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, Kiinamyllynkatu 13, 20520 Turku, Finland

The *rfb* gene cluster of *Yersinia enterocolitica* serotype O:8 (YeO8) 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 YeO8 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 YeO8 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 YeO8 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 YeO8 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/DBJ 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 oligosaccharide 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 sugar transferases, which transfer the 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 (Schnaitman & Klena, 1993). Two different mechanisms of polymerization of the repeating units have been described or speculated in the literature: *rfc*-dependent and *rfc*-independent polymerization. In the former case the *rfc* gene encoding the O-antigen polymerase is needed. In *Salmonella* serogroups A, B and D, the *rfc* 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*, *rfc* 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 *rfc*-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 only 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-side-chain 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 *rfc*-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:

(Man, mannose; Gal, galactose; GalNAc, N-acetyl-galactosamine; 6d-D-Gul, 6-deoxy-D-gulose; Fuc, fucose) (Tomshich *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 *rfb* 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⁻¹) 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 *tac* 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 caesium 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 (Bio101) and the SpinBind DNA extraction unit (FMC BioProducts). Transformations of *E. coli* strains were carried out by the Hanahan method (Hanahan, 1983).

Construction of a genomic cosmid library of YeO8. Chromosomal DNA of YeO8 strain 8081-c was partially digested with *Sau*3AI 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 *Bam*HI-digested pHC79. 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 (DeBruijn & 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 SØ874, which is an isogenic strain of C600 but which

Table 1. Bacterial strains and plasmids

Strains and plasmids	Description	Reference
<i>Y. enterocolitica</i> strain		
8081-c	pYV ⁻ derivative of 8081, serotype O:8	Portnoy <i>et al.</i> (1981)
<i>E. coli</i> strain		
C600	<i>thi thr leuB tonA lacY supE</i>	Appleyard (1954)
DH1	<i>recA endA gyrA thi hsdR17 supE relA</i>	Hanahan (1983)
S17-1	<i>thi pro hsdR hsdM⁺ recA :: RP4-2-Tc::Mu-Km::Tn7</i>	Simon <i>et al.</i> (1983)
SM10 λ pir	(λ pir) <i>thi thr leuB6 lacY tonA glnV recA :: RP4-2 tet::Mu-Km</i>	Miller & Mekalanos (1988)
S ϕ 874	K-12 <i>lacZ trp</i> Δ (<i>sbcB-his-gnd-cps-rfb</i>) <i>upp rel rpsL</i>	Keenleyside <i>et al.</i> (1994)
JM103	Δ (<i>lac pro</i>) <i>thi strA supE endA sbcB15 hsdR4 F' traD36 proAB laqI^a ZM15</i>	
EC42	DH1(pLZ5010)	This study
EC43	DH1(pLZ5005)	This study
EC49	C600(pLZ8001)	This study
EC51	C600(pLZ6010)	This study
EC52	C600(pLZ6020)	This study
EC53	C600(pLZ6017)	This study
Plasmid		
M13mp18 and 19	Sequencing vectors	Norrande <i>et al.</i> (1983)
pHC79	Cosmid vector	Pharmacia
pMMB70	pMMB19::Tn5	Scholz <i>et al.</i> (1985)
pLZ5005	35 kb <i>Sau3A</i> fragment of <i>Y. enterocolitica</i> 8081-c cloned in the <i>Bam</i> HI site of pHC79	This study
pLZ5010	30 kb <i>Sau3A</i> fragment of <i>Y. enterocolitica</i> 8081-c cloned in the <i>Bam</i> HI site of pHC79	This study
pLZ6020	<i>Bam</i> HI deletion derivative of pLZ5010	This study
pLZ6010	<i>Bam</i> HI deletion derivative of pLZ5010	This study
pLZ8001	<i>Sa</i> II deletion derivative of pLZ5010	This study
pLZ6017	<i>Bam</i> HI deletion derivative of pLZ5010	This study
pMMB207	Expression vector	Morales <i>et al.</i> (1991)
pLZ-14	1.4 kb <i>Bam</i> HI fragment of pLZ6020 cloned into the <i>Bam</i> HI site of pMMB207	This study
pLZ-41	As pLZ-14 but the 1.4 kb <i>Bam</i> HI fragment is in the opposite orientation	This study

has a deletion in the chromosomal region from *sbc* 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 YeO8 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-1-naphthol as substrate.

DNA sequencing. Fragments generated by restriction endonuclease digestions of pLZ5010, the cosmid clone carrying the

YeO8 *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 GENEUS (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

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 YeO8 O-antigen. Lanes: 1, C600; 2, C600(pLZ6010); 3, C600(pLZ6020); 4, YeO8 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 YeO8 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 *Bam*HI or *Sal*I digestion and religation. Plasmid pLZ6020, in which the insert ends at the *Bam*HI 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 Sφ874 (data not shown). Deletion derivative pLZ6010, also missing the internal 1.4 kb *Bam*HI fragment, expressed LPS with two bands in DOC-PAGE (Fig. 1a, b, lane 2). Deletions extending to the *Sal*I site (pLZ8001) or pLZ6017 containing only the distal *Bam*HI 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 *Bam*HI 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 Sφ874. Class V was the only class where the O-antigen expression phenotype was different in *E. coli* C600 (positive) and in Sφ874 (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.

