Bacteriophage-encoded glucosyltransferase GtrII of *Shigella flexneri*: membrane topology and identification of critical residues

Adele M. LEHANE, Haralambos KORRES and Naresh K. VERMA¹

School of Biochemistry and Molecular Biology, Faculty of Science Building 41, The Australian National University, Canberra ACT 0200, Australia

The *Shigella flexneri* serotypes differ in the nature of their O-antigens. The addition of glucosyl or O-acetyl groups to the common backbone repeat units gives rise to the different serotypes. GtrII glucosylates rhamnose III of the O-antigen repeat unit, thus converting serotype Y (which has no modifications to the basic O-antigen repeat unit) into serotype 2a, the most prevalent serotype. In the present study, the topology of GtrII has been determined. GtrII has nine transmembrane helices, a re-entrant loop and three large periplasmic regions. Four critical residues (Glu⁴⁰, Phe⁴¹⁴,

Cys⁴³⁵ and Lys⁴⁷⁸) were identified in two of the periplasmic regions. Despite the lack of sequence similarity between GtrII and the Gtrs from other serotypes, three of the critical residues identified are conserved in the remaining Gtrs. This is consistent with some degree of mechanistic conservation in this functionally related group of proteins.

Key words: dual reporter, glucosyltransferase, GtrII, O-antigen, serotype conversion, *Shigella*.

INTRODUCTION

Shigellosis (bacillary dysentery) is a major diarrhoeal disease caused by members of the Gram-negative bacterial genus *Shigella*. All four *Shigella* species cause shigellosis, with *Shigella flexneri* causing the majority of cases and deaths [1]. There are 13 well-recognized *S. flexneri* serotypes, which differ in the nature of their O-antigens. The O-antigen is the outer component of lipopolysaccharide, and, in the case of *S. flexneri*, consists of the tetrasaccharide *N*-acetylglucosamine-rhamnose-rhamnose-rhamnose repeated many times [2]. Most of the *S. flexneri* serotypes arise by the addition of glucosyl groups to one or more of the O-antigen sugars by one of several linkages. Three genes encoded by temperate bacteriophages mediate this process: *gtrA*, *gtrB* and *gtr*_(type) [3]. The former two are conserved among the serotypes; *gtr* is serotype-specific. The encoded proteins are all integral membrane proteins [3]. According to the model proposed by Guan et al. [4], GtrB catalyses the transfer of a glucosyl group from UDP-glucose to the plasma membrane lipid UndP (undecaprenyl phosphate). GtrA then flips UndP-glucose across the membrane, and Gtr attaches the glucosyl group to the O-antigen. Gtr may also recycle the lipid carrier and assist GtrA in flipping UndP-glucose across the membrane [4]. GtrII attaches the glucosyl group to rhamnose III of the O-antigen repeat unit via an α 1,4 linkage, giving rise to serotype 2a [5], the most prevalent serotype [1]. O-antigen modification is thought to take place at a stage in O-antigen biosynthesis when the O-antigen chain is attached to undecaprenyl pyrophosphate in the periplasmic leaflet of the plasma membrane, and has not yet been attached to the lipid A-core polysaccharide precursor and exported to the outer membrane [6,7].

In the present study, the topology of GtrII was determined by creating a series of GtrII-reporter fusion proteins. The dual reporter consisting of alkaline phosphatase and the α -fragment of β -galactosidase [8] was used. With this system, alkaline phosphatase and β -galactosidase activities can be determined simultaneously, thereby eliminating the need for the time-consuming

switching of fusions. The calculation of NARs (normalized activities ratios) corrects for variable expression, thus it is not necessary to consider protein synthesis rates. This is based on the assumption that the level of expression of the fusion protein will affect the absolute activities of alkaline phosphatase and β -galactosidase, but not the ratio of their activities [8]. Alkaline phosphatase is only active in the periplasm because the final folding steps (and thus functional activation) of the enzyme can only occur in the periplasmic environment [9]. In contrast, β -galactosidase activity is restricted to the cytoplasm, as the α -fragment must interact with cytoplasmically located ω -fragments to cause α -complementation [10]. Bacteria in which alkaline phosphatase or β -galactosidase are active will form blue and red colonies respectively on dual-indicator plates containing their substrates, X-phos (5-bromo-4-chloroindol-3-yl phosphate) and Red-Gal (6-chloroindol-3-yl- β -D-galactoside) respectively.

In addition to GtrII, the Gtrs from the *S. flexneri* serotypes 5a, 1a, 4a and X have been identified (reviewed in [3]). Although there is little sequence similarity between them, several conserved residues were identified in the present study and mutated in GtrII. Critical residues were identified in two regions shown by topological analysis to correspond to large periplasmic regions. These residues may be involved in O-antigen glucosylation, a process known to occur in the periplasm. Furthermore, their conservation in all the Gtrs provides the first evidence of mechanistic conservation in this group of proteins.

