

Serotype 1a O-Antigen Modification: Molecular Characterization of the Genes Involved and Their Novel Organization in the *Shigella flexneri* Chromosome

PRADIP ADHIKARI,[†] GWEN ALLISON, BELINDA WHITTLE, AND NARESH K. VERMA*

Division of Biochemistry and Molecular Biology, Faculty of Science, The Australian National University, Canberra, ACT 0200, Australia

Received 15 March 1999/Accepted 7 May 1999

The factors responsible for serotype 1a O-antigen modification in *Shigella flexneri* were localized to a 5.8-kb chromosomal *Hind*III fragment of serotype 1a strain Y53. The entire 5.8-kb fragment and regions up- and downstream of it (10.6-kb total) were sequenced. A putative three-gene operon, which showed homology with other serotype conversion genes, was identified and shown to confer serotype 1a O-antigen modification. The serotype conversion genes were flanked on either side by phage DNA. Multiple insertion sequence (IS) elements were located within and upstream of the phage DNA in a composite transposon-like structure. Host DNA homologous to the *dsdC* and the *thrW proA* genes was located upstream of the IS elements and downstream of the phage DNA, respectively. The sequence analysis indicates that the organization of the 10.6-kb region of the Y53 chromosome is unique and suggests that the serotype conversion genes were originally brought into the host by a bacteriophage. Several features of this region are also characteristic of pathogenicity islands.

Shigella flexneri is an important human pathogen responsible for the majority of cases of endemic bacillary dysentery prevalent in developing nations. It has been estimated that shigellosis affects 200 million people worldwide and causes at least 650,000 deaths among children under 5 years of age per year (29). Poor sanitation and contaminated water supplies contribute to the infection by and spread of *Shigella*. Due to the expense involved in treating and preventing the disease, the development of an effective vaccine is required.

Evidence suggests that the immune response to shigellosis is serotype specific and that the O antigen acts as a protective epitope (22). *S. flexneri* is divided into 13 serotypes based on the structure of the O antigen, a component of the bacterial lipopolysaccharide (LPS) present on the outer membrane of the cell (26). The different *S. flexneri* serotypes, with the exception of serotype 6, contain the basic O-specific repeating tetrasaccharide unit which consists of the following: $\rightarrow 3$ - β -D-GlcNac-(1 \rightarrow 2)- α -L-Rha-(1 \rightarrow 2)- α -L-Rha-(1 \rightarrow 3)- α -L-Rha-(1 \rightarrow) (Fig. 1). The serotype containing the basic O antigen is referred to as serotype Y (26). Different serotypes result from modification of the basic O antigen which occurs through glucosylation and/or O acetylation of one or more sugars within the repeating unit. The factors responsible for the conversion to serotypes 2a, 3b, 5a, and X are encoded by lysogenic bacteriophages (6, 11, 12, 19, 27, 28). The serotype conversion loci in these phages contain three genes (6, 11, 12, 19). The first two genes are highly conserved and interchangeable, while the third gene is unique and encodes the glucosyltransferase, or Gtr, which mediates specific O-antigen modification. The addition of an O-acetyl group is mediated by an O-acetyltransferase encoded by the *oac* gene (27). The *gtrII*, *gtrV*, *gtrX*, and *oac* genes, which are involved in the conversion to serotypes

2a, 5a, X, and 3b, respectively, have recently been characterized (6, 11, 12, 19, 27, 28). In each case, the resident serotype-converting bacteriophages were inducible. Characterization of the phage genomes revealed that the genes involved in serotype conversion are located adjacent to the *int attP* region and that this organization was conserved in all cases. It is thought that phage-encoded serotype conversion factors may be used to develop recombinant, live, oral vaccine strains expressing different serotypes. SFL124 is an attenuated strain of *S. flexneri* serotype Y which has been shown to be safe and effective in human volunteers, and it provided protective immunity against challenge with wild-type serotype Y strains in monkeys (13, 14). SFL124 is a candidate vaccine strain that could be used in the construction of recombinant vaccines expressing different serotypes.

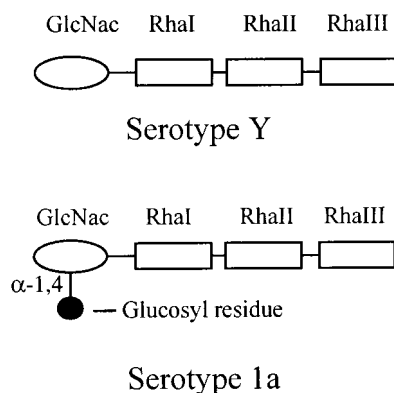
In serotype 1a strains, a glucosyl group is attached to the GlcNac residue of the repeating unit by an α -1,4 linkage (Fig. 1). Previous attempts to induce phage from 1a strains were unsuccessful. A chromosomal cosmid library was prepared from strain Y53 and probed with the *int* gene from SfV. Cosmid pNV394 hybridized to the *int* probe, and it was determined that a 5.8-kb *Hind*III fragment from this cosmid encoded factors which mediated the conversion of a serotype Y strain to serotype 1a (1). We now report on the molecular characterization of the O-antigen modification genes responsible for serotype 1a specificity, including their origins and locations in the genome of *S. flexneri* Y53.

Characterization of the 5.8-kb fragment. Bacterial strains and plasmids used in this study are listed in Table 1. *Escherichia coli* JM109 was used for routine transformation experiments, while SFL124 was used in serotype conversion experiments. Bacterial cultures were grown according to standard procedures in Luria-Bertani broth or agar (24). When necessary, media were supplemented with ampicillin (100 μ g/ml) or kanamycin (50 μ g/ml).

