The Tailspike Protein of Shigella Phage Sf6

A STRUCTURAL HOMOLOG OF SALMONELLA PHAGE P22 TAILSPIKE PROTEIN WITHOUT SEQUENCE SIMILARITY IN THE β-HEXILX DOMAIN*

Alexander Freiberg‡, Renato Morona§, Luisa Van Den Bosch§, Christiane Jung¶, Joachim Behlke¶, Nils Carlini, Robert Seckler**", and Ulrich Baxa‡

From the ‡Physikalische Biochemie, Universität Potsdam, Karl-Liebknecht-Strasse 24–25, D-14476 Golm, Germany, the §Department of Molecular Biosciences, University of Adelaide, Adelaide, South Australia 5005, Australia, ¶Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Strasse 10, D-13122 Berlin, Germany, and the ¶Department of Molecular Biology, SBL Vaccin AB, 105 21 Stockholm, Sweden

Bacteriophage Sf6 tailspike protein is functionally equivalent to the well characterized tailspike of Salmonella phage P22, mediating attachment of the viral particle to host cell-surface polysaccharide. However, there is significant sequence similarity between the two 70-kDa polypeptides only in the N-terminal putative capsid-binding domains. The major, central part of P22 tailspike protein, which forms a parallel β-helix and is responsible for saccharide binding and hydrolysis, lacks detectable sequence homology to the Sf6 protein. After recombinant expression in Escherichia coli as a soluble protein, the Sf6 protein was purified to homogeneity. As shown by circular dichroism and Fourier transform infrared spectroscopy, the secondary structure contents of Sf6 and P22 tailspikes are very similar. Both tailspikes are thermostable homotrimers and resist denaturation by SDS at room temperature. The specific endo-hamnosidase activities of Sf6 tailspike protein toward fluorescence-labeled dodeca- and deca- and octasaccharide fragments of Shigella O-antigen suggest a similar active site topology of both proteins. Upon deletion of the N-terminal putative capsid-binding domain, the protein still forms a thermostable, SDS-resistant trimer that has been crystallized. The observations strongly suggest that the tailspike of phage Sf6 is a trimeric parallel β-helix protein with high structural similarity to its functional homolog from phage P22.

The Shigella flexneri phage Sf6 is morphologically similar to the Salmonella phage P22. Both are members of the class C bacteriophages (1) consisting of an icosahedral head and a short tail containing six tailspike proteins responsible for the binding and hydrolysis of the receptor O-antigen. Phages are classified mainly by their morphology, but an evolutionary relationship of all tailed phages is assumed (2, 3). To verify this relationship, it will be helpful not only to rely on sequence similarities but also on the much more strongly conserved folding topologies of homologous proteins. The gene of Sf6 tailspike protein (TSP) has been cloned on the basis of its high sequence identity to the P22 TSP gene in the part coding for the N-terminal head-binding domain (4). Quite surprisingly no sequence identity was found in the major central and C-terminal parts of the 70-kDa proteins. The central part harbors the O-antigen-binding sites of P22 TSP. High resolution crystal structures of both the head-binding and the O-antigen-binding part of the homotrimeric P22 TSP have been determined (5, 6). The central part of P22 TSP consists of right-handed parallel β-helices, associated side-by-side, whereas the subunits strongly interdigitate in the C-terminal part (5). The short peptide linking the N-terminal domain to the major central and C-terminal part of P22 TSP is thought to be quite flexible (7). Both proteins, the P22 and the Sf6 TSP, are endohamnosidases but function on different O-antigen substrates (8, 9). The end products in both cases are dimers of the repeating unit (in both cases an octasaccharide), but no hydrolysis of Shigella O-antigen treated with P22 TSP was observed and vice versa (4). The interaction of P22 TSP with O-antigen fragments has been investigated by x-ray crystallography and in solution by biophysical techniques (10–12). TSP binds one oligosaccharide per subunit with micromolar affinity, and the binding site for octasaccharide is a groove running parallel to the β-helix axis along its solvent-exposed face. The active site is situated at the reducing end of the octasaccharide product seen in the complex structure (11). Specificity for Salmonella O-antigen is reached by a large contact surface involving all sugar residues in the octasaccharide, explaining the unusually high change in heat capacity upon saccharide binding (12).

