

Reconstructing Cyano-bacteriophage P-SSP7 structure without imposing symmetry

CURRENT STATUS: POSTED

Xiangan Liu

Baylor College of Medicine, Houston, TX 77030, USA

Ryan H. Rochat

Baylor College of Medicine, Houston, TX 77030, USA

Wah Chiu

Baylor College of Medicine, Houston, TX 77030, USA

DOI:

10.1038/nprot.2010.96

SUBJECT AREAS

Biological techniques *Structural biology*

KEYWORDS

virus asymmetric reconstruction, cross common line, MPSA, Multi-Path Simulated Annealing, dynamic masking strategy

Introduction

The 9 Å structure of P-SSP7 was determined by single particle cryo-electron microscopy (cryo-EM) reconstruction without imposing any symmetry. The icosahedral features of the capsid shell of this phage provide a strong signal that greatly facilitates the process of single particle icosahedral orientation determination. Unfortunately, these symmetric features greatly weaken the ease with which the particle's asymmetric orientation can be determined. Icosahedral reconstructions have been a focus of the field for quite some time as they push the envelope of practically achievable resolution via single particle cryo-EM. While there has always been an interest in symmetry free alignment, the ability to achieve asymmetric reconstructions for particles consisting of a large number of symmetric components had been limited by reconstruction algorithms and the sheer magnitude of data required for the process. As of late, the culmination of new and advanced technology in the field of cryo-EM has permitted researchers to perform asymmetric reconstruction¹⁻⁷ at subnanometer resolutions. Outlined below is a series of eight steps (*also see the online methods and supplementary figure 8 of the related article*) that illustrate how Multi-Path Simulated Annealing (MPSA) was used to reconstruct subnanometer resolution phage particle maps without imposing symmetry on the final model.

Equipment

1. CCD frames of electron images of ice embedded P-SSP7 bacteriophages. 2. Known icosahedral alignment parameters of the PSSP-7 particle images.
3. EMAN v1.9 (<http://blake.bcm.tmc.edu/eman/>).
4. Customized MPSA programs (available upon requests) based on EMAN .
5. UCSF Chimera (<http://www.cgl.ucsf.edu/chimera/>).
6. Aimra software (<http://www.amira.com/>)
7. A multiprocessor computer cluster
8. A Linux workstation with >816MB video memory

CRITICAL: An asymmetric reconstruction with the existence of an icosahedrally reconstructed map is not a computationally intensive procedure. It does not require many CPUs, and using more than

30~40 CPUs for asymmetric reconstruction at our cluster slows down its I/O performance. A high performance graphic card will make it possible to view the whole asymmetric map without down sampling. For lower performance graphic card, it is necessary to clip the 3D map to visualize only a portion of the map.

Procedure

Aligning single particles images without the assumption of symmetry is a multi-factorial process. Through a series of iterative steps it is possible to identify the asymmetric features within an image, orient about them accordingly, and then refine the model. The following steps (*also see the online methods and supplementary figure 8 of the related article*) outline the process of asymmetric orientation determination and have the prerequisite that accurate icosahedral alignment parameters (i.e. orientations and centers). Additionally, this methodology relies on the production of a variety of masks and reference projection images, each of which is used at specific points in the refinement to hone in on the true asymmetric orientations.

Step 1 (preparation)

Determine the icosahedral orientation of the 36,000 particles using the MPSA icosahedral orientation determination procedure⁸, and store the alignment parameters in a separate list file (EMAN⁹ format: image number, image name, Euler angles and center). For the icosahedral reconstruction of the P-SSP7 phage, use a box size of 720 pixels (1.17 Å/pixel), which may result in some single-particle images where the phage tail is cut off. For the asymmetric reconstruction it is necessary to include the tail (protein apparatus outside the capsid shell, including nozzle, adaptor and fiber) at one of the 12 5-fold vertices of the phage, corresponding to a box size of 960 pixels (1.17 Å/pixel) in the raw data. Since the target resolution of the final asymmetric reconstruction is well below that which the data was sampled, you must bin the single particle images by a factor of 2, to a size of 480 pixels (2.34Å/pixel) to reduce the computational burden of Fourier synthesis. As binning the data does not alter the computed icosahedral orientations, it is only necessary to modify the particle centering parameters of the list file accordingly. The new center are derived from the following two steps: 1) adjust center 120 pixels in both X and Y directions due to the offset from enlarging the box size from

