Nucleotide and Deduced Amino Acid Sequences of the *lacR*, *lacABCD*, and *lacFE* Genes Encoding the Repressor, Tagatose 6-Phosphate Gene Cluster, and Sugar-Specific Phosphotransferase System Components of the Lactose Operon of *Streptococcus mutans*

EVERETT L. ROSEY[†] AND GEORGE C. STEWART^{‡*}

Department of Microbiology, University of Kansas, Lawrence, Kansas 66045

Received 30 March 1992/Accepted 20 July 1992

The complete nucleotide sequences of *lacRABCDF* and partial nucleotide sequence of *lacE* from the lactose operon of *Streptococcus mutans* are presented. Comparison of the streptococcal *lac* determinants with those of *Staphylococcus aureus* and *Lactococcus lactis* indicate exceptional protein and nucleotide identity. The deduced polypeptides also demonstrate significant, but lower, sequence similarity with the corresponding lactose proteins of *Lactobacillus casei*. Additionally, LacR has sequence homology with the repressor (DeoR) of the *Escherichia coli* deoxyribonucleotide operon, while LacC is similar to phosphokinases (FruK and PfkB) from *E. coli*. The primary translation products of the *lacRABCDFE* genes are polypeptides of 251 (M_r 28,713), 142 (M_r 15,610), 171 (M_r 18,950), 310 (M_r 33,368), 325 (M_r 36,495), 104 (M_r 11,401), and 123 (NH₂-terminal) amino acids, respectively. As inferred from their direct homology to the staphylococcal *lac* genes, these determinants would encode the repressor of the streptococcal lactose operon (LacR), galactose-6-phosphate isomerase (LacA and LacB), tagatose-6-phosphate kinase (LacC), tagatose-1,6-bisphosphate aldolase (LacD), and the sugar-specific components enzyme III-lactose (LacF) and enzyme III-lactose (LacE) of the *S. mutans* phosphoenolpyruvate-dependent phosphotransferase system. The nucleotide sequence encompassing the *S. mutans lac* promoter appears to contain repeat elements analogous to those of *S. aureus*, suggesting that repression and catabolite repression of the lactose operons may be similar in these organisms.

Streptococcus mutans is known to attach to and colonize the tooth pellicle and peridontium of humans, where the utilization of dietary sucrose leads to the formation of dental plaque (20). The fermentative metabolism of these and other oral microorganisms yields lactic acid, which acts directly and indirectly upon the teeth and peridontium, ultimately causing caries and peridontal disease (12). The oral streptococci are also cariogenic in animal models when the diet contains acidogenic carbohydrates other than sucrose. These carbohydrates include lactose and starch, which are major constituents of the human diet, as well as less prevalent carbohydrates such as glucose, fructose, and maltose (18). Since lactose is present in high concentrations in bovine milk and is generally consumed by humans in large quantities throughout life, or at least during the preadolescent or "caries-prone" years, it can be considered a dietary carbohydrate of significant importance in cariogenesis. If the virulence of S. mutans is dependent upon its metabolic potential (12), it appears that a thorough understanding of not only sucrose but also lactose metabolism is essential before effective measures to eliminate dental caries in humans can be designed.

6159

Lactose catabolism by several oral streptococcal strains has been shown to proceed via two mechanistically distinct pathways. For some isolates, such as Streptococcus salivarius 25975 (22), the metabolism of lactose appears to proceed mainly by hydrolysis of the disaccharide via β-galactosidase to glucose and galactose (although significant lactose-phosphotransferase system [PTS] activity is induced upon growth on lactose). The latter hexose is then converted to glucose 6-phosphate via the Leloir pathway (29): D-galactose \rightarrow D-galactose 1-phosphate \rightarrow D-glucose 1phosphate \rightarrow D-glucose 6-phosphate. In S. mutans, low or negligible levels of β-galactosidase activity are detectable (22). In this organism, galactose is phosphorylated during vectorial transport of galactose or lactose by the phosphoenolpyruvate-dependent PTS (13) and is then metabolized via the tagatose 6-phosphate pathway as described for Staphylococcus aureus (5) and later demonstrated for S. mutans (21): D-galactose 6-phosphate \rightarrow D-tagatose 6-phosphate \rightarrow D-tagatose 1,6-bisphosphate \rightarrow D-glyceraldehyde 3-phosphate plus dihydroxyacetone phosphate. The enzymatic activities effecting the degradation of galactose 6phosphate are galactose-6-phosphate isomerase (9), tagatose-6-phosphate kinase (7), and tagatose-1,6-bisphosphate aldolase (8).

We recently reported the molecular organization of the *lacABCD* genes encoding these enzymes in *S. aureus* (35). These genes are part of a heptacistronic operon, *lacAB CDFEG*, in which the terminal three genes, *lacFEG*, encode the sugar-specific transport components enzyme III-lactose (EIII^{Lac}; EIIA^{Lac}_{Sa} as proposed elsewhere [37]) and

^{*} Corresponding author.

[†] Present address: Molecular Biology Research, The Upjohn Co., Kalamazoo, MI 49001.

[‡] Present address: Department of Microbiology and Immunology, School of Medicine, University of South Carolina, Columbia, SC 29208.



FIG. 1. DNA hybridization with pYA501. (A) Chromosomal DNA from S. mutans (lanes 1 to 3) or S. aureus (lanes 4 to 6) was probed with radiolabeled pYA501 under conditions of low stringency. The hybridization probe was the gel-purified 5-kb BamHI insert from radiolabeled pYA501. (B) Hybridization of the probe to either pYA501 (lane 7) or pBO4 (lane 8) under conditions of high stringency. DNAs in each lane were restricted as follows: 1, BamHI; 2, HindIII; 3 and 5, PstI; 4, EcoRI; 6, EcoRI plus PstI; 7, BamHI plus HindIII; 8, EcoRI plus PstI. Phage λ DNA digested with HindIII gave the size standards. Sizes are shown in kilobases.

EII^{Lac} (EIICB_{Sa}^{Lac} [37]) of the phosphoenolpyruvate-dependent PTS as well as phospho- β -galactosidase (10, 11) and are coordinately induced with the tagatose 6-phosphate gene cluster during growth on galactose or lactose. The isolation of mutants deficient in these enzymes supports the conclusion that the tagatose 6-phosphate pathway is the sole route of lactose and D-galactose metabolism in *S. aureus* (6).

Throughout the course of defining the S. aureus tagatose 6-phosphate genes, we were able to partially complement Escherichia coli fda or pfk mutants by using a plasmid, pYA501, which contains a 5-kb chromosomal fragment from S. mutans PS14 (serotype c). This plasmid was previously shown to encode enzymes of the tagatose 6-phosphate pathway (43) and encodes proteins very similar in size to those encoded by the staphylococcal lacABCD genes (24, 35, 36). Because of the similarities in lactose metabolism between these organisms, we have determined the molecular organization of the genes specified by the S. mutans insert

TABLE	E 1. (Comparise	on of Dl	NA and	deduced	amino	acid
sequences	of th	e lactose	operons	from S.	aureus	and S.	mutans

T	%]	DNA identity ^a	% Protein homology				
Locus	Overall	Excluding wobble	Identical	Conservative			
lacR	65.7	77.9	63	75			
lacA	74.2	87.3	76	87			
lacB	74.3	86.0	81	85			
lacC	64.4	75.0	63	71			
lacD	71.3	82.3	72	79			
lacF	72.8	87.4	75	85			
lacE ^b	73.9	88.6	83	87			
Avg	70.9	83.5	73	81			

^a Intercistronic regions were excluded from analysis.

^b Values represent similarities for the NH₂-terminal 369 nucleotides (123 amino acids) only.

present in pYA501. In this report, we show that the streptococcal DNA encodes enzymes highly homologous to the staphylococcal tagatose 6-phosphate pathway enzymes and also that the lactose transport components EIII^{Lac} ($\text{EIIA}^{\text{Lac}}_{\text{Sm}}$) and EII^{Lac} as well as the lac repressor (LacR) are homologous in *S. aureus* and *S. mutans*.

MATERIALS AND METHODS

Bacterial strains, media, and reagents. E. coli DH5 α (3) and JM83r⁻ [F⁻ ara Δ (lac-proAB) rpsL strA thi ϕ 80d lacZ Δ M15 hsdR4 zjj-202::Tn10] (laboratory strain) were cultivated in L broth containing ampicillin (100 µg/ml) or chloramphenicol (10 µg/ml). S. aureus RN4220 (25) was grown in tryptic soy broth (Difco Laboratories), and S. mutans PS14 (serotype c) (43) was grown anaerobically (Gas Pak; Difco) in brain heart infusion (BHI) broth.

Other materials were obtained from the following sources: antibiotics, lysozyme, lysostaphin, mutanolysin, and proteinase K were from Sigma Chemical Co., St. Louis, Mo.; restriction endonucleases, the Klenow fragment of DNA polymerase I, and T4 DNA ligase were from Promega Biotec and Pharmacia, Inc.; $[\alpha^{-32}P]$ dATP was from Du Pont NEN Research Products.

Chromosomal DNA isolation and hybridization. S. aureus chromosomal DNA was isolated by the method of Dyer and Iandolo (17). Chromosomal DNA was isolated from S. mutans PS14 by the following procedure. A 3-ml overnight starter culture was transferred into 200 ml of fresh BHI broth and grown anaerobically overnight. Cells from a 100-ml aliquot were harvested and washed in 5 ml of TE buffer (10 mM Tris-HCl, 1 mM EDTA [pH 8.0]), suspended in 3 ml of TE buffer, and heated at 65°C for 20 min. After the suspension was cooled briefly on ice, 2 ml of TE buffer containing lysozyme (50 mg/ml), mutanolysin (250 U/ml), and RNase A (0.5 mg/ml) was added, and the sample was incubated at 37°C for 60 min. Proteinase K was then added to 50 µg/ml, and incubation at 37°C was continued for an additional 60 min. The resulting cell lysate was then treated as described for S. aureus (17)

DNA fragments generated by endonuclease digestion of total cellular DNAs or appropriate plasmid controls were electrophoresed in 0.8% agarose and transferred to nitrocellulose paper by the method of Smith and Summers (42). Radiolabeled probe DNA was prepared with $[\alpha^{-32}P]$ dATP by nick translation (28). The 5.1-kb BamHI fragment from



FIG. 2. Physical map and nucleotide sequencing strategy for the S. mutans lactose operon. (A) Physical organization of the S. aureus lac operon (3:5 scale) is shown for comparison. (B) The extent and direction of transcription of the S. mutans lacR and lacABCDFE genes contained in pYA501 are indicated by bold arrows, and partial ORFs are indicated by broken arrows. Predicted molecular mass (in kilodaltons) is indicated below each ORF. (C) The subclones generated for sequencing are indicated by arrows, representing the start point, direction, and extent of each sequencing reaction. Putative promoter regions (*) and sequencing reactions primed by synthetic oligonucleotides (\bullet) are indicated. Abbreviations for restriction endonuclease sites: A, AseI; Av, AvaI; B, BamHI; Ba, BalI; Bc, BcII; Bg, BgIII; C, ClaI; E, EcoRI; H, HindIII; Hp, HpaI; P, PstI; Pv, PvuII; S, SalI; Sp, SphI; X, XbaI.

radiolabeled pYA501 was extracted by boiling from 0.8% low-gelling-temperature agarose. Hybridizations were carried out at 65°C for 16 to 20 h. Posthybridization washes (20 min, five times) were done in $0.1 \times$ SSC (15 mM NaCl, 1.5 mM sodium citrate [pH 7.0])–0.05% sodium dodecyl sulfate (SDS) at either 55°C (high stringency) or 37°C (low stringency). The filters were dried (37°C), and radioactive DNA

fragments were visualized by exposure to X-ray (Du Pont Cronex-4) film.

Determination of nucleotide sequence. Plasmid pMK4 (44) and its derivatives, pER2058 and pER2108, were used as vectors for the subcloning of streptococcal DNA restriction fragments from pYA501 (43). Plasmid pER2058 contains a unique *Hind*III site within the polylinker region as a conse-

Comparison	Test	% Amino acid identity									
standard	sequence	LacR	LacA	LacB	LacC	LacD	LacF	LacE ^a			
S. aureus	S. mutans L. lactis L. casei	63 44 ^b	76 70 —	81 85 —	63 61 —	72 73	75 70 47	83 72 51			
S. mutans	S. aureus L. lactis L. casei	63 35 —	76 74	81 77 —	63 61 —	72 78	75 67 43	83 76 55			
L. lactis	L. casei		_	_	_	_	48	56			

TABLE 2. Comparison of deduced amino acid sequences of the lactose proteins from S. aureus, S. mutans, L. lactis, and L. casei

" Values represent similarities for the NH2-terminal 123 residues only.

^b —, corresponding proteins from *L. casei* have not been reported.

Drat

.

