Microbiology (2005), **151**, 1839–1851 **DOI** 10.1099/mic.0.27852-0

Comparative analysis of the exopolysaccharide biosynthesis gene clusters from four strains of *Lactobacillus rhamnosus*

B. Péant,[1,5] G. LaPointe,[1,2] C. Gilbert,[3] D. Atlan,[3] P. Ward[4] and D. Roy[1,2]

Correspondence
Gisèle LaPointe
gisele.lapointe@fsaa.ulaval.ca

[1]STELA Dairy Research Centre, Room 1316, Pavillon Paul-Comtois, Université Laval, Quebec, QC, Canada G1K 7P4

[2]Nutraceuticals and Functional Foods Institute (INAF), Université Laval, Quebec, QC, Canada G1K 7P4

[3]Laboratoire de Microbiologie et de Génétique Moléculaire, UMR CNRS/UCBL/INSA 5122, Université Lyon 1, 69622, Villeurbanne, France

[4]Food Research and Development Centre, Agriculture and Agri-Food Canada, St Hyacinthe, Quebec, Canada J2S 8E3

[5]Hopital Notre-Dame, 560 Sherbrooke Est, Montréal, QC, Canada H2L 4M1

The exopolysaccharide (EPS) biosynthesis gene clusters of four *Lactobacillus rhamnosus* strains consist of chromosomal DNA regions of 18·5 kb encoding 17 ORFs that are highly similar among the strains. However, under identical conditions, EPS production varies considerably among these strains, from 61 to 1611 mg l^{-1}. Fifteen genes are co-transcribed starting from the first promoter upstream of *wzd*. Nevertheless, five transcription start sites were identified by 5′-RACE PCR analysis, and these were associated with promoter sequences upstream of *wzd*, *rmlA*, *welE*, *wzr* and *wzb*. Six potential glycosyltransferase genes were identified that account for the assembly of the heptasaccharide repeat unit containing an unusually high proportion of rhamnose. Four genes involved in the biosynthesis of the sugar nucleotide precursor dTDP-L-rhamnose were identified in the EPS biosynthesis locus, which is unusual for lactic acid bacteria. These four genes are expressed from their own promoter (P2), as well as co-transcribed with the upstream EPS genes, resulting in coordinated production of the rhamnose precursor with the enzymes involved in EPS biosynthesis. This is believed to be the first report demonstrating that the sequence, original organization and transcription of genes encoding EPS production are highly similar among four strains of *Lb. rhamnosus*, and do not vary with the amount of EPS produced.

Received 22 December 2004
Revised 11 February 2005
Accepted 24 February 2005

INTRODUCTION

Interactions between bacteria and human host cells are mediated by cell-surface structures, particularly proteins and polysaccharides. Extracellular polysaccharides can be attached to cell membrane components as lipopolysaccharides (LPS) and lipoteichoic acids, they can form a capsule around the cell as capsular polysaccharides (CPS), or they can be released as exopolysaccharides (EPS) (Roberts, 1996). EPS produced by lactic acid bacteria (LAB) display a great variety of structures, and many are heteropolysaccharides composed of different sugar moieties (glucose, galactose, rhamnose, mannose, *N*-acetylglucosamine, *N*-acetylgalactosamine, glucuronic acid) (De Vuyst & Degeest, 1999; De Vuyst *et al.*, 2001; Ricciardi & Clementi, 2000). In addition to their technological properties in fermented milk products (Ruas-Madiedo *et al.*, 2002; van Kranenburg *et al.*, 1999a), EPS produced by LAB may also have biological roles, such as immunomodulatory (Hosono *et al.*, 1997), antitumour (Kitazawa *et al.*, 1991) and cholesterol-lowering activities (Nakajima *et al.*, 1992). The EPS from *Lactobacillus rhamnosus* RW-9595M stimulates interleukin, tumour necrosis factor and interferon gamma in mouse splenocytes (Chabot *et al.*, 2001).

EPS production by *Lb. rhamnosus* varies greatly among strains (Dupont *et al.*, 2000). The EPS production by strain

Abbreviations: BPGN, bacterial polysaccharide gene nomenclature; CPS, capsular polysaccharide; EPS, exopolysaccharide; LAB, lactic acid bacteria; RACE, rapid amplification of cDNA ends; UP element, upstream enhancer element.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are AY659976 (ATCC 9595), AY659977 (RW-6541M), AY659978 (strain R) and AY659979 (RW-9595M).

RW-9595M (2350 mg l^{-1}; Bergmaier *et al.*, 2003) is among the highest measured to date for LAB, while strain ATCC 9595 produces from 50 to 116 mg l^{-1}. *Lb. rhamnosus* RW-9595M produces an EPS with an exceptionally high rhamnose content, composed of heptasaccharide subunits of L-rhamnose, D-glucose and pyruvate-substituted D-galactose in a molar ratio of 4 : 2 : 1 (Van Calsteren *et al.*, 2002). The structure of the repeating unit remains remarkably stable under different conditions for a number of *Lb. rhamnosus* strains (Van Calsteren *et al.*, 2002). EPS with unusual monosaccharides such as L-rhamnose represent a source of new oligosaccharides and high-value substrates in the synthesis of pharmaceutical and aromatic compounds (Farres *et al.*, 1997; Paul *et al.*, 1986). As polysaccharide structure has a great influence on the technological properties and biological activities of EPS, rapid identification of new structures through genetic screening techniques will accelerate the discovery of LAB strains producing EPS with novel functional properties.

Biosynthesis of heteropolysaccharides starts with the intracellular formation of EPS precursors, the sugar nucleotides uridine-5′-diphosphate (UDP)-glucose, UDP-galactose and deoxythymidine diphosphate (dTDP)-rhamnose, which are the donor monomers for the biosynthesis of most repeating units. Genes directing heteropolysaccharide biosynthesis in LAB have been sequenced for a number of genera and species (Broadbent *et al.*, 2003; Jolly *et al.*, 2001, 2002b; van Kranenburg *et al.*, 1997). Genetic elements required for EPS production include genes encoding regulation, chain-length determination, repeat-unit assembly, polymerization and export (De Vuyst & Degeest, 1999). In spite of accumulating knowledge of EPS gene organization, very little is known about the mechanism regulating EPS biosynthesis.

Despite their economic importance, few genetic studies of EPS production in *Lactobacillus* species have been reported (Jolly *et al.*, 2002a; Lamothe *et al.*, 2002). The objective of this work was to compare the genetic organization of the EPS biosynthesis gene cluster among strains of *Lb. rhamnosus* that vary with respect to their stable capacity for rhamnose-containing EPS production. Comparison of gene structure, sugar precursor biosynthesis and regulatory elements among strains is essential to understand the basis of the diversity in EPS production levels and composition. Our study reveals that the *Lb. rhamnosus* EPS gene organization differs considerably from other LAB, including other lactobacilli, but shows remarkably little variation among the four strains examined. In order to facilitate comparative studies on genes encoding bacterial surface polysaccharides, we have adopted the bacterial polysaccharide gene nomenclature system, which assigns the same symbol to homologous genes according to their function (Broadbent *et al.*, 2003; Reeves *et al.*, 1996).

METHODS

Bacterial strains and media. All bacterial strains and plasmids are listed in Table 1. *Lb. rhamnosus* cultures were incubated at the optimal growth temperature of 37 °C (Dupont *et al.*, 2000) in MRS broth (Quelab) (De Man *et al.*, 1960). *Escherichia coli* strains were grown in Luria–Bertani (LB) medium (Miller, 1992) at 37 °C with aeration. When necessary, media were solidified with 1·5 % agar (Quelab). All stock cultures were maintained at −80 °C in growth medium or BHI supplemented with 15 % (w/v) glycerol. Viable lactobacilli counts were determined by spread-plating 10-fold serial dilutions (in 0·1 %, w/v, sterilized peptone water) on MRS agar, and incubating aerobically at 37 °C for 48 h.

Production, purification and quantification of EPS. *Lb. rhamnosus* strains ATCC 9595 and RW-9595M were selected in order to represent low and high EPS production levels, respectively. Fermentations were carried out in triplicate in 2 l fermenters (Biostat M; B. Braun) with 1·6 l BMM medium (Morishita *et al.*, 1981)

Table 1. Source and relevant characteristics of strains and plasmids used in this study

Strain or plasmid	Source or reference	Relevant characteristics
Strains		
Lb. rhamnosus ATCC 9595	ATCC	116 mg EPS l^{-1}†
Lb. rhamnosus R	Institut Rosell*	498 mg EPS l^{-1}‡
Lb. rhamnosus RW-9595M	Our collection	1611 mg EPS l^{-1}†
Lb. rhamnosus RW-6541M	Our collection	61 mg EPS l^{-1}†
E. coli TOP10	Invitrogen	F$^-$ *mcrA* Δ(*mrr–hsdRMS–mcrBC*) *ϕ*80*lacZ*ΔM15 Δ*lacX74* *deoR recA1 araD139* Δ(*ara–leu*)7697 *galU galK rpsL* (Str^R) *endA1 nupG*
Plasmids		
pCR4-TOPO	Invitrogen	T-overhangs for direct ligation of PCR products
pCR4-TOPO-XL	Invitrogen	T-overhangs for direct ligation of long PCR products

*Institut Rosell, Montreal, Quebec, Canada.

