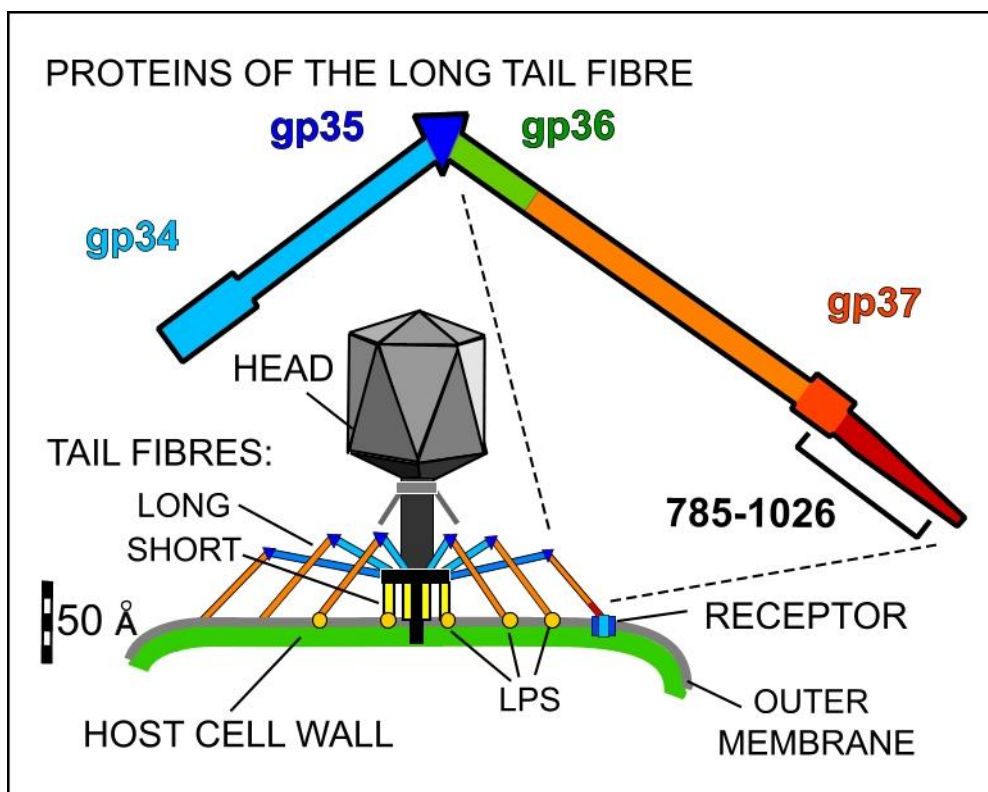


# Bacteriophages get a foothold on their prey

## Long and thin, the receptor-binding needle of bacteriophage T4

Bacterial viruses, bacteriophages or “phages”, have served as a tool to decipher principles of molecular biology for many decades. They are also used in applications such as the phage display of peptide libraries, phage typing to identify bacterial strains, and phage therapy to combat pathogenic bacteria. Most bacteriophages have a cell-attachment and infection organelle (a “tail”) and therefore belong to the *Caudovirales* order. The *Myoviridae* family in this order is characterised by tails with a contractile outer sheath, which drives an inner tail tube through the bacterial cell wall. Once this tube has cut through the cell wall the bacteriophage DNA is injected. This DNA directs the synthesis of the next generation of phages. Bacteriophage T4, which infects *Escherichia coli*, is probably the best known myovirus. As shown in the diagram, it is a wonderful macromolecular machine with a 100Å head (grey), which contains the DNA, long tail fibres (also shown in close up, coloured by gene product) to bind and recognise the host, and short tail fibres (yellow) that lock down the tail. PDB entry 2xgf [view 1] contains the structure of the tip region of the long tail fibre that recognises the host and activates the injection apparatus.



*The proteins that form the bacteriophage T4 long tail fibre and the components of the complete bacteriophage bound to the host outer membrane. LPS, lipopolysaccharide, is a specialised lipid component of this membrane. Gp37 trimer consists of three regions (orange to red). PDB entry 2xgf contains the structure of a part of this trimer, indicated by the bracket.*

The long tail fibres can extend more than 1000Å out from the tail and have a total mass of more than half a million daltons. They consist of four different proteins and assume an articulated leg-like shape. The "thigh" is formed by a homotrimer of the 1289-residue long gene product 34 (gp34, cyan) and the "knee" by a monomer of the 372-residue gp35 (blue). The upper part of the "lower leg" is a trimer of the 221-residue gp36 (green). Finally, the rest of the lower leg and the needle-shaped "foot" is formed by a trimer of the 1026-residue gp37 (orange/red). It is the C-terminal region of the foot domain of gp37 (red) that is responsible for recognising the host bacterium's outer membrane.