In this paper, we report on the characterization of Sf6 TSP using biochemical, spectroscopic, and hydrodynamic techniques. We found that Sf6 TSP is a homotrimeric protein with a stability similar to that of P22 TSP. Circular dichroism and Fourier transform infrared spectroscopy indicated that the secondary structure contents of Sf6 and P22 TSP are very similar, thus suggesting similar three-dimensional structures. In analogy to previous experiments on P22 TSP (13, 14), we produced a C-terminal 60-kDa fragment of the Sf6 tailspike polypeptide lacking the putative capsid-binding domain. This large C-terminal fragment, like the corresponding part of the P22 protein, is a homotrimer and resistant to SDS at room temperature, despite the lack of significant sequence identity between both proteins in these parts. The crystallization of this major C-

* This work was supported by the Deutsche Forschungsgemeinschaft and by the Fonds der Chemischen Industrie. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** To whom correspondence should be addressed: Physikalische Biochemie, Universität Potsdam, Karl-Liebknecht-Str. 24–25, D-14476 Golm, Germany. Tel.: 49-331-977-5240; Fax: 49-331-977-5062; E-mail: seckler@rz.uni-potsdam.de.

‡ The abbreviations used are: TSP, tailspike protein; FTIR, Fourier transform infrared; MES, 4-morpholineethanesulfonic acid; RU, repeating units.
terminal fragment of Sf6 tailspike protein is exceedingly slow, the solution was incubated for 2 days or more at 4 °C to ensure the completeness of precipitation. The precipitate was again resuspended in buffer A, dialyzed against the same buffer, and applied to a DE52 anion exchange column (Whatman) equilibrated with buffer A. Fractions of a linear gradient (0–300 mM NaCl in buffer A) were pooled, brought to 0.8 M ammonium sulfate by addition of a concentrated solution, and applied to a phenyl-Sepharose FF column (Amersham Biosciences). The proteins were eluted with a linear gradient of 0.8–0 M ammonium sulfate in buffer A and were concentrated by ultrafiltration (Amicon). The last impurities were removed by gel filtration on a Superdex 200 column (Amersham Biosciences) in buffer B (20 mM Tris/HisCl, 1 mM EDTA, 200 mM NaCl, pH 7.0). Purified full-length or N-terminally shortened S6 tailspike proteins could be concentrated to about 20 mg/ml by ultrafiltration without showing strong tendency for aggregation. Gel Filtration—Analytical gel filtration was performed at room temperature on a Superdex 200 HR size exclusion column (30 cm × 1 cm; Amersham Biosciences) in buffer B at a flow rate of 0.5 ml/min. Proteins used as molecular mass standards were ferritin (450 kDa), catalase (240 kDa), P22 TSP (215 kDa), P22 TSPAN (180 kDa), aldolase (158 kDa), lactate dehydrogenase (140 kDa), hen egg albumin (45 kDa), Thiolus salivarius pectate lyase (43 kDa), and cytochrome c (12.5 kDa). 50 μl of S6 TSP (1.0 mg/ml) and standard proteins (1.0 mg/ml) dissolved in the buffer B were applied to the column, and protein fluorescence (excitation at 280 nm and emission at 340 nm) was detected.

Analytical Ultracentrifugation—Sedimentation equilibrium measurements were performed using an XL-A analytical ultracentrifuge (Beckman Instruments, Palo Alto, CA) equipped with UV absorbance optics. The proteins were dissolved at concentrations from 0.14 to 0.42 mg/ml in 20 mM Tris/HisCl, pH 7.0, 200 mM NaCl, 1 mM EDTA. To determine the apparent molecular mass (M), radial absorbance distributions at sedimentation equilibrium were recorded at three different wavelengths (280, 285, and 290 nm) and fitted globally to Equations 1 and 2.

