Structure, Vol. 13, 363–372, March, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2004.12.016

# Electron Cryomicroscopy of Biological Machines at Subnanometer Resolution

Review

Wah Chiu,[1,2,*] Matthew L. Baker,[1] Wen Jiang,[1] Matthew Dougherty,[1] and Michael F. Schmid[1]
[1]National Center for Macromolecular Imaging
Verna and Marrs McLean Department of Biochemistry and Molecular Biology
Baylor College of Medicine
Houston, Texas 77030

**Advances in electron cryomicroscopy (cryo-EM) have made possible the structural determination of large biological machines in the resolution range of 6–9 Å. Rice dwarf virus and the acrosomal bundle represent two distinct types of machines amenable to cryo-EM investigations at subnanometer resolutions. However, calculating the density map is only the first step, and much analysis remains to extract structural insights and the mechanism of action in these machines. This paper will review the computational and visualization methodologies necessary for analysis (structure mining) of the computed cryo-EM maps of these machines. These steps include component segmentation, averaging based on local symmetry among components, density connectivity trace, incorporation of bioinformatics analysis, and fitting of high-resolution component data, if available. The consequences of these analyses can not only identify accurately some of the secondary structure elements of the molecular components in machines but also suggest structural mechanisms related to their biological functions.**

## Introduction

Structural biology of macromolecules has become an indispensable branch of molecular biology because the results from structural studies have been able to explain the mechanism of action in macromolecules and/or lead to more targeted experiments for structure/function exploration. So far, structural approaches have been largely based on X-ray crystallography or NMR spectroscopy because these techniques can provide atomic models. However, both of these techniques are limited in their ability to tackle very large biological machines close to or greater than one million daltons. Usually, these large machines are made up of multiple copies of identical or nonidentical molecules. In some cases, the machines can consist of up to hundreds or even thousands of protein molecules. Implicitly, a machine is capable of undergoing motion and/or changes in shape as a characteristic feature of its function to carry out a specific biological process (Alberts, 1998; Alberts and Miake-Lye, 1992) such as signal transduction, genome replication, RNA transcription, translation, chaperonin-assisted protein folding, motility, and viral infection. The study of these large machines has become increasingly popular partly because of the rapid development of proteomics in identifying and purifying ensembles of macromolecules (Gavin et al., 2002; Sali, 2003).

Electron cryomicroscopy (cryo-EM) offers a number of advantages for determining the structures of large biological machines. Specifically, specimens for cryo-EM studies can be analyzed in many forms and shapes, i.e., two- or three-dimensional crystals, one-dimensional filaments or tubular crystals possessing helical symmetry, and single particles with or without symmetry. At present, cryo-EM has been mostly used for studying intact machines, usually complemented by the crystal structures of the individual components. However, cryo-EM can be the sole structural tool in the absence of high-resolution structures of its components. By imaging these machines in different functional or chemical states and characterizing the differences, one can discover conformational switches in the molecular components within the machine (e.g., Gao et al., 2003; Jiang et al., 2003; Saibil, 2000; and Subramaniam and Henderson, 2000). Furthermore, structures of the components within the machine are sometimes different from those in the isolated and crystallized forms. In such cases, cryo-EM can provide an estimate of the structures of machines in their native, interacting states that are directly relevant to their cellular functions.

## Achievable Resolution in Cryo-EM

Because the wavelength of electrons used in electron microscopy is on the order of a fraction of an angstrom, the resolution for electron imaging has never been diffraction limited. In fact, any modern commercially available instrument has a resolving power of <2.5 Å. Though electron imaging is an inherently high-resolution method, its application to biological machines is far from attaining this instrumental resolution. The problem lies in the electron radiation damage to the biological molecules that occurs when a statistically defined image is recorded (Glaeser, 1971; Henderson, 1995). All the experimental and computational methodologies developed for cryo-EM have been directly or indirectly geared to the need to cope with the constraints imposed by the radiation damage effects. To date, low-dose and low-temperature microscopy have been the most effective approaches for preserving the high resolution structural details and minimizing the damage to the biological specimens. The details in preparing cryo-EM samples and recording high-quality images have been established and are routinely used in many laboratories. Nevertheless, the raw images are still noisy and require proper averaging of images of a large number of molecules by computational means to retrieve the correct structures. So far, the machines that have been studied to subnanometer resolutions are those that can be biochemically prepared as structurally homogeneous samples. This structural homogeneity requirement may possibly be relaxed when more advanced image processing methodologies are further developed (Brink et al., 2004).

*Correspondence: wah@bcm.tmc.edu
[2]Lab address: http://ncmi.bcm.tmc.edu

Computer reconstruction of cryo-EM images includes the correction of the instrument transfer function, the assignment of the orientation parameters, and merging the data coherently to calculate a three-dimensional density map. There are numerous software packages for solving cryo-EM structures of specimens of various symmetry types. The first successful application of cryo-EM that resulted in the tracing of the polypeptide backbone was done with images of two-dimensional crystals of bacteriorhodopsin (Henderson et al., 1990), and this approach has subsequently been extended to include several membrane proteins (Kühlbrandt et al., 1994; Murata et al., 2000) as well as $Zn^{2+}$-induced tubulin crystals (Nogales et al., 1998). This type of specimen, two-dimensional crystals, is the easiest for subnanometer resolution cryo-EM analysis because the data are localized in diffraction spots and Fourier averaging methods can be easily applied to retrieve the 3D structures. The next type of specimens that are readily amenable to subnanometer resolution cryo-EM structure determination have helical symmetry, where the data are also localized, this time in the layer lines in diffraction space. Machines such as tobacco mosaic virus (Jeng et al., 1989), bacterial flagella (Yonekura et al., 2003), and acetylcholine receptor (Miyazawa et al., 2003) are examples of structures determined to 4.5 to 9 Å. The most difficult type of specimen for attaining a subnanometer resolution cryo-EM structure is the single particle, either with or without any inherent symmetry. In single particles, the data are delocalized in diffraction space and weaker because they arise from a single particle rather than a coherent average of many. It is more difficult to determine accurately the geometric orientation parameters for each particle necessary for 3D reconstruction, which is equivalent to indexing diffraction spots or layer lines in two- and one-dimensional crystal data. Icosahedral particles, typically viruses, are a subset of single particles that possess high symmetry, comprised of 60 redundant copies of asymmetric units per particle, making them the best case for achieving subnanometer resolution structure determination (for review see, Zhou and Chiu, 2003). Despite the inherent challenges, significant progress has been made in cryo-EM reconstruction of single particle with lesser or no symmetry as exemplified by GroEL (Ludtke et al., 2004), clathrin (Fotin et al., 2004), and ribosome.

