

Isolation and Characterization of Transcription Signal Sequences from *Streptococcus thermophilus*

Daniel K.Y. Solaiman, George A. Somkuti

U.S. Department of Agriculture, ARS, Eastern Regional Research Center, 600 East Mermaid Lane, Wyndmoor, PA 19038, USA

Received: 22 August 1996 / Accepted: 24 September 1996

Abstract. *Streptococcus thermophilus* (ST) chromosomal DNA fragments generated by partial *Sau3A* digestion were cloned into the unique *Bam*HI site upstream from the promoterless chloramphenicol acetyltransferase (*cat*) gene of the *Escherichia coli* (EC) promoter-probe vector pKK520-3. Recombinant plasmids containing ST sequences with transcription-activation activity were isolated from chloramphenicol-resistant (Cm^R) EC transformants. A promoterless *Streptomyces antibioticus* melanin biosynthesis operon (*melC*) was inserted immediately downstream from the ST sequence to identify DNA with strong promoter activity. Several ST transcription-activation sequences, termed ST_ps, were isolated and subcloned, and their nucleic acid sequences determined. The -10 and -35 consensus sequences were identified in these putative ST promoters. Detailed analysis of ST_{P3306} sequence data revealed two partial open reading frames (ORFs) with high degrees of homology to prokaryotic GTP-binding protein and DNA repair enzyme, thus providing valuable information for further study on DNA maintenance in this important lactic acid bacterium.

Lactic acid bacteria (LAB) are used in the dairy and food industry for the production of fermented products. The main function of LAB in the fermentation process is to impart characteristic flavors and textures, and to provide biopreservative activity to the final products. Recently, some species of LAB have been promoted as probiotics and used as supplements in certain nutraceutical food products. The ability to express heterologous genes in these industrially important LAB should thus lead to the development of novel processes and new products in the dairy food fermentation industry.

Streptococcus thermophilus is used in starter cultures for the production of yoghurt and several cheese varieties. Efforts aimed at metabolic engineering of this microbe have led to the development of DNA transfer methodologies [18; reviewed in 9] and the construction of gene cloning vectors [14, 20]. Heterologous genes introduced into ST, however, are not always efficiently expressed [16, 19]. Promoter sequences from the organ-

ism have proven useful in affecting expression of cloned genes in ST [15, 20]. In this paper, we report the cloning and characterization of new promoters from *S. thermophilus*. These promoters add to the collection of ST transcription signals [5, 13] potentially useful for ensuring heterologous gene expression in genetically engineered LAB.

Materials and Methods

Bacteria, growth conditions, and plasmids. *Escherichia coli* DH5 α (BRL Life Technologies, Gaithersburg, MD, USA) and *S. thermophilus* ST128 (laboratory collection) were grown in Luria-Bertani (LB) and tryptone–yeast–lactose broth, respectively, as described [15, 20]. Antibiotics were employed at the concentrations of 100 μ g/ml ampicillin (Ap) and 35 μ g/ml chloramphenicol (Cm) for *E. coli*, and 5–10 μ g/ml Cm and 5–15 μ g/ml erythromycin (Em) for ST cultures. Plasmids used in this study have been described in detail elsewhere [15].

Molecular biology procedures. Restriction endonucleases, T4 DNA ligase, and various DNA-modifying enzymes were purchased from and used according to the protocols supplied by the following vendors: BRL Life Technologies, New England Biolabs (Beverly, MA, USA), and United States Biochemical (Cleveland, OH, USA). Subcloning efficiency-competent DH5 α was transformed with plasmid DNA by a heat-shock procedure described by the supplier. *S. thermophilus* was electrotransformed according to a previously reported protocol [18]. Recombinant plasmid DNAs were isolated from transformed *E. coli* and *S. thermophilus* as described previously [14, 17]. Agarose gel electropho-

