doi:10.1006/jmbi.2001.5278 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 315, 663–676

JMB

Detailed Architecture of a DNA Translocating Machine: The High-resolution Structure of the Bacteriophage ϕ29 Connector Particle

Alicia Guasch[1]†, Joan Pous[1]†, Borja Ibarra[2], F. Xavier Gomis-Rüth[1], José María Valpuesta[2], Natalia Sousa[2], José L. Carrascosa[2] and Miquel Coll[1]*

[1]*Institut de Biologia Molecular de Barcelona, CSIC, Jordi Girona 18-26, E-08034 Barcelona, Spain*

[2]*Centro Nacional de Biotecnología, CSIC Universidad Autónoma de Madrid, Canto Blanco E-28049, Madrid, Spain*

**Corresponding author*

The three-dimensional crystal structure of the bacteriophage ϕ29 connector has been solved and refined to 2.1 Å resolution. This 422 kDa oligomeric protein connects the head of the phage to its tail and translocates the DNA into the prohead during packaging. Each monomer has an elongated shape and is composed of a central, mainly α-helical domain that includes a three-helix bundle, a distal α/β domain and a proximal six-stranded SH3-like domain. The protomers assemble into a 12-mer, propeller-like, super-structure with a 35 Å wide central channel. The surface of the channel is mainly electronegative, but it includes two lysine rings 20 Å apart. On the external surface of the particle a hydrophobic belt extends to the concave area below the SH3-like domain, which forms a crown that retains the particle in the head. The lipophilic belt contacts the non-matching symmetry vertex of the capsid and forms a bearing for the connector rotation. The structure suggests a translocation mechanism in which the longitudinal displacement of the DNA along its axis is coupled to connector spinning.

Keywords: bacteriophage phi29; X-ray difraction; connector particle; DNA packaging machinery

Introduction

DNA translocation is a central biological issue since it is essential to the processing and transfer of genetic information, including cell infection by viruses. It involves a number of basic mechanisms that are probably common to many systems (nucleic acid replication, transcription, translation, conjugation, etc.). Many proteins involved in DNA translocation, or processive movements along the DNA, function as oligomers with a general toroidal morphology.[1] Among these toroidal particles are the bacteriophage head-to-tail connectors, present in one particular vertex of viral heads. The head-to-tail connecting region of bacteriophage ϕ29 is one of the best known. ϕ29 is a small double-stranded DNA bacteriophage that infects *Bacillus subtilis* cells. The viral particles are formed by a prolate icosahedral capsid or head, which contains a 19 kb long DNA, and a tail of complex structure. Between these two structures there is a connecting region called portal vertex or connector. This head-to-tail connector plays an important role in the first steps of head assembly and is the central piece of the DNA packaging machinery. In addition, during viral infection, the DNA is ejected through the tail and hence traverses the connector in the opposite direction (for reviews see references[2,3]).

The packaging of DNA into the head involves, besides the connector, two other essential components that interact in its distal part: an RNA called pRNA,[4] and the ATPase p16, required to provide energy to the translocation machinery. Both RNA and ATPase are of viral origin and, once the DNA has been packaged, they are released from the connector while the collar and tail components assemble from its distal part.

Electron microscopy studies, based on two-dimensional projections and three-dimensional reconstructions, showed that the ϕ29 connector is an oligomeric protein built up from 12 copies of

†Both authors contributed equally to this study and share first authorship.

E-mail address of the corresponding author: mcccri@ibmb.csic.es

0022-2836/02/040663–14 $35.00/0

the 36 kDa gene 10 product.[5] This structure is shaped like a propeller, in two parts: a wide disk, with protruding lobules that interacts with the capsid proteins, and a narrower cylinder that has a longitudinal channel along its axis and contacts the collar and tail proteins.[6,7]

In spite of the absence of sequence homology, viral connectors have similar structural features. The three-dimensional reconstructions of the connector from bacteriophages T3[8] and SPP1[9] have revealed striking similarities to the ϕ29 connector. This common architecture is also seen in other viruses: the connectors from bacteriophages, λ,[10] p22[11] and T4[12] show projection images that are fully consistent with a ring of appendages enclosing a narrower domain with a channel, similar to those obtained from ϕ29,[13,14] T3,[15] or SPP1.[16] These connectors differ only in dimensions (which correlate to the molar mass of the connector proteins), and all of them present 12 outer appendages, except for SPP1, which has 13.

We have previously reported the crystallization of the ϕ29 connector particle in two crystal forms (*C*2 and $P4_2$).[17] The crystals diffracted to 3.6 and 3.0 Å, respectively, the second form showing merohedral twinning. These initial crystallographic studies, together with the cryo electron microscopy data[18] and the atomic force microscopy images obtained from connector aggregates in solution,[19] solved a controversial issue about the symmetry of the ϕ29 connector (reviewed in reference[3]). A self-rotation function calculation demonstrated unambiguously the 12-fold local symmetry of the particle.[20] Recently, Simpson *et al*.[21] reported an unrefined low-resolution crystal structure of the connector, determined from *C*2 crystals. Although our present data fit the general outline of that structure, the previous model differs markedly in both the chain tracing and the side-chain positioning at the proximal part, where five β-strands in each of the 12 protomers are missing, and the main chain is out of register from residue 271 onward. Therefore, the crucial SH3-like domain is not described in the Simpson model.

Here we report the high-resolution 2.1 Å structure of the bacteriophage ϕ29 connector particle, obtained from a new crystal form. The quality of these data allowed us to correctly trace the polypeptide chain and refine the structure. The resulting structural details reveal an amazing DNA translocation device and suggest a simple rigid-body rotary mechanism of the connector coupled to the linear movement of the double helix through its central channel.

Results

Quality of the data

All previous attempts to obtain high-quality diffracting data on the connector particle were unsuccessful. The size of the particle and, most probably, the presence of flexible regions at the N and C terminus of the protein have so far prevented the collection of high-resolution data. This problem was solved by the controlled proteolysis of p10 by the protease Glu-C from *Staphyloccocus aureus* V8, which removes the first 13 residues from the N terminus and the last 18 residues from the C terminus.[22] The *C*2 crystals obtained previously by us[17] and others[21] contain one or several particles in the asymmetric unit. This feature facilitates the phase extension procedure, because 12-fold averaging can be applied, but the monoclinic crystals hardly diffract to 3.5 Å. On the other hand, the orthorhombic $P4_2$ crystals[17] diffracted to higher resolution (about 3 Å) but showed a merohedral twinning problem, where a 50 % mixture of up and down connector orientations was present.

The crystals here reported have a new form, space group *I*422. One connector particle occupies a special position with its 12-fold local axis coincident with the crystallographic 4-fold axis. Therefore, the asymmetric unit includes only three protomers. The particles are tightly packed with many inter-particle contacts and the crystals diffract to 2.1 Å, a remarkably high resolution for a particle of this size.

Structure determination

A 10 Å resolution electron microscopy model[8] was used for initial phasing in the averaging/phase extension procedure. The coincidence of the 12-fold axis of the connector with the crystallographic 4-fold axis and the fact that only 3-fold averaging could be applied resulted in an un-interpretable map. However, because high resolution was available, this map was used for an automatic structure determination with the program Arp/Warp. By alternating Arp/Warp cycles with 3-fold averaging cycles, the map improved substantially and some secondary structure elements showed up, in particular the long helices of the central domain. At this stage, however, the coordinates of the Simpson X-ray model became available (pdb code 1fou) and they were also used for molecular replacement. The model was not good enough to provide a solution using a protein monomer, but a search with the whole connector (see Materials and Methods), gave an identical solution to that obtained with the electron microscopy model. From the resulting maps, given the high resolution of the data, the chain tracing could be corrected. Subsequently, the structure was fully refined, with the addition of solvent molecules (Figure 1(b)).

Structure of the monomer

Each monomer consists of three domains (Figure 2(a)), a central domain that extends to the internal face of the wide proximal end, an SH3-like domain at the exterior of the wide proximal end, and a distal domain at the tip of the narrow end. The central domain contains five α-helices (α1, α2, α3, α5 and α6) and two β-strands (β1 and β11).