720 pixels to 960 pixels, $120 \text{ pixels} = (960-720)/2$; 2) shrink the adjusted center by 2 because the raw image, with size of 960 pixels, was shrunk by 2. For example, an original center (358.148, 359.562) needs to be adjusted to (239.074, 239.781), where $239.074 = (358.148 + 120)/2$, $239.781 = (359.562 + 120)/2$. Using the modified list file described above, and the binned particle images, icosahedrally reconstruct an initial 3D model (480×480×480 pixels, 2.34 Å/pixel) (Fig. 1a) using the EMAN *make3d* program⁹.

Step 2 (initial 3D tail mask and model generation)

During the icosahedral reconstruction of P-SSP7, the tail density located at a special 5-fold vertex is weakened by a factor of at least 12, as there are 12 possible positions for the tail in the icosahedral map. Accordingly, you must lower the display threshold of the initial map (Fig. 1a) from 2.54 to 0.12 in order to visualize the “tail” (Fig. 1b). At this threshold it is possible to see ghost densities of the tail at each of the 12 5-fold vertices. Using the Amira software package produce a tail mask (Fig. 1c) from the ghost density of the tail along the +Z-axis of the initial icosahedral map. This mask should loosely fit the ghost density (i.e. be slightly bigger than the actual size of the ghost density). To make the negative density of tail ghosts visible in Chimera¹⁰, add a positive constant of 0.5 (sufficient to display all densities) to the 3D icosahedral density (Fig. 1b) before applying the mask. Extract the initial tail model (Fig. 1d) from the initial icosahedral model using the previously generated tail mask and then low-pass filter the map in EMAN using the *proc3d* program. Fig. 1e-f show how the ghost tails and capsid look without and with low-pass filtering (20Å), respectively.

Step 3 (initial 2D tail mask generation)

The tail of P-SSP7 occupies a very small amount of volume in the virion. Using a dynamic masking strategy you can make it stand out from the highly symmetric capsid and DNA, by maximally excluding non-target objects in the 2D raw particle images. Assuming the tail is a constant feature in P-SSP7, different particle orientations cause the tail projection in the raw 2D images to occur in different locations and have different shapes (Fig. 2a-b). To optimize computational time, create a database of all possible masks, indexed by projection orientation. As Azimuthal rotation (AZ defined in

EMAN) does not appreciably change the shape of a mask projection, do not consider it as a projection parameter. As such, project the masks for the database along the initial 3D tail mask for only two of the three Euler angles (altitude (*Alt*) and in-plane (*Phi*)), as defined in the EMAN coordinate system. First, make projections of the 3D tail mask for different altitude orientations (Fig. 2c) (0° to 90° at an interval of 3°). In the current implementation of our program, the total computational time is dictated by the time it takes to load all of these masks into memory, and is relatively independent of the process of orientation determination. As there is a trade off between number of masks and computational time, instead of creating a prolific database of tight masks, merge several neighboring masks together to create a set of looser, but more computationally efficient, masks. Accordingly, with the exception of the first projection at 0° , merge every three consecutive projections to the 3rd one (Fig. 2c). Under this scheme you produce 11 projection masks from 0° to 90° (*Alt*) at an interval of 9° . The next step generates in-plane (*Phi*) rotation masks from 0° to 360° . Rotate each of the masks (again with the exception for the 0° (*Alt*) projection) along the Z-axis with the step size 0.5° . Just as before, to minimize computational restraints, merge every 20 rotations and index them to the angle corresponding to the middle (10th) mask (Fig. 2d). Under this scheme 36 in-plane rotation masks are produced for each mask at a given altitude angle. From these two consecutive steps 361 ($1+10\times 36$) 2D tail masks, covering all possible asymmetric orientations of the particle, are generated. During the process of dynamic masking, when choosing a mask from the database for a given particle, round the particle's orientation to the nearest index orientation (ignoring the azimuthal orientation) when picking the 2D tail mask.

