ISSN 1674-1056
http://cpb.iphy.ac.cn
http://iopscience.iop.org/cpb
中国物理 B
Chinese Physics
中国物理
Chinese Physics B

Computing methods for icosahedral and symmetry-mismatch reconstruction of viruses by cryo-electron microscopy

Bin Zhu(朱彬), Lingpeng Cheng(程凌鹏), Hongrong Liu(刘红荣)
Citation: Chin. Phys. B . 2018, 27(5): 056802. **doi:** 10.1088/1674-1056/27/5/056802

Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Quantitative HRTEM and its application in the study of oxide materials

Chun-Lin Jia(贾春林), Shao-Bo Mi(米少波), Lei Jin(金磊)
Chin. Phys. B . 2018, 27(5): 056803. **doi:** 10.1088/1674-1056/27/5/056803

High-resolution electron microscopy for heterogeneous catalysis research

Yong Zhu(朱勇), Mingquan Xu(许名权), Wu Zhou(周武)
Chin. Phys. B . 2018, 27(5): 056804. **doi:** 10.1088/1674-1056/27/5/056804

Lie symmetry theorem of fractional nonholonomic systems

Sun Yi, Chen Ben-Yong, Fu Jing-Li
Chin. Phys. B . 2014, 23(11): 0110201. **doi:** 10.1088/1674-1056/23/11/110201

Symmetries and conserved quantities of discrete wave equation associated with the Ablowitz-Ladik-Lattice system

Fu Jing-Li, Song Duan, Fu Hao, He Yu-Fang, Hong Fang-Yu
Chin. Phys. B . 2013, 22(9): 090201. **doi:** 10.1088/1674-1056/22/9/090201

A symmetry-preserving difference scheme for high dimensional nonlinear evolution equations

Xin Xiang-Peng, Chen Yong, Wang Yun-Hu
Chin. Phys. B . 2013, 22(6): 060201. **doi:** 10.1088/1674-1056/22/6/060201

TOPICAL REVIEW — Electron microscopy methods for the emergent materials and life sciences

Computing methods for icosahedral and symmetry-mismatch reconstruction of viruses by cryo-electron microscopy*

Bin Zhu(朱彬)[1], Lingpeng Cheng(程凌鹏)[2,†], and Hongrong Liu(刘红荣)[1,‡]

[1] College of Physics and Information Science, Synergetic Innovation Center for Quantum Effects and Applications and Key Laboratory of Low-dimensional Quantum Structures, Hunan Normal University, Changsha 410081, China
[2] Technology Center for Protein Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China

(Received 21 November 2017; revised manuscript received 27 March 2018; published online 26 April 2018)

Three-dimensional (3D) reconstruction of icosahedral viruses has played a crucial role in the development of cryo-electron microscopy single-particle reconstruction, with many cryo-electron microscopy techniques first established for structural studies of icosahedral viruses, owing to their high symmetry and large mass. This review summarizes the computational methods for icosahedral and symmetry-mismatch reconstruction of viruses, as well as the likely challenges and bottlenecks in virus reconstruction, such as symmetry mismatch reconstruction, contrast transformation function (CTF) correction, and particle distortion.

Keywords: cryo-electron microscopy, icosahedral virus, computational method, three-dimensional reconstruction, symmetry-mismatch reconstruction

PACS: 68.37.Lp, 68.37.Og, 02.20.–a **DOI:** 10.1088/1674-1056/27/5/056802

1. Introduction

With the development of electron microscopy, algorithms, software, computing power, and particularly, the direct electron detector (DED) camera in recent years, single-particle cryo-electron microscopy (cryo-EM) has become one of the most powerful tools in structural biology.[1,2] Icosahedral viruses have played a critical role in the development of cryo-EM; many new cryo-EM techniques were first established for structural studies of viruses. The main reasons behind the relevance of icosahedral viruses for cryo-EM development are (i) their high symmetry, which is equivalent to a 60-fold increase in data set size, and, hence, more effective noise suppression and improved reconstruction resolution; (ii) their large molecular mass, which can produce a strong image contrast allowing for more accurate determination of particle orientation parameters (three Euler angles, θ, ϕ, and ω), center parameters (two shift coordinates, x and y), and three-dimensional (3D) reconstructions; and (iii) their rigidity and uniformity, which enable near-perfect superposition of structural features in 3D reconstructions.[3]

In the 1970 s, through 3D structural computations of the tomato bushy stunt virus, Crowther *et al.* established the reconstruction theories of icosahedral viruses, including the common-line method used to determine the orientation and center parameters of particles, and the Fourier–Bessel synthesis method used to calculate the 3D structures from two-dimensional (2D) images.[4,5] In 1984, the first high-contrast cryo-EM micrograph was obtained using adenovirus embedded in vitreous ice through plunge freezing,[6] which enabled preserving high-resolution information of macromolecular complexes and expedited high-resolution reconstruction. In 1997, two groups from the Medical Research Council Laboratory and National Institutes of Health,[7,8] respectively, resolved the hepatitis B capsid structure to a sub-nanometer resolution, identifying, for the first time, the secondary structure of capsid proteins from a 3D cryo-EM; their results represented a milestone in cryo-EM and an advancement toward the structural biologist's dream of solving the atomic structure of macromolecules without requiring crystallization.[9] In 2008, several icosahedral virus capsids were resolved to near-atomic resolution (approximately 4 Å) by single-particle cryo-EM,[10–13] enabling the development of backbone models based on cryo-EM maps. In 2010, several full atomic models of icosahedral viruses were constructed *denovo* by cryo-EM utilizing well-resolved side chain densities.[14–16] More recently, determining structures at a near-atomic resolution has become increasingly routine, and many icosahedral viruses have been resolved to a resolution of approximately 3 Å.[17–21]

Many reviews have been published on structural analyses of icosahedral viruses using cryo-EM.[22–24] In this review, we focus only on the methodology for icosahedral and symmetry-mismatch reconstruction of viruses by cryo-EM and the possible challenges.

*Project supported by the National Key R&D Program of China (Grant No. 2016YFA0501100), the National Natural Science Foundation of China (Grant Nos. 91530321, 31570742, and 31570727), and Science and Technology Planning Project of Hunan Province, China (Grant No. 2017RS3033).

†Corresponding author. E-mail: lingpengcheng@mail.tsinghua.edu.cn
‡Corresponding author. E-mail: hrliu@hunnu.edu.cn

© 2018 Chinese Physical Society and IOP Publishing Ltd

2. Icosahedral reconstruction

Figure 1 illustrates the flowchart for cryo-EM imaging and data processing of icosahedral viruses. The principle behind 3D reconstruction using cryo-EM is based on the central section theorem; specifically, that a Fourier transformation of a 2D projection from a 3D object is identical to the central section of a Fourier transformation of the 3D object perpendicular to its projection orientation. According to the weak-phase object approximation (WPOA), the intensity of a cryo-EM image is proportional to its 2D projection from 3D macromolecule samples after convolution with contrast transformation function (CTF). Therefore, once the projection orientation and center parameters of each particle image are assigned, all 2D particle images are merged into a 3D Fourier space and a 3D density map is calculated using inversion Fourier transformation. Because of the random projection of particle images, interpolation must be performed in the 3D Fourier space. Simultaneously, based on the calculated 3D density map, orientation and center parameters can be reassigned more accurately to produce a new 3D map with higher resolution. Through this process, the reconstruction resolution can be improved by repeated iteration. In summary, the data processing of icosahedral viruses includes two major steps: determining the orientation and center parameters, and reconstructing a 3D density map.

