January 1995

Cloning and Characterization of the Galactokinase Gene from *Streptococcus thermophilus*

Azlin Mustapha
University of Nebraska - Lincoln

Robert W. Hutkins
University of Nebraska - Lincoln, rhutkins1@unl.edu

Gerald W. Hirnstein
University of Nebraska - Lincoln

Cloning and Characterization of the Galactokinase Gene from *Streptococcus thermophilus*[1]

AZLIN MUSTAPHA, ROBERT W. HUTKINS, and GERALD W. ZIRNSTEIN
Department of Food Science and Technology
University of Nebraska-Lincoln
Lincoln 68583-0919

ABSTRACT

The objective of this research was to clone and characterize the galactokinase gene (*galK*) from *Streptococcus thermophilus* F410. Partially digested genomic DNA was cloned into pBR322 and transformed into *galK Escherichia coli*, and a galactose-fermenting transformant was isolated. Restriction analysis revealed that the transformant resulted from a *Sau*3A-*Hin*dIII 4.0-kb fragment. Galactokinase activity in the recombinant was 10 times that of the parent strain. Analysis of the DNA sequence showed the presence of a 1.3-kb open reading frame that had high homology with the *galK* gene from other organisms. A putative ribosome-binding site, start and stop codons, and −10 and −35 sequences were identified. The predicted protein had a molecular mass of 49 kDa, which corresponded to the estimated size of a band apparent by SDS-PAGE. Amino acid sequence homologies with other galactokinases ranged from 50 to 62% similarity. Northern blots were performed between the *galK* gene and mRNA from *S. thermophilus*. No hybridization signals were observed for cells grown in glucose, but cells grown in lactose or galactose gave moderate and strong signals. The results suggest that repression of the *galK* gene by glucose may be responsible for the galactose-releasing phenotype in these strains.

(**Key words**: *Streptococcus thermophilus*, galactose, galactokinase)

Abbreviation key: **Gal$^+$** = galactose-fermenting, **Gal$^-$** = galactose-nonfermenting, **GCG** = University of Wisconsin Genetics Computer Group, **IPTG** = isopropyl thio-β-galactoside, **LB** = Luria-Bertani.

INTRODUCTION

Despite the importance of *Streptococcus thermophilus* as a starter culture in the dairy industry, the catabolic pathways for carbohydrate metabolism have only recently been established in this organism. It is now known that carbohydrate fermentation in this organism differs markedly from that of the lactococci and other mesophilic lactic acid bacteria. In fact, metabolism of sugars by *S. thermophilus* is more similar to that in *Escherichia coli* than to that in other lactic acid bacteria. Although lactococci depend on a phosphoenolpyruvate-dependent phosphotransferase system for lactose transport and the enzyme phospho-β-galactosidase and hydrolysis (16), this system is absent in *S. thermophilus*. Moreover, recent studies (21, 22, 24) indicate that many of the genes coding for proteins involved in sugar catabolism in *S. thermophilus* have higher homology with *E. coli* genes than with those of other lactic acid bacteria.

Similar to *E. coli*, *S. thermophilus* utilizes a lactose permease, or LacS, and the enzyme, β-galactosidase, to transport and hydrolyze lactose inside the cell (8, 21, 28, 29). Most wild-type strains of this organism, however, are unable to ferment lactose completely and instead release the galactose portion of lactose back into the medium (19, 25, 27, 28). The galactose-fermenting (**Gal$^+$**) strains of *S. thermophilus* metabolize galactose via the enzymes of the Leloir pathway, but Gal$^+$ strains release this sugar into the medium even when grown on lactose (10, 25, 28) because the rate-limiting Leloir pathway enzyme, galactokinase, is thought to be catabolite repressed (8, 10).

Received October 19, 1994.
Accepted January 6, 1995.
[1]Published as Paper Number 10912 of the Journal Series, Nebraska Agricultural Experiment Station, Lincoln 68583-0919.

However, evidence for regulation at the gene level has not yet been provided.

In this investigation, we report the cloning of the gene coding for galactokinase, *galK*, from *S. thermophilus* and show that transcription of this gene is regulated by glucose and lactose.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

Streptococcus thermophilus F410, isolated from a commercial starter culture, and *S. thermophilus* ATCC 19258 were maintained at 42°C in M17 broth (26) containing 1% galactose. Both strains are Gal^+, but release galactose into the medium during growth on lactose. The *E. coli* strains were maintained in Luria-Bertani (**LB**) broth (14) and grown at 37°C with shaking (80 to 100 rpm). Plasmid vectors used for cloning and sequencing experiments were obtained from commercial sources.

