A Novel Pathway for O-Polysaccharide Biosynthesis in Salmonella enterica Serovar Borreze*

Wendy J. Keenleyside and Chris Whitfield‡

From the Department of Microbiology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

The plasmid-encoded gene cluster for O:54 O-polysaccharide synthesis in Salmonella enterica serovar Borreze (rfbO:54) contains three genes that direct synthesis of a ManNAc homopolymer with alternating β1,3 and β1,4 linkages. In Escherichia coli K-12, RfbA0:54 adds the first ManNAc residue to the Rfe (UDP-GlcNAc:undecaprenylphosphate GlcNAc-1-phosphate transferase)-modified lipopolysaccharide core. Hydrophobic cluster analysis of RfbB0:54 indicates this protein belongs to the ExoU family of nonprocessive β-glycosyltransferases. Two putative catalytic residues and a potential substrate-binding motif were identified in RfbA0:54. Topological analysis of RfbB0:54 predicts four transmembrane domains and a large central cytoplasmic domain. The latter shares homology with a similar domain in the processive β-glycosyltransferases Cps3S of Streplococcus pneumoniae and HasA of Streptococcus pyogenes. Hydrophobic cluster analysis of RfbB0:54 and Cps3S indicates both possess the structural features characteristic of the HasA family of processive β-glycosyltransferases. Four potential catalytic residues and a putative substrate-binding motif were identified in RfbB0:54. In Δrfb E. coli K-12, RfbA0:54 and RfbB0:54 direct synthesis of smooth O:54 lipopolysaccharide, indicating that this O-polysaccharide involves a novel pathway for O-antigen transport. Based on sequence and structural conservation, 15 new ExoU-related and 17 new HasA-related transfers were identified.

Lipopolysaccharide (LPS) is a major component of the outer membrane of Gram-negative bacteria. The hydrophobic lipid A portion forms the outer leaflet of the outer membrane. In enteric bacteria, an O-polysaccharide is attached to lipid A via a short core oligosaccharide. The O-polysaccharide extends away from the cell surface to give a hydrophilic surface layer and represents a major surface antigen (O-antigen). It is characterized by a repeat unit structure, with different epitopes arising from variations in the nature, order, and linkage of the sugar monomers in the O-repeat unit and from the addition of side branches and “decorations” such as acetyl, ketal, and glycosyl residues.

Members of the genus Salmonella are serologically diverse. Approximately 50 O-antigens are recognized in Salmonella, and this diversity, combined with variable flagellar antigens, has resulted in the identification of greater than 2,000 serologically distinct Salmonella strains. Strains are assigned to serogroups based on shared expression of major O-antigen epitopes or factors (1). Synthesis of the O-antigen is directed by the products of the rfb gene cluster, and the structural diversity seen in the LPS O-antigens reflects variation in the rfb gene clusters; this variation is thought to have arisen from repeated lateral gene transfer and recombination events involving rfb genes. Selective pressure from the host immune response is proposed to be the driving force for the continued generation of antigenic diversity (2).

Serogroup O:54 is a heterogeneous group of 13 different Salmonella serovars. It is unique among Salmonella serogroups in that expression of the O:54 antigen requires the presence of a small plasmid. This is the only known plasmid-encoded O-polysaccharide in Salmonella. We have studied the 6.9-kb plasmid pWQ799 from serovar Borreze to determine the role of these plasmids in expression of the O:54 antigen (3). pWQ799 is a novel ColE1-related plasmid carrying the entire rfbO:54 biosynthetic cluster (4). This is the only reported example of a ColE1-related plasmid carrying genes for the synthesis of cell-surface antigens. The plasmid is mobilized in the presence of an appropriate helper plasmid, providing the first defined mechanism for lateral gene transfer of O-antigen biosynthesis genes in Salmonella enterica. Mobilization of plasmids containing rfbO:54 into strains possessing a chromosomally encoded rfb explains the simultaneous expression of two distinct O-antigens in members of this serogroup (3) as well as the serotyping results of plasmid-cured O:54 derivatives.

Enzymes encoded by the rfbO:54 cluster direct the synthesis of an N-acetylmannosamine (ManNAc)-containing O-polysaccharide with the novel disaccharide repeat unit, −(4)-β-D-ManpNAc-(1→3)-β-D-ManpNAc (1→). O:54 polysaccharide synthesis requires a functional chromosomal rfe gene (3). Rfe transfers GlcNAc-1-phosphate to undecaprenol-phosphate to form the first lipid intermediate (lipid 1) in synthesis of the enterobacterial common antigen (5). Two LPS O-polysaccharide pathways have been described, and Rfe may play a role in either of these two pathways (reviewed in Ref. 6). The O:54 O-polysaccharide shares a number of characteristics with the O-polysaccharides synthesized by the Rfc-independent O-polysaccharide biosynthetic pathway including the following: 1) it is a homopolymer; and 2) Rfe is required for synthesis although GlcNAc does not form part of the repeating unit. The Rfc-independent pathway has so far only been identified for a few homopolymeric O-polysaccharides in Klebsiella O1 (7, 8), Escherichia coli O8 and O9 (9–11), and Serratia marsecens O16 (12). In this pathway, Rfe initiates synthesis by transferring GlcNAc-1-P to the carrier lipid to form the acceptor upon which the complete O-polysaccharide chain is subsequently

* This research was supported by funding (to C. W.) from the Natural Sciences and Engineering Research Council of Canada. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Tel.: 519-824-4120; Fax: 519-837-1802; E-mail: cwhitfie@micro.uoguelph.ca.
polymerized. GlnpNAc therefore does not form part of the repeating unit. O-polysaccharides synthesized by this pathway are polymerized in the cytoplasm (13), and then the completed polymer is transported across the plasma membrane by an rfb-encoded dedicated ABC (ATP-binding cassette) transporter, prior to ligation to lipid A core. Analysis of rfbO54 and O:54 expression indicates that transport of the nascent O:54 polysaccharide occurs by a different mechanism. This communication describes the organization of rfbO54 and the characterization of the two glycosyltransferases that direct the synthesis of the O:54 O-polysaccharide. Our studies indicate that synthesis of O:54 LPS follows a novel mechanism distinct from either of the two characterized O-polysaccharide pathways.

