Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA

Jacek Lubkowski[1]*, Frank Hennecke[2†], Andreas Plückthun[2]
and Alexander Wlodawer[1]

Background: Infection of male *Escherichia coli* cells by filamentous Ff bacteriophages (M13, fd, and f1) involves interaction of the phage minor coat gene 3 protein (g3p) with the bacterial F pilus (primary receptor), and subsequently with the integral membrane protein TolA (coreceptor). G3p consists of three domains (N1, N2, and CT). The N2 domain interacts with the F pilus, whereas the N1 domain – connected to N2 by a flexible glycine-rich linker and tightly interacting with it on the phage – forms a complex with the C-terminal domain of TolA at later stages of the infection process.

Results: The crystal structure of the complex between g3p N1 and TolA D3 was obtained by fusing these domains with a long flexible linker, which was not visible in the structure, indicating its very high disorder and presumably a lack of interference with the formation of the complex. The interface between both domains, corresponding to ~1768 Å^2 of buried molecular surface, is clearly defined. Despite the lack of topological similarity between TolA D3 and g3p N2, both domains interact with the same region of the g3p N1 domain. The fold of TolA D3 is not similar to any previously known protein motifs.

Conclusions: The structure of the fusion protein presented here clearly shows that, during the infection process, the g3p N2 domain is displaced by the TolA D3 domain. The folds of g3p N2 and TolA D3 are entirely different, leading to distinctive interdomain contacts observed in their complexes with g3p N1. We can now also explain how the interactions between the g3p N2 domain and the F pilus enable the g3p N1 domain to form a complex with TolA.

Addresses: [1]Macromolecular Structure Laboratory, NCI-Frederick Cancer Research and Development Center, ABL-Basic Research Program, Frederick, MD 21702, USA and [2]Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.

[†]Present address: Cytos Biotechnology AG, Wagistrasse 21, CH-8952 Schlieren, Switzerland.

*Corresponding author.
E-mail: jacek@ncifcrf.gov

Introduction

Filamentous bacteriophages, especially the Ff phages (M13, fd, and f1), have become an important tool for the identification of interacting biomolecules as well as for the improvement of the interactions by evolutionary approaches [1–3]. Peptides and proteins can be displayed on the phage surface by genetic fusion to the phage minor coat gene 3 protein (g3p) and may be selected for their binding properties from large libraries. Because the genetic information is packaged within the phage particle, this process links phenotype and genotype. Selection is accomplished either by binding of the phage particles to immobilized ligands (phage panning [4]) or by linking the infectivity of the phages to the binding of a cognate ligand by the displayed receptor molecule (selectively infective phages (SIP) [5]).

Notwithstanding the great popularity of Ff phages as a selection tool, their mechanism of host cell infection is only very poorly understood. Three to five copies of the protein g3p, which are involved in the infection process, are located at one end of the phage particle [6]. G3p consists of three domains (N1, N2, and CT), which are connected to each other by flexible glycine-rich linkers.

Infection of a male *Escherichia coli* cell is initiated by binding of the N2 domain to the tip of an F pilus (primary receptor) [7] (Figure 1). The pilus is retracted by an unknown mechanism [8]; this retraction guides the bound phage to the cellular envelope. At this stage, interaction of the N1 domain with the C-terminal domain of the integral membrane protein TolA (coreceptor) [9,10] triggers a completely unknown process, finally leading to the entry of the phage genome into the bacterial cytoplasm.

The TolA protein consists of three domains. The N-terminal domain (D1), which anchors the protein within the cytoplasmic membrane, and the C-terminal domain (D3), which appears to be associated with the outer membrane, are connected to each other by the central domain (D2), which is assumed to be an α-helical structure that is long enough to span the entire periplasmic space [11]. Besides being dependent on TolA, phage infection has been shown to rely on the membrane proteins TolQ and TolR, which form a complex with TolA within the cytoplasmic membrane [12–15].

The biological function of the Tol complex is still unclear. A role in membrane transport of macromolecular

Figure 1
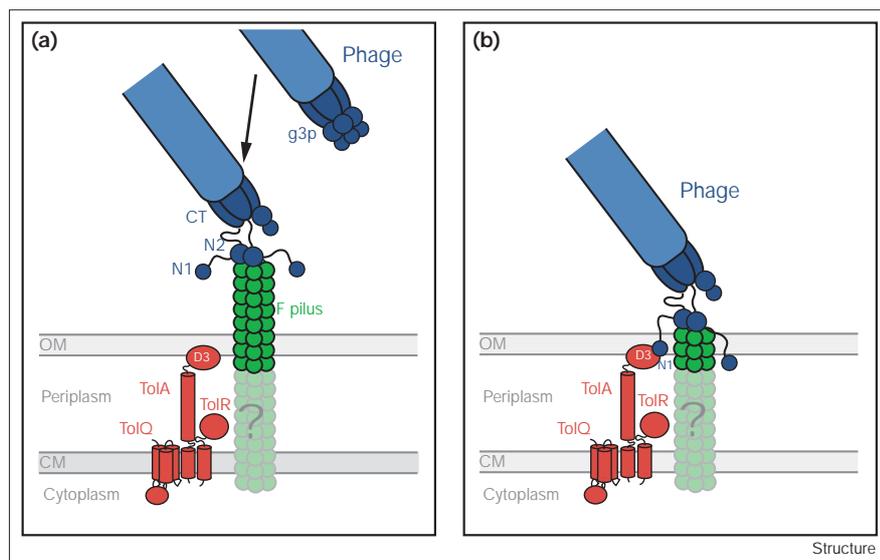
Schematic representation of the initial steps of phage infection. **(a)** Docking of the phage to the tip of an F pilus. **(b)** Binding of the g3p N1 domain to the D3 domain of TolA. For further explanation, refer to the text. CM: cytoplasmic membrane; OM: outer membrane.

compounds has been discussed, but no natural substrate has been found. This proposal has been stimulated by structural and sequence homologies of the TolA–TolQ–TolR complex to the TonB–ExbB–ExbD complex [16], which serves as an uptake system for iron siderophores and vitamin B_{12} [17,18]. A function in maintaining membrane integrity has been ascribed to the Tol proteins, as deletion of any of the genes resulted in leakage of periplasmic proteins and hypersensitivity of the cell to detergents [15,19,20]. Membrane leakiness has also been observed upon overexpressing soluble fragments of TolA containing the D3 domain [21] or by overexpressing fragments containing the g3p N1 domain [22]. These data would be consistent with leakiness occurring whenever TolA, being anchored to the inner membrane via its D1 domain and to a protein in the outer membrane possibly via its D3 domain, is prevented from making this connection [23]. Similar effects have not been reported when the *tonB*, *exbB* and *exbD* genes were deleted.

Both TolA and TonB are also involved in the uptake of phage DNA (TolA: phages M13, f1, fd, IKe, and I2–2 [15]; TonB: phages T1 and ɸ-80 [24]) and colicins (TolA: colicins A, E1, E2, E3, K, L, N, and S4 [25]; TonB: colicins B, D, G, H, Ia, Ib, M, Q and V [15,26]). Colicins are protein toxins lethal to *E. coli* and other bacteria that do not express an immunity protein.

In an attempt to better understand the molecular events during the filamentous phage infection process, we recently solved the crystal structure of the tightly interacting N1 and N2 domains of g3p [27]. Here, we present the crystal structure of the complex between g3p N1 and TolA D3, obtained by fusing these domains with a long,

flexible linker. We believe that, besides enhancing our understanding of the phage-infection process, this structural knowledge will also be advantageous for the improvement of directed molecular-evolution techniques, such as the SIP technology [5].

Results
Overall structure

The structure of the g3p-N1–TolA-D3 construct was solved both at room temperature and under cryogenic conditions, using several data sets. All of these data were used in independent refinements. The resulting structures were very similar, with the root mean square deviation (rmsd) between the room-temperature and the high-resolution, low-temperature structures being 0.37 Å for Cα atoms and 0.38 Å for all atoms. The differences were limited to sidechain conformations of a small number of residues (4, 5, 13, 15, 25, 49, 131, 153, 158, 160, 171, 179 and 184). For these reasons, only the highest resolution, low-temperature structure will be discussed here in detail. The statistics for heavy-atom derivatives and the refinement statistics for the final model are shown in Tables 1 and 2, respectively.

