Genetic organization of the O7-specific lipopolysaccharide biosynthesis cluster of Escherichia coli VW187 (O7:K1)

Cristina L. Marolda,1 Mario F. Feldman1,2 and Miguel A. Valvano1

Author for correspondence: Miguel A. Valvano. Tel: +1 519 661 3996. Fax: +1 519 661 3499. e-mail: mvalvano@julian.uwo.ca

In previous studies the authors cloned and characterized the DNA sequence of the regions at both ends of the O7-specific lipopolysaccharide (LPS) biosynthesis cluster of Escherichia coli VW187 (O7:K1), and identified the biosynthetic genes for dTDP-rhamnose and GDP-mannose, as well as one of the candidate glycosyltransferases. In this work the complete DNA sequence of a 6-9 kb intervening region is presented. Seven new ORFs were identified. All the functions required for the synthesis and transfer of the O7 LPS were assigned on the basis of complementation experiments of transposon insertion mutants, and amino acid sequence homology to proteins involved in LPS synthesis of other bacteria. Of the seven ORFs, two encoded membrane proteins that were homologous to the O-antigen translocase (Wzx) and polymerase (Wxy), two were involved in the biosynthesis of dTDP-N-acetylviosamine, and the remaining three showed homologies to sugar transferases. The O antigen chain length regulator gene wzz was also identified in the vicinity of the O7 polysaccharide cluster. O7-specific DNA primers were designed and tested for serotyping of O7 E. coli strains.

Keywords: Escherichia coli, wbb (rfb) gene cluster, lipopolysaccharide (LPS), O antigen, PCR serotyping

INTRODUCTION

Lipopolysaccharide (LPS) is one of the major components of the bacterial outer membrane (Nikaido & Vaara, 1985). It is an amphipathic molecule composed of lipid A and a core oligosaccharide, and in many species of bacteria also has a polysaccharide chain termed the O-specific antigen (Whitfield & Valvano, 1993). The O antigen is a major contributor to the virulence of many pathogenic bacteria by promoting the activation of complement and formation of the complement membrane attack complex away from its site of insertion in the outer membrane (Joiner, 1988).

Therefore, the understanding of its mechanism of synthesis may unravel new targets for antimicrobial therapy that may contribute to the control of infections by Gram-negative bacteria.

The biosynthesis of LPS is very complex and requires the activity of many genes, most of which are clustered in different regions of the chromosomal map. The synthesis of the O antigen involves the formation of a subunit that is polymerized while remaining attached to undecaprenyl phosphate, and later transferred to the lipid A-core, which is assembled by a separate pathway (Raetz, 1996). Several O antigen gene clusters have been cloned and sequenced (Reeves et al., 1996). During the past 10 years, our laboratory has used the O7 antigen genes of E. coli O7:K1 strains as a model system to investigate the biosynthesis, assembly and regulation of O antigens. Previous work by other investigators has shown that the presence of O7 in E. coli K1 strains contributes to the overall virulence of these micro-organisms, commonly found in extraintestinal infections in humans, especially neonatal meningitis and urinary tract infections (Pluschke & Achtman, 1984).

Abbreviations: ECA, enterobacterial common antigen; FucN, 4-amino-4,6-dideoxy-D-galactose; FucNac, 4-acetamido-4,6-dideoxy-D-galactose; GicNac, N-acetylgalcosamine; Vio, viosamine; VioNAc, N-acetylviosamine (4-acetamido-4,6-dideoxyglucose).

The GenBank accession number for the DNA sequence reported in this paper is AF125322.
The O7 antigen subunit contains five sugars, N-acetylglucosamine (GlcNAc), galactose, mannose, 4-acetamido-4,6-dideoxyglucose (N-acetylviomosamine; ViomNAC) and rhamnose, and its chemical structure has been elucidated (Fig. 1; L’vov et al., 1984). We have cloned and characterized by transposon mutagenesis two cosmids encoding O7 LPS in E. coli K-12 strains (Marolda et al., 1990; Valvano & Crosa, 1989). These laboratory strains usually lack O antigen but produce a complete lipid A-core that serves as an acceptor for O-specific LPS if the genes for its synthesis are supplied in trans (Valvano, 1992). We discovered that the synthesis of the O7 subunit is initiated in the inner membrane by the activity of the weca (rfe) gene product (Alexander & Valvano, 1995), which is an enzyme responsible for the attachment of GlcNAc-1-phosphate to undecaprenyl phosphate via a phosphodiester linkage. weca is part of a cluster of genes involved in the synthesis of enterobacterial common antigen (ECA) that maps outside the O7 LPS gene cluster (Rick & Silver, 1996). We also characterized molecularly and biochemically four genes of the O7 biosynthesis gene cluster that are responsible for the synthesis of dTDP-rhamnose (Marolda & Valvano, 1995), and two genes involved in the GDP-mannose pathway (Marolda & Valvano, 1993). The genes for the dTDP-rhamnose and GDP-mannose pathways are located at each end of the O7 LPS cluster. In this work, we report the complete characterization of the intervening sequence, revealing the existence of two genes for the dTDP-VioNAC pathway, four presumptive sugar transferases for the assembly of the O7 subunit, the O antigen translocate gene, and the O antigen polymerase gene. The sequence information was also used to develop a pair of DNA primers that provide a specific detection of O7 LPS genes by PCR.