**EXPERIMENTAL PROCEDURES**

**Bacterial Strains, Growth Conditions, and Plasmids—**Bacterial strains were grown in Luria-Bertani broth supplemented, where appropriate, with antibiotics (ampicillin, 100 μg/ml; kanamycin, 50 μg/ml; chloramphenicol, 30 μg/ml; tetracycline, 15 μg/ml). For high level expression of rfbA/B, rfbC, rfbE/F, rfbG, rfbH, rfbK, and rfbO54, the transformed bacteria were grown in LB broth to mid-log phase culture of E. coli DH5α (pWQ920) using TRIZol Reagent (Life Technologies, Inc.) according to the manufacturer’s instructions. Primer extension reactions were done using the oligonucleotide primer TS154 (5′-TTTCTAAATGTGGATCTGTTAATCC-3′) that corresponds to the complementary sequence to nucleotides 3678–3700 of plasmid pWQ799. The primer was end-labeled with [γ-32P]ATP (DuPont NEN) and T4 polynucleotide kinase (Boehringer Mannheim, Laval, Quebec, Canada) and then purified using a QiAquick spin column (QIAGEN Inc., Chatsworth, CA). Primer extension experiments were performed using the First-Strand cDNA Synthesis Kit from Pharmacia (Pharma- cia Biotech, Baie D’Urfe, Quebec) following the manufacturer’s recommendations. DNA sequencing was done with the Sequenase version 2.0 sequencing kit (U. S. Biochemical Corp.).

**LPS Extraction and Purification of Core Oligosaccharides—**LPS samples were prepared either from SDS-protease K whole cell lysates, as described by Hitchcock and Brown (23), or by using a modification of the phenol/water extraction method; LPS was collected from both the aqueous and phenol phases (3). For compositional analysis of the LPS core oligosaccharides, lipid A was removed from phenol-puriﬁed LPS by hydrolysis in 1.5% acetic acid at 100 °C for 2 h. Precipitated lipid A was removed by centrifugation, and the supernatant was lyophilized. Samples were resuspended in water to a final concentration of 5 mg/ml, hydrolyzed, and analyzed by high performance anion-exchange chromatography as described previously (24).

**RESULTS**

**Organization of rfbO54—**Previous analysis of the nucleotide sequence and genetic organization of pWQ799 indicated that approximately half of the 6915-base pair plasmid (nucleotides 72–3384) is involved in plasmid replication and mobilization (4). These regions are related to ColE1 and possess an average G + C content of 50–53% (Fig. 1A). This value is typical of Salmonella genomic DNA (26). In contrast, the remaining pWQ799 sequences have a uniformly lower average G + C content of 39%, with no detectable homology to any known ColE1-related sequences. The junctions between the high and low G + C regions in pWQ799 coincide with the 5′ and 3′ ends of the plasmid replicon regions. Abnormally low G + C values relative to those typical for the species is a common observation for genes involved in polysaccharide synthesis (27). These observations therefore suggested that the remaining DNA contained rfbO54 determinants.

Computer analysis for coding regions combined with sequence homology searches in the NCBI database (see below) resulted in the identiﬁcation of three potential open reading frames (Fig. 1B; nucleotides 3547–4364, 4364–5740, and 5740–6873) designated rfbA_{O54}, rfbB_{O54}, and rfbC_{O54}. Each gene is preceded by potential Shine-Dalgarno sequence: AGGA at nucleotides 3532–3535; AAGA at 4356–4359, and AAGA at 5727–5730. The initiation codons of rfbB_{O54} and rfbC_{O54} each overlap with the end of the preceding gene, suggesting that translation of the three genes is coupled. No consensus –10 and –35 E. coli promoter sequences were detected in the region immediately upstream of rfbB_{O54}. To localize the rfbB_{O54} promoter, the 1.14-kb HinII–SspI fragment of pWQ799, containing rfbB_{O54} and a 90-base pair 5′-flanking region, was cloned in front of the promoterless cat gene of pKK232–8. When transformed into E. coli K–12, the resulting plasmid (pWQ822; Fig. 1C) conferred chloramphenicol resistance, indicating the presence of an endogenous promoter. The transcription start site was mapped by primer extended 5′-end extension of the oligonucleotide TS154 using total cellular RNA from E. coli DH5α (pWQ820; Fig. 1C). A single band was obtained in the resulting autoradiogram (data not shown), identifying the start site as nucleo-
Previous studies of O:54 O-polysaccharide expression in strains containing pWQ800 (Fig. 1C) revealed a requirement for a functional rfbE determinant in an E. coli host (3). This gene is part of the chromosomal enterobacterial common antigen (ECA) biosynthesis cluster and encodes the enzyme UDP-GlcNAc-2-epimerase (EC 5.1.3.14, UDP-N-acetyl-D-glucosamine 2-epimerase) (5). The epimerase converts UDP-GlcNAc to UDP-ManNAc, an intermediate in the formation of a precursor for ECA biosynthesis. UDP-ManNAc is also the presumed precursor for O:54 synthesis. Sequence analysis of pWQ799 indicated that the EcoRI site used in cloning of pWQ800 maps within rfbC O:54. Subsequent mutant complementation analysis with an rfbC O:54 subclone has shown that RfbC O:54 is a functional homologue of RffE, and is therefore not essential for O:54 synthesis in members of the Enterobacteriaceae, all of which synthesize ECA (28). The remaining two rfbO:54 gene products, RfbA O:54 and RfbB O:54, are unique and essential for O:54 polysaccharide assembly and their structures and functions are described below.

Structural Analysis of RfbA O:54—Sequence analysis of rfbA O:54 predicts a 274 amino acid protein with a molecular mass of 31.8 kDa and a pl of 9.76. Hydropathic analyses (18) identified one transmembrane helix at the carboxyl terminus of the protein (residues 227–251). To determine the topology of RfbA O:54, a promoterless leader sequence-deficient TnphoA gene was inserted in-frame at a position corresponding to amino acid 21 of RfbA O:54 (Fig. 1B). E. coli CC118 (pWQ800B6) was O:54-deficient and PhoA-negative on indicator media, indicating that this region of the protein is present in the cytoplasm. This cytoplasmic location is in agreement with the function of RfbA O:54 (see below) and with protein topology predictions by the positive-inside rule. This rule allows the prediction of the topology of a bacterial inner membrane protein based on the observation that positively charged amino acids (Arg, Lys) are more abundant in cytoplasmic loops (i.e. 15%) than in periplasmic loops (5%) (29). The Arg + Lys content of this region was 14%.