GGATCCTGCAAGCGGAGAAGACTATGTCGTGCGGGGGGATTATTGACGGTTACATTCTTTT	60	CAACAC
D P A S G E D Y V V R G I I D G Y I L F	20	chrichi
• Mnll• • Alul•		TOTOT
TGATAATUGGATIGI <u>CCTC</u> TTIGATTATAAAACAGATAAATICACTAATAGTUA <u>AGCT</u> AT	120	10101
DNRIVLFDYKTDKFTNSQAI	40	
• Mnli • Alui •		TCTGAC
TAAAGAGCGTTACCGAGGTCAGATGACACTTTATGCTCAAGCTTTGAGCCAATCCTATAA	180	
K E R Y R G Q M T L Y A Q A L S Q S Y N	60	
		PVUII
	240	GCTGGT
	240	<u> </u>
TOGADSILIPPOGEKPEAAF	80	A 0
• • • • • •		
AATTTAATAGCAAACATATTCCAAAAAAGAGTAGTCTAATGAAAAGAACCACTTATAATG	300	TTAGAT
I # *	81	LD
• • • Mnli• • •		Daa
GACAGTATAAAAAAGTTATGCACTGTTTTTATAAG <u>GAGG</u> TTCTTTTCATAAATACTCTTT	360	CAACTA
		E V
· DraI · _ · TagI·		E V
TTTATTTATGAAAAGAACTGTTA <u>TTTAAA</u> CGTAATGTTTGTTTTAIGT <u>TTGAATATCGA</u> A	420	
-35		TTTATC
MnlI		FM
	400	
AAAAATTTTAAAAAATATTATATAATAGAAGAAAAGAAGGAAAAGAGGTATTTATGAAGAA	480	AluI
-10 <u>M K K</u>	3	GCTTAT
Lack Lack		AY
GAAGAAGGAGGAGGAGAAAAAGGAAAAAAAAAAAAAAAA	540	GGAGAT
	22	G D
EERDEEIIKDIWKKGIIKVI	25	3 0
• • • • • • • • • • • • • • • • • • •		
GAAGTTGTGGAACGATTAAAGGTATCAGACATGACTGTCCGT <u>CGCG</u> ATTTAACTGAATTG	600	<u>CG</u> TCAT
E V V E R L K V S D M T V R R D L T E L	43	RH
		AATCAA
GAAGGCTTGGGAGTATTAACACGTATTCACGGTGGGGCTAGAAGCAAT <u>AATATT</u> TTTCAA	000	1012010
EGLGVLTRIHGGARSNNIFQ	63	
• • • <i>Tar</i> ī • • •		
TATAAAGAAATGTCCCATGAAGAAAAGCATTCTCGACAAATAGAAGAAAAGCACTATATT	720	
Y K E M S H E E K H S R O I E E K H Y I	83	ATCTAA
		SK
What Amatt		
• • • MDOI AVAII		
GCACAAAAGGCTGCCGAATTAGTAGAAGAAGGAGATAC <u>GATC</u> TTTTTA <u>GGACC</u> AGGAACA	780	Rsal
GCACAAAAGGCTGCCGAATTAGTAGAAGAAGAAGAAGAAGAAGAAGAATTTTTTAGGCCGGGAACA A Q K A A E L V E E G D T I F L G P G T	780 103	<i>Rsa</i> I AC <u>GTAC</u>
GCACAAAAGGCTGCCGAATTAGTAGAAGAAGAAGAAGAAGAAGAAGAAGAATCTTTTTAGGACCAGGAACA A Q K A A E L V E E G D T I F L G P G T	780 103	<i>Rsa</i> I AC <u>GTAC</u> R T
GCACAAAAGGCTGCCGAATTAGTAGAAGAAGAAGAAGAAGAAGAATTTTTTTGGACCAGGAACAAA Q K A A E L V E E G D T I F L G P G T ASEI A ASEI ASEI ASEI ASEI ASEI ASEI ASEI ASEI	780 103 840	<i>Rs</i> ai AC <u>GTAC</u> R I
GCACAAAAGGCTGCCGAATTAGTAGAAGAAGAAGAAGAAGAAGAAGAAGAACTTTTTTGGACCAGGAACA A Q K A A E L V E E G D T I F L G P G T ActgragaactgctggctggaagaaAttaataagacaacttttgcaagtcatcaccaattgt	780 103 840	<i>Rsa</i> I AC <u>GTAC</u> R T
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123	<i>Rsa</i> I AC <u>GTAC</u> R T GGATTT
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123	<i>Rsa</i> I AC <u>GTAC</u> R T GGATTT D L
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900	RSAI AC <u>GTAC</u> R T GGATTT D L HDAII
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143	Rsal AC <u>GTAC</u> R T GGATTT D L Hpail ACCGGG
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143	RSAI AC <u>GTAC</u> R T GGATTT D L Hpail <u>ACCGGG</u> P G
$\begin{array}{c} GCACAAAAGGCTGCCGAATTAGTAGAAGAAGAAGAAGAAGAAGAAGAATTTTTTT$	780 103 840 123 900 143	Rsal AC <u>GTAC</u> R T GGATTT D L Hpall <u>ACCGG</u> P G
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960	Rsal AC <u>GTAC</u> R T GGATTT D L Hpall <u>ACCGG</u> P G
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163	Rsai AC <u>GTAC</u> R T GGATTT D L Hpail ACCGGG P G ACTTAA
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163	Rsai AC <u>GTAC</u> R I GGATTT D L Hpaii ACCCGG P G ACTTAA L N
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020	Rsai AC <u>GTAC</u> R T GGATTT D L Hpaii ACCGGG P G ACTTAA L N
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183	Real AC <u>GTAC</u> R T GGATTI D L Hoali ACG3GG P G ACTTAA L N TATTGT
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183	Rsai ACGTAC R I GGATTT D L Hpail ACCGGG P G ACTTAA L N TATTGT
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183	RsaI ACGTAC R I GGATTI D L HpaII ACCGGG P G ACTTAA L N TATTGT I V
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080	RSAI ACGTAC R I GGATTI D L Hpail ACCGGG P G ACTTAA L N TATTGT I V Ddel
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203	RSAI ACGTAC R T D L HDAII ACCGGG P G ACTTAA L N TATTGT I N DdeI TGCTAA
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203	RSAI ACCTAC R T D L Hpail ACCGGG P G ACTTAA L N TATTGT I V DdeI TG <u>CTAA</u> A K
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140	RSAI ACCTAC R I GGATTI D L Hpail ACCGGG P G ACTTAA L N TATTGT I V DdeI TGCTAA A K
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223	RSAI ACGTAC R T D L HDAII ACCCCCC P G ACTTAA L N TATTGT I V DdeI TGCTAA A K
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223	RSAI ACGTAC R T D L Hpail ACCGGG P G ACTTAA L N TATTGT I V Ddel TGCTAA A K TGAATT
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223	RSEI ACCTAC R T GGATTI D L Hpail ACCGGG P G ACTTAA L N TATTGT I V DdeI TGCTAA A K TGATTT E F
$\begin{array}{c} \begin{array}{c} \begin{array}{c} GCACAAAAGGCTGCCGGAATTAGTAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223 1140	RSAI ACGTAC R T D L Hpail ACCOGG P G ACTTAA L N TATTGT I V DdeI TGCTAA A K TGAATT E F
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223 1200 243	RSAI ACGTAC R T D L Hpail ACCGGG P G ACTTAA L N TATTGT I V DdeI TGCTAA A K TGAATT E F
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223 1200 243	RSAI ACGTAC R T D L HDAII D L HDAII D L HDAII D L HDAII P G P G P G P G P G P G P G P G P G P G
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223 1200 243 1260	RSAI ACCTAC R T D L Hpail ACCCGG P G ACTTAA L N TATTGT I V Ddei TG <u>CTAA</u> A K TGAATT E F TATGAT M M
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223 1200 243 1260 251	RSEI ACCTAC R T D L Hpail ACCOGO P G ACTTAA L N TATTOT I V Ddei TGCTAA A K TGAATT E F TATGAAT
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223 1200 243 1260 251	RSAI ACGTAC R T D L HDAII ACCOGG P G ACTTAA L N TATTGT I V DdeI TGCTAA A K TGAATT E F TATGAT M M
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223 1200 243 1260 251	RSAL ACCTAC R T D L Hpail ACCGGG P G ACTTAA L N TATTGT I V Ddel TGCTAA A K TGAATT E F TATGAT M M M
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223 1200 243 1260 251 1320	RSAI ACGTAC R T D L HDAII D L HDAII D L HDAII D L HDAII P G P G P G P G P G P G P G P G P G P G
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223 1200 243 1260 251 1320	RSAI ACGTAC R T D L Hpail ACCCGG P G ACTTAA L N TATTGT I V Ddei TGCTAA A K TGAATT E F TATGAT M M AAGGT K V
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223 1200 243 1260 251 1320	RSEI ACGTAC R T D L Hpail ACCOGO P G ACTTAA L N TATTGT I V DdeI TGCTAA A K TGAATT E F TATGAT T E F TATGAT K V
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	780 103 840 123 900 143 960 163 1020 183 1080 203 1140 223 1200 243 1260 251 1320	RSAI ACGTAC R T D L HDAII ACCCCCC P G ACTTAA L N TATTGT I V DOGI TGCTAA A K TGAATT E F TATGAT M M Lac AAAGGT K V GAATGT

