THREE-DIMENSIONAL RECONSTRUCTION OF MACROMOLECULAR ASSEMBLIES USING TRANSMISSION ELECTRON MICROSCOPY AND 3D RECONSTRUCTION SOFTWARES

PREETHI CHANDRAMOULI

(M.Sc Physics, Madras University)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2004

Acknowledgements:

I am deeply indebted to Professor.Hew Choy Leong, Head of the Department, Biological Sciences, National University of Singapore (NUS), for his continued guidance, support and encouragement throughout my Masters, without which it might not have been possible to complete my Masters successfully.

I am extremely thankful to my supervisor at NUS, Dr.Terje Dokland, who was the Principal Investigator, Institute of Molecular and Cell Biology and is currently heading to the Department of Microbiology, University of Alabama, for his supervision and guidance that has helped me to learn so much during my Masters and to complete it satisfactorily.

I am very grateful to Dr.Christopher Akey, Associate Professor of Physiology and Biophysics, Department of Physiology and Biophysics, Boston University, for giving me a chance to work in his lab at Boston University and to help me get an understanding of the working of the ribosome channel complex.

I am very much indebted to Dr. Jean-François Ménétret, Research associate at Dr.Christopher Akey's lab, Department of Physiology and Biophysics, Boston University for taking time to teach me the proper usage of many of the softwares that I used extensively in the course of my Masters work and for helping me understand the working of the ribosome channel complex. His dedication has been a source of inspiration to me.

I am extremely obliged to Ms.Sifang Wang, Research officer in Dr.Terje Dokland's lab, IMCB, for doing the production and purification of bacteriophage P4 procapsid and expanded capsids without which my part of the work wouldn't have been possible.

I am thankful to Dr.Steve Ludtke, Instructor & Co-director of the National Center for Macromolecular Imaging, Baylor College of Medicine, for answering my queries regarding the software EMAN which he designed and which I used extensively for 3D reconstruction in my thesis work.

I am thankful to Ms.Ildikó V. Akey, laboratory supervisor/technician, Dr.Christopher Akey's lab, Department of Physiology and Biophysics, Boston University for helping me get settled into the lab at Boston University.

I am grateful to Dr.Danny Doan, Research fellow in Dr.Terje Dokland's, IMCB, lab for his help initially, in the use of EMAN and in clearing my doubts with regards to some aspects of Linux.

I am obliged to Mr.Rathinam Kumaresan, lab technologist in Dr.Terje Dokland's, IMCB for helping me with the Microscope

I am very thankful to Dr.Kunchithapadam Swaminathan, Assistant Professor, Department of Biological Sciences, NUS for being a source of inspiration to me and for taking time in offering me his guidance.

I am very thankful to my faculty at Department of Biological Sciences, NUS for helping me understand the various nuances of biology and making it an enjoyable experience in the process.

I am much obliged to Ms.Choo Beng Goon Joan, Management support officer and all the other members of Management, Department of Biological Sciences, NUS for promptly helping me with any doubts or problems on the administrative side.

I would like to thank my student colleagues at Department of Biological Sciences, NUS, for making the class a simulating and fun environment.

I wish to thank all my lab members in Dr.Terje Dokland's lab, Dr.Swaminathan's lab and Dr.Christopher Akey's lab for giving me such a good and simulating atmosphere to work in.

Lastly and most importantly, I wish to thank my parents, husband and sister for their support and guidance that has made it possible for me to be where I am.

Table of Contents:

Summary:

Structure based interpretation of biological systems has taken a major step forward in this decade with the advent of superior computational resources and image reconstruction software available to biologists. The basic structures can be solved by three important methods, X-ray crystallography, Nuclear magnetic resonance spectroscopy (NMR) and Electron microscopy (EM). Transmission EM is a direct way of looking at macromolecular structure, and has reached near-atomic resolution for macromolecular structures recently (7,9). Cryo-TEM can be used for looking at the structures of large complexes and relatively short-lived intermediates too.

Study of protein assembly is important towards studying the functional aspect of proteins. The simplest system that could be used for this purpose is a virus. A virus usually consists of repeating units of the same protein thereby decreasing the complexity of the system. The correct assembly of proteins is governed by a number of features like protein-protein interaction, genetic packaging and conformational switches. The structural changes involved during the assembly gives an idea about these control mechanisms. Another important aspect of structure to function annotation is studying complexes of proteins using cryo-EM and fitting their crystal structures, if known, into these maps in order to understand the larger framework in which proteins work.

We have determined the three dimensional structures of a dsDNA phage, the P4 procapsid and expanded capsid at ~14Å and ~27Å respectively using cryo-EM techniques and 3D reconstruction software. We have determined the structural changes accompanying the transition from the procapsid to expanded capsid as a result of cleavage of the capsid protein of P4, gpN. The expanded capsid has a flattened capsomer

unlike the procapsid structure and more like the mature virion. However, the expanded structure is more rounded and not as angular as the mature virion. The crystal structure of another capsid protein gp5 from HK97 phage, belonging to the dsDNA phage family found to have some amount of secondary structure similarity to the P4 capsid protein gpN, has been docked into the P4 procapsid EM model to get an understanding of the packing of the capsid protein at the atomic level, in the capsomers.

Another part of my thesis work comprises the study of the protein conducting channel in the endoplasmic reticulum (ER) membrane consisting of the heterotrimeric sec61p complex and the Translocon-associated protein (TRAP), with ribosome, by adding more data to the existing data set and getting an improved map for the ribosome channel complex. The resolution of the present improved structure is ~20 Å and we are in the process of improving it to resolutions sufficient to localize the secondary structural elements.

Working in areas as varied as viruses and ribosome channel complexes has given me an interesting and illuminating picture of the importance of cryoEM and 3D reconstruction techniques in understanding the intricacies of the molecular machinery in large macromolecular assemblies.

1. Introduction:

1.1 Electron microscopy

1.1.1 Transmission electron microscopy

Transmission electron microscopy (TEM) is a very direct way of determining macromolecular structure, which has reached near-atomic resolution for protein structures recently. In **TEM**, the electrons from the electron beam get scattered from the sample and are recorded either using a fluorescent screen or a CCD camera or a film. The energy of the electrons is typically between 100keV and 300keV, sometimes 400kV, very exceptionally 1MeV. The whole specimen is imaged at the same focus, and the recorded image is a 2-dimensional projection of the 3D electron scattering density of the specimen. It is therefore possible to reconstruct the 3D structure by recording projections of the object from different view directions and putting them together to get the 3D structure. The samples need to be thin, or they will absorb too much of the electron beam. Most biological EM work is done on small (several millimeters) copper discs called grids cast with a fine mesh. The grids are first layered with plastic then carbon is coated on top by evaporating carbon graphite or carbon is evaporated onto mica then placed on the grid. This mesh can vary a lot depending on the intended application, but is usually about 15 squares per millimeter (400 squares per inch), about 3 mm in size. It is on this thin carbon film that the sample will then rest so that it can be examined in the microscope (56,62).

1.1.1.1 Microscope Layout:

The general layout of a transmission electron microscope (Figure 1) is as follows. At the top of the column, there is a high voltage electron emitter that generates a beam of electrons that travel down the column. These electrons pass through the sample and a

series of magnifying magnetic lenses, to where they are ultimately focused at the viewing screen at the bottom of the column. Different lenses can be used to change the magnification and focal point of the image. Apertures along the column can be used to change the contrast and resolution of the image. The column itself is at a very high vacuum to minimize interactions between the electron beam and air molecules (56).

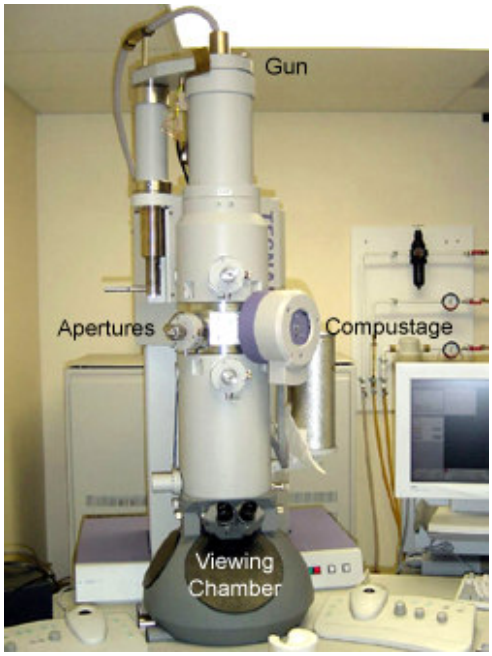


Figure 1: Picture of a Philips® transmission electron microscope with the column and viewing screen readily visible. Source: http://cryoem.berkeley.edu/

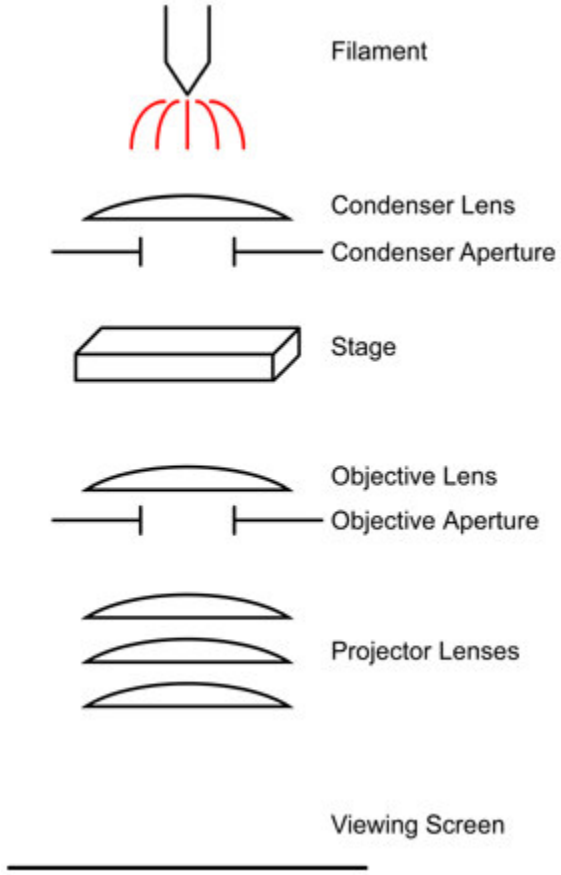


Figure 2. Side view of an electron microscope column, showing the principal components
Source: http://cryoem.berkeley.edu/

1.1.2 Cryo-electron microscopy

Cryo-electron microscopy or Cryo-EM (35,37,39) is a Transmission electron microscopy technique in which, the sample to be viewed is frozen in liquid ethane at about minus 155°C so that a fine layer of vitreous ice is formed on the sample and it is preserved and protected during observation and exposure to vacuum. Liquid Ethane is used instead of liquid nitrogen for cryo-freezing because of its higher heat capacity in comparison to liquid nitrogen. Hence when a sample at room temperature is dipped into liquid ethane the liquid ethane is cold enough to freeze water quickly and correctly without boiling off unlike liquid nitrogen. Quick freezing is required to freeze water into vitreous ice and not cubic or hexagonal ice which forms at much higher temperatures

than vitreous ice. Special grids are used for cryoEM, which are called holey carbon grids (37,39). Similar to normal carbon coated grids, holey grids are covered with a fine layer of carbon. However, as part of the preparation process, the carbon film is deposited in such a way that there are holes of a desired size in the carbon. One of the purposes of these holes is to eliminate any absorption and scattering of the electron beam by the carbon film, which will generate noise and obstruct the signal especially considering that the sample in cryoEM is just frozen and does not have negative stain. The holes also allow for "pockets" of solvent to form. Within these pockets, the specimen remains fully hydrated, even when the sample has been frozen (56).

The advent of Cryo-EM has allowed the imaging of isolated symmetric particles in the absence of stain and fixatives and under conditions, which preserve their symmetry. The advantage of Cryo-EM over traditional staining methods is that the shape of the sample is preserved, as the sample never comes into contact with the adhering surface. There is no stain to distort the sample. Also, low dose methods are normally used. Therefore, the electron beam causes less damage to the sample. Also by cooling the sample, free radicals produced by ionization are trapped, reducing their damage to the sample. The disadvantages are that there is very low contrast as the electron absorption properties of both protein and vitreous ice is the same. Care should be taken that the ice formed is vitreous and not cubic or hexagonal. Cubic ice absorbs electrons easily and the resulting sample is worthless (56,62).

1.2 Imaging and reconstruction

Imaging in cryo-EM is a very involved and intricate process. The sample in cryo-EM is very delicate. Low dose of electrons have to be used to prevent sample damage arising

from inelastic scattering of electrons. Williams and Fisher first developed the method of Minimum Dose Microscopy in 1970 (49). Two types of electron-sample interactions occur. They are elastic and inelastic scattering. The electrons undergoing elastic scattering have a fairly wide angular distribution and do not transfer energy. They give rise to the high-resolution information. The inelastically scattered electrons transfer some of their energy to the sample, which accumulates and can break apart molecular bonds, destroying the sample after some time. Its angular distribution is narrow and gives rise to undesired background term in the image. Therefore, for high-resolution imaging, low dose parameters require that the area to be imaged be not exposed until the picture is actually taken. All image calibration and focusing is done beforehand on a nearby area, in the hope that it's properties are similar to the final imaged area. Also, for the final imaging, very low electron doses of the order of 6-10 electrons per $Å^2$ are used (56, 62).

1.2.1 3D Reconstruction - Principle:

The images received from an electron microscope are two dimensional projections of the randomly oriented particles in the grid. The various projections or 2D images then put together to get a complete 3D image and hence is termed reconstruction. The method of reconstruction is based in the central section theorem that relates the 3D Fourier transform of the reconstruction with the 2D Fourier transform of the projection images. The mathematical theory was first developed by Radon in 1917 (51). This theorem states that in reciprocal space each 2D image produced by the microscope is a central section of the 3D Fourier transform of the object, and perpendicularly oriented when considering the direction of the electron beam with respect to the object. The Fourier Transform is an important image-processing tool, which is used to decompose an image into its sine and

cosine components. Hence, by combining different EM views, one can progressively fill up all the 3D reciprocal space, which by inversed Fourier transform can provide a 3D reconstruction volume of the particle in real space. The output of the transformation represents the image in the Fourier or frequency domain, while the input image is the spatial domain equivalent. In the Fourier domain image, each point represents a particular frequency contained in the spatial domain image. For non-symmetric objects high-resolution 3D reconstruction methods require 10,000 to 100,000 or even more images to process. The process gets simplified when the object to be reconstructed has high symmetry which is not always the case (60,61).

