19. Nucifora, F. C. *et al.* Interference by huntingtin and atrophin-1 with CBP-mediated transcription leading to cellular toxicity. *Science* **291,** 2423–2428 (2001).
20. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293,** 321–331 (1999).
21. Clore, G. M. & Gronenborn, A. M. Multidimensional heteronuclear nuclear magnetic resonance of proteins. *Methods Enzymol.* **239,** 349–363 (1994).
22. Bax, A. *et al.* Measurement of homo- and heteronuclear J-couplings from quantitative J correlation. *Methods Enzymol.* **239,** 79–105 (1994).
23. Grzesiek, S., Kuboniwa, H., Hinck, A. P. & Bax, A. Multiple-quantum line narrowing for measurement of Hα–Hβ J coupling in isotopically enriched proteins. *J. Am. Chem. Soc.* **117,** 5312–5315 (1995).
24. Zwahlen, C. *et al.* Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. *J. Am. Chem. Soc.* **119,** 6711–6721 (1997).
25. Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273,** 283–298 (1997).
26. Pearlman, D. A. *et al.* AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp. Phys. Comm.* **91,** 1–41 (1995).
27. Duggan, B. M., Legge, G. B., Dyson, H. J. & Wright, P. E. SANE (structure assisted NOE evaluation): an automated model-based approach for NOE assignment. *J. Biomol. NMR* **19,** 321–329 (2001).
28. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8,** 477–486 (1996).
29. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14,** 51–55 (1996).
30. Myers, J. K., Pace, C. N. & Scholtz, J. M. Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4,** 2138–2148 (1995).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com).

Acknowledgements

We thank M. Allen for technical help and M. Pique for help with computer graphics. This work was supported by grants from the NIH and by the Skaggs Institute for Chemical Biology. S.J.D. was supported by a fellowship from the NIH.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.E.W. (e-mail: wright@scripps.edu). Coordinates have been deposited with the Protein Data Bank under accession number 1KBH.

Structure of the cell-puncturing device of bacteriophage T4

Shuji Kanamaru*§, Petr G. Leiman*§, Victor A. Kostyuchenko*†, Paul R. Chipman*, Vadim V. Mesyanzhinov†, Fumio Arisaka‡ & Michael G. Rossmann*

** Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907-1392, USA*
† Laboratory of Molecular Bioengineering, Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, 16/10 Miklukho-Maklaya Street, 117997 Moscow, Russia
‡ Department of Life Science, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8501, Japan
§ These authors contributed equally to this work

Bacteriophage T4 has a very efficient mechanism for infecting cells[1]. The key component of this process is the baseplate, located at the end of the phage tail, which regulates the interaction of the tail fibres and the DNA ejection machine[2]. A complex of gene product (gp) 5 (63K) and gp27 (44K), the central part of the baseplate, is required to penetrate the outer cell membrane of *Escherichia coli* and to disrupt the intermembrane peptidoglycan layer, promoting subsequent entry of phage DNA into the host. We present here a crystal structure of the $(gp5–gp27)_3$ 321K complex, determined to 2.9 Å resolution and fitted into a cryo-electron microscopy map at 17 Å resolution of the baseplate-tail tube assembly. The carboxy-terminal domain of gp5 is a triple-stranded β-helix that forms an equilateral triangular prism, which acts as a membrane-puncturing needle. The middle lysozyme domain of gp5, situated on the periphery of the prism, serves to digest the peptidoglycan layer. The amino-terminal, antiparallel β-barrel domain of gp5 is inserted into a cylinder formed by three gp27 monomers, which may serve as a channel for DNA ejection.

Bacteriophage T4 consists of a prolate head with icosahedral ends and a bilayered tail that has a hexagonal baseplate and fibres at its distal end[3]. The phage recognizes a host by using its long tail fibres, and then anchors the baseplate onto the lipopolysaccharide receptors on the cell surface. Attachment to the cell induces a conformational change in the baseplate that initializes a contraction of the external tail sheath, resulting in penetration of the outer cell membrane and intermembrane peptidoglycan by the rigid internal tail tube[2]. This sequence of events brings the tube into contact with the cytoplasmic cell membrane[1]. Subsequently, the phage DNA is injected into the host cell through the tail tube.

