

The structure of bacteriophage T4 gene product 9: the trigger for tail contraction

Victor A Kostyuchenko^{1,2}, Grigorii A Navruzbekov¹, Lidia P Kurochkina¹, Sergei V Strelkov^{1,2†}, Vadim V Mesyanzhinov¹ and Michael G Rossmann^{2*}

Background: The T4 bacteriophage consists of a head, filled with double-stranded DNA, and a complex contractile tail required for the ejection of the viral genome into the *Escherichia coli* host. The tail has a baseplate to which are attached six long and six short tail fibers. These fibers are the sensing devices for recognizing the host. When activated by attachment to cell receptors, the fibers cause a conformational transition in the baseplate and subsequently in the tail sheath, which initiates DNA ejection. The baseplate is a multisubunit complex of proteins encoded by 15 genes. Gene product 9 (gp9) is the protein that connects the long tail fibers to the baseplate and triggers the tail contraction after virus attachment to a host cell.

Results: The crystal structure of recombinant gp9, determined to 2.3 Å resolution, shows that the protein of 288 amino acid residues assembles as a homotrimer. The monomer consists of three domains: the N-terminal domain generates a triple coiled coil; the middle domain is a mixed, seven-stranded β sandwich with a topology not previously observed; and the C-terminal domain is an eight-stranded, antiparallel β sandwich having some resemblance to 'jelly-roll' viral capsid protein structures.

Conclusions: The biologically active form of gp9 is a trimer. The protein contains flexible interdomain hinges, which are presumably required to facilitate signal transmission between the long tail fibers and the baseplate. Structural and genetic analyses show that the C-terminal domain is bound to the baseplate, and the N-terminal coiled-coil domain is associated with the long tail fibers.

Introduction

Coliphage T4 is one of the most structurally complex viruses [1]. The T4 genome consists of 168 kilobase (kb) pairs with about 300 open reading frames, only about 100 of which have had their function identified [2]. In comparison, the smaller tailed phage φ29 has a 19.3 kb genome encoding only 17 genes [3]. The T4 particle has a prolate head with icosahedral caps that contain the genomic DNA. The tail, which is connected to the head after DNA packaging, has a contractile sheath and a baseplate to which are attached the long and short tail fibers (LTFs and STFs, respectively).

The baseplate has a hexagonal shape that is formed of six identical 'wedges', or 'arms', each of which is a complex of ten different gene products, some forming oligomers. In the center of the baseplate there is a 'hub', a complex of eight proteins, including lysozyme [4]. The proximal ends of the LTFs are attached to gene product 9 (gp9), located near each apex of the hexagonal baseplate [5,6] (Figure 1). The LTFs, STFs, and gp9 are involved in the final stages of baseplate assembly [4]. Both LTFs and STFs are fibrous homotrimeric proteins [7,8].

Addresses: ¹Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, 16/10 Miklukho-Maklaya Street, Moscow 117871, Russia and ²Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392, USA.

[†]Present address: ME Müller-Institut, Biozentrum der Universität Basel, CH-4056 Basel, Switzerland.

*Corresponding author.
E-mail: mgr@indiana.bio.purdue.edu

Key words: bacteriophage T4, baseplate, gene product 9, oligomerization, X-ray crystallography

Received: 10 May 1999
Revisions requested: 23 June 1999
Revisions received: 9 July 1999
Accepted: 12 July 1999

Published: 30 September 1999

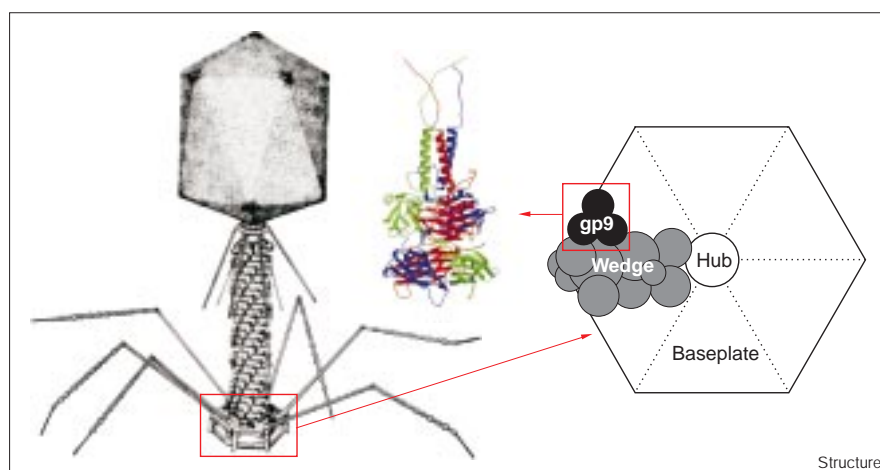
Structure October 1999, 7:1213–1222
<http://biomednet.com/elecref/0969212600701213>

0969-2126/99/\$ – see front matter
© 1999 Elsevier Science Ltd. All rights reserved.

The adsorption of T4 phage on *Escherichia coli* begins with a reversible interaction of the LTFs to lipopolysaccharide (LPS) receptors on the bacterial cell surface. When at least three LTFs have bound to receptors, a signal is transmitted to the baseplate and, subsequently, to the STFs. The STFs, which are usually folded under the baseplate, then unfold and bind irreversibly to specific cell-surface receptors, accompanied by alteration of the shape of the baseplate from a hexagon to a six-pointed, star-like structure [5]. This, in turn, triggers the sheath contraction, resulting in an interaction of the baseplate hub with the cell, penetration of the tail tube across the cell wall, and injection of the genomic DNA into the *E. coli* cytosol [9].

When the LTFs are attached to the baseplate, they can adopt two distinct configurations, designated as 'up' and 'down' [10,11]. In the down position, the LTFs are extended and are able to interact with their cell-surface receptors and initiate infection. At acidic pH or low temperature, the LTFs are retracted into the up configuration and are close to the tail sheath of the virus particle. In the latter case, the virus is not infectious because the distal parts of the LTFs, which recognize the LPS receptors, are

Figure 1



Schematic diagram showing the location of gp9 in the baseplate of bacteriophage T4. (The figure is adapted from [1] with permission from FA Eiserling.)

not able to interact with the bacterial cell surface. Kellenberger *et al.* [12] have suggested that the baseplate determines the collective behavior of the LTFs by fixing the hinge angle, around which the fibers can oscillate freely. Hence, gp9 might be required to regulate the up and down positions of the LTFs.

