Carbohydrate Utilization in Streptococcus thermophilus

Poolman, Bert; Royer, Theresa J.; Mainzer, Stanley E.; Schmidt, Brian F.

Published in:
Journal of Bacteriology

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1990

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Carbohydrate Utilization in *Streptococcus thermophilus*: Characterization of the Genes for Aldose 1-Epimerase (Mutarotase) and UDPglucose 4-Epimerase

BERT POOLMAN,†* THERESA J. ROYER, STANLEY E. MAINZER, AND BRIAN F. SCHMIDT

GENENCOR Inc., South San Francisco, California 94080

Received 12 February 1990/Accepted 26 April 1990

The complete nucleotide sequences of the genes encoding aldose 1-epimerase (mutarotase) (*galM*) and UDPglucose 4-epimerase (*galE*) and flanking regions of *Streptococcus thermophilus* have been determined. Both genes are located immediately upstream of the *S. thermophilus lac operon*. To facilitate the isolation of *galE*, a special polymerase chain reaction-based technique was used to amplify the region upstream of *galM* prior to cloning. The *galM* protein was homologous to the mutarotase of *Acinetobacter calcoaceticus*, whereas the *galE* protein was homologous to UDPglucose 4-epimerase of *Escherichia coli* and *Streptomyces lividans*. The amino acid sequences of *galM* and *galE* proteins also showed significant similarity with the carboxy-terminal and amino-terminal domains, respectively, of UDPglucose 4-epimerase from *Kluyveromyces lactis* and *Saccharomyces cerevisiae*, suggesting that the yeast enzymes contain an additional, yet unidentified (mutarotase) activity. In accordance with the open reading frames of the structural genes, *galM* and *galE* were expressed as polypeptides with apparent molecular masses of 39 and 37 kilodaltons, respectively. Significant activities of mutarotase and UDPglucose 4-epimerase were detected in lysates of *E. coli* cells containing plasmids encoding *galM* and *galE*. Expression of *galE* in *E. coli* was increased 300-fold when the gene was placed downstream of the *lac* promoter. The gene order for the *gal-lac* gene cluster of *S. thermophilus* is *galE-galM-lacS-lacZ*. The flanking regions of these genes were searched for consensus promoter sequences and further characterized by primer extension analysis. Analysis of mRNA levels for the *gal* and *lac* genes in *S. thermophilus* showed a strong reduction upon growth in medium containing glucose instead of lactose. The activities of the *lac* (lactose transport and β-galactosidase) and *gal* (UDPglucose 4-epimerase) proteins of lactose- and glucose-grown *S. thermophilus* cells matched the mRNA levels.

*Streptococcus thermophilus* transports lactose by means of a proton motive force-linked mechanism (33). Lactose enters the cell as a free sugar, and the disaccharide is hydrolyzed into glucose and galactose by β-galactosidase (20, 33). Glucose enters the glycolytic pathway, whereas in the presence of excess lactose, the galactose moiety of lactose is excreted into the medium (39).

The *lac* genes of *S. thermophilus* have recently been cloned, sequenced, and partially characterized (20, 33; C. J. Schroeder, C. Robert, G. Lenzen, L. L. McKay, and A. Mercenier, submitted for publication). The lactose transport gene (*lacS*) encodes a 69,454-dalton (Da) protein consisting of an amino-terminal domain with homology to the melibiose carrier of *Escherichia coli* and a carboxy-terminal domain with homology to enzyme III or enzyme III domains of various phosphoenolpyruvate-dependent phosphotransferase systems from gram-positive and gram-negative organisms. A similar transport protein has been found in *Lactobacillus bulgaricus* (33, 38), and the function(s) of the different domains of the transport proteins is currently under investigation (unpublished data). Contrary to the *E. coli lac* operon, the β-galactosidase gene (*lacZ*) is found downstream of the lactose transport gene in *S. thermophilus* and *L. bulgaricus*. In these last two organisms, the genes for lactose transport and β-galactosidase are separated by a 3-base-pair (bp) intercistronic region (33).

The *gal* genes of *S. thermophilus* have not yet been studied at the molecular level. Enzyme activity measurements have established that the Leloir pathway enzymes UDPglucose 4-epimerase and UDPglucose–hexose-1-phosphate uridylyltransferase are present in *S. thermophilus* but that the organism cannot express high levels of galactokinase (23, 39). The excretion into the medium of the galactose moiety of lactose and the apparent Gal− phenotype of *S. thermophilus* have been attributed to a defect in the induction mechanism for galactokinase. Under appropriate selective pressure, Gal− cultures can become Gal+. However, the Gal+ phenotype of *S. thermophilus* is rapidly lost upon subculturing in milk medium (39). Galactose transport in *S. thermophilus* has been studied to some extent (22). Most likely, galactose enters the cell via the lactose transport protein (33). The lactose transport protein not only catalyzes sugar-transport symport but also homologous and heterologous exchange of β-galactosides (33). This raises the interesting possibility that during lactose metabolism, the transport reaction proceeds as lactose-galactose exchange independent of the proton motive force.