METHODS

Bacterial strains, plasmids and materials. Laboratory strains and plasmids used in this study are described in Table 1. Clinical isolates of different O serogroups were from our laboratory collection. Bacteria were cultured in Luria broth supplemented with antibiotics, as appropriate, at the following final concentrations: ampicillin, 100 µg ml⁻¹; streptomycin, 80 µg ml⁻¹; tetracycline, 20 µg ml⁻¹. Chemicals and antibiotics were obtained from Sigma or Roche Diagnostics. TaqI and Pwo DNA polymerases, restriction enzymes and T4 DNA ligase were purchased from Roche Diagnostics. Oligonucleotide primers were purchased from Procyon/Vetrogen.

Recombinant DNA methods. Isolation and electrophoresis of plasmid DNA were carried out as described elsewhere (Marolda et al., 1990; Valvano & Crosa, 1984). Chromosomal DNA for the PCR reactions was prepared using the InstaGene Matrix from Bio-Rad. Transformations were carried out by the calcium chloride method (Cohen et al., 1972) or by electroporation with a Gene Pulser apparatus (Bio-Rad).

DNA sequencing analysis. DNA was sequenced by the dyeoxy method (Sanger et al., 1977) modified for use with the Pharmacia T7 DNA Sequencing Kit. Overlapping fragments spanning a 6·9 kb central region of the O7 LPS gene cluster (Fig. 2) were cloned into either pGEM3 or pGEM4. The recombinant plasmids pCM29, pCM37, pCM38, pCM39, pCM44, pCM59, pCM93, pCM111 and pLD2 (Fig. 2, Table 1) were sequenced with T7 or SP6 promoter primers as sequencing end-points. Nested deletions from pCM111, pCM29, pCM38, pCM39, pCM44 and pLD2 (Fig. 2, Table 1) were obtained by the ExoIII endonuclease method (Henikoff, 1984), and used to complete gaps and confirm the DNA sequence of some regions. The junctions of the pJHCV32::Tn3HoH1 insertions 21, 47, 58, 128 and 136 (Fig. 2, Table 1) were sequenced using a specific primer annealing at one of the ends of the transposon (Marolda & Valvano, 1993).

DNA and protein sequence analysis. DNA sequences were analysed using programs of the Wisconsin Package Version 9.0 (Genetics Computer Group). Amino acid sequence alignments were performed using CLUSTAL W (Thompson et al., 1994). Transmembrane domains were predicted using the dense alignment surface method (Cserzo et al., 1994). Insertions 21, 47, 58, 128 and 136 (Fig. 2, Table 1) were sequenced using a specific primer annealing at one of the ends of the transposon (Marolda & Valvano, 1993).

PCR. Plasmids pMF21 (carrying weca<sup>o</sup>) and pMF25 (carrying weca<sup>p</sup>) were constructed by amplification of the coding regions of these genes using primers containing unique restriction endonuclease sites to facilitate their cloning into the expression vector pEXT21 (Dykhsoorn et al., 1996). Primers 5′-agaagctctagcataaataacatatcggcg-3′ (SacI site underlined) and 5′-tagctctagtcgaga-3′ (BamHI site underlined), and primers 5′-agctctagtcgcataatatttttagcag-3′ (BamHI site underlined) and 5′-taagctctagtcgactaagc-3′ (HindIII site underlined) were used to amplify weca<sup>o</sup> and weca<sup>p</sup>, respectively. In both cases PCR amplifications were carried out with Pwo since this enzyme has a high fidelity. PCR products were digested with either SacI/BamHI or BamHI/HindIII, gel purified, and ligated to pEXT21 DNA, which was also digested with the same combination of enzymes in each case.

Chromosomal DNA from strains with different O serogroups was used as a template for the PCR reaction with weca<sup>o</sup>- and galf-specific primers. The primers were as follows: weca, 5′-ctattgactcagttct-3′ (located 50 bases downstream of the initiation codon) and 5′-agaattgtcgcagcag-3′ (located 927 bases from the initiation codon, on the complementary strand); galf, 5′-ccggagttcgagtagaatatcgc-3′ (located at the initiation codon) and 5′-ccggagttcctctgta-3′ (located at the stop codon, on the complementary strand). PCR reactions were carried out with TaqI polymerase with an annealing temperature of 55 °C for both primer pairs, using 28 amplification cycles. Ten-microlitre aliquots from the PCR
All transcribed from the promoter located upstream of the region. Therefore, it is likely that the promoter sites were recognized along the sequenced E. coli genome. The putative ribosome-binding site and start codon of the previous gene were very closely spaced, as evidenced by the location of the termination codon of the previous gene between ORFs 5, 6, 7, 8, and 9 within the wbbD and rmlC gene designations. Sequence analysis of the region located between ORF5 and ORF9 (wbbD) and ORF10 (wzy). Using the E. coli σ70 promoter consensus sequence, no putative promoter sites were recognized along the sequenced region. Therefore, it is likely that the wbbE07 genes are all transcribed from the promoter located upstream of the cluster, between galF and rmlB (Fig. 2; Marolda & Valvano, 1998), although the existence of a weak promoter located upstream of wzy cannot be ruled out. All ORFs had a G+C content between 28 and 34 mol%, which is much lower than 51 mol%, the mean G+C content of E. coli (Ochman & Lawrence, 1996). This observation is consistent with the suggested mechanism of acquisition for O-poly saccharide genes via horizontal transfer (Reeves, 1992). The functions of the genes within the analysed region were assigned according to the analysis of the phenotypes of transposon insertion mutants and the homologies with other known proteins, as discussed below. The recently proposed nomenclature for bacterial polysaccharide genes via horizontal transfer (Reeves, 1992).

