 10 000 single particle images, 5000 of these are randomly selected and used to generate one map (Map<sub>1</sub>), whereas the other 5000 are used to generate a second map (Map<sub>2</sub>). When building these two subsets of data, the random selection should be made without replacement to ensure that there is no overlap between the two data sets:

$$\text{FSC}(s, \Delta s) = \frac{\text{Re}\{\sum_{|s, \Delta s|} F_1 F_2^*(s)\}}{\{\sum_{|s, \Delta s|} |F_1(s)|^2 \sum_{|s, \Delta s|} |F_2(s)|^2\}^{1/2}} \quad [2]$$

Note that the resolution of a map is a relative measure, and the primary objective of map assembly is the resulting segmentation and model fitting that follows. In light of a standardized method for reconstruction resolution determination, it has been proposed that the effective ‘resolution,’ as predicted by these approaches, should also take into consideration the resolvability of specific features such as  $\alpha$  helices and  $\beta$  sheets, especially those confirmed by crystallography.<sup>9</sup> Although this method of feature based resolution determination is less quantifiable than other resolution assessments, it represents an alternative, more tangible, interpretation of resolution determination as a function of at what resolution the map can accurately describe structural features. To that end, it is possible that two maps with the same resolution, as determined by the FSC criterion, may have different structural features resolved. Furthermore, although calculating an FSC threshold is a common procedure, it is highly sensitive to any post-processing performed on the map and as such, resolution enhancement via the FSC criterion can occasionally become slightly subjective.

#### 1.16.9.2 Segmentation

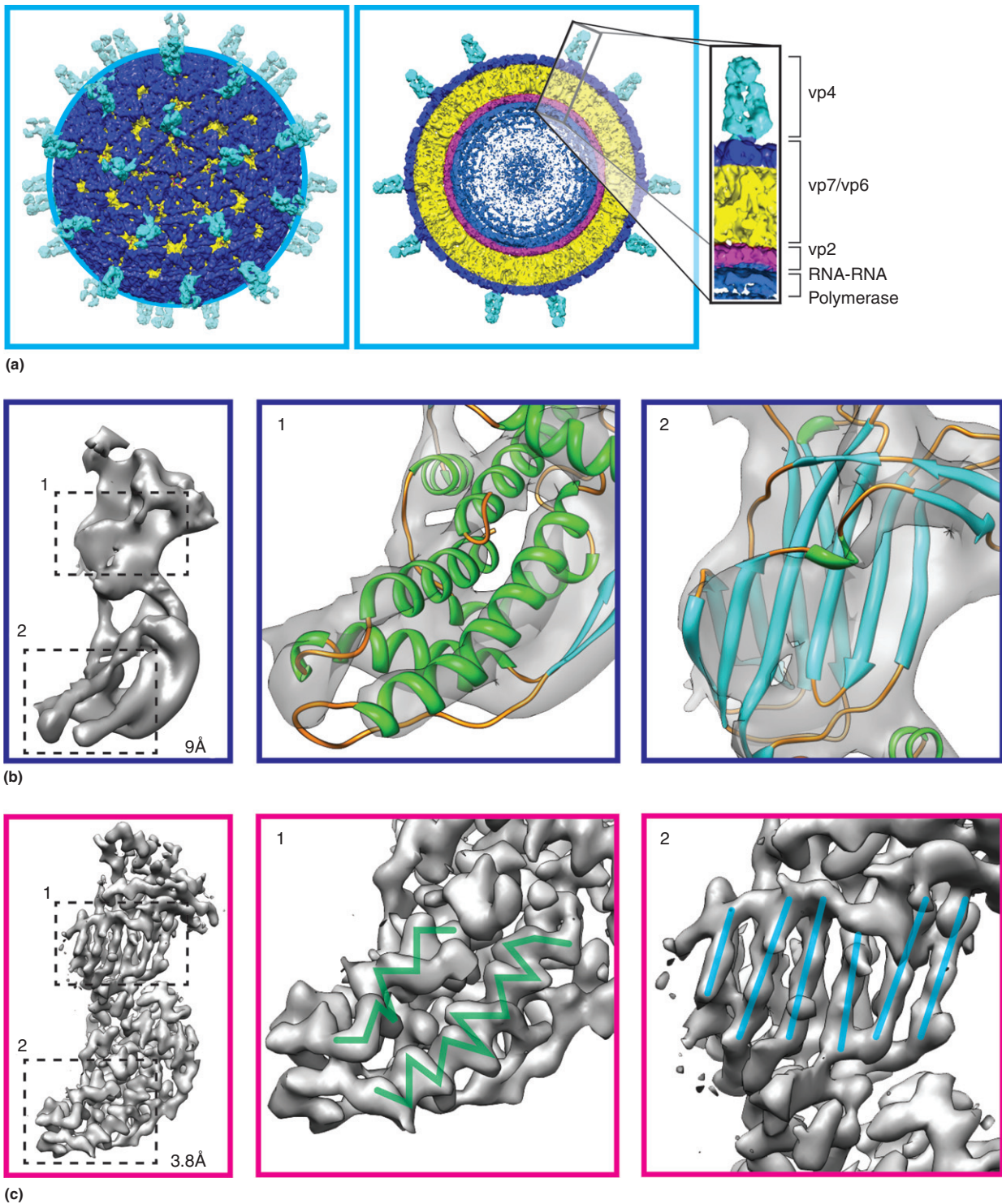
One of the key components in data interpretation is segmentation of a map. Because most complex macromolecules can be broken down into smaller self-assembled protein complexes (a segment), differentiating these pieces from the whole is part of understanding the viral architecture and assembly process. For viruses, when symmetry is assumed in the reconstruction, it is possible to extract just the asymmetric unit and segment it accordingly;<sup>29,48</sup> however, as more viruses are reconstructed without imposed symmetry, it will become necessary to consider the map as a whole.<sup>6,13–15</sup> Although a map can be segmented manually by visual inspection of the density map, simultaneous visualization of the map with fit crystallographic homologs eases the process. Traditionally, segmentation is performed manually,<sup>75,84,85</sup> but tools have been developed to automate certain aspects of the segmentation process. These programs use a variety of approaches, such as watershed<sup>86</sup> and principal component analysis<sup>87</sup> or multi-scale segmentation,<sup>88</sup> to identify individual protein components in the virus. Although these algorithms save a great deal of time during segmentation, their accuracy depends on the resolution and the overlapping density between adjacent molecules.