Cloning of the *galK* Gene

Streptococcus thermophilus F410 was grown in M17 broth containing 5% galactose to an optical density at 600 nm wavelength of .4 to .6. Chromosomal DNA was isolated by the guanidium-thiocyanate method of Pitcher et al. (20). The purified DNA was partially digested with different combinations of *Sau*3AI, *Nde*II, *Eco*RI, *Hin*dIII, and XhoI, and the fragments were ligated into pBR322. Seven different *galK E. coli* hosts were made competent by the $CaCl_2$ method of Maniatis et al. (14) and transformed with the ligation mixtures. The Gal^+ *E. coli* transformants were plated on eosin-methylene blue agar containing 1% galactose and 50 μg/ml of ampicillin. Plates were incubated at 37°C for 48 to 96 h; Gal^+ transformants appeared as bright green colonies. In some experiments, transformants were plated on LB agar containing 5-bromo-4-chloro-3-indoyl-β-D-galactoside, isopropyl thio-β-galactoside (**IPTG**), and ampicillin.

Galactokinase Assays

Cells of *E. coli* were grown in M9 minimal media (14), containing either galactose or glucose as the sole carbon source. Log-phase cells were harvested, washed, and resuspended in 100 m*M* Tris buffer, pH 7.5. A toluene:ethanol (1:9, vol/vol) mixture was added to the cells at a rate of 30 μl/ml of cells, and the mixture was vortexed for 1 min. Permeabilized cells were assayed for galactokinase activity using the NADH-ADP coupled enzyme assay of Thomas and Crow (28) with modifications. Assay mixtures contained permeabilized cells, 2 m*M* ATP, .4 m*M* NADH, 2 m*M* phosphoenolpyruvate, 2 U/ml of pyruvate kinase, 5 U/ml of lactate dehydrogenase, 10 m*M* $MgCl_2$, 1 m*M* dithiothreitol, .5 m*M* NaF, and 100 m*M* triethanolamine, pH 7.5. The reaction was started by the addition of 10 m*M* galactose, and oxidation of NADH was measured by the increase of absorbance at 340 nm. Absorbencies were corrected for endogenous NADH oxidation. Protein concentrations were determined by the Bradford method (2). From triplicate determinations, mean galactokinase activities were expressed as micromoles of galactose phosphorylated per minute per milligram of cell protein.

DNA Sequencing and Analysis

Double-stranded DNA sequencing was done by the dideoxy chain-terminating method of Sanger et al. (23), using 17- to 19-mer synthetic oligonucleotide primers. Sequencing reactions were performed according to the protocol described by the manufacturer (Sequenase 2.0 kit; United States Biochemicals, Cleveland, OH). Both strands were sequenced at least twice from opposite directions. Samples were electrophoresed on a .4-mm thick, 5% polyacrylamide gel using a Sequi-Gen Nucleic Acid Sequencing Cell (Bio-Rad Laboratories, Richmond, CA). Sequences were analyzed using the University of Wisconsin (Madison) Genetics Computer Group (**GCG**) FASTA Program.

SDS-PAGE of the Galactokinase Protein

Cells were grown in LB broth, collected during the log phase by centrifugation, and resuspended in SDS gel-loading buffer (50 m*M* Tris·HCl, 100 m*M* dithiothreitol, 2% SDS, .1% bromphenol blue, and 10% glycerol). Total cell protein was determined by the Lowry method (13). The proteins from cell extracts were sepa-

rated on a denaturing SDS-PAGE 12% acrylamide gel using the method of Laemmli (12). Each lane contained .1 mg of protein.

RNA Isolation

Streptococcus thermophilus strains F410 and 19258 were grown to midlog phase in Elliker broths containing either 1.5% glucose, lactose, or galactose (5). Cells (45 ml) were harvested by centrifugation at 12,000 × *g* for 10 min. at 4°C, and pellets were resuspended in 1 ml of Ultraspec RNA isolation and purification reagent (Biotecx Laboratories, Inc., Houston, TX). Cell suspensions were immediately homogenized for 3 min with .8 volume of .5- to 1-mm glass beads in a Bead-Beater (Biospec Products, Bartlesville, OK). The suspensions were separated from the glass beads by centrifugation (10 min at 12,000 × *g* at 4°C), and the aqueous phase was extracted with .2 volume of chloroform. The RNA was then precipitated with an equal volume of 2-propanol at 4°C for 10 min and centrifuged as described. The RNA pellets were washed twice with 75% ethanol, air dried at 4°C, and solubilized in 5 m*M* EDTA previously treated with diethyl pyrocarbonate.

Northern Analysis

Northern analysis was performed as previously described (7) with minor modifications. Each lane contained 10 μg of total cellular denatured (incubated for 45 min at 55°C in glyoxal loading buffer) RNA. The RNA was fixed by UV crosslinking to a nylon membrane support, and the blots were prehybridized at 65°C for 3 h in hybridization buffer. A .8-kb, *Cla*I-*Eco*RV DNA fragment from pAM1 was used to probe *S. thermophilus* RNA for detection of *galK* transcripts. The probe DNA was prepared and purified by the method of Zhen and Swank (30) and labeled according to the protocol supplied with the DECAprime DNA labeling kit (Ambion, Inc., Austin, TX) using 3000 Ci/mmol of [^{32}P]dCTP. Labeled DNA was separated from unincorporated radioactivity using STE midi Select®-D, G-50 Sephadex microcentrifuge spin columns (5 Prime – >3 Prime, Inc., Boulder, CO). For hybridization, the [^{32}P]-labeled DNA was first denatured by heating to 100°C for 10 min and then immediately cooled in liquid nitrogen before dilution into hybridization buffer at a concentration of 1 to 2 μCi/ml.