It should be pointed out that the resolution of a cryo-EM map has been defined differently for different types of specimen. For crystalline and helical symmetries, the resolution is defined in a similar manner as is done in X-ray crystallography and fiber diffraction, that is, the highest resolution at which discrete signals can be detected above background in diffraction space. In the case of single particle cryo-EM specimens, the resolution is measured by comparing independent reconstructions from two sets of particle images. A common criterion used to compare the two independent maps and assess resolution is the Fourier shell correlation (van Heel, 1987) with a threshold of 0.5. However, there is some variation in picking the resolution threshold (Rosenthal and Henderson, 2003), and caution is warranted in interpreting the structural resolution of a cryo-EM map. Rather than strictly relying on such a resolution criterion, one should look for structural features that can be expected for certain resolution structures as described below.

## Structural Features at Subnanometer Resolution

Because of the residual noise and computational errors in a cryo-EM map determined at subnanometer resolution, it can be a challenge to identify structural features from a complex, large map of a machine, which can be difficult to visualize even with modern graphics hardware and software. However, these maps have a rich information content that can be discovered with proper visualization and structure mining tools, which are described in a number of papers in this volume. The resolvable features may include subunit or domain boundaries, quantitative descriptions of the spatial relationships among these components, and secondary structure elements of polypeptides and RNA molecules. In order to find out the most readily observable features from a molecular structure at this range of resolution, one can take any PDB structure, "blur" it to various resolutions, and then visualize the structure as exemplified in Figure 1. At low resolution, only the overall shape and possibly subunit boundaries are visible. In examining the synthetic data, the 50S ribosome (Ban et al., 2000), helices of rRNA are readily discernable, demonstrating the 22 Å major groove and 12 Å minor groove, at 10–14 Å resolution (Figure 1, left). Between 10 and 6 Å, long $\alpha$ helices in proteins appear as cylinders of density with diameters of 5–6 Å and variable lengths (Figure 1, middle). Also at this resolution, $\beta$ sheets appear as flat, continuous density of variable sizes. Beyond 6 Å, the density corresponding to polypeptide $\alpha$ helices may exhibit the 5.4 Å helical pitch, whereas the increasingly "bumpy" appearance of the density map may be attributed to individual bulky side chains. However, it is not until at least 4.7 Å resolution that the individual strands of sheet may be resolved (Figure 1, right).

Although it is possible to establish theoretical boundaries for structure interpretation, practical limitations may alter the ability to visually or computationally identify structural features. For instance, if a protruding $\alpha$ helix is situated at the external part of the protein and does not have any mass density adjacent to it, this $\alpha$ helix would be detectable even at resolution worse than 10 Å. However, if a helix is floppy and highly mobile, or surrounded closely by other density, it may not be recognized as a helix even at relatively high resolution. Therefore, the extent of recognizable features is really specimen and map dependent as well as resolution dependent. In this review, we have chosen rice dwarf virus and acrosomal bundle as examples to demonstrate some of the different types of structural information that can be extracted from a subnanometer resolution map.

## Two Examples of Biological Machines

Virus particles have long served as the paradigm for studying biological machines because of their complex, multicomponent architecture as well as their ability to carry out many fundamental biological processes (see chapters in a review book by Chiu and Johnson, 2003). Viruses have also driven technological develop-