The translated amino acid sequence of RfbA O:54 shares significant homology with a number of putative bacterial glycosyltransferases (Fig. 2 and Table I). RfbA O:54 shares 32% identity with the predicted product of the rfbEcO7 gene orf275 (30), 26% identity with the hypothetical protein of Haemophilus influenzae, and 25% identity with the AmsE protein of Erwinia amylovora (31). Analysis of the protein alignments of these predicted products revealed that the sequence conservation was relatively uniform throughout the length of the proteins (Fig. 2). A number of protein sequences in the data bases were also identified with significant levels of homology over the N-terminal 192 amino acids of RfbA O:54. Alignment of this region gave identity levels of 21 and 20%, respectively, with the ExoU and ExoO proteins from Rhizobium meliloti (32–34) and 21% with the LgtA glycosyltransferase from Neis-
AmsB Required for synthesis of amylovoran
Orf275 Involved in staphyloxanthin synthesis
Bacillus subtilis

GgaB Involved in secondary teichoic acid synthesis

LgtD Adds GalNAc β1→3 to Gal; involved in LOS synthesis; homologue of LgtA

LgtE Involved in LOS synthesis

RfbAO:54 Adds first ManNAc residue during synthesis of the O:54 O-polsaccharide

RfbA, ExoU, and ExoO includes the region containing domain A in the C-terminal end of the proteins, HCA was used to compare RfbAO:54 with ExoU and the other RfbA O:54 homologous proteins (Fig. 3).

S. enterica sv. Borreze

H. influenzae

Shigella dysenteriae

Yersinia enterocolitica O3

Bacillus subtilis 168

Staphylococcus aureus Newman

Streptococcus thermophilus Sf6

R. meliloti

R. meliloti

R. meliloti

R. meliloti

S. cerevisiae

Anabaena sp. strain PCC 7120

seria gonorrhoeae (35) (Table I).

Previous studies of glycosyltransferases have demonstrated that there is often insufficient sequence similarity for functional predictions using traditional sequence alignments (36). However, transferases which catalyze the formation of glycosidic linkages with the same stereochemistry and with structurally related substrates are predicted to share a similar three-dimensional architecture in the catalytic and binding domains. This would be reflected in the presence of conserved structural regions or sequence motifs for shared mechanistic functions. Such domains can be identified using HCA (36, 37).

This method plots the two-dimensional pattern of protein sequences and allows visual comparison and detection of conserved structural regions or sequence motifs for shared mechanistic domains. Such domains can be identified using HCA (36, 37).

Structural region, domain A, that was present in all five proteins. The transferases all possess a common catalytic activity: formation of a single glycosidic linkage with a β-configuration from α-linked nucleotide diphospho sugar donors. Domain A is therefore speculated to be directly involved in this shared activity. Since the region of similarity detected between RfbA O:54, ExoU, and ExoO includes the region containing domain A in the Rhizobium proteins, HCA was used to compare RfbA O:54 with ExoU and the other RfbA O:54 homologous proteins (Fig. 3).

All of the proteins examined possessed the domain identified by Saxena et al. (36). This domain is characterized by a series of vertical hydrophobic clusters typical of β-strands alternating with clusters characteristic of α-helices (36). Two conserved motifs were identified within the regions analyzed: (DXDD), at the C-terminal end of the β4 region of domain A, and EDY, which lies an average of 94 amino acids downstream of the (DXDD) motif, within another region of structural conservation.

### Table I: Members of the ExoU family of nonprocessive β-glycosyltransferases

| Protein | Putative or known activity | Organism | Accession no. | Ref. |
|---------|----------------------------|----------|---------------|-----|
| RfbA O:54 | Adds first ManNAc residue during synthesis of the O:54 O-polsaccharide | S. enterica sv. Borreze | L39794 | This study |
| Orf275 | Required for synthesis of the O-7 O-polsaccharide | E. coli 07 | L04596 | 30 |
| AmsE | Required for synthesis of amylovoran | Erwinia amylovora | X7921 | 31 |
| AmsB | Required for synthesis of amylovoran | E. amylovora | X7921 | 31 |
| LgtA | Involved in LOS synthesis | Haemophilus influenzae | M94855 | Unpublished |
| LgtD | Adds GlcNAc β1→3 to Gal; Involved in LOS synthesis | Neisseria gonorrhoeae | U15992 | 35 |
| Lai-2 | Adds GlcNAc β1→3 to Gal; Involved in LOS synthesis; homologue of LgtA | S. enterica | U15992 | 74 |
| LgtD | Putative glycosyltransferase involved in LOS synthesis | H. influenzae | L46209, L42023 | 75 |
| LgtD | Putative glycosyltransferase involved in LOS synthesis | H. influenzae | L45506, L42023 | 75 |
| RfpA | Involved in synthesis of type 1 O-polsaccharide | Shigella dysenteriae | S7325 | 76 |
| TrsC | Putative glycosyltransferase involved in outer core LPS synthesis | Yersinia enterocolitica O3 | Z47767, X8327 | 77 |
| GgbB | Involved in secondary teichoic acid synthesis | Bacillus subtilis 168 | U13979 | Unpublished |
| Orf4 | Involved in staphyloxanthin synthesis | Staphylococcus aureus Newman | J340131 | Unpublished |
| EpsE | Putative glycosyltransferase | Streptococcus thermophilus Sf6 | U04830 | 78 |
| ExoU* | Adds Glc β1→6 to Glc; involved in succinoglycan synthesis | R. meliloti | L20758 | 33,34 |
| ExoO* | Adds Glc β1→6 to Glc; involved in succinoglycan synthesis | R. meliloti | L20758 | 33,34 |
| ExoA* | Adds Glc β1→6 to Glc; involved in succinoglycan synthesis | R. meliloti | L20758 | 33,34 |
| ExoM* | Adds Glc β1→6 to acetylated Glc; involved in succinoglycan synthesis | R. meliloti | L20758 | 33,34 |
| ExoW* | Adds Glc β1→3 to Gal; involved in succinoglycan synthesis | R. meliloti | L20758 | 33,34 |
| Dpm1* | Dolichol mannosyl phosphate synthetase | S. cerevisiae | J04184 | 54 |
| Orf2* | Unknown | Anabaena sp. strain PCC 7120 | M31722 | 79 |