Single particle reconstruction (30,9) is the commonly used technique for getting the 3D structure from the 2D images received from a transmission electron microscope. The image of a field of randomly oriented, particles can be processed to yield a three dimensional structure by determining the relative positions of the symmetry elements. This involves putting together the various projections or 2D images to get a complete 3D image and hence is termed reconstruction.

The process of reconstruction is iterative. Experimental projections are first selected from electron micrographs and orientations assigned to them. Particles with similar orientations are then centered, aligned, and classified. From the centered/aligned/classified particles, a preliminary model is generated. There are several methods to do this. They are the Random conical tilt, Common Lines technique and Projection-Matching. To refine this solution, the model is used to better align and classify the original particles. A new model is generated from the refined data, and so on until a satisfactory solution is reached (56).

1.2.2 **Finding the orientations**:

The first step in this process is assignment of Euler angles to each experimental projection. Once these angles are assigned to each experimental projection, the model can be generated.

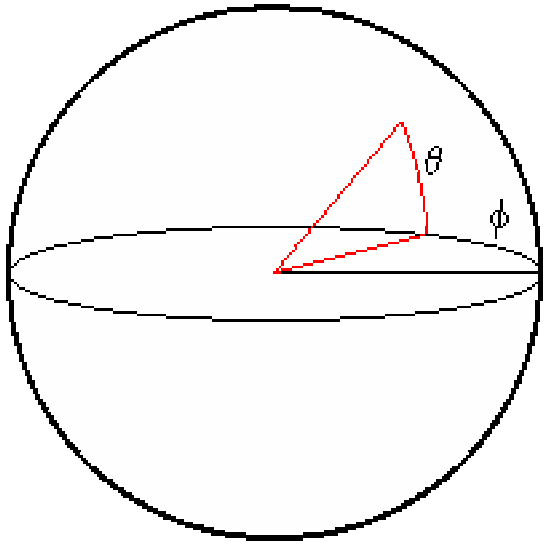


Figure 3. Diagram of an Euler sphere indicating how θ and φ are used to describe any point on the surface of the sphere
(Source:http://cryoem.berkeley.edu/~nieder/em_for_dummies/reconstruction.html)

The transformation between the vectors in the coordinate system of the molecule and those in the coordinate system of the projection is expressed by three Eulerian rotations as per the convention followed by Radermacher (1991). These rotations are characterized by three Euler angles, which indicate a position and orientation in space around a common center. The three Euler angles are,

θ (theta)

Defines the elevation above or below the equator.

φ (phi)

Defines the rotation (azimuth) around the equator.

ψ (psi)

Defines the rotation around the center of the position defined by θ and φ.

In order to define a point on the surface of the sphere from the center of a hollow sphere, two pieces of information are important, the elevation of the spot relative to the equator of the sphere (θ) and what the rotation around the vertical axis of the sphere (φ - azimuth). To generate a reconstruction, θ and φ has to be determined for each experimental particle in order to define its position on the Euler sphere. The center of the Euler sphere represents the position of the 3D model. Assignment of the Euler angles can be done using random conical tilt, common lines method or by using a reference image. (56,62).

1.2.2.1 Random conical tilt:

This method is based on the fact that single macromolecules assume preferred orientations on the specimen grid (38). For this method the specimen field is recorded twice in tilted and untilted positions. For a specific class of particles images, once the untilted particles are aligned with each other, those angles and shifts are applied to the tilted pairs to determine their relative orientations in space. The projections of these aligned particles will form a cone with a fixed angle (the tilt angle) and a random azimuth (rotation) around the cone, matching the in-plane angles found in the alignment of the untilted particles. This method can be applied only if the particles in the specimen are known to take preferred orientations.

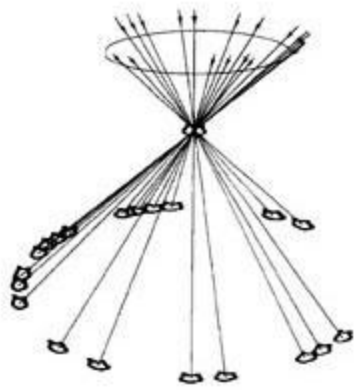


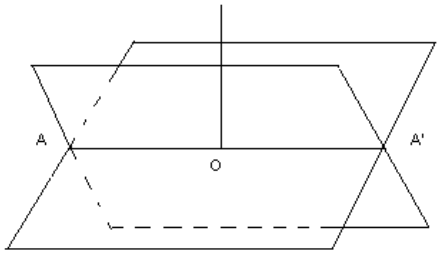
Figure 4 Principle of random conical tilting, showing how many rotated images within a cone come together to form a surface (J Frank ,1996).Originally published in *J.Microsc.* **146**, 113-136
Source: http://cryoem.berkeley.edu/

1.2.2.2 Common Lines approach:

The angular alignment is based on the Central section theorem. It thus follows that two non-parallel projections will have Fourier transforms identical to two non-parallel planes through the origin of the Fourier transform of the 3D volume, i.e. they will intersect in one line through the origin (22,48). For noise free data, their relative orientation can therefore be determined except for the rotation around an axis coinciding with this intersection (common line). For three or more orientations the relative orientation will be fixed except for the handedness (enantiomorphism) of the structure, which cannot be resolved from independent projections alone. The presence of symmetry within a molecule will generate common lines also in a single projection, and for symmetries higher than twofold rotation axis, the symmetry group can be oriented from an individual image. Due to the low signal to noise ratio in real images of non-stained biological macromolecules (typically 0.5-3 MDa molecular weight) it will be necessary to generate a large number of common lines to determine the orientations. For example the common lines for a two-fold, three-fold and five fold respectively are shown below:

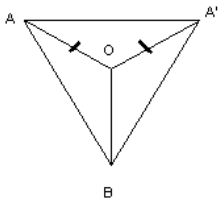
Two fold:

Generates one real line AA’



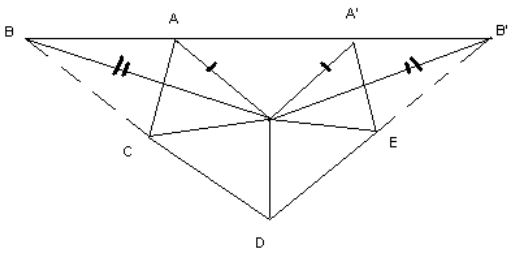
Three fold:

Generates a pair OA and OA’



Five fold symmetry:

Generates two pairs of common lines OB OB’ and OA OA’



Source: Dr.Tony Crowther’s lecture at NCMI workshop, December 2002

Icosahedral reconstruction

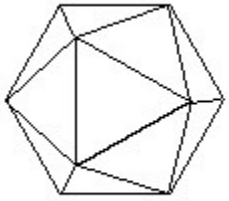


Figure 5: Icosahedron .
Source: http://www.cryst.bbk.ac.uk/BBS/whatis/bbsem.html

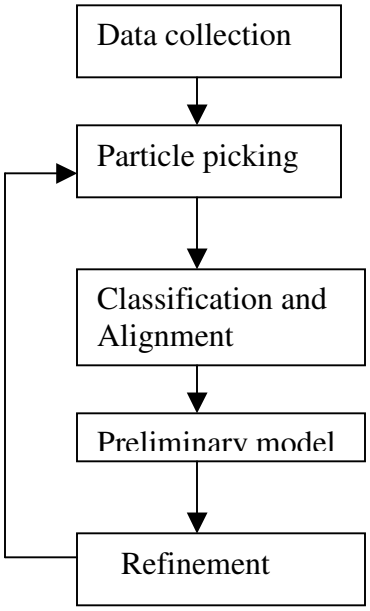
An icosahedron is composed of 20 facets, each an equilateral triangle, and 12 vertices, and because of the axes of rotational symmetry is said to have 5:3:2 symmetry. The Axes of Symmetry are a six 5-fold axes of symmetry passing through the vertices, ten 3-fold axes extending through each face and fifteen 2-fold axes passing through the edges of an icosahedron. The five fold axis is defined by Theta=90º and Phi=±31.717º while the three fold axis is defined by Theta=69.09º and Phi=±0º. Icosahedral reconstructions take advantage of this internal symmetry and repetition to generate a detailed three-dimensional structure from the data set. There are exactly 37 pairs of common lines for an icosahedron. Each of the 6 five folds gives rise to two pairs, each of the 10 three-folds generates 1 pair and each of the 15 two folds a further pair which lie along the projection of the two-fold symmetry axis.

1.2.2.3 Projection-Matching

If a starting model is available then it is possible to take projections from this model and use these to compare and classify the input images. Refinement is then done iteratively to get a better model.

1.2. 3 Reconstruction cycle:

1.2.3.1 Flow chart



1.2.3.2 Particle picking:

The image taken by the microscope of a particular area in a grid is photographed and developed into a micrograph. The micrograph contains several particles in various orientations. In order to reconstruct a 3D molecule from it's 2D projections, it is often necessary to use several thousand experimental projections (particles). The right particles have to be picked out in order to get a good final 3D image. There are several softwares available to do this, which will be discussed later (9,21,22).

1.2.3.3 Particle alignment:

Alignment involves placing the images of particles into a similar orientation on the screen. To calculate the similarities, correlation methods are used. The correlation can be done either between images in the data set (cross correlation) or by comparing an image to itself (autocorrelation). Usually a mixture of the two is used (56,62).

1.2.3.4 Particle classification:

Due to the interaction with the carbon coated grid the particles take up different orientations. Each orientation would make up a class and the particles belonging to each class will then be averaged out. The back and the front projections are usually the most common orientations. The higher the symmetry the lesser will be the number of classes (56).

1.2.3.5 Model building and Refinement:

Using any one of the methods mentioned in 1.2.2 the initial model is built and iteratively refined with expanding particle data sets, to get better models.

1.2.4 Programs used in reconstruction:

There are a number of programs available to do reconstruction. Some examples are the common lines based program MRC-ICOS (22) , EMAN (16) and model based programs like PFT (21). The MRC/ICOS programs are used for icosahedral reconstruction. This program uses a common lines based approach for reconstruction of the model. EMAN is not limited to only icosahedral symmetry and can be used for any kind of symmetry. It is based on single particle reconstruction and also uses both the projection matching and the common lines method for model building. There are several model based programs which given a starting model will refine it iteratively using Projection matching algorithms. The model is used to generate projections, which will then be used to classify and align the input particles and get a more refined model.

1.3 Important experimental parameters in reconstruction:

1.3.1 Contrast transfer function (CTF) correction:

In single particle analysis the amplitude and the phases are both obtained from the images. Amplitude contrast, which is the result of scattered electrons being deflected into the objective aperture, contributes approximately 5-10% to the image, as the electron scattering properties of protein and vitreous ice are very similar. The structure factor is a measure of this atomic scattering. Phase contrast is the result of interaction between the undeflected electron wave and those electrons that are weakly elastically scattered and still pass through the objective aperture. Phase contrast is increased by defocusing the microscope, but only for a range of spatial frequencies. This is dictated by an instrument-dependent function called the point spread function (PSF) that is convoluted with the specimen projection to form the electron image. The PSF is more commonly considered in terms of its Fourier transform, the contrast transfer function (CTF). The CTF is an oscillating function and is dependent on several features of the microscope, most importantly in the case of cryo-TEM, the defocus and accelerating voltage. The advent of Field emission gun (FEG) microscopes has brought about a strong improvement in the temporal and spatial coherence and therefore the contrast transfer function thus enabling high-resolution imaging. The figure below shows the contrast transfer function curves for two microscopes an FEG and a non-FEG (LaB_6) microscope. The strong improvement in contrast transfer by the FEG is to a large extent due to the much smaller incidence angle α, a parameter that describes the range of incidence angles seen by the specimen. Ideal spatial coherence with infinite extent ($\alpha = 0$) would be obtained with a point source for the electrons. FEGs come much closer to this ideal situation than LaB_6 filaments (28).

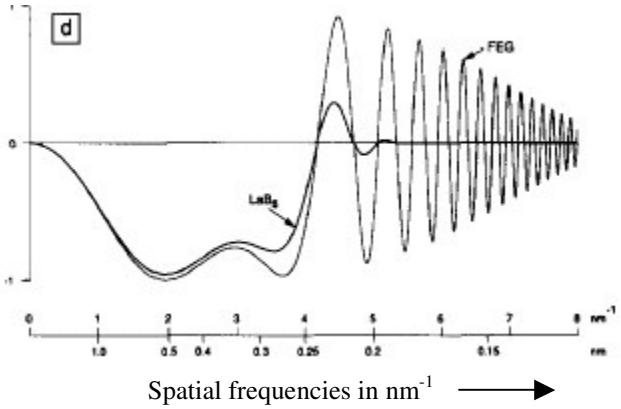
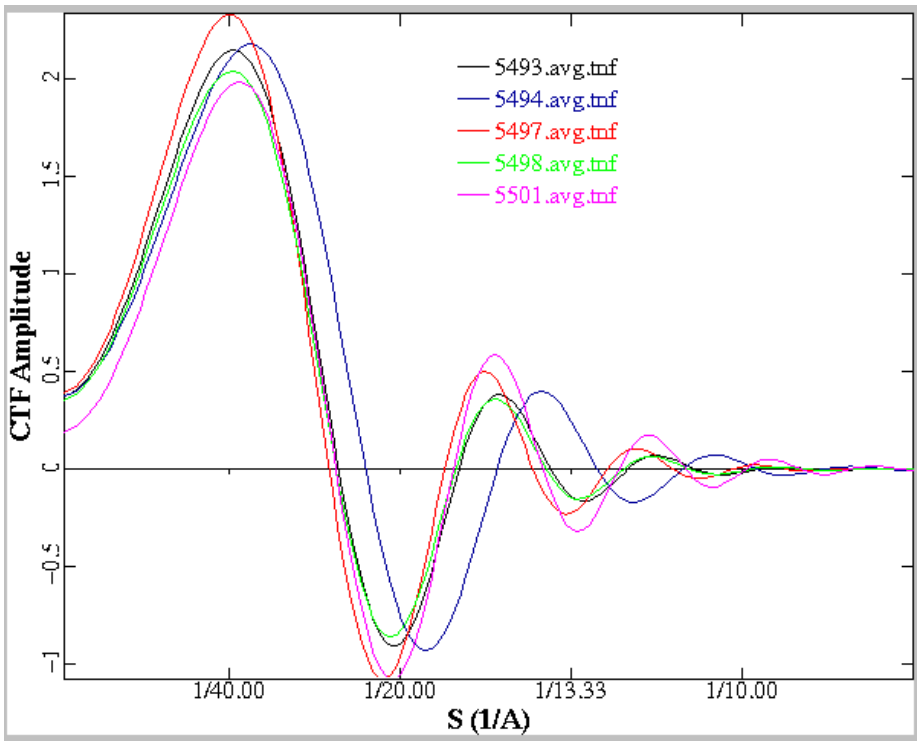


Figure 6: Comparison of contrast transfer functions between an FEG and non-FEG (LaB_6) microscope. Source: M.T. Otten, W.M.J. Coene. High-resolution imaging on a field emission TEM. Ultramicroscopy 48 (1993) 77-91

A consequence of the improved CTF is that within certain resolution ranges density information is inverted i.e. the contrast is reversed. This can profoundly affect the resolution attainable in three-dimensional reconstructions from cryo-microscopy data unless sophisticated CTF corrections are applied. At a large defocus the CTF may restrict resolution to ~ 4 nm, which is too low to gain meaningful structural information. However, a very close to focus image will have very low contrast making the extraction of structural information very difficult. Cryo-TEM is, therefore, a trade off between contrast and resolution. It is important to have a wide range of defoci to get the maximum information for all spatial frequencies as can be seen from the figure below (Figure 7) where information at different defoci complement each other (55,57).