Chemicals

Restriction enzymes were from Bethesda Research Laboratories (Grand Island, NY), Promega (Madison, WI), and United States Biochemicals. The T4 DNA ligase, calf intestinal alkaline phosphatase, 5-bromo-4-chloro-3-indoyl-β-D-galactoside, and IPTG were from Bethesda Research Laboratories. Ampicillin, tetracycline, lysozyme, and other chemicals were from Sigma Chemical Company (St. Louis, MO). Acrylamide, bis-acrylamide, N,N,N′,N′-tetramethylethylene-diamine, and ammonium persulfate were from Bio-Rad Laboratories. The [α-^{35}S]dATP and [^{32}P]dCTP were obtained from Amersham Corporation (Arlington Heights, IL). Oligonucleotide primers were synthesized by the University of Nebraska Center for Biotechnology Oligonucleotide Synthesis Facility (using an Applied Biosystems PCR-MATE DNA Synthesizer).

RESULTS

Cloning of the *galK* Gene

Genomic DNA from *S. thermophilus* F410 was partially digested and cloned into the tetracycline resistance gene of pBR322. From over 5000 colonies screened, only 1 suspect positive colony was obtained. This positive colony exhibited a distinctive dark purple color with a metallic sheen on eosin-methylene blue-galactose agar containing ampicillin and was unable to grow on agar containing ampicillin and tetracycline.

Analysis of the recombinant strain revealed that it had resulted from the cloning of a *Hin*dIII-*Sau*3AI chromosomal fragment from *S. thermophilus* F410 into *E. coli* SA2383. The size of the cloned fragment was estimated by agarose gel electrophoresis to be about 4.0 kb (data not shown). The recombinant plasmid and host were named pAM1 and RW1, respectively. When the recombinant plasmid (pAM1) from the single positive transformant was analyzed, however, the expected *Bam*HI recognition site had been lost. By using *Sph*I, with a recognition site 187 bp downstream from the *Bam*HI site, we were able to isolate a *Hin*dIII-

*Sph*I insert, to clone this fragment into pBR322, and to retransform the original parent. Strong positive signals were again exhibited. The smallest fragment, subcloned from pAM1, that still complemented the *galK E. coli* host was a 1.7-kb *Sph*I-*Eco*RI fragment (Figure 1). Because part of this fragment contained a 187-bp *Sph*I-*Bam*HI fragment that had originated from pBR322, we concluded that the *galK* gene was present within a 1.5-kb *Bam*HI-*Eco*RI chromosomal fragment from *S. thermophilus* F410.

Galactokinase Activity of *E. coli* RW1

Galactokinase activity of the recombinant *E. coli* strain RW1 (containing pAM1) was about 2.5 times higher than that of the wild-type Gal$^+$ *E. coli* strain C600 and more than 10 times higher than the *galK* parental strain, *E. coli* SA2383 (Table 1).

TABLE 1. Mean (n = 3) galactokinase activities of *Escherichia coli* RW1, wild-type *E. coli* C600, and *E. coli* SA2383 parental strain.

Strain	Activity[1,2]	
	$\bar{X}$	SD
E. coli C600, wild-type (Gal$^+$)[3]	139	7
E. coli SA2383, parent (Gal$^-$)	30	2
E. coli RW1 (recombinant strain)	347	14

[1]Nanomoles of galactose phosphorylated per milligram of protein per minute.

[2]Strains C600 and RW1 were grown in M9 medium containing galactose, and SA2383 was grown in M9 containing glucose.

[3]Gal$^+$ = Galactose-fermenting; Gal$^-$ = galactose-nonfermenting.

DNA Sequence Analysis

Sequencing results revealed that the insert fragment of pAM1 consisted of 3896 bp, which corresponded well with the estimated size of 4.0 kb determined by agarose gel electrophoresis. One open reading frame, *orf1*, spanning about 1.3 kb, was identified by the GCG FRAMES (GCG, Madison, WI) program. Comparisons of the insert fragment with DNA sequences in the Gen Bank database revealed significant homology (52 to 59% identity) with the *galK* genes of *Lactobacillus helveticus, E. coli, Haemophilus influenzae*, and *Kluyveromyces lactis*. Based on 1) these DNA homologies, 2) the location of *orf1* predicted by GCG FRAMES, 3) the finding of a putative ribosomal binding site, and 4) comparisons of amino acid sequence homologies (discussed below), a translational start site of *orf1* was predicted at nucleotide position 153 (Figure 2). A putative ribosomal binding site was located at position 135, and putative −10 and −35 sequences were located at positions 117 and 89, respectively. The molar percentage G + C content of *orf1* was 40%, which was consistent with the reported mean value of 38 to 40% for *S. thermophilus* genomic DNA (6).