* Original member of the ExoU family, as defined by Saxena et al. (36).
first motif includes a conserved Asp residue previously identified by Saxena’s group. This residue falls in a loop following a β-sheet, and based on this conserved location and the acidic nature of the residue, Saxena et al. have speculated that this may be a catalytic residue. As noted by Saxena et al. (36), the proteins were also found to contain at least one Asp residue in the loop at the C-terminal end of the β2 strand (Fig. 3). The other protein sequences obtained from the data base search with RfbA0.54 were also examined for the presence of these conserved features (Fig. 4). A total of 15 proteins, in addition to the 7 originally described by Saxena et al. (36), were identified as being members of the ExoU family (Table I).

**RfbA0.54 Is a Monofunctional N-Acetylmannosaminyltransferase**—The observed sequence and structural homology of Rf-
bA o:54 with other glycosyltransferases suggested that rfbA o:54 encodes a ManNAc transferase (EC 2.4.1.X). However, two separate ManNAc transferase activities are required for the generation of a ManNAc homopolymer with alternating β1→3 and β1→4 linkages. In other homopolysaccharide O-antigens where polymerization is initiated by Rfe, the transferase which adds the first residue to the Rfe intermediate is distinct from the enzymes which subsequently polymerize the O-polysaccharide (8, 9). A third transferase activity is therefore predicted for the transfer of the first ManNAc residue to undecaprenol-P-P-GlcNAc, the product of the Rfe reaction (5). To determine which rfb gene product was responsible for the initial ManNAc transfer, plasmids carrying either rfbA o:54 (pWQ823; Fig. 1C) or rfbB o:54 (pWQ819; Fig. 1C) were transformed into different E. coli K-12 backgrounds, and LPS in the whole cell lysates of the transformants was analyzed by SDS-PAGE (Fig. 5). In E. coli DH5a (rfe-1rfE), neither gene was sufficient for O:54 synthesis and expression of a ladder of O-antigen-substituted LPS. Introduction of the two genes together on plasmid pWQ802, either in strain DH5a or the rfb-deletion strain S6874, was sufficient for expression of authentic O:54 O-polysaccharide. These results were confirmed by Western immunoblot using absorbed polyclonal O:54 antisera (data not shown).

Analysis of the LPS core regions in SDS-PAGE profiles revealed that in the presence of rfbA o:54, an additional LPS band was synthesized; this band migrated slightly slower than the lipid A core fraction of the host strain. In contrast, LPS from the rfbB o:54-containing strain was indistinguishable from that of the parental strain. Further analysis determined that the RfbA o:54-mediated band was not synthesized in an rfe- host strain, as would be expected for a biosynthetic pathway initiated by Rfe (Fig. 6A). The structure of the E. coli K-12 core oligosaccharide has been determined (38, 39) and does not include any ManNAc residues. The demonstration of ManNAc in the RfbA o:54-modified core would therefore be indicative of a ManNAc transferase function catalyzed by RfbA o:54. Purified core oligosaccharide from E. coli DH5a (pWQ823) was therefore analyzed by high performance anion-exchange chromatography. A single additional peak corresponding to mannosamine, the acid-hydrolyzed product of ManNAc, was detected in the chromatogram of the RfbA o:54-dependent core oligosaccharide (data not shown).

To determine the size of the RfbA o:54-dependent band, the LPS from E. coli DH5a (pWQ823) was analyzed by SDS-PAGE alongside two LPS samples which each contain a modified core fraction of known size and composition (Fig. 6B). Plasmid pJK2363 contains the Shigella dysenteriae galactopyranosyltransferase gene, rfpB (40). Rfe" K-12 host strains expressing rfpB synthesize a core oligosaccharide modified by the addition of 1,3-linked Gal and 1,6-linked GlcNAc residues.
of the disaccharide \( \text{a-D-Galp-(1-3)-\beta-D-GlcpNAC-(1-\rightarrow) } \) (41). Plasmid pWQ20 contains the \( \text{rfbF}_{\text{KpO1}} \) gene from \textit{Klebsiella pneumoniae} O1, and \textit{E. coli} K-12 strains containing \( \text{rfbF}_{\text{KpO1}} \) produce a core oligosaccharide modified by the addition of the trisaccharide \( \beta\text{-D-Galf}-(1-3)\text{-a-D-Galp-(1-3)-\beta-D-GlcpNAC-(1-\rightarrow) } \) (8). The comigration of the \( \text{RfbA}_{\text{O:54}} \)-modified band with the “core + 2” band of \textit{E. coli} K-12 (pJK2363), combined with the \( \text{Rfe} \) dependence of this band and chemical compositional analysis all indicate that the \( \text{RfbA}_{\text{O:54}} \)-modified band represents lipid A modified by the addition of a single GlcNAc and a single ManNAc residue.