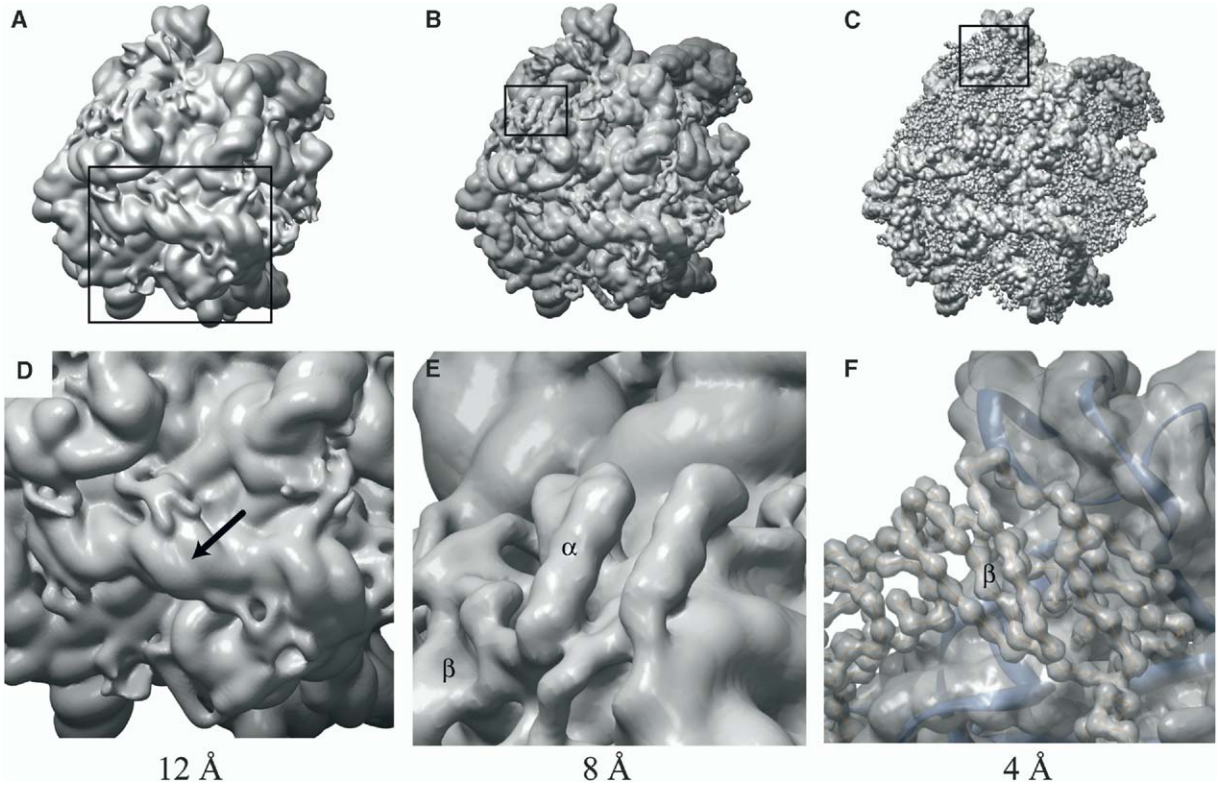


Figure 1. Structural Appearances of Ribosomal Subunit at Different Resolutions

The large ribosomal subunit from *Haloarcula marismortui* (1FFK, Cα only) (Ban et al., 2000) generated from EMAN (Ludtke et al., 1999) at different resolutions: 12 Å (left), 8 Å (middle), and 4 Å (right). Bottom insets show a zoomed-in view illustrating the salient features such as rRNA helices (arrow), α helices, and β strands (marked α and β, respectively). The images shown represent simulated data in idealized conditions (the absence of noise, particle in vacuum, no side chains, etc.) and thus may represent slightly better resolution than actual cryo-EM reconstructions. This is particularly noticeable in the two alpha helices on the surface highlighted in the 8 Å map, which are partially visible in the 12 Å map. The display was generated using UCSF Chimera (Pettersen et al., 2004). This display was generated with UCSF Chimera (Pettersen et al., 2004).

ments in molecular biology and have proved to be useful in deducing fundamental concepts in molecular assemblies (Chiu and Johnson, 2003; King and Chiu, 1997). One particularly interesting virus has been Rice dwarf virus (RDV), a member of reoviridae that infects rice plants and is an important agricultural pathogen. Two icosahedral protein shells (~600 and 700 Å in diameter) surround the 12 segment dsRNA genome. RDV, with a total mass of ~75 MDa, consists of seven structural proteins. The outer shell is primarily composed of 780 copies of P8 (46 kDa), whereas the inner shell is primarily composed of 120 copies of P3 (114 kDa). Additionally, RDV is a transcriptionally competent machine containing the viral-coded RNA-dependent RNA polymerase (Suzuki et al., 1992). When the 6.8 Å cryo-EM map of RDV was obtained (Zhou et al., 2001), no crystal structure for this virus or its components was known. Combining the subnanometer resolution map of RDV and the bioinformatics analysis of the two aforementioned viral structural proteins, it was possible to derive their folds. Subsequently, the crystal structure of RDV was solved (Nakagawa et al., 2003) and showed remarkable agreement with the models derived from the cryo-EM analysis. In short, a cryo-EM study at subnanometer resolution was shown to be able to accurately produce pseudoatomic models for the components of a machine even in the absence of a crystal structure (Zhou et al., 2001).

The acrosomal bundle from *Limulus* sperm is known to be a biological spring, which contains ~90 filaments with a diameter of 1,000 Å and a length of 60 μm (Tilney, 1975). It has rigidity similar to that of a plastic polymer such as polyethylene, with a stiffness (Young's Modulus) of 2 GPa (Gardel et al., 2004). Compositionally, it has an equal stoichiometry of actin, scruin, and calmodulin. It exists as a quasicrystal, obeying the symmetry of space group of $P2_1$ and having a unit cell spacing of 146 × 146 × 765 Å (Sherman et al., 1999). The bundle is coiled in the base of the resting sperm cell of *Limulus* and becomes extended upon activation by $Ca^{2+}$, whereupon it can penetrate through the jelly coat of an egg. Throughout this extension process, the bundle uses stored energy and does not consume ATP. There are many interesting questions inherent in this machine, including how the actin and scruin interact to form a stable large filament and how the bundle can undergo such a large conformational change during the fertilization process. The 9.5 Å cryo-EM structure (Schmid et

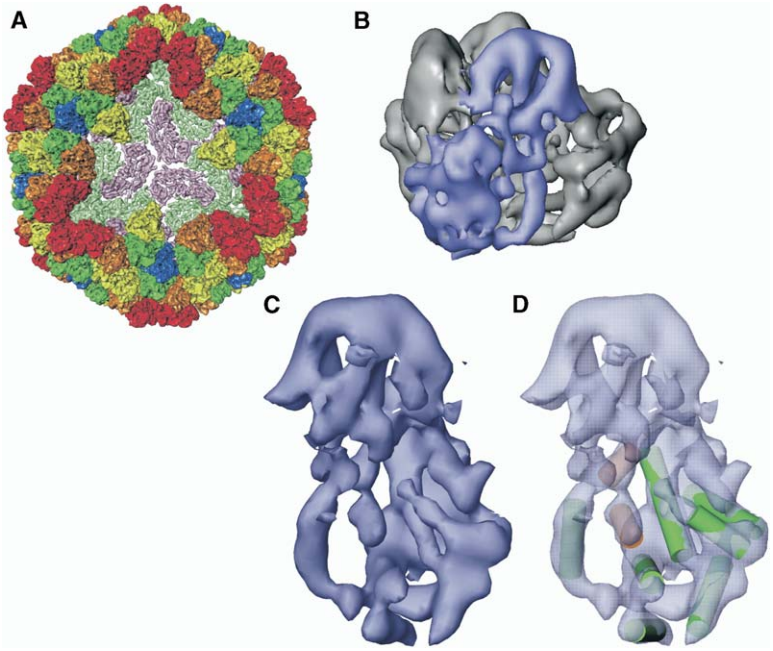


Figure 2. 6.8 Å Cryo-EM Structure of the Double-Shelled Rice Dwarf Virus

(A) Cryo-EM map of rice dwarf virus (RDV) at 6.8 Å (EBI 1060) (Zhou et al., 2001). Center of the map is cut open to show the density of the inner capsid shell. Each of the five trimers in the outer shell is colored differently.

(B) An averaged P8 trimer from the outer capsid of RDV.

(C) A single P8 subunit from the averaged trimer.

(D) P8 with annotated alpha helices. Green cylinders represent high scoring helices and orange cylinders represent lower scoring, less confident helices from *helixhunter* (Jiang et al., 2001).

al., 2004) revealed for the first time how actin molecules are packed in a partially distorted helical filament within the bundle that can be triggered to exert a tremendous force to penetrate through the jelly coat and plasma membrane of an egg.