CAAG	CAG	AAA	AAG	GAA.	ATA	TTT	TTT	AAA	TAA	FACO	CAA	AAA'	IGT	TA	AAA	AAT	AAA	CAAA	AA A	1440
			• D	raI		•							. ``							
TGT	TT	GAT	TTT	TTA	AAC	AAG	TTA	TAG	TGT	ATT	TA	ACTZ	AAAC	GAA	AAA	CGA	AAA	CGCI	TT	1500
		-35	•			• :	rbs		•	-10			•			•			•	
TCIC	SAC	TAA	AAA	ATC	ATT.	ICA	AGG	AGA	ATG	ACTZ	ATGO	GCA	ATT7	\TT/	ATT	GT	ICA	GATO	<u>SCA</u>	1560
											М	A	I	I	I	G	S	D	A	9
Pvu]	II								-		18	CI	V							
Alu	ī		•			•	_		:		D	raI	•			•			•	
GCIC	GT.	AAA	AGA	CIT2		JAIN	51°17	ATT:	AAA	ACA'I	<u> </u>		AQAC	SACA	AC/	VAL	CAT	JAAG	arr 	1620
A	G	ĸ	R	г	ĸ	D	v	T	ĸ	т	F.	Г	ĸ	D	N	N	н	E	v	29
			•			•			•		1	lini	EI .			•	į	Rsal	•	
TTAC	AT	GTT.		GAA	AGG/		JAT.	PIG	GAT.	-	STAC	SAT.	CCA7		FTAC	GCAC	STT	TAC	AT	1680
Г	D	v	т	Е	ĸ	ĸ	D	ь	D	F.	v	D	s	т	Г	A	v	v	н	49
F	sa	I	•			•			•			•	•			•		A	luI	
GAAC	TA	CAA	AAA	AAT	GAT/	AAA	AAT	TTG	GGG	ATTO	CA	TT	SATC	CCI	TAT	GT	GCG	3GCA	GC	1740
Е	v	Q	ĸ	N	D	к	N	L	G	I	A	I	D	A	Y	G	A	G	s	69
		A	luI			•			•							•			•	
TTT	TG	GT <u>A</u>	GCT/	ACT	AAAC	GTC/	AAA	GT	ATG/	ATTG	CAC	CAC	SAAG	TG	CTC	SATC	SAA	CGTI	CA	1800
F	М	v	A	T	ĸ	. <u>v</u> .	ĸ	G	M	I	Α	Α	Е	v	s	D	Е	R	s	89
A 7 11 1				Fnul	hae. DII	•		- 4	HNA. Fni	L DTT	-									
GCT	AT	ATG	ACA	CGCC	GCC	CAT/	\AT/	AATS	GCGC	GTA	TCZ	TT7	CTC	TTO	GT	CTC	GAA	ATTG	TT	1860
A	Y	М	т	R	G	н	N	N	Α	R	I	I	т	L	G	Α	Е	I	v	109
			Ba	11	тт (Ceni	r						4m 7 T					Xmal	II	
GGAG	AT	GAA	CTG	GCC/	AAA	AT	TT	GTT	AAGO	GACT	TTC	TAC	AGG	CT7	AA	TAT	GAC	GCG	GC	1920
G	D	Е	L	Α	ĸ	N	I	v	ĸ	D	F	v T	Е	A	к	Y	D	G	G	129
-	_																	Mnl	I	
COTTO	יאידעי	<u>М</u> С Х С Г	boi MTC	2011	מידיב	• • •	1	~ma	• •		-	201	מממי		• 8 8	• • •	ri 'DA:	28		1980
R	H	Q	Ĩ.	R	v	D	M	L	N	K	M	č	*	GAC			10	Grie		142
AATG	AA	AAT	3881	AATT	IGCI	CAT	rGG'	PTG:	IGAC	CAT	נידאי	GTT	' 'ACT	GAT	GTO	AAZ	ATC	GAA	CT	2040
		м	ĸ	I	A	I	G	C	D	н	I	v	т	D	v	ĸ	M	Е	L	18
		17	-			-		-	-		_		-	-	-					
		4	ac											•						
			•						: -		Dal	·		A	/all	·			n.	2100
דייאדע	נאאי	יאיאמ	TC-TV	AA A/	NGA/	N. SAA	AGG I	I'I'A'	IGAZ	AGT T	CTA	(; A'	(TI'A	CCC	ACC	"I'A'I	GAU		AL.	2100
ATCI	'AA' K	ACA' H	ICT/ L	۹АА/ K	-GAA E	E E	GG' G	PTA: Y	IGA/ E	VGT	CTA L	D	GTA V	0 <u>00</u> 0. G	ACC T	TAT Y	D	F	T	2100
ATCI S	K	ACA: H	L	K	E	E	G	Y	E	V	L	D	V	G G	T	Y	D	F	T	38
ATCI S Rs	K K AI	ACA'	L	K	E	E	G	Y	E	V V	L	D	V	G	T	Y	D	F	T	38
ATCI S Rs AC <u>GI</u>	K K AI	ACA H GCA	L	K	E	E	G	Y Y	E	V V	L	D	V	G G G IGTI	T	Y	D	F	T	2100 38 2160
ATCI S Ra AC <u>GT</u> R	K K A Z A C T	ACA H GCA H	L CTA: Y	K K ICCC P	E	E 	GG G GG G G	TAA Y TAA K	E E AAAJ K	V V GTG V V	L GGC G	GAI D GAA E	GTA V GCA A	G G G IGTI V	T T TAGO	TAT Y CAGI	D IGG IGG	F GAA E	T GC A	2100 38 2160 58
ATCI S Ra AC <u>GT</u> R	K K A ZAC T	ACA H GCAG	L CTA: Y	K K ICCC P	E CATO I	E TTT F	GG G GG G R	TAY Y TAA K K	IGAA E VAAA K	V V AGTG V	L GGC G	D D GAA E	V V AGCA A	G G G IGTI V	T T TAGO S	Y AGI	IGAC D IGG! G	F GAA E <i>Rs</i>	GC A A	2100 38 2160 58
ATCI S AC <u>GI</u> R GGAI	AAD K aI <u>AC</u> T	ACA H BCA H AGG	L CTA: Y	K K ICCC P	E CATO I FATO	E TTT F		TAA Y TAA K SaI	IGAA E AAAA K	V V AGTG V	L G G G G G G G G	D CGAN E	V GCA A AGT	G G G G IGTI V V	T T TAGO S	Y AGI S	IGAC D IGG G	F GAA E Rs TAAA	T GC A GT	2100 38 2160 58 2220
ATCI S AC <u>GT</u> R GGAI D	XAAI K AI YACO T T TTI	ACA H BCAG H AGG G	ETA: L TA: Y IGT/ V	K K ICCC P ATGI C	E CATC I I M	E TTT F C	GG G G G R G G G	TAA Y TAA K SaI TAC	IGAA E AAAA K IGGI G	V V NGTG V NGTG V	L G G G G G G G	GAN CGAN E NATN I	V GCA A AGT	GGG G GTI V ZAAI N	T TAGO S NGCI A	Y AGI S	IGAC D IGG G CAA N	F GAA E Rs TAAA K	GC A GC A GT V	2100 38 2160 58 2220 78
ATCI S Re AC <u>GT</u> R GGAI D HDA		ACA H BCA H AGG G	ETA: L CTA: Y IGT/	K ICCC P ATG1 C	E CATC I IATC M	E TTT F C	GG G IGG G R I G G G	PIA Y TAA K SaI T	IGAA E AAAA K IGGI G	V V AGTG V TGTG V	L GGGC G G G G G	D SAP E NATA	AGCA A AGCA A AGT S	iG <u>GG</u> G IGTI V XAAI N	T TAGC S NGCI A	Y AGI S	IGA D IGG G CAA N	F GAA E Rs TAAA K	T GC A GT V	2100 38 2160 58 2220 78
ATCI S Re ACGI R GGAI D Hpa ACCG		ACA H BCAC H AGG G	L CTA: Y IGT/ V FAG/	K ICCC P ATGI C	E CATC I FATC M SGCP	E TTT F C C		TAAJ K SaI T T	IGAA E AAAA K IGGI G	V V AGTG V V TGTG V	L G G G G G G G G G G G G G G G G G G G	D CGAN E NATN I TCI	V AGCA A AGTA S AGCA	GGGG G IGTI V XAAI N	TAGC S NGCI A	Y AGT S IGCO A	IGA D IGG G CAA N N	F GAA E Rs TAAA K	T GC A GT V GA	2100 38 2160 58 2220 78 2280
ATCI S Re ACGI R GGAI D Hpa ACCG P	K al <u>AC</u> T T L <u>G</u> G	ACA H BCAG H AGG G IGT V	L CTA: Y CTA: Y CTA: Y CTA: R	K ICCC P ATG7 C ATG7 C	E CATO I TATO M SGCA		G G G G G G G G G V G T V	TAAJ K SaI TACT T AAGJ R	IGAZ E NAAZ K IGGI G G NGAI D	V V AGTG V NGTG V V TATG M	L GGC G G G G G ACI T	D GAA E NATA I S	AGCA A AGCA S AGCA A	GGGG G IGTI V XAAI N ICTG L	TAGC TAGC S NGCI A TAI	Y AGI S NGCC A TCZ	IGA D IGG CAA N N AAA K	F F GAA E FAAA K K GAA E	T GC A GT V GA E	2100 38 2160 58 2220 78 2280 98
ATCI S Re AC <u>GT</u> R GGAT D Hpa ACCG P	K A A A A A A A A A A A A A A A A A A A	ACA H BCAG H AGG G IGT V	ETAC L CTAC Y IGT/ V FAG/ R	K ICCC P ATG7 C AACO T	E CATC I TATC M BGCA A	E TTT F C C TTZ C		TAAJ K SaI TACT T AAGJ R	IGAA E K K IGGI G G AGAI D HI	V V AGTG V CGTG V V TATG M	L GGGC G G G ACI T	D CGAN E CATN I TCT	AGCA A AGCA A A A A A A A A	G G G G G G T V XAAT N C T G L	TAGC S NGCT A TAT	TAT Y SAGT S NGCC A	IGAC D IGG G CAA N N AAAJ K	F GAA E RS TAAA K K GAA E	T .GC A GT V .GA E .	2100 38 2160 58 2220 78 2280 98
ATCI S RE ACGI R GGAT D Hpa ACCG P ACTI	K AA AA AA T T T T L L II G G G G	ACA H BCA H AGG G IGT V	L CTA: Y IGT/ V FAG/ R R		E CATC I FATC M SGCA A CGTI	E F F C C C L Alt C		Y Y RAAJ K SaI RACJ R R R R R		V V V V V V V V V V V V V V V V V V V	L GGC G G G G G G C G C C C T T T	D CGAM E CATM I TCI S	V GCA A GCA S GCA A GCA	GGG G G V XAAT N CTG L XAAG	T TAGC S TAGC A TAT Y TTTC	Y CAGI S CAGI S CAGI S CAGI S S CAGI S S S S S	IGG D IGG G CAA N N AAAJ K XTTJ	F GAA E RAAA K K GAA E	GC A II CON	2100 38 2160 58 2220 78 2280 98 2340
ATCI S RE ACGI R GGAI D Hpa ACCG P ACCTI L	K AI AC T T L II GG G G XAA N	ACA' H 3CA(H AGG' G G IGT' V IGC/ A	L CTAC Y IGT/ V FAG/ R R AAAM	K K P ATGJ C AACO T C V	E CATC I FATC M GGCA A CGTI V	E TTT F C C Altr Altr S	GG G G G G G G G G G M G T T T T T T T T	TAAJ K SAI TAAGJ T R GGC G	IGAA E AAAAA K G G G AGAJ D HI CG <u>G</u> G	NGTG V NGTG V CGTG V CATG M NaI NGCT A	L GGC G G G G G ACI T T XATI I	D CGAF E VATF I S VATF I S VATF	GTA V AGCA A AGGT S GGT G	GGG G G G V XAAT N CTG L XAAG K	T TAGO S TAGO A TAT Y TTO L	TTTP	IGAC D IGGS G CAAS N N AAAAA K K IIIII L	F F GAA E TAAA K K GAA E TTTT F	GC A III V GA E GA D	2100 38 2160 58 2220 78 2280 98 2340 118
ATCI S Re ACGI R GGAT D HDA ACCG P ACTT L	K AA A A A A A A A A A A A A A A A A A	ACA' H 3CA(H AGG' G G IGT' V IGC/ A	L CTA: Y IGT/ V FAG/ R R	K R R C R R R R R R R R R R R R R R R R	E CATC I FATC M SGCP A CGTI V	E TTT F C C ATT Alu S	GGT G G G G G G G G V I I T T F	TAA Y K SAI K SAI T T AAGJ R GGGG G	IGAA E AAAAA K IGGI G AGAI D HI CG <u>G</u> G	AGT <u>U</u> V AGTG V TGTG V V TATG M AaI A	L GGC G G G G G G G G C I T T T I	D CGAP E VATA I S VATA I	GTA V AGCA A AGCT S A GGCT G	GGG G G G T V V XAAT N C T G L XAAG K	T T TAGC S TAGC A TAT Y TTC L	Y CAGI S CAGI S CAGI S CAGI S S CAGI S S CAGI S S CAGI S S S CAGI S S S S S S S S S S S S S S S S S S S	IGAC D IGG G CAA: N N AAAJ K K ITTJ L	F GAAA E Rss FAAAA K AGAAA E AGAAA F	GCA GTV GA E GA D .	2100 38 2160 58 2220 78 2280 98 2340 118
ATCI S Res ACGI R GGAT D Hpa ACCG P ACCTI L TATI	K SAL K SAL T TTT/ L SII GG N GG N GT/	ACA' H G G G IGT V IGC/ A AGA'	ICTA L CTA: Y FIGT/ V V FIAG/ R R AAAA/ N	K K FCCCC P ATG7 C ATG7 C T T GTC V	E CATC I FATC M GGCJ A CGTI V	E TTT F STGC C C Alt S S Alt S	GG G G G G G G G G MGT/ V T T T F K G G C	TAN Y RAAJ K SAI T T T AAGJ R G G CCAC	IGAA E AAAAA K G G G AGAA D HI CCGG G G STAI	IGTI V IGTG V IGTG V IGTG M IATG M IATG A XAAA	L GGC G G G G G G G G G G G G G G G G G	D CAP E VATA I TCI S CATA I SACA	GTA V AGCA A AGCA A GGGT G AGAA	GGG G G Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q	T T TAGC S CGCI A TAI Y TTTC L	Y CAGI S GCCC A TCZ S TTZ L	IGAC D IGG G CAA: N N AAAA K TTTI L	F IGAA E Rss ITAAA K AGAA E AGAA E ATTT F Asei I	T .GC A GT V .GA E .GA D .AT	2100 38 2160 58 2220 78 2280 98 2340 118 2400
ATCI S ACGI R GGAI D HDa ACGI P ACTI L I TATI	K AAI K AII T T T L GG G V AA V C T V C T V C C V V V V V V V V V V V V V	ACA' H BCAG H AGG G G G IGT A AGA D	L L CTA: Y V TAGJ R TAGJ R N N S CGCI A	K K FCCCC P ATG7 C ATG7 C T T C T T C T T T T T T T T T T T T	E CATC I FATC M BGCA A CGTI V CGTI V TATI I	E TTT F STGC C ATTZ Alt S S S AL S S K	GGT G G G G G G G G G G G T T T F F G G C C G G G G G G G G G G G G G G	TAA Y FAAA K SAAI T T T AAGJ R G G CCAC Q	IGAA E KAAA K G G G G G G G G G G G G G G G G	KGTG V KGTG V KATG M JAI XGCT A XGCT A XAAA K	L GGC G G G G G G C C C C C C C C C C C	D CGAP E VATP I TCI S CATP I SACP T	GTA V AGCA A AGGT G G G G G G G G G G G G G G G G	GGG G G G TAAT N CTG L CTG K SAAG K GAA	T T TAGC S TAGC A TTAT T TTTC L L AAT	Y Y S NGCC A TCC S TCC S TCC S TCC S S TCC S S TCC S S TCC S S S S	IGAC D IGG G CAA: N AAAA K L J SAA2 K	F IGAA E RSA ITAAA K AGAA E AGAA E ASEI TTTA L	AC T GC A II GC A II GC A II GC A GA GA GA GA GA GA GA GA GA GA GA GA GA G	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138
ATCI S ACGI R GGAI D HDA ACCI L L TATI I	K A A A A A A A A A A A A A A A A A A A	ACA' H BCAG H AGG G G IGT V V IGC I A AGA D	L L CTA: Y R TAGJ R R R AAAM N N S GCT A	K FCCCC P C C ATG7 C T GTC V F F	E CATC I FATC M BGCA A CGTI V FATI I	E TTT F STGC C ATTZ L Alu S S ALU S S ALU	GG G G G G G G K G T T T F K G C C G A	TAN Y FAAJ K SAI FACT T R AAGJ R G G CCAC Q	IGAA E AAAAA K G G G G G G G G G G G G G G G G	KGTG V KGTG V V TGTG V V XATG M DaI SGCT A XAAA K	L GGC G G ACI T I L CCG P	D SAN E VATN I SACN T	UGTA V AGCA A AGGA A GGGT G AGAA E	GGG G G V V V AAT N CTG L V AAG K GAA E	TAGCI S CGCI A STAJ Y STTC L AAAT N	Y S CAGI S CGCC A S TTC S S TTC L S AAG K	GAG D GG CAAS N K AAAJ K TTJ L SAAJ K	F IGAA E Rss TAAA K K AGAA E ATTT F AsseI L L	AC T GC A II GT V GA E GA D TI	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138