ST_{P1205}

```
1 GATCCTGATGAATTGGTTGATTTCCGTGAAATTATGGATAGTATTGTTCAAGATTCTAAT
-10
61 CGTATTTATCGAGATC
```

ST_{P3101}

```
1 GATCGTTTGGATTACCGTCCAAATGCGGTTGCGCGTGGCTTGGCAAGTAAAAACAAC
61 ACTGTAGGAGTTGTCATTCCAAATATGTAATAGCTATTTTGTACTCTAGCTAAAGGT
-35 -10
121 ATTGATGACATTGCAACCATGTATAAGTATAATATTGTTCTTGCTCCAGTGATGATAAT
181 GAGGATC
```

ST_{P3306}

```
ORF2-3306
1 GATCAAATCACAGACCATCCAGAGCGTTTCTGGTATCTGAGATGATTCGTGAAAAAGTC
D Q I T D H P E R F L V S E M I R E K V
61 CTACAGTTGACGCGTGAAGAAATCCGCACTCAGTTGCTGTTGTCATTGATCTATGAAG
L Q L T R E E I P H S V A V V I D S M K
121 CGTGATGAAGAGACTGATACGTTTACATTCGTGCGACCATCATGGTTGAGCGTGATAGC
R D E E T D T V H I R A T I M V E R D S
181 CAAAAAGGTATGTGCATCGGTAAAAAGGTGCTATGCTTAAGAAAATGGTACTCTTGCT
Q K G I V I G K K G A M L K K I G T L A
241 CGTAAAGATATTGAGCTCATGCTTGGGGACAAGGTCTTCTTGAAGTGGGTGAAGGTT
R K D I E L M L G D K V F L E T W V K V
-35 -10
301 AAGAAAATTGGAGGGATAAAAACTTGATTTGGCCGATTTTGGTTATAATGAAAAAGAATA
K K I G G I K T -
S-D ORF1-3306
361 CTAAGAAAAGCTGGGCACTTGCTCGGCTTTCTTTCTAGAAGGAGGGGCAGATGCCTGAGT
M P E
421 TACCTGAGGTAGAAACAGTTAGACGAGGTTAGAAAAGCTGTTACTAGGAAGGACTATTT
L P E V E T V R R G L E K L L L G R T I
481 TGTCACCTGAAGTTAAGGTCCCTAAAATGATC
L S L E V K V P K M I
```

Fig. 1. Nucleic acid sequences of ST_Ps. The consensus -10 and -35 sequences are underlined. Putative Shine-Dalgarno sequence in ST_{P3306} is marked as S-D. Translated amino acid sequences of ORF1-3306 and ORF2-3306 are shown.

resis was performed with the TBE buffer system (0.089 M Tris-base, 0.089 M boric acid, 0.002 M Na-EDTA, pH 8). DNA fragments were isolated either by electroelution with a SixPac GE200 Eluter (Hofer Scientific Instrument, San Francisco, CA, USA) or by centrifugation on GenElute™ spin columns (Supelco Inc., Bellefonte, PA, USA). When DNA with high purity was needed, the sample was processed through the CsCl-gradient ultracentrifugation procedure [2]. Alternatively, Elutip-d columns (Schleicher & Schuell, Keene, NH, USA) were used to purify and concentrate DNA samples.

DNA sequence analysis. Nucleic acid sequence was determined by the dideoxynucleotide chain termination method [11] with AutoRead™ or AutoCycle™ Sequencing Kit (Pharmacia Biotech, Piscataway, NJ, USA). Double-stranded plasmid templates were prepared by CsCl-gradient ultracentrifugation [2] or by using the Perfectprep™ DNA-binding matrix and the accompanying protocol (5' → 3', Inc., Boulder, CO, USA). Sequence information was collected and processed by the ALF DNA Sequencer and Fragment Analysis System (Pharmacia Biotech). Database searches were performed with the BLAST program and the default parameter settings [1]. Multiple sequence alignments were carried out using DNASIS™ WINDOWS 2.1 (Hitachi Software

Engineering America, San Bruno, CA, USA) or the CLUSTAL-W [22] programs.

