The gene cluster directing O-antigen biosynthesis in Yersinia enterocolitica serotype O:8: identification of the genes for mannose and galactose biosynthesis and the gene for the O-antigen polymerase

Lijuan Zhang,1,2 Paavo Toivanen1 and Mikael Skurnik2

Author for correspondence: Lijuan Zhang. Tel: +358 21 3337407. Fax: +358 21 2330008. e-mail: Lzhang@finabo.abo.fi

Department of Medical Microbiology1, and Turku Centre for Biotechnology2, University of Turku, Klinamyllynkatu 13, 20520 Turku, Finland

The rfb gene cluster of Yersinia enterocolitica serotype O:8 (Ye08) strain 8081-c was cloned by cosmid cloning. Restriction mapping, deletion analysis and transposon mutagenesis showed that about 19 kb of the cloned DNA is essential for the synthesis and expression of the Ye08 O-side-chain in Escherichia coli. Deletion analysis generated a derivative that expressed semi-rough LPS, a phenotype typical of an rfc mutant lacking the O-antigen polymerase. The deletions and transcomplementation experiments allowed localization of the rfc gene to the 3'-end of the rfb gene cluster. The deduced Ye08 Rfc did not share significant amino acid sequence similarity with any other protein, but its amino acid composition and hydrophobicity profile are similar to those of identified Rfc proteins. In addition, the codon usage of the rfc gene is similar to other rfc genes. Nucleotide sequence analysis identified three other genes upstream of rfc. Two of the gene products showed 60–70% identity to the RfbM and RfbK proteins that are biosynthetic enzymes for the GDPmannose pathway of enterobacteria. The third gene product was about 50–80% identical to the bacterial GalE protein, UDPglucose 4-epimerase, which catalyses the epimerization of UDPglucose to UDPgalactose. Since mannose and galactose are both present in the Ye08 O-antigen repeat unit, the above three genes are likely to belong to the rfb gene cluster. A gene similar to the gsk gene downstream of rfc, and genes similar to adk and hemH upstream of the rfb gene cluster, were recognized. Thus the rfb gene cluster of Ye08 is located between the adk-hemH and gsk loci, and the order is adk–hemH–rfb–rfc–gsk in the chromosome. Also in other Yersinia spp., the locus downstream of the hemH gene is occupied by gene clusters associated with LPS biosynthesis.

Keywords: Yersinia enterocolitica, rfb gene cluster, lipopolysaccharide (LPS), sequencing, transposon mutagenesis

INTRODUCTION

Lipopolysaccharide (LPS) is the major component of the outer membrane of Gram-negative bacteria. It is the

Abbreviations: DOC-PAGE, deoxycholate polyacrylamide gel electrophoresis; SR, semi-rough; YeO3(8), Yersinia enterocolitica serotype O:3(8).

The GenBank/EMBL/DDBJ accession numbers for the 1245 bp and 5553 bp nucleotide sequences reported in this paper are U25113 and U18674, respectively.

outermost structure encountering the surrounding environment. The basic structure of the LPS has three main components: lipid A, core polysaccharide and O-side-chain (O-antigen). LPS has received intensive attention due to its importance as an immunostimulatory agent (lipid A) or as a virulence factor (O-antigen). The lipid A part is responsible for the endotoxin activity. The O-antigen functions as a barrier against complement-mediated lysis (Joiner, 1985) and resists killing of bacteria
by microbicidal intracellular granules of polymorphonuclear leucocytes (Stinavage et al., 1989).

The O-antigen comprises chains of repeating units of one to five sugars and varies considerably among bacterial species in terms of chemical composition, structure and antigenicity. The genes directing the O-antigen biosynthesis are usually clustered in the rfb locus of the bacterial chromosome (Schnaitman & Klena, 1993). The rfb genes encode the enzymes that synthesize the nucleotide-activated sugar precursors and the corresponding nucleotide sugars to growing repeat units on a lipid carrier on the cytoplasmic face of the inner membrane. Subsequently the complete repeating units are translocated to the periplasmic face of the inner membrane where they are polymerized to long chains. Finally, an O-antigen ligase ligates the fully extended O-antigen to the core part of the lipid-A-core complex to give rise to a typical smooth LPS synthesis are usually clustered in the rfb gene cluster (Schnaitman & Klena, 1993). Two different mechanisms of polymerization of the repeating units have been described or speculated in the literature: rfc-dependent and rfe-independent polymerization. In the former case the rfe gene encoding the O-antigen polymerase is needed. In Salmonella serogroups A, B and D, the rfe gene is located outside the rfb region (Collins & Hackett, 1991), while in Salmonella enterica serogroups C1, C2 and E1, Shigella dysenteriae type I, and Shigella flexneri, rfe is located within the rfb gene cluster (Brown et al., 1991; Lee et al., 1992; Wang et al., 1992; Klena & Schnaitman, 1993; Morona et al., 1994). In the rfe-independent polymerization represented by Escherichia coli serotypes O8 and O9, the O-antigen is synthesized by processive transfer of sugars to a growing chain attached to undecaprenol pyrophosphate (Jann & Jann, 1984; Whitfield & Valvano, 1993). Thus there is no formal requirement for a separate polymerase activity.

