Insights into assembly from structural analysis of bacteriophage PRD1

Nicola G. A. Abrescia[1]*, Joseph J. B. Cockburn[1,2]*, Jonathan M. Grimes[1], Geoffrey C. Sutton[1], Jonathan M. Diprose[1], Sarah J. Butcher[3], Stephen D. Fuller[1], Carmen San Martín[4]†, Roger M. Burnett[4], David I. Stuart[1,2], Dennis H. Bamford[3] & Jaana K. H. Bamford[3]

[1]*Division of Structural Biology, The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7BN, UK*
[2]*Oxford Centre for Molecular Sciences, Central Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QT, UK*
[3]*Institute of Biotechnology and Faculty of Biosciences, Viikki Biocenter, University of Helsinki, P.O. Box 56, Viikinkaari 5, 00014, Finland*
[4]*The Wistar Institute, 3601 Spruce Street, Philadelphia, Pennsylvania 19104, USA*

* These authors contributed equally to this work
† Present address: Biocomputing Unit, Centro Nacional de Biotecnología, Campus Universidad Autónoma, 28049 Madrid, Spain

The structure of the membrane-containing bacteriophage PRD1 has been determined by X-ray crystallography at about 4 Å resolution. Here we describe the structure and location of proteins P3, P16, P30 and P31. Different structural proteins seem to have specialist roles in controlling virus assembly. The linearly extended P30 appears to nucleate the formation of the icosahedral facets (composed of trimers of the major capsid protein, P3) and acts as a molecular tape-measure, defining the size of the virus and cementing the facets together. Pentamers of P31 form the vertex base, interlocking with subunits of P3 and interacting with the membrane protein P16. The architectural similarities with adenovirus and one of the largest known virus particles PBCV-1 support the notion that the mechanism of assembly of PRD1 is scaleable and applies across the major viral lineage formed by these viruses.

PRD1 is a double-stranded DNA bacteriophage and prototype member of the Tectiviridae. It infects Gram-negative bacteria, such as *Escherichia coli* and *Salmonella enterica*. The mature virion has a molecular mass of 66 MDa and comprises over 20 distinct protein species[1]. Protein P3 (43 kDa) is the major component of the capsid, with 240 trimers arranged on an icosahedral lattice with pseudo-T = 25 triangulation. Pentamers of protein P31 (14 kDa) occupy the icosahedral vertices and associate with the trimeric P5 protein (34 kDa) and the receptor-binding protein P2 (64 kDa) to form flexible spikes[2]. The minor capsid protein P30 (9 kDa) is required for virus assembly[3]. The 15-kilobase linear viral genome is packaged into preformed procapsids[4] through a unique vertex[5,6], fuelled by the ATPase P9 (J.K.H.B., unpublished data). PRD1 is very different from the traditional DNA bacteriophages: instead of a tail it uses its internal membrane, acquired from the host during virus assembly[4,7], as an ejection device[8,9]. In addition, there is no critical capsid expansion after DNA packaging[10]. However, PRD1 is remarkably similar to adenovirus, and several lines of evidence, including the structural homology between the major structural proteins, support the hypothesis that the two viruses have a common ancestor[11–13]. Recently, the same structural characteristics have been found in the giant PBCV-1 (Phycodnaviridae family) and CIV viruses (Iridoviridae family)[13–15], indicating that PRD1 belongs to an extensive lineage, whose origin might predate the division of the bacterial, archaeal and eukaryotic domains of life[13]. Previous cryo-electron microscopic (cryo-EM) analysis of the PRD1 virion and X-ray analysis of isolated P3 illuminated aspects of PRD1 architecture[10,11,16,17] but the analysis presented here, using crystals of the PRD1 Sus539 mutant[1,18], reveals in atomic detail the key structural proteins, lays bare the organization of the capsid shell and reveals the membrane. The structure of the membrane is described in the accompanying paper[19]. Here we focus on the architecture of the PRD1 capsid, which suggests a scaleable assembly pathway that might apply to some of the largest known viruses and reveals receptor-binding vertex connections to the viral membrane.

Overall architecture of the virus and structural proteins

The structure determination at about 4 Å resolution for the PRD1 Sus539 mutant is described in Methods. Phase information was obtained by combining an X-ray structure of the major capsid protein[11,20] with a cryo-EM reconstruction of the virus[10,16]. Neglecting magnification errors, the deviation between Cα atoms derived from EM/X-ray hybrid imaging and the final X-ray structure is less than 2 Å within an icosahedral asymmetric unit. The resolution of the resultant electron density map is sufficient to resolve the β-strands in protein P3 (Fig. 1a). In addition, data from crystals of selenomethionated (SeMet) virus provide a SeMet difference Fourier map. Using these two electron density maps we have constructed atomic models for the minor capsid protein P30, the vertex protein P31, the membrane protein P16 and previously undescribed portions of the major capsid protein (see Methods). The analysis is based on 60-fold averaged electron density in which proteins not obeying the icosahedral symmetry are smeared out and are essentially invisible. PRD1 measures about 640 Å between opposite faces and 700 Å between opposite vertices. The Sus539 mutant lacks P2, the receptor-binding protein, whereas the trimeric P5 is not seen in the averaged electron density.

The 240 trimers of protein P3 dominate the PRD1 capsid (Fig. 1b). Each P3 subunit is composed of two 'jellyroll' domains[11,20], giving the trimer a pseudo-hexagonal shape. The icosahedral asymmetric unit of the virus is composed of four P3 trimers (Fig. 1b) arranged in a uniform orientation within the roughly planar facet (root-mean-square (r.m.s.) deviation from planarity 6.9 Å), to give the pseudo-T = 25 lattice characteristic of PRD1 and adenovirus. The cores of the 12 independent copies of P3 are essentially identical to each other and to the single crystal structures of isolated P3 (ref. 11) (r.m.s. deviation in Cα atoms, after superposition, is about 1 Å). There are, however, different conformations for the short amino-terminal and carboxy-terminal extensions, depending on the location of the subunit in the surface lattice of the virus, leading to small differences in the conformation of some loops (Fig. 1c).

We find that the capsid contains 60 copies of the small protein P30 (83 residues). This protein is essential for the assembly of complete particles[3]. P30 is remarkably extended, with essentially no secondary structure and dimensions of about $150 \times 60 \times 10$ Å^3

(Fig. 1d). The electron density of P30 is continuous, and detailed structural interpretation was assisted by the SeMet difference Fourier map, which pinpointed the methionine residues at positions 24, 45 and 69 (see Fig. 1d). The molecule is rich in proline residues (13%) and analysis of its amino acid sequence with PONDR[21] indicates that in the absence of other viral components at least the N-terminal half of the subunit would be disordered. The 60 copies of P30 completely festoon the virus, being suspended about 12 Å above the viral membrane by numerous interactions with P3 (Fig. 1d). N-terminal hooks (the only deviation from a fully extended conformation) lock two P30 subunits into the dimeric building block of the network (dimer interactions are restricted to about the first 30 residues of the subunit). Dimer axes coincide with icosahedral two-fold axes and the protein follows closely the outline of the P3 facet (P30/P3 interactions are described below).