EXPERIMENTAL

Bacterial strains and growth conditions

The *Escherichia coli* strains used in the present study are derivatives of K-12, and the *S. flexneri* strains are derived from the attenuated serotype Y vaccine strain SFL124 [11]. For full details of strains used, see the Supplementary Tables at <http://www.BiochemJ.org/bj/389/bj3890137add.htm>. All strains were grown

Abbreviations used: Exo, exonuclease III; LB, Luria-Bertani; NAR, normalized activities ratio; Red-Gal, 6-chloroindol-3-yl- β -D-galactoside; UndP, undecaprenyl phosphate; X-phos, 5-bromo-4-chloroindol-3-yl phosphate.

¹ To whom correspondence should be addressed (email Naresh.Verma@anu.edu.au).

aerobically at 37 °C in LB (Luria–Bertani) medium. The concentrations of ampicillin, kanamycin or chloramphenicol used for plasmid maintenance were 100 µg/ml, 50 µg/ml and 30 µg/ml respectively. Dual-indicator plates for the detection of alkaline phosphatase and β -galactosidase activities were made as described previously [8], with the appropriate antibiotic. X-phos and Red-Gal were purchased from Sigma and Research Organics respectively.

DNA techniques

Oligonucleotide primers used for PCR were purchased from Invitrogen or Proligo. For details, see the Supplementary Tables at <http://www.BiochemJ.org/bj/389/bj3890137add.htm>. DNA sequencing was performed at the Biomolecular Resources Facility, John Curtin School of Medical Research, Australian National University. DNA sequencing was performed with the ABI 3730 capillary sequence analyser using the Big Dye Version 3.1 sequencing protocol. PCR was performed using Pfu polymerase (Promega), as specified by the manufacturer. Colony PCR was performed according to the protocol in [12]. Restriction enzymes and T4 DNA ligase were obtained from Amersham Biosciences and Promega respectively, and were used as specified by the manufacturers. Plasmids (for details, see the Supplementary Tables at <http://www.BiochemJ.org/bj/389/bj3890137add.htm>) were maintained in JM109 and were prepared using the Qiagen MiniPrep kit. *E. coli* and *S. flexneri* strains were transformed either using RbCl₂ protocols [13] or by electroporation [14].

Templates for topology studies

In order to create *gtrII-phoA/lacZ* fusions using the Exo (exonuclease III) deletion approach, a construct containing *gtrII* and *phoA/lacZ* with a restriction site that leaves an overhang susceptible to Exo digestion closest to the end of *gtrII* and a restriction site that generates an Exo-resistant overhang just upstream of *phoA/lacZ* was required. A similar construct containing *gtrV* rather than *gtrII* (pNV1090) had been constructed previously [15]. The majority of pNV1090, except for the *gtrV* gene, was amplified using a forward primer containing an NsiI site (bold) that anneals directly upstream of *phoA/lacZ* in pNV1090 (pNV1090NsiIF: 5'-ACAATGCATAATTCGATGGGCGAGCTCCAGGC-3') and a reverse primer containing a BamHI site (bold) that anneals upstream of *gtrV* (pNV1090BamHIR: 5'-ACAGGATCCCAGCTTTTGTTCCTTTAGTGAG-3'). *gtrII* was amplified from pNV1038 using a forward primer with a BamHI site (bold) (gtrIIBamF: 5'-ACAGGATCCGACCCAAATACATCATAA-3') and a reverse primer with an NsiI site (bold) that anneals to a region in pNV1038 containing an XbaI site (underlined) (gtrIINsiIR: 5'-ACAATGCATGTGCTGACTCTAGAAACGGTTAG-3'). The two PCR products were digested with BamHI and NsiI, and were ligated to create pNV1216 (Figure 1). Upon digestion, the NsiI site closest to *phoA/lacZ* leaves an Exo-resistant 3' overhang, and the XbaI site closest to *gtrII* provides an Exo-susceptible 5' overhang.

pNV1215 was used for the creation of *gtrII-phoA/lacZ-gtrII* sandwich fusions. It was constructed by digesting the *gtrII* PCR product described above with BamHI and XbaI, and inserting it into pBluescript II KS digested with the same enzymes. The functionality of *gtrII* in pNV1216 and pNV1215 was confirmed by transforming the constructs into SFL1616 (described below).

Construction of *gtrII-phoA/lacZ* fusions

The Exo deletion approach [16] was used to create a series of *gtrII-phoA/lacZ* fusions in which various lengths of the 3'

Figure 1 Template for the creation of *gtrII-phoA/lacZ* fusions

pNV1216 contains *gtrII* and *phoA/lacZ* in tandem with an NsiI site closest to *phoA/lacZ* that leaves an Exo-resistant 3' overhang and an XbaI site closest to the end of *gtrII* that provides an Exo-susceptible 5' overhang. The annealing sites of the sequencing primers M13R and PHOSEQ are shown. Cm^R, chloramphenicol resistance.

end of the *gtrII* gene were deleted. The Promega Erase-a-Base kit was used. pNV1216 was linearized using NsiI and XbaI. Exo was used to progressively delete *gtrII* from its 3' end via the 5' overhang produced by XbaI digestion, according to the method in [17]. Single-stranded DNA was then removed using S1 nuclease; Klenow fragment and the four dNTPs were then used to ensure that the ends were blunt. The ends were ligated and the plasmids were transformed into JM109 and plated on to dual-indicator plates. Plasmid DNA from red, blue and purple colonies was analysed by restriction digests, and the exact point of fusion between *gtrII* and *phoA/lacZ* was determined by sequencing using the PHOSEQ primer [8].

