Article No. jmbi.1999.2720 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 288, 649–657

JMB

Crystal Structure of the Two N-terminal Domains of g3p from Filamentous Phage fd at 1.9 Å: Evidence for Conformational Lability

Philipp Holliger[1]*, Lutz Riechmann[2] and Roger L. Williams[2]

[1]*MRC Centre for Protein Engineering and*

[2]*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK*

Infection of *Escherichia coli* by filamentous bacteriophages is mediated by the minor phage coat protein g3p and involves two distinct cellular receptors, the F′ pilus and the periplasmic protein TolA. Recently we have shown that the two receptors are contacted in a sequential manner, such that binding of TolA by the N-terminal domain g3p-D1 is conditional on a primary interaction of the second g3p domain D2 with the F′ pilus. In order to better understand this process, we have solved the crystal structure of the g3p-D1D2 fragment (residues 2-217) from filamentous phage fd to 1.9 Å resolution and compared it to the recently published structure of the same fragment from the related Ff phage M13. While the structure of individual domains D1 and D2 of the two phages are very similar (rms < 0.7 Å), there is comparatively poor agreement for the overall D1D2 structure (rms > 1.2 Å). This is due to an apparent movement of domain D2 with respect to D1, which results in a widening of the inter-domain groove compared to the structure of the homologous M13 protein. The movement of D2 can be described as a rigid-body rotation around a hinge located at the end of a short anti-parallel beta-sheet connecting domains D1 and D2. Structural flexibility of at least parts of the D1D2 structure was also suggested by studying the thermal unfolding of g3p: the TolA binding site on D1, while fully blocked by D2 at 37 °C, becomes accessible after incubation at temperatures as low as 45 °C. Our results support a model for the early steps of phage infection whereby exposure of the coreceptor binding site on D1 is facilitated by a conformational change in the D1D2 structure, which *in vivo* is induced by binding to the F′ pilus on the host cell and which can be mimicked *in vitro* by thermal unfolding.

Keywords: g3p; phage infection; phage display; conformational change; TolA

*Corresponding author

Introduction

Filamentous bacteriophages (Inoviridae) are a family of single-stranded DNA viruses that infect Gram-negative bacteria. The Ff filamentous bacteriophages (f1, fd, M13) infect *Escherichia coli* bearing an F′ pilus (Webster, 1996; Model & Russel, 1988; Marvin, 1998). Productive infection further requires three products of the tol operon, TolA, TolR and TolQ (Sun & Webster, 1986; Webster, 1991). Infection is mediated by the minor coat protein g3p located at the distal tip of the extended phage capsid. The g3p has a modular structure, comprising three distinct domains (D1, D2 and D3) separated by glycine-rich tetra- and pentapeptide repeats as well as a C-terminal transmembrane segment. Separate domains appear to encode separate functions. Relatively little is known about the functions of the C-terminal domain D3 and the transmembrane segment apart from their requirement for g3p incorporation into the phage particle (Stengele *et al.*, 1990). The role of the two N-terminal domains, D1 and D2, in the infection process is better understood. Phage infection begins with the interaction of the middle domain D2 with the primary infection receptor for Ff phages, the tip of the F′ pilus. Attachment of the phage causes pilus retraction (Jacobson, 1972) by an unknown

E-mail address of the corresponding author: ph1@mrc-lmb.cam.ac.uk

0022-2836/99/190649–9 $30.00/0

mechanism, bringing the phage particle into close proximity of the *E. coli* host cell. Productive infection requires, in the next step, the interaction of the N-terminal domain g3p-D1 with the periplasmic protein TolA (Levengood *et al.*, 1991), the obligatory coreceptor of phage infection. In the absence of pilus, the binding of g3p-D1 to TolA is blocked by g3p-D2. This blockage is released in the presence of pilus (Riechmann & Holliger, 1997; Click & Webster, 1997). Presumably, a conformational change in g3p is taking place, which is triggered directly by an interaction of g3p with the pilus primary receptor. To better understand the molecular events during phage infection, we expressed g3p-D1D2 (residues 2-217) from filamentous phage fd as a recombinant protein in *E. coli*, prepared crystals and solved its structure by X-ray crystallography to 1.9 Å resolution.