The fusion protein consists of two domains originating from different parent molecules, tethered with a glycine-rich, flexible linker. The construct was prepared in this manner to force both domains to be present in the crystal at equimolar concentration and to achieve a high local concentration of both domains, given that their affinity was previously shown to be weak [29]. The N-terminal domain in this construct corresponds to the first (N-terminal) domain of g3p (g3p N1), and the C-terminal domain of the fusion protein is the third (C-terminal) domain present in

Table 1

Statistics for heavy-atom derivatives and data collection.

	Data sets for the native protein				Data sets for heavy-atom derivatives		
	Room	Cryo1	Cryo2	Cryo3	Hg25	Hg50	U50
Unit cell (Å)							
a	88.949	87.924	87.914	88.880	88.791	88.849	88.606
c	64.314	62.705	62.638	63.556	64.187	64.325	63.932
Detector	MAR345	MAR345	MAR345	CCD, ADSC	MAR345	MAR345	MAR345
Temperature (K)	293	100	100	100	293	293	293
Concentration of heavy atom	–	–	–	–	25% of sat.	50% of sat.	50% of sat.
Soaking time (h)	–	–	–	150	110	116	
Number of reflections							
total	72,744	126,605	95,736	190,164	42,886	61,165	28,862
unique	9984	11,367	9729	22,243	5445	6571	5443
Resolution* (Å)							
low	20	20	20	25	25	25	25
high	2.45	2.30	2.40	1.85	3.00	2.80	3.00
high-resolution shell	2.54	2.38	2.49	1.92	3.11	2.90	3.11
Completeness[†] (%)	99.9 (98.7)	99.9 (100.0)	96.4 (94.8)	99.6 (100.0)	99.1 (100.0)	97.6 (99.5)	99.7 (99.8)
R_{scale}[†‡]	0.092 (0.57)	0.070 (0.44)	0.080 (0.45)	0.042 (0.45)	0.100 (0.32)	0.095 (0.33)	0.080 (0.23)
R_{merge}[†§]	–	0.357 (2.45)	0.376 (2.45)	0.332 (2.45)	0.177 (2.80)	0.266 (3.00)	0.215 (3.00)
Number of substitution sites	–	–	–	–	4	4	1
Phasing power[¶]	–	–	–	–	2.48:1.51	2.31:1.60	0.64:1.79
R_{cullis}[#]	–	–	–	–	0.548	0.609	0.706

*The first two numbers are the low- and high-resolution limits applied during data processing, and the third number is the low-resolution limit for the highest resolution shell. [†]Numbers in parentheses correspond to the highest resolution shell. [‡]$R_{scale} = \Sigma|I_n - <I>|/\Sigma<I>$ for multiple observations of the same reflection. [§]$R_{merge} = \Sigma|I_{nat} - I_{der}|/\Sigma<I>$; all data sets are scaled to the native data set collected at room temperature (Room). [¶]Phasing power is defined as Fh_c/E for the isomorphous case and $2FH_c''/E$ for the anomalous case, where E is the rms lack of closure; the reported numbers describe the isomorphous and anomalous phasing power, respectively. [#]$R_{cullis} = \Sigma||FPH_o| \pm |Fp_o| - |FH_c||/\Sigma||FPH_o| \pm |FP_o||$ for centric reflections, where FP, FPH, and FH are the structure factors for the protein, derivative, and heavy atom, respectively.

TolA (TolA D3) (Figure 2). The structure of the g3p N1 domain has been described before, in complex with the g3p N2 domain [27]. The secondary structure of the fusion protein is shown in Figure 3 and the topology is shown in Figure 4. Apart from the short helical motif (α1) located close to the N terminus (residues Val4–Ala9), the g3p N1 domain consists predominantly of β structure. Helix α1 is followed by five β strands (β1–β5) arranged in a barrel-like motif. These five β strands, together with three additional β strands from TolA D3 (β6–β8), form an eight-stranded antiparallel sheet with topology 4x,1,1,1,–3,–1,2x. The strands in the N1 domain are of differing lengths. Strands β2 and β3 are quite short, whereas strands β4 and β5 are the longest in the entire structure and their topology

Table 2

Refinement statistics.

	Low-temperature structure (Cryo3)	Room-temperature structure
Nonhydrogen atoms		
protein	1169	1174
water	159	58
Resolution (Å)	10–1.85	10–2.5
R_{factor}*	0.220/0.227	0.171/0.185
R_{free}[†]	0.283/0.293	0.235/0.251
Rmsd from ideal geometry		
Bonds (Å)	0.007	0.005
Angle distances (Å)	0.024	0.022
Chiral volumes (Å^3)	0.047	0.030
Distances from restraint planes (Å)	0.031	0.025
Residues in most favored regions of Ramachandran plot (%)	88.6	87.2

*$R_{factor} = \Sigma_{hkl}|F_{obs}| - |F_{calc}|/|F_{obs}|$, where F_{obs} denotes the observed structure-factor amplitude and F_{calc} denotes the structure-factor amplitude calculated from the model. [†]7% of the reflections were used to calculate R_{free} for the low-temperature and room-temperature structures.

Figure 2
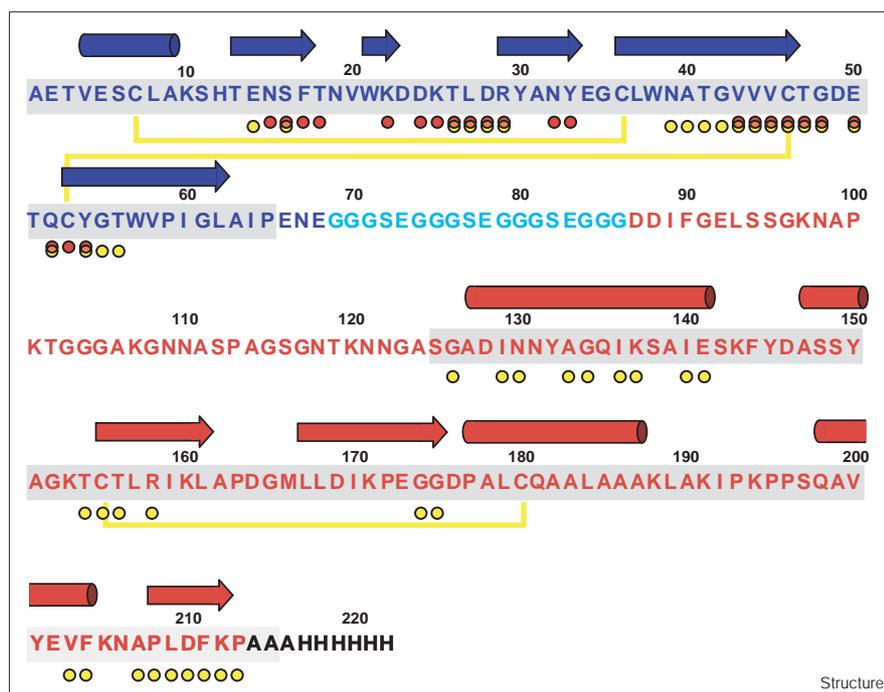
Primary structure of the fusion protein used in the crystallographic studies described here, with every tenth residue numbered. Residues from the g3p N1 domain are in blue, whereas those from the TolA D3 domain are in red. The disulfide connections are represented by yellow lines. Residues located within the glycine-rich linker are in light blue. All residues for which an interpretable electron density was found are shown with a gray background. The secondary-structure elements, colored according to their location in g3p N1 (blue) and TolA D3 (red), are indicated by arrows for β strands or cylinders for helices. The residues participating in the contacts between two domains (interatomic distances shorter than 4.5 Å) are labeled with yellow circles. For comparison, the residues from the g3p N1 domain interacting with whose from the g3p N2 domain [27] are labeled with red circles (see text for more details).

departs visibly from an ideal one. Strands β2, β3, β4 and β5 form three hairpins (hβ2/3, hβ3/4 and hβ4/5). Three bulges [28] (residues Asp23–Leu27/Asp28, residues Gly55–Gly42/Val43, and residues Leu37–Ile60/Gly61, respectively) are present within hβ2/3 and hβ4/5, disrupting the hydrogen-bond network typical of a β sheet. These bulges are probably the result of several topological constraints observed within the N1 domain, introduced by the two disulfide bridges and by the interactions of β4 and β5 with other topological motifs. One disulfide bridge is formed by Cys7 (α1) and Cys36 (β4) and shows the conformation of a left-handed spiral. Several exclusively hydrophobic interactions between α1 and β5 are also found in this region. The second disulfide bridge, which is located close to the tip of hβ4/5, has the conformation of a short right-handed hook and is formed by Cys46 and Cys53. The hairpin hβ4/5 also integrates the domains of both g3p and TolA, because β4, as well as interacting with β3 and β5, extends the β sheet onto β8 (the third strand of TolA D3) via several hydrogen bonds.

