CHAPTER SEVEN

Three-Dimensional Asymmetric Reconstruction of Tailed Bacteriophage

Jinghua Tang,* Robert S. Sinkovits,* *and* Timothy S. Baker*,†

Contents

* Department of Chemistry & Biochemistry, University of California, San Diego, La Jolla, California, USA
† Division of Biological Sciences, University of California, San Diego, La Jolla, California, USA

Abstract

A universal goal in studying the structures of macromolecules and macromo-
lecular complexes by means of electron cryo-microscopy (cryo-TEM) and three-
dimensional (3D) image reconstruction is the derivation of a reliable atomic or
pseudoatomic model. Such a model provides the foundation for exploring in
detail the mechanisms by which biomolecules function. Though a variety of
highly ordered, symmetric specimens such as 2D crystals, helices, and icosahe-
dral virus capsids have been studied by these methods at near-atomic resolu-
tion, until recently, numerous challenges have made it difficult to achieve sub-
nanometer resolution with large ($\geq \sim$500 Å), asymmetric molecules such as the
tailed bacteriophages.

After briefly reviewing some of the history behind the development of asym-
metric virus reconstructions, we use recent structural studies of the prolate
phage ϕ29 as an example to illustrate the step-by-step procedures used to
compute an asymmetric reconstruction at sub-nanometer resolution. In contrast
to methods that have been employed to study other asymmetric complexes, we
demonstrate how symmetries in the head and tail components of the phage can
be exploited to obtain the structure of the entire phage in an expedited,
stepwise process. Prospects for future enhancements to the procedures cur-
rently employed are noted in the concluding section.

1. Introduction: 3D Asymmetric Reconstruction of Tailed Bacteriophage

Electron microscopy and three–dimensional (3D) image reconstruc-
tion have been the preferred tools for more than 40 years for studying large
macromolecular structures that resist crystallization. Even for those viruses
that can be crystallized, an advantage of electron microscopy is that it can be
used to capture more transient, intermediate stages in the viral life cycle
(e.g., Steven *et al.*, 2005) or to visualize the virus complexed with anti-
bodies, receptors, or other molecules and ligands (e.g., Smith, 2003; Stewart
et al., 2003). These techniques were first applied to images of negatively
stained samples of the helical, contractile tail of bacteriophage T4 (DeRosier
and Klug, 1968; see DeRosier's Personal Account in this volume) and to the
icosahedral tomato bushy stunt and human papilloma viruses (Crowther
et al., 1970), and they have since been used to solve the structures of a
rapidly expanding universe of macromolecules and macromolecular

complexes imaged by means of electron cryo-microscopy (cryo-TEM) (Cheng and Walz, 2009; Jonic *et al.*, 2008; also, see other chapters in this volume).

Whenever possible, advantage is taken of the inherent symmetry of the particles since this generally allows one to reach higher resolutions and maximize the signal-to-noise ratio in the final reconstructed map. One of the best examples of exploiting symmetry is found in the study of icosahedral viruses, where each image contains information from 60 equivalent views (Baker *et al.*, 1999; Crowther *et al.*, 1970). Recent advances in both the microscopy and image processing methods have enabled cryo-reconstructions of icosahedral viruses to approach atomic resolution (Chen *et al.*, 2009; Wolf *et al.*, 2010; Zhang *et al.*, 2008, 2010; Zhou, 2008).

1.1. Symmetry mismatch in phage structure

Many bacteriophage contain a multi-subunit, multicomponent tail connected to a unique vertex of the capsid, and it is through this vertex that the genome is packaged during assembly and released during infection. The heads of tailed phages have either a prolate (e.g., T4 and ϕ29) or an isometric (e.g., P22, T7, λ, and ε15) morphology, but all heads possess a fivefold rotational symmetry about an axis that passes through the tail and opposing vertex. The tail complex generally has 6- or 12-fold redundancy in the constituent viral proteins, which results in a symmetry mismatch at the junction between the head and tail (Jiang *et al.*, 2006; Lander *et al.*, 2006; Xiang *et al.*, 2006). The DNA genome is a molecule with a unique sequence and hence cannot adopt a structure that matches any of the local symmetries in the capsid or tail components. Given that there is just one copy of the genome and at least one prominent, symmetry mismatch between the capsid and tail, the phage as a whole must be asymmetric.

Symmetry averaging is often used as part of the reconstruction process, and whether applied to viruses with full icosahedral symmetry or to tailed phages with fivefold symmetric heads, comes at a cost. Only those components that possess the imposed symmetry will be accurately represented in the averaged final map. Other features of the virus structure, such as the packaged genome or unique tail, will be smeared out since they do not share the imposed symmetry. Reconstructions of these viruses could, in principle, be carried out from start to finish without applying any symmetry using techniques similar to those used to study ribosomes (Frank, 2009), in which no symmetry is imposed. But the tailed phages provide a unique set of challenges since intrinsic local symmetries in parts of the phage can complicate the reconstruction process. For example, an attempt to process images of P22 phage, in which only C_1 symmetry was assumed from the start, failed to lead directly to a valid reconstruction (Chang *et al.*, 2006). Thus, it can be beneficial to exploit the inherent local symmetry in components of the

phage when designing a reconstruction strategy that will yield a reliable density map from images of particles whose global structure is asymmetric.

