

Bacteriophage P22 tailspike: structure of the complete protein and function of the interdomain linker

Anaït Seul,^{a,‡} Jürgen J. Müller,^{b,‡}
Dorothee Andres,^{a,§} Eva
Stettner,^a Udo Heinemann^{b,c,*}
and Robert Seckler^{a,*}

^aPhysikalische Biochemie, Universität Potsdam, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam, Germany, ^bMax-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Strasse 10, 13125 Berlin, Germany, and ^cInstitut für Chemie und Biochemie, Freie Universität, Takustrasse 6, 14195 Berlin, Germany

‡ These authors contributed equally to this work.

§ Present address: Harvard University, Department of Chemistry and Chemical Biology, 12 Oxford Street, Cambridge, MA 02138, USA

Correspondence e-mail:
heinemann@mdc-berlin.de,
seckler@uni-potsdam.de

Attachment of phages to host cells, followed by phage DNA ejection, represents the first stage of viral infection of bacteria. *Salmonella* phage P22 has been extensively studied, serving as an experimental model for bacterial infection by phages. P22 engages bacteria by binding to the sugar moiety of lipopolysaccharides using the viral tailspike protein for attachment. While the structures of the N-terminal particle-binding domain and the major receptor-binding domain of the tailspike have been analyzed individually, the three-dimensional organization of the intact protein, including the highly conserved linker region between the two domains, remained unknown. A single amino-acid exchange in the linker sequence made it possible to crystallize the full-length protein. Two crystal structures of the linker region are presented: one attached to the N-terminal domain and the other present within the complete tailspike protein. Both retain their biological function, but the mutated full-length tailspike displays a retarded folding pathway. Fitting of the full-length tailspike into a published cryo-electron microscopy map of the P22 virion requires an elastic distortion of the crystal structure. The conservation of the linker suggests a role in signal transmission from the distal tip of the molecule to the phage head, eventually leading to DNA ejection.

Received 10 January 2014

Accepted 5 February 2014

PDB references: pre-PBDzip, 2vky; PBDzip-Y108W, 2vnl; TSP-Y108W, 2xc1

1. Introduction

One of the key processes in cell infection by a virus is the delivery of the viral genome into the host cell. Most bacterial viruses (bacteriophages) possess a capsid (or phage head) enclosing a double-stranded DNA genome and a tail structure (Ackermann, 2009). The tail mediates binding to the host cell during infection and delivers the viral DNA across the bacterial envelope into the host cell. This process of DNA ejection by tailed bacteriophages has been studied *in vitro* (Molineux, 2006). In such experiments, DNA ejection is triggered specifically upon interaction of phage tail proteins with bacterial cell-surface components. Despite recent advances in the elucidation of virus and tail structures by cryo-electron microscopy (cryo-EM) and X-ray crystallography, molecular details of the triggering mechanisms are not well understood (Casjens & Molineux, 2012; Davidson *et al.*, 2012; Leiman & Shneider, 2012).

We have recently succeeded in triggering DNA ejection *in vitro* from purified particles of phage P22 (Andres *et al.*, 2010). *Salmonella* phage P22 is a structurally well characterized virus

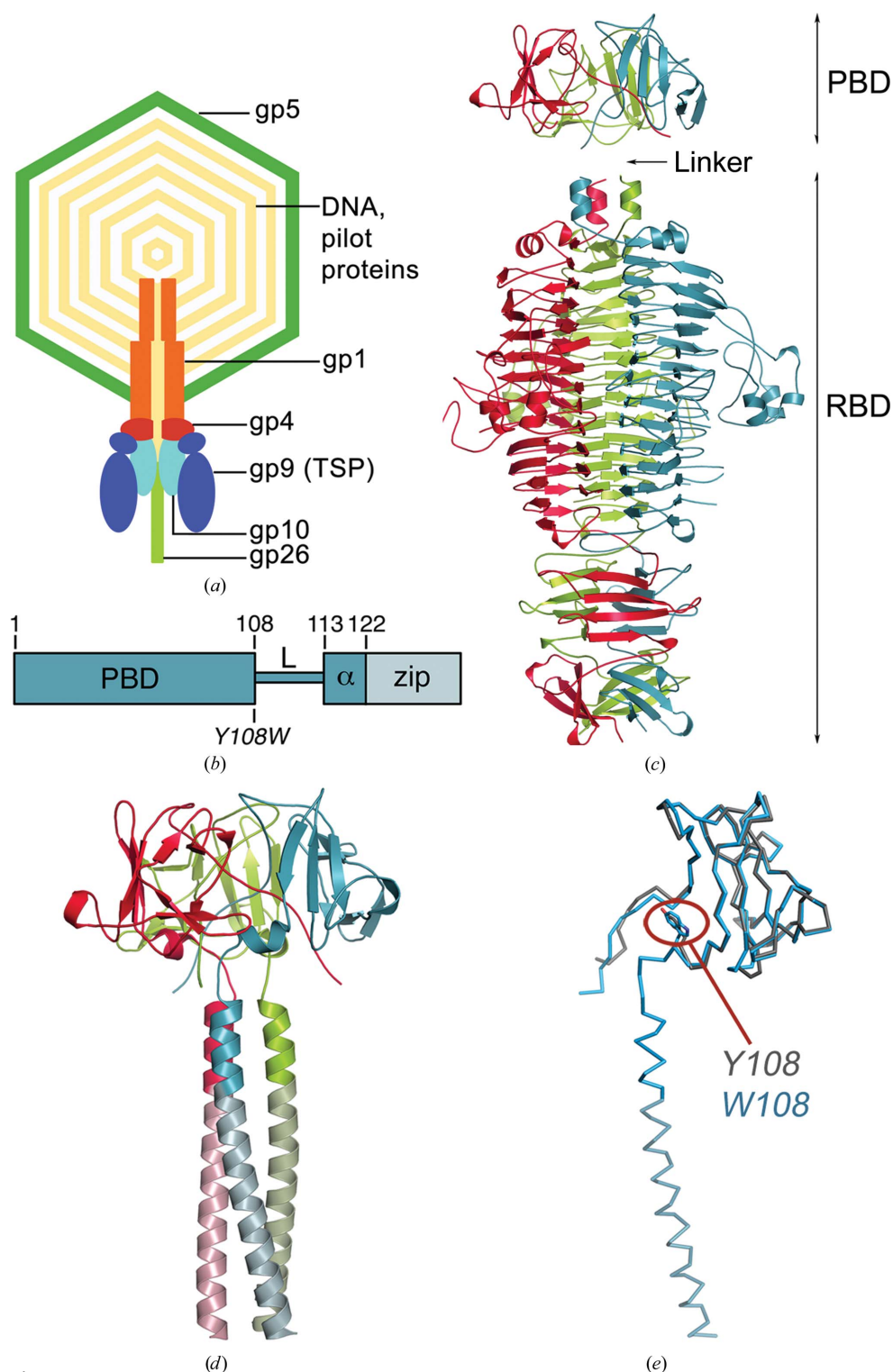


Figure 1

PBD constructs with isoleucine zippers. (a) Schematic cross-section of a P22 virion drawn according to the cryo-EM structure presented in Tang *et al.* (2011). The proteins (gps) forming the capsid and tail are labelled. (b) Schematic representation of the zipper constructs PBDzip and PBDzip-Y108W. L, linker region; α , α -helix at the upper part of the RBD; zip, isoleucine zipper sequence. The position of the Y108W mutation is marked. (c) Ribbon tracing of the structures of the PBD (PDB entry 1lkt) and RBD (PDB entry 1tyu). The three subunits are depicted in red, blue and green. An arrow points to the position of the missing linker connecting the two domains. (d) Ribbon tracing of the structure of PBDzip-Y108W. Subunits are coloured as in (c). The portion of the isoleucine zipper is shown in lighter colours. (e) Superposition of the core C $^{\alpha}$ atoms (residues 14–105) of PBDzip-Y108W (blue) with the PBD (PDB entry 1lkt; grey). The root-mean-square deviation of corresponding C $^{\alpha}$ positions is 0.65 Å. Only one subunit is shown, in the same orientation as in (d).