Regardless of what software is used, accurate segmentation remains a challenging task. For example, a 9-Å resolution map of the bacteriophage  $\epsilon 15$ <sup>14</sup> did not reveal the presence of two separate coat proteins in the capsid. However, when a 4.5-Å map of  $\epsilon 15$  was obtained, a second coat protein was discovered after the C $\alpha$  backbone trace of that protein revealed that it was in fact two distinct proteins.<sup>4</sup>

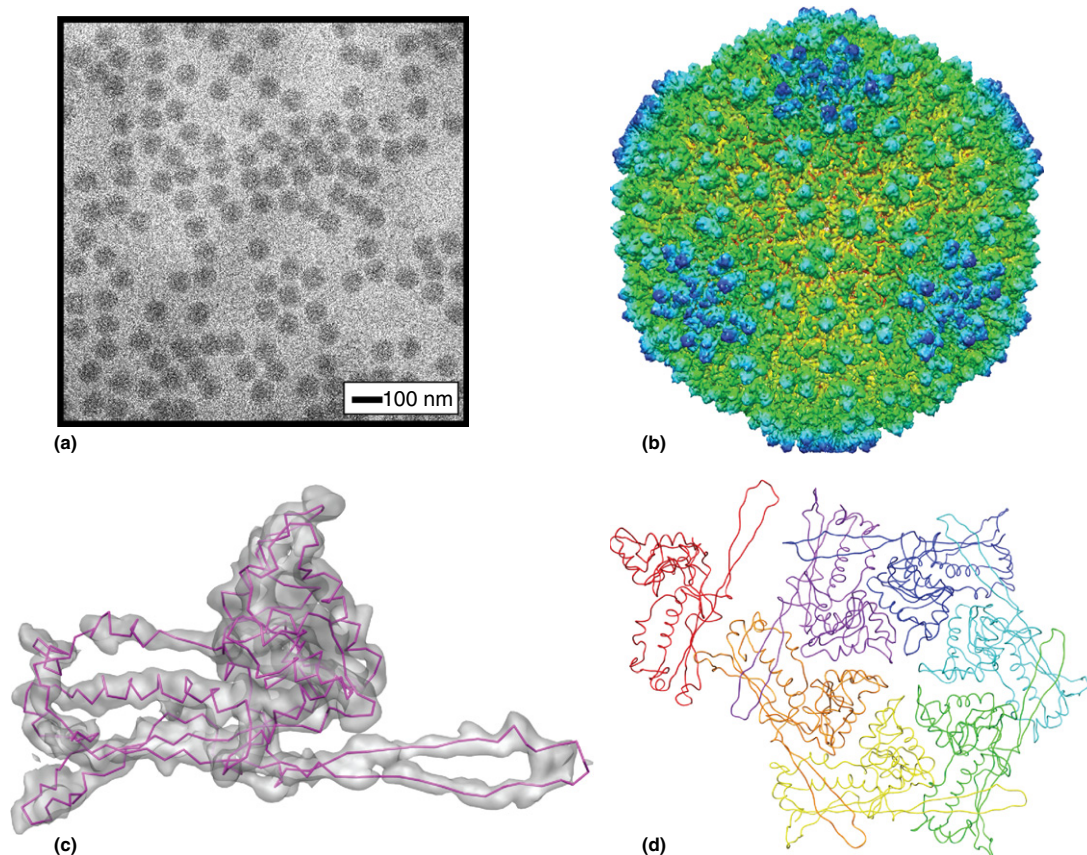
#### 1.16.9.3 Model Fitting

At a resolution of 7–9 Å, it is possible to visually identify long  $\alpha$  helices and  $\beta$  sheets that appear as rodlike structures and flat broad densities, respectively (Figure 18(b)).<sup>29</sup> When the resolution of the map is less than 4.5 Å, it becomes possible to





**Figure 18** Visible features as a function of map resolution. (a) 9.5-Å resolution reconstruction of human rotavirus.<sup>103</sup> Cross section (cyan box) of the virus capsid shows the intricate layering of the viral capsid proteins. (b) At 9 Å, it is possible to see large  $\alpha$  helices separated from each other in the structure of vp6; however, at this resolution,  $\beta$  sheets are not separated and appear as large, flat surfaces in the reconstruction. The crystal structure of vp6 [PDB ID 1QHD] fit to this density shows the locations of the  $\alpha$  helices (1) and  $\beta$  sheets (2). (c) At 3.8 Å, very small  $\alpha$  helices are resolved (1), and it becomes possible to visualize the pitch of these helices in addition to the individual  $\beta$  strands that comprise a  $\beta$  sheet (2) [EMID-1461]. The crystal structure of vp6 has not been fit to this density. To ease the visualization of the pitch of the  $\alpha$  helices and  $\beta$ -strand separation, green and blue lines have been overlaid on the density.



**Figure 19** 4.5-Å reconstruction of  $\epsilon 15$  phage. (a) Electron micrograph of ice-embedded  $\epsilon 15$  phage collected on a 300-kV microscope at 101 K. (b) Radially colored 4.5-Å single particle icosahedral reconstruction of  $\epsilon 15$  phage [EMID-5003]. (c) Subunit averaging of the capsid protein gp7 (gray density) further enhances the resolution of this protein, making it possible to visually identify bulky side chains. (d) The backbone trace of the cryo-EM-derived gp7 density can be used to generate an atomic model of the asymmetric unit of  $\epsilon 15$  phage (each of the seven subunits is uniquely colored).

identify short  $\alpha$  helices and visualize  $\beta$  strand separation (Figure 18(c)). In the past few years, a few viruses have been reconstructed to near-atomic resolution ( $< 4 \text{ \AA}$ ),<sup>4–11</sup> at which level it is possible to resolve individual side chains (Figure 19(c)). When reconstructions reach these ‘ultra-high’ resolutions, it is possible to start building accurate atomic models of massive macromolecules (Figure 19(d)).

Whereas it is easy to visually identify  $\alpha$  helices and  $\beta$  sheets in a subnanometer resolution map, the process of manually fitting structures to these densities is obtrusively difficult but possible.<sup>29</sup> A variety of programs have been developed to quantify the density maps through pattern recognition algorithms (Helixhunter,<sup>89</sup> Foldhunter,<sup>89</sup> SSEhunter,<sup>77</sup> and Gorgon<sup>90</sup>) and have been successfully implemented to identify secondary structure elements (SSEs) in 3-D volume data. By identifying the location, length, and relative positions of  $\alpha$  helices and  $\beta$  sheets in a map, it is possible to search these parameters across structures in the Protein Data Bank (PDB) and identify potential homologs.<sup>77</sup> When the map resolution approaches 4.5 Å and beyond, it becomes feasible to trace the backbone of individual protein subunits either *de novo* or through homology modeling.<sup>30</sup> Furthermore, the ability to identify SSEs within a 3-D volume suggests that these

programs may eventually provide a means to determine feature-based assessment of the resolution of a map (Figure 18).

#### 1.16.9.4 Visualization

For visualization of a map, many tools are available, both free/open source (Chimera,<sup>91</sup> Coot,<sup>92</sup> and Pymol (<http://www.pymol.org>)) and commercially licensed (Amira (<http://www.amiravis.com>)). Although Coot and Pymol have good visualization tools for atomic resolution density maps of moderate-size proteins, their usefulness is limited for larger whole-virus maps composed of hundreds to thousands of proteins. Therefore, Chimera and Amira are typically used to visualize these maps because they can handle large amounts of data more efficiently. Amira is a complete package for visualization and segmentation, but it has a few limitations in terms of cost and a relatively steep learning curve. Alternatively, Chimera is freely available and a good tool for simultaneous visualization of both density maps and PDB models.<sup>4,6,14</sup> Both these programs can be used to make publication-worthy animations of specific features of a map. However, high-end computer animation and visual effects



software such as Maya (<http://usa.autodesk.com>) can be used to enhance the quality of these animations.