PCR was also used to create *gtrII-phoA/lacZ* fusions. In this approach, the majority of pNV1216 was amplified using the forward primer pholacHF (5'-TTGGGCCCTGTTCTGGAAACCGGG-3'), which anneals at the beginning of the *phoA/lacZ* sequence, and a reverse primer that anneals at the point of interest in *gtrII*. The PCR products were treated with DpnI (Fermentas) to remove template DNA, then self-ligated and transformed into JM109. The colonies were grown on dual-indicator plates, and sequencing using the PHOSEQ primer [8] was used to confirm the presence of the intended in-frame fusions in plasmids prepared from coloured colonies.

Construction of *gtrII-phoA/lacZ-gtrII* sandwich fusions

The sandwich fusion approach was invented by Ehrmann et al. [18]. HpaI sites were introduced into the *gtrII* sequence in pNV1215 using site-directed mutagenesis (see below). Mutated constructs were identified using HpaI digests. Retention of function was confirmed as described below. *phoA/lacZ* was excised from pMA632 [8] using an EheI/Ecl136II double digest and ligated with the HpaI-digested constructs. The ligation mixtures were transformed into JM109, and colonies showing colouration on dual-indicator plates were investigated further using restriction digests. The insertion of *phoA/lacZ* at the intended site in the correct orientation was verified by sequencing using the PHOSEQ [8] or M13R primer.

Assays of alkaline phosphatase and β -galactosidase activities

JM109 strains bearing pNV1216 (background control) and fusion constructs (Table 1) were grown and induced with IPTG (isopropyl β -D-thiogalactoside) as described previously [8], and the

Table 1 Analysis of GtrII/dual reporter fusions

C-terminal fusions include those created by Exo deletion and PCR.

Sample ID	AA*	AP†	BG†	NAR (AP/BG)‡	Location§	Colour
C-terminal fusions						
B1491	Asn ¹¹	14.7	20.4	1:1	c.1	Red
B1492	Asn ³⁶	16.4	0.0	> 100:1	p.2	Blue
B1494	Gly ⁵²	40.3	0.2	> 100:1	p.2	Blue
B1495	Leu ⁹⁹	0.0	8.8	1: > 100	c.3	Red
B1489	Arg ¹⁴⁵	0.0	30.8	1: > 100	c.5	Red
B1490	Gly ¹⁵¹	0.0	46.6	1: > 100	c.5	Red
B1496	Ile ¹⁸⁸	0.2	5.9	1:29	c.5	Red
B1485	Ser ²²⁴	0.5	0.0	NA	p.6	Blue
B1497	Arg ²⁸⁴	5.2	29.0	1:6	c.7	Red
B1498	Asn ³³⁹	4.5	100.0	1:22	t.9-c.9	Red
B1504	Gly ³⁵⁰	22.3	6.1	4:1	t.9-p.10	Blue
B1499	Lys ³⁷³	9.5	1.5	6:1	p.10	Blue
B1500	Leu ³⁸⁶	82.0	5.5	15:1	p.10	Blue
B1501	Lys ⁴¹²	100.0	23.3	4:1	p.10	Blue
B1503	Asn ⁴⁶⁰	4.8	0.0	> 100:1	p.10	Purple
B1443	Pro ⁴⁸³	1.5	0.2	7:1	p.10	Purple
Sandwich fusions						
B1460	Leu ¹¹⁹	12.3	0.0	> 100:1	p.4	Blue
B1461	Val ¹⁶⁶	8.27	0.0	> 100:1	re (p.5)	Blue
B1478	Val ³⁰⁶	0.1	0.0	NA	p.8	Blue

* Amino acid (AA) position of the final residue of GtrII followed by alkaline phosphatase (AP)/ β -galactosidase (BG).

† Percentages of AP and BG activities relative to the maximum activity in the set.

‡ Normalized AP/BG activities ratio (NAR) rounded to the nearest integer. NA indicates that the AP and BG activities are too low to generate a reliable NAR.

§ Location of the fusion on the final topological model of GtrII (Figure 2); c, cytoplasm; p, periplasm; re, re-entrant loop; t, transmembrane helix.

|| Colony colouration on dual-indicator plates.

alkaline phosphatase and β -galactosidase activities were determined as described previously [19,20]. Background activities were subtracted from experimental data. NARs were calculated as follows after obtaining the alkaline phosphatase and β -galactosidase activities for each fusion:

$$\text{NAR} = (\text{alkaline phosphatase activity/highest alkaline phosphatase activity}) / (\beta\text{-galactosidase activity/highest } \beta\text{-galactosidase activity}).$$

The NARs were calculated using alkaline phosphatase and β -galactosidase activities that were averaged from two independent experiments, each of which included duplicates.

