The complete cps gene cluster from Streptococcus thermophilus NCFB 2393 involved in the biosynthesis of a new exopolysaccharide

Eva Almirón-Roig, Francis Mulholland, Michael J. Gasson and Annette M. Griffin

INTRODUCTION

Streptococcus thermophilus is a thermophilic lactic acid bacterium used as a starter culture in the production of yoghurt and Swiss or Italian-type cheeses. In yoghurt, the presence of exopolysaccharide (EPS) produced by strains of Str. thermophilus leads to an improvement in the viscosity and texture, an increased resistance to mechanical handling and a decreased susceptibility to ‘wheying-off’. The chemical structures of five different EPSs produced by different strains of Str. thermophilus have been determined (Bubb et al., 1997; Doco et al., 1990; Lémoine et al., 1997; Low et al., 1998). These were all found to be heteropolymers containing differing amounts of four sugars: glucose, galactose, rhamnose and N-acetylgalactosamine. In one case, fucose was also part of the repeat unit. The pathway for EPS biosynthesis in Str. thermophilus is postulated to resemble the wzy-dependent systems reported for O-antigen biosynthesis in Salmonella typhimurium and Escherichia coli (Griffin et al., 1996, Stingele et al., 1996; Whitfield, 1995). These pathways involve the synthesis of individual repeat units on a lipid carrier (undecaprenyl phosphate) by the sequential activities of glycosyltransferase enzymes. Lipid-linked repeat units are then transported across the plasma membrane and exported from the bacterial cell (Whitfield, 1995). Recently, an operon directing the biosynthesis of EPS in the Sfi6 strain of Str. thermophilus has been identified (Stingele et al., 1996), while four complete genes involved in EPS biosynthesis in strain NCFB 2393 have been cloned and sequenced (Griffin et al., 1996). Interestingly, the predicted glycosyltransferase CpsE from NCFB 2393 differed from its homologue in the Sfi6 strain, indicating that these strains may produce EPSs with different chemical structures. We report here the cloning and nucleotide sequence analysis of an ~ 11 kb DNA fragment containing the remaining cps genes and genes flanking this cluster in Str. thermophilus NCFB 2393. We compare the sugar composition of the EPS produced by the NCFB 2393 strain to that reported for Sfi6 and several other strains.
of *Str. thermophilus*. Expression in *E. coli* was used to identify the glycosyltransferase activity encoded by the NCFB 2393 *cpsE* gene.

**METHODS**

**Strains, culture conditions and fermentation.** The bacterial strains and vectors used in this work are listed in Table 1. *E. coli* was grown in LB broth, Terrific broth or LB agar (Sambrook *et al.*, 1989) supplemented with 200 µg ampicillin ml−1 when applicable and incubated at 37 °C aerobically for 18 h. *Str. thermophilus* was grown in M17 medium (Terzaghi & Sandine, 1975) supplemented with 1% sucrose (SM17) and incubated at 42 °C without shaking. For plates, agar was added at 15 g l−1. For the production of EPS, *Str. thermophilus* was grown in MMST (Cerning *et al.*, 1988) using a 1 l fermenter (419 Modular Fermenter; Wood Enterprises), connected to a fermenter meter unit (H. R. Flow Inducer MHRE200; Watson-Marlow). Incubation was at 40 °C for 24 h with constant agitation (300 r.p.m.). The pH was maintained at 5.0 with 5 M NaOH.

**EPS isolation and characterization.** EPS was isolated as described by Stingele *et al.* (1996) except that 3 vols acetone were used instead of one and precipitation of EPS was carried out for 18 h at room temperature. EPS was further purified by passage through a Sephacryl S-400 column (M. Meldgaard, Chr. Hansen’s Laboratories). The molecular mass of the EPS was estimated by high-performance size-exclusion chromatography as follows: 1 mg freeze-dried EPS was added to 1 ml 10 mM ammonium bicarbonate (Sigma) and the solution filtered through a 0.2-µm pore size filter (Millipore). The solution was loaded immediately onto a Polysep column (Phenomenex) for Hewlett Packard HPLC, pre-standardized with a series of pullulan synthetic polysaccharide samples of known molecular mass (Shodex). Neutral sugar analysis of the EPS was performed by GC as described by Fox *et al.* (1989). The aminosugar content of the EPS was analysed by the method of Ludowieg & Benmamna (1967).