1.2. History of asymmetric virus reconstructions by single-particle cryo-TEM

The widespread success with icosahedral virus reconstructions has also stimulated the development of techniques to study nonicosahedral viruses (Johnson and Chiu, 2007). After generalizing our polar Fourier transform method (Baker and Cheng, 1996) to handle cyclic symmetries, we obtained the first reconstruction of a tailed bacteriophage with a prolate head (Tao *et al.*, 1998). In that study of ϕ29, fivefold symmetry was imposed to enhance features in the head structure. However, since the ϕ29 tail is not fivefold symmetric, its structure was smeared out in the reconstruction. Later, a two-step processing scheme was developed to preserve symmetry mismatched components in the entire phage (Morais *et al.*, 2001). In step one, a fivefold-averaged reconstruction was computed in which the head but not tail was symmetrized. Then, the fivefold symmetry constraint of the head was removed and each particle image was compared to five separate, related projections of the model. The projection with the highest correlation to each particle image was used to assign the view orientation for that particle. This new set of particle orientations provided the necessary information needed to combine the images and compute a new reconstruction in which no symmetry was enforced. The second step of the procedure was repeated and led to the first asymmetric reconstruction of the complete ϕ29 phage at 33-Å resolution (Morais *et al.*, 2001).

Since this initial ϕ29 asymmetric reconstruction, several other 3D density maps of entire tailed phages have been determined at progressively higher resolutions (Table 7.1). In all these studies, even those at lower resolutions ($\sim$20 Å), the head and tail structures were clearly resolved. Our studies of ϕ29 have led to two, sub-nanometer resolution, asymmetric reconstructions that have made it possible to resolve features corresponding to helices in the head–tail connector as well as a highly condensed, toroid-like DNA structure embedded within a cavity at the connector–tail junction (Tang *et al.*, 2008b).

1.3. Alternative strategies to determine the structures of asymmetric viruses

Not all phages or asymmetric complexes can be readily solved using the strategy we have outlined above. For example, T4 phage with its long, contractile tail, remains a significant challenge for single particle, asymmetric reconstruction methods. For T4, a concerted, "divide and conquer" approach was used to solve the head (Fokine *et al.*, 2004) and tail

Table 7.1 Chronological history of asymmetric cryo-reconstructions of entire tailed bacteriophage

Phage	Head shape	Software[a]	Resolution (Å)	Reference
ϕ29	Prolate	PFT	33	Morais *et al.*, 2001
T7	Isometric	XMIPP	24	Agirrezabala, 2005
ε15	Isometric	EMAN	20	Jiang *et al.*, 2006
P22	Isometric	EMAN	20	Chang *et al.*, 2006
P22	Isometric	SPIDER	17	Lander *et al.*, 2006
ϕ29	Prolate	EMAN	16	Xiang *et al.*, 2006
ϕ29	Prolate	EMAN, FREALIGN, AUTO3DEM	7.8	Tang *et al.*, 2008b
N4	Isometric	EMAN	29	Choi *et al.*, 2008

[a] Software indicates main program(s) used to perform reconstruction. It is assumed that, in all or most cases, additional scripts were required for image preprocessing, data manipulation, file format conversions, and other tasks.

(Kostyuchenko *et al.*, 2005) structures separately. Asymmetric reconstruction techniques have also been used to examine virus–host interactions at limited resolution. Studies of polio and Semliki Forest virus attachment to liposomes were aided by manually adding high intensity dots in the images at points in the membrane where virus particles attached (Bubeck *et al.*, 2008). These dots served as fiducial markers to help in determining the relative orientations of individual virus particles and to compute from those images a reconstruction that indicated a unique vertex is involved in the delivery of the genome in both types of virus.

1.4. Cryo-TEM of ϕ29

The sample preparation and microscopy steps required to produce a set of images from which the reconstructed structure of an asymmetric virus can be obtained are identical to those used to study icosahedral particles (e.g., Baker *et al.*, 1999). We recently computed asymmetric reconstructions of two different ϕ29 particles (Tang *et al.*, 2008b). These included fiberless (gp8.5-) virions and fiberless "ghosts," which are particles formed by inducing virions *in vitro* to lose their dsDNA genome and the two molecules of viral gene product 3 (gp3) that are covalently linked to the ends of the linear genome. The same procedures were used to determine the 3D structures of both types of particles (to 7.8 and 9.3 Å, respectively), and we limit our discussion here to the procedures used to study the virion. Briefly, the microscopy involved first taking purified samples of fiberless ϕ29 virions

and vitrifing them over holey, carbon-coated grids (Chapter 3, Vol. 481). Images of these samples were then recorded on Kodak SO163 electron image film at a nominal magnification of 38,000× and an electron dose of $\sim 20e^-/Å^2$ in an FEI CM200 FEG microscope operated at 200 keV. Micrographs that exhibited minimal astigmatism and specimen drift, and with the objective lens under-focused by 1–5 μm, were digitized at 7 μm intervals on a Zeiss PHODIS scanner and bin-averaged to yield an effective pixel size of 3.68 Å. A total of 12,682 particle images were selected from 74 micrographs for further processing.