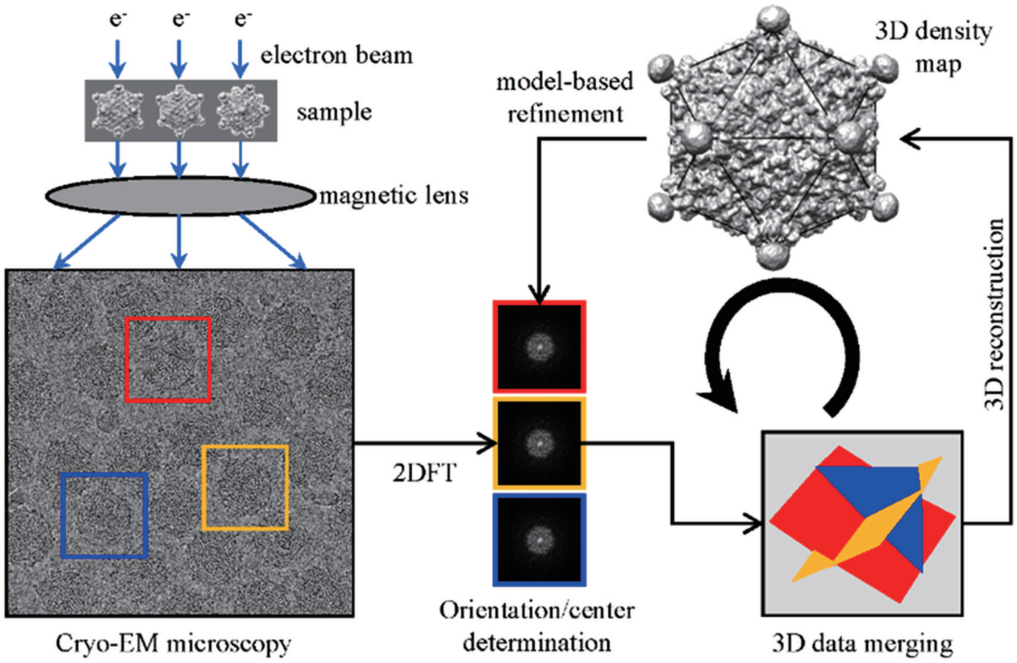


Fig. 1. (color online) Schematic of cryo-EM imaging and data processing of icosahedral viruses.

2.1. General methods for single-particle reconstruction

The determination of the orientation and center parameters begins with an initial 3D model, which is generated from either a low-resolution negative-stain 3D reconstruction, an appropriately filtered x-ray model, an EM map of a homolog,[25] or even a random sphere. The accuracy of the orientation and center parameters of each particle image directly determines the resolution of the reconstruction. Two approaches are used to determine these parameters, which are described in the two subsequent sections.

2.1.1. Orientation and center determination by projection matching

In Fig. 2, an illustration of the basic idea behind projection matching method is shown. Starting from an initial 3D model, many 2D projections are generated at a given angular step size in an asymmetric unit. Subsequently, for each particle image, cross-correlation coefficients are calculated between the image and all projections respectively, and the orientation of the best matched projection is assigned to the particle orientation. The precision of the particle orientation depends on the angular step of the projection. However, the number of projections increases cubically with the step size of the projection. Therefore, a trade-off exists between the computational time needed for reconstruction and the desired resolution.

To accelerate computation, the orientation and center alignment is usually performed in two steps: 2D classification (assigning particles with the same view to one class) and 3D refinement (locally optimizing around given parameters). The first step, 2D classification, is used to determine the three in-plane parameters (x, y, ω); subsequently, all particle images within the same class are averaged to improve the signal-to-noise ratio (SNR); and the second step, 3D refinement, is used to simultaneously determine the two out-of-plane param-

eters (θ, ϕ) and refine the three in-plane parameters. Two approaches are used to perform 2D classification. The first approach, which is implemented in EMAN[26,27] and Relion,[28] does not rely on the initial references. It determines the optimal rotational angle in-plane, then refines the center with the rotational angle, and finally uses the optimal in-plane parameters for classification. The second approach, which is implemented in Frealign,[29] first performs a rough calculation of the center, then searches the rotational relationship between raw particle and template with assigned parameters from 0° to 90°. Angles in the other quadrants, 90° to 360°, are calculated at the same time by rotating the particles 90° without interpolation.

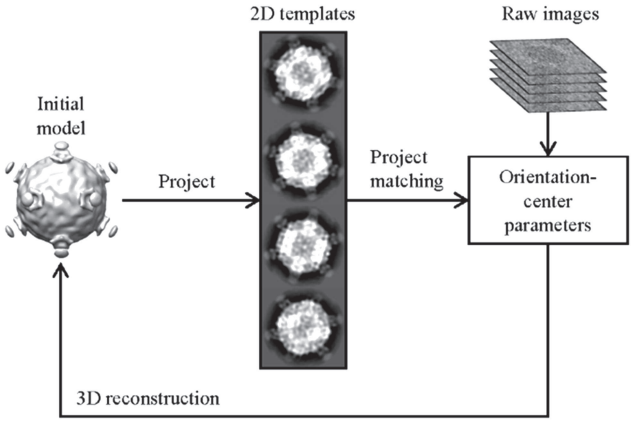


Fig. 2. Schematic of the projection matching method. Two-dimensional projections are generated from an initial model with known out-of-plane parameters (θ, φ), and subsequently, for each raw particle image, the remaining three in-plane parameters (ω, x, y) are assigned by image matching.

An additional approach is polar Fourier transform (PFT), which uses an alternative sequence to assign the orientation and center parameters.[30] Each raw particle image and all projections are first resampled in the polar coordinates, and the initial particle center is determined using the cross-correlation coefficient between the raw particle image and the average image for all projections. Subsequently, the in-plane rotational angle ω is determined using the best rotational correlation between the image and all projections. Finally, an accurate particle center is calculated using the cross-correlation between the particle images rotated with ω and the optimal projection.