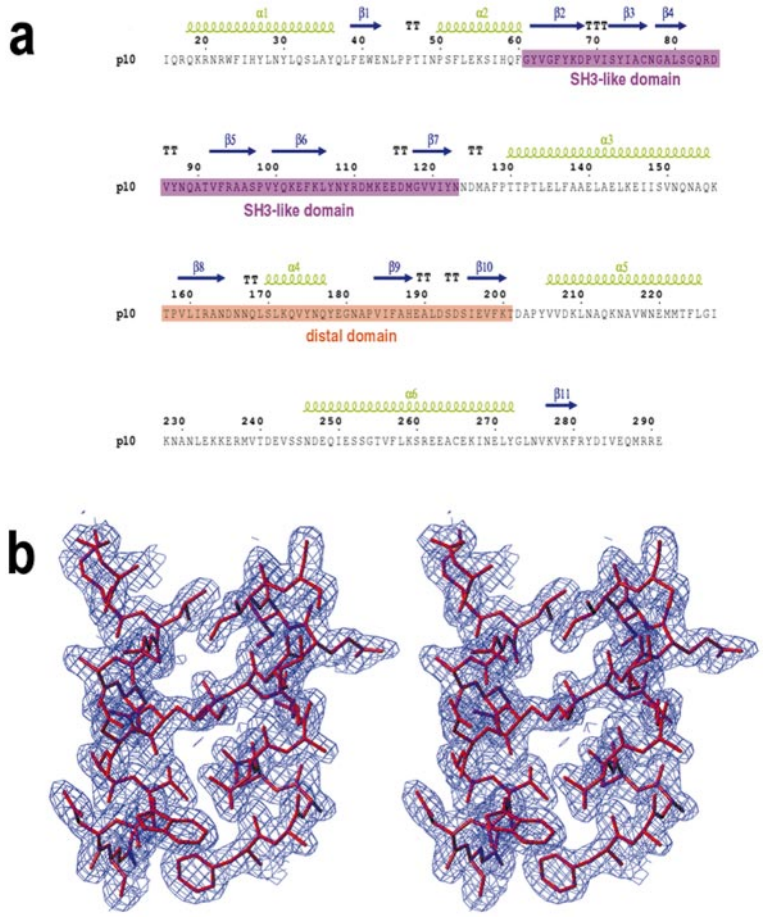

Figure 1. (a) Primary and secondary structure of bacteriophage ϕ29 connector protein. Helices (α1- α6) are displayed in green and strands (β1-β11) as blue arrows, as determined with DSSP.[57] Distinct subdomains of the structure are shadowed in pink and magenta. (b) Stereo pair of the final sigmaA-weighted $2F_o - F_c$ electron density map contoured at the 1σ level, superimposed over the final refined model shown as red sticks, around the region of interaction between helices α5 and α6.

Figure 1(a) shows the residues included in each secondary structure element. Helices α1, α3 and α5 form a bundle in which α3 and α5 are anti-parallel and α1 forms an angle of 35° with the other two, running anti-parallel to α3. Helix α2 is located at the C terminus of α1 and the N terminus of α3, almost perpendicular to both, while helix α6 is placed at the C terminus of α5. Helix α6 has a kink between residue Val256 and Phe257. This causes a deviation of the helical axis of about 30°. Interestingly, the axes of helices α5 and the second part of α6 (after the kink) are almost coincident and thus, this part of α6 appears as a continuation of α5. Finally strands β1 and β11 form an anti-parallel β-ribbon, at the proximal part of the central domain.

A total of 14 residues, from 231 to 244, in the loop between α5 and α6, are not seen in the electron density maps. We checked the molecular weight of the crystallized protein by mass spectroscopy and ruled out proteolytic cleavage at this site. Therefore, this loop is a flexible region, facing the interior of the channel, and it might play a role in closing the gate after DNA packaging. The rest of the loops of this central domain are well defined and include two β-turns, between β1 and α2 and between β7 (of the proximal SH3-like domain) and α3.

The SH3-like domain is inserted between α2 and α3 of the central domain, and forms the protruding lobules of the proximal wide part of the particle. It consists of six β-strands (β2 to β7) with an SH3-like β-barrel topological arrangement, where β2, β3 and β7 form an anti-parallel three-stranded β-sheet which lies perpendicular to the second anti-parallel three-stranded β-sheet formed by β4, β5 and β6 (Figure 2(b)). The two sheets enclose a compact hydrophobic core that includes six aromatic residues, three of them (Tyr66, Phe104 and Phe93) perfectly stacked. The connections of β2 with β3, β3 with β4 and β5 with β6 are rather short, with a β-turn between β2 and β3. On the other hand, the connections from β4 to β5 and β6 to β7 are ten-residue long loops or hairpins, both including β-turns. A structural homology search with the program DALI [23] did not reveal any similarity with other known structures. Nevertheless, it is clear from Figure 2(b) and superposition studies that the topology is similar to the SH3 protein-protein recognition domains. However, it does not form a cleft for polyproline binding, because loop β6-β7 would occupy part of the cleft space and loop β2-β3 is too short to form one of the walls of the cleft, in contrast to the long equivalent RT-Src loop of typical SH3 domains.

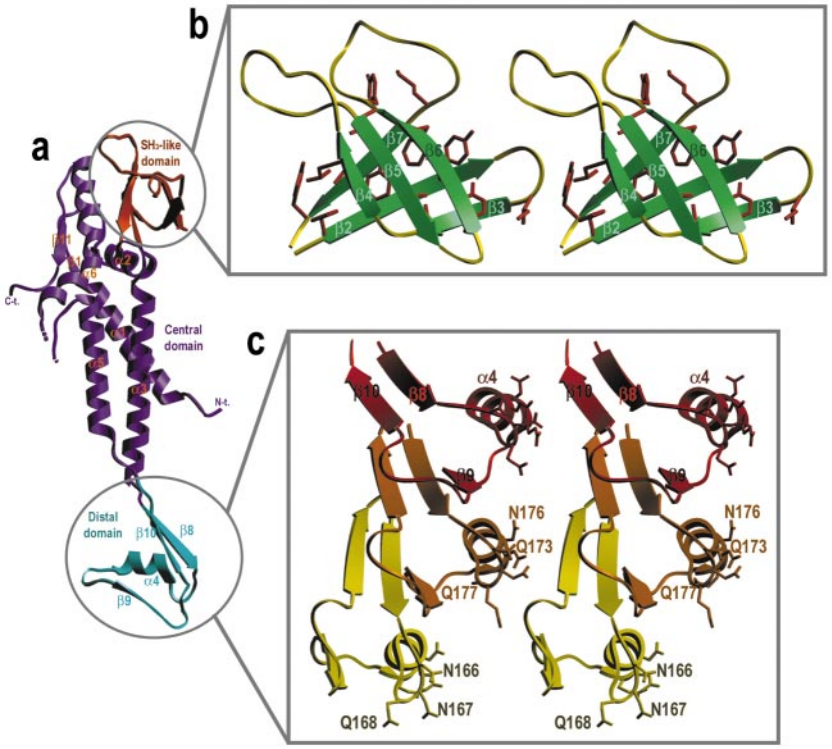

Figure 2. (a) Ribbon cartoon of a bacteriophage ϕ29 connector protein monomer displaying the labelled secondary structure elements. The three domains of the connector monomer are shown in different colours. (b) and (c) Close up stereo views of the two distinct apical domains of the bacteriophage ϕ29 connector protein monomer, the SH_3-like domain (b) and the distal domain (c). The regular secondary structure elements are labelled. In (b) the hydrophobic residue side-chains making up the filling of the β-sandwich are displayed as orange sticks. In (c) three vicinal distal domains are displayed in red, orange and yellow, respectively. In the latter two, the side-chains of the residues making up the Gln/Asn belt are displayed as sticks. Figure created with SETOR.[58]

The distal domain, at the narrow tip of the connector, consists of three β-strands (β8, β9 and β10), one α-helix (α4) and a β-turn located between the last two β-strands (Figure 2(c)). Two strands (β8 and β10) of one monomer and one strand (β9) of an adjacent monomer form a mixed β-sheet where β8 and β9 are parallel and β10 anti-parallel. Helix α4 runs anti-parallel to β8 and also belongs to the adjacent monomer (Figure 2(c)). In this way the monomers are strongly interconnected.

The architecture of the p10 protein monomer does not share any similarity with other known protein structures. Only the disposition of the three-helix bundle at the central domain is reminiscent of other three-helix bundles, like that found in the N-terminal domain of the low-density lipoprotein receptor related protein RAP (pdb code 1lre).

Structure of the dodecamer

Twelve protomers assemble in a propeller-like super-structure that has an external diameter of 146 Å at its proximal part and 77 Å at its distal end (Figure 3(a) and (b)). The height of the connector is 75 Å. The three helices forming the bundle of the central domain are not parallel with the 12-fold axis of the particle, but laterally inclined, about 45°. This implies that the proximal part of monomer n, including the SH3-like protuberance, projects over the distal domain of monomer $n+2$, as seen when looking down the 12-fold axis of the particle.