The centre of the baseplate, or hub, consists of at least two phage proteins, gp5 and gp27 (refs 4, 5), which form a stable complex *in vivo* and *in vitro*. Phage mutants lacking either of these proteins produce hubless baseplates and, as a result, are tailless and non-infectious. The hub has lysozyme activity attributed to the lysozyme domain of gp5, which is required to digest the cellular intermembrane peptidoglycan layer during phage tail contraction[6]. The gp5 lysozyme domain has 43% sequence identity[7] with the cytoplasmic T4 lysozyme[8] (T4L). Unlike T4L, gp5 lysozyme activity is essential for phage growth and infection[9]. When incorporated into the phage baseplate or stored at high concentration, gp5 undergoes maturational cleavage between Ser 351 and Ala 352, but both resultant

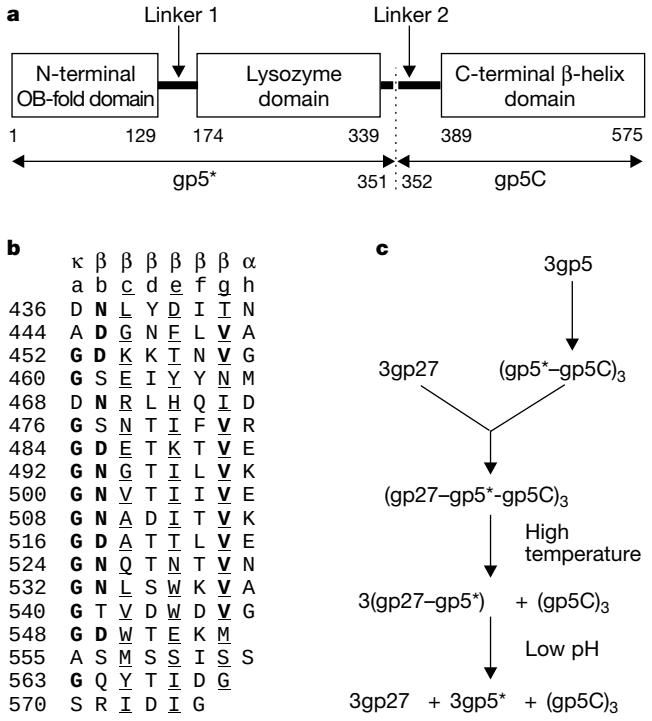

Figure 1 Assembly of $(gp27–gp5^*–gp5C)_3$. **a**, Domain organization within the gp5 monomer. The maturation cleavage is indicated by the dotted line. Initial and final residue numbers are shown for each domain. **b**, Alignment of the octapeptide units composing the intertwined part of the C-terminal β-helix domain of gp5. Conserved residues are in bold; residues facing the inside are underlined. The main-chain dihedral angle configuration of each residue in the octapeptide is indicated at the top by κ (kink), β (sheet), and α (helix). **c**, Assembly of gp5 and gp27 into the hub and needle of the baseplate.

parts, the N-terminal part (gp5*) and the C-terminal part (gp5C) (Fig. 1a), remain in the phage particle[10]. gp5C is an SDS-resistant trimer, rich in β-structure[10] and containing 11 VXGXXXXX sequence repeats[7,10] (Fig. 1b). The complex of gp5 with gp27 obtained *in vivo* and *in vitro* contains both gp5* and gp5C, thus comprising a heterononameric assembly: $(gp27–gp5^*–gp5C)_3$ (Fig. 1c). At elevated temperatures, this complex dissociates into three (gp27–gp5*) heterodimers and a $(gp5C)_3$ homotrimer, showing that gp5C is necessary for trimerization of the entire complex[10] (Fig. 1c). However, dissociation of gp5* from gp27 can occur only at low pH (S.K., Y. Takeda and F.A., unpublished data). The lysozyme activity of the trimeric gp5*, when stabilized by gp5C, is an order of magnitude less than the activity of the monomeric gp5* (ref. 10).

The structure of the $(gp27–gp5^*–gp5C)_3$ complex (Fig. 2a) resembles a torch (or flashlight) with a length of 190 Å in which the gp27 trimer forms its cylindrical 'head' and the gp5C trimer forms the 'handle'. The gp27 trimer makes a hollow cylinder about 60 Å long, with internal and external diameters of about 30 and 80 Å, respectively, encompassing the three N-terminal domains of gp5* to which the trimeric gp5C 'torch handle' is attached. The