### Analysis of Density Maps of Large Machines

A cryo-EM map of a biological machine is usually large and noisy, as exemplified by the density map of RDV, European Bioinformatics Institute (EBI) accession number EBI-1060, and of the acrosomal bundle (EBI-1088), which were reconstructed to 6.8 and 9.5 Å, respectively (Figures 2A and 3A and Movies S1 and S2 available with this article online). After such maps were obtained, a significant amount of analysis was carried out, including denoising, segmentation of the individual subunits in the asymmetric unit, averaging redundant copies of the chemically identical subunits within the asymmetric unit and establishing the spatial relationships among the subunits, and searching for secondary

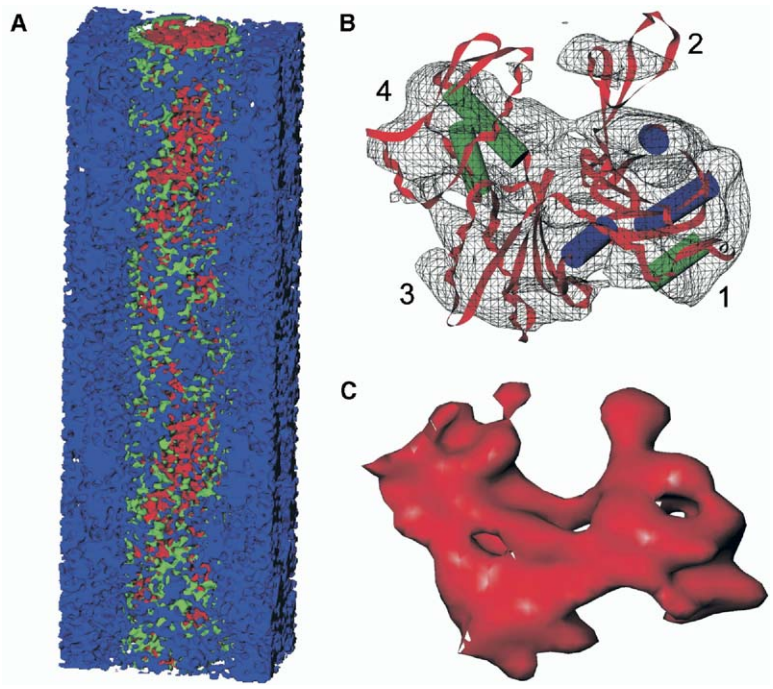


Figure 3. 9.5 Å Cryo-EM Structure of the Acrosomal Bundle from *Limulus* Sperm

(A) Cryo-EM map of acrosomal bundle at 9.5 Å (EBI-1088) (Schmid et al., 2004). A full unit cell is shown in a centered view. Thus, there is one filament shown radially colored plus one quarter section of each of four identical filaments at the corners.

(B) Final averaged actin subunit with four subdomains (labeled 1–4) after *foldhunter* search and local symmetry averaging with annotated helices (blue ones found by *helixhunter*, green ones recognized after *foldhunter* averaging).

(C) Coordinates used by Holmes for the F-actin model blurred to equivalent resolution and nonisotropic B factor (effective resolution is thus 9 Å in the vertical direction, about 12 Å in the horizontal plane).

structure elements. We will describe the different approaches performed on these density maps.

### Segmentation and Postmap Averaging

The first task in understanding a multicomponent machine is to define the individual molecular components. In many cases, the individual components may actually form higher-ordered structures within the complex, such as an asymmetric unit in a virus. Defining the boundary of each subunit can be a challenge, both from the standpoint of resolution and resolvability. At higher resolutions, it is easier to define the molecular boundary of the components. However, intimate interactions between the neighbor molecules and the residual noise may result in poor resolvability of the components. In segmenting the individual components of a machine, both computational and manual routines can be employed to varying levels of success. Segmentation of the individual components may also be facilitated by the availability of the atomic resolution structure of the component or of a homologous protein.

Often, the asymmetric unit of the machine consists of multiple copies of the same polypeptide. There is no guarantee that each of the identical polypeptides will have identical structures. If the subunits are related to each other via local symmetry operations and if those symmetry operations are not assumed in the reconstruction procedure, one may ask quantitatively how closely those subunits follow the symmetry relationship (He et al., 2001). In such circumstances where individual components share a high degree of similarity, inter- and/or intrasubunit averaging may result in a better-resolved component.

It should be pointed out that averaging is very useful to reveal the structurally invariant parts of the subunit but may blur out the structurally variable regions, which can be critical at the intermolecular interfaces. Additionally, the extent of the symmetry agreement can be a function of resolution. Specifically, a subunit may appear very symmetrical around a local symmetry axis at low resolution but will be less so at a higher resolution. As suggested decades ago, the quasiequivalence of a viral assembly would cause their symmetry-related subunits to have similar, but not identical structures in order to form a stable and closed particle (Caspar and Klug, 1962). In determining the symmetry axis and symmetry agreement of a subunit, further analysis of subunit locations, segmentation, and alignment of the individual subunits may result not only in an averaged subunit but also a physical description of subunit location, orientation, and interactions. The process of segmentation and averaging is iterative and can best be illustrated through two examples.

As mentioned previously, no crystal structure for RDV was available when the cryo-EM map was obtained (Zhou et al., 2001). However, the map was well resolved and allowed for the preliminary segmentation and subsequent assessment of the P8 trimers in an asymmetric unit. In establishing the symmetry axis and relationship, it was then possible to align the different types of trimers, resulting in an averaged trimer (Figure 2B). This trimer was then used as a template to find and extract all P8 trimer locations. Because of the local 3-fold symmetry and the enhanced signal from the averaged trimer, it was also possible to segment a single P8 monomer from the trimer (Figure 2C and Movie S3).

The acrosomal bundle is a $P2_1$ crystal with 14 subunits of actin, scruin, and calmodulin in each asymmetric unit. It has been shown previously that the scruin wraps around each actin filament, and scruin-scruin interfilament interactions are critical for crosslinking the filaments into the bundle (Schmid et al., 1994). Based on this knowledge, a single actin monomer was segmented and used as a template to search for the rest of the actin in the asymmetric unit. The remaining subunits were then segmented out and averaged to produce an averaged actin monomer. Because the first round of segmentation of the single actin subunit was approximate and did not use any crystal structure of the actin to avoid bias in the analysis, a number of iterations of segmentation and averaging were used in order to obtain a more accurate segmentation of the subunit resulting in an averaged actin subunit (Figure 3B). This procedure was also carried out with the scruin molecule, which resolved into two domains as was predicted by sequence analysis.