Results and Discussion

Isolation of ST transcription signals. *S. thermophilus* ST128 chromosomal DNA fragments containing transcription activation activity were isolated with the *cat*-based promoter-probe vector pKK520-3 [4]. The ST chromosomal DNA was first partially digested with *Sau3A* and then shotgun-cloned into the unique *Bam*HI-site of pKK520-3. Chloramphenicol-resistant *E. coli* transformants were screened for plasmids that contained inserts in the range of ca. 100–800 bps. Eighteen such recombinants, designated as pKST-series plasmids, were pooled and double-digested with *Bgl*II and *Hind*III. The *Bam*HI/*Hind*III fragment (1.4 kb) of pMCU23a [15] containing a

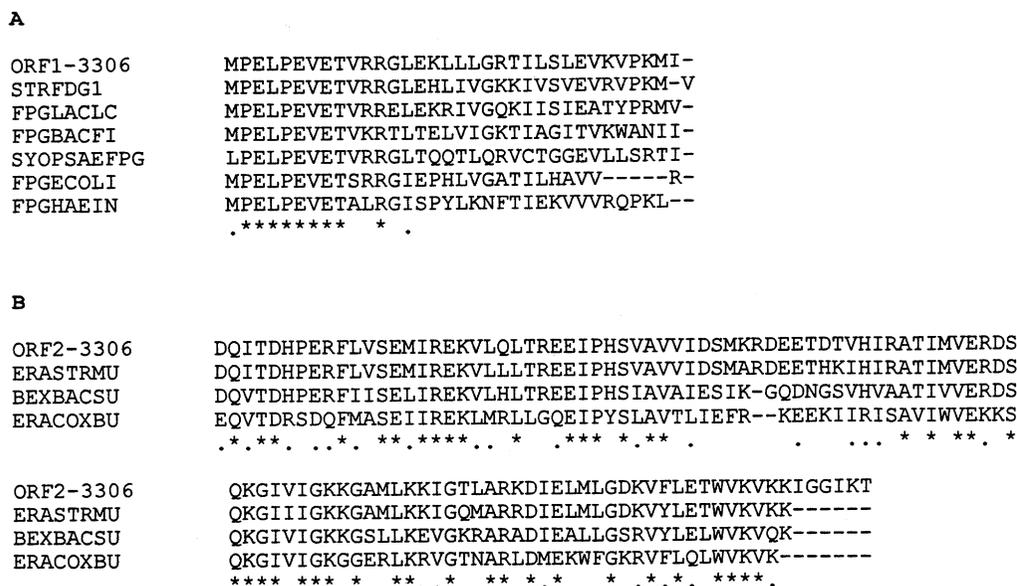


Fig. 2. Multiple sequence alignments of ORF1-3306 and ORF2-3306. The stars indicate perfect matches in all sequences compared, and the dots mark the conservedly substituted residues. (A) ORF1-3306 aligned with the N-terminal amino acid sequences of the formamidopyrimidine-DNA glycosylases from *Streptococcus mutans* (STRFDG1; GenBank D26071), *Lactococcus lactis* (FPGLACLC; GenBank X74298), *Bacillus firmus* (FPGBACFI; GenBank X53930), *Synechococcus* sp. PCC 6301 (SYOPSAEFPG; GenBank M99432), *E. coli* (FPGECOLI; GenBank X06036), and *Haemophilus influenzae* (FPGHAEIN; GenBank L45584). (B) ORF2-3306 sequence was aligned with the C-terminal sequences of the GTP-binding proteins or Era homologs from *Streptococcus mutans* (ERASTRMU; GenBank L03428), *Bacillus subtilis* (BEXBACSU; GenBank U18532), and *Coxiella burnetii* (ERACOXBU; GenBank L27436).

promoterless *mel* operon from *Streptomyces antibioticus* was directionally cloned downstream from the ST sequences in the *Bg*III/*Hind*III-digested pKST recombinant plasmids. The resultant pKST_{mel} plasmids were introduced into *E. coli* and screened for melanin pigment production. Five clones containing pKST_{mel}-1201, -1205, -2201, -3101, and -3306 were initially selected for further studies. Promoter ST_{P2201} associated with the pKST_{mel}2201 has previously been characterized [15; Somkuti and Solaiman, unpublished data].