Fig. 2. Restriction maps of *E. coli* clones. The maps were constructed using the following enzymes: Sc (*ScaI*), C (*ClaI*), P (*PstI*), S (*SalI*) and B (*BamHI*). The cloning vector pHC79 is shown as black bars. pLZ5005 and pLZ5010 are the initial cosmid clones carrying two copies of pHC79 (see text). pLZ6020, pLZ6010, pLZ8001 and pLZ6017 are deletion derivatives of pLZ5010. Expression of YeO8 O-antigen was assessed by colony-immunoblotting (CIM, ++ and +) and by DOC-PAGE, where the presence of an O-antigen ladder is indicated by S (smooth), SR (semi-rough) or R (rough).

Fig. 3. Genetic characterization of the YeO8 *rfb* region. The restriction map of pLZ5010 is shown at the top. Locations of individual Tn5 insertions are indicated by vertical lines. O-antigen expression and phenotypic classes are indicated below. The sequenced region and predicted ORFs are indicated below the restriction map of pLZ5010. The arrow indicates the transcriptional direction.

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 *rfa* analogue is not present in pLZ6010.

The difference between pLZ6010 and pLZ6020 is the absence of a 1.4 kb *Bam*HI fragment in pLZ6010, suggesting that the *rfa* gene is located in it. To confirm this, C600(pLZ6010) was transcomplemented with pLZ14 and 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 *Bam*HI fragment was determined. Inspection of the sequence revealed two incomplete ORFs. The first ORF was a 3'-end of a gene and