Based on analyses of DNA and amino acid sequences, the size of *orf1* was predicted to be 1319 bp. This size translated into a protein monomer of 439 amino acid residues with a calculated molecular mass of 49,168 Da. The amino acid homologies between the predicted amino acid sequence of the *S. thermophilus* galactokinase and galactokinases from other organisms ranged from 50 to 62% similarities and from 27 to 45% identities (Table 2). An alignment of the predicted sequence with other galactokinases (Figure 3) revealed several regions of significant amino acid homology.

Figure 1. Linear restriction map of the 4.0-kb *Sph*I-*Hin*dIII fragment cloned from *Streptococcus thermophilus*. Upstream and downstream regions containing pBR322 DNA are shown by the hatched lines. The 1.3-kb *galK* gene is shown by the heavy line. Restriction enzyme abbreviations are P, *Sph*I; B, *Bam*HI; V, *Eco*RV; C, *Cla*I; E, *Eco*RI; S, *Sal*I; X, *Xho*I; and H, *Hin*dIII.

Analysis of *orf1* Flanking Regions

Comparisons of the flanking regions of *orf1* with other DNA sequences in the Gen Bank did not reveal significant homologies with other reported *lac* or *gal* genes. Surprisingly, no homologies were found between the region downstream of the *galK* gene in this work and *galM*, *galE*, or *galT* genes (coding for aldose-1-epimerase, uridine-diphosphogalactose-4-epimerase, and galactose-1-phosphate uri-

```
                                                        -35
1    ggatcatgcc cgtttttctt gaaagcgaac aaggattaga caaagtagta
                                                        -35
51   gcttttcttg tcagtccaga tgaaaaaatc aaagtaagtg tagcaacaga
                          -10                    RBS
101  tgcgccaagt atgtggtatt tacagcaaag ctttggagaa aatacacctg
       start galK
151  aaatgcgagg aaaaaatctg agtgatcatg gtgggatcac gtttgaaacg
201  caaattgctc ctggggcaga acagttctcg gattttggag atatcaaacg
251  ttccaccaaa gtcccctttg aaacaacaac agaatttaaa atcgaaacaa
301  gaaaggattt ttaataatgg aaatcaaacg tagcttaaat acaaatttgc
351  agaaatcttc ggacctaaac agacagctca atacttctca cccggaagaa
401  tcaacttgat cggtgaacac actgactata tggcgggcat gtcttccctg
451  cttcgatcac ttatggtacc aaggtgtcgc agcttctcga aaagacaata
501  aggtcttagt gtatcaacga attttgaaga tgacggcgta acagctttac
551  gttgaatgag ttagaatatg ataaacaagc tggttgggca aactatgtca
601  aaggaatgat cctaaagtta aaagaagcag ggtatacatt tgatcatggc
651  ttcgaattat tagtggaagg aacgatccct aatggtgctg gcttatcaag
701  tagcgcttct ttagaattat tagtagggtgt cgtattagaa gatttgtttg
751  atctagctat agatcgtctt gcattagtac aaacaggaaa aaaagtagag
801  aatgaattta tcggtgtcaa ctcagggatc atggatcaat tcgcaatcgg
851  ctttggagaa gtggacaaag cgatcttact tgatacaaat accttaaaat
901  atgagatggt tccagtgaaa ttagatggct atgcagtggt tatcatgaac
951  acaaacaaac gtcgtgaatt agcggattct aaatacaatg aacgtcgaag
1001 tgaatgcgaa gaggcgctta aacgtctgca aacacaatta acaatcgatg
1051 cattaggtga cctagacagc gaaacctttg aagctcatac agatttgatc
1101 aacgatgata ctttgattcg ccgagcacgc catgcagtaa cagaaaacca
1151 gcgcacacta gaagcaaaag cagaattaga aaaaggaaac ttagctgcgt
1201 ttggcaagtt attgaatgct tctcattact cgttgcgtca tgattacgaa
1251 gtaactggga ttgaattaga cactttggta gatgctgccc aaaaacaaga
1301 aggtgtctta ggcgctcgaa tgactggtgc tggattcggc ggatgtgcca
1351 tcgcattggt taaagaagaa aagcattcca gaatttgaaa ataatgtcta
1401 cgatgcttac ttgaaagtta ttggttatgc accagagttt tatgtcgctc
1451 atatcggtaa tggcacgact aaaatcgatc ctgatgtagc cactgcttga
                          *
```