**General DNA manipulations.** DNA digestion and ligation was performed as recommended by the enzyme manufacturer. Electroporation into *E. coli* TG1 cells was carried out as described previously (Dower *et al.*, 1988). Plasmid DNA from *E. coli* was extracted using SV-Plus miniprep columns from Promega or by centrifugation through caesium chloride/ethidium bromide gradients (Sambrook & Sambrook, 1989). Total DNA was extracted from *Str. thermophilus* as described by Lewington *et al.* (1987). PCR's were carried out on Hybaid Omnigene thermocyclers. Template concentration was 600 ng per 50 µl reaction for chromosomal DNA; each primer was added to 0.1 µM. AmpliTaq polymerase enzyme (Perkin Elmer; 5 U µl−1) was used at 0.2 µl per 50 µl reaction. Annealing and extension conditions were dependent on the nature of primers and the expected product; these were as predicted by Oligo 4.0 for Macintosh.

**DNA sequencing and sequence analysis.** A primer-walking strategy was used to obtain the DNA sequence from the lambda inserts containing the *cpsE* genes. For the sequencing reactions, the Dye Terminator DNA Sequencing Kit (Applied Biosystems) was used and primer was added at 3-2 pmol per 20 µl reaction. Lambda DNA template for sequencing was used at 2–3 µg per 20 µl reaction. Sequencing reactions were electrophoresed and analysed using a Perkin Elmer 373A Fluorescence DNA sequencer. Primer design was checked in Oligo 4.0 for Macintosh. Sequence data were analysed using the Wisconsin GCG package Version 10, Laser Gene Package for the Macintosh version 1.58 (DNAStar), TFASTA (Pearson *et al.*, 1997), BLASTP (Altschul *et al.*, 1990), CLUSTAL W (Higgins *et al.*, 1994) and PHDsec (Columbia University; http://www2.ebi.ac.uk). Hydrophobicity domains were determined using DNA STRIDER (version 1.1). Prediction of transmembrane domains and cytoplasmic/extracytoplasmic loops was also carried out using six different programs available from the internet at the Expasy site (PHDhtm, ‘DAS’, HMMTOP, TopPred2, TMpred and TMHMM).

**Cloning of the *Str. thermophilus* *cpsE* genes.** A Lambda DASH II (Stratagene) library was constructed according to the manufacturer’s instructions. This was screened for the presence of the *cpsE* gene, using PCR primers designed from the 3’ end of a 4 kb insert in pAG14 as described previously (Griffin *et al.*, 1993). Lambda DNA from positive clones was extracted as described by Sambrook *et al.* (1989) or using the λ-extraction kit (Qiagen), and their inserts sequenced and analysed.

**Cloning and expression of *cpsE* in *E. coli*.** Primers for the amplification of the *cpsE* gene designed from positions 4707– 4726 and 6188–6206 were as follows: E60 5’-GGGATGA-TGCGGTTCCCTTA-3’ and E61 5’-CCCTTGAAGCCAAGATAT-3’, respectively. These were used to amplify a 1.48 kb fragment from lambda clone 9L3, using a proof-reading polymerase mixture (Boehringer). This fragment, which contained the entire *cpsE* gene and the upstream RBS (but no promoter), was cloned into the vector pcR2.1 (Stratagene), in both the sense and antisense orientations in respect of the promoter for the β-galactosidase gene (Plac) (confirmed by nucleotide sequencing). The resulting plasmids were called p13 and p34 respectively. *E. coli* TG1 cells were transformed with plasmids p13 and p34 by electroporation. Diluted (1/10) cultures of the clones containing the *cpsE* gene constructs were grown in 100 ml Terrific Broth until mid-exponential phase (OD540 = 0.475). At this stage induction was carried out with 1 mM IPTG. Cultures of TG1 containing the pCR2.1 vector or no vector at all were treated in the same way and used as negative controls. Membrane extracts of all the *E. coli* clones were prepared as described by Kolkman *et al.* (1996). The efficiency of the membrane extraction was checked by loading 12 µl aliquots on a denaturing 10% polyacrylamide gel as described by Laemmli (1970) and gels stained with Coomassie brilliant blue (Gibco-BRL). The protein concentration was determined using a simplification of the Lowry method (Peterson, 1977).

**Glucosyltransferase assay and TLC.** Membrane extracts (40 µl of 5 mg ml−1) from clones containing the *cpsE* gene construct as well as negative controls were incubated with UDP-[14C]glucose as described by Kolkman *et al.* (1996). One-third of the lipid fraction was used for scintillation counting and the remainder was vacuum dried and analysed by TLC. For TLC, the dried reacted samples were resuspended in 100 µl N-butanol and mild hydrolysis was carried out by adding 100 µl 50 mM trifluoroacetic acid and incubating at 95 °C for 20 min. The hydrolysed samples were vacuum dried and resuspended in 10 µl 40% 2-propanol, 8 µl of which was run on a silica gel plate (Merck HPTLC Alufolien Kieselgel 60). The developer used was a mixture of butanol/ethanol/water (5:3:2 by vol.). Standards of monosaccharides (40 µg each) were run simultaneously. Detection of radioactivity was done by exposing an X-ray sensitive film directly to the TLC plate, previously sprayed with a β-enhancer spray (EnHance, New England Nuclear), for 5 d at −80 °C. The standards were visualized by spraying with 5% sulfuric acid in ethanol and heating at
Table 1. Bacterial strains and vectors used in this study