ATCI S RE ACGI R GGAI D HDA ACCG P ACTI L TATI I I GGCI	K SAI KAC T T T L II GG G V C N C T V N C T V N C C V N C C V N C C C C C C C C C C C C C	ACA' H 3CA(H AGG G G G G G G G G G C A AGA D D AGA D AGA AGA A AGA A A G A G	L L CTA: Y V V TAGJ R R AAAA N N GCT A	K FCCCC P C ATG7 C AACC T F C V F F F	E CATC I TATC M SGCP A CGTI V TATI I	E E TTT F STGC C C ATTZ L All S S AAAA K	GG G G G G G G G G V I I T T F K G C C A	TAX Y RAAA SAI T T T AAGJ R G G CCAC Q	IGAA E 	KGTG V KGTG V KATG M MAI K CAAA K K	L GGGI G G ACI T VATI I CCCG P		GTA V AGCA A AGCA S CCA A GCA E AGCA	GGG G G V V V AAT V AAT N CTG L V AAG K GAA E MDC GAT	TAGCI STAGCI A STATI STATI L AAAT N DI	Y S CAGI S CACC A S TCZ S S TTZ L S CAAGI S S CAAGI S S CAAGI S S CACC S S S S S S S S S S S S S S S	IGAC D IGG G CAA N N AAAA K I I J AAAA K TTT	F IGAA E Rss TAAA K AGAA E ATTT F Asei L TTTA L	AC T GC A MAIN OF ACTION O	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2460
ATCT S Rest ACCT R GGAT D Hpaa ACTT L ACTT L TATTI I D C TGCTI A	KAA KAA T T T L S G G S AA V K K K K	ACA' H BCAG H AGG G G G G G G T G C T A AGA D A G A G A G A G T C T T I I I I I I I I I I I I I I I I	ICTA L CTA: Y FAG7 R R AAAA N N GCCI A L L AAAA K	K ICCCC P ATG7 C T T IGTC V F F ACA1 H	E CATC I FATC M SGCP A CGTI V FATI I FTTA L	E E TTT F STGC C C Alt S S AAA K K S AAA K	AGGI G G G G AGTI V I I T T F AGCCC A AGCI A	TAX Y RAAJ K SAI T T AAGJ R G G CCAC Q ACAJ H	IGAA E AAAAA K G G G G G G G G G G G G G G G C G G C G C C G C	V V GTG V V TATG M Mai X CT A X X A X X A X A X A X A X A X A X A	L GGC G G ACT T I ACT I CCCG P XAAA K	CAP	GTA V AGCA A AGGT G AGGT G A GGAA E A GCA A	GGG G G V V AAT N CTG L CTG K GAA E MDC GAT D	TAGC TAGC S CGCI A STAT Y STAT L AAAT N S CCC P	Y CAGI S CAGI S CAGI S S TCZ S S TCZ S S TCZ S S CAAGI S S CAAGI S S S CAGI S S S S S S S S S S S S S S S S S S S	D D G G CAA N N AAA K L J SAA K TTY F	F RGAA E RSS RAAA K AGAA E AGAA F ASSEI L TTTT F	ACT. GCA MATT. GA D. ATT. GA D. GA D	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2460 158
ATCT S Res AC <u>GT</u> R GGAT D Hpaa ACTT L TATTI L TATTI I D d TG <u>CT</u> A	K AA K AA T T T L C G C C C C C C C C C C C C C C C C C	ACA H B B CA H A G G G G G G G G G G G G G G G G G	L CTA: Y FIGT/ V FIGT/ N FIGT/ A A A A A A A A A A A A A A A A A A A	K F C ATG7 C ATG7 C T T GTC V F F ACA1 H	E CATC I FATC M GGC/J A CGTI V FATI I FTTA L	E TTT F STGC C Alti S S Alti S S Alti S S S S S S S S S S S S S	GGT G G G G G G G G V I I T T T F G C C C A G C I I	ITA: Y FAAJ K SaI FAC T T AAGJ R CCAC G CCAC Q ACAT H	IGAA E K K G G G G G G G G G G G G G G C G G C G C G C G C G C C G C C C G C	V AGTG V TATG M AGTG V V TATG M A A A A A A A A A A A A A A A A A A	L GGC G G G G G G G G C C C G G C C C G C	цал D Сам E VATM I VATM I SACM T CAM Q	AGCA A AGCA A AGCA A A A A A A A A A A A	GGG G G V V XAAT N CTG L XAAG K GAA D	TAGC TAGC S CGCI A STAI Y STTC L AAAI N SI SCCG P	Y CAGI S MGCC A S TCJ S S TTJ L S CAI H	IGG G G CAA: N AAAJ K K TTY K F	F IGAA E Rss Raa K AGAA E AGAA E AGAA E ATTT F L	ACT CA AIT CA AI	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2460 158
ATCT S Res ACCT R GGAI D Hpa ACCC P ACTT L I TATTI I I D C C TATTI I A	K A A A A A A A A A A A A A	ACAY H 3CAG H AGG G G G G G G G G G G G G G G G G	L L CTA: Y V V TAGJ R R N N S CTA: Y V V V L S CTA: Y V V V V V V V V V V V V V V V V V V	K F C C ATG7 C AACC T T GTC V F F ACA1 H	E CATC I FATC M GGCJ A CGTI V FATI I FTTA L <i>F</i>	E TTT F STGC C Alti S C Alti S C Alti S S S S S S S S S S S S S	GGT G G G G G G G G G G G G G G G G G G	ITA: Y FAAJ K SaI FAC T T AAGJ R G G CCAC Q ACAI H	IGAA E K K K G G G G G G G G G G G G G G C G G C G G C G C G C G C G C C G C C G C C G C C C C G C	V AGTG V CGTG V CATG M AAI A CAAA K CAAA K CQAT D	L GGGC G GGGT G G ACT T I S CCCG P XAAA K		AGCA A AGCA A AGCA A A A A A A A A A A A	GGG G G V XAAT N CTG L XAAG K GAA D D	TAGC TAGC S CGCI A STAI Y STTC L AAT N S CCC P	TAT Y S S S S S S S S S S S S S S S S S S	IGG G G CAA: N AAAJ K TTT K F	F NGAA E RSS TAAA K AGAA E AGAA F ASSEI L TTTT F	ACT CA AIT CA AI	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2460 158
ATCT S Res ACT R GGAI D Hpa ACCT D Hpa ACCT I L TATT I I D C TGCT A TGAA	K A A A A A A A A A A A A A A A A A A A	ACAY H BCAG H AGG G G G TGT V V TGC A A G A G A G T C T C T C T C T C T C T C T C T C T	L CTA: Y V TAGJ/ R R N CTAGJ/ R N CTAGJ/ K	K R C P ATG7 C AACC T C V TTT7 F ACA1 H H	E CATC I FATC M GGCA A CGTI V FATI I FTTA L E FTGG	E TTT F STGC C ATTZ L Alu S S C AGAZ K K S AGAZ	GGT G G G G G G G G G G G G G G G G G G	ITA: Y RAAJ K SaI T T T AAGJ R G G CCAC G CCAC Q ACAJ H	E AAAAA K G G G G G G G G G G G G G G G G	V AGTG V CGTG V CATG M AAI A CAAA K CAAA K CAAA K CAAT D	CTZ G G G G G G G G C ACT T T C CCG K C ACC		AGCA A AGCA A AAGI S AGCA A C GCA A C GCA A C GCA A C GCA A C C C A C C C A C C A C C A C C A C C A C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C C A C C C C A C	GGG G G V XAAT N CTG L XAAG K GAA D T GTA	TAGC S A CCI A STAG TAI Y STAT A AAT N S CCC P S S CCC P S S CCC P	TAT Y S S S CAG S S S S S S S S S S S S S S S S S S S	GAC D G CAA: N AAAA K TTY K TTY F	F IGAA E Raa FAAA K AGAA E ASSEI TTT F TTTT F SATC	AC T GC A II GC A II GC V GA E GA D AT I GA D AT	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2400 138 2460 158
ATCT S REACT R GGAT D HDS ACCC P ACCC P ACCC P ACCC I D ACCC T A C TGAT A C C TGAA E	K AA K AA T T T L C G C AA V C V N C T T C C C C N V C V N C T T T T T T T T T T T T T T T T T T	ACAY H BCAG H AGG G G TGT V V IGCI A AGA D D IGCI I I I I I I I I I I I I I I I	L CTA: Y V TAGJ R R N CTAGJ R R N CTAGJ R R N CTAGJ K L CTA: Y V V TAGJ K K L CTA: Y V V TAGJ V R R N S TAGT/ V V V TAGT/ V V V V V V V V V V V V V V V V V V V	K R C P ATG7 C AACC T C V T T F F ACA1 H H AAAC K	E CATC I FATG M SGC2 A CGTI V V TATI I FTTA L E FTGG W	E TTT F STGC C VTTZ L Alu S S S S S S S S S S S S S S S S S S S	GGI G G G G G G G G G G I I F M G C G I F T I F N G C G I R I R I R I R I R I R I R I R I R I	ITA: Y RAAA K SaI T T T AAGA R R G CCAC G CCAC Q ACAA H DIII SGG	E AAAAA K G G G G G G G G G G G G G G G G	V AGTG V CGTG V CGTG V V TATG A A CGAT D CGAT D CTAT	L GGGI G GGGI T T I SATI I I I I I I I I I I I I I I I I I I	CAP CAP E CAP I CAP I SACA T CAP CAP CAP CAP	AGCA A AGCA A AGGT G A A A A A A A A A A A A A A A	GGC G G CAAT N CTG L CTG C CTG K GAA G C T C C C C C C C C C C C C C C C C C	TAGCI S TAGCI A STAI Y STAI Y STAI N DI S S C C C P S S S S S S S S S S S S S S	TAT Y SAGI S S TCZ S TCZ S S TCZ S S TCZ S S TCZ S S S TCZ S S S S S S S S S S S S S S S S S S S	GAC D G CAA: N AAAJ K L SAAJ K TTY F	FITT F RSA RAGAA K AGAA E ATTTT F L TTTT F TTTT F	AC T GC A MAI GC A MAI GC V GA E GA D AT I GA D AT	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2400 158 2520 171
ATCT S REACT R GGAI D HDA ACTT L TATTI I D ACTTI L TGAT A TGAA E	K AA K AA T T L L C C C C C C C C C C C C C C C C	ACAY H 3CAY H AGGT G G G G G G G G G G G G G G G G G	L CTA: Y V TAGJ R R N CTAGJ R R N CTAGJ R R N CTAGJ K L CTA: Y V V TAGJ R R N N CTA: Y V V TAGJ V R TAGJ V TAGJ V R TAGJ V V V V V V V V V V V V V V V V V V V	K F C C T C C T C C T C C T C C T C C T C C T C C C T C C C T C C C C T C	E CATC I FATG M SGCP A CGTI V V FATI I FTTPA L E FTTPA V W	E CTTT F STGC C C All CAGA S CAAN K S CAAN K S CAAN S C C C C C C C C C C C C C	GGI G G G G G G G G G G G I I F M G C G I F T I F N G C G C G C G C G G C G G C G G G C G	ITA: Y RAAJ K SAI T T AAGJ R CCAC G CCAC Q ACAJ H JDIII XQGI G	E AAAAA K . G G G G G G G G G G G G G G G G G G G	V AGTG V CGTG V CGTG V V CGTG A A CAAA K CGAT D CTAT Y	L GGGI G GGGI T I SATI I SATI I SATI I SATI I SATI I SATI I SATI I SATI I SATI I SATI I SATI I	CAP CAP E CAP E CAP T SACA T CAP CAP CAP CAP CAP	V AGCA A AAGT S CGCA A AGGT A A GGCA A A CGCA A A GGCA A A GGCA A A GGCA A A CGCA A C C A A C C A A A C A A A C A A A C A A A C A A A C A A A C A A A C A A C A A C A A C A A C A A C A A C A A C A A C A A C A A C A A C A A C A A C A A C A A C A A C A A C A C A A C A A C A C A A C A C A A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C C A C C A C A C C A C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C C A C C A C C A C C A C C A C C A C C A C C C A C C C A C C C A C C C A C C C A C C C A C C A C C C A C C C A C C C A C C C C A C C C A C C C A C C C C C A C C C C A C C C A C	GGG G G CAAT N CAAT N CAAT L CAAG K GAA D CAA C CAA C CAA C CAA C C C C C C C	TAGCI S TAGCI A STAI Y STAI Y STAI N DI S CCCO P S S S S S S S S S S S S S S S S S S	TCZ S CGCC A TCZ S TCZ S TCZ S TCZ S TCZ S CGCC A S TCZ S S TCZ S S TCZ S S TCZ S S TCZ S S S TCZ S S S S S S S S S S S S S S S S S S S	GAC G G CAA? N AAAA K K TTY F GTT	FITT F IGAA E Raaa K AGAA E AGAA E AGAA E ATTT F F SATC	ACT GA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2460 158 2520 171
ATCT S REST R GGAT D HD ACTT L HD ACTT L TATTI I D C TATTI I TGCT A TGAAA E TATG	K AAA K AAA T T L L L L L L L L L L L L L	ACAY H 3CAY H AGGC G G G G G G G G G G G G G G G G G	L CTA: Y V FAGJ R FAGJ R R FAGJ R R R R R R R R R R R R R R R R R R R	K K C P C C C C T C C T C C T C C T C C T C C T C C C C C C C C C C C C C	E CATC I FATC M BGCA A CGTI I V FATI I FTTA L A STGG W	E CTTT F STGC C C All CAGA S S All All	GGTI G G G G G G G G G G G G G G G G G G	TAX Y RAAA SAI T T T T AAGA G G CCAC Q ACAT H JDII G G G G G G G G G G G G G G G G G G	E AAAAA K . INGGI G . AAGAI D . HI G G G . INGGI C G G . INGAC D . INGAC	V V V V V V V V V V V V V V V V V V V	L GGCC G G G G G G G G G G G G T T I S C C G C G C G G C G G C G G C G G C G G C C G C C G C G C C G C C G C C G C C G C C G C C G C C C G C C C G C		U GCA A GCA A GCA A GCA A GCA A GCA A GCA	GGC G G G C T C C C C C C C C C C C C C	TAGC S A STAT Y STAT V S TA P S C C C P P N D S C A C A C A C A C A C A C A C A C A C	TAT Y S AGO A S TCZ S TCZ S S TCZ S S TCZ S S TCZ S S S TCZ S S S TCZ S S S S S S S S S S S S S S S S S S S	GAC D G CAA? N AAAA K TTY F GTT GTT	FITT F RGAA E RSS TAAA K K GAA E VITT F L STTT F CITT F	ACT GCA III GA D GA	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2400 138 2460 158 2520 171