Defocus of micrographs:

5493.avg.tnf = 2.4 µ
5494.avg.tnf = 1.9 µ
5497.avg.tnf = 2.4 µ
5498.avg.tnf = 2.3 µ
5501.avg.tnf = 1.3 µ

Figure 7:Graph represents the CTF amplitudes at different defoci. Figure taken from "ctfit" program in EMAN

1.4 Bacteriophage P2 and P4

Bacteriophages P2 and P4 are two genetically unrelated dsDNA bacteriophages. P4 is a satellite phage of P2 as it is constructed from the structural proteins encoded by P2 (Bertani capsid protein, arranged with T =7 symmetry). The smaller P4 capsid has T =4 symmetry and contains 235 copies of gpN-derived protein (4) as compared to 415 copies in P2. P4 contains 12 pentamers and 30 hexamers.

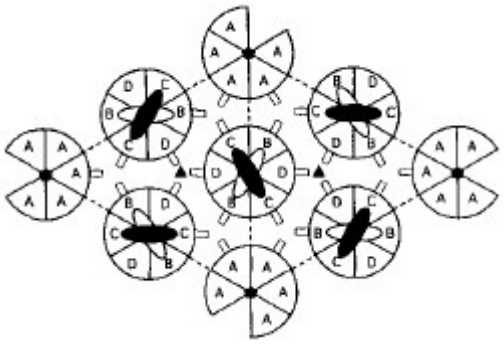


Figure 8: This is the view while looking down the two-fold view. The symmetrically different subunits are A,B,C,D wherein five A subunits form the pentamer and two of B,C,D subunits each form a hexamer. Pentagons and triangles indicate the five fold and three fold axes.
Source: Terje Dokland (2000). Freedom and restraint: themes in viral capsid assembly. Structure 8: pp R157-62

The presence of external scaffolding protein Sid (24) is responsible for the smaller size of the P4 capsid (36). Both P2 and P4 capsids contain a connector protein gpQ, which is a dodecameric ring-shaped protein and serves as a portal vertex for DNA transport (26). In the double-stranded (ds) DNA bacteriophages, cleavage is usually an integral part of the maturation process, occurring concomitantly with the removal of scaffolding proteins, DNA packaging, and accompanied by large-scale structural transitions in the capsid (14). An empty precursor capsid of the P2 and P4, called the procapsid is formed initially during assembly. The procapsid contains mostly uncleaved gpN, as well as scaffolding proteins (25). During assembly, the scaffolding is removed, the DNA is packaged through the portal vertex, and the capsids undergo striking conformational changes (24). These processes are accompanied by cleavage of the structural proteins (25), in this case gpN (40.2 kDa). It is cleaved to the smaller protein species h1 (39.0 kDa), h2 (38.6 kDa), and N* (36.7 kDa) (30). The predominant form of gpN found in mature capsids of both P2 and P4 is N*, although P4 capsids also contain about 10% of h1 and h2 (30). In vivo, this cleavage is presumed to be dependent on the internal scaffolding protein gpO (31 kDa), which itself is cleaved to an amino-terminal 17-kDa fragment that remains inside

the mature capsid (26). Much is not known about gpO protease activity. The 31.9-kDa gpQ protein also undergoes N-terminal cleavage during capsid maturation, leaving the 36.7-kDa cleavage product Q* (26).

1.5 Protein-conducting channel in the endoplasmic reticulum (ER) membrane:

Membrane bound ribosomes associate with the Sec61p complex. Secretory proteins pass through the channel and enter the ER lumen. In the case of the membrane proteins, hydrophobic segments are incorporated in the lipid bilayer while their luminal and cytoplasmic loops are exposed on either side of the membrane (17). The sec61p complex is the minimally required protein to reconstitute translocation in vitro. The heterotrimeric sec61p complex consists of an α subunit with ten transmembrane (TM) domains as well as smaller β and γ subunits, each with a single TM domain (17). Electron microscopy studies show that the channel in both the prokaryotes and eukaryotes forms a ring-like structure (23,19). The size of the channel suggests multiple copies of sec61p. The other complexes associated with the channel are the Translocon-associated protein (2) and the Oligosaccharyl transferase (OST) (31,32).

Cryo-electron microscopy and single particle methods were used to study the 3D structures of the ribosome channel complex with or without nascent chains in both yeast and mammalian systems (11). These studies showed that the junction between the ribosome and the channel was not tight; but instead made of a ~15A gap, three strong connections and a weaker one. The existence of the gap suggests the possibility that the cytosolic loops of membrane proteins may egress laterally and be esposed to the cystosol. The hydrophobic segment of the membrane protein exits laterally through the channel walls while the secretory proteins exit through the channel pore into the ER lumen.

Beside the sec61p complex, the TRAP and OST also form the channel and may contribute to the lumenal protrusion seen in the native channel. The ribosome channel complex structure was improved to 17 Å resolution from 27 Å and all known high-resolution ribosomal subunit models were then docked within this map of the ribosome–channel complex (5). These studies helped in identifying candidate regions on the large subunit that may participate in forming connections to the ER channel and the possible role of L1 stalk as a gate to release the E-site tRNA from the ribosome. The tRNA molecule is retained in the E-site after programmed mammalian ribosomes are treated with puromycin. Finally, six expansion segments were localized on the large subunit and, when combined with previous work provides insight into possible roles for these rRNA components. The TRAP and OST in the membrane channel do not seem to be directly in contact with the ribosome but instead seem to be associated laterally to the channel. The lumenal protrusion seems to be more due to TRAP as it is consistent to the size of TRAP. The narrowest region of the translocation pore may be smaller than 10–12 Å in diameter, and span a distance of 15Å (5).

2 Methods:

2.1 Purification of P4 procapsid and expanded capsids:

A new method for the production of P4 procapsids in vitro and in vivo was devised and greatly facilitated the study of procapsids in vitro (13,4). The chimeric plasmid pLucky7 was constructed by fusing the Sid-expressing pVE2 and the gpN-expressing pTL1 plasmids, so that the two genes were under independent control of the T7 promoter. Induction of pLucky7 yielded high-level expression of both Sid and gpN protein with Sid in slight excess. The proteins expressed from pLucky7 assembled effectively to form procapsids, which were recovered by centrifugation at 40,000 rpm in a Ti50.2 rotor (190,000 *g*) for 1 hour. After resuspending in procapsid buffer, the particle mixture was run on a 5–20% sucrose gradient for 2 h at 34,000 rpm in an SW41 rotor. Gradient fractions were analyzed by SDS–PAGE and EM. The procapsid-containing band was diluted 5 times, concentrated by pelleting at 190,000 *g* for 1 h, and resuspended in procapsid buffer (100 mM Tris, pH 8.0, 50 mM NaCl, 10 mM MgCl2)(4).

Trypsin treatment of the procapsids at a mass ratio of 1:400, corresponding to 0.36 BAEE enzyme units per milligram procapsids, for 24 h at 4°C, resulted in a cleavage pattern similar to that observed in spontaneously cleaved material. The main cleavage product was a doublet band at around 36 kDa, equivalent to the size of N*. Little cleavage was seen in Sid at this time point. At longer times (from 2 to 10 days), most of the cleaved gpN was present in the smaller of the two bands in the doublet, and more cleavage was seen in Sid as well. Treatment with trypsin at a mass ratio of 1:50 (3.0BAEE units per mg procapsids) for up to 2 h at 22°C yielded similar results. By comparison, Sid in solution is completely degraded by trypsin at 1:50 mass ratio after 20 min at 22°C (S. Wang and

T. Dokland, unpublished results). Part of the Sid cleavage seen upon trypsin treatment of procapsids is probably due to the release of protease-sensitive Sid, while Sid bound to procapsids is at least partially protected against proteolysis. Sucrose gradient purification of the trypsin-cleaved material resulted in a faster fraction containing uncleaved gpN and Sid and a slower fraction containing mainly cleaved gpN and no Sid (1).

2.2 Cryo-EM for P4 procapsids:

The procapsid particle samples recovered from the sucrose gradients were diluted 5 times, pelleted at 190,000 *g* for 1 h, and resuspended in procapsid buffer at a concentration of 9 mg/ml. The sample was diluted to 1 mg/ml concentration in procapsid buffer immediately before microscopy, applied to quantifoil copper grids, plunge-frozen in liquid ethane, and observed in an FEI Tecnai 12 electron microscope operated at 120 kV, equipped with a Gatan cryospecimen holder. The images were collected on Kodak SO-163 film at a magnification of X 50,000. The particles measure about 390-400A. The defocus of the micrographs ranged from 0.9 μm to 2.5 μm.

2.3 Reconstruction of P4 procapsid using EMAN:

EMAN (16) was used for reconstruction. EMAN is a set of image/volume processing tools to perform single particle reconstructions. It can be used for parallel processing and contains subroutines to run the desired process in as many as 32 processors at a time. The basic steps in the reconstruction using EMAN are illustrated in the flow chart in figure 9.

2.3.1 Scanning and Digitization:

The negatives were scanned using an EverSmart scanner from CreoScitex. The scanning resolution is 220 dots/millimeter. Once the images were scanned they were converted into a file format with extension “smv”. This format is very useful to conserve space and

reduces it by one-third as it stores the RGB image by taking in a single channel, red, blue or green. These files were then converted into spider format using a small script and subsequently converted into the "MRC" format (stands for medical research council) using Proc2d program in EMAN (16) and inverted. In cryo-EM negatives, the proteins appear darker then the background and hence inversion is done to follow the conventions of EMAN where the higher density, in this case protein, appears lighter than lower density regions.

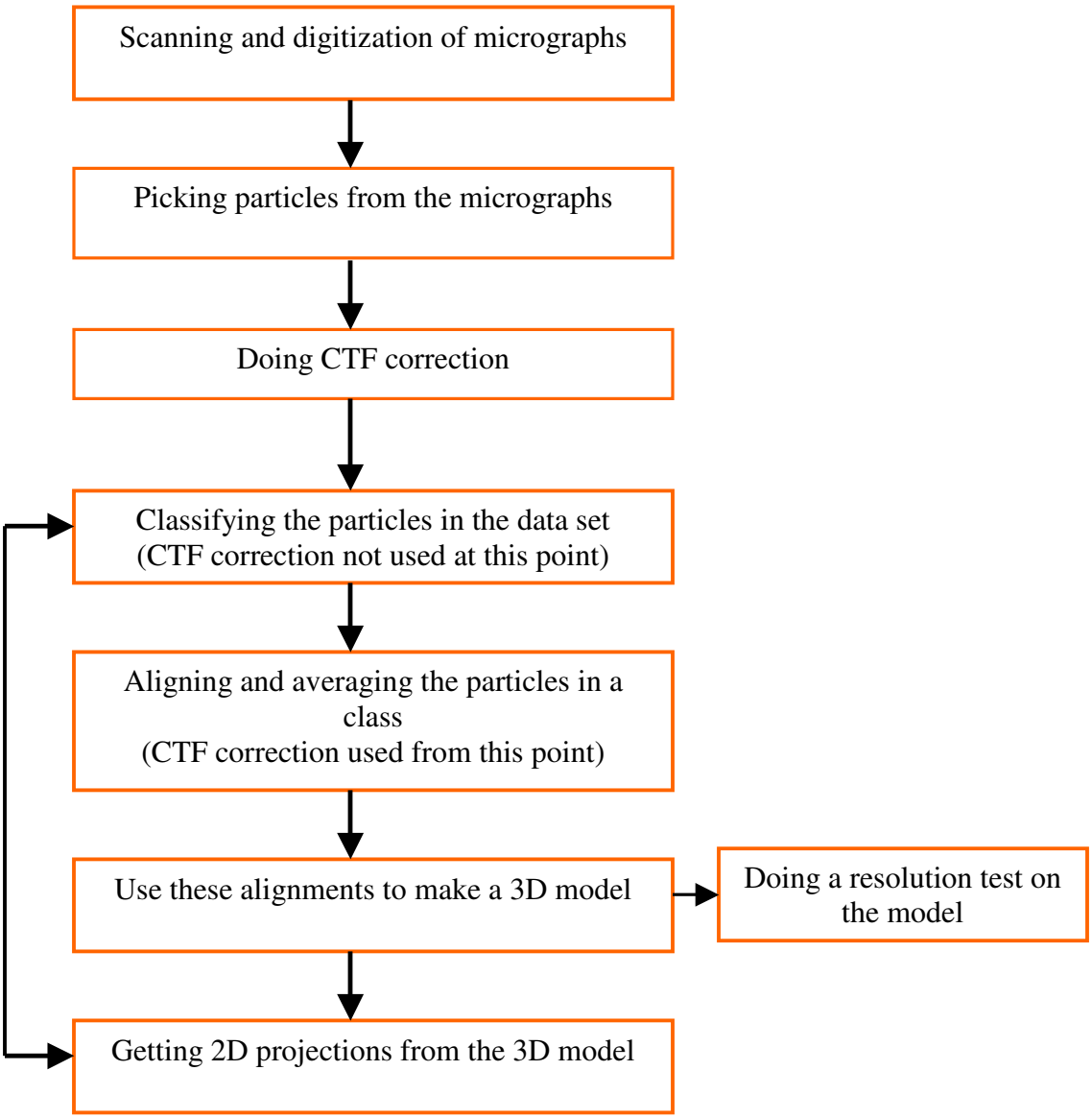


Figure 9: Flow chart of reconstruction in EMAN

2.3.2 Picking the Particles (Boxing out)

Particles were picked out from the scanned micrographs by using Boxer (16) available in EMAN. The semi-automated feature in boxer allows the user to choose a few good particles based on which it searches, using correlation based algorithms, the rest of the micrograph and picks out similar particles. This is not a foolproof method due to the varying contrasts and contaminations in a micrograph. Hence the selected particles have to be manually scanned to remove out any bad particles. This is a very important step in reconstruction, as selection of bad particles will lead to an incorrect 3D model.