The \( \text{RfbA}_{\text{O:54}} \)-modified lipid A core fraction was not present in \textit{E. coli} \textit{Sb874} (pWQ823), indicating that synthesis of this LPS fraction requires one or more function(s) encoded by the K-12 \( \text{rfb} \) gene cluster (Fig. 6A). Expression of \( \text{O:54} \) LPS in \textit{Sb874} (pWQ802) indicates that this requirement is overcome by the simultaneous activities of \( \text{rfbA}_{\text{O:54}} \) and \( \text{rfbB}_{\text{O:54}} \). Ligation of newly polymerized O-chains to lipid A core occurs at the periplasmic face of the plasma membrane and must therefore follow trans-plasma membrane transport of the undecaprenol-bound intermediate. A similar \( \text{rfbK}_{\text{K-12}} \)-dependent transport activity has been demonstrated for the \( \text{RfbF}_{\text{KpO1}} \) and \( \text{RfpB} \)-mediated core modifications (8, 42). The \textit{E. coli} K-12 O-antigen transporter, \( \text{RfhX} \), is believed to be responsible for the transport of the lipid-bound \( \text{RfbF}_{\text{KpO1}} \) and \( \text{RfpB} \) products (8, 42, 43). Given the nature of this transport event and the involvement of both host and cloned plasmid functions, it was possible that the appearance of a single modified core band in \textit{E. coli} DH5\( \alpha \) (pWQ823) resulted from a substrate size limitation imposed by \( \text{RfhX} \), rather than the monofunctional transferase activity of \( \text{RfbA}_{\text{O:54}} \). In this case, lipid-linked intermediates with higher degrees of polymerization would be synthesized but would remain in the cytoplasm, attached to carrier lipid. To address this possibility, the O-poly saccharide ligase-deficient strain \textit{E. coli} CS2334 was transformed with pWQ823. Under such conditions any O-poly saccharide that is formed, but not transported, should accumulate in the cytoplasm. \textit{E. coli} CS2334 (pWQ823) was phenol-extracted, and because the polymeric \( \text{O:54} \) LPS partitions primarily into the organic phase (3), both phases were examined. The linkage between undecaprenol and carbohydrate polymer is phenol-labile (43) and, as a consequence, extracted O-haptenic material remains in the supernatant following a 100,000 \( \times \) g centrifugation step. Supernatants from both the aqueous and phenol phases were size-fractionated on a Sephadex G-50 column, and fractions containing amino sugars were analyzed by \( \text{H} \) NMR. Although extracted high molecular weight ECA was detected in these experiments, no ManNAc-containing polymer was present (data not shown). These data confirm that \( \text{RfbA}_{\text{O:54}} \) transfers a single ManNAc residue.

Sequence Analysis of \( \text{RfbB}_{\text{O:54}} \)—\( \text{RfbB}_{\text{O:54}} \) is predicted to be a 53.3-kDa protein composed of 459 amino acids, with a calculated pl of 8.32. Hydropathic analysis predicted four transmembrane helices between residues 11–40, 325–340, 385–406, and 416–438. This protein is therefore expected to be an integral membrane protein. Based on the relative distribution of positive amino acid residues, the hydrophilic region defined by residues 340 and 384 is predicted to lie in the periplasm (6% Arg + Lys). This location was confirmed by construction of an in-frame PhoA-positive fusion (pWQ800B22) at amino acid 368 (Fig. 1B).

The demonstrated activity of \( \text{RfbA}_{\text{O:54}} \) combined with the observation that \( \text{RfbA}_{\text{O:54}} \) and \( \text{RfbB}_{\text{O:54}} \) together are sufficient for synthesis of polymeric \( \text{O:54} \) LPS suggests that \( \text{RfbB}_{\text{O:54}} \) is a processive \( \beta \)-glycosyltransferase (EC 2.4.1.X). Searches of the data bases with the \( \text{RfbB}_{\text{O:54}} \) sequence identified only one protein, IcaA (44), with significant sequence homology over its entire length (22%). IcaA is a putative glycosyltransferase from the \( \text{ica} \) (intercellular adhesion) locus of \textit{Staphylococcus epidermidis}. This protein is believed to play a role in the synthesis of the linear \( \beta -1,6 \)-linked glucosaminoglycan involved in intercellular adhesion. The predicted size and topology of this protein resembles that of \( \text{RfbB}_{\text{O:54}} \). A number of other proteins were identified that possess regions of conservation with the large central hydrophilic domain of \( \text{RfbB}_{\text{O:54}} \) (Table II). Most of these proteins are glycosyltransferases involved in the synthesis of bacterial cell surface polysaccharides. Two of these proteins, Cps3S and HasA, are also predicted to be integral membrane proteins with similar hydropathy plots and four predicted transmembrane domains. Cps3S is a glycosyltransferase that directs the synthesis of the type 3 capsule of \textit{Streptococcus pneumoniae} (45, 46). The enzyme is bifunctional and processive, catalyzing the formation of \( \text{[-(3)-\beta-D-GlcA-(1-\rightarrow)-\beta-D-Glc-(1-\rightarrow)}_n \) (45). HasA is the hyaluronic acid synthase from Group A \textit{Streptococcus pyogenes} (47, 48). This protein is also a bifunctional processive \( \beta \)-glycosyltransferase, catalyzing the formation of a polysaccharide with the structure \( \text{[-(4)-\beta-D-GlcA-(1-\rightarrow)-\beta-D-Glc-(1-\rightarrow)}_n \) (47).

Using HCA, Saxena \textit{et al.} (36) compared the plots of HasA with a number of known processive \( \beta \)-glycosyltransferases. They reported a correlation between the presence of two conserved structural regions (domains A and B) and the shared catalytic activity. Domain A is common to both the ExoU and HasA families of \( \beta \)-glycosyltransferases; whereas domain B, located a short distance downstream of domain A, is unique to the HasA family. A single conserved Asp residue in region II and a conserved sequence motif (QXXRW) in region IV were both reported to characterize domain B. Comparison of the HCA plots of the large hydrophilic domains of \( \text{RfbB}_{\text{O:54}} \), Cps3S, and IcaA, with that of HasA, confirmed the presence of both domains in these proteins (Fig. 7). All three proteins possessed the conserved Asp and (QXXRW) motif of domain B and with the exception of Cps3S, the two conserved Asp residues characteristic of domain A were also present in \( \text{RfbB}_{\text{O:54}} \) and IcaA (Fig. 7). Although the Asp-\( \beta \)4 is conserved in Cps3S, there is no
Asp immediately next to the β2 sheet of the protein; however, an Asp residue is located in the middle of β2 (Fig. 8). A search of the data bases identified a total of 17 other proteins, in addition to the 6 originally identified by Saxena et al., that possess the conserved features of the HasA family (Table II).

**DISCUSSION**

Plasmid pWQ799 carries the entire O:54 biosynthetic cluster and consists of three genes, rfbA<sub>0.54</sub>, rfbB<sub>0.54</sub>, and rfbC<sub>0.54</sub>. Together these genes direct the synthesis of a unique *Salmonella* O-polysaccharide. The O:54 polysaccharide is the only known homoplymeric O-antigen in *Salmonella* (49), and while the presence of N-acetylated amino sugars is not an unusual feature for *Salmonella* O-polysaccharides, ManNAC is found only in O:54. It is therefore not surprising that the predicted products of rfb<sub>0.54</sub> showed no homology to other *Salmonella* protein sequences in the data bases.

