doi:10.1016/S0022-2836(02)01201-9 *J. Mol. Biol.* (2003) **325**, 11–24

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Structural Studies of Bacteriophage α3 Assembly

Ricardo A. Bernal[1], Susan Hafenstein[2], Norman H. Olson[1] Valorie D. Bowman[1], Paul R. Chipman[1], Timothy S. Baker[1] Bentley A. Fane[2] and Michael G. Rossmann[1]*

[1]*Department of Biological Sciences, Purdue University 1392 Lilly Hall, West Lafayette IN 47907-1392, USA*

[2]*Department of Veterinary Science and Microbiology University of Arizona Building 90, Room 201, Tucson AZ 85721, USA*

**Corresponding author*

Bacteriophage α3 is a member of the *Microviridae*, a family of small, single-stranded, icosahedral phages that include ϕX174. These viruses have an ssDNA genome associated with approximately 12 copies of an H pilot protein and 60 copies of a small J DNA-binding protein. The surrounding capsid consists of 60 F coat proteins decorated with 12 pentameric spikes of G protein. Assembly proceeds *via* a 108 S empty procapsid that requires the external D and internal B scaffolding proteins for its formation.

The α3 "open" procapsid structural intermediate was determined to 15 Å resolution by cryo-electron microscopy (cryo-EM). Unlike the ϕX174 "closed" procapsid and the infectious virion, the α3 open procapsid has 30 Å wide pores at the 3-fold vertices and 20 Å wide gaps between F pentamers as a result of the disordering of two helices in the F capsid protein. The large pores are probably used for DNA entry and internal scaffolding protein exit during DNA packaging. Portions of the B scaffolding protein are located at the 5-fold axes under the spike and in the hydrophobic pocket on the inner surface of the capsid. Protein B appears to have autoproteolytic activity that cleaves at an Arg-Phe motif and probably facilitates the removal of the protein through the 30 Å wide pores.

The structure of the α3 mature virion was solved to 3.5 Å resolution by X-ray crystallography and was used to interpret the open procapsid cryo-EM structure. The main differences between the α3 and ϕX174 virion structures are in the spike and the DNA-binding proteins. The α3 pentameric spikes have a rotation of 3.5° compared to those of ϕX174. The α3 DNA-binding protein, which is shorter by 13 amino acid residues at its amino end when compared to the ϕX174 J protein, retains its carboxy-terminal-binding site on the internal surface of the capsid protein. The icosahedrally ordered structural component of the ssDNA appears to be substantially increased in α3 compared to ϕX174, allowing the building of about 10% of the ribose-phosphate backbone.

© 2002 Elsevier Science Ltd. All rights reserved

Keywords: bacteriophage α3; three-dimensional structure; procapsid; morphogenesis; ϕX174

Introduction

Microviridae are small, icosahedral, (+) single-stranded, circular DNA bacteriophages that include ϕX174, G4, and α3.[1] The morphogenesis of these phages proceeds through an empty procapsid intermediate (Figure 1). ϕX174 has been the most studied, with the structures of the procapsid and virion having been determined by cryo-electron microscopy (cryo-EM) and X-ray crystallography.[2–4] However, differences between the cryo-EM and X-ray procapsid structures raise questions regarding the true nature of the biologically significant intermediate. The cryo-EM structure at 26 Å resolution has prominent pores at all the 3-fold axes of symmetry ("open" procapsid) through which DNA is presumed to package. In contrast, these pores are blocked by three symmetry-related α-helices of protein F in the crystal

Abbreviations used: cryo-EM, cryo-electron microscopy; GST, glutathione-*S*-transferase; NCS, non-crystallographic symmetry; r.m.s., root-mean-square; ssDNA, single-stranded DNA.

E-mail address of the corresponding author: mgr@indiana.bio.purdue.edu

 © 2002 Elsevier Science Ltd. All rights reserved

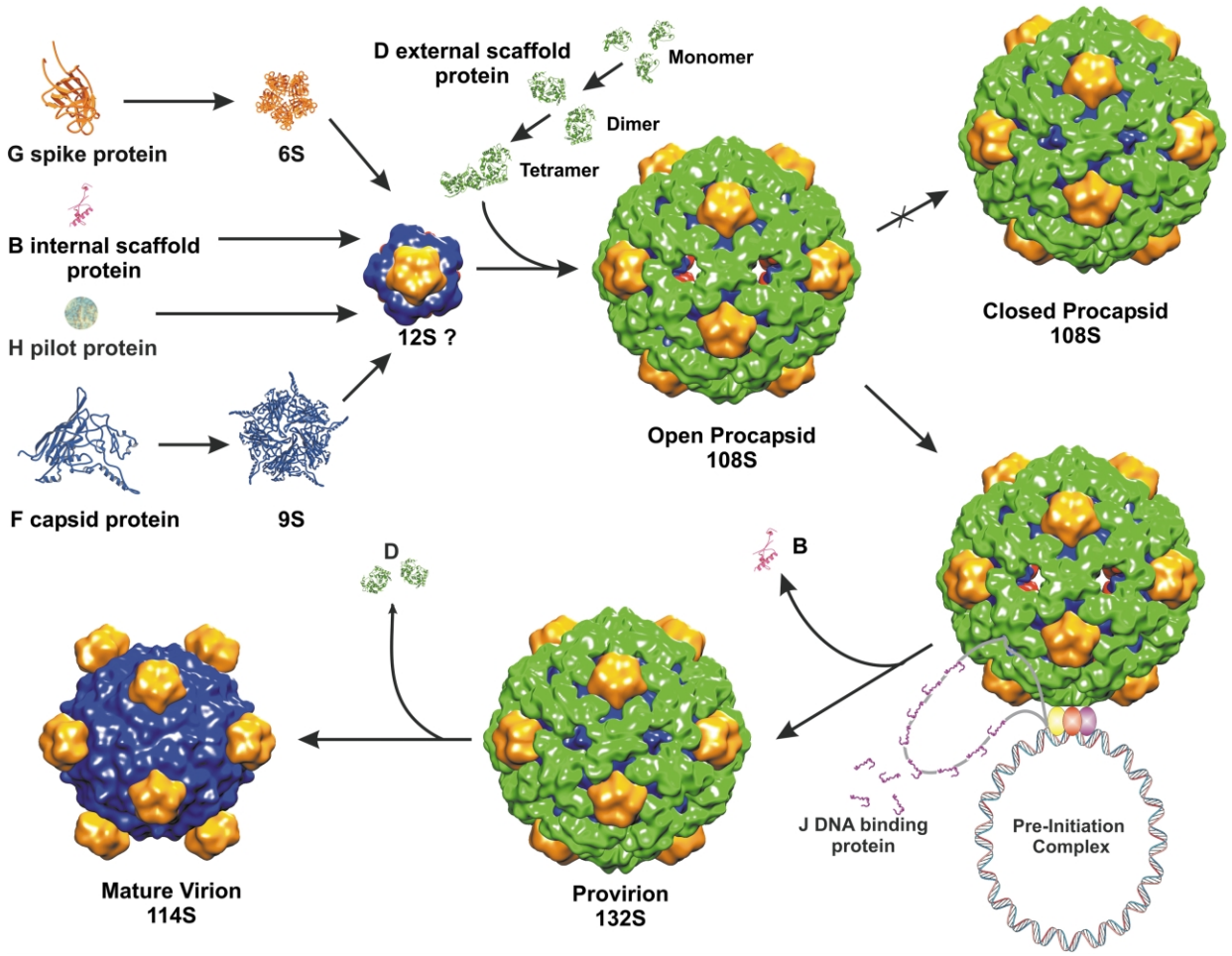


Figure 1. *Microviridae* assembly pathway. The first morphogenetic intermediates in the assembly pathway are the 9 S and 6 S particles, respective pentamers of the F capsid and G spike proteins.[1] The internal and external scaffolding proteins direct the assembly of these pentameric intermediates into an empty protein shell called the procapsid.[39–41] The single-stranded DNA (ssDNA) is concurrently synthesized and packaged along with the basic DNA-binding protein J, which displaces the internal scaffolding protein B by competition for the same hydrophobic-binding pocket on the internal surface of the F capsid protein.[3,4,42] The resulting particle, or provirion, sheds the external scaffolding protein lattice forming the mature virion.

structure, the "closed" procapsid. This structure is similar to that of the mature virion, except for the presence of the scaffolding proteins, suggesting that the crystallization process in ammonium sulfate may have induced a conformational change mimicking the maturation process.