Pentamers of protein P31 plug the 60-Å diameter holes in the P3/P30 shell at the five-fold vertices. P31 forms a compact eight-stranded β-barrel with jellyroll topology (Fig. 1e), elaborated only by 13-residue and 8-residue extensions at the N and C termini respectively. Residues 5–118 of the 126-amino-acid subunit have been modelled with the aid of the SeMet difference Fourier map. P31 resembles half of a P3 subunit and can be superposed almost equally well on either of the P3 jellyrolls (82 and 83 Cα atoms superposed for domains 1 and 2, respectively, in each case with an

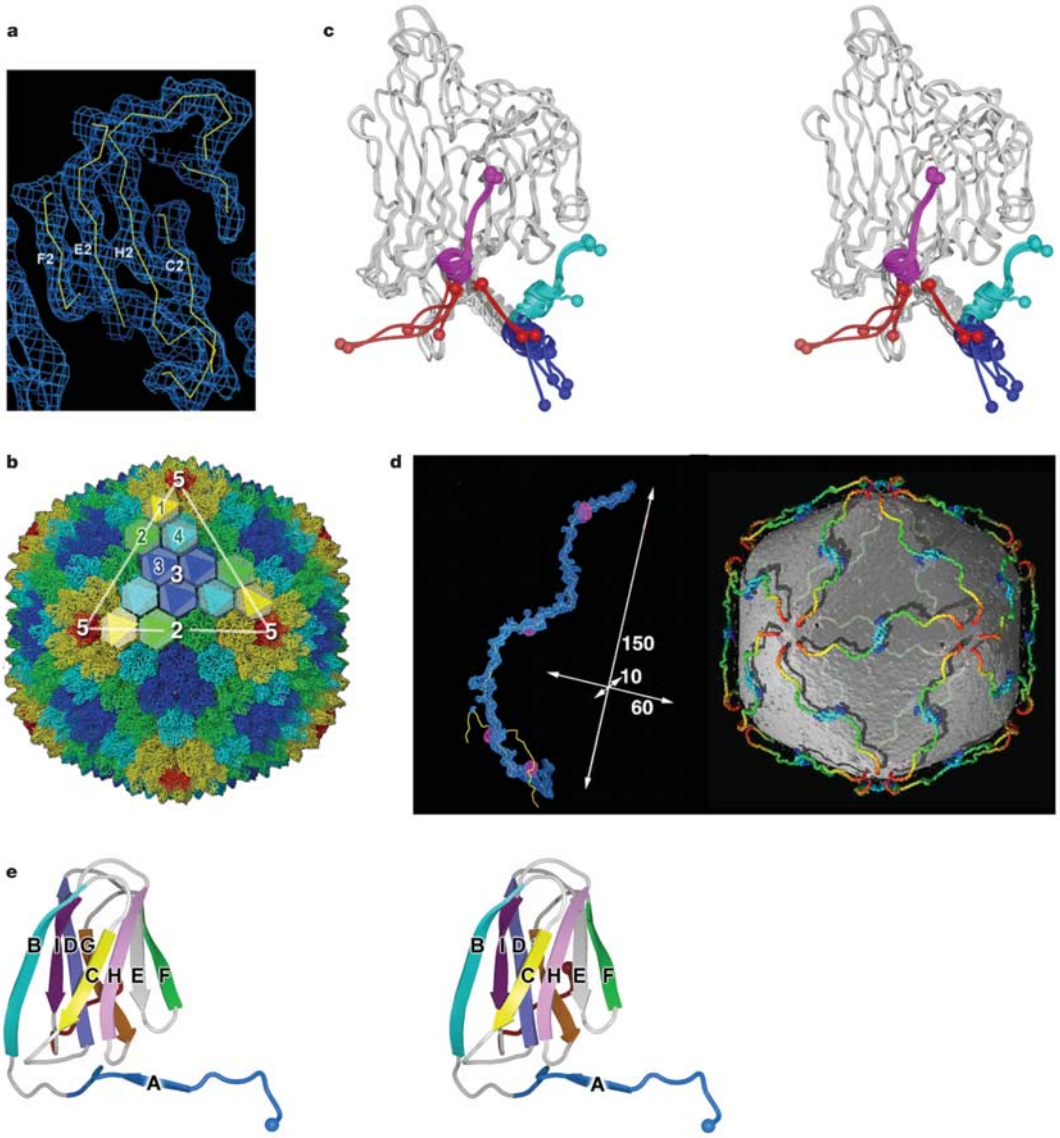

Figure 1 Architecture and structural components of bacteriophage PRD1. **a**, Electron density of cell 2 contoured at 1.5σ. The strands of the P3 second jellyroll are clearly resolved. **b**, The building blocks of a facet (the triangular area defined by white lines). The icosahedral asymmetric unit contains 12 P3 subunits arranged as four trimers, with orientations represented by triangles labelled 1 to 4 and coloured yellow, green, blue and cyan respectively (the underlying grey hexagons show the trimer morphology; numbers and symbols follow the adenovirus convention[39]). P3 trimers outside the marked facet are coloured similarly and shown as coil, as is P31, drawn in red. **c**, Stereo Cα traces of the superposed 12 unique copies of P3, showing switching at the N (blue and cyan) and C (red and magenta) termini. **d**, Protein P30. Left: royal blue, $2F_o - F_c$ electron density contoured at 1.2σ; magenta, SeMet difference density contoured at 8.5σ (the peaks, from the top, correspond to SeMet residues 69, 45 and 24). Part of a two-fold related P30 is shown in yellow. Dimensions are in ångströms. Right: 60 copies of P30 (coloured blue through green to red from the N terminus to the C terminus) wrap around the electron density of the membrane. The small holes in the membrane close to the five-fold axes are the transmembrane helices of P16. **e**, Stereo view of P31. Jellyroll strands are labelled B to G; the β-strand at the N terminus is labelled A. N-terminal and C-terminal segments are coloured blue and red respectively. β-strand residues: B, 22–31; C, 34–41; D, 48–55; E, 60–66; F, 69–76; G, 79–87; H, 90–97; I, 101–109.

r.m.s. deviation of 2.9 Å as determined with program SHP[22]), although there is no sequence similarity between them. Finally, beneath each P31 pentamer lie five copies of protein P16. Residues 7–56 and 94–116 are visible in the electron density, with residues 7–28 forming a transmembrane helix. The assignment of P16 is described in Methods.