†EPS production was quantified after purification from BMM medium (pH controlled at 6·0) with 40 g glucose l^{-1} as the carbon source (this study).

‡EPS production was quantified after purification from BMM medium (pH controlled at 6·0) with 20 g glucose l^{-1} as the carbon source and incubation for 24 h (Dupont *et al.*, 2000).

containing 2 g tryptone l^{-1} and 40 g glucose l^{-1}. The pH was adjusted to 6·0 with HCl, and maintained by the addition of 7 M NH_4OH. After inoculating at 1 % (v/v) with a 16 h preculture in MRS medium, incubation was carried out at 37 °C under constant agitation of 100 r.p.m. Samples of 50 ml were taken at 0, 16, 20, 24 and 40 h. Growth was measured by serial dilution plating on MRS agar, and by OD_{600} measurement. An OD_{600} of 1·0 represents a dry mass (cells and attached EPS) of 1·4 g l^{-1} for ATCC 9595, and 2·1 g l^{-1} for RW-9595M.

EPS isolation and purification were carried out by ethanol precipitation, as described by Cerning *et al.* (1994), with a few modifications. Samples were heated for 15 min at 100 °C, then cells were eliminated by centrifugation at 12 000 ***g*** for 45 min at 4 °C. EPS were precipitated from the supernatant with 3 vols 95 % ethanol at 4 °C for 16 h, and collected by centrifugation at 12 000 ***g*** for 20 min. EPS pellets were dissolved in deionized water, and dialysed over a period of 3 days at 4 °C, with two water changes per day. Total sugars were measured by the phenol/sulfuric acid method (Dubois *et al.*, 1951), with glucose as a standard, and the results are expressed in mg glucose l^{-1}.

DNA isolation and manipulation. *Lb. rhamnosus* cells from 16–18 h cultures (2 or 10 ml) were harvested by centrifugation (8000 ***g***, 10 min), and DNA was extracted as described previously (Péant & LaPointe, 2004; Vincent *et al.*, 1998). PCR was performed using standard conditions (Ausubel, 1995) with *Taq* polymerase (Promega) and the primers listed in Table 2. Platinum *Pfx* DNA polymerase (Invitrogen Life Technologies) was used to amplify larger fragments (>2 kb). PCR products were purified using the QIAquick PCR Purification Kit (Qiagen).

For chromosome walking, the Universal GenomeWalker kit was used according to the manufacturer's recommendations (BD Biosciences). For inverse PCR experiments, 200 ng DNA was digested with rare-cutting restriction enzymes, purified, and ligated using T4 DNA ligase (Boehringer Mannheim) for 10 h at 16 °C. The amplification reaction was carried out using the r*Tth* DNA polymerase (Applied Biosystems) in a reaction volume of 100 µl. PCR cycling was done with the following conditions: 92 °C for 2 min, 30 cycles of 92 °C for 45 s, 60 °C for 1 min, and 68 °C for 4 min, followed by 6 min at 68 °C, and holding at 16 °C. Amplifications were visualized on a 0·8 % agarose gel after ethidium bromide staining by UV transillumination (Gel-Doc 1000; Bio-Rad). PCR amplicons less than 4 kb were cloned into pCR 4-TOPO, while those 4 kb and over were cloned into pCR XL-TOPO, according to the manufacturer's recommendations (Invitrogen).

Nucleotide sequencing and sequence analysis. The DNA sequence of both strands was determined by the sequencing service of Université Laval (Life and Health Sciences Pavilion) using the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems), with either universal primers or the primers listed in Table 2. DNA sequence analysis and similarity searches were carried out using the BLAST network service at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/), National Institutes of Health (Altschul *et al.*, 1997). BLASTX software (http://www.ncbi.nlm.nih.gov/blast/) was used to conduct similarity searches of the nucleotide and protein databases. Membrane-spanning regions of translated gene products were predicted using the TMpred program (http://www.ch.embnet.org/software/TMPRED_form.html).

Gene nomenclature. Genes were named according to the bacterial polysaccharide gene nomenclature (BPGN) system (www.microbio.usyd.edu.au/BPGD/default.htm). This nomenclature provides a single name for genes of a given function, and it can be applied to all species (Reeves *et al.*, 1996). The BPGN system has not been used widely for LAB, except for some *Streptococcus thermophilus* strains (Broadbent *et al.*, 2003). The designation '*we**' is used for naming genes involved in EPS biosythesis, while '*wc**' is used for capsule genes. The designation '*wz**' was proposed for genes with homologous predicted functions in regulating polysaccharide processing and polymerization; these genes are found in many polysaccharide gene clusters. Finally, genes implicated in nucleotide sugar precursor biosynthesis are named for the pathway, such as '*rml*' for the dTDP-L-rhamnose pathway.

Analysis of gene transcription. Total cellular RNA was extracted from 10 ml cultures of *Lb. rhamnosus* strains ATCC 9595 and RW-9595M after 6 h (exponential phase) or 24 h (stationary growth phase) incubation by using the RNeasy Midi Kit (Qiagen). All RNA isolation steps were performed according to the manufacturer's instructions, except that 50 U mutanolysin ml^{-1} (Sigma) was used in addition to lysozyme to degrade the cell walls, and incubation time was extended to 1 h. The isolated RNA was treated with RNase-free DNase I (Qiagen) at 25 °C for 1 h, followed by a second purification using an RNeasy column.

For RT-PCR (One-step kit; Qiagen), reverse transcription was performed with primers (Table 2) derived from the *Lb. rhamnosus* ATCC 9595 EPS biosynthesis locus to obtain an amplicon length between 1·5 and 3 kb. As positive controls, each region was amplified with the same primers using chromosomal DNA of *Lb. rhamnosus* ATCC 9595 as the template. As negative controls, reaction product without the reverse transcription stage was used as the template. Identification of the 5′ end of the EPS gene-specific mRNA, and characterization of putative promoter activities, were carried out using the 5′-RACE kit (Invitrogen Life Technologies) and the primers listed in Table 2. All experimental steps were performed according to the manufacturer's instructions. This method captures the 5′-end information of mRNA through synthesis of first-strand cDNA initiated from a gene-specific reverse primer. An anchor sequence was then added to the 3′ end of the cDNA using terminal deoxynucleotide transferase, followed by direct amplification of tailed cDNA using the nested gene-specific primers and the anchor-specific primer provided. PCR products were cloned and sequenced as described above.

RESULTS

Growth and EPS production by strains ATCC 9595 and RW-9595M

Mean specific growth rates (h^{-1}) were 0·266±0·006 for ATCC 9595, and 0·296±0·027 for RW-9595M. These values are comparable between strains, considering the precision limits of the plate-count method. Glucose was exhausted in the medium after the total incubation period (Fig. 1). RW-6541M produced less EPS than ATCC 9595, whereas strain R exhibited an EPS production between those of RW-9595M and ATCC 9595 (Table 1). The maximum EPS production at 40 h incubation was 1611 mg l^{-1} for strain RW-9595M, while most of the EPS production occurred by 24 h (Fig. 1). For ATCC 9595, the maximum production of 116 mg l^{-1} was attained after 40 h incubation. Strain R produced a maximum of 498 mg l^{-1} after 24 h incubation, while strain RW-6541M produced only 61 mg l^{-1} after 72 h incubation.

Cloning and sequencing of the EPS biosynthesis loci from four *Lb. rhamnosus* strains

Specific nested primers (Table 2) were designed from the C-terminal region of the priming glycosyltransferase gene