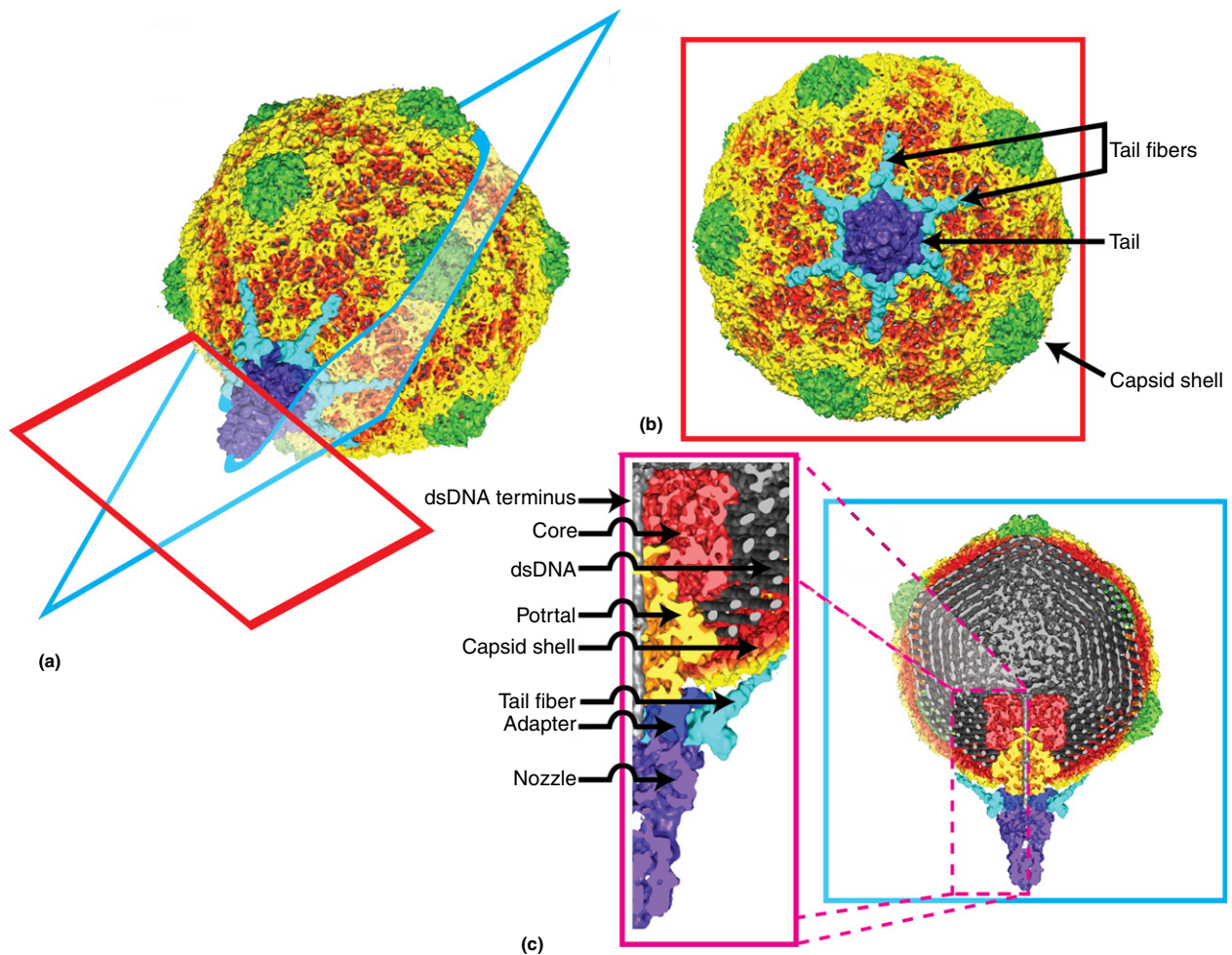
Regardless of the size of the map with which one is working, either a cropped region containing the asymmetric

unit or the whole map, visualization and segmentation should be performed on a multicore multiprocessor computer with large amounts of both physical and video memory. Even on a dual-core/quad-processor 2.67-GHz computer with 16 GB of

**Table 1** Icosahedral virus reconstructions with resolutions  $<4.5 \text{ \AA}$  deposited in the Electron Microscopy Data Bank

EMIB	Fit PDB ID	Resolution ( $\text{\AA}$ )	Sample
1461	3GZU	3.8	Rotavirus VP6 protein
1508	3CNF	3.9	Cytoplasmic polyhedrosis virus
1571	3GZT	3.8	VP7 reconstituted rotavirus DLP
1713	2XD8	4.6	Cyanophage P-SSP7
5003	3C5B	4.5	Bacteriophage $\epsilon$ 15
5156	3IYJ	3.6	Bovine papillomavirus type 1 (BPV1), full virion
5160	3IYL	3.3	Aquareovirus
5172	3IYN	3.6	Human adenovirus type 5
1824	2XYY	4.0	P22 procapsid
1826	2XYZ	3.8	P22 mature virion

Electron Microscopy Data Bank: <http://www.ebi.ac.uk/pdbe/emdb>.



**Figure 20** 3-D reconstruction of cyanophage P-SSP7 without imposed symmetry [EMID-1715]. (a) Oblique view of segmented P-SSP7 showing two viewing planes (red, top view; cyan, coronal slice). (b) Top view showing the arrangement of the tail-fibers with respect to the capsid shell. (c) Coronal slice of the P-SSP7 capsid showing the individual subunits of the virus, namely those responsible for dsDNA genome packaging and release (nozzle, adapter, portal, core, and dsDNA terminus). Adapted from Liu, X.; Zhang, Q.; Murata, K.; Baker, M. L.; Sullivan, M. B.; Fu, C.; Dougherty, M. T.; Schmid, M. F.; Osburne, M. S.; Chisholm, S. W.; Chiu, W. Structural changes in a marine podovirus associated with release of its genome into *Prochlorococcus*. *Nat. Struct. Mol. Biol.* **2010**, *17*, 830–836. Copyright by Nature.

physical memory and 900 MB of video memory, visualization of a whole virus map can cause the computer to freeze due to the massive amount of data being processed by the GPU. For this reason, it is important to consider the physical limitations of one's desktop workstation and, if need be, downsample one's data to a level where it can be processed smoothly.

### 1.16.10 Biological Examples

Although hundreds of tools are available for scientific research, cryo-EM has carved out a niche in the field. The pipeline involved in cryo-EM is the confluence of many fields, from biochemistry to physics and computer science. With regard to virus research, cryo-EM has provided answers to questions that were well beyond the scope of other structural techniques. From solving viral portal structures to elucidating the mechanism by which structured virus are formed, cryo-EM and cryo-ET can provide structural details of viruses in both a biophysical and a structural context.<sup>93</sup>