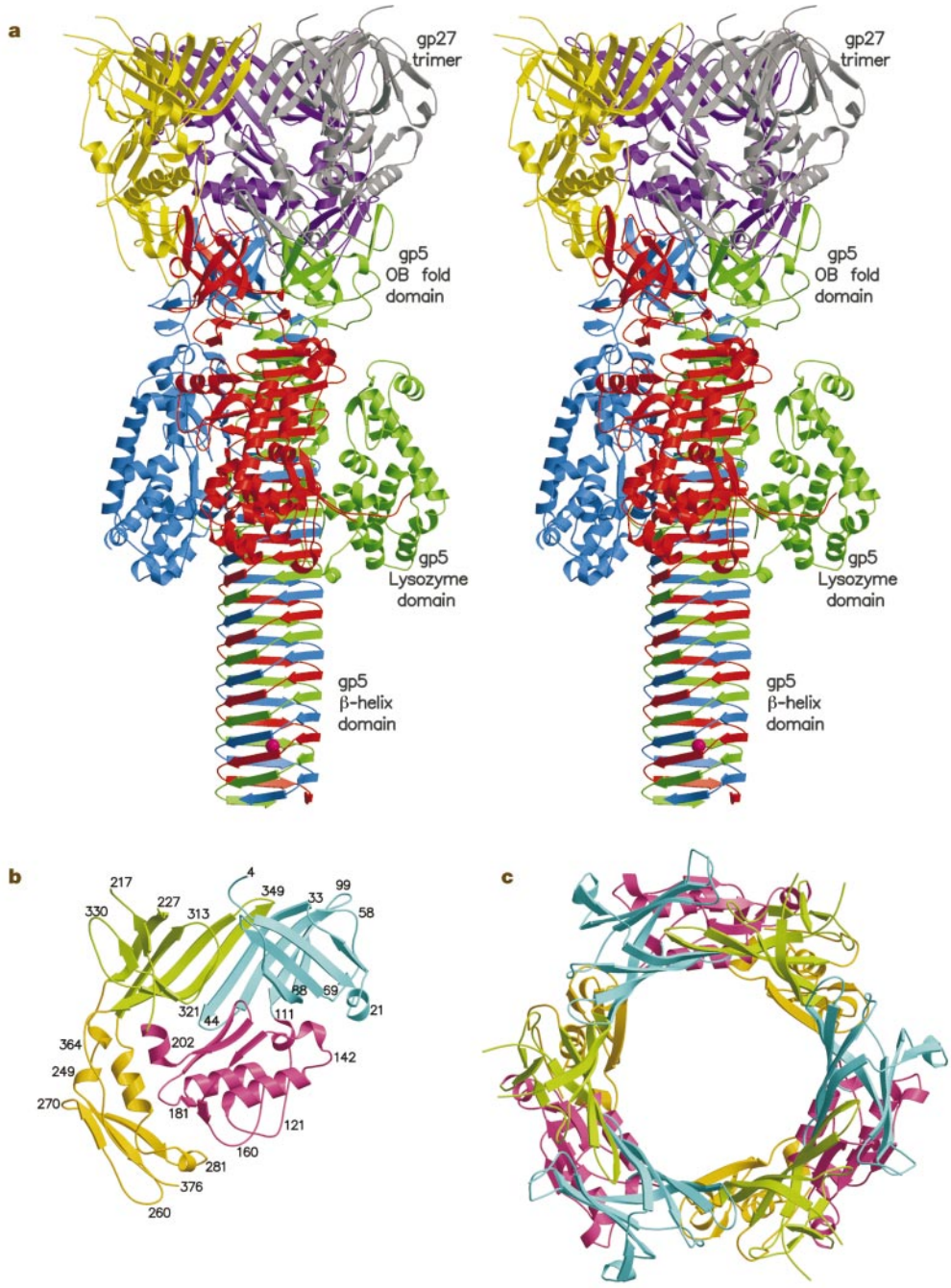

Figure 2 Structure of the $(gp27–gp5^*–gp5C)_3$ complex. **a**, Ribbon stereo diagram of the complex. There is one polypeptide chain of gp5 and one of gp27 in the asymmetric unit. The three-fold symmetry axis of the complex coincided with the crystallographic three-fold axis. The three gp5 monomers are coloured red, green and blue. The three gp27 monomers are coloured yellow, grey and purple. The K^+ ion within gp5C is shown in pink. The PO_4 ion is hidden behind the lysozyme domain. **b**, The structure of the gp27 monomer with its four domains, coloured cyan, pink, light green and gold, along the polypeptide chain. **c**, Top view of the gp27 cylinder, showing that the cyan and green domains form a hexagonal torus. Ribbon diagrams were drawn with MOLSCRIPT[22] and Raster3D[23].

