Isolation, characterization and nucleotide sequence of the *Streptococcus mutans* lactose-specific Enzyme II (*lacE*) gene of the PTS and the phospho- β -galactosidase (*lacG*) gene

ALLEN L. HONEYMAN* and ROY CURTISS, III

Department of Biology, Washington University, St Louis, MO 63130, USA

(Received 17 March 1993; revised 2 June 1993; accepted 10 June 1993)

The *lacE* and *lacG* genes from *Streptococcus mutans* have been isolated and characterized, and their nucleotide sequence has been determined. The *lacE* gene encodes the lactose-specific Enzyme II component of the phosphoenolpyruvate-dependent phosphotransferase system (PTS). The *lacG* gene encodes the phospho- β -galactosidase which cleaves the lactose phosphate that is formed by the lactose PTS. The *S. mutans lacE* and *lacG* genes are located in the same operon as the tagatose genes. *S. mutans* metabolizes lactose via the tagatose phosphate pathway. The deduced LacE and LacG proteins of *S. mutans* display high homology with the corresponding proteins from *Lactococcus lactis*, *Staphylococcus aureus* and *Lactobacillus casei*.

Introduction

Streptococcus mutans, the causative agent of dental caries (Hamada & Slade, 1980), transports various dietary carbohydrates via the phosphoenolpyruvate-dependent phosphotransferase system (PTS) (Postma & Lengeler, 1985). The transported carbohydrates are metabolized to produce large amounts of lactic acid (Drucker & Melville, 1968) which initiates the demineralization of tooth enamel and the start of caries formation. Although sucrose is the major dietary carbohydrate involved with caries formation, other dietary carbohydrates, including lactose, have been shown to induce caries in studies using animal models (Guggenheim *et al.*, 1966; Frostell *et al.*, 1967). Lactose, which is present in the diet of most individuals in high concentrations due to the consumption of various dairy products, may therefore contribute to the production of dental caries throughout one's lifetime.

Lactose is metabolized by *S. mutans* (Hamilton & Lebttag, 1979) and other Gram-positive organisms via the tagatose phosphate pathway (Bisset & Anderson, 1973). In *Staphylococcus aureus* (Breidt *et al.*, 1987),

Lactococcus lactis (de Vos *et al.*, 1990), *Lactobacillus casei* (Alpert & Chassy, 1988, 1990) and *S. mutans* (Rosey & Stewart, 1992), the enzymes involved in the transport of lactose are the Factor III^{Lac} (*lacF*) and Enzyme II^{Lac} (*lacE*) gene products of the PTS. Mutations in either of these loci in *Staph. aureus* eliminates the utilization of lactose (Bisset & Anderson, 1974). Lactose is vectorially transported by the PTS to form lactose phosphate, which is cleaved by the enzyme phospho- β -galactosidase to form glucose and galactose 6-phosphate. The glucose moiety produced by the cleavage of lactose phosphate may be directly utilized via the glycolytic pathway while the galactose 6-phosphate moiety is sequentially converted by the tagatose pathway to tagatose 6-phosphate, tagatose 1,6-bisphosphate, and glyceraldehyde 3-phosphate and dihydroxyacetone phosphate prior to utilization in glycolysis. This is in contrast to the Leloir pathway of galactose metabolism found in *E. coli* where galactose is sequentially converted to galactose 1-phosphate, UDP-galactose, UDP-glucose and glucose 1-phosphate before utilization in glycolysis (Maxwell *et al.*, 1962). However, preliminary reports indicate that *S. mutans* may contain the genes for galactose utilization via the Leloir pathway (D. Ajdic, R. R. B. Russell & J. J. Ferretti, personal communication; Hamilton & Lebttag, 1979; A. L. Honeyman & R. Curtiss, III, unpublished results) in addition to the genes for utilization by the tagatose pathway (Hamilton & Lebttag, 1979).

We have previously described the isolation of the

* Author for correspondence. Tel. +1 314 935 4662; fax +1 314 935 4432.

Abbreviation: PTS, phosphotransferase system.

The GenBank accession number for the sequence reported in this paper is L18993.

tagatose genes from *S. mutans* (Smorawinska *et al.*, 1983) and the characterization of their gene products (Jagusztyn-Krynica *et al.*, 1992). G. C. Stewart and co-workers have shown that the organization of the lactose genes in *Staph. aureus* (which includes the tagatose genes) are arranged within a single operon structure (Rosey *et al.*, 1991). A similar operon structure also exists in *Lactococcus lactis* (van Rooijen *et al.*, 1991). Therefore, we hypothesized that the genes for the transport and cleavage of lactose, the lactose Enzyme II (*lacE*) and phospho- β -galactosidase (*lacG*) genes, are located on the *S. mutans* chromosome adjacent to the previously isolated tagatose genes (Smorawinska *et al.*, 1983). In this report, we substantiate this hypothesis and describe the isolation, characterization and nucleotide sequence of the *lacE* and *lacG* genes of *S. mutans*.

Methods

Bacterial strains and media. *S. mutans* serotype *c* strain UA130 was used as the recipient strain for streptococcal DNA transformations (Murchison *et al.*, 1986). *E. coli* host strain CC118 (Manoil & Beckwith, 1985) was used for all recombinant DNA manipulations.

S. mutans strains were grown in Brain-Heart Infusion (BHI) broth or on BHI agar plates (Difco). *E. coli* strains were grown in Tryptic Soy Broth (TSB) or on Tryptic Soy Agar (TSA) plates (Difco). Antibiotics used for selection with *E. coli* were ampicillin, 100 $\mu\text{g ml}^{-1}$, and chloramphenicol, 20 $\mu\text{g ml}^{-1}$, and for selection with *S. mutans*, erythromycin, 10 $\mu\text{g ml}^{-1}$.