Step 4 (initial 2D tail reference projections)

The masked density of a raw image is comparable to the references projected from the tail model. We empirically use 400 projections as cross common line¹¹ references during asymmetric orientation alignment. Compared to the number of references typically used during icosahedral reconstruction, this is sufficient to obtain a subnanometer resolution reconstruction. To save time, distribute this process to 10 CPUs, using EMAN's *runpar* program. After each node simultaneously generates 40

random projections, merge these 10 projection sets together into one (this process takes only a matter of minutes to complete).

Step 5 (masking potential asymmetric features)

Starting with a single icosahedral orientations (from step 1), MPSA internally generates 60 equivalent icosahedral orientations for each particle image. As each of these orientations is a candidate for the true asymmetric orientation of a particle, round each of these icosahedral orientations to the nearest index orientation in the 2D tail mask database (from step 3), extract the corresponding mask and apply it to the particle image.

Step 6 (determining asymmetric orientations)

In MPSA, icosahedral orientations⁸ are determined through the use of cross common line¹¹ correlation between a raw particle image and reference projections with 60-fold symmetry enforced (60 common lines per comparison between the particle image and a reference projection). Here the asymmetric orientations are determined using the same convention, but with no symmetry enforcement (one common line per comparison). Compute the cross common line correlations between each of the masked regions generated in step 5 and the 400 2D tail model reference projections. Compute the residual (a value inversely related to the correlation) using the formula $\sum_i (0.5 - \cos(\varphi_i))$ where i runs all selected points along the common line, and φ_i is the phase difference between the two Fourier transforms. The distribution of this function changes from iteration to iteration and depends on the data range used for calculating the common line correlations. For the first iteration, focus on a subset of frequency space between $1/200$ - $1/60 \text{ \AA}^{-1}$ for calculating the residuals of the cross common line correlations. After calculating the residuals for the 60 candidate orientations, choose the orientation that had the smallest residual as the particle's asymmetric orientation. For the first iteration ($1/200$ - $1/60 \text{ \AA}^{-1}$) the best residuals should range from 0.42 to 0.47, which will improve to 0.32 to 0.42 by the fifth iteration ($1/200$ - $1/18 \text{ \AA}^{-1}$).

Step 7 (reconstructing a 3D density map without symmetry imposed)

Following the schematic outlined in steps 5 and 6, process each of the 36,000 particles to determine

their initial asymmetric orientations. Once the approximate asymmetric orientations of all the particles are determined, reconstruct a 3D density map, free of any symmetry, using the EMAN *make3d* program⁹.

Step 8 (refining the asymmetric 3D density map)

After the first iteration of the asymmetric reconstruction algorithm, the tail's relative intensity becomes much stronger as the capsid shell is no longer over-represented in the reconstruction. Alternatively the portal density is not well differentiated from the high density DNA that fills the capsid shell. For the 2nd iteration continue to use just the tail to generate masks and reference projections, however starting in the 3rd iteration, and for subsequent iterations, use both the tail and portal densities (Fig. 2b) (which reside within the capsid shell) as the asymmetric target, which makes the orientation determination easier due to the larger volume of the model used. To refine the asymmetric reconstruction, repeat steps 2-7 for data with different resolution cut-offs (45Å, 35Å, 25Å, and 18Å for the 2nd, 3rd, 4th and 5th iterations respectively). For the final couple of iterations, extend the resolution cut-off to 12Å and compare the orientations determined at this level to those from the 5th iteration, keeping only those particles that were consistent from the 5th to final iteration. From the initial set of 36,000 particles, approximately 15,000 of the particles should be retained in the final reconstruction, yielding a final asymmetric map at 9Å.

Timing

As this is an interactive process, the majority of the time is spent making the masks and filtering the data. Computationally speaking, it takes only a few hours to align all 36,000 of the particles when using 30 CPUs on a contemporary cluster. In total t

Critical Steps

Generate a proper 3D mask. Ideally the 3D mask you use should be slightly larger than the target asymmetric feature, if the mask is either too small or too large, it could result in the misalignment of your data.

Troubleshooting

Slow alignment

Depending upon the architecture of your cluster, as you begin to use more and more CPUs (>40 CPUs in our case) I/O quickly becomes the limiting factor in the speed with which this process will complete. If you observe diminishing returns in computation times, consider reducing number of CPUs you are using as it may be related to sluggish I/O. For the current version of MPSA (for asymmetric reconstructions) the maximum number of CPUs should be limited to 30~40 on our current clusters.