The interactions between monomers are different in the three parts of the structure. For example (as indicated above), in the distal domain one of the β-strands of the three-stranded β-sheet belongs to the adjacent monomer and, therefore, multiple hydrogen bonds form between monomers (Figures 2(c) and 4(c)). In addition, there are several hydrophobic contacts, involving eight non-polar residues: Tyr175A-Pro183B, Leu171A-Ile185B, Asp165A-Ala187B, Thr199A-Ile196B and Leu160A-Ile196B (A, B and C denote adjacent monomers). Furthermore, there is an H-bond between Asp194A and Ser100B. Interestingly, residues of three consecutive monomers (Thr157A, Tyr175B and Pro183C) converge at one point establishing hydrophobic contacts between each other.

In the central domain most of the contacts are formed by residues belonging to the helices (Figure 4(b)). Helices α5 of two consecutive monomers run parallel along almost their entire length, establishing multiple contacts. Helix α3 contacts an adjacent α3 with its C-terminal half, while its N-terminal half contacts α1. The N-terminal part of the kinked α6 helix interacts with the last turns of α1 and α5 of a neighbouring protomer.

In the proximal part, the C-terminal half of helix α6 contacts the C terminus of helix α2 and residues of strand β1 (Figure 4(a)). More important are the interactions at the SH3-like domain. Loop β5-β6, together with the loop β1-α2 (which does not belong to the SH3-like domain, but which is also in the proximal part of the connector) define a hydrophobic cleft with strand β3 at the bottom. Loop β4-β5 of a neighbouring molecule intrudes into this cleft establishing hydrophobic interactions. Tyr87A is at the tip of the loop and forms several hydro-

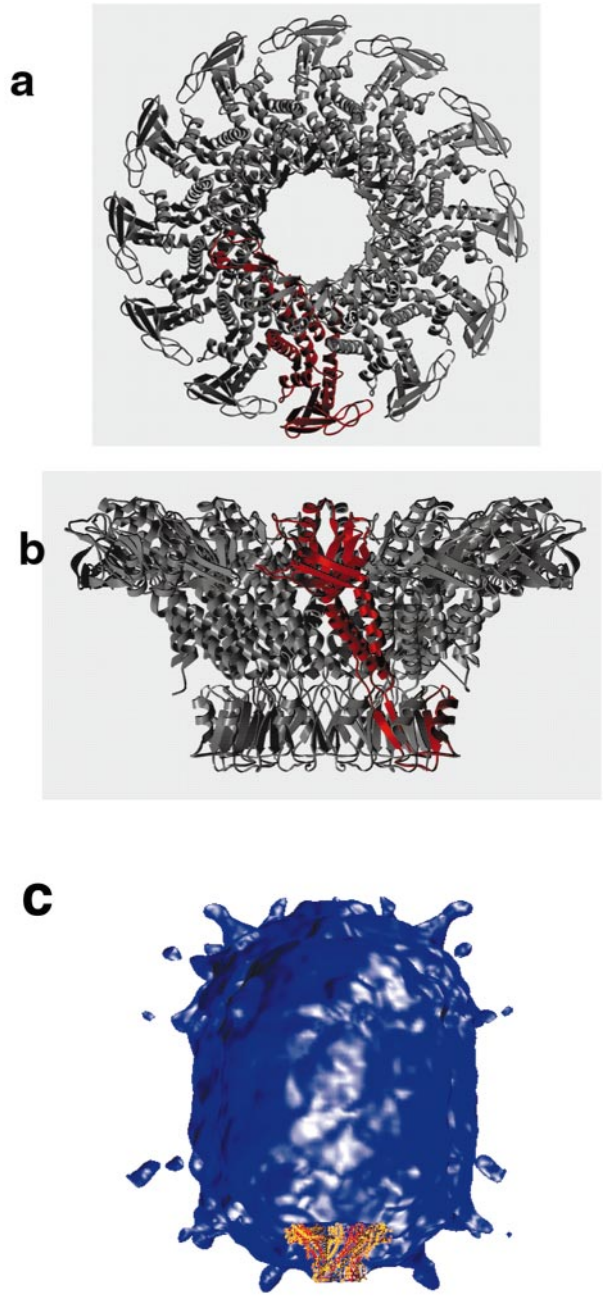

Figure 3. (a) and (b) Ribbon representations of a whole dodecamer of the bacteriophage ϕ29 connector protein. One of the 12 monomers is displayed in distinct red. The structure is shown in its axial (a) and lateral (b) views. Figure made with VOLUMES/BOBSCRIPT[59,60] and rendered with Raster-3D.[61] (c) Location of the connector structure in the viral proheads. The connector X-ray structure was fitted into the three-dimensional reconstruction of the ϕ29 prohead obtained by cryo-electron microscopy at 3 nm resolution[34]. The fits of both volumes and their graphic representation were generated using the VMD program.[62]

phobic contacts with Ala75B and Ile74B, both from β3. It is edge-to-face stacked on Phe52B. Furthermore, its OH group forms H-bonds with the carbonyl oxygen of Thr47B and the amide nitrogen of Ala75B. Val86A of the same β4-β5 loop contacts Ile74B and Val99B, while Gln89A forms an H-bond with Asn49B. Further contacts are observed between loop β6-β7 and loop β1-α2 of a neighbouring protomer, including Arg110A-Pro46B and Tyr109A-Thr47B.

The buried surface at a monomer-monomer interface is 5519 Å^2, which means that, in the dodecamer, only one third of the total area of each monomer is accessible. Thus, a large proportion of the surface is involved in the inter-protomer contacts, resulting in a strongly bound particle. This is in accordance with previous biochemical data indicating that single monomers are never observed, even at high salt concentrations. The connector structure is disassembled[24] only when incubated at high chaotropic agent concentrations (2.5 M guanidinium hydrochloride).

Internal channel and surfaces

A channel traverses the particle longitudinally. Its width is about 35 Å at the distal end and widens up at the proximal end. The width at the proximal end is difficult to estimate accurately, however, since the disordered and hence not modeled residues 231-244 and the excised C terminus are present in that area. The proximal end of the channel may thus be much narrower than is shown in this model.

In Figure 5(a), a space filling representation highlights different features of the exterior surface of the particle. Going from the proximal wide end to the distal narrow end, we find first the protuberance that corresponds to the SH3-like domain with positively and negatively charged polar residues. At the subsequent concave surface, there is a belt of hydrophobic residues, including Phe24, Ile25, Leu28, Phe60, Phe128, Pro129, and Pro132. Then, another area of mixed positively and negatively charged residues follows. This part corresponds to the N terminus of the protein, where 14 residues were excised. Further down, a ring with exposed Asn and Gln residues forms the external surface of the distal domain. It encompasses a large concentration of residues with side-chains having amide moieties, including Asn166, Asn167, Gln168, Gln173, Asn176 and Gln177. All these residues are found at the exposed face of helix α4 (Figure 2(c)). Finally, at the apical end of the connector, surrounding the entrance to the channel, there is an acidic ring formed by the side-chains of Glu189, Asp192 and Asp194.

Figure 5(b) shows the electrostatic potential interior surface of the connector, which appears to be highly electronegative. However, closer inspections reveals two electropositive rings, due to the presence of residues Lys200 and Lys209, the first in the middle of the narrow channel and the second at the beginning of the wide channel. The distance between the two rings (i.e. from Lys200B C^{α} to Lys209A C^{α}) is 20 Å, while the distance between adjacent lysine residues of the same ring is 9-10 Å.

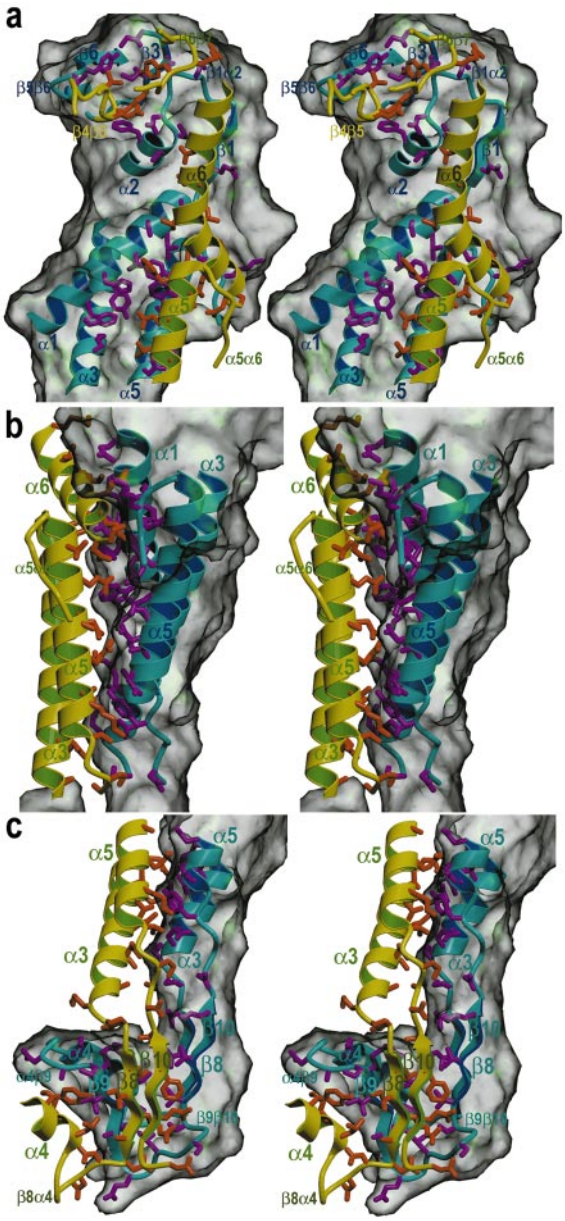

Figure 4. Close up stereo view of a protein monomer in light blue (ribbons) and magenta (selected side-chains), superimposed over its semi-transparent Connolly surface, and the segments of a vicinal monomer (yellow ribbons, orange side-chains) taking part in inter-monomer contacts. The three subfigures display the SH_3-like domain and the upper part of the central domain (a), the central helical domain (b) and the distal domain with the lower part of the central domain (c). Figure prepared with VOLUMES/BOBSCRIPT[59,60] and rendered with Raster-3D.[61]