TolA D3, the larger domain that was introduced into the fusion protein, was defined according to the protein sequence as the C-terminal stretch (TolA residues 295–421) immediately following the alanine-rich center portion of this protein. However, only the residues corresponding to the fragment TolA 333–421 (here, numbered sequentially Ser125–Pro213) are visible in the structure. This part consists of three β strands and four helical motifs.

The β strands form an antiparallel arrangement and are an extension of the sheet formed by g3p N1. The longest α helix in the entire molecule (α2, Ala127–Glu141) is the first structural element of TolA D3. This 15-residue helix also participates in the interactions with g3p N1 via hydrogen bonds and hydrophobic contacts with strands β1, β4, and β5. The second helix in TolA D3 is a very short 3_{10} helical motif located between α2 and β6. The second longest helix in the molecule is α3, which follows strand β7 and, like helix α2, is followed by a rather short α helix (α4, with seven residues). All helices in the TolA D3 domain are located on the same side of the β sheet. The contacts between TolA and g3p N1 are quite extensive, resulting in stabilization of the molecular core. This effect is reflected in the distribution of the temperature factors, which are low for the residues located at the interface between the two domains and higher elsewhere. The most flexible fragments of the molecule are the terminal residues, a γ turn formed by Ser125, Gly126 and Ala127, a β turn formed by residues Phe17, Thr18, Asn19 and Val20, and part of the helix α4. As a result of the flexibility and resulting structural disorder, we could not observe any defined structure for the last three residues of g3p N1 (Glu66–Glu68), the interdomain linker (Gly69–Gly86) and the beginning of the TolA stretch used (Asp87–Ala124).

Crystal packing

The molecules forming the crystal participate in only a limited number of crystal contacts. This situation is in

agreement with the Matthews coefficient for this structure being 3.22 Å³/Dalton, which is rather high and indicates loose packing. Moreover, nearly all the crystal contacts involve residues of the TolA D3 domain. Most pronounced are the interactions between the residues within helix α2 of one molecule and those within helix α3 of a symmetry mate. Another group of significantly less apparent interactions is observed between residues belonging to strands β6 and β7 and between several N-terminal residues of a symmetry-related molecule.

Comparison between complexes of g3p N1 with TolA D3 and g3p N2

Both the previously described structure of a two-domain construct of g3p [27] and the fusion protein described here share the common N-terminal domain g3p N1, the isolated form of which was previously studied by nuclear magnetic resonance (NMR) spectroscopy [29]. The structure of this domain is very similar in complex to g3p N2 and TolA, with rms deviations of 0.40 Å for all atoms and 0.38 Å for Cα atoms only. The most significant difference in the mainchain trace is observed for hairpin hβ2/3, which in the fusion protein is moved by more than 1 Å toward the TolA D3 domain. Such a conformation cannot be achieved in the g3p(N1–N2) complex because of a collision with strands β9 and β10 of the N2 domain. Several residues located in other regions of the g3p N1 domain accommodate different sidechain conformations in the two proteins. These residues include Glu5, a segment from His12 to Val20, Asp49, Glu50, and Tyr54. In contrast to the double conformation of the His12 sidechain found in g3p(N1–N2), we observe only a single position of this residue in the structure described here. Part of the apparent structural variability of the g3p N1 domain probably results from different crystallization conditions and crystal packing, and may also be due to differences in the resolution of respective X-ray data.

In addition to these structural variations we observed a difference in the chemical nature of an element of this subunit. In a previous study we observed the unusual oxidized Trp21 in g3p(N1–N2), which was possibly a product of air oxidation during crystal growth catalyzed by an unknown factor as the mass spectroscopy data of that protein sample before crystallization was exactly as expected [27]. In contrast, the electron density (multiple isomorphous replacement [MIR] as well as the final $2F_o-F_c$) in the current structure unambiguously corresponds to a normal, unmodified tryptophan residue (Figure 5). This observation supports our previous interpretation that the observed unusual phenomenon [27] was not relevant to the activity of the protein.

Interdomain interactions
As seen in Figures 2, 3, 4, and 6, the topologies of the domains TolA D3 and g3p N2 are very different. A closer

Figure 3
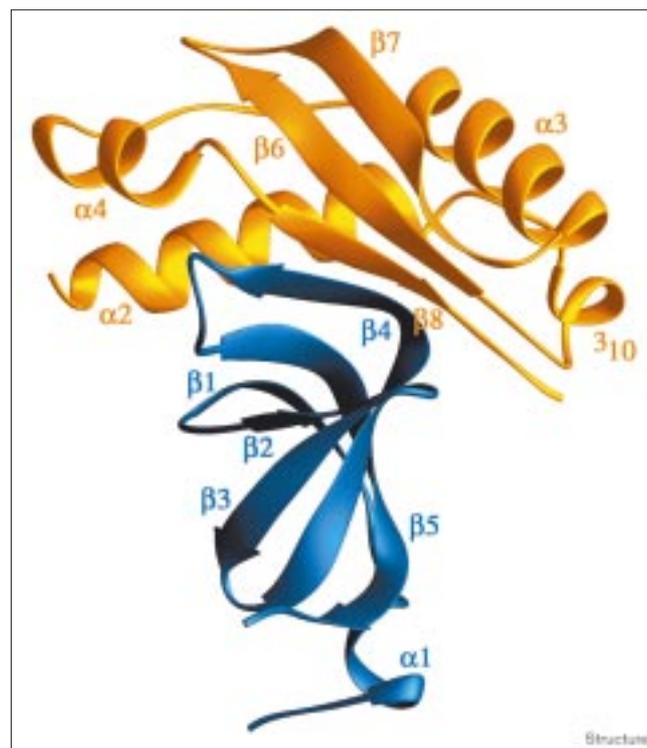
Secondary-structure diagram of the fusion protein g3p-N1–TolA-D3. The N1 domain of g3p is shown in blue, and the D3 domain of TolA is shown in orange. The secondary structure elements are labeled. The figure was prepared with the program RIBBONS [63].

look at Figure 6, however, shows that the interfaces between the pairs of domains are somewhat similar. This similarity is also reflected by the fact that in both cases the accessible surface that is buried as a result of domain association is comparable: 1768 Å² for the fusion protein described here (892 Å²: g3p N1; 876 Å²: TolA D3) and 2154 Å² for g3p(N1–N2) (1070 Å²: g3p N1; 1084 Å²: TolA D3). The fragments of g3p N1 participating in the interdomain contacts in both proteins are shown in Figure 2. Despite the differences in the partner domains, the subset of g3p N1 residues involved in interdomain interactions is strikingly similar. This similarity was also previously suggested by Riechmann and Holliger on the basis of their NMR studies [9]; however, no detailed structural description was provided in the previous work.