Figure 2. Nucleotide sequence of *Streptococcus thermophilus* F410 *galK*. The start site, the ribosomal binding site (RBS), and –10 and –35 sequences are underlined; the termination site is indicated by an asterisk.

```
            1                                                  50
   Ecogalk  .......... .......... .......... .......... ..........
 Higaltkmr  .......... .......... .......... .......... ..........
    Stgalk  MRGKNLSDHG GITFETQIAP GAEQFSDFGD IKRSTKVPFE TTTEFKIETR
  Lhgalktm  .......... .......... .......... .......... ..........

            51                                                 100
   Ecogalk  .....MSLKE KTQSLFANAF GYPATHTIQA PGRVNLIGEH TD.YNDGFVL
 Higaltkmr  .....MTPIQ NAQQIFNRQH KNLPEITVYA PGRVNIIGEH TD.YNDGFVM
    Stgalk  KDF**WKSNV A*IQICRNLR T*TDSSILL. .TRKNQLDR* TH*LYGGHVF
  Lhgalktm  ......MNKE ELLKKYEATS NEKAKDVFFS PGRINVIGEH TD.YNGGHVF
                                              ::# # :::: ## :: # #

            101                                                150
   Ecogalk  PCAIDYQTVI SCAPRDDRKV RVMAADYENQ LDEFSLDA.P IVAHENYQWA
 Higaltkmr  PCAINFGTAV SGTKRDDHIW NVYAADLD.E TDEFSLNV.E IPKSEH.KWA
    Stgalk  PASITYG..T KVSQLLEKTI RS*CINEF*R *RRNSFTLNE LEYDKQAGWA
  Lhgalktm  PAPISLGVYG VYGPRDDKKV RLYSGNVDGD IVEFDIDDTN VEKDEDRFWA
            #  #   :         :::    :              :::           :  ##

            151                                                200
   Ecogalk  NYVRGVVKHL QLRNNSF.GG VDMVISGNVP QGAGLSSSAS LEVAVGTVLQ
 Higaltkmr  NYVRGVVKFI QERYPHFQQG ANLVISGNVP LSSGLSSSAA LEVAVGKFCQ
    Stgalk  NYVKGMILKL KEAGYTFDHG FELLVEGTIP NGAGLSSSAS LELLVGVVLE
  Lhgalktm  NYFKGMITYL REKYDGIDHG FNLYIEANLP SGSGLSSSAA IEMLMGIILK
            ##: #    :  :    :  #   : : :: #  : ########  :#  :#  :

            201                                                250
   Ecogalk  QLYHLPLDGA QIALHGQEAE NQFVGCNCGI MDQLISALGK KDHALLIDCR
 Higaltkmr  QLGNLPLSHT DIALNGQKAE NQFVGANCGH MDQLISALGQ ENHLLMIDCR
    Stgalk  DLFDLAIDRL ALVQTGKKVE NEFIGVNSGI MDQFAIGFGE VDKAILLDTN
  Lhgalktm  DEFNLDVDRV SLAKMGQRTE NEFIGLNSGI MDQFACIMGK KNSAIFLDCN
             :  #  :    ::  #:  # # # # # #: ####     :     :   #:

            251                                                300
   Ecogalk  SLGTKAVSMP .KGVAVVIIN SNFKRTLVGS EYNTRREQCE TGARFFQQP.
 Higaltkmr  SLETTPTPVP .QDVAVIIVN SNVPHDLVTG EYNTRRQQCE EAAKFFGVK.
    Stgalk  TLKYEMVPVK LDGYAVVIMN TNKRRELADS KYNERRSECE EALKRLQTQL
  Lhgalktm  TLKYEYLPLA LGDYEIIIMA TNHPHTLADS AYNNRVAECG RALKKLQQKL
             #      :        :: # :  #    #  :  ## #:  #:  : :  :

            301                                                350
   Ecogalk  ...ALRDVTI EEF...NAVA HELDPIVAKR VRHILTENAR TVEAASALEQ
 Higaltkmr  ...ALRDVSV EQFQKREAEL TALSPLAAKR ARHVVTENQR VLDAVEALKN
    Stgalk  TIDALGDLDS ETFEAHTDLI NDDTLI..RR ARHAVTENQR TLEAKAELEK
  Lhgalktm  DIKALGELDN DTFDEYSYLI NDETEI..KR ARHAVSENQR TLRATQAMKD
                ## :    : #               :  ## :## : :##:# :: #  :

            351                                                400
   Ecogalk  GDLKRMGELM AESHASMRDD FEITVPQIDT LVEIVKAVIG DKGGVRMTGG
 Higaltkmr  NDLTCLGKLM EASHDSMRDD FEITVPQIDY LVELVQLVIG KSGGTRMTGG
    Stgalk  GNLAAFGKLL NASHYSLRHD YEVTGIELDT LVDAAQKQEG VLGA.RMTGA
  Lhgalktm  GDLEKLGRLI NASHESLHYD YEVTGKELDT LAEASWKQPG VLGA.RMIGG
             : :#   #:#   :### # # #  # #    #: #: :      #   #  #####:

            401                                                450
   Ecogalk  GFGGCIVALI PEELVPAVQQ AVAEQYEAKT GIKETFYVCK
 Higaltkmr  GFGGCIVALA PHDKVDAVRK IIADNYEKTT GLKETFYVCT
    Stgalk  GFGGCAIALV KEEKHSRI*K *CLRCL.... .LESYWLCT
  Lhgalktm  GFGGSAIAIV KKSEAENFKK NVGKIYRDAV GYDASFYDAE
            #####:  #:            :             ::   : :
```

Figure 3. Amino acid sequence alignment of open reading frame ORF1 (*galK* gene product) with other galactokinases, as determined by the Genetics Computer Group PILEUP program. Dots represent gaps in the alignments. The numbers correspond to the amino acid position of *Streptococcus thermophilus* galactokinase. Regions containing identical amino acids (#) and conserved amino acids (:) are indicated. The species presented are *Escherichia coli*, Ecogalk; *Haemophilus influenzae*, Higalktmr; *Streptococcus thermophilus*, Stgalk; and *Lactobacillus helveticus*, Lhgalktm.