The electron density is well defined for most of these lysine side-chains. The diameters of the rings are 36 and 41 Å, respectively. In addition, the disordered 231-244 area contains two lysine residues, Lys234 and Lys235. Both residues 231 and 244 are at 20 Å from the second lysine ring towards the proximal end. Therefore, there are two (and probably three) positively charged rings along the channel, at 20 Å intervals. Both Lys200 and Lys209 have their electronegative counterparts in close proximity, Asp202 and Asp208. It is interesting that the position of the aspartate side-chains is quite similar in both rings, with an alternation of Asp and Lys residues around a channel section. In the Lys200 ring, however, we observe a clear salt bridge, at least in two out of the three monomers of the crystal asymmetric unit, between the lysine and one of the flanking aspartate residues, while the other is at a distance of 6.3 Å. This is because the lysine side-chains are not oriented radially towards the centre of the channel. In contrast, in the Lys209-ring they point to the centre of the channel and are equidistant, 5 to 6 Å away, from the two flanking aspartate residues, without establishing any interaction with them.

Discussion

Requirements for DNA packaging

The packaging of DNA into preformed proheads is a mechanism exhibited by complex double-stranded DNA bacteriophages, as well as other animal viruses (adenovirus, herpesvirus).[25,26] The packaging process[27] has two important aspects: First, there is a processive movement of the DNA through the packaging machinery.[7] The other characteristic feature of the viral DNA packaging process is the condensation of the DNA inside the viral head to crystal-like concentrations. The translocation of a long, rather stiff DNA polymer into a preformed container imposes severe limitations on the function of the packaging machinery and the movement of the DNA, as the acquisition of the solenoid-like ordered parallel arrangement of the DNA inside the viral prohead generates pressure and a high concentration of mass and charges. Thus, the viral DNA packaging machinery has to deal with the general problems of translocation of nucleic acids by protein pumps, and at the same time has to find suitable solutions to release the tensions accumulated by the forced concentration of a highly charged polymer in a limited container.

The connector shows a unique architecture

It was reasonable to believe that the movement of the DNA would share some common mechanistic features with other nucleic acid processive systems,[1] such as the sliding clamps,[28] certain helicases[29,30] or conjugation systems.[31] However, the ϕ29 connector analysed here has a completely novel architecture, which suggests differences in the way the system functions. For example, ring helicases and the conjugation protein TrwB have a basic RecA/F1-ATPase architecture, including a core formed by a parallel doubly wound β-sheet with helices on both sides that encompasses a nucleotide binding domain. These proteins may contain additional domains, but neither these nor

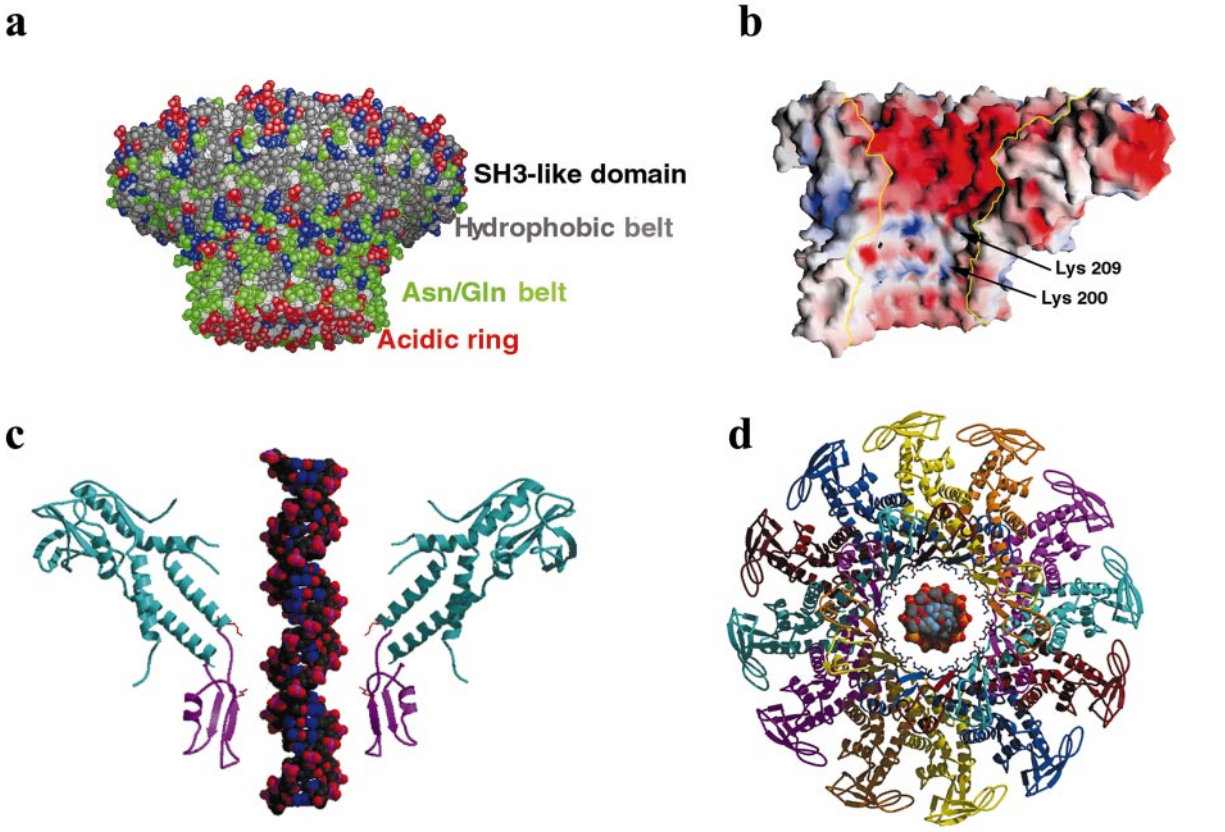

Figure 5. (a) Atom-sphere model of the p10 connector particle. Colour coding of the side-chains: green for Asn/Gln, red for Asp/Glu, grey for hydrophobic residues, and dark blue for Lys/Arg. The model was generated with GRASP.[63] (b) Molecular surface coloured according to electrostatic potential (red for negative, blue for positive) belonging to six vicinal protein monomers generated with GRASP.[63] The position of the two interior lysine rings (made up by Lys209 and Lys200 of each monomer, respectively) is indicated. (c) Ribbon plot showing a longitudinal cut of the connector particle. Note that the colours of the distal domains are different from the central and SH3-like domains since they belong to different monomers. A ds*B*-DNA molecule is shown traversing the central channel. The side-chains of Lys200 and Lys209 are indicated. (d) View along the connector axis with the modelled DNA. The side-chains of Lys200 and Lys209 are displayed as a balls and sticks.

the central nucleotide binding domain share any resemblance with the connector p10 protein monomer. Regarding the size of the whole particle, the connector is also unique, larger than all the others. Ring helicases are normally hexamers, while all the functional connectors are formed by 12 subunits.[7] The shape of the connector is unique too, with its distal narrow cylinder and a wide proximal disc of a diameter twice as large as the first, quite different from the more ellipsoidal or cylindrical shape of helicases and TrwB. A notable functional difference with ring helicases is that the connectors do not contain ATPase activity, and thus need ancillary ATP hydrolases to function, which is similar to what happens in sliding clamps. The protomer of a sliding clamp has a completely different fold from a connector protomer, but their oligomeric structure shows striking coincidences with the ϕ29 connector particle: the central channel is similarly defined by 12 α-helices that are inclined with respect to the particle axis, giving a propeller-like appearance.[28] The inclination is such that the helices would lie perpendicular to the sugar-phosphate backbone of a putative DNA traversing the channel. Kong *et al.*[28] suggest that this feature prevents the entry of the protein into the DNA groove, facilitating its rapid motion. A further coincidence between sliding clamps and the connector is the diameter of the channel, which is obviously determined by the diameter of the DNA double helix.