### Identifying Secondary Structure Elements from Cryo-EM Maps

As shown in Figure 1, secondary structure features of molecular components can be detected in structures determined at subnanometer resolution. Alpha helices have a characteristic cylindrical shape and can be recognized visually when appropriate display thresholds are selected. Alternatively, a computational search with an $\alpha$ helix model of several turns in length can be used as a template to probe for cylindrical density throughout the entire map of the machine. Such an automatic computational approach has been confirmed by simulation to be >90% accurate in detecting helices of three or more turns (Jiang et al., 2001). Although such a procedure is relatively robust and accurate, helix assignment should ultimately be assessed through visual inspection and comparison with known biochemical or bioinformatics data. The identification of helices is not only useful in describing the fold of a component but also can be an important method for identifying homologous structures. By providing a geometric description of helix locations and orientations, it is possible to query the library of known folds, derived from the PDB, to identify potentially homologous structures based solely on $\alpha$ helices (Baker et al., 2003; Jiang et al., 2003; Zhou et al., 2001).

Figure 2D shows the detected $\alpha$ helices in the averaged outer capsid shell protein, P8, of RDV. Nine helices were clearly visible in the lower domain of P8; however, the connections between these helices were less visible. Based on sequence and density analysis, a topological model for RDV P8 was proposed (detailed in a subsequent section). Once the crystal structure of RDV was subsequently determined (Nakagawa et al., 2003), it was then possible to validate the helix assignment. The locations of all nine $\alpha$ helices with three or more helical turns were correctly assigned in the cryo-

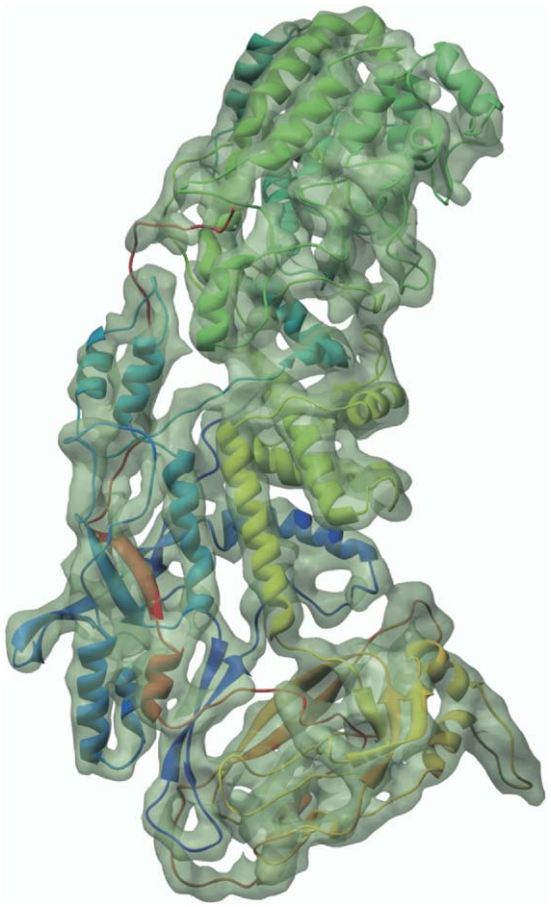

Figure 4. Cryo-EM Map of the Inner Capsid Shell Protein of Rice Dwarf Virus Determined by EMAN Algorithm

Cryo-EM map of inner capsid shell protein (P3) of RDV reconstructed with the EMAN software (Ludtke et al., 2004) and its superimposition with the crystal structure (Nakagawa et al., 2003). This demonstrates the matches of the cryo-EM density and the crystal structure in corresponding helices, sheets, and loops.

EM map (Figure 2D) to within 2 Å helix centroid rms deviation.

Figure 3B displays another example where α helices were seen in the averaged actin subunit of the acrosomal bundle at lower resolution. Compared with the actin subunit model derived from crystallography and fiber diffraction (Holmes et al., 1990) blurred to the equivalent resolution with a nonisotropic B factor (Figure 3C), its overall shape matches that of the averaged map. In the averaged map, only six α helices could be identified. The other five α helices of actin are not resolvable in this map, probably because of a number of factors including limited resolution, nonisotropic B factor, and possible postmap averaging errors.

In addition to α helices, it is feasible to detect the structural signature of β sheet as shown in Figure 1 (middle). As mentioned before, α helices are relatively cylindrical, whereas β sheets are more planar at subnanometer resolutions. However, sheets may adopt a variety of twists and bends and may have vastly different shapes and size (Kong and Ma, 2003). Such variations make both manual and automated approaches difficult. As RDV was at sufficiently high resolution, it was possible to visually assess the upper domain as potentially β sheet (Figure 2C). This assignment was later confirmed by the crystal structure of RDV (Figure 6A).

## Toward Full Structural Models

After describing the helices and sheets, the next major challenge in deriving an atomic model for the components of biological machines is the ability to assign the connectivity of the secondary structure elements in the map. Figure 4 shows the segmented density of the inner capsid protein of RDV from a recent density map obtained with a different computational algorithm (W.J., unpublished data). Not only are the helices and sheets visible but also many of the loop densities are well represented in the cryo-EM density map, as illustrated when the RDV crystal structure is superimposed (Figure 4). Of course, these features cannot be interpreted with certainty at this resolution unless the crystal structure of its component or a homolog is known.

In the study of RDV, the interpretation of the cryo-EM map was combined with other bioinformatics approaches to derive the connectivity of the secondary structure elements. Multiple secondary structure prediction algorithms were used to produce a consensus assignment of α helices and β sheets. The lengths of the helices, as determined from the cryo-EM map, were then matched to the predicted helices from the consensus of the secondary structure predictions. Finally, by analyzing the connecting density and the proximity of the different helix segments, a model for the lower domain of the P8 was derived (Figure 2D). The resulting assignment, when compared to the crystal structure, illustrated the correct assignment of helices and connectivity in all but two small, tentatively assigned helices, for which the order was swapped (Figure 6A).

## Fitting Atomic Data

Although cryo-EM structures can yield significant information about the fold of the individual components of a biological machine, it is still advantageous to fit an X-ray or NMR structure to the complex. Such fitting may reveal much of the same type of information as discussed above, i.e., subunit location and orientation, subunit contacts, and location. However, it is possible that the structure of the component may be different when complexed with the other components. There are numerous available algorithms for both rigid body and flexible fitting of components into the larger structure (Gao et al., 2003; Jiang et al., 2001; Volkmann and Hanein, 1999; Wriggers and Birmanns, 2001).