DNA isolation and manipulation. Streptococcal chromosomal DNA was isolated in the following manner. A 5 ml starter culture was grown overnight in BHI broth in a 15 ml plastic screw-cap tube. This was used to inoculate 500 ml of pre-warmed BHI broth in a 500 ml flask. This culture was grown to stationary phase overnight at 37 °C. The cells were harvested by centrifugation and washed in TE buffer (10 mM-Tris/HCl, 1 mM-EDTA, pH 8.0). Following the wash step, the cells were resuspended in 6 ml of TE buffer and incubated for 1 h at 65 °C. The cells were then cooled on ice prior to the addition of 1 ml 10 mg lysozyme ml^{-1} and 200 μl 5000 units mutanolysin ml^{-1} (Sigma). The cell suspension was incubated for 1 h at 37 °C. Pronase (1 ml of 10 mg ml^{-1} stock solution) was added and incubation continued for 1 h. The cells were lysed by the addition of 1 ml 20% (w/v) Sarkosyl. An additional 1 ml of Pronase (Sigma) was added prior to a 1 h incubation at 37 °C. The cell lysate was brought to a volume of 17.6 ml and 19.4 g caesium chloride and 3.0 ml ethidium bromide (10 mg ml^{-1}) were added. The DNA solution was placed in heat-sealable tubes and centrifuged at 50000 r.p.m. for 20 h in a Sorvall TTF65.13 rotor. The DNA band was removed from the tube with a syringe, extracted with water-saturated butanol, and dialysed overnight against TE buffer. The DNA solution was brought to a volume of 9.2 ml and 6.8 g guanidine hydrochloride added. The DNA solution was incubated at 55 °C for 1.5 h prior to being layered onto a two-step caesium chloride gradient. The top layer of the gradient was 38% (w/v) caesium chloride while the bottom layer was 60% (w/v). Each layer contained 20 μl ethidium bromide (10 mg ml^{-1}). The gradient was centrifuged at 30000 r.p.m. for 24 h at 4 °C in a Beckman SW41 rotor. The DNA band was removed with a Pasteur pipette, extracted with water-saturated butanol, and dialysed overnight against TE buffer prior to use.

Plasmid DNA isolations for either large scale or miniprep protocols were done as described by Sambrook *et al.* (1989) using the PEG

precipitation method. Quick-lysis preparations of plasmid DNA for initial screening procedures were done according to a protocol provided by Stratagene with their mung bean nuclease protocol. All DNA manipulations and enzyme procedures were done according to the manufacturer's instructions and under NIH recombinant DNA guidelines.

DNA sequence analysis. The complete nucleotide sequence of the 2.9 kb *Bam*HI–*Hind*III fragment contained in pYA3120 (this study, Fig. 1) was determined on both DNA strands. Sequenase-modified T7 DNA polymerase (Tabor & Richardson, 1987) from USB was used in the double-stranded sequencing technique of Chen & Seeborg (1985). A nested deletion series was generated by the procedure of Henikoff (1984) from either end of the cloned fragment for use as the template DNA for the sequencing reactions. The nucleotide sequence 5' of the *Bam*HI site (bp 138 to 388) was determined using pYA3119 (this study, Fig. 1) as the template DNA and a synthetic oligonucleotide from Midland Certified Reagents (Midland, Texas, USA) as the primer. The sequence in this region (250 bp) was determined on one strand only and agrees exactly with the data of Rosey & Stewart (1992). The nucleotide sequence of bp 1–138 was taken from Rosey and Stewart to complete the open reading frame of the *lacE* gene. Nucleotide sequence analysis was done using the following computer programs: Genepro (Riverside Scientific), Seqnce (Access Biosystems), GCG (University of Wisconsin) (Devereux *et al.*, 1984) and TFASTA (Pearson & Lipman, 1988).

Phospho- β -galactosidase assays. Assays for phospho- β -galactosidase activity were done according to the procedure described by Miller (1972) for β -galactosidase except that the substrate ONPG-phosphate (Sigma) was substituted for ONPG.

Results and Discussion

Isolation of the *S. mutans lacE* and *lacG* genes

Our previous results have indicated that the genes involved with galactose and tagatose utilization were contained on the plasmid pYA501 (Smorawinska *et al.*, 1983). Based upon the structure of the lactose operons in *Staph. aureus* and *L. lactis*, we hypothesized that the *S. mutans* PTS lactose Enzyme II (*lacE*) and phospho- β -galactosidase (*lacG*) genes were located on the chromosome adjacent to the 5.1 kb DNA fragment contained in pYA501. To isolate the adjacent DNA, a chromosome walking technique using the non-replicative shuttle vector pVA891 was used (Macrina *et al.*, 1983). The plasmid pVA891 contains the origin of replication and the chloramphenicol resistance marker from pACYC184 and an erythromycin resistance marker for selection in streptococci. This plasmid will not replicate in *S. mutans*. Streptococcal DNA fragments cloned into pVA891 can facilitate recombination with the homologous site on the streptococcal chromosome. Selection for erythromycin resistant *S. mutans* transformants allows the isolation of colonies in which Campbell insertions into the chromosome have occurred.