![Figure](https://via.placeholder.com/150)

Using the program Polymole (22). In these equations \( \rho \) is the solvent density; \( \delta_p \) is the partial specific volume of the protein; \( \omega \) is the angular velocity; \( c \) is the absolute temperature; \( a_p \) is the radial absorbance, and \( a_p \) is the corresponding value at the meniscus position. The molecular mass was determined by extrapolation of the apparent data to infinite dilution according to Equation 3.

Crystallization—Crystals were grown in hanging drops in cell culture plates (24 wells sealed with cover slips). Protein solutions of about 12 mg/ml were dialyzed against 10 mM sodium phosphate, pH 7.0, and 0.1% (Novagen). A clone having the PCR fragments in an N-terminal fragment of S6 tailspike protein was precipitated in equal volumes (1 μl) of protein and precipitant solution and were suspended over 0.5 ml of precipitant solution at 20 °C. By using 0.1 M MES, pH 6.0, 18% PEG 8000 as the precipitant solution, crystals appeared within 2 weeks.
RESULTS

Purification and Solubility—Expression of the gene coding for Sf6 TSP in E. coli in the absence of Sf6 phage heads or any other Sf6 components resulted in over-produced material in the soluble fraction of cell lysates. After purification as described under “Experimental Procedures,” the recombinant Sf6 TSP was judged to be at least 98% pure, because no additional bands were detectable on silver-stained SDS gels at high sample loads.

State of Association—Gel filtration and analytical ultracentrifugation were used to determine the molecular size of Sf6 TSP. The elution volume of Sf6 TSP was almost identical to that of P22 TSP. Both proteins eluted slightly earlier than expected for a trimer on the basis of the column calibration made with globular proteins (Fig. 1A). This discrepancy may be explained by the somewhat elongated shape of the two tailspike proteins. However, on the basis of the gel filtration results, a tetrameric association state cannot be ruled out. Therefore, we determined the molecular mass of Sf6 TSP independently by analytical ultracentrifugation (Fig. 1B). Sedimentation equilibrium runs were done at different initial concentrations, and the resulting apparent molecular masses were extrapolated to infinite dilution. The molecular mass so determined was 201.2 kDa, close to the value of 202 kDa expected for a homotrimer and a uni-unimolecular model N → I → U for P22 TSP, where N is the native protein, I is the unfolding intermediate, and U is the unfolded protein. The time constants resulted in $k_1 = 0.033 \text{ min}^{-1}$ for Sf6 TSP, and $k_1 = 0.79 \text{ min}^{-1}$ and $k_2 = 0.047 \text{ min}^{-1}$ for P22 TSP.

Similar to P22 TSP, Sf6 TSP was found to be resistant to denaturation by SDS at room temperature. Unheated samples migrated with an apparent molecular mass of about 180 kDa on SDS-polyacrylamide gels, whereas a band at about 67 kDa, the molecular mass expected for the monomer, was observed when the samples had been heated to 99 °C for 3 min prior to electrophoresis.

The SDS resistance of the Sf6 TSP trimer allowed us to use SDS gel electrophoresis in order to analyze the time course of thermal denaturation of the protein. The kinetics of thermal denaturation in the presence of 2% SDS at 72 °C are compared for the Sf6 and P22 TSPs in Fig. 2. Only two bands were observed for Sf6 TSP, corresponding to the native protein and the denatured monomer (Fig. 2B). This is in contrast to the heat denaturation of P22 TSP, where an additional intermediate band is observed (Fig. 2A). It has been shown previously (25) that the N-terminal domain is unfolded in this intermedi-
TSP (thin dashed lines) infrared spectra in the amide band region of Sf6 (solid line) and P22 TSP (broken line). Single peaks obtained by fitting the non-deconvoluted spectrum of Sf6 TSP are shown in thin dashed lines (compare Table I).