Yersinia enterocolitica is an invasive Gram-negative enteropathogen that most commonly causes enterocolitis in humans. There are about 60 different serotypes and the serotypes are mainly determined by the variability of the O-antigen. The genetics of the O-antigen and its role in virulence have been studied mainly in Y. enterocolitica serotype O:3 (YeO3) (Al-Hendy et al., 1991a, b; Zhang et al., 1993). The O-antigen of YeO3 is an essential virulence factor in an orally infected murine model (Al-Hendy et al., 1992) and it plays also a role in resistance to killing by the alternative complement pathway (M. Skurnik, A. Al-Hendy & P. Toivanen, unpublished data). The O-sidechain of YeO3 is a homopolymer of 6-deoxy-L-altrose (Hoffman & Lindberg, 1980; Wartenberg et al., 1983). We demonstrated that the dTDP-6-deoxy-L-altrose biosynthetic pathway is similar to the dTDP-L-rhamnose biosynthesis pathway of Salmonella. Studies on YeO3 and Klebsiella pneumoniae serotype O1 indicate that the transport process of translocation of completed O-antigen across membranes to the cell surface is coupled to ATP hydrolysis (Zhang et al., 1993; Bronner et al., 1994), and that the O-antigen polymerization is rfe-independent (Bronner et al., 1994).

In this work we have started the molecular genetic analysis of the O-antigen of Y. enterocolitica serotype O:8 (YeO8). The structure of the O-antigen repeating unit of YeO8 is:

\[
[\rightarrow 4\beta-D-Man 1 \rightarrow 3\beta-D-Gal 1 \rightarrow 3\alpha-D-GalNAc 1 \rightarrow ]
\]

1 1 3 2

6d-D-Gul l-Fuc

(Man, mannos; Gal, galactose; GalNAc, N-acetyl-galactosamine; 6d-D-Gul, 6-deoxy-D-gulose; Fuc, fucose) (Tomshish et al., 1987). To understand the molecular genetics of the biosynthetic pathways of the O-antigen we cloned and characterized the rfb gene cluster of YeO8. We report here that the polymerization of the O-antigen in YeO8 is rfc-dependent and we identified three rfe genes involved in the GDPmannose and UDPgalactose biosynthetic pathways.

METHODS

Bacterial strains and plasmids. The bacterial strains and plasmids used in this study are listed in Table 1.

Media and growth conditions. Y. enterocolitica and E. coli strains were routinely grown in Luria-Bertani (LB) medium (10 g Bacto tryptone, 10 g yeast extract, 5 g NaCl l-') at either 25°C or 37°C. For the strains harbouring pMMB207-based plasmids, 0.2 mM IPTG was added to media to induce the lac promoter. Antibiotics were added to plates and culture media at the following concentrations (µg ml-1): ampicillin, 50; kanamycin, 50; chloramphenicol, 50.

DNA techniques. DNA isolations, restriction enzyme digestions and DNA ligations were performed as described by Ausubel et al. (1989). Restriction enzymes, alkaline phosphatase and T4 ligase were obtained from Promega or Boehringer Mannheim and were used according to the manufacturers’ instructions. Plasmid DNA purification was performed by cesium chloride/ethidium bromide density gradient centrifugation or by using the Qiagen plasmid kit. DNA fragments for subcloning or probing purposes were isolated from agarose gels by the use of GenCleanII (BiolOl) and the SpinBind DNA purification kit. DNA fragments were subcloned into pUC18 or pUC19 with appropriate restriction enzymes and cloned into E. coli strains DH5α and BL21 (DE3). The resulting plasmids were used for DNA sequencing.

Construction of a genomic cosmid library of YeO8. Chromosomal DNA of YeO8 strain 8081-c was partially digested with Sau3A1 and fractionated using the 10–40% sucrose gradient centrifugation as described by Ausubel et al. (1989). The fractions containing 20–50 kb fragments were pooled, treated with calf intestinal phosphatase (Boehringer Mannheim) and ligated to BamHI-digested pUC19. The ligation products were then introduced into E. coli DH1 by the lambda in vitro DNA packaging system (Stratagene) according to the manufacturer’s instructions. The ampicillin-resistant colonies were pooled to generate a genomic cosmid library of YeO8.