1.5. Strategy for determining the ϕ29 structure

The general image reconstruction scheme we used followed a model-based refinement procedure. This entailed aligning raw particle images relative to a series of projections of an existing 3D model to estimate the origin and orientation of each particle. The particle origin is defined by the (x,y) pixel coordinates of the position of the center of the particle in the image, and the particle orientation is defined by three angles (θ, ϕ, ω) that specify the direction from which the particle is viewed in the image (See Baker et al., 1999 for definition of these angles.). With this set of five particle parameters, a new 3D reconstruction can be computed and used as the model for the next iteration of the process.

The specific reconstruction strategy that we adopted in our ϕ29 study involved two major branches as shown schematically in Fig. 7.1. This strategy took into account that fact that ϕ29 has a prolate head ($\sim$480 × 600 Å) and a tail of comparable length ($\sim$400 Å). The first branch of the processing scheme constructs a reliable starting model and achieves optimal particle boxing. This includes calculating separate head and tail models and then combining them into a single, hybrid model. The second branch takes this hybrid model, computes projections, and compares these to each raw image to assign to it an origin and orientation. The images are used to compute a new reconstruction and several cycles of alignment and image screening (to weed out "bad" particles) are carried out until no further improvement in resolution is achieved.

1.6. Image reconstruction software

Owing to the complexity of performing asymmetric reconstructions, we found it necessary to use different image reconstruction packages at various stages of the process. In the ϕ29 project and as described below, we used EMAN (Ludtke *et al.*, 1999), FREALIGN (Grigorieff, 2007), AUTO3-DEM (Yan *et al.*, 2007b), and BSOFT (Heymann, 2001) to ultimately reach sub-nanometer resolution. It is important to stress that our choice of programs more reflects our familiarity with the capabilities of the software
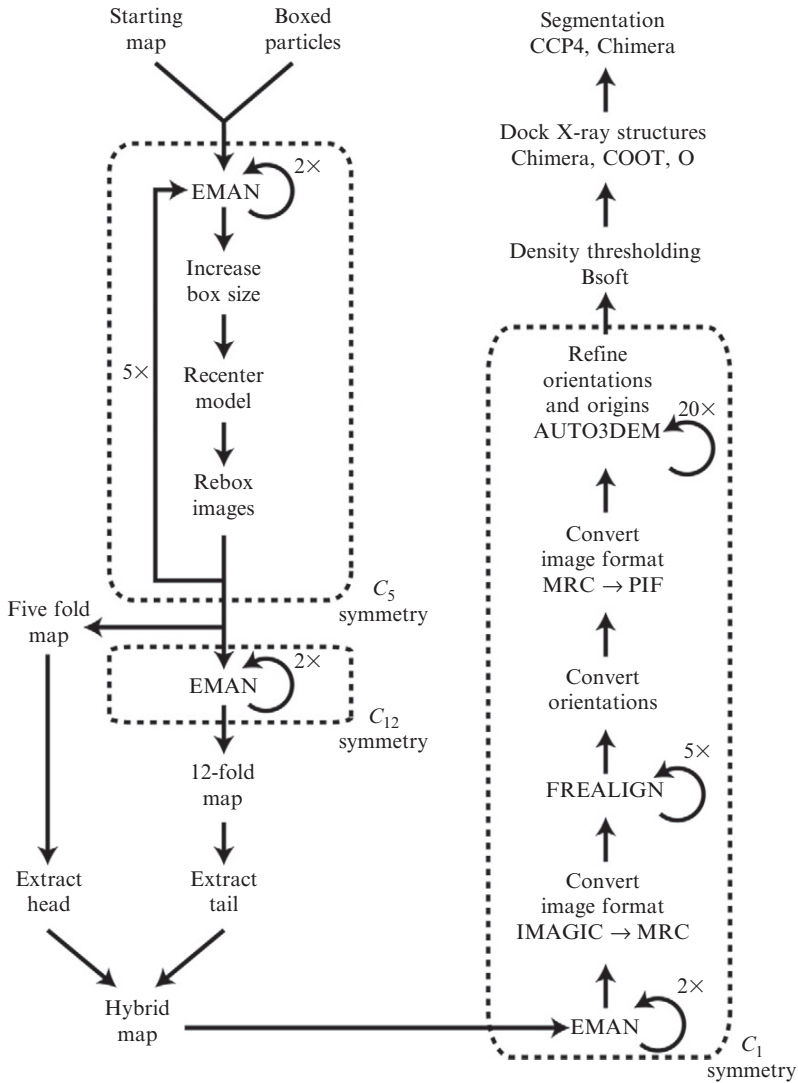
Figure 7.1 ϕ29 image reconstruction flowchart. Left hand side of flowchart shows steps taken to generate hybrid-starting model built from fivefold symmetric phage head and 12-fold symmetric tail. Right hand side illustrates steps in fully asymmetric reconstruction. Numbers next to loops indicate the number of iterations used in stages of the ϕ29 reconstruction and will likely vary for other phage species.