The 5.8-kb *Hind*III chromosomal fragment from *S. flexneri* serotype 1a strain Y53 was sequenced by generating successive deletions with the Erase-a-Base kit (Promega) and filling in the gaps by primer walking. The Genetics Computer Group

* Corresponding author. Mailing address: Division of Biochemistry and Molecular Biology, Faculty of Science, The Australian National University, Canberra, ACT 0200, Australia. Phone: 61 2 6249 2666. Fax: 61 2 6249 0313. E-mail: Naresh.Verma@anu.edu.au.

[†] Present address: Immunobiology Section, Therapeutic Goods Administration Laboratories, Woden, ACT, Australia.

FIG. 1. O-antigen structure of *S. flexneri* serotypes Y and 1a.

(University of Wisconsin) programs and programs available through the Australian National Genomic Information Service were used to analyze sequence data. Within the 5.8-kb fragment, a total of four complete open reading frames (ORFs) and one incomplete ORF were predicted (Table 2). Sequences homologous to IS600 were found on both ends of the fragment.

ORFs *orf1*, *orf2*, and *orf3* are transcribed in the same direction (Table 2). Putative ribosomal binding sites were identified upstream of each ORF. A promoter was identified within a reasonable distance upstream of *orf1* (−35 region, nucleotides [nt] 796 to 801; −10 region, nt 811 to 816), and a potential rho-independent transcriptional terminator was identified downstream of *orf3* (nt 3690 to 3715). The general organization of *orf1*, *orf2*, and *orf3* and the locations of putative transcriptional and translational signals suggest that it is likely that

these 3 ORFs form an operon. A database search revealed that the proteins encoded by *orf1* and *orf2* exhibit very high degrees of homology (88 to 99% identity) to proteins encoded by genes within the serotype conversion loci of *S. flexneri* bacteriophages SfII (19), SfV (11), and SfX (6) (Table 2). Homologues of these genes are also found in the *E. coli* K-12 genome (2, 19). Database comparisons revealed that there are no significant nucleotide or protein sequences homologous to *orf3*, suggesting that *orf3* is unique to *S. flexneri* 1a. The general organization of this putative operon is similar to that in phages SfII, SfV, and SfX, in which two conserved genes are followed by a gene which encodes the specific glucosyltransferase.

The regions up- and downstream of *orf1*, *orf2*, and *orf3* were analyzed. The region upstream of *orf1* shows a significant level of similarity to attachment sites (*attP*) of several bacteriophages. The nucleotide sequence between positions 601 and 646 is identical to the *attP* core sequence of SfV (11), SfII (19), P22 (15), and DLP12 (17) (Table 2). Two potential ORFs present on the complementary strand, *orf4* and *orf5'*, were identified downstream of *orf3* (Table 2). A putative Shine-Dalgarno sequence was identified upstream of *orf4*; however, a promoter sequence was not evident. The hypothetical protein encoded by *orf4* is almost identical to that encoded by *orf3* of SfV (Table 2), which is thought to play a role in tail fiber assembly of the phage particle (12). *orf5'* is interrupted by IS600 and is predicted to encode a protein of 170 amino acids. The first 66 amino acids of the Orf5 protein have a high degree of homology (94% identity) to the last 66 amino acids of the protein encoded by *orf2* of SfV (Table 2), which has no known function (12). The Orf5 protein also exhibits various degrees of homology to hypothetical proteins encoded by the *E. coli* genome (data not shown). One of these proteins is encoded by section 214 of the chromosome, which is located downstream

TABLE 1. Strains and plasmids used in this study

Strain or plasmid	Relevant characteristics	Reference or source
Strains		
<i>E. coli</i>		
JM109	<i>supE44 gyrA96 recA1 relA1 endA1 thi hsdR17 Δ(lac-proAB)</i> e14 [−] (F' <i>traD36 proAB lacI^q lacZΔM15</i>)	30
B789	JM109(pNV700)	This study
B799	JM109(pNV710)	This study
B800	JM109(pNV711)	This study
B801	JM109(pNV712)	This study
<i>S. flexneri</i>		
Y53	Serotype 1a	SIDC ^a
Y340	Serotype 1a	SIDC
Y829	Serotype 1a	SIDC
SFL124	<i>ΔaroD</i> , serotype Y	16
SFL1243	SFL124(pNV711)	This study
SFL1244	SFL124(pNV712)	This study
Plasmids		
pUC18	Cloning vector, Ap ^r	30
pBluescript II KS	Cloning vector, Ap ^r	Stratagene
pCR2.1	Cloning vector, Km ^r	Invitrogen
pNV394	pPR691::Y53 genomic DNA	1
pNV462	pUC19::5.8-kb <i>Hind</i> III fragment from pNV394	1
pNV700	pBluescript::5.8-kb <i>Eco</i> RI fragment from pNV394	This study
pNV710	pBluescript::4.8-kb <i>Sal</i> I fragment from pNV394	This study
pNV711	pUC18::gtrI	This study
pNV712	pUC18::orf1 bgt gtrI	This study

^a SIDC, Swedish Institute for Infectious Disease Control, Stockholm, Sweden.

