The Shigella flexneri bacteriophage Sf6 tailspike protein (TSP)/endorhamnosidase is related to the bacteriophage P22 TSP and has a motif common to exo- and endoglycanases, and C-5 epimerases

James E. H. Chua, Paul A. Manning and Renato Morona

The temperate bacteriophage Sf6 infects Shigella flexneri strains of serotype X or Y, converting them into serotypes 3a or 3b, respectively. The tailspike protein (TSP) of Sf6 possesses endo-1,3-α-D-rihamnosidase (endorhamnosidase) activity which results in cleavage of the lipopolysaccharide O-antigen receptor during the adsorption of the phage to the cell surface. When used in Southern hybridization, a P22 gene 9 (encoding P22 TSP) DNA probe hybridized with restriction fragment PstI-7 of Sf6. DNA sequencing and analysis of PstI-7 and the adjacent PstI-8 fragment revealed an open reading frame (ORF1) of 1872 bp (624 amino acids) bearing amino acid sequence homology to the bacteriophage P22 TSP N-terminal head-binding domain. High conservation of key residues was suggestive of similar secondary and tertiary N-terminal protein structure and a similar function of the Sf6 TSP in this region. In addition, an amino acid sequence motif (DFGX,DGX,AX,A) was identified between residues 164 and 184 which was also found to exist in various prokaryotic and eukaryotic exo-/endoglycanases, C-5 epimerases and bacteriophage proteins. Expression of ORF1 from a T7 promoter produced a 67 kDa protein (detected by 35S-methionine labelling and SDS-PAGE). Assay of heat-treated cytoplasmic extracts containing the ORF1-encoded protein by incubation with whole Sh. flexneri Y cells demonstrated that O-antigen hydrolysis activity was present; ORF1 therefore encodes Sf6 TSP. Sf6 TSP exhibited specific and preferential activity for long-chain Sh. flexneri serotype X or Y O-antigen, cleavage of which resulted in the release of oligosaccharide fragments, consistent with octasaccharides in size, as detected by fluorophore-assisted carbohydrate electrophoresis (FACE).

Keywords: tailspike protein, cloning, motif, O-antigen, endorhamnosidase activity

INTRODUCTION

Shigella flexneri is a Gram-negative, non-motile, rod-shaped, intracellular bacterial pathogen that invades the human colonic epithelium, causing bacillary dysentery. The presence of group- and type-specific O somatic antigens on the LPS component of the bacterial outer membrane, reflected in serological differences, form the basis for the distinction of related serotypes within the species (Simmons & Romanowska, 1987). The LPS of the Sh. flexneri outer membrane is an important component closely associated with the virulence of the organism (Van den Bosch et al., 1997). The basic O-antigen structure of Sh. flexneri is composed of repeated tetrasaccharide units consisting of N-acetylglucosamine linked to three rhamnose residues (Fig. 1). Strains possessing this O-antigen structure have the group 3,4 antigen and are assigned the Y serotype (Carlin &
**Fig. 1.** O-antigen repeat unit of **Sh. flexneri** Y serotype. Structure of the O-antigen tetrasaccharide repeat (between vertical lines) of **Sh. flexneri** Y serotype LPS showing the location of the 1,3-α-linkage cleaved by the bacteriophage Sf6 TSP/endorhamnosidase. Rha, rhamnose; GlcNAc, N-acetylglucosamine.

Lindberg, 1987). Bacteriophage Sf6 mediates the conversion of Y or X strains of **Sh. flexneri** into the 3b and 3a serotypes, respectively (Clark et al., 1991). A Sf6-borne gene (oac) encodes an O-antigen acetyltansferase which adds an acetyl group to O-2 of the third rhamnose residue on the repeated O-antigen tetrasaccharide (Clark et al., 1991).

**Sh. flexneri** bacteriophage Sf6 possesses an isometric head, which is 53 nm in diameter and encloses a double-stranded DNA genome slightly less than 40 kb in length (Gemski et al., 1975; Clark et al., 1978). The morphological characteristics of phage Sf6 are typical of the C phages, according to the classification of Bradley (1967), and are very similar to **Salmonella** bacteriophage P22. Bacteriophage Sf6 infects through making initial contact by adsorption to a primary receptor in the group 3,4 O-antigen before gaining access to its secondary receptor on the cell surface, at which the phage releases its genomic DNA into the **Sh. flexneri** host (Lindberg et al., 1978). Phage adsorption to the bacterial O-antigen is accompanied by an endo-1,3-α-L-rhamnosidase (endorhamnosidase) O-antigen cleavage activity facilitating phage access to the bacterial outer-membrane surface. This endorhamnosidase activity is also exhibited by other O-antigen converting phages, such as **Salmonella** phages P22 and 410, and by coliphage Ω8 (Kanegasaki & Wright, 1973; Takeda & Uetake, 1973; Wallenfels & Jann, 1974). The endorhamnosidase activity of bacteriophage P22 is associated with the tailsplkes, which are composed of a trimer of the P22 gene-9-encoded tailspike protein (TSP) (Israel et al., 1972; Israel, 1978; Baxa et al., 1996), and this is also likely to be the case for Sf6 TSP. The endorhamnosidase of Sf6 acts by cleaving a specific 1,3-α-linkage between the second and third rhamnose residues of the O-antigen tetrascarharide repeat unit of susceptible **Sh. flexneri** strains, releasing an octasaccharide cleavage product as shown by methylation and reducing end-group sugar analyses (Lindberg et al., 1978) (Fig. 1). Such endorhamnosidase activity may be utilized by the phage to clear a path through the bacterial LPS to gain access to the secondary receptor on the cell surface (Lindberg et al., 1978). Alternatively, as described for the P22 tailspike (Baxa et al., 1996), it may act to release progeny phage particles from cellular debris at the end of a lytic cycle.