than it does with providing an optimized strategy. These packages and others such as SPIDER (Frank *et al.*, 1996) and XMIPP (Sorzano *et al.*, 2004) contain many useful features that require significant expertise in order to use to their full potential (see also Chapter 15 in this volume).
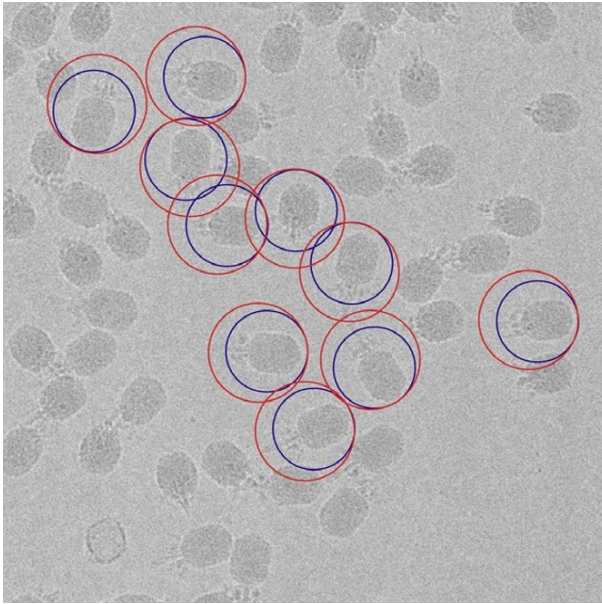
Figure 7.2 Progressive boxing of phage particles. The small blue circles indicate the initial boxing of the particles, with box size chosen to capture head and proximal portion of tail. As the reconstruction progresses, the box size is gradually expanded to include more of the tail (large red circles). Note that the red and blue circles are not concentric and that the top of the phage head remains a constant distance from the edge of the circle.

2. PARTICLE BOXING

The purpose of boxing is to window out individual particles from their surroundings and to minimize the contribution of extraneous noise in the images to the final 3D reconstruction. Ideally, the boxing procedure centers each particle in its own box and excludes all neighboring particles. The defocus used in generating phase contrast during microscopy causes particle details to spread outside the particle boundary (Rosenthal and Henderson, 2003), and hence the pixel dimensions of the box need to extend well enough beyond this boundary to preserve structural information. The choice of the box size depends on resolution and amount of defocus. For the $\phi 29$ image data, we chose a box size approximately 50% larger than the width of the prolate head (Fig. 7.2, red circles).

The type of boxing employed depends on the shape of the phage head and the length of the tail. If the head is isometric and the tail is short relative to the size of the capsid, which is true for some phage like P22 (Lander *et al.*, 2006), the particle can simply be boxed so that the head center coincides with the center of the box. Alternatively, if the tail dimension is comparable to that of the head, then the center of the box should coincide with the center of the whole particle to avoid having to make the box size excessively large.

2.1. Initial tight boxing of the phage head

Given that we did not have a starting model of the entire $\phi 29$ virion, we decided to adopt a systematic, stepwise approach to solving its structure. This involved first constructing a model just for the head from a subset of ~ 1000 images masked tightly to include primarily the phage heads. Subsequently, we incrementally increased the size of the box to include more and more of the tail with the eventual goal of including the entire image of each phage particle in the final reconstruction as described below (Section 3). Hence, we started by boxing out just the particle heads and thereby excluded most of the tail and most of the neighboring particles (Fig. 7.2, blue circles). This tight boxing strategy helped to enhance the reliability with which initial particle origin and orientation parameters could be determined. Such a procedure is unnecessary with icosahedral particles since the origin of spherically symmetric objects is generally fairly easy to define quite accurately. The procedure does limit the resolution of the initial 3D reconstruction, but the overriding concern at this point is to obtain a reliable starting model. All boxing was carried out with the program RobEM (http://cryoem.ucsd.edu/programs.shtml) and images were converted from PIF to IMAGIC format using BSOFT (Heymann, 2001) before the next step of processing was performed using EMAN (Ludtke *et al.*, 1999).

2.2. Reboxing to center the particle

Whether particle boxing is performed manually or automatically, it is difficult to assure that each particle will be centered to 1 pixel accuracy in its box. Experience shows that reconstruction quality can be improved by periodically reboxing the particles using the latest set of origin positions obtained during the iterative refinement process (Gurda *et al.*, 2010). More significantly, with asymmetric particles it is important to make sure that the origin of each boxed particle correlates with the defined origin of the most current reconstructed model. Hence, after each cycle of EMAN as described later (Section 3.1), the identified origin of the particle within the box is used to reextract without interpolation the particle image from the micrograph such that this origin lies within 1 pixel of the center of the new box.

2.3. Expanding the box to include the tail

After a reliable head map was obtained (i.e., one in which pentameric and hexameric gp8 capsomers were clearly resolved), all particles were boxed anew from the raw micrographs using a circular mask, typically larger by 10–15 pixels in radius, to include more of the tail structure in each image. This necessitated that the density map of the head just calculated be padded

with voxels set to the background density to match the size of the newly boxed particles and used as the model to align the images. Padding entailed adding voxels to expand all three dimensions of the cubic density map by equal amounts. The long axis of the phage was kept centered in the map but voxels in the axial direction were primarily added toward the tail side. This procedure assured that the particle (head + currently included portion of tail) was centered within the box containing the reconstructed 3D map.

The padded density map was used to generate a new set of projected images from which the origin and orientation parameters for each image could be redetermined. These were then used to compute a new reconstruction and to recenter and rebox the particles. At this point the mask was expanded, the map repadded with the background density, and the origins and orientations redetermined. This cycle of steps was repeated four additional times, at the end of which the entire phage particle was included in the density map.