The 600 bp *Hind*III–*Bam*HI fragment located at the end of the *Bam*HI fragment in pYA501 was cloned into pVA891 and the resulting plasmid was designated pYA3117 (Fig. 1). This plasmid was used to transform *S.*

Fig. 1. Organization of the *lac* operon of *S. mutans*. The various ORFs are shown above a partial restriction map of the region. The DNA fragments contained in the listed plasmids are indicated by the dark lines shown below the restriction map.

mutans strain UA130, and erythromycin resistant colonies were selected. Chromosomal DNA was isolated from several transformants and screened by Southern hybridization (Southern, 1975) to determine that only one copy of pYA3117 had integrated into the chromosome at the correct site (data not shown). One such *S. mutans* isolate was designated SMS206 and was used for further study.

Chromosomal DNA was isolated from SMS206, digested with various restriction endonucleases which would only cut outside of the origin of replication and the chloramphenicol resistance marker, and ligated to itself at low DNA concentrations which favour intramolecular ligation. The ligated DNA was used to transform *E. coli* strain CC118 and the transformation mixture grown in TSB with selection for chloramphenicol resistance. This procedure should allow the isolation of chromosomal DNA adjacent to the 600 bp *Hind*III–*Bam*HI fragment. Chromosomal DNA from this region should contain the *S. mutans* *lacE* and *lacG* genes.

Two independent plasmid isolates were obtained in this manner. One isolate, generated with the restriction endonuclease *Ava*I, contained a 1.5 kb chromosomal DNA fragment originating from the *Hind*III site in pYA501 and pYA3117 and was designated pYA3118. The other plasmid isolate, generated with the restriction endonuclease *Sph*I, contained a 6 kb chromosomal DNA fragment originating from the same *Hind*III site and was designated pYA3119. The chromosomal inserts contained in both plasmids are depicted in Fig. 1. Based upon the organization of the *Staph. aureus* and *L. lactis* *lac* operons and their gene products, the organization of the tagatose genes of *S. mutans*, and the predicted size of the *S. mutans* Enzyme II and phospho-β-galactosidase proteins, we hypothesized that approximately 3 kb of chromosomal DNA adjacent to the *Bam*HI site in pYA501 was needed to encode both genes. If this hypothesis were correct, pYA3118 would not contain the entire phospho-β-galactosidase gene and would not exhibit phospho-β-galactosidase activity, while pYA3119, which contained a larger chromosomal DNA

Table 1. Phospho-β-galactosidase activity of *E. coli* containing various clones of the streptococcal *lacG* gene

Enzyme activity is expressed as Miller units [nmol ONP min⁻¹ (mg protein)⁻¹]. The values shown are means of three independent assays, ±SD.

Strain/plasmid	Enzyme activity
CC118	9.7 ± 1.5
CC118 pACYC184	11.7 ± 4.5
CC118 pYA3118	9.0 ± 3.5
CC118 pYA3119	26.7 ± 1.5
CC118 pBluescript KS+	11.0 ± 1.7
CC118 pYA3120	1149.3 ± 172.2

fragment, would contain the entire phospho-β-galactosidase gene.

To confirm the initial hypothesis about the structure of the lactose operon of *S. mutans* and to determine if the *S. mutans* phospho-β-galactosidase gene had been isolated, both plasmids were tested for the ability to confer to their *E. coli* host cells phospho-β-galactosidase enzyme activity. *E. coli* does not possess a native phospho-β-galactosidase activity. The results from the enzyme activity assays are shown in Table 1. As predicted, only cells containing pYA3119 exhibited phospho-β-galactosidase activity. These data confirmed that the phospho-β-galactosidase gene had been cloned and that our assumption about the location of the *lacG* gene in relation to the tagatose genes of *S. mutans* was correct. Subsequent restriction endonuclease analysis of pYA3119 revealed the presence of a 2.9 kb *Bam*HI–*Hind*III restriction fragment believed to contain the coding region for the lactose Enzyme II and phospho-β-galactosidase gene. This fragment was subcloned into pBluescript KS+ (Stratagene) and the resulting plasmid was designated pYA3120 (Fig. 1). This construct was also tested for the ability to confer phospho-β-galactosidase activity to the host cells. This construct displayed enzymic activity (Table 1), indicating that the entire coding region for the *S. mutans* phospho-β-galactosidase

RsaI 12 24 36 48 60
GTACAAAGAGGAGCTAAATAATGAAACATTGATTGCTCAAATGAAAAGGGAAA
M N T L I A Q I E K G K

72 84 96 108 120
GCCTTTCTTTGAAAAATTTCCGTAATATTATTACGTGCTATTCGTGACGGGTTTAT
P F F E K I S R N I Y L R A I R D G F I

132 144 156 168 180
CTGGCAATGCCGTCATCCTATTTTCGAGTATTTTCTGTGATCGCTTATGTTCTCTAA
S A M P V I L F S S I F L L I A Y V P N

192 204 216 228 240
TATCTTTGGTTTACTTGGCCAAAAGGTATTGAAACATGTTGATGACACCCCTATAACTA
I F G F T W P K G I E N M L M T P Y N Y

252 264 276 288 300
TACGATGGGATTATTGGTTTCTAGTGGCTGGTACAACTGCTAAATCACTAACGGATTC
T M G I I G F L V A G T T A K S L T D S

312 324 336 348 360
GATGAACCGCAATTAGAAAAGACCAACAGATTAAATTCATTCAACAATGTTAGCTTC
M N R Q L E K T N Q I N F I S T M L A S

372 384 BamHI 396 408 420
TATGGCCGGTTTCTTAATCATGGCAGCGGATCCAGCTAAAGAAGTGGTTTCTTGAAGTC
M A G F L I M A A D P A K E G G F L S A