Failure to converge

If the asymmetric map does not converge (i.e. the asymmetric feature is not clearly resolved) after 7 or 8 iterations, make sure the following requirements were met.

1. The initial ghost tail was masked properly.
2. The database of 3D masks was created properly.
3. The data ranged used for cross common line correlation was not increased too fast from iteration to next iteration at the beginning.

Misalignment

If the internal portal and DNA terminus show DNA fingerprinting, this indicates that some particles were not aligned correctly. It is important to check for consistency of the particle's alignment parameters, we did so by monitoring this behavior in the last few iterations. After removing those particles with inconsistency in their alignment parameters, the map's quality should be improved.

Artifacts from the masked images

It is imperative to note that when implementing this algorithm you must make sure that after masking the targeted features, the mean value of the edge in the masked image must be zero so as to avoid artifacts in Fourier space.

Anticipated Results

In the case of P-SSP7 we expect to see a 6-fold or pseudo 6-fold tail structure for phage. Furthermore, after 3 or 4 iterations, the phage portal structure should begin to stand out as well. For a generic samples, you should expect to see the asymmetric feature coarsely resolved after a couple iterations. As this process is built on refining the location and shape of the asymmetric feature from one iteration to the next, if the feature is not resolved in intermediate maps, it becomes harder to properly align