B-DNA fits smoothly into the channel

The connector defines a central channel of appropriate size to accommodate a double-stranded DNA molecule. The diameter of the channel is 35 Å at its narrowest part, while the DNA *B*-form has an average diameter of 23 Å. This leaves about 6 Å between the DNA and the channel wall. This wall is mostly electronegative (Figure 5(b)), as is the tip of the connector where the DNA enters the channel. The electronegative nature of the channel surface is not incompatible

with its function since it does not have to bind tightly to DNA, but rather forces it to pass smoothly through at 140 bp s^{-1}.[32,33] The DNA would traverse the channel floating away from the repulsive walls. There is room for a layer of water molecules, between the DNA and the protein. It is notable that the recently solved TrwB protein, involved in single-stranded DNA translocation, also displays a rather electronegative channel.[31]

The exceptions to the extended electronegative nature of the channel surface are the two lysine rings (200 and 209) (and putatively two more at Lys234 and Lys235). When fitting the DNA into the channel, if a particular backbone phosphate is aligned with one of the Lys200, another lysine of the same ring would be aligned with another phosphate from the other DNA strand (Figure 5(c) and (d)). This second lysine is not opposite the first one, across the channel, but that belonging to the fifth monomer. Keeping these two residues of ring 200 aligned with two phosphates, we again observe that at the second lysine ring (209) two lysine side-chains, five monomers apart, lie in front of two phosphate groups. This is because the 20 Å that separates the two lysine rings is about the same distance that separates two phosphates of the DNA across the major groove, in the helix axis direction. In consequence, at any particular moment during its translocation, the DNA molecule seems to be able to interact electrostatically at four points of the channel (and presumably two more at the next disordered lysine ring).

Sliding groove and SH3-like ridge

The X-ray structure of the ϕ29 connector fits very well in the three-dimensional reconstructions from cryo-electron microscopy of both proheads[34] (Figure 3(c)) and purified connectors[5] (not shown). Although the docking does not reveal details due to the limited resolution and the averaging imposed in the microscopy reconstruction, it is clear that the wider part of the connector, comprising some elements of the central domain and the whole SH3-like domain, is buried inside the cavity of the head, while the α-helical bundle of the central domain is inserted into the five-folded prohead vertex. In particular, the hydrophobic belt at the external concave area or groove of the connector is facing the corresponding areas of the head protein building the vertex.

As indicated in Results, the SH3-like domain does not have the canonical RT-Src loop for polyproline recognition. Its topology, however, is similar to SH3 domains, with two sets of three-stranded β-sheets arranged at 90 degrees to each other, forming a barrel, which encloses a tight hydrophobic core. A comparison with other structures with the same topology, but not strictly SH3, reveals that it closely resembles the DNA-binding domain of the HIV1 integrase.[35] This protein, also a viral protein like ϕ29 p10, forms dimers using one of the β-sheets as a dimerization surface. In the ϕ29 connector the SH3-like domains from the various monomers also contact each other but in a different way. They assemble in a circling head-to-tail fashion, using the protruding loops β4-β5 and β6-β7, which fit into the groove defined by β5-β6 and β1-α2 of the next monomer (Figures 4(a) and 5(c)). Thus, the distinct arrangement of the loops in the SH3-like domain facilitates protein-protein contacts, which finally hold together the protruding lobules of the wide proximal end. The SH3-like domain is therefore a sort of protein recognition motif, since it serves to assemble the compact protuberances that form the outer crown of the connector.

Spinning mechanism

Hendrix[36,37] suggested a way to correlate translocation of DNA and its packaging into the viral prohead based on the symmetry mismatch between the 5-fold vertex of the prohead and the 6*n*-folded symmetry of the tail components. The hypothesis that the rotation of the connector is actively involved in the actual DNA translocation was strongly supported by the description of 12-fold symmetry in several head-to-tail connectors either extracted from viral particles or assembled from overexpressed protein.[14,15,38–40] A longitudinal movement of the DNA, coupled to the connector rotation, has been proposed for both SPP1[16] and ϕ29[21] phage connectors.

The surface features as defined now in the high-resolution connector structure are consistent with the spinning hypothesis. As previously discussed,[8] the rotation of the connector with respect to the non-matching 5-fold symmetry head vertex could take advantage of the hydrophobic groove used as a sliding furrow. The adjacent crown of SH3-like domain would retain the connector inside the capsid while spinning.

Is the rotation of the connector enough to drive the DNA into the head? The finding of the two (or three, if we also assume the disordered loop to interact) lysine rings at regular 20 Å spacings in an otherwise electronegatively coated channel, together with the helical arrangement of the negatively charged DNA phosphate groups, strongly suggests that both movements are coupled. A one base-pair longitudinal displacement of the DNA would result in the weakening of four (or six) lysine-phosphate interactions. The restoring of the four interactions, but this time involving four new lysine residues of adjacent monomers, would occur only if the connector rotated. In other words, the energy barrier for jumping from a particular set of lysine-phosphate interactions to the next set would be overcome by the rotary force of the connector. For this model to work, the rotation of the connector has to be clockwise when looking from the wide domain, i.e. in the opposite direction of the DNA displacement. This is determined by the right-handed nature of the double helix, as the right-handed groove of a bolt determines the clock-

wise rotation of the nut in order to drive the bolt in the direction of the phage head.

A characteristic feature of this model is that the rotation of the connector has to match the symmetry of the DNA rather than the symmetry of the connector itself. In that sense, a rotation of 6° would be appropriate. Such a rotation would misalign the position of the lysine residues with respect to their phosphate counterparts but place the next monomer lysine residues at exactly the right place for interacting with the next set of phosphates, providing a DNA translocation of one base-pair into the prohead (Figure 6).

A spinning model has the beauty of simplicity. It does not need molecular rearrangements within the connector and thus the speed does not depend on time-limited inter-molecular movements. Other more complicated models, like the one put forward by Simpson *et al.*,[21] assume, besides the rotation, a domain displacement within the p10 protomers and a spring-like elongation and contraction of the connector length, which would require the relative slipping of the central α-helices between monomers. The ϕ29 connector quaternary structure analysed here appears to be very compact, with a high percentage of protomer surface involved in inter-subunit contacts. This arrangement gives little ground for extensive structural rearrangements, leaving only as a possible hinge candidate the region that connects the central helix bundle domain with the narrow β-structured RNA-binding tip (Figure 3(b)). However, even here, we do not observe a flexible connecting region, i.e. there are no Gly residues. A superposition of the three monomers of the asymmetric unit does not reveal any substantial differences. Although we cannot rule out a movement between domains or between subunits, we do not find any indication for such a movement in the solved high-resolution structure.

Localization and possible role of the pRNA

The packaging reaction at the ϕ29 system requires a competent prohead, which contains, beside the capsid, the pRNA and the ATPase p16.[4] These are indispensable components of the packaging machinery since they act, in our model, as a stator and ancillary motor, respectively. The pRNA required for the packaging of ϕ29 DNA has been located in the narrow tip of the connector.[21,34] The docking of the connector structure in the portal region reconstructed from proheads containing pRNA clearly shows the pRNA moiety forming a protruding ring that embraces the distal domain of interlaced β-strands at the apical region of the p10 dodecamer (Figure not shown). A peculiar aspect of this region is the presence of a ring built by helices α4 that presents a high content of exposed glutamine and asparagine residues. This unusual arrangement offers an uncharged polar surface rich in hydrogen bond donors and acceptors that could play a role in the interaction with the pRNA.

The pRNA is a unique component of the ϕ29 packaging machinery that interacts with the outer end of the connector in a multimeric form.[21,34] Extensive analysis of this pRNA based on functional assays of directed mutant pRNA molecules indicated the importance of the interaction between complementary loops in adjacent pRNA molecules (inter-subunit complementation.[41] The fact that two or three different serially complementary mutant pRNAs are required for restoring packaging activity in DNA packaging assays *in vitro* led to the conclusion that the pRNA should be arranged in a multimer containing 2*n* and 3*n* subunits. This conclusion is supported by the finding that a set of six serially complementary mutants are fully active in packaging, strongly suggesting the hexameric arrangement of the pRNA as the functional oligomer that is assembled in proheads competent for DNA packaging.[41,42] A different pRNA arrangement has recently been proposed,[21] where a pentamer has been found associated with the presence of pRNA in proheads reconstructed from electron microscopy.