and is morphologically classified as a podovirus, a DNA bacteriophage with a short non-contractile tail (Fig. 1a). Its tail structure, attached to a dodecameric portal ring formed by gp1, is assembled from only four gene products (gps): gp4 and gp10 form a double layer of a sixfold-symmetric tube-like structure into which the needle-shaped gp26 trimer is inserted, serving as a plug to the central opening (Chang *et al.*, 2006; Lander *et al.*, 2006). Six trimeric gp9 molecules, called tailspikes, are attached to the tail hub in the last step of phage particle assembly (King *et al.*, 1973) and mediate binding to the viral receptor: the repetitive O-antigen carbohydrate structure of *Salmonella* lipopolysaccharide (Baxa *et al.*, 1996).

The tailspikes of P22 and its podoviral relatives display modular structures. The smaller N-terminal particle-binding domain (PBD) of the homotrimeric tailspike protein (TSP) binds the elongated molecule to the phage head. This segment of around 110 amino-acid residues is highly conserved in sequence between different phages of the same morphology (Casjens & Thuman-Commike, 2011). X-ray crystallography of the P22 PBD revealed a dome-like trimer in which each subunit consists of two central β -sheets with five and three strands, respectively, connected by extensive loops (Steinbacher *et al.*, 1997). The much larger C-terminal part of TSP mediates binding to the receptor (receptor-binding domain; RBD). It carries an endoglycosidase activity that degrades the O-antigen polysaccharide, thus allowing access of the virus particle to the outer membrane surface prior to DNA ejection (Andres *et al.*, 2010). The RBD from phage P22 and its morphological relatives Sf6, a *Shigella* phage, and HK620, an *Escherichia coli* phage, fold in the

absence of the PBDs and assemble to form highly thermostable, SDS-resistant and enzymatically active trimers. Their crystal structures (Steinbacher *et al.*, 1997; Barbirz *et al.*, 2008; Müller *et al.*, 2008) revealed that the overall dimensions and general architecture of the RBD are well conserved despite their divergence in sequence (Casjens & Thuman-Commike, 2011). The major structural feature of each subunit is a right-handed parallel β -helix of 13–14 rungs. In the trimeric protein, the β -helices are associated side by side such that their helical axes are oriented almost parallel to the threefold axis.

The two domains of the tailspike, the PBD and RBD, are connected by a short linker peptide that is thought to be very flexible because no electron density could be observed in this region for any of the crystallized fragments (Steinbacher *et al.*, 1997; Barbirz *et al.*, 2008; Müller *et al.*, 2008). No crystal structure of a complete TSP containing both domains could be obtained to date. Moreover, if the crystal structures of the tailspike domains are fitted separately into the 7.8 Å resolution cryo-EM map of the P22 viral particle (Tang *et al.*, 2011), their threefold rotational axes are tilted at the linker by about 13° relative to each other. Despite the obvious conformational flexibility, the sequence of the peptide segment is conserved between podoviruses. In the conserved sequence Tyr-Asp-Pro-Asp-Gln-Phe/Tyr, the two tyrosine residues, Tyr108 and Tyr113 in P22, are the last residues in the linker region for which electron density can be observed in the X-ray structures of the PBD and RBD, respectively (numbering for the mature protein lacking the N-terminal Met; Steinbacher *et al.*, 1994, 1997). A flexible linker in the viral attachment protein facilitates the engagement of several receptors at once: an important process during infection, as at least three tailspike molecules are needed to infect the host cell (Israel, 1978). In addition, such a conserved, flexible linker could play an important role during infection, either in signal transmission or in the conformational changes necessary for DNA ejection (Steinbacher *et al.*, 1997). Towards the C-terminus, the linker region is followed by a parallel α -helical bundle (residues 114–120 in P22 TSP; Steinbacher *et al.*, 1997). Very similar three-helix bundles have been observed at the N-terminal ends of the PBDs of Sf6 TSP (residues 113–121; Müller *et al.*, 2008) and HK620 (residues 112–120; Barbirz *et al.*, 2008) and in Det7 TSP, a myoviral P22 TSP homologue in which the RBD is fused to a different particle-binding N-terminal domain (Walter *et al.*, 2008).

To obtain structural information on the linker peptide, we created a new fusion-protein construct which contains the PBD, the linker peptide and the natural three-helix bundle elongated by an isoleucine zipper (Weissenhorn *et al.*, 1997). A single-residue mutation in the linker peptide allowed crystallization of both the new fusion construct and, for the first time, the full-length P22 TSP, revealing the flexible nature of the linker. Combined with biochemical data, our results suggest that a conformational change in the PBD upon the encounter of the TSP distal tips with the outer membrane surface may play a role in triggering ejection and that the linker peptide may be involved in transmission of the signal to the phage head *via* the PBD.

2. Materials and methods

2.1. Cloning and protein purification

For cloning of the first isoleucine zipper construct, pre-PBDzip, the sequence of the PBD (residues 1–123) was amplified from pTSP8 (Miller *et al.*, 1998) with the primers PBD-fw and PBD-rv (Supplementary Table S1¹). The isoleucine zipper sequence pIIGCN4 (Weissenhorn *et al.*, 1997) was provided by Winfried Weissenhorn and was amplified using the primers Zip-fw and Zip-rv (Supplementary Table S1). Both were cloned into separate pBC KS+ plasmids cut with *EcoRV* for amplification and successively transferred to pET-17b using *NdeI* and *HindIII* (PBD) or *BclI* and *NotI* (pIIGCN4), yielding pre-PBDzip. As this construct contained a helical mismatch between the isoleucine zipper and the short coiled-coil sequence C-terminal to the PBD, a second construct, PBDzip, was made from pre-PBDzip by consecutive steps of site-directed deletion and mutagenesis (primers Del01, Del02 and Del03; Supplementary Table S1). PBDzip-Y108W and TSP-Y108W were produced by site-directed mutagenesis (QuikChange, Stratagene) from PBDzip and pTSP1 (Miller *et al.*, 1998), respectively, using primer Y108W-fw (Supplementary Table S1) and its reverse complement.

Purification of wild-type TSP, TSP-Y108W and TSP-D395N (Baxa *et al.*, 1996) was carried out as described for the wild-type protein (Miller *et al.*, 1998). The fusion proteins pre-PBDzip, PBDzip and PBDzip-Y108W were expressed and purified using a modified protocol: expression was carried out in *E. coli* BL21 (DE3) cells at 30°C for 5–6 h. Impurities were precipitated from the cleared lysate using 10.25% ammonium sulfate. The protein was precipitated from the supernatant with 60% ammonium sulfate and dialyzed against buffer (40 mM Tris-HCl, 2 mM EDTA pH 8), followed by hydrophobic interaction chromatography using a Phenyl Sepharose column. Further impurities were removed by anion-exchange chromatography (Q Sepharose).