Conformational switching in P3 underlies capsid assembly

In closed assemblies, geometric constraints prevent an exactly symmetric arrangement of more than 60 subunits. The switching of subunit conformation during the assembly process provides a mechanism for building more complex systems (see, for example, refs 23 and 24), and this was suggested for PRD1 (ref. 17). The 720 subunits of P3 present in each PRD1 particle occupy 12 distinct local environments within each facet (Fig. 2), which define the conformations of the N-terminal and C-terminal extensions. Whereas in the intact virus these extensions are generally well ordered, in the crystal structure of isolated P3 both termini were disordered (about 13 and 11 residues respectively)[11,20] implying that there is little energetic penalty in switching between conformers. The N termini of P3 fall into two distinct conformations (denoted NI and NII) that converge at residue 18 (between 9 and 12 Å above the outer leaflet of the membrane). Residues 18–34 form an α-helix that runs away from the membrane. In one set of conformations (NII, shown in dark blue in Figs 1c and 2) this helix is longer, beginning at residue 8, and is preceded by a short extended section of chain that contacts the membrane (Fig. 2). Three of the first six amino acids are glutamine residues, which probably attach to the

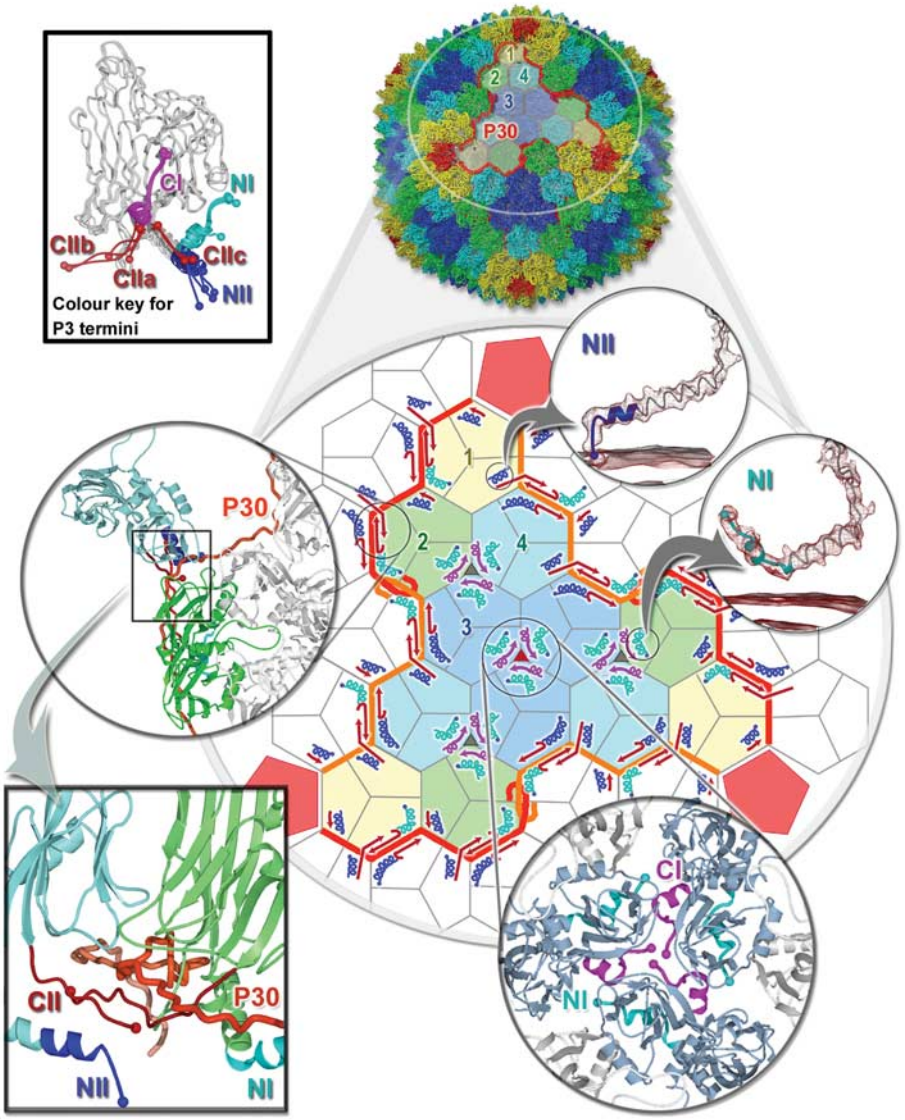

Figure 2 Switching of the P3 termini. Top left, key summarizing the conformations. Top centre, overview of the particle. The white oval marks the portion shown below. Centre, schematic diagram of a facet with neighbouring P3 trimers (white). Trimers are shown as colour-coded hexagons. The P31 pentamer is shown as a bright pink pentagon. Two-fold related P30 subunits are drawn in red and orange. Three-fold and pseudo-three-fold axes are depicted as small triangles, red and grey respectively. For each P3 subunit, the N terminus is shown as a straight or kinked helix and the C terminus as a small arrow, coloured according to the key. Type CIIa termini (the short hooks in the main diagram) form a two-stranded β-structure with P30. Type CIIb termini are further ordered, taking part in a three-stranded β-sheet; the other two strands are contributed by P30 and by the C termini of P3 subunits from the neighbouring facet (type CIIc, shown as doglegs in the main diagram). The close-ups linked by straight zoom-lines are viewed to match the diagram, and grey arrows indicate an orthogonal view. Close-ups on the right include electron density (contoured at 0.6σ) for the N termini and membrane outer leaflet.