The resolution of the images was around 2.73Angstorm/pixel. This can be calculated by taking into consideration the resolution of the scanner (220dots/mm or 4.545 microns), magnification of the microscope (50,000) and the binning factor (3) applied while scanning the micrographs. The formula is:

(Resolution (in microns) * binning factor)/magnification

$(4.545 * 3)/(5*10^4)=2.727$ A/Pixel

2.3.3 Correction of Contrast Transfer Function:

A number of experimental and instrumental factors must be correctly included in the data analysis. These factors include the contrast transfer function (CTF) of the microscope (46,50), the effective envelope function (*E*) (35, 40, 41, 44, 45, 47) and a background noise function (*N*) (16,20,34). The envelope function is defined by B-factor and is a measure of a combination of factors like the spatial and temporal coherence of a microscope, the sample drift and scanner optics. These factors can be approximated as a single Gaussian envelope function, the width of which is described by B-factor (8). The B-factor is experimentally dependent on the objective lens defocus setting of the

microscope. The experimental B-factor is very critical for obtaining an atomic resolution structure. EMAN (16) uses a program called ctfit to fit in the B-factor, defocus, ctf amplitude and the Noise functions in a power spectrum and uses these corrected ctf parameters in the image reconstruction (53).

This is done using the “ctfit” program in EMAN (16). 'Ctfit' determines the parameters of the CTF in each micrograph by fitting parameters in a predefined 10-parameter model to the power spectrum of data from the micrograph. These parameters are then used during the reconstruction procedure to make corrections for the CTF. A power spectrum is produced for each micrograph and the 10 parameters can be adjusted using sliding panels in a graphical interface window to best fit the default curve (Figure 10). The y-axis represents the intensity while the x-axis is the reciprocal of resolution in Angstrom.

The measured data is usually a product of the structure factors and the CTF. The structure factor as discussed before contributes to about 5-10% of the final structure. Most of the structure information is from the phase contrast. To have more accurate structure factor data we would need to use the structure factors of the gpN protein itself. For this we would have to use small angle and for reaching higher resolutions large angle diffraction experiments of gpN. Experimentally, this would require a large quantity of the protein. To derive the structure factor file from EM data itself we would require some amount of very close to focus data in order to facilitate collection of data from a larger frequency range.

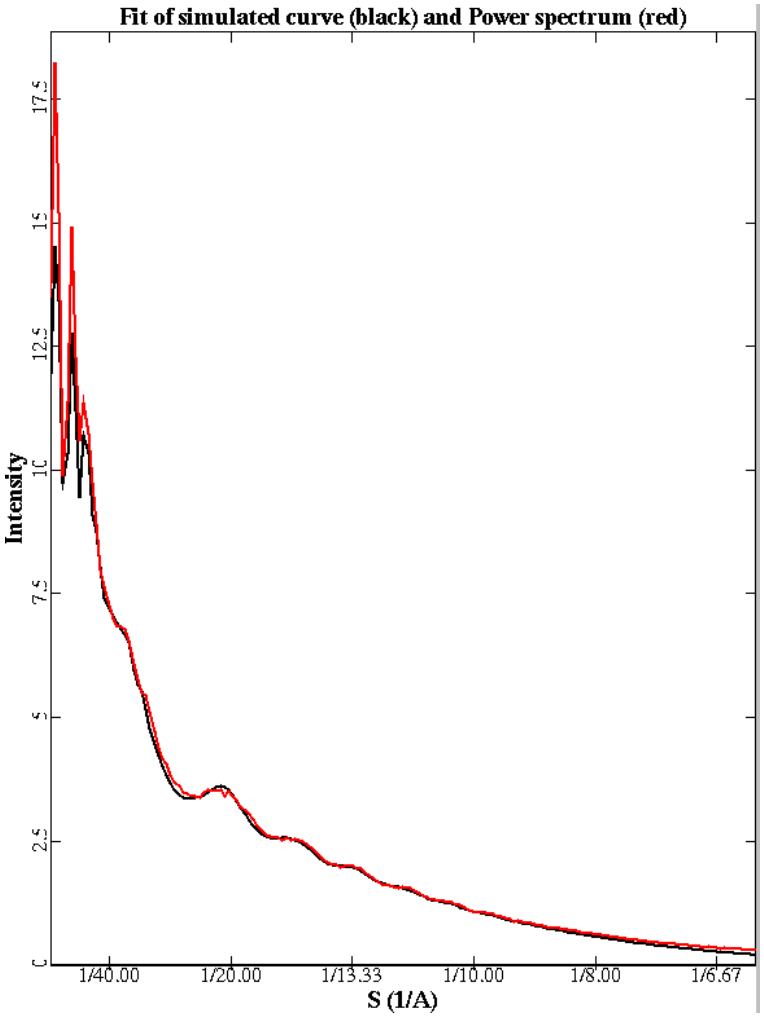


Figure 10: Fitting a simulated curve into the experimental spectrum of the micrograph

2.3.4 Generating a preliminary 3D model and refinement:

An initial model was generated from the best particles, about 442 raw particles from 5 micrographs after doing four cycles of refinement and then refined iteratively with an increasing subset of original images to get a better model using the "refine" command in EMAN. Eight processors were used for the process of reconstruction. The final model was generated from 2755 particles after doing about 35 cycles of refinement excluding the first 4 cycles. Several parameters have to be specified in the refine command, chief among which is the angular step. This is given in degrees and defines the unit triangle in which the projections are generated with a relatively uniform projection density. The

other parameters are “classkeep”, which denotes the threshold value for keeping particles in a class, “classiter”, indicating the number of iterative loops, a circular mask radius for the class average and the symmetry of the particle, icosahedral, in this case. To apply ctf corrections the “ctfcw” option is used which performs CTF correction with a wiener filter using the structure factor file. Initially the angular step was taken as 3.8 and then gradually decreased to 2.5 for a finer refinement. The “classkeep” was decreased from 0.8 to less than 0.001, which would discard as many as 50% of the original data set before the final reconstruction, to increase the stringency of the reconstruction process. Computational power is a rate-determining factor in the reconstruction. Hence the reconstruction was initially done at a sampling of 5.44 Å/pixel to speed up the process and then increased to 2.73 Å/pixel, once an optimum resolution for the 5.44 Å/pixel sampling space was reached (which depends upon the scanner and could be about 2 to3 times the sampling space). About 20 cycles were done at 5.44 Å/pixel and 15 cycles at 2.73 Å/pixel. In the last 15 cycles the “defilt” option in refine was also used to improve resolution. This option will make the rotational average structure factor of the particle the same as the projection. In the last two cycles the “refine” option in refine was used, which enables half pixel accuracy in alignment. Both the defilt and refine options considerably improved the resolution but decreased the computational speed to a large extent and hence were limited to the last few cycles only (53).

2.3.5 Resolution test:

In order to calculate the final resolution “eotest” program in EMAN (16) is used. This calculates the Fourier shell coefficient (fsc) between two reconstructions from even and odd halves of the data. There are different ways of calculating the resolution. A more

conservative but foolproof resolution test is to take the resolution at which the fsc falls to 0.5. The output of eotest is a file containing the fsc for the entire range of pixels ranging from 1 to half the boxsize (216/2=108 in the case of procapsids). Hence the resolution is calculated by multiplying, twice the sampling space, with maximum pixel value and dividing it by pixel value at which fsc is 0.5. The reason we take twice the sampling space is due to Shannon Nyquist theorem. This theorem states that to avoid aliasing, that is non-overlap of the periodic transform of the object density, the finest spatial period must be at least twice the sampling space (Dr.Tony Crowther's lecture at NCMI workshop, December 2002).

Again this is an underestimation of resolution as it compares the models from two halves of the data.

2.4 CryoEM for P4 expanded particles:

The material from the upper band of the sucrose gradient separated trypsin-cleaved particles was observed in a native, unstained state by cryo-EM (37, 39). The samples were pelleted in a Beckman Airfuge, resuspended in 10 mM Tris pH 8.0, 15 mM MgCl2, placed on Quantifoil copper grids, vitrified by plunging in ethane, and observed in an FEI Tecnai F20 field emission gun electron microscope operated at 200 kV, equipped with a Gatan cryo-specimen holder. The images were collected on Kodak SO-163 film at a magnification of X50,000 (1).

2.5 Reconstruction of P4 expanded particles using the MRC-ICOS programs:

The MRC-ICOS (22) set of programs and PFT (21) was used for reconstruction and a simple program "Micos" written by Dr.Terje Dokland was used to set up scripts and organize particle lists. The flow chart for the reconstruction process is given in Figure 11.

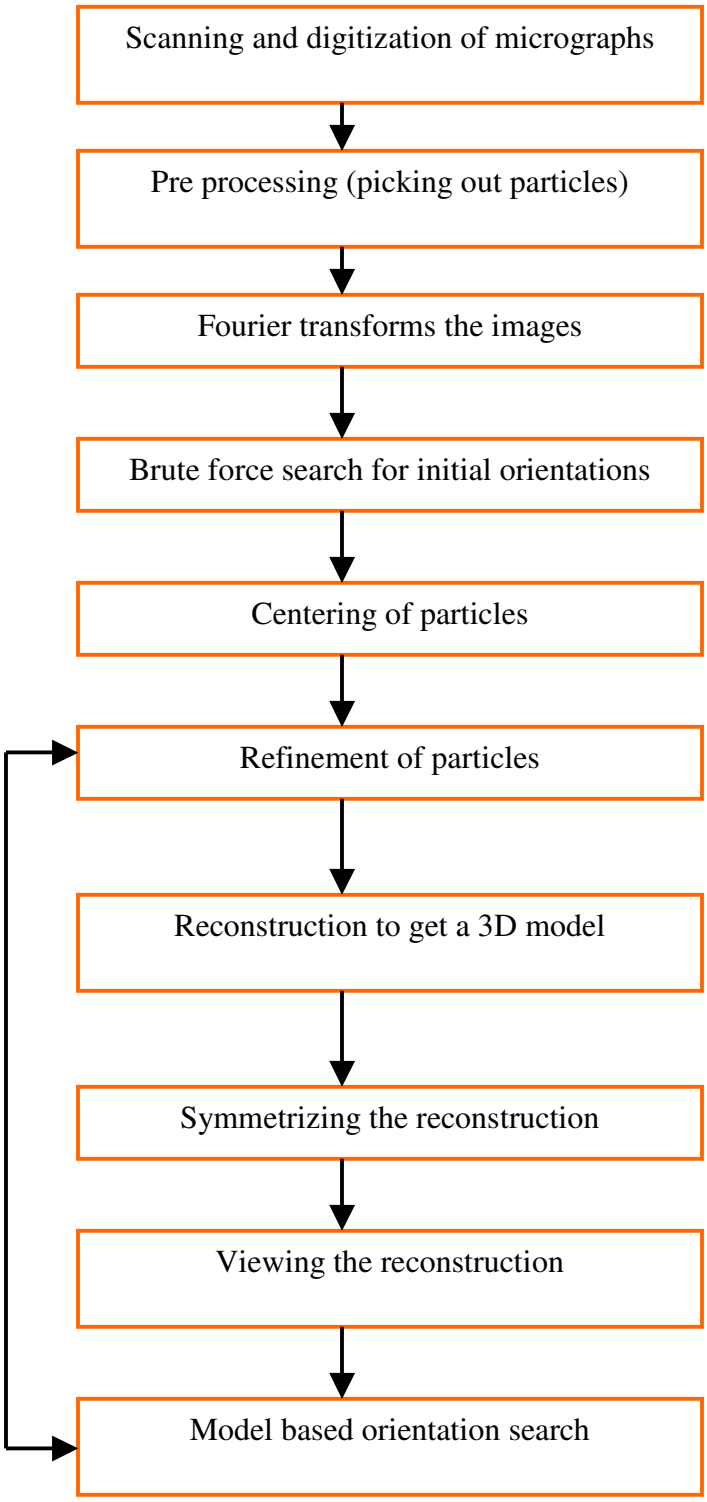


Figure 11: Flow chart representing reconstruction-using MRC-ICOS

2.5.1 Scanning and digitization:

The cryomicrographs were scanned on a FAG Hi-scan rotating drum scanner (Dunvegan, Geneva) in 16-bit gray scale mode at a raster step of 1000 dpi or 25.4 μm, corresponding to 5.08 Å in the specimen. A linear stretch of the gray values was applied using Adobe Photoshop before conversion to 8-bit gray scale.

2.5.2 Particle picking:

Particles of expanded appearance were selected for reconstruction from 30 micrographs using BOXER from the EMAN package (16). The procedure for boxing is as explained in section 2.3.1.

2.5.3 Fourier Transformation:

The boxed particles were separated into individual particle files using "proc2d" in EMAN. Fourier Transformation of these images was done with FFTRANS in the MRC-ICOS program suite. All subsequent processing steps are done on these Fourier transforms.

2.5.4 Brute force search of orientations and Centering:

The program EMICOFV was used to generate orientations for these Fourier transforms. It works by searching all orientations (theta, phi, omega) in the asymmetric unit in one-degree steps for the lowest common-lines residual after weighting for degeneracy. The program computes various residuals and outputs the ten best orientations (those giving lowest residuals) as well as a residual map for each. By inspection of these, the best orientation, which is not affected by artifacts such as degeneracy, is picked. These particles were then centered using EMICOORG. EMICOORG1 minimizes the common-

lines residual to determine the phase origin (X, Y) for the particle in the orientation (theta, phi, omega) given by FV (52).

2.5.5 Reconstruction and Symmetrizing:

Reconstruction of this set of orientations was done with EMICOMAT, EMICOBG, EMICOLG, and EMICOFB (MAT_FB). These set of programs apply the icosahedral symmetry, do an inverse Fourier transform and generate three-dimensional maps of the averaged particle in various settings. In generating the final reconstruction, EMICOFB only imposes the D5 symmetry to the map. To apply the full icosahedral symmetry, the program SYMMETRIZE is used, and applies the three-fold symmetry missing from the map. SYMMETRIZE also allows interpolation and rotation of the map (52).