Results and Discussion

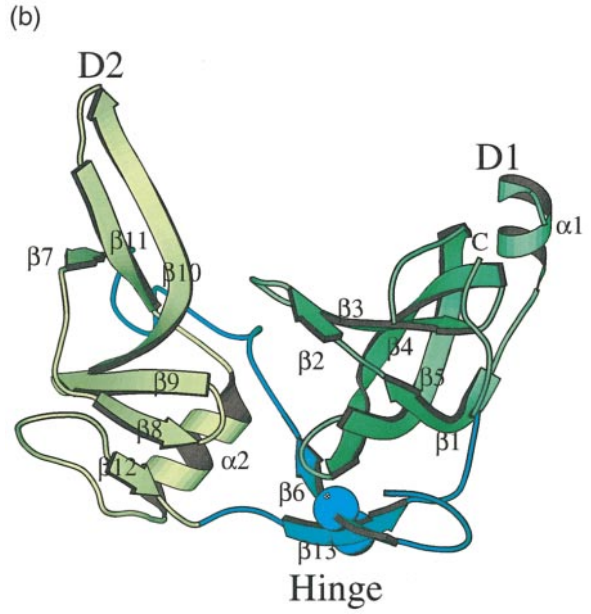
The sequence and overall structure of g3p-D1D2 from phage fd (D1D2-fd) is shown in Figure 1(a) to (c). The D1D2-fd structure was solved using multiple isomorphous replacement (MIR) phasing with two heavy-metal derivatives. Details of crystallisation and structure determination are summarised in Table 1. The asymmetric unit of the crystal consists of two molecules of D1D2-fd (referred to as A and B). Molecules A and B are related to each other by a 2-fold axis. The surface area buried in the packing of A and B in the unit cell is extensive and buries a total of 2200 Å^2 (1100 Å^2 per molecule). Interaction interfaces of this size are typical of biologically relevant interactions (Janin & Chothia, 1990; LoConte *et al.*, 1998). However, we could find no evidence of dimer formation of D1D2-fd, as judged by gel-filtration chromatography or light-scattering at protein concentrations (10 mg/ml) close to those used in the crystallisation set-up. Dimer formation may therefore be an artifact of the high-salt crystallisation conditions and may not reflect an interaction of biological relevance. Indeed, no such dimers are observed in the crystal packing of the D1D2 fragment from the closely related Ff phage M13 (D1D2-M13; Lubkowski *et al.*, 1997), which crystallised under low-salt conditions.

The loop from residue G158 to V162 is not resolved in molecule B only. In molecule B, this region is free from any crystal contacts, while in molecule A it is packed against the end of helix α2. This interaction appears to better order this region (G158-V162) in molecule A. The His_6 termini in both A and B are clearly visible up to the fifth histidine residue (H223) in the electron density maps and have been incorporated into the model. In contrast, the Gly-rich linker segment from residue G69-T87 as well as adjacent residues N67-E68 and K88-P90 are not visible in either molecule A or B (Figure 1(a)). This linker region is also not visible in the D1D2-M13 structure and thus represents a *bona fide* flexible element of g3p structure. No other breaks in molecules A or B are apparent (when contoured at 1σ).

The D1D2 fragment of g3p from phage fd (D1D2-fd) folds into a horseshoe-shaped structure with a hydrophobic groove running between the two lobes. The N-terminal domain D1 forms the majority of one lobe. This lobe is formed by a seven-stranded anti-parallel β-sheet. Five of the strands (β1-β5) from D1 form a β-barrel against which the remaining two strands of the sheet (β6 and β13) from the second domain, D2, are packed. The Gly-rich linker connecting D1 to D2 is not visible in the electron density map (see above). The remainder of D2 forms the other lobe comprising a six-stranded mixed β-sheet folded into another β-barrel. Both domains also contain a single α-helix packed onto the outside of the β-barrel. In the D1 domain, the helix is located at the N terminus and packs against the ends of strands β4 and β5. The α-helix in the D2 domain lies between strands β11 and β12, and packs across the ends of strands β9, β8 and β12. Several features result in distortion of the β-sheets of the two domains. Domain D1 has two intra-domain disulphide bonds, Cys7-Cys36 and Cys46-Cys53. One β-bulge within D1 (Gly42/Val43/Gly55), results in a sharp bend of strand β4, while a second bulge (Leu37/Ile60/Gly61) results in a turn of the polypeptide chain at the end of strand β5, allowing the C-terminal residues of D1 to cover a hydrophobic face of its β-sheet. A β-bulge in the D2 domain (Gln167/Thr152/Gly153) produces a sharp bend in the middle of strand β2. There is a single disulphide bridge, Cys188-Cys201, within D2. The structure also contains two *cis*-proline residues, Pro161 and Pro213.

The overall C^α trace of the individual domains D1 and D2 of the phage fd D1D2 fragment (D1D2-fd) is very similar to that of phage M13 (D1D2-M13; Lubkowski *et al.*, 1997: see Figure 2), as expected from the high degree of sequence similarity (D1D2-fd differs from D1D2-M13 in just two residues; see below). The root-mean-square (rms) deviation for the C^α atoms between the individual domains within the two structures is less than 0.7 Å for residues 22 to 99 (mainly D1) and 0.65 Å for residues 101 to 191 (D2).

Despite the high degree of similarity, there are some localised regions within the domains where the main-chain conformation differs extensively between the two structures. In D1, the main difference involves the region from residues A40 to V44. In particular, G42 in the centre of the region adopts a conformation very different in D1D2-fd as compared to the D1D2-M13 structure (Figure 2). The structure of the isolated D1-fd domain (residues 2-67) has also been solved by heteronuclear NMR (Holliger & Riechmann, 1997). Poor NOE restraints were obtained for the region from residues A40 to G42 in the NMR structure, suggesting that this region may be an inherently flexible segment of the D1 structure. There are only two positions where the D1D2-fd sequence differs from the M13