T4 phage mutants that lack both gp9 and LTFs have most of their baseplates in the 'star' conformation and have contracted tail sheaths, even before any interaction with the host. Phages that contain gp9 but lack the LTFs, however, are non-infectious and have hexagonal baseplates. This suggests that gp9, besides connecting the LTFs to the baseplate and signaling the attachment of LTFs to the target cell, also inhibits the spontaneous baseplate transition in the absence of LTFs [13]. It was also demonstrated [14] that T4 particles are infectious even in the absence of LTFs, showing that the attachment of the LTFs to the cell is not essential.

It is our goal to study the mechanisms of conformational transitions in the baseplate that occur during the initial stages of virus infection. Especially, we would like to understand how gp9 controls the structural reorganization of the phage tail. Here we present the three-dimensional structure of recombinant gp9 at atomic resolution (Figure 2). The protein forms a trimer, with each monomer having three distinct domains. The N-terminal, α -helical domain forms a parallel coiled-coil structure, and the C-terminal domain is an eight-stranded, antiparallel β sandwich slightly resembling the 'jelly-roll' β barrel found in many viral capsid proteins [15]. The middle domain, a seven-stranded, mixed β sandwich, has a novel fold not previously observed. There are large loops that connect the domains and provide flexibility, which is presumably required for the signal transmission and conformational transitions in the baseplate that occur during the initial steps of infection.

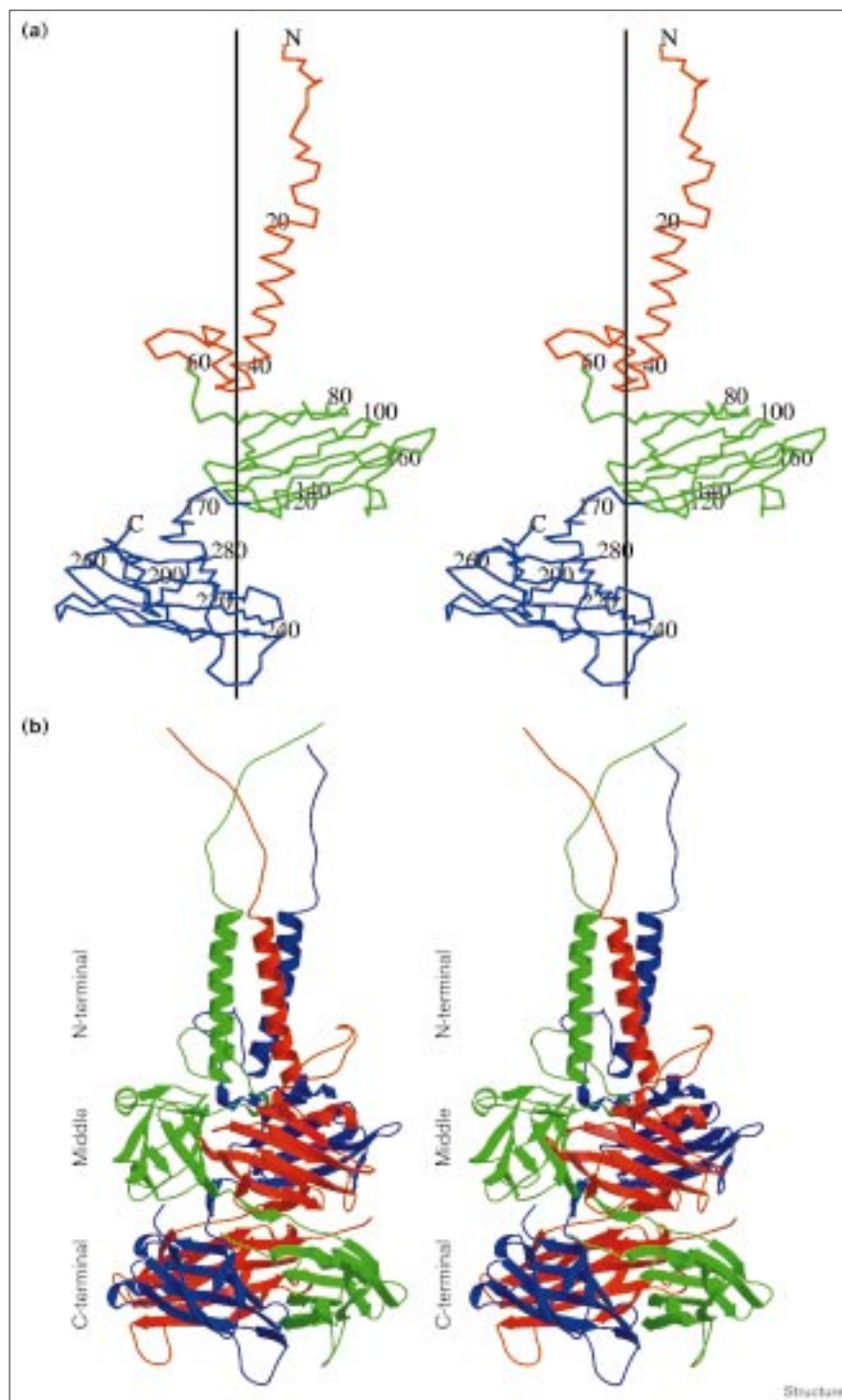
Results and discussion

The gp9 structure was determined to 2.3 Å resolution using multiple isomorphous replacement methods, single wavelength anomalous dispersion and averaging of electron density related by noncrystallographic symmetry. The gp9 monomer consists of 288 amino acid residues [16,17] and folds into three domains. It forms an elongated trimer with overall dimensions of $\sim 60 \text{ \AA} \times 60 \text{ \AA} \times 130 \text{ \AA}$ (Figure 2b). The areas of contact between adjacent monomers related by the threefold axis are 570 \AA^2 , 1480 \AA^2 and 1670 \AA^2 for the N-terminal, middle and C-terminal domains, respectively, as calculated with the ASC program [18]. These values are comparable to those between monomers in, for instance, picornaviruses [19] and suggest that the observed crystallographic trimer is of biological significance. This conclusion is reinforced by the observation that there are two independent trimers in the crystal unit cell, showing the tendency for the formation of trimers. Also, mutants of gp9 that do not form trimers are unstable and degrade quickly. It was previously suggested that gp9 usually exists in a tetrameric form [20], but this would have resulted in a symmetry mismatch with the attached trimeric LTFs.