(a)	
1	110
E-CpsB	..maqskLyPVMAGGSGSRLWPLSRvLYPKQFLcLkGdLTLMLQTTicRLngvecesPvVICNEqHRFIVAEQLRqLnKLtenIILEPaGRNTAPAIALAALaAkrhspe
S-CpsB	mimrqtkLyPVMAGGSGSRLWPLSRvLYPKQFLcLkGdLTLMLQTTicRLngvecesPvVICNEqHRFIVAEQLRqLnKLtenIILEPaGRNTAPAIALAALaAtrqhtd
C1-RfbM	mitPIIMAGGSGSRLWPLSRtLYPKQFLcLdGsgsMLQTTicRinnlnasdpivVICNEqHRFIVAEQLKelsKssgdIILEPvGRNTAPAVALAALaLkclkk...
Y-RfbM	mLlPVMIMAGGSGSRLWPLSRthLYPKQFLaLtsLsmFQeTl1RLLeishpessPvVICNaeHRFIIAEQLRqLnLhsgIVLEPIGRNTAPAIALAALaLtsn..g
B-RfbM	msFlPVMIMAGGSGSRLWPLSRthLYPKQFLaLtsLsmFQeTl1RLLeishpessPvVICNdrHRFIVAEQLRReidKLannIILEPvGRNTAPAIALAALaAlqnadn
Consensus -----L-PVIMAGG-GSRLWPLSR--YPKQFL-L-G-L-MLQ-TI-RL-----P-VICNE-HRFIVAEQLR---KL---IILEP-GRNTAPAIALAAL-A-----	
111	220
E-CpsB	sDPLMLVLAADHVIaDEdAFraAVrnAmpYAEAGKLVTFGIVPdIPETGYGYIRRG...vsageqdmvAFeVaqFVEKPNLETAqaYvaSGEYYWNSGMFLFRagRYLee
S-CpsB	cDPLMLVLAADHVIaEaFrdaVrgAmpYADaGKLVTFGIVPdIPETGYGYIRRG...vvpqatdavAFeVaqFVEKPNLETAqaYvaSGDYWNSGMFLFRagRYLee
C1-RfbM	..naLLLVLAADHIIkDEetFckIqdaRkYAEAGKLVTFGIVPtmPETGYGYIRRGalfsaendcslAFrVaeFVEKPNLETAqaYldSGEYYWNSGMFLFRadRYLee
Y-RfbM	eDPLLVLAADHVIqDkPAfIsAIqlAeplAEAGKLVTFGIVPntPETGYGYIRRG...qieggAFqVdsFVEKPNLETAqaYldSGDYWNSGMFLFRadRYLee
B-RfbM	adPLLVLAADHVIqDEIAftkAVrhAeeYAAGKLVTFGIVPthaETGYGYIRRG...igndAYaVaeFVEKPNLETAqaYldSGDYWNSGMFLFRadRYLee
Consensus -DPLLVLAADHVI-DE-AF--AV--A--YAE-GKLVTFGIVP--PETGYGYIRRG-----AF-VA-FVEKP-LETA--Y--SG-YYWNSGMFLFR-RYL-E	
221	330
E-CpsB	LkKyRPDILdAcekAmsavdpDLnFIRVDEEaFlaCPESIDYAVMERtaDAVVVPMdAGWSdVGSWSLWEISahtaeGNVhGdVInHkteNsYVYaEsgLVttVGlk
S-CpsB	LkKfRPDILaAcekAmsavdpDLDFIRVDEEaFlaCPESIDYAVMERtaDAVVVPMdAGWSdVGSWSLWEISahtpeGNVhGdVIsHkteNsYVYaEsgLVttVGlk
C1-RfbM	L..KiRPDIYKACslAmesavtDLDFIRVDEESfcaCPDESIDYAVMERtaDAVVVPLnAGWSdVGSWSLWEISdkdsngNVtkGdVIsHnadNcYlhaEtgLVtaVGvk
Y-RfbM	LeKhRPDIfeACKsAiadngkDLDFIRlDEaafysCPDESIDYAVMERtsDAVVVPMnAGWSdVGSWSLWEIndkdshGNVrgdVlTedtNsYIYsqnrLVatVGIn
B-RfbM	LkylsPEIYKACekAvghinpDLDFIRIDKEefmsCPDSIDYAVMEtqhAVVPMsAGWSdVGSWSLWDISnkdqhrNVlkgDIfaHaecnYIYsEdmFIsaIGVs
Consensus L-K-RPDI--AC--A-----DLDFIRVDEE-F--CP-ESIDYAVME-T-DAVVVPM-AGWSdVGSWSLWEIS-----GNV--GDV--H--N-Y-Y-E--LV--VGv-	
331	440
E-CpsB	DLVVQTKDAvLIAdrnaVQDVKKVVEqikadgRhEhrVrEVYRPWgKYDsIdaGdRyqVkrITVKPGEglsVqMHHHRAEHwVvGaTAKVTIdgDikLlGEnESIYI
S-CpsB	DLVVQTKDAvLIAdrnaVQDVKKVVEqikadgRhEhmHrEVYRPWgKYDsIdaGeRyqVkrITVKPGEglsVqMHHHRAEHwVvGaTArVTingEvklLlGEnESIYI
C1-RfbM	DLIVVQTKDAvLVantncVQDVKKIVEkiklenRhehithrEVYRPWgKYDsIdfGeRyqVkrITVKPGEgiseQvHyHRAEHwIIVaGTAKITikgEvklitENESYI
Y-RfbM	DLIIvETKDAILVanknkVQsVkeIVgqIKlgsRlEylqHkEVYRPWgShDaIaeGvRyHvqhVTIKPGqrtaQiHHHRAEHwVvGaTArVyratEslVtENESYI
B-RfbM	nLVIVQTTdAILVankdtVQDVKKIVDylKrndRnEyqHqEVFRPWgKYnVIdSgkNylVrcITVKPGEkfvaQmHHHRAEHwIIVlsGTArVTkeqyqMysENESYI
Consensus DLVVQTKDA-LVA---VQDVKKIVE--K--R-E--H-EVYRPWgKYD-ID-G-RY-V--ITVKPGE---Q-HHHRAEHwVv-GTA-VT---E-L--ENES-YI	
441	490
E-CpsB	PLGatHcLENPGKiPLDLIEVRSGsYLeEDDVRFadryGrv*.....