particles in the subsequent iteration.

References

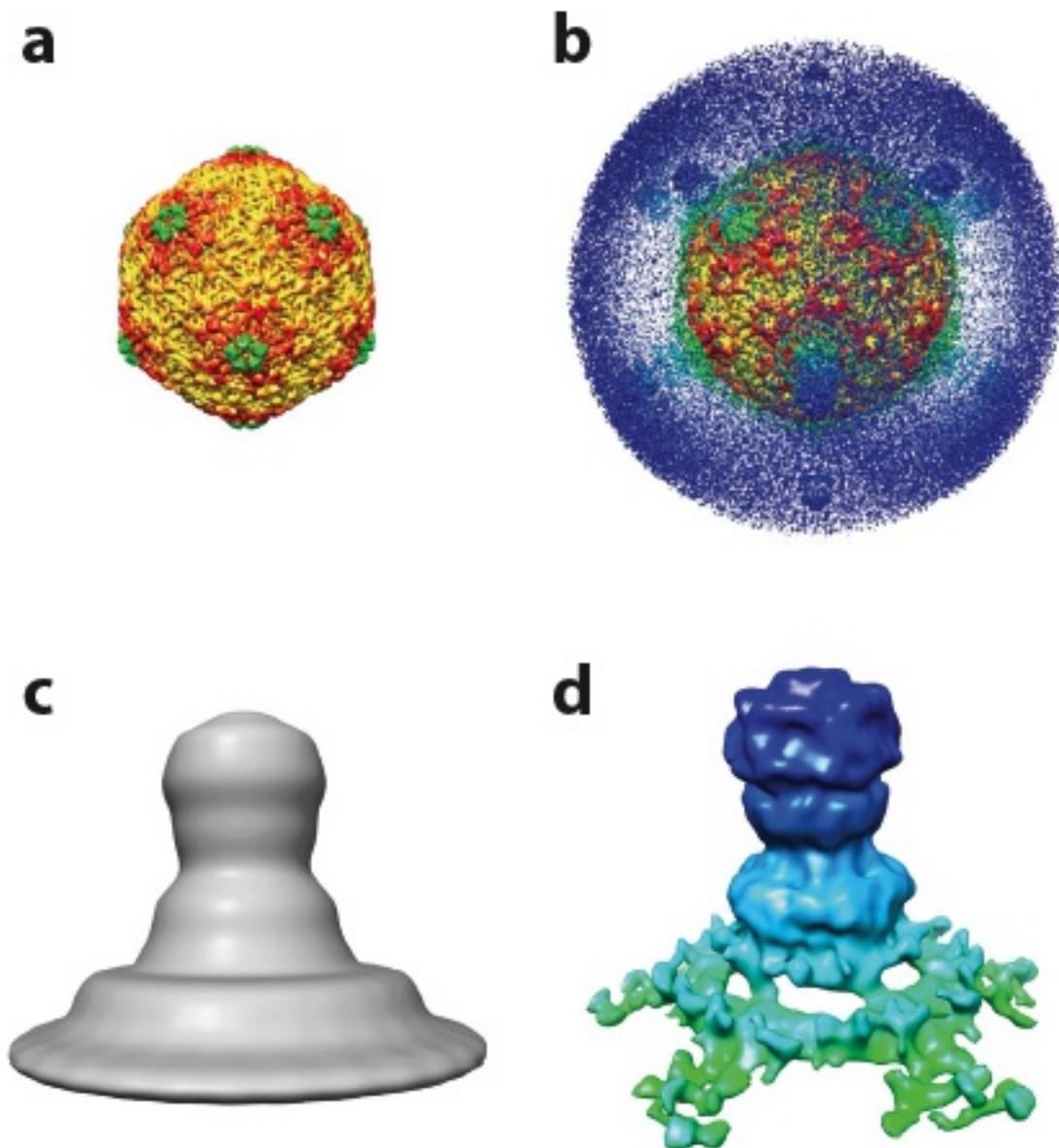
1. Jiang, W. et al. Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature* **439**, 612-6 (2006).
2. Chang, J., Weigele, P., King, J., Chiu, W. & Jiang, W. Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. *Structure* **14**, 1073-82 (2006).
3. Lander, G.C. et al. The structure of an infectious P22 virion shows the signal for headful DNA packaging. *Science* **312**, 1791-5 (2006).
4. Tang, J. et al. DNA poised for release in bacteriophage phi29. *Structure* **16**, 935-43 (2008).
5. Leiman, P.G. et al. The structures of bacteriophages K1E and K1-5 explain processive degradation of polysaccharide capsules and evolution of new host specificities. *J Mol Biol* **371**, 836-49 (2007).
6. Agirrezabala, X. et al. Maturation of phage T7 involves structural modification of both shell and inner core components. *EMBO J* **24**, 3820-9 (2005).
7. Plisson, C. et al. Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. *EMBO J* **26**, 3720-8 (2007).
8. 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* **160**, 11-27 (2007).
9. Ludtke, S.J., Baldwin, P.R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**, 82-97 (1999).
10. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).