<table>
<thead>
<tr>
<th>Strain/vector</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FI9070</td>
<td>E. coli TG1/pAG14</td>
<td>A. Griffin</td>
</tr>
<tr>
<td>FI9298</td>
<td>E. coli TG1/p13</td>
<td>This study</td>
</tr>
<tr>
<td>FI9299</td>
<td>E. coli TG1/p34</td>
<td>This study</td>
</tr>
<tr>
<td>TG1</td>
<td>Δ(lac-proAB), supE, thi, bsdΔ5, F’[traΔ36 proA’ B’ lac’ lacZΔM15]</td>
<td>Gibson (1984)</td>
</tr>
<tr>
<td>XL1-Blue</td>
<td>Δ(mcrCB–bsdMR–mrr)173, endA1, supE44, thi-1, gyrA96, relA1, lac (P2 lysogen)</td>
<td>Stratagene</td>
</tr>
<tr>
<td>MRA (P2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCFB 2393</td>
<td>EPS-producing strain from dairy starter</td>
<td>IFR culture collection</td>
</tr>
<tr>
<td><strong>Vectors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p13 (pFI2172)</td>
<td>5.3 kb; pCR2.1 containing cpsE from NCFB 2393 under P&lt;sub&gt;lac&lt;/sub&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>p34 (pFI2173)</td>
<td>5.3 kb; as p13 but the insert is in the opposite orientation to P&lt;sub&gt;lac&lt;/sub&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pAG14</td>
<td>6.7 kb; pUC18 containing 4 kb EcoRI fragment from NCFB 2393 for EPS biosynthesis</td>
<td>Griffin et al. (1996)</td>
</tr>
<tr>
<td>PCR2.1</td>
<td>3.9 kb; Amp and Km resistance, TA cloning vector</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>Lambda DASH II</td>
<td>i replacement vector. Accepts fragments of 9–23 kb</td>
<td>Stratagene</td>
</tr>
</tbody>
</table>

110 °C for 10 min. The assays described here were repeated and identical results were obtained.

**RESULTS AND DISCUSSION**

**Analysis of EPS produced by NCFB 2393**

Using the fermentation conditions outlined in the Methods, a maximum of 300 mg EPS l<sup>−1</sup> was obtained from NCFB 2393 cultures. Prior passage of NCFB 2393 (on a daily basis) through MMS<sub>r</sub> medium, at least five times, was required to achieve this level of EPS. Chemical analysis of the EPS showed that it consists of heptasaccharide repeating units containing the sugars GalNAc, Gal, Rha and Glc in a ratio of 1:1:2:3. Such a ratio has been reported for the Sfi6 strain (Doco et al., 1998). Indeed all of these strains produce EPS with similar features and the same ratio of sugars, presence/absence of branching and the type of linkages between sugar units of the EPS repeat unit. The molecular mass of the sample was estimated by high-performance size-exclusion chromatography to be ~2 × 10<sup>6</sup> Da. This was lower than the published 1 × 10<sup>6</sup> Da reported for the Sfi6 strain (Doco et al., 1990) and the 2 × 10<sup>6</sup> Da reported by Lemoine et al. (1997) for the Sfi12 and Sfi39 strains. Since the functional properties of a polysaccharide are influenced both by the chemical composition and the molecular mass (Sutherland, 1996), the NCFB 2393 polymer identified here is likely to display novel functional properties. Additional experiments are under way to demonstrate this.

**Nucleotide sequence analysis of the cps cluster**

In a previous study, we identified, cloned and nucleotide sequenced a 4074 bp chromosomal fragment containing part of the cps gene cluster (Griffin et al., 1996). To identify the remaining genes, primers designed from the 3’ end of the NCFB 2393 cpsE gene were used to screen a lambda library of *Str. thermophilus* genomic DNA. This was used to identify three positive lambda clones; i9L3, i3K1 and i8FJ1 (see Fig. 1). Nucleotide sequencing of the inserts from the lambda clones was performed to obtain the complete sequence data for a 1462 kb region. Analysis of these sequence data revealed the presence of 16 ORFs including the 3’ end of *cpsA* and the *cpsBCDE* genes previously reported by Griffin et al. (1996) with strong conformity with the streptococcal consensus sequence of *cpsA* (see accession no. Y17900) and putative promoters were identified upstream of six genes (see accession no. Y17900) and putative promoters were identified upstream of six genes (see Fig. 1 and accession no. Y17900). An AT rich region, 5’-AAAACGTTTTTTTGGTTTTTTTTTTGAAAAAAA-3’, similar to the consensus sequence of an upstream promoter (UP) element, 5’-NNAAAAT(AT)|T(AT)T(TT)-T(TT)NNAAAANNN-3’ (Estrem et al., 1998) was identified upstream of the -35 box of the *cpsA* gene between positions -72 and -42. UP elements have been reported to stimulate promoter activity up to 30-fold, and are found both in Gram-negative and Gram-positive bacteria (Estrem et al., 1998). Thus strong transcription may occur from the *cpsA* promoter, resulting in the generation of a single long transcript. The short intergenic spacing between *cps* genes suggestive of translational coupling, supports this view. In *Staphylococcus*
Sequence analysis of the cps genes