TABLE 2. Amino acid homology between the *Streptococcus thermophilus* galactokinase and galactokinases from other sources.[1]

Organism	Similarity	Identity
	(%)	
Lactobacillus helveticus	62	45
Haemophilus influenzae	58	35
Escherichia coli	55	35
Kluyveromyces lactis	50	27

[1]Values were calculated by the Genetics Computer Group GAP program.

dyltransferase, respectively) from *E. coli, S. thermophilus*, or *L. helveticus*.

Interestingly, a homology of 56.7% was found by the GCG FASTA program between a 462-bp region at position 1293 to 1735 of the *galK* gene in this work with a proposed *galK* regulatory sequence of *Clostridium pasteurianum* (3). However, this homology was with the opposite strand of our DNA sequence and in the reverse direction. Because those workers (3) did not sequence the *C. pasteurianum galK* gene, the role or the actual location of this regulatory region relative to the *galK* could not be determined in this study.

SDS-PAGE of the Galactokinase Protein

An *Eco*RI-*Sph*I fragment from pAM1 containing the *galK* gene from *S. thermophilus* F410 (Figure 1) was inserted downstream from the *lacZ* promoter of pUC19, resulting in pAM2, and transformed into *E. coli* DH5α. When cell extracts of the *E. coli* DH5α containing pAM2 were electrophoresed on SDS-PAGE, a protein band was observed of about 49 kDa (lane 1, Figure 4). This size corresponded with the predicted size of 49,168 Da calculated from the DNA sequence of the cloned *galK* gene. The intensity of this band increased when cells were induced with IPTG (lane 2). The 49-kDa protein was not detected in extracts of *E. coli* DH5α transformed with pUC19.

Northern Analysis

Northern analysis of total RNA isolated from midlog cultures of *S. thermophilus* strains F410 and 19258 grown in Elliker broth with 1.5% of glucose, lactose, or galactose are shown in Figure 5. There was no apparent hybridization between the *galK* DNA gene probe derived from strain F410 and total RNA isolated from strain F410 grown in glucose (lane 5). Hybridization signals were moderately strong (at 2.7 and 1.3 kb) for cells grown with lactose (lane 6), and signals were stronger for

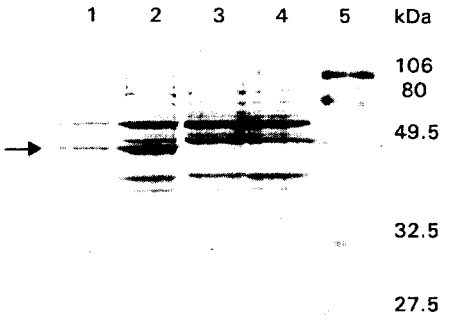

Figure 4. The SDS-PAGE of the expressed galactokinase protein from *Streptococcus thermophilus* F410. Lane 1, noninduced pAM2; lane 2, pAM2 induced by isopropyl thio-β-galactoside (IPTG); lane 3, noninduced control pUC19; lane 4, control pUC19 induced by IPTG; and lane 5, molecular mass marker.

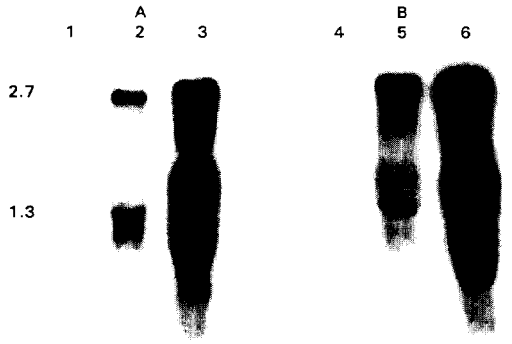

Figure 5. Northern analysis of total RNA isolated from midlog cultures of *Streptococcus thermophilus* 19258 (A) and F410 (B) grown in Elliker broth containing 1.5% glucose (lanes 1 and 4), lactose (lanes 2 and 5), or galactose (lanes 3 and 6).

galactose-grown cells (lane 7). A similar pattern was observed when strain 19258, grown in glucose, lactose, and galactose (lanes 1, 2, 3), respectively, was probed with the *galK* DNA of strain F410. Interference by the rRNA bands of 2.9 kb and 1.5 kb made the determination of the *galK* transcript size difficult to estimate, but the predominant hybridization signal in Figure 5 and subsequent blots (not shown) indicated a transcript size of approximately 1.3 to 1.4 kb, which was in reasonably good agreement with the predicted 1319-bp size of the *galK* gene.