The N-terminal domain (residues 1–60) consists of an extended polypeptide chain followed by helices $\alpha 1$ and $\alpha 2$ (Figures 2,3a). Helices $\alpha 1$ from three crystallographically related monomers form a parallel, trimeric, coiled-coil structure. Helix $\alpha 1$ has a typical amphipathic character that separates hydrophobic and polar groups to opposite sides of the helix. This is generated by a heptad repeat pattern (*abcdefg*, where *a* and *d* are the conserved hydrophobic internal residues) typical of coiled-coil structures [21] (Figure 3b). Although it is unusual to have large aromatic residues (specifically, Phe35 and Phe42) in the interior of a coiled coil, the parameters of the superhelix

Figure 2

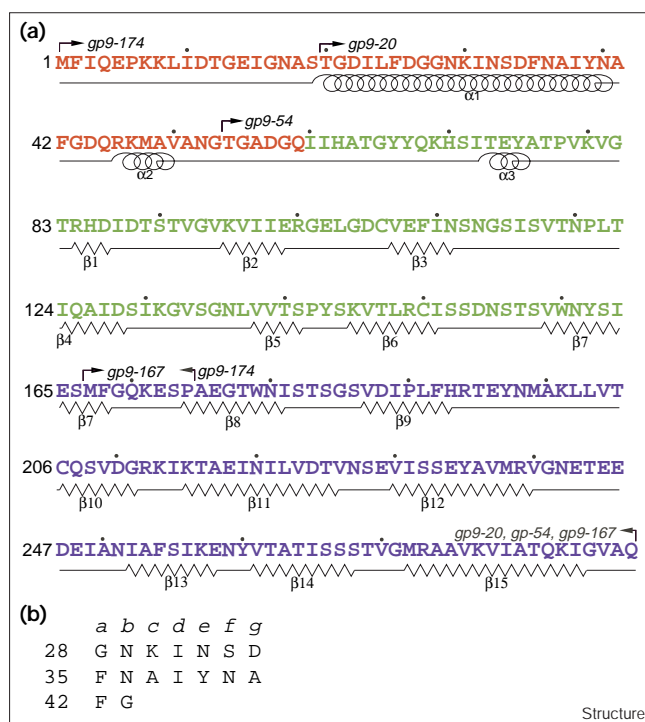
Stereoviews of the gp9 structure. (a) The gp9 C α backbone. Every 20th residue is labeled. The N-terminal, middle and C-terminal domains are colored red, green and blue, respectively. (b) Ribbon diagram of the gp9 trimer. The individual monomers are in different colors. The blue monomer is in the same orientation as the monomer shown in (a). (The programs Molscript [47] and Raster3D [48] were used to construct the picture.)



are comparable to those of coiled coils of more usual amino acid composition (Table 1). The N-terminal domain is connected to the middle domain by a long flexible region (residues 43–80) which has only a few hydrogen-bond restraints.

The middle domain (residues 61–166, strands β 1 to β 7) is a seven-stranded β sandwich and represents a unique protein fold (Figure 4a). The DALI algorithm and database of Holm and Sander [22] shows that the best comparison of the middle domain with any other structure has a

Figure 3



(a) The amino acid sequence of gp9. The secondary structure elements are indicated below the sequence entries. The domains are colored as in Figure 2a. The start and finish of mutants gp9-20, gp9-54, gp9-167 and gp9-174 are shown. (b) The gp9 sequence from residues 28 to 43 arranged as a sequence of *abcdefg* heptad repeats, consistent with the observed coiled-coil structure.

Z score of 2.1 on a scale where $Z > 2.0$ suggests a possible structural similarity. However, on manual inspection, even the best comparison did not give a credible superposition. Residues 167–173 connect the middle domain with the C-terminal domain (Figure 4b) by an extended polypeptide chain which makes hydrogen bonds with neighboring monomers to form a β annulus (Table 2; Figure 4d). The N-terminal and middle domains are on the same side of the trimer threefold axis. However, the C-terminal domain has been rotated by $\sim 120^\circ$ around the triad axis relative to the first two domains (Figure 2b). This may be because Pro174 creates a hinge between the middle and the C-terminal domains. Similar cases where proline residues redirect a polypeptide chain were noted by Bergdoll *et al.* [23]. The change in position of the C-terminal domain gives rise to a swapped domain topology [24] of the three adjacent monomers in a gp9 trimer.

The C-terminal domain (residues 167–288, strands $\beta 8$ – $\beta 15$) is an eight-stranded, antiparallel β barrel (Figure 4b) with a slight resemblance to the jelly-roll virus capsid structure found in numerous plant, animal and bacterial viruses [15,25] (the DALI algorithm *Z* score was 4.3). The usual

nomenclature for the structures of viral capsid proteins is such that the β strands are named B, C, ..., I sequentially along the polypeptide chain. This produces the opposing sheets BIDG and CHEF with the strands running antiparallel to each other within each sheet (Figure 4c). Using the same nomenclature for the C-terminal domain of gp9, the opposing sheets are represented by BIDEF and CHG. Thus, the BID and CH strands have a common topology (Figures 4b,c) with a viral capsid protein structure.

A viral capsid protein like structure, as occurs in the C-terminal domain of gp9, suggests that this domain might form a more integral part of the baseplate. Similarly, the partially coiled-coil structure of the gp9 N-terminal domain is more likely to be associated with the trimeric fiber of LTFs. The association of the LTFs with gp9 is aided by the virally coded, chaperone-like protein gp63 [11]. Possibly, the N-terminal extensions of gp9 require some special consideration to facilitate LTF binding. The C-terminal apex of the gp9 trimer has a diameter of ~ 60 Å. The charge distribution, looking along the threefold axis, shows three symmetry-related clusters of high negative charge arising from the sequence Glu-Thr-Glu-Glu-Asp-Glu in the $\beta 12$ – $\beta 13$ loop. This charge distribution might be required to bind the gp9 trimer to the baseplate. This proposed directional orientation of the gp9 trimer was confirmed by mutational analysis.

Four gp9 deletion mutants (gp9-20, gp9-54, gp9-167 and gp9-174) were constructed. The first three of these mutants lack the first 19, 53 and 166 amino acid residues, respectively (Figure 3a). Mutant gp9-20 lacks the non-helical part of the N-terminal domain. Mutant gp9-54 contains both the middle and the C-terminal domains,