Owing to the large size of these fibrous protein components, structure determination by high-resolution NMR spectroscopy is not possible. The flexibility-inducing hinges also preclude cryo-EM averaging of single particle images. This same flexibility, combined with the asymmetric shape also makes crystallisation difficult. However, structural information can be obtained using a divide-and-conquer approach, in which domains are expressed and crystallised separately. This has been done for the receptor-binding region of gp37, which contains residues 785-1026 [ref 1] (red section in the diagram).

### **Three protein strands wrap around each other and around irons**

The structure of the receptor-binding region of gp37 (PDB entry 2xgf) [view 1] revealed three domains. A globular collar domain, an intervening extended fibrous section called the needle domain, and finally a head domain that contains the residues implicated in binding to the host. The collar domain is the only domain in which each monomer forms a separate structure and contains the N- and C-termini of PDB entry 2xgf [view 2]. Each polypeptide then contributes to a highly interwoven structure in which they traverse 175 Å to the terminal head domain and back, forming a long six-stranded anti-parallel beta-barrel. At the centre of the interwoven strands are seven iron ions [view 3]. Each iron has octahedral coordination provided by six histidine residues (two from each chain) [view 3]. These central iron ions hold the strands together, but they were also essential in solving the crystal structure. Clever use of X-rays tuned to the electronic structure of these ions allowed the crystallographers to locate their positions accurately and this knowledge enabled them to obtain the first electron density maps of the structure.

### **Phage specificity from lipid- or membrane protein-binding**

The business end of the fibre is the small head domain [view 4], which interacts with the outer membrane of *E. coli*. Target receptors for the domain are either the specialised membrane lipid lipopolysaccharide (LPS) or the trimeric outer-membrane protein OmpC. No experimental structures of gp37 bound to LPS or OmpC are yet available, but inspection of the gp37 tip structure provides some clues about possible interactions. LPS is likely to bind to aromatic or positively-charged residues and several of these are present in the gp37 structure.

However, it is also striking how compatible the diameter of the head domain is with the outer cavity of trimeric OmpC (PDB entry 2j1n). OmpC is a major outer-membrane protein of *E. coli*. Each of its three identical subunits folds into a beta-barrel that acts as a pore for small hydrophilic molecules. Taken together, however, the group of three subunits forms a bowl-shaped cavity when viewed from the outer face of the membrane. The head domain of the fibre can be neatly docked into this cavity. One example of the docking of the head domain of gp37 with the OmpC trimer is shown here [view 5]. Docked structures such as this suggest that the gp37 head domain can bind to the host OmpC [ref. 2]. The predicted

binding of the gp37 shown here does not require it to be vertical and the oblique arrangement observed [view 5] is compatible with the arrangement of the long tail fibres stretching out from the base of the phage.

### **Clues from sequence and structure about fibre protein evolution in phages**

Sequence comparisons of the gp37 proteins from other phages suggest that the needle-shaped structural domain is conserved in a family including Tula, Tulb, and lambda phages. However, the sequences of the other family members contain eight pairs of His residues so they probably contain eight iron ions. Their receptor-binding head domains also have different sequences and appear to be larger, suggesting they may bind to different receptors.

Searches for structures similar to the domains of gp37 show that the trimeric arrangement in the collar domain is also conserved in the structure of the short fibre protein gp12 (PDB Entry 1ocy). The similarity extends into a short metal-binding domain but in gp12 there is only a single metal-binding site - for zinc rather than iron. It seems possible therefore that the long-fibre gp37 needle domain structure has evolved by repeated duplication of the metal-binding motifs. This would explain how gp37 proteins in different phages have developed fibres with different numbers of ions bound.

The structure of the receptor-binding domain of the T4 long tail fibre opens the door to modifying it to target different receptors. In the future this may lead to engineered bacteriophages that specifically recognise and eliminate pathogenic bacteria.

### **Acknowledgement**

This Quips article was developed in collaboration with Mark van Raaij.