2.2. Crystallization, X-ray structure determination and refinement

Crystals were grown at 19°C using the hanging-drop vapour-diffusion method. Pre-PBDzip and PBDzip-Y108W were concentrated to 8.2 mg ml⁻¹ in 50 mM HEPES pH 6.5 and crystallized in 20% 2-propanol, 0.1 M sodium acetate pH 4.6, 0.2 M CaCl₂ (4 μ l drops over 2 ml reservoir). Crystals of about 0.4 mm edge length appeared within 5 d and were flash-cooled in liquid nitrogen after adding 30%(v/v) glycerol to the drop as a cryoprotectant. TSP-Y108W crystallized under several conditions within 1–3 d. For the final condition, the protein concentration was 15 mg ml⁻¹ in 10 mM HEPES pH 7 and the reservoir consisted of 0.2 M ammonium acetate, 0.1 M trisodium citrate dihydrate pH 5.6, 30% polyethylene glycol 4000 (3 μ l drops over 750 μ l reservoir). Crystals were flash-cooled in the crystallization solution containing 3% glycerol.

¹ Supporting information has been deposited in the IUCr electronic archive (Reference: MV5103).

Diffraction data from pre-PBDzip crystals were collected using a Rigaku RU H2B rotating-anode generator at a wavelength of 1.542 Å under cryo-cooling (110 K) using a MAR Research MAR345 image-plate detector. Data for PBDzip-Y108W and TSP-Y108W were collected at wavelengths of 0.9537 and 0.9184 Å, respectively, under cryo-cooling (100 K) using a MAR Research MarMosaic CCD detector on the Protein Structure Factory beamline BL14.1 of Freie Universität Berlin at BESSY, Berlin, Germany (Heinemann *et al.*, 2003). Pre-PBDzip and PBDzip-Y108W form crystallographic trimers (space groups $P6_3$ and $H3$, respectively), whereas TSP-Y108W crystallizes with one trimer in the asymmetric unit (space group $P2_12_12_1$). All data sets were processed with *XDS* (Kabsch, 2010). The statistics of data collection and processing are summarized in Table 1.

The structures of pre-PBDzip and PBDzip-Y108W were determined using molecular replacement with the coordinates of the PBD (PDB entry 1lkt; Steinbacher *et al.*, 1997) and the zipper domain (PDB entry 1ebo; Weissenhorn *et al.*, 1998) using the programs *MOLREP* (Vagin & Teplyakov, 2010) and *Phaser* v.2.1 (McCoy *et al.*, 2007). Refinement of the coordinates was carried out using *REFMAC5* (Murshudov *et al.*, 2011) and interactive modelling in *O* (Jones *et al.*, 1991). Owing to the structural flexibility of PBDzip-Y108W, the protein chain in the linker region Asp112–Ile116 could only be traced after using the *Bias Removal Server* (Reddy *et al.*, 2003; Supplementary Fig. S1). The structure of TSP-Y108W was determined using the coordinates for the trimer of a P22 RBD structure containing residues 112–666 (PDB entry 2vfm; M. Becker, J. J. Mueller, T. Weikl, U. Heinemann & R. Seckler, unpublished work) in *AMoRe* (Navaza, 1994). *Phaser* v.2.1 was used to position the PBD using the model 1lkt truncated to residues Ala6–Lys108. The linker region was constructed with *ARP/wARP* (Lamzin & Wilson, 1993) and *O* and refined with *REFMAC5*. The refinement statistics are summarized in Table 2. Superposition of structures was performed using *LSQ* in *Coot* (Emsley & Cowtan, 2004) or *LSQKAB* in *CCP4* (Kabsch, 1976). Structural figures were generated with *PyMOL* (Schrödinger) or *UCSF Chimera* (Pettersen *et al.*, 2004).

Table 1
Data collection and processing.

Values in parentheses are for the outer shell.

	Pre-PBDzip (PDB entry 2vky)	PBDzip-Y108W (PDB entry 2vnl)	TSP-Y108W (PDB entry 2xc1)
Diffraction source	Rigaku RU H2B rotating anode	Beamline 14.1, BESSY	Beamline 14.1, BESSY
Wavelength (Å)	1.542	0.9537	0.9184
Temperature (K)	110	100	100
Detector	MAR Research image plate	MAR Research CCD	MAR Research CCD
Space group	$P6_3$	$H3$	$P2_12_12_1$
Unit-cell parameters (Å)	$a = b = 57.81, c = 107.12$	$a = b = 58.01, c = 156.11$	$a = 86.36, b = 121.56, c = 208.26$
Mosaicity (°)	0.21	0.22	0.23
Resolution range (Å)	19.70–2.05 (2.20–2.05)	50.00–1.80 (1.93–1.80)	48.30–1.65 (1.85–1.65)
No. of unique reflections	12716 (1748)	17904 (3300)	237282 (51669)
Completeness (%)	99.6 (99.5)	98.5 (95.5)	90.4 (68.3)
Multiplicity ($\langle I/\sigma(I) \rangle$)	4.2 (4.2)	7.3 (7.3)	6.3 (2.8)
$R_{r.i.m.}^\dagger$	0.034 (0.115)	0.054 (0.409)	0.109 (0.561)
Overall B factor from Wilson plot (Å ²)	30.9	33.3	20.0‡

† Estimated $R_{r.i.m.} = R_{merge} [N/(N-1)]^{1/2}$, where N is the data multiplicity. ‡ The side-by-side packing of strands in the β -sheets gives rise to a 4.6 Å maximum.

Table 2
Structure solution and refinement.

Values in parentheses are for the outer shell.

	Pre-PBDzip (PDB entry 2vky)	PBDzip-Y108W (PDB entry 2vnl)	TSP-Y108W (PDB entry 2xc1)
Resolution range (Å)	20.00–2.05 (2.16–2.05)	19.37–1.80 (1.86–1.80)	49.45–1.65 (1.693–1.650)
Completeness (%)	99.8	99.7	91.7
No. of reflections, working set	12111 (1748)	17174 (1666)	228771 (9322)
No. of reflections, test set	622 (93)	934 (94)	12041 (491)
Final R_{cryst}	0.145 (0.134)	0.190 (0.231)	0.165 (0.325)
Final R_{free}	0.187 (0.199)	0.229 (0.256)	0.210 (0.349)
No. of non-H atoms			
Protein	1197	1188	15254
Ion	0	0	6
Ligand	0	18	125
Water	173	131	2265
Total	1370	1337	17650
R.m.s. deviations			
Bonds (Å)	0.014	0.013	0.012
Angles (°)	1.450	1.287	1.366
Average B factors (Å ²)			
Protein	36.3	40.5	17.7
Ion	—	—	40.2
Ligand	—	50.9	36.0
Water	49.8	47.0	21.8
Ramachandran plot†			
Most favoured (%)	100	98.6	97.3
Allowed (%)	0	1.4	2.4

† Calculated with *RAMPAGE* (Lovell *et al.*, 2003). Six Ramachandran outliers, Ile504 and Ser513 in all chains, are found in PDB entry 2xc1. They are localized at loop ends and are caused by crystal-packing effects.