Assuming the rotation of the connector, there are two possible scenarios for the pRNA. In one of them, it rotates with the connector. In the other, it is fixed to the head and does not follow the connector movement. The latter is an attractive hypothesis since the pRNA could act as a stator fixed to the head vertex. A 5-fold symmetry would suit this case better. The alternative scenario is also acceptable: the pRNA would be fixed to the connector, constituting a washer to hold it in place, and would be the equivalent piece of the SH3-like crown, but on the outside of the capsid wall. A 6-fold symmetry would suit this role better, to be compatible with the 12-fold symmetry of the connector. In this scenario, the head vertex would be the sole stator.

Localization and role of ATPase p16

The other basic component of the packaging reaction is protein p16.[43] In our model, the energy necessary for the connector spinning is provided by this ancillary 39 kDa ATP hydrolase. This protein interacts transiently with the prohead during packaging, but it is not present in the final viral particle.[44] The pRNA is required to couple the ATPase activity to packaging.

There are no direct data on the location of p16 in the packaging machinery, although Simpson *et al.*[21] provide evidence for its presence in the pRNA domain of proheads competent for packaging. We have found that p16 cosediments with proheads in gradients (B.I., J.M.V. and J.L.C., unpublished results). Our results also indicate that p16 interacts with the connector. The interaction with the isolated connector is dependent on the presence of the amino-terminal region, as connectors lacking the 13 N-terminal amino acid residues cannot interact with p16. Taken together, these results support the hypothesis that p16 contacts the narrow part of the

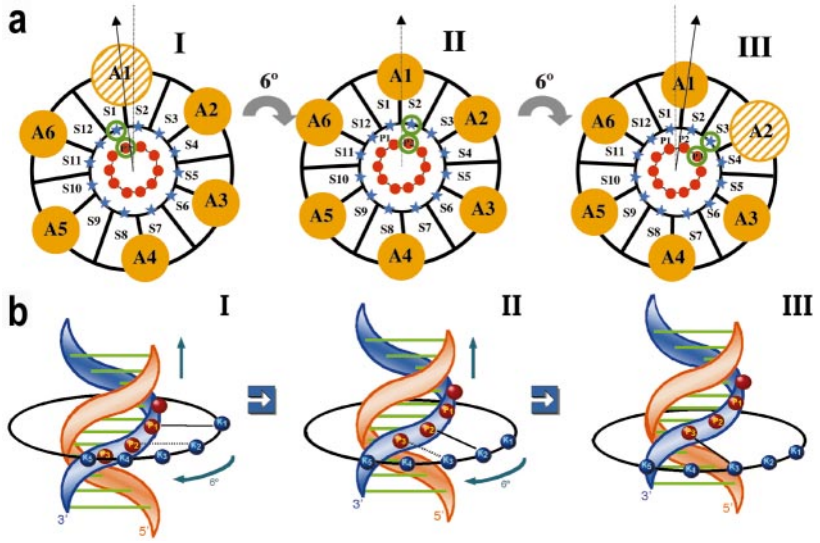

Figure 6. The spinning connector model. (a) Schematic arrangement of a portal vertex in projection view from the viral prohead. A1 to A6 represent the ATPase (p16), S1 to S12 represent the connector subunits, and P1, 2, 3, etc. represent the DNA phosphates of a DNA chain. For clarity, the model is restricted to the description of one single DNA chain and one lysine ring. The starting position (I) aligns a Lys (blue star) of subunit S1 with the phosphate P1. Firing of the p16 ATPase (A1) rotates the connector 12° clockwise while the DNA moves linearly along the connector axis by 2 bp. A transient interaction among the corresponding Lys from S2 and the phosphates P2 (II) occurs at 6° rotation. The final stage (III) is characterized by the interaction of Lys S3 and the phosphate P3, which is accomplished after a further clockwise rotation of the connector by 6°. This movement positions the next p16 ATPase (A2) in an equivalent position with respect to the DNA, to that of A1 in step I. (b) Schematic representation of a side view of the model, corresponding to the same steps as in (a). The lysine residues are depicted as blue spheres labelled K1, K2,.... In step I there is a strong interaction between lysine K1 and phosphate P1 (continuous line), while the interaction of lysine K2 with phosphate P2 is still weak because P2 is on a lower plane than the ring of lysine residues (broken line). In step II the 6° rotation of the connector places lysine K2 in an optimal position to interact with phosphate P2, provided there is a longitudinal displacement of the DNA by 1 bp. In step III there has been a further 6° rotation of the connector and another 1-bp displacement of the DNA, implying that lysine K3 is now aligned with phosphate P3. The movement of two base-pairs of DNA from I to III is, thus, correlated with a rotation of the connector by 12°.

connector, which remains accessible from outside the prohead. As the pRNA is arranged in a ring around the outer end of the dodecameric connector, this region comprises a complex set of concentric structures with different symmetries: the DNA inside the channel (10_1-fold), the connector walls (12-fold), and the outer ring of pRNA coupled to p16 (most probably, 6-fold). At the same time, the other side of the connector (12-fold) is embedded in the 5-fold vertex of the prohead.

At present there are insufficient data to confirm a direct interaction of p16 with the DNA. Therefore, we cannot discern whether the ATPase stroke pushes the DNA directly while the connector rotates passively, or whether the stroke only drives the connector rotation. In any case, as explained above, the connector would have to rotate at least 6° until the next base-pair establishes an electrostatic interaction with the following connector subunit. If the pRNA-p16 complex were a hexamer, the connector would have to rotate twice 6° (i.e. 12°) to place the p16/pRNA into an equivalent position with respect to the DNA. This would correlate the three main factors in the model: the linear translocation of two DNA base-pairs, a clockwise rotation of 12° of the connector, and the activation of one p16 ATPase. Each packaging step would then involve two base-pairs of DNA, two p10 subunits and one ATPase/pRNA subunit. Every ATPase would then be sequentially fired, followed by the clockwise rotation of the connector. According to our model, the activation of one ATPase would induce the translocation of two base-pairs into the prohead, the rotation of the connector by 12°, and the placing of the next ATPase in an equivalent position with respect to the DNA. Even if the connector rotates passively to the DNA translocation, its rotation would control the sequential firing of the ATPase.

The hydrolysis of one ATP molecule to package two base-pairs fits with the experimental data obtained for packaging ϕ29 DNA and other viruses *in vitro*.[41,45,46] In fact, this ATP consumption is slightly lower than that associated with the walking of helicases along the DNA (one to two ATP molecules per base-pair); but in the latter case these proteins separate DNA strands besides moving it. However, despite the structural and mechanical differences mentioned above, the translocation of DNA by connectors and helicases or sliding clamps could have in common the hydrolysis of ATP as a power stroke to processively move the DNA in a confined channel created by the toroid topology of these protein assemblies.

Other alternatives for the packaging of DNA into the viral heads could involve the rotation of the DNA and the connector (reviewed in reference[36]). Yet, any model involving rotation of the DNA would be penalized by the additional tension produced during its packaging into the prohead container. The sum of the strain imposed

by the coiling of the DNA within the head, following the inner wall of the prohead, together with the twisting forces induced by the rotation of DNA during packaging would lead to intolerable strain in the DNA inserted into the viral head (supercoiling and partial strand separation) and, eventually, to the arrest of the packaging process.

Further studies on the structure of the complexes of the various components of the packaging machinery and the connector will be required to provide deeper insight into the molecular events involved in the processive translocation of DNA, but the present high-resolution structure already suggests a simple and elegant mechanism that can now be validated with other complementary techniques.

Materials and Methods

A detailed description of the protein expression, proteolized connector particle preparation, crystallization and preliminary crystallographic analysis has been published elsewhere.[17,20] A new crystal form was obtained using the vapour diffusion method, at 20 °C, from sitting drops containing 70 % MPD (2-methyl-2,4-pentanediol), 0.1 M Tris-HCl buffer (pH 8.0). Typically, crystals appear after two days and reach a maximum size of approx. 0.5 mm × 0.1 mm × 0.1 mm. These crystals were picked up with a cryo-loop and directly flash-cryocooled by immersion in liquid propane. X-ray diffraction data were collected using synchrotron radiation at the ESRF ID14-EH2 beam-line in Grenoble, France. The crystals diffract to 1.9 Å resolution, although the very high-resolution shell decayed rapidly after collecting one frame. Complete data were collected up to 2.1 Å resolution (Table 1). Crystals are tetragonal, with space group *I*422, and have unit cell dimensions of $a = b = 155.3$ Å and $c = 160.4$ Å. The calculated V_m value[47] is 2.8 Å/Da, assuming one quarter of a connector particle per asymmetric unit. Data were evaluated with MOSFLM[48] and SCALA, of the CCP4 suite.[49] All attempts to obtain suitable heavy-atom derivatives were unsuccessful. Crystals obtained from Se-Met preparations did not diffract appropriately either.