Characterization of ST_P's sequences. The ST_P-containing *Eco*RI/*Sph*I fragments of the pKST_{mel}'s were subcloned into the same restriction sites of pUC19 to generate the corresponding pUCP_x plasmids. Sequences of the inserts were determined with the M13 universal and reverse primers provided in the sequencing kits. When necessary, the inserts were shortened or further digested and subcloned to allow for complete sequencing of the entire inserts. Figure 1 shows the sequences determined for ST_{P1205} (76 bp), ST_{P3101} (187 bp), and ST_{P3306} (504 bp). The ST_{P1201} (sequence not shown) is identical to the previously described ST_{P2201} [15, Somkuti and Solaiman, unpublished data]. The G + C contents of these sequences were 32.9%, 37.4%, and 41.3% for ST_{P1205}, ST_{P3101}, and ST_{P3306} respectively. These values agree with the reported low G + C contents (37–40%) of

S. thermophilus chromosome [7], suggesting that the isolated ST_Ps were indeed originated from ST DNA. Putative -10 and -35 hexanucleotide blocks were assigned on these sequences based on the consensus promoter sequences of Gram-positive bacteria [6]. The -10 sequences of the ST_Ps closely conform to the TATAAT consensus, while the -35 regions are slightly divergent from the conserved TTGACA. As with most prokaryotic promoters, the two hexanucleotides in ST_{P3101} were separated by 17 bases. The ST_{P1205} and ST_{P3306}, however, contained 13- and 14-base interconsensus spacings, respectively, that have also been reported for a few LAB promoters [5, 8].

In order to further characterize the isolated ST sequences, potential ORFs with ATG start codons were tentatively assigned to sequences downstream from the -10 region of the ST_Ps. Database searches with the amino acid sequences of these putative ORFs revealed that the ORF1 of ST_{P3306}, designated as ORF1-3306, is highly homologous to the conserved N-terminal sequences of prokaryotic formamidopyrimidine-DNA glycosylases (Fpg; Fig. 2A). This enzyme is thought to be involved in the excision repair of alkylated DNA [3]. As expected with the coding sequence of a protein, a putative Shine-Dalgarno sequence was identified upstream from the ORF1-3306 (Fig. 1). The ORF1s (not shown) of the

other two cloned promoters did not yield meaningful results in the database searches.

The sequences upstream from the -10 region of the three ST_Ps were further analyzed by first translating them into ORFs in the three possible reading frames. Results of database searches with these putative ORFs showed that ORF2-3306 (Fig. 1) exhibited strong homology with the C-terminal sequences of a class of microbial proteins variously known as GTP-binding protein and Era homolog (Fig. 2B). Although the exact function of these proteins in vivo is not clear, their coding sequences are frequently located adjacent to the genes coding for nucleic acid-modifying enzymes [21, 23]. This pattern recurred in *S. thermophilus*, as shown by the organization of ST_{P3306} in which the genes for GTP-binding protein homolog and the putative DNA repair enzyme were adjacent to each other. It thus seems highly possible that this protein is also somehow involved in DNA metabolism. As part of the chromosomal region that codes for nucleic acid metabolism enzymes, ST_{P3306} is thus valuable as a potentially useful transcription activation element for the enhanced expression of genes encoding such biochemical activities as the peptidases [10] and β -galactosidase [12], which play important roles in determining the attributes of dairy fermentation products. Furthermore, the sequence information contained in ST_{P3306} provides important leads for future study and manipulation of the DNA repair and maintenance mechanisms crucial to the genotypic stability of genetically engineered *S. thermophilus* and possibly other LAB.

ACKNOWLEDGMENT

We thank Dennis H. Steinberg for providing excellent technical expertise.