#### 1.16.10.1 Icosahedral Reconstructions

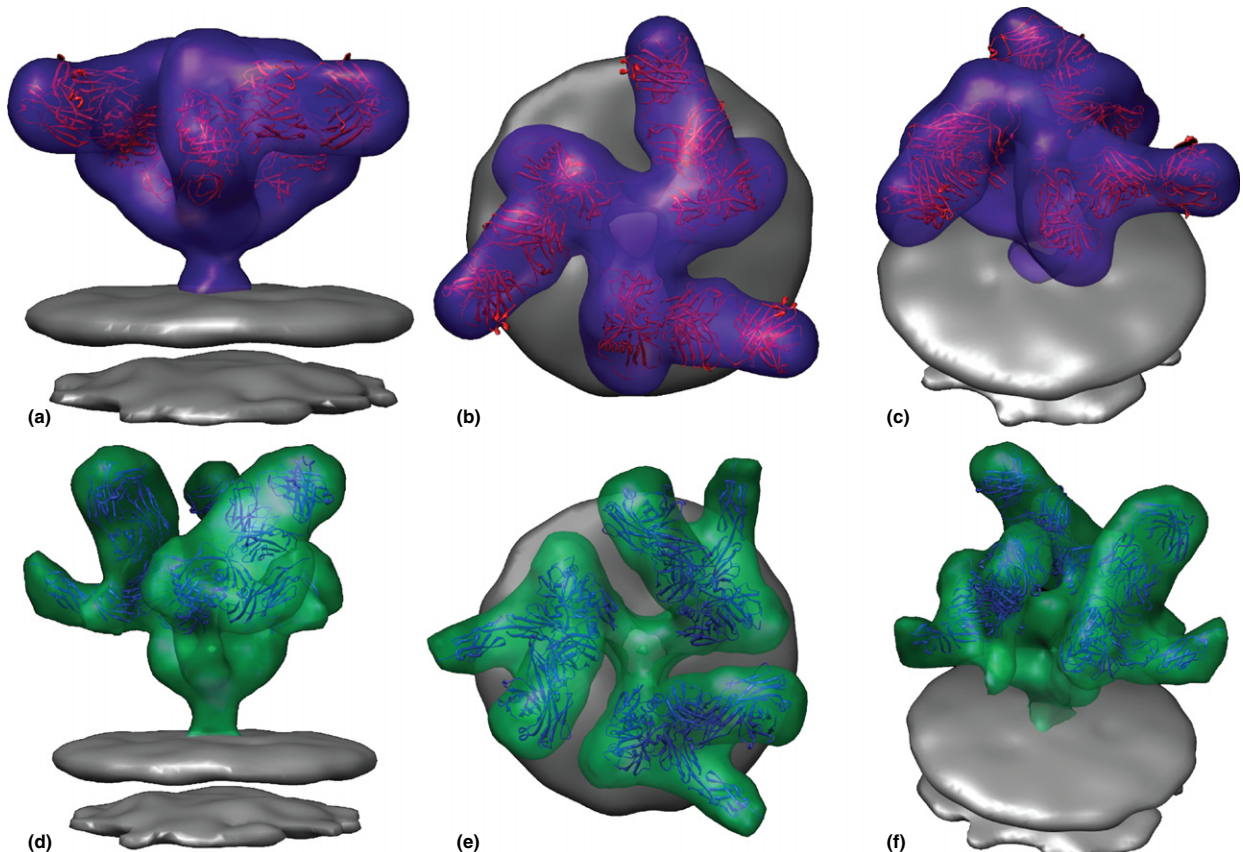
The icosahedral structures of several virus particles have been determined to better than 4.5 Å resolution (Table 1). These

resolutions have been further improved by averaging individual subunits from within a single asymmetric unit together.<sup>10</sup> Although it is reasonable to assume that the multiple similar subunits within a single asymmetric unit are structurally identical, backbone tracing of these subunits has shown this to be false.<sup>4,5</sup>

A 3.3-Å resolution reconstruction of aquareovirus is proof that cryo-EM is capable of producing 3-D density maps from which accurate full atomic models of large macromolecular complexes in their native state can be generated.<sup>10</sup> As the quality of cryo-EM maps improves to near X-ray crystallographic resolutions, the ability to unambiguously trace the C $\alpha$  backbone and identify the locations of individual amino acid residues will unravel many of the mysteries surrounding the structural implication of virus assembly and infection. Furthermore, the ability to identify both large and small side chains, in the context of the entire virus, opens the door to structural analysis of genetically modified viruses.

#### 1.16.10.2 Reconstructions without Imposing Symmetry

Although icosahedral symmetry is often assumed during virus reconstruction, many proteins in the capsid shell are not icosahedrally arranged. Perhaps the most important of these non-icosahedrally arranged proteins, for the viral life cycle, is the genome packaging apparatus (portal) located at one of the

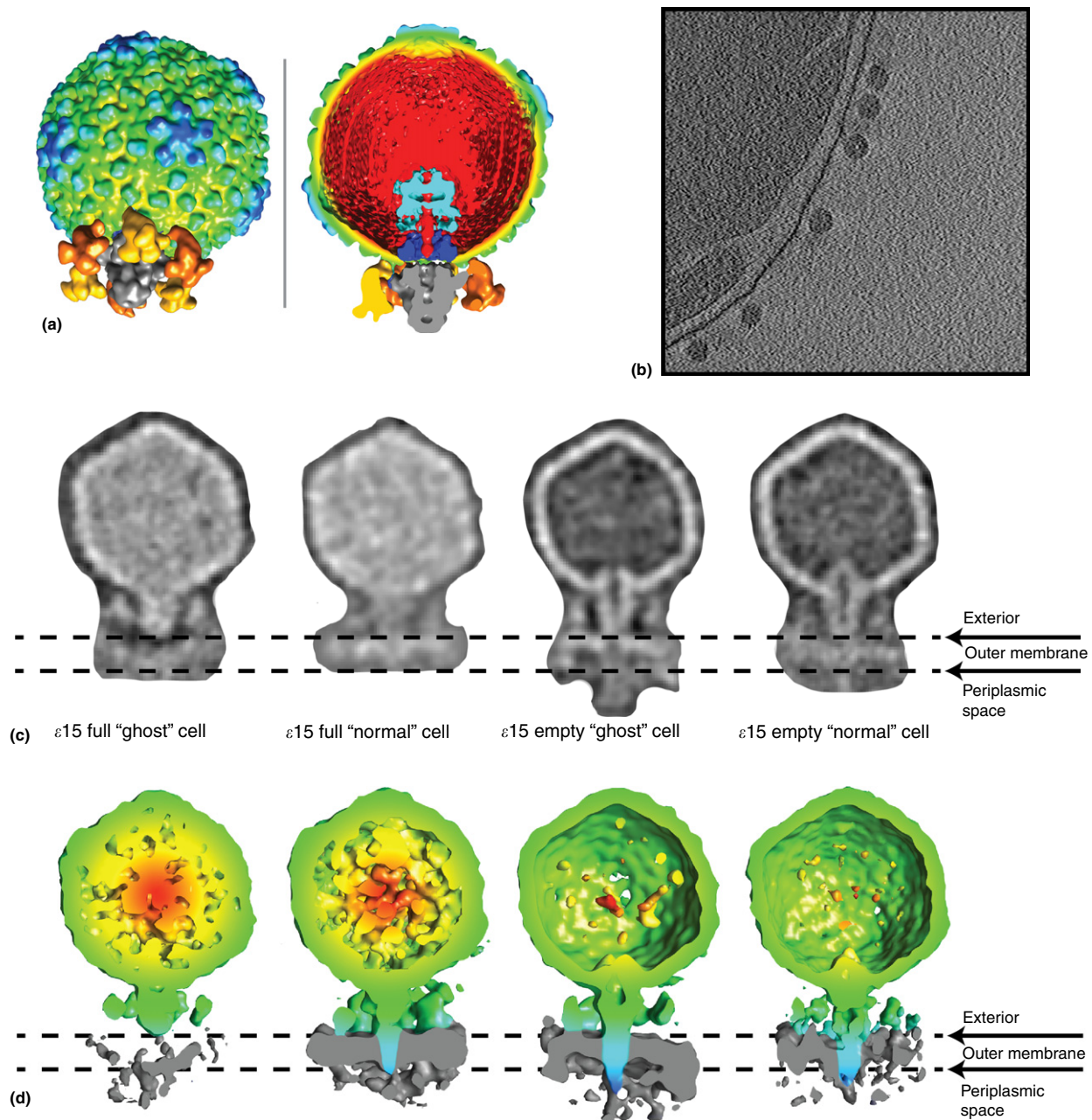


**Figure 21** Structure of the HIV gp120 surface glycoprotein in different conformational states as determined from cryo-ET. (a)–(c) The structure of the HIV surface glycoprotein gp120 in complex with b12-Fab (red, PDB ID, 2NY7), EMID-5018 (blue) and EMID-5022 (gray). (d)–(f) The structure of the HIV surface glycoprotein gp120 in complex with CD4 and 17b-Fab (blue, PDB ID, 1GC1), EMID-5020 (green) and EMID-5022 (gray). Adapted from Liu, J.; Bartesaghi, A.; Borgnia, M.; Sapiro, G.; Subramaniam, S. Molecular architecture of native HIV-1 gp120 trimers. *Nature* **2008**, *455*, 109–113. Copyright by Nature.