### Table 1. Laboratory strains and plasmids used in this study

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Relevant properties*</th>
<th>Source/reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLM4</td>
<td>lacZ trp Δ(sbcB–rfd) upp rel rpsL recA</td>
<td>Marolda &amp; Valvano (1993)</td>
</tr>
<tr>
<td>DH5</td>
<td>F' Φ80lacZ M15 endA recA hsdR (rK mC) supE thi gyrA relA Δ(lacZYA-argF)U169</td>
<td>Laboratory stock</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pEXT21</td>
<td>Cloning vector, SpR</td>
<td>Dykxhoorn et al. (1996)</td>
</tr>
<tr>
<td>pCM29</td>
<td>5.3 kb HindIII fragment cloned into pGEM3, ApR</td>
<td>This work</td>
</tr>
<tr>
<td>pCM37</td>
<td>5.2 kb HindIII fragment cloned into pGEM3, ApR</td>
<td>This work</td>
</tr>
<tr>
<td>pCM38</td>
<td>1.3 kb EcoRI fragment cloned into pGEM3, ApR</td>
<td>This work</td>
</tr>
<tr>
<td>pCM39</td>
<td>2.0 kb EcoRI fragment cloned into pGEM3, ApR</td>
<td>This work</td>
</tr>
<tr>
<td>pCM44</td>
<td>2.4 kb EcoRI fragment cloned into pGEM3, ApR</td>
<td>This work</td>
</tr>
<tr>
<td>pCM59</td>
<td>2.2 kb EcoRV fragment cloned into pGEM3, ApR</td>
<td>This work</td>
</tr>
<tr>
<td>pCM93</td>
<td>5.4 kb EcoRI fragment cloned into pGEM4, ApR</td>
<td>This work</td>
</tr>
<tr>
<td>pCM111</td>
<td>5.6 kb EcoRI fragment cloned into pSF6, SpR</td>
<td>This work</td>
</tr>
<tr>
<td>pGEM3</td>
<td>Cloning and sequencing vector, ApR</td>
<td>Promega Biotech</td>
</tr>
<tr>
<td>pJHCV32</td>
<td>Cosmid clone containing wbE07, TcR</td>
<td>Valvo &amp; Cosa (1989)</td>
</tr>
<tr>
<td>pJHCV32::TnHoHo1-21</td>
<td>wzy::TnHoHo1-21, TcR ApR</td>
<td>Marolda et al. (1990)</td>
</tr>
<tr>
<td>pJHCV32::TnHoHo1-47</td>
<td>wzy::TnHoHo1-47, TcR ApR</td>
<td>Marolda et al. (1990)</td>
</tr>
<tr>
<td>pJHCV32::TnHoHo1-58</td>
<td>wzy::TnHoHo1-58, TcR ApR</td>
<td>Marolda et al. (1990)</td>
</tr>
<tr>
<td>pJHCV32::TnHoHo1-128</td>
<td>wzy::TnHoHo1-128, TcR ApR</td>
<td>Marolda et al. (1990)</td>
</tr>
<tr>
<td>pJHCV32::TnHoHo1-134</td>
<td>wzy::TnHoHo1-134, TcR ApR</td>
<td>Marolda et al. (1990)</td>
</tr>
<tr>
<td>pJHCV32::TnHoHo1-136</td>
<td>wzy::TnHoHo1-136, TcR ApR</td>
<td>Marolda et al. (1990)</td>
</tr>
<tr>
<td>pJHCV32::TnHoHo1-145</td>
<td>wzy::TnHoHo1-145, TcR ApR</td>
<td>Marolda et al. (1990)</td>
</tr>
<tr>
<td>pLD2</td>
<td>2.3 kb HindIII fragment cloned into pGEM3, ApR</td>
<td>This work</td>
</tr>
<tr>
<td>pMF21</td>
<td>1.5 kb fragment containing wzy cloned into pEXT21, SpR</td>
<td>This work</td>
</tr>
<tr>
<td>pMF25</td>
<td>1.16 kb fragment containing wzy cloned into pEXT21, SpR</td>
<td>This work</td>
</tr>
<tr>
<td>pSF6</td>
<td>Cloning vector, SpR</td>
<td>Selvaraj et al. (1984)</td>
</tr>
</tbody>
</table>

* Ap, ampicillin; Sp, spectinomycin; Tc, tetracycline.

**Results**

**DNA sequence analysis of the central region of the wbE07 gene cluster**

Sequence analysis of the region located between rmlC and wbbD identified seven ORFs, all transcribed in the same direction (Fig. 2, Table 2). ORF5, 6, 7, 8, and 9 were very closely spaced, as evidenced by the location of the termination codon of the previous gene between or within the putative ribosome-binding site and start codon of the next gene (Table 2). A space of 140 bp was found between ORF9 (wbbB) and ORF10 (wzy). Using the E. coli σ70 promoter consensus sequence, no putative promoter sites were recognized along the sequenced region. Therefore, it is likely that the wbE07 genes are all transcribed from the promoter located upstream of the cluster, between galF and rmlB (Fig. 2; Marolda & Valvano, 1998), although the existence of a weak promoter located upstream of wzy cannot be ruled out. All ORFs had a G+C content between 28 and 34 mol% (Table 2), which is much lower than 51 mol%, the mean G+C content of E. coli (Ochman & Lawrence, 1996). This observation is consistent with the suggested mechanism of acquisition for O-polysaccharide genes via horizontal transfer (Reeves, 1992). The functions of the genes within the analysed region were assigned according to the analysis of the phenotypes of transposon insertion mutants and the homologies with other known proteins, as discussed below. The recently proposed nomenclature for bacterial polysaccharide genes (Reeves et al., 1996) was utilized for gene designations.