Table 2. Primers

Primer name	Sense*	Sequence	Purpose
BP100HaeF	F	ATCGACTCACTGGCGAAAGCG	Upstream of EPS region (*wzd*) of ATCC 9595
BP109cF	F	ATGTAGTTAAAGAAGAGCAACCG	Amplification, sequencing and RT-PCR from *wzd*
BP115CDRace3	R	TCAGACCGACAGTAAATGCTAGAACC	Amplification from *wzd* and identification of transcription start for P1
95F2657	F	AGCCGAACCAGAAACGACG	RT-PCR from *wzd*
BP102epMR	R	AGCACGGATAAAATTCTGAAGCAATGCC	RT-PCR from *wzx*
BP84BgP95F1	F	ATGAATGGCAACAGGCAAACCG	Amplification and sequencing from *welF*
BP87BgP95R2	R	ATAGCGTTCATCGAATCCTCCGACC	Amplification and sequencing from *welF*
95R5298	R	TTCAATTGTCTTCAGTTCTGTGCG	RT-PCR from *welF*
95F6007	F	TGGGGATGGAAGTTGTTCTAAAGG	RT-PCR from *welF*
BP99AmRhamR	R	TTCGAATCATCAGGTACTTGCCCG	RT-PCR from *welG*
BP95-8385F	F	AAGAACCATCGAAATCTAGTTGACGC	RT-PCR from *welG*
BP96Rham2R	R	ATACAACACAAACCGCGCCATTCC	RT-PCR from *welH*
BP97Rham1F	F	ATTCGACTGGCAAAAAAATGAGCC	Amplification and sequencing from *welI*
BP82XH95R1	R	ATAAAGCTGGCCAGACGCGGC	Amplification, sequencing and RT-PCR from *welI*
BP81XH95F	F	TTAGTGCATGATAATGTTTGTGTCGGCG	RT-PCR from *welI*
BP70F	F	TTCGGAATAGCCATTTTTCGTC	Amplification and sequencing from *wzy*
BP71R	R	AAAGTCTCATACGAAGCTCCCC	Amplification and sequencing from *wzy*
BP106epGR	R	TTTCTACACCCGTTGCCACC	RT-PCR from *welJ*
GTFXF1	F	TGGTTTTGTGACAGCCATTGAGA	RT-PCR from *welJ*
BP78gR	F	CCATCTCAAAACAGTTGGTTCCCG	Amplification, sequencing and RT-PCR from *rmlA*
RMLA90R	R	TGGGTAATAAATCATTGGCTTGTCAT	RT-PCR from *rmlA*
95R12463	R	ATCACGCGGTGTTGAAATCAC	RT-PCR from *rmlA*
95F13099	F	GGCACAGCCGCTCAAGAAG	RT-PCR from *rmlA*
BP64	F	GCGAACAGGATAACAAGACCGTT	Amplification and sequencing from *rmlB*
BP65	R	GGTTGAAATGTGGTGGAAGCGAA	Amplification, sequencing and RT-PCR from *rmlB*
95R15551	R	TTGACGATACCGGCTTAACCG	RT-PCR from *rmlD*
BP119ERace3	R	AATTTTAAACATTGCCCCCTCAAC	Amplification from middle of *welE* and identification of transcription start for P3
EF31	F	ACTGCATCTAAGCACAGACAAACTACCGTT	Sequencing and RT-PCR on *welE*
BP36	R	GGTCTTGGTCCAACCAAACTCAT	RT-PCR from *welE*
BP114CDRace1	R	CAAAACAGGCGGAGCATCTAGTAC	Identification of transcription start for P1
BP116CDRace2	R	AACCAGTCCGAGATTTACAAGTCCTAG	Identification of transcription start for P1
BP117MRace1	R	CTCATCTTGTTTTTCCTTTTCCTAACG	Identification of transcription start for P2
BP118MRace2	R	TTATTAGCAACTGCATAGAGCCCG	Identification of transcription start for P2
BP113eRace1	R	CACAATCCCGTAACACCAGG	Identification of transcription start for P3
BP120ERace2	R	AAAACTTAATTCGTTTCGTCCCGA	Identification of transcription start for P3
BP111aRace2	R	AAGTTGCTTTCTGGATGCTCAGG	Identification of transcription start for P4 and P5
BP112aRace3	R	AACTTGGACAAGCCGCTGCC	Identification of transcription start for P4 and P5
BP121ARace3	F	AACGGCATCTACCATCTGTTCTAGG	Identification of transcription start for P4 and P5
BP122ARace2	F	CAAGTTCTTAACCGACCCTAGCG	Identification of transcription start for P4 and P5
BP123ARace1	F	TGGATAACATCATGTGCCCT	Identification of transcription start for P4 and P5

*F, forward; R, reverse.

catalysing the first step in polysaccharide synthesis for strain ATCC 9595 (Provencher *et al.*, 2003) (GenBank accession no. AF323527). These primers were used for isolating one PCR product covering 3·5 kb upstream of the priming glycosyltransferase (*welE*) and 1·5 kb downstream. Inverse PCR was subsequently used to obtain the complete coding region of the EPS gene locus of *Lb. rhamnosus* ATCC 9595. Alignment of the DNA sequences obtained from the amplicons resulted in an 18·7 kb contiguous sequence (Fig. 2). Equally spaced primers (Table 2) were then designed along the length of this sequence for amplification and cloning of five fragments of about 4 kb each from strains RW-9595M, R and RW-6541M.

The DNA sequence of 18·7 kb was completely determined on both strands. The G+C content of this region

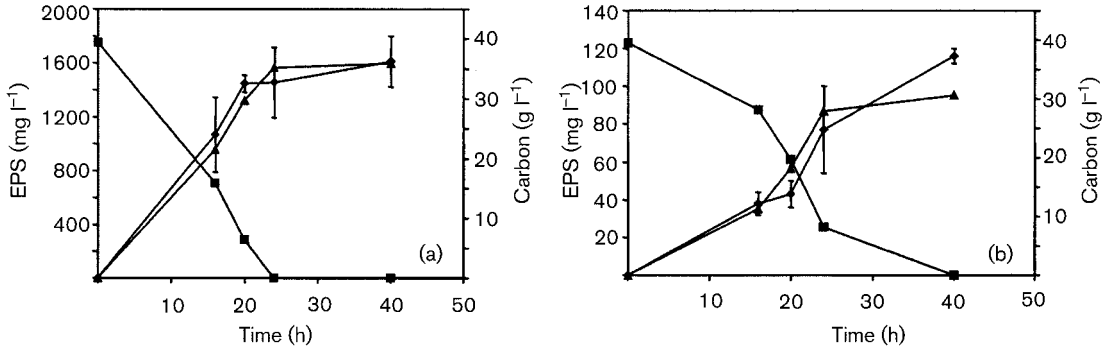

Fig. 1. EPS production (◆), residual glucose (■) and lactic acid accumulation (▲) during incubation of *Lb. rhamnosus* RW-9595M (a) and ATCC 9595 (b) in BMM medium containing 40 g glucose l^{-1} controlled at pH 6·0.

varied from 41·6 (strain R) to 42·1 mol% (RW-9595M), which is close to the typical 42–48 mol% G+C content reported for the *Lb. rhamnosus* genome (Sriranganathan *et al.*, 1985).

Sequence analysis and comparison

ORF analysis revealed the presence of 18 putative ORFs (Fig. 2). Overall pairwise comparison revealed that the four DNA regions are highly homologous, at 99 %. RW-9595M and ATCC 9595 have 18 bp different out of 18 749 bp, while RW-6541M and R show a difference of 34 out of 18 750 bp. Between strains of these two groups, there are 139–150 bp different. Upstream of the first ORF for 456 bp, the DNA sequences are 100 % identical for ATCC 9595 and RW-9595M, as well as for RW-6541M and R. Between these two groups, this segment is 98 % identical.

Of the predicted gene products, 17 are similar to proteins involved in the biosynthesis of various bacterial polysaccharides (Table 3). Based on these sequence similarities, a putative biological function could be attributed to most of the predicted proteins. Within the gene region encoding EPS biosynthesis, one ORF (*orf1*) encodes a truncated transposase of the IS*1165* family (Johansen & Kibenich, 1992). All ORFs were in the same orientation except for *orf1* and *wzr* (Fig. 2).

In the 5′ region, the first two ORFs (*wzd* and *wze*) have predicted amino acid sequences that most resemble proteins involved in putative polysaccharide-chain-length determination in *Leuconostoc mesenteroides* (39 % for Wzd versus Wzz) and *Lactobacillus delbrueckii* subsp. *bulgaricus* (*Lb. bulgaricus*; 42 % for Wze versus EpsC; Lamothe *et al.*, 2002). Prediction of membrane-spanning regions shows two putative transmembrane helices for Wzd [amino acids 18 (inside)–36 (outside) and amino acids 219 (outside)–238 (inside)]. The comparative analysis of Wze with various prokaryotic proteins known to autophosphorylate on tyrosine shows a Walker A [(AG)X_4GK(ST)] ATP-binding motif (Walker *et al.*, 1982), two Walker B (hhhD, where h represents a hydrophobic amino acid) magnesium-binding motifs and a series of tyrosine residues, termed the tyrosine cluster, located in the C-terminus. In the 3′ region of the cluster, gene *wzb* is located 154 bp downstream of *wzr*. Wzb has homology with the phosphotyrosine-protein phosphatase EpsD from *Lb. bulgaricus* (Lamothe *et al.*, 2002) and CpsB from *Streptococcus pneumoniae*, and is also proposed to be involved in chain-length determination (Morona *et al.*, 2002). The three gene products (*wzd*, *wze* and *wzb*) show 98–100 % identity among the four *Lb. rhamnosus* strains examined (Table 4), while the 19 bp intergenic region between *wzd* and *wze* is also identical.