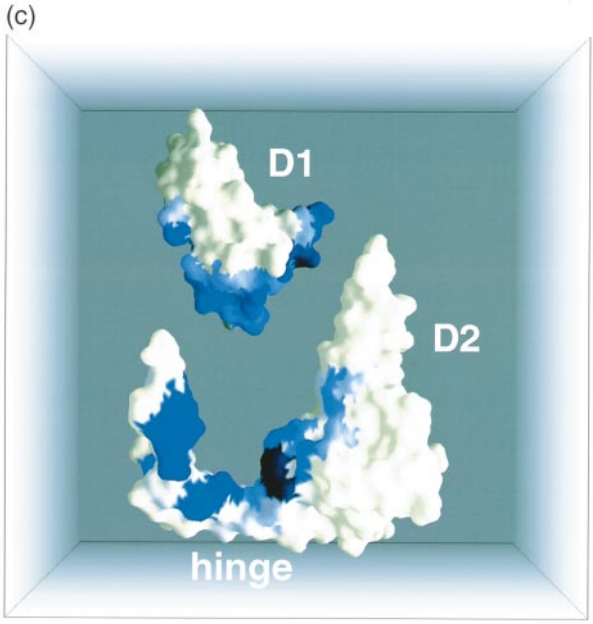


Figure 1. (a) Sequence of the D1D2-fd. The sequence is overlaid with secondary structure elements. Rounded boxes and arrows represent α-helices and β-sheets, respectively. Residues visible in the electron density map are in bold print. The hinge domain region is shown in italic print and underlined by a broken line. Hinge residues G99 and S208 are highlighted. Circles underneath residues indicate that residues are buried in the D1D2 structure but not in isolated domains. Shading reflects degree of burial of residues in the D1, D2 interface (see (c)), white reflects the lowest degree, grey an intermediate and black the highest degree of burial. Domains D1 and D2* are shown within square brackets. (b) Ribbon representation of D1D2 fragment from phage fd (D1D2-fd): domain D1 (2-66) is coloured yellow, the homologous β-barrel core of D2 (D2*) is coloured green, and the proposed hinge sub-domain (HD) including strands β6 and β13, is coloured cyan. Secondary structure elements are labelled throughout. Hinge residues G99 and S208 are shown as spheres. (c) Space-filling representation of D1D2-fd. D1 (2-66) is moved away from the rest of the molecule. The colour coding for residues represents the degree of burial as measured by the difference in solvent accessibility (using a probe sphere of 1.4 Å) in the individual D1 and D2 domains and in the D1D2-fd. Degree of burial varies from white (fully exposed) over blue to black (completely buried).

Table 1. Statistics for crystallographic data collection and phase refinement

Data set	X-ray source	Wavelength (Å)	d_{min} (Å)	Measurements	Unique reflections	Completeness (%)	R_{merge} (%)
Native	ESRF ID14-3	0.947	1.8	343,352	58,927	99.8	7.9
K_2PtCl_4	CuKα	1.54	3.0	41,141	12,466	98.4	8.7
PIP	CuKα	1.54	3.0	42,257	12,424	99.0	4.9
SeMet	ESRF ID14-3	0.947	2.8	67,448	16,089	99.1	7.6

MIR Phase determination

					Phasing power		
					Isomorphous		Anomalous
Data set	R_{iso} (%)	Sites	R_{Cullis}	Resolution (Å)	Centric	Acentric	
K_2PtCl_4	0.33	3	0.91	3.2	0.98	1.10	1.1
PIP	0.21	8	0.97	3.0	0.81	0.78	0.78
SeMet	0.39	6	1.04	2.8	0.53	0.48	0.59

$R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$
$R_{iso} = \Sigma ||F_{deriv}| - |F_{native}|| / \Sigma |F_{native}|$.
$R_{Cullis} = \Sigma ||F_{PH} \pm F_P| - F_{H(calc)}| / \Sigma |F_{PH} \pm F_P|$, shown for isomorphous, acentric differences.

sequence (S11P and L198P). Despite the fact that both substitutions introduce a proline, only the L198P substitution appears to affect the main chain conformation: the loop from residue S192 to F199 in D1D2-fd has a conformation very different than the same loop in D1D2-M13 (Figure 2) and this is likely to be caused by the L198P substitution. However, crystal packing interactions in this region, while limited, are more extensive in D1D2-fd than in D1D2-M13 and crystal packing can therefore not be ruled out as a possible explanation for the altered loop conformation.

In the D1D2-M13 structure (Lubkowski *et al.*, 1997) residue 21 was assigned to be oxidised tryptophan (1,3-dihydro-indole-2-one, W_{ox}), because the electron density of the indole ring was inconsistent with an sp^2 hybridisation state. This unusual modification had previously not been observed in other structures and is also not found in our structure. The electron density in D1D2-fd agrees with an unmodified tryptophan residue at position 21. Indeed, the ^{1}H-chemical shifts as determined by heteronuclear NMR for the D1-fd are also consistent with an unmodified tryptophan residue at position 21 (Holliger & Riechmann, 1997). In the absence of other g3p structures, it is difficult to decide if the W_{ox} in D1D2-M13 is a natural post-translational modification unique to the M13 protein and therefore a genuine difference between the two proteins, or if it reflects differences in the protein preparations. Both D1D2-fd and D1-fd were expressed by secretion to the periplasm as in the natural folding pathway of g3p, while D1D2-M13 was refolded from cytoplasmic inclusion bodies (Lubkowski *et al.*, 1997).