**O-antigen translocase (wzx).** The deduced amino acid sequence of ORF5 predicts a polypeptide of 53.8 kDa, which has 12 transmembrane segments, as indicated by the dense alignment surface method. This ORF shows a low level of homology with other cytoplasmic membrane proteins such as ExoT (accession S40176), GumJ...
Fig. 2. Partial restriction map and genetic organization of the DNA region containing the wbE cluster and flanking genes. The thick line indicates the sequenced area presented in this work and from previous reports (Marolda & Valvano, 1993, 1995). Genes for the O7 biosynthesis cluster are indicated by open boxes, as follows: rmlBDAC, dTDP-rhamnose synthesis (Marolda & Valvano, 1995); wzx, O-antigen translocase; vioAB, dTDP-N-acetylviosamine synthesis; wbbABCD, putative glycosyltransferases; wzy, O-antigen polymerase; manBC, GDP-mannose synthesis (Marolda & Valvano, 1993); wzz, O-chain length regulator. Numbers in the boxes correspond to the different ORFs identified in this study. Numbered bars underneath the boxes indicate the location of the transposon insertions in pJHV32 (Table 1). Hatched boxes indicate flanking genes not involved in O7 LPS synthesis: galF, UDP-glucose pyrophosphorylase regulator (Marolda & Valvano, 1996); gnd, gluconate-6-phosphate dehydrogenase; ugd, UDP-glucuronic acid dehydrogenase; hisEIF, histidine biosynthesis. The boundaries of plasmids used in this study (Table 1) are also indicated. S, SalI; E, EcoRI; H, HindIII; Ev, EcoRV; K, KpnI.

Table 2. Characteristics of the ORFs found

<table>
<thead>
<tr>
<th>ORF</th>
<th>Gene</th>
<th>Translation start*</th>
<th>Spacing†</th>
<th>Position (nt)‡</th>
<th>Size (aa)</th>
<th>Mol. mass (kDa)</th>
<th>pI</th>
<th>G + C (mol%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>wzx</td>
<td>ttGAGaatgtaattATGat</td>
<td>11</td>
<td>4788–6215</td>
<td>476</td>
<td>53.8</td>
<td>9.9</td>
<td>29.8</td>
</tr>
<tr>
<td>6</td>
<td>vioA</td>
<td>taAGAgataaatATGaa</td>
<td>6</td>
<td>6221–7333</td>
<td>371</td>
<td>41.6</td>
<td>8.1</td>
<td>34.3</td>
</tr>
<tr>
<td>7</td>
<td>vioB</td>
<td>gggtAGataaatATGga</td>
<td>7</td>
<td>7338–7913</td>
<td>192</td>
<td>20.6</td>
<td>4.9</td>
<td>34.3</td>
</tr>
<tr>
<td>8</td>
<td>wbbA</td>
<td>attGGAtattGTAaa</td>
<td>7</td>
<td>7913–8656</td>
<td>248</td>
<td>29.5</td>
<td>9.1</td>
<td>28.6</td>
</tr>
<tr>
<td>9</td>
<td>wbbB</td>
<td>tgtGGGatattATGga</td>
<td>9</td>
<td>8659–9318</td>
<td>220</td>
<td>25.6</td>
<td>6.5</td>
<td>28.6</td>
</tr>
<tr>
<td>10</td>
<td>wzy</td>
<td>agAGGAagattattATGag</td>
<td>13</td>
<td>9478–10626</td>
<td>383</td>
<td>43.7</td>
<td>9.8</td>
<td>29.5</td>
</tr>
<tr>
<td>11</td>
<td>wbbC</td>
<td>tttGCAGattatATGgg</td>
<td>12</td>
<td>10700–11743</td>
<td>348</td>
<td>39.6</td>
<td>7.6</td>
<td>30.4</td>
</tr>
</tbody>
</table>

* Potential ribosome-binding sequences and start codons (bold) are capitalized. The stop codon of the previous ORF (bold) is underlined.
† Spacing between putative ribosome-binding site sequence and start codon.
‡ Nucleotide positions correspond to the entire O7 gene cluster depicted in Fig. 2, as indicated in GenBank accession AF125322.