The cpsE, F, G, H and I genes encode proteins with predicted molecular masses of 518, 446, 426, 213 and 37.8 kDa respectively. When the amino acid sequences of these were compared to the databases, all showed homology to sugar transferases from a variety of different organisms (see Table 2). Given that the NCFB 2393 EPS contains the sugars GalNAc, Gal, Rha and Glc in a ratio of 1:1:2:3, we expected to find one GalNAc transferase, one galactosyltransferase, two rhamnosyltransferases and three glucosyltransferases (Table 2), therefore it may be the single galactosyltransferase expected in this system. However, C-terminal region of NCFB 2393 CpsE also contained a conserved catalytic motif DxD (residues 342–344) preceded by three hydrophobic amino acids (MFYM). This domain was found to be conserved by a diverse range of glycosyltransferase families, both prokaryotic and eukaryotic (Wiggins & Munro, 1998). A second putative catalytic region containing a pair of charged residues, Glu and Asp (residues 386 and 438 respectively) previously reported to be within a catalytic site in β-glycosyltransferases (Saxena et al., 1995; Van Kranenburg et al., 1999) was also identified.

The topological model of CpsE (Fig. 3), depicts CpsE spanning the cytoplasmic membrane five times (segments TM I, II, III, IV and V) and having two large loops, the first one extracytoplasmic, the second one intracytoplasmic (C-terminal region). The transmembrane domains are composed mainly of hydrophobic residues, which are predicted to adopt an α-helix conformation. The C-terminal domain, containing the active site, is in the cytoplasm. This contains three domains, A, B and C, reported to be common to glycosyltransferases by Wang et al. (1996). Domains A and B are reported to be associated with the interaction of the glycosyltransferase with the membrane lipid, while block C seems to be specific for the transferred sugar. The model presented here is in agreement with reports that EPS biosynthesis occurs at the internal side of the cytoplasmic membrane, with subsequent translocation of the repeat unit across the membrane (Kolkman et al., 1997; Stingele et al., 1996).

Analysis of CpsF-CpsG. CpsF and CpsG displayed greatest similarity to a rhamnosyltransferase and an N-acetylgalactosyltransferase respectively (Table 2), therefore it is reasonable to assign these enzyme activities to these proteins. Biochemical analysis will be required to confirm these predictions. CpsH displayed high similarity to a number of different glycosyltransferases (Table 2), therefore it seemed logical to predict that this represents the second of the three expected glycosyltransferases. CpsI had highest homology to galactosyltransferases (Table 2), therefore it may be the single galactosyltransferase expected in this system. However,
Table 2. Amino acid homologies for the gene products of the \textit{S. thermophilus} NCFB 2393 \textit{cps} cluster and flanking regions