Site-directed mutagenesis

Site-directed mutagenesis was performed as described in the protocol from the QuikChange Site-Directed Mutagenesis kit (Stratagene). Briefly, the plasmid was amplified by PCR using primers containing the desired mutation(s). The methylated non-mutated parental DNA was digested with DpnI, and the PCR product was transformed into XL-1 Blue competent cells [21], which repair the nicks in the mutated plasmid.

Functional determination of Gtrs

The function of a *gtr* gene is evident when a serotype Y *S. flexneri* strain is converted into the serotype containing the relevant *gtr* gene (2a for *gtrII*). The function of *gtrII* was tested by transforming the constructs into SFL1616, a serotype Y *S. flexneri*

strain containing *gtrA* and *gtrB*. *gtrA* and *gtrB* from bacteriophage SfV were amplified as one fragment from pNV323 using the primers *gtrABamHIF* (5'-GGTGGATCCGGTGCCGATAAT-AGGAGT-3') and *gtrBPstIR* (5'-CGTCTGCAGCATGAGCA-TCTTCTGCCC-3'), which contain BamHI and PstI sites (bold) respectively. The *gtrA/gtrB* fragment and pACYC177 were digested with BamHI and PstI and were ligated to create pNV1241. pNV1241 was transformed into the serotype Y candidate vaccine strain SFL124 to create SFL1616.

Several SFL1616 colonies transformed with the construct of interest were streaked on to LB agar plates containing the appropriate antibiotics. Once grown, these lawns were used in slide agglutination tests using *S. flexneri* type II antibody (Denka Seiken). Cells were mixed directly into a drop of antibody on a glass slide, and the slide was rocked gently while being monitored for agglutination. For a negative control, cells were mixed in saline (0.9 % NaCl) instead of antibody. Immunogold labelling followed by electron microscopy was used to confirm modification (or lack thereof) of the O-antigen. This was performed essentially as described previously [22] using carbon-coated copper grids. The primary antibody was *S. flexneri* type II antibody (Denka Seiken) and the secondary antibody was anti-rabbit IgG conjugated to 10 nm gold particles (British Biocell International). They were diluted 1/10 and 1/9 respectively in PBS with 0.4 % BSA. Negative staining was performed for 30 s using 0.5 % uranyl acetate. Grids were viewed under a transmission electron microscope (Hitachi H-7100FA, Electron Microscopy Unit, Research School of Biological Sciences, Australian National University).

RESULTS

Topology of GtrII

The rapid generation of random fusions using Exo deletion was the first approach to create *gtrII-phoA/lacZ* fusions. The template used was pNV1216 (Figure 1). pNV1216 was linearized using NsiI and XbaI, and Exo was used to progressively delete *gtrII* from its 3' end. Coloured colonies corresponding to in-frame *gtrII-phoA/lacZ* fusions resulting in alkaline phosphatase (blue or purple) or β -galactosidase (red) activity were obtained and analysed by colony PCR (*gtrIIBamF* and *PHOSEQ* primers) or restriction digests (*BamHI/HindIII*) to estimate the amount of the *gtrII* sequence remaining. Analyses of the red colonies revealed the presence of fusions at many points in the *gtrII* sequence. In contrast, analyses of constructs prepared from blue and purple colonies revealed only two size groups, corresponding to fusions close to the N- and C-termini of GtrII (results not shown).

A total of 36 (nine blue, nine purple and 18 red) fusions present within the coding region of *gtrII* were sequenced using the *PHOSEQ* [8] or M13R primer to determine the precise point of fusion. A total of 13 in-frame fusions were examined further with enzyme assays to calculate NARs [8]. The NARs are generally consistent with colour observations. The red fusion close to the N-terminus (after residue Asn¹¹) has an NAR of 1:1; however, inappropriate alkaline phosphatase activity during the enzyme assay has been documented previously when *phoA/lacZ* is fused to very small lengths of a protein [19].

The locations of the protein loops not covered using the Exo deletion approach were determined using either the sandwich fusion or the PCR-based approach. The fusions Leu¹¹⁹, Val¹⁶⁶ and Val³⁰⁶ were created using the sandwich fusion approach, which involves the introduction of *phoA/lacZ* into the complete *gtrII* sequence at the point of interest. JM109 colonies bearing these three fusion constructs showed blue colouration on dual-indicator plates. The NARs for fusions Leu¹¹⁹ and Val¹⁶⁶ are > 100:1

Figure 2 GtrII topology model and critical residues

The GtrII topology model after the creation of *gtrII-phoA/lacZ* fusions by the Exo deletion and PCR-based approaches, and *gtrII-phoA/lacZ-gtrII* sandwich fusions. Closed circles, C-terminal cytoplasmic fusions showing β -galactosidase activity; open circles, C-terminal periplasmic fusions showing alkaline phosphatase activity; open squares, sandwich fusions showing alkaline phosphatase activity; NA: the alkaline phosphatase and β -galactosidase activities measured in the enzyme assays are too low to generate a reliable NAR, but colour observations were in accordance with the proposed model. The fusion labels in boxes correspond to the last GtrII residue before the dual reporter. The positions of the mutated GtrII residues (shaded circles) that are critical for function are shown. The O-antigen modification mediated by GtrII was tested for by transforming the mutated constructs into SFL1616, then performing agglutination tests and immunogold labelling using the anti-*S. flexneri* type II antibody. Each residue was converted into alanine.