We previously mapped two transposon insertions to this gene, Tn3HoHo1-128 and Tn3HoHo1-145 (Fig. 2; Marolda et al., 1990). The insertion 128 was positioned after the first base of the ATA codon for the amino acid Ile157. A plasmid carrying this insertion was transformed into strain CLM4 and the O7 LPS expression was examined by Western immunoblotting. No O7-specific LPS bands were detected (Fig. 3, lane B), in
contrast to the control experiment using plasmid pJHCV32 (Fig. 3, lane A), which carries an almost intact O7 LPS cluster (Fig. 2), and directs the expression of low levels of O7 LPS (Marolda & Valvano, 1998). Very weak O7 LPS bands could be observed after overloading the gel, and by using chemoluminescence to develop the Western blot (data not shown). This suggested that the wzx gene is not involved in the formation of the O7 subunit and functions at a later step in the biosynthesis of the O antigen. These observations are consistent with a mutation in the O-antigen translocase, since O7 antigen subunits would be made but not reach the periplasmic side of the cytoplasmic membrane and therefore, resulting in a pronounced reduction of O7 polymer attached to the lipid A-core oligosaccharide. However, a similar phenotype could be due to polar effect of the Tn\(^{3}\) HoHo1-128 insertion on the downstream genes of wzx. In our experience, Tn3HoHo1 often gives rise to non-polar mutations, probably due to the presence of promoter within the large transposon element (Marolda et al., 1990; C. L. Marolda & M. A. Valvano, unpublished). Also, we have observed that in the E. coli K-12 strain DH5\(\alpha\), pJHCV32::Tn3HoHo1-128 directs the expression of a diminished but readily detectable amount of O7 polysaccharide (Marolda et al., 1990), possibly due to complementation of the translocase function by the wzx gene of the E. coli K-12 O antigen cluster, which is present in DH5\(\alpha\) and absent from CLM4. The wzx coding sequence has an abundance of codons atypical for E. coli, suggesting that the Wzx protein is poorly translated. Lukomski et al. (1996) showed that the wzx gene, which also contains an unusually high number of rare codons, cannot complement a wzx mutation when cloned in high-copy-number vectors. We then cloned the 5.6 kb EcoRI insert of pCM111 into the low-copy-number vector pSF6, resulting in plasmid pCM138 (Fig. 2). We also constructed plasmid pMF21 (Fig. 2), containing the wzx gene under the control of the lac promoter of pEXT21, another low-copy-number expression vector derived from pSF6 (Dykxhoorn et al., 1996). Both of these plasmids complemented the insertion 128 as indicated in Fig. 3, lanes C and D. These results confirm that the insertion 128 is non-polar on the downstream genes of the \(wzx_{E\text{c}O}\) cluster.

**Genes for the biosynthesis of dTDP-VioNAc.** ORFs 6 and 7 encode proteins with no apparent transmembrane regions, and therefore they are probably cytoplasmic enzymes involved in the synthesis or modification of nucleotide sugar precursors for the assembly of the O7 subunit. The genes manBC and rmlBDAC, which encode enzymes for the synthesis of GDP-mannose and dTDP-rhamnose, respectively, have been previously identified and characterized (Marolda & Valvano, 1993, 1995). We do not expect to find genes in the \(wzx_{E\text{c}O}\) cluster encoding enzymes for the synthesis of UDP-GlcNAc and UDP-Gal, since they map elsewhere in the bacterial chromosome (Berlyn, 1998). The only nucleotide sugar precursor of the O7 subunit for which there are no enzyme genes assigned is 4-acetamido-4,6-dideoxy-glucose (N-acetylviosamine; VioNAc). Acetamido-sugar-containing nucleotides have been isolated in the past from several strains of E. coli (Dietzler & Strominger, 1973). dTDP-4-acetamido-4,6-dideoxy-d-galactose was isolated from extracts of E. coli K-12 Y10, while dTDP-VioNAc was identified in extracts from E. coli B (Matsuhishi & Strominger, 1966). 4-Acetamido-4,6-dideoxygalactose, also known as 4-acetamidofucose (FucNAc), is a component of the ECA, widely dis-
The biosynthesis of acetamido sugar nucleotides requires dTDP-4-keto-6-deoxyglucose, an intermediate product in the dTDP-rhamnose pathway. By means of a pyridoxal-phosphate-dependent transamination, an amino group from L-glutamate is used to form dTDP-4-amino-6-deoxyhexoses. The stereospecificity of the transaminase determines whether the product has the N-gluc or N-galacto configuration (Matsuhashi & Strominger, 1966). These amino sugar nucleotides are then acetylated by specific enzymes that utilize acetyl-CoA, yielding the acetamido forms of the nucleotide sugars (Dietzler & Strominger, 1973). The predicted protein product of ORF6 is similar to the members of a family of enzymes catalysing sugar transaminations, which are required in the synthesis pathway of 2,6-, 3,6- and 4,6-dideoxyhexoses (Thorson et al., 1993). Amongst them, the protein encoded by ORF6 is homologous to the putative perosamine synthases from Vibrio cholerae (encoded by wbeE, accession S28471; Stroehler et al., 1995), and E. coli O157 (encoded by per, accession AF061251; Wang & Reeves, 1998), a gene of the ECA biosynthesis cluster of E. coli presumably involved in the synthesis of dTDP-FucNAc (wecE, accession g2367285; Rick & Silver, 1996), and the first ORF of a gene cluster thought to be involved in the biosynthesis of 4-aminoarabinose in Salmonella and E. coli (yfbE/ORF1, accession P77690; Gunn et al., 1998). We propose that ORF6 corresponds to the transaminase for the synthesis of dTDP-VioNAc, and should be designated as vioA.

The product of ORF7 is similar to several proteins in the databases, especially those of the NodL-LacA family of acetyltransferases. These proteins include, among others, WbeO from V. cholerae (accession S28479; Stroehler et al., 1995), WbbJ from E. coli K-12 (accession P37750; Yao & Valpan, 1994), CapG from Sinorhizobium meliloti (accession P28266; Beav & Kondorosi, 1992) and CatO from Agrobacterium tumefaciens (accession P23364; Tennenkei & Matzura, 1991). The proteins from this family are involved in the transfer of acetyl groups from acetyl-CoA and share a region of homology along the C-terminal 50–70 amino acids (Fig. 4). One of the main features of this common region is the presence of an hexapeptide motif, also termed the isoleucine patch (Dicker & Seetharam, 1992), which is present in some transferases (Vaara, 1982). Each motif starts with an isoleucine, leucine or valine residue and often contains a glycine in the second position. This repeating motif is present in all the proteins of this family (Fig. 4). We propose that the product of ORF7 is involved in the N-acetylation of dTDP-Vio to complete the formation of dTDP-VioNAc as indicated in Fig. 5, and have designated this gene as vioB.