2.3. Tailing competition assays and tailing kinetics

Preparation of the tailless phage heads and tailing assays were performed as described by Haase-Pettingell & King (1997) with modifications (Andres *et al.*, 2010). For tailing competition assays, proteins were diluted to 25 µg ml⁻¹ in tailing buffer (4 mM MgCl₂, 50 mM Tris pH 7.6), and 0.4 µg wild-type TSP was mixed with PBDzip or TSP-D395N at the respective molar ratios. 1.28 × 10⁸ phage heads were added

and the volume was adjusted to 250 μl with dilution fluid (1 g peptone, 14 g NaCl, 2 mM MgSO_4 per litre). The reactions were incubated at 30°C for 1 h, serially diluted in 1:10 steps, and 100 μl of each dilution was plated on DB7155 bacterial lawns (Winston *et al.*, 1979). The PFU ml^{-1} were calculated after overnight incubation at 30°C. Tailing kinetics were performed at 4°C. Phage heads ($1 \times 10^{11} \text{ ml}^{-1}$) and tailspike solutions (7 $\mu\text{g ml}^{-1}$) were mixed at a 195:1 ratio of tailspike molecules:phage heads in tailing buffer. At defined time points, aliquots of the reaction were drawn and serial dilutions were plated as described above.

2.4. Folding kinetics of TSP-Y108W

Assessment of folding kinetics was performed as described by Danner *et al.* (1993), with the following modifications. Natively folded proteins at a concentration of 1.7 mg ml^{-1} were diluted 1:10 in denaturing buffer (6 M urea, 50–65 mM sodium phosphate pH 3) and incubated at room temperature for 1.5 h. The samples were cooled to 10°C and rapidly diluted according to the protocol. The protein was kept at 10°C for 1.5 h to allow the formation of partially folded monomers. To start the trimerization process, protein solutions were transferred to 30°C and samples were drawn at defined time points, mixed with SDS–PAGE sample buffer and kept at 10°C. After the last time point, 20.4 ng of each sample was loaded onto SDS–PAGE gels, which were run, silver-stained and evaluated densitometrically using *Gelscan* 5.1 (BioSciTec GmbH, Frankfurt, Germany).

2.5. Accession codes

Coordinates and structure factors have been deposited in the Protein Data Bank as PDB entries 2vky (pre-PBDzip), 2vnl (PBDzip-Y108W) and 2xc1 (TSP-Y108W).

3. Results

3.1. The structure of the tailspike linker region within the fusion protein PBDzip-Y108W

The conserved flexible linker peptide between the two tailspike domains PBD and RBD might play an important role during infection. Owing to its flexibility, no structural information about it was available to date. The isolated PBD is mostly monomeric in solution (A. Seul, unpublished data), but crystallizes as a trimer (Steinbacher *et al.*, 1997). In order to stabilize the linker region and force trimerization of the PBD, we designed a construct which comprises the PBD, the linker region and the α -helical bundle at the start of the C-terminal domain elongated by the isoleucine zipper pIIGCN4 (Weissenhorn *et al.*, 1997; Figs. 1*b* and 1*c*). The first such construct, pre-PBDzip, crystallized, but contained a helical mismatch between the isoleucine zipper and the α -helical bundle, leading to structural distortion. We therefore constructed another protein, PBDzip. PBDzip did not yield diffraction-quality crystals. As the PBD does not contain any tryptophan residues, Tyr108 was mutated to tryptophan to perform fluorescence spectroscopy for protein stability measurements.

Tyr108 is located at the N-terminal end of the linker region and is the last residue visible in the electron density of the PBD (Steinbacher *et al.*, 1997). This mutated zipper construct, PBDzip-Y108W, crystallized readily, with crystals diffracting to 1.8 Å resolution.

The PBD monomer core regions (Pro14–Val105) and the α -helical bundle on top of the RBD are nearly identical to the structures of the isolated domains (Steinbacher *et al.*, 1997; Figs. 1*d* and 1*e*). The region around Trp108 forms a short α -helix that was not previously observed and closes off the bottom of the PBD. In the trimer, two of these helices from neighbouring subunits embed the N-terminus dangling from the third subunit. The N-terminus therefore has a slightly more elongated shape than in the isolated PBD structure. The cores of PBDzip-Y108W and isolated PBD can be superimposed with an r.m.s. C^α deviation of 0.65 Å. Even though the linker peptide seems to be stabilized by the tyrosine-to-tryptophan mutation, the electron density of the backbone in the linker region could only be traced after using the *Bias Removal Server* (Reddy *et al.*, 2003), indicating that the linker peptide still retains residual flexibility. In the structures of both the isolated PBD and RBD (Steinbacher *et al.*, 1997), the linker region may have been invisible in the corresponding electron densities owing to the fact that it is located at the end of each of the truncated domains or because the diffraction experiments were performed at room temperature. However, in the PBDzip-Y108W structure the linker is embedded between the PBD and the isoleucine zipper. Its obvious mobility therefore indicates a role as a flexible element between the two tailspike domains.

3.2. Binding of the zipper construct to phage heads

The structure of the zipper construct PBDzip-Y108W superimposes very well with that of the isolated wild-type PBD. However, it is possible that the PBDzip construct does

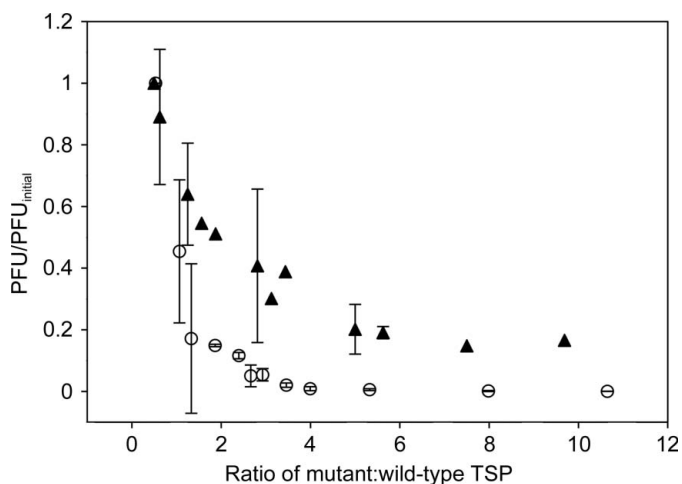


Figure 2 Tailing competition assays. Different ratios of TSP variants (PBDzip, empty circles, or TSP-D395N, filled triangles) to wild-type TSP were added to tailless phage heads. After incubation, the mixture was plated on bacterial lawns and the number of PFUs was determined. The experiment was repeated two times with different preparations of phage heads.

not retain the particle-binding function, as residues or signals from the RBD might be necessary for binding. We therefore performed tailing competition assays to establish whether the isoleucine zipper constructs are indeed able to bind to phage heads. *In vitro* tailing assays simulate the binding of TSP to phage heads, which is the final step of P22 phage assembly. In these assays, a tailless phage (phage head) is used. It is complemented by the addition of TSP and plated onto *Salmonella* host cells. Functional phages infect and lyse the bacteria and form plaques. For our tailing competition assays, different molar ratios of a mixture of full-length TSP and PBDzip were used. Functional PBDzip is expected to compete with TSP in binding to the phage head and thus to produce phages that are unable to infect the host cell because they lack the RBD, resulting in a reduced number of plaque-forming units (PFU) compared with TSP alone. As a control, we used a mixture of TSP and TSP-D395N, a mutant that binds phage heads but is deficient in hydrolysis of the receptor and thus

also unable to infect the host cells (Baxa *et al.*, 1996). The results of this experiment are shown in Fig. 2. Both PBDzip and TSP-D395N show a reduced number of PFUs as the ratio of mutant to wild-type TSP increases. The effect is slightly more pronounced for PBDzip, showing that the smaller zipper construct competes with TSP more efficiently than does the TSP-D395N mutant.