ATCT S Rs ACCT R GGAN D HD2a P ACCT L L TATTI L D D C T TATTI A TGAA E TATG M	K ali T T L III G C C C C C C C C C C C C C	ACAY H 3CAY H 3CAY H G G G G G G G C C C C C C C C C C C C	L CTA: Y V TAGJ R TAGJ R N CTAGJ R N CTAGJ K L CTAGJ K K L CTAGJ R N N CTAGJ R N N CTAGJ R N N CTAGJ R N N CTAGJ R N N N CTAGJ R N N N N N N N N N N N N N N N N N N	K K C P ATG7 C AACC T C AACC T F ACA1 H AAAC K V V	E CATC I FATC M BGCA A CGTI I V FATI I FTTA L A STGG W TACG T	E CTTT F STGC C All CAL All CAL S S S AL S S S S S S S S S S S S S	GGTI G G G G G G G G G G G G G G G G G G	TAX Y TAAA K SAI T T AAGA R G G CCAC Q ACAN H DIII SCGT G G G G G G G G G G G G G G G G G G	E AAAAA K G G G G G G G G G G G G G G G G G G	V V V V V V V V V V V V V V V V V V V	L GGC G G G G G G G G T T I S C C G G K K C A C A C A C A C D	CAP I SACP I SACP I SACP I SACP I SACP I SACP I I	GCA A GCA A GCA A GCA A GCA A GCA A GCA A GCTA	GGC G G C C C C C C C C C C C C C C C C	TAGO STAT STAT STAT STAT STAT STAT N STAT N STAT N P SCO O P	TAT Y CAGT S CAGT S CAT S CAT H K CAT H H L L L	GAC D G CAAN G CAAN N N AAAAA K K TTY K F GGTI GGTI GGTI	FITT F RGAAA E RAAA K GAAA E MITTT F L STITT F CATC C GAC D	ACT .GCA MATT .GCA .GCV .GAE .GAD .TI .GCD .AT .TL	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2400 158 2520 171 2580 20
ATCTI S Res ACGTI R GGAI D Hpaa ACTTI L TATTI I D C TATTI I TGAA E TATGA	K all T T L L L L L L L L L L L L L	ACAY H 3CAY H 3CAY H 3CAY G G G G G G G G G G G G G G G G G G G	L CTA: Y V FAG2 R N N CTA: Y V V FAG2 R N N CTA: Y V V V TAG2 R N N CTA: Y V V TAG2 R R N N CTA: Y V V V TAG2 R R N N N CTA: Y V V V V V V V V V V V V V V V V V V	K FCCCC P ATG7 C AACC T C AACC T F F ACA1 H AAAC K K V	E CATC I I TATC M GGCA A CGTI I V V TATI I FITA L E STGG W V V CATG T	E TTT F STGC C Alt TAGA S Alt TAGA K S S S S S S S S S S S S S	GGT G G G G G G G G G G G G G G G G G G	TAX Y TAAA K SAI T T AAGA R G G G CCAC G G CCAC Q ACAT H JDIIJ G G G G G G G G G G G G G G G G G G	E AAAAA K G G G G G G G G G G G G G G G G	V V V V V V V V V V V V V V V V V V V	L GGCC G G G G G G G G G G G G G G G G G	CGAN E CGAN E CATA I SACA T CATA I SACA T CATA I I CATA I I	UGCA A AGCA A AGGT G A A A A A A A A A A A A A A A	GGC G G C C C C C C C C C C C C C C C C	T TAGC S CGCI A STAI Y STAI Y STAI P S AGC P S DS CCO Q	TAT Y CAGT S CAGT S TCZ S TCZ S TCZ S TCZ S TCZ S TCZ S CAGT S CAGT S CAGT S S CAGT S S CAGT S S CAGT S S CAGT S S CAGT S S S CAGT S S S S C C C C S S S S S S S S S S S	GAC D G CAAN G CAAN N N AAAAA K K TTY F GAT D	FITT F RGAA E RAA K K GAA E MITTA F SATC. D	ACT .GCA MATT .GCA .ATT .GCA .ATT .GCA .TTL	2100 38 2160 58 2220 78 2280 98 2340 118 2460 138 2460 158 2520 171 2580 20
ATCTI S RESCONT R GGATI D HDA ACCCC P ACTTI L TATTI I I CCCT TATTI I CCCT TATTI I CCCT TATTI I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT TATTI I I CCCT I CCCT I CCCT I CCCT I D CCCT I I CCCT I I CCCCT I I CCCT I I CCCCCT I I CCCT I I CCCCT I I CCCTI I I I	K AAA K AAA T T T L L L L L L L L L L L L L L L	ACAY H BCAC H AGGT G G G G IGCT A AGAT I C C C	L L CTA: Y N CTAGJ R R N N CTAGJ N N CTAGJ N C CTACJ C CTACJ C CTACJ C CTACJ C CTACJ C CTACJ C CTAC C CTAC C CTAC C CTAC C C C C C C	K FCCCC P ATG7 C ATG7 C AACC T T T F ACA1 H AAAC K AGT1 V	E CATC I FATC M SGC2 A SGC2 A CGTI V TATI I FTTA L FTTA L K STGG W TACG	E TTTT F STGC C Ali S S AAAA E S S AAAA E S AAAA E S AAAA S S AAAA S S AAAA S S AAAA S S A A A A A A A A A A A A A	GGT G G G G G G G G G G G G G G G G G G	TA: Y FAAJ K SAI TAAGJ R TAAGJ R TAAGJ R CCAC Q ACAJ H JDIJ JCCI P SCCI P	E AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	V V V V V V V V V V V V V V V V V V V	L GGCC G G G G G G G G G G G G G G G G G	CAAP I CAAP I CAAP I CAAP I SACA I SACA I CAAP I CAAP I CAAP I I CAAP I I CAAP I I CAAP I I CAAP I I CAAP I SACAI I SACI I SI I I I I I I I I I I I I I I I I	UGCA A AGCA A AGCA A A A A A A A A A A A	GGC G G I CTG L CTG L CTG L CTG K GAA GAA GAA GAA TAT Y HD	TAGC S A CCC A TAGC A TAGC A TAGC C C C C C C C C C C C C C C C C C C	TAT Y CAGT S CAGT S CAT S CAT S CAT H CAT L CAT L CAT L CAT L CAT L CAT L CAT S C CAT S C C C C C C C C C C C C C C C C C C	GAC D GG G CAA? N K K TTY K F GAT D GAT	CTTT F RGAA E RASA K K AGAA K K AGAA F L STTTA F CTTTT F CATC	ACT OGA HETV OGE OGD OTTI OGD OTTI OTTI OTTI	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2400 138 2520 171 2580 20
ATCTI S R ACCTI R GGAN D HpaaGAT D ACTTI L L TATTI I D C TATTI A AAGG P ACTTI L L TATTI I A C C C P T C C C P T C C C C C C C C C C	K AAA K AAA T T T L L L L L L L L L L L L L L L	ACAY H BCAG H AGG G G IGT V IGCI A AGAY I C C I G C I C I G C I C I C	L L CTA: Y N CTA: Y N CTA: Y N CTA: N N CTA: N N CTA: N N CTA: Y CTA: Y CTA: Y N C CTA: Y N C CTA: Y N C CTA: Y N C CTA: Y N C C C C C C C C C C C C C C C C C C	K K R C P AIGJ C AIGJ C V T F ACAJ H AAAC X V C V C V C V C V C V C V C V C V C V C V C V V V V V V V V V V V V V	E CATC I FATCO M GGCJ A SGCJ A CGTI V TATI I CATI I K STGG W TACG T CAAT	E TTT F STGC C Alt S TTT Alt S TAA K S S S S S S S S S S S S S	GGG GGG GGG GG GG GG GG GG CG GG CG GG CG GG CG GG CG GG CG GG CG GG CG GG CG GG CG GG CG GG CG GG G	TA: Y FAAJ K Sali T T T AAGJ R T T AAGJ R C C ACAI B I I G G C CAC Q ACAI B I I G G G C CAC Q C ACAI T T T T T T T T T T T T T T T T T T T	E AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	V V V V V V V V V V V V V V V V V V V	L GGC G G G G G G G G G T T C ATT I C C AT I C C AT I C C AT I C C C C C C C C C C C C C C C C C C	CAP CAP	AGCA AGCA AGCA AGGT G AGCA A AGCA A ACA	GCC G G V XAAT N CTG L XAAG K GAT D CC C C C	TAGC S S S S S S S S S S S S S S S S S S S	TAT Y XAGI S S CA TC/ S TT/ L XAAG K S CAI H S CAI L MM	GAC D GG G CAA? N AAAA K TTY L SAAA K TTY F GG T GG T GG T GG T GG T GG T	FITT F IGAA E Res IAAA K K GAA E ATTT F SATC D GGAC D	ACT . GCA AITV . GE . GD . ATI . GD . AT . TL . TT.	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2460 158 2520 171 2580 20 20
ATCTI S R ACCTI R GGAN D Hpa ACCG P ACTTI L TATTI I I D C C C TATTI A TGAA E TATGAA E TATGAA K	K AAA K AAA T T L L L L L L L L L L L L L L L L	ACAY H BCAG H AGG G G IGT V V IGCI A AGA I C I C I C I C I C I C I C I C I C I	ICTIA L CTA: Y V IGTI V IGTI R IGTI R IGTI R IGTI R IGTI R I IGTI R I I I I I I I I I I I I I I I I I I	K K R C C AIGT C AIGT T C AIGT T F ACAI H AAAG K V C C C C C C C C C C C C C	E CATC I FATC M GGCJ A CGTI V TATT I FATT I FATT I FATG T CAAT N	E TTT F TGC C TTT All TAGE S TAAH K GAM K GAM K C C C C C C C C C C C C C	GGG GGG GGG GGG GGG GG GG GG V TTT F GGC A CG CG G A CG CG G G A CG CG G G A CG C G G C G G G G	TA: Y FAAJ K Sali T T AAGJ R CAG C CACA C C C C C C C C C C C C C C	E AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	V AGTG V V GTG V V TATG A C TATG X ACG T	L GGC G G G G G G G G T T C A T I C C A T I C C A C H C A T I C A C I C A C I C C C C C C C C C C C	CAN CATA CATA CATA CATA CATA CATA CATA C	UGCA A AGCA A AGCA A A GGA A A GGCA A A CACA T	GCC G G V XAAT N CTG L XAAG K GAA D C GAA D TAT Y HD C C C P	TAGC S S CCI S S TAGC S TAGC S TAGC S S S S S S S S S S S S S S S S S S S	Y XAGI S S CAGI S S S S S S S S S S S S S	GAU D GG G CAAX N AAAA K TTV F GAT D GAT D JI AAAA K	GAAA E Raba K K GAAA E ATTT F CATC C CATC C C C C C C C C C C C C C C	ACT . GCA HITV . A E . A D . AT I . A D . AT . T L . T L	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2460 158 2520 171 2580 20 2640 40
ATCTI S RESC R GGAT D HD0 ACTTI D ACTTI L TATTI I D CCT TATTI I TGAT A CCC P ACTTI L TATTI I A CCC P ACTTI L TATTI I A CCT A CCC A CCT A CCT A CCT A CCCC A CCT A CCCC A CCT A CCCC A CCT A CCCC A CCT A CCCC A CCT A CCCC A CCT A CCCC A CCT A CCCCC A CCCC A CCCCC A CCCCC A CCCC A CCCCC A CCCCC A CCCCC A CCCCC A CCCCC A CCCCC A CCCCCC	K AAA K AAA T T L L L L L L L L L L L L L L L L	ACAY H BCAG H C C C C C C C C C C C C C C C C C C	ICTI L CTA: Y V V TAGI R TAGI R N N CCTI R N N CCTI R V V V V V V V V V V V V V V V V V V	K R R R R R R R R R R R R R	E CATC I FATC M BGCA A SGCI V CATI I FATTI I FATTI I K STGG T V V CAATI N	E TTT F TGC C TTT All All All All All All All	GGGI GGGI G G G G G G G G G G G G G G G	TA: Y FAAJ K Sali T T AAGJ R CAGJ G CCAC Q ACAJ H JDII C CATI I I	E ANAAN K K GGT G G G G G G G G G G G G G G G G G G	V AGTG V AGTG V CTGTC V V CATG A CAAAA K CAAAA K CAAAA K CAAAA K CAATT I ACCG T	L GGCI G G ACI T T I ACCG P AAA K CACC H GAI D CAI H		GCA AGCA A AGCA S GCA A GCA A GCA A C GCA A C GCA A C GCA A C GCA A C GCA A C GCA A C GCA A C GCA A C C A C A	GGC G G V VAAT N CTG L CTG L CTG K GAT D TAT T TAT Y HD C C C C P	ACC T TAGC S CGC ¹ A TAI T TAI T TAI T TAI T TAI T TAI T TAI T TAI T TAI T T T T	Y CAGI S CAGI S S S S S S S S S S S S S	GAU D GGC G CAAN K N TTV F C GTT C GTT C GTT C GTT C GTT C GTT C GTT C GTT C GTT C GTT C GTT C GTT C GTT C GTT C G C C A C C C C C C C C C C C C C C C	FITT F RGAA E RAAA K WGAAA E WTTT F L VTTT F L CTTT F CATC C QGAC G	ACT GA AITV GE GAD ATI GAD AT TL TL.	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2460 158 2520 171 2580 20 2640 40
ATCTI S RE ACCTI R GGAI D HDA ACCTI L L TATTI I D CCTI I TATTI I TATTI A AAAG K GAAT	XAA K XAA XAA T T L L L L L L L L L L L L L L L L L	ACAY H BCAG H G G G G G G G G G G G G G G G G G G	ICTA: CTA: Y IGTA: Y V IGTA: Y V IGTA: Y V IGTA: Y V IGTA: Y V IGTA: Y V IGTA: Y V IGTA: Y V IGTA: Y V IGTA: Y V IGTA: Y V IGTA: Y V IGTA: N IGTA:	K K C C C C C C C C C C C C C C C C C C	E CATC I TATC M SGCA A SGT V V TATI I V TATI I V TACG T N N CTGC	E TTTF F TGC All ALL AGAN C TTTS C ALL ALL ALL ALL AGAN C C C C C C C C C C C C C	GGGI GGGI GGGI GGGI GGGI GGGI GGGI GGG	TAAN K SAAN K SAAN K SAAN K SAAN C C C C C C C C C C C C C C C C C C	E AVAAN E AVAAN K GGI G G G AGAA D HI CGGC G G TAAC N GGA D FICCO CGAA E GAA CGAA CGAA CGAA CGAA CGAA CGAA CGAA CGAA CGGC CGAA CGGC	V V V V V V V V V V V V V V V V V V V	L GGC G G ACI T I I CCC P K CAC H GAI GAI GAI		GCA A GCA A AGCA A GCA A GCA A GCA A C GCA A C C T C A C C T C A C C T C A C C C C	GCC G G V V VAAT N CTG L CTG C C C C C C C C C C C C C C C C C C	ACC T TAGC S CCT A TAT T T T T T T T T T T T T T T T	Y CAGINA SAG	GAU D G G CAAN K TTV L SAAN K TTV F GGTI GGTI GGTI GGTI GGTI GGTI GGTI GG	FITT F IGAA E ITAAA K AGAA E I I I I I I I I I I I I I I I I I I	ACT GAILTV GAE GAD STI SAD STI STI STI SGG	2100 38 2160 58 2220 78 2280 98 2340 118 2400 138 2460 158 2520 171 2580 20 2640 40 27000