Transposon mutagenesis of pLZ6020 carrying the YeO8 rfb gene cluster. Transposon mutagenesis with Tn5 was done as described before using the temperature-sensitive mobilizable plasmid pMMB19 to deliver Tn5 (DeBrujin & Lupski, 1984). Kanamycin/ampicillin-resistant transposon mutants were selected at 42°C and were further studied. Plasmids from the mutants were retransformed for phenotypic analysis into E. coli C600 and Sf674, which is an isogenic strain of C600 but which...
Table 1. Bacterial strains and plasmids

<table>
<thead>
<tr>
<th>Strains and plasmids</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Y. enterocolitica strain</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8081-c</td>
<td>pYV- derivative of 8081, serotype O:8</td>
<td>Portnoy et al. (1981)</td>
</tr>
<tr>
<td><strong>E. coli strain</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C600</td>
<td>thi thr leuB tonA lacY supE</td>
<td>Appleyard (1954)</td>
</tr>
<tr>
<td>DH1</td>
<td>recA endA gyrA thi hisR17 supE relA</td>
<td>Hanahan (1983)</td>
</tr>
<tr>
<td>S17-1</td>
<td>thi pro hisd bsdM+ recA::RP4-2-Tc::Mu-Km::Tn7</td>
<td>Simon et al. (1983)</td>
</tr>
<tr>
<td>SM10 apr</td>
<td>thi thr leuB6 lacY tonA glnV recA::RP4-2 tet::Mu-Km</td>
<td>Miller &amp; Mekalanos (1988)</td>
</tr>
<tr>
<td>Sφ874</td>
<td>K-12 lacZ trp Δ(ishB-bii-gud-cps-rfb) upp rel rpsL</td>
<td>Keenleyside et al. (1994)</td>
</tr>
<tr>
<td>JM103</td>
<td>Δ(lac pro)thi strA supE endA sbcB</td>
<td></td>
</tr>
<tr>
<td>EC42</td>
<td>DH1(pLZ5010)</td>
<td>This study</td>
</tr>
<tr>
<td>EC43</td>
<td>DH1(pLZ5005)</td>
<td>This study</td>
</tr>
<tr>
<td>EC49</td>
<td>C600(pLZ8001)</td>
<td>This study</td>
</tr>
<tr>
<td>EC51</td>
<td>C600(pLZ6010)</td>
<td>This study</td>
</tr>
<tr>
<td>EC52</td>
<td>C600(pLZ6020)</td>
<td>This study</td>
</tr>
<tr>
<td>EC53</td>
<td>C600(pLZ6017)</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Plasmid</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M13mp18 and 19</td>
<td>Sequencing vectors</td>
<td>Norrander et al. (1983)</td>
</tr>
<tr>
<td>pHC79</td>
<td>Cosmid vector</td>
<td>Pharmacia</td>
</tr>
<tr>
<td>pHMBB70</td>
<td>pMBB19::Tn5</td>
<td>Scholz et al. (1985)</td>
</tr>
<tr>
<td>pLZ5005</td>
<td>35 kb Sau3A fragment of Y. enterocolitica 8081-c cloned in the BamHI site of pHC79</td>
<td>This study</td>
</tr>
<tr>
<td>pLZ5010</td>
<td>30 kb Sau3A fragment of Y. enterocolitica 8081-c cloned in the BamHI site of pHC79</td>
<td>This study</td>
</tr>
<tr>
<td>pLZ6017</td>
<td>BamHI deletion derivative of pLZ5010</td>
<td>This study</td>
</tr>
<tr>
<td>pLZ6001</td>
<td>BamHI deletion derivative of pLZ5010</td>
<td>This study</td>
</tr>
<tr>
<td>pLZ6017</td>
<td>BamHI deletion derivative of pLZ5010</td>
<td>This study</td>
</tr>
<tr>
<td>pM3B207</td>
<td>Expression vector</td>
<td>Morales et al. (1991)</td>
</tr>
<tr>
<td>pLZ-14</td>
<td>1.4 kb BamHI fragment of pLZ6020 cloned into the BamHI site of pLZ6020</td>
<td>This study</td>
</tr>
<tr>
<td>pLZ-41</td>
<td>As pLZ-14 but the 1.4 kb BamHI fragment is in the opposite orientation</td>
<td>This study</td>
</tr>
</tbody>
</table>

has a deletion in the chromosomal region from thi to rfb (Table 1).

**LPS isolation and immunoblotting.** LPS isolation and silver staining were performed as described previously (Helander, 1985; Zhang & Skurnik, 1994). LPS was analysed in deoxycholate (DOC)-PAGE as described by Krauss et al. (1988).