In the case of acrosomal bundle, when the actin crystal structure was compared with the averaged actin subunit, it was found that the separation between half of the actin (subdomains 1 and 2) with respect to the other half (subdomains 3 and 4) is smaller than that of the existing crystal structure. This suggests that the actin in the bundle corresponds to a nucleotide bound state (Figure 3B). This finding is consistent with the biochemical evidence that ADP is present in the bundle (Schmid et al., 2004). The current analysis has yet to yield an atomic model of the molecule and requires fur-

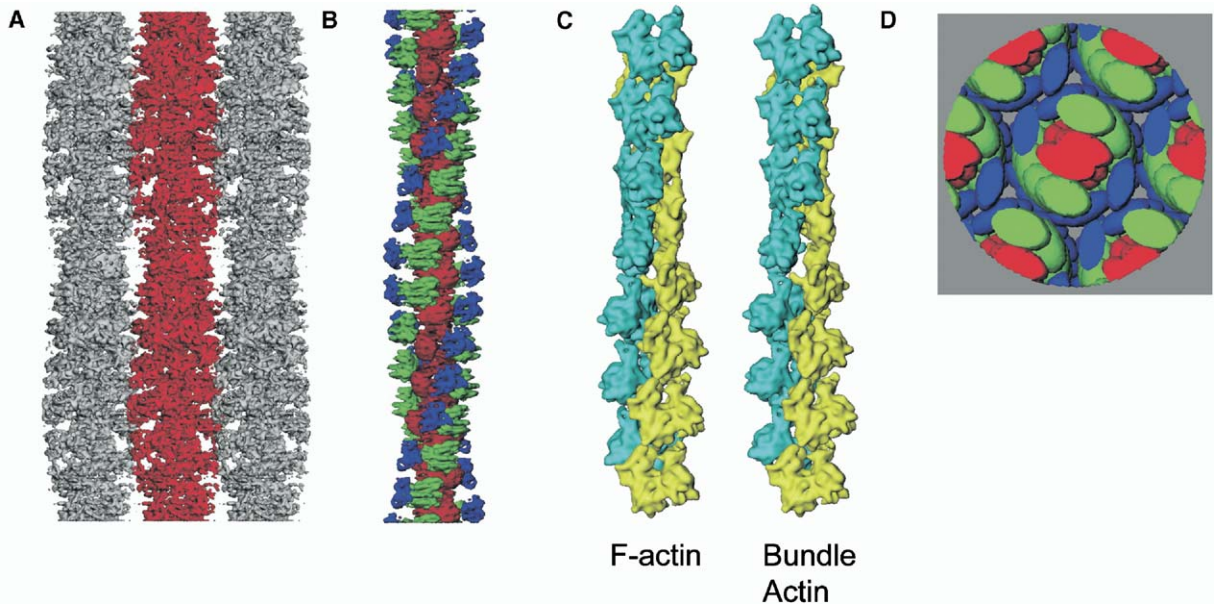


Figure 5. Molecular Interactions among Adjacent Filaments in the Acrosomal Bundle

(A) Three adjacent filaments in the bundle, masked out without averaging from the raw map (Figure 3).

(B) Correlation average of the actin (in red) and the two domains of scruin, S domain (in blue) and E domain (in green), placed back into their proper locations in the unit cell of the bundle.

(C) Actin filament model in the bundle. Left, the perfect F-actin helix, shown for comparison; right, the actin conformations in the asymmetric unit of the acrosomal bundle.

(D) Schematic diagram of the molecular interactions in the filaments. One central and six neighboring filaments are shown, with the centroids of the actin (red) and scruin subdomains (blue and green), looking down the bundle axis. The blue domains are responsible for most or all of the interfilament interactions.

ther refinement to incorporate the variable conformation of each of the actin molecules in the final model.

**Structural Mechanism Based on Cryo-EM Structures**

The purpose of studying an entire machine is to understand the relationship between its parts as it affects the function of the machine. The structure of the acrosomal bundle illustrates nicely the valuable information that can be derived from the cryo-EM structure analysis. Figure 5A shows the packing of the actin-scruin complex within the bundle in the unaveraged map, whereas Figure 5B shows the refined model of a single filament of the bundle (Schmid et al., 2004), which is different from the standard F-actin (Holmes et al., 1990). This deviation of actin from exact helical symmetry (Figure 5C and Movie S4) reveals the detailed "quaternary structure" arrangement of the actin molecules along the filament, presumably driven by their interactions with their associated scruin molecules, which in turn have to interact with the scruin molecules of the adjacent filaments to maintain a stable bundle. Figure 5A shows a side view of three adjacent filaments of the acrosomal bundle that can only begin to suggest the complexity of interactions in the bundle crystal, involving six neighboring filaments around each filament in the crystal. Figure 5D shows that one of the two domains of the scruin molecule (represented in green) interacts more closely with the actin, and so its position conforms closely to the distortions seen in the actin filament (Movie S5). The other domain (shown in blue) appears to interact more closely with the scruin of the neighboring filament, and thus it occupies the locations needed for crystal packing. This sort of division of labor is reminiscent of the different functions and interactions that are displayed, for instance, by the DNA binding domains and the activation domains of transcriptional coactivators (Remenyi et al., 2004) that are both necessary for the activity of such machines.

Throughout the asymmetric unit, there are variations in the detailed interactions for each of the constituent molecules, which show their ability to engage in different types of interactions, even with the same molecules, and hints at the flexibility that must be present for structural conformation transition in the different physiological states of the bundle. This type of protein-protein interaction information can only be obtained through direct structural determination of the intact machine. Based on this observation, the transformation of the acrosomal bundle from coiled to straight form would likely involve a specific but small variation in actin orientations along the filament (Movie S4). That is to say, distortion of the filament is necessary to allow this to happen. A similar mechanism may be adopted in the actin filaments involved in a variety of cellular functions with other actin binding proteins.