11. Crowther, R.A., DeRosier, D.J. & Klug, A. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proc Roy Soc Lond A* **317**, 319-340 (1970).

Acknowledgements

This research is supported by NIH (P41RR002250, R01AI0175208), Welch Foundation (Q1242) and a training fellowship from the Keck Center NLM Training Program in Biomedical Informatics of the Gulf Coast Consortia (NLM Grant No. T15 LM007093).

Figures



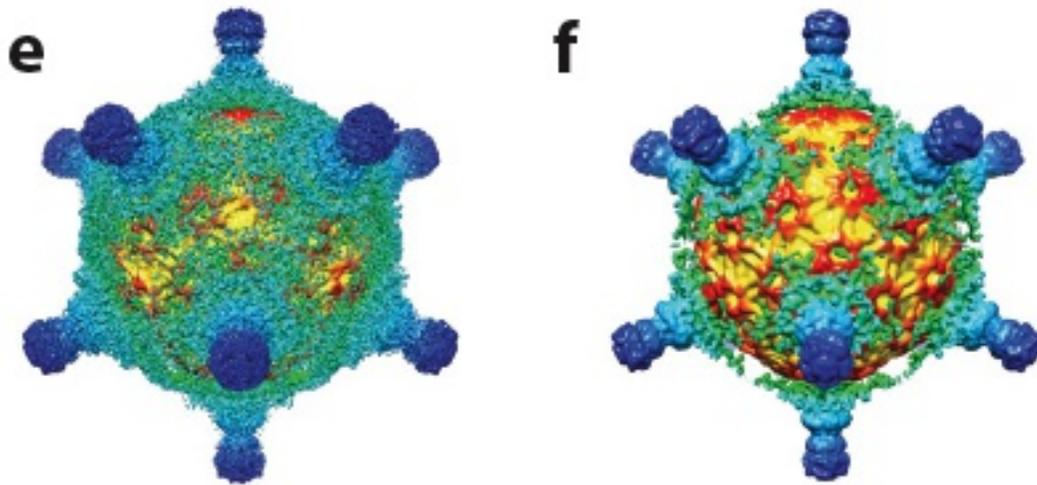
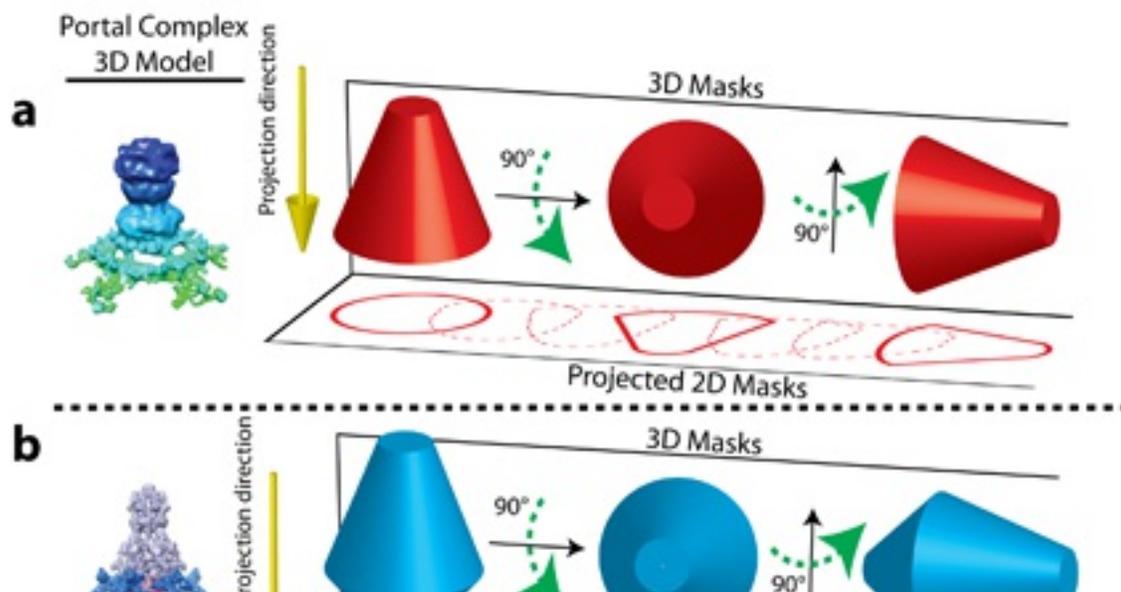


Figure 1

Initial 3D tail mask and model generation. (a) Icosahedral reconstruction of P-SSP7 viewed at normal threshold (2.54). (b) After lowering the threshold from 2.54 to 0.12 it is possible to see 12 ghost tail densities situated on the five-fold vertices of the capsid shell. (c) 3D tail mask manually generated from the ghost tail density (b) at +Z-axis using Amira software. (d) Initial 3D tail density masked from the ghost density after adding a constant of 0.5 and low-pass filtering to 20Å. (e) The tails and capsid from (b) after adding a constant of 0.5 and then removing noise using the masks from (c). The 3D density map shown at very low threshold. (f) The density map from (e) low-pass filtered (20 Å) and viewed at very low threshold.