<table>
<thead>
<tr>
<th>Product from strain 2393</th>
<th>Highest homology to</th>
<th>Similarity (%)</th>
<th>Identity (%)</th>
<th>Species</th>
<th>Function of homologue</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeoD</td>
<td>DeoD</td>
<td>99</td>
<td>98</td>
<td>\textit{Str. thermophilus} S\textit{66}</td>
<td>Nucleotide metabolism</td>
<td>Q56037</td>
</tr>
<tr>
<td>CpsE</td>
<td>Cps23E</td>
<td>75</td>
<td>57</td>
<td>\textit{Str. pneumoniae} 23F</td>
<td>Und-1P-Glc transferase</td>
<td>AF030373</td>
</tr>
<tr>
<td></td>
<td>Cps14E</td>
<td>74</td>
<td>55</td>
<td>\textit{Str. pneumoniae} S14</td>
<td>Und-1P-Glc transferase</td>
<td>P72513</td>
</tr>
<tr>
<td></td>
<td>OrfY</td>
<td>63</td>
<td>43</td>
<td>\textit{Str. thermophilus} S\textit{66}</td>
<td>Und-1P-Glc transferase</td>
<td>U40830</td>
</tr>
<tr>
<td></td>
<td>Cps23F</td>
<td>45</td>
<td>26</td>
<td>\textit{Bacillus subtilis} 60015</td>
<td>Spore coat biosynthesis</td>
<td>P46915</td>
</tr>
<tr>
<td></td>
<td>EpsF</td>
<td>51</td>
<td>25</td>
<td>\textit{Str. thermophilus} S\textit{66}</td>
<td>α-GalNAc transferase</td>
<td>U40830</td>
</tr>
<tr>
<td></td>
<td>VipC</td>
<td>46</td>
<td>24</td>
<td>Sal. typhimurium</td>
<td>LPS biosynthesis</td>
<td>Q04975</td>
</tr>
<tr>
<td></td>
<td>Rfas (WaaK)</td>
<td>48</td>
<td>23</td>
<td>Sal. typhimurium</td>
<td>1,2-GlcNAc transferase</td>
<td>P26470</td>
</tr>
<tr>
<td></td>
<td>Hi1244</td>
<td>64</td>
<td>48</td>
<td>\textit{Haemophilus influenza}</td>
<td>Unknown</td>
<td>P44134</td>
</tr>
<tr>
<td></td>
<td>Cps19K</td>
<td>50</td>
<td>25</td>
<td>\textit{Str. pneumoniae} 19F</td>
<td>Epimerase</td>
<td>U09239</td>
</tr>
<tr>
<td></td>
<td>Cps19E</td>
<td>51</td>
<td>24</td>
<td>\textit{Str. pneumoniae} 19F</td>
<td>Glucosyltransferase</td>
<td>U09239</td>
</tr>
<tr>
<td></td>
<td>Cps14E</td>
<td>50</td>
<td>22</td>
<td>\textit{Str. pneumoniae} S14</td>
<td>Glucosyltransferase</td>
<td>P72513</td>
</tr>
<tr>
<td></td>
<td>Cps14J</td>
<td>54</td>
<td>36</td>
<td>\textit{Galactosyltransferase}</td>
<td>Galactosyltransferase</td>
<td>X85787</td>
</tr>
<tr>
<td></td>
<td>EpsI</td>
<td>55</td>
<td>33</td>
<td>\textit{Str. thermophilus} S\textit{66}</td>
<td>Glucosyltransferase</td>
<td>U40830</td>
</tr>
<tr>
<td></td>
<td>EpsG</td>
<td>57</td>
<td>33</td>
<td>\textit{L. lactis} NIZO B\textit{40}</td>
<td>Glucosyltransferase</td>
<td>U93364</td>
</tr>
<tr>
<td></td>
<td>Gtf-s</td>
<td>47</td>
<td>21</td>
<td>\textit{Streptococcus sobrinus}</td>
<td>Dextranucrase precursor</td>
<td>P29336</td>
</tr>
<tr>
<td></td>
<td>91 kDa protein</td>
<td>49</td>
<td>20</td>
<td>\textit{Saccharomyces cerevisiae}</td>
<td>Unknown</td>
<td>P53243</td>
</tr>
<tr>
<td></td>
<td>RfbG (WbbG)</td>
<td>48</td>
<td>18</td>
<td>\textit{S. flexneri} 2a</td>
<td>Di-rhamnosyltransferase</td>
<td>X71970</td>
</tr>
<tr>
<td></td>
<td>Nup133</td>
<td>54</td>
<td>29</td>
<td>\textit{Sac. cerevisiae}</td>
<td>Pore protein</td>
<td>P36161</td>
</tr>
<tr>
<td></td>
<td>CaaX</td>
<td>54</td>
<td>26</td>
<td>\textit{Sac. cerevisiae}</td>
<td>Prenyl endoprotease</td>
<td>P47154</td>
</tr>
<tr>
<td></td>
<td>RfbX (Wxz)</td>
<td>47</td>
<td>20</td>
<td>\textit{S. flexneri} 2a</td>
<td>LPS transport</td>
<td>X71970</td>
</tr>
<tr>
<td></td>
<td>Yeast protein</td>
<td>46</td>
<td>17</td>
<td>\textit{Deharyomyces occidentalis}</td>
<td>K⁺-affinity transporter</td>
<td>P50505</td>
</tr>
<tr>
<td></td>
<td>Rfas (WaaS)</td>
<td>52</td>
<td>27</td>
<td>\textit{E. coli} K-12</td>
<td>LPS outer core assembly</td>
<td>P27126</td>
</tr>
<tr>
<td></td>
<td>Undescribed ORF</td>
<td>95</td>
<td>91</td>
<td>\textit{Str. thermophilus} CNRZ 368</td>
<td>Unknown</td>
<td>Y13713</td>
</tr>
<tr>
<td></td>
<td>CYOB</td>
<td>53</td>
<td>29</td>
<td>\textit{E. coli} K-12</td>
<td>Cyt/ubiquit oxidase</td>
<td>P18401</td>
</tr>
<tr>
<td></td>
<td>GR78</td>
<td>41</td>
<td>24</td>
<td>\textit{Klebsiella lactis}</td>
<td>Glc-regulated precursor</td>
<td>P20210</td>
</tr>
<tr>
<td></td>
<td>HSP70</td>
<td>40</td>
<td>23</td>
<td>\textit{Glycine max} (soybean)</td>
<td>Heat Shock protein 70</td>
<td>P26413</td>
</tr>
<tr>
<td></td>
<td>Orf14.9</td>
<td>97</td>
<td>96</td>
<td>\textit{Str. thermophilus} S\textit{66}</td>
<td>Unknown</td>
<td>U40830</td>
</tr>
<tr>
<td></td>
<td>VanZ</td>
<td>59</td>
<td>33</td>
<td>\textit{Enterococcus faecium}</td>
<td>Antibiotic resistance</td>
<td>Q06242</td>
</tr>
<tr>
<td></td>
<td>Nu5M</td>
<td>57</td>
<td>26</td>
<td>\textit{Caenorhabditis elegans}</td>
<td>NADH-ubiquinone OR</td>
<td>P24896</td>
</tr>
<tr>
<td></td>
<td>Ypud</td>
<td>51</td>
<td>27</td>
<td>\textit{B. subtilis}</td>
<td>Unknown</td>
<td>P17616</td>
</tr>
</tbody>
</table>