To address some of these questions, studies were initiated with the related bacteriophage α3, with the aim of obtaining a more stable open procapsid. Bacteriophage α3 has an overall amino acid identity of 59% with ϕX174[5–7] (Table 1), the most conserved protein being the F capsid protein with an identity of 72% (Figure 2). The least conserved is the spike protein, which has an identity of only 31%. The α3 virion and procapsid structures reported here at 3.5 Å and 15.0 Å resolution, respectively, demonstrate that the procapsid to virion transition includes a radial collapse of the capsid pentamers around the genome. In addition, there are major conformational changes at the 2- and 3-fold axes of symmetry. Furthermore, the improved resolution shows that the density around the 5-fold axes is a mixture of the H pilot and B internal scaffolding proteins. This suggests that the *Microviridae* internal scaffolding protein, like that of P22[8,9] and ϕ29 (M. Morais, personal communication) may play some role in minor vertex protein incorporation.

Table 1. Amino acid identity for proteins F and G

	Protein F (%)	Protein G (%)
α3-ϕX174	72	31
α3-G4	64	30
G4-ϕX174	67	41

Results and Discussion

The mature α3 virion

The structure of the α3 virion was determined to 3.5 Å resolution by X-ray crystallography. The

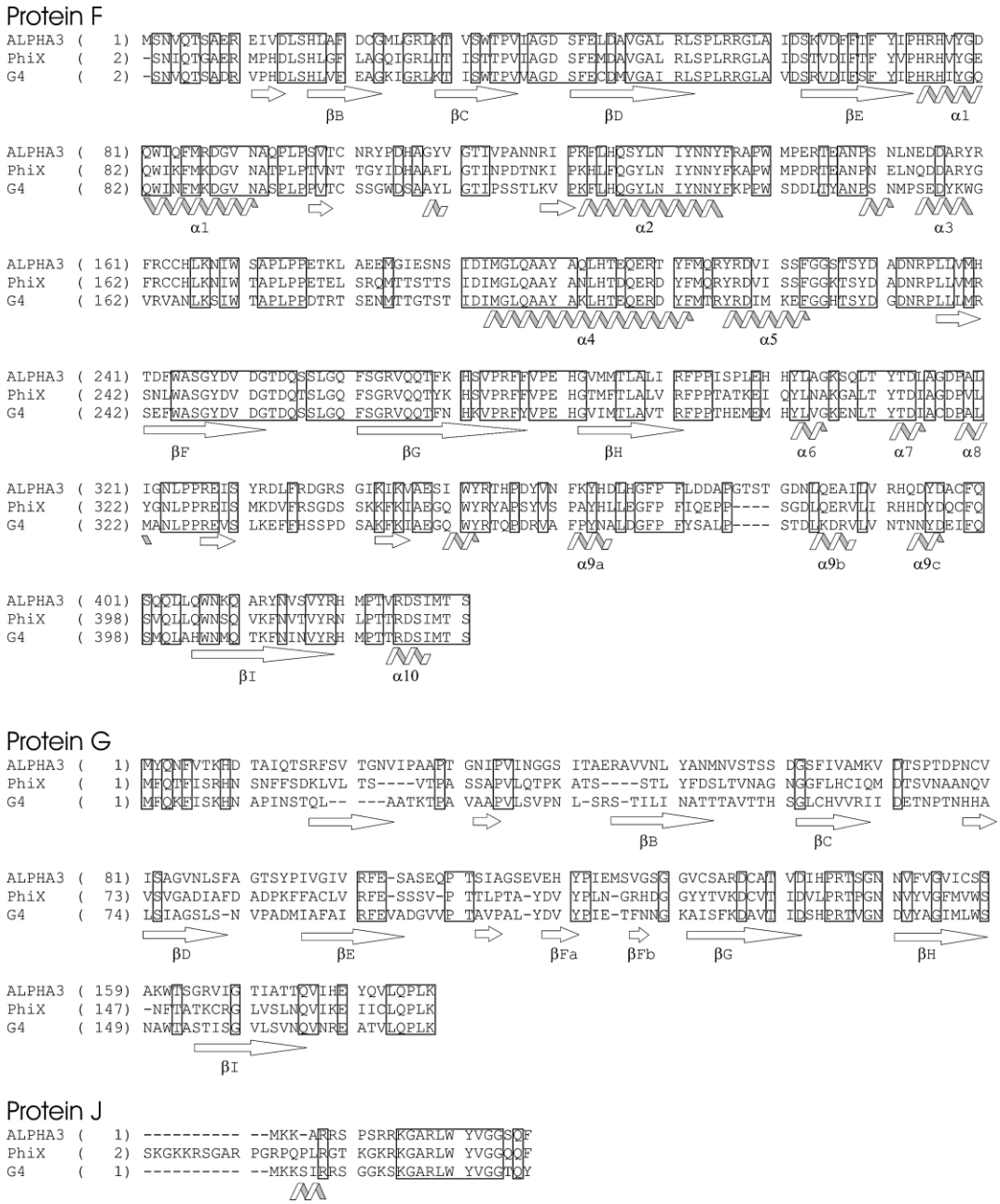


Figure 2. Amino acid alignment based on the structural superpositions of proteins F, G, and J for bacteriophages α3, ϕX174, and G4. Alpha helices α1 to α10 (/\) in F and β-strands βB to βI (⇨) are labeled to be consistent with the nomenclature established for ϕX174. Completely conserved residues are boxed.

virion is icosahedral with 60 copies each of the F capsid, G spike, and J DNA-binding proteins. In addition, there are at most 12 copies of the H pilot protein. As with bacteriophage ϕX174, the main feature of the F capsid protein is an eight-stranded antiparallel β-barrel (Figures 3 and 4). However, the F β-barrel has two large insertions containing 12 α-helices. The longest of these helices is helix 4, which is positioned near the 3-fold axes of symmetry. The spike protein also contains the β-barrel motif but, unlike the capsid protein, it has no large insertions. This antiparallel β-barrel motif is found in many plant and animal viruses, such as tomato bushy stunt virus[10] and human rhinovirus.[11,12]

Most of the differences between the ϕX174 and α3 structures are in the spike protein G due to a greater divergence of the primary structures (Figure 2). The first eight and the last ten residues of this protein extend away from its β-barrel and interact with other F and G subunits. If the G and F proteins within each pentamer are numbered $i = 1$ to 5 in a clockwise manner as viewed from outside the virus, the amino terminus of subunit G_i makes three main-chain hydrogen bonds with the neighboring symmetry-related amino terminus of subunit G_{i+1}. These terminal extensions form a β-annulus around the 5-fold axes of symmetry and thus contribute to the formation of the

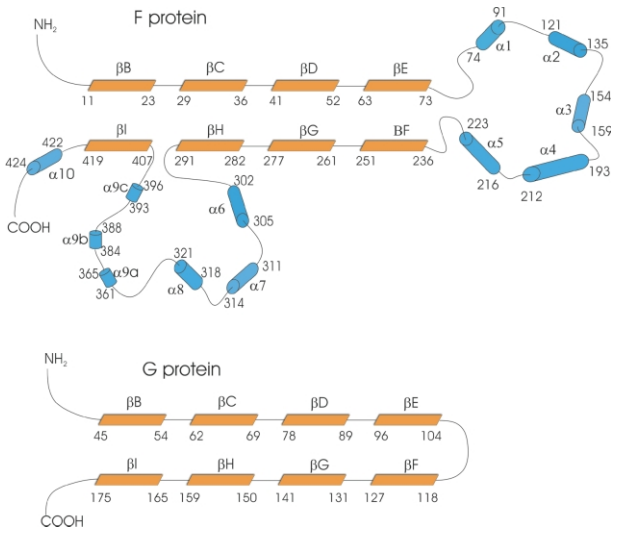


Figure 3. Secondary structure assignment for α3 F (upper panel) and G (lower panel) proteins based on the program PROCHECK.[31] Boxes (orange) and cylinders (blue) represent β-strands and α-helices, respectively. β-Strands B, C, D, and E form hydrogen bonds with β-strands I, H, G and F, respectively.

pentameric spike (Figure 5). The carboxy terminus of subunit G_i makes five hydrogen bonds with subunit F_{i+1} and five hydrogen bonds with subunit F_{i+2} (Figure 6). Thus, a total of 50 hydrogen bonds anchor the pentameric spike to the capsid shell (Table 2).

If the α3 and ϕX174 structures are superimposed by aligning the icosahedral symmetry axes, the root-mean-square (r.m.s.) deviations between equivalent C^{α} atoms of the F and G proteins are 0.8 Å and 2.7 Å, respectively. However, separate least-squares structural alignments of the F and G proteins result in respective r.m.s. deviations of only 0.5 Å and 0.7 Å (Table 3). The significant reduction in the r.m.s. difference between G pentamer C^{α} atoms is a consequence of a 3.5° clockwise rotation about the 5-fold axes, relative to the icosahedral framework if viewed from the outside (Table 4). This demonstrates that the F, but not the G, proteins maintain their relationship to the icosahedral symmetry axes.