This study was undertaken in an attempt to find the 5′ end, and possible regulatory sequences or genes, of the *S. thermophilus lac* operon. However, DNA sequencing data and gene expression studies showed the presence of additional genes coding for enzymes involved in carbohydrate metabolism, i.e., a mutarotase and a UDPglucose 4-epimerase. Thus far, *S. thermophilus* genes other than those involved in lactose and galactose metabolism have not been cloned and sequenced or characterized genetically.

* Corresponding author.
† Present address: Department of Microbiology, State University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands.
MATERIALS AND METHODS

Bacterial strains. The E. coli strains used in this study are listed in Table 1. S. thermophilus A147 was obtained from the Centre International de Recherche Daniel Carasso, BSN group. The organisms were grown as described previously (33).

Nomenclature. We have chosen to denote S. thermophilus genes by using three lowercase letters followed by a capital letter. For the lac and gal genes of S. thermophilus, the E. coli nomenclature was used when genes encoded homologous proteins with similar functions, e.g., lacZ for the β-galactosidase gene and galE for the UDPglucose 4-epimerase gene. For nonhomologous genes encoding proteins with similar functions, a different capital letter was used, e.g., lacS rather than lacY for the lactose transport gene of S. thermophilus. The mutarotase gene of S. thermophilus was arbitrarily named galM, since it has no known counterpart in E. coli.

DNA manipulations. Chromosomal DNA was isolated essentially as described previously (38), except that after lysis of the cells an equal volume of isopropanol was added for spoiling of the DNA. Upon removal from the isopropanol phase, the DNA was dried and dissolved in 1 to 5 ml of TE (10 mM Tris hydrochloride [pH 7.5], 1 mM disodium EDTA) containing 0.5% sodium dodecyl sulfate and 100 μg of proteinase K per ml. After incubation for 30 to 60 min at 50°C, the suspension was extracted several times with phenol-chloroform. Plasmid DNA was isolated from E. coli strains by the alkaline lysis method (6). DNA sequencing was performed as described previously (33). Synthetic oligonucleotides were synthesized by the Genentech Organic Synthesis Group (Genentech Inc., South San Francisco, Calif.). Other methods, including enzymatic reactions, electrophoresis, and Southern hybridizations, were performed according to standard procedures (27).

Transformation. Routinely, E. coli strains were transformed by the calcium chloride-rubidium chloride method, as described previously (27). For transformation with complex ligation mixes (three- and four-way ligations), cells were transformed by electroporation (14) (electrotransformation (9)). Electrottransformation was also used for E. coli E274, SA500, SA599, and SA1287, which are only poorly transformable by other procedures. For electrottransformation, cells were grown to mid-exponential phase, harvested by centrifugation, washed twice with distilled H2O, and suspended to an A600 of 100 to 400 in 10% (vol/vol) glycerol. Samples (50 to 100 μl) were quickly frozen in a dry ice-ethanol bath and stored at -70°C. The cells were transformed with the Gene Pulser apparatus (Bio-Rad Laboratories, Richmond, Calif.) by using cuvettes with a 0.2-cm electrode gap. The voltage and capacitance settings were 2,500 V and 25 μF, respectively. The resistance of the pulse controller was set at 200 Ω.

Cloning of galE. The procedure for cloning a DNA fragment upstream of galM is schematically shown in Fig. 1 (41). Briefly, chromosomal DNA was digested to completion with PstI. The cleaved DNA was fractionated by agarose gel electrophoresis, and, on the basis of Southern hybridization experiments, fragments of about 4.5 kilobases (kb) were electroeluted. The DNA fragments were ligated at a concentration of <0.1 ng/μl. With these dilute DNA suspensions, the formation of DNA circles is favored over the formation of DNA concatamers (11). Subsequently, the DNA ligase was inactivated and the DNA was digested with NcoI. Following phenol-chloroform extraction and ethanol precipitation in the presence of 100 ng of carrier tRNA per μl, DNA (1 to 5 ng) was used for the polymerase chain reaction. The approximate positions of the forward (pne) and reverse (pne) primers are indicated in Fig. 1. The sequences of pne and pne are 5'ACCATCAGGTTTGCACCCGCCACATTA and 5'GGTATCGGAATAT'TTlAGTATTGCAGGTAC, respectively. A portion (10 to 20 μl) of the polymerase chain reaction-amplified DNA was extracted with chloroform prior to digestion with EcoRI and PstI. The EcoRI-PstI (1.5 kb) fragment together with the 4.2-kb EcoRI fragment of pEKS8 was ligated into pKK223-3 that had been linearized with EcoRI and PstI. The ligation mix was used to transform...
E. coli HB101. The transformed organisms were spread on MacConkey agar (Difco Laboratories, Detroit, Mich.) plates containing 1% (wt/vol) lactose, 50 μg of carbenicillin per ml, and 25 μg of streptomycin per ml.