2.1.2. Reconstruction by direct Fourier transformation

In cryo-EM, the 3D structure can be described as its Coulomb potential function $\rho(r)$, which can be calculated using Fourier inversion form structure factors $F(X, Y, Z)$. In Cartesian coordinates, the expression is written as follows:

$$\rho(x,y,z) = \iiint F(X,Y,Z)\,\mathrm{e}^{-2\pi\mathrm{i}(xX+yY+zZ)}\,\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z, \quad (1)$$

where (x, y, z) and (X, Y, Z) denote the coordinates of the points in real space and Fourier space, respectively. Theoretically, 3D interpolation can be performed, and all structural factors can be obtained on lattice points in the 3D Fourier space by solving equations; however, as the reconstruction resolution increases,[31] the order of equations becomes too large to be solved. In fact, each lattice point is calculated in 3D Fourier space, $F(X, Y, Z)$, by summing its neighbor points[29] following

$$F(X,Y,Z) = \frac{\sum_j w_j^2 b^2 C_j P_{i,j}}{\sum_j (w_j b c_j)^2 + f}, \quad (2)$$

where $P_{i,j}$ is a lattice point in the 2D Fourier transformation of a particle image j; C_j is the CTF for image j; b is the box transformation; w_j is a weighting factor describing the quality of the image; and f is the Wiener filter constant. Figure 3 depicts an illustration of a 2D interpolation in Cartesian coordinates.

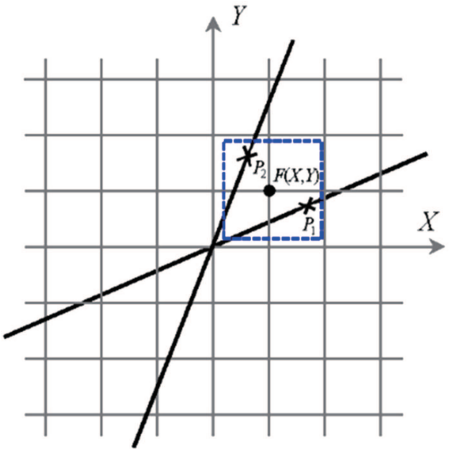


Fig. 3. (color online) Two-dimensional interpolation in Cartesian coordinates. According to Eq. (2), the value of the lattice point, $F(X,Y)$, is calculated by its neighboring points (within the blue box) from the available lines (black).

2.2. Special methods for icosahedral reconstruction

2.2.1. Orientation and center determination using the common-line method

Besides the two general methods, projection matching and PFT, another method can be used to determine the particle orientation and center of icosahedral viruses. In the 1970 s, Crowther *et al.* first proposed the common-line method[4,5] to determine the orientation and center parameters of EM particle images, and the Fourier–Bessel synthesis method to calculate a 3D density map. They used these methods to reconstruct two icosahedral viruses: the tomato bushy stunt virus and human wart virus.[5] In Fig. 4, the generation of common lines is illustrated. According to the central section theorem, the 2D Fourier transformation of a particle image is equal to a central plane in a 3D Fourier volume. For any two different central sections in the Fourier space, the intersecting line is called a pair of cross-common lines (from two planes). Theoretically,

all values on such paired lines are identical. For icosahedral viruses, each plane has 60 equivalent planes, because there are 60 symmetrical operations on the icosahedron; therefore, 60 pairs of cross-common lines exist for every two different planes. For an individual icosahedral particle image, 37 pairs of self-common lines exist, comprising 12 from the 6 five-fold symmetry axes of the icosahedron (one pair between 72° and −72°, and another pair between 144° and −144°); 10 from the 10 three-fold symmetry axes (between 120° and −120°); and 15 from the 15 two-fold symmetry axes (between 0° and 180°).

For an individual virus particle image, an approximate center can be determined using a center-of-mass calculation or a cross-correlation calculation between the particle image and itself rotated by 180° (self-correlation); subsequently, the possible orientation can be exhaustively searched in an asymmetric unit of the icosahedron using the minimal phase residual (the difference in the phase angle of the two points on the pair of common-lines with the same resolution) between self-common line pairs. Because self-common lines are produced by the inherent symmetry of the particle, this method can be used to generate an initial icosahedral model, which is highly effective for viruses with nonsmoothed capsids, such as adenovirus[16,32,33] and cypovirus.[34]

Once an initial 3D model is constructed, several 2D projections with known orientations in an asymmetric unit of an icosahedron can be generated as references. Subsequently, for each particle image, the minimal phase residual of the cross-common lines between the particle image and each of the references can be searched exhaustively to determine the possible orientation and center parameters. Because the cross-common lines are highly sensitive to variations of the particle orientation and center, the search for the orientation and center parameters can converge quickly. Using a similar method, based on the convergence of the orientation and center parameters, more accurate orientation and center parameters can then be obtained by local refinement around the known parameters.

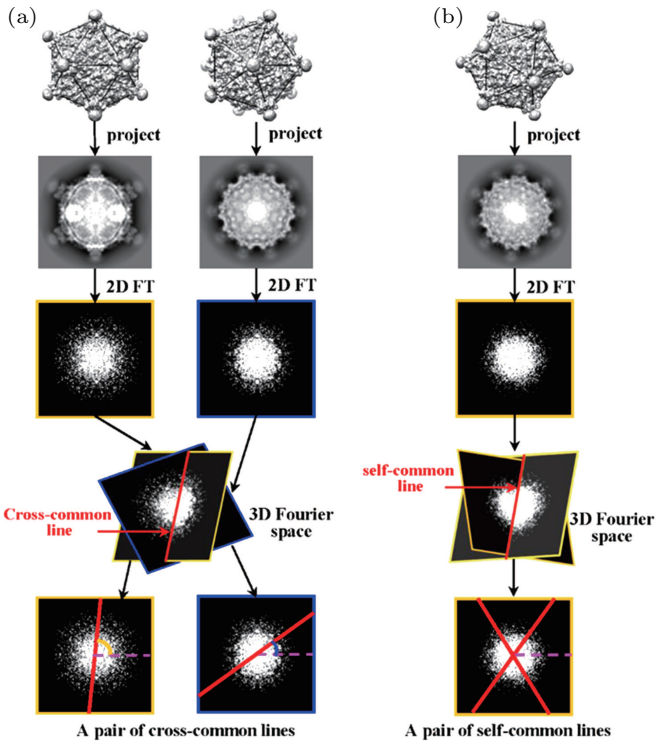


Fig. 4. (color online) Illustration of (a) cross-common line and (b) self-common line. (a) Fourier transformations of two projections (images) intersecting a line in 3D Fourier space, and mapping of the intersecting line to the two Fourier transformations (bottom of (a)) to produce a pair of cross-common lines. (b) Two central sections form one Fourier transformation with two symmetry operations intersecting a line in 3D Fourier space to produce a pair of self-common lines. According to the symmetry operation of an icosahedron, 60 pairs of cross-common lines exist for any two different Fourier transformation, and 37 pairs of self-common lines exist for one Fourier transformation.

Table 1. Selected high resolution of viruses in EMDB.