In this study we report on the identification, cloning and sequencing of the gene encoding the bacteriophage Sf6 TSP/endorhamnosidase. Characterization of the Sf6 TSP showed similarity with a number ofexo- and endoglycosases, C-5 epimerases and proteins from bacteriophages and eukaryotic viruses. The TSP was overexpressed, partially purified and its endorhamnosidase activity was investigated.

**METHODS**

**Bacterial strains, culture conditions and culture media.** Strains used in this study included: *Escherichia coli* DH5α, E2096 (E. coli DH5(pGP1-2)) and E2096(pSf6orf1); **Sh. flexneri** PE577 (serotype Y) (Morona et al., 1994) and various other serotypes as detailed in the legend of Fig. 7; and **Salmonella typhimurium** LT2 (from K. Sanderson, Salmonella Generic Stock Center, University of Calgary, Canada). *E. coli* strain PY13579 harbouring pJS28, which encodes the P22 TSP (Sauer et al., 1982), was also obtained from K. Sanderson. All strains were grown at 37°C in Luria–Bertani (LB) broth (Morona et al., 1994), except derivatives of E2096. These were grown at 30°C since they harbour pGP1-2, which encodes a temperature-inducible T7 RNA polymerase. Strains were grown in 2× YT medium (Miller, 1972) when cytoplasmic extracts were to be prepared. Antibiotics were used at the following concentrations where appropriate (μg ml⁻¹): ampicillin (Ap), 100; chloramphenicol (Cm), 25; kanamycin (Km), 50; and rifampicin (Rif), 200. IPTG and X-Gal were used at 40 μg ml⁻¹ in LB plates.

**Bacteriophage, plasmids and DNA methods.** Sf6c phage (herein called Sf6) was recently described by Van den Bosch et al. (1997). Sf6 was propagated and DNA prepared as previously described (Clark et al., 1991; Van den Bosch et al., 1997). Plasmids used in this study include: pBS-SK (Ap⁹) and pBC-KS (Cm⁶) (Stratagene), pGEMT and pGEMZY (Promega), pGP1-2 (Tabor & Richardson, 1985), pSf6Pst7 (PstI-7 fragment of Sf6 genome cloned into pBS-SK), pJS28 (P22 gene 9 under lacUV5 promoter, Ap⁹; Sauer et al., 1982; Schwarz & Berger, 1989), pSf6Pst71 (0.7 kb PstI–HindIII subclone of PstI-7 inserted into pBC-KS, Cm⁶), pSf6Pst72 (1.7 kb PstI–EcoRV subclone of PstI-7 inserted into pBC-KS, Cm⁶), pSf6Pst73 (1.0 kb EcoRV subclone of PstI-7 inserted into pBS-SK, Ap⁹), pSf6Pst74 (2.2 kb HindIII–PstI subclone of PstI-7 inserted into pBC-KS, Cm⁶) and pSf6orf1 (Sf6 ORF1 in pGEMT, Ap⁹).
Plasmid DNA was prepared by the alkaline lysis technique, and restriction enzyme digestion, agarose gel electrophoresis, ligation, electroporation and transformation were performed as described by Morona et al. (1994, 1995).

**DNA sequencing and PCR.** DNA sequencing was carried out using dye-terminator, or dye-primer sequencing reactions (Applied Biosystems) using the M13, -21 forward and reverse primers and other primers as listed below. The complete DNA sequence of ORF1 was obtained with dye-primer sequencing of PstI-7 subclones, whilst internal primers were used in dye-terminator sequencing. PCR amplification was performed using standard protocols with Amplitaq DNA polymerase (Hoffman-La Roche) as described recently (Mavris et al., 1997). PCR products were cloned into pGEMT (Promega) according to the manufacturer's instructions. Oligodeoxynucleotides were obtained from Biotech International.

**Oligonucleotide primers.** Oligonucleotide primers used in PCR and dye-terminator sequencing reactions included: 2188 (5'-TAAATACAGATCATCAG-3'); 2189 (3'-TTCAACAGATCTCTTTG-5'); 2111 (140 bp downstream of ORF1, 5'-GTATAGACCGTCTGCTG-3'); 2232 (100 bp upstream of ORF1, 5'-TTGACGAGCGGAGGCTG-3'); 2235 (internal sequencing primer for ORF1, 5'-CTCGGAGCCAAAAGCC-3'); 2236 (internal sequencing primer for ORF1, 5'-TTCTTCTAGATACGGCGGC-3'); 2237 (internal sequencing primer for ORF1, 5'-TATTGATAACTGGCCAGCC-3').

**DNA and protein sequence analysis.** Raw sequencing data from the 373A automated sequencer were analysed using the Applied Biosystems SeqEd program version 6.0. Sequencing data were analysed using DNAsis and FROGIS (Hitachi Software Engineering). DNA sequences and translated amino acid sequences obtained were searched against the NCBI sequence database through on-line BLAST and BLAST2 software (Altschul et al., 1997). CLUSTAL W was also used to perform amino acid sequence alignments (Thompson et al., 1994).

**Southern blotting methods.** Southern blotting procedures were performed as recently described (Morona et al., 1995). Plasmid pJS28 provided the bacteriophage P22 gene 9 DNA template for a PCR reaction using primers 2188 and 2189 to synthesize a digoxigenin (DIG) -labelled ddUTP probe. DIG-ddUTP was incorporated into PCR products in accordance with the protocol supplied by Boehringer Mannheim. The labelled PCR product (50 µl) was then used as a probe in Southern hybridizations as recently described (Mavris et al., 1997). For the hybridization of bacteriophage P22 gene 9 DNA to PstI-digested Sf6 genome, stringency was lowered by the use of 0.1x SSC washes at 50 °C. Hybridization and detection were performed with DIG labelling and detection kits (Boehringer Mannheim), according to the manufacturer's instructions.