SSDI • Dral •

quence of deletion of a 1.2-kb *Cla*I fragment encompassing a second *Hind*III restriction site of pMK4. The pUC9 polylinker of pER2058 was replaced with that from pUC19 to generate pER2108. Plasmid templates were purified by alkaline lysis (4) followed by alkaline denaturation. Both strands from the 5.1-kb *Bam*HI insert in pYA501 were determined by the dideoxy chain termination method (39). Synthetic oligonucleotide primers were obtained from the University of Kansas Biochemical Research Service Laboratory. Nucleotide sequence accession number. The sequences de-

FIG. 3. Nucleotide sequence of the S. mutans insert in pYA501 and deduced amino acid sequences of ORF X and lacRABCDFE. Sequences capable of forming a stable stem-loop structure ($\Delta G^0 = -17.4$) immediately upstream of the lacR promoter are overlined. An 11-bp sequence common to the L. lactis promoter region (48) is boxed, and RBS sequences and consensus promoter regions are indicated by double underlines. The direct and inverted repeat elements near the lac promoter are denoted by arrows and arrowheads, respectively.

TGGGAAACTTGGTGAATTTTTTAGAAGCGGAACTTGATAAGTCTGCCATAAAGCATTCTTT 2760 G K L G E F L E A E L D K S A I K H S F PStI TTATAAGATTT<u>CTGCAG</u>AGACAAGAAATTGTATTGCTATTTACATGGTGGCTATCAAAC 2820 YKISAETRNCIAILHGGYQT 100 SSPI · AVAII · Hinfi AGAAATATTAGAACAAGGACCTTATGTTTCCGCTAAAGAATCTAAAGGGTTTCTTGAATT 2880 E I L E Q G P Y V S A K E S K G F L E F 120 TTTTGAAAAATTACTTCCAAAATTAGAAGTTGTCGCAATTTCAGGAAGTCTTCCAAAAGG 2940 F E K L L P K L E V V A I S G S L P K G 140 V P V D Y Y S Q M I A I C K Q H Q V P I 160 Alui TGTTTTGGATTGTTCAGGTCAGGCTTTGTTGGAAGTGTTAAACGGTGC<u>AGCT</u>AAACCAAC 3060 V L D C S G Q A L L E V L N G A A K P T 180 TGTCATCAAGCCCAATACAGAAGAATTATCTCAAATTATGGAACGGGAGATTACAAATGA 3120 IKPNTEELSQIMEREITND 200 Drai Sphi. TGTTGCTGT<u>TTTAAAGCATGC</u>TTTGGCTAGCCCTATCTTTTCAGGAATTGATTGGATTAT 3180 V A V L K H A L A S P I F S G I D W I I 220 TGTCTCACTTGGTTCTCAAGGTGCTTTTGCCAAGCATGGTCAAACATTTTATAAGGTCAC 3240 V S L G S Q G A F A K H G Q T F Y K V T 240 Hinfi Alui TATTCCTAAAATAGCAGTCGTTAATCCAGTTGGTTCAGGG<u>GATTC</u>AACGGT<u>AGCT</u>GGAAT 3300 I P K I A V V N P V G S G D S T V A G I 260 PStI TACATCGGCTCTTG<u>CTGCAG</u>GAGCAAGTGATGAGAAATTGCTCAAAAAAGCGAATACACT 3360 T S A L A A G A S D E K L L K K A N T L 280 HindII Ddei • Hpai Ddei • Hpai Adei • 3420 G M L N A Q E K L T G H V N L E N Y D N 300 MnllDral MMII EBS TTTATACCAACAAATIGAGGTAGCGAGGTTTAAAAAGATGACATTAACACAAGAAAAGC 3480 L Y Q Q I E V A E V \$ <u>M T L</u> T Q E K 310 HhaI DdeI XmnI AlUI AlUI GCAGTTATATGGAAAAAACTTAGTGAATGGAATTATTTAGGCTTTAGCTTTTGACC 3540 R S Y M E K L S D E N G I I S À L À F D 27 •Drai Hhai AGCGTGGTGC<u>TTTAAA</u>ACGCTTGATG<u>GCGC</u>AGTATCAAACGCAAGAACCAACGATTGCTC 3600 Q R G A L K R L M A Q Y Q T Q E P T I A Alui Ddei AAATGGAAGAGCTGAAGGT<u>CTTAG</u>TAGCAGAAGAATTAACACCTTATGCTTCATCCATGC 3660 Q M E E L K V L V A E E L T P Y A S S M 67 TGCTTGATCCAGAATATGGTCTTCCAGCAGCAAAACATTTGGATAAAAATGCAGGTTTGC 3720 LLDPEYGLPAAKHLDKNAGL 87 TCCTTGCTTATGAGAAGACTGGTTATGATACAACAAGCACTAAACGCTTGCCAGATTGTC 3780 L L A Y E K T G Y D T T S T K R L P D C 107 TGGTTGAATGGTCAGCCAAACGTTTGAAAAAACAAGGTGCAGATGCTGTTAAATTCTTGC 3840 L V E W S A K R L K K Q G A D A V K F L 127 нра I Нілоді I ТСТАСТАТБАТБТТБАТББТБАТБАБСАБСАЛАЛАСАВССТТАТАТТБААС 3900 LYYDVDGDEEVNQQKQAYIE 147 Xmri • • Alui • GAATTGGTTCTGAATGTAAGGCA<u>GAAGATATTC</u>CCTTTTTCCTTGAAATTCT<u>AGCT</u>TATG 3960 RIGSECKAEDIPFFLEILAY 167 Mn11 ACGAAACCATTACTGATGCCGCAAGCGTTGAGTATGCTAAAGTAAAGCCTCATAAAGTGC D E T I T D A A S V E Y A K V K P H K V 4020

	~ > m	000	• X	mni	200	D	del	230	220	•	mmm	000	• •	יחארי	~mm	~		200	MnlI	4090
L	D	A	M	K	v	F	S	D	E	R	F	G	I	D	v	L	K	v	E	207
		Hpa Hin	I ITT				Ta	σī		•			Hae	ттт					•	
TT	CCT	GTT	AAC	ATG.	AAG	TAT	GTO	<u>S</u> AA	GGT	TTT	GGT	GAT	GGC	CA	ATT	JTT	CAC	ACT	CAAG	4140
v	P	v	N	м	K	¥.	v	Е	G	F	G	D	G	P	I	v	н	т	Q	227
MЬ	οI		•			Dra	I			Mbo	r		•			•			•	
AT	CAA	GCA	GCA	AAC	TTC	TTT	AAA	CAA	CAA	GAT	CAA	GCA	ACA	CCG	CTT	CCC	TAT	ATT	TATT	4200
D	Q	A	A	N	F	F	ĸ	Q	Q	D	Q	A	т	P	L	P	Y	I	Y	247
			•	_		Dd	eI			•			•			•		Hin	fI ·	
TG	AGT	GCA	GI	GTT	ren	GCD	AAG:	TTA:	TTC	CAN	GAT.	ACG	CFR	STT.	PTR	JCA.		JAG	CAG	4260
ь	5	A	G	v	5	A	r	Ц	r	Q	υ	T	ь	v	F.	A	r	Б	3	207
~~	~~~		•		~~~~	-	~~~~~	1 000	~~~	•			·	~~~~~	~~~		-			4220
GG	SCT:	NI N	F	N	G	v	T.	C	GGA	R	۵CC. ک	ACA T	W	A L	G	S S	v	K	D	287
9	ĥ		F	14	9	•	-		9	n	ĥ	•		^	9	Ο.		•		207
እጥ	እጥጥ			aan	222	aci.	CA	300	ന്ദറ	• ^}@	TCC	C TITA	RS	BI ACA		H1. YCN	nfI TTC	AAG	SSPI AATA	4380
Y	I	E	K	G	E	λ	λ	A	R	0	W	L	R	T	E	G	F	K	N	307
-	-					-						_		-	-	-				
TT	GAT	GAG	CTG	аат.	AAG	GTT	TTA	AAA	GCA	ACO	GCA	ACG	ICT.	rggo	JAAG	JAA	AGA'	TAG	AAAG	4440
ī	D	E	L	N	ĸ	v	L	ĸ	λ	т	A	т	S	W	Е	Е	R			325
									н	inf	т									
GT	GGA	AAA	JAG	ATG	AAC	AGA	GAA	GAN	cci	ACT	ĊТТ	TTG	GGG	FTR	GAA	ATT	GTA	3CC	TATG	4500
-				M	N	R	Е	E	A	Т	L	L	G	F	Е	I	v	A	Y	16
				18	CI	2			Hi	ndI	II					Hр	aII		_	
CN	001	CATY	202	021	TC-TT		TT 34	אזיידיי	A 445	luI	באדית	חגג	• •	2020	H NG	aeI	II	R	sal•	4560
2	G	חס	A A	R	s	K	L	L	E	A	L	N	A	A	0	A	G	E	Y	36
••	-	-	••					-	-		-			••	•		-			
AT.	AGG	GCA	GAAG	A1) GAG	UI CTT	A. JTA	JUI GCT	CGG	GCA	• GAT	GAC	TGT	ATT	TTC	JAT	scc	CAT	A AAA	GCTC	4620
D	R	A	Е	E	L	v	λ	A	A	D	D	с	I	v	D	A	Н	ĸ	A	56
					~	лт .									-	. .				
AA	ACA	AGC	CTT	TTG	GCT	AAG	зала	3CA(CAN	GGC	GAT	GAT	ATT	JAA	TT.	AGT	GTT	ACT	TTAA	4680
Q	т	s	L	L	A	K	Е	A	Q	G	D	D	I	Е	L	s	v	т	L	76
			•	Mho	т															
TG	CAO	GGA	CAA	GAT	CAC	CTG	ATG	ACA	ACC.	ATA	CTT	CTA	AAA	3AT'	rTG/	ATG.	AAA	CAT	TTAA	4740
M	H	G	Q	D	H	Ļ,	M	т	т	I	L	L	K	D	L	M	ĸ	H	L	96
		R	saI		nn. r	bs•/	Alu:	I		•			•			•			•	
TT	GAA	TTG	FAC	AAA	AGA	GGG	AGC	[AA	AAT	AAT.	ATG.	AAC	ACA!	FTG/	TT	3CT	CAA	ATT	GAAA	4800
I	Е	L	Y	ĸ	R	G	s				M	N	T	L	I	λ	Q	I	Е	104
											18	C	3							9
	~~`								Fn	UDI	I <i>S</i>	spl						~~~~		4960
AG	GGA	AAG	DCP.	FIC:	P	JAAJ P	R	T	s	D D	N	T	V TAT	L.	D	ACT:	A1-14	B	D	4000
r	9	r	£			Ĩ.,		•	5	n	14	-	-		n	^	÷.,		5	
~~		2000	ŵ	203	i Marcene	Hpa:	II 2TC2	۵ ۳ ΓΥ	- د د חר	•	Tag	I	• •	עדידי	-	M	bol	201	• •	4920
G	F	T I	s	یمی ۸	M	P	v	I	L	F	S	S	T	F	L.	L	I	A	Y	49
Ū	-	-	-			-	<u></u> 1	Bal	I 	-	-	-	-	•	-	-	-		-	
тт	CCT	AAT	ATC:	FFR	GT	· FTT/	ACT	Ha (CCA	AAA	GT	ATT	3ÅAJ	AC	TG'	rTG/	ATG		CCCT	4980
v	P	N	I	F	G	F	т	W	P	ĸ	G	I	Е	N	M	L	M	т	P	69
													Des	ьт						
AT.	AAC	TAT	٠ مُن	ATG	GGG	ATT	ATT	GT.	FTC	CTA	JTG	JCT	GT	ŝάλ	CT	GCT	AAA!	ICA(CTAA	5040
Y	N	Y	т	M	G	I	I	G	F	L	v	A	G	т	т	A	K	s	L	89
		Taqî f	۲ <u>.</u>)							
cod	GAT	ŤĊGJ	TG	AAC	caco	CAAS	FTA	3AA	AAG	ACC.	AAC	CAG	ATT/	AT	TC	ATT.	ICA	ACA	ATGT	5100
т	D	S	М	N	R	Q	L	Е	ĸ	т	N	Q	I	N	F	I	s	т	м	109
	• · · ·			Hp	II						B	amH	I							
A. TA	IUI GCT	TCT	• <i>Ha</i> ATGO		LL GT	rrc.	TA	ATC	ATG	• GCA) 3CG	SAT(5140
L	λ	s	M	λ	G	F	L	I	M	λ	A	D	P							123

FIG. 3-Continued.

scribed in this work have been assigned GenBank accession number M80797.

RESULTS AND DISCUSSION

Hybridization of pYA501 to *S. aureus lac* DNA. Previous studies in which DNA from *S. mutans* was cloned into an *E. coli galKTE* mutant (43) demonstrated that the streptococcal DNA contained in pYA501 encoded enzymes of the tagatose 6-phosphate rather than the Leloir pathway. We have recently characterized the genes of the tagatose 6-phosphate

pathway from S. aureus (35) and found that they encode proteins very similar in size to those predicted by minicell analysis of the streptococcal tagatose 6-phosphate gene cluster (24). Thus, we reasoned that both organisms may possess closely related genetic determinants encoding the enzymes of the tagatose 6-phosphate pathway. To test this hypothesis, restriction digests of total DNA from S. aureus and S. mutans were tested for hybridization to the S. mutans insert-containing plasmid pYA501 probe under conditions of low stringency (Fig. 1A). In addition, when the

Ribosome Binding Site	Spacer	ΔG°
UCUUU CUCCA UA GGAAA GAGGU-AUUU AUG A LacR	7	-14.8
uuccuc aaggaga aug acu aug LacA	2,8	-12.8
UUUC-CUCC··· UAAAGACAAAUAGAG GAGGAA AUG AAA AUG LacA U LacB	4,10	-18.0
U U UC UUCC CC··· UGAG AAGG-GGAGUUAUCAUU AUGAUG LacB U LacC	12,15	-15.6
CCUCCA GGAGGUU <u>UAA</u> AAAG AUG <i>Lac</i> C LacD	11	-16.6
UCUUUCC CC···· UAGAAAGG GGAAAAGAG AUG LacD ^U LacF	9	-17.0
UUUCCUC···· GAGGGAGC <u>UAA</u> AAUAAU AUG	11	-9.4

FIG. 4. RBS sequences of the *lac* operon. The sequences corresponding to the *lacRABCDFE* mRNA immediately upstream of the initiation codon are presented below the complementary sequence from the 3' terminus of the *B. subtilis* 16S rRNA (written $3' \rightarrow 5'$). Termination codons of the preceding ORF at each junction are underlined, and probable initiation codons are indicated in boldface and aligned with a box designating the downstream ORF. Calculation of the predicted free energy of base pairing (ΔG^0) of the RBSs with the 16S rRNA and determination of spacer distances (in nucleotides) were done as described in the text.

5-kb streptococcal insert of pYA501 was gel purified and used to probe pBO4 (33), which contains cloned *S. aureus* lactose genes, bands known to contain the *lacABCD* determinants were detected (Fig. 1B). The detection of hybridizing restriction fragments of predicted sizes from the *S. aureus* chromosome (and pBO4) suggested that there was considerable DNA homology (>60% under the conditions of stringency employed) between the *lac* regions of these organisms.

Nucleotide sequence analysis. The nucleotide sequence of the 5.1-kb BamHI insert in pYA501 was determined. Examination of the nucleotide sequence revealed seven open reading frames (ORFs) which, based on homology with the previously characterized S. aureus lac operon (10, 34–36), encode LacR, LacA, LacB, LacC, LacD, LacF, and the NH₂-terminal 123 amino acids of LacE. The lac genes from

both organisms are directly analogous, with the lactose operons of *S. mutans* and *S. aureus* being superimposable, as depicted by their respective physical maps (Fig. 2). The nucleotide and deduced amino acid sequences of the *S. mutans lac* region are presented in Fig. 3.

Several interesting features of the streptococcal *lac* region may be illuminated here. First, there is a substantial 81amino-acid ORF at the 5'-proximal end of the 5.1-kb BamHI fragment. This is presumably a truncated ORF that normally extends beyond the BamHI insert terminus of pYA501. The function of this ORF is unknown, and it bears no homology to known lac genes. Thus, we suggest that it is likely to be unrelated to lactose metabolism. The lack determinant extends from the initiation codon ATG (position 472) to the ocher (TAA) codon (position 1224). This ORF encodes a 251-amino-acid polypeptide with a calculated molecular mass of 28,713 Da, a value very similar to that calculated for the lacR gene product from S. aureus. LacR, which functions as the repressor of the staphylococcal lactose operon, has a relative molecular mass of 28,534 Da (34). The lacA and lacB ORFs (positions 1534 to 1959 and 1988 to 2500, respectively) encode the subunits of galactose-6-phosphate isomerase. The DNA sequence indicates that the lacA and lacB genes would encode proteins of 142 (M_r 15,610) and 171 $(M_r 18,950)$ amino acids, respectively. The LacA and LacB proteins from S. aureus are identical in length and nearly identical in size $(M_r, 15, 425 \text{ and } M_r, 18, 953, \text{ respectively})$ (35) to the analogous proteins from S. mutans.