Virus	EMD	Resolution/Å	Orientation determination method	Reconstruction algorithm
Human rhinovirus B14[40]	8762	2.26	cross common-line	direct Fourier inversion
Adeno-associated virus[41]	6470	2.8	projection matching	direct Fourier inversion
Grapevine fanleaf virus[42]	3246	2.8	multivariate Statistical Analysis	angular reconstitution
Bombyx mori cypovirus[43]	6374	2.9	common lines	direct Fourier inversion
Cowpea mosaic virus[44]	3014	3.04	maximum likelihood classification	direct Fourier inversion
Bombyx mori cypovirus[45]	9564	3.3	cross common-line	direct Fourier inversion
Adeno-associated virus[46]	8100	3.8	polar Fourier transform	direct Fourier inversion
Bombyx mori cypovirus[47]	5926	3.8	cross common-line	Fourier spherical bessel synthesis
Brome mosaic virus[48]	6000	3.8	cross common-line	direct Fourier inversion
Cytoplasmic polyhedrosis virus[14]	5233	3.9	cross common-line	Fourier spherical bessel synthesis
Bacteriophage epsilon15[13]	5003	4.5	projection matching	direct Fourier inversion
Bovine adenovirus 3[32]	2273	4.5	cross common-line	Fourier spherical bessel synthesis

Compared to the different methods for determination of particle orientation and center, the goal of the projection matching method is to achieve convergence quickly to determine a particle's orientation and center parameters by globally searching for all possibilities through 3D refinement. To reduce the computational cost, local refinements can be performed instead, which can be started from several parameter tuples that have demonstrated better correlation coefficients in

the global search. This is called a heuristics search. Subsequently, the search can be optimized by performing local refinements near the best point. However, heuristics searches may result in a local optimum instead of a global optimum, which could be achieved by 3D refinement.. The search for the orientation and center parameters in PFT is divided into three independent steps, and, because the angle calculation is performed in one-dimensional space, the computation speed is high; however, for the same reason, the precision of the calculation is limited. Owing to its fast convergence, the common-line method is widely used to resolve the 3D structures of icosahedral viruses (Table 1). The common-line method has been implemented in IMIRS,[35] JPSR,[36] SIMPLE,[37] and SPIDER.[38] To avoid exhaustive searches and to accelerate calculations, a multipath simulated annealing optimization algorithm with the Monte Carlo optimization scheme has been used to globally determine both the particle center and orientation simultaneously.[39]

2.2.2. Reconstruction methods for icosahedral viruses

In addition to the direct Fourier transformation method, there are two other methods for icosahedral reconstruction, Fourier–Bessel synthesis and Fourier-spherical-Bessel synthesis.

In cylindrical coordinates, the 3D density map can be expanded as

$$\rho(r,\phi,z)=\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}g_n(r,Z)\,\mathrm{e}^{\mathrm{i}n\phi}\,\mathrm{e}^{2\pi\mathrm{i}zZ}\,\mathrm{d}Z, \tag{3}$$

where $g_n(r, Z)$ is an undetermined coefficient, and (r, ϕ, z) denotes the coordinates in real space. The Fourier–Bessel synthesis method comprises the following steps. (i) All 2D Fourier transformations of the particle images are merged to a 3D Fourier space, and the 3D Fourier space is resampled on a series of equally spaced planes of constant Z (Fig. 5); subsequently, each Z plane is divided into a number of equally spaced concentric annuli. (ii) On each annulus, the structure factors are expanded as

$$F(R,\Phi,Z)=\sum_{n}G_n(R,Z)\,\mathrm{e}^{\mathrm{i}n\Phi}, \tag{4}$$

where $G_n(R,Z)$ is the expanding coefficient that can be obtained using the least squares method. (iii) The undetermined coefficients are determined by a Fourier–Bessel transformation as follows:

$$g_n(r,Z)=\int_0^{\infty}G_n(R,Z)J_n(2\pi R)\,\mathrm{d}R, \tag{5}$$

where J_n denotes the cylindrical Bessel function. (iv) g_n is substituted into Eq. (3) to calculate the 3D density map.

In spherical coordinates, the Fourier inversion can be written as

$$\rho(\boldsymbol{r},\theta,\varphi)=\iiint F(\boldsymbol{R})\,\mathrm{e}^{-2\pi\mathrm{i}\boldsymbol{R}\cdot\boldsymbol{r}}R^2\sin\Theta\,\mathrm{d}R\,\mathrm{d}\Theta\,\mathrm{d}\Phi, \tag{6}$$

where (r, θ, φ) and (R, Θ, Φ) denote the coordinates of the points in real space and Fourier space, respectively. First, we merge all 2D Fourier transformations of the particle images to a 3D Fourier space and resample the 3D Fourier space into a series of spherical shells (Fig. 6). Second, for each spherical shell in the 3D Fourier space, we expand the structure factors with icosahedral symmetry-adapted functions (ISAFs)

$$F(\boldsymbol{R})=\sum_{l}\sum_{\mu}f_{l,\mu}(R)S_{l,\mu}(\Theta,\Phi), \tag{7}$$

where f is the expanding coefficient. Each ISAF has the icosahedral symmetry and is a linear combination of spherical harmonics

$$S_{l,\mu}(\Theta,\Phi)=\sum_{m=-l}^{l}A_{l,m}^{\mu}Y_{l,m}(\Theta,\Phi), \tag{8}$$

where A is the expanding coefficient, and Y is the spherical harmonic. ISAFs are a complete orthonormal set of basis functions in spherical coordinates. The expanding coefficients f_l, μ in Eq. (7) can be obtained using the least squares method. Finally, we substitute Eq. (7) into Eq. (6). Taking advantage of the Fourier-spherical-Bessel transformation, the 3D structure is calculated as

$$\rho(r,\theta,\varphi)=4\pi\sum_{l}(-i)^l\left[\sum_{\mu}\left(\int f_{l,\mu}(R)\,j_l(2\pi Rr)R^2\,\mathrm{d}R\right)\cdot S_{l,\mu}(\theta,\varphi)\right], \tag{9}$$

where j_l denotes the spherical Bessel function.

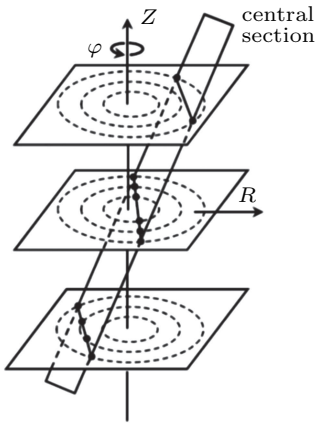


Fig. 5. Schematic of 3D interpolation in cylindrical coordinates shows how a central section cuts the various annuli. Three-dimensional Fourier space is sampled on planes of constant Z and annuli of constant R within each plane.