3.3. Structure of the full-length tailspike TSP-Y108W

Previous attempts to crystallize a full-length tailspike from any of the phages studied have not been successful, probably owing to the flexible linker between the PBD and RBD. The different crystallization behaviour of PBDzip and PBDzip-Y108W suggested that the Tyr-to-Trp mutation could render the linker region more rigid and thus favour crystallization of the full-length protein as well. We therefore mutated the full-length tailspike accordingly (construct TSP-Y108W) and well diffracting crystals were obtained. The structure was determined to 1.65 Å resolution (Fig. 3*a*). Both the PBD and RBD (Steinbacher *et al.*, 1997; M. Becker, J. J. Mueller, T. Weikl, U. Heinemann & R. Seckler, unpublished work; PDB entry 2vfp) display virtually the same structure as the isolated domains, but the strict crystallographic symmetry of the former structures is replaced by non-crystallographic symmetry in TSP-Y108W. The linker is formed as in the zipper construct PBDzip-Y108W. The linker peptide retains some flexibility, as reflected in the *B* factors of the atoms. The *B* factors, or displacement factors, reflect the uncertainties of atomic positions, and the *B* factors of the C α atoms therefore indicate the flexibility of a protein chain (Karplus & Schulz, 1985). The *B* factors of the C α atoms in the linker region are the highest in the whole protein (Fig. 3*b*) and thus represent the region with the greatest mobility.

The PBD is tilted by 5.5° towards the intersubunit cleft maintained within the RBD, probably owing to crystal contacts, leading to slightly different chain conformations and distances between the three monomers. This movement opens

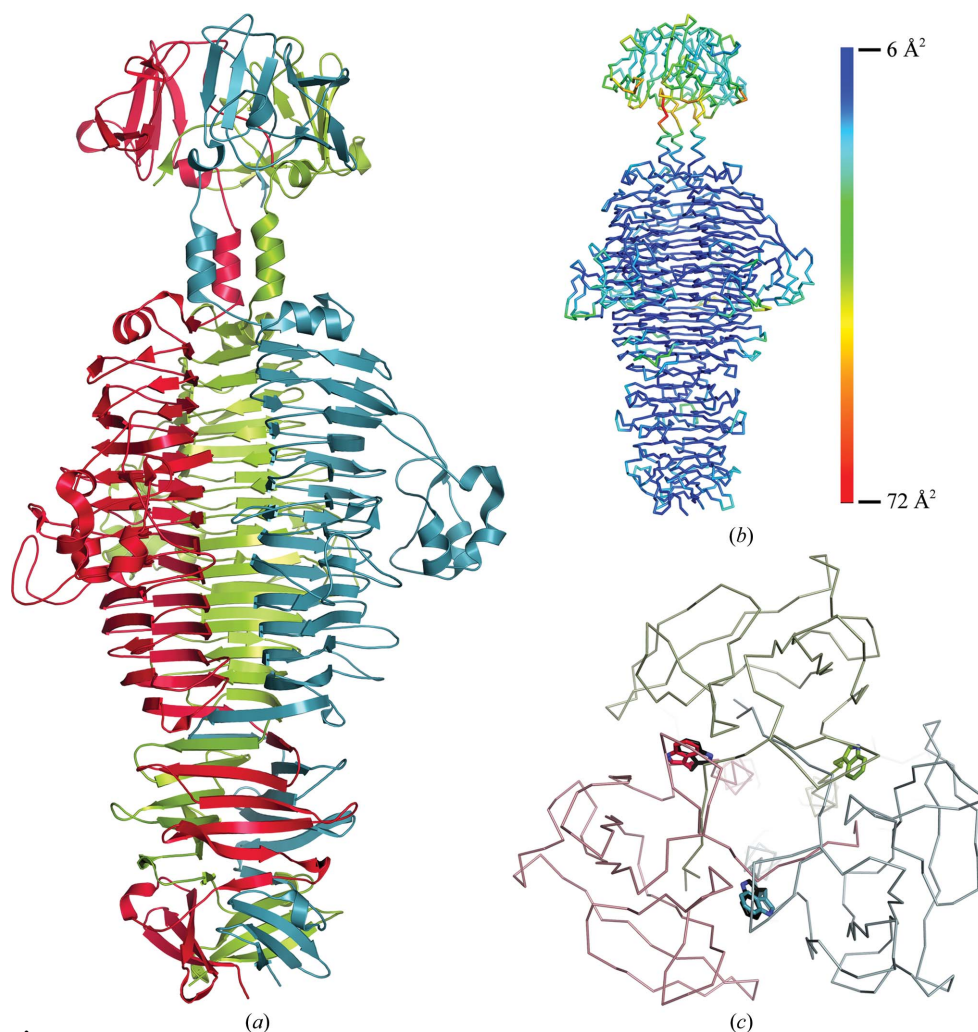


Figure 3

Structure of a full-length tailspike. (*a*) Ribbon tracing of the full-length TSP-Y108W structure. Subunits are shown in red, blue and green. (*b*) Backbone tracing of TSP-Y108W. The C α atoms are coloured according to *B* factors, ranging from dark blue (6 Å², lowest) to red (72 Å², highest). (*c*) Backbone tracing of TSP-Y108W viewed from the top along the trimer axis. Subunits are coloured light red, light blue and light green. The side chains of Trp108 are shown; in two subunits (red and blue) they adopt alternate conformations shown in black.

hydrophobic pockets between the C α atoms of Pro12 and Lys80 of chains *A* and *B*, which harbour the Trp108 residues of chains *C* and *A*, respectively. At the high resolution of 1.65 Å, Trp108 can be seen to adopt two alternative conformations with different occupancies in each monomer (Fig. 3c). The interior of the PBD is accessible to solvent and highly hydrated (Supplementary Fig. S2).

3.4. Implications of increased linker rigidity

Rigidification of the linker region owing to the Tyr-to-Trp mutation is the likely cause of the different crystallization behaviours of wild-type and Y108W TSP. Tyr108 is highly conserved between tailspikes of different phages, pointing

towards an essential role of this residue at this position (Supplementary Fig. S3). To assess the question whether the mutation has an effect on tailspike function *in vivo*, we performed kinetic tailing assays with full-length TSP and TSP-Y108W. No significant difference between wild-type and mutant could be observed in binding to phage heads in these assays (Supplementary Fig. S4), indicating that the Y108W mutation, and thus probably the flexibility of the linker, have only minor effects in the head-binding process.

The folding pathway of P22 TSP has been extensively studied *in vivo* (Goldenberg & King, 1982; Goldenberg *et al.*, 1982, 1983) and *in vitro* (Fuchs *et al.*, 1991; Danner & Seckler, 1993; Danner *et al.*, 1993; Miller *et al.*, 1998). After partial folding of the monomers, the protrimer is assembled, which then folds into the native SDS-resistant TSP trimer. This final step is rate-limiting and involves the complete folding of the PBD. Numerous mutants are known that affect the folding, but not the function, of TSP, *e.g.* the temperature-sensitive for folding (tsf) mutants (King *et al.*, 1996). As the replacement of the highly conserved Tyr108 by tryptophan still yields functional phages according to tailing assays, we therefore studied the folding and unfolding of TSP-Y108W in comparison with wild-type TSP *in vitro*. The mutation did not change the thermal denaturation kinetics in experiments performed according to Miller *et al.* (1998) (data not shown). However, the formation of native trimers from protrimers was strongly retarded in folding assays for the mutant protein. Wild-type TSP and Y108W were denatured by urea and renatured using rapid dilution. During renaturation, samples were drawn at defined time points and analyzed by SDS-PAGE (Fig. 4*a*). Monomer/protrimer and native trimer can be distinguished easily, as the native trimer is SDS-resistant at room temperature, while the protrimer is not and thus runs on the gel at the molecular weight of the monomer. For wild-type TSP, two different species can be observed on the SDS-PAGE gel (Fig. 4*a*, top): the band at lower molecular weight corresponds to the monomer and protrimer (M/PT) and the band at higher molecular weight, increasing over time, to the native trimer (T). A faint third band below the trimer band is a thermal unfolding artifact resulting from the native trimer, which is produced by insufficient cooling during SDS-PAGE and adsorption of protein to the eppendorf cup walls (Mitraki *et al.*, 1993). For TSP-Y108W, a third species (IT) between the monomer/protrimer and native trimer can be observed (Fig. 4*a*, bottom). It initially increases over time and then decreases again while more native trimer is formed, and thus seems to represent an intermediate between the protrimer and native trimer. Running at a molecular weight close to that of the native trimer, this intermediate species is partly SDS-resistant and is therefore most likely to be a trimeric intermediate. A similar intermediate is found during the unfolding of wild-type TSP, comprising an SDS-resistant C-terminal portion and an unfolded PBD (Chen & King, 1991).