Comparing the individual interatomic contacts contributing to the stabilization of the pairs of domains is enlightening. Because of the dramatic difference in topologies and shapes of TolA D3 and g3p N2, some contacts with g3p N1 do not have direct equivalences in the two structures. Thus, the interdomain contacts involving residues Glu14, Asn39, Ala40, Thr41, Gly42, Gly55 and Thr56 are only made by the g3p N1 domain of the fusion protein

Figure 4
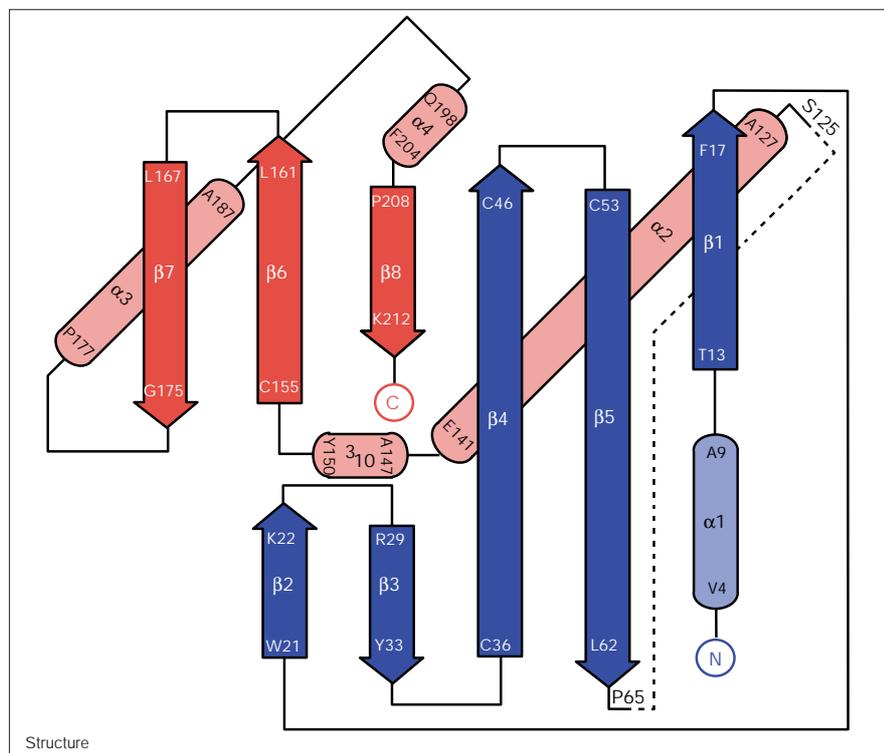
Topological diagram of the g3p-N1–TolA-D3 fusion protein with the β strands indicated by arrows (darker shades) and the helices by cylinders (lighter shades). Elements identified in the g3p N1 domain are shown in blue, whereas the elements from the TolA D3 domain are in red. The interdomain linker is represented by a dotted line.

g3p-N1–TolA-D3 (Figure 2, yellow dots). All these contacts occur predominantly between residues in helix α2 and strand β8 of TolA D3, secondary structure elements that do not have an equivalence in the N2 domain of g3p. Similarly, in the N1–N2 fragment of g3p [27], we find unique interdomain contacts involving Asn15, Phe17,

Figure 5
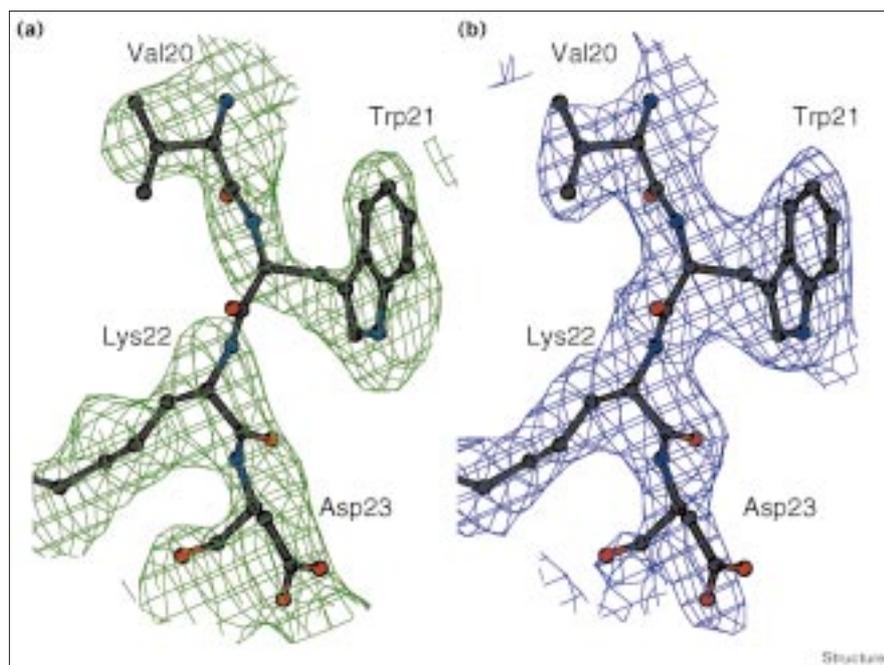
Electron-density maps for the area of Trp21, contoured at the 1σ level. **(a)** An experimental map after calculations with SHARP [59] shown in green. **(b)** The final refined $2F_o–F_c$ map shown in blue. On the basis of the density presented here, it is very clear that in this complex Trp21 is not oxidized, contrary to what was previously found in g3p(N1–N2) [27], emphasizing that the latter was probably formed by air oxidation during the crystallographic procedures. However, this map fragment is not fully representative, as significantly better electron density has been seen in other parts of the molecule. The figure was prepared with the program BOBSCRIPT, a modification of MOLSCRIPT [64].

Figure 6

Ribbon diagrams illustrating the interface between the two domains in **(a)** the fusion protein g3p-N1–TolA-D3 and **(b)** the phage minor coat protein g3p(N1–N2). The domain common to both proteins, g3p N1, is shown in yellow and is oriented identically in (a) and (b). The interacting domains, TolA D3 and g3p N2, are represented by thinner ribbons, colored in blue. In both cases, the red surfaces represent the solvent-accessible surfaces for the residues of the g3p N1 domain that interact with residues of the other domains. Despite the dramatic differences in the topology of the g3p N2 and TolA D3 domains, it is evident that they use a very similar binding site of the g3p N1 domain during the formation of the complex.
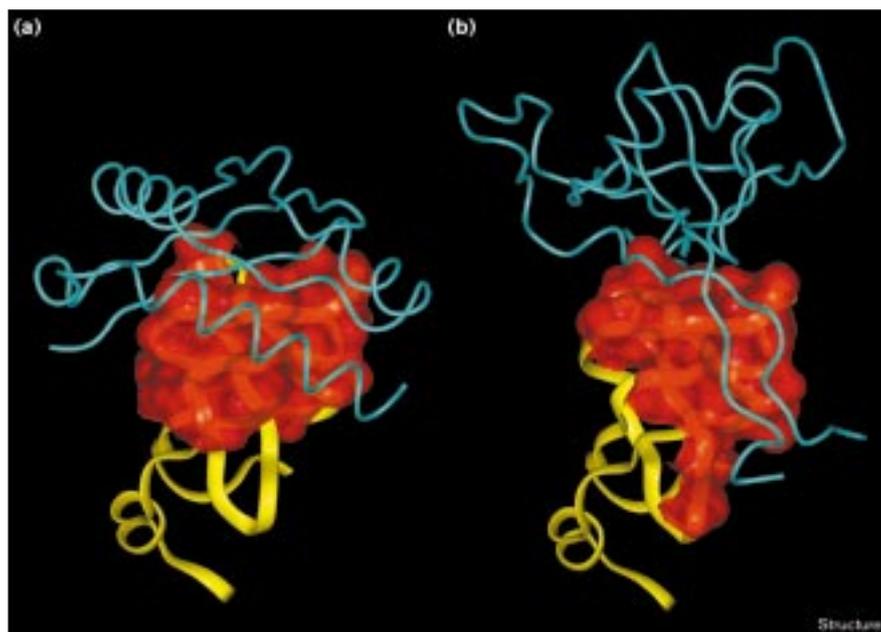
Thr18, Lys22, Asp24, Lys25, Asn32, Tyr33 and Cys53 (Figure 2, red dots). In this case, the contacts are with the termini of g3p N2, which form a long, extended, claw-like motif, and also with the fragments of g3p N2, which do not have structural and topological equivalences in TolA D3. Importantly, even though such contacts are unique to each molecule, a majority of residues from g3p N1 involved in these interactions are adjacent to those forming equivalent interdomain contacts. The first example here is Ser16, which forms hydrogen bonds Oγ···X in both molecules. In the case of g3p-N1–TolA-D3, the interaction occurs with the sidechain nitrogen atom of Asn130 (with an Oγ···N distance of ~2.7 Å), whereas in g3p(N1–N2) an equivalent hydrogen bond is formed with the mainchain oxygen atom of Pro211 (~2.8 Å). Another common fragment participating in the interdomain contacts in g3p N1 is formed by Thr26, Leu27, Asp28 and Arg29. Although the nature of the interactions in both molecules is similar for Thr26 and Asp28 (hydrogen bonds) and partially similar for Leu27 (hydrophobic contacts), we observe contacts that are energetically rather unfavorable for Leu27 and Arg29 in g3p(N1–N2). The same residues in the g3p-N1–TolA-D3 complex seem to participate exclusively in energetically stabilizing interactions (hydrogen bonds in the case of Arg29, and hydrophobic interactions for Leu27).