2.4. Additional considerations about boxing

Typically, the early stages of the reconstruction process are accelerated significantly by making use of images that are two- or fourfold, bin-averaged. This is feasible because the primary goal at this stage is to obtain a low-resolution starting model whose size and shape are approximately correct. Hence, it is not necessary to use data at full pixel resolution. It is worth noting that, though automatic boxing routines can be used to speed up the boxing process, it was helpful with $\phi29$ to take the extra time to manually screen the entire data set of images and select the best ones to include in the processing. Ultimately, even the most sophisticated alignment algorithms cannot compensate for poor particle selection. Finally, though our reboxing strategy added additional steps to the entire image processing procedure, it proved to be quite effective in assuring that the particles were optimally centered in the boxes.

3. GENERATING A STARTING MODEL OF THE COMPLETE $\phi29$ PHAGE

Considerable effort is often required at the onset of a new project to generate a reliable, nonbiased model for initiating refinement of particle origin and orientation parameters. If the head essentially has icosahedral symmetry (i.e., neglecting the presence of the tail), the random model method (Yan *et al.*, 2007a) provides a relatively straightforward means to obtain a suitable starting model. To this then, one can either graft to one vertex of the icosahedral reconstruction an available reconstructed density

map of the tail (Lander *et al.*, 2006) or a very simple, cylindrically averaged 3D model constructed from the image of a single, clearly visible tail (Jiang *et al.*, 2006). Even though the tail in the latter instance would not have the correct rotational symmetry, it is good enough to jump start refinement.

Because the ϕ29 head has a prolate rather than isometric shape, the random model method could not be used to generate a starting model for the reconstruction process. However, we were able to use the ϕ29 prohead structure (Morais *et al.*, 2005) as a starting model for the head of the mature phage since the two are similar in size and shape and both have fivefold axial symmetry (Fig. 7.3A).

3.1. Fivefold, symmetrized model of ϕ29 head

The first 3D reconstruction of the head and truncated tail was obtained by combining all the tightly boxed particle images (Section 2.1) and imposing fivefold axial symmetry in EMAN. Then, as described in Sections 2.1 and 2.2, we obtained a series of four phage reconstructions that progressively encompassed more of the tail (Fig. 7.3B) and ultimately included the entire phage (Fig. 7.4A).

3.2. Hybrid head–tail model of ϕ29

Given that the capsid portion of the ϕ29 head has fivefold rotational symmetry, its structure was preserved in all of the fivefold averaged reconstructions generated as just described (Section 3.1). There is ample evidence
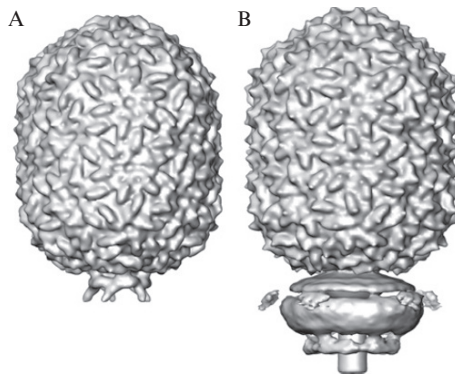

Figure 7.3 Starting model for phage head and connector. (A) Shaded surface representation of ϕ29 prohead map used as the starting point for the reconstruction of the complete phage. (B) Initial, fivefold averaged phage reconstruction obtained with tightly boxed particle images (blue circles, Fig. 7.2). Fivefold averaging yields clearly defined features in the head but features of the connector and tail are smeared out.
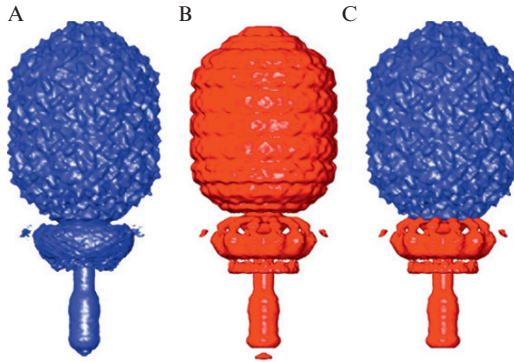
Figure 7.4 Construction of asymmetric, hybrid model. (A) Shaded surface represen-
tation of complete phage reconstruction with fivefold symmetry enforced during
processing. (B) Same as (A) for phage with 12-fold symmetry. Since the head and tail
do not share the same symmetry, these reconstructions smear out the tail and head
densities in panels (A) and (B), respectively. (C) Hybrid model obtained by combining
head from fivefold reconstruction and tail from 12-fold reconstruction. Colors in
the hybrid map highlight contributions from the two symmetrized maps. At this point
the symmetry mismatch between the head and tail was unknown and no effort
was made to impose a particular rotational alignment between the two segments.