To analyze the time course of native trimer formation more precisely, the gel bands were analyzed densitometrically (Fig. 4*b*). For wild-type TSP, the native trimer band appeared with a half-time of 7–9 min, as determined by previous studies

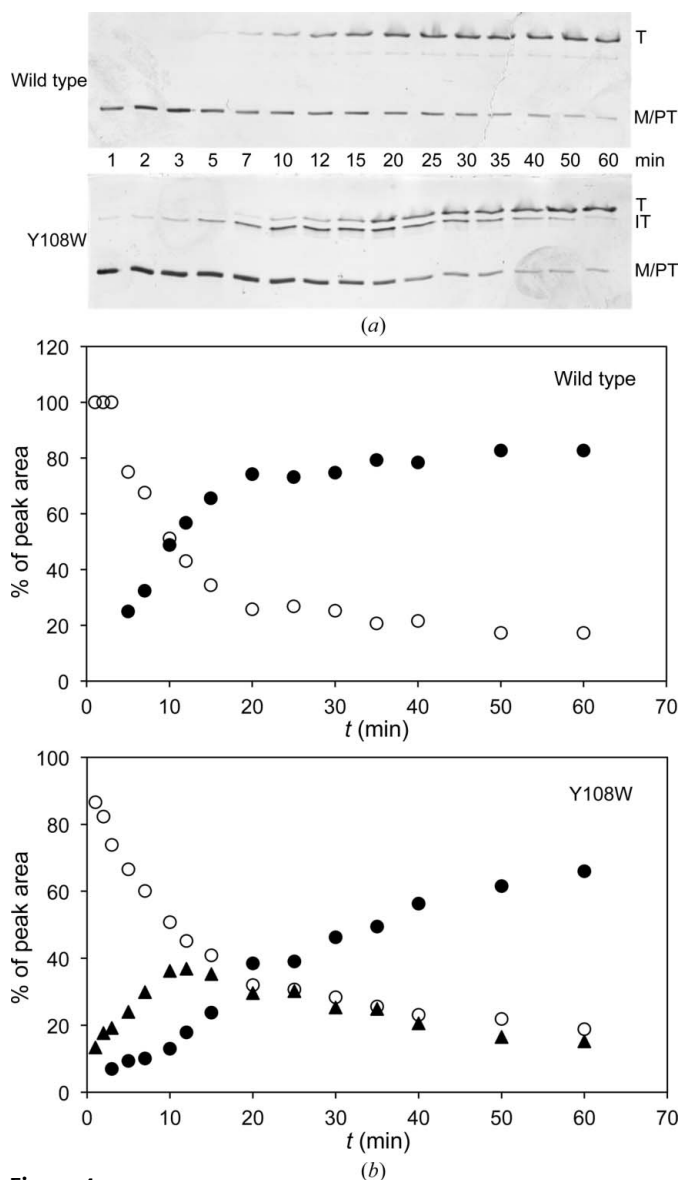
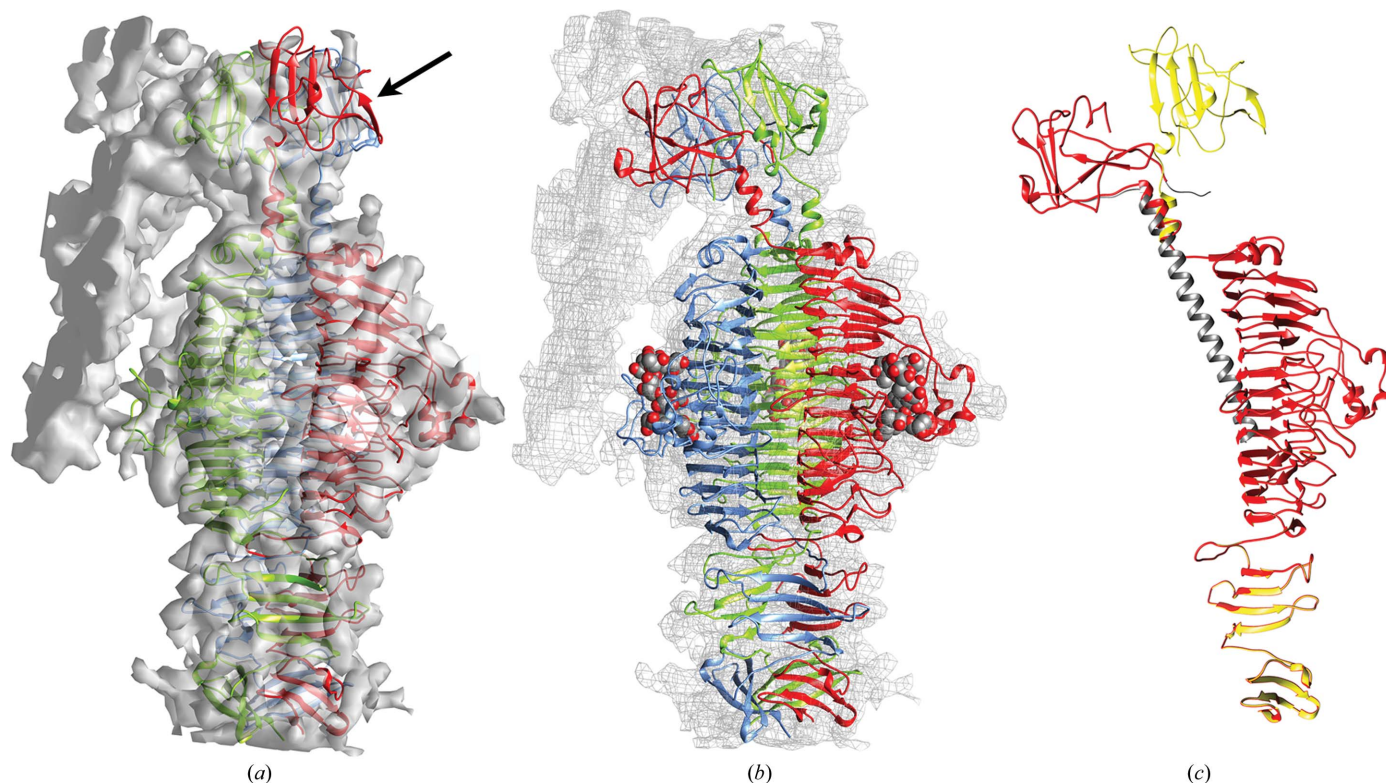


Figure 4 Folding assays of TSP and TSP-Y108W. (*a*) Trimer formation of wild-type TSP and TSP-Y108W. M/PT, monomer and protrimer; IT, trimeric intermediate; T, native trimer. (*b*) Time courses of trimer formation for wild-type TSP (top) and TSP-Y108W (bottom). Percentages of peak areas are plotted for native trimers (filled circles), trimeric intermediate (triangles) and monomers/protrimers (empty circles).