The ORF corresponding to *lacC* extends from an ATG initiation codon (position 2522) to the ocher codon (position 3451), encoding a 310-amino-acid protein (M_r 33,368) directly analogous to LacC (M_r 33,856) of *S. aureus* (36). The latter polypeptide demonstrates tagatose-6-phosphate kinase activity in *S. aureus* (35). The *lacD* ORF (positions 3459 to 4433) encodes a 325-amino-acid polypeptide (M_r 36,495) which is one residue shorter than its counterpart (M_r 36,599) from *S. aureus* (36). Staphylococcal LacD demonstrates tagatose-1,6-bisphosphate aldolase activity (35).

Immediately following the genes for galactose 6-phosphate catabolism are the genes lacF and lacE, which specify the phosphoenolpyruvate-dependent PTS transport components EIII^{Lac} and EII^{Lac}, respectively. The lacF ORF (positions 4452 to 4763) specifies a 104-amino-acid protein (M, 11, 401)which is one residue larger than EIII^{Lac} $(M_r, 11, 372)$ from S. aureus (10). The final streptococcal ORF was determined to be truncated at the BamHI site delineating the 3' end of the insert in pYA501. This ORF extends from the initiation ATG codon (position 4773) through the BamHI site (position 5140), encoding the NH₂-terminal 123 amino acids of EII^{Lac}. Although the entire ORF for *lacE* has not been determined, there is extensive homology to S. aureus lacE (EII^{Lac}) at both the protein (83% identity) and DNA (74%) levels for that portion which has been characterized. Although there is a lack of supporting genetic evidence, the organization of the lac genes discussed above strongly predicts that the gene encoding phospho- β -galactosidase (lacG) of S. mutans is located directly downstream of lacE, as in S. aureus.

Translation of the *S. mutans lac* genes. The initiation codons for each of the *lac* genes are preceded by a putative ribosome-binding site (RBS) sequence (Fig. 4). The free energies for association of the *lac* mRNA with the 3' end of *Bacillus subtilis* 16S rRNA (19) range from -9.4 to -18.0 kcal/mol, as determined by best-fit analysis by the rules of Tinoco et al. (45). The high affinity for *B. subtilis* 16S rRNA exhibited by the *lac* genes is similar to that demonstrated by

the *S. aureus lac* RBS sequences (32) and is within the range of values for the more stringent RBSs reported for grampositive genes (30).

The distance between the RBS and the downstream initiation codon was calculated by measuring the first base to the right of the preferred E. coli RBS sequence, AGGA (or its equivalent), through the base adjacent to the initiation codon (30). In cases of multiple possible initiation codons (i.e., lacABC), assignment of the probable translation initiation site was based on both optimal spacing and alignment with the homologous staphylococcal proteins. For the seven lac ORFs, the spacer distance was generally 8 to 12 nucleotides, a distance also observed for the *lac* genes from S. aureus (32) and for genes of B. subtilis (30). The B. subtilis 16S rRNA sequence was used in these calculations because neither the S. mutans nor the S. aureus sequence has been reported. The high similarity between the S. mutans and S. aureus lac RBS sequences suggests that the rRNA sequences from both organisms as well as those from B. subtilis are similar. The demonstration that numerous staphylococcal genes are efficiently expressed in B. subtilis (23) provides biological support for this notion.

A strong bias toward A- or U-rich codons in S. aureus has been well established (27, 32, 40) and is thought to reflect the low G+C content of this organism. However, it is known that there is a strong bias against certain codons (e.g., AUA for Ile) in some staphylococcal genes despite their A- or U-rich content (32, 46). This bias is quite apparent for the *lac* determinants from both S. aureus and S. mutans (data not shown). The finding that the codon usage patterns and RBS sequences for the genes of the *lac* operons are very similar suggests that the translational apparati of these organisms are closely related.

The high level of similarity between the S. mutans and S. aureus lactose operons is summarized in Table 1. The G+C content of the S. mutans lac region is 36.6%, while that of the S. aureus lac region is 34.4%. Direct amino acid homology for the LacRABCDFE proteins ranges from 63 to 83%, while the region exhibits 64 to 74% DNA identity. However, if only the first two positions of each codon are compared, the DNA identity increases to 75 to 89% for the seven lac ORFs. Thus, there appears to have been sufficient genetic drift to allow optimization for codon preferences in both organisms.

It is of interest that the nucleotide sequence similarity between these organisms decreases significantly upstream of the *lacR* determinant. Such evidence implies that the lactose operon has become inserted in one or the other organism at a nonhomologous chromosomal site. The presence of a putative stable stem-loop structure 5' of *lacR* (Fig. 3) suggests that the *S. mutans lac* region may have arisen through acquisition of the genes from *S. aureus* or another common ancestor, with the insertion occurring within a gene set containing associated potential RBS sequences.

Homogramic conservation of the lactose proteins. Recent studies on lactose metabolism in *L. lactis* MG1820 have demonstrated that the functions of the lactococcal proteins are directly analogous to those of the *lac* gene products from *S. mutans* and *S. aureus* (50). The *lac* operon of *L. lactis* is arranged identically to the operons of *S. mutans* and *S. aureus* (Fig. 2) except that *lacR* is transcribed divergently from the remaining *lac* determinants and an additional ORF (*lacX*) is located downstream of *lacG* in *L. lactis* (48). The *L. lactis lacRABCDFE* proteins (15, 48, 50) are very similar in length and predicted molecular mass to those from *S. mutans* and *S. aureus*. The molecular masses of analogous

proteins from all three organisms vary by less than ± 610 Da. As shown in Fig. 5 and Table 2, the Lac proteins from S. mutans, S. aureus, and L. lactis exhibit extensive direct homology. Furthermore, the PTS transport components LacF (EIII^{Lac}) and LacE (EII^{Lac}), which have recently been characterized in Lactobacillus casei (1, 2), are homologous in these four organisms, albeit at a distinctly lower level of relatedness. The percent identities for pairwise comparisons of LacRABCDFE from S. aureus, S. mutans, and L. lactis and for L. casei LacFE are presented in Table 2. Although the transcriptional organization of the lacEGF genes is different in L. casei, they appear to be part of an operon containing at least two additional upstream ORFs (2). The sizes of these ORFs and their encoded products have not yet been reported, and thus it is not known whether these determinants may function to catabolize galactose 6-phosphate, as does the *lacABCD* gene cluster, or regulate expression of the *lac* operon, as does LacR, in the above grampositive cocci.

LacC and LacR are similar to phosphokinases and DeoR from E. coli. In addition to the homology discussed above, the LacR and LacC proteins also show significant homology to repressors and kinases from E. coli (Fig. 5). LacR from S. mutans, S. aureus, and L. lactis has homology ($\approx 25\%$ identity) with the repressor (DeoR) of the E. coli deoxyribonucleotide operon (47) as well as with several other repressor proteins, including GutR (48), FucR (48), and GlpR (26) from E. coli and AccR (51) from Agrobacterium tumefaciens. LacC has approximately 25% direct similarity with 1-phosphofructokinase (FruK) (31) and the minor 6-phosphofructokinase (PfkB) (14) from E. coli. In addition, LacC has 25% identity with the 316-amino-acid 1-phosphofructokinase encoded by fruK from Rhodobacter capsulatus (52).

It is of interest that both LacR from S. aureus (34) and L. lactis (48) and DeoR are transcriptional repressors which bind inducing species that are phosphorylated compounds (galactose 6-phosphate or deoxyribose 5-phosphate). One region of substantial homology exists between residues 20 and 42 of DeoR, which forms the helix-turn-helix domain of this DNA-binding protein (16). The latter half of this motif is especially well conserved in all of the above LacR-related repressor molecules. Although transcriptional repression by S. mutans LacR has not been tested, the high overall identity to S. aureus LacR (63%), more specifically to the highly conserved helix-turn-helix domain of these repressors, strongly predicts that expression of the streptococcal lac operon is controlled transcriptionally by the product of lacR.

Unique COOH-terminal amphipathic helix encoded by lacF. The carboxyl-terminal 16-amino-acid segment of S. aureus EIII^{Lac} (LacF) has been shown by Saier et al. (38) to constitute an amphipathic helix which is not present in other PTS proteins. A helical wheel projection of this region was shown to position predominantly basic residues in the entirely charged seven-residue hydrophilic half of the helix, while the hydrophobic half comprised mostly leucyl residues. Interestingly, this region is very highly conserved in EIII^{Lac} from S. mutans (14 of 17 identical residues) and is unaltered in the charged hydrophilic face of the helix. Strong genetic evidence (41) coupled with the predominance of basic amino acids within the COOH-terminal amphipathic helical sequence has led to the suggestion that this region facilitates binding of EIII^{Lac} to the negatively charged backbone of DNA, thereby mediating transcriptional activation of *lac* operon expression in *S. aureus* (38). It is not known whether EIII^{Lac} from *S. mutans* also performs a similar function. In light of the above evidence, it is perplexing that

ir:		
87	[2 D
L	м	

H helix-turn-helix		
Sa MNKHERLDEIAKLWNKKGTIRINEIVECLNVSDMTVRRDLHELENKGULTK	IHGGARSNSI	61
Sm MKKEERLEEIIKLINKEGTIRVIPWERLKVSDMTVRRDLHELEGIGVLTR	IHGGARSNNI	61
LI MKESLHMNKKRRLEKILDMLKIIGTIFIKEIIDELDISDMTARRDLDALEAIGULTR	I <mark>HGGADLLS</mark> S	67
Ec DeoR METRREERIGQULDELKRSDKLHUKDAAAILGVSEMTHRRDLNHSAPVVLLG	GYIVLEPRSA	63
Sa FQYKEIISHKEKHIIRQIAEKRHIAKKAASLIEIGDTLFFGPGTTVELLAEEMNH-H	TLTIITNCLP	125
Sm FQYKEMSHEEKHSRQIEEKHYIAQKAAEIMEGDTHFLGPGTTVELLAEEINK-T	TLOVITNCLP	125
LI KKELEKTHHEKKELNTKEKIIIAKKAQSLILKGRKGYKIRVIINGLPVFLKDG-D	THIGPGTTL	131
Ec DeoR SHVL-19DQKSRLVEEKRRAAKLAAILVEPDQTLFFDCGTTFPWIIEAIDNEI	PHIAVCYSLN	125
Sa VYKIILLEKOTAHFRVMLIGGEMRHITEAFMGEMANAMLEKLEFSKM-FFSSNAVN	KGAVMTS-LLL	188
Sm VFGILSGKOSETIFRVHLIGGEMRSITOSFIGETINIVLEKMHFSKM-FFSGNGVK	GNEVMTS-SF	188
LI VGLALETGAFVGSMASTNLKAMRFAKILNDSETIDLILLGGEYREIAFVRANAVF	HNSTATM <u>-SD</u>	195
Ec DeoR TFLALKEKFHCRAFLGGEFHASNAIFKPIDFQQTLINNFCPD-IAFMSAAGVH	VSKCATCFNL	187
Sa LEAYTQOLALSNSTEKYLLIDHTKNGKEDFTSFODINHLTAVMD-YE-DEHKVE	TIKTY IEVVD	251
Sm CEAYTOKMALGRAIEKYFLIDGSKUGKEDFTSFYDISOLTALITDCO-DDNUCK	LSKYTEIIN	251
LI KEGVIQOLALNNAMEKFLIMDSTKFDRYDFFNFYNIDOLDTILTINDISPOHLEE	ESCYTIILKAD	261
Ec DeoR HELPVKHWAMSMAQKHVLMVDHSKFGKVRPARMGDLKRFDIVVSDCC-PEDEMVK	YACTORIKLMY	252
Lac A		
Sa MAIIIGSDHAGKRLRHVIRSYLLDNKYLVVDVTEGQEVDFVDATLAVAKUVQSCHGN Sm MAIIIGSDAAGKRLKDVIRIFIKDNNHEVLDVTERKDLDFVDSTLAVVHEVQRNDKN LI MAIVVGAULRGTRLKDVKKNFLVEEGFEVIDVIKDGQ-DFVDVTLAVASEVNKDHQN	LGIVIDA FCII GSFI LGIAIDAYGAGSFI LGIVIDAYGACAFI	91 91 91 91 91 91
Sa VATKIKGMIAAEVSDERGGYMTRGHNNGRMITMGGEIVGDULAKNVVKGFVEGKYDG Sm VATKVKGMIAAEVSDERSAYMTRGHNNAFUITLGAEIVGDELAKNIVKDFVEAKYDG LI VATKIKGMVAAEVSDERSAYMTRGHNNARMITVGAEIVGDELAKNIAKAFVNGKYDG	GRHQIRVDMLNKM GRHQIRVDMLNKM GRHQ <mark>V</mark> RVDMLNKM	142 142 142 142
LacB		
Sa MKIALGCDHIVTDIKMRVSEFLKSKCHEVIDVGTYDFTRTHYPIFGKKVGEOVV	SGNADL	60
Sm MKIAIGCDHIVTDVKMEISKHLKEEGYEVLDVGTYDFTRTHYPIFGKKVGEAVG	SGFADL	60
Ll MFIAIGCDHIVTDVKMAVSEFLKSKGYEVLDFGTYDHVRTHYPIYGKKVGEAVV	SGQADL	60
Sa GVCICGTGVGINNAVNKVPGVRSALVRDMTSALYAKEELNANVIGFGGRIIGEL	IMCDII	120
Sm GVCMCGTGVGIGNAANKVPGVRIALVRDMTSALYGKEELNANVVGFGGAIIGKL	ILFDIV	120
Ll GVCICGTGVGINNAVNKVPGVRSALVRDMTSALYAKEELNANVIGFGGMINGCL	IMNDII	120
Sa DAFINAEYKPTEENKKLIAKIKHLEISNALQADPHFFDEFLEKWDRGEYHD Sm DAFIKAQYKPTEENKKLIAKIKHLEAHNDKOADPHFFDEFLEKWDRGEYHD Ll FAFIFAEYKPTEENKKLIAKIEHMETHNAHQALEFFFIEFLEKWDRGEYHD		171 171 171

FIG. 5. Homology between the S. mutans, S. aureus, L. lactis, and L. casei LacRABCDFE proteins and similarities with DeoR and phosphofructokinases (PfkB and FruK) from E. coli. The deduced amino acid sequences of the lac-related genes from S. mutans (Sm), S. aureus (Sa), L. lactis (Ll), and L. casei (Lc) are aligned to maximize identity. Alignments of E. coli (Ec) DeoR with LacR and of PfkB and FruK with LacC are also presented. Identical amino acids are boxed. Percent identities for pairwise comparisons are summarized in Table 2. Only the NH₂-terminal region of LacE, as specified by the streptococcal insert in pYA501, was included in this analysis.