2.5.6 Model based refinement:

Model based orientation search on this initial model was done using PFT (21). The algorithm is the polar Fourier transform method developed by Drs. T.S. Baker and R. H. Cheng (Purdue University) and compares the projections of the model with the original images iteratively. It then classifies and aligns the original images based on these projections and outputs a reconstruction, which will be the starting model for the next cycle. In this way it progresses iteratively producing a more refined model in each cycle. SIMPLEX, a program to refine the particles was also used at this point but no further improvement of resolution was seen, possibly because the resolution limit was reached for the available number of particles. For viewing the reconstruction – PTONE was used. PTONE takes MRC format image files and generates half tone output in postscript format.

2.5.7 Determining the resolution:

The resolution was calculated by dividing the particles into even and odd halves and comparing the reconstructions from these two halves. The Fourier shell correlation (FSC) was calculated between these two reconstructions and the point at which it falls off steeply is considered as the most likely resolution for the model. This is an under estimation of the resolution as it compares two images formed from half the original data set and not the complete set.

2.4 Purification of ribosome channel complexes:

The ribosome channel complexes (2) were prepared by binding ribosomes to the purified Sec61p complex reconstituted in membranes, which created a more natural interaction (5). The canine non-translocating, native ribosome–channel complex was prepared by taking puromycin-treated ribosomes and ribosome-stripped membranes from rough microsomes in high salt conditions to re-assemble "native" channel complexes in which the ribosome has a tRNA in the exit-site.

Purified ribosomes were then added back to stripped ER membranes and the resulting membranes separated from unbound ribosomes by floatation on a sucrose gradient. These membranes were solubilized in deoxy-BigCHAP (DBC) and the ribosome channel complexes were concentrated by centrifugation (2).

2.5 Cryo-EM and three-dimensional reconstruction of Ribosome channel complex:

CryoEM:

After resuspension, the RCCs were adsorbed to a carbon-coated EM grid in buffer and flash frozen. Interactions of the channel with the continuous carbon support seemed to be a limiting factor. Therefore the grid surface was treated with poly-lysine, to prefer

interactions of the RCC through its ribosomal portion and avoid damage to the channel thus allowing the study of the channel in greater detail (2). ”Boxer” in EMAN was used to select particles from a set of micrographs scanned using an EverSmart scanner from CreoScitex. The scanning resolution is 220 dots/millimeter. The 8-bit data were binned 3X3. Boxer has an option called “autobox from reference” using which the particles can be boxed based on the projections in a reference file. An initial model at 17 Å (5) was used to generate these projections using “project3d” software in EMAN. The ribosome channel complex is an asymmetric particle and hence about 40 to 50 reference projections are required to box the whole micrograph with a reasonable amount of accuracy. The references should be able to describe all possible unique orientations of the particles but at the same time should not be prohibitively long. The project3d in EMAN has a unique algorithm to choose an optimal set of particles for use in autobox from reference (53). On using this option, a small area in the micrograph is chosen and a panel with four sliders appears. These have to be suitably adjusted until all the good particles in the chosen area are selected and then the program is run to autobox from the whole micrograph. About fifteen thousand particles were selected in this method. CTF correction was applied to these micrographs as discussed in section 2.3.The structure factor file used was derived from the micrographs that were close to defocus.

3: Results and Discussion:

3.1 Procapsid structure

Procapsids were produced by co-expression of gpN and Sid from the chimeric plasmid pLucky7 and purified on sucrose gradients as previously described in section2.1. Such procapsids consist of a 420 Å diameter (Figure 12), *T*= 4 capsid surrounded by an external scaffold made of 120 copies of Sid and are identical to capsids produced by combining purified gpN and Sid in vitro (13).

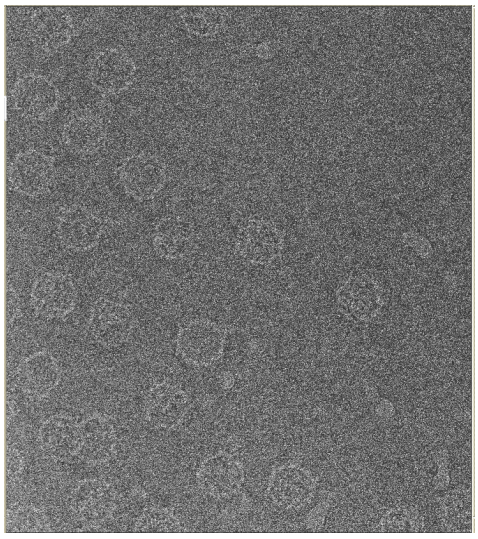


Fig 12: Cryo-EM micrograph showing P4 procapsids
Source: Wang et al. In vitro assembly of bacteriophage P4 procapsids from purified capsid and scaffolding proteins. Virology 275, 133–144

These particles were observed by CryoEM as explained in section 2.2. Initially five of the best micrographs were picked and particles boxed from these and ctf correction applied. These were used for generating a preliminary model in EMAN (16). Four hundred and forty two particles were aligned to generate the initial model at around 36 Å. Micrographs with a good distribution of P4 particles, which are well formed, and without too much ice and superfluous densities and other contaminations like dust, were chosen. Particles were chosen from these micrographs by using the semi-automated boxing

feature in boxer. A total of 30 micrographs were chosen and boxed to a size of 256x256. The structure factor data until 31.5 Å was got from the EM data on P4. Below 31.5 Å, the structure factor data from GROEL x-ray diffraction experiments were used (9). GROEL data was considered because between proteins (with the exception of membrane proteins) the structure factor data usually does not vary widely. Also, as explained before the structure factor contributes to 5-10% of the final resolution which may make a difference only when we go to very high resolutions like below 10 Å. Hence it may suffice to use the GROEL data for the missing structure factor data in the capsid protein. Initially to speed up the refinement process the refinement was done at a sampling space of 5.44 Å/pixel. After 20 cycles the original sampling resolution of 2.73 Å /pixel was considered at this point for further refinement. The resolution of the initial model was found to be around 36 Å and after doing 20 cycles of refinement at higher sampling of 5.44 Å/pixel the resolution improved to 17.408 Å. After this, 15 cycles of refinement were carried out at 2.73 Å/pixel sampling. Several cycles of refinement with different refinement conditions like “refine” and “dfilt”(53) were carried out as explained in section 2.3.4. A final resolution of 14.38 was reached with 2755 particles (Figure 13) as calculated by eotest (16). Ideally the curve should fall off gradually towards higher resolutions but in this case the curve rises up again after falling steeply at 14.38 Å. This could be due to missing data at around x=40, caused by the overlaps of the CTF zeroes at different defoci. This signifies the importance of having data from a large range of defocus.

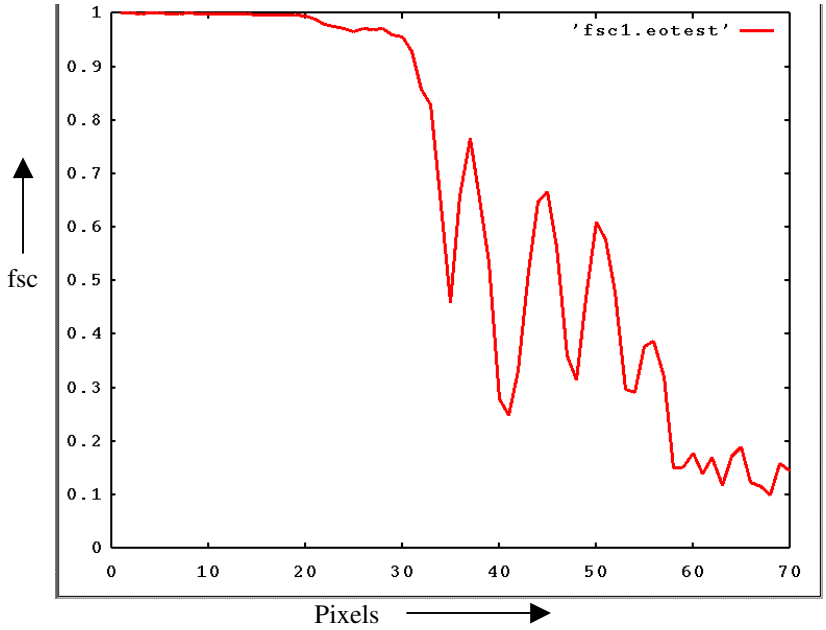


Figure 13: Output from "eotest". Resolution at which the fsc=0.5 on y-axis, corresponds to 40 pixels on the x-axis. The resolution is calculated to be 14.38 Å as discussed in section 2.3.5

The final structure of the procapsid at 14.38 Å (Figure 14) has more detail than the previous structure at 21 Å (4).

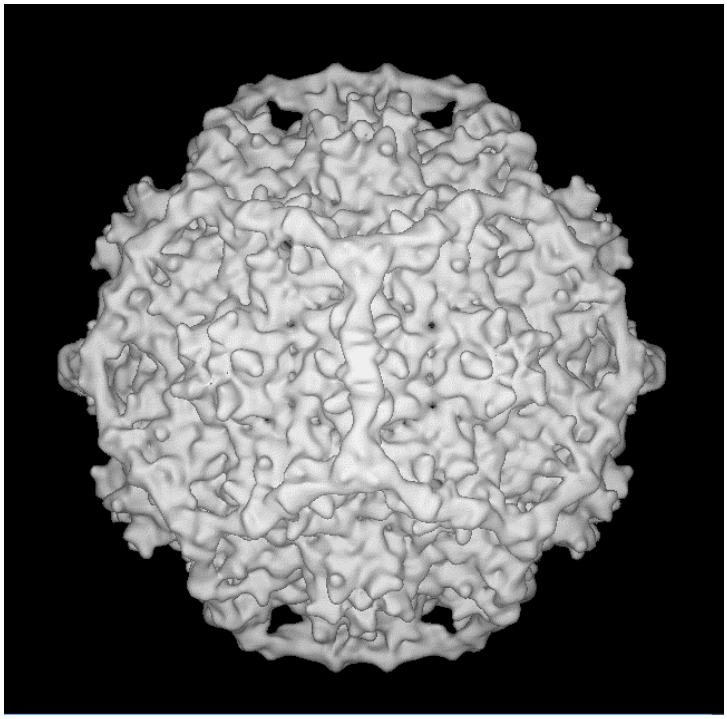


Figure 14: P4 procapsid at 14.38 Å looking down the 2-fold axis

The reconstruction clearly shows the T = 4 arrangement of gpN in the shell, organized as 12 pentamers and 30 hexamers. There are 120 copies of Sid (4) arranged in dimeric and trimeric parts. The handedness of the scaffold and the three-fold symmetry of the trimeric part are clearly seen (Figure 14). The presence of two distinct domains in gpN had been mentioned in an earlier paper (29) with one domain comprising the capsomer and the other the trimeric connection between the capsomers with a hinge region connecting them. The hinge region is supposed to conform the flexibility to take on several different conformations. The previous structure at 21 Å (4) also showed the gpN subunits in both the hexamers and the pentamers as having a protruding domain that is shaped similar to a slightly twisted slab, and a shell domain that made up the bulk of the capsid density. This feature is also seen in the present higher resolution model at 14.38 Å. The hexamers are not sixfold symmetric, but have an elongated appearance characteristic of dsDNA phage procapsids with a two-fold symmetry (Figure 13). Another feature that is seen at this resolution is that certain intercalating densities apparently belonging to Sid seem to restrict the interactions amongst the subunits (Figure 15). The significance of this is not immediately understood. We would require higher resolution structure of the procapsid to see finer connections between the Sid and the subunits. The interaction of Sid with two domains of the hexamer is seen in the current model (Figure 14, Figure 17) but the likely interaction with two other domains () is not apparent in the structure. The pentamers have a slightly different arrangement of the gpN protein (Figure 16) and the domains in gpN may be more clustered together in the pentamer as is evident from the raised portion in the center of the pentamer.

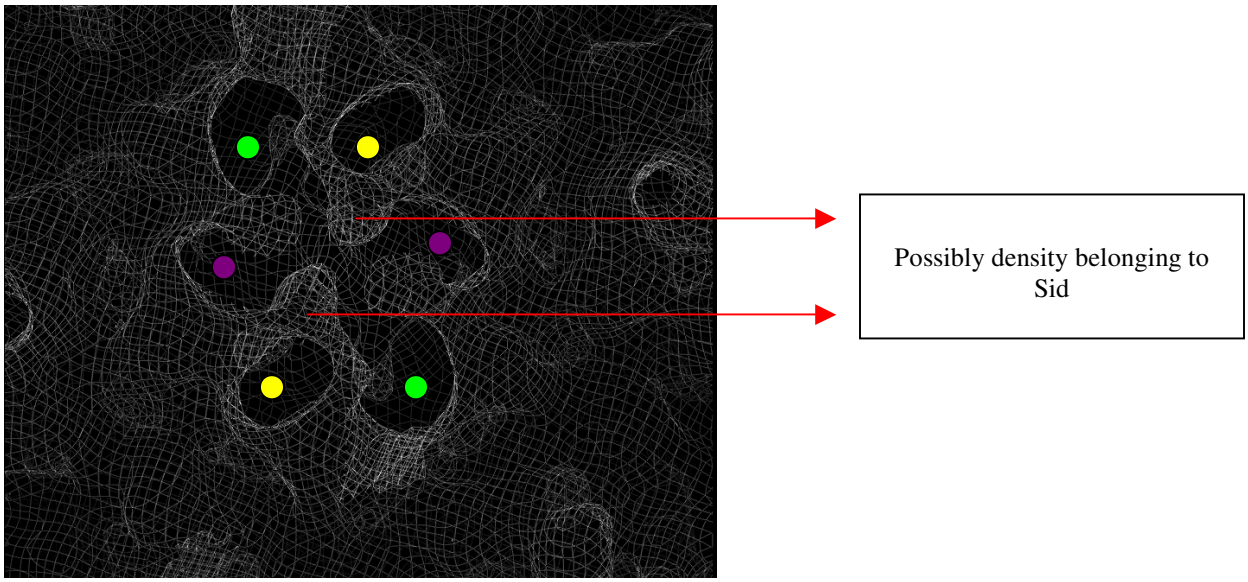


Figure 15:A view of the hexamer in “O” after slabbing sufficiently to see the outlines of the shapes of the different hexameric subunits. Similar colors dotted subunits, represent the same symmetric subunit