Polymarker chain reaction. Amplification of DNA (29, 35) was carried out for 30 cycles in the DNA Thermal Cycler (Perkin Elmer-Cetus Instruments, Norwalk, Conn.) after an initial denaturation cycle at 94°C for 10 min. Subsequent cycles were carried out for 90 s at 94°C (denaturation of DNA), 60 s at 50°C (annealing of primers), and 4 min at 70°C (primer extension). The reaction mixture (total volume, 100 μl) consisted of 10 mM Tris hydrochloric, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 0.1% (wt/vol) gelatin, 1 to 5 ng of DNA, 500 μM deoxynucleoside triphosphates, ~1 μM forward primer, and ~1 μM reverse primer. To prevent evaporation and refluxing of the solution, 20 to 50 μl of mineral oil was layered on top of the reaction mixture. Thermus aquaticus DNA polymerase (5 U; Perkin Elmer-Cetus) was added immediately after the initial denaturation step. To analyze the amplification products, 2 μl of the reaction mixture was electrophoresed on a 0.8% agarose gel.

Isolation of RNA and Northern (RNA) hybridizations. S. thermophilus A147 was grown to an A₆₀₀ of 0.2 to 0.8 in MR5 medium (12) containing 1% (wt/vol) lactose or glucose as the sole carbohydrate source. Cultures (100 to 200 ml) were quickly harvested by centrifugation, suspended in 5 ml of ice-cold TE, and frozen in liquid nitrogen. The frozen cells were ground in a coffee grinder, and the powder was added immediately to 5 ml of 2× extraction buffer (prepared by mixing predissolved solutions of para-aminosalicylic acid sodium salt [2.4 g per 7.5 ml] and trisopropylamine-sulphonic acid [0.4 g/7.5 ml] to which 10 ml of 1 M Tris hydrochloride, pH 8.5, 1.25 M NaCl, and 0.25 M dodium EGTA were added). The suspension was vortexed and extracted with phenol-chloroform until protein was no longer detectable at the interface. The RNA was ethanol precipitated, and the pellet was dissolved in 50 μl of diethylpyrocarbonate-treated H₂O containing 1 U of RNase inhibitor (RNasin) per μl.

For Northern analysis, 2 to 5 μl of RNA was dissolved in a mixture of 50% formamide, 6% formaldehyde, 50 mM 3-(N-morpholino)propanesulfonic acid—NaOH (pH 7.0), 3 mM sodium acetate, and 0.1 mM dodium EDTA (total volume, 20 μl). The RNA was denatured for 4 min at 65°C, chilled on ice, and electrophoresed on 0.8% (wt/vol) agarose containing 2.2 M formaldehyde (2%). RNA transfer to Nytran (Schleicher & Schuell, Inc., Keene, N. H.) filters and nick translation of DNA fragments were carried out as described previously (27).

Primer extension assay. For mapping of the lacS and galM promoters, oligonucleotide pext (5′-GAGCTGTTTACCAGAACGACGAGTGGTA AATGACAAACCAATTATTAAAAAAATTATTAT) was used for mapping the region immediately upstream of lacZ. The primers (0.2 to 0.4 pmol) were end labeled by standard procedures (27), extracted with phenol-chloroform, ethanol precipitated, and dissolved in 5 μl of 100 mM Tris hydrochloride (pH 8.0) buffer containing 25 mM sodium EDTA, 0.1 mM NaCl, and 5 U of RNasin. RNA was added (up to 5 μl) and the mixture was incubated for 2 min at 90°C and allowed to cool to room temperature. Subsequently, 10 μl of 200 mM Tris hydrochloride (pH 8.3) buffer containing 20 mM MgCl₂, 100 mM KCl, 20 mM β-mercaptoethanol, 200 μM of deoxynucleoside triphosphates per ml, 200 μg of actinomycin D per ml, 2 U of RNasin per μl, and 10 U of avian myeloblastosis reverse transcriptase (Boehringer GmbH, Mannheim, Federal Republic of Germany) was added. The mixture was incubated for 60 min at 42°C, after which the reaction was stopped by phenol-chloroform extraction. Following ethanol precipitation, the pellet was dissolved in 5 μl of sequencing loading buffer. The cDNAs were separated on a 5% (wt/vol) polyacrylamide–8 M urea sequencing gel parallel to sequencing runs of the appropriate DNA fragments in M13mpl8/mpl9. The primers used for primer extension, i.e., pext and pexm, were also used for sequencing the promoter regions.

Transport assay. S. thermophilus A147 cells were grown to an A₆₀₀ of about 0.7 in lactose- and glucose-MRS (see above). The cells were harvested by centrifugation, washed twice with 100 mM K-PIPES (K-piperazine-N,N'-bis(2-ethanesulfonic acid), pH 6.0, buffer containing 5 mM MgSO₄, and suspended to an A₆₀₀ of about 20 in the same buffer with or without 1 mM thiomethyl-β-d-galactopyranoside (TMG). The cells were incubated at room temperature for about 2 h and concentrated by centrifugation. For [¹⁴C]TMG/TMG exchange experiments, 5 μl of concentrated cells (A₆₀₀ of about 200) were diluted 40-fold into buffer containing [¹⁴C]TMG (final concentration, 45 μM). Further manipulations were the same as those described previously (34).