The central portion of the locus is occupied by five of the six genes (*welF*, *welG*, *welH*, *welI*, *welJ* and *welE*) encoding potential glycosyltransferases (Tables 4 and 5). The *welE* gene is located farther along in the 3′ region, apart from the other glycosyltransferase genes, and displays 56 % identity with the predicted product Lp_1233 from *Lactobacillus plantarum* (Kleerebezem *et al.*, 2003). A variable degree of

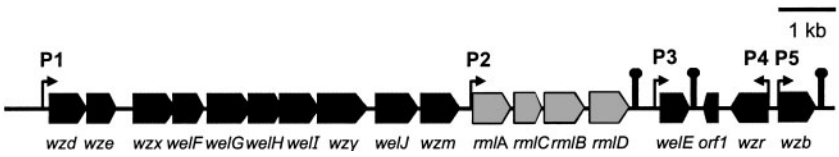

Fig. 2. Organization of the EPS biosynthesis gene cluster from *Lb. rhamnosus* strains RW-9595M, ATCC 9595, R and RW-6541M. Promoters (numbered from P1 to P5) and putative terminators are represented as flags and hairpins, respectively. Arrows represent potential ORFs and gene designations are indicated under the arrows. Grey arrows indicate ORFs involved in dTDP-rhamnose precursor biosynthesis.

Table 3. Profiles of predicted ORFs of the EPS biosynthesis gene locus of *Lb. rhamnosus* RW-9595M

ORF	G+C content (mol%)	No. of codons	Similar protein	Potential function	Bacterium with the most relevant identity	Protein sequence identity (%)	Accession no.
wzd	40·1	304	Wzz	Chain-length determinant protein	*Leuconostoc mesenteroides* subsp. *mesenteroides*	39	ZP_00063785
wze	42·7	248	EpsC	Putative role in chain length determination	*Lactobacillus delbrueckii* subsp. *bulgaricus*	42	AAG44707
wzx	38·5	396	Wzx	Putative involvement in EPS synthesis	*Streptococcus thermophilus*	31	AAL32506
welF	35·8	260	BT1340	Putative LPS glycosyltransferase	*Bacteroides thetaiotaomicron*	30	AAO76447
welG	37·2	359	WbbK	Putative glucose transferase	*Escherichia coli*	24	AAC75093
welH	39·2	272	RfbF	Rhamnosyl transferase	*Shigella flexneri*	28	AAP17467
welI	37·6	339	RgpB	Rhamnosyltransferase	*Lactococcus lactis* subsp. *lactis*	31	AAK04299
wzy	34·6	363	Eps11O	Putative polysaccharide polymerase	*Streptococcus thermophilus*	20	AAN63798
welJ	39·2	358	Cps1B	Glycosyltransferase	*Lactobacillus plantarum*	33	NP_784844
wzm	37·9	329	DVUA0070	Polysaccharide pyruvyl transferase (N-terminus) and *O*-methyltransferase (C-terminus)	*Desulfovibrio vulgaris* subsp. *vulgaris*	16	AAS94385
rmlA	44·1	290	Efae 03000460	COG1209 : dTDP-glucose pyrophosphorylase	*Enterococcus faecium*	72	ZP_00287334
rmlC	44·7	190	Ooen1401	COG1898 : dTDP-4-dehydrorhamnose 3,5-epimerase	*Oenococcus oeni*	77	ZP_00070380
rmlB	45·6	341	RmlB	dTDP-D-glucose-4,6-dehydratase	*Streptococcus gordonii*	75	AAN64546
rmlD	42·8	280	Orfde9	dTDP-4-keto-L-rhamnose reductase	*Enterococcus faecium*	57	AAC35923
welE	40·4	222	Lp_1233	Priming glycosyltransferase	*Lactobacillus plantarum*	56	NP_784894
orf1	43·7	106	IS1165	COG3464 : transposase (IS*1165*)	*Leuconostoc mesenteroides* subsp. *cremoris*	38	S23759
wzr	46·0	297	Ooen02000936	COG1316 : transcriptional regulator	*Oenococcus oeni*	46	ZP_00319536
wzb	47·7	254	EpsD	Putative role in chain-length determination	*Lactobacillus delbrueckii* subsp. *bulgaricus*	51	AAG44708

identity (38–100 %) was found between the WelE predicted product and the undecaprenyl-phosphate glycosyl-1-phosphate transferase (also known as the priming glycosyltransferase) found in several capsular and EPS-producing bacteria. The function of the priming glycosyltransferase has been demonstrated for EpsD from *Lactococcus lactis* (van Kranenburg *et al.*, 1999b). The predicted product of *welF* shows closest identity (30 % amino acids) to a putative glycosyltransferase from *Bacteroides thetaiotaomicron* (Xu *et al.*, 2003), while WelH and WelI show moderate identity with rhamnosyltransferases involved in LPS biosynthesis in *Shigella flexneri* SflI (28 % for WelH versus RfbF) (Wei *et al.*, 2003) and *Lactococcus lactis* subsp. *lactis* strain IL1403 (31 % for WelI versus RgpB) (Bolotin *et al.*, 2001), as well as to those involved in capsule biosynthesis by *St. pneumoniae* (Morona *et al.*, 1997), and in EPS biosynthesis by *St. thermophilus* CNRZ368 (Bourgoin *et al.*, 1999). According to the sequence alignments, WelF, WelH and WelI are members of the glycosyltransferase family 2 (Table 5) (Campbell *et al.*, 1997). WelG has the highest identity (24 % identical amino acids) with a putative glycosyltransferase from *E. coli* K-12 (Blattner *et al.*, 1997). WelJ exhibits most identity (33 % amino acids) with glycosyltransferase Cps1B from *Lb. plantarum* WCSF1 (Kleerebezem *et al.*, 2003), but also shows moderate homology with galactosyltransferase gene products found in *S. thermophilus* strains with different *eps* operon types (31 % with Eps3M, type III *eps* operon; GenBank accession no. AAL23736). WelG and WelJ are predicted to be members of the glycosyltransferase family 1 (Table 5) (Campbell *et al.*, 1997). All six putative glycosyltransferases could therefore be involved in the sequential biosynthesis of the repeating unit. The glycosidic linkage catalysed by each enzyme can be proposed according to the structure of the previously determined heptasaccharide (Fig. 3, Table 5). Among the four strains examined, there are few amino acid differences in the glycosyltransferase sequences (Table 4). Some of these changes result in a conserved amino acid (such as Ile to Leu or Val), and will not be detailed here. There are three instances of loss of an Asp or a Glu, residues that have been proposed as catalytic residues in glycosyltransferase sequences (Campbell *et al.*, 1997; Kapitonov & Yu, 1999). Strain R shows one amino acid difference in two glycosyltransferase gene products (Asp to Asn in WelG, and Glu to Gly in WelJ), while RW-6541M shows one amino acid change (Asp to Gly in WelI) in comparison with the corresponding sequences from the other three strains.

Table 4. Comparison of the predicted amino acid sequences of ORFs of the EPS gene clusters of four strains of *Lb. rhamnosus*

The number of different amino acids is given in parentheses.

Protein	Percentage identity with the amino acid sequence of the predicted protein product from RW-9595M		
	ATCC 9595	R	RW-6541M
Wzd	100 (0)	99 (1)	99 (1)
Wze	99 (1)	98 (5)	98 (4)
Wzx	99 (3)	92 (37)*	99 (4)
WelF	100 (0)	99 (1)	98 (4)
WelG	100 (0)	97 (8)	98 (7)
WelH	100 (0)	99 (2)	99 (2)
WelI	99 (1)	99 (1)	99 (2)
Wzy	100 (0)	100 (0)	100 (0)
WelJ	100 (0)	99 (2)	99 (1)
Wzm	100 (0)	98 (5)	98 (4)
RmlA	100 (0)	99 (2)	100 (0)
RmlC	98 (2)	98 (3)	98 (3)
RmlB	100 (0)	99 (1)	99 (1)
RmlD	99 (1)	99 (2)	99 (1)
WelE	100 (0)	99 (2)	99 (2)
Wzr	99 (2)	99 (4)	99 (4)
Wzb	100 (0)	100 (0)	99 (1)

*Thirty-four amino acids are truncated by a stop in Wzx of strain R.

Gene *wzm* could code for a pyruvyltransferase, as the predicted protein product has 16 % overall amino acid identity with a putative combined pyruvyltransferase/*O*-methyltransferase (DVUA0070) from *Desulfovibrio vulgaris* subsp. *vulgaris* (Heidelberg *et al.*, 2004). Wzm shows 27 % identity over the N-terminal 300 aa of DVUA0070. Among the four strains, Wzm shows 98–100 % identity (Table 4).

Gene *wzx* is located 377 bp downstream of *wze* in the 5′ region of the locus, and this 379 bp intergenic region is 98·6 (five nucleotides different) to 100 % identical among the four strains. The predicted amino acid sequence of *wzx* displays moderate identity (31 % amino acids) with Wzz from *S. thermophilus* strain MR-2C (Broadbent *et al.*, 2003). Similar proteins found in *eps* gene clusters from LAB are predicted to act as the EPS repeat-unit transporter, such as EpsM from *S. thermophilus* Sfi6, for example (Stingele *et al.*, 1996). Prediction of membrane-spanning regions shows 13 putative transmembrane helices (data not shown). Wzx is 98 % identical between RW-9595M, ATCC 9595 and RW-6541M, whereas the same gene product from strain R shows only 92 % identity with the other strains, because of a point mutation (G to A) inserting a stop codon which truncates the final 34 aa. Wzy has nine highly probable transmembrane helices, and shows some similarity (20 % identity) with the putative polysaccharide polymerase Eps11O from the *S. thermophilus* type XI *eps* operon. Wzy is 100 % identical among the four strains. The 16 bp intergenic region between *wzy* and *welJ* is also 100 % identical.