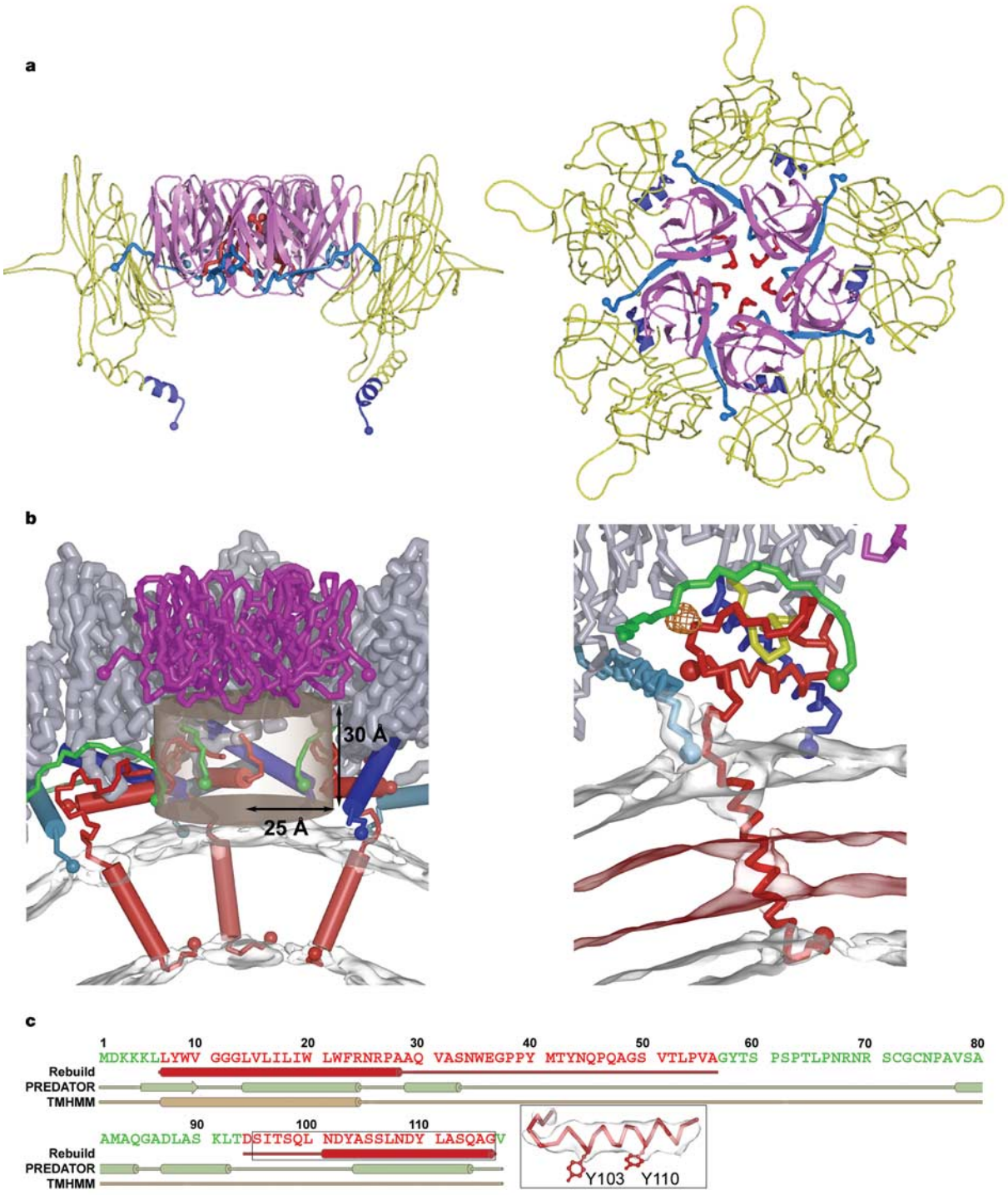

Figure 3 Vertex organization. **a**, Protein P31 and peripentonal P3 subunits. P31 is shown in lilac. The P31 N and C termini are shown as royal blue and red coils respectively. Peripentonal P3 subunits are drawn as yellow coils (residues involved in conformational switching at the N terminus, which contact the membrane, are dark blue). Left, view orthogonal to the five-fold axis. Right, view along the five-fold axis, showing the interdigitation of the N terminus of P31 (royal blue) at the interface between peripentonal P3 monomers. **b**, P16 and its interactions. Left, five copies of P16 (red) are arranged around each vertex, between the peripentonal P3 trimers (grey), of which three are shown. Each copy of P16 contributes disordered residues to the 25 Å × 30 Å cylindrical cavity under the vertex base protein P31 (magenta). The first six residues of P16 beneath the membrane are disordered but are shown for completeness. Right, each copy of P16 interacts with two subunits of the adjacent peripentonal P3 trimer clockwise around the five-fold axis and the P30 C-terminal region (green). The N terminus of the peripentonal P3 subunit furthest from the vertex (light blue) contacts P16 at the membrane. The loop connecting the two jellyrolls of the P3 subunit nearest the vertex (yellow) is clipped between P16 residues 43–54 and the P16 C-terminal helix, which is flanked on the other side by the corresponding P3 N terminus (dark blue). Protein termini are represented by balls. The outer and inner leaflet headgroup regions of the membrane are contoured in grey at 0.6σ and 0.3σ, respectively, with the transmembrane helix density contoured in red at -0.3σ. The selenomethionine peak for P16 residue 41 (which identified the protein) is contoured in orange at 8.5σ (see Methods). **c**, Secondary structure elements of our P16 model (red), secondary structure prediction (PREDATOR[49], green) and transmembrane helix prediction (TMHMM[47], brown), aligned with the amino acid sequence (amino acids visible in the electron density are in red). The electron density for the P16 C-terminal helix, contoured at 1.5σ, shows two large side chains consistent with the tyrosine residues at positions 103 and 110 (see Methods).

headgroups of the outer leaflet of the membrane. These membrane-anchoring P3 subunits are found along the edge of the icosahedral facets and might help to stabilize the fold in the membrane that occurs at this intersection[19]. In the other N-terminal conformation (NI, shown in cyan in Figs 1c and 2), the residues preceding 18 form a short α-helix twisted away from the membrane (Fig. 2) to interact with another subunit of the P3 trimer and, for subunits within the body of the facet, with the C termini (residues 385–390) of P3 subunits from neighbouring trimers. This conformational switching allows the P3 N termini to stabilize either the viral membrane or interactions with neighbouring P3 molecules. The switching of the P3 C termini is central to the assembly process, because these structures mediate P3 trimer–trimer interactions. As with the N-terminal switching, the four subunits within the body of the icosahedral face have essentially the same structure (denoted CI in Figs 1c and 2). In this conformation, after a short helix (residues 383–389) the final six residues extend to form a trimeric interaction with termini from neighbouring trimers, locking a set of three trimers together (Fig. 2). The electron density and the presence of acidic residues suggest that this structure might be further stabilized by a divalent cation lying on the local inter-trimer three-fold axis. In conformation CII (brick red on Figs 1c and 2), adopted by six subunits, the helix is replaced by an extended chain. These extensions (which constitute three subsets, CIIa–c; see Fig. 2) form β-structures with P30, which threads between the P3 trimers of neighbouring facets. P30 is clamped by P3 in four β-sheets along the edge of the facet (Fig. 2), stabilizing the facet interface. The final two P3 C-terminal structures (short red arrows in Fig. 2, found in the peripentonal P3 trimers and subunits abutting them) are disordered, eliminating C-terminal interactions between the peripentonal trimers and the rest of the structure.