TABLE 2. Sequence analysis of the 5.8-kb *Hind*III fragment of pNV394

ORF or feature	Nt position ^a	% GC	No. of amino acids encoded	Database search results	
				Feature or protein (source; accession no.)	% Identity
<i>attP</i>	601–646 (3342–3387)	NA ^b	NA	<i>attP</i> core (SfV; U82619, U82620)	100
				<i>attP</i> core (SfII; AF021347)	100
				<i>attP</i> core (P22; X04052)	100
				<i>attP</i> core (DLP12; M27155)	100
<i>orf1</i>	848–1210 (3589–3951)	42.7	120	Orf2 (SfII; AF021347)	98
				Orf6 (SfV; U82619, U82620)	99
				GtrA (SfX; AF056939)	89
				o120 (<i>E. coli</i> ; ECAE000323)	88
<i>orf2</i>	1207–2127 (3948–4868)	42.0	306	Bgt (SfII; AF021347)	97
				Orf5 (SfV; U82619, U82620)	97
				GtrB (SfX; AF056939)	95
				o306 (<i>E. coli</i> ; ECAE000323)	89
<i>orf3</i>	2130–3650 (4871–6391)	33.4	506	No significant homology	
<i>orf4</i> ^c	4387–3884 (7128–6625)	39.7	167	Orf3 (SfV; U82619, U82620)	91
				Hypothetical 18.3-kDa protein (<i>E. coli</i> ; ECAE000324)	68 (133-amino-acid overlap)
<i>orf5</i> ^c	4808–4296 (7549–7037)	50.1	170	Orf2 (SfV; U82619, U82620)	94 (66-amino-acid overlap)

^a The position relative to the 5.8-kb *Hind*III fragment is indicated. The nucleotide position relative to the complete 10.6-kb sequence is indicated in the parentheses.

^b NA, not applicable.

^c Present on complementary strand.

of section 213, where the *orf1* and *orf2* homologues of *E. coli* are found (2).

The general organization of this region in the Y53 chromosome is remarkably similar to that of other *S. flexneri* phages. The *attP* sites of SfII and SfV are located immediately upstream of the glucosyltransferase genes (11, 19). This organization is also conserved in the *Salmonella typhimurium* bacteriophage P22, which is involved in serotype conversion (19). In SfV, *orf3* and *orf2* are located downstream of the glucosyltransferase genes (11, 12). These data suggest that the DNA in this region of the Y53 chromosome is of phage origin and may be involved in serotype conversion.

Functional analysis of *orf1*, *orf2*, and *orf3*. To determine if the putative three-gene operon within the 5.8-kb *Hind*III fragment is involved in O-antigen modification, this region of the Y53 chromosome was introduced into SFL124. PCR was used to amplify *orf3*, using primers GTRF (5' AATGGATCCAACCTTTCCTCTTCGCGT [nt 1871 to 1889]) and GTRR (5' CATAAGCTTGTTGTATACGGCAACCAC [complementary to nt 3759 to 3740]). To amplify *orf1*, *orf2*, and *orf3*, primers GTRALL (5' TATGGATTCATCCCATCAACACTGCGC [nt 677 to 694]) and GTRR were used. The nucleotide positions of the primers refer to their positions within the 5.8-kb fragment only, and restriction sites incorporated into the primers are underlined. Plasmid pNV462 (Table 1) was used as a template for both reactions. The *orf3* PCR product was digested with *Bam*HI and *Hind*III and cloned into pUC18 to form pNV711. A 3.1-kb fragment containing the three ORFs was first cloned into pCR2.1 (Invitrogen) and then subcloned into the *Bam*HI and *Hind*III sites of pUC18 to form pNV712. In both constructs, the *lac* promoter of the vector was in the correct orientation with respect to the ORF(s).

SFL124 was transformed with pNV711 and pNV712, resulting in SFL1243 and SFL1244, respectively. LPS was extracted from the control and recombinant strains by the whole-cell

lysate method described elsewhere (9). LPS was separated by electrophoresis on a sodium dodecyl sulfate–12% polyacrylamide gel, transferred onto a nitrocellulose membrane (MSI), and analyzed by using serotype-specific monoclonal antibodies and chemiluminescence detection (kit from Boehringer Mannheim). The primary antibodies used were MASF-I and MASF-Y5, which are specific for serotype 1a and Y O antigen, respectively (3, 4) (Fig. 2). The LPS of wild-type 1a strain Y53 is recognized by MASF-I only; the LPS of SFL124 is recognized by MASF-Y5. The O antigen expressed by SFL1243 is recognized by both antibodies. This suggests that *orf3* alone mediates partial conversion of the SFL1243 LPS, which contains both Y-specific (parental) and 1a-specific O antigens on its surface. The LPS of SFL1244, however, is recognized by MASF-I only, indicating that all three ORFs together mediate complete serotype conversion. The same results were obtained when Western immunoblotting was repeated with alkaline phosphatase detection (data not shown). These results demonstrate that *orf1*, *orf2*, and *orf3* are directly involved in serotype 1a O-antigen modification.

The proteins encoded by *orf1* and *orf2* are highly conserved among serotype-converting bacteriophages (Table 2). The Orf1 protein is largely hydrophobic and contains four potential transmembrane regions. This protein exhibits ca. 45% identity to the RfbI protein, which is encoded within the *rfb* locus of *S. flexneri* and has an undetermined function (20). Guan et al. (6) proposed that the Orf1 protein may play a role in the translocation of lipid-linked glucose across the cytoplasmic membrane. The Orf2 protein homologue of SfII, called Bgt, was reported to have homology to other bactoprenol glucosyltransferases (19). Functional analysis of the Bgt homologue in SfX, called GtrB, confirmed bactoprenol glucosyltransferase activity and showed that it catalyzes the formation of the lipid P-glucose precursor (6). The Bgt proteins contain a largely hydrophilic N-terminal region and two potential transmembrane