C-terminal parts of the three gp5C polypeptide chains fold into a β-helical prism in which the faces have a slight left-handed twist. The middle lysozyme domains surround the N end of the prism. Generally, there was good agreement on fitting the (gp27–gp5*–gp5C)$_3$ structure into a 17 Å resolution three-dimensional image reconstruction of the baseplate-tail tube assembly, obtained with cryo-electron microscopy (cryoEM). As expected, the complex forms the central hub of the baseplate, with the β-helical prism domain forming a needle pointing towards the potential host surface (Fig. 3). The three-fold symmetry of the (gp27–gp5*–gp5C)$_3$ complex contrasts with the assumed six-fold symmetry of the baseplate, and may, in part, account for the absence of density representing the lysozyme domain. However, it is also probable that the lysozyme is somewhat disordered in its position, made possible by the long, flexible linker regions 1 and 2 (Fig. 1a).

The gp27 monomer, which consists of 319 amino acids[11], is folded into four domains (Fig. 2b). The first (residues 2–111) and third (residues 207–239 plus 307–368) domains of the three symmetry-related gp27 monomers create a torus. Each of these two domains is formed by seven- or eight-stranded antiparallel β-barrels. These can be superimposed on each other, with a root-mean-square deviation (r.m.s.d.) of 2.4 Å between equivalenced C_α atoms, by means of a 57° rotation approximately about the crystallographic threefold axis, thus creating a cylinder with a pseudo six-fold symmetry. Although there is no significant sequence similarity (4% sequence identity) between these domains, their external surfaces are mostly hydrophobic, providing a nucleus for the assembly of the six-fold-symmetric baseplate (Fig. 2c). Thus, the gp27 trimer forms an interface between the hexameric baseplate and trimeric gp5. The internal and external surfaces of the gp27 cylinder match the dimensions of the tail tube, suggesting that the gp27 trimer serves as an extension to the 90 Å-diameter tail tube (Fig. 3b).

The second domain (residues 112–206) and the fourth domain (residues 240–306 plus 359–376) of gp27 bind to the two adjacent N-terminal domains of gp5*, thus promoting assembly of the