While the individual domains in D1D2-fd and D1D2-M13, apart from the minor differences discussed above, are very similar (< 0.7 Å), the overall rms for the two structures is significantly higher (1.27 Å) than that for the individual domains. This is caused by a rigid-body ''rotation'' of D2 relative to D1 compared to the D1D2-M13 structure. If the D1 domains of the two structures are aligned for maximum overlap, this apparent movement of D2 causes a maximum divergence of 5 Å among main-chain atoms within D2 from the two phages

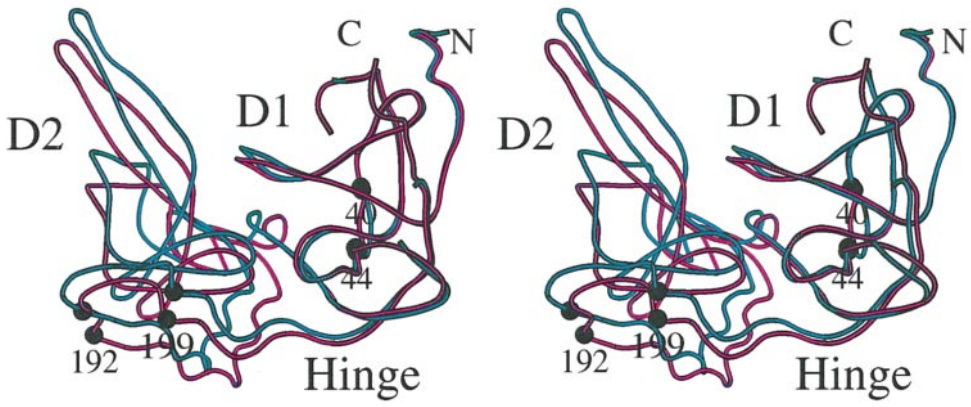


Figure 2. Stereo representation showing a superimposition of a backbone worm of the g3p D1D2-M13 (cyan) onto g3p D1D2-fd structure (magenta). The structures are superimposed so as to have maximum overlap for the D1 domain (residues 2-66), indicating the different relative positions of the D2 domain in the two structures with respect to the D1 domain. The N and C termini of D1D2-fd are labelled. Residues A40, V44 and S192, F199, are highlighted

(Figure 2). The rotation occurs about a ''hinge'' centred on residues G99 and S208. These residues lie in strands β6 and β13 of D2, which form a short two-stranded anti-parallel β-sheet at the centre of two extended polypeptide strands. The hinge strands (β6 and β13) connect the first lobe (containing D1) to the second lobe (containing the majority of D2). This hinge region should not be confused with the Gly-rich linker (residues 69-87) that separates D1 and D2 in the sequence of g3p. Herewith we will refer to this region involving residues K88 to P123 and Y203 to A217 as the hinge sub-domain (HD; Figure 1(b)).

Several features make the hinge sub-domain (HD) stand out from the rest of the structure. It has been proposed that the core of D2 (residues 124-202; D2* (Figure 1(a)) may have arisen from D1 by gene-duplication (Lubkowski *et al.*, 1997) and indeed, both D1 and D2* are compact β-barrel domains of identical topology with a short α-helix packing onto the outside of the barrel. In contrast, the hinge sub-domain seems unlikely to have arisen as part of this gene duplication. Rather, D2* appears to be an insertion into the previously continuous sequence of the HD. The fact that hinge sub-domain makes the majority of its contacts with D1 (12 compared to 6 with D2*, Figure 1(a) and (c)) also suggests that it may have predated the gene-duplication that gave rise to D2*. Intriguingly, most of the buried contacts between HD and D1 (Figure 1(a) and (c)) involve residues predicted by NMR to be part of the TolA-binding site on D1 (Riechman & Holliger, 1997). This raises the possibility that the HD is involved in giving some conformational flexibility to the D2 domain and is a major component of the blockage of the TolA-binding site on D1.

These considerations lead us to speculate that the hinge domain may play a pivotal role in the infection process, presumably by allowing and facilitating the conformational change in D1D2 required to free the TolA coreceptor binding site on D1 after contact with the pilus. The differences between the D1D2-fd and D1D2-M13 structures suggest how such a conformational change might proceed. Although the rotation of D2 away from D1 in D1D2-fd is a significant movement, it is not sufficient to expose the TolA-binding site on D1. However, the rotation clearly leads to a widening of the groove running between the two domains, where the postulated pilus-binding site is located. Assuming that pilus binding requires or promotes a further widening of the groove, a more substantial movement of D2 with respect to D1 would be required. Such a movement, if continued about the G99, S208 hinge, would soon lead to considerable conformational strains within the HD. As the HD itself covers a large part of the TolA-binding site on D1, it is tempting to speculate that this strain may eventually lead to a breaking of HD interactions with D1.

Because of the difficulties in obtaining pili in sufficient quantity and purity for biochemical or structural investigations, we reasoned that it may be possible to mimic the above process *in vitro* using agents that influence protein conformation. Thus we probed the D1D2-fd structure by exposing wild-type phage particles to a range of different conditions and measuring phage infectivity both on pilus-bearing bacteria (TG1) as well as *E. coli* devoid of pilus (HB2156). The same phage aliquots were simultaneously analysed for TolA binding by ELISA (Riechmann & Holliger, 1997). While various detergents proved unsuccessful (not shown), we found that the blockage of TolA binding (and therefore the wild-type conformation in D1D2) is remarkably sensitive to changes in temperature. The TolA-binding site on D1 becomes exposed at temperatures as low as 45 °C, as judged both by ELISA and by the increase in pilus-independent infectivity in HB2156 cells (Figure 3(a)). Similarly, wild-type phage loses infectivity due to protease cleavage of g3p after incubation at temperatures above 47 °C but is resistant at 37 °C (Kristensen & Winter, 1998). Thus, the transition temperature for both the exposure of the TolA-binding site on g3p and the exposure of proteolytic sites (due to at least local unfolding) within g3p are very similar and coincide with the apparent starting point of unfolding of the intact D1D2 domain as measured by CD (Figure 3(b)). In contrast, the recombinant D1 domain is significantly more resistant to thermal unfolding with a t_m of 66 °C as judged by CD (Figure 3(b)). This suggests that the conformational changes around 45 °C leading to the exposure of the TolA-binding site (and sensitivity to proteases) take place in the D2 portion of the g3p molecule.