it also contained an AX₁LDXD motif reported to be found in the β-glucosyltransferase and cellulose synthase superfamilies (Becker et al., 1997; Wiggins & Munro, 1998). This could imply that CpsI could be the third glucosyltransferase, thus a detailed biochemical analysis will be required to establish which sugar CpsI transfers. While CpsE–CpsL clearly encode five of the seven sugar transferases, the identity of the remaining two transferases is not so clear. Good candidate sugar transferases are CpsJ and CpsL. CpsJ had homology to dextranucrase, a processive glucosyltransferase enzyme. This could indicate that CpsJ may transfer more than one sugar unit (and so account for the two remaining transferases). Transfer of more than one sugar unit has been reported for glycosyltransferases from other systems, for example RfbG (WbbG) from \textit{Shigella flexneri} transfers two consecutive rhamnosyl units (Morona et al., 1995).
Analysis of \textit{cpsK} and \textit{cpsL}. The \textit{cpsK} and \textit{cpsL} genes encode proteins with predicted molecular masses of 14.0 and 35.0 kDa respectively. When the amino acid sequence of \textit{cpsK} was compared to the databases (see Table 2), similarity to several proteins involved in transport was detected. \textit{CpsL} displayed significant
homology to only one protein: WaaS from the LPS biosynthetic pathway of *E. coli*. A specific function has not been assigned to WaaS, however it has been suggested that it may act as an accessory protein required for the assembly of outer-core oligosaccharide, interacting either with glycosyltransferases or with the inner core LPS moiety (Pradel et al., 1992). Alternatively, Heinrichs et al. (1998) speculate that WaaS may be a rhamnosyltransferase, thus CpsL may be a rhamnosyltransferase. The position of a glycosyltransferase down-stream of a putative export gene has been reported before for the *cps/eps* clusters of *L. lactis* and *Str. pneumoniae* (Van Kranenburg et al., 1999; Morona et al., 1999). More experiments are required to resolve the function of *cpsL*.

**Expression of *cpsE* in *E. coli***

The NCFB 2393 *cpsE* gene was cloned into a high copy number vector and expressed in *E. coli* TG1. Expression from plasmid p13 (containing *cpsE* under the control of \( P_{lac} \)) resulted in a protein of approximately 42 kDa, which was smaller than the expected molecular mass of 51.8 kDa (Fig. 4a). The reason for this is unclear. A glycosyltransferase assay using membrane extracts from *E. coli* cells containing either p13, p34 (*cpsE* cloned in the opposite direction to \( P_{lac} \)) or pCR2.1 (vector alone) revealed enzymic activity, detected as the incorporation of UDP-[\(^{14}\)C]Glc in extracts from TG1/p13 and TG1/p34 (Table 3). No protein band was detected in extracts from p34, therefore we were surprised to detect glycosyltransferase activity in these. This activity may have resulted from expression from a non-specific promoter within the vector, yielding a level of protein too low to be detected by SDS-PAGE. Use of extracts from *E. coli* containing the vector alone confirmed that the glycosyltransferase activity detected was genuinely due to expression of the NCFB 2393 *cpsE* gene. No significant difference in glycosyltransferase activity was detected between the induced and uninduced clones, suggesting that the chromosomal \( lacI^q \) gene in the background *E. coli* TG1 is insufficient to completely repress expression from \( P_{lac} \) in pCR2.1. TLC analysis demonstrated that CpsE transferred only 14-C-labelled glucose to the lipid fraction (Fig. 4b), demonstrating that the conversion of glucose to galactose by the activity of an epimerase that was reported to occur in some strains of *E. coli* by Kolkman et al. (1997), did not occur here. In addition, only labelled monosaccharide was detected in this study, demonstrating that CpsE transfers only a single glucose residue and is not a processive glycosyltransferase. Results from these assays combined with the data for the homology of NCFB 2393 CpsE to *Xanthomonas campestris* GumD (Ielpi et al., 1993), imply that NCFB 2393 CpsE transfers a glucose 1-phosphate from UDP-glucose to undecaprenyl phosphate.