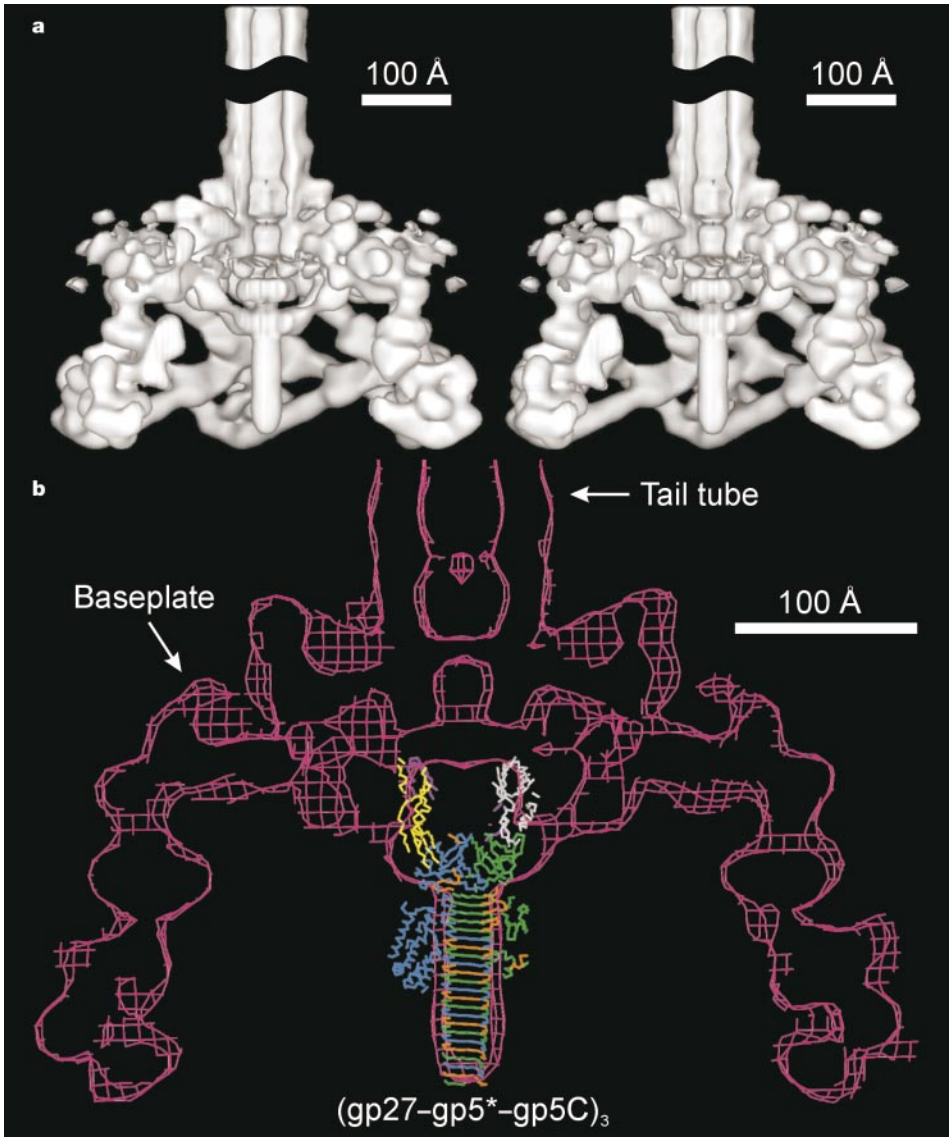

Figure 3 CryoEM reconstruction of the T4 baseplate-tail tube assembly. **a**, Stereo view of the surface-shaded representation. The top quarter of the baseplate has been removed to show the internal features. **b**, Cross-sectioned density fitted with the atomic (gp27–gp5*–gp5C)$_3$ complex structure. The three-fold-symmetric crystallographic structure fitted into the six-fold averaged cryoEM density is shown in the same colours as Fig. 2a. Three-fold averaging indicated that the lysozyme domain of gp5 may have moved into the baseplate or may be disordered. For clarity, the contour level of the surface-shaded figure in **a** is higher than the contour level of the cross-sectioned density in **b**. Thus, some of the density representing gp27 in **a** is missing compared with **b**.

oligomeric (gp27–gp5*–gp5C)$_3$ complex. The interface between gp27 and gp5* is formed mostly by complementary polar and charged residues, with the surface of gp27 being mostly positively charged and the surface of gp5* being mostly negatively charged. The surface contacts between gp27 monomers are formed primarily by a few hydrogen bonds, consistent with the lack of trimer formation when gp27 is expressed by itself (S.K., Y. Takeda and F.A., unpublished data).

Gp5* consists of both an N-terminal and a lysozyme domain (Fig. 1a). The N-terminal domain is a five-stranded, antiparallel β-barrel that has the oligonucleotide/oligosaccharide-binding (OB) fold[12]. The OB domain of enterotoxin LT (Protein Data Bank accession number 1LTT) can be superimposed onto the N-terminal domain of gp5 with an r.m.s.d. of 2.4 Å between the 50 structurally equivalenced C_α atoms. Unlike the equivalent nucleotide-binding site, the gp5 oligosaccharide-binding site has appropriate aromatic and charged residues, but in the assembled structure of (gp27–gp5*–gp5C)$_3$, the putative polysaccharide-binding site is occluded by intersubunit contacts. Therefore, the function of the OB domain might be to facilitate binding of gp5* to oligosaccharides of the periplasmic peptidoglycan layer for subsequent digestion by the lysozyme domain, suggesting that gp5* separates from the baseplate during tail contraction (see below).

The gp5 lysozyme domain, located on the periphery of the central β-helix, is connected to the N-terminal domain by linker 1 and to the C-terminal β-helical domain by linker 2, which contains the cleavage site between gp5* and gp5C (Fig. 1a). Seven residues upstream and nine residues downstream of the cleavage site are disordered in the crystal structure. Structural comparison with T4L[13] shows that residues (Pro 363–Ala 364–Asp 365) of linker 2 of the three-fold-related neighbouring subunit bind, correctly oriented, into three of the four peptide-binding sites used by the peptidoglycan substrate (Fig. 4), but leave the catalytic polysaccharide binding cleft open to solution. The other side of the polysaccharide binding cleft is sterically blocked by the β-helix (Fig. 4). Both of these blockages explain the lack of activity of trimeric (gp5*–gp5C)$_3$ compared with monomeric gp5* (ref. 10).

The C_α atoms of the gp5 lysozyme domain and of T4L can be superimposed on each other with an r.m.s.d. of 1.1 Å, except for five inserted residues in gp5. This analysis also shows that both lysozymes have the same residues in their active sites: Glu 11, Asp 20 and Thr 26 in T4L correspond to Glu 184, Asp 193 and Thr 199 in gp5, respectively, establishing that the enzymatic mechanism is the same and that the molecules probably have a common evolutionary origin.