The exposure of the TolA-binding site as judged by ELISA is closely parallelled by phage infectivity *via* the pilus-independent infection pathway (Figure 3(a)). Therefore, the change in proteolytic sensitivity of g3p, the exposure of the TolA-binding site and the correlated increase in infectivity by the pilus-independent pathway all point to critical conformational changes within g3p. Conversely, infectivity by the pilus-dependent pathway seems largely unaffected by temperature (Figure 3(c)). This is particularly intriguing as the pilus-binding site on g3p is thought to be localised within D2, the presumably least stable component of the g3p-D1D2 fragment (Figure 3(b)). It is plausible that unfolding of g3p-D1D2 and especially the D2 domain is highly reversible and that unfolded D2 may quickly revert to an ordered conformation when presented by its cognate pilus ligand. Precedents for such behaviour include transcription factors like GCN4, the DNA-binding domain of which is unfolded at 37 °C as judged by NMR analysis but becomes ordered upon contact with its cognate DNA target sequence (Weiss *et al.*, 1990; Spolar & Record, 1994), as well as the anti-termination factor N from phage lambda, which appears fully unfolded in the absence of its RNA

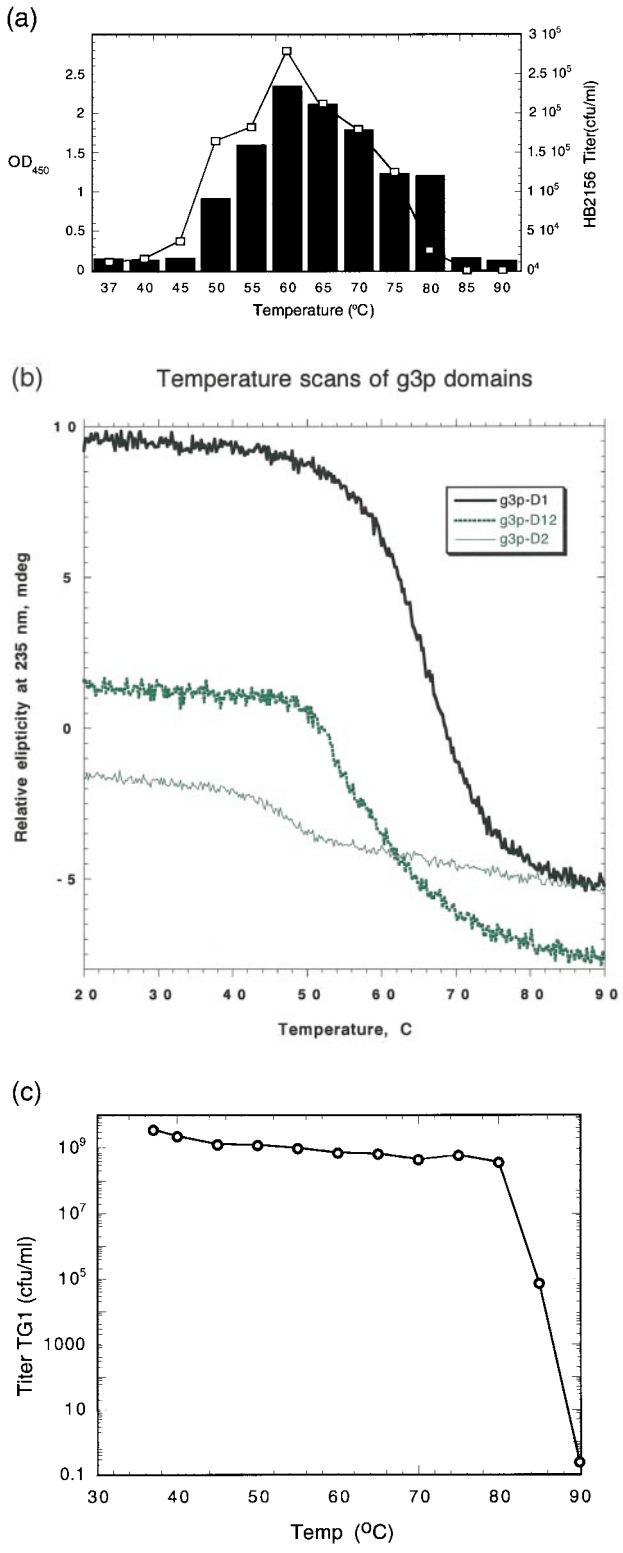


Figure 3. (a) Temperature-dependence of infection titers (colony-forming units (cfu)/ml) for phage fd and *E. coli* strain HB2156 (no pilus; open squares) overlaid on the temperature-dependence of the ELISA signal of phage fd for TolA binding (filled bars). (b) CD melting curves of recombinant g3p fragments: g3p-D1 (thick line), g3p-D2 (thin line) and g3p-D1D2 (broken line). Relative elipticity is plotted against temperature. (c) Temperature-dependence of infection titers for phage fd and *E. coli* strain TG1 (pilus bearing; open circles). Infection titers (cfu/ml) are shown on a logarithmic scale.

target sequence (Mogridge *et al.*, 1998; van Gilst *et al.*, 1997). However, it cannot be ruled out that the pilus-binding site on D2 is independent of the folded state of the rest of the D2 domain, for example, if it were formed by the disulphide-stabilised loop C188 to C201 or by a linear peptide sequence within D2.