432 444 456 468 480
CTTTATGGGAACTAAGGCGTTTAACTGCTTTATGTCAGCTTTTACTGTTAAATGT
F M G T K G L L T A F I A A F I T V N V

492 504 516 528 540
CTCAAGATTGTGTTAAAAATAAGTTACCATTCGTATGCTGAGGAGGTACCGCCAAA
Y K I C V K N N V T I R M P E E V P P N

552 564 576 588 600
TATCTCTCAAGTGTAAAGATATTTTCCATTGCGCTTTTCCATTATTTTGTATGC
I S Q V F K D I F P F A F S I I I L Y A

612 624 636 648 660
TATCCAATTAGCTATTAAAGCAGTTATTGGAGTAAATGTGGCTCAGTCAATTGGAACCTT
I Q L A I K A V I G V N V A Q S I G T L

672 684 696 708 720
GCTTGTCTCTTTTTCAGCAGCGGATGGCTATCTTGGTATTACGATTATTTTGGAGC
L A P L F S A A D G Y L G I T I I F G A

732 744 756 768 780
TTACGCTTGTCTGGTTTGTGCGTATTCTGCTCTTCTATTGTTGAGCCGGCTATTGC
Y A L F W F V G I H G P S I V E P A I A

792 804 816 828 840
AGCCATTACATATTCAAATGTTGAATTAATGCACATTTGATTCTGCTGGTCAACATGC
A I T Y S N V E L N A H L I H A G Q H A

852 864 876 888 900
TGATAAGTGATTACCTCAGGAACCTCAGATGTTATTGTAACCTATGGGTGGTACTGGTGC
D K V I T S G T Q M F I V T M G G T G A

912 924 936 948 960
AACATTAGTTGTTCTTTCATGTTCTGTTGTTATGTAATCAAAACGCAATAAGGCTAT
T L V V P F M F M W L C K S K R N K A I

972 984 996 1008 1020
TGGACGTGCTTCTGTTGTCGAACCTTCTTGGCGTCAATGAACCTATCTTGTGAGGC
G R A S V V P T F F G V N E P I L F G A

1032 1044 1056 1068 1080
ACCTATTGTTTAAATCCAGTCTTTTCATCCCTTTTCATTCTTGGCCCAATTGTTAATGT
P I V L N P V F F I P F I L A P I V N V

1092 1104 1116 1128 1140
TTGGATTTTTAAAGTTTGTGTTGATACCTTGGAAATGAACAGTTTCTTGTCTAATCTTCC
W I F K F F V D T L G M N S F F A N L P

1152 1164 1176 1188 1200
ATGGACAACACCGGGCAATAGGCATTGCTTAGGAACCTGATTGCTGTTTATCTTT
W T T P G P I G I V L G T G F A V L S F

1212 1224 1236 1248 1260
TGTCTTGCAGCACTTTTAATCCTTGTAGATAGTCTATTATTATCTTTTGTAAAGT
V L A A L L I L V D T V I Y Y P F V K V

1272 1284 1296 1308 1320
TTATGATGAACAGATTTTGGCTGAAGAAGCTGAAGTAAAGTTCATCTGATGCTTTAAA
Y D E Q I L A E E A E G K S S S D A L K

1332 1344 1356 1368 1380
AGAAAAAGTAGCAGCTAACTTTGATACGAAAAAGCTGATGCTATTCTTGAAGGTGCTGA
E K V A A N F D T K K A D A I L E G A E

1392 1404 1416 1428 1440
ATCTAAAGAGAGCCAGCAACTCATGCTATCACTGAAGAAACAAATGTTCTCGTTTATG
S K E E P A T H A I T E E T N V L V L C

1452 1464 1476 1488 1500
TGCCGGCGCGGTACCAGCGGTCTTTTAGCTAAATGCTTTAAATAAGCTGCAGAAGAATA
A G G G T S G L L A N A L N K A A E E Y

1512 1524 1536 1548 1560
CGGAGCCCCGTGTTAAGGCTGACGCTGGAAGTTATGAGAGCTCACCGCAAAATTTAGACCA
G A P V K A A A G S Y G A H R E I L D Q

1572 1584 1596 1608 1620
ATACCAATTAGTCATCTTGGCTCCGCAAGTGTCTTCTAATATGAGGATATGAAGGCAGA
Y Q L V I L A P Q V A S N Y E D M K A E

1632 1644 1656 1668 1680
AACGGATAAACTGGCATCAAATTAGCTAAGACGGAAGGTGCTCAATATATGTTGTGAC
T D K L G I K L A K T E G A Q Y I G L T

1692 1704 1716 1728 1740
GCGTGATGGCAAGGAGCGCTTGGCTTTGCTGGAAGAACAAATTTAAAGATTGAGAGGTCAC
R D G K G A L A F V E E Q F K D E

1752 1764 1776 1788 1800
ACTATGAGCAAAACATTACCTAAAGATTTTATCTTTGGTGGCAACGACGCTTATTCAG
M S K T L P K D F I F G G A T A A Y Q

1812 1824 1836 1848 1860
GCTGAAGGTGCCACTCATGCGATGGGAAGGGTCTGTTGCTTGGGATAAATATTTAGAA
A E G A T H A D G K G P V A W D K Y L E

1872 1884 1896 1908 1920
GATAACTATTGGTATACAGCTGAACCTGCCAGCGATTCTATCATCAATATCTGTTGAT
D N Y W Y T A E P A S D F Y H Q Y P V D

1932 1944 1956 1968 1980
TTAAAGTTGGCAGAAGAATTTGGCGTCAACGGTATTGCGATTTCATTGCTTGGTTCAGG
L K L A E E F G V N G I R I S I A W S R