Four genes (*rmlA–rmlD*) were found in the central region of the locus, 191 bp downstream of *wzm*, and 826 bp upstream of *welE*. Considering the high degree of identity (Table 3), the predicted functions of these four gene products are respectively glucose-1-phosphate thymidylyltransferase (also known as dTDP-glucose pyrophosphorylase), dTDP-D-glucose 4,6-dehydratase, dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase and dTDP-6-deoxy-L-lyxo-4-hexulose reductase (Table 3). These enzymes would be responsible for the production of the rhamnose precursor essential for

Table 5. Properties of the proposed products encoded by the glycosyltransferase genes

Protein	Putative sugar specificity	Glycosidic linkage	Family*
WelE	Glucose	–	Phosphoglycosyltransferase†
WelF	Rhamnose	α(1,3)	Glycosyltransferase 2 family‡
WelG	Glucose	α(1,3)	Glycosyltransferase 1 family§
WelH	Rhamnose	α(1,2)	Glycosyltransferase 2 family
WelI	Rhamnose	α(1,3)	Glycosyltransferase 2 family
WelJ	Galactose	α(1,2)	Glycosyltransferase 1 family

*Family designation is based on similarities with glucosyltransferases deposited in GenBank or in the CAZy database (http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html).

†Phosphoglycosyltransferase proteins transfer the first sugar nucleotide to the undecaprenyl phosphate lipid carrier (Pfam database; Bateman *et al.*, 2004; accession no. PF02397).

‡Family 2 (CAZy) contains proteins transferring sugar from UDP-glucose, UDP-*N*-acetylgalactosamine, GDP-mannose, TDP-rhamnose or cytidine-5′-abequose to a range of substrates, including cellulose, dolichol phosphate and teichoic acid.

§Family 1 (CAZy) proteins transfer activated sugars to a variety of substrates, including glycogen, fructose 6-phosphate and LPS.

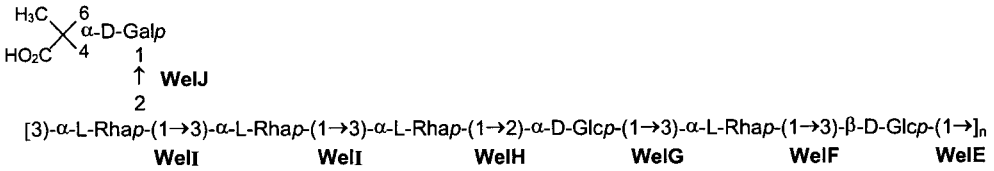

Fig. 3. Structure of the repeating unit of the EPS produced by *Lb. rhamnosus* RW-9595M and R (Van Calsteren *et al.*, 2002). Gene products predicted to carry out the incorporation of each monosaccharide residue are indicated. Glc, glucose; Rha, rhamnose; Gal, galactose; *p*, pyranose.

EPS biosynthesis. Southern hybridization analysis using these four genes as a probe revealed that the *Lb. rhamnosus* genome has only one copy of these genes (data not shown). Among the four strains, the four gene products are 98–100 % identical (Table 4). The intergenic regions of *rmlAC* (14 bp) as well as *rmlBD* (61 bp) are 100 % identical among the four strains.

In the 3′ region of the EPS locus, gene *wzr* is organized in the opposite transcriptional sense. The predicted product Wzr shows closest identity (46 %) to the putative transcription regulator Ooen02000936 from *Oenococcus oeni*, and shows 31 % identity with EpsA from *Lb. bulgaricus* (cell-envelope-related transcriptional attenuator; GenBank accession no. AAG44705) (Lamothe *et al.*, 2002). The *wzr* sequence from RW-9595M differs from the remaining three strains by a Pro residue in position 44 instead of Ser, and Phe instead of Leu in position 139. As these changes in amino acids could result in changes in protein secondary structure (for phenylalanine) or post-translational modification (Ser phosphorylation), they may have an impact on the function of this potential regulator. A short interrupted ORF, designated *orf1*, is located just upstream of *wzr*. The translated product of *orf1* is similar to a transposase of *Leuconostoc mesenteroides* subsp. *cremoris* IS1165 (38 % identity over 94 aa of translated *orf1*) (Johansen & Kibenich, 1992).

Transcriptional analysis of the EPS biosynthesis gene cluster

Analysis of the DNA sequence suggests that the EPS gene cluster of *Lb. rhamnosus* is organized into five transcriptional units (Fig. 2). Five putative promoters with −35 and −10 sequences were identified with 23, 50, 63, 33 and 25 bp spacing between the ends of the −10 sequence and the start codon of ORFs *wzd*, *rmlA*, *welE*, *wzr* and *wzb*, respectively. For each of the five promoters, the sequences were 100 % identical among the four *Lb. rhamnosus* strains. Overlapping reading frames could be identified at the end of *wzx* and the beginning of *welF* (4 bp overlap), at the end of *welG* and the beginning of *welH* (11 bp overlap), at the end of *welH* and the beginning of *welI* (20 bp overlap), and finally at the end of *welI* and the beginning of *wzy* (1 bp overlap). This analysis suggests the transcriptional coupling of these six genes. Putative rho-independent transcription terminator sequences were identified downstream of *rmlD* (−17·6 kcal mol^{-1}, −73·6 kJ mol^{-1}), *welE* (−12·2 kcal mol^{-1}, −51·0 kJ mol^{-1}) and *wzb* (−29·5 kcal mol^{-1}, −123·4 kJ mol^{-1}). The putative terminator of *welE* could also be the putative terminator of *wzr*, which is transcribed in the opposite sense. To confirm some of these hypotheses, total RNA was isolated from ATCC 9595 at two different growth phases (OD_{600} 0·6 and 3), and used for RT-PCR and 5′ RACE experiments.

Successive overlapping RT-PCR reactions were carried out to determine the whether the EPS genes are co-transcribed in an identical fashion for strains ATCC 9595 and RW-9595M. Specific amplifications were obtained for each RT-PCR carried out (Fig. 4), indicating that all genes from *wzd* to *welE* (inclusive) are transcribed as a single 15·4 kb polycistronic mRNA from a promoter sequence upstream of *wzd*. In particular, an amplification fragment was obtained by the RT-PCR which covered the 191 bp intergenic region between *wzm* and *rmlA* (Fig. 4c, lane 5). This region is identical between RW-9595M and ATCC 9595, and varies by only 3 or 4 nucleotides from the same region in strains RW-6541M and R. Another amplified fragment covered the region from *rmlD* to *welE* (Fig. 4a and b, lane 4). The 827 bp intergenic space varied by 3 nucleotides (between RW-9595M and ATCC 9595) to 5 nucleotides (RW-9595M with R and RW-6541M). In the 3′ region, two single-gene (*wzr* and *wzb*) RT-PCR reactions gave positive results, but no amplification was obtained in this region when two or mores genes were combined (data not shown). This 157 bp intergenic region varied by 0–2 nucleotides among the four strains.

Determination of transcriptional start sites

Experiments by 5′ RACE were performed on total RNA isolated from the early stationary growth phase of a *Lb. rhamnosus* ATCC 9595 culture. The transcription start of each of the five putative promoters was detected in front of the putative RBS site using antisense oligonucleotides derived from the 5′ sequence of *wzd*, *rmlA*, *welE*, *wzr* and *wzb* (Table 2). Upstream of each transcription start, putative consensus sequences for the −35 and −10 regions were located (Fig. 5), and these are in good agreement with the consensus sequence of vegetative promoters identified in *Lb. rhamnosus* (McCracken *et al.*, 2000). Conserved bases corresponding to the consensus AT-rich UP element (upstream enhancer element; Estrem *et al.*, 1998) were