DISCUSSION

It was previously thought that the galactose-nonfermenting phenotype (**Gal⁻**), which is common to most strains of *S. thermophilus*, was due to the absence of enzymes catabolizing galactose, the absence of a specific galactose transport system, or both (8, 25). However, galactose transport as well as enzymes of the Leloir pathway involved in galactose catabolism have been demonstrated for both Gal⁺ and Gal⁻ strains of *S. thermophilus*, suggesting that the Gal⁻ phenotype may be due to negative regulation of *gal* genes rather than to the actual absence of the relevant genes (9, 10, 25, 28). The successful isolation and cloning of the *galK* gene from *S. thermophilus* F410 in this study have provided additional evidence to support this view.

Shotgun cloning of *S. thermophilus* F410 genomic DNA fragments, generated by partial digests, resulted in the successful cloning of a 3.9 kb *Hin*dIII-*Sau*3AI fragment that complemented a *galK E. coli*. Sequence analysis of the cloned fragment indicated high DNA homologies between a region of about 1300 bp and the *galK* genes of *L. helveticus* (18), *E. coli* (4), *H. influenzae* (15), and *K. lactis* (17). Deduced amino acid sequence of the *galK* gene revealed equally high homologies with the same four galactokinase proteins (Table 2). Clusters of high homology were distributed throughout the entirety of the sequences (Figure 3).

Poolman et al. (22) reported the presence of other genes in *S. thermophilus* that code for enzymes involved in galactose metabolism. The *galE* and *galM* genes were both located upstream of the *lacSZ* gene cluster (coding for the lactose transporter and β-galactosidase) in *S. thermophilus* A147. Location of the *galK* gene in *L. helveticus* has been proposed to be upstream of *galT* on a single operon (18). In *E. coli*, the galactose genes are arranged on a single operon of the order *galETK* (1). Interestingly, however, no homologies were significant between the region downstream from the *galK* in this work and other lactose or galactose structural genes in the Gen Bank, indicating that the *galK* gene in *S. thermophilus* may be unlinked from these other genes metabolizing carbohydrate.

Northern analysis indicated that the transcription of the *galK* gene in both strains of *S. thermophilus* tested was catabolite-repressed by glucose. The lack of any *galK* transcripts in cells grown in glucose (Figure 5; lanes 1 and 5) suggests a repression mechanism with very tight control over *galK* expression in strains F410 and 19258. This result is consistent with the observation that *S. thermophilus* F410 and 19258 ferment glucose preferentially over galactose, although both strains are Gal⁺ in the absence of glucose or lactose (data not shown).

Recently, Hutkins and Ponne (11) and Poolman et al. (21) showed that most strains of *S. thermophilus* prefer to excrete galactose into the medium during growth on lactose. They suggested that excretion of galactose might occur because galactokinase was unable to compete with the efflux reaction, mediated by LacS, the lactose carrier. However, our results clearly demonstrate that the transcription of the *galK* gene is dramatically reduced in the presence of glucose and that decreased expression of galactokinase may also be responsible for the galactose-releasing and Gal⁻ phenotype that is typical of most strains of *S. thermophilus* when grown on lactose. Because *S. thermophilus* is used in the manufacture of Mozzarella and other cheeses in which the presence of galactose is undesirable, efforts aimed at depressing *galK* may result in the availability of improved *S. thermophilus* starter culture strains.

ACKNOWLEDGMENTS

We acknowledge the National Dairy Promotion and Research Board and the University of Nebraska Center for Biotechnology for their support of this project. We thank Suzette Spratt for technical assistance.

REFERENCES

1 Adhya, S. 1987. The galactose operon. Page 1503 *in* *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. Vol. 2. F. C. Neidhart, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger, ed. Am. Soc. Microbiol., Washington, DC.

2 Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72:248.

3 Daldal, F., and J. Applebaum. 1985. Cloning and expression of *Clostridium pasteurianum* galactokinase gene in *Escherichia coli* K-12 and nucleotide sequence analysis of a region affecting the amount of the enzyme. J. Mol. Biol. 186:533.

4 Debouck, C., A. Riccio, D. Schumperli, K. McKenny, J. Jeffers, C. Hughes, M. Rosenberg, M. Heusterspreute, F. Brunel, and J. Davison. 1985. Structure of the galactokinase gene of *Escherichia coli*, the last(?) gene of the *gal* operon. Nucleic Acids Res. 13:1841.

5 Elliker, P. R., A. W. Anderson, and G. Hannesson. 1956. An agar culture medium for lactic acid streptococci and lactobacilli. J. Dairy Sci. 39:1611.

6 Hardie, J. M. 1986. Genus *Streptococcus* Rosenbach 1884. Page 1043 *in* Bergey's Manual of Systematic Bacteriology. Vol. 2. P.H.A. Sneath, N. S. Mair, M. E. Sharpe, and J. G. Holt, ed. Williams and Wilkins, Baltimore, MD.