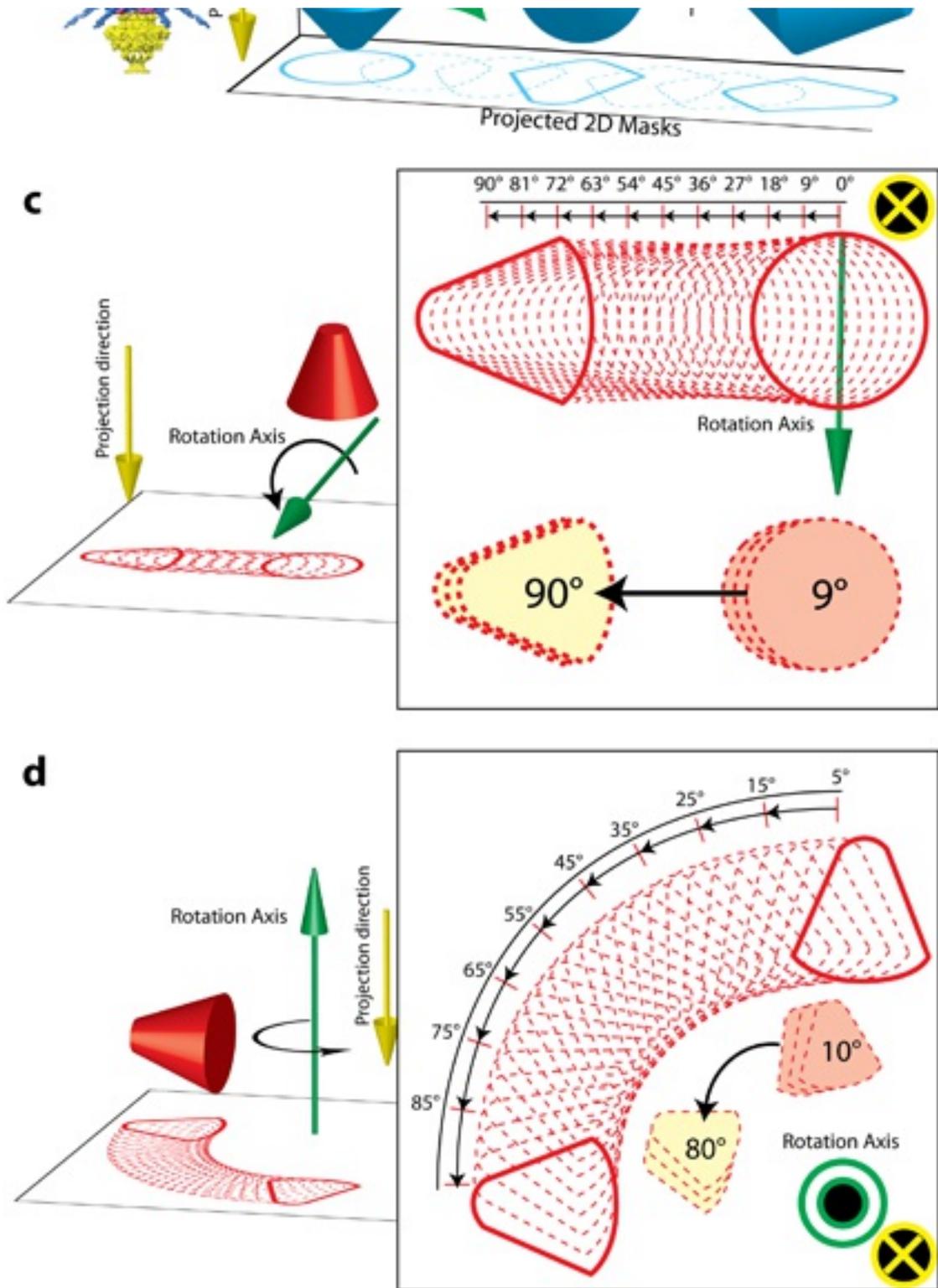


Figure 2

2D tail mask database generation. (a) The initial tail ghost is shown on the left. The corresponding 3D mask is schematically shown on the right (red cone). The 2D masks generated in two consecutive steps: (I) projecting 3D masks distributed along the altitude

($_Alt_$) rotation angle, (II) rotating the 2D masks from the previous step with different in-plane rotation angles. (b) In later iterations, both tail and portal show up clearly. Both of them were treated as the target feature for our asymmetric reconstruction. The corresponding 3D mask is schematically shown in blue, while the 2D mask generation is similar to (a). (c) 2D masks in step (I) are generated by projecting the 3D mask along the altitude ($_Alt_$) rotation plane with a 3° interval and superimposing every three masks with each other across 90° starting at 3° (the 0° is not merged with any other masks) for a total of 11 merged masks. (d) 2D masks in step (II) are generated by rotating each of the masks from (c) in-plane ($_Phi_$) (except for the 0° mask) with step size of 0.5° . Every 20 masks (covering a range of 10°) are superimposed into a single merged mask, resulting in a total of 36 merged masks for the entire 360° range of $_Phi_$. A total of 361 ($1+10\times 36$) merged masks are produced by this process.

Structural changes in a marine podovirus associated with release of its genome into Prochlorococcus

by Xiangan Liu, Qinfen Zhang, Kazuyoshi Murata, +7
Nature Structural & Molecular Biology (19 April, 2010)