**Fig. 4.** Alignment of the acetyltransferase homologues of VioB. Only the common region of 50 amino acids near the C-terminus in all these proteins is shown. Asterisks indicate the first residue of the isoleucine patch. Residues identical or similar in all sequences are indicated on a dark background. RHIME, R. (S.) mellioti; AGRTU, A. tumefaciens; ECOLI, E. coli; VCHOL, V. cholerae; STAAU, Staphylococcus aureus.

**Fig. 5.** Metabolic pathway and genes involved in the synthesis of dTDP-4-acetamido-4,6-dideoxyglucose (dTDP-N-acetylvi-amosine).
which mapped to this ORF and conferred a phenotype consistent with a defect in the polymerization of the O7 subunit. This was suggested by the presence of a single O7-specific band reacting with the O7 antiserum in the region of the gel corresponding to the migration of lipid A-core band with a single O antigen subunit attached (Fig. 3, lane E; Marolda et al., 1990). These results suggested that ORF10 corresponds to the O7 antigen polymerase, and it was designated as \( wzy \). We sequenced the end-points of each transposon insertion and positioned them within the codons for the amino acids Ile19 (insertion 136), Asp79 (insertion 58), Ser80 (insertion 21) and Val199 (insertion 47) of the \( wzy \) gene (Fig. 2). The putative \( wzy \) gene was cloned in the vector pEXT21, resulting in plasmid pMF25. Western immunoblots of LPS extracts from CLM4 cells containing pJHVC32::Tn3HoHo1-L36 and pMF25 showed the restoration of the O7 LPS phenotype as in LPS extracts from CLM4 containing pJHVC32 (Fig. 3, lanes F and G), and thus confirmed that ORF10 is the \( wzy \) gene.

**Glycosyltransferases (wbbABCD).** We expected to find four transferases since we had already established that \( wecA \) is required for the initiation of the synthesis of the O7 subunit (Alexander & Valvano, 1994). In a previous report, we identified \( wbbD \) (ORF275, Marolda & Valvano, 1993) as a putative glycosyltransferase gene. A BLAST search revealed that WbbD is homologous to a diverse family of bacterial glycosyltransferases, transferring sugar from UDP-glucose, UDP-N-acetylglactosamine, GDP-mannose or GDP-aborque to a range of substrates. The strongest homologies are with AmsE (accession Q46635) and ORF6 (accession S27582) from the \( lsg \) locus of *Haemophilus influenzae*. ORF6 of *H. influenzae* is involved in the synthesis of lipooligosaccharide (McLaughlin et al., 1992) but its specific function remains unknown. AmsE is involved in the synthesis of amylovoran (Bugert & Geider, 1995), an exopolysaccharide of *Erwinia amylovora*, which contains a repeating unit made of one glucuronic acid and four galactose residues decorated with pyruvate and acetate groups (Smith et al., 1990). Tentatively, we have assigned \( wbbD \) as a galactosyltransferase involved in adding galactose to GlcNac-P-P-undecaprenol. ORF11 has a strong conservation with mannosyltransferases, especially with WbaU (accession P26402), WbaW (accession X61917), WbdB (accession I76776) and ORF469 (accession AB010150), which are involved in the synthesis of O antigen in two *Salmonella enterica* serogroups, *E. coli* O9 and *E. coli* O8, respectively. These findings suggest that ORF11 encodes the \( \alpha \)-mannosyltransferase responsible for adding the third sugar to the O7 repeat (Fig. 1), and it has been designated as \( \text{wbbC} \). No significant homologies in the database were found for ORF8 and ORF9. In *E. coli* and *Salmonella* O antigen clusters that have been sequenced, the order of transferase genes is in inverse relation to the order of synthesis of the subunit. Therefore, we tentatively assigned ORF8 as \( \text{wbbA} \) and ORF9 as \( \text{wbbB} \), encoding the N-acetyllivosaminyl- and rhamnosyltransferases, respectively (Fig. 2).

**Identification of the O antigen chain regulator \( wzz \)**

Using transposon mutagenesis, we identified an insertion located outside the O7 LPS gene cluster that conferred a unimodal O antigen polysaccharide distribution (Fig. 3, lane I). This phenotype is consistent with a mutation in the \( wzz \) gene, which encodes a protein involved in the regulation of the O-side chain length (Batchelor et al., 1992). To confirm these results and determine the gene structure of the region between \( gnd \) and \( his \) we performed spotted sequencing using the DNA inserts in plasmids pCM37, pCM59 and pCM93 (Fig. 2). The results confirmed the existence of two genes in this region, \( ugd \) and \( wzz \), both highly homologous to the *E. coli* K-12 counterparts. \( ugd \) encodes the UDP-glucose dehydrogenase, which is an enzyme involved in the synthesis of UDP-glucuronic acid, a precursor for the synthesis of the colanic acid capsular polysaccharide (Stevenson et al., 1996). By restriction endonuclease analysis, the insertion 134 was mapped to approximately the middle of the \( wzz \) gene (Fig. 2), in agreement with the O7 LPS phenotype determined by pJHVC32::Tn3HoHo1-L36.