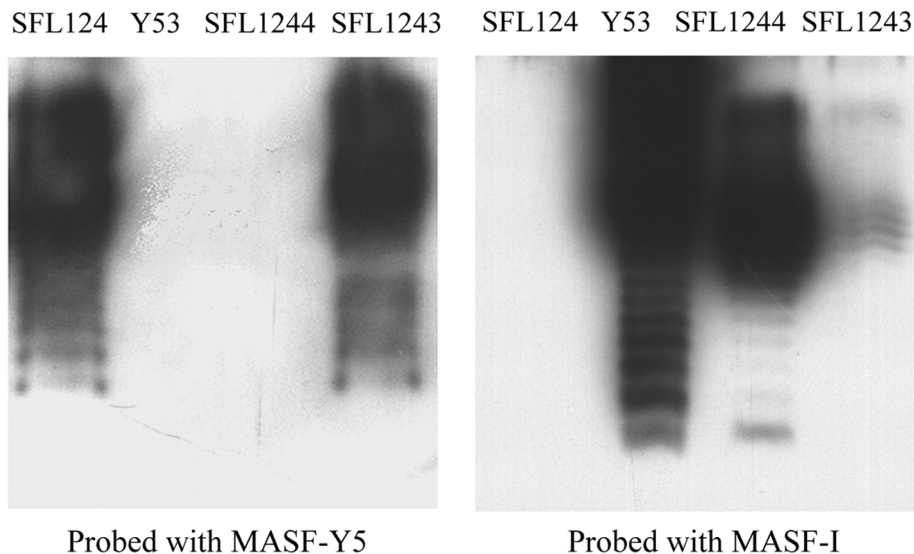


FIG. 2. Western immunoblot of LPS preparations. Membranes were probed with MASF-Y5 or MASF-I as the primary antibody, and O antigens were detected by chemiluminescence. SFL124 (serotype Y) and *S. flexneri* Y53 (serotype 1a) were used as controls. LPS prepared from SFL1243 (SFL124 containing *orf3*) is recognized by both Y- and 1a-specific monoclonal antibodies, whereas LPS from SFL1244 (SFL124 containing *orf1*, *orf2*, and *orf3*) is recognized only by MASF-I.

regions in the C terminus (19). Based on its homology and its role in O-antigen modification, *orf2* has been renamed *bgt*. *orf3* is unique to *S. flexneri* Y53 and encodes the serotype 1a-specific glucosyltransferase, so it has subsequently been renamed *gtrI*. Based on amino acid sequence, the homology of GtrI to the other Gtrs is less than 20% (data not shown). Analysis of hydrophathy plots, however, suggests that despite the differences in amino acid sequence, the glucosyltransferases have similar secondary structures consisting of 10 or 11 putative transmembrane segments (data not shown). The occurrence of low-level nucleotide and amino acid homology, yet similar secondary structure, appears to be a common phenomenon in proteins involved in certain steps of LPS and O-antigen synthesis (25).

The results of the seroconversion experiments described here are consistent with other studies with SFL124 in that the three-gene cassette from SfV or SfX was also required for the complete conversion of serotype Y to serotype 5a or X, respectively (6, 11, 12). Mavris et al. (19), however, reported that only *bgt* and *gtrII* or *gtrX* were required to mediate complete conversion to serotype 2a or X, respectively, in serotype Y strain PE577. Analysis of various strains of different serotypes indicated that more than one copy of *bgt* exists in any one strain, including *E. coli*, and it was proposed that the *bgt* homologue in PE577 is not functional (19). SFL124 may not produce functional Orf1 and Bgt proteins and would therefore require both for complete serotype conversion. In the absence of these genes, host-encoded factors, such as RfbI and other proteins with homology to Bgt, may compensate for their absence but may not be as specific or efficient, thus resulting in partial conversion.

Characterization of the Y53 chromosome up- and downstream of the serotype conversion genes. While the organization of *attP*, the O-antigen modification genes, and the *orf4 orf5* region appears to have been conserved, the general organization of the 5.8-kb *HindIII* fragment is unique in that the phage DNA appears to be flanked by copies of IS600. To determine the nature and extent to which the phage DNA was interrupted, the sequence was extended up- and downstream.

The 5.8-kb *HindIII* fragment in pNV462 was originally sub-

cloned from cosmid pNV394 (1). Since it appeared that duplicate copies of IS600 flanked the serotype conversion genes, pNV394 could not be used as a template for primer walking. Consequently, restriction mapping and Southern hybridization were used to identify appropriate fragments that overlapped with the known sequence. A 5.8-kb *EcoRI* fragment (1.2-kb overlap) and a 4.8-kb *SaII* fragment (2.9-kb overlap) were identified as being up- and downstream of the *gtr* genes. These fragments were cloned into pBluescript II KS (Stratagene) to form pNV700 and pNV710, respectively, which were then introduced into *E. coli* JM109, resulting in recombinant strains B789 and B799, respectively. The sequences of the fragments were determined by primer walking. In total, the sequences of 2,742 nt upstream and 2,012 nt downstream of the *HindIII* fragment in pNV394 were determined, and the results are summarized in Table 3 and Fig. 3.