Secondary Structure—The secondary structure content of proteins is commonly determined by far-UV circular dichroism or Fourier transform infrared spectroscopy. Both were used to compare Sf6 and P22 TSP (Fig. 3). In both methods, shape and amplitudes of the spectra were similar between Sf6 and P22 TSP in the regions that are indicative for β-structure. Both methods reveal a very high content of β-structure and suggest that both proteins are essentially devoid of α-helices. Specific differences between the two tailspike proteins are the amplitude of the far-UV CD peak at about 195 nm and the exact position of the main peak in the infrared spectra, which was observed at 1638 and 1635 cm⁻¹ for Sf6 and P22 TSP, respectively. Because FT-IR is a better method for estimating the secondary structure content for all-β-proteins (26), only the IR spectra were analyzed quantitatively. The minimum fit model (18) for the non-deconvoluted spectra of both P22 and Sf6 TSPs was realized by 5 Voigt bands with predominantly Gauss character. For fitting the deconvoluted spectra, which showed a higher amide I band resolution, 10 Voigt bands with predominantly Gauss character gave the best result. The frequency of the single bands assigned to the different structural elements was very similar to the frequencies found for other proteins (18). The results summarized in Tables I and II show that the secondary structure contents of Sf6 and P22 TSPs are identical within experimental error, regardless of the model used.

Absorbance, fluorescence, and near-UV CD spectra of Sf6 TSP did not show close similarities to the corresponding spectra of P22 TSP, although the content of aromatic side chains in the two proteins is similar. Nevertheless, these methods verified the well-defined tertiary structure of the native Sf6 TSP. Fluorescence emission of the native protein showed a maximum at 342 nm. Denaturation in 6 M guanidinium chloride shifted the maximum to 355 nm, as expected for tryptophan fluorescence in aqueous solution but did scarcely influence the fluorescence emission amplitude. The near-UV circular dichroism spectrum revealed well defined peaks at 278, 285, and 293 nm. Obviously, the environment of the aromatic side chains in Sf6 and P22 TSP is rather different, as could be expected from the different sequences in the C-terminal part.

Active Site Structure—Shigella O-antigen serotype Y consisting of repeating units of α-L-Rhap-(1,2)-α-L-Rhap-(1,3)-α-L-Rhap-(1,3)-β-D-GlcNAc-(1,2)-16 is hydrolyzed by Sf6 TSP in the α-L-Rhap-(1,3)-α-L-Rhap bond, with the main product containing 2 repeating units (octasaccharide) (8). For a more detailed analysis of the enzymatic activity of Sf6 TSP, we performed hydrolysis assays with fragments of the Shigella O-antigen labeled at their reducing ends with the fluorescent dye amino-methyl-coumarin (Fig. 4). Octasaccharide (2 RU), dodecasaccharide (3 RU), and decasaccharide were used as substrates. The latter results from the nonreducing end of the O-antigen polysaccharide chains and contains 2 RU with an α-L-Rhap-(1,2)-α-L-Rhap-(1,3)-unit at the nonreducing end (16). Hydrolysis could be followed by separating samples after different reaction times by reversed phase high pressure liquid chromatography. For all three substrates, the only fluorescence-labeled product was tetrasaccharide. No coumarin-labeled octasaccharide was produced from labeled decasaccharide or dodecasaccharide. Labeled octasaccharide...
and decasaccharide were hydrolyzed very slowly. At 2.2 μM oligosaccharide, the observed initial rates of enzymatic turn-over were 3 × 10^{-8} and 1 × 10^{-4} s^{-1}, respectively, more than 100-fold lower than the rate for labeled decasaccharide, which was 0.13 s^{-1}. Still, decasaccharide was hydrolyzed significantly faster than octasaccharide (Fig. 4). These features point to a minimum architecture of the binding and active site of the Sf6 tailspike endo-ramnosidase, where at least two RU are necessary for efficient binding, and the hydrolysis reaction takes place at the reducing end of these two repeating units.