In the transition from the open procapsid to the mature virion, coat and spike protein pentamers undergo a respective 5.2 Å and 8.1 Å radial collapse around the genome (see below). This process requires the spike proteins to lose contacts with the external scaffolding proteins and establish novel contacts with the underlying capsid proteins. As the pentameric spike moves radially inward, it is forced to rotate counter-clockwise because the carboxy termini of the G proteins are already associated with the capsid. The rotation of the spike stops when contacts are made with F capsid protein residues. In the case of the α3 virion, Thr75 and Asp76 of each G monomer hydrogen bond to capsid protein residues Tyr159 and Gln402, respectively. Similar, but not identical, contacts occur in ϕX174 and G4 (Figure 6; Table 5), accounting for the relative rotation of the spikes in these three phages. The rotation of the spikes about the 5-fold axes is most likely a consequence of morphogenesis.

The positively charged J protein enters the assembly pathway during DNA packaging by associating with the genome[13] *via* charge–charge interactions.[14,15] Biophysical characterization of particles packaged with J protein mutants suggests that protein–DNA interactions play a role in mediating the final stages of assembly.[14,15] *Microviridae* J DNA-binding proteins contain a positively charged amino terminus and a highly conserved hydrophobic carboxy terminus, which binds to a pocket on the interior of the F capsid protein. In ϕX174 there is an additional amino-terminal basic 13-residue sequence[5–7] (Figure 2) that binds to a neighboring 5-fold related F_{i-1} capsid protein (Figure 4(D)). Only the carboxy-terminal 13 hydrophobic residues can be structurally equivalenced and are almost completely conserved between the α3, ϕX174, and G4 proteins. The first 11 residues of the α3 J protein (corresponding to the middle region of ϕX174) contain a one-turn helix and have an r.m.s. deviation of 4.8 Å between equivalenced C^{α} atoms when compared to ϕX174. In contrast, the last 13 amino acid residues superimpose with an r.m.s. deviation of only 0.5 Å, indicating that the amino termini of these two proteins have different structures.

ϕX174 may have evolved a more elaborate J protein to compensate for a less basic capsid protein inner surface. When the inner surface charge of the α3 and ϕX174 coat proteins was compared, the α3 protein was found to be slightly more basic. This may also explain *in vivo* complementation studies where G4 could efficiently utilize the ϕX174 J protein, but ϕX174 was not able to use the shorter G4 protein.[16] A similar phenomenon has been recently documented for the ϕX174 and α3 J proteins (S.H., data not shown).

The sugar phosphate backbone of ten icosahedrally ordered nucleotides could be identified in the α3 virion electron density map in the proximity of the positively charged amino terminus of the J protein. The icosahedrally ordered α3 DNA corresponds to approximately 10% of the genome. The density for these nucleotides is about one-fourth the height of the protein density, but twice the height of the noise. In contrast, only about four icosahedrally ordered nucleotides were reported in the ϕX174 virion structure, and these were not directly associated with the J protein.[17] However, it should be noted that the amount of ϕX174 ordered DNA may have been under reported, due to the lack of confidence in interpreting low density features (the interpreted density had a larger peak to noise ratio).

α3 procapsid

The α3 open procapsid is an inherently unstable assembly intermediate that is triggered to progress

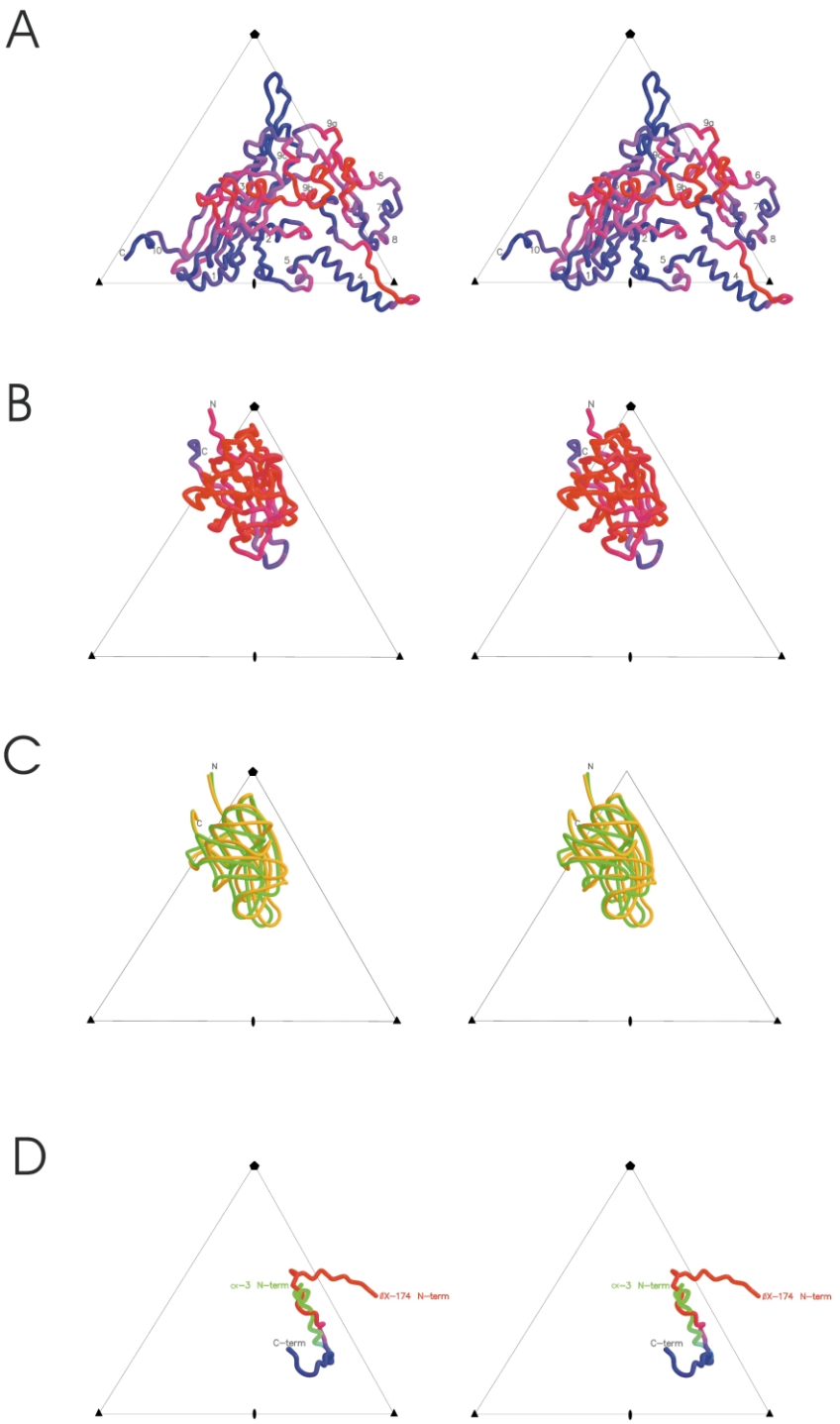


Figure 4. Stereo representation of the bacteriophage α3 structural proteins illustrating the level of conservation between α3, ϕX174, and G4. More conserved areas are shaded blue and less conserved regions are shaded red. Conservation was defined by $\Sigma_{i-4,i+4}f(i)$ where i is a residue and $f(i) = 1$ if there is complete conservation among the three phages or 0 otherwise. (A) Protein F has a highly conserved β-barrel, while surface residues are less conserved. (B) The spike protein is the least conserved of all the proteins. (C) Superposition of the spike protein from bacteriophage α3 (orange) and ϕX174 (green). The bacteriophage α3 spike protein has a 3.5° anticlockwise rotation about the 5-fold axis compared to the spike protein of ϕX174. The amino and carboxy termini of the spike protein extend away from the β-barrel and make extensive contacts with symmetry-related F and G proteins. (D) Superposition and conservation of the DNA-binding protein J. In blue is the highly conserved carboxy-terminal region of α3 and ϕX174. The less conserved sequence and structural amino-terminal regions are colored green for α3 and red for ϕX174. Note that the 13 amino-terminal residues in ϕX174 are missing in α3.

along the assembly pathway upon DNA packaging. A more stable closed procapsid can be produced by high concentrations of ammonium sulfate.[4] However, this closed particle probably represents an off-pathway product. The α3 open procapsid was purified in the absence of high ionic strength compounds, but could not be crystallized due to rapid dissociation. Therefore, a structure determination was undertaken using cryo-EM, which preserves the particle integrity upon vitrification in liquid ethane. The resulting 15 Å resolution reconstruction (Figure 7) is considerably more detailed than the homologous 26 Å resolution ϕX174 reconstruction.