In theory, interpolation in various reconstruction methods could easily be done using the direct Fourier transform method, but the uneven distribution of the 2D Fourier transformations may lead to different accuracies for different interpolated lattice points in Fourier space. In the Fourier–Bessel synthesis, the calculation is fast thanks to the one-dimensional

interpolation in the Fourier space. However, the basis functions used to interpolate the 3D Fourier space only have a five-fold symmetry, and not an icosahedral symmetry, which leads to a decline in accuracy; also, uneven data distribution in cylindrical coordinates leads to data loss. Compared with the other two methods, interpolation in the Fourier-spherical-Bessel synthesis is performed on each 2D spherical shell in the 3D Fourier space; this results in much more of the available data to be included in the step of linear-squares fitting. Moreover, the symmetry of the ISAFs is identical with that of the reconstructed object; therefore, the Fourier-spherical-Bessel synthesis method can improve the accuracy of interpolation and significantly reduce the influence of noise. Because the basis functions become increasingly complex as the desired resolution is improved, this method increases computing costs and likely increases calculation errors.

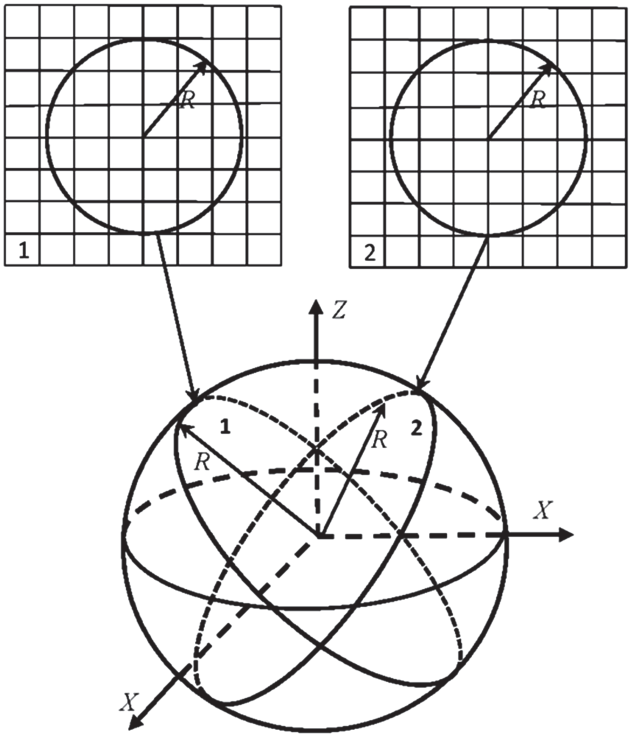


Fig. 6. Schematic of interpolation in spherical coordinates shows how to merge the Fourier transformations of particle images into a 3D Fourier space.

3. Symmetry-mismatch reconstruction of icosahedral virus

Symmetry mismatch is ubiquitous in virus structural organization. For example, adenoviruses comprise 12 three-fold fibers projecting from the 12 five-fold vertices of their icosahedral capsid, and many bacteriophages consist of an icosahedral head and a long tail on one unique vertex; most spherical viruses consist of a nonicosahedral genome encapsidated within an icosahedral capsid. Reconstruction in 3D of an icosahedral capsid presents the great advantage of a 60-fold symmetry, which can be used to significantly improve the reconstruction resolution. However, the 60-fold symmetry becomes a major disadvantage in symmetry-mismatch reconstruction because all nonicosahedral features are smoothed upon imposition of the icosahedral symmetry. This explains why, despite nearly five decades of extensive study of capsid structures of icosahedral viruses, there are few structural studies of the viral genomes within the capsids.

To explain the symmetry-mismatch structure, a 2D symmetry-mismatch object is used as an example: a nonequilateral triangle encapsidated by a square. When the squares are lined up (crystallized), the triangles become random (Fig. 7(a)). Therefore, it is impossible to solve the symmetry-mismatch problem by crystallography. Theoretically, single-particle cryo-EM can solve this problem by asymmetric reconstruction, but the strong information from the square (high symmetry) influences the triangle (asymmetry) orientation determination (Fig. 7(b)).

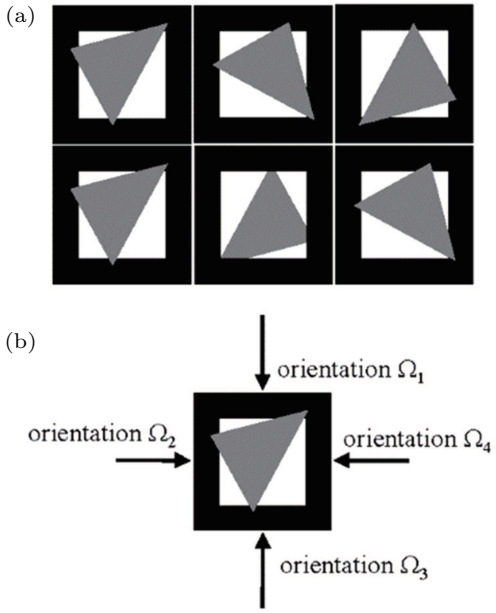


Fig. 7. A nonequilateral triangle encapsidated by a square illustrating a symmetry-mismatch structure. (a) When the squares are lined up, the triangles become random. (b) If the object is projected from an orientation Ω_1, the projection orientation of the square has three other equivalent orientations (Ω_2, Ω_3, and, Ω_4). In contrast, only the projection orientation Ω_1 is correct for the triangle that has no symmetry.

Three major methods are available to study the symmetry-mismatch structures of icosahedral viruses. The first is based on icosahedral reconstruction; its main process comprises three steps: (i) icosahedral reconstruction is performed to obtain a 3D density map as an initial model; (ii) the icosahedral symmetry is relaxed to a C1 symmetry and the orientation of each particle image is exhaustively searched; and (iii) asymmetric reconstruction is performed without any imposed symmetry. Some bacteriophage tail structures have been obtained using this method; however, the resolution of asymmetric reconstructions tends to be much lower than that

of icosahedral reconstructions; possibly because the icosahedral information in particle images causes errors when determining the asymmetric orientation. The second method is based on localized reconstruction from raw particle images; its main process comprises three steps: (i) the icosahedral orientation and center of each particle image are determined using a conventional icosahedral data processing method; (ii) the positions of 12 vertices of the icosahedron are calculated according to the known icosahedral orientation and center, and the vertices images from each 2D raw particle image are boxed; and (iii) these new images are classified and reconstructed without any imposed symmetry to obtain the asymmetric structures. Briggs *et al.* and Morais *et al.* used this method to reconstruct the nonicosahedral structures of a Kelp fly virus vertex[40] and a bacteriophage tail;[41] IIca *et al.* used a similar method to reconstruct the VP4 spike structure in rotavirus.[42] These authors did not obtain high-resolution structures of nonsymmetric regions of the viruses, which is probably due to bias in the nonsymmetric structural information when aligning the overlapping symmetric structural information in raw cryo-EM images.