A long stretch of residues extending from Val43 to Gly48, located on strand β4 of g3p N1, also participates in the contacts between the two domains of g3p-N1–TolA-D3. The leading four residues of this fragment (Val-Val-Val-Cys)

have hydrophobic sidechains; however, in both molecules Val44 and Cys46 interact with the second domain via hydrogen bonds involving only mainchain atoms. Interestingly, in both cases the residues from the other domain in the complex are located in the β strand, which extends the interdomain β sheet, even though it is the first strand in the g3p N2 domain and the last strand in the TolA D3 domain of the fusion protein. The other interdomain contacts by the residues within this stretch involve longer distances and are thus likely to contribute less to the association of the two domains. The remaining common residues in the interface formed by two domains are Glu50, Gln52 and Tyr54. In both molecules, the sidechain of Glu50 is primarily stabilized by the intradomain hydrogen bonds to Arg29. Therefore, relatively distant and rather repulsive contacts with the residues from the second domain seem to be less significant for interdomain stabilization in g3p(N1–N2). However, two other residues form well-defined and energetically favorable interactions stabilizing the interdomain association (the hydrogen bond of Gln52 and hydrophilic interactions of Tyr54). We do not observe comparable favorable interactions involving these residues in the complex with TolA.

Comparison of TolA D3 with other proteins
Using the program DALI [30], we performed a comparison of the TolA D3 domain with the database of unique protein folds. No significantly similar structural motifs were detected by this procedure. The highest score (z = 2.3) involved a superposition of 56 residues from one of the chains of the 547-residue 14mer structure of the

bacterial chaperonin GroEL ([31], PDB code 1der). The apparent similarity was limited to helices α2, 3₁₀, α3, and α4. However, no similarity was detected in the β-sheet region, and no conservation of the amino acid sequence was found. The apparent similarity with phosphatase 2c ([32], PDB code 1a6q) was limited to the two major antiparallel helices α2 and α3, as was the similarity to fragment H of ubiquinol cytochrome c oxidoreductase (cytochrome bc1 [33], PDB code 1bbc-H), which by itself consists of only two helices. On the basis of these comparisons, we must conclude that this comparatively small domain of TolA represents a novel fold. Interestingly, in a homology search of the SwissProt database with the program BLAST [34], using the TolA domain as a query sequence, the highest scoring non-TolA sequence was that of the C-terminal domain of TonB from *Hemophilus ducreyi*.

Discussion

The phage-infection process results in the transfer of genetic material across the cellular envelope and consists of at least three steps. The first step, binding of the phage to the tip of an F pilus, is mediated by the g3p N2 domain [7]. The pilus is then retracted by an unknown mechanism, thereby guiding the bound phage to the cellular envelope. It is not known whether the pilus undergoes continuous alternating synthesis and depolymerization or whether phage binding triggers the retraction [8,35]. Next, an interaction occurs between the phage g3p N1 domain and a coreceptor within the bacterial envelope [12–15]. TolA has recently been shown to be the coreceptor that interacts with g3p N1 [9]. Together with the recently solved structure of the g3p(N1–N2) fragment [27], the g3p-N1–TolA-D3 structure now sheds light on the molecular events during the initial phase of the phage-infection process.

In unbound phage particles, N1 and N2 domains are engaged in tight interactions with each other. Experiments in which purified N1, N2, and N1–N2 compete with whole phages for infection [36] suggested that the N1–N2 complex is the most potent inhibitor, indicating either that the binding site to the pilus is formed by the associated N1–N2 complex or that the complex dissociates, but both N1 and N2 contribute to binding to the pilus. In contrast, it is now clear from the structure of the TolA–N1 complex that TolA D3 binds at a site overlapping that of g3p N2. Thus, during phage binding to the F pilus, the N1 domain must be displaced from the N2 domain and is thus available for binding of the coreceptor in this later step of the infection process. Nevertheless, although the binding sites for N2 and TolA are overlapping, neither structural homology nor even a similarity of secondary structure between the N1-binding interfaces of N2 and TolA is seen (Figure 6).

The further steps of the phage infection process — in particular the role of membrane proteins TolQ and TolR, which are strictly required [12–15] — are still completely unclear. The localization of both the F pili and the Tol complex at adhesion zones of the inner and outer membrane [37] may be of mechanistic importance for the transfer of phage DNA across the cell envelope.

Apart from the Ff phages, a number of other phages and colicins use the TolQRA complex, as well as the structurally homologous TonB–ExbB–ExbD complex, as an entry port [15,26]; the question arises whether the same uptake mechanism is shared by all of these molecules. The highest similarity exists between the Ff phages and the related filamentous phages IKe and I2-2, which carry membrane-penetration domains highly homologous to g3p N1 (although in a different arrangement within the protein architecture of g3p) [38–40]. In contrast to the Ff phages, the latter phages are N-pilus specific and share no detectable sequence homology with the Ff phages within the presumed pilus-binding domain. However, the use of the same coreceptor (TolQRA) suggests a common DNA-uptake mechanism. Previous studies have suggested [5,9,27,36] that an ancestral phage containing only the membrane-penetration domain, capable of infecting *E. coli* at low efficiency [7,36], may have existed, from which N-specific and F-specific phages may have been derived in the course of further evolution, most likely by gene duplication.

TolA is involved not only in the entry of bacteriophages into cells, but also in the uptake of group A colicins [15,26]. Like g3p, colicins consist of an N-terminal translocation domain and a second domain that interacts with a primary receptor on the cell surface, an outer-membrane protein. The toxic functionality of colicins is located in the C-terminal domain. Beyond the similarity in modular architecture, however, no sequence homology exists between the translocation domains of colicins and bacteriophages, although both groups of proteins contain glycine-rich stretches. Although there is now evidence from the structures of g3p(N1–N2) [27] and the g3p-N1–TolA-D3 complex described here that the glycine-rich stretches in g3p are flexible disordered linkers between protein domains (much like the artificial glycine-rich linkers popular in single-chain antibody fragments [41]), a very different function has been ascribed to these stretches in group A colicins. An analogy was postulated to the TonB box of group B colicins [42], a hydrophobic, extended but ordered stretch of sequence of the colicin that is presumed to interact with TonB [43]. One such glycine-rich sequence of group A colicins has been suggested to constitute a TolA box [44], interacting with the coreceptor within the *E. coli* periplasm. More recently, however, the pentapeptide Asp-Gly-Ser-Gly-Trp, located in the first 39 residues of several TolA-dependent colicins, has been proposed to constitute the TolA box [45]. If any of these stretches indeed constitutes a TolA-binding site in colicins, these TolA-dependent colicins must interact with TolA in a manner completely different from g3p.