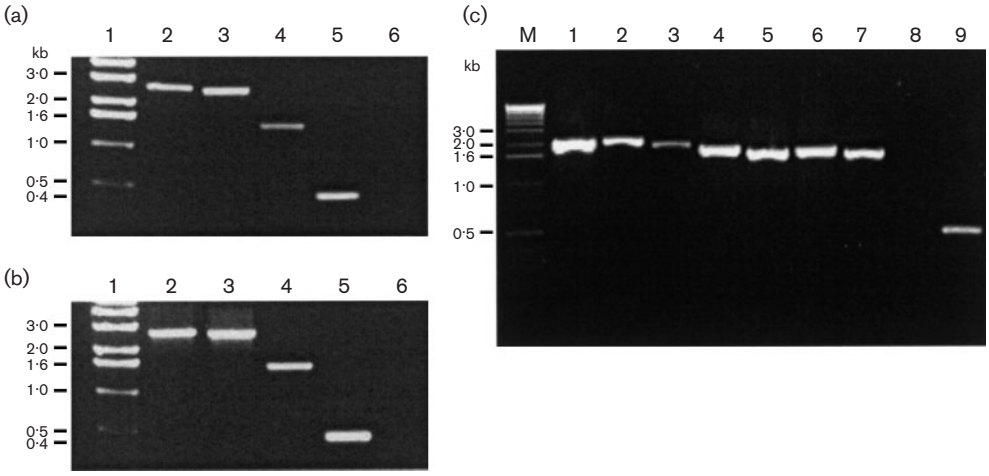

Fig. 4. Transcription analysis by RT-PCR of EPS genes of *Lb. rhamnosus* strains ATCC 9595 (a) and RW-9595M (b) carried out on RNA isolated after incubation for 6 h at 30 °C. Lanes: 1, the 1 kb ladder (Invitrogen); 2, *wzd–welF*; 3, *rmlA–rmlD*; 4, *rmlD–welE*; 5, *welE* alone; 6, negative control (PCR without the RT step). (c) RT-PCR of *wzd–welE* transcripts from *Lb. rhamnosus* ATCC 9595. Lanes: 1, *wzd–wzx*; 2, is *wzx–welG*; 3, *welG–welI*; 4, *welI–welJ*; 5, *welJ–rmlA*; 6, *rmlA–rmlB*; 7, *rmlB–rmlD*; 8, negative control; 9, *welE*. Lane M: 1 kb ladder (Invitrogen Life Technologies).

identified upstream of the −35 hexamer of each putative promoter.

The putative P1 promoter region is at the appropriate distance from the transcriptional starting point determined by 5′ RACE (Fig. 5). Analysis of the region upstream of the first promoter, P1, of the *Lb. rhamnosus* EPS gene clusters reveals three significant motifs. First, a sequence was found almost 200 bases upstream of promoter P1 that differed by two nucleotides (ATAAAAGTTTTACA) from the *cre* site (ATAAAACGTTTACA) of the lactose operon of *Lactobacillus casei* (Gosalbes *et al.*, 1997). Secondly, an UP element (Estrem *et al.*, 1998) is positioned at one DNA turn further upstream than in the other *Lactobacillus* spp. promoters identified to date (McCracken *et al.*, 2000). Third, a putative regulation site was identified overlapping the −35 hexamer. A symmetrical TGA half-site motif is localized 5 bp upstream of the −35 hexamer, and a TCA half-site motif is localized in the −35 hexamer. This motif (TTTGActagctTCACA) displays only one nucleotide difference from the CRP consensus sequence of *E. coli* (TG**TGA**NNNNNN**TCA**CA) (Gaston *et al.*, 1989), some deviation from the FLP consensus sequence of *Lb. casei* ($C^{A}/_{C}$**TGA**NNNN**TCA**$^{G}/_{T}$G) (Gostick *et al.*, 1998), and three nucleotides different from the FNR consensus sequence from *E. coli* (T**TGA**TNNNNA**TCA**A) (Spiro & Guest, 1990).

DISCUSSION

This is the first comparative analysis of EPS-related genes for *Lb. rhamnosus*, to our knowledge. *Lb. rhamnosus* strain RW-9595M has been shown to produce up to 2350 EPS mg l^{-1} after 32 h of pH-controlled (pH 6) growth at 37 °C in supplemented whey permeate medium (Bergmaier *et al.*, 2003), which is the highest quantity recorded to date for heteropolysaccharide production by LAB. This strain is a spontaneous variant of *Lb. rhamnosus* ATCC 9595. There is very low probability that the EPS produced by *Lb.*

```
Consensus sequence                GxTTraca                    TAtaaT        +1

P1     TAGTATTTTGTTGAGAAGTTTTGACTAGCTTcACATGAGTAGGAT...GAATTATATAAgA.....TGGGATTAGGTGGG
P2 GAACTTTTTTTAAATGATAGGGAAAAGCAAACTgGtaATGATGGCTTT..GTTGGGGTATtgTTT....TGTTGGTCATATGA
P3 AGATTTCACACTGGTTAGTGTACAAATTAAGGTTGctACTACATTAGG..AATGTTATAagATTA....CTTGAACAGAGAAT
P4       TTGTGTTTGTTAACGTTTCACAATAGATcAgCgAGCTGGCAGA.CTTTGCGGTATAATATG....CAAAGATTACGGA
P5                 ACTGTTACAAAAAAATTTGACcTTGGGAGCGC..TTCATTAgATAcTT.....AGATAGTATTGGAA
```

Fig. 5. Comparative analysis of nucleotide sequences and transcriptional start sites of promoters located in the EPS biosynthesis locus from *Lb. rhamnosus* ATCC 9595. Sequences of −35 and −10 hexamers are underlined and compared with their respective consensus sequences **G**x**T**Traca and **TA**taa**T** (**T** > 75 %, T 60–75 %, t 40–59 %, r represents adenosine or guanosine, and x represents any nucleotide) (McCracken *et al.*, 2000). Transcription start sites are indicated under +1 as underlined single bold A or G nucleotides. Dots are used for alignment of −10 hexamer sequences and +1 start sites. The sequence similar to CRP (Gaston *et al.*, 1989) is boxed.

rhamnosus ATCC 9595 differs in structure from the EPS secreted by the variant (M. R. Van Calsteren, personal communication), especially given the high conservation in glycosyltransferase gene products among the strains examined. The presence of a few amino acid differences does not have an effect on enzyme specificity, as the same EPS structure is produced by strains RW-9595M, R (Van Calsteren *et al.*, 2002) and RW-6541M (M. R. Van Calsteren, personal communication). There is thus a high correlation between glycosyltransferase gene presence and EPS repeat-unit structure that can be used in future screening of *Lb. rhamnosus* strains for novel EPS structures. The variation in EPS production levels observed among the *Lb. rhamnosus* strains studied here may originate from differences in central metabolism and the availability of sugar precursors. Studies are presently under way to correlate differences in EPS production among *Lb. rhamnosus* strains with differences in the activity of enzymes of central metabolism and precursor biosynthesis.

The genetic organization and transcription of the EPS gene clusters from *Lb. rhamnosus* show very important differences from those of the gene clusters involved in the synthesis of capsular and free EPS in Gram-positive bacteria such as *S. thermophilus* (Stingele *et al.*, 1996), *Lactococcus lactis* subsp. *cremoris* (van Kranenburg *et al.*, 1997), *Lb. bulgaricus* (Lamothe *et al.*, 2002) and *Lactobacillus helveticus* (accession no. AX009415) (Germond *et al.*, 1999). Fifteen genes are co-transcribed on one long mRNA, as has been reported for other EPS biosynthesis gene clusters. However, transcription of the EPS genes of *Lb. rhamnosus* may be driven by five different promoters, as five start sites were determined. Although these start sites are appropriately spaced from promoter sequences, there is a possibility that they may be formed by mRNA processing. This indicates the potential for a complex system for regulating the expression of these genes in *Lb. rhamnosus*.

Six genes which probably encode the glycosyltransferases responsible for the structural organization of the heptasaccharide repeating unit are co-transcribed with genes involved in the polymerization and export of the EPS. In all EPS biosynthesis gene clusters reported to date, the priming glycosyltransferase gene is located just upstream of all other glycosyltransferase genes. In the *Lb. rhamnosus* strains examined, the *welE* gene encoding the first glycosyltransferase is located after the other glycosyltransferase genes, but is also transcribed from its own promoter, P3. This suggests an independent expression of the priming glycosyltransferase that may lead to higher amounts of mRNA, and thus to an improved initiation of EPS repeat-unit biosynthesis through the translation of more priming glycosyltransferase proteins.

The carbohydrate linkage arrangement of the heptasaccharide repeating unit from *Lb. rhamnosus* strains R and RW-9595M was determined previously (Van Calsteren *et al.*, 2002). The order of the monosaccharides represents the proposed sequential addition of sugars during biosynthetic assembly of the repeating unit (Fig. 3), beginning with WelE, which transfers the first glucose onto the lipophilic carrier molecule. The glycosyltransferases carrying out the next reactions form an ordered biosynthetic complex that reflects the order of the genes. Only three putative rhamnosyltransferases (WelF, WelH and WelI) were found for the addition of four rhamnose residues to the repeating unit. We hypothesize that the same rhamnosyltransferase (WelI) could accomplish the addition of two rhamnose residues in the fifth and sixth positions by catalysing the formation of identical $\alpha(1\rightarrow3)$ glycosidic linkages. In fact, previous studies reported two mannosyltransferases (MtfA and MtfB) from *E. coli* that transfer mannose residues, respectively forming two $\alpha(1\rightarrow3)$ and three $\alpha(1\rightarrow2)$ bonds in the O9 antigen biosynthesis pathway (Kido *et al.*, 1995). Once the hexasaccharide of the backbone is assembled, WelJ, the $\alpha(1\rightarrow2)$ galactosyltransferase, could add one branching sugar to form the final heptasaccharide repeat unit. Finally, Wzm may be the pyruvyltransferase that adds the pyruvate substituent to this galactose residue.