In immunoblotting, tissue culture supernatant of the mouse mAb Fu26-1F1-1 specific for the Ye08 O-antigen (Gripenberg-Lerche et al., 1994), diluted 1:20 in 2% (w/v) BSA in PBS (0.14 M NaCl; 0.0069 M K2HPO4; 0.0025 M KH2PO4), was used as the first antibody. Rabbit peroxidase-conjugated immunoglobulins against mouse immunoglobulins (P260, Dakopatts) were used as the second antibody at a 1:200 dilution in 2% BSA-PBS. The bound peroxidase was visualized with 4-chloro-l-naphthol as substrate.

**DNA sequencing.** Fragments generated by restriction endonuclease digestions of pLZ5010, the cosmid clone carrying the Ye08 rfb region, were subcloned into M13mp18 and M13mp19 for single-stranded sequencing of both strands using the Sequenase version 2.0 kit (United States Biochemical). PCR fragments over the junction sites where overlapping clones were not available were sequenced utilizing the Taq DyeDeoxy Terminator Cycle Sequencing kit (Applied Biosystems). Oligonucleotide primers were synthesized using a PCR-mate 391 DNA synthesizer (Applied Biosystems).

**Computer analysis.** The computer system GENES (version 2.0) (Harr et al., 1983, 1985) and the sequence analysis software package of the Genetics Computer Group (GCG), version 7.1 (Devereux et al., 1984), were used in this study.

**RESULTS**

**Molecular cloning of the rfb gene cluster of YeO8**

About 1000 c.f.u. of the cosmid library were screened by colony immunoblotting using mAb Fu26-1F1-1. Two positive clones, designated *E. coli* DH1(pLZ5005) and
L. ZHANG, P. TOIVANEN and M. SKURNIK

Fig. 1. O-antigen profiles in DOC-PAGE. (a) Silver-stained DOC-PAGE analysis of isolated LPS samples. (b) Immunoblotting using mAb Fu26-1F1-1 specific for the Ye08 O-antigen. Lanes: 1, C600; 2, C600(pLZ6010); 3, C600(pLZ6020); 4, Ye08 8081-c; 5, DH1(pLZ5010); 6, DH1(pLZ5005); 7, Salmonella minnesota R60 (Ra chemotype); 8, S. minnesota R5 (Rc chemotype); 9, S. DH1(pLZ5010), were obtained. LPS was isolated from the clones and analysed in SDS-PAGE. Both clones had a typical ladder-like LPS profile similar to Ye08 O-antigen in silver-stained gels and reacted in immunoblots with mAb Fu26-1F1-1 (Fig. 1a, b, lanes 5 and 6). Restriction maps of plasmids pLZ5005 and pLZ5010 were constructed and are shown in Fig. 2. pLZ5005 contains a larger insert and includes the region cloned in pLZ5010. Restriction mapping revealed that both clones contained two copies of the cloning vector pH79 ligated head to tail and this was confirmed by Southern blotting using a pH79-specific probe (data not shown).

Mapping of the O-antigen coding region in pLZ5010

Four deletion derivatives, pLZ6020, pLZ6010, pLZ8001 and pLZ6017, were generated from pLZ5010 by BamHI or SalI digestion and religation. Plasmid pLZ6020, in which the insert ends at the BamHI site at a position of 20.5 kb, still expressed wild-type O-antigen on the cell surface of E. coli C600 (Fig. 1a, b, lane 3), and in S874 (data not shown). Deletion derivative pLZ6010, also missing the internal 1.4 kb BamHI fragment, expressed LPS with two bands in DOC-PAGE (Fig. 1a, b, lane 2). Deletions extending to the SalI site (pLZ8001) or pLZ6017 containing only the distal BamHI fragment (Fig. 2) were O-antigen negative (data not shown). These results suggested that the right side border of the rfb region is close to the BamHI site at 20.5 kb (Fig. 2).

Deletions on the left side of the cloned fragment could not be constructed due to lack of suitable restriction sites in pLZ6020. Therefore transposon mutagenesis using Tn5 was applied to localize the left border of the rfb region. The insertion sites are shown in Fig. 3.

Based on the O-antigen expression in E. coli C600 the transposon insertion mutants were divided into five phenotypic classes (Fig. 3). Class I, with insertions between restriction map positions 1.6 and 1.9 kb, expressed O-antigen normally and we assume that the left border of the rfb region is located close to 1.9 kb. Class II, with insertions between 2.3 and 5.4 kb, and class IV, with an insertion at 11.3 kb, were completely O-antigen negative. Class III, with an insertion at 6.9 kb, expressed O-antigen normally in both E. coli C600 and S874. Class V was the only class where the O-antigen expression phenotype was different in E. coli C600 (positive) and in S874 (negative). One insertion was in the vector and it was O-antigen positive as expected (data not shown).