Although it is not yet possible to address the transcriptional mechanism in RDV because the structure of the polymerase complex is not revealed in this analysis, it is possible to come to a better understanding of the capsid stability and assembly. Despite the lack of se-

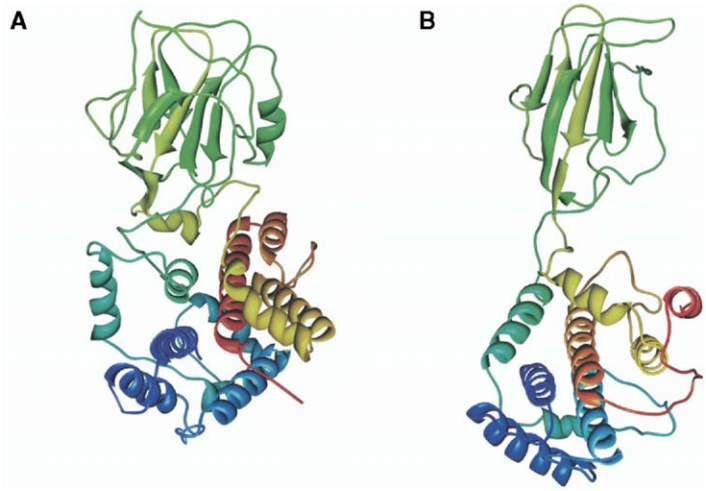


Figure 6. Crystal Structures of Rice Dwarf Virus and Bluetongue Virus

Crystal structures of the outer capsid protein, P8 (A) of RDV (Nakagawa et al., 2003) and the corresponding bluetongue virus structural protein (B) (Grimes et al., 1998). The relative spatial orientation between the upper and lower domains is different in these two proteins though they have similar fold.

quence similarity, it was shown that RDV (Figure 6A) is structurally similar to other double-shelled capsids from nonturreted reoviruses (Figure 6B). In the double shelled capsids of reoviridae, a symmetry mismatch between the two shells is used to establish capsid stability and provide a portal for RNA extrusion about the 5-fold axes. In essence, the two capsid layers provide a mechanism for protecting the genomic RNA through a series of polymorphic interactions between capsid layers while maintaining a location about the 5-fold axes for RNA transcripts to be extruded. Furthermore, much of the topology of the individual structural proteins is the same among the reoviridae (Figure 6). However, in RDV P8, the orientation of the upper and lower domain is twisted with respect to the corresponding bluetongue virus capsid protein (Grimes et al., 1998). This difference may in fact confer physiological differences in host range and possible arrangement of additional capsid proteins.

## Remaining Challenges

Given the progress in structural genomics, the number of known folds will keep increasing (Sali, 1998). Machines under investigation will therefore likely have the structures of some of their components partially or entirely known. However, the structures of components can differ between the isolated state and the intact machine. As alluded to earlier, objective methods for "morphing" the crystal structure with constraints based on the cryo-EM density maps and stereochemistry will be a critical step in the structure analysis. The acrosomal bundle demonstrates the need of such tools for cryo-EM. In addition to flexible fitting, tools for homology modeling of related components are also becoming increasingly important in the analysis of subnanometer resolution machines. Recent work has laid the groundwork for using cryo-EM density maps to build and assess homology models based on low sequence similarity and alternative sequence alignments (Topf et al., 2005). However, there is much work ahead to develop additional computational tools to build reliable atomic models.

The examples used here illustrate the application of global and local symmetry during the data analysis. Global symmetry such as crystallographic, helical, icosahedral, and rotational symmetry relating the spatial relationship of the asymmetric units is used and imposed in creating the reconstruction. Local symmetry is found within the asymmetric unit, not assumed in the reconstruction, and is usually discovered after the map is obtained. Although care must be taken in applying local symmetry, its proper evaluation and use was able to increase the signal-to-noise ratio of the averaged map and, thus, the ability to identify the invariant parts of molecular machines and precisely how they fit into the structure of the machine as a whole. Many biological machines will lack any global or local symmetry. In such case, the amount of data has to be dramatically increased in order to compensate for the nonredundant arrangement of the molecular components in the machine.

The methodology shown here suggests that subnanometer resolution density maps of biological machines can be mined to extract new structural information. Many of the steps used in the analyses were done manually or visually. Current research is being carried out in some laboratories to automate these steps to facilitate the throughput and eliminate subjectivity in the estimates. Cryo-EM is gradually emerging as a mainstream tool in structural determination of biological machines of great interest in proteomic research.

### Supplemental Data

Supplemental Data include five movies and can be found with this article online at http://www.structure.org/cgi/content/full/13/3/363/DC1/.

### Acknowledgments

We would like to thank Paul Matsudaira at MIT and Z. Hong Zhou at the University of Texas Houston Medical School for their collaborations on the biological examples cited in this paper. This research has been supported by grants from the National Institutes of Health and National Science Foundation (P41RR02250, R01GM070557,

R01AI38469, P20RR020647, and EIA-0325004). M.L.B. is supported by the training grant (5T15 LM07093) through the Keck Center for Computational Biology of the Gulf Coast Consortia.

Received: December 3, 2004
Revised: December 22, 2004
Accepted: December 25, 2004
Published: March 8, 2005

## References

Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell *92*, 291–294.

Alberts, B., and Miake-Lye, R. (1992). Unscrambling the puzzle of biological machines: the importance of the details. Cell *68*, 415–420.

Baker, M.L., Jiang, W., Bowman, B.R., Zhou, Z.H., Quiocho, F.A., Rixon, F.J., and Chiu, W. (2003). Architecture of the herpes simplex virus major capsid protein derived from structural bioinformatics. J. Mol. Biol. *331*, 447–456.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science *289*, 905–920.

Brink, J., Ludtke, S.J., Kong, Y., Wakil, S.J., Ma, J., and Chiu, W. (2004). Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. Structure *12*, 185–191.

Caspar, D., and Klug, A. (1962). Physical principles in the construction of regular viruses. Cold Spring Harb. Symp. Quant. Biol. *27*, 1–24.

Chiu, W., and Johnson, J.E. (2003). Virus structure. Preface. Adv. Protein Chem. *64*, xi–xii.

Fotin, A., Cheng, Y., Grigorieff, N., Walz, T., Harrison, S.C., and Kirchhausen, T. (2004). Structure of an auxilin-bound clathrin coat and its implications for the mechanism of uncoating. Nature *432*, 649–653.

Gao, H., Sengupta, J., Valle, M., Korostelev, A., Eswar, N., Stagg, S.M., Van Roey, P., Agrawal, R.K., Harvey, S.C., Sali, A., et al. (2003). Study of the structural dynamics of the *E. coli* 70S ribosome using real-space refinement. Cell *113*, 789–801.

Gardel, M.L., Shin, J.H., MacKintosh, F.C., Mahadevan, L., Matsudaira, P., and Weitz, D.A. (2004). Elastic behavior of cross-linked and bundled actin networks. Science *304*, 1301–1305.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141–147.