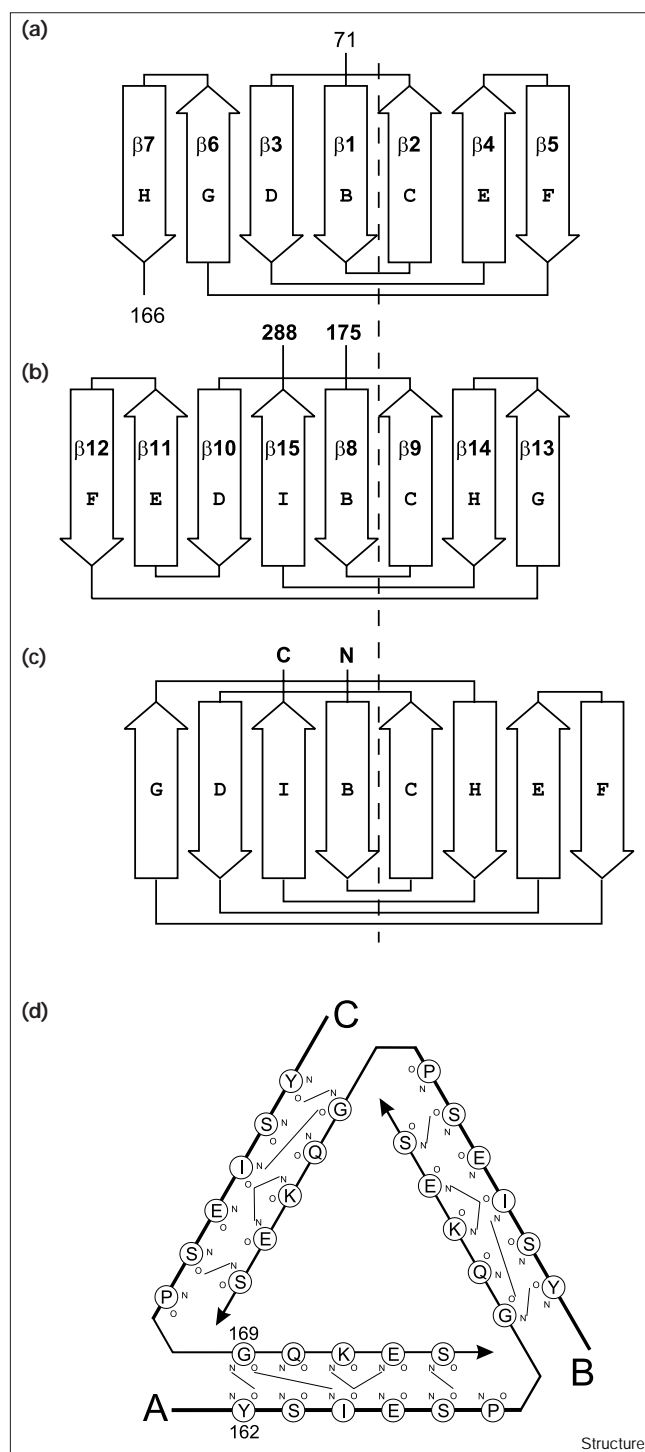
Table 1

Parameters of the coiled-coil region of gp9 and comparison with other typical triple coiled-coil structures.

Parameter	gp9	GCN4	Fibrin E	Mannose-binding protein
Radius of superhelix (Å)	6.9	7.57	6.2	6.82
Number of residues per superhelix	105.6	135.54	112.4	99.81
Superhelix pitch	153.4	200.54	165.2	143.72
Radius of curvature (Å)	93.1	142.19	114.9	83.51
Superhelix cross-angle (°)	31.6	26.67	23.4	33.22
Pairwise interhelical distance (Å)	11.9	10.70	11.0	11.80

Parameters for the leucine zipper structure of the yeast transcriptional activator GCN4 were from [44]; values for fibrin E were obtained from [45] and those for mannose-binding protein were from [46].

Figure 4



Topology diagrams for (a) the middle domain and (b) the C-terminal domain. (c) Topology of a jelly-roll fold, as occurs in many viral capsid proteins. (d) The β -annulus connecting the middle and C-terminal domains. The amino acids and hydrogen bonding between subunits A, B and C are shown.

whereas gp9-167 consists of only the C-terminal domain. Mutant gp9-174 consists of the N-terminal and middle

Table 2

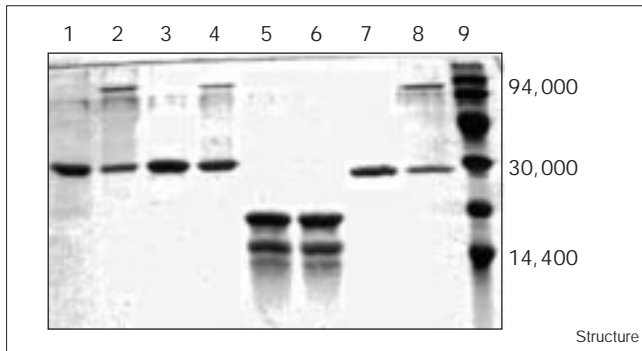
Contacts between monomers.

	Molecule 1	Molecule 2
Hydrogen bonds		
Between monomers within a trimer		
Lys69	N ζ	Leu103 O
Gly82	O	Ala63 N
Gly82	O	Thr64 N
Thr140	O γ	Glu196 O ϵ
Tyr162	O	Gly169 N
Ile164	O	Lys171 N ζ
Ser166	O γ	Ser171 N
Gln170	N	Glu176 O ϵ
Gln170	O	Lys277 N ζ
Thr178	O γ	Gly285 N
Lys201	N ζ	Glu218 O ϵ^*
Leu203	O	Gln282 N ϵ
Thr205	O γ	Gln282 O ϵ
Glu218	O ϵ	Asn220 N δ
Glu218	O ϵ	Tyr235 OH
Intertrimeric		
Arg194	NH ₂	Ser228 O γ
Asn227	O	Asn259 N δ
Ile256	O	Glu258 N
Gln125	O ϵ	Asn136 O δ
Hydrophobic interactions		
Between monomers within a trimer		
Thr20		Thr20
Leu24		Leu24
Ile31		Ile31
Phe35		Phe35
Ile38		Ile38
Ala41		Ile61
Phe42		Phe42
Gly82		His62
Cys106		Thr64
Val118		His193
Arg149		Met167
Arg149		Phe168
Asn161		Phe168
Tyr162		Phe168
Ile164		Glu172
Ile221		Met238
Leu222		Met238
Ile231		Met238
Ala280		Ile279
Ile284		Thr205
Intertrimeric		
Ala49		Val270

*Salt bridge, judged by the distance between interacting atoms.

domains, but lacks the entire C-terminal domain. Sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) of recombinant wild-type (WT) gp9 using 1% SDS (Figure 5), showed evidence of trimer formation in spite of the presence of SDS (Figure 5, lane 4). However, when gp9 is heated to 100°C for 5 min, the trimers appear to have dissociated to monomers (Figure 5, lane 3). Similar results were obtained for the three mutants with truncated N termini (Figure 5, lanes 1,2,7,8), whereas there is no indication that the gp9-174 mutant forms trimers (Figure 5, lanes 5,6), although it is unstable and degrades