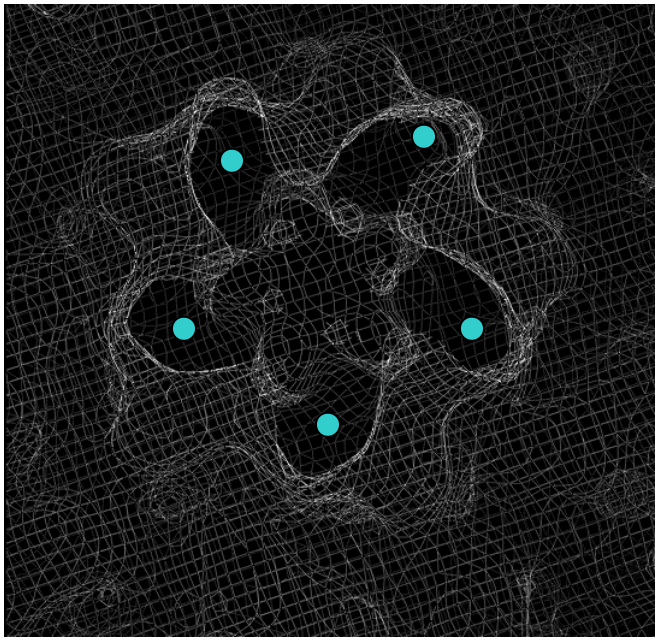


Figure 16:View of pentamer containing the same symmetric subunit in 5 copies

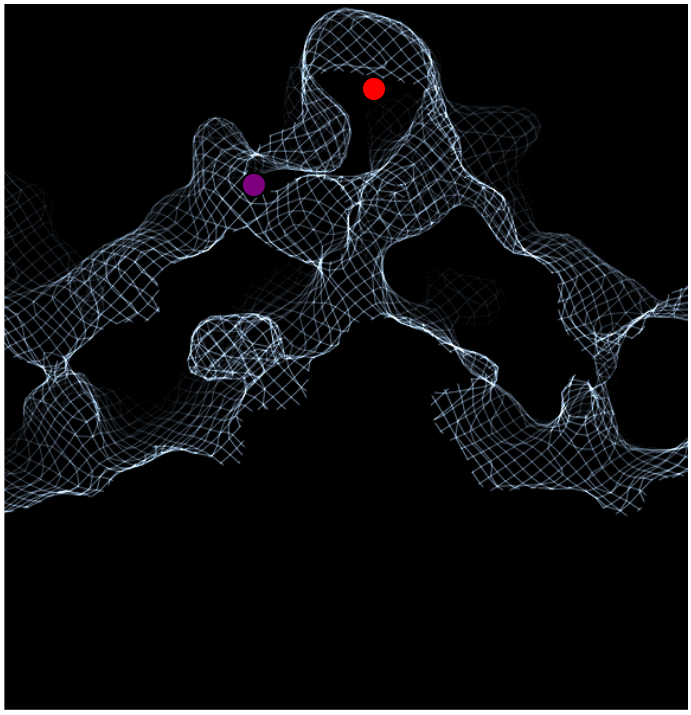


Figure 17: Shows the continuous density between Sid and a hexamer subunit. The colored dot indicates the same subunits as in figure 13 while the red dot stands for possible Sid density

3.2 Fitting the HK97 capsid head II protein gp5 into the P4 procapsid:

Studying a protein structure in isolation cannot give a complete understanding of its interactions with other proteins or DNA/RNA. EM maps of protein complexes are very important towards getting a larger picture of the protein interactions by docking available atomic models of the respective proteins or homologous proteins into these maps. This leads to the complete functional elucidation of the complex and the proteins therein. Hence fitting atomic models of protein into EM maps of their complexes is gaining importance.

A vast majority of dsDNA-tailed phages have common ancestry and they undergo profuse exchange of functional genetic elements drawn from a large shared pool. Phages frequently have specific functional properties in common (for example, the cross linked head proteins of L5 and HK97 and ribosomal frame shifting in tail genes) even when sequence similarity is not evident. The combination of these functional, organizational, and sometimes sequence similarities suggests that a significant number of dsDNA bacteriophages are in fact related, having had and continuing to have access to a large pool of functional genetic elements (18). The HK97 capsid head II was solved by crystallography (15) to a resolution of 3.6 Å. HK97 like P4 is a double stranded DNA phage virus but with a T=7 symmetry. The HK97 capsid protein Gp5 undergoes a 102-residue proteolysis of its Δ domain to form procapsid II from its earliest precursor procapsid I. Procapsid II then forms the head II capsid. The head II capsid protein's conformation was adapted and modeled to fit prohead II (6).

Crystallization attempts of P4 procapsids and expanded capsids, similar to the crystallization of the HK97 capsid Head-II were not very successful. Crystallization of P4

procapsids did not produce good quality crystals in the past and in case of the expanded capsids the heterogeneous sizes of the capsids in the sample is not pure enough for crystallization.

In order to investigate the possibilities of docking the proheadII pseudo atomic model in P4, the sequences of the two, capsid proteins gpN (P4) and gp5 (HK97) were compared to check for any significant sequence similarity. An output of ClustalW (59) showing the multiple alignment of the two sequences is given below. The secondary structure prediction using Predict Protein software (58), for each of the sequences was also done.

CLUSTALW Results

```
gigpn          1 --MRQETRFKFNAYLSRVAELNGIDAGDVSKKFTVEPSVTQTLMNTMQESSDFLTRINIV
gigp5          1 MSELALIQKAIEESQQKMTQLFDAQKAEIESTGQVSKQLQSDLMKVQEELTKSGTRLFDL
consensus      1 ms              rv  L     gdv     V   v   LM    E s   TRi  v

gigpn         59 PVSEMKGEKIGIG-----------VTGSIASTTDTAGGTERQPKDFSKLASNKYECDQIN
gigp5         61 EQKLASGAENPGEKKSFSERAAEELIKSWDGKQGTFGAKTFNKSLGSDADSAGSLIQPMQ
consensus     61       G      kksfseraaeev  S      T Gg   q    S   S       in

gigpn        108 FDFYIRYKTLDLWARYQDFQLRIRNAIIKRQSLDFIMAGFNGVKRAETSDRSSNPMLQDV
gigp5        121 IPGIIMPGLRRLTIRDLLAQGRTSSNALEYVREEVFTNNADVVAEKALKPESDITFSKQT
consensus    121     I      L  R    Q R     i     d        V        S

gigpn        168 AVGWLQKYRNEAPARVMSKVTDEEGRTTSEVIR-------------VGKGGDYASLDALV
gigp5        181 ANVKTIAHWVQASRQVMDDAPMLQSYINNRLMYGLALKEEGQLLNGDGTGDNLEGLNKVA
consensus    181 A          A   VM             vi glalkeegqllng G G     L  l

gigpn        215 MDATNNLIEPWYQEDPDLVVIVGRQLLADKYFPIVNKEQDNSEMLAADVIISQKRIGNLP
gigp5        241 TAYDTSLNATGDTRADIIAHAIYQVTESEFSASGIVLNPRDWHNIALLKDNEGRYIFGGP
consensus    241       L          l   v      d     v         lA       k I   P

gigpn        275 AVRVPYFPADAMLITKLENLSIYYMDDSHRRVIEENPKLDRVENYESMNIDYVVEDYAAG
gigp5        301 QAFTSNIMWGLPVVPTKAQAAGTFTVGGFDMASQVWDRMDATVEVSREDRDNFVKNMLTI
consensus    301             li    n    y             klD          D  V

gigpn        335 CLVEKIKVGDFSTPAKATAEPGA--
gigp5        361 LCEERLALAHYRPTAIIKGTFSSGS
consensus    361    Eki vg f   A   a    gs
```

Secondary structure Prediction:

GpN

```
          ....,....1....,....2....,....3....,....4....,....5....,....6....,....7....,....8....,....9....,....1
AA        MRQETRFKFNAYLSRVAELNGIDAGDVSKKFTVEPSVTQTLMNTMQESSDFLTRINIVPVSEMKGEKIGIGVTGSIASTTDTAGGTERQPKDFSKLASNK
PHD_sec     HHHHHHHHHHHHHHHH               HHHHHHHHHHHHHHHHHH EEEEE      EEEE EEEEEEEE

          ....,....11...,....12...,....13...,....14...,....15...,....16...,....17...,....18...,....19...,....20
AA        YECDQINFDFYIRYKTLDLWARYQDFQLRIRNAIIKRQSLDFIMAGFNGVKRAETSDRSSNPMLQDVAVGWLQKYRNEAPARVMSKVTDEEGRTTSEVIR
PHD_sec    EEEE     HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHEE                  HHHHHHHHHHHHHHH HHHHHH          EEEEE

          ....,....21...,....22...,....23...,....24...,....25...,....26...,....27...,....28...,....29...,....30
AA        VGKGGDYASLDALVMDATNNLIEPWYQEDPDLVVIVGRQLLADKYFPIVNKEQDNSEMLAADVIISQKRIGNLPAVRVPYFPADAMLITKLENLSIYYMD
PHD_sec   E     HHHHHHHHHHHHHHH HHH      EEEEEEHHHH              HHHHHHHHHHHHHHH              EEEEEEE   EEEE

          ....,....31...,....32...,....33...,....34...,....35...,....36
AA        DSHRRVIEENPKLDRVENYESMNIDYVVEDYAAGCLVEKIKVGDFSTPAKATAEPGA
PHD_sec     EE         EEEE      EEEEE     EEEEEEEE
```

Gp5

```
          ....,....1....,....2....,....3....,....4....,....5....,....6....,....7....,....8....,....9....,....1
AA        MSELALIQKAIEESQQKMTQLFDAQKAEIESTGQVSKQLQSDLMKVQEELTKSGTRLFDLEQKLASGAENPGEKKSFSERAAEELIKSWDGKQGTFGAKT
PHD_sec     HHHHHHHHHHHHHHHHHHHH
```

has 39.74% helices, 15.84% extended sheets and 44.42% loops. The sequence in gpN aligned to the domain P in gp5 is also principally helical and hence this domain could be conserved in gpN. Sequence in gpN aligned to Domain A, shows some variance but also seem to have two helices and an extended sheet like domain A. The domain E is not conserved in gpN and hence the sequence from 144 to 182 is cut off from the pdb structure of gp5 before fitting into the EM map of P4 procapsid and it is interesting to note that the sequence aligned to domain E in gpN is principally helical and this gives rise to the interesting speculation of the presence of a totally different domain in P4. Based on this assumption, the PDB structure of the prohead II without the N-terminal region and domain E, was docked into the ~14A map of P4 procapsid with scaffolding protein Sid, using O (63). The seventh chain in the proheadII belonging to the pentamer was removed and docking done with only the hexamer. The reason for this is that the packing between hexamers and pentamers differ for phages with different symmetries, as is the case here. The chain belonging to the pentamer was then fit separately into a pentamer in the EM map by visual inspection and using a program called “apply60f “ was generated for the rest of the icosahedron.

Figure 20 shows the pdb structure of proheadII without the N-Terminal region and the non-homologous sequence to domain E (144 to 182 Amino acids), docked into the hexamer of P4 Procapsid. This is the view looking down the pseudo six-fold axis or icosahedron two-fold axis. The 6 chains are colored in different colors. The P4 EM density is shown in grey.

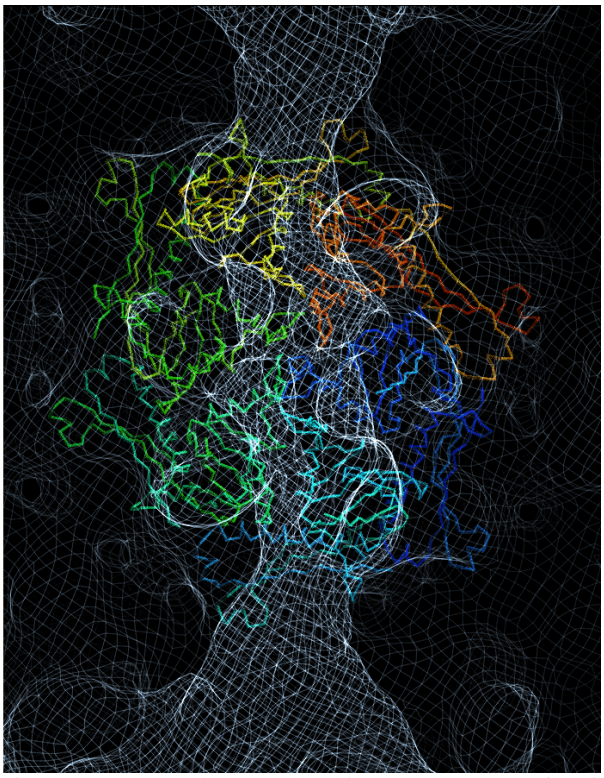


Figure 18 shows a hexamer looking down the two-fold icosahedral axis using “O”.

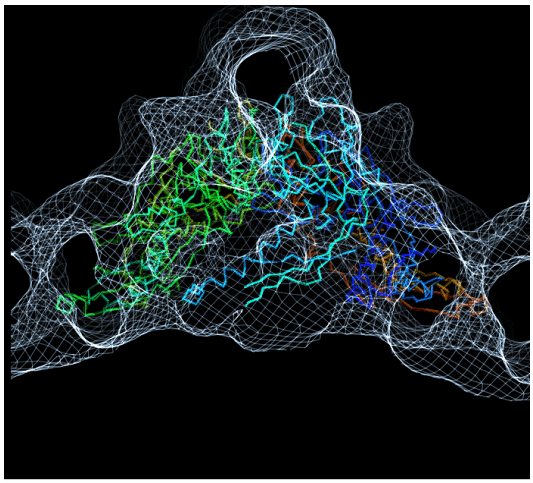


Figure 19 shows the view of the hexamer after rotating around the x-axis by 90 degrees after slabbing adequately to make out the inner features, using “O”.

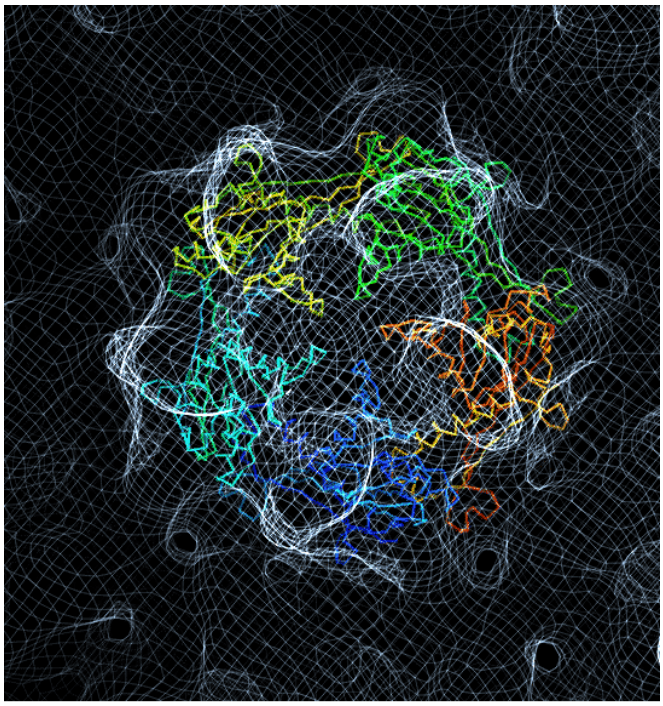


Figure 20 shows a view of a pentamer down the five-fold using “O”.