Interactions at the vertex

Five subunits of P31 pack with their jellyrolls vertical to form the pentameric base of the vertex spike (Fig. 3a). The pentamer is stabilized by residues 9–14 (strand A in Fig. 1e) clipping onto strand B of the adjacent P31 β-barrel and by side-to-side interactions between the BIDG sheet of one subunit and strands G and F of its neighbour. A series of interactions engage the P31 pentamer with the P3 subunits that surround it. The nine N-terminal residues of P31 interdigitate between the B1–C1 and B2–C2 β-strands of five-fold related P3 subunits, and the ACHEF face of P31 nestles between the two jellyrolls of a peripentonal P3 subunit (Fig. 3a).

There is a symmetry mismatch between the pentameric P31 and the trimeric spike protein P5 (a similar symmetry mismatch is found in adenovirus[25]). P31 and the N-terminal domain of P5 share 38% sequence identity[26,27], which might allow a mixed pentameric association of P31 with the N-terminal domains of P5 (ref. 28). However, there is little evidence for this in the electron density maps: P31 is generally well defined, although the attenuation of the SeMet difference Fourier peak for residue Met9 would be consistent with up to two of the five subunits being P5. The C termini of P31 lie close to the five-fold axes, in line with direct P5–P31 interactions at these axes[17,27].

Below P31, in a position where unidentified density was seen in cryo-EM analysis[17], lies the 117-residue integral membrane protein P16 (ref. 29) (Fig. 3b, c), anchored into the membrane through a single transmembrane helix (residues 7–28). Residues 1–6 are disordered beneath the membrane, with the triple lysine sequence 3–5 presumably stabilizing the negative charge of the viral DNA. There is no electron density for residues 57–93, indicating that this portion (denoted $loop_{57–93}$) does not obey icosahedral symmetry. The trajectory of the polypeptide chain entering and leaving $loop_{57–93}$ indicates that it resides in a cylindrical cavity under the vertex (Fig. 3b), which it largely fills (see Supplementary Methods). P16 engages in extensive interactions with the P3 trimers adjacent to the vertex and the C-terminal region of the cementing protein P30 (Fig. 3b). The vertex complexes and peripentonal P3 trimers are labile in PRD1 virions lacking P16 (ref. 29), whereas removal of the phage membrane from PRD1 particles lacking DNA, by treatment with sodium dodecyl sulphate, removes the vertex complex and peripentonal P3 trimers[30]. Thus P16 acts as a second cementing protein stabilizing the vertices, a function analogous to that of adenovirus polypeptide VI (ref. 31). The observed lability of the vertex might facilitate the formation of a pore in the capsid for egress of the PRD1 DNA injection tube[32]. The disordered $loop_{57–93}$ probably forms non-icosahedrally symmetric interactions with the vertex base, presumably explaining arms of electron density seen connecting the P16 region and capsid vertex in the cryo-EM analysis[17]. The breakdown in icosahedral symmetry might reflect the symmetry mismatch between the pentameric vertex base and the trimeric spike shaft. Perhaps P16 acts as the component that transmits the receptor-binding signal to the membrane, priming the metastable system for DNA ejection[2]. In addition to P16, 11 further protein species have been reported to be associated with the viral membrane[1,4,32–35]; in particular, two integral membrane proteins and two capsid-associated proteins are located at a unique vertex responsible for DNA packaging[5,6]. We see no evidence of these proteins in the averaged electron density maps and unaveraged electron density maps show no features to suggest a preferred orientation of the special vertices within the crystal lattice. Visualization of the remaining membrane proteins is therefore probably beyond the reach of the current data.

Table 1 **Data collection statistics**

Location	Reflections	Resolution (Å)	$I/\sigma(I)$*	R_{merge}†	Completeness (%)
Cell 1	3829761‡	100–4.2	1.8	0.33	45
	2383561§	5.05–4.98	1.1	0.63	42
		4.27–4.20	0.9	0.65	7
Cell2	1088596‡	95.3–4.2	2.6	0.23	16
	899751§	5.05–4.98	1.3	0.45	8
		4.48–4.43	1.0	0.50	3
SeMet	5180365‡	67.4–3.7	1.49	0.42	43
	3345164§	4.27–4.17	0.96	0.59	32
		3.76–3.70	0.70	0.67	3

*σ values were adjusted to reflect the observed discrepancies between multiply measured data.
†$R_{merge} = \Sigma_h\Sigma_i|(|I_h|) - |I_{h,i}||/\Sigma_h\Sigma_i\langle I_h\rangle$, where h are the unique reflection indices, $I_{h,i}$ are the intensities of symmetry related reflections and $\langle I_h\rangle$ is the mean intensity.
‡Measured reflections.
§Unique reflections.

Table 2 **Phase refinement statistics for the merged native-SeMet cell 1 and cell 2 data sets**

Resolution (Å)	Merged cell 1			Cell 2		
	Completeness (%)	Averaging *R*-factor*	Correlation coefficient†	Completeness (%)	Averaging *R*-factor*	Correlation coefficient†
100–4.2	57	0.33	0.55	16	0.18	0.90
5.01–4.97	54	0.36	0.25	7	0.21	0.80
4.23–4.21	17	0.34	0.26	1	0.18	0.77

*R-factor $= \Sigma_h||F_o|_h - |F_c|_h|/\Sigma|F_o|_h$ where h are the unique reflection indices, F_o are the observed structure factors and F_c are the structure factors obtained from inversion of the solvent-flattened map (SHELLSCALE; D.I.S., unpublished program).
†Correlation coefficient $= \Sigma_h(\langle|F_o|_h\rangle - |F_o|_h\rangle(\langle|F_c|_h\rangle - |F_c|_h\rangle/\Sigma_h[(\langle|F_o|_h\rangle - |F_o|_h\rangle)^2 - (\langle|F_c|_h\rangle - |F_c|_h\rangle)^2]^{1/2}$ where h are the unique reflection indices, F_o are the observed structure factors and F_c are the structure factors obtained from inversion of the solvent-flattened map (SHELLSCALE; D.I.S., unpublished program).