Identification of the putative O-antigen polymerase gene

In colony immunoblotting, C600(pLZ6010) reacted with mAb Fu26-1F1-1 (data not shown), and in DOC-PAGE analysis its LPS phenotype showed the characteristics of the so-called semi-rough (SR) LPS (Fig. 1a, lane 2). SR-

minnesota R595 (Re chemotype). Used as size standards, the Ra chemotype LPS molecule has a complete core with eleven sugar residues, Rc has a core with seven, and Re has three sugar residues.
rj2

gene cluster of Y. enterocolitica serotype 0:8

LPS contains one single O-subunit attached to the lipid-A–core complex due to lack of Rfc protein, the O-antigen polymerase (Naide et al., 1965; Yuasa et al., 1970). This phenotype is clearly seen by the differences illustrated by the LPS profiles in Fig. 1(a) (lanes 1–3). The LPS of C600 (Fig. 1a, lane 1) migrates in the gel as a single band composed of lipid A plus core. The LPS (smooth-type LPS) of C600(pLZ6020) (Fig. 1a, lane 3) migrates as multiple bands (ladder-like); each band of the ladder contains LPS molecules with one additional O-unit. The LPS of C600(pLZ6010) migrates as two bands: the lower band represents the lipid-A-core complex, and the upper band migrates to the position corresponding to LPS containing a single O-subunit plus core-lipid A (Fig. 1a, lane 2); this latter band reacted positively with Fu26-1F1-1 in immunoblotting (Fig. 1b, lane 2).

An alternative explanation for the phenotype of C600(pLZ6010) could be the loss of a distal glycosyltransferase gene, which would lead to one incomplete O-subunit being attached to core-lipid A; this incomplete O-subunit would lack the distal sugar residue. If this was the case in C600(pLZ6010), the upper band would migrate a little faster than the first O-side-chain band of the smooth-type of LPS. However, this is not seen here (Fig. 1a, b, lane 2). Instead the band migrates identically with the first O-side-chain band of C600(pLZ6020) in both DOC-PAGE and immunoblotting. Thus we believe that the rfc analogue is not present in pLZ6010.

The difference between pLZ6010 and pLZ6020 is the absence of a 1.4 kb BamHI fragment in pLZ6010, suggesting that the rfc gene is located in it. To confirm this, C600(pLZ6010) was transcomplemented with pLZ41 carrying the 1.4 kb fragment in either orientation downstream of the tac promoter (Table 1). Plasmid pLZ14 but not pLZ41 nor the vector alone was able to transcomplement C600(pLZ6010) to produce ladder-like O-antigen (data not shown).

Nucleotide sequence analysis

The sequence of the 1.4 kb BamHI fragment was determined. Inspection of the sequence revealed two incomplete ORFs. The first ORF was a 3′-end of a gene and
it encoded a truncated polypeptide of 145 amino acids in length. The amino acid sequence shared significant identity to a number of bacterial UDPglucose 4-epimerases (GalE, EC 5.1.3.2) (Fig. 4c). Thus the second ORF was considered as the putative \( \text{rfl} \) gene based on the deletion and transcomplementation data. To complete the

Fig. 4(a, b). For legend see facing page.
The *rfb* gene cluster of *Y. enterocolitica* serotype 0:8

![Figure 4](image)

**Fig. 4.** Multiple alignments of amino acid sequence. (a) Multiple alignment of Ye08 RfbM (Y-RfbM) to the RfbM and CpsB proteins; (b) multiple alignment of Ye08 RfbK (Y-RfbK) to the RfbK and CpsG proteins; (c) multiple alignment of Ye08 GalE (Y-GalE) to other GalE proteins of bacteria; (d) multiple alignment of Ye08 Rfc (Y-Rfc) to *E. coli* FepE and *Y. pseudotuberculosis* Cld (Ypstb-Cld) proteins. E-CpsB (E-CpsG), *E. coli* CpsB (CpsG); S-CpsB (S-CpsG), *Salmonella* CpsB (CpsG); C1-RfbM (C1-RfbK), group C1 *Salmonella enterica* RfbM (RfbK); B-RfbM, group B *Salmonella enterica* RfbM; H-GalE, GalE of *H. influenzae*; E-GalE, GalE of *E. coli*.

**DISCUSSION**

**Identification of the rfb gene cluster**

In this work we report the cloning and preliminary characterization of the *rfb* and *rfs* genes of Ye08. The cloned *rfb* gene cluster properly expressed Ye08 O-antigen in *E. coli*, suggesting that the genes necessary for O-antigen repeat unit synthesis and polymerization are recognized from the sequence. Based on the similarity given by the database search, the two ORFs were designated *adh* and *hemG*, encoding adenylate kinase and ferrochelatase, respectively (Fig. 3; Table 2).