RfbA<sub>0.54</sub> is an N-acetylmannosaminyltransferase. This enzyme transfers the first ManNAC residue to undecaprenylphosphoryl-GlcNAc, the product of the Rfe reaction. The involvement of Rfe (UDP-GlcNAc:undecaprenylphosphate GlcNAc-1-phosphate transferase) in the initiation of O:54 biosynthesis resembles the situation observed with other homopolysaccharide O-antigens (7, 10, 12, 50), where biosynthesis is initiated by Rfe but GlcNAc does not form part of the O-unit. It is unclear if this type of initiation reaction is necessarily limited to O-polysaccharides. RfbA<sub>0.54</sub> shares significant sequence homology with a number of glycosyltransferases in the data bases. Two of the proteins, ExoO and ExoU from *R. meliloti*, are β-glycosyltransferases that catalyze β-1,6 linkages from a-1-linked nucleotide diphospho sugar donors. Alignment of the HCA plots derived for RfbA<sub>0.54</sub> and 4 RfbA<sub>0.54</sub>-related proteins with the previously characterized plot of ExoU (36) revealed that all of the proteins possess the same N-terminal structural region (domain A) identified by Saxena and co-workers (36) in ExoU and a number of functionally related transferases. This domain is believed to be responsible for the nonprocessive β-glycosyltransferase activity of proteins within this family. Given their common activities and the conservation of domain A among the members of the ExoU family of glycosyltransferases, it is likely that this region represents at least part of the catalytic domain of these enzymes. We identified two highly conserved motifs among the aligned HCA plots. The first (DX-DD), falls in the β4 region of domain A and includes a conserved Asp residue previously identified by Saxena and co-workers (36). Saxena et al. (36) also identified a second conserved Asp residue in domain A, in the loop at the C-terminal end of the β2 strand of domain A. HCA alignments revealed that all of the proteins possess either one (RfbA<sub>0.54</sub>, rfb725, Lsg6, and AmsE) or two Asp residues (ExoU) in this region, although the position is not strictly conserved and an Asp may occupy the first and/or the second site within this loop.

Very little is known about anabolic glycosyl transfer reactions, but it has been proposed that, mechanistically, this type of reaction may be viewed as the reverse reaction of the glycosyl transfer reaction performed by O-glycosidases, the difference being that the result is the extension rather than hydrolysis of an oligosaccharide or polysaccharide chain. By analogy with the extensively characterized polysaccharide hydrolyase systems (reviewed in Refs. 51, 52), this hypothesis predicts that formation of a β-glycosyl linkage from an α-linked sugar nucleotide donor would involve the same type of catalytic event as that of the inverting glycoside hydrolase. Hydrolysis of glycosidic bonds by inverting glycosidases results in a net inversion of configuration at the anomeric center of the reducing sugar product. The catalytic mechanism involves two acidic active site amino acids that act as acid-base catalysts. Among the cellulases and xylanases, in every instance where the catalytic residues have been identified, the amino acid has been either an aspartate or a glutamate (52). One of these residues is believed to act as the acid catalyst to protonate the substrate, whereas the other is thought to act as the base catalyst by deprotonating water. The two catalytic residues are located in flexible loop regions in the active site cleft, between substrate-binding subsites. Characterization of RfbA<sub>0.54</sub>-related proteins in data bases has identified 15 new members of the ExoU family, in addition to the 7 originally described by Saxena et al.
These new members all possess two conserved Asp residues within the putative catalytic domain (Fig. 4). The second, in the terminal loop of β4, falls in a more strictly conserved region of the proteins and is surrounded by additional acidic residues. In RfbA₀:₅₄, sequence and HCA plot alignments predict that the catalytic residues are Asp-41 and either Asp-93, Asp-95, or Asp-96.

HCA also identified a conserved motif (EDY) downstream of domain A, in a region containing amino acid clusters with a shape typical of α-helices (21). Carbohydrate ligands are believed to interact with binding subsites through a combination of ionic interactions, hydrogen bonding, and stacking interactions between aromatic amino acids and the hydrophobic patches of carbohydrate monomers (reviewed in Ref. 52). The conservation of the (EDY) motif and its position within a region of structural conservation, a short distance from a putative catalytic residue, suggests this motif might represent part of a binding subsite. Multiple sequence alignment revealed that a similar motif occurs in other members of this family (Fig. 4) although the order of the first two acidic amino acids may be reversed and in some cases the third amino acid may be a Lys. Although Tyr is an aromatic amino acid, both it and Lys are structurally similar, in that both have long chains and may participate in hydrogen bonding through either a terminal hydroxyl (Tyr) or amino (Lys) group. The acidic nature of Asp and Glu would also allow hydrogen bonding of these residues with a carbohydrate ligand. Given the predominance of aromatic amino acids surrounding the (EDY) motif, it is possible that the surrounding residues participate in stacking interactions with the hydrophobic patch of the glycosyl monomer. The only protein that did not possess the (EDY) motif was Dpm1. This enzyme catalyzes the transfer of mannose from GDP-mannose to dolichol phosphate (54). The fact that the ligand recognized by Dpm1 is not a carbohydrate is further support for the hypothesis that this motif is involved in substrate binding.