The gp5 and T4L enzymes have the same natural substrate, the *E. coli* periplasmic cell wall, the major component of which—(NAG-NAM)-L-Ala-D-*iso*-Glu-DAP-D-Ala (ref. 13)—contains sugar and peptide moieties. However, gp5 and T4L have activity optima at pH 5.8 and pH 6.8, respectively. The peptide-binding surfaces of gp5 and T4L have several amino-acid differences that cause the peptide-binding surface in T4L to be more positively charged than that in gp5 at neutral pH, possibly accounting for the differences in pH activity optima.

Three gp5C polypeptide chains wind around a central crystallographic three-fold axis to create an equilateral triangular prism that is 110 Å long and 28 Å in diameter. Each face has a slight left-handed twist (−3° per β-strand), as is normally observed in β-sheets. The width of the prism face narrows gradually from 33 Å at the N end to 25 Å at the C end, thus creating a pointed needle. This narrowing is caused by a decrease in size of the external side chains and by the internal methionines 554 and 557, which break the octapeptide repeat near the tip of the helix (Fig. 1c). The first five β-strands (residues 389–435) form an antiparallel β-sheet, creating one of the three faces of the prism. A similar β-sheet prism domain has been observed in the central portion of the phage P22 tailspike protein[14,15]. However, unlike the hydrophobic interior of the P22 tailspike, the gp5 structure is mostly polar in its interior. The succeeding 18 β-strands of gp5 form a three-start β-helix together with the other 2 three-fold-related polypeptides, generating six complete turns of the β-cylinder. The intertwined, C-terminal part of the β-helical prism (residues 436–575) is a remarkably smooth continuation of its three monomeric N-terminal parts (residues 389–435). Another example of a three-fold intertwined fibrous β-structure is the adenovirus fibre[16]. However, the lack of superimposable structural similarity suggests that these viral fibres may be the consequence of convergent evolution to useful, stable, oligomeric fibrous structures.

The octapeptide sequence of the helical intertwined part of the prism (residues a–h, Fig. 1b) has dominant glycines at position a, asparagines or aspartic acids at position b, valines at position g, and polar or charged residues at position h (Fig. 1b). Residues b–g form extended β-strands (Ramachandran angles $\phi \simeq -129°$, $\psi \simeq 128°$) that run at an angle of 75° with respect to the helix axis. The glycines at position a ($\phi \simeq -85°$, $\psi \simeq -143°$, an allowed region of the Ramachandran diagram) and residues at position h ($\phi \simeq -70°$, ψ

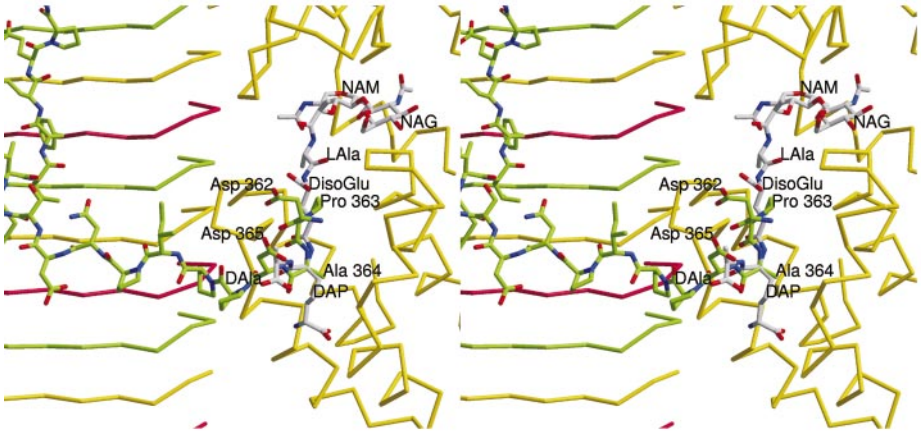

Figure 4 Structural comparison with T4L[13] shows that residues Pro 363-Ala 364-Asp 365 of linker 2 of the three-fold-related neighbouring subunit bind, correctly oriented, into three of the four peptide-binding sites used by the peptidoglycan substrate. Polypeptide chains of gp5 are represented by virtual C_α–C_α bonds, except for those in linker 2, which are represented as a stick model, with green for carbon atoms. The peptidoglycan substrate as bound to T4L[13] is also shown as a stick model, with grey for carbon atoms. The three individual polypeptide chains of gp5 are coloured yellow, red and green. DAla, D-Ala; LAla, L-Ala; DisoGlu, D-*iso*-Glu; NAG, *N*-acetylglucosamine; NAM, *N*-acetylmuramic acid; DAP, diaminopimelic acid.