A self-rotation function, calculated with GRLF[50] showed a 12-fold axis parallel to the crystal *c*-axis. This and packing considerations suggested that the particle was located so that its centre of mass was approximately at $x = 0$, $y = 0$, $z = 0.25$. A starting model for phasing was obtained from the electron microscopy reconstructions.[3] The model was positioned as mentioned above and adjusted by two-dimensional correlation fitting, using the *hk*0 reflections between 40 and 12 Å (our programs). A phase extension protocol was carried out using DM,[49] as follows. The phase extension started from a map calculated at 8 Å and the increment steps were of 0.1 Å. The DM-generated values for the figure-of-merit (FOM) were not used. Instead, FOM-values were set, at each extension cycle, to a lower value in the high-resolution range, to avoid an overestimation of the likelihood. The typical values for the manually set FOM ranged between 0.6 and 0.9 for stabilized resolution shells and 0.1 to 0.3 for the new shell increments. The mask used for the phase extension was made initially from a 30° cylindrical sector of the EM model. This mask was manually edited (our programs) after three or four cycles of phase extension according to the maps. The extension was continued in this way until 2.3 Å. Although the global shape of the molecule connector was identifiable in the resulting maps, no clear secondary structure elements could be located. This map, however, was used to build a pseudo-atom model with MAPMAN,[51] filling the asymmetric unit. This model was then divided in three 30° cylindrical sectors, which were then superimposed, resulting in an averaged pseudo-atom model (our programs) that was used both to generate a new mask and to further improve the phases with ARP/WARP,[52] using the whole resolution of the data to 2.1 Å. Each cycle consisted of an unrestrained water search and refinement using REFMAC. After each ARP/WARP cycle the model was reconstructed from an averaged pseudo-atom model as before.

As in this protocol the FOM-values and the crystallographic *R*-factors are unreliable, checks were made between cycles by trying different combinations of FOM-values, looking for features in maps, and mainly checking the correlation between the cylindrical sectors of the map. This protocol was successful, and after a few runs zones of well connected electron density appeared, with several alpha helices identified.

At this stage, the coordinates of the Simpson *et al.* model (PDB code 1fou) became available in the PDB. They were used as an initial model for molecular replacement with AMoRe.[53] No solution, however, could be obtained using a monomer as a starting model. A

Table 1. Data collection and refinement statistics

Data collection	
Wavelength (Å)	0.932
Temperature (K)	100
Resolution range (Å)	42-2.1
No. of observed reflections	327,940
No. of unique reflections	56,472
Completeness (%)	98.9 (94.1)
R_{merge} [a]	0.10 (0.40)
$\langle I \rangle / \langle \sigma I \rangle$	5.4 (2.1)
Multiplicity	5.8 (2.4)
Refinement	
Resolution range used (Å)	42.0-2.1
No. of reflections in working set	53,611
No. of reflections in test set	2861
No. of protein atoms in the asymmetric unit (a.u.)	6517
No. of solvent molecules in the a.u.	523
No. of MPD molecules in the a.u.	4
R [b], R_{free} [c]	0.205, 0.247
R.m.s deviations from target values	
Bond lengths	0.006 Å
Bond angles	1.13 degrees
Average B *factors (Å²)*	
Protein atoms	35.4
Water molecules	51.9
All atoms	37.0
Wilson plot	29.3

The outermost resolution shell (2.21-2.10 Å) values are shown in parentheses.

[a] $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where I_i is the *i*th measurement of reflection *hkl* and their symmetry related reflections.

[b] $R = \Sigma_{hkl}||F_o(hkl)| - |F_c(hkl)||/\Sigma_{hkl}|F_o(hkl)|$, where $F_o(hkl)$ and $F_c(hkl)$ are the observed and calculated structure factors, respectively.

[c] R_{free} = *R*-value calculated for 5 % of reflections not used for refinement.

different strategy was them adopted: a whole connector model, with 12 monomers, was fitted into the density of the phase extension/WARP map. A one-dimensional translation search was then carried out along the *c*-axis, followed by rigid-body refinement. This gave a solution with an *R*-factor of 44.6 and a correlation of 54.6 (next peak 37.0 and 50.6, respectively). SigmaA-weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps calculated with this solution showed that it was correct, although in many areas, and in particular in the proximal disc-like part, the density did not coincide with the model. These areas were manually rebuilt with TURBO,[54] including the tracing of five new β-strands and the correct placing of the amino acid sequence, which in the search model was out of register from residue 271 onwards. The structure was refined with CNS[55] and further manual interventions. Initially, 3-fold NCS restraints were applied. These were relaxed and finally suppressed, once the *R*-free fell below 26% at full resolution. At this stage, water molecules were added to the model. Atoms not seen in the $2F_o - F_c$ and $F_o - F_c$ sigmaA-weighted maps were set to zero occupancy. This mainly affected the N and C-terminal residues, and some residues before β10. The sequence of 15 amino acid residues between residues 230 and 245 of each monomer was not seen on the maps and was not included, as explained in Results. The final model is of excellent geometry as confirmed by PROCHECK.[56] The final refinement statistics are summarized in Table 1.

Atomic coordinates

The atomic coordinates have been deposited with the PDB under accession number 1h5w.

Acknowledgements

This study was supported by grants from the Ministerio de Educación y Cultura of Spain (PB98-1631 and 2FD97-0518 to M.C., PB97-1225-C02 to J.M.V., and PB96-0818 to J.L.C.) and the Generalitat de Catalunya (1999SGR-188 and Centre de Referència en Biotecnologia) to M.C. Synchrotron data collection was supported by the EU grants ERBFMGECT980134 and HPRI-CT-1999-00017 to the EMBL-DESY and by the ESRF. J.P. acknowledges a fellowship from the Ministerio de Educación y Cultura. We are indebted to Rosa Pérez-Luque for assistance in crystallization experiments and Ignacio Fita and Jaime Martín-Benito for useful suggestions.

References

1. Hingorani, M. M. & O'Donnell, M. (1998). Toroidal proteins: running rings around DNA. *Curr. Biol.* **8**, R83-R86.
2. Casjens, S. & Hendrix, R. W. (1988). Control mechanisms in dsDNA bacteriophage assembly. In *The Bacteriophages* (Calendar, R., ed.), vol. 1, pp. 15-90, Plenum Press, New York.
3. Valpuesta, J. M. & Carrascosa, J. L. (1994). Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. *Quart. Rev. Biophys.* **27**, 107-155.
4. Guo, P. X., Erickson, S. & Anderson, D. (1987). A small viral RNA is required for *in vitro* packaging of bacteriophage phi 29 DNA. *Science,* **236**, 609-614.
5. Valpuesta, J. M., Fernandez, J. J., Carazo, J. M. & Carrascosa, J. L. (1999). The three-dimensional structure of a DNA translocating machine at 10 Å resolution. *Struct. Fold. Des.* **7**, 289-296.
6. Tao, Y., Olson, N. H., Xu, W., Anderson, D. L., Rossmann, M. G. & Baker, T. S. (1998). Assembly of a tailed bacterial virus and its genome release studies in three dimensions. *Cell,* **95**, 431-437.
7. Carrascosa, J. L. & Valpuesta, J. M. (1999). Bacteriophage connectors: structural features of a DNA translocating motor. *Recent Res. Dev. Virol.* **1**, 449-465.
8. Valpuesta, J. M., Sousa, N., Barthelemy, I., Fernandez, J., Fujisawa, H., Ibarra, B. & Carrascosa, J. L. (2000). Structural analysis of the bacteriophage T3 head-to-tail connector. *J. Struct. Biol.* **131**, 146-155.
9. Orlova, E. V., Duve, P., Beckmann, E., Zemlin, F., Lurz, R., Trautner, T. A. *et al.* (1999). Structure of the 13-fold symmetric portal protein of bacteriophage SPP1. *Nature Struct. Biol.* **6**, 842-846.
10. Kochan, J., Carrascosa, J. L. & Murialdo, H. (1984). Bacteriophage lambda preconnectors. Purification and structure. *J. Mol. Biol.* **174**, 433-447.
11. Bazinet, C., Benbasat, J., King, J., Carazo, J. M. & Carrascosa, J. L. (1988). Purification and organization of the gene 1 portal protein required for phage P22 DNA. *Biochemistry,* **27**, 1849-1856.
12. Driedonks, R. A., Engel, A., ten Heggeler, B. & van Driel, R. (1981). Gene 20 product of bacteriophage T4. Its purification and structure. *J. Mol. Biol.* **152**, 641-662.
13. Carrascosa, J. L., Carazo, J. M., Ibañez, C. & Santisteban, A. (1985). Structure of phage phi 29 connector protein assembled *in vivo*. *Virology,* **141**, 190-200.
14. Carrascosa, J. L., Viñuela, E., Garcia, N. & Santisteban, A. (1982). Structure of the head-tail connector of bacteriophage phi 29. *J. Mol. Biol.* **154**, 311-324.
15. Carazo, J. M., Donate, L. E., Herranz, L., Secilla, J. P. & Carrascosa, J. L. (1986). Three-dimensional reconstruction of the connector of bacteriophage phi 29 at 1.8 nm resolution. *J. Mol. Biol.* **192**, 853-867.
16. Dube, P., Tavares, P., Lurz, R. & van Heel, M. (1993). The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO J.* **12**, 1303-1309.
17. Guasch, A., Parraga, A., Pous, J., Valpuesta, J. M., Carrascosa, J. L. & Coll, M. (1998). Purification, crystallization and preliminary X-ray diffraction studies of the bacteriophage phi29 connector particle. *FEBS Letters,* **430**, 283-287.
18. Valpuesta, J. M., Carrascosa, J. L. & Henderson, R. (1994). Analysis of electron microscope images and electron diffraction patterns of thin crystals of Φ29 connectors in ice. *J. Mol. Biol.* **240**, 281-287.
19. Muller, D. J., Engel, A., Carrascosa, J. L. & Velez, M. (1997). The bacteriophage phi29 head-tail connector imaged at high resolution with the atomic force microscope in buffer solution. *EMBO J.* **16**, 2547-2553.
20. Guasch, A., Pous, J., Parraga, A., Valpuesta, J. M., Carrascosa, J. L. & Coll, M. (1998). Crystallographic analysis reveals the 12-fold symmetry of the bacteriophage phi29 connector particle. *J. Mol. Biol.* **281**, 219-225.
21. Simpson, A. A., Tao, Y., Leiman, P. G., Badasso, M. O., He, Y., Jardine, P. J. *et al.* (2000). Structure of