S-CpsB	PLGatHcLENPGKiPLDLIEVRSGsYLeEDDVRFadryGrv.....
C1-RfbM	PvGvkHcLENPGKiALELIEVRSGaYLGEDDIVRFsdKyGrn.....
Y-RfbM	avGvaHsIENPGKIPLEIEIVRTGSYLeEDDIVRIehki.....
B-RfbM	PpntIHaLENPGmtPLKLIETIGSGtYLGEDDIIRLeqrsGfskewtners
Consensus P-G--H-LENPGK-PL-LIEVRSG-YL-EDDIVRF---G-----	
(b)	
1	110
C1-Rfbk	MnnLtcFKAYDIRGrLGEELNEDIAWRIGRAYGEYIKPKtIVlGGDVRLTSEaLklAlakGLQDAGVDVLDIGMSGTEEYFATFHLgVDGGIEVTASHNPmDYNGMKLV
S-CpsG	MckLtcFKAYDIRGrLGEELNEDIAWRIGRAYGEYIKPKtIVlGGDVRLTSEaLnlAlakGLQDAGVDVLDIGMSGTEEYFATFHLgVDGGIEVTASHNPmDYNGMKLV
E-CpsG	MkkLtcFKAYDIRGrLGEELNEDIAWRIGRAYGEYIKPKtIVlGGDVRLTSEtLklAlakGLQDAGVDVLDIGMSGTEEYFATFHLgVDGGIEVTASHNPmDYNGMKLV
Y-Rfbk	MsaLscFKAYDIRGrLGEELNEDIAWRIGRAYGEYIKPKrVvGGDVRLTSEsLklAlarGLQDAGVDVLDIGMSGTEEYFATFHLdVDGGIEVTASHNPmDYNGMKLV
Consensus M--L-CFKAYDIRG-LGEELNEDIAWRIGRAYGE--KPK--V-VGGDVRLTSE-L--AL--GLQDAGVDVLDIGMSGTEEYF-TFHL-VDDGIEVTASHNP-DYNGMKLV	
111	220
C1-Rfbk	RegArPISGDTGLRDVQrLAEagdfPPVndaarGsYrqIsrlrDaYIDHLLaYIsvnnLtlPLkLVvNsGNGAAGpVIDAIEaRLkAlGaPVeFIKIHNTPDGTFPnGIPNP
S-CpsG	RegArPISGDTGLRDVQrLAEagdfPPVndaarGsYrqIsrlrDaYIDHLLgYIsvnnLtlPLkLVfNaGNGAAGpVIDAIEaRLkAlGaPVeFIKIHNTPDGTFPnGIPNP
E-CpsG	RegArPISGDTGLRDVQrLAEandfPPVdetkrGrYqqInlrDaYIDHLLfYIsvnnLtlPLkLVfNsGNGAAGpVVDaIEaRfKAlGaPVeFIKIHNTPDGTFPnGIPNP
Y-Rfbk	RknAcPISGDTGLRInqHLaEinkfPPVpepmkGsYdkIsiidDyIDHLMgYVnpslFkPlrLVvNsGNGAAGhVIDAIEKRFTALdiPitLIKIHNTPDGTFPnGIPNP
Consensus R--A-PISGDTGLR-VQ-LAE---PPPV-----G-Y--I---D-YIDHLL-YI---L-PL-LV-N-GNGAAG-VIDA-E-R--AL--PV--IKIHNT-PDG-TF-PnGIPNP	
221	330
C1-Rfbk	LLPECRdDtrkAviehGADMGIAFDGDFDRCLFfdekgGfIEGYIYVGLLAEaFleKhpgakIIHDPRltWNTeavVtaagGtpVMSKTGHAFIKERMRTEDAIYGGEMS
S-CpsG	LLPECRdDtrkAviehGADMGIAFDGDFDRCLFfdekgGfIEGYIYVGLLAEaFleKhpgakIIHDPRltWNTeavVtaagGtpVMSKTGHAFIKERMRTEDAIYGGEMS
E-CpsG	LLPECRdDtrkAviehGADMGIAFDGDFDRCLFfdekgGfIEGYIYVGLLAEaFleKhpgakIIHDPRltWNTvVvtaagGtpVMSKTGHAFIKERMRTEDAIYGGEMS
Y-Rfbk	LLPECRqDtdAvlanGADMGIAFDGDFDRCLFFnnhGdfIEGYIYVGLLAEfIfLkGdfskIiHDPRltWNTIdVnssGGkAImSKTGHAfIKERMRTEDAIYGGEMS
Consensus LLPECR-DT--AV---GADMGIAFDGDFDRCLF---G-FIEGYIYVGLLAE-FL-K---KIIHDPRlt-WNT--VV---GG--VMSKTGHAFIKERMRT-EDAIYGGEMS	
331	440
C1-Rfbk	AHHYFRDFAYCDSGMIPWLLVaElVclKqgsLgeLvRDRMaAFPASGEINsrLaePaaAIArVeahFaeeAqaVDrtdGLSmsFADWRFNLRsSNTepVVRNLVESRgDi
S-CpsG	AHHYFRDFAYCDSGMIPWLLVaElVclKqgsLgeLvRDRMaAFPASGEINsrLaePaaAIArVeahFaeeAqaVDrtdGLSmsFADWRFNLRsSNTepVVRNLVESRgDi
E-CpsG	AHHYFRDFAYCDSGMIPWLLVaElVclKdktLgeLvRDRMaAFPASGEINskLaqPveAlnRveghFseAlaVDrtdGLSmtFADWRFNLRsSNTepVVRNLVESRgDv
Y-Rfbk	AHHYFRDFAYCDSGMIPWLLIsEiIsVgKgsLeqLVKERMAKYPASGEINSLIdPisAIArVraaYeakAiKIdDiDgISMeFAEWRFNLRsSNTesVVRNLVESRcDt
Consensus AHHYFRDFAYCDSGMIPWLLV-E-V--K---L--LV-DRM-AFPASGEINS-L--P--AI-RV---F--A--VD--DG-SM-FADWRFNLR-SNTE-VVRNLVESR-D-	
441	457
C1-Rfbk	pLMearTkeiLqLlns.
S-CpsG	pLMearTrtlLalLnq.
E-CpsG	pLMearTrtlLtlLn*
Y-Rfbk	kLMneeTsnILrILre.
Consensus -LM---T---L--L---	