$\simeq -30°$, typical for α-helices) kink the polypeptide chain by about 130° clockwise. The conserved valines at position g always point to the inside of the β-helix and form a 'knob into holes' arrangement with the main-chain atoms of the glycines at position a and the aliphatic part of the side chains of residues at position c. Asp 436 replaces the normal glycine in position a and is at the start of the β-helix. This substitution may be required for folding of the β-helix, because the Asp 436 O_δ atom makes a hydrogen bond with O_γ of Ser 427 from a three-fold-related polypeptide chain. The side-chain oxygen atoms of Asp 468, which also occupies position a, form hydrogen bonds with residues in the lysozyme domain.

The inside of the β-helix is progressively more hydrophobic toward its C-terminal tip (Fig. 1b). The middle part of the helix has a pore, which is filled with at least 42 water molecules bound to polar and charged side chains. The helix is stabilized by two ions situated on the helix symmetry axis: a phosphate ion coordinated by three Lys 454 residues and a potassium ion coordinated by three Glu 552 residues. These ions were identified by the presence of density peaks and their amino-acid environments. Most probably, these ions are incorporated into the β-helix during folding, as no such ions are present in the purification and crystallization solutions. All these features of the β-helix create a highly stable structure, consistent with its resistance to 10% SDS or 2 M guanidine HCl. The surface of the β-helix is highly negatively charged. This charge is necessary to repel the phosphates of the lipid bilayer when the β-helix penetrates through the outer cell membrane.

The $(gp27–gp5^*–gp5C)_3$ complex can be fitted into the cryoEM image reconstruction of the T4 baseplate-tail tube assembly (Fig. 3b). The gp27 cylinder is shown to be an extension of the tail tube, which is continued by the three N-terminal domains of gp5. The latter associates with the N end of the β-helix that terminates the extended tube. The three lysozyme domains are located around the β-helix under the baseplate, but vary in their exact positions. The diameter of the gp27 cylinder is of a size that can accommodate a double-stranded DNA helix, thus forming a channel that would allow the DNA to penetrate the tail tube as far as the N-terminal domain of gp5.

On attachment of the baseplate to the cell surface, the tail sheath contracts, exerting a force onto the tail tube towards the cell membrane. This force would then be transmitted through the gp27 cylinder and the N-terminal domain of gp5 to the β-helix, causing the latter to puncture the outer membrane. As the contraction of the tail sheath progresses, the β-helix would be able to span the entire 40 Å width of the outer membrane, thereby enlarging the pore in the membrane. Subsequently, the three lysozyme domains would reach the peptidoglycan layer to digest the cell wall, allowing penetration of the tail tube to the inner membrane for injection of the phage DNA into the host. □

Methods

Separate expression vectors containing gene 5 and gene 27 and a coexpression vector were constructed (S.K., Y. Takeda and F.A., unpublished data). The complex was purified with the aid of a gp5 C-terminal His tag. Crystals were formed in hanging drops at 12 °C from 5.2% PEG 8000, 0.65 M Tris buffer at pH 8.5, and 35% glycerol. They belonged to space group *R*32, with cell dimensions of $a = 138.1$ and $c = 388.9$ Å. Selenomethionine (SeMet)-substituted gp5 was insoluble, but, fortunately, the SeMet-substituted gp27 was soluble. A chimaeric complex containing separately expressed native gp5 and the SeMet-substituted gp27 was purified and crystallized. Four-wavelength multiple anomalous dispersion data were collected using the chimaeric crystals of the complex that diffracted to 2.9 Å resolution. Fifteen of the sixteen selenium atoms in the gp27 monomer were found with the help of the program SHELX[17]. These sites were refined and used to obtain phases with the program MLPHARE[18]. The phases were then improved by solvent flipping using the program CNS[19]. The resultant electron density map was easily interpreted with the graphics program XtalView[20]. Refinement of the atomic model with the program CNS[19] converged to the working *R*-factor of 21% and R_{free} of 28%.

Mutant T4 phages lacking the ability to assemble the tail sheath and head were used to produce the baseplate-tail tube assembly. Three-dimensional image reconstructions of the baseplate-tail tube assembly were computed by SPIDER[21]. The baseplate was assumed to be six-fold symmetric. A total of 418 images were employed to obtain a map at about 17 Å resolution.

Received 1 October; accepted 3 December 2001.