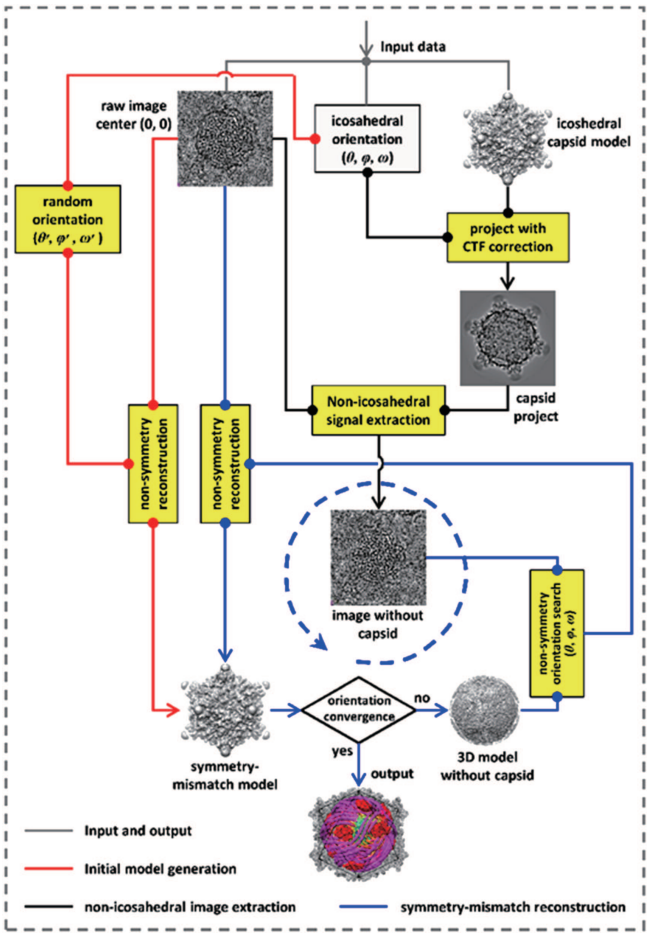


Fig. 8. (color online) Flow chart of a symmetry-mismatch reconstruction. The procedure comprises three steps: nonicosahedral information extraction (black lines), initial asymmetry model generation (red lines), and symmetry-mismatch orientation determination and reconstruction (blue lines). The input data (grey lines) include raw particle images whose centers are shifted to image centers (0, 0), their icosahedral orientations for each particle image, and the 3D density map of the icosahedral capsid.

In our study conducted in 2015, we presented a new symmetry-mismatch reconstruction method based on 2D particle image intensity extraction. Using this method, we resolved the complete structure of the genome within an icosahedral capsid, together with its associated polymerases to a near-atomic resolution. Our symmetry-mismatch reconstruction method, which took advantage of the near-atomic resolution density map of the icosahedral capsid and of the calculated icosahedral orientation and center parameters for all raw particle images, comprised three modules: nonicosahedral information extraction, initial asymmetry model generation, and symmetry-mismatch orientation determination and reconstruction. Figure 8 presents a flow chart of our new method, and the complete process is described in subsequent paragraphs.

3.1. Nonicosahedral information extraction

According to WPOA, regardless of CTF convolution, the intensity of a cryo-EM microscopy image can be written as follows:[43]

$$I(x,y) = 1 + 2\sigma\phi(x,y), \tag{10}$$

where σ is the interaction constant between the electron beam and the object, and $\phi(x, y)$ is a 2D projection of a 3D object, which can be written as

$$\phi(x,y) = \int \rho(x,y,z)\,\mathrm{d}z, \tag{11}$$

where $\rho(x, y, z)$ is the density of the 3D object. The density of an icosahedral virus can be written as

$$\rho = \rho_{\mathrm{icos}} + \rho_{\mathrm{non-icos}}, \tag{12}$$

where ρ_{icos} and $\rho_{\mathrm{non-icos}}$ denote the icosahedral and nonicosahedral components, respectively. By substituting Eqs. (11) and (12) in Eq. (10), a 2D particle image of an icosahedral virus can be written as

$$\begin{aligned} I(x,y) &= 1+2\sigma\left[\int \rho_{\mathrm{icos}}(x,y,z)\,\mathrm{d}z + \int \rho_{\mathrm{non-icos}}(x,y,z)\,\mathrm{d}z\right] \\ &= I_{\mathrm{icos}}(x,y) + I_{\mathrm{non-icos}}(x,y), \end{aligned} \tag{13}$$

where I_{icos} and $I_{\mathrm{non-icos}}$ denote the icosahedral and nonicosahedral intensities, respectively. If ρ_{icos} is known, the icosahedral intensity for each particle image can be calculated, $I_{\mathrm{cal_icos}}$, and subsequently, the nonicosahedral information can be extracted as

$$I_{\mathrm{non-icos}}(x,y) = I(x,y) - I_{\mathrm{cal_icos}}(x,y). \tag{14}$$

Because of the extremely low SNR of the raw particle image and the distinct grayscale between I_{icos} and $I(x,y)$, we

could not extract the nonicosahedral information directly according to Eq. (14). To normalize the icosahedral information in the raw image $I(x, y)$ and calculated images $I_{\text{cal_icos}}(x,y)$, we fit a linear relationship between the calculated capsid image $I_{\text{cal_icos}}(x,y)$ and the capsid information in the raw particle image $I(x, y)$. This can be written as

$$I_{\text{capsid}}(x,y) = a_0 + a_1 I_{\text{cal-capsid}}(x,y), \tag{15}$$

where a_0 and a_1 are coefficients that need to be determined. To calculate accurately a_0 and a_1, we chose for the fitting an annulus region that only contains capsid information in each viral particle image (Fig. 9(a)), and the inside and outside radii of the annulus were determined by comparing the one-dimensional radially averaged intensity distribution of the well-centered raw particle image and the calculated capsid projection image, respectively (Fig. 9(b)). From the radial intensity distribution profile of the raw particle image, we identified a valley between r_1 and r_2 that is free of nonicosahedral information (Fig. 9), which could be defined as

$$r_1 < \sqrt{x^2 + y^2} < r_2 \tag{16}$$

in the original raw image and the calculated capsid projection. Comparing the radially averaged curves of the raw image and the calculated capsid image (Figs. 9(c) and 9(d)), the valleys were determined to be in the identical circular region. Using only the pixels within this circular region of the raw particle image and the calculated capsid image, we determined the coefficients a_0 and a_1 by least square fitting of the grayscale. Thus, the nonicosahedral image (Fig. 9(e)) of each particle is obtained using

$$\begin{aligned} &I_{\text{genome}}(x,y) + I_{\text{noise}}(x,y) \\ &= I_{\text{image}}(x,y) - [a_0 + a_1 I_{\text{capsid}}(x,y)]. \end{aligned} \tag{17}$$

The missing valley in the radially averaged profile of the genome image (Fig. 9(f)) confirms that the capsid information is removed.