(Table 1), which are consistent with their predicted periplasmic locations (Figure 2). Alkaline phosphatase activity was low for JM109 carrying the Val³⁰⁶ fusion, but no β -galactosidase activity was detected. This is consistent with the predicted periplasmic location of Val³⁰⁶, but the alkaline phosphatase activity was too low to generate a reliable NAR. Since the creation of sandwich fusions is the most reliable approach to determine topology [8] and the loop in question only contains four to five residues, no further fusions were attempted in this loop. Furthermore, the red fusions Arg²⁸⁴ and Asn³³⁹ in the adjacent loops provide very strong support for the periplasmic location of Val³⁰⁶ (Figure 2).

Three fusions were created by PCR to confirm the locations of two remaining loops. The PCR approach results in a partially deleted *gtrII* gene fused at its end to *phoA/lacZ*, but in contrast with the Exo deletion approach, it allows the point of fusion to be chosen. The fusions created were Arg¹⁴⁵, Gly¹⁵¹ and Ser²²⁴. NARs of 1: > 100 for fusions Arg¹⁴⁵ and Gly¹⁵¹ are consistent with cytoplasmic localization (Table 1). There are too few residues between Gly¹⁵¹ and Ile¹⁸⁸ to form two complete transmembrane helices (only 36), but the central blue fusion Val¹⁶⁶ eliminates the possibility that the hydrophobic segment between them is a cytoplasmic loop. These fusions provide evidence for a re-entrant loop after helix IV (Figure 2). The possibility that the red fusions are present in the membrane cannot be excluded, since this has been observed previously (see for example [15]). However, in the case of fusion Ile¹⁸⁸, this would mean the presence of two positively charged residues in the membrane as well (Arg¹⁸³ and Lys¹⁸⁶).

Colonies bearing fusion Ser²²⁴ took several days at 37°C to develop blue colouration. The alkaline phosphatase and β -galactosidase activities in the enzyme assays were too low to generate reliable NARs. However, the red fusions obtained in the adjacent loops (Ile¹⁸⁸ and Arg²⁸⁴) provide further evidence that this highly charged loop is periplasmic (Figure 2). Two further

fusions were attempted unsuccessfully at different points in this loop (see the Discussion).

The reliability of NARs in providing data about reporter membrane localization is related to the size and diversity of the set of fusions [8]. The reliability of the data can be tested by choosing the second highest reference point in calculating the NARs [8]. This was performed (results not shown), and the NARs remained consistent with the previously determined localization of the reporter. The NAR for fusion Asn¹¹ was changed from 1:1 to 1:2, which provides further evidence for its cytoplasmic location.

The topology of GtrII is similar to that of GtrV, which was determined previously [15]. The major differences are that GtrII has a much larger periplasmic loop between helices V and VI, and a larger periplasmic C-terminus. The hydrophobicity profiles of GtrX and GtrI closely parallel those of GtrV and GtrII respectively. The topology of GtrIV appears to be quite different, although the N-terminal periplasmic loop and re-entrant loop might be conserved. The topological models of all the Gtrs are shown in Figure 3.

Identification of critical GtrII residues

Residues to mutate were chosen largely based on conservation between the Gtrs. Bioinformatics was of limited use in finding conserved residues between them, as a result of the low sequence homology and different lengths of the proteins. The potential motifs Glu-Gln-Xaa-Xaa-Lys, Lys-Lys and Phe-Phe were identified manually. Accordingly, the GtrII residues Lys⁴⁴⁷, Lys⁴⁷⁸ and Phe⁴¹⁴ were mutated. Based on previous studies, the GtrII residues Glu⁴⁰ and Cys⁴³⁵ were selected for mutagenesis. A critical glutamic acid residue (Glu⁴²) has been identified in the N-terminal periplasmic loop of GtrV (H. Korres, unpublished work), and a glutamic acid residue exists in approximately the same position in the remaining Gtrs. Chen et al. [23] discovered that a cysteine

Figure 3 Topological models of the Gtrs

Topological models of all the Gtrs identified to date. The topology of GtrV was determined previously [15]; the topologies of GtrI, GtrIV and GtrX shown here are based on results from computer prediction programs. Two tightly packed transmembrane segments predicted in GtrI and GtrX in the vicinity of the re-entrant loops in GtrV and GtrII were assumed to correspond to re-entrant loops.

residue in the large periplasmic C-terminus of GtrII (Cys⁴³⁷) is essential for function. A second cysteine residue exists in this region, Cys⁴³⁵, and was mutated to investigate whether Cys⁴³⁷ is catalytic or forms a structurally important disulphide bond with Cys⁴³⁵.