that the connectors and tail portions of all tailed bacteriophage, including
ϕ29, do not possess fivefold symmetry like the capsids (Jiang *et al.*, 2006;
Lander *et al.*, 2006; Xiang *et al.*, 2006). The presence of a head–tail
symmetry mismatch was also evident in the fivefold averaged ϕ29 recon-
struction since features in the head were more distinct and not smeared as in
the tail (Fig. 7.4A). The entire neck and tail of ϕ29 was shown to have
quasi–six–fold symmetry (Peterson *et al.*, 2001) and the existence of 12
distinct appendages was revealed in some of the very earliest electron
micrographs of negatively stained phage (Anderson *et al.*, 1966). Hence,
to establish the rotational symmetry, if any, in the structure of the ϕ29 tail,
we used EMAN to recompute the phage reconstruction with imposed
three-, six-, and 12-fold symmetries instead of fivefold. Inspection of
these three reconstructions showed, as expected, a smeared capsid structure
but tails with features more distinct than in the fivefold averaged map. This
strategy works since the head portions of the images have an equally poor
agreement with all projections of the model and the assignment of particle
orientations becomes influenced most by the tail structure. In addition, the
ring of appendages near the head–tail junction appeared most distinct in the
12-fold symmetrized map (Fig. 7.4B).

At this point we constructed a hybrid model in the following way. We
used the *subregion selection* tool in Chimera (Pettersen *et al.*, 2004) to segment

out the fivefold symmetrized head portion of the first reconstruction (Fig. 7.4A) and the 12-fold symmetrized tail portion of the second reconstruction (Fig. 7.4B). These separately segmented volumes were then combined into a single, hybrid density map (Fig. 7.4C). No attempt was made to impose any particular rotational alignment of the tail and head segmented maps in constructing the hybrid model. Regardless, the exact same hybrid model was used to produce the final virion and ghost $\phi 29$ reconstructions, which have tail structures that differ in several ways including the asymmetric arrangements of appendage conformations that break the 12-fold axial symmetry present in the hybrid model (Tang *et al.*, 2008b).

4. Asymmetric Reconstruction of the Entire $\phi 29$ Phage

A hybrid density map generated from full resolution image data was used as input to the second branch of the reconstruction scheme (Fig. 7.1), which began with two cycles of refinement carried out in EMAN. This produced an asymmetric reconstruction of the virion at ~ 30-Å resolution. At this point, the use of unbinned data and the abandonment of symmetry averaging led to no further improvement in resolution and resulted in significant computational overhead. This failure to reach higher resolutions does not imply there are limitations in the capabilities of EMAN, but more likely reflects our own lack of expertise with the software.

Our group has developed a set of programs as part of AUTO3DEM that can be run efficiently on computer clusters and other parallel computers (Yan *et al.*, 2007b). Most of the underlying programs were designed to handle cyclic, dihedral, and cubic point group symmetries, but the global search procedure implemented in PPFT is strictly only applicable to spherical particles since it requires particle images and projections of the maps to be reinterpolated into a series of concentric, circular annuli. We then turned to FREALIGN (Grigorieff, 2007) to initiate a refinement process that could be transitioned relatively easily to AUTO3DEM.

4.1. Origin and orientation determination and refinement with FREALIGN

The version of FREALIGN available at the time this work on $\phi 29$ was being performed required that the particle images be combined in a single stack file. Hence, particle images were first converted from IMAGIC to MRC format using the EMAN *proc2d* command. Also, to accommodate FREALIGN requirements, the MRC image file header was converted with the MRC program *image_convert.exe*.

The final map generated from EMAN was used as input to FREALIGN, which provides several modes of operation. We used its global search procedure (Mode 3) to assign initial origin and orientation parameters for each particle image. Also, the *matching projection* function was enabled to produce side-by-side comparisons of particle images and corresponding projections of the current reconstruction. This was carried out for the entire data set of images and, though somewhat tedious, gave a useful way to validate the assigned orientation parameters and to detect and screen out potentially "bad" particles that failed to align properly to the model.

The accuracy of the particle origins and orientations was further verified because the resultant cryo–reconstruction computed from the particle images with these assigned parameters had distinct, easily interpretable features. This was followed by standard refinement and map calculation in FREALIGN (Mode 1) for several more cycles (Fig. 7.1). The asymmetric reconstruction of the entire phage that emerged from this procedure reached a resolution of ~ 25 Å. Here, we had ample confidence in the current set of assigned origin and orientation parameters to carry out further processing steps with AUTO3DEM.

4.2. Origin and orientation refinement in AUTO3DEM

After obtaining estimates of the origins and orientations of all particles and a reliable 3D model of the entire phage with FREALIGN, we switched to AUTO3DEM for subsequent processing. AUTO3DEM runs in either serial or distributed–memory parallel mode and hence, access to a computer cluster can lead to a dramatic improvement in algorithm performance and significantly reduced computation time. The program PO^2R (Ji *et al.*, 2006) in AUTO3DEM carries out the Fourier-based origin and orientation refinement process, and program P3DR (Marinescu and Ji, 2003) computes the 3D density map from a selected ("best") set of particle images. Both programs can be run with or without symmetry constraints and hence are suitable for analysis of asymmetric particles like $\phi 29$.