**gaIE and rfCE sequences.** The regions flanking the *BamHI* fragment between map positions 15.5 and 21 kb were determined. Altogether four complete and one incomplete ORFs were identified from the sequence as indicated in Fig. 3. A summary of the properties of the ORFs is shown in Table 2. Based on similarity to known proteins, the two ORFs upstream of *galE* were designated *r-M* and *rjbK* and the ORF downstream of *rfc* was designated *gsk* (encoding the inosine-guanosine kinase gene). Each ORF is preceded by a potential ribosome-binding site with the exception of the *gsk* gene (data not shown).

No clear -35 and -10 promoter motifs were identified from the sequenced region. We also sequenced the first 1.3 kb of DNA upstream of the *rfzl* gene cluster. Two ORFs were recognized from the sequence. Based on the similarity given by the database search, the two ORFs were designated *adk* and *hemH*, encoding adenylate kinase and ferrochelatase, respectively (Fig. 3; Table 2).
L. ZHANG, P. TOIVANEN and M. SKURNIK

**Table 2. YeO8 rfb region encoded proteins: properties and similarities to proteins in GenBank/EMBL and H. influenzae databases**

<table>
<thead>
<tr>
<th>YeO8 proteins</th>
<th>Size (amino acids/kDa)</th>
<th>pI</th>
<th>mol % G+C of the ORF</th>
<th>Similar polypeptide</th>
<th>Database accession no.</th>
<th>Degree of identity</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>RfbM</td>
<td>465/51-9</td>
<td>5-5</td>
<td>43.4</td>
<td>GDPmannose pyrophosphorylase (EC 2.7.7.13)</td>
<td>X59886</td>
<td>61 %</td>
<td>CpsB of S. enterica</td>
</tr>
<tr>
<td>RfbK</td>
<td>456/50-8</td>
<td>5-32</td>
<td>37.3</td>
<td>Phosphomannomutase (EC 5.4.2.8)</td>
<td>M84642</td>
<td>59 %</td>
<td>RfbM of C1 S. enterica</td>
</tr>
<tr>
<td>GalE</td>
<td>335/37-0</td>
<td>6-85</td>
<td>36.5</td>
<td>UDPglucose 4-epimerase (EC 5.1.3.2)</td>
<td>L11721</td>
<td>60 %</td>
<td>CpsB of E. coli</td>
</tr>
<tr>
<td>Rfc</td>
<td>362/40-8</td>
<td>9-26</td>
<td>32.4</td>
<td>Ferric enterobactin transport protein (FepE) Cld</td>
<td>X74129</td>
<td>25 %</td>
<td>E. coli</td>
</tr>
<tr>
<td>Adk</td>
<td>&gt; 60/6-9</td>
<td>10-53</td>
<td>45.1</td>
<td>Adenylate kinase</td>
<td>U13865</td>
<td>36 %</td>
<td>Y. pseudotuberculosis</td>
</tr>
<tr>
<td>HemH</td>
<td>322/36-7</td>
<td>8-37</td>
<td>47.3</td>
<td>Ferrochelatase</td>
<td>H10351</td>
<td>63 %</td>
<td>H. influenzae</td>
</tr>
<tr>
<td>Gsk</td>
<td>&gt; 79/ &gt; 9-2</td>
<td>9-92</td>
<td>43.3</td>
<td>Inosine-guanosine kinase (EC 2.7.1.73)</td>
<td>L19201</td>
<td>40 %</td>
<td>E. coli</td>
</tr>
</tbody>
</table>

located in the cloned fragment. Deletion analysis, transposon mutagenesis and nucleotide sequencing showed that the size of the rfb region was about 19 kb, starting at map position 1-9 kb and ending at 21 kb (Fig. 3).

LPS phenotypes of deletion derivatives and Tn5 insertion mutants gave information on the organization of the rfb gene cluster. Tn5 insertions identified five phenotypic classes (Fig. 3). Normal expression of O-antigen in both C600 and S@74 by the class I11 phenotypic mutant having an insertion at 6.9 kb suggests that the YeO8 rfb region would be composed of two transcriptional units. A more detailed analysis of the rfb region is needed to elucidate this question. At present the exact number of genes and functions of the gene products between map positions 1.9 and 15 kb cannot be given. Instead, based on the LPS profiles and sequence data, four genes associated with O-antigen biosynthesis were identified from the region between 15 and 21 kb.