Fig. 4(a, b). For legend see facing page.

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-epimer-

ases (GalE, EC 5.1.3.2) (Fig. 4c). Thus the second ORF was considered as the putative *rfe* gene based on the deletion and transcomplementation data. To complete the

Fig. 4. Multiple alignments of amino acid sequence. (a) Multiple alignment of YeO8 RfbM (Y-RfbM) to the RfbM and CpsB proteins; (b) multiple alignment of YeO8 RfbK (Y-RfbK) to the RfbK and CpsG proteins; (c) multiple alignment of YeO8 GalE (Y-GalE) to other GalE proteins of bacteria; (d) multiple alignment of YeO8 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*.

galE and *rfe* sequences, the regions flanking the *Bam*HI 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 *rfbM* and *rfbK* and the ORF downstream of *rfe* 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 *rfb* gene cluster. Two ORFs were

recognized from the sequence. Based on the similarity given by the database search, the two ORFs were designated *ade* and *hemH*, encoding adenylate kinase and ferrochelatase, respectively (Fig. 3; Table 2).

DISCUSSION

Identification of the *rfb* gene cluster

In this work we report the cloning and preliminary characterization of the *rfb* and *rfe* genes of YeO8. The cloned *rfb* gene cluster properly expressed YeO8 O-antigen in *E. coli*, suggesting that the genes necessary for O-antigen repeat unit synthesis and polymerization are

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

YeO8 proteins	Size (amino acids/kDa)	pI	mol % G + C of the ORF	Similar polypeptide	Database accession no.	Degree of identity	Organism
RfbM	465/51.9	5.5	43.4	GDPmannose pyrophosphorylase (EC 2.7.7.13)	X59886	61 %	CpsB of <i>S. enterica</i>
					M84642	59 %	RfbM of C1 <i>S. enterica</i>
					L11721	60 %	CpsB of <i>E. coli</i>
RfbK	456/50.8	5.32	37.3	Phosphomannomutase (EC 5.4.2.8)	X59886	71 %	CpsG of <i>S. enterica</i>
					M84642	71 %	RfbK of C1 <i>S. enterica</i>
					L11721	70 %	CpsG of <i>E. coli</i>
GalE	335/37.0	6.85	36.5	UDPglucose 4-epimerase (EC 5.1.3.2)	Z47767	84 %	YeO3
					HI0351	63 %	<i>H. influenzae</i>
					X06226	50 %	<i>E. coli</i>
					L09188	53 %	<i>Neisseria meningitidis</i>
Rfc	362/40.8	9.26	32.4	Ferric enterobactin transport protein (FepE)	X74129	25 %	<i>E. coli</i>
Adk	> 60/6.9	10.53	45.1	Adenylate kinase	U13865	36 %	<i>Y. pseudotuberculosis</i>
					Z47767	92 %	YeO3
					HI0349	62 %	<i>H. influenzae</i>
HemH	322/36.7	8.37	47.3	Ferrochelatase	L19201	40 %	<i>E. coli</i>
					Z47767	95 %	YeO3
Gsk	> 79/> 9.2	9.92	43.3	Inosine-guanosine kinase (EC 2.7.1.73)	L19201	50 %	<i>E. coli</i>
					Z47767	85 %	YeO3
					D00798	82 %	<i>E. coli</i>

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 SØ874 by the class III 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 *rfe* 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 SØ874, since both the *rfb* and *cps* gene clusters are deleted in SØ874 (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 *rfe*) must be transcribed. This suggests that either there is a separate transcriptional unit for the *galE* and *rfe* 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 *rfe* 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 YeO8 O-antigen makes the location of the *galE* gene within the *rfb* region logical. To our knowledge, YeO8 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 YeO3 it is located in the *trs* locus involved in the LPS outer-core biosynthesis, upstream of the *gsk* gene (Skurnik *et al.*, 1995).

YeO8 O-antigen polymerization is *rfc*-dependent

The deletion from pLZ6020 of the 1.4 kb *Bam*HI fragment between 19 and 20.5 kb resulted in pLZ6010, leading to the identification of a sixth phenotypic class that expresses SR-LPS, 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 by 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 *Bam*HI fragment is able to transcomplement the *rfc* defect. These data indicate that the C-terminal 57 amino acids of the YeO8 Rfc protein may not be essential for the function of the polymerase. The biosynthesis of O-antigens in YeO8 and YeO3 is different. The YeO3 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). We show here that the YeO8 O-antigen polymerization is *rfc*-dependent and the pathway may follow the general scheme in which the O-antigen is synthesized first as units on a lipid carrier, which, once complete, are translocated to the periplasmic face of the plasma membrane, polymerized by Rfc and ligated to a preformed lipid-A-core acceptor. After ligation the completed LPS molecule is translocated to the cell surface by an unknown mechanism.