1992 2004 2016 2028 2040
ATTTTCCAAAAGGGTATGGAGCCGTTAAACCCCAAGGCTTGGCTTTTATCATAATCTT
I F P K G Y G A V N P K G L A F Y H N L

2052 2064 2076 2088 2100
TTTGTCTGAATGCCATAAACGTCATGTGGAACCTTTTGAACACTTTCATCTTGTATACC
F A E C H K R H V E P F V T L H H F D T

2112 2124 2136 2148 2160
CCAGAGGCTTTTCATTCAAAATGGTGACTTCTCAATCGTGAAAATATAGAGCATTTTGTC
P E A L H S N G D F L N R E N I E H F V

2172 2184 2196 2208 2220
AACTATGCTGAATTTTGTCTTAAAGAAATTTTCAAGATGAATATTGGACAACTTCAAT
N Y A E F C F K E F S E V N Y W T T F N

2232 2244 2256 2268 2280
GAAATGGTCCCATTTGGTATGCTGAGTATCTGCTTGGAAATTTCCGCGGGCATCCAA
E I G P I G D G Q Y L V G K F P P G I Q

2292 2304 2316 2328 2340
TACGATCTTGGCAAGGTTTCCAATCCCACCAACATGATGTTGACATAGCAAGGCA
Y D L A K V F Q S H H N M M V A H S K A

2352 2364 2376 2388 2400
GTCAAACCTTTCAAAGATGGTGGCTACTCTGGCGAAATTTGGTGTGATGCTCTGCCA
V K L F K D G G Y S G E I G V V H A L P

2412 2424 2436 2448 2460
ACCAAAATCTCTTATGATCCAAACCAACCCAGCAGATATCGTGGCGCAGAATTAGAAGAT
T K Y P Y D P N N P A D I R A A E L E D

2472 2484 2496 2508 2520
ATTATTCATAATAAATTCATCTTGGATGGGACTTATCTGGGTAAGTATTCTGAGAAAAT
I I H N K F I L D A T Y L G K Y S E K T

2532 2544 2556 2568 2580
ATGGAAGGGTCAATCATATTTTGGCTGCAATGGTGGTCAATAGACCTTCTGAGGAA
M E G V N H I L A V N G G Q L D L R E E

2592 2604 2616 2628 2640
GATTTTGTCTCTTGAAGCTGCTAAAGATCTCAATGATTCTTGGGTATCAATATTAT
D F A A L E A A K D L N D F L G I N Y Y

```

2652      2664      2676      2688      2700
ATAAGTGGTGGATGCGTCTTTTGTGGTGAACCGAAATTACGCATAATGCTAAAGGC
I S D W M R A F D G E T E I T H N A K G

2712      2724      2736      2748      2760
GAAAAAGGAAGTTCTAAGTACCAAGATTAGGGTGTCCGTCGACAGAGCGCCAGTAAAT
E K G S S K Y Q I K G V G R R E A P V N

2772      2784      2796      2808      2820
GTTCCCAAAACAGATTGGGATTGGATCATTTATCCGCAAGGACTTTACGACCAATCATG
V P K T D W D W I I Y P Q G L Y D Q I M

2832      2844      2856      2868      2880
CGAGTTAAGCAGGACTATCCTAATTACAGAAGATTATATCACTGAAAAATGGTCTTGGT
R V K Q D Y P N Y K K I Y I T E N G L G

2892      2904      2916      2928      2940
TATAAGGATGAATTTGTGAATCATCTGTTTATGATGATGCCGATTGATTATGTTAA
Y K D E F V N H T V Y D D A R I D Y V K

2952      2964      2976      2988      3000
AAACACTTAGAAGTTCTTTTCTATAGCAGCGGTGCAATGTCAAGGTTACTTT
K H L E V L S D A I A D G A N V K G Y F

3012      3024      3036      3048      3060
ATCTGGTCTCTGATGGATGCTCTCTCATGTCAAATGGTTATGAGAAGCGCTATGGTCTC
I W S L M D V F S W S N G Y E K R Y G L

3072      3084      3096      3108      3120
TTTATGTTGATTTTGTATGACGAGCGTTATCCCAAGAAAGTCCCTACTGGTATAAA
F Y V D F D T Q E R Y P K K S A Y W Y K

3132      3144      3156      3168      3180
AAGTTAGCTGAGACCCCAATTAATGACTGATGCCGTTATTTAATCGCATTGGTAATCA
K L A E T Q I I D E

3192      3204      3216      3228 HindIII
TCTGCGTTGTTTGTGGGAGGCTTGCTGCTCTGACTGAATTAATCAAAAAGCTT

```

Fig. 2. Nucleotide sequence of the *S. mutans* *lacE* and *lacG* genes. The deduced amino acid sequences of the two proteins are also shown. The putative ribosome binding sites are underlined. The sequence of bases 1-138 was taken from Rosey & Stewart.

gene was contained on the 2.9 kb *Bam*HI–*Hind*III restriction fragment contained in this plasmid. The plasmid pYA3120 also displayed a much greater increase in enzymic activity over background levels when compared to plasmid pYA3119. We attribute this increase in enzymic activity to the difference in the plasmid copy number of the two different replicons (pACYC184 and pBluescript KS+).