### **Further exploration**

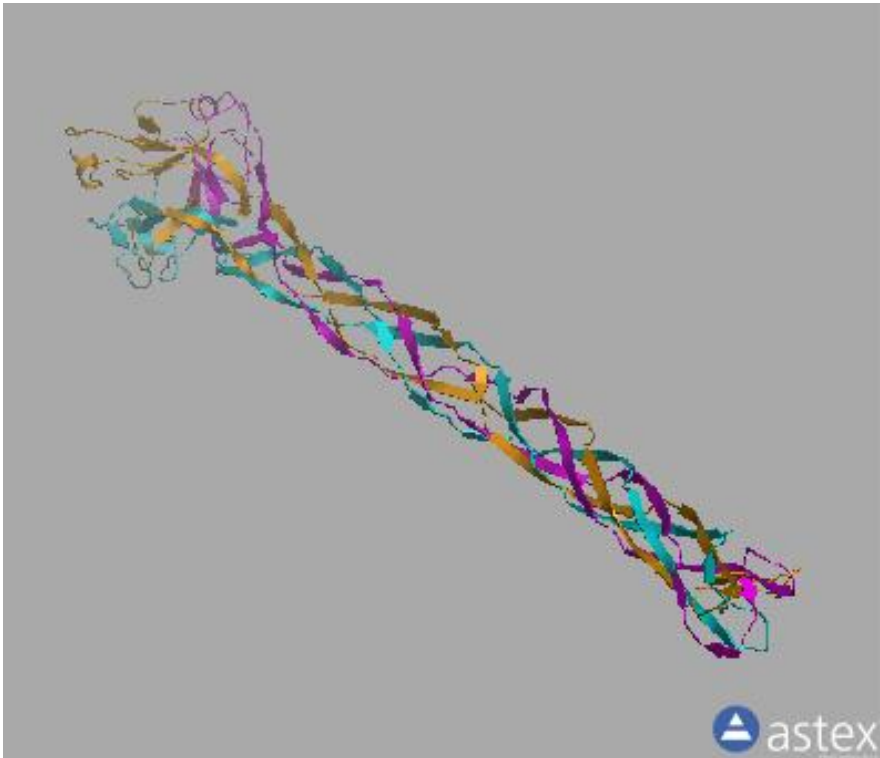
The iron coordination in gp37 can be analysed using the PDBeMotif service as explained in the mini-tutorial. The simplified docking process shown in [view 5] uses output from the Hex 6.3 program. Hex uses spherical polar Fourier correlations to provide rapid docking of protein surfaces – it can be downloaded from <http://hex.loria.fr/> [ref 3].

### **References:**

Ref. 1. Bartual SG, Garcia-Doval C, Alonso J, Schoehn G, van Raaij MJ. Two-chaperone assisted soluble expression and purification of the bacteriophage T4 long tail fibre protein gp37. *Protein Expr Purif.* 2010; 70(1):116-21.

Ref. 2. Bartual SG, Otero JM, Garcia-Doval C, Llamas-Saiz AL, Kahn R, Fox GC, van Raaij MJ. Structure of the bacteriophage T4 long tail fiber receptor-binding tip. *Proc Natl Acad Sci U S A.* 2010; 107(47):20287-92.

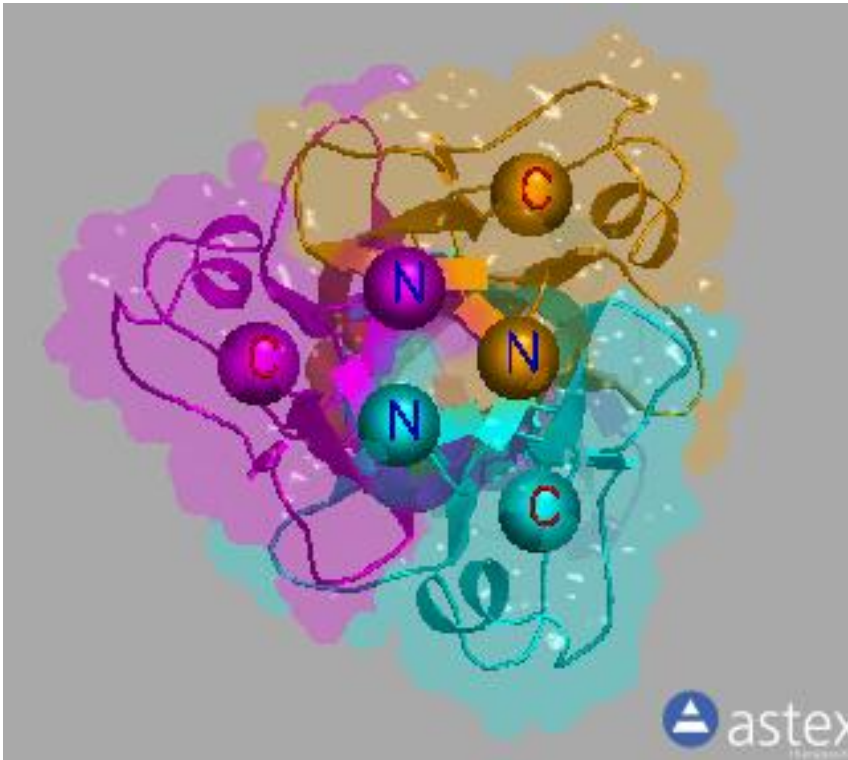
Ref. 3. Ritchie, D.W., Kozakov, D. Vajda, S. Accelerating and focusing protein–protein docking correlations using multi-dimensional rotational FFT generating functions. *Bioinformatics.* 2008; 24 (17): 1865-73.