The rfbMK genes involved in GDPmannose biosynthesis

The two genes involved in the synthesis of GDPmannose from mannose 6-phosphate are rfbM and rfbK, located near the right border of the rfb gene cluster of *Salmonella* (Jiang et al., 1991). In *Salmonella*, the cps gene cluster containing genes directing the biosynthesis of the M antigen, a capsular polysaccharide comprising glucose, galactose and fucose (Markovitz, 1977), is near the rfb region. Two isogenes of rfbMK, cpsB and cpsG, also direct the biosynthesis of GDPmannose because in capsule biosynthesis the precursor GDPfucose is synthesized from GDPmannose (Markovitz, 1977). Based on the nucleotide sequence similarities the rfbM and rfbK genes in the rfb gene cluster of YeO8 were identified upstream of the galE and rfc genes. The high similarity between RfbM and CpsB, and between RfbK and CpsG, argue for the same function for these proteins (Fig. 4a, b). Supporting the above conclusion are the phenotypes of the class V Tn5 insertions that mapped within the rfbM gene (Fig. 3). Although *E. coli* K12 does not have rfbMK genes in the rfb region, the cpsB and cpsG genes are present (Stevenson et al., 1994). Complementation by the cpsB gene explains why these mutants expressed O-antigen in C600 but not in S@74, since both the rfb and cps genes clusters are deleted in S@74 (Keenleyside et al., 1994). On the other hand, complementation of class V mutants by chromosomal genes of *E. coli* also indicates that the genes downstream of rfbM (galE and rfc) must be transcribed. This suggests that either there is a separate transcriptional unit for the galE and rfc genes or that the Tn5 insertion is leaky. The absence of clear promoter motifs upstream of galE favours the latter explanation, especially since we have also seen this phenomenon earlier (Zhang et al., 1993).

**galE is within the rfb gene cluster**

The gene upstream of the rfc gene is highly similar to the galE gene of other bacteria, and about 50-80% identity at
amino acid level was observed to different GalE proteins (Fig. 4c). The GalE protein catalyses epimerization of UDPglucose to UDPgalactose (Adhya, 1989). The presence of galactose in the Ye08 O-antigen makes the location of the galE gene within the rfb region logical. To our knowledge, Ye08 is the first case where the galE gene is located within the rfb region. UDPgalactose has other roles in the metabolism of S. enterica and E. coli and the galE gene is located outside the rfb cluster (Reeves, 1994) at 17 min on the E. coli K12 and 18 min on Salmonella chromosomal maps (Bachmann, 1990; Sanderson et al., 1995). In Neisseria gonorrhoeae the galE gene was found in the capsule gene complex region D (Robertson et al., 1993), and recently the DNA sequence of this region showed that cryptic rhamnose biosynthesis genes are adjacent to the galE gene (Robertson et al., 1994). In Haemophilus influenzae, galE is located in a lipo-oligosaccharide-associated locus lic3 upstream of adk (Maskell et al., 1991), and in Ye03 it is located in the trs locus involved in the LPS outer-core biosynthesis, upstream of the gsk gene (Skurnik et al., 1995).

**Ye08 O-antigen polymerization is rfc-dependent**

The deletion from pLZ6020 of the 1.4 kb BamHI fragment between 19 and 20.5 kb resulted in pLZ6010, leading to the identification of a sixth phenotypic class that expresses SRLPS, a typical feature of rfc mutants. This helped us to localize the rfc gene to the right border of the rfb gene cluster. Sequence information revealed that the Rfc protein expressed by pLZ6020 is truncated at 57 C-terminal amino acids. Nevertheless, pLZ6020 expresses a ladder-like O-side-chain. This is also consistent with the finding that the subcloned 1.4 kb BamHI fragment is able to complement the rfc defect. These data indicate that the C-terminal 57 amino acids of the Ye08 Rfc protein may not be essential for the function of the polymerase. The biosynthesis of O-antigens in Ye08 and Ye03 is different. The Ye03 O-antigen is a homopolymer of 6-deoxy-L-altrose, and polymerization of this type of O-antigen is rfc-independent and occurs in the cytoplasm. The O-units are sequentially added to the nonreducing terminus of the acceptor in a processive mechanism (Whitfield, 1995), and an ‘ABC’ transporter exports polymerized homopolymeric polysaccharide across the plasma membrane (Zhang et al., 1993; Whitfield, 1995).