Literature Cited

1. Altschul SF, Gish W, Miller W, Myers EG, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
2. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1987) *Current protocols in molecular biology*. New York: Wiley-Interscience
3. Boiteux S, O'Connor TR, Laval J (1987) Formamidopyrimidine-DNA glycosylase of *Escherichia coli*: cloning and sequencing of the *fpg* structural gene and overproduction of the protein. *EMBO J* 6:3177–3183
4. Brosius J (1992) Compilation of superlinker vectors. *Methods Enzymol* 216:469–483
5. Constable A, Mollet B (1994) Isolation and characterisation of promoter regions from *Streptococcus thermophilus*. *FEMS Microbiol Lett* 122:85–90
6. Graves MC, Rabinowitz JC (1986) In vivo and in vitro transcription of the *Clostridium pasteurianum* ferredoxin gene. *J Biol Chem* 261:11409–11415
7. Hardie JM, Whiley RA (1992) The genus *Streptococcus*—Oral. In: Balows A, Trüper HG, Dworkin M, Harder W, Schleifer K-H (eds) *The prokaryotes*, vol. 2, 2nd edn. New York: Springer-Verlag, chapter 64
8. Koivula T, Sibakov M, Palva I (1991) Isolation and characterization of *Lactococcus lactis* subsp. *lactis* promoters. *Appl Environ Microbiol* 57:333–340
9. Mercenier A (1990) Molecular genetics of *Streptococcus thermophilus*. *FEMS Microbiol Rev* 87:61–77
10. Meyer J, Howald D, Jordi R, Fürst M (1989) Location of proteolytic enzymes in *Lactobacillus lactis* and *Streptococcus thermophilus* and their influence on cheese ripening. *Milchwissenschaft* 44:678–681
11. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467
12. Schroeder DJ, Robert C, Lenzen G, McKay LL, Mercenier A (1991) Analysis of the *lacZ* sequences from two *Streptococcus thermophilus* strains: comparison with *Escherichia coli* and *Lactobacillus bulgaricus* β -galactosidase sequences. *J Gen Microbiol* 137:369–380
13. Slos P, Bourquin JC, Lemoine Y, Mercenier A (1991) Isolation and characterization of chromosomal promoters of *Streptococcus salivarius* subsp. *thermophilus*. *Appl Environ Microbiol* 57:1333–1339
14. Solaiman DKY, Somkuti GA (1993) Shuttle vectors developed from *Streptococcus thermophilus* native plasmid. *Plasmid* 30: 67–78
15. Solaiman DKY, Somkuti GA (1995) Expression of *Streptomyces melC* and *choA* genes by a cloned *Streptococcus thermophilus* promoter ST_{P2201}. *J Ind Microbiol* 15:39–44
16. Solaiman DKY, Somkuti GA, Steinberg DH (1992) Construction and characterization of shuttle plasmids for lactic acid bacteria and *Escherichia coli*. *Plasmid* 28:25–36
17. Somkuti GA, Steinberg DH (1986) General method for plasmid DNA isolation from thermophilic lactic acid bacteria. *J Biotechnol* 3:323–332
18. Somkuti GA, Steinberg DH (1987) Genetic transformation of *Streptococcus thermophilus* by electroporation. In: Neijssel O, Van der Meer R, Luyben K (eds) *Proceedings of the 4th European Congress on Biotechnology*, Vol. 1, p 412, Amsterdam
19. Somkuti GA, Solaiman DKY, Steinberg DH (1993) Cloning of a tyrosinase gene in *Streptococcus thermophilus*. *Biotechnol Lett* 15:773–778
20. Somkuti GA, Solaiman DKY, Steinberg DH (1995) Native promoter-plasmid vector system for heterologous cholesterol oxidase synthesis in *Streptococcus thermophilus*. *Plasmid* 33:7–14
21. Song B-H, Neuhaud J (1989) Chromosomal location, cloning and nucleotide sequence of the *Bacillus subtilis cdd* gene encoding cytidine/deoxycytidine deaminase. *Mol Gen Genet* 216:462–468
22. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
23. Zuber M, Hoover TA, Powell BS, Court DL (1994) Analysis of the *rnc* locus of *Coxiella burnetii*. *Mol Microbiol* 14:291–300