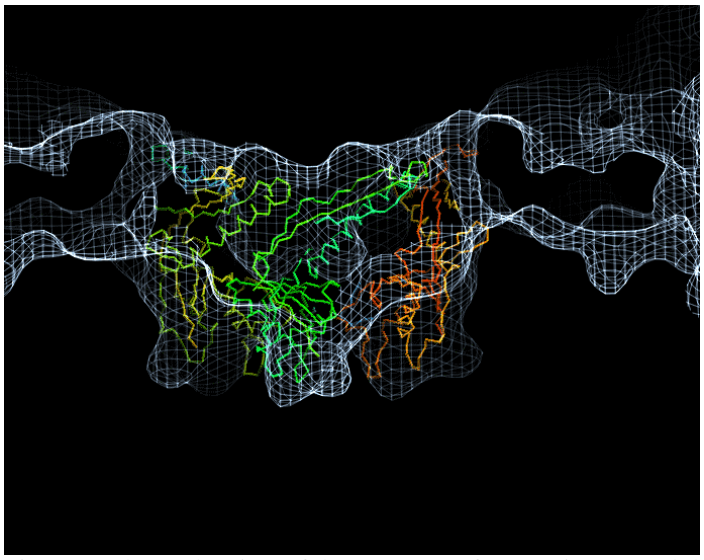


Figure 21 shows the view of the pentamer after rotating around the x-axis by -90 degrees after slabbing adequately to make out the inner features, using “O”.

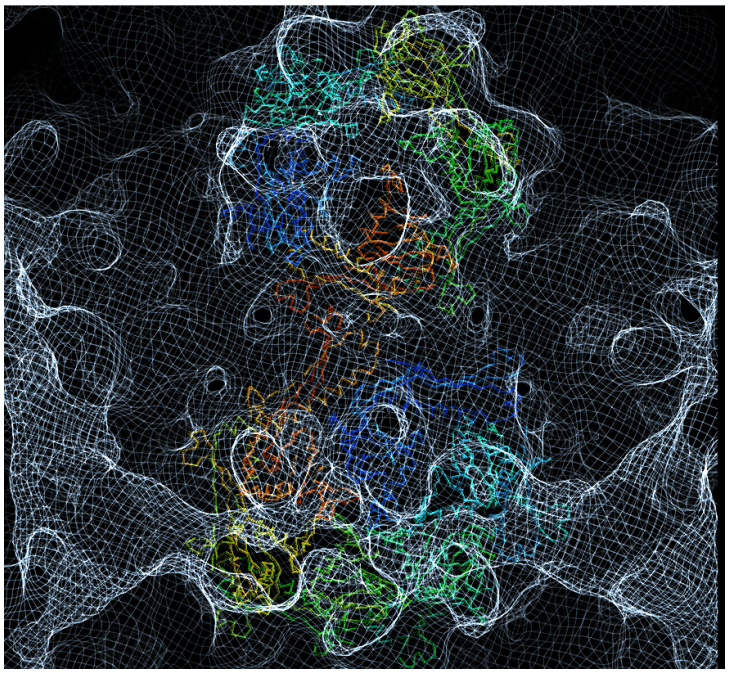


Figure 22 shows both the pentamer and hexamer using “O”.

The fitting of the proheadII into the P4 EM map shows that there are some common conserved domains in the capsid proteins of the two phages. However there are also gaps in the EM map not filled by the atomic model indicating the presence of extra domains in gpN. One possibility is the sequence in gpN aligned to the domain E in gp5. The N-terminal region in gpN is also different as compared to gp5 and is not known to undergo any proteolysis unlike gp5 in the formation of P4 procapsids. Also, in the hexamer there is some gap in the twisted slab like regions after fitting domain A in gp5. This could be because the homologous domains in gpN may be positioned slightly differently. The differential arrangement of the subunits in the hexamer maybe a result of intercalation of Sid within the hexamer (Figure 15). In the pentamer the fitting seems better because the chain belonging to the pentameric unit was moved slightly to fit the map better before using apply60f to generate similar chains in the remaining pentameric subunits. Also there is no interference of Sid in this region.

A previous study (29) had indicated the presence of two domains in gpN with one domain comprising the capsomer and the other the trimeric connection between the capsomers with a hinge region connecting them. It is possible that the sequence homologous to domain A forms the domain comprising the capsomer and the sequence homologous to domain P, the shell region along with the N terminal region and the sequence aligned to domain E.

3.3 Structure of Expanded capsid:

Upon storage of these purified procapsids for about 2 months at 4°C, some of the procapsid samples were found to have undergone spontaneous proteolytic degradation of gpN. Presumably the cleavage of gpN had been caused by an unknown contaminating protease, as other samples stored in the same way were unaffected. When the spontaneously cleaved samples were rerun on sucrose gradients as explained above, uncleaved gpN and Sid sedimented in the position expected for normal procapsids, while cleaved gpN was found in a more predominantly slowly sedimenting fraction (1). Negative stain EM as well as cryo-EM confirmed that the Sid-containing band contained procapsids of normal appearance (Figure 12), while the band containing cleaved protein comprised mostly expanded, thin-shelled particles (Figure 23). Hence it is possible that the spontaneous cleavage of gpN had caused an expansion of the procapsids with associated loss of Sid to form expanded capsids.

These expanded particles were studied using cryo-EM as described in section 2.4. The expanded particles showed a wide variation in size from 250 to 350 Å in diameter. About 20–30 particles could be found in each negative. These particles were thin-shelled, about 450 Å in diameter and had a somewhat angular shape with a smooth outline (Figure 23).

In total 316 particles were chosen from 30 negatives based on the fact that they were more uniform in shape. Preliminary orientations were generated for 92 of the best particles using the program EMICOFV, and a map was calculated to 35 Å resolution. Initially only 92 of the best particles were considered and a model generated. In successive iterations more and more particles were included, with the final number of particles considered coming to 247, with a PFT correlation of 0.4. EMAN (16), described in detail in the above section was used initially to get a model for 156 best, expanded particles. Although this procedure yielded a reconstruction with $T = 4$, due to concerns that the number of particles used maybe insufficient for the EMAN classification the more traditional approach with the common lines-based MRC/ICOS programs (21,22) was taken. These programs are more forgiving of having small numbers of particles, probably due to the use of exact orientations and interpolation in the Fourier transforms (48). In the end, the reconstruction produced by this procedure was quite similar to that determined by EMAN, demonstrating the equality of these two fundamentally different reconstruction algorithms (1).

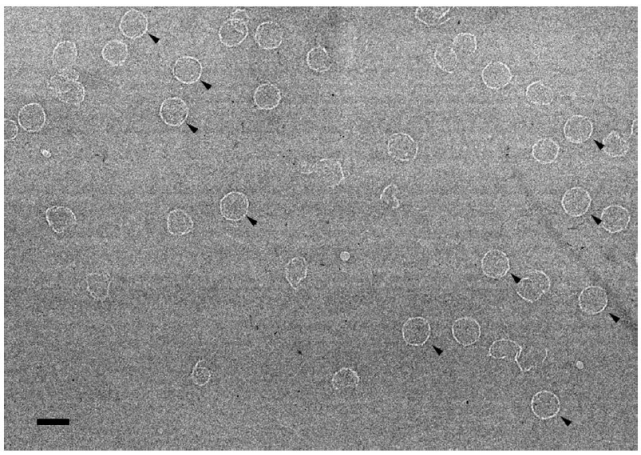


Figure 23: Cryo-EM micrograph of P4 expanded capsids
Source: Wang et al. Cleavage leads to expansion of bacteriophage P4 procapsids in vitro. Virology 314: pp1-8

The Fourier shell correlation (fsc) between two reconstructions calculated to 24 Å resolution for the P4 expanded capsid data set, from two separate halves of the data dropped steeply at about 29 Å resolution, suggesting that this is the effective resolution limit of the data (figure 24).

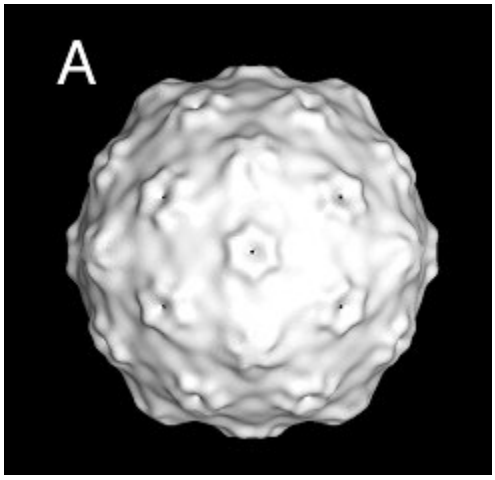


Figure 24: expanded capsid at 29 Å resolution.
Source: Wang et al. Cleavage leads to expansion of bacteriophage P4 procapsids in vitro. Virology 314: pp1-8

Failure to reach higher resolution is probably mainly due to slight structural differences in the particles chosen for reconstruction.

Hence, capsid expansion and scaffold removal can be triggered in vitro by cleavage of gpN with trypsin in the absence of genome packaging or other structural proteins. Thus, the role of gpN cleavage may be to destabilize the gpN–gpN interactions and make the particle more susceptible to expansion. Due to this expansion the Sid scaffolding is removed. After trypsin treatment there were two cleavage sites in gpN an N-terminal and C-terminal cleavage site. The N-terminal cleavage site is almost identical (within two residues) of the native cleavage site for gpN (33) while the C-terminal could be an artifact of trypsin treatment. To be sure that the expanded capsids are not an aberrant by

product of this C-terminal cleavage N-mutants without this cleavage site should be assembled and analyzed (1).

Another interesting study would be to try and block the protease involved in the cleavage of gpN and thereby block the maturation and DNA assembly. Unfortunately the exact protease in vivo is not known, and indeed the gpN cleavage may be autocatalytic. It is also known that gpO is involved in this in vivo though, gpO in itself may not be a protease. Producing gpO in vitro has also been quite difficult. Hence checking out these possibilities is beyond the scope of this present investigation.

3.4 Comparison between the P4 Procapsid and P4 expanded capsid and conclusions on the fitting of proheadII capsid into P4 procapsid:

From this it is clear that the P4 capsid undergoes rearrangement of subunits in the process of maturation, which is important for the unfolding of protein-protein interactions and DNA packing subsequently. An exact comparison between the expanded capsid and procapsid of P4 is not possible as they are at very different resolutions.

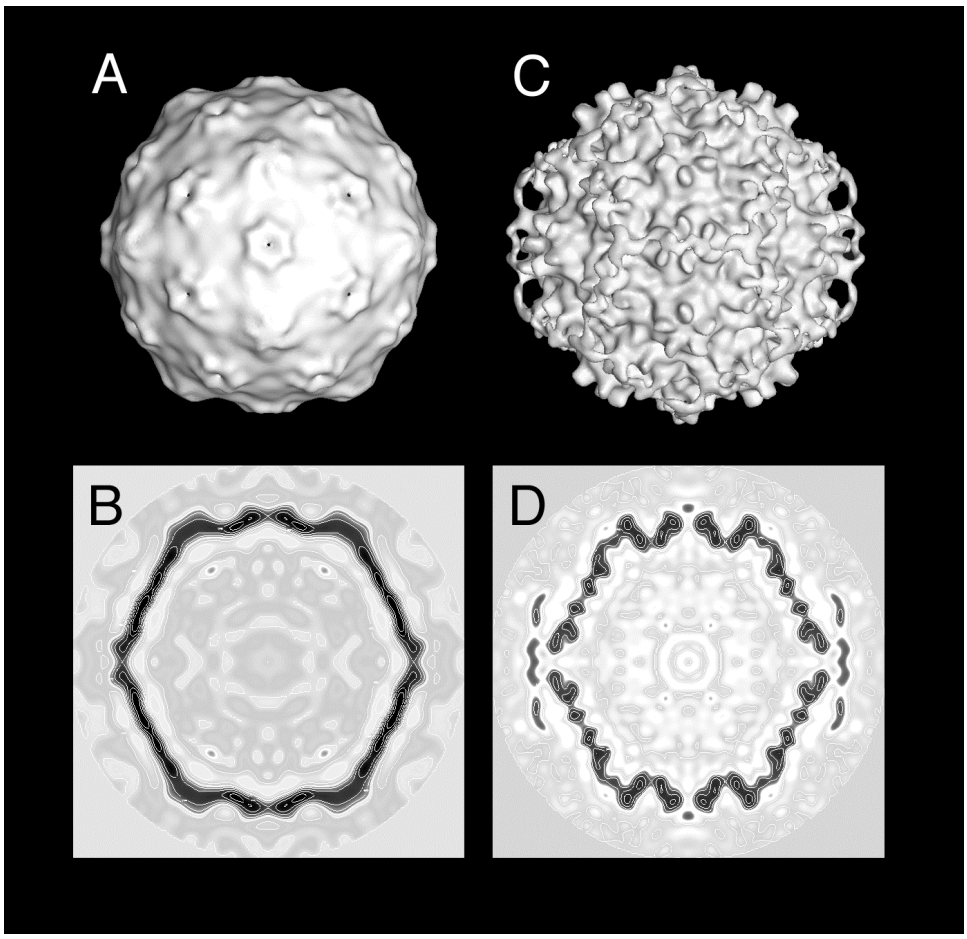


Figure 25: Comparison between P4 procapsid and P4 expanded capsid
Source: Wang et al. Cleavage leads to expansion of bacteriophage P4 procapsids in vitro. Virology 314: pp1-8

The P4 Procapsid hexamers have an elongated appearance characteristic of the dsDNA phage procapsids and has a two-fold symmetry. Unlike the procapsid structure, the expanded capsid has a flattened, symmetric capsomer and a thinner shell, more closely resembling the structure of the mature virion (29). The hexamers almost have a perfect six-fold symmetry. However, the mature virion appears to have a more distinctly angular shape than the expanded procapsid. It could be that although the shell had expanded it had not gone through the complete angularization process typical of mature virions. Thus,

P4 capsid maturation might proceed through several steps, most likely involving the DNA and other proteins in vivo (1).