The capsid and spike proteins from the α3 virion and the scaffolding proteins from the closed procapsid were fitted into the open procapsid density using the program EMfit.[18] There were no contacts, defined as atoms in opposing subunits being separated by less than 4 Å, between symmetry-related subunits. In comparison, the number of atoms in one F subunit of the virion X-ray structure in contact with other F subunits was 6.3% of the total number of atoms in the protein. Therefore, the

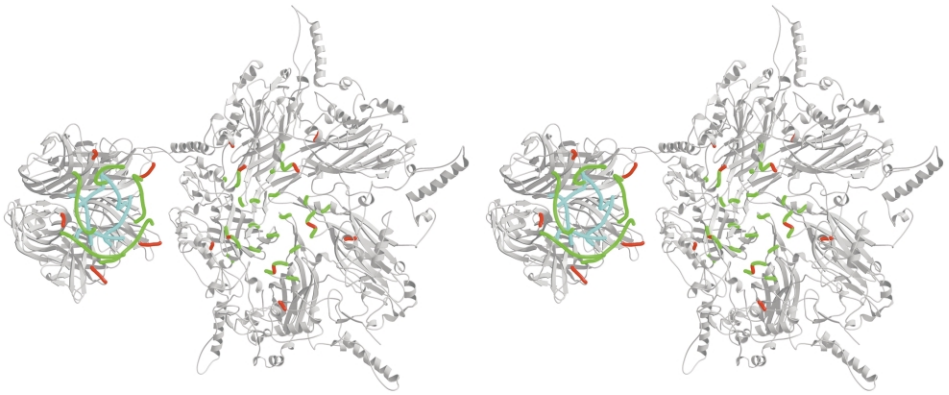


Figure 5. Stereo ribbon diagram of the interface between proteins F (right) and G (left). The two protein pentamers have been pulled apart to highlight the contact residues. The amino-terminal residues of the spike proteins highlighted in blue only make contact with neighboring spike proteins. The carboxy-terminal residues of the G proteins, highlighted in green, make ten hydrogen bonds per monomer with matching residues in the F capsid proteins, also highlighted in green. Two additional hydrogen bonds per monomer between Thr75 and Asp76 of G with Tyr159 and Gln402 of F, respectively, are highlighted in red. These latter residues probably limit the amount of rotation of the spike in the transition from procapsid to mature virus.

pixel size had to be reduced by 5% in order to obtain the same percentage of contacts in the open procapsid.

When capsid protein coordinates were fitted into the procapsid density, all atoms are within positive density except for residues 172–231. This region consists of helices 4 and 5, found near the 3- and 2-fold axes of symmetry, respectively (Figure 4(A)). Hence, protein F, with helices 4 and 5 removed, was fitted into the density using main-chain and side-chain atoms resulting in a *sumf*†,[18] value of 52% (Figure 8(A); Table 6). Removal of these two helices created an apparent 20 Å gap surrounding the F pentamers (Figure 7). Similarly a pentamer of G proteins was fitted into the density resulting in a *sumf* value of 47% (Figure 8(B); Table 6). All density covered by the atoms of the F and G proteins was then set to zero, leaving the remaining density to be interpreted as D and B scaffolding and H pilot proteins.

ϕX174 D1–D2 and D3–D4 external scaffolding protein dimers were used as models to interpret the open procapsid D scaffolding density. This was justified because a D1–D2 dimer can be superimposed onto a D3–D4 dimer with an r.m.s. deviation of less than 2.5 Å. It has been reported that the D proteins may exist as dimers in solution and assemble into tetramers during morphogenesis[4] (Table 6). Fitting individual D proteins was not possible because the density at this resolution is somewhat featureless and the coordinates are placed at locations that will cause monomers to overlap.

After the F, G, and D proteins had been fitted into the cryo-EM density of the open procapsid, the only density remaining above 1σ was inside the F capsid, located at the icosahedral 2- and 5-fold axes of symmetry. This density is presumably the B scaffolding protein. In the X-ray structure of the closed procapsid, the B scaffolding protein was disordered between residues 9 and 60. An attempt was made to manually fit the structure of the ordered component of B (residues 1–8, 61–120) into the difference density of the open procapsid. However, residues 61–80, which form an α-helix in the closed procapsid X-ray structure, were out of density. Therefore, they were independently fitted into a tube-like feature of the cryo-EM map that runs close to the nearest 5-fold axis of symmetry (Figure 8(D)). A secondary structure analysis of the ϕX174 B protein predicts a long central helix between residues 56 and 77, closely matching the long helix in the closed procapsid structure. The tube-like feature connects to the density around the 5-fold symmetry axes previously associated with protein H. Hence this 5-fold associated density could represent five copies of the missing amino-terminal portion of protein B. However, at 15 Å resolution it was not possible to follow more of the polypeptide chain.

The density feature around the 5-fold axes of symmetry occupies 41,000 Å^3 and is sufficient to accommodate five copies of residues 9–60 of the B protein, which would require 40,700 Å^3 (see Materials and Methods). Similar density near the 5-fold axes in the closed procapsid structure had been interpreted as one copy of protein H. This is also consistent with the molecular weight of one H protein that would occupy about 42,400 Å^3. This, together with the above observation that the amino termini of the B protein are close to the density feature at the 5-fold axes suggests that this density may represent either protein B or H. Partial occupancy of B is supported by the reduced height of the carboxy-terminal section of protein B relative to the other protein density in the ϕX174 closed procapsid and the α3 open procapsid.

† *sumf* is the average density of all the fitted atoms normalized by placing the largest density of the map equal to 100.

Table 2. Polar protein contacts

Protein G_1 residues			Residues in F making polar contacts with G_1		
α3	ϕX174	G4	α3	ϕX174	G4
	Q3	Q3		$Q254_3$	$Q254_2$
T75	N67	T68	Y159	Y158	M397
D76			Q402		
E179	E167	T171	$R56_2$	$E366_2$	$Q395_2$
Q181	I169	Q174	$D396_2$, $F399_2$	$S396_2$, $R255_2$	$S396_2$, $S255_2$
Q184	Q172	P175	$S256_3$	$R56_3$, $T255_3$, $S396_3$	$R56_3$, $Q398_2$
P185			$R57_3$, $Q403_2$		
K187	K175	K177	$R57_3$, $S256_3$, $L258_3$, $Q400_2$	$R56_2$, $T255_3$, $L257_3$, $Q395_3$	$R56_3$, $S255_3$, $L257_3$, $Q395_2$

Protein J residues		Residues in protein F making polar contacts with J	
α3	ϕX174	α3	ϕX174
	G3		$L236_5$, $V237_5$, $M238_5$
	K4		$R239_5$
	K5		$R239_5$
	R6		$R239_5$
	S7		$K269_5$
	G8		$T267_5$, $Y268_5$
	R10		$T267_5$, $K269_5$
	G12		$T267_5$
R5	L17	A60, D62	D61
R6	R18	W244, D62	D61
S7	G19	D357, W244	$K407_5$
	T20		S356, $L17_5$
	K21		$K407_5$
R10	G22	H355, D357	D13, $S15_5$, $K407_5$, $N409_5$
	R24		$T413_5$
W17		K167, A172, I350, N168, I169	
Y18		L174	
V19		P139, W140	
G20	G33	T210, Y211, F212, Q214, P139	Y210, Q213
G21		P139	
S22		R215	
Q23	Q36	Y135, C164, A138, K167	K166
F24	F37	Y135, F68, R291	Y134, R290

Subscripts denote symmetry-related subunits numbered clockwise. Selection criterion based on a distance of 3.0 Å.

Table 3. Superposition of C^{α} coordinates of various virion proteins

Superposition of proteins	Number of equivalent C^{α} atoms	r.m.s.d. (Å) by least-squares refinement	r.m.s.d. (Å) by alignment of icosahedral axes
α3/ϕX174 all	618	0.65	2.85
α3/ϕX174 protein F	416	0.49	0.85
α3/ϕX174 protein G	175	0.71	2.72
α3/ϕX174 G pentamer	875	0.72	2.72
α3/ϕX174 protein J	13	0.36	0.53
α3/G4 all	602	0.68	2.70
α3/G4 protein F	417	0.76	0.85
α3/G4 protein G	175	0.79	2.29
α3/G4 G pentamer	865	0.78	2.29
G4/ϕX174 all	600	0.47	1.51
G4/ϕX174 protein F	417	0.40	0.59
G4/ϕX174 protein G	173	0.65	3.27
G4/ϕX174 G pentamer	860	0.66	3.27

The relative mobility in SDS-PAGE gels of the B protein, derived from ϕX174 and α3, changes with the age of the preparation, suggesting that the B protein is unstable (data not shown). Furthermore, it was observed that purified glutathione *S*-transferase (GST)-B fusion protein degrades with time, generating fragments of distinct sizes. This observation is consistent with earlier data on B proteolysis.[19] Analysis of the GST-B fusion protein cleavage products indicated that one of three fragments was removed from the protein (Figure 9). The sizes of these fragments are consistent with a repeating motif in the primary structure. An Arg-Phe sequence repeats at approximately 13 amino acid intervals from the carboxy-terminal end of protein B (Figure 9). The absence of ordered B protein structure prior to Arg61 in the closed procapsid may be the result of a cleavage event, which may have allowed the amino-terminal portion to diffuse away. In contrast, the open procapsid, which probably represents a true assembly intermediate, appears to contain an undigested B scaffolding protein as observed by SDS-polyacrylamide gel electrophoresis.