12 fivefold vertices in dsDNA viruses (Figure 20).<sup>6,14</sup> During icosahedral reconstruction, asymmetric features such as the portal are averaged out across the entirety of the map. To adequately resolve these features, it is necessary to implement a reconstruction scheme that does not assume symmetry. As a result of this asymmetric approach, reconstructions require

approximately 60 times as many particles to achieve an equivalent icosahedral resolution. A comparison of the structures of viral particles both with and without DNA has elucidated one mechanism of DNA release.<sup>6</sup> In the case of P-SSP7, a valve in the nozzle protein (through which the viral DNA travels) opens and triggers this release. This event is



**Figure 22** Cryo-ET of live infected cells. (a) Symmetry-free model of the dsDNA filled  $\epsilon 15$  bacteriophage, reconstructed to 20 Å from single particle images [EMID-1175]. (b) Slice through the reconstructed tomogram showing  $\epsilon 15$  phages attached to the surface of the bacterium *Salmonella anatum*. These phages can be extracted from the reconstructed tomogram and averaged together to generate symmetry-free models of the different stages of  $\epsilon 15$  infection (c). (c and d) Subtomogram averaging of single particle  $\epsilon 15$  particles in different conformational states reveals the mechanism of infection. The four structures shown in panel c are 2-D projections of the four 3-D densities (d), which are of full and empty  $\epsilon 15$  phages in the presence of physiologically 'normal' and 'ghost' (cell membrane only) cells [EMID-5216, EMID-5217, EMID-5218, EMID-5219]. The tube, which extends across the outer membrane of *Salmonella*, is clearly visible in the 2-D projections (c) and segmented accordingly in the 3-D models (d). Adapted from Chang, J. T.; Schmid, M. F.; Haase-Pettingell, C.; Weigele, P. R.; King, J. A.; Chiu, W. Visualizing the structural changes of bacteriophage epsilon15 and its *Salmonella* host during infection. *J. Mol. Biol.* **2010**, *402*, 731–740.

synchronized with a conformational change in the tail spikes, in addition to the loss of core proteins (Figure 20(c)). To date, asymmetric reconstructions have yet to reach the resolutions seen when icosahedral symmetry is assumed ( $<4 \text{ \AA}$ ). Nevertheless, the potential to extend these reconstructions to higher resolutions remains feasible as larger data sets become available and reconstruction algorithms continue to mature.

### 1.16.10.3 Tomographic Reconstruction

One prerequisite for high-resolution single particle reconstruction is structural homogeneity within a sample. Although subnanometer resolution reconstructions have become the rule, not the exception, for single particle cryo-EM, its use for studying pleomorphic particles such as HIV, influenza, and severe acute respiratory syndrome is limited. Because HIV is an enveloped virus, its outer surface consists of a variety of glycoproteins and so there is no guarantee that the individual particles are the same shape or size, even though they appear to be so in cryo-EM. As a result of this size heterogeneity and the possibility for variation in the number of glycoproteins in the envelope, cryo-ET is the only feasible approach for studying these particles. As discussed previously, the use of cryo-ET limits the maximum achievable resolution for a reconstruction. However, a few studies have pushed this resolution limit to  $20 \text{ \AA}$  and below for a few structural elements of HIV.<sup>26,33</sup> To address this issue, the process of CTF correction, often employed during single particle cryo-EM, has been used on tomographic data to resolve the Gag lattice of HIV to approximately  $17 \text{ \AA}$  (Figure 21).<sup>26</sup> Cryo-ET has also been used to resolve the individual gp120 trimers on the surface of the HIV virus.<sup>33</sup>

### 1.16.11 Future Prospects

As described in this chapter, the structural features of most viruses are lost in the noisy images recorded from the electron microscope. Fortunately, after extensive data processing and reconstruction, it becomes possible to resolve these features – in some instances at near-atomic resolutions. Although most of this work has been augmented by image processing techniques that computationally enhance image contrast, advances in fabrication techniques have enabled the microscopist to do so directly by altering the optics of the electron microscope. One such approach, conceived more than a half century ago, works by introducing a  $\pi/2$  phase shift in the scattered electrons with respect to the unscattered electrons.<sup>94</sup> Although not possible in electron microscopy with materials at that time, this is now accomplished by introducing a thin carbon-film phase plate in the back focal plane of the objective lens. This technique, also known as Zernike phase contrast electron microscopy, is technically challenging due to the difficulty in implementing and maintaining such a device, but the possibilities of its application are tremendous. Although many devices have been proposed for this application,<sup>55,95,96</sup> thin carbon-film phase plates have already begun to show promising results.<sup>16,97,98</sup> As a result of the high contrast in these images, it has been shown that one can reduce the data

required for a subnanometer reconstruction by a factor of 3 compared with conventional imaging.<sup>16</sup> This approach is particularly advantageous for cryo-ET because a reconstruction from a single tomogram can provide as much detail as reconstruction from approximately 17 000 single particle images of  $\epsilon 15$  phage.<sup>16</sup> For the purpose of studying virus particles whose structural components have proven intractable to analysis given current methodologies, this new technology is likely to play a key role in addressing questions plagued by structural heterogeneity.

Although structural studies of viruses have traditionally been confined to biochemically purified particles, visualizing these viruses infecting their host cells is now possible by cryo-ET.<sup>99</sup> Studies such as that by Chang et al.<sup>99</sup> open the door for investigating the structural rearrangements that occur at the level of the virus throughout the process of infection. The ability to directly visualize a tube spanning the outer membrane of a bacteria (Figure 22) marks the beginning of an era in structural virology in which it is possible to study the interaction of specific virus structures with those of an entire cell.

Cryo-EM and cryo-ET are rapidly evolving fields, not only with regard to the questions that one can address with the two technologies but also with regard to their maturation toward becoming a routine scientific tool. During approximately the past 5 years, exciting advances have occurred with regard to how electron microscopy data are produced and collected, in addition to how these data are processed and reconstructed into high-resolution models. For icosahedral reconstructions, the resolutions of these models are quickly approaching that of X-ray crystallography, making it possible to identify primary structural features, firsthand, from within large macromolecular complexes in near-native environments.<sup>5,11,100–102</sup> Furthermore, advances in data processing algorithms have allowed reconstructions, when icosahedral symmetry is not assumed, to finally reach subnanometer resolutions. Although the resolutions of cryo-ET reconstructions have yet to surpass the subnanometer threshold, the breadth of questions that can be addressed with this technology far surpasses that of the single particle approach. Because cryo-ET provides a means to discern heterogeneity at the virus-cell level, it is possible to process the data accordingly. Although these advances appear to have pushed cryo-EM up against the seemingly insurmountable resolution limit set by crystallography, the fact that they have done so in the span of just a few years gives promise for future advancements in the field. Looking forward, considering how the field has evolved in such a short time, both cryo-EM and cryo-ET are likely to have many more surprises in store, not just in technology advancements but also in the scientific discoveries that will result from their application.