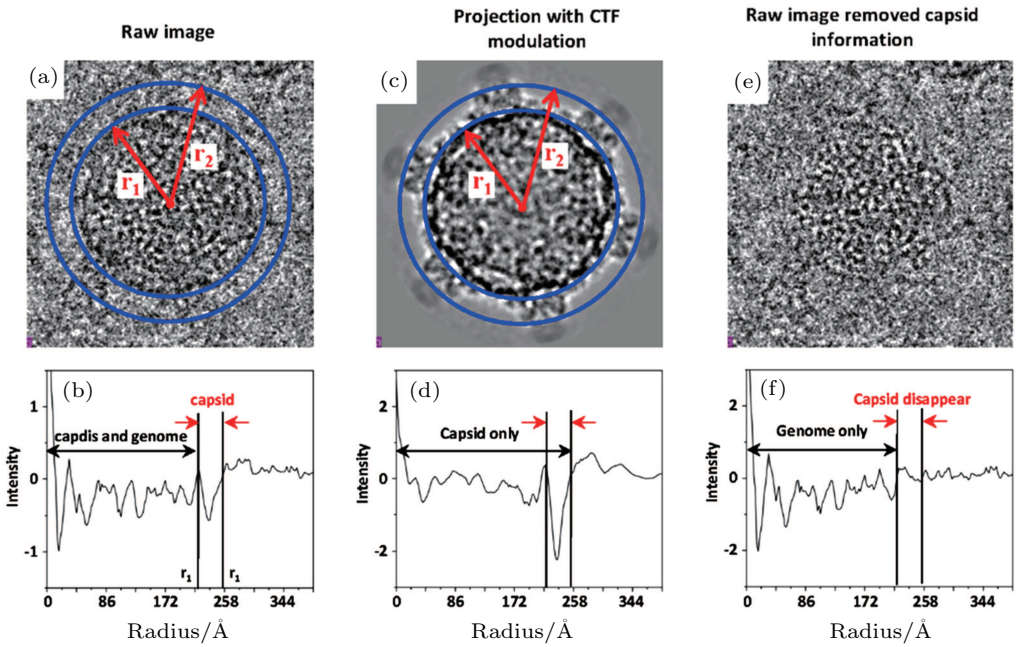


Fig. 9. (color online) The process of genome image extraction. (a) The raw image is well aligned after a high resolution map of the capsid is reconstructed. (b) The annulus information of capsid shell can be determined from the raw image. (c) The particle's projection can be projected from high resolution map with CTF correction. (d) Subsequently, the annulus information of capsid shell can be determined from projection after masking of the inner information. (e) The information on the annulus of capsid shell is used to fit the same signal on the raw image and the projection, and the genome image is extracted by subtracting the capsid from the raw image. (f) The capsid information disappears on the genome image.

3.2. Initial nonicosahedral model generation

For each raw particle image's icosahedral orientation, we randomly selected one of the 60 equivalent orientations as an initial orientation. Subsequently, we reconstructed a full virus structure (including the icosahedral and nonicosahedral components) without any symmetry imposed using raw particle images. Finally, we obtained an initial model of the nonicosahedral component by masking icosahedral capsid densities.

3.3. Symmetry-mismatch orientation determination and 3D reconstruction

For each new nonicosahedral image produced in the nonicosahedral information extraction step, we projected the model of the nonicosahedral component to generate 60 projection images according to the 60 equivalent orientations of the icosahedron. Subsequently, we searched for the projection that best matched with the nonicosahedral image (the highest cross-correlation coefficient) and assigned the corresponding

orientation to the new image for further analysis. Next, we reconstructed a full virus structure without any imposed symmetry using the raw particle images with their newly assigned orientations, improving the model of the nonicosahedral structure by masking the capsid structure. Finally, the improved model was used as a new reference to perform iterations of the symmetry-mismatch orientation search and asymmetry reconstruction until the orientations of individual nonicosahedral images stabilized and no further improvements to the nonicosahedral structure could be obtained.

Compared with the other two methods for symmetry-mismatch reconstruction, the major advantage of our method is that we use the extracted images to determine the symmetry-mismatch orientation and the raw particle image to reconstruct the asymmetric structure. Therefore, we avoid high-symmetry bias when determining the asymmetric orientation, and we push the resolution of the full symmetry-mismatch of the icosahedral structure to a near-atomic level.

4. Conclusion and perspectives

With the development of DED cameras, the resolution of cryo-EM reconstructions for low or nonsymmetric complexes has experienced a revolution; the highest resolution so far was 1.8 Å.[1] However, determining the near-atomic structures of icosahedral viruses was already routine before the arrival of DED cameras, and essential improvements in icosahedral reconstruction have not occurred since, particularly for large viruses. Except for the change in the research focus of cryo-EM, the key reason for this is that the 3D reconstruction of icosahedral viruses suffers from methodological challenges.

4.1. Geometric distortion of large virus particles

To improve the reconstruction resolution of viruses, the capsid is always considered to be perfect icosahedron. However, the capsid is not, in fact, a true spherical shell; some geometrical local distortions must occur among the 20 triangular facets of an icosahedron. Moreover, tests have indicated that electron microscopes suffer from elliptical distortion of 2%–3%.[44,45] Distortions can hinder further improvement of the resolution, particularly for large particles, since distortions tend to become increasingly severe as the diameter of particles increases.

4.2. Ewald sphere effect

CTF appears as a vibrating curve in the Fourier space of a cryo-EM image, specifically in the high-frequency part; a small error in the defocus value could lead to a complete phase reversal. Currently available methods for CTF correction assume a constant defocus value for an entire micrograph or particle, while, in fact, a focus gradient exists within a virus particle, and the difference in defocus values between the top and the bottom of the particle is equal to the diameter of a virus particle. Overcoming this problem using current reconstruction algorithms is difficult.

4.3. Symmetry-mismatch reconstruction

All icosahedral viruses contain a symmetry-mismatch core, composed of a genome and its associated proteins, encapsidated within a capsid shell. Little structural information is currently known about the core. Our symmetry-mismatch reconstruction method is only the beginning; to resolve the atomic structure of the inner cores of viruses, more robust algorithms need to be developed.