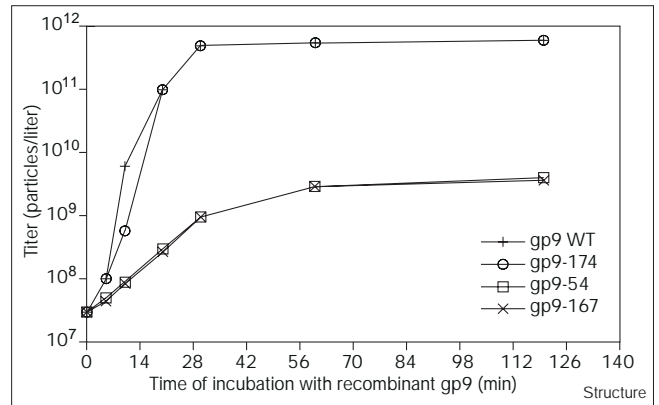
Figure 5



1% SDS-PAGE of gp9 and truncated mutants. Lane 1, gp9-20 mutant, heated to 100°C. Lane 2, gp9-20 mutant, not heated. Lane 3, wild-type gp9, heated to 100°C. Lane 4, wild-type gp9, not heated. Lane 5, gp9-174 mutant, heated to 100°C. Lane 6, gp9-174 mutant, not heated. Lane 7, gp9-54 mutant, heated to 100°C. Lane 8, gp9-54 mutant, not heated. Lane 9, molecular weight markers; numbers on the right show the molecular weight of the marker sample proteins in Daltons.

readily. The C-terminal domain appears to be essential for gp9 trimerization. Indeed, the C-terminal domain has the largest surface area of contact with its neighbors. The gp9 truncated mutants were used for an *in vitro* complementation assay of T4 gp9⁻ defective particles (Figure 6). Preincubation of the T4 gp9⁻ particles with either gp9-167 or gp9-54 mutants significantly decreases the titer of assembled infectious T4 particles. This result is consistent with an SDS-PAGE analysis (GAN and LPK unpublished data), confirming that these truncated mutants assembled as trimers and probably are able to incorporate into the baseplate. Full-sized recombinant gp9, added later, was

Figure 6

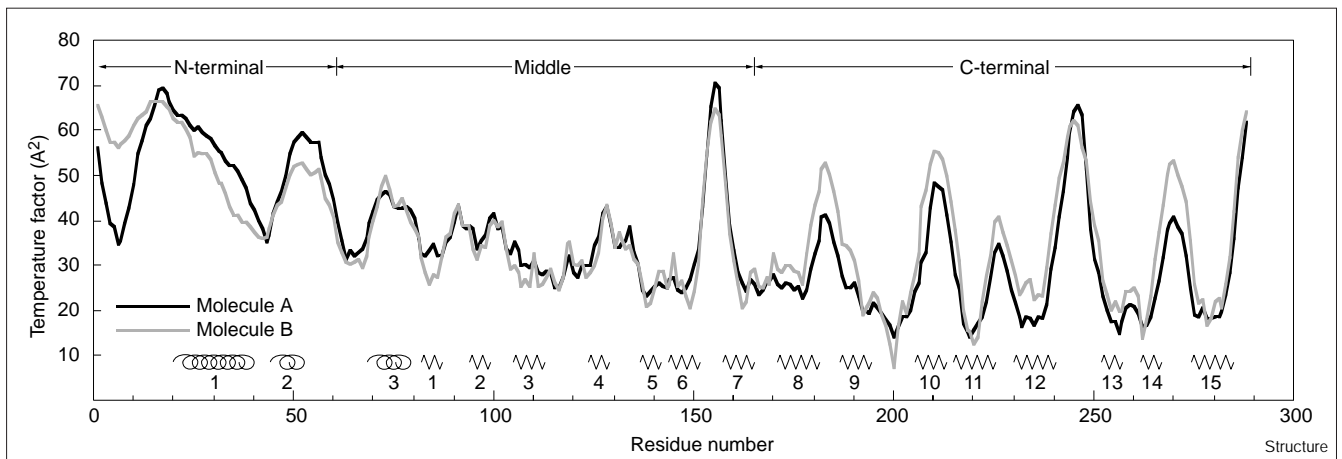


In vitro complementation assay of T4 gp9⁻ phage with different gp9 truncated mutants. All the mutants were preincubated for 30 min with gp9⁻ phage before addition of the recombinant gp9.

not able to complement baseplate, presumably because the mutant protein occupied its place on the baseplate. However, the gp9-174 mutant does not affect *in vitro* complementation, as it cannot bind to the baseplate in the absence of the C-terminal domain and does not form trimers. Also, this mutant is very unstable and degrades in several days.

One of the functions of gp9 is to allow for the up and down movements of the LTFs. The region formed by the loops between helix $\alpha 2$ and strand $\beta 1$ of the middle domain probably provides the flexibility required for this function of the gp9 trimer. In contrast, the loop $\beta 7$ - $\beta 8$

Figure 7



Temperature factor variation along the polypeptide chain for the two independent monomers of gp9 in the crystal cell. Molecule A is shown in black and molecule B in gray. Secondary structural elements are shown below the graph: coils represent α helices and zigzag lines

represent β strands. Note the decrease in temperature factor for the middle and C-terminal domains as a result of more intermolecular lattice contacts for these domains. The regions of high temperature factor correspond to loops between β strands.

between the middle and C-terminal domains has extensive hydrogen bonding forming a rather rigid β annulus (Figure 4d). The assignment of flexibility and rigidity in these respective loops is consistent with the crystallographic temperature factors (Figure 7). The magnitude of the temperature factors decreases gradually from the N terminus to the C terminus with the exception of several flexible loop regions. The middle and C-terminal domains participate in crystal packing and thus are less mobile than the unrestricted N-terminal domain.

Another function of gp9 is to act as a molecular trigger that controls the hexagon-to-star reorganization of the T4 baseplate in the initial stages of *E. coli* infection. The conformational change induced in the LTFs by interactions with their receptors might alter the gp9 conformation, thus activating the baseplate transition. The attachment of at least three LTFs is required to activate the baseplate [9], thus restraining the movement of the T4 particle to be only perpendicular to the cell surface. The strain induced on gp9 might destabilize the wedge, causing the baseplate to alter its conformation from a hexagonal to a star-like shape.

Biological implications

The baseplate of bacteriophage T4 is the control center of viral infection, communicating between the tail fibers, which sense host-cell receptors, and the tail sheath, which contracts and ejects the viral DNA into the host. Gene product 9 (gp9) forms the attachment site for a long tail fiber on the baseplate. Thus, gp9 is required to carry the message from the long tail fibers to initiate a major conformational transition in the baseplate from a hexagonal to a star-like shape, which then triggers the tail contraction.