Genes involved in chain-length determination are not organized at the beginning of the cluster, but are divided into two regions: *wzd* and *wze* are in the 5′ region of the cluster, while *wzb* is located at the end of the cluster. Given the similarity in predicted protein products of these three genes with other gene products involved in chain-length determination, the model of regulating EPS biosynthesis at the post-transcriptional level described for *S. pneumoniae* (Bender *et al.*, 2003; Morona *et al.*, 2000, 2003) is probably applicable to a certain degree to *Lb. rhamnosus* as well. All CpsD-specific motifs and residues involved in the regulating model of *S. pneumoniae* were also found in Wze (Walker A and Walker B ATP-binding motifs, tyrosine cluster). These motifs are usually found in protein kinases involved in the biosynthesis of EPS by Gram-negative and Gram-positive bacteria. Bender *et al.* (2003) showed that deletion of *cps2C* or *cps2D* led to short-chain polymers instead of long polysaccharide molecules in *S. pneumoniae*.

Four genes (*rmlA–rmlD*) exhibit significant similarity to *rml* gene products involved in the anabolism of dTDP-L-rhamnose from α-D-glucose 1-phosphate. EPS biosynthesis gene clusters of LAB do not generally contain these four genes, while *cps* clusters often do (Morona *et al.*, 1997). For *S. pneumoniae*, *rml* gene expression seems to be controlled by the promoter of the *cps* cluster (Iannelli *et al.*, 1999). For *Lb. rhamnosus*, we observed that the transcription of these four genes could be controlled by two different promoters, as they are cotranscribed on a large 15·2–15·4 kb mRNA from *wzd* to *welE*, and also independently from promoter P2. Expression of *rmlA–D* genes might be activated independently of the EPS biosynthesis genes to form the cell wall polysaccharides, but cotranscription of these genes with the EPS biosythesis gene cluster might increase the dTDP-L-rhamnose pool available for biosynthesis of EPS. Location of genes involved in rhamnose precursor biosynthesis within this locus, and their coordinated expression,

suggest the importance of this monosaccharide for biosynthesis of the *Lb. rhamnosus* EPS structure, where four out of seven monosaccharides of the repeat unit are rhamnose.

The four *Lb. rhamnosus* strains produce very different quantities of EPS under the same conditions (from 61 to 1611 mg l^{-1}), yet the gene clusters show remarkably little variation, even in all the promoter regions, including the region upstream of promoter P1. The potential *cre* sequence found upstream of P1 shows significant differences from the consensus *cre* site (TGWNANCGNTNWCA) (Weickert & Chambliss, 1990), the target of the CcpA regulator mediating catabolite repression and carbon flow in Gram-positive bacteria (Henkin, 1996). The central CG nucleotides of the *cre* site are essential for interacting with the CcpA complex (Weickert & Chambliss, 1990), but are degenerated in the sequence found upstream of the P1 promoter. As a consequence, catabolite repression does not appear to be a likely mechanism for regulating gene expression of the *Lb. rhamnosus* EPS biosynthesis locus.

In conclusion, comparative analysis of EPS biosynthesis gene clusters from strains that produce different EPS levels is the first step to providing new insight into this industrially relevant phenotype. *Lb. rhamnosus* strains show remarkable variation in EPS production levels, but little variation in EPS gene sequences. The presence of six genes encoding potential glycosyltransferases responsible for the structural organization correlated well with the assembly of the heptasaccharide repeating unit structure previously determined by our team. In addition, four genes were identified within the EPS biosynthesis locus that might result in coordinated production of the rhamnose precursor with the enzymes involved in EPS biosynthesis. Future studies will aim for functional analysis of the EPS gene products and their interactions.

ACKNOWLEDGEMENTS

The authors thank the Natural Sciences and Engineering Research Council of Canada (Research Partnerships Program–Research Network on Lactic Acid Bacteria), Agriculture and Agri-Food Canada, Novalait of Quebec, Dairy Farmers of Canada and the Institute Rosell for financial support. We gratefully acknowledge Dr M. Frenette and Dr E. Emond for valuable contributions and discussions, as well as I. Pelletier for technical assistance. The authors are also grateful to A. Parent for carrying out EPS production and quantification, and to C. Provencher for contributing to this project from the beginning.

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Ausubel, F. M. (1995). *Short Protocols in Molecular Biology: a Compendium of Methods from Current Protocols in Molecular Biology*, 3rd edn. New York: Wiley.

Bateman, A., Coin, L., Durbin, R. & 10 other authors (2004). The Pfam protein families database. *Nucleic Acids Res* **32**, D138–141.

Bender, M. H., Cartee, R. T. & Yother, J. (2003). Positive correlation between tyrosine phosphorylation of CpsD and capsular polysaccharide production in *Streptococcus pneumoniae*. *J Bacteriol* **185**, 6057–6066.

Bergmaier, D., Champagne, C. P. & Lacroix, C. (2003). Exopolysaccharide production during batch cultures with free and immobilized *Lactobacillus rhamnosus* RW-9595M. *J Appl Microbiol* **95**, 1049–1057.

Blattner, F. R., Plunkett, G., III, Bloch, C. A. & 14 other authors (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474.

Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S. D. & Sorokin, A. (2001). The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res* **11**, 731–753.

Bourgoin, F., Pluvinet, A., Gintz, B., Decaris, B. & Guedon, G. (1999). Are horizontal transfers involved in the evolution of the *Streptococcus thermophilus* exopolysaccharide synthesis loci? *Gene* **233**, 151–161.

Broadbent, J. R., McMahon, D. J., Welker, D. L., Oberg, C. J. & Moineau, S. (2003). Biochemistry, genetics, and applications of exopolysaccharide production in *Streptococcus thermophilus*: a review. *J Dairy Sci* **86**, 407–423.

Campbell, J. A., Davies, G. J., Bulone, V. & Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem J* **326**, 929–939.

Cerning, J., Renard, C. M. G. C., Thibault, J. F., Bouilliane, C., Landon, M., Desmazeaud, M. & Topisirovic, L. (1994). Carbon source requirements for exopolysaccharide production by *Lactobacillus casei* CG11 and partial structure analysis of the polymer. *Appl Environ Microbiol* **60**, 3914–3919.

Chabot, S., Yu, H.-L., De Léséleuc, L., Cloutier, D., Van Calsteren, M.-R., Lessard, M., Roy, D., Lacroix, M. & Oth, D. (2001). Exopolysaccharides from *Lactobacillus rhamnosus* RW-9595M stimulate TNF, IL-6 and IL-12 in human and mouse cultured immunocompetent cells, and IFN-gamma in mouse splenocytes. *Lait* **81**, 683–697.

De Man, J. C., Rogosa, M. & Sharpe, M. E. (1960). A medium for the cultivation of lactobacilli. *J Appl Bacteriol* **23**, 130–135.

De Vuyst, L. & Degeest, B. (1999). Heteropolysaccharides from lactic acid bacteria. *FEMS Microbiol Rev* **23**, 153–177.

De Vuyst, L., De Vin, F., Vaningelgem, F. & Degeest, B. (2001). Recent developments in the biosynthesis and applications of heteropolysaccharides from lactic acid bacteria. *Int Dairy J* **11**, 687–707.

Dubois, M., Gilles, K., Hamilton, J. K., Rebers, P. A. & Smith, F. (1951). A colorimetric method for the determination of sugars. *Nature* **168**, 167.

Dupont, I., Roy, D. & LaPointe, G. (2000). Comparison of exopolysaccharide production by strains of *Lactobacillus rhamnosus* and *Lactobacillus paracasei* grown in chemically defined medium and milk. *J Ind Microbiol Biotechnol* **24**, 251–255.

Estrem, S. T., Gaal, T., Ross, W. & Gourse, R. L. (1998). Identification of an UP element consensus sequence for bacterial promoters. *Proc Natl Acad Sci U S A* **95**, 9761–9766.

Farres, J., Caminal, G. & Lopez-Santin, J. (1997). Influence of phosphate on rhamnose-containing exopolysaccharide rheology and production by *Klebsiella* I-714. *Appl Microbiol Biotechnol* **48**, 522–527.

Gaston, K., Kolb, A. & Busby, S. (1989). Binding of the *Escherichia coli* cyclic AMP receptor protein to DNA fragments containing consensus nucleotide sequences. *Biochem J* **261**, 649–653.

Germond, E., Lamothe, G. & Stingele, F. (1999). Lactic acid bacteria producing exopolysaccharides. European Patent WO9962316.

Gosalbes, M. J., Monedero, V., Alpert, C. A. & Pérez-Martinez, G. (1997). Establishing a model to study the regulation of the lactose operon in *Lactobacillus casei*. *FEMS Microbiol Lett* **148**, 83–89.