The fitting of the proheadII atomic model into the P4 procapsid EM map illustrates that the construction principles within a capsomer are essentially identical for most dsDNA phages. However, the gaps in the fitting indicate that in spite of some common domains, the capsid proteins have evolved differently within the same dsDNA phage family. The gaps in the P4 EM model, based on the multiple sequence alignment and secondary structure prediction may be due to the presence of some other domains like the predicted helical domain in gpN corresponding to domain E in gp5 or due to the N-terminal region not used in the fitting. Also, the gap in the twisted slab like regions of the hexamer after fitting domain A in gp5 could indicate that the homologous domains in gpN may be shifted slightly with respect to each other due to the interaction of Sid with two of the subunits (Figure 15). In an earlier paper (4) it was discussed that in P4 phage, unlike P2 phage only 4 subunits get cleaved from N to N* form and that Sid interaction with the N-terminal of gpN protein in the remaining two subunits, prevent them from getting cleaved. Based on the assumption that the present fitting of gp5 is correct and that gpN and gp5 share optimum homology on a structural basis, in the present model it appears that Sid does interact with two subunits but the interaction is most likely with the sequence in gpN homologous to domain A and this is close to C-terminal. It is clear that a higher resolution structure is required to know how and where Sid interacts with the subunits in the hexamer.

Improving the resolution of the Ribosome channel complex:

The present structure (Figure 26) of the ribosome channel complex using poly-lysine treated grid surface is at 20 Å (2).

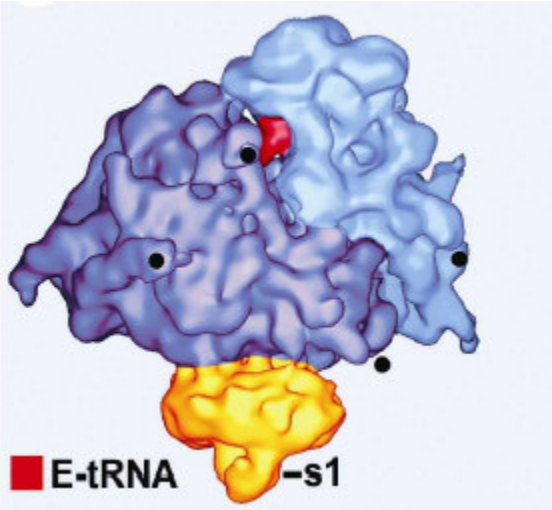


Figure 26: Ribosome channel complex. The dark blue region is the large ribosomal subunit while light blue represents the small ribosomal subunit. The golden region represents the channel. S1 is the stalk of TRAP. The black dots represent truncated RNA helices due to interaction with poly lysine grid.
Source: Ménétret et al, . Architecture of the Ribosome-Channel Complex Derived from Native Membranes (To be published)

The boxed particles would be used to improve this current model to a higher resolution with the aim being to reach a resolution at which the localized helices will begin to be apparent and would allow the study of the channel in greater detail. The main purpose of this exercise was to study the robustness of EMAN in dealing with both symmetrical (P4 phage) and non-symmetrical (ribosome channel complex) particles and to grasp the importance of 3D reconstruction in studying protein complexes be it virus assembly or ribosome channel complexes.

3.6 Future prospects:

Higher resolution maps below 10 Å are required for both the P4 phage and the ribosome channel complex to identify helices, which can be used to do a better fitting of the atomic

models. A higher resolution map can help in identifying the densities for the more flexible regions too. Also, the gpN structure could be modeled from gp5 using threading, as there is sufficient homology between the gp5 and gpN proteins and used to fit into the higher resolution EM map of the P4 procapsid as well as expanded capsid.

References:

1. Sifang Wang, Preethi Chandramouli, Sarah Butcher, and Terje Dokland (2003). Cleavage leads to expansion of bacteriophage P4 procapsids in vitro.Virology 314: pp1-8

2. Jean-François Ménétret, Ramanujan S. Hegde, Sven U. Heinrich, Steven J. Ludtke, Tom A.Rapoport and Christopher W. Akey. Architecture of the Ribosome-Channel Complex Derived from Native Membranes (To be published)

3. Fons, R.D., Bogert, B.A. and Hegde, R.S. (2003/in press). The translocon-associated protein complex facilitates the initiation of substrate translocation across the ER membrane. J. Cell Biol.

4. T. Dokland, S. Wang & B. H. Lindqvist (2002). The structure of P4 procapsids produced by co-expression of capsid and external scaffolding protein. Virology 298: pp 224-231

5. David Gene Morgan, Jean-Francois Ménétret, Andrea Neuhof, Tom A. Rapoport and Christopher W. Akey. Structure of the Mammalian Ribosome–Channel Complex at 17Å Resolution. J.Mol. Biol. (2002) 324, 871–886

6. J. F. Conway,W. R. Wikoff, N. Cheng, R. L. Duda,R. W. Hendrix, J. E Johnson, A.C. Steven. Virus Maturation Involving Large Subunit Rotations and Local Refolding SCIENCE (2001) 292,745-748

7. Yusupov, M. M., Yusupova, G. Zh., Baucom, A.,Lieberman, K., Earnest, T. N., Cate, J. H. D. & Noller,H. F. (2001). Crystal structure of the ribosome at 5.5 resolution. Science, 292, 883–896

8. Ali Saad, Steven J. Ludtke, Joanita Jakana, Frazer J. Rixon, Hiro Tsuruta, and Wah Chiu (2001). Fourier Amplitude Decay of Electron Cryomicroscopic Images of Single Particles and Effects on Structure Determination. Journal of Structural Biology 133:pp 32–42

9. Ludtke, S.J., Jakana, J., Song, J.L., Chuang, D.T., Chiu, W., 2001. A 11.5 A single particle reconstruction of GroEL using EMAN. J. Mol. Biol. 314, 253–262

10. Beckmann, R., Spahn, C.M., Eswar, N., Helmers, J., Penczek, P.A., Sali, A., Frank, J., Blobel G. (2001). Architecture of the protein-conducting channel associated with the translating 80S ribosome. Cell 107,361-372

11. Ménétret, J. F., Neuhof, A., Morgan, D. G., Plath, K., Rademacher, M., Rapoport, T.A., & Akey, C.W. The structure of ribosome-channel complexes engaged in protein translocation. Molec. Cell 6 (2000), Pages 1219-1232

12. P.A. Thuman-Commike, W. Chiu. Reconstruction principles of icosahedral virus structure determination using electron cryomicroscopy . Micron 31 (2000) 687–711

13. Wang, S., Palasingam, P., Nøkling, R. H., Lindqvist, B. H., and Dokland, T. (2000). In vitro assembly of bacteriophage P4 procapsids from purified capsid and scaffolding proteins. Virology 275, 133–144

14. Terje Dokland (2000). Freedom and restraint: themes in viral capsid assembly. Structure 8: pp R157-62

15. Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE. Topologically linked protein rings in the bacteriophage HK97 capsid. Science. 2000 Sep 22; 289(5487):2129-33

16. Ludtke, S.J., Baldwin, P.R., Chiu, W., 1999. EMAN: Semiautomated software for high-resolution single-partice reconstructions. J. Struct. Biol. 128, 82–97.

17. Matlack, K. E. S., Mothes, W., and Rapoport, T. A. (1998). Protein translocation-tunnel vision. Cell *92*,381-390.

18. Roger W. Hendrix, Margaret C. M. Smith, R. Neil Burns, Michael E. Ford, and Graham F. Hatfull. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. Proc Natl Acad Sci U S A. 1999 March 2; 96 (5):2192–2197

19. Beckmann, R., Bubeck, D., Grassucci, R., Penczek, P., Verschoor, A., Blobel, G., and Frank, J. (1997). Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. Science *19*, 2123-2126

20. Zhu, J., Penczek, P. A., Schroder, R., and Frank, J. (1997) Threedimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: Procedure and application to the 70S Escherichia coli ribosome, *J.Struct. Biol.* **118,** 197–219.

21. Baker, T. S. and R. H. Cheng (1996). A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. J.Struct. Biol. 116:120-130

22. Fuller, S.D., Butcher, S.J., Cheng, R.H., Baker, T.S., 1996. Three-dimensional reconstruction of icosahedral particles—the uncommon line. J. Struct. Biol. 116, 48–55.

23. Hanein, D., Matlack, K. E. S., Jungnickel, B., Plath, K., Kalies, K.-U., Miller, K. R., Rapoport, T. A., and Akey, C. W. (1996). Oligomeric rings of the Sec61p complex induced by ligands required for protein translocation. Cell *87*, 721-732.

24. Marvik, O.J., Dokland, T.E., Nøkling, R.H., Jacobsen, E., Larsen, T., Lindqvist, B.H., 1995. The capsid size-determining protein Sid forms an external scaffold on phage P4 procapsids. J. Mol. Biol. 251, 59–75.

25. Marvik, O.J., Sharma, P., Dokland, T., Lindqvist, B.H., 1994. Bacteriophage P2 and P4 assembly: alternative scaffolding proteins regulate capsid size. Virology 200, 702–714

26. Rishovd, S., Marvik, O.J., Jacobsen, E., Lindqvist, B.H., 1994. Bacteriophage P2 and P4 morphogenesis: identification and characterization of the portal protein. Virology 200, 744–751.

27. Görlich, D., and Rapoport, T. A. (1993). Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. Cell *75*, 615-630.

28. Max T. Otten and Wim M.J. Coene (1993). High-resolution imaging on a field emission TEM. Ultramicroscopy 48 (1993) 77-91

29. Dokland, T., Lindqvist, B.H., Fuller, S.D., 1992. Image reconstruction from cryo-electron micrographs reveals the morphopoietic mechanism in the P2-P4 bacteriophage system. EMBO J. 11, 839–846.

30. Penczek, P., Radermacher, M. and Frank, J. (1992) Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy* **40**, 33-53 .

31. Görlich, D., Hartmann, E., Prehn, S. and Rapoport, T.A. (1992a) A protein of the endoplasmic reticulum involved early in polypeptide translocation. Nature. 357, 47-52.

32. Görlich, D., Prehn, S., Hartmann, E., Kalies, K.U., and Rapoport, T.A. (1992b). A mammalian homolog of Sec61p and SecYp is associated with ribosomes and nascent polypeptides during translocation. Cell 71, 489–503.

33. Rishovd, S., Lindqvist, B.H., 1992. Bacteriophage P2 and P4 morphogenesis: protein processing and capsid size determination. Virology 187,548–554.

34. Glaeser, R. M., and Downing, K. H. (1992) Assessment of resolution in biological electron crystallography, *Ultramicroscopy* **47,** 256–265

35. Brink, J., and Chiu, W. (1991) Contrast analysis of cryo-images of *n*-paraffin recorded at 400 kV out to 2.1 Å resolution, *J. Microsc.* **161,** 279–295

36. Bertani, L.E., Six, E., 1988. The P2-like phages and their parasite, P4, in:Calendar, R. (Ed.), The Bacteriophages, Vol. 2, 2 Vols. Plenum Press, New York, pp. 73–143.

37. Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowall, A.W., Schultz, P., 1988. Cryo-electron microscopy of vitrified specimens. Q. Rev.Biophys. 21, 129–228

38. Radermacher, M., Wagenknecht, A. Verschoor, A. and Frank, J. (1987) Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*. *J.Microsc.* **146**, 113-136.

39. Adrian, M., Dubochet, J., Lepault, J., McDowall, A.W., 1984. Cryoelectron microscopy of viruses. Nature (London) 308, 32– 36.

40. Downing, K. H., and Chiu, W. (1982) Effect of stray magnetic field on image resolution in transmission electron microscopy, *Ultramicroscopy* **5,** 351–356.

41. Wade, R. H., and Frank, J. (1977) Electron microscope transfer functions for partially coherent axial illumination and chromatic defocus spread, *Optik* **49,** 81–92.

42. Barrett, K.J., Marsh, M.L., Calendar, R., 1976. Interactions between a satellite bacteriophage and its helper. J. Mol. Biol. 106,683–707.

43. Taylor, K., and Glaeser, R. M. 1976. Electron Microscopy of frozen-hydrated biological specimens. J. Ultrastr. Res. 55. 448-456

44. Frank, J. (1976) Determination of source size and energy spread from electron micrographs using the method of Young's fringes, *Optik* **44,** 379–391.

45. Frank, J. (1973) The envelope of electron microscopic transfer functions for partially coherent illumination, *Optik* **38,** 519–536.

46. Thon, F. (1971) Phase contrast electron microscopy, *in* (U. Valdre,Ed.) Electron Microscopy in Material Sciences, pp. 571–625, Academic Press, New York.

47. Hanszen, K. J. (1971) The optical transfer theory of the electron microscope: Fundamental principles and applications, *in* (R.Barer and V. E. Cosslett, Eds.), *Advances in optical and electron microscopy* Vol. 4, pp. 1–84, Academic Press, New York.

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

49. Williams, R. C., and Fisher, H. W. (1970). Electron Microscopy of TMV under conditions of Minimal Beam Exposure. J. Mol. Bil. 52,121-123

50. Erickson, H. P., and Klug, A. (1970) The Fourier transform of an electron micrograph: Effects of defocussing and aberrations, and implications for the use of underfocus contrast enhancement, *Phil. Trans. R. Soc. Lond. B* **261,** 105–118

51. Radon, 1917, "Über die Bestimmung von Funktionen durch ihre Integral-werte längs gewisser Mannig-faltigkeiten. Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig", Math. Phys. Klasse, 69, 262-277.

52. http://www-db.embl-heidelberg.de /jss/servlet/de.embl.bk.wwwTools.GroupLeftEMBL/ExternalInfo/fuller/EMBL_Virus_Structure.html

53. http://ncmi.bcm.tmc.edu/~stevel/EMAN/doc

54. http://www.wadsworth.org/spider_doc/spider/docs/techs/ctf/ctf.html

55. http://www.strubi.ox.ac.uk/

56. http://cryoem.berkeley.edu/~nieder/em_for_dummies/cryo_em.html for information about cryo-EM and the techniques involved. © 1999, Snaggled Works, Hanspeter Niederstrasser /

57. http://www.vir.gla.ac.uk/staff/bhellad/cryoEM.htm

58. PHDsec: http://www.embl-heidelberg.de/predictprotein/ B Rost & C Sander: JMB, 1993, 232, 584-599.

59. http://www.ebi.ac.uk/clustalw/

60. http://www.cnb.uam.es/~bioinfo/paper/etbiospe/node4.html

61. http://w1.411.telia.com/~u41112276/strul3.html

62. Frank, J., 1996 Three-Dimensional Electron Microscopy of Macromolecular Assemblies. Academic Press Inc., San Diego, California

63. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A 47,110-119.