We wondered whether, despite the lack of sequence homology, there might exist a common fold among the translocation domains of the group A colicins and filamentous phages. Such structural homology has been shown to exist for the N1 and N2 domains of g3p, which share only very limited sequence homology. However, secondary-structure prediction using the PHD program (http://www.embl-heidelberg.de/predictprotein/predictprotein.html) revealed no striking homologies in the arrangement of secondary-structure elements between the translocation domains of colicins and filamentous phages. This result has been confirmed by the recently solved crystal structure of colicin N [25], indicating that the presumed translocation domain of colicin N does not adopt a regular structure. It remains to be shown by more direct means, however, which parts of the structure interact with TolA. Clearly, translocation domains with completely different structures have independently evolved for binding to the same coreceptor.

It thus still remains an open question whether a common translocation mechanism is shared by filamentous phages and colicins. The most obvious difference is that DNA is taken into the cell in the case of bacteriophages, whereas colicins are proteins. However, a number of colicins are inserted into the cytoplasmic membrane (forming ion channels or a further translocation domain to inject an enzymatic activity into the cytoplasm), and the major phage coat protein has been detected within the cytoplasmic membrane after phage infection — it might even be reused in the assembly of new phages from the infected cell [46,47]. Given that TolA is situated at junction regions of the two membranes, a common mechanism for membrane insertion of these proteins is at least conceivable. Such an event might include the formation of a pore large enough for DNA, consisting of TolA, TolR, TolQ, and the membrane anchor of the C-terminal domain of g3p. The C-terminal amino acids of g3p presumably contact the DNA in the phage and might remain bound to it when it contacts the cytoplasmic membrane. Nevertheless, additional components like TolB (required by colicins A, E2–E9, K and N) or TolC (required by colicin E1) may further distinguish colicins and filamentous phages in terms of infection events [26,44].

We have also shown here that the strategy of covalently linking two proteins, which were already known to have a very weak affinity for each other [9], was successful in obtaining a crystal structure of the complex. By designing a long flexible linker, normally present between the N1 and N2 domains of g3p, we imposed no constraints on their relative orientation, but instead forced them to be present at stoichiometric amounts and in close vicinity at all times during crystal formation. We believe that this strategy should be broadly useful in such situations, and thus perhaps of general importance, given the very large number of weak complexes that are known to be involved

in cellular interactions and regulation, both at the surface of the cell and in its interior.

Biological implications

The infection of *E. coli* by the filamentous Ff bacteriophages (M13, fd, and f1) proceeds via the formation of a complex between the phage minor coat protein g3p and the bacterial integral membrane protein TolA. The processes triggered by this interaction are completely unknown; however, their result is the entry of the phage genome into the bacterial cytoplasm. Although the function of g3p, which is crucial for phage infectivity, is known and has been extensively explored in molecular biology, the function of TolA forming the complex with other Tol proteins (TolQ and TolR) within the cytoplasmic membrane is poorly understood.

The crystal structure of the complex between the N-terminal domain of g3p (g3p N1) and the C-terminal domain of TolA (TolA D3), obtained by fusing these two domains with a long flexible linker, shows in detail how the two proteins interact with each other. It is evident that the g3p N1 domain uses the very same epitope during recognition and binding to its coreceptor, TolA, as found previously in the two-domain (N1–N2) fragment of g3p. Thus, the interaction between g3p and TolA is accomplished by two well-defined compact protein domains, whereas group B colicins interact with TonB, probably via a hydrophobic, extended region (TonB box) and group A colicins have been proposed to interact with TolA via a hydrophobic stretch or even a glycine-rich stretch. These differences illustrate that divergent evolutionary paths have led to the appearance of macromolecules that can enter *E. coli*.

Materials and methods
Protein production and purification
We assembled the gene encoding a fusion protein (Figure 2) comprising residues 1–86 of mature g3p [48] and residues 295–421 of TolA [49] by the polymerase chain reaction (PCR) using the oligonucleotides N1for (GGAGATATATCCATGGCTGAAACTGTTGAAAG), N1Back (ACCCTCGGATCCGCCACCCTCAGAGCCACCACCCT-CATTTTCAGG), TolACTfor (GGTGGCGGATCCGAGGGTGGCG-GTTCTGAGGGTGGCGGTGATGATATTTTCGGTG) and TolAHis (GAATTCAAGCTTACTAGTGATGGTGATGGTGATGTGCGGCCG CCGGTTTGAAGTCCAATG). This construct encodes a protein starting with an N-terminal methionine residue (which is excised by methionyl aminopeptidase [50]), residues 1–86 of mature g3p (which comprise the N1 domain and the first glycine-rich linker ((Gly$_3$SerGlu)$_3$Gly$_3$) of g3p, residues 295–421 of TolA, and a C-terminal tail with the sequence Ala$_3$His$_6$ (Figure 2). The fusion gene was cloned into the *Nco*I–*Hind*III sites of expression vector pTFT74 [51,52] and expressed in *E. coli* BL21(DE3) [53]. The recombinant protein was purified from cytoplasmic inclusion bodies in the presence of 8 M urea by using the coupled IMAC-AIEX protocol [54] and refolded by dialysis against 0.2 M Tris-HCl, 0.4 M arginine, 0.2 M guanidinium hydrochloride, 0.1 M (NH$_4$)$_2$SO$_4$, 2 mM EDTA, 1 mM oxidized glutathione, 0.2 mM reduced glutathione (pH 8.5). After dialysis against 50 mM Tris-HCl (pH 7.5), the protein was subjected to AIEX chromatography using a perfusion chromatography HQ column on a BioCAD 60 system (Perseptive Biosystems). The eluate was concentrated to 12 mg/ml

and subjected to gel filtration on a Sephacryl S-100 26/60 column (Pharmacia) in 50 mM Tris-HCl, 150 mM NaCl (pH 7.5). Processing of the N-terminal methionine residue, resulting in the correct N terminus of g3p, was verified by mass spectroscopy. The yield obtained from 7.5 liters of *E. coli* culture was 75 mg of pure protein.

Crystallization

Crystals of g3p-N1–TolA-D3 were obtained either by the hanging-drop or sitting-drop vapor-diffusion methods at 15°C. Crystals could also be grown at room temperature and at 5°C; however, their quality was lower in these cases. Before crystallization, protein stock solutions were transferred into 50 mM Hepes buffer (pH 7.5) containing 2 mM DTT and dialyzed against the same Hepes–DTT solution over three days, with four buffer changes. Subsequently, the protein solutions were concentrated to ~8–12 mg/ml. Drops were formed by mixing equivalent volumes of protein and precipitant solutions. The precipitant contained 24.6% (v/v) PEG4000, 0.08 M Tris buffer (pH 8.5), 0.15 M sodium acetate, and 2 mM DTT. The first crystals usually appeared after 6–8 days and reached their final size (0.3 mm × 0.3 mm × 0.2 mm) after an additional 10–15 days. They belong to the tetragonal system, space group $P4_32_12$ (the enantiomorph was established during structure determination), and the unit-cell dimensions at room temperature are $a = b = 88.949$ Å, $c = 64.314$ Å, whereas at 100K they are $a = b = 88.880$ Å, $c = 63.556$ Å. Each crystallographic asymmetric unit contains one molecule of the fusion protein g3p-N1–TolA-D3.

X-ray analysis

A number of different X-ray diffraction data sets were collected from native crystals of g3p-N1–TolA-D3, both at room temperature as well as at 100K. All X-ray data for the heavy-atom derivatives were collected at room temperature. The room-temperature data were collected using a standard X-ray generator, whereas the low-temperature data were also acquired at a synchrotron source. For cryo-experiments, crystals were flash-frozen in a stream of cold nitrogen generated by the Oxford CryoSystems device. Before flash-freezing, crystals were transferred to a cryoprotecting solution, corresponding to the precipitant solution containing 7.5% (v/v) glycerol, and equilibrated for 2–5 min. The laboratory X-ray source was a Rigaku Ru200 generator operated at 50 kV and 100 mA. The intensities were recorded with a MAR345 image-plate detector. The final data set for the native protein was collected at 100K, using an ADSC four-element charge coupled device (CCD) detector on beamline F2 of the CHESS synchrotron facility at Cornell University. The X-ray wavelength during data collection was 0.98 Å. Statistics for data collection are shown in Table 1.