The amino acid sequence of Rfc shows significant similarity to both FepE (ferric enterobactin transport protein, 25 % identity) of *E. coli* (Table 2) and to putative Cld (O-antigen chain length determinant, 36 % identity) of *Yersinia pseudotuberculosis* (Table 2, Fig. 4d). It is also

Fig. 5. Multiple alignment of hydrophobicity plots. The hydrophobicity profiles of Rfc proteins were generated using the program PEPPLOT (GCG). The scale on top indicates amino acid residues. (a) YeO8 Rfc; (b) *Shig. dysenteriae* Rfc; (c) *Shig. flexneri* Rfc; (d) group B *S. enterica* Rfc.

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 YeO8 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 YeO8 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 YeO8 *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 YeO8 *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 YeO8

In *E. coli* and *S. enterica*, the *rfb* region is closely linked to the *his* 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 adenylate kinase and ferredoxin, 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 YeO8 *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 (Ypsth) 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 YeO8 *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*_{YeO8} and in the *trs*_{YeO3} operon but not in *rfb*_{Ypsth}. *rfc* is located between *galE* and *gsk* in *rfb*_{YeO8}. At the same location there is nothing in the *trs* operon, while in *rfb*_{Ypsth} 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 SØ874 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

- Adhya, S. (1989). The galactose operon. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, vol. 1, pp. 1503–1512. Edited by F. C. Neidhardt, J. L. Ingraham, K. Brooks Low, B. Magasanik, M. Schaechter & H. E. Umbarger. Washington, DC: American Society for Microbiology.
- Al-Hendy, A., Toivanen, P. & Skurnik, M. (1991a). The effect of growth temperature on the biosynthesis of *Yersinia enterocolitica* O:3 lipopolysaccharide: temperature regulates the transcription of the *rfb* but not of the *rfa* region. *Microb Pathog* 10, 81–86.
- Al-Hendy, A., Toivanen, P. & Skurnik, M. (1991b). Expression cloning of *Yersinia enterocolitica* O:3 *rfb* gene cluster in *Escherichia coli* K12. *Microb Pathog* 10, 47–59.
- Al-Hendy, A., Toivanen, P. & Skurnik, M. (1992). Lipopolysaccharide O side chain of *Yersinia enterocolitica* O:3 is an essential virulence factor in orally infected murine model. *Infect Immun* 60, 870–875.
- Appleyard, R. K. (1954). Segregation of new lysogenic types during growth of doubly lysogenic strain derived from *Escherichia coli* K-12. *Genetics* 39, 440–447.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, O. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1989). *Current Protocols in Molecular Biology*. New York: John Wiley.
- Bachmann, B. J. (1990). Linkage map of *Escherichia coli* K-12, edition 8. *Microbiol Rev* 54, 130–197.
- Brenner, D. J. (1978). Characterization and identification of *Enterobacteriaceae* by DNA hybridization. In *Progress in Clinical Pathology*, pp. 71–117. Edited by M. Stefanini. New York: Grune & Stratton.
- Bronner, D., Clarke, B. R. & Whitfield, C. (1994). Identification of an ATP-binding cassette transport system required for translocation of lipopolysaccharide O-antigen side-chains across the cytoplasmic membrane of *Klebsiella pneumoniae* serotype O1. *Mol Microbiol* 14, 505–519.
- Brown, P. K., Romana, L. K. & Reeves, P. R. (1991). Cloning of the *rfb* gene cluster of a group C2 *Salmonella* strain: comparison with the *rfb* regions of groups B and D. *Mol Microbiol* 5, 1873–1881.
- Collins, L. V. & Hackett, J. (1991). Molecular cloning, characterization, and nucleotide sequence of the *rfc* gene, which encodes an O-antigen polymerase of *Salmonella typhimurium*. *J Bacteriol* 173, 2521–2529.
- DeBruijn, F. J. & Lupski, J. R. (1984). The use of transposon Tn5 mutagenesis in the rapid generation of correlated physical and genetic maps of DNA segments cloned into multicopy plasmids – a review. *Gene* 27, 131–149.
- Devereux, J., Haeblerli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12, 387–395.