Conformational lability is a feature of several other viral infectivity proteins. Conformational changes can be critical for infection and are triggered by specific events encountered during the infection pathway. These events include the pH change after entry into the lysosome in the case of influenza hemagglutinin (Bullough *et al.*, 1994), the binding to the CD4 primary receptor in the case of HIV-1 gp120 (Sattentau *et al.*, 1993) and presumably pilus binding in the case of g3p of filamentous phage. While the precise molecular events during phage infection that lead to the unblocking of the TolA-binding site on D1 remain to be determined, it seems likely that the here-described conformational lability of D1D2 plays an important part in the process. In a plausible scenario, the binding of the F′ pilus to its binding site in the inter-domain groove of D1D2 leads to a widening of the groove facilitated by the hinge domain. This opening up eventually may lead to a more substantial unfolding of the hinge subdomain or even other parts of the D2 structure, thus releasing the blockage of the TolA-binding site and triggering further steps of the infection process.

The possibility that a partial unfolding of g3p plays a part during phage infection has interesting analogies in the proposed models of colicin translocation, in particular group A colicines. These colicines share with filamentous phages the TolA requirement (as well as other components of the tol locus) for entry into the cell (Sun & Webster, 1987). Structures of colicines from both group B (colicin Ia; Wiener *et al.*, 1997) and group A (colicin N fragment; Vetter *et al.*, 1998) together with biochemical data suggest that colicin translocation may involve substantial unfolding of the colicin structure. This leads to the colicin molecule being stretched across the periplasmic space while anchored in both the outer and the inner *E. coli* membranes (Vetter *et al.*, 1998). g3p could become similarly extended across the periplasmic space during the infection process (Riechmann & Holliger, 1997) and the structure of D1D2 suggests that g3p may well be designed to do so: the hinge sub-domain is flanked on both sides by the extended Gly-rich linker peptides of g3p. Thus an unfolding of the hinge sub-domain may expose the TolA-binding site on D1 and release the g3p domains like beads on a string, allowing them to span a large distance (>200 Å) while simultaneously remaining bound to their respective receptors.

A better understanding of filamentous phage infection has implications for biotechnology, concerning phage display technology (Winter *et al.*, 1994), as well as for medicine, concerning the

association of bacterial virulence genes with mobile genetic elements like the filamentous phage ctxϕ (Waldor & Mekalanos, 1996) in the case of cholera. While further work is required, some aspects of the molecular events during the early steps of the infection process of filamentous phages are beginning to emerge. The challenge will now be to integrate these into a coherent picture and to elucidate the later events of the infection process leading to the import of the phage genome into the host cytoplasm.

Materials and Methods

Protein expression, purification, crystallisation and CD analysis

The *E. coli* strains TG1 (Gibson, 1984) and HB2156 (Riechmann & Holliger, 1997) were used for propagation of plasmids, expression of the g3p-fragments and phage infection. Phage preparation and infection experiments were as in (Riechmann & Holliger, 1997). The g3p-D1D2 fragment (comprising amino acid residues 2 to 217 of the mature fd-g3p) were PCR amplified with primers (5′-CAT GCC ATG ACT CGC GGC CCA GCC GGC CAT GGC AGA AAC TGT TGA AAG TTG TTT AGC A-3′) and (5′-GAG TCA TTC TGC GGC CGC ATT GAC AGG AGG TTG AGG CAG GTC-3′) using *Taq* polymerase (HT Biotechnology Ltd.), digested with *Sfi*I and *Not*I (New England Biolabs), and ligated into the *Sfi*I and *Not*I digested pUC119His6 (Holliger & Riechmann, 1997). The correct sequence of the resulting construct was confirmed. The g3p-D1 and g3p-D2 fragments were prepared and purified as described (Riechmann & Holliger, 1997).

For expression of g3p-D1D2, TG1 cells harbouring the expression plasmid were grown in 2 × TY with 100 μg/ml ampicillin and 0.1 % (w/v) glucose at 37 °C overnight and then induced for five hours with 1 mM IPTG at 30 °C. Selenomethionine incorporation was essentially as described (Budisa *et al*., 1995; Doublié, 1997). B834 cells with the expression plasmid (Novagen) were grown overnight in M9 medium with 100 μg/ml ampicillin supplemented with 2 % (w/v) glucose, 2 μg/ml thiamine (SIGMA) and 50 μg/ml seleno-DL-methionine (SIGMA, M9Se), after overnight growth, cells were spun down and resuspended in an equal volume of M9Se without glucose and induced for five hours with 0.4 mM IPTG. g3p-D1D2 was purified from the periplasmic extract, exchanged into Tris-buffered saline (TBS), pH 7, and purified using immobilised metal-affinity chromatography IMAC (Hochuli *et al*., 1988). Yields were 20 mg/l. Protein was further purified by gel-filtration on a Superose-12 column (Pharmacia) by fast protein liquid chromatography (FPLC) and concentrated using Centricon-10 concentrators (Amicon) to a final concentration of 30 mg/ml. Crystals grew within a few days to 0.1 mm length, at 21 °C in 4.3 M sodium formate, 0.1 M sodium cacodylate (pH 6.2) in a hanging drop vapour diffusion set-up and were propagated by hair-seeding. CD analyses were performed as described (Holliger & Riechmann, 1997).