FREALIGN uses an Euler angle convention (θ_e, ϕ_e, ψ_e) to specify particle orientation, whereas AUTO3DEM uses the (θ_a, ϕ_a, ω_a) convention first described by Finch and Klug (Finch and Klug, 1965) and implemented in the original icosahedral processing programs developed by Crowther (Crowther *et al.*, 1970). We employed the following relationships to convert orientation parameters from the FREALIGN convention to that used in AUTO3DEM:

$$\theta_a = 180 - \theta_e$$
$$\phi_a = \varphi_e - 180$$
$$\omega_a = \psi_e - 90$$

Once the above conversions were made, we employed the full capabilities of AUTO3DEM refinement to help improve the asymmetric reconstruction of $\phi 29$ as much as possible with the set of images that were available. AUTO3DEM works to progressively improve the resolution of any reconstruction by automatically optimizing numerous refinement parameters. For example, the images are typically band-pass filtered in Fourier space to limit the data used in comparing images to model projections. The upper Fourier limit (i.e., highest spatial frequency) included in the next cycle of refinement calculations is slowly increased as long as the resulting reconstruction shows improvement over the previous one. Improvement can be monitored in a variety of ways, but typically includes conventional Fourier shell correlation procedures (van Heel and Schatz, 2005). Other important parameters adjusted automatically by AUTO3DEM or manually by the user include the step sizes used to define the range of origin and orientation parameters to be tested. Typically, the origin and orientation intervals start out at about 1.0 pixel and 1–2°, respectively, and these are generally reduced when refinement stalls (i.e., no longer yields improvement in reconstruction resolution). At the end of refinement the origin and orientation step sizes might drop to 0.1 pixel and 0.1°, respectively. At even smaller step sizes, the process can lead to over-refinement and unreliable resolution estimates, as the noise in the data can drive refinement (Stewart and Grigorieff, 2004).

Another empirical parameter that we employed with some success is use of the inverse temperature factor (Fernandez et al., 2008; Havelka et al., 1995) to enhance high spatial frequency details in the reconstruction and during particle refinement (Tang et al., 2008a). We generally did not make use of this strategy until the reconstructed density map reached about 10–12 Å resolution, at which point an initial inverse temperature factor of $1/100$ Å^{-2} would be employed during refinement. As refinement progressed and the resolution improved, the sharpening factor would be increased progressively to about $1/400$ Å^{-2}. It proved important, if not essential, to carefully inspect the 3D density map calculated at the end of each refinement cycle to ensure that the signal–to–noise of reliably represented features (e.g., tubes of density ascribed to α-helices) did not decrease as this would signify the refinement was being driven by the dominant, high frequency noise in the data. A carefully monitored, trial and error approach and inverse temperature factors even as low as $1/1200$ Å^{-2} were used during the $\phi 29$ asymmetric refinement.

An additional strategy we used in the $\phi 29$ asymmetric reconstruction study was to mask out the genome density in the reconstructed density map, which is a method that typically improves the refinement of icosahedral viruses (Chen et al., 2009; Yan et al., 2007b; Zhang et al., 2008). This procedure leads to enhancement of reliably represented features in the projected images of the model and improves the accuracy in determining particle parameters. In a similar manner, we used with some success a density threshold procedure to remove some of the more obvious, random

noise in the reconstructed $\phi29$ density map. In density maps of icosahedral viruses computed from thousands or more particle images, the noise level outside the particle is generally quite low owing to the benefits of the 60-fold symmetry averaging that is an inherent part of the 3D reconstruction process. Hence, in an asymmetric reconstruction computed from a comparable number of particle images, the average noise level is significantly higher throughout the map. We tried a number of different masking and threshold procedures to zero noise outside the outer envelope of the phage particle. These are akin to solvent flattening used during the phase extension step in X-ray crystallographic studies (Wang, 1985). The technique that worked best in this instance involved the following steps. A second map was first recalculated at a lower resolution, typically about 20 Å. The two maps were then read into Chimera, which could take the low-resolution map and define a surface that enclosed the final, sub-nanometer resolution map. The Chimera *mask* tool was used to set values of all voxels that lie outside the defined surface to zero. This provided an additional enhancement of the signal-to-noise in the projections of the model used to correlate with the raw particle images for refinement of their parameters.

5. ANALYSIS AND INTERPRETATION OF $\phi29$ RECONSTRUCTION

Careful analysis and interpretation of a reconstruction are critical for understanding the biology of the viral system, and these also help guide the refinement process by distinguishing genuine structural features from the noise. The availability of X-ray crystallographic data for individual components is particularly valuable as it helps in validating the reliability of the cryo-reconstruction, determining the significance of various structural features, and defining the boundaries between individual viral components.

5.1. Model docking

Fitting X-ray crystal structures into cryo-TEM reconstructions to produce pseudoatomic models of macromolecular complexes has become a powerful tool in the arsenal of analysis procedures ever since it was first introduced in studies of viruses (Stewart *et al.*, 1993), virus–antibody (Smith *et al.*, 1993; Wang *et al.*, 1992) and virus–receptor (Olson *et al.*, 1993) complexes, and acto-myosin filaments (Rayment *et al.*, 1993). While useful for a wide variety of macromolecular systems, the combination of cryo-TEM and X-ray crystallographic structures can have a particularly profound impact on the study of tailed phages for two main reasons. First, owing to their shape and size, intact tailed phages resist crystallization and this will likely

continue to thwart attempts at obtaining diffraction quality single crystals of a complete asymmetric phage. Second, the head–tail symmetry mismatch precludes the imposition of any symmetry and hence limits the resolutions that can be achieved. The docking of X-ray models of individual phage components into cryo-reconstructions currently offers the best means for obtaining a pseudoatomic model of the complete phage.