- Gripenberg-Lerche, C., Skurnik, M., Zhang, L., Söderström, K. & Toivanen, P. (1994). Role of YadA in arthritogenicity of *Yersinia enterocolitica* serotype O:8; experimental studies in rat. *Infect Immun* **62**, 5568–5575.
- Grosjean, H. & Fiers, W. (1982). Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* **18**, 199–209.
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557–580.
- Harr, R., Häggström, M. & Gustafsson, P. (1983). Search algorithm for pattern match analysis of nucleic acid sequences. *Nucleic Acids Res* **11**, 2943–2957.
- Harr, R., Fällman, P., Häggström, M., Wahlström, L. & Gustafsson, P. (1985). GENEUS, a computer system for DNA and protein sequence analysis containing information retrieval system for the EMBL data library. *Nucleic Acids Res* **14**, 273–284.
- Helander, I. M. (1985). Isolation and electrophoretic analysis of bacterial lipopolysaccharides. In *Enterobacterial Surface Antigens: Methods for Molecular Characterization*, pp. 263–274. Edited by T. K. Korhonen, E. A. Dawes & P. H. Mäkelä. New York: Elsevier Science Publishers (Biomedical Division).
- Hoffman, J. & Lindberg, B. (1980). Structural studies of the O-specific side-chains of the lipopolysaccharide from *Yersinia enterocolitica* Ye 128. *Carbohydr Res* **78**, 212–214.
- Jann, K. & Jann, B. (1984). Structure and biosynthesis of O-antigen. In *Handbook of Endotoxin*, pp. 138–186. Edited by E. T. Reitschel. Amsterdam: Elsevier Science.
- Jiang, X.-M., Neal, B., Santiago, F., Lee, S. J., Romana, L. K. & Reeves, P. R. (1991). Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). *Mol Microbiol* **5**, 695–713.
- Joiner, K. A. (1985). Studies on the mechanism of bacterial resistance to complement-mediated killing and on the mechanism of action of bactericidal antibody. *Curr Top Microbiol Immunol* **121**, 99–133.
- Keenleyside, W., Perry, M., Maclean, L., Poppe, C. & Whitfield, C. (1994). A plasmid-encoded *rfbO*:54 gene cluster is required for biosynthesis of the O:54 antigen in *Salmonella enterica* serovar borreza. *Mol Microbiol* **11**, 437–448.
- Kessler, A. C., Brown, P. K., Romana, L. K. & Reeves, P. R. (1991). Molecular cloning and genetic characterization of the *rfb* region from *Yersinia pseudotuberculosis* serogroup IIA, which determines the formation of the 3,6-dideoxyhexose abequose. *J Gen Microbiol* **137**, 2689–2695.
- Kessler, A. C., Haase, A. & Reeves, P. R. (1993). Molecular analysis of the 3,6-dideoxyhexose pathway genes of *Yersinia pseudotuberculosis* serotype IIA. *J Bacteriol* **175**, 1412–1422.
- Klena, J. D. & Schnaitman, C. A. (1993). Function of the *rfb* gene cluster and *rfe* gene in the synthesis of O antigen by *Shigella dysenteriae*. *Mol Microbiol* **9**, 393–402.
- Krauss, J. H., Weckesser, J. & Mayer, H. (1988). Electrophoretic analysis of lipopolysaccharides of purple nonsulfur bacteria. *Int J Syst Bacteriol* **38**, 157–163.
- Lee, S. J., Romana, L. K. & Reeves, P. R. (1992). Sequence and structural analysis of the *rfb* (O antigen) gene cluster from a group C1 *Salmonella enterica* strain. *J Gen Microbiol* **138**, 1843–1855.
- Markovitz, A. (1977). Genetics and regulation of bacterial capsular polysaccharide biosynthesis and radiation sensitivity. In *Surface Carbohydrates of the Prokaryotic Cell*, pp. 415–462. Edited by I. W. Sutherland. New York: Academic Press.
- Maskell, D. J., Szabo, M. J., Butler, P. D., Williams, A. E. & Moxon, E. R. (1991). Molecular analysis of a complex locus from *Haemophilus influenzae* involved in phase-variable lipopolysaccharide biosynthesis. *Mol Microbiol* **5**, 1013–1022.
- Miller, V. L. & Mekalanos, J. J. (1988). A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. *J Bacteriol* **170**, 2575–2583.
- Morales, V. M., Bäckman, A. & Bagdasarian, M. (1991). A series of wide-host-range low-copy-number vectors that allow direct screening for recombinants. *Gene* **97**, 39–47.
- Morona, R., Mavris, M., Fallarino, A. & Manning, P. A. (1994). Characterization of the *rfc* region of *Shigella flexneri*. *J Bacteriol* **176**, 733–747.
- Naide, Y., Nikaido, H., Mäkelä, P. H., Wilkinson, R. G. & Stocker, B. A. D. (1965). Semirough strains of *Salmonella*. *Proc Natl Acad Sci USA* **53**, 147–153.
- Norlander, J., Kempe, T. & Messing, J. (1983). Construction of improved M13 vectors using oligodeoxyribonucleotide-directed mutagenesis. *Gene* **26**, 101–106.
- Portnoy, D. A., Moseley, S. L. & Falkow, S. (1981). Characterization of plasmids and plasmid-associated determinants of *Yersinia enterocolitica* pathogenesis. *Infect Immun* **31**, 775–782.
- Reeves, P. (1994). Biosynthesis and assembly of lipopolysaccharide. In *New Comprehensive Biochemistry: Bacterial Cell Wall*, vol. 27, pp. 281–316. Edited by J.-M. Ghuysen & R. Hakenbeck. Amsterdam: Elsevier Science.
- Robertson, B. D., Frosch, M. & van Putten, J. P. M. (1993). The role of *galE* in the biosynthesis and function of gonococcal lipopolysaccharide. *Mol Microbiol* **8**, 891–901.
- Robertson, B. D., Frosch, M. & van Putten, J. P. M. (1994). The identification of cryptic rhamnose biosynthesis genes in *Neisseria gonorrhoeae* and their relationship to lipopolysaccharide biosynthesis. *J Bacteriol* **176**, 6915–6920.
- Sanderson, K. E., Hessel, A. & Rudd, K. E. (1995). Linkage map of *Salmonella typhimurium*. *Microbiol Rev* **59**, 241–303.
- Schnaitman, C. A. & Klena, J. D. (1993). Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol Rev* **57**, 655–682.
- Scholz, P., Haring, V., Scherzinger, E., Lurz, R., Bagdasarian, M. M., Schuster, H. & Bagdasarian, M. (1985). Replication determinants of the broad host-range plasmids RSF1010. In *Plasmids in Bacteria*, pp. 243–259. Edited by D. R. Helinski, S. N. Cohen, D. B. Clewell, D. A. Jackson & A. Hollaender. New York: Plenum Publishing.
- Simon, R., Priefer, U. & Pühler, A. (1983). A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in Gram-negative bacteria. *Biotechnology* **1**, 784–791.
- Skurnik, M., Venho, R., Toivanen, P. & Al-Hendy, A. (1995). A novel locus of *Yersinia enterocolitica* O:3 involved in lipopolysaccharide core biosynthesis. *Mol Microbiol* **17**, 575–594.
- Stevenson, G., Neal, B., Liu, D., Hobbs, M., Packer, N. H., Batley, M., Redmond, J. W., Lindquist, L. & Reeves, P. (1994). Structure of the O-antigen of *Escherichia coli* K-12 and the sequence of its *rfb* gene cluster. *J Bacteriol* **176**, 4144–4156.
- Stevenson, G., Kessler, A. & Reeves, P. (1995). A plasmid-borne O-antigen chain length determinant and its relationship to other chain length determinants. *FEMS Microbiol Lett* **125**, 23–30.
- Stinavage, P., Artin, L. E. & Spitznagel, J. K. (1989). O antigen and lipid A phosphoryl groups in resistance of *Salmonella typhimurium* LT-2 to nonoxidative killing in human polymorphonuclear neutrophils. *Infect Immun* **57**, 3894–3900.