Enzyme assays. Enzyme assays were performed on lysates of S. thermophilus A147 and different E. coli strains transformed with various plasmid constructions. Cell-free lysates were prepared as described previously (32). Mutarotase activity was determined polarographically at 30°C in a coupled enzyme system by using α-d-glucose (Sigma Chemical Co., St. Louis, Mo.) as the substrate and β-d-glucose-oxidoreductase (Boehringer GmbH) as the coupling enzyme (28). The assay mixture consisted of 100 mM Tris hydrochloride, 10 mM EDTA, 0.002% NaN₃, 200 μM α-d-glucose, 0.5 mg of β-d-glucose oxidase per ml, and 0.2 to 0.4 mg (total protein) of cell extract per ml. UDPglucose-4-epimerase, UDPglucose–hexose-1-phosphate uridylyltransferase, and β-galactosidase activities were determined by standard procedures (see references 45, 39, and 38, respectively; see also footnote to Table 2).

Miscellaneous. Protein was measured by the method of Lowry et al. (26), with bovine serum albumin as the standard. Minicells were isolated and purified by sucrose gradient centrifugation, and proteins were labeled in vivo with [⁵⁷S]methionine, as described previously (33).

RESULTS

Nucleotide sequence of the mutarotase gene. In a previous paper, the cloning, sequencing, and characterization of the lactose transport gene (lacS) of S. thermophilus were described (33). The chimeric plasmid pEK58 encoding the lactose transport protein consists of the expression vector pKK223-3 and a 4.2-kb EcoRI insert of S. thermophilus DNA. Expression of pEK58-encoded proteins in a minicell-producing E. coli strain indicated the presence of an open reading frame (ORF), corresponding with a 38- to 39-kDa protein, in addition of the lacS ORF. Construction of a deletion derivative (pEK58-7) of pEK58 showed that the 38- to 39-kDa protein was not essential for lactose transport (33). The nucleotide sequence and the deduced amino acid sequence of the ORF are described below.

Translation of the ORF predicts a polypeptide of 348 amino acids corresponding to a calculated molecular mass of 39,095 Da (Fig. 2). The translation initiation site at position 1204 (Fig. 2) is proposed on the basis of the start of the ORF.
(at bp 1195), the position of the putative ribosome-binding site (GAGGA at bp 1193), and the amino acid sequence homology between the ORF (galM protein) and mutarotase of A. calcoaceticus (mutA protein) (Fig. 3B). Furthermore, the calculated molecular mass of the galM protein is in good agreement with the molecular mass determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Fig. 4). The intercistronic region between galM and lacS is 105 bp and contains the promoter region of the lactose transport gene (see below). Upstream of galM, a 72-bp intercistronic region is found which contains sequences with some similarity to −10 and −35 promoter regions of gram-positive and gram-negative organisms (19). The putative −10 promoter region overlaps with a perfect inverted repeat which upon formation of a stem-loop structure in the mRNA corresponds with a free-energy change of −18.9 kcal/mol (see Fig. 7). Since galM is expressed in E. coli in constructions in which the −10 and −35 promoter regions at the inverted repeat position is the only E. coli consensus promoter present (pEKS8, pPK1/anti-tac, and pEPK1/anti-tac), galM is most likely transcribed from this promoter (at least in E. coli). Characterization of promoter regions functional in S. thermophilus is described below. Finally, an almost perfect direct repeat is present about 120 bp upstream of the inverted repeat (Fig. 2).

**Amino acid sequence homology.** The amino acid sequence of the galM protein has been compared with those of proteins in the Dayhoff protein bank. Significant similarity (29.7% amino acid identity) was found with mutarotase (aldose 1-epimerase) of A. calcoaceticus (18) and the carbonyl-terminal domain of UDPglucose 4-epimerase of K. lactis (43) and S. cerevisiae (10) (19.3 to 21.6% amino acid identity) (Fig. 3B and Fig. 5).

**Mutarotase activity.** Mutarotase activity was measured in lysates prepared from E. coli E274 (galM galE), E274(pPK1) (galM+ galE+), E274(pHPl) (galM galE+), E274(pEKS8) (galM+ galE+), and E274(pEKS8) (galM+ galE+) (Fig. 1). Enzyme-catalyzed mutarotase rates were corrected for spontaneous mutarotation (28). The net rates of mutarotation were 7, 44, and 38 nmol of β-glucose formed per min per mg of protein for E274, E274(pPK1), E274(pHPl), and E274(pEKS8), respectively, indicating that galM is functionally expressed in E. coli independent of the presence of galE. The galM protein did not have detectable UDPglucose 4-epimerase activity.

**FIG. 2.** DNA sequence of the mutarotase (galM) and UDPglucose 4-epimerase (galE) genes and flanking regions determined by the dye-excision chain termination method (37). Deduced amino acids sequences of proteins are given below the nucleotide sequence. Relevant restriction sites, ribosome-binding sites (RBS), promoter regions, and putative start and stop sites of translation are given above the DNA sequence. A direct repeat at the end of galE and an inverted repeat in the intercistronic region between galE and galM are indicated by arrowheads.
Since the size of the protein is about half the size of the UDPglucose 4-epimerases (gal10) of K. lactis and S. cerevisiae, it was hypothesized that the gene encoding a protein homologous to the amino terminus of UDPglucose 4-epimerase of K. lactis and S. cerevisiae might be located upstream of galM. To test this hypothesis, an additional fragment of S. thermophilus chromosomal DNA was cloned.