X-ray diffraction and structure solution

The fd g3p-D1D2 fragment crystallised with $P2_12_12$ symmetry and unit cell dimensions $a = 125.2$ Å, $b = 78.3$ Å, $c = 62.7$ Å. Diffraction data were collected at 100 K and are summarised in Table 1. For data collection, crystals were flash-frozen in nylon loops using a nitrogen gas stream. For the freezing, crystals were transferred to a cryoprotectant solution consisting of crystallisation solution with 20 % glycerol. Each data set was collected from a single crystal and processed using the program MOSFLM (Leslie, 1992). Diffraction data for the two isomorphous heavy-atom derivatives were collected using a rotating anode X-ray source. For the K_2PtCl_4 and di-μ-iodobis-(ethylene diamine)-diplatinum (II) nitrate (PIP; Strem Chemicals) derivatives, crystals were soaked for 16 hours in the crystallisation solution containing 20 mM K_2PtCl_4 or 10 mM PIP, respectively. Most crystallographic calculations were carried out using the CCP4 program suite (CCP4, 1994). An electron density map was calculated using multiple isomorphous (MIR) phases based on the K_2PtCl_4 and the PIP derivatives and a selenomethionine-substituted protein crystal. Initial sites for the heavy-atom derivative were located by the program SOLVE (Terwilliger & Eisenberg, 1983, 1987; Terwilliger *et al*., 1987; Terwilliger & Berendzen, 1996) and additional sites were located by difference Fourier analyses. The program SHARP (de La Fortelle & Bricogne, 1997) was used to refine all heavy-atom parameters using both anomalous and isomorphous differences. The MIRAS map showed two molecules in the asymmetric unit. Following solvent flattening with the program SOLOMON (Abrahams & Leslie, 1996) using a solvent content of 45 %, the electron density was easily interpretable. In the course of building the model of D1D2-fd, a structure became available for the D1D2-M13. The model of D1D2-M13 (pdb entry 1g3p; Lubkowski *et al*., 1997) was placed into the D1D2-fd unit cell using molecular replacement with the program AMORE (Navaza, 1994). The initial manual model rebuilding of D1D2-fd employed both the MIR map and the SIGMAA weighted $2\ m|F_o| - D|F_c|$ electron density map calculated from the molecular replacement model. The MIR phases were used as external phases in subsequent rounds of maximum likelihood refinement with the program REFMAC (Murshudov *et al*., 1997). The model was built using the program O (Jones *et al*., 1991). Refinement used all data between 13 Å and 1.9 Å (47,624 reflections). Non-crystallographic symmetry (NCS) restraints on coordinates and *B*-factors were maintained throughout the refinement for the two molecules in the asymmetric unit except for residues involved in intermolecular contacts and residues that were visible in only one of the two molecules in the asymmetric unit. The NCS restraints were initially tight and in later stages of refinement loosened only if the free *R*-factor decreased. The final conventional *R*-factor was 25.7 % with a free *R*-factor of 29.9 % (based on 1499 random reflections) for a model consisting of 3160 protein atoms and 303 water molecules. The average *B*-factor was 40 Å^2. The rms deviations from ideal geometry were 0.006 Å and 1.5 ° for bond lengths and bond angles. The rms deviation between the two molecules in the asymmetric unit is 0.33 Å. According to analysis by PROCHECK (Laskowski *et al*., 1993), 86 % of the residues are in the most favourable regions of the Ramachandran plot and none is in a disallowed area. The overall average *G*-factor from PROCHECK was 0.08. Analysis of the heavy-atom positions relative to the coordinates of the final refined model showed that the platinum sites

were near the SD atom in residue Met135 of each of the two molecules in the asymmetric unit. A third platinum site was found between non-crystallographically related ND1 atoms of His12. The Se sites on Met176 and Met135 of the Se-Met protein were related only approximately by the non-crystallographic symmetry. The side-chain conformation of Met135 differs in the two molecules in this region due to a crystal contact.

Brookhaven Protein Data Bank

Atomic coordinates for the refined D1D2-fd model have been deposited in the Brookhaven Protein Data Bank with the accession code 2g3p.

Acknowledgements

We thank the staff of beamlines Daresbury SRS Station 7.2, and W. Burmeister and other staff at ESRF ID14-EH3.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallog. sect. D,* **52**, 30-42.

Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. & Huber, R. (1995). High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *Eur. J. Biochem.* **230**, 788-796.

Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature,* **371**, 37-43.

CCP4. (1994). Collaborative Computing Project number 4: a suite of programs for protein crystallography. *Acta Crystallog. sect. D,* **50**, 760-763.

Click, E. M. & Webster, R. E. (1997). Filamentous phage infection: required interactions with the TolA protein. *J. Bacteriol.* **179**, 6464-6471.

de La Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472-494.

Doublié, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523-530.

Gibson, T. J. (1984). *Studies on the Epstein-Barr Virus Genome*, University of Cambridge.

Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R. & Stüber, D. (1988). Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Bio/Technology,* **6**, 1321-1325.