and decasaccharide were hydrolyzed very slowly. At 2.2 μM oligosaccharide, the observed initial rates of enzymatic turn-over were 3 × 10^{-8} and 1 × 10^{-4} s^{-1}, respectively, more than 100-fold lower than the rate for labeled decasaccharide, which was 0.13 s^{-1}. Still, decasaccharide was hydrolyzed significantly faster than octasaccharide (Fig. 4). These features point to a minimum architecture of the binding and active site of the Sf6 tailspike endo-ramnosidase, where at least two RU are necessary for efficient binding, and the hydrolysis reaction takes place at the reducing end of these two repeating units.

DISCUSSION

Fig. 4. Substrate specificity of the Sf6 TSP endorhamnosidase. A, hydrolysis of 7-α-methylcoumarin (Amc)-labeled S. flexneri serotype Y O-antigen decasaccharide (■), decasaccharide (●), and octasaccharide (◆) each 2.2 μM with 0.45 μM Sf6 TSP at 40 °C. The amount of 7-α-methylcoumarin-labeled tetrasaccharide, as determined by high pressure liquid chromatography after different times of reaction, is given. In no case was 7-α-methylcoumarin-labeled octasaccharide observed as a product. B, schematic representation of the active site of Sf6 TSP as it follows from the hydrolysis experiments with different 7-α-methylcoumarin-labeled O-antigen fragments. Two repeating units of the polysaccharide chain are bound with comparably high affinity. The site of hydrolysis indicated by the arrow resides at the reducing end of these two repeating units.

The molecular mass at infinite dilution derived from sedimentation equilibrium runs (Fig. 1B, inset) amounted to 165.6 kDa. As the polypeptide molecular mass calculated from the amino acid sequence amounts to 55,278 Da, the ultracentrifugation result confirms the trimeric structure of Sf6 TSPΔN. Whereas full-length Sf6 TSP did not crystallize under any condition examined so far, Sf6 TSPΔN readily crystallized in a rapid vectorial screen for crystallization conditions (27).

According to the idea of modular evolution of bacteriophages (28, 29), all tail phages with double-stranded DNA genomes may be seen as one gene pool, exchanging functionally related gene groups by recombination events with each other and with their respective host bacteria. This theory is supported by considerable sequence data (30). Based on sequence comparison, it has also been postulated that single genes or even parts of genes, probably corresponding to single protein domains, were exchanged between different phages or acquired from host cells (31–33). Between tailspike proteins of class C bacteriophages similar to *Salmonella* phage P22, sequence similarities could only be detected in the N-terminal 100 amino acid residues, probably corresponding to the domain anchoring the tailspikes to the phage particle. There are four tailspike protein sequences published so far with sequence identities of about 70–80% in the N-terminal region. In addition to P22 and Sf6 TSP, they are open reading frame 36 of phage APSE-1 (34) and gene 9 of phage HK620 (49). Furthermore, the TSPs of *Salmonella* phages e34 and e341 have shown to be able to bind tail-less P22 heads (35, 36). Thus, these proteins probably are also homologs of P22 and Sf6 TSP, regarding their N-terminal domains. Although all tailspike polypeptides are of similar size, no sequence similarity has been detected between any of these proteins in their major C-terminal parts beyond residue 110.

**Fig. 5. Formation of SDS-resistant trimers by TSPΔN.** Protein expression from plasmids containing or lacking the insert that codes for amino acid residues 108–623 of Sf6 TSP (TSPΔN). Soluble cell extracts were boiled (100 °C +) or not boiled (100 °C −) after addition of SDS prior to electrophoresis. The arrows indicate the positions of TSPΔN in the 1st and 2nd lanes. The bands are not visible in the controls depicted in the 3rd and 4th lanes.