Cloning of galE. To isolate a DNA fragment upstream of galM, a new method was employed in which the polymerase chain reaction was used to enrich a chromosomal DNA library for the appropriate DNA fragment prior to cloning (see Materials and Methods; Fig. 1) (30, 36, 41). For the construction of pEPK1, a 1.5-kb EcoRI-PstI polymerase chain reaction-derived fragment and the 4.2-kb EcoRI fragment of pEK8 were ligated into pKK223-3 (EcoRI-PstI) (Fig. 1). The ligation mixture was used to transform E. coli HB101 (lacY1), and lac+ clones were selected on lactose-MacConkey agar plates. Plasmid DNA was isolated from a few of these colonies and the construction of pEPK1 was confirmed by restriction analysis. Subsequently, pEPK1 was used to transform E. coli E274 (galE) and the transformed organisms were spread on galactose-MacConkey agar plates. None of the transformants had a gal+ phenotype, indicating either that the UDPglucose 4-epimerase gene of S. thermophilus was not present on this DNA fragment or that the gene was not (functionally) expressed in E. coli. Since in the construction of pEPK1 the tac promoter runs in the opposite direction of lacS and galM (and possibly the

UDPglucose 4-epimerase gene [galE]), plasmids in which the tac promoter was placed upstream of these genes were constructed (Fig. 4). Plasmid constructions pPK1/tac and pEPK1/tac (but not pPK1/tac) subcloned gal+ phenotypes in these constructions and that the tac promoter was required for sufficient expression of the gene in E. coli E274 (Table 2).

Nucleotide sequence of the UDPglucose 4-epimerase gene. The nucleotide sequence and the deduced amino acid sequence of galE are shown in Fig. 2. Errors in the sequence could occur, since galE was constructed from a 1.5-kb EcoRI-PstI fragment obtained after amplification by the polymerase chain reaction (Fig. 1) and the fidelity of DNA synthesis in vitro by Taq DNA polymerase is suspect (15, 40). Therefore, the 500-bp EcoRV-EcoRI fragment (Fig. 2) from two independent galE clones, comprising the amino-terminal part of galE, was subcloned into M13mp18/19. The DNA sequences from both these clones were identical, indicating that no errors were introduced by the Taq DNA polymerase (data not shown).

The ORF for galE most likely consists of 996 bp which correspond to a protein monomer of 332 amino acids with a calculated molecular mass of 37,023 Da. The translation initiation site at position 136 (Fig. 2) is proposed on the basis of the start of the ORF (at bp 133), the position of the putative ribosome-binding site (AGGAG at bp 124), and the amino acid sequence homology between the galE protein, the corresponding enzymes of E. coli and Streptomyces
**FIG. 4.** Expression of galM and galE in minicell-producing E. coli P678-54. Upper half: 35S-labeled proteins were resolved by sodium dodecyl sulfate–10% polyacrylamide gel electrophoresis, after which the proteins were identified by autoradiography. Lane 1, pPK1/tac; lane 2, pPK1/anti-tac; lane 3, pHPl/tac; lane 4, pEPK1/anti-tac. Bottom half: Plasmid constructions containing galM or galE or both. Symbols are the same as those defined in the legend to Fig. 1.

lividans, and the amino-terminal domain of UDPglucose 4-epimerase of K. lactis (Fig. 3A and Fig. 5). Upstream of galE, another ORF (orfX) was found, consisting of at least 1,008 bp (data not shown). The intercistronic region between orfX and galE was only 32 bp long. A search for −10 and −35 promoter regions by using consensus sequences reported for gram-positive and gram-negative bacteria (19) did not reveal strong promoter sequences upstream of galE, consistent with the observation that without an exogenous promoter galE is weakly expressed in E. coli. The sequence data together with the Northern analysis of S. thermophilus mRNA (see below) suggest that galE and orfX might be transcribed as a single message with the promoter being located upstream of orfX. To check whether orfX possibly coded for UDPglucose–hexose-1-phosphate uridylyltransferase (galT), the other gal enzyme known to be present in S. thermophilus (39), pPK1/tac, was used to transform E. coli SA1287 (galEΔX::galTamb79) and SA599 (galTamb79). Complementation was not observed and UDPglucose–hexose-1-phosphate uridylyltransferase activity could not be detected in lysates prepared from these organisms, indicating that orfX does not contain a complete ORF and/or proper ribosome-binding site or that orfX does not encode the uridylyltransferase.

**Amino acid sequence homology.** UDPglucose 4-epimerase of S. thermophilus was homologous to the corresponding enzymes of E. coli (25) (38% amino acid identity), S. lividans (1) (42% amino acid identity), and the amino-terminal domain of K. lactis UDPglucose 4-epimerase (43) (40% amino acid identity) (Fig. 3A and 5). UDPglucose 4-epimerase from S. cerevisiae, which has been only partially sequenced (10), is also similar to S. thermophilus galE at the amino terminus. No similarity was found between orfX and galT from E. coli or between orfX and any other protein in the Dayhoff protein bank.