RfbB₀:₅₄ is predicted to be an integral membrane protein with four transmembrane helices, a large central hydrophilic domain and a periplasmic loop. Searches of the data bases...
identified a number of proteins possessing regions of conservatism with the central hydrophilic domain. Among these, the proteins with known activities were all glycosyltransferases involved in synthesis of cell surface or secreted polysaccharides. The demonstrated activity of RfbBO:54 combined with the fact that rfbAB are sufficient for polymeric O:54 LPS in an E. coli Δrfb host suggest that RfbBO:54 is a processive ManNAc transferase that acts after RfbAO:54 to polymerize the O:54 polysaccharide. Analysis of the sequence alignments obtained with RfbBO:54 identified regions of homology with two known processive β-glycosyltransferases, Cps3S from S. pneumoniae (45, 46) and HasA from Group A S. pyogenes (47, 48). Both proteins have four predicted transmembrane domains with a large central hydrophilic region, and HCA of these hydrophilic proteins have four predicted transmembrane domains with a central hydrophilic domain. Among these, the identified an number of proteins possessing regions of conservatism with the central hydrophilic domain. Among these, the proteins with known activities were all glycosyltransferases involved in synthesis of cell surface or secreted polysaccharides. The demonstrated activity of RfbBO:54 combined with the fact that rfbAB are sufficient for polymeric O:54 LPS in an E. coli Δrfb host suggest that RfbBO:54 is a processive ManNAc transferase that acts after RfbAO:54 to polymerize the O:54 polysaccharide. Analysis of the sequence alignments obtained with RfbBO:54 identified regions of homology with two known processive β-glycosyltransferases, Cps3S from S. pneumoniae (45, 46) and HasA from Group A S. pyogenes (47, 48). Both proteins have four predicted transmembrane domains with a large central hydrophilic region, and HCA of these hydrophilic domains revealed that RfbBO:54 and Cps3S both possess the same multidomain structure and conserved residues previously identified in the HasA family of processive β-glycosyltransferases recently described by Saxena et al. (53). There are two conserved structural regions in proteins within this family. The first region (domain A) is common to both the ExoU and HasA families, and the second (domain B) is specific to the HasA family. Domain B is subdivided into four regions and has a conserved Asp residue at the C-terminal end of region II and a conserved Asp residue at the C-terminal end of region III. A search of the data bases for proteins possessing the conserved features of this family of proteins identified 17 new members (Table II), including RfbBo:54, Cps3S, and IcaA. Of the proteins that were identified, most are known processive enzymes, including Cps3S; five are NodC proteins involved in the polymerization of the β,1-4-linked Nod factors of Rhizobia; eight are chitin synthases, involved in the polymerization of β,1-4-linked GlcNAc polysaccharide chains of yeast chitin; and one is a cellulose synthase, involved in the polymerization of the β,1-4-linked Glc chains of cellulose. The identification of RfbBo:54 as a member of this family provides further evidence for the processive nature of this enzyme; however, it must be acknowledged that the data do not preclude the possible involvement of RfbBo:54 in the elongation process.

It is interesting to note that synthesis of the E. coli K5 capsule also involves a processive glycosyltransferase, Kfc (55). In this polysaccharide, the repeating unit consists of alternating β- and α-linkages, [→4)-β-D-GalpA-(1→4)-α-D-Glc-NAc-(1→56). Kfc must therefore catalyze both α- and β-linkages from α-linked donor molecules. Sequence alignment between RfbBo:54 and Kfc revealed that the K5 transferase possesses domain A but not domain B (data not shown), providing further support for the role of each domain in the formation of β-glycosidic bonds.

Saxena et al. (36) have proposed a model for the processive mechanism of polymerization in the HasA family of proteins. The model accounts for the characteristic multidomain architecture of the HasA family of proteins and, as with the monofunctional β-glycosyltransferases, is based on the hypothesis that inverting anabolic glycosyltransferases use the same catalytic mechanism as inverting glycoside hydrolases (51). According to this model, domains A and B represent different conserved structural domains that together allow these proteins to catalyze two β-glycosidic bonds, either simultaneously or sequentially. The subsequent loss of the two UDP groups from the catalytic sites is proposed to provide the driving force for the chain to move through the catalytic cleft until the terminal sugar interacts with the last binding subsite, allowing two more UDP sugars to enter. The simultaneous formation of two glycosidic linkages provides a simple mechanism for the generation of the 2-fold screw axis that arises from a disaccharide repeat with two β-glycosidic linkages, without invoking a concomitant rotation of either the enzyme or the substrate. It also provides an effective mechanism for maintaining the fidelity of a heteropolysaccharide disaccharide repeat. However, by analogy with the inverting glycosidases, such an activity would involve a total of four conserved acidic amino acid residues, two in each domain. Sequence and HCA plot alignments identified only one conserved acidic residue in domain B. Based on sequence conservation in the chitin synthases, Nagashita et al. (57) have also speculated that this residue is involved in catalysis. To further substantiate this hypothesis, these workers used site-specific mutagenesis to replace this conserved Asp (Asp-562) with the longer Glu and observed a 100% reduction in enzyme activity. They also replaced Asp-562 with an Asn, to determine whether the hydroxyl group of Asp was necessary for activity. Not surprisingly, a complete loss of activity was observed in the mutant. Because the inverting mechanism predicts two catalytic residues in domain B, the sequences of all of the known transferases in the HasA family were examined for additional...
conserved amino acids. The search was limited to the region extending from the end of β4 to approximately 25 residues past the (QXXRW) motif. The C-terminal end of the search region corresponds to the start of a predicted transmembrane domain in RfbB_{O:54}. Cps3S, HasA, IcaA, and NodC. A highly conserved proline was identified a short distance in front of the (QXXRW) motif, in the junction between regions III and IV (Fig. 8). This Pro is predicted to lie within a loop at the C-terminal end of a β-sheet. In all of the proteins, an Asp or Glu was found 2–4 residues before the conserved Pro. It is possible that the conserved carboxylate adjacent to this Pro represents the second catalytic residue of domain B. In RfbB_{O:54}, the catalytic residues are therefore speculated to be Asp-92 or Asp-94, Asp-151, Asp-244, and either Glu-268 or Asp-269.

The position of the (QXXRW) motif relative to the fourth potential catalytic residue in these proteins suggests that the motif may represent part of a binding subsite. This possibility is supported by the predicted interactions of the residues in the motif with a carbohydrate ligand: hydrophobic interactions between the aromatic Trp residue and the hydrophobic patch of a glycosyl monomer, and hydrogen-bonding interactions between the guanidinium side chain of Arg and the glycosidic hydroxyls. Noting that this motif was highly conserved in the chitin synthases, Nagahashi et al. (57) also speculated on the possible function of this motif in enzyme activity. To confirm a structure-function relationship, they used site-specific mutagenesis to individually replace each residue in the motif, and then measured enzyme activity and K_m values in the resulting mutant proteins. In each case, a conservative change resulted in a reduction in activity and either an increase or decrease in binding affinity. While these results were interpreted as evidence for a role in catalysis, they could equally be interpreted as evidence for a role in binding of the substrate or in hydrogen-bonding interactions with a catalytic residue, thereby maintaining the correct orientation for catalysis.