The quaternary structure of gp9 was found to be a trimer. The intimate structural association of the monomers within the trimer (total buried surface area 3720 Å²), the presence of two closely similar but independent trimers in the crystal cell, and mutational data show that the observed trimers are biologically significant. This contradicts earlier data [20] suggesting that there are four gp9 monomers per wedge in the baseplate. The gp9 polypeptide chain folds into three domains. The N-terminal domain forms an α -helical, triple coiled-coil. The other two domains are β sandwiches with the C-terminal domain having some resemblance to the eight-stranded, antiparallel fold (jelly roll) found in many viral capsid proteins. Whether this is the result of evolutionary divergence from a common primordial viral capsid fold is uncertain. However, the structure does suggest that the C-terminal domain is part of the baseplate, whereas the N-terminal domain is associated with the trimeric long tail fibers. This assembly organization has been confirmed by mutagenesis studies of gp9 in which different domains have been deleted.

The function of gp9 to transmit signals from the long tail fibers may be helped by the flexible hinge between the N-terminal and middle domains. An asymmetric impulse from the long tail fibers might reorient the coiled coil of the N-terminal domains with respect to the common threefold axis of the other two domains, setting up a strain within the middle and C-terminal domains. This might be a conformational switch that directs further steps of the infection process.

The structure of gp9 has also been a stimulus for mutational studies aimed at determining the mode of association of gp9 trimers into T4 baseplates, and provides a basis for further mutational studies of folding and oligomerization.

Materials and methods

Gene expression and protein purification

The *E. coli* BL21 (DE3) strain (Novagen) was used for gp9 expression under the control of the T7 phage promoter [26]. CR63 and Be/1 strains were used for *in vitro* complementation assays. Gene 9 and its fragments were amplified by 30 cycles of the polymerase chain reaction (PCR) using Vent polymerase (New England Biolabs), T4 genomic DNA as a template, and oligonucleotide primers to generate the *Nco*I–*Bam*HI cloning sites. Three gp9-truncated mutants (gp9-20, gp9-54 and gp9-167) were constructed which lack the N-terminal 19, 53 and 166 residues, respectively. Another gp9 mutant (gp9-174) was constructed that lacks the entire C-terminal domain. The fragments with N-terminal truncations were cloned into the pET-23d expression vector (Novagen), and the mutant with the C-terminal truncation was cloned into the pET-22b(+) expression vector. All expression vectors were verified by DNA sequencing. To induce expression of gp9, isopropylthio- β -D-galactoside was added to a final concentration of 1 mM. After induction, the cells were grown for 3 h at 37°C. Cells were pelleted by centrifugation, resuspended in 0.05 M Tris-HCl buffer pH 7.8, and disrupted by sonication, as described for the expression and purification of bacteriophage T4 fibrin [27]. Polyethylenimine (10%) and 4 M NaCl were added to the supernatant to a final concentration of 0.2% and 0.2 M, respectively. The mixture was then incubated on ice for 1 h, with subsequent centrifugation for 10 min at 21,000 g. A saturated solution of ammonium sulfate was added slowly to the supernatant to 25% saturation and incubated on ice for 1 h. After centrifugation, the pellet was dissolved in 8–10 ml of 0.05 M Tris-HCl buffer pH 7.8, applied to a hydroxyapatite column (Bio-Rad), and eluted with 10 mM sodium phosphate buffer pH 7.8. The relevant fractions were analyzed by SDS–PAGE according to the method of Laemmli [28].

In vitro complementation assay

Cell extract containing T4 gp9⁻ particles was prepared using the gene 9⁻ C215 am-mutant, containing a nonsense codon for Gln282 that terminates mRNA translation (VVM, unpublished results), and the non-suppressive Be/1 *E. coli* strain as described by Edgar & Wood [29]. Cell extract was divided into 0.2 ml portions and stored at –70°C. For *in vitro* complementation, the extract was thawed at room temperature and 0.02 ml of the protein solution (or cell extract containing recombinant protein) was added and incubated for 2 h at 30°C. The mixture was diluted with 0.14 M NaCl and then plated on permissive *E. coli* CR63 (Sull suppressor) cells.

Crystallization

The crystallization of gp9 has been described previously [30]. Other, better diffracting crystals were obtained using the Crystal Screen Kit (Hampton Research). Initial crystals were obtained in hanging drops that contained equal volumes of 20 mg/ml protein and 28% v/v polyethyleneglycol 400, 0.2 M CaCl₂ in 0.1 M Hepes-Na buffer pH 7.5 as the precipitant. Crystals of 0.3–0.5 mm in diameter were obtained by

Table 3

Data statistics.							
Data set	Resolution limits (Å)	Completeness (%) [*]	R _{merge} ^{*†}	R _{diff} [‡]	Number of sites	Phasing power [§]	R _{Cullis} [#]
Native	40.0–2.3	97.5 (96.3)	0.07 (0.17)	–	–	–	–
HgCl ₂	40.0–2.5	99.8 (98.3)	0.08 (0.24)	0.32	4	1.55	0.73
HgEtPO ₄	40.0–2.5	99.8 (98.3)	0.08 (0.27)	0.40	4	1.41	0.74

^{*}Numbers in parentheses are for the highest resolution shell. [†]R_{merge} = $\sum_h \sum_i |I_{hi} - I_h| / \sum_h \sum_i I_{hi}$. [‡]R_{diff} = $\sum_h |I_{h,heavy} - I_{h,native}| / \sum_h I_{h,heavy}$.

[§]Phasing power = $\sum |F_H| / \sum |F_{PH} - |F_P + F_H||$. [#]R_{Cullis} = $\sum |F_{PH} - |F_P + F_H|| / \sum |F_{PH} - F_P|$.

the streak-seeding technique [31] in pre-equilibrated drops of the same composition; these crystals were used in the structure determination.

Data collection

Crystals were flash-frozen at 110K using the mother liquor as the cryo-protectant. Diffraction from the frozen crystals extended to 1.8 Å resolution, although the data collected were limited to 2.3 Å resolution (Table 3). All data were collected on a Rigaku R-Axis IIC image-plate detector mounted on a Rigaku rotating-anode X-ray source. Each data set was collected from a single crystal. The diffraction data were processed using the DENZO program [32]. The native data set was scaled and post-refined with the SNP program [33] and the heavy-atom derivative data were scaled and post-refined with the SCALEPACK program [32]. The crystals were in space group R32, with cell dimensions of the native crystal being a = 94.33 Å, c = 440.94 Å (in the hexagonal setting). There are two molecules per asymmetric unit, with a Matthews' coefficient [34], V_M, of 3.1 Å³/Da. Two heavy-atom derivatives were used in the structure determination. These were produced by soaking crystals in 10% saturation of ethyl mercury phosphate (C₂H₅PO₄Hg) in mother liquor for 2 h and by soaking in a 20 mM solution of mercury chloride (HgCl₂) in mother liquor for seven days, respectively.