**PCR amplification and cloning of ORF1.** Primers 2111 and 2232 were used in PCR with Sf6 DNA as the template to amplify a 21 kb region, spanning the PstI-7 and PstI-8 junction, which contained the entire ORF1 (Fig. 2). The primers were placed to include approximately 100 bp of flanking sequence upstream of the start codon as well as approximately 140 bp of sequence downstream of the termination codon. The PCR product was ligated to pGEMT and transformed into DH5α. A clone (pSf6orf1) harbouring the construct, inserted such that expression was driven by the T7 promoter, was selected by restriction enzyme analysis and confirmed by DNA sequencing.

**l-[35S]Met labelling and SDS-PAGE.** Induction of T7 RNA polymerase and radiolabelling of proteins was performed according to Tabor & Richardson (1985) as detailed recently (Morona et al., 1994), except that E2096 (pSf6orf1) and E2096(pGEM5Zf) cells were washed and resuspended in 1 ml M9 salts (Miller, 1972) with 0.01% (w/v) methionine assay medium (Difco 0423-15-2) prior to incubation with Rif, and labelled with l-[35S]methionine (SIGMA, Amersham) for 5 min at 30 °C. Cells were then pelleted in a microfuge, resuspended in sample buffer (Laemmli et al., 1977), heated at 100 °C for 3 min and subjected to SDS-PAGE on a linear 15% polyacrylamide gel (Laemmli et al., 1977). The gel was stained in 0.06% (w/v) Coomassie brilliant blue G250 [dissolved in 5% (v/v) perchloric acid], destained with 5% (v/v) acetic acid, dried and labelled proteins were detected by autoradiography (Morona et al., 1995). Molecular size markers (Pharmacia) were: phosphorylase B (94 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), soybean trypsin inhibitor (20.1 kDa) and a-lactalbumin (14.4 kDa).

**Preparation of cytoplasmic extracts.** E2096(pSf6orf1) was grown at 30 °C for 16 h in LB with Ap and Km, then

---

**Sf6 tailspike gene and protein**

---

**Fig. 2.** (a) PstI restriction map of bacteriophage Sf6 genome. PstI fragments are numbered in order of decreasing size as determined by agarose gel electrophoresis. An additional 100 bp PstI fragment, designated PstI-10, located between PstI-7 and PstI-2, was identified during the course of this study (data not shown). The Sf6 integrase (int) and O-acetylase (oac) genes are located in the previously sequenced PstI-8 fragment (Clark et al., 1991). (b) Expanded view of the region to show the location of ORF1 relative to PstI-8 and PstI-7 fragments. The orientation of ORF1 (orf1) is indicated. Also shown is the location of the insert in pSf6orf1 which was amplified by PCR and cloned into pGEMT. Relevant restriction enzyme sites are: H, HindIII; P, PstI; S, SmaI; V, EcoRV.
subcultured and grown to an OD₆₀₀ of 0.5. Induction of protein expression was achieved by shifting the culture to 42 °C. After 2 h, a cytoplasmic extract (CE) was prepared based on a procedure provided by K. Sanderson (University of Calgary, Canada). Cells were pelleted, resuspended in 100 mM Tris/HCl pH 7.5, 10 mM EDTA [2 ml (g cell pellet)] and lysed by the addition of lysozyme [2 mg (g cell pellet)] at 37 °C for 1 h. The mixture was incubated at 65 °C for 2 h to complete lysis and denature heat-labile proteins. The crude lysate was subsequently centrifuged at 17000 r.p.m. in a JA20 rotor (Beckman J2-MI) for 70 min. The amber supernatant was then decanted, incubated at 4°C overnight and centrifuged again at 17000 r.p.m. for a further 20 min to remove precipitants. The supernatant (crude cytoplasmic extract) was either used directly, or concentrated and reconstituted in 0.1 ml TM buffer, was used instead of CE as a positive control. Extracts were also prepared the same method.

Ultrafree-15 centrifugal filter device (Millipore) according to the manufacturer's instructions. Extracts were also prepared the same method.

Preparation of formalin-fixed cells. Bacterial strains grown for 18 h in 10 ml LB at 37 °C were collected by centrifugation (5000 r.p.m., 10 min; Centra 4X, IEC) and washed twice with saline. Cells were then treated with 1% formalin solution in saline at 37 °C with agitation for 1 h. The formalin-fixed cells was then decanted, incubated at 4°C overnight and centrifuged again at 17000 r.p.m. for a further 20 min to remove precipitants. The supernatant (crude cytoplasmic extract) was either used directly, or concentrated and reconstituted in 10 mM Tris/HCl pH 7.5, 10 mM MgCl₂ (TM buffer) using an Ultrafree-15 centrifugal filter device (Millipore) according to the manufacturer's instructions. Extracts were also prepared from PY13579, E2096(pGEMS5ZfI) and DH5α(pBS-SK) using the same method.

Endorhamnosidase activity assay. An aliquot of 0.1 ml formalin-fixed cells was mixed with equal volumes of CE and TM buffer and the suspension was incubated at 37 °C for 30 min. Alternatively, 10¹¹ f.u. of Sf6 phage, resuspended in 0.1 ml TM buffer, was used instead of CE as a positive control. Fixed cells treated with CE of plasmid-vector-containing strains were used as negative controls. After the incubation period, the cells were pelleted as above and washed twice in saline. Cells were then treated with 1% formalin solution in saline at 37 °C for 1 h. The formalin-fixed cells treated samples. After the incubation period, the cells were pelleted as above and washed twice in saline. Cells were then treated with 1% formalin solution in saline at 37 °C for 1 h. The formalin-fixed cells was then decanted, incubated at 4°C overnight and centrifuged again at 17000 r.p.m. for a further 20 min to remove precipitants. The supernatant (crude cytoplasmic extract) was either used directly, or concentrated and reconstituted in 10 mM Tris/HCl pH 7.5, 10 mM MgCl₂ (TM buffer) using an Ultrafree-15 centrifugal filter device (Millipore) according to the manufacturer's instructions. Extracts were also prepared from PY13579, E2096(pGEMS5ZfI) and DH5α(pBS-SK) using the same method.