**number AF128887.** Throughout this publication, the numbering of amino acids is according to the corrected sequence starting with Met, as it is not known whether the initiating Met is cleaved off post-translationally in *E. coli*. The C-terminal part of Sf6 TSP expressed in *E. coli* was soluble, turned out to be SDS-resistant at room temperature (Fig. 5), and was purified to homogeneity (>98%, cf. above). The molecular mass at infinite dilution derived from sedimentation equilibrium runs (Fig. 1B, inset) amounted to 165.6 kDa. As the polypeptide molecular mass calculated from the amino acid sequence amounts to 55,278 Da, the ultracentrifugation result confirms the trimeric structure of Sf6 TSPΔN. Whereas full-length Sf6 TSP did not crystallize under any condition examined so far, Sf6 TSPΔN readily crystallized in a rapid vectorial screen for crystallization conditions (27).

**DISCUSSION**

According to the idea of modular evolution of bacteriophages (28, 29), all tail phages with double-stranded DNA genomes may be seen as one gene pool, exchanging functionally related gene groups by recombination events with each other and with their respective host bacteria. This theory is supported by considerable sequence data (30). Based on sequence comparison, it has also been postulated that single genes or even parts of genes, probably corresponding to single protein domains, were exchanged between different phages or acquired from host cells (31–33). Between tailspike proteins of class C bacteriophages similar to *Salmonella* phage P22, sequence similarities could only be detected in the N-terminal 100 amino acid residues, probably corresponding to the domain anchoring the tailspikes to the phage particle. There are four tailspike protein sequences published so far with sequence identities of about 70–80% in the N-terminal region. In addition to P22 and Sf6 TSP, they are open reading frame 36 of phage APSE-1 (34) and gene 9 of phage HK620 (49). Furthermore, the TSPs of *Salmonella* phages e34 and e341 have shown to be able to bind tail-less P22 heads (35, 36). Thus, these proteins probably are also homologs of P22 and Sf6 TSP, regarding their N-terminal domains. Although all tailspike polypeptides are of similar size, no sequence similarity has been detected between any of these proteins in their major C-terminal parts beyond residue 110.
The two parts of P22 TSP are independent folding units and have independent functions, comprising the binding to the phage head for the N-terminal domain and the binding and hydrolysis of the receptor on the bacterial surface for the C-terminal part, respectively (6, 13). The specificity of the TSP largely determines the host range, and P22 heads complemented with TSP from other phages could infect different host cells (35). Thus, the data available to date might suggest that the tailspikes of many class C phages share a common N-terminal head-binding domain combined with unrelated C-terminal host-recognizing domains.