References

[1] Merk A, Bartesaghi A, Banerjee S, Falconieri V, Rao P, Davis M I, Pragani R, Boxer M B, Earl L A, Milne J L S and Subramaniam S 2016 *Cell* **165** 1698
[2] Kuhlbrandt W 2014 *ELife* **3** e03678
[3] Grigorieff N and Harrison S C 2011 *Current Opinion in Structural Biology* **21** 265
[4] Crowther R A 1971 *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **261** 221
[5] Crowther R A, Amos L A, Finch J T, De Rosier D J and Klug A 1970 *Nature* **226** 421
[6] Adrian M, Dubochet J, Lepault J and McDowall A W 1984 *Nature* **308** 32
[7] Bottcher B, Wynne S A and Crowther R A 1997 *Nature* **386** 88
[8] Conway J F, Cheng N, Zlotnick A, Wingfield P T, Stahl S J and Steven A C 1997 *Nature* **386** 91
[9] De Rosier D J 1997 *Nature* **386** 26
[10] Ludtke S J, Baker M L, Chen D H, Song J L, Chuang D T and Chiu W 2008 *Structure* **16** 441
[11] Zhang X, Settembre E, Xu C, Dormitzer P R, Bellamy R, Harrison S C and Grigorieff N 2008 *Proc. Natl. Acad. Sci. USA* **105** 1867
[12] Yu X, Jin L and Zhou Z H 2008 *Nature* **453** 415
[13] Jiang W, Baker M L, Jakana J, Weigele P R, King J and Chiu W 2008 *Nature* **451** 1130
[14] Cheng L, Sun J, Zhang K, Mou Z, Huang X, Ji G, Sun F, Zhang J and Zhu P 2011 *Proc. Natl. Acad. Sci. USA* **108** 1373
[15] Zhang X, Jin L, Fang Q, Hui W H and Zhou Z H 2010 *Cell* **141** 472
[16] Liu H, Jin L, Koh S B, Atanasov I, Schein S, Wu L and Zhou Z H 2010 *Science* **329** 1038
[17] Yu G, Li K, Huang P, Jiang X and Jiang W 2016 *Structure* **24** 1984
[18] Schur F K, Obr M, Hagen W J, Wan W, Jakobi A J, Kirkpatrick J M, Sachse C, Krausslich H G and Briggs J A 2016 *Science* **353** 506
[19] Zhao H, Li K, Lynn A Y, Aron K E, Yu G, Jiang W and Tang L 2017 *Proc. Natl. Acad. Sci. USA* **114** 3601
[20] Kostyuchenko V A, Lim E X, Zhang S, Fibriansah G, Ng T S, Ooi J S, Shi J and Lok S M 2016 *Nature* **533** 425
[21] Sirohi D, Chen Z, Sun L, Klose T, Pierson T C, Rossmann M G and Kuhn R J 2016 *Science* **352** 467
[22] Grigorieff N and Harrison S C 2011 *Current Opinion in Structural Biology* **21** 265
[23] Rosenthal P B 2015 *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences* **370**
[24] Jiang W and Tang L 2017 *Current Opinion in Structural Biology* **46** 122
[25] Beckmann R, Bubeck D, Grassucci R, Penczek P, Verschoor A, Blobel G and Frank J 1997 *Science* **278** 2123
[26] Ludtke S J, Baldwin P R and Chiu W 1999 *Journal of Structural Biology* **128** 82
[27] Tang G, Peng L, Baldwin P R, Mann D S, Jiang W, Rees I and Ludtke S J 2007 *Journal of Structural Biology* **157** 38
[28] Scheres S H 2012 *Journal of Structural Biology* **180** 519
[29] Grigorieff N 2007 *Journal of Structural Biology* **157** 117
[30] Baker T S and Cheng R H 1996 *Journal of Structural Biology* **116** 120
[31] Crowther R A, De Rosier D J and Klug A 1970 *Proc. Roy Soc. Lon. A* **317** 319

[32] Cheng L, Huang X, Li X, Xiong W, Sun W, Yang C, Zhang K, Wang Y, Liu H, Huang X, Ji G, Sun F, Zheng C and Zhu P 2014 *Virology* **450–451** 174
[33] Cao C, Dong X, Wu X, Wen B, Ji G, Cheng L and Liu H 2012 *Journal of Virology* **86** 12322
[34] Yang C, Ji G, Liu H, Zhang K, Liu G, Sun F, Zhu P and Cheng L 2012 *Proc. Natl. Acad. Sci. USA* **109** 6118
[35] Liang Y, Ke E Y and Zhou Z H 2002 *Journal of Structural Biology* **137** 292
[36] Guo F and Jiang W 2014 *Methods in Molecular Biology* **1117** 401
[37] Elmlund D and Elmlund H 2012 *Journal of Structural Biology* **180** 420
[38] Shaikh T R, Gao H, Baxter W T, Asturias F J, Boisset N, Leith A and Frank J 2008 *Nature Protocols* **3** 1941
[39] Liu X, Jiang W, Jakana J and Chiu W 2007 *Journal of Structural Biology* **160** 11
[40] Briggs J A, Huiskonen J T, Fernando K V, Gilbert R J, Scotti P, Butcher S J and Fuller S D 2005 *Journal of Structural Biology* **150** 332
[41] Morais M C, Tao Y, Olson N H, Grimes S, Jardine P J, Anderson D L, Baker T S and Rossmann M G 2001 *Journal of Structural Biology* **135** 38
[42] Ilca S L, Kotecha A, Sun X, Poranen M M, Stuart D I and Huiskonen J T 2015 *Nature Communications* **6** 8843
[43] Cowley J M 1995 *Diffraction Physics, 3rd revised edition* (North-Holland personal library)
[44] Grant T and Grigorieff N 2015 *Journal of Structural Biology* **192** 204
[45] Yu G, Li K, Liu Y, Chen Z, Wang Z, Yan R, Klose T, Tang L and Jiang W 2016 *Journal of Structural Biology* **195** 207
[46] Dong Y, Liu Y, Jiang W, Smith T J, Xu Z and Rossmann M G 2017 *Proc. Natl. Acad. Sci. USA* **114** 8017
[47] Spear J M, Noble A J, Xie Q, Sousa D R, Chapman M S and Stagg S M 2015 *Journal of Structural Biology* **192** 196
[48] Hemmer C, Orlov I, Ackerer L, Marmonier A, Hleibieh K, Schmitt-Keichinger C, Vigne E, Gersch S, Komar V, and Belval L, in *European Microscopy Congress* 2016*: Proceedings* (Wiley Online Library)
[49] Yu X, Jiang J, Sun J and Zhou Z H 2015 *ELife* **4** e07901
[50] Hesketh E L, Meshcheriakova Y, Dent K C, Saxena P, Thompson R F, Cockburn J J, Lomonossoff G P and Ranson N A 2015 *Nature Communications* **6** 10113
[51] Li X, Zhou N, Chen W, Zhu B, Wang X, Xu B, Wang J, Liu H and Cheng L 2017 *Journal of Molecular Biology* **429** 79
[52] Drouin L M, Lins B, Janssen M, Bennett A, Chipman P, McKenna R, Chen W, Muzyczka N, Cardone G, Baker T S and Agbandje-McKenna M 2016 *Journal of Virology* **90** 8542
[53] Zhu B, Yang C, Liu H, Cheng L, Song F, Zeng S, Huang X, Ji G and Zhu P 2014 *Journal of Molecular Biology* **426** 2167
[54] Wang Z, Hryc C F, Bammes B, Afonine P V, Jakana J, Chen D H, Liu X, Baker M L, Kao C, Ludtke S J, Schmid M F, Adams P D and Chiu W 2014 *Nature Communications* **5** 4808