**Figure 5**

Fit of the full-length tailspike into the 7.8 Å resolution cryo-EM map of the P22 virion. (a) RBD portion of the TSP-Y108W trimer fitted into the map (Tang *et al.*, 2011) by *UCSF Chimera* (Pettersen *et al.*, 2004). In contrast to the good fit of the RBD, the N-terminal PBD cannot be fitted into the EM density without distortion (arrow). (b) Linker region and PBD distorted for best fit into the map. The coordinates for the O-antigens (viral receptors) are taken from PDB entry 1tyu (drawn as sphere models). (c) Chain A of the crystal structure of TSP-Y108W (yellow), of the model generated by elastic and inelastic deformation (red) and of the template model pre-PBDzip for the inelastic distortion (grey; Supplementary Fig. S5b).

(Fuchs *et al.*, 1991; Danner *et al.*, 1993). In contrast to this, formation of the native trimer of TSP-Y108W is delayed and slowed down significantly, while the putative trimeric intermediate accumulates at first and is then slowly depleted.

3.5. A distorted model of TSP-Y108W fits the cryo-EM map

A cryo-EM map of the P22 virion at 7.8 Å resolution has recently been published (Tang *et al.*, 2011). The TSP-Y108W crystal structure cannot be fitted into this cryo-EM map unless distortions of the TSP trimer are allowed. To assess possible kinking and rotational movements of the PBD relative to the RBD facilitated by the inherent flexibility of the domain junction and the TSP trimer, a normal-mode analysis of TSP-Y108W was carried out with *EINémo* (Suhre & Sanejouand, 2004; results not shown). It was realised that elastic deformation according to this analysis was insufficient to fit the full-length TSP into the EM map. Instead, combining two normal modes (kink and rotation) with a translation of the PBD was required for a satisfactory model fit to the EM map (Fig. 5). The translational component is associated with a local unwinding of the short α -helix Asn104–Trp108 in the innermost PBD. An unwinding of this kind was indeed experimentally observed in the crystal structure of the trimeric zipper construct pre-PBDzip (Supplementary Fig. S5). In this construct, an error in the helical repeat sequence between the

isoleucine zipper and the short α -helical segment C-terminal to the PBD leads to a helical mismatch (Supplementary Fig. S6) and thus to mechanical strain on the residues below and in the linker region.

4. Discussion

A homogeneous sample of high purity is a prerequisite for protein crystallization in most cases (McPherson, 2004). Multi-domain proteins containing linker regions can display a considerable degree of conformational flexibility. The resulting structural heterogeneity in the sample can inhibit crystal growth. Various methods have been developed to stabilize flexible proteins for crystallization, *e.g.* introducing binding partners such as antibodies (Koide, 2009) or small stabilizing molecules (Sousa, 1995) and mutation and deletion of protein regions (Price & Nagai, 1995). The result is a reduced conformational flexibility of the protein and thus a significantly more homogeneous sample. In this work, we have shown that a single mutation from tyrosine to tryptophan in a flexible linker region can stabilize a protein sufficiently to be crystallized. In spite of the stabilization that led to the determination of a high-quality three-dimensional structure, the linker region remains the most flexible region of the molecule. We have also shown that the mutant protein is still in a

functional conformation by binding and infection experiments.

The full-length mutant tailspike, TSP-Y108W, can bind to phage heads and is still able to produce infectious phages, but its folding process is significantly slowed down and shows a folding intermediate which is most likely to be trimeric and not observed in the wild-type protein. It is possible that this intermediate is also present during the folding of the wild-type protein, but exhibits such a short lifetime that this conformational state is very sparsely populated and is therefore not observable. As the folding of wild-type TSP follows a complex pathway and is relatively slow (Goldenberg & King, 1982; Goldenberg *et al.*, 1982, 1983; Fuchs *et al.*, 1991; Danner & Seckler, 1993; Danner *et al.*, 1993; Miller *et al.*, 1998), it is obvious that an even slower folding of the TSP would represent a serious functional disadvantage for P22. This may explain why the Y108W mutation has not been observed in nature to date. The tilt of the PBD towards the RBD has been observed in all cryo-EM reconstructions of P22 and its portal complex available to date (Tang *et al.*, 2005, 2011; Chang *et al.*, 2006; Lander *et al.*, 2006, 2009), and it is obvious that the linker between the two domains needs to be flexible in order for the protein to fit into the cryo-electron density maps. However, it is not clear why the linker sequence shows such a high degree of conservation. It has been suggested that for interdomain linkers flexibility alone is not the most important factor, but that linker sequences serve as transmitters of allosteric signals (Ma *et al.*, 2011). According to this theory, the linker sequence encodes successive conformational states that act in propagating a signalling event from one domain to the other. To date, there have been no conformational changes observed in TSP upon binding of the phage to the bacterial surface. However, such signalling processes have been shown to occur even without visible conformational changes in the protein backbone (Tsai *et al.*, 2008).

It is not yet clear how DNA release from the phage is triggered during infection by P22, but it has been suggested that TSP and the tail needle gp26 act together as surface-pressure sensors upon encountering the bacterial surface (Andres *et al.*, 2010). During this process, the tip of the tail needle is likely to control the kinetics of the DNA ejection process (Leavitt *et al.*, 2013). The transmission of the pressure sensor signal from the C-terminal tip of the elongated TSP to the head-binding N-terminal domain through the linker would explain its conserved sequence. Further analysis of the linker residues is necessary to investigate its role. If such signal transmission occurs, we speculate that Tyr108 plays a key role in it. For the mutated tryptophan in this position, alternate conformations were observed in two of the three subunits (see Fig. 3c), and are likely to reflect different strains exerted on the PBD owing to its tilt relative to the RBD. The smaller Tyr108 side chain in the wild-type TSP which superimposes very well with the mutant tryptophan (see Fig. 1e) could therefore adopt two different conformations as well, thus acting as a part of the intra-trimeric signalling cascade that transfers a signal leading to DNA release from the phage head.

An even stronger signal for triggering DNA release may be provided by the asymmetrical unwinding of the small α -helix Asn104–Tyr/Trp108 that is required for TSP binding to the phage base (see above). In this state, conformational energy is stored within the TSP, because Trp108 (Tyr108 in wild-type TSP) is forced out of the hydrophobic pocket formed by the neighbouring chain in which it is accommodated in the crystal structure. Release of this conformational energy may be triggered by contact with the bacterial surface. The observed conservation of the linker sequence in TSPs supports this model. In-depth analysis of the flexible region by site-directed mutagenesis will yield further information about the importance of the various linker residues and their influence on assembly kinetics and genome ejection.

We have shown that the N-terminal domain of TSP competes efficiently with the full-length protein in binding to phage heads. The isolated PBD even displayed a better binding compared with the full-length protein. From our assay, we cannot determine the reason for this behaviour, although it appears to be likely that the smaller PBD can bind more easily because of less steric hindrance than that experienced by the full-length protein. The tailing competition assay showed that the trimeric N-terminal domain of TSP is sufficient for binding to the phage head and that no residues or signals from the RBD are necessary for head engagement. These results add to our understanding of the architecture of the P22 portal and tail formed from the recent high-resolution cryo-EM reconstructions of P22 and its tail machine (Lander *et al.*, 2009; Tang *et al.*, 2011), the crystallographic structures of the P22 portal complex of gp1 and gp4 (Olia *et al.*, 2011), and the structure of the tail needle gp26 (Olia *et al.*, 2007).