1. Goldberg, E., Grinius, L. & Letellier, L. in *Molecular Biology of Bacteriophage T4* (ed. Karam, J. D.) 347–356 (American Society for Microbiology, Washington, DC, 1994).
2. Coombs, D. H. & Arisaka, F. in *Molecular Biology of Bacteriophage T4* (ed. Karam, J. D.) 259–281 (American Society for Microbiology, Washington, DC, 1994).
3. Eiserling, F. A. & Black, L. W. in *Molecular Biology of Bacteriophage T4* (ed. Karam, J. D.) 209–212 (American Society for Microbiology, Washington, DC, 1994).
4. Kikuchi, Y. & King, J. Genetic control of bacteriophage T4 baseplate morphogenesis. III. Formation of the central plug and overall assembly pathway. *J. Mol. Biol.* **99,** 695–716 (1975).
5. Vanderslice, R. W. & Yegian, C. D. The identification of late bacteriophage T4 proteins on sodium dodecyl sulfate polyacrylamide gels. *Virology* **60,** 265–275 (1974).
6. Nakagawa, H., Arisaka, F. & Ishii, S. Isolation and characterization of the bacteriophage T4 tail-associated lysozyme. *J. Virol.* **54,** 460–466 (1985).
7. Mosig, G., Lin, G. W., Franklin, J. & Fan, W. H. Functional relationships and structural determinants of two bacteriophage T4 lysozymes: a soluble (gene e) and a baseplate-associated (gene 5) protein. *New Biol.* **1,** 171–179 (1989).
8. Matthews, B. W. & Remington, S. J. The three dimensional structure of the lysozyme from bacteriophage T4. *Proc. Natl Acad. Sci. USA* **71,** 4178–4182 (1974).
9. Takeda, S., Hoshida, K. & Arisaka, F. Mapping of functional sites on the primary structure of the tail lysozyme of bacteriophage T4 by mutational analysis. *Biochim. Biophys. Acta* **1384,** 243–252 (1998).
10. Kanamaru, S., Gassner, N. C., Ye, N., Takeda, S. & Arisaka, F. The C-terminal fragment of the precursor tail lysozyme of bacteriophage T4 stays as a structural component of the baseplate after cleavage. *J. Bacteriol.* **181,** 2739–2744 (1999).
11. Bova, R. *et al.* Bacteriophage T4 gene 27. *Nucleic Acids Res.* **18,** 3046 (1990).
12. Murzin, A. G. & Chothia, C. Protein architecture: new superfamilies. *Curr. Opin. Struct. Biol.* **2,** 895–903 (1992).
13. Kuroki, R., Weaver, L. H. & Matthews, B. W. A covalent enzyme-substrate intermediate with saccharide distortion in a mutant T4 lysozyme. *Science* **262,** 2030–2033 (1993).
14. Steinbacher, S. *et al.* Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. *Science* **265,** 383–386 (1994).
15. Seckler, R. Folding and function of repetitive structure in the homotrimeric phage P22 tailspike protein. *J. Struct. Biol.* **122,** 216–222 (1998).
16. van Raaij, M. J., Mitraki, A., Lavigne, G. & Cusack, S. A triple β-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401,** 935–938 (1999).
17. Sheldrick, G. M. in *Crystallography of Biological Macromolecules* International Tables for Crystallography Vol. F (eds Rossmann, M. G. & Arnold, E.) 734–738 (Kluwer Academic, Dordrecht, 2001).
18. Otwinowski, Z. in *Isomorphous Replacement and Anomalous Scattering. Proc. CCP4 Study Weekend, 25–26 January 1991* (eds Wolf, W., Evans, P. R. & Leslie, A. G. W.) 80–86 (Science and Engineering Research Council, Daresbury, UK, 1991).
19. Brünger, A. T. *et al.* Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
20. McRee, D. E. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125,** 156–165 (1999).
21. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
22. Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24,** 946–950 (1991).
23. Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277,** 505–524 (1997).

Acknowledgements

We thank T. S. Baker for establishing the electron microscopy facilities at Purdue, where we collected the cryoEM data, A. A. Simpson for advice and help in data collection, and B. W. Matthews for discussion of the results. We thank the staff of BioCARS for their help and advice in the data collection at the Advanced Photon Source beam lines 14-BM-C and 14-BM-D. We thank S. Wilder for help in preparation of the manuscript. The work was supported by a National Science Foundation grant to M.G.R.; Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan to F.A.; a Howard Hughes Medical Institute grant to V.V.M.; a reinvestment grant from Purdue University; and a Keck Foundation award for the purchase of a Philips CM300 electron microscope.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.G.R. (e-mail: mgr@indiana.bio.purdue.edu). The coordinates of the $(gp27–gp5^*–gp5C)_3$ complex have been deposited in the Protein Data Bank, under an accession number of 1K28.