Phasing, model building and refinement

The structure was solved using multiple isomorphous replacement with anomalous scattering methods. Mercury-binding sites of the derivatives were found using difference Patterson maps and checked with Bijvoet difference Patterson maps as well as by feedback experiments. Heavy-atom site parameters were refined with the MLPHARE program [35] using both isomorphous and anomalous scattering data between 10.0 and 3.0 Å resolution (Table 4). Phasing was then extended to 2.5 Å resolution. A Patterson map of the native compound has a very large peak, about half of the origin in height, at (0,0,0.4441). This implies that

there should be a noncrystallographic twofold axis parallel to a crystallographic axis, say, (x = y, z = 0), showing that the two independent molecules in the unit cell must be related by approximately (x, y, z) and (y, x, 0.4441–z) noncrystallographic symmetry (NCS). The NCS relation was subsequently confirmed by interpretation of preliminary electron-density maps and by positions of the mercury atoms. Phase information generated by MLPHARE was improved by solvent flattening and NCS averaging with the DM [36] program, helped by the programs MAPMAN [37], O [38], BONES2PDB and NCSMASK [39], which were used for mask generation. The NCS operation was refined by DM. The resultant averaged electron-density map made it possible to trace ~80% of the amino acid residues and to fit them to the sequence of gp9. The initial model was fitted in the electron-density map with the help of the program O. The remaining portions of the molecule were built with respect to an averaged (2F_o–F_c) map. The CNS program [40] was used for crystallographic refinement, with the maximum-likelihood target function [41]. The quality of the model was monitored by R_{working} and R_{free} [42,43]. A set of observed reflections representing 5% of all the data was used to determine R_{free}. These reflections were chosen with the program xIDATAMAN [37] and were omitted from all refinement procedures, but were included in map calculations. Initially, the NCS constraints were used to refine the model against the 8.0–2.5 Å resolution data. Temperature factors were refined with restraints on bond lengths and angles.

After initial positional and temperature-factor refinement, a (2F_o–F_c) map was used for model rebuilding. The resolution was extended to include data between 40.0 and 2.3 Å resolution, bulk-solvent correction was applied, and no NCS restraints were used. After several cycles of positional (conjugated gradient energy minimization) and temperature-factor refinement and further manual rebuilding, the R_{working} and R_{free} values were 0.30 and 0.32, respectively. Removal of the NCS restraints substantially reduced the R_{free}.

Table 4

Heavy-atom parameters.				
Site	1	2 [†]	3	2 [†]
Compound	C ₂ H ₅ HgPO ₄	C ₂ H ₅ HgPO ₄	HgCl ₂	HgCl ₂
Coordinates	(0.196, 0.062, 0.047)	(0.082, 0.022, 0.119)	(0.465, 0.101, 0.008)	(0.112, 0.024, 0.126)
Molecule 1				
occupancy [*]	0.98 (0.62)	0.29 (0.25)	0.82 (0.44)	0.23 (0.20)
B factor (Å ²)	9	35	18	22
Molecule 2				
occupancy [*]	0.67 (0.53)	0.18 (0.17)	0.51 (0.32)	0.27 (0.22)
B factor (Å ²)	10	14	39	28
Ligand residue	Cys206	Cys106	Cys150	Cys106

^{*}Occupancies determined from anomalous dispersion data are in parentheses. [†]Site 2 is the same site for the two heavy-atom compounds.

Water molecules were then assigned to peaks that were higher than 1.2 rms deviations from zero density in a ($2F_o - F_c$) map, higher than 2.5 rms deviations from zero density in a ($F_o - F_c$) map, and in positions suitable for hydrogen bonding with amino acid residues of the two molecules. Water molecules with temperature factors greater than 60 \AA^2 after refinement were rejected. The remaining water molecules were examined manually. This procedure resulted in 292 water and two Hepes (4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid) molecules in the asymmetric unit. All positions and temperature factors were then refined again, giving the final R_{working} and R_{free} of 0.23 and 0.27. The final rms deviations from idealized bond lengths and angles were 0.007 \AA and 1.3° , respectively. A Ramachandran plot showed that 97.9% of the residues were in the most favored and additionally allowed regions, whereas 1.4% were in the disallowed region.

Accession numbers

The refined coordinates and structure amplitudes have been deposited in the Protein Data Bank with accession numbers 1QEX and R1QEXSF.

Acknowledgements

This work was supported by a grant from the National Science Foundation (MCB-9603571) to MGR, by a grant from the Russian Foundation of Basic Research (99-04-48430) and from the 'Universities of Russia' program (2608) to VVM. VVM and MGR are recipients of an International Howard Hughes Medical Institute grant (75195-520803).