Table 4. Rotation and translation of G pentamer for superposition

	Rotation (°)	Translation (Å)
α3-ϕX174	3.6	−0.2
α3-G4	2.3	−0.4
G4-ϕX174	1.1	0.2

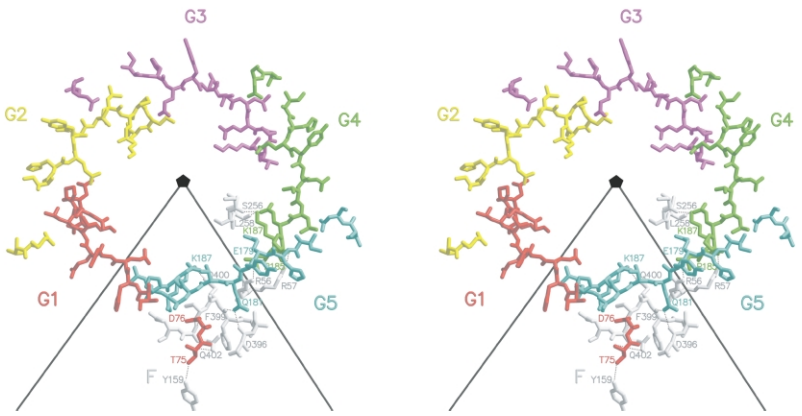


Figure 6. Stereo diagram of the G and F contact residues at the interface between the capsid and spike proteins. In gray are residues belonging to the F_1 capsid protein. In red are the residues belonging to the G_1 spike protein. Symmetry-related residues that belong to G_2, G_3, G_4, and G_5 are in yellow, magenta, green, and light blue, respectively. Hydrogen bonds are represented by dotted lines.

In P22 and ϕ29, scaffolding proteins exit intact and recycle in further rounds of assembly.[20,21] However, in most bacteriophages and in the *Herpesviridae*, cleavage of the scaffolding protein facilitates its removal during morphogenesis.[22] Therefore, the autoproteolytic activity of protein B may be a necessary component of morphogenesis, facilitating the escape of suitably sized fragments of protein B through the 30 Å diameter pores located at the 3-fold axes of symmetry. During the DNA packaging, protein J displaces the carboxy terminus of protein B by competition for the same hydrophobic-binding pocket on the internal surface of the F capsid protein.[3] Since the B protein is intact in the open procapsid, it follows that B protein displacement may activate the protein autoproteolytic activity.

Materials and Methods

Purification and crystallization

Mature virion was generated in *Escherichia coli slyD* cells infected with bacteriophage α3 *am(E)W4*. The host cell *slyD* mutation confers resistance to E protein-mediated lysis.[23] Both the host cell and phage mutations were needed to effectively block lysis. Cells (12 l) at 1.0×10^8 cfu/ml were infected at a multiplicity of infection (moi) of 5 at 33 °C and allowed to incubate for six hours. Infected bacterial pellets were resuspended in 400 ml buffer A (50 mM Tris (pH 7.5), 25 mM EDTA, 50 mM NaCl, 0.1 mM β-mercaptoethanol) and lysed by the addition of 0.2 g lysozyme. After a one-hour incubation at 20 °C, the sample was subjected to three rapid freeze/thaw cycles. After removal of bacterial debris by centrifugation at 16,000*g*, particles were precipitated by the addition of 8.5% (w/v) PEG 8000. The phage was further purified in two consecutive 0–40% (w/v) sucrose gradients. The purified sample was concentrated on 100 kDa cutoff Millipore centrifuge concentrator units and a buffer exchange was performed into buffer B (10 mM Tris (pH 7.5), 1 mM EDTA, 300 mM NaCl, 0.1 mM β-mercaptoethanol, 0.02% (w/v) sodium azide). The virion was crystallized by vapor diffusion using sitting drops at 45 °C. A 10 µl drop of sample was mixed with an equal volume of reservoir buffer containing 3–5% PEG 8000, 100 mM sodium citrate (pH 5.0), 40% (v/v) glycerol, 0.02% sodium azide, and 0.1% β-mercaptoethanol.

Bacteriophage α3 procapsids were purified from *E. coli slyD* cells infected with α3 *am(E)W4*, *am(J)M1*, *am(J)S7*. The amber mutations in gene *J* prevent the production of the DNA-binding protein, thus arresting morphogenesis after procapsid formation. Two amber mutations were used to prevent reversion during the long incubation periods needed to generate sufficient quantities. Moi and infection conditions were identical with those described above. Purification of the procapsid is essentially the same as for the virion with a few minor modifications. After lysis of the bacterial cells, the procapsid was precipitated by the addition of 7% PEG 8000, followed by centrifugation at 16,000*g* for ten minutes at 4 °C. The procapsid sample was loaded onto a 5–35% sucrose gradient and centrifuged at 121,000*g* in an SW28 rotor for four hours at 4 °C. Samples collected from the gradients were centrifuged at 226,000*g* at 4 °C for two hours in a Ti50 rotor to pellet the procapsid. Procapsids were resuspended overnight in 100–500 µl buffer containing 10 mM Tris (pH 7.5), 1 mM EDTA, 0.1 M NaCl.

Table 5. Residues that stop the spike rotation

	Protein F	Protein G
α3	Y159, Q402	T75, D76
ϕX174	Y158	N67
G4	M397, Q398	T68, N66

X-ray data collection and processing

Data were collected at BioCARS beam line 14 BM-c at the Advanced Photon Source. A single frozen crystal was used to collect 180° of data with an oscillation angle of 0.25° on an MAR detector (345 mm diameter) and a crystal-to-detector distance of 535 mm. Images were processed with DENZO[24] and scaled together using the program SNP.[25] The space group of the bacteriophage α3 virion crystals was $P2_1$ with cell dimensions of 290.3 Å, 332.1 Å, 337.7 Å, and β = 94.1°. The Matthews' coefficient, V_M, is 2.6 $Å^3\,Da^{-1}$ assuming one particle in the unit cell. Thus, the icosahedral symmetry produces 60-fold non-crystallographic symmetry (NCS) redundancy. The final R_{merge} was 8.0% for all the data (11.5% in the highest resolution bin) with an average of 3.5 measurements per reflection. The overall completeness was 75% (55% in the highest resolution range) after removal of reflections with less that 4σ. Initial