### Acknowledgments

This work was supported by grants R01AI0175208 and P41RR002250 from the National Institutes of Health (NIH) and grant Q1242 from the Robert Welch Foundation. RHR is supported by NIH training grants (CM07330 through the



Medical Scientist Training Program and T15LM007093 through the Gulf Coast Consortia).

## References

- [1] Crowther, R.; Amos, L.; Finch, J.; De Rosier, D.; Klug, A. Three dimensional reconstructions of spherical viruses by Fourier synthesis from electron micrographs. *Nature* **1970**, *226*, 421–425.
- [2] DeRosier, D. A.; Klug, A. Reconstruction of three-dimensional structures from electron micrographs. *Nature* **1968**, *217*, 130–134.
- [3] Chen, D. H.; Baker, M. L.; Hryc, C. F.; Dimairo, F.; Jakana, J.; Wu, W.; Dougherty, M.; Haase-Pettingell, C.; Schmid, M. F.; Jiang, W.; Baker, D.; King, J. A.; Chiu, W. Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1355–1360.
- [4] Jiang, W.; Baker, M.; Jakana, J.; Weigele, P.; King, J.; Chiu, W. Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. *Nature* **2008**, *451*, 1130–1134.
- [5] Liu, H.; Jin, L.; Koh, S. B. S.; Atanasov, I.; Schein, S.; Wu, L.; Zhou, Z. H. Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. *Science* **2010**, *329*, 1038–1043.
- [6] Liu, X.; Zhang, Q.; Murata, K.; Baker, M. L.; Sullivan, M. B.; Fu, C.; Dougherty, M. T.; Schmid, M. F.; Osburne, M. S.; Chisholm, S. W.; Chiu, W. Structural changes in a marine podovirus associated with release of its genome into *Prochlorococcus*. *Nat. Struct. Mol. Biol.* **2010**, *17*, 830–836.
- [7] Sachse, C.; Chen, J.; Coureux, P.-D.; Stroupe, M.; Fändrich, M.; Grigorieff, N. High-resolution electron microscopy of helical specimens: A fresh look at tobacco mosaic virus. *J. Mol. Biol.* **2007**, *371*, 812–835.
- [8] Wolf, M.; Garcea, R. L.; Grigorieff, N.; Harrison, S. C. Subunit interactions in bovine papillomavirus. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 6298–6303.
- [9] Yu, X.; Jin, L.; Zhou, Z. 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature* **2008**, *453*, 415–419.
- [10] Zhang, X.; Jin, L.; Fang, Q.; Hui, W. H.; Zhou, Z. H. 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell* **2010**, *141*, 472–482.
- [11] Zhang, X.; Settembre, E.; Xu, C.; Dormitzer, P.; Bellamy, R.; Harrison, S.; Grigorieff, N. Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1867–1872.
- [12] Agirrezabala, X.; Martín-Benito, J.; Castón, J.; Miranda, R.; Valpuesta, J.; Carrascosa, J. Maturation of phage T7 involves structural modification of both shell and inner core components. *EMBO J.* **2005**, *24*, 3820–3829.
- [13] Chang, J.; Weigele, P.; King, J.; Chiu, W.; Jiang, W. Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. *Structure* **2006**, *14*, 1073–1082.
- [14] Jiang, W.; Chang, J.; Jakana, J.; Weigele, P.; King, J.; Chiu, W. Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature* **2006**, *439*, 612–616.
- [15] Lander, G. C.; Tang, L.; Casjens, S. R.; Gilcrease, E. B.; Prevelige, P.; Poliakov, A.; Potter, C. S.; Carragher, B.; Johnson, J. E. The structure of an infectious P22 virion shows the signal for headful DNA packaging. *Science* **2006**, *312*, 1791–1795.
- [16] Murata, K.; Liu, X.; Danev, R.; Jakana, J.; Schmid, M. F.; King, J.; Nagayama, K.; Chiu, W. Zernike phase contrast cryo-electron microscopy and tomography for structure determination at nanometer and subnanometer resolutions. *Structure* **2010**, *18*, 903–912.
- [17] Tang, J.; Olson, N.; Jardine, P. J.; Grimes, S.; Anderson, D. L.; Baker, T. S. DNA poised for release in bacteriophage phi29. *Structure* **2008**, *16*, 935–943.
- [18] Baker, T. S.; Olson, N. H.; Fuller, S. D. Adding the third dimension to virus life cycles: Three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 862–922.
- [19] Jiang, W.; Chiu, W. Cryoelectron microscopy of icosahedral virus particles. *Methods Mol. Biol.* **2007**, *369*, 345–363.
- [20] Thuman-Commike, P.; Chiu, W. Reconstruction principles of icosahedral virus structure determination using electron cryomicroscopy. *Micron* **2000**, *31*, 687–711.
- [21] Zhou, Z.; Chiu, W. Determination of icosahedral virus structures by electron cryomicroscopy at subnanometer resolution. *Adv. Protein Chem* **2003**, *64*, 93–124.
- [22] Finch, J. T.; Klug, A. Three-dimensional reconstruction of the stacked-disk aggregate of tobacco mosaic virus protein from electron micrographs. *Philos. Trans. R. Soc. London B Biol. Sci* **1971**, *261*, 211–219.
- [23] Baumeister, W.; Grimm, R.; Walz, J. Electron tomography of molecules and cells. *Trends Cell Biol.* **1999**, *9*, 81–85.
- [24] Frank, J. *Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell*, 2nd edn.; Springer: New York, 2006.
- [25] Förster, F.; Medalia, O.; Zauberman, N.; Baumeister, W.; Fass, D. Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 4729–4734.
- [26] Briggs, J.; Riches, J.; Glass, B.; Bartonova, V.; Zanetti, G.; Kräusslich, H.-G. Structure and assembly of immature HIV. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11090–11095.
- [27] Chang, J.; Schmid, M.; Rixon, F.; Chiu, W. Electron cryotomography reveals the portal in the herpesvirus capsid. *J. Virol.* **2007**, *81*, 2065–2068.
- [28] Schmid, M.; Booth, C. Methods for aligning and for averaging 3D volumes with missing data. *J. Struct. Biol.* **2008**, *161*, 243–248.
- [29] Zhou, Z.; Dougherty, M.; Jakana, J.; He, J.; Rixon, F.; Chiu, W. Seeing the herpesvirus capsid at 8.5 Å. *Science* **2000**, *288*, 877.
- [30] Baker, L. A.; Smith, E. A.; Bueler, S. A.; Rubinstein, J. L. The resolution dependence of optimal exposures in liquid nitrogen temperature electron cryomicroscopy of catalase crystals. *J. Struct. Biol.* **2010**, *169*, 431–437.
- [31] Bamms, B. E.; Jakana, J.; Schmid, M. F.; Chiu, W. Radiation damage effects at four specimen temperatures from 4 to 100 K. *J. Struct. Biol.* **2010**, *169*, 331–341.
- [32] McEwen, B. F.; Downing, K. H.; Glaeser, R. M. The relevance of dose-rate in tomography of radiation-sensitive specimens. *Ultramicroscopy* **1995**, *60*, 357–373.
- [33] Liu, J.; Bartesaghi, A.; Borgnia, M.; Sapiro, G.; Subramaniam, S. Molecular architecture of native HIV-1 gp120 trimers. *Nature* **2008**, *455*, 109–113.
- [34] Frank, J. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*, 2nd edn.; Oxford University Press: Oxford, 2006.
- [35] Glaeser, R. M. *Electron Crystallography of Biological Macromolecules*, Oxford University Press: Oxford, 2007.
- [36] Fuller, S. D.; Argos, P. Is Sindbis a simple picornavirus with an envelope? *EMBO J.* **1987**, *6*, 1099–1105.
- [37] Zhou, Z.; Chiu, W.; Haskell, K.; Spears, H.; Jakana, J.; Rixon, F.; Scott, L. Refinement of herpesvirus B-capsid structure on parallel supercomputers. *Biophys. J.* **1998**, *74*, 576–588.
- [38] Jiang, W.; Li, Z.; Zhang, Z.; Baker, M.; Prevelige, P.; Chiu, W. Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nature Struct. Biol.* **2003**, *10*, 131–135.
- [39] Lanio, S. High-resolution imaging magnetic energy filters with simple structure. *OPTIK* **1986**, *73*, 99–107.
- [40] Tsuno, K.; Kaneyama, T.; Honda, T.; Ishida, Y. Non-isochromaticity of an omega filter in a 200 kV transmission electron microscope. *J. Electron Microscop.* **1998**, *47*, 611–619.
- [41] Cong, Y.; Baker, M. L.; Jakana, J.; Woolford, D.; Miller, E. J.; Reissmann, S.; Kumar, R. N.; Redding-Johanson, A. M.; Bath, T. S.; Mukhopadhyay, A.; Ludtke, S. J.; Frydman, J.; Chiu, W. 4.0-Å resolution cryo-EM structure of the mammalian chaperonin TRiC/CCT reveals its unique subunit arrangement. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4967–4972.
- [42] Chen, D.-H.; Jakana, J.; Liu, X.; Schmid, M.; Chiu, W. Achievable resolution from images of biological specimens acquired from a 4 k × 4 k CCD camera in a 300-kV electron cryomicroscope. *J. Struct. Biol.* **2008**, *163*, 45–52.
- [43] Zhang, J.; Baker, M. L.; Schroder, G. F.; Douglas, N. R.; Reissmann, S.; Jakana, J.; Dougherty, M.; Fu, C. J.; Levitt, M.; Ludtke, S. J.; Frydman, J.; Chiu, W. Mechanism of folding chamber closure in a group II chaperonin. *Nature* **2010**, *463*, 379–383.
- [44] Booth, C.; Jakana, J.; Chiu, W. Assessing the capabilities of a 4 k × 4 k CCD camera for electron cryo-microscopy at 300 kV. *J. Struct. Biol.* **2006**, *156*, 556–563.
- [45] Brink, J.; Chiu, W. Applications of a slow-scan CCD camera in protein electron crystallography. *J. Struct. Biol.* **1994**, *113*, 23–34.
- [46] Sherman, M. B.; Brink, J.; Chiu, W. Performance of a slow-scan CCD camera for macromolecular imaging in a 400 kV electron cryomicroscope. *Micron* **1996**, *27*, 129–139.
- [47] Fan, G.; Ellisman, M. Digital imaging in transmission electron microscopy. *J. Microscop.* **2000**, *200*, 1–13.
- [48] Booth, C.; Jiang, W.; Baker, M.; Zhou, Z.; Ludtke, S.; Chiu, W. A 9 angstroms single particle reconstruction from CCD captured images on a 200 kV electron cryomicroscope. *J. Struct. Biol.* **2004**, *147*, 116–127.