The mutations E40A, F414A, C435A, K447A and K478A were introduced by site-directed mutagenesis. Sequencing using the PHOSEQ [8] or M13R primer was used to verify the introduction of the desired mutation. The template used for mutagenesis was a construct in which *gtrII* is fused to *phoA/lacZ* (pNV1260 containing fusion Pro⁴⁸³). *gtrII* functionality was confirmed. The use of this construct as the template enables rapid testing of whether non-functional mutated proteins are assembled in the

Figure 4 Functional analysis of mutants

Electron micrographs ($\times 60\,000$ magnification) of *S. flexneri* strains treated with anti-(*S. flexneri* type II) antibody and a secondary antibody conjugated to 10 nm gold particles. The strain identities are shown below the micrographs. The presence of many gold particles on the bacterial surface corresponds to modification of the O-antigen by GtrII.

plasma membrane in JM109. If they are, purple colouration (as seen for the template) should be produced.

Constructs encoding the mutated proteins were transformed into SFL1616 (a serotype Y *S. flexneri* strain containing *gtrA* and *gtrB*), and agglutination tests were used to detect the O-antigen modification mediated by GtrII. The E40A, F414A, C435A and K478A mutations were found to destroy function: SFL1616 transformed with the constructs encoding these mutants did not agglutinate with type II antibody. The K447A mutant was functional. These results were confirmed using immunogold labelling and electron microscopy using the same antibody (Figure 4). Antibody binding is only seen in the positive control SFL1627 (SFL1616 carrying the non-mutated template, pNV1260) and in SFL1639 (SFL1616 carrying the K447A *gtrII* mutant). The positions of the critical GtrII residues are shown in Figure 2.

To determine whether the non-functional mutated proteins were assembled in the plasma membrane, JM109 containing the

constructs were grown on dual-indicator plates. Purple colouration (as seen for the template construct, pNV1260) revealed that the mutated proteins are assembled in the membrane, with the C-terminus of GtrII located in the periplasm.

DISCUSSION

The topology of GtrII was determined in the present study by creating a series of *gtrII-phoA/lacZ* and *gtrII-phoA/lacZ-gtrII* fusions. This revealed that GtrII has a cytoplasmic N-terminus, nine transmembrane helices, a re-entrant loop, a large periplasmic C-terminal region and two large periplasmic loops between helices I and II, and between helices V and VI. The fusion at the N-terminus of GtrII resulted in red colouration, which is indicative of cytoplasmic localization; the NAR of 1:1 is probably a result of inappropriate alkaline phosphatase activity in the enzyme assay, which has been documented previously for fusions in which the reporter gene is fused to a small protein region [19]. However, it is possible that this fusion is in the first transmembrane helix. Two periplasmic purple fusions were obtained, Asn⁴⁶⁰ and Pro⁴⁸³ with NARs of > 100:1 and 7:1 respectively. Purple colouration has been associated with transmembrane fusions in previous studies [8,15]. However, in the light of their NARs, speculations that fusions Asn⁴⁶⁰ and Pro⁴⁸³ are close to or within the membrane are not warranted.

The Exo deletion approach did not lead to the identification of fusions in most of the periplasmic loops (except for the large N- and C-terminal regions). Furthermore, in contrast with previous studies [8,15], few transmembrane fusions were obtained. The tight packing of the helices in GtrII may make fusions in such areas unstable. The failure to identify fusions in the large periplasmic loop between helices V and VI was surprising because of its large size. Colouration in colonies bearing the PCR-constructed fusion in this loop took several days to develop. If fusions in this area were generated by Exo deletion, they may have been overlooked. This PCR-constructed fusion is one of two fusions for which the reporter enzyme activities are too low to generate a reliable NAR. The fusion protein is probably unstable and less abundant in the membrane. Two further fusions were attempted in this loop: a sandwich fusion after residue Gln²³³ and a PCR-constructed fusion after residue Phe²⁵¹. However, no coloured colonies were obtained, presumably as a result of stability problems.

The other fusion for which a NAR could not be generated is the sandwich fusion between helices VII and VIII. It is not known why the intense blue colour produced by this fusion (which did not require a prolonged incubation) could not be translated into a reliable NAR, although this phenomenon has been reported previously for red fusions [8]. However, the red fusions on either side and the two hydrophobic segments between them require that this fusion is periplasmic.

The red fusions Gly¹⁵¹ and Ile¹⁸⁸, and the blue fusion Val¹⁶⁶, provide convincing evidence for the presence of a re-entrant loop, since there are too few residues between Gly¹⁵¹ and Ile¹⁸⁸ to form two complete transmembrane helices. A re-entrant loop was also discovered in the recently determined topology of GtrV [15]. Re-entrant loops have been documented in the bacterial potassium channel KcsA [24], in eukaryotic glutamate transporters [25], and in various other channels (reviewed in [26]). Given the association of re-entrant loops with permeation and the transporter- or channel-like secondary structure of GtrII, it is reasonable to hypothesize that GtrII may couple the flipping of UndP (and/or UndP-glucose) across the membrane to the proton motive force or electrochemical gradient. Alternatively, the re-entrant loop may provide the protein with the conformational flexibility required to

transfer a glucosyl group from a membrane lipid to the O-antigen, as proposed for GtrV [15].