The nucleotide sequence of *S. mutans* *lacE* and *lacG* genes

The nucleotide sequence of the 2.9 kb fragment contained in pYA3120 was determined. Nucleotide sequence analysis revealed the presence of two large open reading frames (ORFs). The deduced protein from the first ORF exhibited extensive homology to the LacE protein from several organisms (discussed below) but did not contain the amino-terminal end of the protein. To determine the nucleotide sequence of this region, an oligonucleotide was synthesized to use as a DNA sequencing primer with pYA3119 as the template DNA. This allowed the nucleotide sequence of the amino-terminal end of the

lacE gene, which is located 5' to the *Bam*HI site in pYA3120, to be determined. The nucleotide sequence of this region of the *lacE* gene has been previously reported by Rosey & Stewart (1992) from another *S. mutans* serotype *c* strain, PS14 (Murchison *et al.*, 1981). However, this strain is not transformable (Murchison *et al.*, 1986), which eliminates the possibility of using it to isolate a chromosomal DNA fragment adjacent to the restriction fragment contained in pYA501 using the technique described here. For this reason, we used the transformable strain UA130 (Murchison *et al.*, 1986) to isolate the *lacE* and *lacG* genes from *S. mutans*. In this report, the nucleotide sequence of 3235 bp is given and is shown in Fig. 2. This extends from a *Rsa*I site (bp 1–4) located in pYA3119 to the *Hind*III site (bp 3230–3235) found in pYA3120. The first 393 bp of this sequence were also reported by Rosey & Stewart (1992). We have also determined the nucleotide sequence of 250 bp within this region and our results are identical with those of Rosey & Stewart (1992). Our base number 1 corresponds to their base number 4748.

Nucleotide sequence analysis of the entire 3235 bp revealed the presence of two complete ORFs. The first ORF extends for 1704 bp (bases 26–1729) and would encode a protein consisting of 568 amino acids with a molecular mass of 61 473 Da. The second ORF extends for 1404 bp (bases 1744–3147) and would encode a protein consisting of 468 amino acids with a molecular mass of 53 758 Da. The deduced amino acid sequences of these proteins are shown in Fig. 2.

Immediately 5' to both ORFs, sequences which may function as ribosome binding sites may be found. These sequences, AAGAGG and AGAGG, resemble ribosome binding sites found in *Bacillus subtilis* (Moran *et al.*, 1982) and resemble putative ribosome binding sites from other streptococci (de Vos, 1987). These putative ribosome binding sites are underlined in Fig. 2. Because the sequence of the 3' end of the 16S ribosomal RNA from *S. mutans* has not been determined, the actual free energies of association between the streptococcal rRNA and the putative mRNA cannot be calculated.

Analysis of the nucleotide sequence for other putative transcription control elements did not reveal the presence of either putative promoter or terminator structures. The lack of any potential promoter sequence is consistent with the data of Rosey & Stewart (1992), which indicates that these genes appear to be part of a multicistronic operon controlled by DNA sequences 5' to those reported here. The lack of an apparent transcriptional terminator 3' to the *lacG* gene may be due to the paucity of information on streptococcal terminators or the fact that the *S. mutans* lactose operon may contain an additional ORF 3' to those reported here, as was reported for the *L. lactis* *lac* operon which contains an

[illegible]

additional ORF 3' to the *lacE* and *lacG* genes (de Vos *et al.*, 1990). The function for this gene, *lacX*, has not been determined. Analysis of the nucleotide sequence of the 85 bp 3' to the *S. mutans lacG* gene does not reveal the presence of any potential ribosome binding site which would precede an additional ORF.

The S. mutans lac gene products show extensive homology with other lactose enzymes

To determine the potential gene function of the two *S. mutans* ORFs, the deduced polypeptides were screened for homology with the translated products of all GenBank entries using the TFASTA algorithm (Pearson & Lipman, 1988). The deduced protein from the first ORF exhibited significant homology to the LacE proteins from *L. lactis* (de Vos *et al.*, 1990), *Staph. aureus* (Breidt *et al.*, 1987) and *L. casei* (Alpert & Chassy, 1990). The *S. mutans* LacE protein exhibited 80%, 76% and 48% identity, respectively, with these proteins. When conserved amino acid substitutions are considered, the homology values between these proteins are 97%, 96% and 82%, respectively. These homology alignments are shown in Fig. 3. Based upon the extensive homology with these PTS lactose Enzyme II molecules, we conclude that this ORF encodes the *S. mutans* lactose Enzyme II and that this locus should be designated *lacE*.

One local area of high homology between these proteins is the region surrounding cysteine residue 472 in the *S. mutans* protein. The homologous residue in the LacE protein of *L. casei* has been shown to be the site of phosphorylation during the vectorial transport of lactose of the PTS (Alpert & Chassy, 1990). Mutants which have had this residue changed to serine by *in vitro* mutagenesis do not phosphorylate lactose. Therefore, based upon the high conservation of residues in this region, we propose that this residue is the site of phosphorylation in the *S. mutans* LacE protein.

Although a high degree of homology is conserved around this cysteine residue in all of the LacE proteins, it is interesting to note that the *L. casei* protein appears to contain a 14 residue insertion adjacent to this site. This apparent insertion would involve residues 462–475 of the *L. casei* protein and is depicted in Fig. 3. This region was identified because the TFASTA alignment algorithm determined that two possible alignments existed between the *S. mutans* and *L. casei* LacE proteins. One alignment showed homology between the amino acid residues on the amino-terminal side of this apparent

insertion while the other alignment displayed homology between the residues on the carboxyl-terminal side of the insertion. To determine if this apparent 14 amino acid residue insertion was part of an insertion element, a homology analysis was conducted on the insert alone. Neither the amino acid sequence of this apparent insertion nor its DNA coding sequence exhibits significant homology to any GenBank entry.