**Design of O7-specific PCR primers for serotyping**

The availability of the complete DNA sequence of the O7 cluster permitted us to design O7-specific primers. The sequence of \( wzz \) was targeted for this purpose since this gene has great variability among different O antigens. We designed a pair of \( wzz \) primers and used them in PCR reactions with chromosomal DNA templates prepared from *E. coli* strains of different O serogroups. As a control we used primers specific for the \( \text{galf} \) gene, which is highly conserved in most *E. coli* strains examined to date (Marolda & Valvano, 1996; C. L. Marolda & M. A. Valvano, unpublished results). The results of these experiments showed that the control strain VW187, as well as other O7 isolates from different clonal groups, were all PCR positive as evidenced by the detection of a 775 bp band (data not shown). In contrast, chromosomes from *E. coli* strains belonging to other O types, including O1, O2, O4, O6, O8, O14, O16, O18, O25, O28, O39, O39, O45, O68, O75, O78, O100, O111, O116, O124, O136, O138, O141 and O157, did not afford any PCR product. This was also the case for the laboratory *E. coli* strain B (strain ATCC 113301) and C (strain C1a), which do not form O antigen. A 900 bp fragment corresponding to the amplification of part of the \( \text{galf} \) gene was found in all cases, demonstrating that the lack of detection of O7-specific amplification products was not due to technical problems with the PCR reactions. On the basis of DNA–DNA hybridization, analysis of outer-membrane protein profiles, and LPS migration patterns, we have previously reported that *Shigella boydii* strains that had been classified as serotype O12 constitute two different clonal groups, SB1 and SB2 (Valvano & Marolda, 1991). The strains belonging to the SB1 clone produce LPS that is identical to O7 and they have genes highly homologous to the \( \text{wbb}_{\text{E.o}} \) genes in *E. coli* VW187 (Valvano

2491
& Marolda, 1991). PCR assays with the O7-specific primers confirmed these results since chromosomal DNA of the strain N9BC12 (SB1 clonal group) used as template gave an amplicon similar to that afforded by the chromosomal DNA of the other E. coli O7 strains. In contrast, chromosomal DNA of strains SB504-58 and SB123 belonging to the SB2 clonal group did not afford any PCR product. Therefore, we concluded that our primers are specific and can serve for the identification of strains carrying O7-specific LPS.

DISCUSSION

In the present work we completed the characterization of the \(\text{ub}_{E. coli}^{O7}\) cluster, and assigned functions to each of the identified genes. The cluster consists of 14 genes, six of which encode functions for the synthesis of GDP-mannose (Marolda & Valvano, 1993), dTDP-rhamnose (Marolda & Valvano, 1995) and dTDP-VioNAc. The synthesis of the latter utilizes some of the steps involved in the synthesis of dTDP-rhamnose since the precursor, dTDP-4-keto-6-deoxy-d-glucose, is used in both pathways (Marolda & Valvano, 1995; Matsuhashi & Strominger, 1964). Although at the present time we do not have direct biochemical evidence for the enzyme activities of VioA and VioB, the sequence alignments with homologous proteins provide strong support to the assignment of VioA as a nucleotide sugar amino-

transferase and VioB as an acetyltransferase. This is in good agreement with the reported biosynthetic pathway for 4-acetamido-4,6-dideoxyhexoses, which requires two enzymatic activities, and is initiated by an amination reaction followed by the N-acetylation of the amino group (Dietzler & Strominger, 1973; Matsuhashi & Strominger, 1966). A pair of genes encoding similar proteins is also found in other O polysaccharide gene clusters such as those of V. cholerae (Stroher et al., 1995) and E. coli O157 (Wang & Reeves, 1998). The O antigen in these strains contains the acetamido-dideoxy-sugar N-acetylpersamine. Also, the \(\text{uweE}\) gene product has a strong homology with VioA. This gene is located in the ECA biosynthesis cluster, and it is probably involved in the synthesis of the precursor dTDP-4-acetamido-4,6-dideoxygalactose (dTDP-FucNAc; Rick & Silver, 1996). Interestingly, the \(\text{uweD}\) gene located next to \(\text{uweE}\) encodes a polypeptide with a similar mass to that of the acetyltransferases but has no homologues in the databases. Since not all acetyltransferases belong to the NodL-LacA family, it is tempting to speculate that \(\text{uweD}\) may encode an acetyltransferase that completes the synthesis of the dTDP-FucNAc.

The genes for the synthesis of the remaining precursors of the O7 subunit, UDP-galactose and UDP-GlcNAc, were not expected to be found in the \(\text{ub}_{E. coli}^{O7}\) cluster. These genes are usually present in other locations of the E. coli chromosome, since these precursors are used in a variety of metabolic pathways, and they are also involved in the synthesis of other cell envelope molecules such as peptidoglycan, core oligosaccharide, colanic acid and ECA (Berlyn, 1998). These findings are in agreement with similar results obtained by other investigators suggesting that only O-antigen-specific genes are present in the \(\text{ub}\) clusters (Reeves et al., 1996).