P30 and a model for viral assembly

P30 cements P3 along the edge of the facet (Fig. 2), with the C-terminal tail of P30 reaching the penton where it also interlocks with protein P16 (at residues 70–79 of P30). Large viruses frequently use a scaffolding protein to drive assembly and control the capsid size[36], for example gp8 of bacteriophage P22 and gp7 of bacteriophage φ29, both of which are elongated proteins with a substantial α-helical content[36,37]. P30 of PRD1 is even more elongated, being almost completely extended, allowing the dimer to span the entire edge of an icosahedral facet (about 300 Å). In effect P30 acts as a molecular tape-measure, a function attributed to protein YscP in the assembly of the bacterial injectisome, and also responsible for the determination of phage tail length (ref. 38 and references therein). The linear nature of the 83-residue P30 would allow a virus of eightfold the mass of PRD1 to be assembled with a minimal 'P30' of less than 200 residues, and very large viruses (ref. 14, and S. D. Benson, J. K. H. Bamford, D. H. Bamford and R. M. Burnett, unpublished data) could be assembled with a tape-measure protein of only 300 amino acids. The efficient and accurate assembly of such complex structures with many thousands of subunits presents numerous challenges, but the limited conformational switching seen in PRD1 P3 suggests a plausible pathway in which assembly is governed by extensive β-sheet interactions between P3 and P30. The use of preformed P3 trimers and P30 dimers probably provides early quality-control steps in the assembly pathway. A plausible assembly model could involve the initial attachment of two copies of P3 trimer 2 (defined in Fig. 2) around the P30 dimer axis, followed by the attachment of trimers 3 and 1 (probably with fairly independent kinetics) to form the binding face to which trimer 4 then binds. Such an assembly of eight trimers of P3 and a dimer of P30 would represent 1/30 of the virus, and these preformed units might then clip together to form the complete virus shell. P31, P16 and perhaps other proteins are likely to be involved in the final stages of the assembly process. Particle formation is also aided by two non-structural assembly factors, P10 and P17 (ref. 32), with a large amount of P10 being associated with the viral membrane during capsid assembly[3]. We suggest that P10 is involved in the formation of correctly sized membrane vesicles, which the assembling procapsid selects through a size-exclusion mechanism to form an empty virion into which the genome is packaged through a single special vertex[6,32].

Generic features of assembly of a major virus lineage

The structure of the major capsid protein of the massive PBCV-1 (ref. 15) validated the proposal[13] that the Phycodnaviridae and Iridoviridae belong to the PRD1/adenovirus structure-based lineage, and analysis (S. D. Benson, J. K. H. Bamford, D. H. Bamford and R. M. Burnett, unpublished data) suggests that the 'double-barrelled' major capsid protein is present in other virus families including the Asfarviridae, the Ascoviridae and the very large Mimivirus. However, the set of core properties that defines the PRD1/adenovirus lineage might extend beyond the architecture of the major structural protein. The PRD1 vertex protein, P31, resembles a single jellyroll domain of P3, and presumably these proteins originated from a common ancestor constructed from a single jellyroll. The divergence of P3 and P31 probably occurred before the separation of adenovirus from PRD1, so the adenovirus penton base might possess a much elaborated jellyroll (molecular mass 63 kDa, in comparison with 13 kDa for P31), echoing the relationship between P3 of PRD1 and the adenovirus hexon[11]. The pseudo-hexagonal arrangement of structurally homologous trimeric molecules into triangular facets is another core property of the PRD1/adenovirus lineage, which probably reflects underlying similarities in the assembly mechanism[13–15]. Thus, adenovirus might possess a homologue of P30. Although postulated to have an analogous function[31], polypeptide IIIa is larger than expected for a P30 homologue and is a less attractive candidate than polypeptide VIII, which is proline-rich, is almost exactly the expected length (110 residues) and possesses an N-terminal sequence that PONDR[21] predicts to be disordered. The larger capsids such as PBCV-1 are, like PRD1 and adenovirus, assembled principally out of triangular plates composed of trimers of a single protein arranged with *p*3 symmetry[39], with neighbouring plates related by icosahedral two-fold axes, suggesting that all these viruses might use a tape-measure mechanism to control assembly. Further structural analyses are required to test this hypothesis.

Discussion

Bacteriophage PRD1 adopts a simple and effective strategy for protein shell assembly. The bulk of the capsid is cemented together by protein P30, and the vertex is stabilized by the integral membrane protein P16. The latter seems to act as a metastable sensor, linking the symmetry-mismatched vertex base and receptor-binding spike shaft with the membrane and providing a mechanism for the controlled extrusion of a membrane–protein–DNA tube to initiate host entry. Particle assembly seems to be regulated by P30, which acts as a molecular tape-measure, modulating a two-state conformational switch in the major capsid protein (P3) to control both the assembly of an initial P30–P3 core structure and the proliferation of P3–P3 contacts. This allows 60 subunits of the small, essentially linear, P30 protein to control the assembly of 720 P3 subunits. This model can be easily scaled to very large systems, and circumstantial evidence indicates that it might underpin some of the most complex self-assembling systems found in biology. □

Methods

Crystallographic data collection and processing

The growth, purification and crystallization of PRD1 Sus539 mutant has been described[1,18]. For SeMet labelling, the virus was grown on M9 glucose–thiamine medium[40] and 60 mg ml^{-1} L-selenomethionine (Calbiochem) added 20 min after infection. Diffraction data were collected at ~21 °C on beamlines ID13 and ID14 of the European Synchrotron Radiation Facility (ESRF), Grenoble[1,18]. Data were analysed with DENZO[41] and the resolution limit was determined with TRIM_DENZO (D.I.S., unpublished program). Data were scaled with a special version of SCALEPACK[41]. For the native data and SeMet data, reflections with fractional partialities of 0.7 or more and 0.5, respectively, were scaled to full intensity and incorporated into the data set as fully recorded reflections (program POST: D.I.S. and J.M.D., unpublished program). The data from native crystals fall into two sets with the same spacegroup $P2_12_12_1$ but different unit cell dimensions[18]. The SeMet-substituted crystals were isomorphous with native cell 1. The crystallographic asymmetric unit contains an entire virus particle. Data statistics are summarized in Table 1 and Supplementary Figure. Intensities were converted to structure factor amplitudes with TRUNCATE[42].

Phasing and phase refinement

Initial phases were obtained by molecular replacement with the use of a quasi-atomic model of the viral capsid (PDB entry 1HB7) derived from fitting a 14 Å cryo-EM reconstruction[16]. The position and orientation of this model were refined against the experimental data using X-PLOR 3.1 (ref. 43). This was performed separately for data from the cell 1 and cell 2 crystal forms. An initial estimate of the particle orientation was obtained from the self-rotation function calculated from the cell 1 data set. Patterson correlation refinement[44] was performed against 60–15-Å resolution data to optimize the orientation (final correlation coefficient (cc) of 0.021 for cell 1 and 0.027 for cell 2). The position and orientation of the particle were then refined at 30 Å with the E1E1 target function[43] (cc = 0.196 for cell 1, 0.195 for cell 2), giving an initial set of non-crystallographic symmetry (NCS) operators for the two cell types. P3 trimers inside the icosahedral asymmetric unit were refined at 6 Å resolution and NCS operators were re-refined at 8 Å resolution[43] (cc = 0.375 for cell 1, 0.546 for cell 2). Further refinement of the trimer positions and the NCS operators with data to 5 Å gave final correlation coefficients of 0.230 and 0.423 for cell 1 and cell 2, respectively. A $2F_o - F_c$ map was calculated for each cell type (to 4.2 Å resolution) with the corresponding model phases, which were then refined by iterative solvent flattening and 60-fold averaging; F_c was calculated from the starting model or from Fourier transformation of the modified map (see Table 2) (program GAP; D.I.S. and J.M.G., unpublished program, and CCP4 programs[42]). For missing reflections, including those at low resolution, appropriately weighted F_c values were used.