The phage genes in the Y53 chromosome are flanked by direct copies of IS600, referred to as IS600u and IS600d, indicating the up- and downstream copies, respectively. The nucleotide sequences of IS600u and IS600d are 99.7% identical. Both copies are also highly homologous to IS600 from *Shigella sonnei* (18) and other IS600-related sequences (Table 3). IS600 insertion elements contain two ORFs, *orfA* and *orfB*, which encode proteins of 100 and 272 amino acids, respectively (18). The OrfA proteins of IS600u and IS600d are identical, while the OrfB proteins differ by one amino acid. These proteins are nearly identical to the OrfA and OrfB proteins of *S. sonnei*. Each insertion sequence (IS) element is flanked by 29-bp imperfect inverted repeats which are also conserved among IS600 sequences. An additional insertion element, IS629, is located 44 bp upstream of IS600u (Fig. 3). IS629 is composed of 1,306 nt and has significant homology to several insertion elements in the genetic databases, and its highest degree of homology is with *S. sonnei* IS629 (18) (Table 3). Similar to IS629 of *S. sonnei*, Y53 IS629 is predicted to contain two ORFs which are present on the complementary strand (Table 3). The protein encoded by the putative *orfA* is 99% identical to that encoded by *S. sonnei*. A 4-bp deletion beginning at nt 1082, however, resulted in a frameshift mutation in the *orfB*-encoded protein that resulted in premature termination and likely affected the

TABLE 3. Sequence analysis of the regions up- and downstream of the 5.8-kb *Hind*III fragment

Feature	Nt position (% GC)	Database search results	
		Origin, feature, and gene (accession no.)	% Identity, overlap size ^a
<i>dsdC</i> ^b	353–54 (44.7)	<i>E. coli dsdC</i> (ECAE000324)	93.4, 361 nt 94.9, 99 aa (120 aa total)
IS629 ^b	1660–355 (53.9)	<i>S. sonnei</i> IS629 (X51586, X05953, X05954) <i>E. coli</i> IS3411 (M19532)	98.9, 1,310 nt 98.6, 1,310 nt
IS600u	1705–2968 (49.5)	<i>S. flexneri</i> plasmid pMYSH6000 (U82621) <i>S. flexneri</i> SBA1336 IS600-like element (U97492) <i>S. sonnei</i> IS600 (X05952) <i>Shigella dysenteriae</i> cytotoxin gene (M24352, M21947)	99.4, 1,264 nt 99.4, 1,264 nt 99.3, 1,264 nt 99.3, 1,264 nt
IS600d	7550–8813 (49.1)	<i>S. flexneri</i> SBA1336 IS600-like element (U97492) <i>S. sonnei</i> IS600 (X05952) <i>S. dysenteriae</i> cytotoxin gene (M24352, M21947)	99.2, 1,264 nt 99.1, 1,264 nt 99.1, 1,264 nt
<i>int'</i>	8814–9782 (46.8)	SfV <i>int</i> (U82619, U82620) P22 <i>int</i> (X04052) SfII partial <i>int</i> (AF021347)	98.7, 972 nt 97.5, 323 aa (387 aa total) 76.9, 971 nt 89.4, 321 aa (387 aa total) 98.7, 620 nt 99.5, 205 aa
<i>thrW</i> tRNA gene ^b	9872–9797 (50.9)	<i>E. coli thrW</i> (ECAE000132)	100, 76 nt
<i>attR</i>	9797–9842	Refer to <i>attP</i> (Table 2)	100, 45 nt <i>attP</i> (Table 2)
<i>proA</i> ^b	10604–9987 (50.9)	<i>E. coli proAB</i> (ECAE000132)	99, 616 nt 98.5, 206-aa overlap with ProA (417 aa total)

^a aa, amino acid.
^b Present on complementary strand.

function of this insertion element. The first 68 amino acids encoded by the 5' end of *orfB* are nearly identical to the amino-terminal end of the OrfB protein of *S. sonnei*. IS600 and IS629 are frequently found in *Shigella* and other related bacteria, where they are often associated with virulence determinants (reviewed in reference 23). The organization of the multiple IS elements is similar to that of a composite transposon and suggests that this region of the Y53 chromosome may be mobile.

The sequence immediately upstream of IS629 has significant homology to the *dsdC* gene of *E. coli* K-12 (Table 3). This gene encodes the D-serine dehydratase (deaminase) transcriptional activator and is located at 53 min on the *E. coli* chromosome (2). The protein homology in this region is limited to the last 99 (out of 120 total) amino acids of DsdC, suggesting that the 5' end of the gene was interrupted by IS629. This end of the

insertion element, however, encodes 20 amino acids in the same frame as the partial protein. Furthermore, putative transcriptional (–10) and translational signals are located upstream of the ATG start codon in IS629. This may suggest that even though the insertion element disrupted the *dsdC* gene, it may be possible for the gene to be transcribed and translated although the amino-terminal end of the protein would differ from that in *E. coli*.

The region immediately downstream of IS600d has significant homology to the integrase genes (*int*) of SfV, SfII, P22, and other bacteriophages (Table 3). This region contains an incomplete ORF encoding a partial protein of 323 amino acids, suggesting that IS600d interrupted the 5' end of the *int'* gene. The partial ORF is transcribed in the same direction as the serotype conversion genes. A 46-nt sequence immediately downstream of *int'* is highly homologous to the *attP* sites of SfV

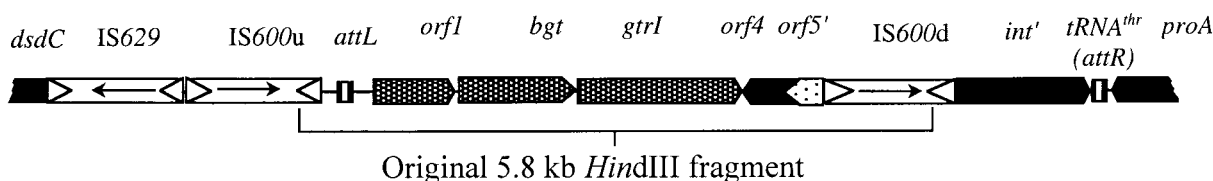


FIG. 3. The organization of the O-antigen modification genes in *S. flexneri* 1a. Arrows represent the direction of each IS. The O-antigen modification gene cluster is shaded.