*Table 3. Scintillation counts for the *cpsE* clones after 90% efficiency was reached (maximum of 10 min)*

<table>
<thead>
<tr>
<th>Clone</th>
<th>c.p.m.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TG1/p13 <em>cpsE</em> (+)</td>
<td>10034</td>
</tr>
<tr>
<td>TG1/p13 induced <em>cpsE</em> (+)</td>
<td>9078</td>
</tr>
<tr>
<td>TG1/p34 <em>cpsE</em> (–)</td>
<td>5663</td>
</tr>
<tr>
<td>TG1/p34 induced <em>cpsE</em> (–)</td>
<td>5551</td>
</tr>
<tr>
<td>TG1/pCR2.1</td>
<td>72</td>
</tr>
<tr>
<td>TG1/pCR2.1 induced</td>
<td>218*</td>
</tr>
<tr>
<td>TG1</td>
<td>32</td>
</tr>
<tr>
<td>TG1 induced</td>
<td>37</td>
</tr>
</tbody>
</table>

* Possibly unincorporated radiolabelled glucose.

**Analysis of genes outside the *cps* region**

Comparison of the predicted amino acid sequences to the databases revealed that those at the S’ and 3’ end of the region sequenced were not involved in EPS biosynthesis (Fig. 1 and Table 2). The positions of *deoD*...
end of the cluster, \textit{cpsJJKL}, is only of 30 mol\%.
These genes are located next to IS1193, suggesting that foreign
genes could have been introduced into the NCFB 2393
\textit{cps} locus by horizontal transfer. Recombination could
have been facilitated by the presence of mobile elements
such as IS1193 and ISS1, or by the presence of conserved
genes within these clusters. Additionally, given the
similarity of some of our genes to the \textit{cps} genes from \textit{Streptococcus pneumoniae},
horizontal transfer between different strains of streptococci might also have occurred at some
time during the evolution of these two species.

**ACKNOWLEDGEMENTS**

The authors would like to thank A. Jay and L. Adams for the chemical and high-performance size-exclusion chromatography analysis of the EPS; Dr A. Colhuane and Dr S. Andrews for useful advice with the computer analysis of \textit{CpsE}; and Dr J. Payne for help with the sequencing of the \textit{cps} operon. A.G. is supported by an EC contract no. ERB1
B104CT96739. E.A. is supported by an EC-TMR fellowship
no. ERBF MIBICT961816.

**REFERENCES**

Altshchul, S. F., Gish, W., Miller, W., Myers, E. & Lipman, D. J.

Becker, A., Rüberg, S., Küster, H., Roxlau, A., Keller, M., Ivashina,
cluster of \textit{Rhizobium meliloti} directing the biosynthesis of

Bourgoin, F., Guédon, G., Pebay, M., Roussel, Y., Panis, C. &
Decaris, B. (1996). Characterization of a mosaic ISS1 element and
evidence for the recent horizontal transfer of two different types
of ISS1 between \textit{Streptococcus thermophilus} and \textit{Lactococcus lactis}. \textit{Gene} 178, 15–23.

Characterization of a novel insertion sequence, IS1194, in
\textit{Streptococcus thermophilus}. \textit{Plasmid} 40, 44–49.

Bourgoin, F., Pluvinet, A., Gintz, B., Decaris, B. & Guédon, G.
(1999). Are horizontal transfers involved in the evolution of
\textit{Streptococcus thermophilus} EPS synthesis loci? \textit{Gene} 233,
151–161.

Bubb, W., Urashima, T., Fujiwara, R., Shinnai, T. & Ariga, H.
(1997). Structural characterization of the exocellular polysaccharide
produced by \textit{Streptococcus thermophilus} \textit{OR901}. \textit{Carbohydr Res} 301, 41–50.

Exocellular polysaccharide production by \textit{Streptococcus thermophilus}.

 Coffey, T. J., Enright, M. C., Daniels, M., Morona, J. K., Morona,
Recombinational exchanges at the capsular polysaccharide biosynthetic
locus lead to frequent serotype changes among natural isolates of

exocellular polysaccharide produced by \textit{Streptococcus thermophilus}.