Endorhamnosidase activity assay. An aliquot of 0.1 ml formalin-fixed cells was mixed with equal volumes of CE and TM buffer and the suspension was incubated at 37 °C for 30 min. Alternatively, 10¹¹ f.u. of Sf6 phage, resuspended in 0.1 ml TM buffer, was used instead of CE as a positive control. Fixed cells treated with CE of plasmid-vector-containing strains were used as negative controls. After the incubation period, the cells were pelleted as above and washed twice in MilliQ purified water (Millipore). Small-scale LPS preparations were made by proteinase K treatment of whole-cell lysates (Hitchcock & Brown, 1983). After electrophoresis on SDS/20% polyacrylamide gels, LPS was detected by silver staining as described previously (Morona et al., 1991; Van den Bosch et al., 1997). Endorhamnosidase activity was shown by the cleavage of O-antigen as evidenced by a shift in intensity of bands representing predominantly long-chain LPS molecules in the untreated samples, to bands representing predominantly short-chain LPS molecules in endorhamnosidase-treated samples.

**Fluorophore-assisted carbohydrate electrophoresis (FACE).** The procedure followed was that described in detail by Jackson (1991). In this technique, the reducing end sugar of an oligosaccharide is reacted with a highly charged, fluorescent molecule, 8-aminonaphthalene trisulfonic acid (ANTS). The derivatized oligosaccharides are then separated by electrophoresis on a polyacrylamide gel and detected by fluorescence. In brief, a 100 µl aliquot of supernatant from a standard endorhamnosidase assay (described above) was deproteinized by ethanol precipitation, dried in vacuo in a Savant SpeediVac, resuspended in reaction buffer and incubated with ANTS and sodium cyanoborohydride for 16 h at 37 °C, exactly as described by Jackson (1991). The sample was dried in vacuo, resuspended in 100 µl sample buffer (Jackson, 1991) and 10 µl was electrophoresed on a 0.75 mm thick 30% polyacrylamide gel using a Bio-Rad Mini Protein apparatus. The gel was placed on a 312 nm UV transilluminator to detect fluorescent bands, which were photographed using Polaroid 667 film.

**RESULTS**

**DNA sequencing of Sf6 PstI-7**

The genome of *Sh. flexneri* bacteriophage Sf6 has been shown to be digested into nine fragments using the restriction enzyme PstI (Clark et al., 1991). These fragments were designated PstI-1 to PstI-9, in descending size order. During this study a previously unidentified 100 bp fragment (PstI-10), located between PstI-7 and PstI-2, was detected and is incorporated into the PstI restriction map of the Sf6 genome (Fig. 2). Preliminary DNA sequencing of Sf6 has identified homologues to 14 bacteriophage P22 genes (1, 7, 10, 13, 14, 15, 16, 19, 20, 26, N, sieB, xis/int) and the organization of the genes is very similar to that in the P22 genome (C. Kuhlewein,
**Sf6 tailspike gene and protein**

| SF6       | 151 -RTQHDKNKEAISILDGVIDEQTYQAIQNAIDAVASLPSGELFIPASNQAVGYIVG- - - - - - - - - - - - - - - - - - - - 210 |
| K1        | 22 -RNLDKLETVVSLKDFGAKGDDGTNDQDAVRAIGSGKRIGATGTYKVSSLPFMDFY- - - - - - - - - - - - - - - - - - - - - 81 |
| YEFG      | 147 -TAKTSAPQIOVNRDFGAGDDKLTNKAIAQGIDSCP--GCRVEPAGTYSGWALSLKSD- - - - - - - - - - - - - - - - - - - - - 206 |
| C51       | 1 -MDYNVKGDDGVSDTATQAIDAAHAA---GGTVYLPAGEYRVSSGEEPSDGCLTI- - - - - - - - - - - - - - - - - - - - - - - - - 60 |
| C52       | 836 DSQNVIFEPAVFNAKDFGALGDSDDRPAIQAIDAAA---GGTVYLPAGEYRVSP- - - - - - - - - - - - - - - - - - - - - - - - - 895 |
| ATA       | 58 -VRASLRTPTTVSVSFAGDKGTDDQTVNAFKACSSNGAVNLVTVPKNTYLLKSI- - - - - - - - - - - - - - - - - - - - - - - - - 117 |

---

**Fig. 4.** Identification of a motif (DFGX,DGX,AX,A) in Sf6 ORF1 which is present in enzymes active on a variety of polysaccharides. The Sf6 ORF1 region between amino acid positions 150 and 211 is shown aligned with the indicated protein. The amino acids forming the motif are in bold type and indicated with asterisks. The sequences are: K1, bacteriophage K1 endonerasin (PIR accession number: S52299); YEFG, *Yersinia enterocolitica* exopolysaccharase (GenBank: AF059505); C51 and C52, *Azotobacter vinelandii* mannuronan C-5 epimerase (GenBank: L39096); ATA, *Arabidopsis thaliana* putative polygalacturonase (GenBank: AC002339). Additionally, the proteins described in Table 1, which have similarity with this region of Sf6 ORF1, also have sequences which are a subset of the motif shown here.