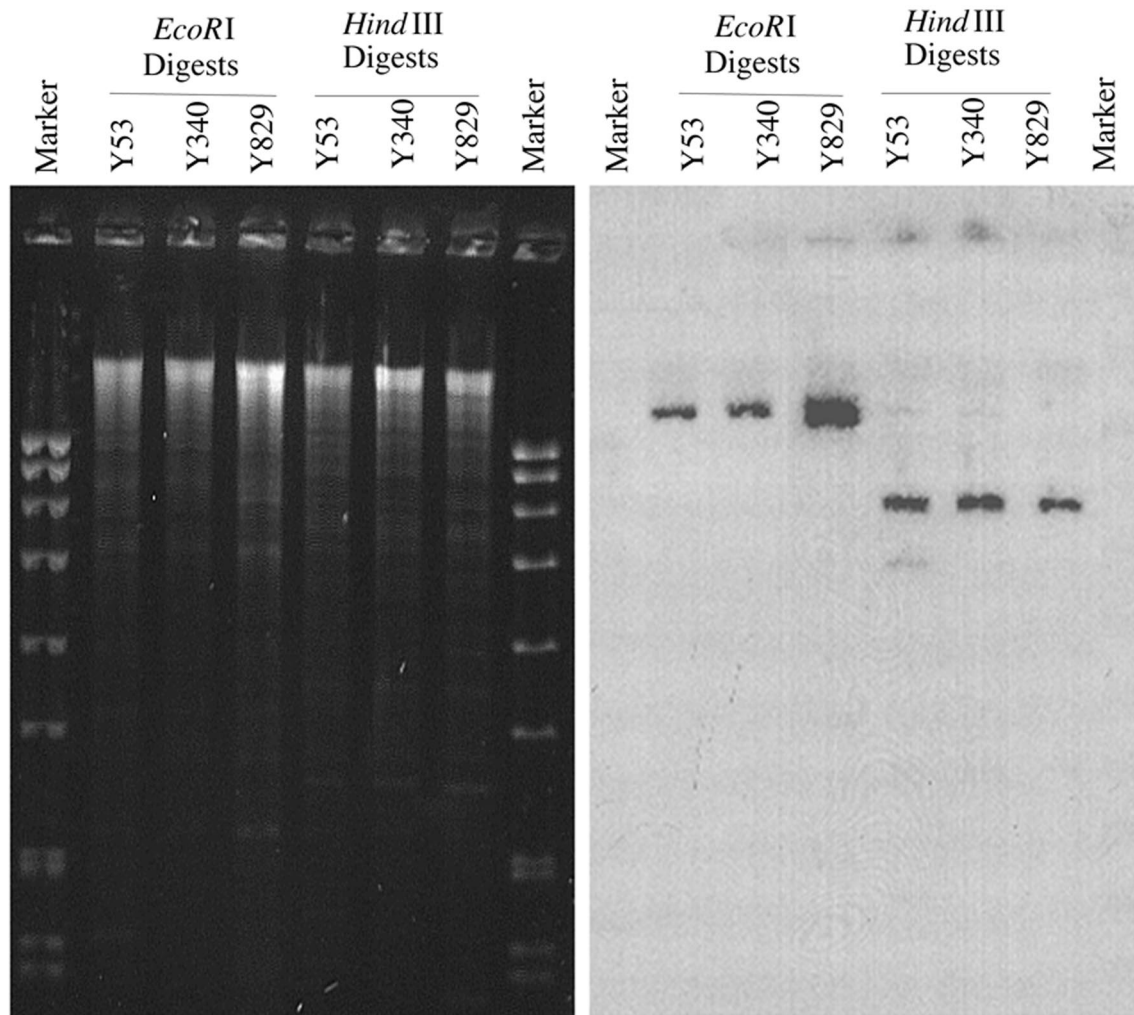


FIG. 4. Determination of copy number and organization of *gtrI* in wild-type *S. flexneri* 1a strains. The chromosomal DNA from the strains indicated was digested with *EcoRI* and *HindIII* and subjected to Southern hybridization with ^{32}P -labelled *gtrI* as a probe. The molecular weight marker was *EcoRI*-digested SPP1 bacteriophage DNA (Progen).

and related bacteriophages and is identical to the *attP* site located upstream of *orf1* (Table 2). The two *attP* sites are separated by ca. 6.5 kb. When a temperate phage integrates into the chromosome, the *attL* and *attR* sites are separated by the entire phage genome, which is ca. 40 kb for *S. flexneri* phages. The *int* and glucosyltransferase genes, which are adjacent on the phage genome, are located at opposite ends of the phage DNA once integrated into the chromosome but are still transcribed in the same direction. The organization of the *attP* sites, glucosyltransferase gene, and *int'* in the Y53 chromosome is reminiscent of a lysogen, although it appears that a large portion of the phage genome was deleted. It is possible that the associated IS elements played a role in the deletion when, for example, the insertion of direct repeats of IS600 into both *orf5* and *int* could have resulted in the homologous recombination and deletion of the intervening phage DNA. Based on this analogy, the *attP* sites upstream of *orf1* and downstream of *int'* have been renamed *attL* and *attR*, respectively, and they define the boundaries of the phage DNA in this area of the Y53 chromosome (Fig. 3).