Our data and the LPS phenotype of the cld mutant (Kessler et al., 1991) argue that Ye08 Rfc and Cld are both O-antigen polymerases. How FepE fits into this picture is an enigma. The amino acid composition of Rfc shows that it has a high content of nonpolar amino acids which accounts for 42% of the total amino acids, indicating that Rfc is a hydrophobic protein. The hydrophobicity plot shows that the Ye08 Rfc has at least 12 potential membrane-spanning segments and that the middle region is less hydrophobic than that of other known Rfc proteins (Fig. 5). The codon usage of the Ye08 Rfc gene is interesting. Since the gene has a low G + C content (32.4%) the wobble position is highly (75%) occupied by A or T. In addition, the rfc seems to have unusual codon usage. The codons AUA (Ile) and AGA/AGG/CGG (Arg) are considered as modulating codons because the cognate tRNAs are very minor species in E. coli (Grosjean & Fiers, 1982), i.e. these codons are used in poorly translated genes. In the Ye08 rfc gene, these codons are highly represented, i.e. > 50% of the Ile and Arg codons are modulators. This situation is similar worth mentioning that Cld and FepE are 36% identical (Stevenson et al., 1995). Our data and the LPS phenotype of the cld mutant (Kessler et al., 1991) argue that Ye08 Rfc and Cld are both O-antigen polymerases. How FepE fits into this picture is an enigma. The amino acid composition of Rfc shows that it has a high content of nonpolar amino acids which accounts for 42% of the total amino acids, indicating that Rfc is a hydrophobic protein. The hydrophobicity plot shows that the Ye08 Rfc has at least 12 potential membrane-spanning segments and that the middle region is less hydrophobic than that of other known Rfc proteins (Fig. 5). The codon usage of the Ye08 rfc gene is interesting. Since the gene has a low G + C content (32.4%) the wobble position is highly (75%) occupied by A or T. In addition, the rfc seems to have unusual codon usage. The codons AUA (Ile) and AGA/AGG/CGG (Arg) are considered as modulating codons because the cognate tRNAs are very minor species in E. coli (Grosjean & Fiers, 1982), i.e. these codons are used in poorly translated genes. In the Ye08 rfc gene, these codons are highly represented, i.e. > 50% of the Ile and Arg codons are modulators. This situation is similar
in the other rfc genes (Collins & Hackett, 1991; Morona et al., 1994). This characteristic is common to all identified rfc genes and it is believed to be the reason for the poor expression of Rfc in *in vitro* transcription and translation experiments (Collins & Hackett, 1991; Morona et al., 1994).

### Chromosomal location of the rfb gene cluster of Ye08

In *E. coli* and *S. enterica*, the rfb region is closely linked to the bis and gnd genes on the bacterial chromosome, which are immediately upstream and downstream of rfb, respectively, at 44 min on the chromosomal map of *E. coli* K12 and at 42 min on the *S. enterica* LT2 map (Bachmann, 1990). However, this location is apparently not used by *Yersinia*. The deduced polypeptides of the two genes upstream of the rfb gene cluster are highly identical to the products of adk and hemH encoding adenylyl kinase and ferrochelatase, respectively. In *E. coli* K12, the adk and hemH genes are located at approximately 11 min on the chromosome (Bachmann, 1990), at a considerable distance from the rfb region. The gene downstream of Ye08 rfc is gsk, which is located at 13 min on the *E. coli* chromosome (Bachmann, 1990). In YeO3, adk and hemH are located upstream of a gene cluster (trs gene cluster) directing the biosynthesis of LPS outer-core oligosaccharide (Skurnik et al., 1995). The galE and gsk genes are located downstream of the trs gene cluster. In addition, the rfb gene cluster of *Y. pseudotuberculosis* M85 (Ypsbt) is located downstream of hemH (Kessler et al., 1993). Thus it is obvious that the locus in the *Yersinia* chromosome between the hemH and gsk genes is occupied by gene clusters associated with LPS biosynthesis. Because we only have a partial sequence of the Ye08 rfb gene cluster, extensive sequence comparisons with the other *Yersinia* sequences are premature. On the other hand one would not expect too much similarity between these three gene clusters since each directs the biosynthesis of oligosaccharides of different composition and structure. At present only limited similarities upstream of the gsk gene can be recognized. galE is present in rfb*Ye08* and in the trsYpsbt operon but not in rfb*Ypsbt*. rfc is located between galE and gsk in rfb*Ye08*. At the same location there is nothing in the trs operon, while in rfb*Ypsbt* there is the cld gene which we believe is the rfc gene. Thus traces of similar organization can be recognized between the three organisms, indicating that they may have a common ancestral origin where a gene cluster directing the biosynthesis of an oligosaccharide was inserted between the hemH and gsk genes. During evolution the genes in the cluster have changed more or less completely with the help of deletions and interspecific gene transfer, much like what has been speculated for the *Salmonella* O-antigen variation. It is generally believed that the separation of *Yersinia* and *Salmonella* is more ancient than that of *Salmonella* and *E. coli* (Brenner, 1978). Different chromosomal locations of the rfb gene cluster in *Yersinia* and *Salmonella* indicate that the rfb gene clusters have inserted independently and at different time-points into these two organisms. Sequence analysis of the whole rfb gene cluster will allow more precise conclusions about the relationship, homology and history of the rfb genes.

### ACKNOWLEDGEMENTS

The authors are indebted to Dr C. Whitfield for generously providing *E. coli* strain S8874 and Dr H. Mayer for providing LPS preparations of Ra, Rc and Re chemotypes of *S. minnesota*. This study was supported by the Sigrid Jusélius Foundation, The Academy of Finland and The Turku University Foundation.

### REFERENCES


287


Received 21 July 1995; revised 25 September 1995; accepted 28 September 1995.