**Table 1.** Proteins exhibiting homology to Sf6 ORF1

<table>
<thead>
<tr>
<th>Region in Sf6 ORF1</th>
<th>Region in homologous protein</th>
<th>Identity (%)</th>
<th>Name</th>
<th>Accession number*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-137</td>
<td>1-137</td>
<td>64</td>
<td>Bacteriophage P22 TSP</td>
<td>SP: P12528</td>
</tr>
<tr>
<td>7-328</td>
<td>49-365</td>
<td>33</td>
<td><em>Erwinia amylovora</em> AmsF</td>
<td>SP: Q46636</td>
</tr>
<tr>
<td>30-73</td>
<td>109-154</td>
<td>30</td>
<td><em>Yersinia enterocolitica</em> exopolysaccharase</td>
<td>G: AF059505</td>
</tr>
<tr>
<td>35-226</td>
<td>283-473</td>
<td>27</td>
<td><em>Ampelomyces quisqualis</em> exo-1,3-β-glucanase</td>
<td>G: AF029354</td>
</tr>
<tr>
<td>135-197</td>
<td>6-68</td>
<td>28</td>
<td>Bacteriophage K1 endonerasin</td>
<td>PIR: S52299</td>
</tr>
<tr>
<td>159-186</td>
<td>155-182</td>
<td>53</td>
<td><em>Yersinia enterocolitica</em> exopolysaccharase</td>
<td>G: AF059505</td>
</tr>
<tr>
<td>162-289</td>
<td>4-152</td>
<td>43</td>
<td><em>Azotobacter vinelandii</em> mannuronan C-5 epimerase</td>
<td>G: L39096</td>
</tr>
<tr>
<td>165-193</td>
<td>851-879</td>
<td>43</td>
<td><em>Azotobacter vinelandii</em> mannuronan C-5 epimerase</td>
<td>G: L39096</td>
</tr>
<tr>
<td>162-202</td>
<td>68-102</td>
<td>38</td>
<td><em>Arabidopsis thaliana</em> putative polygalacturonase</td>
<td>G: AC002339</td>
</tr>
<tr>
<td>148-190</td>
<td>185-227</td>
<td>34</td>
<td><em>Escherichia coli</em> eliminase</td>
<td>G: X96495</td>
</tr>
<tr>
<td>159-214</td>
<td>150-214</td>
<td>34</td>
<td><em>Erwinia chrysanthemi</em> exo-α-polygalacturonosidase</td>
<td>G: M31308</td>
</tr>
<tr>
<td>149-226</td>
<td>93-160</td>
<td>24</td>
<td>Bacteriophage phi-29 pre-neck appendage protein</td>
<td>G: M14782</td>
</tr>
<tr>
<td>155-225</td>
<td>97-158</td>
<td>29</td>
<td>Bacteriophage 103 pre-neck appendage protein</td>
<td>G: X99260</td>
</tr>
<tr>
<td>257-321</td>
<td>673-735</td>
<td>32</td>
<td>Bacteriophage 103 pre-neck appendage protein</td>
<td>G: X99260</td>
</tr>
<tr>
<td>347-449</td>
<td>739-850</td>
<td>21</td>
<td>Bacteriophage 103 pre-neck appendage protein</td>
<td>G: X99260</td>
</tr>
<tr>
<td>547-624</td>
<td>870-955</td>
<td>29</td>
<td><em>Candida albicans</em> agglutinin-like protein</td>
<td>G: L25902</td>
</tr>
<tr>
<td>535-612</td>
<td>187-263</td>
<td>33</td>
<td><em>Semliki Forest virus</em> structural polyprotein</td>
<td>G: X78109</td>
</tr>
<tr>
<td>535-618</td>
<td>180-267</td>
<td>32</td>
<td>O'nyong-nyong virus structural polyprotein</td>
<td>G: M20303</td>
</tr>
</tbody>
</table>

*SP, SWISS-PROT database; G, GenBank database; PIR, PIR database.

J. E. H. Chua, R. Manadhar, P. A. Manning and R. Morona, unpublished results). These data and the morphological similarity between the two phages suggested to us that the Sf6 gene encoding the TSP may have some homology to gene 9, which encodes the TSP of P22.
J.E.H. CHUA, P.A. MANNING and R. MORONA

Fig. 5. Identification of the Sf6 ORF1 product. E. coli K-12 E2096 harbouring pSF6orf1 (lane 1) or pGEM5Zf (lane 2) were grown and induced, and expressed proteins labelled with L-[35S]methionine, as described in Methods. Samples were electrophoresed in an SDS-polyacrylamide gel and processed for autoradiography. The position of the molecular mass standards is indicated.

A DIG-labelled probe derived from the bacteriophage P22 TSP gene (gene 9) was used in Southern hybridization with PstI-digested Sf6 genomic DNA. The labelled probe hybridized to a single 2.9 kb fragment of Sf6 which corresponds to Pstl-7 (data not shown). Restriction enzyme analyses of Pstl-7 cloned in pBS-SK (pSF6Pst7) using PstI, HindIII and EcoRV led to the construction of a restriction map of PstI-7 (Fig. 2b). Based on the restriction map of Pstl-7 (Fig. 2b), subclones of Pstl-7 (pSF6Pst71, 72, 73, 74) were constructed for DNA sequencing. Pstl-8, which lies adjacent to Pstl-7, had been previously sequenced (Clark et al., 1991). Analysis of the DNA sequences obtained revealed an ORF of 1872 bp (ORF1) extending from Pstl-7 into Pstl-8 (Fig. 2b). ORF1 encodes a protein (624 amino acids) with a predicted molecular mass of 66.2 kDa. A putative rho-independent transcription terminator separated ORF1 from the convergently transcribed O-acetylase gene (oac) (Clark et al., 1991). The DNA sequence preceding and near the 5' end of ORF1 has significant homology (69% identity over 340 bp, beginning from the 5' end of the gene) to the bacteriophage P22 gene 9 (encoding P22 TSP) sequence.

Analyses of the ORF1 protein amino acid sequence

The predicted amino acid sequence encoded by ORF1 shows regions of marked homology to various proteins. The region bearing significant homology (64% identity, 74% similarity) to P22 TSP spanned the first 137 amino acids at the N termini of ORF1 and P22 TSP (Fig. 3). This corresponded to the head-binding domain of the P22 TSP (Steinbacher et al., 1994, 1996, 1997). Other regions of ORF1 exhibited similarity to other phage TSPs, phage structural proteins and non-phage proteins. These similarities are discussed in detail below. Of particular interest was the detection of a motif sequence (DFGX,DGX,A,X,A) located between residues 164 and 184 which was present in various prokaryotic and eukaryotic exo-/endoglycanases, such as the exopolygalacturonase of Yersinia enterocolitica, and in C-5 epimerases of Azotobacter vinelandii (Fig. 4). Table 1 summarizes the proteins found to have amino acid sequence similarity with the ORF1 protein.