**Expression of galE and UDPglucose 4-epimerase activity.** Different plasmids containing galM or galE or both were introduced into the minicell-producing E. coli P678-54. galM was expressed as a 39-kDa protein with the constructions

<table>
<thead>
<tr>
<th>Strain</th>
<th>Orientation of P tac-galE</th>
<th>IPTG</th>
<th>Activity (µmol/min per mg of protein)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E274</td>
<td>tac</td>
<td>−</td>
<td>0.0</td>
</tr>
<tr>
<td>E274(pPK1)</td>
<td>Anti-tac</td>
<td>+</td>
<td>14.9</td>
</tr>
<tr>
<td>E274(pPK1)</td>
<td>Anti-tac</td>
<td>−</td>
<td>0.05</td>
</tr>
<tr>
<td>E274(pPK1)</td>
<td>Anti-tac</td>
<td>−</td>
<td>17.9</td>
</tr>
<tr>
<td>E274(pPK1)</td>
<td>Anti-tac</td>
<td>−</td>
<td>0.06</td>
</tr>
<tr>
<td>E274(pPH1)</td>
<td>tac</td>
<td>−</td>
<td>15.1</td>
</tr>
</tbody>
</table>

a IPTG, Isopropyl-β-D-thiogalactopyranoside.

b Activity was measured as UDPgalactose 4-epimerase activity. The assay mixture consisted of 100 mM glycine-NaOH (pH 8.6), 10 mM NAD+, 1 mM UDPgalactose, 0.3 U of UDPglucose dehydrogenase per ml, and, depending on the final UDPgalactose 4-epimerase activity, 2 to 100 µg of cell-free lysate per ml. Measurements were done at room temperature.
pPK1 and pEPK1, and expression was more or less independent of the orientation of galM towards the tac promoter (Fig. 4, lanes 1, 2, and 4). galM was absent in the deletion derivative pH1P1/tac (Fig. 4, lane 3). A 37-kDa protein, corresponding with the size of galE (see above), was observed in all plasmid constructions. Expression of galE was greatly increased when the gene was placed downstream of the tac promoter (Fig. 4, lanes 1 and 3). Finally, a faint band corresponding to expression of lacS protein, a protein with an apparent molecular mass of 55 kDa (33), was observed in pEPK1 only (Fig. 4, lane 4). For UDPglucose 4-epimerase activity measurements, lysates were prepared from E. coli E274 alone or transformed with different plasmid constructions (Table 2). UDPglucose 4-epimerase activity could not be detected in E. coli E274. Low but significant activity was present in lysates prepared from E. coli E274 (pPK1/anti-tac). This activity, i.e., 50 to 60 nmol of UDPglucose formed per min per mg of protein, was insufficient to complement E. coli E274 (see above). UDPglucose 4-epimerase activity was increased about 300-fold when galE was placed downstream of the tac promoter, and this activity was irrespective of expression of galM (compare E274/pPK1-tac with E274/pHP1-tac). Activities were not affected by the presence of isopropyl-β-D-thiogalactopyranoside during growth of the organisms, indicating that the tac promoter is hardly regulated or not regulated in E. coli E274. Although it is difficult to quantitate the protein expression levels precisely, there seems to be a discrepancy between the difference in protein levels (Fig. 4) and enzyme activities (Table 2) in tac and anti-tac constructions of galE. It is not clear whether the discrepancy is due to differences in the strains used.

mRNA levels and 5'-end mapping of mRNA. Total RNA isolated from lactose- and glucose-MRS-grown S. thermophilus cells was analyzed by Northern hybridizations. To identify mRNAs, appropriate gene fragments, i.e., 1,450-bp PvuII, 646-bp BstXI-EcoRV, and 540-bp PvuII fragments for lacS, galM, and galE, respectively, were nick-translated and used as probes. For lactose-MRS-grown cells, major mRNAs of about 5.5 and 3.5 kb were identified for lacS and galE, respectively (data not shown). Very little mRNA was detected in glucose-MRS-grown cells. galM mRNA could not be detected (reliably) by Northern analysis, irrespective of whether cells were grown in lactose- or glucose-MRS.

The size of the lacS mRNA suggests that lacS is transcribed together with the β-galactosidase gene (lacZ) as a single message. To locate the transcriptional start point of the lac genes, the primer extension method was used. Two major transcripts were observed (Fig. 6A, vertical bars); one (Fig. 6A, m1) starts 10 bp downstream of the inferred −10 sequence (indicated as +1), whereas the other (Fig. 6A, m2) is located 34 bp downstream of the −10 region. For the latter transcript, no sequences similar to promoters from gram-positive and gram-negative bacteria could be found. In accordance with the Northern analysis, the amount of DNAs synthesized with the mRNA template from glucose-MRS-grown cells was reduced at least 10-fold compared with lactose-MRS-grown cells (Fig. 6A). The two major transcripts as well as the effect of glucose on the transcript levels were observed with four different preparations of RNA, i.e., RNA isolated from cells harvested at an A600 of 0.2 to 0.8 (data not shown). Primer extension analysis in the 300-bp region immediately upstream of lacZ did not reveal a major transcriptional start point, supporting the contention that lacZ and lacS are transcribed from a single promoter(s) located upstream of lacS.