Initially, molecular-replacement calculations were performed using the program AMoRe [55], with the previously published N-terminal domain of g3p [27] used as the search model. The molecular replacement was completed in both enantiomorphic space groups, under the same conditions (resolution limits, Patterson integration radius, etc.). The calculations were repeated several times, for different resolution ranges. We observed an outstanding solution in the space group $P4_32_12$, which appeared independently of data resolution and was characterized by very good molecular packing. Depending on the resolution of X-ray data, this solution was higher than the remaining peaks by 3–6% in terms of the correlation coefficient. As the predicted identities of the residues of the N-terminal domain of g3p interacting with the C-terminal domain of TolA have been suggested by Riechmann and Holliger [9], we also confirmed that the packing resulting from the molecular replacement would not rule out such interactions. Although we realized from the beginning that the contribution of the smaller domain would not be sufficient to properly phase the rest of the structure, this approach nevertheless provided very valuable information, making the subsequent isomorphous-replacement calculations much easier. Molecular replacement fixed the space group enantiomorph unambiguously and helped in solving the positions of heavy atoms. In addition, the knowledge of the location of the g3p N1 domain helped in the interpretation of the final electron-density map.

Data from crystals of two heavy-atom derivatives were utilized in the multiple isomorphous replacement with anomalous scattering (MIRAS) calculations. The heavy-atom sites were determined for the mercury derivative first, using the program SHELXS-97 [56] We found consistent positions for three sites by either direct or Patterson methods, utilizing both isomorphous and anomalous differences. These strong peaks were also visually identified on the Harker sections, calculated by PHASES [57] for isomorphous and for anomalous differences. The position of the uranyl site was determined with SHELXS-97 from the isomorphous differences for this derivative. Even though it was not as apparent as the mercury sites, we confirmed its position by applying the difference Fourier methods to the X-ray data for both derivatives. Such an approach also removed the origin ambiguity.

Subsequently, we proceeded with the traditional MIRAS phase refinement in PHASES, extending the resolution to 3.3 Å. The resulting figure of merit (FOM) was 0.63. This step was followed by density modification and histogram matching approaches, performed with the program DM (part of the CCP4 suite [58]), which resulted in additional improvement of statistics (FOM 0.76, and free R factor equal to 0.47). Although some details were still ambiguous, the electron-density maps were already largely interpretable and we could clearly see the electron density corresponding to the N-terminal domain of g3p, as determined by molecular replacement. Before we began model building, we performed phase refinement according to the maximum-likelihood principle as implemented in SHARP [59]. In this approach, we used all available X-ray data and allowed the program to select the suitable cutoffs. The phases obtained by this procedure were then modified using DM in a manner similar to that described above, which resulted in values of 0.85 and 0.41 for the FOM and free R factor, respectively. The electron-density maps obtained by this procedure were readily interpretable (Figure 5). Placing the N-terminal domain of g3p was quite straightforward, making only small local corrections. In the case of the TolA fragment, assignment of the appropriate sequence to the corresponding electron density was not completely obvious. We could, however, clearly see the helical features in the electron density corresponding to this fragment. To support model building, we tried to correlate the helical features of the density with the helices that were theoretically predicted on the basis of the amino acid sequence. These calculations were performed with the program PHD (http://www.embl-heidelberg.de/predictprotein/predictprotein.html), which predicted the helical conformation for residues Asn130–Phe143, Pro177–Lys191 and Gln198–Phe204. As seen in the final structure, these predictions were very good.

Structure refinement

The structure resulting from the model-building step described above was initially refined with X-PLOR [60], whereas all subsequent refinement steps were performed with SHELXL-97 [61]. Residues 1–65 from the N1 domain of g3p were all modeled, although both termini of this domain had high temperature factors. Residues 125–215 belonging to the D3 domain of TolA were also fitted. The structure refined well at 2.5 Å, with tight geometry (Table 2) and with all mainchain torsion angles occupying either the most favorable or additionally allowed regions of the Ramachandran plot. A limited number of solvent molecules were fitted into the electron density. There was no indication of any ordered protein for the linker region between the two domains.

The structure was subsequently refined with higher-resolution data, which were collected on frozen crystals using synchrotron radiation. Unlike the R factors for the room-temperature model, those of the low-temperature model remained significantly higher, although no electron density that would suggest a different tracing than for the room-temperature structure could be seen. Surprisingly, the residues that did not fit the density very well in the final map are mostly found in the N1 domain of g3p, although the corresponding residues in the previous structure did not show any problems. The residues affected are found in loops 22–25 and 47–48. The relatively high crystallographic R factors were observed not only for the data set collected with synchrotron radiation, but also for

two other low-temperature data sets that were collected on a laboratory-based X-ray source (Table 1). Two such data sets were refined independently, with the results similar to those for the synchrotron data set, although the resolution and quality of the latter data were very similar to those of the room-temperature data set. We conclude that scattering by the stretch of ~60 disordered amino acids could be responsible for this phenomenon, which has also been observed in refinements of other proteins in our laboratories and those of others [62]. An example of the final $2F_o$–F_c electron-density map is shown in Figure 5.

Accession numbers

The coordinates of the fusion protein structure described in this paper and the corresponding experimental structure factors have been deposited with the Protein Data Bank with the accession codes 1tol and r1tolsf, respectively.

References

1. Smith, G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315-1317.
2. Phizicky, E.M. & Fields, S. (1995). Protein-protein interactions: methods for detection and analysis. *Microbiol. Rev.* **59**, 94-123.
3. Cortese, R., *et al.*, & Felici, F. (1996). Selection of biologically active peptides by phage display of random peptide libraries. *Curr. Opin. Biotechnol.* **7**, 616-621.
4. Felici, F., Luzzago, A. Monaci, P., Nicosia, A., Sollazzo, M., & Traboni, C. (1995). Peptide and protein display on the surface of filamentous bacteriophage. *Biotechnol. Annu. Rev.* **1**, 149-183.
5. Spada, S., Krebber, C. & Plückthun, A. (1997). Selectively infective phages (SIP). *Biol. Chem.* **378**, 445-456.
6. Gray, C.W., Brown, R.S. & Marvin, D.A. (1981). Adsorption complex of filamentous fd virus. *J. Mol. Biol.* **146**, 621-627.
7. 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.
8. Jacobson, A. (1972). Role of F pili in the penetration of bacteriophage fl. *J. Virol.* **10**, 835-843.
9. 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.
10. Click, E.M. & Webster, R.E. (1997). Filamentous phage infection: required interactions with the TolA protein. *J. Bacteriol.* **179**, 6464-6471.
11. 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.
12. 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.
13. 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.
14. 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.
15. Webster, R.E. (1991). The *tol* gene products and the import of macromolecules into *Escherichia coli*. *Mol. Microbiol.* **5**, 1005-1011.
16. Braun, V. & Herrmann, C. (1993). Evolutionary relationship of uptake systems for biopolymers in *Escherichia coli*: cross-complementation between the TonB-ExbB- ExbD and the TolA-TolQ-TolR proteins. *Mol. Microbiol.* **8**, 261-268.