- [49] Lander, G. C.; Stagg, S. M.; Voss, N. R.; Cheng, A.; Fellmann, D.; Pulokas, J.; Yoshioka, C.; Irving, C.; Mulder, A.; Lau, P. W.; Lyumkis, D.; Potter, C. S.; Carragher, B. Appion: An integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **2009**, *166*, 95–102.
- [50] Ludtke, S.; Nason, L.; Tu, H.; Peng, L.; Chiu, W. Object oriented database and electronic notebook for transmission electron microscopy. *Microsc. Microanal.* **2003**, *9*, 556–565.
- [51] Ludtke, S.; Baldwin, P.; Chiu, W. EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **1999**, *128*, 82–97.
- [52] Tang, G.; Peng, L.; Baldwin, P.; Mann, D.; Jiang, W.; Rees, I.; Ludtke, S. EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **2007**, *157*, 38–46.
- [53] Brink, J.; Gross, H.; Tittmann, P.; Sherman, M. B.; Chiu, W. Reduction of charging in protein electron cryomicroscopy. *J. Microsc.* **1998**, *191*, 67–73.
- [54] Erickson, H.; Klug, A. Measurement and compensation of de-focusing and aberrations by Fourier processing of electron micrographs. *Philos. Trans. R. Soc. London B* **1971**, *261*, 105–118.
- [55] Thon, F. Phase Contrast Electron Microscopy. In *Electron Microscopy in Material Sciences*; Academic Press: London, 1971; pp. 571–625.
- [56] Saad, A.; Ludtke, S.; Jakana, J.; Rixon, F.; Tsuruta, H.; Chiu, W. Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination. *J. Struct. Biol.* **2001**, *133*, 32–42.
- [57] Yang, K.; Wills, E.; Baines, J. The putative leucine zipper of the UL6-encoded portal protein of herpes simplex virus 1 is necessary for interaction with pUL15 and pUL28 and their association with capsids. *J. Virol.* **2009**, *83*, 4557–4564.
- [58] Liu, X.; Jiang, W.; Jakana, J.; Chiu, W. 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.* **2007**, *160*, 11–27.
- [59] Rossmann, M. G.; Mesyanzhinov, V. V.; Arisaka, F.; Leiman, P. G. The bacteriophage T4 DNA injection machine. *Curr. Opin. Struct. Biol.* **2004**, *14*, 171–180.
- [60] Tao, Y.; Olson, N. H.; Xu, W.; Anderson, D. L.; Rossmann, M. G.; Baker, T. S. Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell* **1998**, *95*, 431–437.
- [61] Frank, J.; Radermacher, M.; Penczek, P.; Zhu, J.; Li, Y.; Ladjadj, M.; Leith, A.; SPIDER, WEB Processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **1996**, *116*, 190–199.
- [62] Marabini, R.; Masegosa, I. M.; San Martin, M. C.; Marco, S.; Fernandez, J. J.; de la Fraga, L. G.; Vaquerizo, C.; Carazo, J. M. Xmipp: An image processing package for electron microscopy. *J. Struct. Biol.* **1996**, *116*, 237–240.
- [63] van Heel, M.; Harauz, G.; Orlova, E.; Schmidt, R.; Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **1996**, *116*, 17–24.
- [64] Stewart, A.; Grigorieff, N. Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy* **2004**, *102*, 67–84.
- [65] Yan, X.; Sinkovits, R.; Baker, T. AUTO3DEM: An automated and high throughput program for image reconstruction of icosahedral particles. *J. Struct. Biol.* **2007**, *157*, 73–82.
- [66] Hohn, M.; Tang, G.; Goodyear, G.; Baldwin, P.; Huang, Z.; Penczek, P.; Yang, C.; Glaeser, R.; Adams, P.; Ludtke, S. SPARX, a new environment for cryo-EM image processing. *J. Struct. Biol.* **2007**, *157*, 47–55.
- [67] Liang, Y.; Ke, E.; Zhou, Z. IMIRS: A high-resolution 3D reconstruction package integrated with a relational image database. *J. Struct. Biol.* **2002**, *137*, 292–304.
- [68] Fuller, S.; Butcher, S.; Cheng, R.; Baker, T. Three-dimensional reconstruction of icosahedral particles: The uncommon line. *J. Struct. Biol.* **1996**, *116*, 48–55.
- [69] Chen, J. Z.; Settembre, E. C.; Aoki, S. T.; Zhang, X.; Bellamy, A. R.; Dormitzer, P. R.; Harrison, S. C.; Grigorieff, N. Molecular interactions in rotavirus assembly and uncoating seen by high-resolution cryo-EM. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10644–10648.
- [70] Parent, K. N.; Khayat, R.; Tu, L. H.; Suhanovsky, M. M.; Cortines, J. R.; Teschke, C. M.; Johnson, J. E.; Baker, T. S. P22 coat protein structures reveal a novel mechanism for capsid maturation: Stability without auxiliary proteins or chemical crosslinks. *Structure* **2010**, *18*, 390–401.
- [71] Glaeser, R. Electron crystallography: Present excitement, a nod to the past, anticipating the future. *J. Struct. Biol.* **1999**, *128*, 3–14.
- [72] Rosenthal, P.; Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **2003**, *333*, 721–745.
- [73] Grigorieff, N. Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 Å in ice. *J. Mol. Biol.* **1998**, *277*, 1033–1046.
- [74] Lucić, V.; Förster, F.; Baumeister, W. Structural studies by electron tomography: From cells to molecules. *Annu. Rev. Biochem.* **2005**, *74*, 833–865.