Table 3. Effects of lactose and glucose on expression of lac and gal genes in S. thermophilus

<table>
<thead>
<tr>
<th>Protein</th>
<th>Gene</th>
<th>Activity (nmol/min per mg of protein)</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-Galactosidase</td>
<td>lacZ</td>
<td>10,800</td>
</tr>
<tr>
<td>Lactose transport</td>
<td>lacS</td>
<td>18a</td>
</tr>
<tr>
<td>UDPglucose 4-epimerase</td>
<td>galE</td>
<td>2,070</td>
</tr>
<tr>
<td>UDPglucose–hexose-1-phosphate uridylyltransferase</td>
<td>galT</td>
<td>492</td>
</tr>
</tbody>
</table>

* Lactose transport was measured as [14C]TMG/TMG exchange activity (see Materials and Methods).

For galM, a minor transcript, starting 19 bp downstream of a putative −10 sequence, was detected (Fig. 6B, vertical bar). The galM transcript was observed in lactose- but not glucose-MRS-grown cells (Fig. 6B). The size of the galE transcript combined with the observation that a DNA probe derived from galM did not hybridize with the transcript suggests that the true promoter is located upstream of orfX.

Enzyme and lactose transport activities. Since the mRNA levels for the lac and gal genes of S. thermophilus were highly reduced in glucose-MRS-grown cells, 10-fold or more compared with lactose-MRS-grown cells, the corresponding enzyme and transport activities were measured. Table 3 shows that the activities for β-galactosidase, lactose transport, UDPglucose 4-epimerase, and UDPglucose–hexose-1-phosphate uridylyltransferase in lactose-MRS-grown cells are 10- to 20-fold higher than in glucose-MRS-grown cells. These data together with the effect of glucose on the mRNA levels indicate that expression of the lac and gal genes is regulated at the level of transcription. Mutatorase activities in cell lysates prepared from lactose- and glucose-MRS-grown cells are too low to be measured accurately, i.e., on top of the rate of spontaneous mutatorate.

DISCUSSION

The cloning of galE was achieved by using a special polymerase chain reaction for which a region of interest can be amplified when only 5' or 3' flanking sequences are known (30, 41). The advantage of the method over classical cloning procedures is that transformants do not have to be screened by laborious hybridization protocols in order to identify the desired recombinant clone. A potential disadvantage of the method relates to the relatively low fidelity of Taq DNA polymerase (lack of detectable exonucleolytic proofreading activity) (40), which can result, among other errors, from base substitution errors in the amplified DNA fragment. However, such errors have not been observed in the fragment (500-bp EcoRV-EcoRI) needed for the construction of galE (see Results).

In aqueous solution, many saccharides exist in two different isomeric forms, i.e., α and β, differing in specific rotation. The anomeric interconversion of α-glucose and other aldoses which can occur spontaneously is catalyzed by mutarotase. Spontaneous mutarotation is acid-base catalyzed and consequently strongly affected by pH and chemical composition of the aqueous medium. On the basis of the kinetics of spontaneous mutarotation, pH dependence, and catalysis by amino acids (in particular histidine [and other
FIG. 6. Determination of the 5' ends of lacS (A) and galM (B) transcripts. Primer extension on mRNA isolated from glucose-MRS- and lactose-MRS-grown *S. thermophilus* A147 is indicated by g and l, respectively. Equal amounts of total RNA were loaded in all lanes. To calibrate the transcriptional start points, M13 clones of the corresponding fragments were sequenced in parallel. Relevant nucleotide sequences of the promoter regions are given on either side of the figure. Putative promoter regions and transcriptional start points are indicated by vertical bars. For the *galM* promoter region, the inverted repeat is indicated by arrows.

Solutes containing an imidazole group)), it has been proposed that one component of the active site of enzyme-catalyzed mutarotation involves a histidine residue (4, 5). In the alignment of *galM* with *mutA* and *gal10p* proteins (Fig. 3B), two histidine residues are conserved, i.e., His-105 and His-180 (indicated by arrow), of which the latter is present in a region highly conserved between the four proteins. This suggests that His-105 or (more likely) His-180 furnishes the imidazole group needed for catalysis.

Mutarotase of *A. calcoaceticus* has been shown to contain a leader peptide sequence which directs the secretion of the enzyme into the periplasmic space (17, 18). Interestingly,
the first amino acid of the galM protein aligns with the first amino acid of the mature mutarotase protein (Fig. 3B).

An interesting outcome of this study is the sequence homology between galM (and mutA) and the carboxy-terminal domain of UDPglucose 4-epimerase (gal10) of yeasts and that between galE and the amino-terminal domain of the same enzymes (Fig. 5). On the basis of these similarities, we speculate that gal10 enzymes possess mutarotase activity in addition to UDPglucose 4-epimerase activity. The presence of an apparent hybrid epimerase in K. lactis and S. cerevisiae could suggest that galM and galE are expressed functionally only as a hetero-multimeric enzyme complex. On the other hand, the data indicate that galM and galE can be expressed as functionally active proteins independent of each other.