References

- Eiserling, F.A. & Black, L.W. (1994). Pathways in T4 morphogenesis. In *Molecular Biology of Bacteriophage T4*. (Karam, J.D., ed.), pp. 209-212. American Society for Microbiology, Washington, DC.
- Kutter, E., *et al.* & Mosiz, G. (1993). Genomic map of bacteriophage T4. In *Genomic Maps*. (O'Brien, S.J., ed.), pp. 1-27. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Mellado, R.P., *et al.*, & Anderson, D.L. (1976). Genetic analysis of bacteriophage $\phi 29$ of *Bacillus subtilis*: integration and mapping of reference mutants of two collections. *J. Virol.* **19**, 495-500.
- Kikuchi, Y. & King, J. (1975). Genetic control of bacteriophage T4 baseplate morphogenesis. III. Formation of the central plug and overall assembly pathway. *J. Mol. Biol.* **99**, 695-716.
- Crowther, R.A., Lenk, E.V., Kikuchi, Y. & King, J. (1977). Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. *J. Mol. Biol.* **116**, 489-523.
- Watts, N.R. & Coombs, D.H. (1990). Structure of the bacteriophage T4 baseplate as determined by chemical cross-linking. *J. Virol.* **64**, 143-54.
- Cerritelli, M.E., Wall, J.S., Simon, M.N., Conway, J.F. & Steven, A.C. (1996). Stoichiometry and domain organization of the long tail-fiber of bacteriophage T4: a hinged viral adhesin. *J. Mol. Biol.* **260**, 767-780.
- Makhov, A.M., *et al.*, & Steven, A.C. (1993). The short tail-fiber of bacteriophage T4: molecular structure and a mechanism for its conformational transition. *Virology* **194**, 117-127.
- Goldberg, E., Grinius, L. & Letellier, L. (1994). Recognition, attachment, and injection. In *Molecular Biology of Bacteriophage T4*. (Karam, J.D., ed.), pp. 347-356. American Society for Microbiology, Washington, DC.
- Kellenberger, E., *et al.*, & Lauffer, M.A. (1965). Function and properties related to the tail fibers of bacteriophage T4. *Virology* **26**, 419-440.
- Wood, W.B., Eiserling, F.F. & Crowther, R.F. (1994). Long tail fibers: genes, proteins, structure, and assembly. In *Molecular Biology of Bacteriophage T4*. (Karam, J.D., ed.), pp. 282-290. American Society for Microbiology, Washington, DC.
- Kellenberger, E., Stauffer, E., Haner, M., Lustig, A. & Karamata, D. (1996). Mechanism of the long tail-fiber deployment of bacteriophages T-even and its role in adsorption, infection and sedimentation. *Biophys. Chem.* **59**, 41-59.
- Crowther, R.A. (1980). Mutants of bacteriophage T4 that produce infective fibreless particles. *J. Mol. Biol.* **137**, 159-174.
- Simon, L.D., Swan, J.G. & Flatgaard, J.E. (1970). Functional defects in T4 bacteriophages lacking the gene 11 and gene 12 products. *Virology* **41**, 77-90.
- Liljas, L. (1999). Virus assembly. *Curr. Opin. Struct. Biol.* **9**, 129-134.
- Selivanov, N.A., Prilipov, A.G., Efimov, V.P., Marusich, E.I. & Mesyanzhinov, V.V. (1990). Cascade of overlapping late genes in bacteriophage T4. *Biomed. Sci.* **1**, 55-62.
- Prilipov, A.G., Selivanov, N.A., Efimov, V.P., Marusich, E.I. & Mesyanzhinov, V.V. (1989). Nucleotide sequences of bacteriophage T4 genes 9, 10 and 11. *Nucleic Acids Res.* **17**, 3303.
- Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C. & Scharf, M. (1995). The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.* **16**, 273-284.
- Arnold, E. & Rossmann, M.G. (1990). Analysis of the structure of a common cold virus, human rhinovirus 14, refined at a resolution of 3.0 \AA . *J. Mol. Biol.* **211**, 763-801.
- Kikuchi, Y. & King, J. (1975). Genetic control of bacteriophage T4 baseplate morphogenesis. I. Sequential assembly of the major precursor, *in vivo* and *in vitro*. *J. Mol. Biol.* **99**, 645-672.
- Lupas, A. (1996). Coiled coils: new structures and new functions. *Trends Biochem. Sci.* **21**, 375-82.
- Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.
- Bergdoll, M., Remy, M.H., Cagnon, C., Masson, J.M. & Dumas, P. (1997). Proline-dependent oligomerization with arm exchange. *Structure* **5**, 391-401.
- Bennett, M.J., Schlunegger, M.P. & Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci.* **4**, 2455-2468.
- Rossmann, M.G. & Johnson, J.E. (1989). Icosahedral RNA virus structure. *Annu. Rev. Biochem.* **58**, 533-573.
- Studier, F.W., Rosenberg, A.N., Dunn, A.H. & Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Meth. Enzymol.* **185**, 60-89.
- Efimov, V.P., *et al.*, & Mesyanzhinov, V.V. (1994). Fibrin encoded by bacteriophage T4 gene *wac* has a parallel triple-stranded α -helical coiled-coil structure. *J. Mol. Biol.* **242**, 470-486.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.
- Edgar, R.S. & Wood, W.B. (1966). Morphogenesis of bacteriophage T4 in extracts of mutant-infected cells. *Proc. Natl Acad. Sci. USA* **55**, 498-505.
- Strelkov, S.V., *et al.*, & Mesyanzhinov, V.V. (1993). Crystallization and preliminary crystallographic characterization of bacteriophage T4 baseplate protein encoded by gene 9. *J. Mol. Biol.* **234**, 493-495.
- Ducruix, A. & Giege, R. (1992). *Crystallization of Nucleic Acids and Proteins: A Practical Approach*. Oxford University Press, New York.
- Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
- Bolotovskiy, R., Steller, I. & Rossmann, M.G. (1998). The use of partial reflections for scaling and averaging X-ray area detector data. *J. Appl. Crystallogr.* **31**, 708-717.
- Matthews, B.W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.
- Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In *Isomorphous Replacement and Anomalous Scattering Proceedings of the CCP4 Study Weekend, 25-26 January 1991*. (Wolf, W., Evans, P.R. & Leslie, A.G.W., eds), pp. 80-86. Science and Engineering Research Council, Daresbury, UK.
- Cowtan, K.D. (1994). 'DM': an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newslitt. Protein Crystallogr.* **31**, 34-38.
- Kleywegt, G.J. & Jones, T.A. (1996). *xdMAPMAN* and *xdIDATAMAN* - programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr. D* **52**, 826-828.
- Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110-119.
- Collaborative Computational Project Number 4 (1994). The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763.
- Brünger, A.T., *et al.*, & Warren, G.L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905-921.
- Adams, P.D., Pannu, N.S., Read, R.J. & Brünger, A.T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl Acad. Sci. USA* **94**, 5018-5023.
- Brünger, A.T. (1992). Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.
- Kleywegt, G.J. & Brünger, A.T. (1996). Checking your imagination: applications of the free *R* value. *Structure* **4**, 897-904.

44. Gonzalez Jr., L., Brown, R.A., Richardson, D. & Alber, T. (1996). Crystal structures of a single coiled-coil peptide in two oligomeric states reveal the basis for structural polymorphism. *Nat. Struct. Biol.* **3**, 1002-1010.
45. Tao, Y., Strelkov, S.V., Mesyanzhinov, V.V. & Rossmann, M.G. (1997). Structure of bacteriophage T4 fibrin: a segmented coiled coil and the role of the C-terminal domain. *Structure* **5**, 789-798.
46. Sheriff, S., Chang, C.Y. & Ezekowitz, R.A. (1994). Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple α -helical coiled-coil. *Nat. Struct. Biol.* **1**, 789-793. [Published erratum appears in *Nat. Struct. Biol.* (1996), **3**, 103].
47. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.
48. Merrit, E.A. & Bacon, D.J. (1997). Raster3D: a photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.

Because *Structure with Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed (accessed from <http://biomednet.com/cbiology/str>). For further information, see the explanation on the contents page.