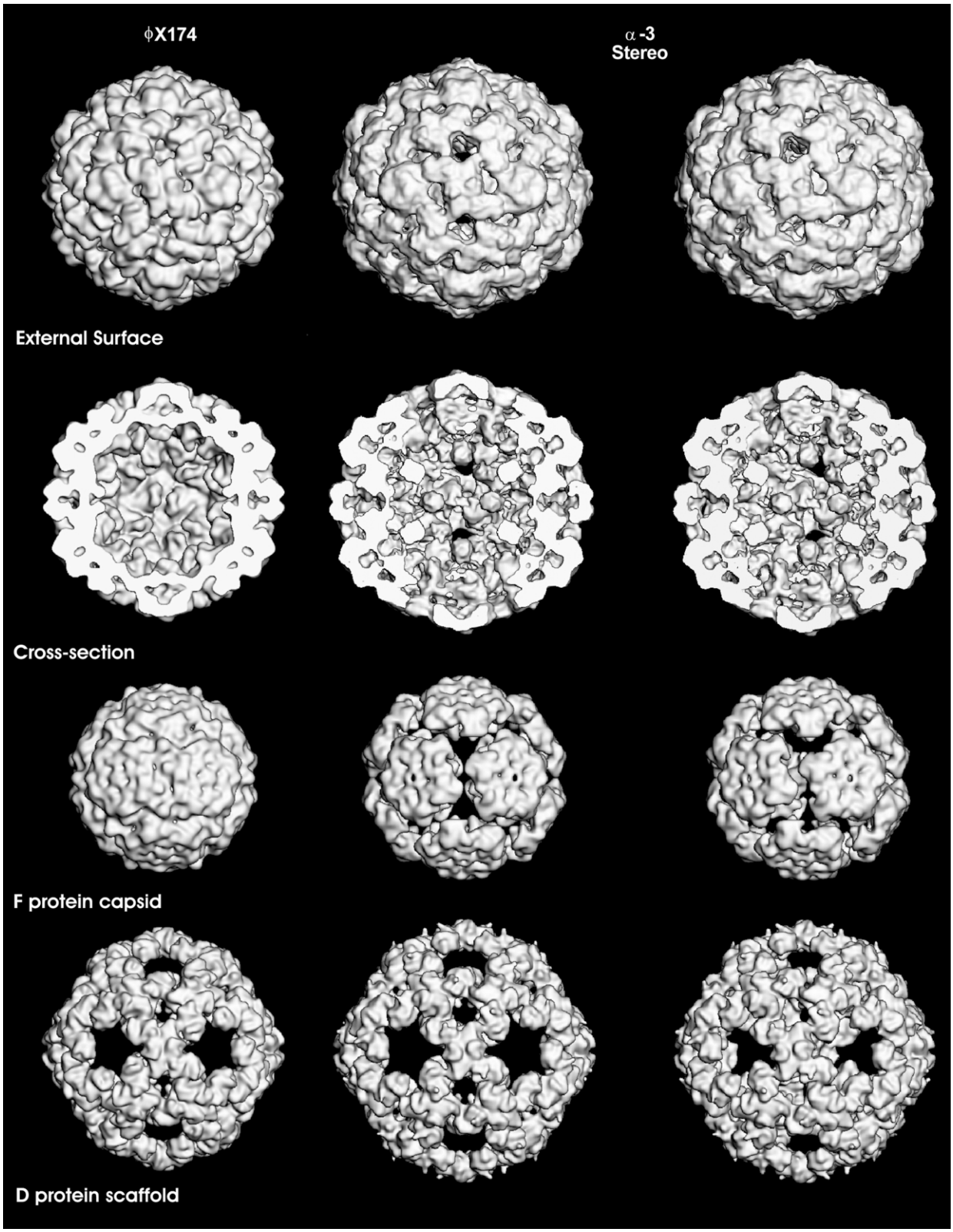


Figure 7. Comparison of the ϕX174 and α3 procapsids. The particles illustrated on the left-hand column correspond to the bacteriophage ϕX174 closed procapsid structure calculated to 15 Å resolution from the X-ray atomic coordinates. On the right are shown stereo diagrams of the α3 cryo-EM reconstruction also at 15 Å resolution. The second row is a cross-section that allows a clear view of the interior of the procapsid and the additional density feature around the 5-fold axes of symmetry present only in the α3 open procapsid. The third row consists of a particle where all proteins except for F were removed to illustrate the apparent 20 Å wide gaps in the α3 procapsid that are created by the disorder of helices 4 and 5. The fourth row shows the same for the D proteins. The α3 particles in the last two rows were calculated from coordinates that resulted from fitting the X-ray structural components derived from the closed procapsid into the open procapsid cryo-EM density.

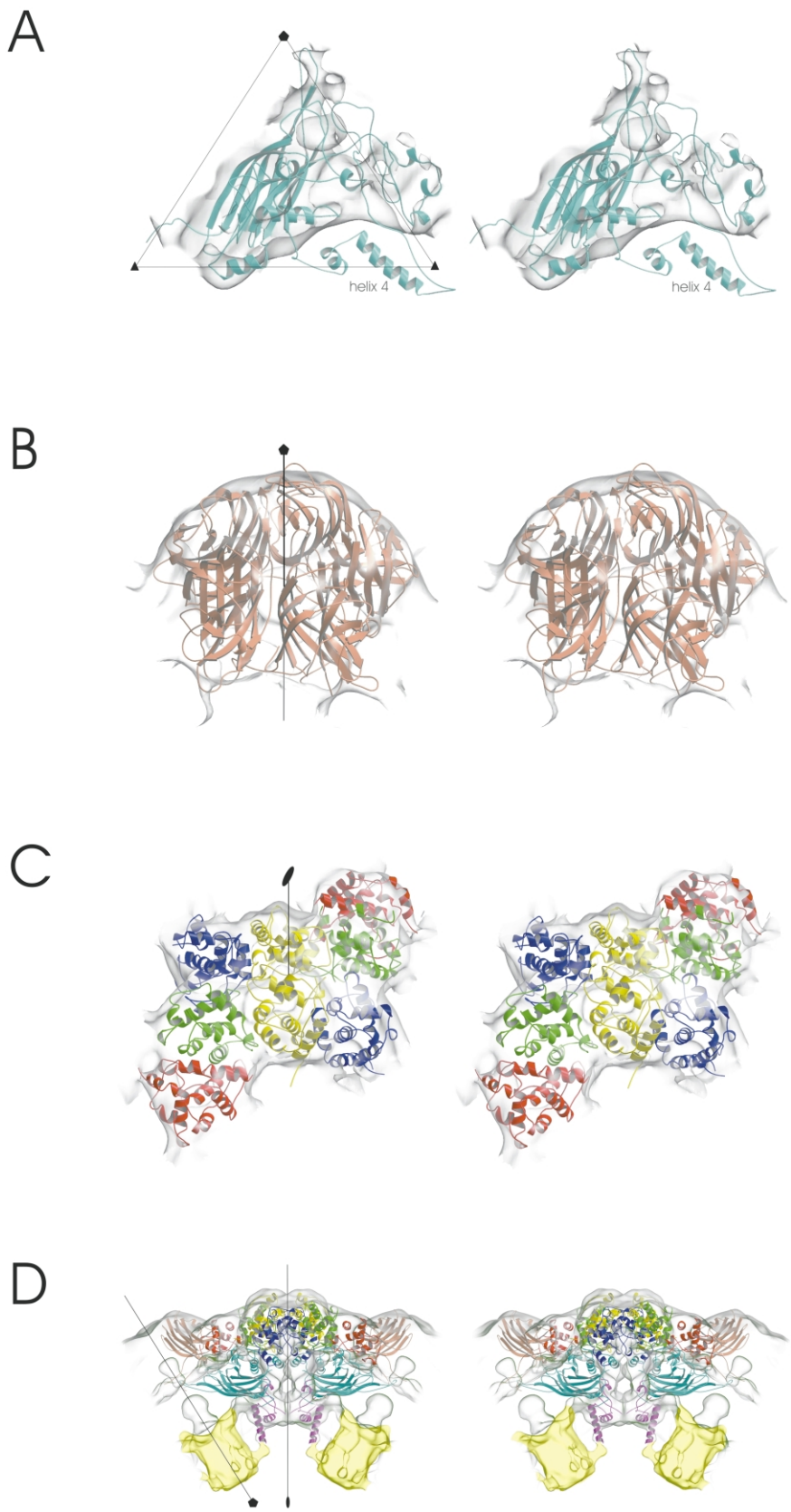


Figure 8. Stereo diagrams of the procapsid protein coordinates fitted into the open procapsid density. The surface of the cryo-EM density is contoured at a level of 0.5σ and suitably slabbed to see into the interior of the density. (A) Helices 4 and 5 are clearly shown to be out of density after fitting the capsid protein F into the density. (B) Protein G was fitted into the density as a pentamer. (C) The D external scaffold proteins were fitted as dimers of D1(red)–D2(green) and D3(yellow)–D4(violet). (D) The internal scaffolding protein B was manually fitted into the open procapsid cryo-EM structure where its long α-helix fits into a tube of density that leads toward the uninterpreted density around the 5-fold axis of symmetry (yellow). A side view of the open procapsid is shown with the outer surface of the particle on the top and the interior of the particle on the bottom. Proteins F (light blue), G (brown), D1 (red), D2 (green), D3 (yellow), D4 (violet) and B (magenta) were included along with their 2-fold-related subunits. Icosahedral 5- and 2-fold symmetry axes are shown as lines.

orientations of the α3 particles in the unit cell were determined with a self-rotation function using the program GLRF.[26,27] The particle orientation had one of its 2-fold axes only 0.65° away from being parallel with the crystallographic 2_1 screw axis. This allowed the particle position in the unit cell to be determined using a self-Patterson function, resulting in a large peak on the Harker section corresponding to (0.2498, 0.2500, 0.2498) in fractions of the cell edges. The virion structure was determined to 3.5 Å resolution by molecular replacement

Table 6. EMfit results

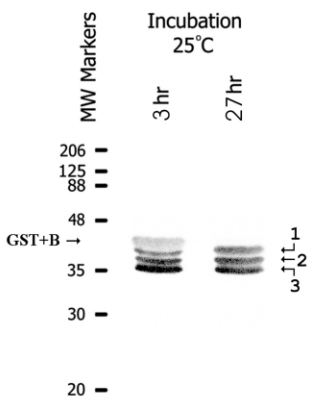
Object used for fitting	sumf (%)		Clash[a] (%)	(−) Density[a] (%)	Rotation[b] (deg.)	Translation[b] (Å)
	Before fit	After fit				
α3 virion F	47.8	51.8	6.1	4.0	1.0	5.2
α3 virion G	27.8	46.8	0.0	1.7	2.5	8.1
ϕX174 virion F	42.2	51.3	12.7	2.8	2.1	5.2
ϕX174 virion G	28.8	47.3	0.0	1.3	1.1	8.2
ϕX174 closed procapsid F	50.9	51.6	2.6	2.6	1.3	2.2
ϕX174 closed procapsid G	24.8	47.5	0.0	1.1	1.8	2.7
ϕX174 closed procapsid D1D2	43.4	45.3	0.0	4.3	1.3	2.1
ϕX174 closed procapsid D3D4	44.6	45.2	1.2	4.5	1.0	1.7

[a] Clash is the percentage of symmetry-related atoms that are <3.0 Å apart. Negative (−) density describes the percentage of atoms whose position corresponds to an area of negative density in the cryo-EM map.
[b] Anticlockwise rotation and outward radial translation (as determined by the program homology) relate the position of the protein in the native particle and after fitting into procapsid.

real-space averaging using the ϕX174 virion structure as a phasing model[28] with calculated phases to 3.5 Å resolution. The overall correlation coefficient of the back-transformed map was 0.86 (0.56 at 3.5 Å resolution) and an overall *R* factor of 0.19 (0.29 at 3.5 Å resolution).