5. Related literature

The following references are cited in the Supporting Information for this article: Larkin *et al.* (2007), Lakowski (1995) and Reddy *et al.* (2003).

The authors wish to thank Gabriel Müller for cloning of the zipper constructs, Mandy Schietke for performing the tailing assays, Yvette Roske for X-ray data collection on TSP-Y108W crystals, Winfried Weissenhorn for providing the construct of the isoleucine zipper pIIGCN4 and Bernhard Rupp for his support in using the *Bias Removal Server*.

References

- Ackermann, H. W. (2009). *Methods Mol. Biol.* **501**, 127–140.
- Andres, D., Hanke, C., Baxa, U., Seul, A., Barbirz, S. & Seckler, R. (2010). *J. Biol. Chem.* **285**, 36768–36775.
- Barbirz, S., Müller, J. J., Uetrecht, C., Clark, A. J., Heinemann, U. & Seckler, R. (2008). *Mol. Microbiol.* **69**, 303–316.
- Baxa, U., Steinbacher, S., Miller, S., Weintraub, A., Huber, R. & Seckler, R. (1996). *Biophys. J.* **71**, 2040–2048.
- Casjens, S. R. & Molineux, I. J. (2012). *Adv. Exp. Med. Biol.* **726**, 143–179.
- Casjens, S. R. & Thuman-Commike, P. A. (2011). *Virology*, **411**, 393–415.
- Chang, J., Weigele, P., King, J., Chiu, W. & Jiang, W. (2006). *Structure*, **14**, 1073–1082.

- Chen, B. & King, J. (1991). *Biochemistry*, **30**, 6260–6269.
- Danner, M., Fuchs, A., Miller, S. & Seckler, R. (1993). *Eur. J. Biochem.* **215**, 653–661.
- Danner, M. & Seckler, R. (1993). *Protein Sci.* **2**, 1869–1881.
- Davidson, A. R., Cardarelli, L., Pell, L. G., Radford, D. R. & Maxwell, K. L. (2012). *Adv. Exp. Med. Biol.* **726**, 115–142.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Fuchs, A., Seiderer, C. & Seckler, R. (1991). *Biochemistry*, **30**, 6598–6604.
- Goldenberg, D. P., Berget, P. B. & King, J. (1982). *J. Biol. Chem.* **257**, 7864–7871.
- Goldenberg, D. & King, J. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 3403–3407.
- Goldenberg, D. P., Smith, D. H. & King, J. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 7060–7064.
- Haase-Pettingell, C. & King, J. (1997). *J. Mol. Biol.* **267**, 88–102.
- Heinemann, U., Büssov, K., Mueller, U. & Umbach, P. (2003). *Acc. Chem. Res.* **36**, 157–163.
- Israel, V. (1978). *J. Gen. Virol.* **40**, 669–673.
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Kabsch, W. (1976). *Acta Cryst.* **A32**, 922–923.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Karplus, P. A. & Schulz, G. E. (1985). *Naturwissenschaften*, **72**, 212–213.
- King, J., Haase-Pettingell, C., Robinson, A. S., Speed, M. & Mitraki, A. (1996). *FASEB J.* **10**, 57–66.
- King, J., Lenk, E. V. & Botstein, D. (1973). *J. Mol. Biol.* **80**, 697–731.
- Koide, S. (2009). *Curr. Opin. Struct. Biol.* **19**, 449–457.
- Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* **D49**, 129–147.
- Lander, G. C., Khayat, R., Li, R., Prevelige, P. E., Potter, C. S., Carragher, B. & Johnson, J. E. (2009). *Structure*, **17**, 789–799.
- Lander, G. C., Tang, L., Casjens, S. R., Gilcrease, E. B., Prevelige, P., Poliakov, A., Potter, C. S., Carragher, B. & Johnson, J. E. (2006). *Science*, **312**, 1791–1795.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
- Laskowski, R. A. (1995). *J. Mol. Graph.* **13**, 323–330.
- Leavitt, J. C., Gogokhia, L., Gilcrease, E. B., Bhardwaj, A., Cingolani, G. & Casjens, S. R. (2013). *PLoS One*, **8**, e70936.
- Leiman, P. G. & Shneider, M. M. (2012). *Adv. Exp. Med. Biol.* **726**, 93–114.
- Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
- Ma, B., Tsai, C. J., Halilović, T. & Nussinov, R. (2011). *Structure*, **19**, 907–917.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- McPherson, A. (2004). *Methods*, **34**, 254–265.
- Miller, S., Schuler, B. & Seckler, R. (1998). *Protein Sci.* **7**, 2223–2232.
- Mitraki, A., Danner, M., King, J. & Seckler, R. (1993). *J. Biol. Chem.* **268**, 20071–20075.
- Molineux, I. J. (2006). *Virology*, **344**, 221–229.
- Müller, J. J., Barbirz, S., Heinle, K., Freiberg, A., Seckler, R. & Heinemann, U. (2008). *Structure*, **16**, 766–775.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Navaza, J. (1994). *Acta Cryst.* **A50**, 157–163.
- Olia, A. S., Casjens, S. & Cingolani, G. (2007). *Nature Struct. Mol. Biol.* **14**, 1221–1226.
- Olia, A. S., Prevelige, P. E. Jr, Johnson, J. E. & Cingolani, G. (2011). *Nature Struct. Mol. Biol.* **18**, 597–603.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
- Price, S. R. & Nagai, K. (1995). *Curr. Opin. Biotechnol.* **6**, 425–430.
- Reddy, V., Swanson, S. M., Segelke, B., Kantardjiev, K. A., Sacchettini, J. C. & Rupp, B. (2003). *Acta Cryst.* **D59**, 2200–2210.
- Sousa, R. (1995). *Acta Cryst.* **D51**, 271–277.
- Steinbacher, S., Miller, S., Baxa, U., Budisa, N., Weintraub, A., Seckler, R. & Huber, R. (1997). *J. Mol. Biol.* **267**, 865–880.
- Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R. & Reinemer, P. (1994). *Science*, **265**, 383–386.
- Suhre, K. & Sanejouand, Y.-H. (2004). *Nucleic Acids Res.* **32**, W610–W614.
- Tang, J., Lander, G. C., Olia, A. S., Olia, A., Li, R., Casjens, S., Prevelige, P. Jr, Cingolani, G., Baker, T. S. & Johnson, J. E. (2011). *Structure*, **19**, 496–502.
- Tang, L., Marion, W. R., Cingolani, G., Prevelige, P. E. Jr & Johnson, J. E. (2005). *EMBO J.* **24**, 2087–2095.
- Tsai, C. J., del Sol, A. & Nussinov, R. (2008). *J. Mol. Biol.* **378**, 1–11.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- Walter, M., Fiedler, C., Grassl, R., Biebl, M., Rachel, R., Hermo-Parrado, X. L., Llamas-Saiz, A. L., Seckler, R., Miller, S. & van Raaij, M. J. (2008). *J. Virol.* **82**, 2265–2273.
- Weissenhorn, W., Calder, L. J., Dessen, A., Laue, T., Skehel, J. J. & Wiley, D. C. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 6065–6069.
- Weissenhorn, W., Carfi, A., Lee, K. H., Skehel, J. J. & Wiley, D. C. (1998). *Mol. Cell*, **2**, 605–616.
- Winston, F., Botstein, D. & Miller, J. H. (1979). *J. Bacteriol.* **137**, 433–439.