View 1

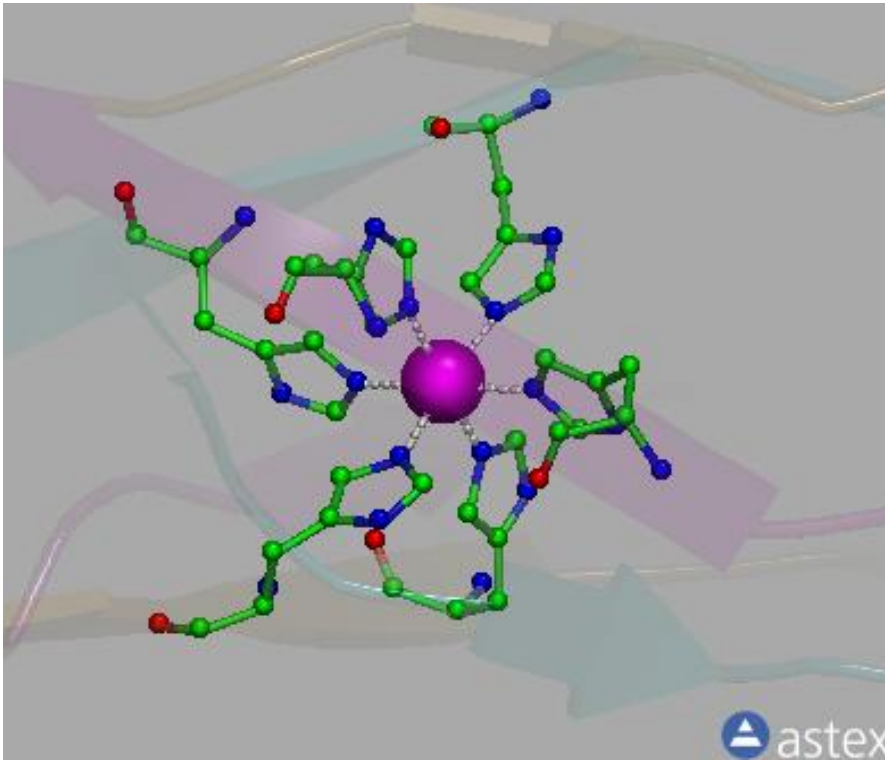
**Structure of the receptor-binding region of gp37 from bacteriophage T4.**

The structure (PDB entry [2xgf](#)) has three domains: a collar domain (left), a needle domain (middle) and a head domain (right). The entire structure consists of a highly intertwined arrangement of three polypeptide chains.