- Tomshich, S. V., Gorshkova, R. P. & Ovodov, Y. S. (1987). Structural studies on lipopolysaccharide from *Y. enterocolitica* serovar O:8. *Khim Priir Soedin*, 657–664.
- Wang, L., Romana, L. K. & Reeves, P. R. (1992). Molecular analysis of a *Salmonella enterica* group E1 *rfb* gene cluster: O antigen and the genetic basis of the major polymorphism. *Genetics* **130**, 429–443.
- Wartenberg, K., Knapp, W., Ahamed, N. M., Widemann, C. & Mayer, H. (1983). Temperature-dependent changes in the sugar and fatty acid composition of lipopolysaccharides from *Yersinia enterocolitica* strains. *Zentralbl Bakteriol Abt Orig Hyg* **253**, 523–530.
- Whitfield, C. (1995). Biosynthesis of lipopolysaccharide O antigens. *Trends Microbiol* **3**, 178–185.
- Whitfield, C. & Valvano, M. A. (1993). Biosynthesis and expression of cell-surface polysaccharides in Gram-negative bacteria. *Adv Microb Physiol* **35**, 135–246.
- Yuasa, R., Nakane, K. & Nikaïdo, H. (1970). Structure of cell wall lipopolysaccharide from *Salmonella typhimurium*: structure of lipopolysaccharide from a semirough mutant. *Eur J Biochem* **15**, 63–71.
- Zhang, L. & Skurnik, M. (1994). Isolation of an R⁻M⁺ mutant of *Yersinia enterocolitica* serotype O:8 and its application in construction of rough mutants utilizing mini-Tn5 derivatives and lipopolysaccharide-specific phage. *J Bacteriol* **176**, 1756–1760.
- Zhang, L., Al-Hendy, A., Toivanen, P. & Skurnik, M. (1993). Genetic organization and sequence of the *rfb* gene cluster of *Y. enterocolitica* serotype O:3: similarities to the dTDP-L-rhamnose biosynthesis pathway of *Salmonella* and to the bacterial polysaccharide transport systems. *Mol Microbiol* **9**, 309–321.
-

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