7 Herold, K. M., and P. G. Rothberg. 1988. Evidence for a labile intermediate in the butyrate induced reduction of the level of c-*myc* RNA in SW837 rectal carcinoma cells. Oncogene 3:423.

8 Hutkins, R. W., and H. A. Morris. 1987. Carbohydrate metabolism by *Streptococcus thermophilus*: a review. J. Food Prot. 50:876.

9 Hutkins, R. W., H. A. Morris, and L. L. McKay. 1985. Galactose transport in *Streptococcus thermophilus*. Appl. Environ. Microbiol. 50:772.

10 Hutkins, R. W., H. A. Morris, and L. L. McKay. 1985. Galactokinase activity in *Streptococcus thermophilus*. Appl. Environ. Microbiol. 50:777.

11 Hutkins, R. W., and C. Ponne. 1991. Lactose uptake driven by galactose efflux in *Streptococcus thermophilus*: evidence for a galactose-lactose antiporter. Appl. Environ. Microbiol. 57:941.

12 Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (Lond.) 227:680.

13 Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall. 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. 193:265.

14 Maniatis, T., E. F. Fritsch, and J. Sambrook. 1989. Molecular Cloning, A Laboratory Manual. 2nd ed. Cold Spring Harbor Press, Cold Spring Harbor, NY.

15 Maskell, D. J., M. J. Szabo, M. E. Deadman, and E. R. Moxon. 1992. The *gal* locus from *Haemophilus influenzae*: cloning, sequencing, and the use of *gal* mutants to study liposaccharide. Mol. Microbiol. 6: 3051.

16 McKay, L. L. 1982. Regulation of lactose metabolism in dairy streptococci. Page 153 *in* Developments in Food Microbiology. R. Davies, ed. Appl. Sci. Publ. Ltd., Essex, England.

17 Meyer, J., A. Walker-Jonah, and C. P. Hollenberg. 1991. Galactokinase encoded by GAL1 is a bifunctional protein required for the induction of the GAL genes in *Kluyveromyces lactis* and able to suppress the GAL3 phenotype in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 11:5454.

18 Mollet, B., and N. Pilloud. 1991. Galactose utilization in *Lactobacillus helveticus*: Isolation and characterization of the galactokinase (*galK*) and galactose-1-phosphate uridyl transferase (*galT*) genes. J. Bacteriol. 173:4464.

19 O'Leary, V. S., and J. H. Woychik. 1976. Utilization of lactose, glucose, and galactose by a mixed culture of *Streptococcus thermophilus* and *Lactobacillus bulgaricus* in milk treated with lactase enzyme. Appl. Environ. Microbiol. 32:89.

20 Pitcher, D. G., N. A. Saunders, and R. J. Owen. 1989. Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Lett. Appl. Microbiol. 8:151.

21 Poolman, B., T. J. Royer, S. E. Mainzer, and B. F. Schmidt. 1989. Lactose transport system of *Streptococcus thermophilus*: a hybrid protein with homology to the melibiose carrier and enzyme III of phosphoenolpyruvate-dependent phosphotransferase systems. J. Bacteriol. 171:244.

22 Poolman, B., T. J. Royer, S. E. Mainzer, and B. F. Schmidt. 1990. Carbohydrate utilization in *Streptococcus thermophilus*: characterization of the genes for aldose 1-epimerase (mutarotase) and UDP glucose-4-epimerase. J. Bacteriol. 172:4037.

23 Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463.

24 Schroeder, C. J., C. Robert, G. Lenzen, L. L. McKay, and A. Mercenier. 1991. Analysis of the *lacZ* sequences from two *Streptococcus thermophilus* strains: comparison with the *Escherichia coli* and *Lactobacillus bulgaricus* β-galactosidase sequences. J. Gen. Microbiol. 137:369.

25 Somkuti, G. A., and D. H. Steinberg. 1979. Adaptability of *Streptococcus thermophilus* to lactose, glucose, and galactose. J. Food Prot. 42:885.

26 Terzaghi, B. E., and W. E. Sandine. 1975. Improved medium for lactic streptococci and their bacteriophage. Appl. Environ. Microbiol. 29:807.

27 Thomas, T. D., and V. L. Crow. 1983. Lactose and sucrose utilization by *Streptococcus thermophilus*. Fed. Eur. Microbiol. Soc. Microbiol. Lett. 17:13.

28 Thomas, T. D., and V. L. Crow. 1984. Selection of galactose-fermenting *Streptococcus thermophilus* in lactose-limited chemostat cultures. Appl. Environ. Microbiol. 48:186.

29 Tinson, W., A. J. Hillier, and G. R. Jago. 1982. Metabolism of *Streptococcus thermophilus*. I. Utilization of lactose, glucose, and galactose. Aust. J. Dairy Technol. 38:8.

30 Zhen, L., and R. T. Swank. 1993. A simple and high yield method for recovering DNA from agarose gels. Biotechniques 14:894.