Calculation of the SeMet difference map

The SeMet and native cell 1 data were scaled in resolution shells with the use of SHELLSCALE (D.I.S., unpublished program). Difference Fourier maps were calculated using cell 1 phases and difference attenuation weighting to down-weight terms with large experimental errors[45]. The maps were 60-fold averaged. The mean (μ) and standard

deviation (σ) of the electron density was evaluated in hollow shells (outer and inner radii 30 and 15 Å, respectively) around each methionine residue in the P3, P30, P31 and P16 models, as well as the height of the difference peak (ρ_{max}) and the signal-to-noise ratio $(\rho_{max} - \mu)/\sigma$. Varying the partiality threshold and resolution showed that the difference map calculated to 4.2 Å with partiality threshold 0.5 gave the highest mean signal-to-noise ratio (9.4).

Model building

The P3 N- and C-terminal regions were built into the averaged map of cell 2 by using O[46]. Major movements of the loop regions of the different P3 subunits were fitted manually to the electron density. The SeMet difference Fourier showed five other strong peaks in the icosahedral asymmetric unit in addition to those attributable to P3. The location of three of these and the quality of the native map led to the unequivocal identification and tracing of the minor capsid protein P30 (residues 1–83). The electron density close to the five-fold axes was also interpreted. A jellyroll fold was apparent and a preliminary model was manually fitted as a rigid body, followed by retracing to match the observed map. The unique SeMet peak at position 9 identified this protein as the pentameric vertex protein P31 (residues 5–118).

The electron density maps revealed two apparently disconnected polypeptide fragments located between the peri-pentonal P3 trimers (Fig. 3b). Fragment I is a roughly 50-residue peptide containing a transmembrane helix about 20 residues long and a methionine residue (identified in the SeMet difference Fourier) about 11 residues from the point of contact with the membrane. Fragment II consists of about 23 amino acids, terminating in a roughly 17-residue α-helix. Analysing the relative positions of methionine residues and predicted transmembrane helices (program TMHMM[47]) in all the candidate sequences from the virus showed that fragment I belonged to the N-terminal region of the 117-residue integral membrane protein P16, predicted to possess an 18-residue transmembrane helix between residues 7 and 24. Placing an α-helix into the transmembrane electron density (Fig. 3b, right) and positioning the P16 residue Met 41 in the vicinity of the SeMet peak enabled the construction of a C_α trace for P16 residues 7–56. To give the best fit to the electron density, the transmembrane helix was extended from residue 24 to residue 28. Fragment II was assigned as the P16 C terminus on the basis of protein packing density considerations in the cavity under the vertex (see Supplementary Methods) and the alignment of tyrosines at positions 103 and 110 (the largest amino acids in the C-terminal sequence) to two particularly pronounced bumps seven residues apart (Fig. 3c). All structures were brought into conformity with expected backbone conformations and side chains were built with program CALPHA[48].

Received 9 March; accepted 21 September 2004; doi:10.1038/nature03056.