X-ray structure refinement

The F, G, and J peptide chains were built into the 3.5 Å resolution electron density map using the program O.[29] The subsequent structure was refined using five cycles of simulated annealing followed by five cycles of strict NCS conjugate gradient minimization using the program CNS.[30] The ten nucleotides were not included in the refinement because there is density only for the sugar phosphate backbone and not the bases. The final *R* factor is 0.23 (0.28 at 3.5 Å resolution) with all of the residues in the most-favored regions of the Ramachandran plot as calculated by the program PROCHECK.[31] Because of the 60-fold NCS redundancy, there was less than 0.3% difference between $R_{working}$ and R_{free}.



GST-(1) **MEQLTKNQ**AV ATSQEAVQNQ NEPQLRDENA HNDKSVHGVL NPTYQAGLRR DAVQPDIEAE

3* 2 1

(61) **RKKRDEIEAG KSYCSRRFGG ATCDDKSAQI YARFDKNDWR IQPAEFYRFH DAEVNTFGYF**

* Previously reported cleavage site [15].

Figure 9. Proteolysis of the GST-B fusion protein. (Top) SDS-PAGE gel of purified GST-B fusion protein after incubation at room temperature. Proteolytic fragments begin to appear one hour post-induction. Thus, the incubation time indicated for each lane is calculated from the time of induction and includes the two hours needed for purification. The first lane represents the sample immediately following purification while the sample in the second lane was incubated an additional 24 hours at room temperature. The full-length fusion protein is indicated by the label GST-B. Proteolytic fragments are labeled 1, 2, and 3. (Bottom) Sequence of the ϕX174 B protein with the possible cleavage sites labeled the same as in the gel. Residues that were ordered in the high resolution structure of the closed procapsid are highlighted in bold. Each of the potential cleavage sites has a common motif composed of an Arg and a Phe, which are underlined.

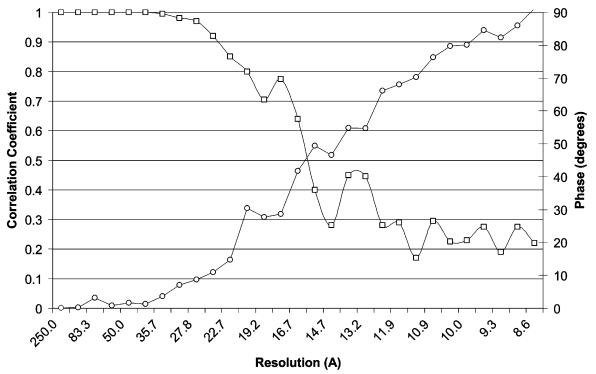


Figure 10. Fourier cross-correlation for the α3 open procapsid reconstruction data. The data were divided in half and corresponding individual reconstructions were cross-correlated. The correlation (□) dropped below 0.5, and the average difference between phase angles (○) was greater than 45° at about 15 Å resolution.

Cryo-EM

Cryo-EM micrographs were taken on an FEG200 electron microscope at defocus levels ranging from 1.6 μm to 3.2 μm under focus. The micrographs were digitized on an SCAI flatbed densitometer. Individual particles were selected, boxed, and normalized using the Purdue program ROBEM. The defocus level was estimated from the averaged Fourier transform of all the boxed particles. Initial particle orientations for a preliminary low-resolution reconstruction were calculated using a 26 Å resolution ϕX174 procapsid reconstruction.

Particle orientations and centers were found using the program EMPFT (Polar Fourier Transform),[32] which uses cross-correlation procedures to compare particle images against a database of reference projected views produced from a 3D model map. The program EM3DR was used to merge all images for computing a Fourier Bessel transformation to produce a 3D density map of the particle in a standard orientation. The phases and amplitudes of the calculated reconstruction were corrected using the contrast transfer function (CTF).[33] The resolution was slowly extended to 10 Å for determining the particle orientations and for map calculation. Of 3433 boxed particles, 2378 were included in the final reconstruction and, presumably, the rejected particles were damaged to a greater or lesser extent. The effective resolution of the reconstruction was calculated by dividing the images into two sets and calculating individual reconstructions from each half dataset. The resulting reconstructions were Fourier transformed and cross-correlated. The resolution, determined as the point where the correlation coefficient falls below 0.5 and the mean phase difference becomes greater than 45°, was found to be 15 Å (Figure 10). A possible reason for the inability to obtain better than 15 Å resolution may have been due to the tendency of particles to orient themselves non-randomly (Figure 11). This led to a reconstruction that was lacking data in certain directions. No tilted grids had been used in the data collection as it had been anticipated that the particles would have a random distribution.

Volume determination

The volume (V) that a protein molecule would occupy in solution was calculated from the molecular mass (M_r) assuming a specific density (ρ_s) of 1.37 g/cm^3, and Avogadro's number (n) according to the formula:[34]

$$V = \frac{M_r \rho_s}{n} \cong (1.212 M_r)\,\text{Å}^3$$

This was compared to the volume of the feature near the 5-fold density, which was determined by counting the number of pixels with a density above a cutoff of 0.4σ (where the highest noise density was 0.2σ).

Fitting X-ray coordinates into cryo-EM density

X-ray coordinates of the α3 capsid and spike proteins were initially fitted manually into the cryo-EM density. The program EMfit was used to obtain a more quantitative fit.[18] The criteria of fit used by the program are the average value of the density at all atomic sites in the fitted protein, the lack of atoms in negative density, and the absence of icosahedral symmetry-related atomic clashes. The overall measure of fit as expressed by R_{crit} is a combination of all these into a single best fit criterion. After a protein was fit into the cryo-EM map, density

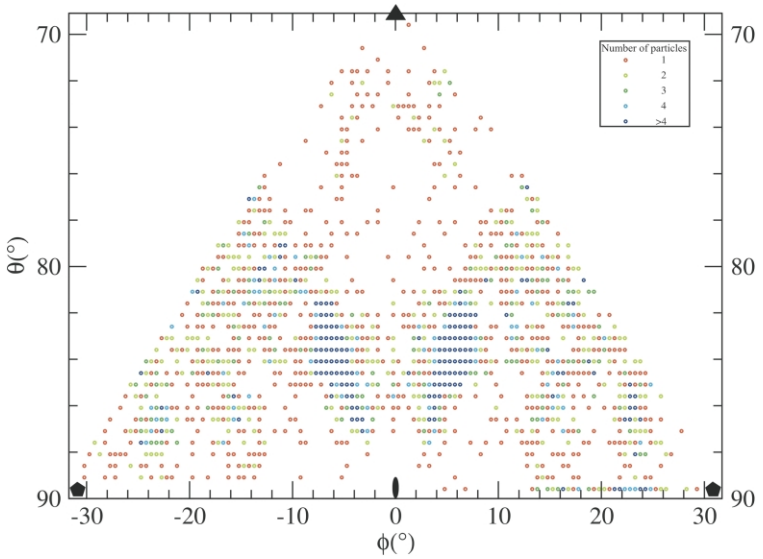


Figure 11. Icosahedral asymmetric unit showing the non-random orientation of the individual particles used for the cryo-EM image reconstruction of the α3 open procapsid. Each point represents the orientation of a particle relative to the standard orientation given by the triangular asymmetric unit. The orientations are plotted as polar angles with ϕ as the x-axis and θ as the y-axis. The angle ω is not plotted.

within a radius of 4.0 Å around all atomic positions was set to zero before fitting the next protein.