Detection and characterization of the ORF1 protein

L-[-35S]Methionine was used to label the primary translation product of E2096(pSF6orf1). A single band unique to E2096(pSF6orf1) at 67 kDa was detected (Fig. 5, lane 1). The size of this protein corresponds closely to the predicted size of the translational product of ORF1. No
Fig. 7. Specificity of Sf6 ORF1 protein/TSP. Formalin-fixed cells were prepared from the indicated strains and incubated either with CE from the control strain E2096(pGEM5Zf) (even-numbered lanes) or with Sf6 TSP containing CE from E2096(pSf6orfl) (odd-numbered lanes). LPS was then prepared from the cells and analysed by SDS-PAGE and silver-staining. Lanes: 1 and 2, E576 (serotype X); 3 and 4, PE645 (serotype 3b); 5 and 6, PE843 (serotype 3a); 7 and 8, PE839 (serotype 1a); 9 and 10, PE840 (serotype 1b); 11 and 12, PE566 (serotype 4a); 13 and 14, PE572 (serotype 4b); 15 and 16, serotype PE655 (serotype 2a); 17 and 18, PE842 (serotype 2b); 19 and 20, PE647 (serotype 5a); 21 and 22, PE790 (serotype 2a).

We then investigated whether the ORF1 protein had endorhamnosidase activity. A CE was prepared from E2096(pSf6orfl), as described in Methods. Analysis of the CE by SDS-PAGE and Coomassie blue staining revealed the presence of a protein band corresponding in size to that detected by [35S]methionine labelling and which was absent in the E2096(pGEM5Zf) CE (data not shown). Endorhamnosidase activity assays were performed using CE from E2096(pSf6orfl), with appropriate controls. The E2096(pSf6orfl) CE exhibited endorhamnosidase activity as evidenced by the drastic shortening of Sh. flexneri serotype Y (PE577) LPS O-antigen chains (Fig. 6, lane 3) when compared with the LPS of negative-control CE-treated cells (Fig. 6, lane 1) and with untreated PE577 LPS O-antigen (Fig. 6, lane 8). This endorhamnosidase activity was similar to that produced by incubation with whole bacteriophage Sf6, used as a positive control (Fig. 6, lane 2). In addition, PE577 cells treated with CE containing bacteriophage P22 TSP had an unaltered LPS profile (Fig. 6, lane 4). Likewise, E2096(pSf6orfl) CE did not affect the LPS O-antigen of Sa. typhimurium LT2 (Fig. 6, lane 6); the latter was susceptible to P22 TSP/endorhamnosidase cleavage (Fig. 6, lane 5) and resulted in the production of LPS with shortened O-antigen chains as compared with the sample treated with a control CE (Fig. 6, lane 7). These data show that ORF1 encodes the Sf6 TSP/endorhamnosidase. Additionally, we observed that the addition of concentrated CE from E2096(pSf6orfl) to a suspension of formalin-fixed cells of PE577 resulted in the immediate agglutination of the cells. This observation provided a rapid and simple assay for detecting Sf6 TSP in CEs.

Serotype-specific activity of Sf6 endorhamnosidase

The E2096(pSf6orfl) CE containing the Sf6 TSP/endorhamnosidase was assayed against a panel of Sh. flexneri strains of different serotypes (Fig. 7). Endorhamnosidase-treated PE576 cells (serotype X) showed significant O-antigen cleavage (Fig. 7, lane 1), as expected from the host range of Sf6. Strains PE645 and PE843 (serotypes 3b and 3a, respectively) possess O-antigen modifications arising from the lysogenization of Sh. flexneri Y and X strains, respectively, by bacteriophage Sf6. No O-antigen cleavage was observed when either strain was treated with Sf6 TSP/endorhamnosidase (Fig. 7, lanes 3 and 5). O-antigen cleavage was not observed in any of the other serotypes of Sh. flexneri tested, except PE647 (serotype 5a). In the latter case, only a low level of O-antigen cleavage was observed (Fig. 7, lane 19). Hence, this and the above data demonstrate that ORF1 encodes the Sf6 TSP/endorhamnosidase with a specificity which agrees with the known host range of Sf6 (Clark et al., 1991).

Detection of the oligosaccharide released by Sf6 TSP activity on Sh. flexneri O-antigen

As the endorhamnosidase activity of Sf6 phage on Sh. flexneri Y O-antigen results in release of octasaccharide fragments (Fig. 1), we investigated the ability of Sf6 TSP to release oligosaccharides from O-antigen by using FACE (Jackson, 1991).

Cells of Sh. flexneri Y (PE577) were incubated with buffer, control CE, Sf6 TSP containing CE, or Sf6 phage. The supernatant was recovered and oligosaccharides present detected as described in Methods. The supernatant from control cells showed no spontaneous release of oligosaccharides (Fig. 8, lane 5). The cells treated with either Sf6 phage or cytoplasmic extracts containing Sf6 TSP released a small oligosaccharide whose size was close to that of maltoheptose (Fig. 8, lane 4). This observation provided a rapid and simple assay for detecting Sf6 TSP in CEs.
the indicated type were treated with Sf6 phage or CE from treatment of whole cells by FACE. Formalin-fixed cells (lo9) of ANTS. A sample (corresponding supernatant was recovered and incubated with derivatized with control CE from E2096(pGEM5ZF); 2, PE577 incubated with Sf6 phage; 3, PE577 incubated with CE containing Sf6 TSP from E2096(pSf6orf1); 4, PE655 (serotype 2a) incubated with SF6 TSP containing CE from E2096(pSf6orf1); 5, PE577 untreated. Lanes 6 and 7 contain maltosehexose (6 glucose repeat) and maltolheptose (7 glucose repeat), respectively, which have been derivatized with ANTS and served as size markers. Free ANTS is indicated with an arrow.

lane 1). Hence, we have shown that the major product of Sf6 TSP/endorhamnosidase activity on Sh. flexneri Y LPS O-antigen is an oligosaccharide which is likely to be the expected octasaccharide.