the bacteriophage phi29 DNA packaging motor. *Nature,* **408**, 745-750.
22. Donate, L. E., Valpuesta, J. M., Rocher, A., Méndez, E., Rojo, F., Salas, M. & Carrascosa, J. L. (1992). Role of the amino-terminal domain of bacteriophage ϕ-29 connector in DNA binding and packaging. *J. Biol. Chem.* **267**, 10919-10924.
23. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.
24. Urbaneja, M. A., Rivas, S., Carrascosa, J. L. & Valpuesta, J. M. (1994). An intrinsic-tryptophan-fluorescence study of phage phi 29 connector/nucleic acid interactions. *Eur. J. Biochem.* **225**, 747-753.
25. Casjens, S., van Vugt, R., Tilly, K., Rosa, P. A. & Stevenson, B. (1997). Homology throughout the multiple 32-kilobase circular plasmids present in Lyme disease spirochetes. *J. Bacteriol.* **179**, 217-227.
26. Wikoff, W. R. & Johnson, J. E. (1999). Virus assembly: imaging a molecular machine. *Curr. Biol.* **9**, 296-300.
27. Black, L. W. (1989). DNA packaging in dsDNA bacteriophages. *Annu. Rev. Microbiol.* **43**, 267-292.
28. Kong, X. P., Onrust, R., O'Donnell, M. & Kuriyan, J. (1992). Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell,* **69**, 425-437.
29. Sawaya, M. R., Guo, S., Tabor, S., Richardson, C. C. & Ellenberger, T. (1999). Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. *Cell,* **99**, 167-177.
30. Fass, D., Bogden, C. E. & Berger, J. M. (1999). Crystal structure of the N-terminal domain of the DnaB hexameric helicase. *Struct. Fold. Des.* **7**, 691-698.
31. Gomis-Ruth, F. X., Moncalian, G., Perez-Luque, R., Gonzalez, A., Cabezon, E., de la Cruz, F. & Coll, M. (2001). The bacterial conjugation protein TrwB resembles ring helicases and F1-ATPase. *Nature,* **409**, 637-641.
32. Shibata, H., Fujisawa, H. & Minagawa, T. (1987). Characterization of the bacteriophage T3 DNA packaging reaction *in vitro* in a defined system. *J. Mol. Biol,* **196**, 845-851.
33. Catalano, C. E., Cue, D. & Feiss, M. (1995). Virus DNA packaging: the strategy used by phage lambda. *Mol. Microbiol.* **16**, 1075-1086.
34. Ibarra, B., Caston, J. R., Llorca, O., Valle, M., Valpuesta, J. & Carrascosa, J. L. (2000). Topology of the components of the DNA packaging machinery in the phage phi29 prohead. *J. Mol. Biol.* **298**, 807-815.
35. Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R. & Davies, D. R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science,* **266**, 1981-1986.
36. Hendrix, R. W. (1998). Bacteriophage DNA packaging: RNA gears in a DNA transport machine. *Cell,* **94**, 147-150.
37. Hendrix, R. W. (1978). Symmetry mismatch and DNA packaging in large DNA bacteriophages. *Proc. Natl Acad. Sci. USA,* **75**, 4779-4783.
38. Carazo, J. M., Santisteban, A. & Carrascosa, J. L. (1985). Three-dimensional reconstruction of bacteriophage phi 29 neck particles at 2×2 nm resolution. *J. Mol. Biol.* **183**, 79-88.
39. Donate, L. E., Herranz, L., Secilla, J. P., Carazo, J. M., Fujisawa, H. & Carrascosa, J. L. (1988). Bacteriophage T3 connector: three-dimensional structure and comparison with other viral head-tail connecting regions. *J. Mol. Biol.* **201**, 91-100.
40. Valpuesta, J. M., Serrano, M., Donate, L. E., Herranz, J. & Carrascosa, J. L. (1992). DNA conformational change induced by the bacteriophage phi 29 connector. *Nucl. Acids Res.* **20**, 5549-5554.
41. Guo, P., Zhang, C., Chen, C., Garver, K. & Trottier, C. (1998). Inter-RNA interaction of phage ϕ29 pRNA to form a hexameric complex for virial DNA transportation. *Mol. Cell,* **2**, 149-155.
42. Zhang, F., Lemieux, S., Wu, X., St.-Arnaud, D., McMurray, C. T. & Anderson, D. (1998). Function of hexameric RNA in packaging of bacteriophage f-29 DNA *in vitro*. *Mol. Cell,* **2**, 141-147.
43. Bjornsti, M. A., Reilly, B. E. & Anderson, D. (1982). Morphogenesis of bacteriophage phi29 of *Bacillus subtilis*: oriented and quantized in vitro packaging of DNA protein gp3. *J. Virol.* **45**, 383-396.
44. Viñuela, E., Camacho, A., Jimenez, F., Carrascosa, J. L., Ramirez, G. & Salas, M. (1976). Structure and assembly of phage phi29. *Phil. Trans. Roy. Soc. Lond. B, Biol. Sci.* **276**, 29-35.
45. Guo, P., Peterson, C. & Anderson, D. (1987). Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage phi 29. *J. Mol. Biol.* **197**, 229-236.
46. Morita, M., Tasaka, M. & Fujisawa, H. (1993). DNA packaging ATPase of bacteriophage T3. *Virology,* **193**, 748-752.
47. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.
48. Leslie, A. G. W. (1991). Macromolecular data processing. In *Crystallographic Computing* (Moras, D., Podjarny, A. D. & Thierry, J. C., eds), vol. 5, pp. 27-38, Oxford University Press, Oxford.
49. CCP4. (1994). The CCP4 suite:programs for protein crystallography. *Acta Crystallog. sect. D,* **50**, 760-763.
50. Tong, L. (1993). REPLACE, a suite of computer programs for molecular-replacement calculations. *J. Appl. Crystallog.* **26**, 748-751.
51. Kleywegt, G. J. & Jones, T. A. (1999). Software for handling macromolecular envelopes. *Acta Crystallog. sect. D,* **55**, 941-944.
52. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458-463.
53. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A,* **50**, 157-163.
54. Roussel, A. & Cambilleau, C. (1989). Turbo-Frodo. In *Silicon Graphics Geometry Partners Directory*, pp. 77-79, Silicon Graphics, Mountain View, CA.
55. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al*. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905-921.
56. Laskowsky, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of a protein structure. *J. Appl. Crystallog.* **24**, 946-950.
57. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* **22**, 2577-2637.

58. Evans, S. V. (1993). SETOR: hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.* **11**, 134-138.
59. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
60. Esnouf, R. M. (1997). An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 133-138.
61. Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallog. sect. D,* **50**, 869-873.
62. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD-visual molecular dynamics. *J. Mol. Graph.* **14.1**, 33-38.
63. Nicholls, A., Bharadwaj, R. & Honig, B. (1993). GRASP: graphical representation and analysis of surface properties. *Biophys. J.* **64**, A166-.

Edited by R. Huber

(Received 17 September 2001; received in revised form 13 November 2001; accepted 16 November 2001)