The hydropathy profile of the deduced *S. mutans* LacE protein indicates that the amino-terminal 70% of the protein is highly hydrophobic while the carboxyl-terminal region is more hydrophilic. This implies that the amino-terminal region of the molecule is involved with the penetration of and transport of lactose across the cell membrane. This hydropathy profile is similar to those previously reported for other Enzyme II^{lac} proteins (Alpert & Chassy, 1990; Breidt *et al.*, 1987; de Vos *et al.*, 1990).

The protein sequence deduced from the second ORF exhibited homology with the phospho- β -galactosidases from *L. lactis* (de Vos & Gasson, 1989), *Staph. aureus* (Breidt & Stewart, 1987) and *L. casei* (Porter & Chassy, 1988). These proteins, the *lacG* gene products, had 89%, 82% and 53% identity, respectively, with the deduced protein from *S. mutans*. These homology values increased to 99%, 97% and 85%, respectively, when conserved amino acids were considered. These alignments are shown in Fig. 4. Thus, we conclude that this ORF is the *lacG* locus of *S. mutans*. This conclusion is consistent with the phospho- β -galactosidase activity associated with the plasmids pYA3119 and pYA3120.

The high degree of homology between the *S. mutans lacE* and *lacG* proteins and those of *L. lactis*, *Staph. aureus* and *L. casei* support the hypothesis that the *lac* genes may have been transferred between organisms as a modular genetic element (de Vos *et al.*, 1990). The *lac* operon in *L. lactis* is flanked by a sequence which has homology to an insertion sequence element which may have served to facilitate this type of transfer. Our homology results appear to indicate that the lactose genes of *S. mutans*, *L. lactis* and *Staph. aureus* are more closely related on the evolutionary scale, while those of *L. casei* are more distantly related. The homologies described here between the *lac* proteins are distributed evenly throughout the length of these molecules. These results are consistent with the hypotheses that these organisms recently acquired these genes from an ancestral donor or that evolutionary divergence of these loci has been restricted in these organisms.

Fig. 3. Homology alignments between the LacE proteins of *S. mutans*, *L. lactis*, *Staph. aureus* and *L. casei*. The character · indicates identical residues, + indicates conservative substitutions, and letters indicate non-conservative amino acid substitutions which have replaced the corresponding *S. mutans* residue.

The overall operon structure of the S. mutans lactose operon is similar to that of Staph. aureus and L. lactis

Our data, combined with those of Rosey & Stewart (1992), show that an apparent operon structure for the *lacRABCD FEG* genes exists in *S. mutans*. This organization is identical to the operon structure of the *lac* genes from *Staph. aureus* (Rosey *et al.*, 1991). The *lac* operon of *L. lactis* is also organized in the same way except that transcription of the *lacR* gene is in the opposite direction. These results, the high DNA homology and the overall structure of the *lac* operons, add support to the hypothesis that the *lac* genes may have been transferred between organisms as a modular genetic element. Initial reports indicate that a similar operon structure may also exist in *L. casei* but the exact function of the ORFs 5' to the *lacE* and *lacG* genes have not been determined. Also, it has not been determined if the *S. mutans* or *Staph. aureus lac* operons contain a *lacX* gene similar to that found in *L. lactis* (de Vos *et al.*, 1990).

In this report, we have described the isolation, characterization and nucleotide sequence of the *lacE* and *lacG* genes of *S. mutans*. Based upon homology with other *lac* systems, the LacE protein is presumed to be the permease component of the lactose PTS responsible for the vectorial transport of lactose. The LacG protein is the phospho- β -galactosidase responsible for the cleavage of the lactose phosphate generated by the lactose PTS. Further studies are being initiated to understand the regulation of these loci.

We thank Evert Rosey and George Stewart for communication of data prior to publication. This work was supported by National Research Service Award F32-DE05518 to A.L.H. and by National Institutes of Health grant DE06673.