View 2

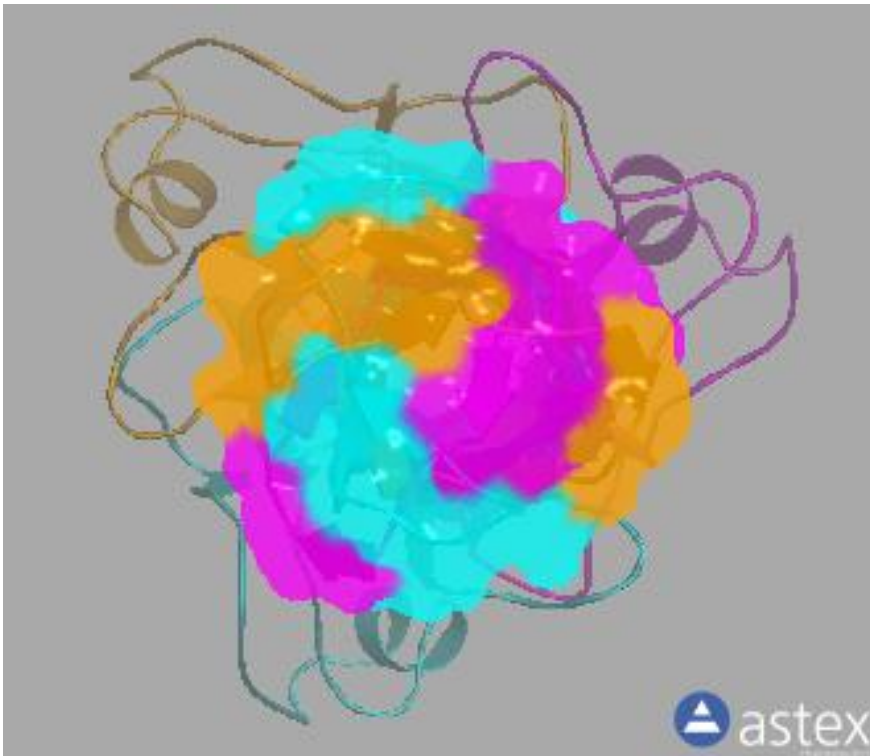
**The collar domain of gp37 (residues 811-860 and 1016-1026 from all three chains).**  
This all-beta domain contains both the N and C termini of the three chains of the trimer.



View 3

**The needle domain binds seven iron ions.**

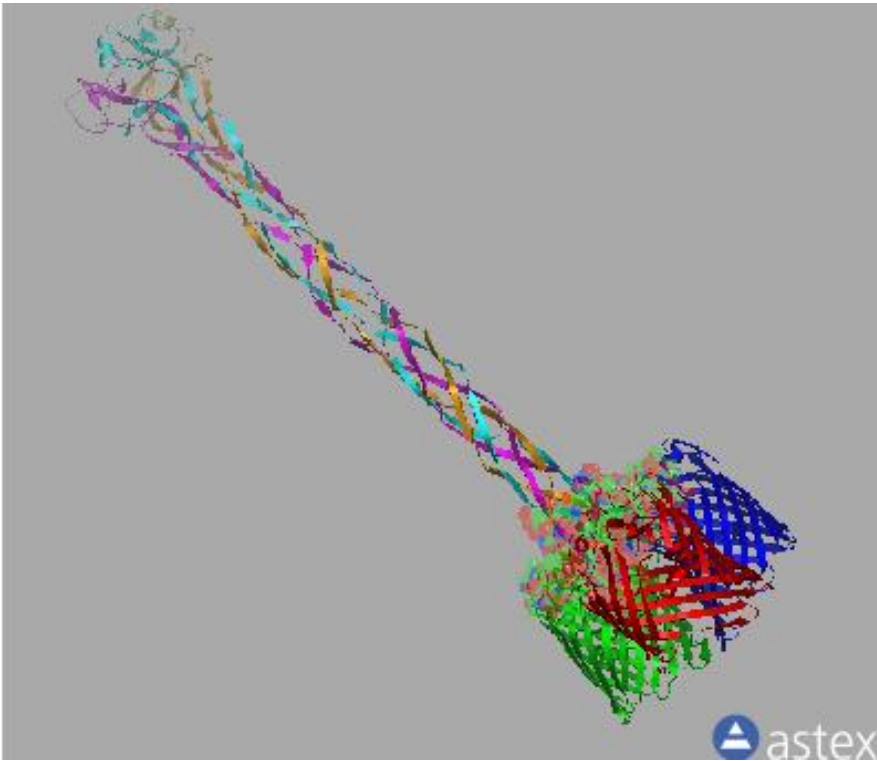
Each of the irons is octahedrally coordinated by six histidine sidechains, two from each chain.



View 4

**Receptor-binding head domain of gp37.**

The head domain contains sharp turns in the three polypeptide chains, thereby forming a surface that is available to bind the OmpC receptor.



View 5

**Predicted binding of gp37 with OmpC.**

The outer membrane protein OmpC is a trimeric protein made up of three  $\beta$ -barrels. Its outer surface has a bowl-shaped cavity that matches the surface of the gp37 head domain. This simulated docking shows how the head domain could bind to the cavity.