 Duwat, P., Cochu, A., Ehrlich, D. & Gruss, A. (1997). Charac-

Copies of ISS1 and ISS193 are widely present in strains of \textit{Streptococcus thermophilus} and \textit{L. lactis} and it has been
suggested that these could allow horizontal transfer to
occur between these two organisms during co-culture in
dairy manufacture (Bourgoin \textit{et al.}, 1996, 1999). Horiz-
ontal transfer and generation of new capsular poly-
saccharides or LPS as a consequence of homologous
recombination between insertion elements has also been
reported to occur in the CPS and LPS loci of strains of
\textit{Streptococcus pneumoniae} and \textit{Sal. enterica} (Morona \textit{et al.},
1997; Muñoz \textit{et al.}, 1997; Xiang \textit{et al.}, 1994). The \textit{G} + \textit{C} content of the NCFB 2393 \textit{cps} cluster is 35.6 mol\%; this
is close to the average \textit{G} + \textit{C} content reported for \textit{Streptococcus thermophilus} of 37–40 mol\% (Farrow & Collins, 1984).
Interestingly, the \textit{G} + \textit{C} content of four genes at the 3‘
and \textit{orf}14.9 are conserved when compared to \textit{Str. thermophilus} strain S6b (Stingele \textit{et al.}, 1996); also in the
CNZR 368 strain these genes are associated with the
EPS locus (Bourgoin \textit{et al.}, 1999).

A partial copy of the ISS1 transposase was identified
next to \textit{orf}14.9 in \textit{Str. thermophilus} NCFB 2393. This
contains only one long terminal repeat (at the 3’ end)
identical to that found in the CNZR 368 ISS1. The
NCFB 2393 ISS1 has 64\% identity (78\% similarity) to
its homologue from \textit{Str. thermophilus} CNZR 368
(Bourgoin \textit{et al.}, 1999); 32\% identity (58\% similarity)
to the IS257 transposase from \textit{Sta. aureus} \textit{Tn}4003
(accession no. P14506) and 28\% identity (54\% simi-
larity) to the ISS1 transposase from \textit{L. lactis} IL946
(accession no. L35176). Copies of ISS1 members are
widespread among industrial strains of \textit{Str. thermo-
philus} and \textit{L. lactis}, where they have been found in
different positions on the chromosome including the
EPS gene clusters (Bourgoin \textit{et al.}, 1999; Duwat \textit{et al.},
1997; Mercenier & Lemoin, 1989; Stingele \textit{et al.}, 1996).

A second IS element, IS1193, was also identified; this
contained two 19 bp imperfect inverted repeats and two
8 bp perfect direct repeats, flanking a transposase gene
(1246 bp). The NCFB 2393 IS1193 displayed 88\% identity
(91\% similarity) at the amino acid level to IS1193 from \textit{Str. thermophilus} CNZR 368 (Bourgoin \textit{et al.},
1999), 62\% identity (75\% similarity) to TnpA from
\textit{Str. pneumoniae} \textit{WU2} (accession no. U66845), and
57\% identity (72\% similarity) to IS1167 from \textit{Str.
\textit{pneumoniae} RX1} (accession no. M31680). A short ORF
\textit{orfM}, 234 bp) has been identified within IS1193 in the
opposite orientation to the transposase gene. \textit{orfM},
including its putative RBS, is also present in \textit{Str.
\textit{thermophilus} CNZR} 368, although this was not
reported by the authors. Recently, a new IS element
from strain CNZR 368, IS1194, was identified in the \textit{eps}
cluster (Bourgoin \textit{et al.}, 1998). IS1194 and other
members of the IS5 subfamily all contain two ORFs, the
shorter one (of unknown function) in the opposite
direction to and overlapping the larger ORF which
encodes the transposase gene. \textit{orfM} might be a
remaining part of some element needed for the trans-
position of all IS5 members, or might be the vestige of a
gene picked up during transposition between organisms.

Copies of ISS1 and IS1193 are widely present in strains of
\textit{Str. thermophilus} and \textit{L. lactis} and it has been
suggested that these could allow horizontal transfer to
occur between these two organisms during co-culture in
dairy manufacture (Bourgoin \textit{et al.}, 1996, 1999). Horiz-
ontal transfer and generation of new capsular poly-
saccharides or LPS as a consequence of homologous
recombination between insertion elements has also been
reported to occur in the CPS and LPS loci of strains of
\textit{Str. pneumoniae} and \textit{Sal. enterica} (Morona \textit{et al.},
1997; Muñoz \textit{et al.}, 1997; Xiang \textit{et al.}, 1994). The \textit{G} + \textit{C} content of the NCFB 2393 \textit{cps} cluster is 35.6 mol\%; this
is close to the average \textit{G} + \textit{C} content reported for \textit{Str.
thermophilus} of 37–40 mol\% (Farrow & Collins, 1984).
Interestingly, the \textit{G} + \textit{C} content of four genes at the 3‘


Received 20 March 2000; revised 30 June 2000; accepted 26 July 2000.