In many organisms, the genes for galactokinase (galK), UDPglucose 4-epimerase (galE), and UDPglucose-hexose-1-phosphate uridylyltransferase (galT) are present in a single operon, constituting the Leloir pathway for galactose metabolism (1, 2, 44). Alternatively, it has been shown for Erwinia stewartii that galE is not linked to galK and galT; rather, galE is linked to genes encoding enzymes involved in the biosynthesis of extracellular polysaccharides (13). galE is expressed constitutively, whereas galK and galT are inducible in E. stewartii. The fact that orfX is not homologous to galT or galK from E. coli could argue against the presence of the complete gal operon upstream of the lac operon. On the other hand, in accordance with an operonlike structure, galE and galT appear to be regulated similarly in S. thermophilus (Table 3; Fig. 6). Future experiments will be directed towards characterizing the function of orfX and cloting genes involved in the regulation of carbohydrate metabolism.

UDPglucose 4-epimerase is required during galactose metabolism to convert UDPgalactose to UDPglucose, which enters the glycolytic pathway. Since most strains of S. thermophilus are gal mutants (23, 39), the role of UDPglucose 4-epimerase in this organism could be to convert UDPglucose into UDPgalactose for the biosynthesis of extracellular polysaccharides (8, 13) or galactosyl lipids (31).

The gene order and possible regulatory elements of the gal and lac genes are depicted in Fig. 7. Although the inverted repeat, indicated as a stem-loop structure, downstream of lacZ does not fully resemble a typical rho-independent terminator, the region could well be involved in transcription termination, considering the size of the lac transcript and the fact that an 860-bp region downstream of lacZ is noncoding (Schroeder et al., submitted). Putative lac and galM promoter regions are indicated by $p_r$ and $p_{ai}$, respectively. On the basis of the primer extension assay, $p_r$ $-35/-10$ promoter region seems to be functional in S. thermophilus, although a second transcript (Fig. 6A, m2), for which no consensus promoter sequence can be identified, appears to be present too. A possible start of the galM transcript is found 19 bp downstream of a putative $-35/-10$ promoter (Fig. 6B, m). Whether this promoter region is indeed used or whether an as yet unrecognized promoter region is used by RNA polymerase from S. thermophilus remains unclear. Also, it cannot be excluded that galM is transcribed from a promoter located upstream of galE and orfX. In that case, the role of the inverted repeat could be to decrease the expression of the distally located galM (21). Finally, the level of the putative galM transcript in S. thermophilus is too low to direct the synthesis of sufficient mutarotase activity to be detected under the conditions employed.

The codon usage in the mutarotase and UDPglucose 4-epimerase genes was compared with the codon usage in the lactose transport gene (33) and the β-galactosidase gene (Schroeder et al., submitted) of S. thermophilus (Table 4). In general, the codon usage is very similar in the four genes. A strong preference for A or U in the third position can be observed, resembling the average codon usage in E. coli genes (16, 42). Major exceptions are the codons for Leu and Pro.

**FIG. 7.** gal-lac gene cluster of S. thermophilus. The gene order and nucleotide sequences of intercistronic regions of the gal and lac genes of S. thermophilus A147 are shown. The nucleotide sequences for the region between lacS and lacZ were taken from reference 33, whereas the sequence for the region 3′ of lacZ was taken from Schroeder et al. (submitted). Sequences are given as DNA rather than RNA.
TABLE 4. Condon usage in the β-galactosidase (lacZ), lactose transport (lacS), mutarotase (galM), and UDPglucose 4-epimerase (galE) genes of *S. thermophilus*

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino acid</th>
<th>lacZ</th>
<th>lacS</th>
<th>galM</th>
<th>galE</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUC</td>
<td>Phe</td>
<td>28</td>
<td>20</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>UUG</td>
<td>Leu</td>
<td>22</td>
<td>22</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>UAU</td>
<td>Ile</td>
<td>15</td>
<td>17</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>UGU</td>
<td>Ser</td>
<td>58</td>
<td>27</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>UCA</td>
<td>Ser</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>UGU</td>
<td>Ser</td>
<td>12</td>
<td>16</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>UCC</td>
<td>Pro</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
<td>23</td>
<td>10</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>UCA</td>
<td>Ser</td>
<td>20</td>
<td>15</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>CAC</td>
<td>Thr</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>ACG</td>
<td>Thr</td>
<td>24</td>
<td>20</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>GCU</td>
<td>Ala</td>
<td>36</td>
<td>16</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>18</td>
<td>14</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>UAU</td>
<td>Tyr</td>
<td>39</td>
<td>8</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>UAA</td>
<td>Tyr</td>
<td>23</td>
<td>10</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>UAG</td>
<td>Pro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UGU</td>
<td>Pro</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>UCA</td>
<td>Ser</td>
<td>37</td>
<td>12</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>GGU</td>
<td>Trp</td>
<td>33</td>
<td>16</td>
<td>22</td>
<td>4</td>
</tr>
<tr>
<td>CGU</td>
<td>Arg</td>
<td>24</td>
<td>14</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>CGC</td>
<td>Arg</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>GGU</td>
<td>Gly</td>
<td>50</td>
<td>42</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>GGC</td>
<td>Gly</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>GCA</td>
<td>Arg</td>
<td>13</td>
<td>14</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>GGA</td>
<td>Gly</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

ACKNOWLEDGMENTS

We thank S. Adhya for providing us with various *E. coli* strains and A. Mercenier for sharing unpublished data. The research of B. Poolman has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

LITERATURE CITED