The other genes identified in the \(\text{ub}_{E. coli}^{O7}\) cluster encode proteins involved in the assembly of the O7 subunit. The presence of \(\text{wzy}\), encoding a polymerase, confirms that the O7 antigen is synthesized by a \(\text{wzy}\)-dependent pathway, and therefore it is likely that O7 polymerization occurs on the periplasmic side of the inner membrane, as has been proposed for this class of O antigens (Whitfield, 1995). Consistent with this idea is the fact that the Wzy protein has predicted topological features similar to those described for the Shigella flexneri homologue, including a relatively large periplasmic loop (Daniels et al., 1998). The \(\text{ub}_{E. coli}^{O7}\) cluster also contains a \(\text{uwx}\) gene, which appears to play an as yet unidentified role in the translocation of the O7 antigen subunit across the inner membrane (Liu et al., 1996). An interesting observation made with \(\text{uwx}\) genes and their products is the lack of strong amino acid sequence similarities despite the apparent common function in the translocation of the O antigen subunit. Nevertheless, we assigned the \(\text{uwx}_{E. coli}^{O7}\) gene based on weak amino acid sequence similarities with other proteins also predicted to be involved in the translocation of oligosaccharides, as well as the predicted structural topology as an integral membrane protein with multiple transmembrane domains. More importantly, we observed that a non-polar transposon insertion in this gene (insertion 128) expressed a complete O7 LPS in E. coli K-12 strains such as DH5α, LE392 and HB101, albeit at a lower level than wild-type expression by the cosmid pJHCV32 (Marolda et al., 1990). This observation ruled out the possibility that the insertion 128 was in a gene for the synthesis or the transfer of an essential precursor of the O7 antigen subunit. Many E. coli K-12 strains cannot make O antigen because of an insertion in a rhamnosyltransferase gene (\(\text{rfb}_{50}\) mutation; Liu & Reeves, 1994). However, when this mutation is complemented, E. coli K-12 makes O16 LPS (Stevenson et al., 1994; Yao & Valvano, 1994), suggesting that the remaining genes of the O16 LPS cluster are functional. Since a \(\text{uwx}\) gene is also present in the \(\text{ub}_{E. coli}^{O16}\) cluster, we suggest that this gene can at least partially complement the phenotype of insertion 128. This explains the lack of significant amounts of O7 LPS in strain CLM4, which has a large deletion eliminating the O16 LPS gene cluster (Marolda & Valvano, 1993). The O7 and O16 subunits are structurally unrelated (L’vov et al., 1984; Stevenson et al., 1994) except for the fact that both contain GlcNAc, which is the first sugar added to undecaprenol phosphate (Alexander & Valvano, 1994; Yao et al., 1992). Preliminary evidence in our laboratories indicates that a cloned \(\text{uwx}_{E. coli}^{O16}\) gene can also complement the insertion 128 in strain CLM4, and provide the basis for a functional analysis of Wzx proteins (M. Feldman, C. L. Marolda, M. A. Monteiro, A. J. Parodi & M. A. Valvano, unpublished).

In a previous study we determined that the synthesis of the O7 subunit is initiated by the function of \(\text{uweA},\)
resulting in the transfer of GlcNAc-1-phosphate to undecaprenyl phosphate (Alexander & Valvano, 1994). We also determined that this step is very common in the synthesis of O polysaccharides containing GlcNAc in the O subunit (Alexander & Valvano, 1994). \textit{wecA} is the first gene of the biosynthesis cluster for ECA, which is present at a distant location in the chromosome with respect to the \textit{wb} cluster (Rick & Silver, 1996). Based on this information, we predicted that the \textit{wb}_{E07} cluster should contain genes for the remaining four glycosyltransferases to complete the synthesis of the O7 subunit. The galactosyl- and mannosyltransferase genes, \textit{wbbD} and \textit{wbbC}, respectively, were assigned based on homologies of the predicted amino acid sequences with known glycosyltransferases. The two remaining genes, \textit{wbbA} and \textit{wbbB}, are predicted to be involved in the transfer of VioNAc and rhamnose to complete the O7 subunit. However, no homologues were found in the database, and these assignments are tentative until we can confirm them biochemically, work already in progress in our laboratories.

The completion of the genetic characterization of the \textit{wb}_{E07} cluster also permitted us to design a set of primers for detection of O7-specific sequences. Using this primer set we were able to confirm our previous assignment of clonal groups in \textit{Shigella boydii} strains serotype 12, some of which produce an O antigen identical to O7 (Valvano & Marolda, 1991). We also showed that \textit{E. coli} B and C, a pair of O-antigen-deficient strains used in genetic studies like \textit{E. coli} K-12, do not contain O7-specific genes. Therefore, our primers may be useful, alone or in combination with other O antigen-specific primers, to develop a molecular diagnostic kit for determination of O types. Recently, it has been shown that the use of rapid genomic techniques for cloning and sequencing of \textit{wb} clusters is feasible (Wang & Reeves, 1998). This progress makes possible the rapid generation of a database of gene sequences that can be used for the design of specific primers for diagnostic purposes and epidemiological surveillance of \textit{E. coli} infections.

ACKNOWLEDGEMENTS

We thank Lydia Dafoe for technical help, and the colleagues referenced or mentioned in Table 1 for supplying strains and plasmids. M.F.F. was a visiting graduate student supported by a doctoral fellowship of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) and by grants from CONICET, ANPCT and UBACYT to A.J. Parodi. This work was supported by a grant from the Medical Research Council of Canada.

REFERENCES


Wang, L. & Reeves, P. R. (1998). Organization of the Escherichia
O7-specific LPS biosynthesis cluster of *E. coli*


Received 9 February 1999; revised 5 May 1999; accepted 17 May 1999.