Glaeser, R.M. (1971). Limitations to significant information in biological electron microscopy as a result of radiation damage. J. Ultrastruct. Res. *36*, 466–482.

Grimes, J.M., Burroughs, J.N., Gouet, P., Diprose, J.M., Malby, R., Zientara, S., Mertens, P.P.C., and Stuart, D.I. (1998). The atomic structure of the bluetongue virus core. Nature *395*, 470–477.

He, J., Schmid, M.F., Zhou, Z.H., Rixon, F., and Chiu, W. (2001). Finding and using local symmetry in identifying lower domain movements in hexon subunits of the herpes simplex virus type 1 B capsid. J. Mol. Biol. *309*, 903–914.

Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Q. Rev. Biophys. *28*, 171–193.

Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E., and Downing, K.H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. J. Mol. Biol. *213*, 899–929.

Holmes, K.C., Popp, D., Gebhard, W., and Kabsch, W. (1990). Atomic model of the actin filament. Nature *347*, 44–49.

Jeng, T.W., Crowther, R.A., Stubbs, G., and Chiu, W. (1989). Visualization of alpha-helices in tobacco mosaic virus by cryo-electron microscopy. J. Mol. Biol. *205*, 251–257.

Jiang, W., Baker, M.L., Ludtke, S.J., and Chiu, W. (2001). Bridging the information gap: computational tools for intermediate resolution structure interpretation. J. Mol. Biol. *308*, 1033–1044.

Jiang, W., Li, Z., Zhang, Z., Baker, M.L., Prevelige, P.E., Jr., and Chiu, W. (2003). Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. Nat. Struct. Biol. *10*, 131–135.

King, J., and Chiu, W. (1997). The procapsid to capsid transition in double-stranded DNA bacteriophages. In Structural Biology of Viruses, W. Chiu, R.M. Burnett, and R. L. Garcea, eds. (New York: Oxford Press), pp. 288–311.

Kong, Y., and Ma, J. (2003). A structural-informatics approach for mining beta-sheets: locating sheets in intermediate-resolution density maps. J. Mol. Biol. *332*, 399–413.

Kühlbrandt, W., Wang, D.N., and Fujiyoshi, Y. (1994). Atomic model of plant light-harvesting complex by electron crystallography. Nature *367*, 614–621.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: Semi-automated software for high resolution single particle reconstructions. J. Struct. Biol. *128*, 82–97.

Ludtke, S.J., Chen, D.H., Song, J.L., Chuang, D.T., and Chiu, W. (2004). Seeing GroEL at 6 Å resolution by single particle electron cryomicroscopy. Structure *12*, 1129–1136.

Miyazawa, A., Fujiyoshi, Y., and Unwin, N. (2003). Structure and gating mechanism of the acetylcholine receptor pore. Nature *424*, 949–955.

Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J.B., Engel, A., and Fujiyoshi, Y. (2000). Structural determinants of water permeation through aquaporin-1. Nature *407*, 599–605.

Nakagawa, A., Miyazaki, N., Taka, J., Naitow, H., Ogawa, A., Fujimoto, Z., Mizuno, H., Higashi, T., Watanabe, Y., Omura, T., et al. (2003). The atomic structure of rice dwarf virus reveals the self-assembly mechanism of component proteins. Structure *11*, 1227–1238.

Nogales, E., Wolf, S.G., and Downing, K.H. (1998). Structure of the alpha beta tubulin dimer by electron crystallography. Nature *391*, 199–203.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Remenyi, A., Scholer, H.R., and Wilmanns, M. (2004). Combinatorial control of gene expression. Nat. Struct. Mol. Biol. *11*, 812–815.

Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. *333*, 721–745.

Saibil, H.R. (2000). Conformational changes studied by cryo-electron microscopy. Nat. Struct. Biol. *7*, 711–714.

Sali, A. (1998). 100,000 protein structures for the biologist. Nat. Struct. Biol. *5*, 1029–1032.

Sali, A. (2003). NIH workshop on structural proteomics of biological complexes. Structure *11*, 1043–1047.

Schmid, M.F., Agris, J.M., Jakana, J., Matsudaira, P., and Chiu, W. (1994). Three-dimensional structure of a single filament in the Limulus acrosomal bundle: scruin binds to homologous helix-loop-beta motifs in actin. J. Cell Biol. *124*, 341–350.

Schmid, M.F., Sherman, M.B., Matsudaira, P., and Chiu, W. (2004). Structure of the acrosomal bundle. Nature *431*, 104–107.

Sherman, M.B., Jakana, J., Sun, S., Matsudaira, P., Chiu, W., and Schmid, M.F. (1999). The three-dimensional structure of the Limulus acrosomal process: a dynamic actin bundle. J. Mol. Biol. *294*, 139–149.

Subramaniam, S., and Henderson, R. (2000). Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. Nature *406*, 653–657.

Suzuki, N., Tanimura, M., Watanabe, Y., Kusano, T., Kitagawa, Y.,

Suda, N., Kudo, H., Uyeda, I., and Shikata, E. (1992). Molecular analysis of rice dwarf phytoreovirus segment S1: interviral homology of the putative RNA-dependent RNA polymerase between plant- and animal-infecting reoviruses. Virology *190*, 240–247.

Tilney, L.G. (1975). Actin filaments in the acrosomal reaction of Limulus sperm. J. Cell Biol. *64*, 289–310.

Topf, M., Baker, M.L., John, B., Chiu, W., and Sali, A. (2005). Structural modeling of components in protein assemblies by comparative modeling and electron cryo-microscopy. J. Struct. Biol. *149*, 191–203.

van Heel, M. (1987). Similarity measures between images. Ultramicroscopy *21*, 95–100.

Volkmann, N., and Hanein, D. (1999). Quantitative fitting of atomic models into observed densities derived by electron microscopy. J. Struct. Biol. *125*, 176–184.

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.

Yonekura, K., Maki-Yonekura, S., and Namba, K. (2003). Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. Nature *424*, 643–650.

Zhou, Z.H., and Chiu, W. (2003). Determination of icosahedral virus structures by electron cryomicroscopy at subnanometer resolution. Adv. Protein Chem. *64*, 93–124.

Zhou, Z.H., Baker, M.L., Jiang, W., Dougherty, M., Jakana, J., Dong, G., Lu, G., and Chiu, W. (2001). Electron cryomicroscopy and bioinformatics suggest protein fold models for rice dwarf virus. Nat. Struct. Biol. *8*, 868–873.