Gostick, D. O., Green, J., Irvine, A. S., Gasson, M. J. & Guest, J. R. (1998). A novel regulatory switch mediated by the FNR-like protein of *Lactobacillus casei*. *Microbiology* **144**, 705–717.

Heidelberg, J. F., Seshadri, R., Haveman, S. A. & 32 other authors (2004). The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. *Nat Biotechnol* **22**, 554–559.

Henkin, T. M. (1996). The role of CcpA transcriptional regulator in carbon metabolism in *Bacillus subtilis*. *FEMS Microbiol Lett* **135**, 9–15.

Hosono, A., Lee, J., Ametani, A., Natsume, M., Hirayama, M., Adachi, T. & Kaminogawa, S. (1997). Characterization of a water-soluble polysaccharide fraction with immunopotentiating activity from *Bifidobacterium adolescentis* M101-4. *Biosci Biotechnol Biochem* **61**, 312–316.

Iannelli, F., Pearce, B. J. & Pozzi, G. (1999). The type 2 capsule locus of *Streptococcus pneumoniae*. *J Bacteriol* **181**, 2652–2654.

Johansen, E. & Kibenich, A. (1992). Isolation and characterization of IS1165, an insertion sequence of *Leuconostoc mesenteroides* subsp. *cremoris* and other lactic acid bacteria. *Plasmid* **27**, 200–206.

Jolly, L. & Stingele, F. (2001). Molecular organization and functionality of exopolysaccharide gene clusters. *Int Dairy J* **11**, 733–745.

Jolly, L., Newell, J., Porcelli, I., Vincent, S. J. & Stingele, F. (2002a). *Lactobacillus helveticus* glycosyltransferases: from genes to carbohydrate synthesis. *Glycobiology* **12**, 319–327.

Jolly, L., Vincent, S. J., Duboc, P. & Neeser, J. R. (2002b). Exploiting exopolysaccharides from lactic acid bacteria. *Antonie Van Leeuwenhoek* **82**, 367–374.

Kapitonov, D. & Yu, R. K. (1999). Conserved domains of glycosyltransferases. *Glycobiology* **9**, 961–978.

Kido, N., Torgov, V. I., Sugiyama, T., Uchiya, K., Sugihara, H., Komatsu, T., Kato, N. & Jann, K. (1995). Expression of the O9 polysaccharide of *Escherichia coli*: sequencing of the *E. coli* O9 *rfb* gene cluster, characterization of mannosyl transferases, and evidence for an ATP-binding cassette transport system. *J Bacteriol* **177**, 2178–2187.

Kitazawa, H., Toba, T., Itoh, T., Kumano, N., Adachi, S. & Yamaguchi, T. (1991). Antitumoral activity of slime-forming encapsulated *Lactococcus lactis* subsp. *cremoris* isolated from Scandinavian ropy sour milk, 'viili'. *Anim Sci Technol* **62**, 277–283.

Kleerebezem, M., Boekhorst, J., van Kranenburg, R. & 17 other authors (2003). Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc Natl Acad Sci U S A* **100**, 1990–1995.

Lamothe, G. T., Jolly, L., Mollet, B. & Stingele, F. (2002). Genetic and biochemical characterization of exopolysaccharide biosynthesis by *Lactobacillus delbrueckii* subsp. *bulgaricus*. *Arch Microbiol* **178**, 218–228.

McCracken, A., Turner, M. S., Giffard, P., Hafner, L. M. & Timms, P. (2000). Analysis of promoter sequences from *Lactobacillus* and *Lactococcus* and their activity in several *Lactobacillus* species. *Arch Microbiol* **173**, 383–389.

Miller, J. H. (1992). *A Short Course in Bacterial Genetics: a Laboratory Manual and Handbook for Escherichia coli and Related Bacteria*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Morishita, T., Deguchi, Y., Yajima, M., Sakurai, T. & Yura, T. (1981). Multiple nutritional requirements of lactobacilli: genetic lesions affecting amino acid biosynthesis pathways. *J Bacteriol* **148**, 64–71.

Morona, J. K., Morona, R. & Paton, J. C. (1997). Characterization of the locus encoding the *Streptococcus pneumoniae* type 19F capsular polysaccharide biosynthetic pathway. *Mol Microbiol* **23**, 751–763.

Morona, J. K., Paton, J. C., Miller, D. C. & Morona, R. (2000). Tyrosine phosphorylation of CpsD negatively regulates capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. *Mol Microbiol* **35**, 1431–1442.

Morona, J. K., Morona, R., Miller, D. C. & Paton, J. C. (2002). *Streptococcus pneumoniae* capsule biosynthesis protein CpsB is a novel manganese-dependent phosphotyrosine-protein phosphatase. *J Bacteriol* **184**, 577–583.

Morona, J. K., Morona, R., Miller, D. C. & Paton, J. C. (2003). Mutational analysis of the carboxy-terminal (YGX)4 repeat domain of CpsD, an autophosphorylating tyrosine kinase required for capsule biosynthesis in *Streptococcus pneumoniae*. *J Bacteriol* **185**, 3009–3019.

Nakajima, H., Hirota, T., Toba, T., Itoh, T. & Adachi, S. (1992). Structure of the extracellular polysaccharide from slime-forming *Lactococcus lactis* subsp. *cremoris* SBT 0495. *Carbohydr Res* **224**, 245–253.

Paul, F., Morin, A. & Monsan, P. (1986). Microbial polysaccharides with actual potential industrial applications. *Biotechnol Adv* **4**, 245–259.

Péant, B. & LaPointe, G. (2004). Identification and characterization of a conserved nuclease secreted by strains of the *Lactobacillus casei* group. *J Appl Microbiol* **96**, 367–374.

Provencher, C., LaPointe, G., Sirois, S., Van Calsteren, M. R. & Roy, D. (2003). Consensus-degenerate hybrid oligonucleotide primers for amplification of priming glycosyltransferase genes of the exopolysaccharide locus in strains of the *Lactobacillus casei* group. *Appl Environ Microbiol* **69**, 3299–3307.

Reeves, P. R., Hobbs, M., Valvano, M. A. & 8 other authors (1996). Bacterial polysaccharide synthesis and gene nomenclature. *Trends Microbiol* **4**, 495–503.

Ricciardi, A. & Clementi, F. (2000). Exopolysaccharides from lactic acid bacteria: structure, production and technological applications. *Ital J Food Sci* **12**, 23–45.

Roberts, I. S. (1996). The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu Rev Microbiol* **50**, 285–315.

Ruas-Madiedo, P., Hugenholtz, J. & Zoon, P. (2002). An overview of the functionality of exopolysaccharides produced by lactic acid bacteria. *Int Dairy J* **12**, 163–171.

Spiro, S. & Guest, J. R. (1990). FNR and its role in oxygen-regulated gene expression in *Escherichia coli*. *FEMS Microbiol Rev* **6**, 399–428.

Sriranganathan, N., Seidler, R. J. & Sandine, W. E. (1985). Nucleic acids of species of *Lactobacillus*. *J Dairy Sci* **68**, 1077–1086.

Stingele, F., Neeser, J. R. & Mollet, B. (1996). Identification and characterization of the *eps* (exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6. *J Bacteriol* **178**, 1680–1690.

Van Calsteren, M. R., Pau-Roblot, C., Begin, A. & Roy, D. (2002). Structure determination of the exopolysaccharide produced by *Lactobacillus rhamnosus* strains RW-9595M and R. *Biochem J* **363**, 7–17.

van Kranenburg, R., Marugg, J. D., van Swam, I. I., Willem, N. J. & de Vos, W. M. (1997). Molecular characterization of the plasmid-encoded *eps* gene cluster essential for exopolysaccharide biosynthesis in *Lactococcus lactis*. *Mol Microbiol* **24**, 387–397.

van Kranenburg, R., Boels, I. C., Kleerebezem, M. & de Vos, W. M. (1999a). Genetics and engineering of microbial exopolysaccharides for food: approaches for the production of existing and novel polysaccharides. *Curr Opin Biotechnol* **10**, 498–504.

van Kranenburg, R., Vos, H. R., van Swam, I. I., Marugg, J. D., Kleerebezem, M. & de Vos, W. M. (1999b). Exopolysaccharide biosynthesis in *Lactococcus lactis* NIZO B40: functional analysis of the glycosyltransferase genes involved in synthesis of the polysaccharide backbone. *J Bacteriol* **181**, 338–340.

Vincent, D., Roy, D., Mondou, F. & Dery, C. (1998). Characterization of bifidobacteria by random DNA amplification. *Int J Food Microbiol* **43**, 185–193.

Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**, 945–951.

Wei, J., Goldberg, M. B., Burland, V. & 14 other authors (2003). Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect Immun* **71**, 2775–2786.

Weickert, M. J. & Chambliss, G. H. (1990). Site-directed mutagenesis of a catabolite repression operator sequence in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **87**, 6238–6242.

Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V. & Gordon, J. I. (2003). A genomic view of the human–*Bacteroides thetaiotaomicron* symbiosis. *Science* **299**, 2074–2076.