As the three-dimensional structure of proteins is generally much more conserved than their amino acid or the corresponding nucleotide sequences, a structural characterization, like the one attempted in the present paper, may reveal evolutionary relatedness that remains undetected by mere sequence comparisons. Our biophysical data strongly suggest that the overall folds of Sf6 and P22 TSP are very similar. Both proteins are homotrimers, as shown by gel filtration and analytical ultracentrifugation. The C-terminal parts, apparently unrelated in sequence, resemble each other in their SDS resistance and in their stability against thermal denaturation, with only a 1.5-fold difference in the unfolding rate constants at the same temperature. The close similarity is quite surprising, as even a single point mutation in P22 TSP can decrease or increase the denaturation rate constants 10- and 5-fold, respectively (14). The secondary structure contents of Sf6 and P22 TSP, as calculated from FT-IR, are essentially identical (Tables I and II); in addition, far-UV CD and FT-IR spectra are comparable in shape, indicating a similar secondary structure of the proteins. There is a small shift to lower frequencies (about 2–4 cm⁻¹) in all parts of the amide I spectrum of P22 TSP relative to the spectrum of Sf6 TSP but also relative to previously determined spectra of β-helical proteins (37). Although such a shift could be indicative of slightly stronger hydrogen bonding, it might also be the result of incomplete hydrogen-deuterium exchange. Previous data of Khurana and Fink (37) indicate that β-helix proteins do not have a special signature in infrared absorbance. Interestingly, however, the similarity of the spectra of the two TSPs observed here is much closer than that of the spectra of different β-helix proteins (LpxA, PelC, and P22 TSP) measured in the previous study. As observed previously (37), the β-sheet content and the total amount of regular secondary structure are significantly overestimated by FT-IR when compared with the x-ray structure of P22 TSP. This may be explained by a high content of hydrogen-bonded turns and loops in P22 TSP. The total amount of secondary structure elements also varied with the fit procedures used. It is common to fit deconvoluted FT-IR spectra (18) which requires the consideration of more bands because of the higher band resolution. This may lead to higher α-helix contents. For P22 and Sf6 TSP the α-helix content increases by ~7–10% in favor of a decrease of turn structure, when compared with the values obtained from fits of non-deconvoluted spectra (Tables I and II). The β-sheet content decreases by ~15–20% in favor of a new band to be assigned to unordered or 3_α structure and now becomes more similar to the crystal structure value (Table II). However, independent of the fit model used, the secondary structure contents of P22 and Sf6 TSP are very similar. Taken together, the hydrodynamic and spectroscopic data prove that both tailspike proteins are highly thermostable homotrimers of similar shape and closely similar secondary structure.

Furthermore, our investigation of the enzymatic activities toward fluorescence-labeled enterobacterial lipopolysaccharide fragments strongly suggests that the active site topologies of both proteins are quite similar. Both tailspikes are endorhamnosidases, and just as observed with P22 TSP (11), an efficiently cleaved oligosaccharide substrate of Sf6 TSP must contain two full repeats of the O-antigen toward the non-reducing end from the cleavage site. The differential oligosaccharide specificity of both endoglycosidases readily explains why octasaccharide is the major accumulating product in the hydrolysis of lipopolysaccharide receptors by both phages. It has been shown also for a number of other phages recognizing and hydrolyzing O-antigen that the end products are not monomers but rather dimers or trimers of the repetitive O-antigen unit (38, 39), suggesting that the active site topology of phage endoglycosidases is conserved far beyond the two enterobacterial phages studied here. A “glycanase” motif has been detected in the polypeptides sequences of Sf6 TSP and other polysaccharide-degrading and -modifying enzymes but not in P22 TSP (4). Its position around residue 174 in Sf6 TSP, i.e., far from the sequence positions of active site residues in P22 TSP (10, 11), originally suggested a dissimilar architecture of the two tailspike proteins. This glycanase motif, but not an N-terminal domain homologous to P22 or Sf6 TSP, was also detected in the TSP/endosialidase of bacteriophage K1. In recently determined crystal structures of endoglycosidases, however, the polypeptide segments corresponding to the glycanase sequence motif form a strand-helix-strand structural motif capping the N-terminal end of the right-handed β-helical fold common to the enzymes (40, 41). Thus, this motif is not involved in the active site but rather is a structural feature of β-helices, further supporting a parallel β-helix architecture of Sf6 TSP. The recombinant production of the C-terminal part of Sf6 TSP resulted in a natively folded, homotrimeric, and SDS-resistant protein, thus resembling P22 TSPΔN. We conclude that the central host cell receptor-binding domains of Sf6 and P22 TSP are indeed homologous and not unrelated domains. The most parsimonious explanation for this finding is that both the proteins descend from one ancestor protein, which already had an N-terminal head-binding domain and a C-terminal adhesin domain. Different selective pressure might then have led to different conservation of sequence similarity in the two domains. The N-terminal domain has to interact with other proteins (head connector) probably with a large binding surface producing a large free energy of binding, because the interaction between TSPs and phage heads is basically irreversible (42). The C-terminal domain, however, is just constrained by protein stability and substrate specificity. Even mutations in the receptor-binding site could have been selected, if they increased or changed the host range. However, based on our data alone we cannot exclude that N- and C-terminal domains have different ancestors and different evolutionary ages and came together by reshuffling during phage evolution. This explanation finds some support by the finding that the lytic Salmonella phage SP6 encodes for a tail protein with 58% identity to the C-terminal domain of P22 TSP but totally missing the N-terminal domain (43) and by the finding of the glycanase motif in the endosialidase of bacteriophage K1, as mentioned above, which also has no N-terminal head-binding domain. Although we cannot exclude this possibility, our findings emphasize the importance of structural in addition to sequence information, when considering evolution mechanisms of proteins and protein domains.