The major difference between the topologies of GtrII and GtrV is that GtrII has a large periplasmic loop between helices V and VI. The role of this loop remains to be determined; its size and periplasmic localization would be consistent with some involvement in the O-antigen modification process. Also, the periplasmic C-terminus of GtrII is larger than that of GtrV. The hydropathy profiles of GtrX and GtrI are very similar to those of GtrV and GtrII respectively. The results of the present study are consistent with these four Gtrs sharing a high degree of structural conservation. GtrIV appears to be topologically dissimilar, although the N-terminal periplasmic loop (and possibly the re-entrant loop) may be conserved. The apparent conservation of the N-terminal topology (including the re-entrant loop) points to the existence of an N-terminal domain involved in functions conserved between all the Gtrs (for example, interactions with the donor substrate and conformational changes). The more structurally variable C-terminus may determine which O-antigen sugar is modified and by what linkage. GtrII and GtrV differ in both respects, while the pairs GtrV and GtrX, and GtrII and GtrI, differ only in the choice of acceptor sugar (the linkage is the same), and appear to have more highly conserved C-terminal topologies.

Converting topological information into an understanding of the mechanism of action of GtrII will require further work. However, three regions were identified that could be involved in the attachment of the glucosyl group to the O-antigen (which occurs in the periplasm): the periplasmic loops between helices I and II, and V and VI, and the large periplasmic C-terminus.

As a further step to elucidate the mechanism of O-antigen glucosylation, critical residues were identified in the N-terminal periplasmic loop and large C-terminal periplasmic region. Four critical GtrII residues were identified: Glu⁴⁰, Phe⁴¹⁴, Cys⁴³⁵ and Lys⁴⁷⁸. The mutated proteins were fused to the dual reporter, allowing rapid confirmation that the non-functional mutants were assembled in the plasma membrane in JM109.

The requirement for the GtrII residues Cys⁴³⁵ (the present study) and Cys⁴³⁷ [23] for function is consistent with the formation of a disulphide bond between them. This disulphide bond may be required for the positioning of other critical residues. The absence of a pair of cysteine residues in the C-terminal regions of GtrV, GtrX and GtrIV indicates that this is not a conserved feature between the Gtrs.

In contrast, the critical glutamic acid residue identified in the N-terminal periplasmic loop of GtrII has potential equivalents in the remaining Gtrs; the GtrV equivalent has already been shown to be critical (H. Korres, unpublished work). This points to a possible conserved element in the mechanisms of action of the Gtrs. These glutamic acid residues may catalyse the addition of glucosyl groups to the O-antigen. Many glycosyltransferases are thought to use glutamic acid or aspartic acid to assist in deprotonating the nucleophilic hydroxy group of the acceptor sugar [27].

Based on their conserved localization, it is likely that the dilysine and diphenylalanine motifs close to the C-terminus, which were shown in the present study to be critical in GtrII, are critical in all the Gtrs. Further occurrences of these motifs are scattered throughout the sequences of all the Gtrs. GtrI, GtrIV and GtrV all have four dilysine motifs; GtrII and GtrX have one each. GtrI, GtrV and GtrX have three diphenylalanine motifs, whereas GtrII and GtrIV have two. The motifs that are not positionally conserved may be evolutionary remnants that are no longer required for function. The roles of the critical dilysine and diphenylalanine motifs remain to be determined. Recently, it has been shown that C-terminal diphenylalanine and dilysine motifs

in a plant protein interact with each other and mediate protein–protein interactions [28]. No attempts have been made to identify proteins that interact with the Gtrs to date. A likely possibility not excluded in the model proposed by Guan et al. [4] is that the Gtrs may interact with GtrA to flip UndP-glucose across the plasma membrane.

The heterogeneous nature of the Gtrs is in sharp contrast with the high level of conservation of GtrA and GtrB, but the present study supports the notion that conserved residues involved in conserved functions do exist. It is not surprising that some sequence conservation was lost as each Gtr evolved to recognize a different acceptor and attach a glucosyl group via a specific linkage, but the extent seems surprising, since they all use the same donor substrate and catalyse a similar reaction. It has also been observed among eukaryotic glycosyltransferases that recognize identical donor or acceptor substrates that few regions of sequence homology exist, and that enzymes that are structurally related often catalyse the same or a similar reaction [29].

The present study provides the first experimental evidence that the Gtrs share structural similarity and have conserved elements in their mechanisms of action. The sequence divergence between them may be attributable to a small number of critical residues relative to residues that function only in structural support. Furthermore, the topological analysis of GtrII revealed which protein regions may be involved in O-antigen glucosylation, and the identification of critical residues in these regions provides the basis for the localization of the active site and the elucidation of the mechanism of action.

We thank Sally Stowe, Lily Shen and Cheng Huang at the Electron Microscopy Unit for their technical assistance.