There are presently two known pathways for O-antigen biosynthesis. These pathways are fundamentally different. Key criteria distinguishing the two are the cellular location of the polymerization reaction and the mode of export across the plasma membrane. In RfE-dependent synthesis, polymerization involves block-wise addition of single O-repeat units and occurs at the periplasmic face of the plasma membrane. Individual O-units are assembled on undecaprenol-P in the cytoplasm and then transported across the plasma membrane, presumably by the O-unit transporter, RfhX (58). Ligation to lipid A core is catalyzed by RfaL. RfE-independent O-poly saccharide biosynthesis is currently limited to homopolymeric O-polysaccharides. In this pathway, synthesis is initiated by RfE-dependent transfer of GlcNAc-1-P to undecaprenol-P. The complete O-polysaccharide chain is then polymerized in the cytoplasm prior to being delivered to the site of ligation by an rfb-encoded dedicated ABC transporter. At first glance, the pathway for O:54 polysaccharide synthesis appears similar to the RfE-independent pathway of O-antigen synthesis. In both cases, synthesis is initiated by RfE and the O-repeat is a homopolymer. In addition, by analogy with other more well-characterized HasA-related proteins, RfbB_{O:54} is expected to polymerize the O:54 polysaccharide in the cytoplasm, where pools of activated precursors are available. It is at this point in the pathway that O:54 biosynthesis diverges from the RfE-independent pathway; there is no dedicated ABC transporter for export of the polymerized O:54 polysaccharide. This is the only known O-polysaccharide system that does not encode either an ABC transporter or an RfhY O-unit transporter. Despite the absence of either of these components, smooth O:54 LPS is expressed in a Δrfb E. coli K-12 host strain containing only RfbA_{O:54} and rfbB_{O:54}. We have previously shown that, in the absence of the cognate ABC transporter, an E. coli K-12 strain containing the remaining rfb genes from K. pneumoniae O1 (RfE-independent) accumulates O-antigen in the cytoplasm (13). There is, therefore, no alternate, generic O-poly saccharide export system in E. coli. Consequently O:54 synthesis represents a new pathway for O-antigen assembly, involving a different mechanism for delivering nascent O-poly saccharide to the LPS-O-antigen ligase.

Intriguingly, no export system has yet been identified for polymers produced by a number of other processive β-glycosyltransferases in the HasA family. These polysaccharides include bacterial cellulose (Acetobacter xylinum and Agrobacterium tumefaciens (53, 59, 60)), hyaluronic acid (S. pyogenes (47, 48, 61)), the type 3 capsule of S. pneumoniae (45, 46), alginate (P. aeruginosa (62)), and chitin (S. cerevisiae, Candida albicans, Enterococci nidulans and Neospora caninum (63–68)). In contrast, transmembrane transport of the rhizobial Nod factors occurs through the action of a dedicated ABC transporter (69, 70). However, synthesis of Nod signal factor differs somewhat from that of cell-surface polysaccharides, as a carrier lipid has not yet been identified; the polymer that is synthesized is much shorter and is generally thought to be secreted (71); and the oligosaccharide product is substituted with acyl, acetyl, and sulfate groups in a strain-specific manner (72). The Nod proteins are all highly homologous, and their predicted topology differs from that of RfbB_{O:54}, HasA, and Cps3S. Nod proteins may therefore represent a subfamily of the HasA family. The predicted topology of RfbB_{O:54}, with a periplasmic loop following the cytoplasmic glycosyltransferase domain, suggests the possibility that the protein possesses two separate activities, catalyzing the polymerization of the O:54 polysaccharide and coupling this with transport in a vectorial reaction, similar to that suggested for chitin synthases (73). The C-terminal transmembrane domains would therefore be predicted to form a pore or channel through which the growing chain is extruded. Similarities in size, two-dimensional architecture, and hydropathy plots point to the possibility of a similar transferase/transport function for Cps3S, IcaA, and HasA.

Synthesis of the O:54 polysaccharide clearly represents a third and new pathway for O-antigen assembly. With the identification of potential catalytic residues in RfbB_{O:54} and the speculated role for the periplasmic loop in transmembrane transport, the O:54 O-polysaccharide provides a relatively simple system for testing this putative export function and for examining the mechanism of catalysis of β-glycosyltransferases. These analyses will also serve to characterize a novel O-polysaccharide biosynthetic pathway.

Acknowledgments—We thank Dr. A. J. Clarke for his help in the analysis and presentation of the multiple sequence alignments and for critically reviewing the manuscript. We also thank Malcolm Perry at the National Research Council of Canada for performing the 1H NMR, Paul Amor for the O:54 polyclonal antisemur, Christine Dodgson for construction of the TnphoA fusions, and Karen Sutherland for technical assistance.

REFERENCES

1. Lindberg, A. A., and Le Minor, L. (1984) Methods Microbiol. 14, 1–141
2. Reeve, P. R. (1992) FEBS Lett. 310, 509–516
3. Keenleyside, W. J., Perry, M., MacLean, L., Poppe, C., and Whitfield, C. (1994) Mol. Microbiol. 11, 437–448
4. Keenleyside, W. J., and Whitfield, C. (1995) J. Bacteriol. 177, 5247–5253
5. Meier-Dieter, U., Barr, K., Starman, R., Hatch, L., and Rick, P. D. (1992) J. Biol. Chem. 267, 746–753
6. Whitfield, C. (1995) Trends Microbiol. 3, 178–185
7. Clarke, B. R., and Whitfield, C. (1992) J. Bacteriol. 174, 4614–4621
8. Clarke, B. R., Bronner, D., Keenleyside, W. J., Severn, W. B., Richards, J. C., and Whitfield, C. (1995) J. Bacteriol. 177, 5411–5418
9. Kido, N., Togov, V. I., Sugiyama, T., Uchiya, K., Sugihara, T., Kato, N., and Jann, K. (1995) J. Bacteriol. 177, 2178–2187
10. Rick, P. D., Hubbard, G. L., and Barr, K. (1994) J. Bacteriol. 176, 2877–2884
11. Weisgerber, C., and Jann, K. (1982) Eur. J. Biochem. 127, 165–168