The *attR* sequence is identical to the 3' end of the *thrW* tRNA gene. In fact, the DNA sequence downstream of and

including *attR* is remarkably similar to the *thrW proA* region of the *E. coli* chromosome (Table 3), which is located at ca. 6 min (2). Integration of lysogenic phage into the host chromosome frequently occurs in tRNA genes. For example, phage DLP12 is integrated into the arginine-accepting tRNA in *E. coli* (17) and P22 and SfV insert into the threonine-accepting tRNA in *Salmonella* and *S. flexneri*, respectively (7, 15). The finding that the cryptic Sf1 phage integrated into the *thrW proA* site is also consistent with earlier data suggesting that the *proAB* locus was the site of integration of *S. flexneri* seroconverting phage (21). It is of interest, however, that the *dsdC* gene upstream of *attL* does not map adjacent to the *proAB* locus in the *E. coli* chromosome. While it is possible that the organization of these genes in *S. flexneri* is different than that in *E. coli*, it is also tempting to speculate that the IS elements may be involved in a chromosomal rearrangement resulting in the placement of these genes adjacent to one another.

The sequence analysis indicates that the organization of the 10.6-kb region of the Y53 chromosome is unique and suggests that the serotype conversion genes were originally brought into the host by a bacteriophage. Several features of this region of the Y53 chromosome are also characteristic of pathogenicity

islands, which are distinct chromosomal segments that contain clusters of genes responsible for a particular virulent phenotype (reviewed in references 5 and 8). There are eight criteria used to define pathogenicity islands (8), and the 10.6-kb region of the Y53 chromosome exhibits many of these. A ca. 9-kb region represents a distinct genetic unit which is flanked by insertion elements on one end and a cryptic integrase gene and tRNA gene on the other end. The phage DNA contained within this unit has a dramatically lower GC content than that in *S. flexneri* (Tables 2 and 3). The overall GC content of the region flanked by copies of IS600 is 40%, whereas the GC content of *Shigella* is 50% (2). The GC content of the flanking IS elements and the *thrW proAB* genes is much closer to that expected for the host. The O antigen has been shown to be an important virulence factor of *S. flexneri* (10). The O-antigen modification genes, by generating antigenic variation, enhance the virulence properties of *S. flexneri* since the existence of several serotypes means that the host has to mount a specific immune response to each one. The suggestion that the maintenance of antigenic variation is beneficial to *Shigella* is also supported by the fact that, in strain Y53, the genes in the prophage region not directly involved in O-antigen modification have been either deleted or inactivated by ISs.

Organization and distribution of *gtrI* in natural isolates of *S. flexneri* 1a. The presence of multiple insertion elements flanking the putative O-antigen-modification genes gives the entire region a composite transposon-like structure, suggesting that this region of the Y53 chromosome may be mobile. Based on the restriction map of the 5.8-kb *Hind*III fragment, a Southern hybridization experiment was designed to investigate the distribution and organization of the serotype conversion loci in other serotype 1a strains. Chromosomal DNA from three natural isolates of *S. flexneri* 1a was prepared by the procedure outlined by Bastin et al. (1) and was digested with restriction enzymes. Chromosomal fragments were separated by electrophoresis and subjected to Southern hybridization. The radio-labelled *gtrI* gene was chosen as a probe because it is specific for *S. flexneri* 1a. *Eco*RI and *Hind*III were chosen because *Eco*RI cuts within *orf1*, hence allowing the determination of the number of copies of the *gtrI* locus present in the genome, whereas *Hind*III cuts within IS600, which should reveal the organization of the O-antigen modification gene clusters in different strains. One *Eco*RI fragment hybridized with the probe in all three strains, suggesting that only one copy of *gtrI* was present (Fig. 4). In addition, all three strains tested contained a 5.8-kb *Hind*III fragment that hybridized to *gtrI*, which suggests that the composite transposon-like organization of the O-antigen modification genes was conserved in the strains analyzed.

Conclusions. The genes encoding type I antigen specificity in *S. flexneri* serotype 1a have been characterized and their functions in serotype conversion were confirmed. The general organization of the region encoding 1a O antigen has numerous unique characteristics, which raises a number of questions regarding the evolution of serotype conversion genes in *Shigella* and stresses their importance in virulence.

Nucleotide sequence accession number. The nucleotide sequence of the complete 10.6-kb region of the *S. flexneri* Y53 chromosome has been deposited in the GenBank database under accession no. AF139596.

We thank N. Carlin for the monoclonal antibodies and A. Lindberg for the *Shigella* strains.

This project was supported by a grant from the National Health and Medical Research Council of Australia.