**DISCUSSION**

Sh. flexneri bacteriophage Sf6 is related to the well-characterized Salmonella bacteriophage P22 in terms of morphology, genome organization and DNA sequence. Of particular relevance to this study, bacteriophage P22 possesses an endorhamnosidase activity associated with its TSP which plays a similar role to that of the Sf6 TSP in the infection process (Israel et al., 1972; Israel, 1978; Steinbacher et al., 1997). The Sf6 ORF1 DNA sequence was homologous to the P22 gene 9 sequence, explaining the detection of PstI-7 by Southern hybridization with the P22 gene 9 probe. The region of nucleotide sequence homology to P22 TSP involved not just the coding region, but also a 16 bp sequence immediately prior to the translational start, incorporating the proposed ribosome-binding site. In addition to its similarity with P22 TSP (Fig. 3), the Sf6 TSP also had similarity with sequences near the N-terminal region of bacteriophage K1 TSP/endoasialidase (Table 1) (Long et al., 1995). Like P22 TSP, the K1 TSP possesses enzymic activity and has been shown to form active trimers as part of the assembly process to produce mature phage particles (Petter & Vimr, 1993; Schwarz & Berget, 1989; Steinbacher et al., 1997). The N-terminal amino acid sequences of phage TSPs can be highly conserved between related bacteriophages (Petter & Vimr, 1993), an observation which is consistent with our observation that the P22 TSP gene could be used to detect the Sf6 TSP gene. Hence, Sf6 TSP is a member of a family of bacteriophage endoglycanases.

The region of homology (64% identity) between Sf6 TSP and P22 TSP (residues 1–137) is a domain previously shown to be involved in the attachment of the tailspike to the phage capsid (Chen & King, 1991; Steinbacher et al., 1994, 1996, 1997). The three-dimensional structure of the phage head-binding domain of P22 TSP has recently been solved to a resolution of 2.3 Å (0.23 nm), enabling the elucidation of detailed secondary structural features and the identification of the interactions between key residues which contribute to the tertiary structure of the protein (Steinbacher et al., 1997). The head-binding domain of P22 TSP consists of two β-sheets: A (5 strands termed A1–A5) and B (3 strands, B1–B3) (Fig. 3). The residues that make up these strands are very highly conserved between the Sf6 and P22 TSPs (82% similarity, Fig. 3), suggesting that the N-terminal structure of Sf6 TSP is almost identical to that of P22 TSP and that it acts as the head-binding domain. Furthermore, key residues involved in intra-/inter-subunit interactions were also highly conserved. Such conserved residues include those involved in inter-subunit hydrogen-bond formation (residues 7–9 of one subunit with 81–83 of another; residue Arg20 of one subunit with Gly76 of another), intra-subunit hydrophobic contact (residues Lys80–Tyr108) and a cluster of three phenylalanine residues (Phe22 of each subunit) which cap the roof of the dome formed by trimerrization (Steinbacher et al., 1997). The conserved N-terminal residues 7–9 form an interdigitating strand, hydrogen-bonding with residues 81–83 of a neighbouring subunit in an interaction typical of virus structures (Harrison et al., 1996). In addition, the N-terminal region of AmsF of Erwinia amylovora (Bugert & Geider, 1995) was also found to be homologous (33% identical) to the N-terminal domain of Sf6 TSP (Table 1). AmsF was thought to be a periplasmic protein involved in either the polymerization or processing of repeat units of amylovoran, a polysaccharide that confers virulence to the organism. Our observation suggests that AmsF may have originated in part from bacteriophage sources.

A remarkable finding was the detection of a large variety of bacterial and eukaryotic enzymes and proteins bearing amino acid sequence similarity to the region between residues 151 and 210 of Sf6 TSP (Table 1). In particular, all homologous sequences contained a motif sequence (DFGX,XDGX,AAX,A) which can be aligned with the Sf6 TSP sequence spanning residues 165–183 (Fig. 4, Table 1). Many homologous sequences originate from enzymes involved in degradation of polysaccharide chains (endo- or exoglycanases), similar to the endorhamnosidase activity of Sf6 TSP. This motif will be referred to as the ‘glycanase motif’.

Bacteriophage proteins containing the glycanase motif include the bacteriophage K1 TSP (Fig. 4, Table 1) and the pre-neck appendage protein of phages 103 and phi..
The motif is not in the P22 TSP sequence. The position of the motif differs between Sf6 and bacteriophage K1 TSP. The glycanase motif, which is situated in the region of Sf6 TSP immediately following the putative head-binding N-terminal domain, aligns with the beginning of the N-terminus in bacteriophage K1 TSP. Whilst possessing the glycanase motif, the K1 TSP lacks sequences corresponding to the phage head-binding domains of P22 and Sf6. Hence, it is possible that a separate head-binding, adaptor protein is utilized by bacteriophage K1.

Other enzymes which contain the glycanase motif, summarized in Table 1, are produced by bacteria (Erwinia chrysanthemi, A. vinelandii, E. coli, Y. enterocolitica), mycoparasites (Ampelomyces quisqualis), plants (Arabidopsis thaliana) and various bacteriophages. Whilst many of these enzymes possess polysaccharide-chain-degrading glycanase activity, some non-glycanase enzymes also bear this motif. A notable example is the A. vinelandii AlgE (C-5 epimerase), which possesses a modular structure and is involved in the control of extracellular polysaccharide capsule alginate structure modifications (Ertesvåg et al., 1995).

A. vinelandii produces at least five separate C-5 epimerases (termed AlgE1–AlgE5), each being composed of one or two A modules (thought to be involved in substrate binding and catalysis) and single or multiple R modules (involved in Ca\(^{2+}\) binding). One copy of the glycanase motif can be found in each A module (Fig. 4, Table 1). This raises the possibility that the glycanase motif may be situated near, or within, a region of Sf6 TSP that is involved in substrate binding or catalysis, as is suggested in AlgE. The possibility of a role in substrate binding is further enhanced by the presence of this motif in such a wide range of enzymes known to catalyse polysaccharide-chain hydrolysis (e.g. polygalacturonidases of various organisms). An important observation is that such glycanase motif-bearing galacturonidases (active on polygalacturonans) can bind to alginate, the substrate of C-5 epimerase (Gupta et al., 1993), illustrating that substrate binding does not necessarily accompany enzyme activity.