17. Braun (1995). Energy-coupled transport and signal transduction through the gram-negative outer membrane via TonB-ExbB-ExbD-dependent receptor proteins. *FEMS Microbiol. Rev.* **16**, 295-307.
18. Moeck, G.S. & Coulton, J.W. (1998). TonB-dependent iron acquisition: mechanisms of siderophore-mediated active transport. *Mol. Microbiol.* **28**, 675-681.
19. Lazzaroni, J.C., Fognini-Lefebvre, N. & Portalier, R. (1989). Cloning of the *excC* and *excD* genes involved in the release of periplasmic proteins by *Escherichia coli* K12. *Mol. Gen. Genet.* **218**, 460-464.
20. Lazzaroni, J.C. & Portalier, R.C. (1981). Genetic and biochemical characterization of periplasmic-leaky mutants of *Escherichia coli* K-12. *J. Bacteriol.* **145**, 1351-1358.
21. Levengood-Freyermuth, S.K., Click, E.M. & Webster, R.E. (1993). Role of the carboxyl-terminal domain of TolA in protein import and integrity of the outer membrane. *J. Bacteriol.* **175**, 222-228.
22. Rampf, B., Bross, P., Vocke, T. & Rasched, I. (1991). Release of periplasmic proteins induced in *E. coli* by expression of an N-terminal proximal segment of the phage fd gene 3 protein. *FEBS Lett.* **280**, 27-31.
23. Amouroux, C., Lazzaroni, J.C. & Portalier, R. (1991). Isolation and characterization of extragenic suppressor mutants of the *tolA-876* periplasmic-leaky allele in *Escherichia coli* K-12. *FEMS Microbiol. Lett.* **62**, 305-313.
24. Hancock, R.W. & Braun, V. (1976). Nature of the energy requirement for the irreversible adsorption of bacteriophages T1 and phi80 to *Escherichia coli*. *J. Bacteriol.* **125**, 409-415.
25. 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.
26. Lazdunski, C.J. (1995). Colicin import and pore formation: a system for studying protein transport across membranes? *Mol. Microbiol.* **16**, 1059-1066.
27. Lubkowski, J., Hennecke, F., Plückthun, A. & Wlodawer, A. (1998). The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. *Nat. Struct. Biol.* **5**, 140-147.
28. Hutchinson, E.G. & Thornton, J.M. (1996). PROMOTIF – A program to identify and analyze structural motifs in proteins. *Prot. Sci.* **5**, 212-220.
29. Holliger, P. & Riechmann, L. (1997). A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. *Structure* **5**, 265-275.
30. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.
31. Boisvert, D.C., Wang, J., Otwinowski, Z., Horwich, A.L. & Sigler, P.B. (1996). The 2.4 Å structure of the bacterial chaperonin GroEL complexed with ATP gamma S. *Nat. Struct. Biol.* **3**, 170-177.
32. Das, A.K., Helps, N.R., Cohen, P.T. & Barford, D. (1996). Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 Å resolution. *EMBO J.* **15**, 6798-6809.
33. Zhang, Z., *et al.*, & Kim, S.H. (1998). Electron transfer by domain movement in cytochrome bc1. *Nature* **392**, 677-684.
34. Altschul, S.F., *et al.*, & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.
35. Frost, L.S., Ippen-Ihler, K. & Skurray, R.A. (1994). Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* **58**, 162-210.
36. Krebber, C., Spada, S., Desplancq, D., Krebber, A., Ge, L. & Plückthun, A. (1997). Selectively-infective phage (SIP): a mechanistic dissection of a novel *in vivo* selection for protein-ligand interactions. *J. Mol. Biol.* **268**, 607-618.
37. Guihard, G., Boulanger, P., Benedetti, H., Lloubes, R., Besnard, M. & Letellier, L. (1994). Colicin A and the Tol proteins involved in its translocation are preferentially located in the contact sites between the inner and outer membranes of *Escherichia coli* cells. *J. Biol. Chem.* **269**, 5874-5880.
38. Peeters, B.P.H., Peters, R.M., Schoenmakers, J.G.G. & Konings, R.N.H. (1985). Nucleotide sequence and genetic organizaton of the genome of the N-specific filamentous bacteriophage Ike. Comparison with the genome of the F-specific filamentous phages M13, fd and fl. *J. Mol. Biol.* **181**, 27-39.
39. Strassen, A.P.M., Schoenmakers, E.F.P.M., Yu, M., Schoenmakers, J.G.G., & Konings, R.N.H. (1992). Nucleotide sequence of the genome of the filamentous bacteriophage I2-2: module evolution of the filamentous phage genome. *J. Mol. Evol.* **34**, 141-152.

40. Endemann, H., Bross, P. & Rasched, I. (1992). The adsorption protein of phage IKe. Localization by deletion mutagenesis of domains involved in infectivity. *Mol. Microbiol.* **6**, 471-478.

41. Huston, J.S., *et al.*, & Oppermann, H. (1991). Protein engineering of single-chain Fv analogs and fusion proteins. *Methods Enzymol.* **203**, 46-88.

42. Schramm, E., Mende, J., Braun, V. & Kamp, R.M. (1987). Nucleotide sequence of the colicin B activity gene cba: consensus pentapeptide among TonB-dependent colicins and receptors. *J. Bacteriol.* **169**, 3350-3357.

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

44. Schendel, S.L., Click, E.M., Webster, R.E. & Cramer, W.A. (1997). The TolA protein interacts with colicin E1 differently than with other group A colicins. *J. Bacteriol.* **179**, 3683-3690.

45. Garinot-Schneider, C., Penfold, C.N., Moore, G.R., Kleanthous, C. & James, R. (1997). Identification of residues in the putative TolA box which are essential for the toxicity of the endonuclease toxin colicin E9. *Microbiology* **143**, 2931-2938.

46. Wickner, W. (1975). Asymmetric orientation of a phage coat protein in cytoplasmic membrane of *Escherichia coli. Proc. Natl Acad. Sci. USA* **72**, 4749-4753.

47. Armstrong, J., Hewitt, J.A. & Perham, R.N. (1983). Chemical modification of the coat protein in bacteriophage fd and orientation of the virion during assembly and disassembly. *EMBO J.* **10**, 1641-1646.

48. Ebright, R., Dong, Q. & Messing, J. (1992). Corrected nucleotide sequence of M13mp18 gene III. *Gene* **114**, 81-83.

49. Levengood, S.K. & Webster, R.E. (1989). Nucleotide sequences of the *tolA* and *tolB* genes and localization of their products, components of a multistep translocation system in *Escherichia coli. J. Bacteriol.* **171**, 6600-6609.

50. Hirel, P.H., Schmitter, M.J., Dessen, P., Fayat, G. & Blanquet, S. (1989). Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the sidechain length of the penultimate amino acid. *Proc. Natl Acad. Sci. USA* **86** , 8247-8251.

51. Ge, L., Knappik, A., Pack, P., Freund, C. & Plückthun, A. (1995). Expressing antibodies in Escherichia coli. In *Antibody Engineering.*, (Borrebaeck, C.A.K., ed.), pp. 229-266, Oxford University Press, Oxford.

52. Freund, C., Ross, A., Guth, B., Pluckthun, A. & Holak, T.A. (1993). Characterization of the linker peptide of the single-chain Fv fragment of an antibody by NMR spectroscopy. *FEBS Lett.* **320**, 97-100.

53. Studier, F.W. & Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113-130.

54. Plückthun, A., *et al.*, & Riesenberg, D. (1996). Producing antibodies in *Escherichia coli*: from PCR to fermentation. In *Antibody Engineering: A Practical Approach* (McCafferty, J., Hoogenboom, H.R. & Chiswell, D.J., eds), pp. 203-252, IRL Press, Oxford.

55. Navaza, J. (1994). An automated package for molecular replacement. *Acta Cryst. A* **50**, 157-163.

56. Sheldrick, G.M. (1997). Patterson superposition and *ab initio* phasing. In *Methods in Enzymology.* (Carter, C.W., Jr. & Sweet, R.M., eds), pp. 628-641, Academic Press, New York, NY.

57. Furey, W. & Swaminathan, S. (1990). PHASES – A program package for the processing and analysis of diffraction data for macromolecules. *Acta Crystallogr.* **18**, 73.

58. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763.

59. de La Fortelle, E.B.G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In *Methods in Enzymology.* (C.W. Carter, Jr. & R.M. Sweet., eds), pp. 472-494, Academic Press, NY.

60. Brünger, A.T. (1992). *X-PLOR version 3.1: A System for X-ray Crystallography and NMR.* Yale University Press, New Haven, CT.

61. Sheldrick, G.M. & Schneider, T.R. (1997). SHELXL: high-resolution refinement. In *Methods in Enzymology.* (C.W. Carter, Jr. & R.M. Sweet., eds), pp. 319-343, Academic Press, New York, NY.

62. Zdanov, A., Schalk-Hihi, C., Gustchina, A., Tsang, M., Weatherbee, J. & Wlodawer, A. (1995). Crystal structure of interleukin-10 reveals the functional dimer with an unexpected topological similarity to interferon γ. *Structure* **3**, 591-601.

63. Carson, M. (1991). RIBBONS 4.0. *J. Appl. Crystallogr.* **24**, 958-961.

64. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.