- [75] Kremer, J.; Mastronarde, D.; McIntosh, J. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **1996**, *116*, 71–76.
- [76] Schmid, M.; Paredes, A.; Khant, H.; Soyer, F.; Aldrich, H.; Chiu, W.; Shively, J. Structure of *Halothiobacillus neapolitanus* carboxysomes by cryo-electron tomography. *J. Mol. Biol.* **2006**, *364*, 526–535.
- [77] Baker, M.; Ju, T.; Chiu, W. Identification of secondary structure elements in intermediate-resolution density maps. *Structure* **2007**, *15*, 7–19.
- [78] Blow, D. *Outline of Crystallography for Biologists*; Oxford University Press: New York, 2002.
- [79] Serysheva, I. I.; Hamilton, S. L.; Chiu, W.; Ludtke, S. J. Structure of  $\text{Ca}^{2+}$  release channel at 14 Å resolution. *J. Mol. Biol.* **2005**, *345*, 427–431.
- [80] Harauz, G.; Van Heel, M. Exact filters for general geometry three dimensional reconstruction. *OPTIK* **1986**, *73*, 146–156.
- [81] Saxton, W.; Baumeister, W. The correlation averaging of a regularly arranged bacterial cell envelope protein. *J. Microsc.* **1982**, *127*, 127–138.
- [82] Liao, H. Y.; Frank, J. Definition and estimation of resolution in single-particle reconstructions. *Structure* **2010**, *18*, 768–775.
- [83] van Heel, M.; Schatz, M. Fourier shell correlation threshold criteria. *J. Struct. Biol.* **2005**, *151*, 250–262.
- [84] Hessler, D.; Young, S.; Ellisman, M. A flexible environment for the visualization of three-dimensional biological structures. *J. Struct. Biol.* **1996**, *116*, 113–119.
- [85] Li, Y.; Leith, A.; Frank, J. Tinkerbell: A tool for interactive segmentation of 3D data. *J. Struct. Biol.* **1997**, *120*, 266–275.
- [86] Volkman, N. A novel three-dimensional variant of the watershed transform for segmentation of electron density maps. *J. Struct. Biol.* **2002**, *138*, 123–129.
- [87] Frangakis, A.; Hegerl, R. Segmentation of two- and three-dimensional data from electron microscopy using eigenvector analysis. *J. Struct. Biol.* **2002**, *138*, 105–113.
- [88] Pintilie, G. D.; Zhang, J.; Goddard, T. D.; Chiu, W.; Gossard, D. C. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **2010**, *170*, 427–438.
- [89] Jiang, W.; Baker, M.; Ludtke, S.; Chiu, W. Bridging the information gap: Computational tools for intermediate resolution structure interpretation. *J. Mol. Biol.* **2001**, *308*, 1033–1044.
- [90] Baker, M. L.; Abeyasinghe, S. S.; Schuh, S.; Coleman, R. A.; Abrams, A.; Marsh, M. P.; Hryc, C. F.; Ruths, T.; Chiu, W.; Ju, T. Modeling protein structure at near atomic resolutions with Gorgon. *J. Struct. Biol.* **2011**, *174*, 360–373.
- [91] Pettersen, E.; Goddard, T.; Huang, C.; Couch, G.; Greenblatt, D.; Meng, E.; Ferrin, T. UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- [92] Emsley, P.; Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126–2132.
- [93] Johnson, J.; Chiu, W. DNA packaging and delivery machines in tailed bacteriophages. *Curr. Opin. Struct. Biol.* **2007**, *17*, 237–243.
- [94] Boersch, H. *Über die Kontraste von Atomen im Elektronenmikroskop*. *Z. J. Phys. Sci.* **1947**, *2*, 615.
- [95] Cambie, R.; Downing, K. H.; Typke, D.; Glaeser, R. M.; Jin, J. Design of a microfabricated, two-electrode phase-contrast element suitable for electron microscopy. *Ultramicroscopy* **2007**, *107*, 329–339.
- [96] Danev, R.; Nagayama, K. Transmission electron microscopy with Zernike phase plate. *Ultramicroscopy* **2001**, *88*, 243–252.
- [97] Danev, R.; Nagayama, K. Single particle analysis based on Zernike phase contrast transmission electron microscopy. *J. Struct. Biol.* **2008**, *161*, 211–218.
- [98] Rochat, R. H.; Liu, X.; Murata, K.; Nagayama, K.; Rixon, F. J.; Chiu, W. Seeing the portal in herpes simplex virus type 1 B capsids. *J. Virol.* **2011**, *85*, 1871–1874.
- [99] Chang, J. T.; Schmid, M. F.; Haase-Pettingell, C.; Weigele, P. R.; King, J. A.; Chiu, W. Visualizing the structural changes of bacteriophage epsilon15 and its *Salmonella* host during infection. *J. Mol. Biol.* **2010**, *402*, 731–740.

- [100] Roberts, M. M.; White, J. L.; Grutter, M. G.; Burnett, R. M. Three-dimensional structure of the adenovirus major coat protein hexon. *Science* **1986**, *232*, 1148–1151.
- [101] Rux, J. J.; Kuser, P. R.; Burnett, R. M. Structural and phylogenetic analysis of adenovirus hexons by use of high-resolution X-ray crystallographic, molecular modeling, and sequence-based methods. *J. Virol.* **2003**, *77*, 9553–9566.
- [102] Zubieta, C.; Schoehn, G.; Chroboczek, J.; Cusack, S. The structure of the human adenovirus 2 penton. *Mol. Cell* **2005**, *17*, 121–135.
- [103] Li, Z.; Baker, M.; Jiang, W.; Estes, M.; Prasad, B. Rotavirus architecture at subnanometer resolution. *J. Virol.* **2009**, *83*, 1754–1766.