Based on the DNA sequence of ORF1, plasmid pSf6orf1 was constructed and then transformed into E2096 to facilitate T7 RNA-polymerase-dependent gene expression. A protein corresponding in size with the ORF1 translation product was detected by labelling with 1\(^{-}[\text{S}]\)methionine (Fig. 5) and Coomassie blue staining (data not shown). The demonstration of endo-

rhamnosidase activity by visualizing silver-stained LPS profiles from PE577 cells treated with CE prepared from induced E2096(pSf6orf1) (Fig. 6) provided direct evidence that ORF1 encoded the Sf6 TSP/endo-
rhamnosidase. In common with bacteriophage P22 TSP (Sauer et al., 1982), Sf6 TSP was heat-stable as preparation of CE containing the protein involved heat treatment at 65 °C for 2 h. Additionally, we observed that CE containing SF6 TSP were able to agglutinate PE577 cells. A hypothesis to explain this observation is that, like P22 TSP (Steinbacher et al., 1994), Sf6 TSP is able to trimerize and at high concentrations is able to cross-link bacterial cells.

O-antigen cleavage catalysed by either the Sf6 TSP in cytoplasmic extracts or whole Sf6 phage gave rise to an LPS profile on SDS-polyacrylamide gels consisting predominantly of LPS molecules with O-antigen chains of approximately five tetrasaccharide repeat units (Fig. 6, lanes 2 and 3). This was also true for P22 TSP cleavage of Sa. typhimurium LT2 LPS (Fig. 6, lane 5). These findings agree with the kinetic studies of Baxa et al. (1996), who calculated that the P22 TSP could only cleave one glycosidic bond per 30 s at 14 °C. We propose that long O-antigen polysaccharide chains are the preferred substrate for the Sf6 TSP/endorhamnosidase and cleavage rates are several orders of magnitude faster for this substrate. That this was not observed by Baxa et al. (1996) could be because they used octa- and dodeca-
saccharides as substrates and not the long-chain molecules that exist on Salmonella bacterial cell surfaces. Our observations may have implications for the role of TSP in phage infection and for the possibility of the existence of alternative substrate-binding and catalysis sites in addition to, or instead of, those documented by Steinbacher et al. (1996, 1997). The greater rate of hydrolysis against the preferred long-chain O-antigen substrate would support a drilling function of TSP, as originally suggested by Lindberg et al. (1978), but subsequently rejected by Baxa et al. (1996). The three-dimensional structure of a P22 TSP–substrate (octa-/dodecasaccharide) complex demonstrated the association of substrate binding and enzyme catalysis within the C-domain region between residues 300 and 400 (Steinbacher et al., 1997). However, this does not exclude either additional and/or alternative sites being involved when the substrate is long-chain O-antigen polysaccharides. One such site in Sf6 TSP may well be associated with the conserved 'glycanase motif' (Fig. 4).

We used FACE to detect a small oligosaccharide which was the major oligosaccharide product released into the supernatant of Sh. flexneri Y cells treated with either Sf6 phage or with Sf6 TSP. This oligosaccharide is likely to be the expected octasaccharide resulting from endo-
glycanase cleavage at alternative RhaII-1,3-RhaIII gly-
cosidic linkages (Fig. 1). The oligosaccharide migrated similarly to maltolheptose (linear heptamer of glucose) and this may be a reflection of the conformation of the octasaccharide compared with the standard. These data are consistent with, and confirm, the hypothesis that Sf6 and Sf6 TSP have the previously described endo-
rhamnosidase activity (Lindberg et al., 1978).

Under the conditions used, we found that the Sf6 TSP was not active on Sa. typhimurium LT2 LPS, the substrate of the P22 TSP (Fig. 7, lane 7). Likewise, the P22 TSP exhibited no activity for Sh. flexneri PE577 LPS (Fig. 7, lane 4), highlighting the specificity of the individual enzymes of the corresponding bacteriophages for their natural host. Furthermore, the Sf6 TSP/ endorhamnosidase was only significantly active on Sh. flexneri of serotype Y or X LPS (Figs 6 and 7). This
corresponds to previous findings that only organisms of these two serotypes are sensitive to Sf6 (Clark et al., 1991). Slight activity was observed on PE649 (serotype 5a) (Fig. 7, lane 19); this can be accounted for by the fact that this serotype possesses the only O-antigen repeat structure besides serotypes X and Y which lacks any acetyl or glucosyl substitutions on RhaII (Gemski et al., 1975; Carlin & Lindberg, 1987). As TSP cleavage normally occurs at the 1,3-α-linkage between the unsubstituted RhaII and RhaIII of Sh. flexneri serotype X or Y, the lack of substitutions at RhaIII of serotype 5a may provide a more sterically favourable environment for TSP-mediated O-antigen cleavage. The insensitivity of Sh. flexneri serotype 3a and 3b LPS to Sf6 TSP provides direct support for O-antigen modification as a mechanism of superinfection exclusion. The modification of the O-antigen by lysogenic Sf6 via oac expression renders the O-antigen resistant to further TSP cleavage, denying access by superinfecting Sf6. Future studies will focus on using the Sf6 TSP in studies on Sh. flexneri O-antigen biosynthesis and function.

ACKNOWLEDGEMENTS

This study was funded by a Project Grant from the National Health and Medical Research Council of Australia. Mr Rakesh Manandhar is thanked for constructing pSf6Ps? and Professor Ken Sanderson for providing strains and methods.

REFERENCES


Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R. &


Received 8 December 1998; revised 16 March 1999; accepted 31 March 1999.