Holliger, P. & Riechmann, R. (1997). A conserved infection pathway for filamentous phages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. *Structure,* **5**, 265-275.

Jacobson, A. (1972). Role of the F-pili in the penetration of bacteriophage f1. *J. Virol.* **10**, 835-843.

Janin, J. & Chothia, C. (1990). The structure of protein-protein recongition sites. *J. Biol. Chem.* **265**, 16027-16030.

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 Crystallog. sect. A,* **47**, 110-119.

Kristensen, P. & Winter, G. (1998). Proteolytic selection for protein folding using filamentous bacteriophages. *Fold. Des.* **3**, 321-328.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.

Leslie, A. G. W. (1992). *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography*, Daresbury Laboratory, Warrington, UK.

Levengood, S. K., Beyer, W. F., Jr & Webster, R. E. (1991). TolA: a membrane protein involved in colicin uptake contains an extended helical region. *Proc. Natl Acad. Sci. USA,* **88**, 5939-5943.

LoConte, L., Chothia, C. & Janin, J. (1999). The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177-2198.

Lubkowski, J., Hennecke, F., Pluckthun, A. & Wlodawer, A. (1997). The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. *Nature Struct. Biol.* **5**, 140-147.

Marvin, D. A. (1998). Filamentous phage structure, infection and assembly. *Curr. Opin. Struct. Biol.* **8**, 150-158.

Model, P. & Russel, M. (1988). Filamentous bacteriophages. In *The Bacteriophages* (Calendar, R., ed.), pp. 375-456, Plenum Press, New York.

Mogridge, J., Legault, P., Li, J., Van Oene, M. D., Kay, L. E. & Greenblatt, J. (1998). Independent ligand-induced folding of the RNA binding domain and functionally distinct antitermination regions in the phage lambda N protein. *Mol. Cell,* **1**, 265-275.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D,* **53**, 240-255.

Navaza, J. (1994). AMORE - an automated procedure for molecular replacement. *Acta Crystallog. sect. A,* **50**, 157-163.

Riechmann, L. & Holliger, P. (1997). The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. *Cell,* **90**, 351-360.

Russel, M., Whirlow, H., Sun, T. P. & Webster, R. E. (1988). Low-frequency infection of F^- bacteria by transducing particles of filamentous bacteriophages. *J. Bacteriol.* **170**, 5312-5316.

Sattentau, O. J., Moore, J. P., Vignaux, F., Traincard, F. & Poignard, P. (1993). Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding. *J. Virol.* **67**, 7383-7393.

Spolar, R. S. & Record, M. T. (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science,* **263**, 777-784.

Stengele, I., Bross, P., Garces, X., Giray, J. & Rasched, I. (1990). Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites. *J. Mol. Biol.* **212**, 143-149.

Sun, T. P. & Webster, R. E. (1986). fii, a bacterial locus required for filamentous phage infection and its relation to colicin-tolerant tolA and tolB. *J. Bacteriol.* **165**, 107-115.

Sun, T. P. & Webster, R. E. (1987). Nucleotide sequence of a gene cluster involved in entry of E colicins and single-stranded DNA of infecting filamentous bacteriophages into *Escherichia coli*. *J. Bacteriol.* **169**, 2667-2674.

Terwilliger, T. C. & Berendzen, J. (1996). Correlated phasing of multiple isomorphous replacement data. *Acta Crystallog. sect. D,* **52**, 749-757.

Terwilliger, T. C. & Eisenberg, D. (1983). Unbiased three-dimensional refinement of heavy-atom parameters by correlation of origin-removed Patterson functions. *Acta Crystallog. sect. A,* **39**, 813-817.

Terwilliger, T. C. & Eisenberg, D. (1987). Isomorphous replacement: effects of errors on the phase probability distribution. *Acta Crystallog. sect. A,* **43**, 6-13.

Terwilliger, T. C., Kim, S. H. & Eisenberg, D. (1987). Generalized method of determining heavy-atom positions using the difference Patterson function. *Acta Crystallog. sect. A,* **43**, 1-5.

van Gilst, M. R., Rees, W. A., Das, A. & von Hippel, P. H. (1997). Complexes of N antiterminationprotein of phage lambda with specific and nonspecific RNA target sites on the nascent transcript. *Biochemistry,* **36**, 1514-1524.

Vetter, I. R., Parker, M. W., Tucker, A. D., Lakey, J. H., Pattus, F. & Tsernoglou, D. (1998). Crystal structure of a colicin N fragment suggests a model for toxicity. *Structure,* **6**, 863-874.

Waldor, M. & Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science,* **272**, 1910-1914.

Webster, R. E. (1991). The tol gene products and the import of macromolecules into *Escherichia coli*. *Mol. Microbiol.* **5**, 1005-1011.

Webster, R. E. (1996). Biology of the filamentous bacteriophage. In *Phage Display of Peptides and Proteins* (Kay, B. K. & Winter, J., eds), Academic Press, New York.

Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C. & Struhl, K. (1990). Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature,* **347**, 575-578.

Wiener, M., Freymann, D., Gosh, P. & Stroud, R. M. (1997). Crystal structure of colicin Ia. *Nature,* **385**, 461-464.

Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. (1994). Making antibodies by phage display technology. *Annu. Rev. Immunol.* **12**, 433-455.

Edited by J. M. Thornton

(Received 7 December 1998; received in revised form 9 March 1999; accepted 15 March 1999)