Regarding the tailspike proteins of class C bacteriophages, the polypeptide sequences of their N-terminal head-binding domains are much more strongly conserved than the sequences of the C-terminal and central parts, although a common β-helical architecture of the latter is strongly suggested by our results. The right-handed parallel β-helix fold is not only of interest for phage evolution but generally as a polysaccharide
binding architecture that might find use in biotechnology. Exceptionally high sequence diversity despite close structural homology may be a characteristic feature of the right-handed parallel β-helix fold, in which loops and turns of variable length alternate with short β-strands, and a large fraction of the structurally conserved residues is solvent-exposed. No repeats are readily recognized in the polypeptide sequences of such proteins; the alignment of their sequences is difficult in the absence of a crystal structure, and most attempts to recognize the fold from amino acid sequences have failed (44, 45). In principle, however, repetitive structures should be more readily assignable to non-homologous sequences than globular proteins; the alignment of their sequences is difficult in the absence of a crystal structure, and most attempts to recognize the fold from amino acid sequences have failed (44, 45). In principle, however, repetitive structures should be more readily assignable to non-homologous sequences than globular structures (46), and the recently developed BETAWRAP prediction method does appear promising in that respect (47). Relying on β-strand interactions learned from non-helical β-structures and allowing for variability in the length of individual β-helical turns, the algorithm distinguishes β-helical from other structures in the protein structural data base. When subjected to BETAWRAP, the polypeptide sequence of Sf6 scores slightly higher than the sequence of P22 TSP.

As the N-terminal 110 residues are about 80% identical between P22 and Sf6 TSP, the three-dimensional structures of the two domains must be very similar. Their stability, however, appears to be significantly different. In thermal denaturation analyzed by SDS-gel electrophoresis, Sf6 TSP appears to unfold in a single step process with no obvious intermediate, whereas P22 TSP accumulates a thermal denaturation intermediate with unfolded N-terminal domains. Partial proteolysis experiments with trypsin and chymotrypsin have shown that the N-terminal domain of Sf6 TSP can be totally digested even at room temperature and in the absence of denaturants.2 The N-terminal domain of P22 TSP, in contrast, is very stable against proteases under the same conditions and is only digested after the thermal denaturation intermediate has been accumulated during a preincubation at high temperature (13). This indicates that the N-terminal domain of Sf6 TSP is denatured by SDS already at room temperature, i.e. under the conditions in SDS electrophoresis and that the SDS-resistant trimer band of Sf6 TSP is the equivalent of the intermediate band of P22 TSP.

Crystallization experiments with the complete P22 TSP have not yielded crystals of high enough quality for x-ray structure determination, possibly due to the flexible link between the N- and C-terminal parts of the protein or due to their differential stability, leading to structural heterogeneity (7). Similarly, crystals of Sf6 TSP were readily obtained, but only after deletion of the N-terminal domain. Future work will be aimed at the determination of a high resolution x-ray structure expected to shed light on the evolution of bacteriophages and parallel β-helix proteins.

Acknowledgments—We thank Stephanie Barbirz for fruitful discussions and Mandy Schietke for excellent technical assistance.

2 S. Barbirz and R. Seckler, unpublished results.