REFERENCES

- Bastin, D. A., A. Lord, and N. K. Verma. 1997. Cloning and analysis of the glucosyl transferase gene encoding type I antigen in *Shigella flexneri*. FEMS Microbiol. Lett. 156:133-139.
- Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *E. coli* K-12. Science 277:1453-1474.
- Carlin, N. I. A., and A. A. Lindberg. 1986. Monoclonal antibodies specific for *Shigella flexneri* lipopolysaccharides: clones binding to type I and type III: 6,7,8 antigens, group 6 antigen, and a core epitope. Infect. Immun. 53:103-109.
- Carlin, N. I. A., and A. A. Lindberg. 1987. Monoclonal antibodies specific for *Shigella flexneri* lipopolysaccharides: clones binding to type IV, V, and VI antigens, group 3,4 antigen, and an epitope common to all *Shigella flexneri* and *Shigella dysenteriae* type 1 strains. Infect. Immun. 55:1412-1420.
- Groisman, E. A., and H. Ochman. 1996. Pathogenicity islands: bacterial evolution in quantum leaps. Cell 87:791-794.
- Guan, S., D. A. Bastin, and N. K. Verma. 1999. Functional analysis of the O antigen glucosylation gene cluster of *Shigella flexneri* bacteriophage SFX. Microbiology 145:1263-1273.
- Guan, S., and N. K. Verma. 1998. Serotype conversion of a *S. flexneri* candidate vaccine strain via a novel site-specific chromosome-integration system. FEMS Microbiol. Lett. 166:79-87.
- Hacker, J., G. Blum-Oehler, I. Muhldorfer, and H. Tschape. 1997. Pathogenicity islands of virulent bacteria: structure function and impact on microbial evolution. Mol. Microbiol. 23:1089-1097.
- Hitchcock, P. J., and T. M. Brown. 1983. Morphological heterogeneity among *Salmonella* lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. J. Bacteriol. 154:269-277.
- Hong, M., and S. M. Payne. 1997. Effect of mutations in *S. flexneri* chromosomal and plasmid-encoded lipopolysaccharide genes on invasion and serum resistance. Mol. Microbiol. 24:779-791.
- Huan, P. T., D. A. Bastin, B. L. Whittle, A. A. Lindberg, and N. K. Verma. 1997. Molecular characterization of the genes involved in O-antigen modification, attachment, integration and excision in *Shigella flexneri* bacteriophage SFV. Gene 195:217-227.
- Huan, P. T., B. L. Whittle, D. A. Bastin, A. A. Lindberg, and N. K. Verma. 1997. *Shigella flexneri* type-specific antigen V: cloning, sequencing and characterization of the glucosyl transferase gene of temperate bacteriophage SFV. Gene 195:207-216.
- Karnell, A., B. A. Stocker, S. Katakura, F. P. Reinhold, and A. A. Lindberg. 1992. Live oral auxotrophic *Shigella flexneri* SFL124 vaccine with a deleted *aroD* gene: characterization and monkey protection studies. Vaccine 10:389-394.
- Karnell, A., H. Sweiha, and A. A. Lindberg. 1992. Auxotrophic live oral *Shigella flexneri* vaccine protects monkeys against challenge with *S. flexneri* of different serotypes. Vaccine 10:167-174.
- Leong, J. M., S. Nunes-Dueby, C. F. Lesser, P. Youderian, M. M. Susskind, and A. Landy. 1985. The phi-80 and P22 attachment sites: primary structure and interaction with *E. coli* integration host factor. J. Biol. Chem. 260:4468-4477.
- Lindberg, A. A., A. Karnell, B. A. Stocker, S. Katakura, H. Sweiha, and F. P. Reinhold. 1988. Development of an auxotrophic oral live *Shigella flexneri* vaccine. Vaccine 6:146-150.
- Lindsey, D. F., D. A. Mullin, and J. R. Walker. 1989. Characterization of the cryptic lambdaoid prophage DLP12 of *Escherichia coli* and overlap of the DLP12 integrase gene with the tRNA gene *argU*. J. Bacteriol. 171:6197-6205.
- Matsutani, S., H. Ohtsubo, Y. Maeda, and E. Ohtsubo. 1987. Isolation and characterisation of IS elements repeated in the bacterial chromosome. J. Mol. Biol. 196:445-455.
- Mavris, M., P. A. Manning, and R. Morona. 1997. Mechanism of bacteriophage Sfil-mediated serotype conversion in *Shigella flexneri*. Mol. Microbiol. 26:939-950.
- Morona, R., M. Mavris, A. Fallarino, and P. A. Manning. 1994. Characterization of the *rfa* region of *Shigella flexneri*. J. Bacteriol. 176:733-747.
- Petrovskaya, V. G., and T. A. Licheva. 1982. A provisional chromosome map of *Shigella* and the regions related to pathogenicity. Acta Microbiol. Acad. Sci. Hung. 29:41-53.
- Phalipon, A., M. Kaufmann, P. Michetti, J. M. Cavillon, M. Huerre, P. Sansonetti, and J. P. Kraehenbuhl. 1995. Monoclonal immunoglobulin A antibody directed against serotype-specific epitope of *Shigella flexneri* lipopolysaccharide protects against murine experimental shigellosis. J. Exp. Med. 182:769-778.
- Rajakumar, K., C. Sasakawa, and B. Adler. 1997. Use of a novel approach, termed island probing, identifies the *Shigella flexneri* *she* pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. Infect. Immun. 65:4606-4614.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

25. **Schnaitman, C. A., and J. D. Klena.** 1993. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol. Rev.* **57**:655–682.
26. **Simmons, D. A., and E. Romanowska.** 1987. Structure and biology of *Shigella flexneri* O antigens. *J. Med. Microbiol.* **23**:289–302.
27. **Verma, N. K., J. M. Brandt, D. J. Verma, and A. A. Lindberg.** 1991. Molecular characterization of the O-acetyl transferase gene of converting bacteriophage Sf6 that adds group antigen 6 to *Shigella flexneri*. *Mol. Microbiol.* **5**:71–75.
28. **Verma, N. K., D. J. Verma, P. T. Huan, and A. A. Lindberg.** 1993. Cloning and sequencing of the glucosyl transferase-encoding gene from converting bacteriophage X (SfX) of *Shigella flexneri*. *Gene* **129**:99–101.
29. **World Health Organization.** 1991. Research priorities for diarrhoeal disease vaccines: memorandum from a WHO meeting. *Bull. W. H. O.* **69**:667–676.
30. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**:103–119.