If an X-ray structure of a viral component is available, it is generally a relatively straightforward procedure to dock the atomic model manually into the reconstructed density map using a variety of interactive programs such as O (Jones *et al.*, 1991), COOT (Emsley and Cowtan, 2004), and Chimera (Pettersen *et al.*, 2004). After obtaining a reasonable fit of model to density via manual procedures, one can then quantitatively refine the fit by translating and rotating the atomic model as a rigid body until the correlation coefficient between the model and density map is maximized. This can be accomplished in numerous programs such as RSREF (Chapman, 1995), SITUS (Wriggers and Birmanns, 2001), EMfit (Rossmann *et al.*, 2001), CoAn (Volkmann and Hanein, 2003), and Chimera (Pettersen *et al.*, 2004).

Rigid body docking of the ϕ29 connector (gp10) crystal structure (Guasch *et al.*, 2002) into reconstructions of the prohead, ghost, and mature virion shows that the top of the connector fits well into the cryo-TEM density in all three cases, but the lower portion of the connector only fits well into the prohead (Fig. 7.5). Ghosts and virions both contain additional proteins that were added sequentially during assembly onto the bottom of the connector to create a functional tail, and it is clear from the rigid body fits that the attachment of the tail and packaging of the genome induce conformational changes in the connector. At the time of this study (Tang *et al.*, 2008b), the existing software was not capable of modifying the crystal structure to better fit the density.

Flexible fitting methods have been developed to permit models to be modified in various ways to achieve better fits to the cryo-TEM density.
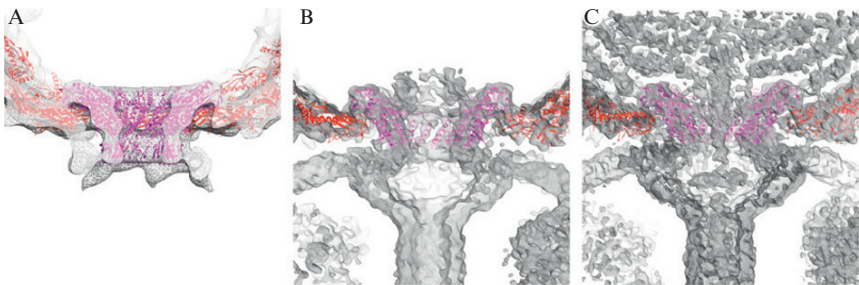

Figure 7.5 Rigid body fit of gp10 connector crystal structure (magenta ribbon model) and gp8 capsid subunit homology model (red), into ϕ29 density maps (gray). The top portion of the connector fits well into the prohead (A), ghost (B), and virion (C) reconstructions, whereas the lower portion only fits well into the prohead.

Tao, Y., Olson, N. H., Xu, W., Anderson, D. L., Rossmann, M. G., and Baker, T. S. (1998). Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell* **95,** 431–437.

Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W., and Sali, A. (2008). Protein structure fitting and refinement guided by cryo-EM density. *Structure* **16,** 295–307.

Trabuco, L. G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16,** 673–683.

van Heel, M., and Schatz, M. (2005). Fourier shell correlation threshold criteria. *J. Struct. Biol*. **151,** 250–262.

Volkmann, N., and Hanein, D. (2003). Docking of atomic models into reconstructions from electron microscopy. *Methods Enzymol*. **374,** 204–225.

Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol*. **115,** 90–112.

Wang, G. J., Porta, C., Chen, Z. G., Baker, T. S., and Johnson, J. E. (1992). Identification of a Fab interaction footprint site on an icosahedral virus by cryoelectron microscopy and X-ray crystallography. *Nature* **355,** 275–278.

Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W., and Johnson, J. E. (2000). Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289,** 2129–2133.

Wolf, M., Garcea, R. L., Grigorieff, N., and Harrison, S. C. (2010). Subunit interactions in bovine papillomavirus. *Proc. Natl. Acad. Sci. USA* **107,** 6298–6303.

Wriggers, W., and Birmanns, S. (2001). Using situs for flexible and rigid-body fitting of multiresolution single-molecule data. *J. Struct. Biol*. **133,** 193–202.

Xiang, Y., Morais, M. C., Battisti, A. J., Grimes, S., Jardine, P. J., Anderson, D. L., and Rossmann, M. G. (2006). Structural changes of bacteriophage ϕ29 upon DNA packaging and release. *EMBO J*. **25,** 5229–5239.

Yan, X., Dryden, K. A., Tang, J., and Baker, T. S. (2007a). Ab initio random model method facilitates 3D reconstruction of icosahedral particles. *J. Struct. Biol*. **157,** 211–225.

Yan, X., Sinkovits, R. S., and Baker, T. S. (2007b). AUTO3DEM–an automated and high throughput program for image reconstruction of icosahedral particles. *J. Struct. Biol*. **157,** 73–82.

Zhang, X., Jin, L., Fang, Q., Hui, W. H., and Zhou, Z. H. (2010). 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell* **141,** 472–482.

Zhang, X., Settembre, E., Xu, C., Dormitzer, P. R., Bellamy, R., Harrison, S. C., and Grigorieff, N. (2008). Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl. Acad. Sci. USA* **105,** 1867–1872.

Zhou, Z. H. (2008). Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Curr. Opin. Struct. Biol*. **18,** 218–228.