References

- ALPERT, C.-A. & CHASSY, B. M. (1988). Molecular cloning and nucleotide sequence of the factor III^{lac} gene of *Lactobacillus casei*. *Gene* **62**, 277–288.
- ALPERT, C.-A. & CHASSY, B. M. (1990). Molecular cloning and DNA sequence of *lacE*, the gene encoding the lactose-specific Enzyme II of the phosphotransferase system of *Lactobacillus casei*. *Journal of Biological Chemistry* **265**, 22561–22568.
- BISSET, D. & ANDERSON, R. L. (1973). Lactose and D-galactose metabolism in *Staphylococcus aureus*: pathway of D-galactose 6-phosphate degradation. *Biochemical and Biophysical Research Communications* **52**, 641–647.
- BISSET, D. & ANDERSON, R. L. (1974). Genetic evidence for the physiological significance of the D-tagatose 6-phosphate pathway of lactose and D-galactose degradation in *Staphylococcus aureus*. *Journal of Bacteriology* **119**, 698–704.
- BREIDT, F., JR & STEWART, G. C. (1987). Nucleotide and deduced amino acid sequences of the *Staphylococcus aureus* phospho-beta-galactosidase gene. *Applied and Environmental Microbiology* **53**, 969–973.
- BREIDT, F., JR, HENGSTENBERG, W., FINKELDEI, U. & STEWART, G. C. (1987). Identification of the genes for the lactose-specific components of the phosphotransferase system in the *lac* operon of *Staphylococcus aureus*. *Journal of Biological Chemistry* **262**, 16444–16449.
- CHEN, E. J. & SEEBURG, P. H. (1985). Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA* **4**, 165–170.
- DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.
- DE VOS, W. M. (1987). Gene cloning and expression in lactic streptococci. *FEMS Microbiological Reviews* **46**, 281–295.
- DE VOS, W. M. & GASSON, M. J. (1989). Structure and expression of the *Lactococcus lactis* gene for phospho-beta-galactosidase (*lacG*) in *Escherichia coli* and *L. lactis*. *Journal of General Microbiology* **135**, 1833–1846.
- DE VOS, W. M., BOERRIGTER, I., VAN ROOYEN, R. J., REIDHE, B. & HENGSTENBERG, W. (1990). Characterization of the lactose-specific enzymes of the phosphotransferase system in *Lactococcus lactis*. *Journal of Biological Chemistry* **265**, 22554–22560.
- DRUCKER, D. B. & MELVILLE, T. H. (1968). Fermentation end-products of cariogenic and non-cariogenic streptococci. *Archives of Oral Biology* **13**, 565–570.
- FROSTELL, G., KEYES, P. H. & LARSON, R. H. (1967). Effect of various sugars and sugar substitutes on dental caries in hamsters and rats. *Journal of Nutrition* **93**, 65–76.
- GUGGENHEIM, B., KONIG, K. G., HERZOG, E. & MUHLEMANN, H. R. (1966). The cariogenicity of different dietary carbohydrates tested on rats in relative gnotobiosis with a streptococcus producing extracellular polysaccharide. *Helvetica Odontologica Acta* **10**, 101–113.
- HAMADA, S. & SLADE, H. D. (1980). Biology, immunology, and cariogenicity of *Streptococcus mutans*. *Microbiological Reviews* **44**, 331–384.
- HAMILTON, I. R. & LEBTAG, H. (1979). Lactose metabolism by *Streptococcus mutans*: evidence for induction of the tagatose 6-phosphate pathway. *Journal of Bacteriology* **140**, 1102–1104.
- HENIKOFF, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–359.
- JAGUSZTYN-KRYNICKA, E. K., HANSEN, J. B., CROW, V. L., THOMAS, T. D., HONEYMAN, A. L. & CURTISS, R., III (1992). *Streptococcus mutans* serotype c tagatose 6-phosphate gene cluster. *Journal of Bacteriology* **174**, 6152–6158.
- MACRINA, F. L., EVANS, R. P., TOBIAN, J. A., HARTLEY, D. L., CLEWELL, D. B. & JONES, K. R. (1983). Novel shuttle plasmid vehicles for *Escherichia-Streptococcus* transgeneric cloning. *Gene* **25**, 145–150.
- MANOIL, C. & BECKWITH, J. (1985). *TnphoA*: a transposon probe for protein export signals. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 8129–8133.
- MAXWELL, E. S., KURAHASHI, K. & KALCKAR, H. M. (1962). Enzymes of the Leloir pathway. *Methods in Enzymology* **5**, 174–189.
- MILLER, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- MORAN, C. P., JR, LANG, N., LEGRICE, S., LEE, G., STEPHENS, M., SONENSHEIN, A. L., PERO, J. & LOSICK, R. (1982). Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Molecular and General Genetics* **186**, 339–346.
- MURCHISON, H., LARRIMORE, S. & CURTISS, R., III (1981). Isolation and characterization of *Streptococcus mutans* mutants defective in adherence and aggregation. *Infection and Immunity* **34**, 1044–1055.
- MURCHISON, H. H., BARRETT, J. F., CARDINEAU, G. A. & CURTISS, R., III (1986). Transformation of *Streptococcus mutans* with chromosomal and shuttle plasmid (pYA629 DNAs. *Infection and Immunity* **54**, 273–282.
- PEARSON, W. R. & LIPMAN, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 2444–2448.
- PORTER, E. V. & CHASSY, B. M. (1988). Nucleotide sequence of the beta-D-phosphogalactoside galactohydrolase gene of *Lactobacillus casei*: comparison to analogous *pbg* genes of other Gram-positive organisms. *Gene* **62**, 263–276.
- POSTMA, P. W. & LENGELER, J. W. (1985). Phosphoenolpyruvate:carbohydrate phosphotransferase system of bacteria. *Microbiological Reviews* **49**, 232–269.
- ROSEY, E. L. & STEWART, G. C. (1992). Nucleotide and deduced amino

- acid sequences of the *lacR*, *lacABCD*, and *lacFE* genes encoding the repressor, tagatose 6-phosphate gene cluster, and sugar-specific PTS components of the lactose operon of *Streptococcus mutans*. *Journal of Bacteriology* **174**, 6159–6170.
- ROSEY, E. L., OSKOUIAN, B. & STEWART, G. C. (1991). Lactose metabolism by *Staphylococcus aureus*: characterization of *lacABCD*, the structural genes of the tagatose 6-phosphate pathway. *Journal of Bacteriology* **173**, 5992–5998.
- SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- SMORAWINSKA, M., HSU, J. C., HANSEN, J. B., JAGUSZTYN-KRYNICKA, E. K., ABIKO, Y. & CURTISS, R., III (1983). Clustered genes for galactose metabolism from *Streptococcus mutans* cloned in *Escherichia coli*. *Journal of Bacteriology* **153**, 1095–1097.
- SOUTHERN, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503–517.
- TABOR, S. & RICHARDSON, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 4767–4771.
- VAN ROOIJEN, R. J., VAN SCHALKWIJK, S. & DE VOS, W. M. (1991). Molecular cloning, characterization, and nucleotide sequence of the tagatose 6-phosphate pathway gene cluster of the lactose operon of *Lactococcus lactis*. *Journal of Biological Chemistry* **266**, 7176–7181.