Protein B

A GST-B fusion was created by cloning a *Bam* HI-*Eco* RI fragment containing the φX174 protein B into the pGEX-2T plasmid (Pharmacia Biotech). Transformed BL21 (DE3) cells were incubated at 37 °C and allowed to reach an absorbance of 1.0 before being cooled to room temperature. Induction of GST-B fusion protein was initiated by the addition of IPTG to a final concentration of 0.1 mM. After 20 minutes, the cells were centrifuged at 7000*g* and resuspended in two pellet volumes of lysis buffer (50 mM Tris (pH 7.5), 10 mM EDTA, 150 mM NaCl, 0.1 mM β-mercaptoethanol, 10 μg ml^{-1} lysozyme). After one-hour incubation at 20 °C, the sample was subjected to three rapid freeze/thaw cycles and then centrifuged at 16,000*g* to remove the bacterial debris. The supernatant was loaded onto a glutathione Sepharose 4B column (bed volume of 2 ml). The column was washed with 40 ml of wash buffer (50 mM Tris (pH 7.5), 10 mM EDTA, 150 mM NaCl, 0.1 mM β-mercaptoethanol) and eluted with wash buffer containing 10 mM glutathione.

Atomic coordinates

The coordinates of the crystallographically-determined virion and the cryo-EM-determined open procapsid structures have been deposited with the Protein Data Bank (PDB) (virion accession code 1M06, open procapsid accession code 1M0F).

Acknowledgements

We thank Shuji Kanamaru, Cheryl Towell, and Sharon Wilder for their help in the preparation of this manuscript. We are grateful to Stephen Fuller and Erika Mancini for discussions about the cryo-EM reconstruction and to Chuan Xiao for the use of his parallelized versions of EMPFT, EM3DR, and other programs. We also thank the BioCARS staff of the Advanced Photon Source for providing excellent support for their X-ray data collection facilities. This research was supported by an NIH Biophysics Training Grant fellowship and GAANN fellowship to R.A.B., NIH grant GM33050 to T.S.B., NSF grant MCB-9986266 to M.G.R., and NSF grant MCB-9982284 to B.A.F. Figures were generated using the graphics programs VMD,[35] Molscript,[36] Bobscript,[37] and Raster 3D.[38]

References

1. Hayashi, M., Aoyama, A., Richardson, D. L. & Hayashi, M. N. (1988). Biology of the bacteriophage φX174. In *The Bacteriophages* (Calendar, R., ed.), pp. 1–71, Plenum Press, New York.
2. Ilag, L. L., Olson, N. H., Dokland, T., Music, C. L., Cheng, R. H., Bowen, Z. *et al.* (1995). DNA packaging intermediates of bacteriophage φX174. *Structure*, **3**, 353–363.
3. Dokland, T., McKenna, R., Ilag, L. L., Bowman, B. R., Incardona, N. L., Fane, B. A. & Rossmann, M. G. (1997). Structure of a viral procapsid with molecular scaffolding. *Nature*, **389**, 308–313.
4. Dokland, T., Bernal, R. A., Burch, A., Pletnev, S., Fane, B. A. & Rossmann, M. G. (1999). The role of scaffolding proteins in the assembly of the small, single-stranded DNA virus φX174. *J. Mol. Biol.* **288**, 595–608.
5. Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A. *et al.* (1977). Nucleotide sequence of bacteriophage φX174 DNA. *Nature*, **265**, 687–695.
6. Kodaira, K., Nakano, K., Okada, S. & Taketo, A. (1992). Nucleotide sequence of the genome of the bacteriophage α3: interrelationship of the genome structure and the gene products with those of the phages, φX174, G4 and φK. *Biochim. Biophys. Acta*, **1130**, 277–288.
7. Godson, G. N., Barrell, B. G., Staden, R. & Fiddes, J. C. (1978). Nucleotide sequence of bacteriophage G4 DNA. *Nature*, **276**, 236–247.
8. Greene, B. & King, J. (1994). Binding of scaffolding subunits within the P22 procapsid lattice. *Virology*, **205**, 188–197.
9. Greene, B. & King, J. (1999). *In vitro* unfolding/refolding of wild type phage P22 scaffolding protein reveals capsid-binding domain. *J. Biol. Chem.* **274**, 16135–16140.
10. Winkler, F. K., Schutt, C. E., Harrison, S. C. & Bricogne, G. (1977). Tomato bushy stunt virus at 5.5 Å resolution. *Nature*, **265**, 509–513.
11. Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H. J. *et al.* (1985). Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature*, **317**, 145–153.
12. Rossmann, M. G. & Johnson, J. E. (1989). Icosahedral RNA virus structure. *Annu. Rev. Biochem.* **58**, 533–573.
13. Fujisawa, H. & Hayashi, M. (1976). Viral DNA-synthesizing intermediate complex isolated during assembly of bacteriophage φX174. *J. Virol.* **19**, 409–415.
14. Jennings, B. & Fane, B. A. (1997). Genetic analysis of the φX174 DNA binding protein. *Virology*, **227**, 370–377.
15. Hafenstein, S. & Fane, B. A. (2002). φX174 genome–capsid interactions influence the biophysical properties of the virion: evidence for a scaffolding-like function for the genome during the final stages of morphogenesis. *J. Virol.* **76**, 5350–5356.
16. Fane, B. A., Head, S. & Hayashi, M. (1992). Functional relationship between the J proteins of bacteriophages φX174 and G4 during phage morphogenesis. *J. Bacteriol.* **174**, 2717–2719.
17. McKenna, R., Ilag, L. L. & Rossmann, M. G. (1994). Analysis of the single-stranded DNA bacteriophage φX174, refined at a resolution of 3.0 Å. *J. Mol. Biol.* **237**, 517–543.
18. Rossmann, M. G., Bernal, R. A. & Pletnev, S. V. (2002). Combining electron microscopic with X-ray crystallographic structures. *J. Struct. Biol.* **136**, 190–200.
19. Richardson, D. L., Jr, Aoyama, A. & Hayashi, M. (1988). Proteolysis of bacteriophage φX174 prohead

protein gpB by a protease located in the *Escherichia coli* outer membrane. *J. Bacteriol.* **170**, 5564–5571.
20. King, J. & Casjens, S. (1974). Catalytic head assembling protein in virus morphogenesis. *Nature*, **251**, 112–119.
21. Casjens, S. & King, J. (1974). P22 morphogenesis. I: catalytic scaffolding protein in capsid assembly. *J. Supramol. Struct.* **2**, 202–224.
22. Newcomb, W. W., Trus, B. L., Cheng, N., Steven, A. C., Sheaffer, A. K., Tenney, D. J. *et al.* (2000). Isolation of herpes simplex virus procapsids from cells infected with a protease-deficient mutant virus. *J. Virol.* **74**, 1663–1673.
23. Roof, W. D., Horne, S. M., Young, K. D. & Young, R. (1994). *SlyD*, a host gene required for φX174 lysis, is related to the FK506-binding protein family of peptidyl-prolyl *cis–trans*-isomerases. *J. Biol. Chem.* **269**, 2902–2910.
24. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276A**, 307–326.
25. Bolotovsky, R., Steller, I. & Rossmann, M. G. (1998). The use of partial reflections for scaling and averaging X-ray area detector data. *J. Appl. Crystallog.* **31**, 708–717.
26. Rossmann, M. G. & Blow, D. M. (1962). The detection of sub-units within the asymmetric unit. *Acta Crystallog.* **15**, 15–31.
27. Tong, L. A. & Rossmann, M. G. (1990). The locked rotation function. *Acta Crystallog. sect. A*, **46**, 783–792.
28. Rossmann, M. G. (1990). The molecular replacement method. *Acta Crystallog. sect. A*, **46**, 73–82.
29. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
30. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
31. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
32. Baker, T. S. & Cheng, R. H. (1996). A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *J. Struct. Biol.* **116**, 120–130.
33. Baker, T. S., Olson, N. H. & Fuller, S. D. (1999). Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiol. Mol. Biol. Rev.* **63**, 862–922.
34. Harpaz, Y., Gerstein, M. & Chothia, C. (1994). Volume changes on protein folding. *Structure*, **2**, 641–649.
35. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 27–38.
36. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
37. Esnouf, R. M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallog. sect. D*, **55**, 938–940.
38. Merritt, E. A. & Bacon, D. J. (1997). Raster3D photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.
39. Fujisawa, H. & Hayashi, M. (1977). Assembly of bacteriophage φX174: identification of a virion capsid precursor and proposal of a model for the functions of bacteriophage gene products during morphogenesis. *J. Virol.* **24**, 303–313.
40. Mukai, R., Hamatake, R. K. & Hayashi, M. (1979). Isolation and identification of bacteriophage φX174 prohead. *Proc. Natl Acad. Sci. USA*, **76**, 4877–4881.
41. Siden, E. J. & Hayashi, M. (1974). Role of the gene beta-product in bacteriophage φX174 development. *J. Mol. Biol.* **89**, 1–16.
42. Dokland, T., McKenna, R., Sherman, D. M., Bowman, B. R., Bean, W. F. & Rossmann, M. G. (1998). Structure determination of the φX174 closed procapsid. *Acta Crystallog. sect. D*, **54**, 878–890.

Edited by W. Baumeister

(Received 18 July 2002; received in revised form 18 October 2002; accepted 18 October 2002)