1. Bamford, J. K. *et al.* Diffraction quality crystals of PRD1, a 66-MDa dsDNA virus with an internal membrane. *J. Struct. Biol.* **139,** 103–112 (2002).
2. Rydman, P. S. *et al.* Bacteriophage PRD1 contains a labile receptor-binding structure at each vertex. *J. Mol. Biol.* **291,** 575–587 (1999).
3. Rydman, P. S., Bamford, J. K. & Bamford, D. H. A minor capsid protein P30 is essential for bacteriophage PRD1 capsid assembly. *J. Mol. Biol.* **313,** 785–795 (2001).
4. Mindich, L., Bamford, D., McGraw, T. & Mackenzie, G. Assembly of bacteriophage PRD1: particle formation with wild-type and mutant viruses. *J. Virol.* **44,** 1021–1030 (1982).
5. Stromsten, N. J., Bamford, D. H. & Bamford, J. K. The unique vertex of bacterial virus PRD1 is connected to the viral internal membrane. *J. Virol.* **77,** 6314–6321 (2003).
6. Gowen, B., Bamford, J. K., Bamford, D. H. & Fuller, S. D. The tailless icosahedral membrane virus PRD1 localizes the proteins involved in genome packaging and injection at a unique vertex. *J. Virol.* **77,** 7863–7871 (2003).
7. Bamford, D. H., Caldentey, J. & Bamford, J. K. Bacteriophage PRD1: a broad host range dsDNA tectivirus with an internal membrane. *Adv. Virus Res.* **45,** 281–319 (1995).
8. Grahn, A. M., Daugelavicius, R. & Bamford, D. H. Sequential model of phage PRD1 DNA delivery: active involvement of the viral membrane. *Mol. Microbiol.* **46,** 1199–1209 (2002).
9. Grahn, A. M., Daugelavicius, R. & Bamford, D. H. The small viral membrane-associated protein P32 is involved in bacteriophage PRD1 DNA entry. *J. Virol.* **76,** 4866–4872 (2002).
10. Butcher, S. J., Bamford, D. H. & Fuller, S. D. DNA packaging orders the membrane of bacteriophage PRD1. *EMBO J.* **14,** 6078–6086 (1995).
11. Benson, S. D., Bamford, J. K., Bamford, D. H. & Burnett, R. M. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* **98,** 825–833 (1999).
12. Belnap, D. M. & Steven, A. C. 'Deja vu all over again': the similar structures of bacteriophage PRD1 and adenovirus. *Trends Microbiol.* **8,** 91–93 (2000).
13. Bamford, D. H., Burnett, R. M. & Stuart, D. I. Evolution of viral structure. *Theor. Popul. Biol.* **61,** 461–470 (2002).
14. Yan, X. *et al.* Structure and assembly of large lipid-containing dsDNA viruses. *Nature Struct. Biol.* **7,** 101–103 (2000).
15. Nandhagopal, N. *et al.* The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proc. Natl Acad. Sci. USA* **99,** 14758–14763 (2002).
16. San Martín, C. *et al.* Combined EM/X-ray imaging yields a quasi-atomic model of the adenovirus-related bacteriophage PRD1 and shows key capsid and membrane interactions. *Structure* **9,** 917–930 (2001).
17. San Martín, C. *et al.* Minor proteins, mobile arms and membrane-capsid interactions in the bacteriophage PRD1 capsid. *Nature Struct. Biol.* **9,** 756–763 (2002).
18. Cockburn, J. J., Bamford, J. K., Grimes, J. M., Bamford, D. H. & Stuart, D. I. Crystallization of the membrane-containing bacteriophage PRD1 in quartz capillaries by vapour diffusion. *Acta Crystallogr. D* **59,** 538–540 (2003).
19. Cockburn, J. J. *et al.* Membrane structure and interactions with protein and DNA in bacteriophage PRD1. *Nature* doi:10.1038/nature03053 (this issue).
20. Benson, S. D., Bamford, J. K., Bamford, D. H. & Burnett, R. M. The X-ray crystal structure of P3, the major coat protein of the lipid-containing bacteriophage PRD1, at 1.65 Å resolution. *Acta Crystallogr. D* **58,** 39–59 (2002).
21. Li, X., Romero, P., Rani, M., Dunker, A. K. & Obradovic, Z. Predicting protein disorder for N-, C-, and internal regions. *Genome Informatics* **10,** 30–40 (1999).
22. Stuart, D. I., Levine, M., Muirhead, H. & Stammers, D. K. Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 Å. *J. Mol. Biol.* **134,** 109–142 (1979).
23. Liddington, R. C. *et al.* Structure of simian virus 40 at 3.8 Å resolution. *Nature* **354,** 278–284 (1991).
24. Grimes, J. M. *et al.* The atomic structure of the bluetongue virus core. *Nature* **395,** 470–478 (1998).
25. Chiu, W., Burnett, R. M. & Garcea, R. L. in *Structural Biology of Viruses* (eds Chiu, W., Burnett, R. M. & Garcea, R. L.) 209–238 (Oxford Univ. Press, New York, 1997).
26. Caldentey, J., Tuma, R. & Bamford, D. H. Assembly of bacteriophage PRD1 spike complex: role of the multidomain protein P5. *Biochemistry* **39,** 10566–10573 (2000).
27. Sokolova, A. *et al.* Solution structure of bacteriophage PRD1 vertex complex. *J. Biol. Chem.* **276,** 46187–46195 (2001).
28. Xu, L., Benson, S. D., Butcher, S. J., Bamford, D. H. & Burnett, R. M. The receptor binding protein P2 of PRD1, a virus targeting antibiotic-resistant bacteria, has a novel fold suggesting multiple functions. *Structure* **11,** 309–322 (2003).
29. Jaatinen, S., Viitanen, S., Bamford, D. H. & Bamford, J. K. The integral membrane protein P16 of bacteriophage PRD1 stabilizes the adsorption vertex structure. *J. Virol.* **78,** 9790–9797 (2004).
30. Luo, C., Butcher, S. & Bamford, D. H. Isolation of a phospholipid-free protein shell of bacteriophage PRD1, an *Escherichia coli* virus with an internal membrane. *Virology* **194,** 564–569 (1993).
31. Stewart, P. L., Fuller, S. D. & Burnett, R. M. Difference imaging of adenovirus: bridging the resolution gap between X-ray crystallography and electron microscopy. *EMBO J.* **12,** 2589–2599 (1993).
32. Bamford, D. & Mindich, L. Structure of the lipid-containing bacteriophage PRD1: disruption of wild-type and nonsense mutant phage particles with guanidine hydrochloride. *J. Virol.* **44,** 1031–1038 (1982).
33. Bamford, J. K. *et al.* Genome organization of membrane-containing bacteriophage PRD1. *Virology* **183,** 658–676 (1991).
34. Rydman, P. S. & Bamford, D. H. Bacteriophage PRD1 DNA entry uses a viral membrane-associated transglycosylase activity. *Mol. Microbiol.* **37,** 356–363 (2000).
35. Rydman, P. S. & Bamford, D. H. The lytic enzyme of bacteriophage PRD1 is associated with the viral membrane. *J. Bacteriol.* **184,** 104–110 (2002).
36. Thuman-Commike, P. A. *et al.* Mechanism of scaffolding-directed virus assembly suggested by comparison of scaffolding-containing and scaffolding-lacking P22 procapsids. *Biophys. J.* **76,** 3267–3277 (1999).
37. Morais, M. C. *et al.* Bacteriophage phi29 scaffolding protein gp7 before and after prohead assembly. *Nature Struct. Biol.* **10,** 572–576 (2003).
38. Journet, L., Agrain, C., Broz, P. & Cornelis, G. R. The needle length of bacterial injectisomes is determined by a molecular ruler. *Science* **302,** 1757–1760 (2003).
39. Burnett, R. M. The structure of the adenovirus capsid. II. The packing symmetry of hexon and its implications for viral architecture. *J. Mol. Biol.* **185,** 125–143 (1985).
40. Sambrook, J. & Russell, D. W. *Molecular Cloning: a Laboratory Manual* 3rd edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2001).
41. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
42. Collaborative Computational Project, Number 4. The CCP 4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
43. Brunger, A. T. *X-PLOR, Version 3.1: A system for X-ray crystallography and NMR* (Yale Univ. Press, New Haven, Connecticut, 1992).
44. DeLano, W. L. & Brunger, A. T. The direct rotation function – rotational Patterson correlation search applied to molecular replacement. *Acta Crystallogr. D* **51,** 740–748 (1995).
45. Diprose, J. M. *et al.* Translocation portals for the substrates and products of a viral transcription complex: the bluetongue virus core. *EMBO J.* **20,** 7229–7239 (2001).
46. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
47. Sonnhammer, E. L. L., von Heijne, G. & Krogh, A. in *Proc. Sixth International Conference on Intelligent Systems for Molecular Biology* (ed. Glasgow, J.) 175–182 (AAAI Press, Menlo Park, California, 1998).
48. Esnouf, R. M. Polyalanine reconstruction from Cα positions using the program CALPHA can aid initial phasing of data by molecular replacement procedures. *Acta Crystallogr. D* **53,** 665–672 (1997).
49. Frishman, D. & Argos, P. Seventy-five percent accuracy in protein secondary structure prediction. *Proteins* **27,** 329–335 (1997).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank M.-L. Perälä and S. Ollila for technical assistance, the staff at the European Synchrotron Radiation Facility for beamline support, and E. Mancini for assistance with data collection. This investigation was supported by research grants from the Academy of Finland to J.K.H.B. and S.J.B., research grants and the Finnish Centre of Excellence Program (2000–2005) from the Academy of Finland to D.H.B., a grant from the National Science Foundation to R.M.B., the Biotechnology and Biological Sciences Research Council and the Medical Research Council, UK and a grant from the Human Frontiers Science Program to D.I.S., R.M.B. and D.H.B. S.D.F. is supported by the Wellcome Trust, J.M.G. by the Royal Society and D.I.S. by the UK Medical Research Council.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.I.S. (dave@strubi.ox.ac.uk) or J.K.H.B. (jaana.bamford@helsinki.fi). Coordinates for proteins P3, P30, P31 and P16 within the icosahedral asymmetric unit have been deposited in the Protein Data Bank under accession code 1w8x.