REFERENCES

- Kotloff, K. L., Winickoff, J. P., Ivanoff, B., Clemens, J. D., Swerdlow, D. L., Sansonetti, P. J., Adak, G. K. and Levine, M. M. (1999) Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bull. World Health Org.* **77**, 651–666.
- Kenne, L., Lindberg, B., Petersson, K. and Romanowska, E. (1977) Basic structure of the oligosaccharide repeating unit of the *Shigella flexneri* O-antigen. *Carbohydr. Res.* **56**, 363–370.
- Allison, G. E. and Verma, N. K. (2000) Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *Trends Microbiol.* **8**, 17–23.
- Guan, S., Bastin, D. A. and Verma, N. K. (1999) Functional analysis of the O antigen glucosylation gene cluster of *Shigella flexneri* bacteriophage SfX. *Microbiology* **145**, 1263–1273.
- Mavris, M., Manning, P. A. and Morona, R. (1997) Mechanism of bacteriophage SfII-mediated serotype conversion in *Shigella flexneri*. *Mol. Microbiol.* **26**, 939–950.
- Schnaitman, C. A. and Klena, J. D. (1993) Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol. Rev.* **57**, 655–682.
- Whitfield, C. (1995) Biosynthesis of lipopolysaccharide O antigens. *Trends Microbiol.* **3**, 178–185.
- Alexeyev, M. F. and Winkler, H. H. (1999) Membrane topology of the *Rickettsia prowazekii* ATP/ADP translocase revealed by novel dual pho-lac reporters. *J. Mol. Biol.* **285**, 1503–1513.
- Manoil, C. and Beckwith, J. (1986) A genetic approach to analyzing membrane protein topology. *Science* **233**, 1403–1408.
- Welply, J. K., Fowler, A. V. and Zabin, I. (1981) β -Galactosidase α -complementation: overlapping sequences. *J. Biol. Chem.* **256**, 6804–6810.
- Lindberg, A. A., Karnell, A., Stocker, B. A., Katakura, S., Sweiha, H. and Reinholt, F. P. (1988) Development of an auxotrophic oral live *Shigella flexneri* vaccine. *Vaccine* **6**, 146–150.
- Schuch, R. and Maurelli, A. T. (1997) Virulence plasmid instability in *Shigella flexneri* 2a is induced by virulence gene expression. *Infect. Immun.* **65**, 3686–3692.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Dower, W. J., Miller, J. F. and Ragsdale, C. W. (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* **16**, 6127–6145.
- Korres, H. and Verma, N. (2004) Topological analysis of glucosyltransferase GtrV of *Shigella flexneri* by a dual reporter system and identification of a unique reentrant loop. *J. Biol. Chem.* **279**, 22469–22476.
- Sugiyama, J. E., Mahmoodian, S. and Jacobson, G. R. (1991) Membrane topology analysis of *Escherichia coli* mannitol permease by using a nested-deletion method to create mtlA-phoA fusions. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9603–9607.
- Henikoff, S. (1987) Unidirectional digestion with exonuclease III in DNA sequence analysis. *Methods Enzymol.* **155**, 156–165.
- Ehrmann, M., Boyd, D. and Beckwith, J. (1990) Genetic analysis of membrane protein topology by a sandwich gene fusion approach. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7574–7578.
- Manoil, C. (1991) Analysis of membrane protein topology using alkaline phosphatase and β -galactosidase gene fusions. *Methods Cell. Biol.* **34**, 61–75.
- Miller, J. H. (1992) *A Short Course in Bacterial Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Jerpseth, B., Greener, A., Short, J. M., Viola, J. and Kretz, P. L. (1992) *Strategies Newsletter*, vol. 5, pp. 81–83, Stratagene, La Jolla.
- Huan, P. T., Bastin, D. A., Whittle, B. L., Lindberg, A. A. and Verma, N. K. (1997) Molecular characterisation of the genes involved in O-antigen modification, attachment, integration and excision in *Shigella flexneri* bacteriophage SfV. *Gene* **195**, 217–227.
- Chen, J. H., Hsu, W. B., Chiou, C. S. and Chen, C. M. (2003) Conversion of *Shigella flexneri* serotype 2a to serotype Y in a shigellosis patient due to a single amino acid substitution in the protein product of the bacterial glucosyltransferase *gtrII* gene. *FEMS Microbiol. Lett.* **224**, 277–283.
- Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T. and MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **280**, 69–77.
- Grunewald, M., Bendahan, A. and Kanner, B. I. (1998) Biotinylation of single cysteine mutants of the glutamate transporter GLT-1 from rat brain reveals its unusual topology. *Neuron* **21**, 623–632.
- Harris-Warrick, R. M. (2000) Ion channels and receptors: molecular targets for behavioral evolution. *J. Comp. Physiol. A* **186**, 605–616.
- Unligil, U. M. and Rini, J. M. (2000) Glycosyltransferase structure and mechanism. *Curr. Opin. Struct. Biol.* **10**, 510–517.
- Contreras, I., Ortiz-Zapater, E. and Aniento, F. (2004) Sorting signals in the cytosolic tail of membrane proteins involved in the interaction with plant ARF1 and coatamer. *Plant J.* **38**, 685–698.
- Breton, C. and Imberty, A. (1999) Structure/function studies of glycosyltransferases. *Curr. Opin. Struct. Biol.* **9**, 563–571.

Received 14 January 2005/3 March 2005; accepted 14 March 2005

Published as BJ Immediate Publication 14 March 2005, DOI 10.1042/BJ20050102