LacC

Sa Sm Ll Ec FruK Ec PfkB	MILTITINPSVDISYPITALKIDDVNRVCEVSKTAGGKGLNVTRVLAQVGEPULASGEIGGE MALTAMENPSTDIAYQUDDLKVDTVNRVTETHKTEGGKGLNVTRVLSQLGDVUASGULGGK MILTATINPSVDISYPIETLKIDTVNRVKDVSKTAGGKGLNVTRVLYESGUKVTATGFLGGK MSRRVATUTLNFAYDLVGFCPETERGEVNUVKTTGLHAAGKGUNVAKVLKULGUDVTVGGFLGKD MVRTYTLTIAPSTDSATITEQUYPEENCAVPHRCSNEGG-GUNVARATAHLGGSATAIFPAGGA	62 62 65 63
Sa	LQCFIAKKLDHADIKHAFYNIKGETRNCIAILHEGQQTEILEQGPEIDNCEAAGFIKHFEQLU	125
Sm	LGEFIFFAELDKSAIKHSFYKISAETRNCIAILHGGYQTEILEQGPYVSAKESKGFLEFFEKLU	125
Ll	IGEFIESFIFEQSFVSFAFYKISGNTRNCIAILHEGNOTEILEQGPIIISHEEAEGFIDHYSNUI	125
Ec FruK	NODGFQQIFSELGIANREQVVQGRTFINVKL-TEKDGEVIDFNVSGFEVTPADWEREVTDSISWU	129
Ec PfkB	IGEHIVSLUADENVPVATVEAKDWTRONLHVHVEASGEQYRFVMFGAALNEDEFRQLEEQVIE-I	127
Sa	ERVEAVAISGSLPKGINGDYMAGIIEROCNKGVPVILDCSGAILGTVLENPMKPTVIKPNISELY	190
Sm	PRIEVVAISGSLPKGVPVDYMSOMIAICKOHOVHIVLDCSGOAILEVINGAAKPTVIKPNIEELS	190
Ll	KOSEVVIISGSLPSGIPNDYMERLIGIASDEGVAVVLDCSGAPIETVLRSSAKPTAIKPNNEELS	190
Ec FruK	GOFDMYCVSGSLPSGVSHPAFIDWMTRLRSOCPOIIFISSREAIVAGLKAA-PW-LVKPNRRELE	192
Ec PfkB	ESGAILVISGSLPPGVKIEKITOLISIIRKNKGSAASSTVIGOGISAALAIG-NIELVKPNOKELS	191
Sa	OLLNOPILESLEGIKGAVSCPLFEGIEWIIVSLGAQGAFAKHNHTFYEVNIPIIISVLNPVGSGDS	255
Sm	OIMEREIITNDVAVLKHALASEIFSGIDWIIVSLGSOGAFAKHOOTFYKVIIPKIAVVNPVGSGDS	255
Ll	OLLCKEVTKDIEELKDVLKESLFSGIEWIVVSLGRNGAFAKHODVFYKVDIPDIEVVNPVGSGDS	255
Ec FruK	IWAGRKIEEMKDVIEAAHALREOOIA-HVVISLGAEGATWVNASGEWIAKPPSVDVVSTVOAGDS	256
Ec PfkB	ALVNRELTOPDDVFKAAQEIVNSGKAKRVVVSLGFOGALGVDSENCIQVVEPALKSQSTVOAGDR	256
Sa	TVAGITSAIIINHENDHDLLKKANTLGMLNAQEACTGYVNINNYDDIFNQIEVLEV	310
Sm	TVAGITSALAAGASDEKLLKKANTLGMLNAQEKLTGHVNIENYDNLYQQIEVAEV	310
Ll	TVAGIASALNSKKSDADLLKHAMTLGMLNAQETMTGHVNMTNYETIINSQIGVKEV	310
Ec FruK	MVGGLIYGLIIMRESSEHTILFILAIAVAALAVSQSNVGITDRPQLAAMMARVDLQPFN	312
Ec PfkB	LVGAMTLKLAENFSLEEMVRFGVAAGSAATLNQGTRLCSHDDTQKIYAYLSR	308
LacD		
Sa	MSKSNONIASIFQLSNNEGIISALAFDORGALKRMMAKHOIHEPTVAQIHOLKVLVAEELTOYAS	65
Sm	MILTOEKRSYMEKLSDENGIISALAFDORGALKRLMAQYOTOEPTIAOMEELKVLVAEELTEYAS	65
Ll	MVLTEOKRKSLEKLSDKNORISALAFDORGALKRLMAQYODIEPTVAOMEELKVLVADELTKYAS	65
Sa	SULLDPEYGLPASIARNKDOGLLLAYEKTGYDWNAKORLPDCLVEWSAKRLKEOGANAVKFLLYY	130
Sm	SMLLDPEYGLPAAKHLDKNAGLLLAYEKTGYDTISTKRLPDCLVEWSAKRLKKOGADAVKFLLYY	130
Ll	SMLLDPEYGLPATKALDKEAGLLLAFEKTGYDTSSTKRLPDCLDVWSAKFIKEOGADAVKFLLYY	130
Sa	DVDDAEHINIOKKAYIERIGSECVAEDIPFFLEMIIYIDNIFINGSVEFAKVKPRKVNEAMKLIFS	195
Sm	DVDGDEFVNQQKQAYIERIGSECKAEDIPFFLEILAYDEHITDAASVEYAKVKPHKVLDAMKVFS	195
Ll	DVDSSDEINQQKQAYIERVGSECVAEDIPFFLEILAYDEHISDAGSVEYAKVKPRKVIEAMKVFS	195
Sa	EPRFNMDVLKVEVPVNMKYVEGFALGEVVYTREEAAOHFKLODAATHLPYIYLSAGVSAELFOET	260
Sm	DERFGIDVLKVEVPVNMKYVEGFGDCPIVHIODCAANFFKCODCATHLPYIYLSAGVSAKLFCDT	260
Ll	DPRFNIDVLKVEVPVNVKYVEGFADGEVVYSRAEAADFFKACEFATNLPYIYLSAGVSAKLFOET	260
Sa	ĨĹŀĸſŦĂĦĔġġġaĸĸſŊġvlcgrat <mark>wġġġ</mark> wowyiecgedaarewlrttgfkniillinkvlkutatswkoRk	326
Sm	LvſŗġĸesgġŊſŊġvlcgratwagsvkdyiekgeaaarġwlrtßgfknidelnkvlkatatsweer	325
Ll	LcſŗaĦſsgakfŊġvlcgratwagsveryikegekaarewlrttgffnidelnkvlvtrasfautokv	326



FIG. 5—Continued.

the sequence of the COOH termini of EIII^{Lac} from both organisms is highly homologous (14 of 17 residues identical) while the regulatory region sequences differ markedly (see below). It should be informative to test whether these proteins are interchangeable between these gram-positive hosts.

Comparison of the S. mutans and S. aureus lac promoter regions. The transcriptional start sites of the L. lactis lac operon have recently been identified (49). However, the organization of the regulatory region in this organism is clearly different from that in S. aureus and S. mutans, lacking the striking repeat elements of the staphylococcal and streptococcal promoters. The presence of several analogous control elements upstream of the lac structural genes suggests that similar mechanisms for transcriptional regulation of lac expression may exist in S. mutans and S. aureus. Both promoter regions contain a 10-bp directly repeated sequence, the half-sites of which are centered about the -30and -80 regions of the lac promoter (Fig. 6). Although these elements are not conserved at the nucleotide level, their length and spacing relative to the proposed -35 and -10hexanucleotide promoter elements are very similar. In *S. aureus*, there exists a 12-bp perfect inverted repeat sequence centered about the -24 and -60 regions of the *lac* promoter. An analogous 11-bp sequence is also found in *S. mutans*, although the half-sites of this inverted repeat are centered about the -55 and -80 regions. Thus, it appears that this element has been shifted approximately 20 bp upstream and positioned 13 bp closer together in *S. mutans* than in *S. aureus*.

Through gene fusion and deletion analysis of the staphylococcal promoter, it was demonstrated that deletion of the -80 region resulted in diminution of *lac* repressor binding, while simultaneous -60 and -80 deletions abolished repressor binding (34). Additionally, the -60 region was postulated to be involved in the binding of a negative-acting catabolite repressor, as catabolite repression of *lac* operon fusions was relieved in the context of a multicopy plasmid containing the



FIG. 6. Comparison of the putative S. mutans lac promoter with that of S. aureus. The nucleotide sequence of the streptococcal lac promoter region (Fig. 3; positions 1396 to 1475) is presented above that of the defined lac promoter of S. aureus (34). Direct repeat elements are indicated by arrows, and perfect inverted repeats are denoted by arrowheads.

staphylococcal -60 promoter region. While the similarities in the *S. mutans* and *S. aureus* promoter regions raise the possibility that repression and catabolite repression of *lac* expression may be similar in these organisms, the precise role(s) of these repeat elements (if any) in streptococcal *lac* expression clearly awaits further studies.

ACKNOWLEDGMENTS

We thank Roy Curtiss III for pYA501 and for communicating data prior to publication.

This work was supported by University of Kansas General Research Fund allocation 3190-XX-0038 and by Public Health Service grant GM37990 from the National Institutes of Health.

REFERENCES

- Alpert, C. A., and B. M. Chassy. 1988. Molecular cloning and nucleotide sequence of the factor III^{lac} gene of *Lactobacillus casei*. Gene 62:277-288.
- Alpert, C. A., and B. M. Chassy. 1990. Molecular cloning and DNA sequence of *lacE*, the gene encoding the lactose-specific enzyme II of the phosphotransferase system of *Lactobacillus casei*: evidence that a cysteine residue is essential for sugar phosphorylation. J. Biol. Chem. 265:22561-22568.
- Bethesda Research Laboratories. 1986. BRL pUC host: E. coli DH5α competent cells. Focus 8:9.
- Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513–1523.
- Bisset, D., and R. L. Anderson. 1973. Lactose and D-galactose metabolism in *Staphylococcus aureus*: pathway of D-galactose 6-phosphate degradation. Biochem. Biophys. Res. Commun. 52:641-647.
- Bisset, D., and R. L. Anderson. 1974. Genetic evidence for the physiological significance of the D-tagatose 6-phosphate pathway of lactose and D-galactose degradation in *Staphylococcus aureus*. J. Bacteriol. 119:698-704.
- Bisset, D., and R. L. Anderson. 1980. Lactose and D-galactose metabolism in *Staphylococcus aureus*. III. Purification and properties of D-tagatose-6-phosphate kinase. J. Biol. Chem. 255:8745-8749.
- Bisset, D., and R. L. Anderson. 1980. Lactose and D-galactose metabolism in *Staphylococcus aureus*. IV. Isolation and properties of a class I D-ketohexose-1,6-diphosphate aldolase that catalyzes the cleavage of D-tagatose 1,6-diphosphate. J. Biol. Chem. 255:8750–8755.
- Bisset, D., W. C. Wenger, and R. L. Anderson. 1980. Lactose and D-galactose metabolism in *Staphylococcus aureus*. II. Isomerization of D-galactose 6-phosphate to D-tagatose 6-phosphate by a specific D-galactose 6-phosphate isomerase. J. Biol. Chem. 255:8740-8744.
- Breidt, F., Jr., W. Hengstenberg, U. Finkeldei, and G. C. Stewart. 1987. Identification of the genes for the lactose-specific components of the phosphotransferase system in the *lac* operon of *Staphylococcus aureus*. J. Biol. Chem. 262:16444–16449.
- Breidi, F., Jr., and G. C. Stewart. 1987. Nucleotide and deduced amino acid sequences of the *Staphylococcus aureus* phosphoβ-galactosidase gene. Appl. Environ. Microbiol. 53:969-973.
- 12. Brown, A. T. 1975. The role of dietary carbohydrates in plaque formation and oral disease. Nutr. Rev. 33:353-361.
- 13. Calmes, R. 1978. Involvement of phosphoenolpyruvate in the catabolism of caries-conducive disaccharides by *Streptococcus mutans*: lactose transport. Infect. Immun. 19:934–942.
- 14. Daldal, F. 1984. Nucleotide sequence of a gene, *pfkB*, encoding the minor phosphofructokinase of *Escherichia coli* K12. Gene 28:337-342.
- de Vos, W. M., I. Boerrigter, R. J. van Rooijen, B. Reiche, and W. Hengstenberg. 1990. Characterization of the lactose-specific enzymes of the phosphotransferase system in *Lactococcus lactis*. J. Biol. Chem. 265:22554–22560.
- Dodd, I. B., and J. B. Egan. 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nu-

cleic Acids Res. 18:5019-5026.

- Dyer, D. W., and J. J. Iandolo. 1983. Rapid isolation of DNA from *Staphylococcus aureus*. Appl. Environ. Microbiol. 46: 283-285.
- Frostell, G., P. H. Keyes, and R. H. Larson. 1967. Effect of various sugars and sugar substitutes on dental caries in hamsters and rats. J. Nutr. 93:65-76.
- Green, C. J., G. C. Stewart, M. A. Hollis, B. S. Vold, and K. F. Bott. 1985. Nucleotide sequence of the *Bacillus subtilis* ribosomal RNA operon, *rrnB*. Gene 37:261-266.
- Hamada, S., and H. D. Slade. 1980. Biology, immunology, and cariogenicity of *Streptococcus mutans*. Microbiol. Rev. 44:331– 384.
- 21. Hamilton, I. R., and H. Lebtag. 1979. Lactose metabolism by *Streptococcus mutans*: evidence for induction of the tagatose 6-phosphate pathway. J. Bacteriol. 140:1102–1104.
- Hamilton, I. R., and G. Y. C. Lo. 1978. Co-induction of β-galactosidase and the lactose-P-enolpyruvate phosphotransferase system in *Streptococcus salivarius* and *Streptococcus mutans*. J. Bacteriol. 136:900-908.
- Hudson, M. C., and G. C. Stewart. 1986. Differential utilization of Staphylococcus aureus promoter sequences by Escherichia coli and Bacillus subtilis. Gene 48:93–100.
- 24. Jagusztyn-Krynicka, E. K., J. B. Hansen, V. L. Crow, T. D. Thomas, A. L. Honeyman, and R. Curtiss III. 1992. Streptococcus mutans serotype c tagatose 6-phosphate pathway gene cluster. J. Bacteriol. 174:6152-6158.
- 25. Kreiswirth, B. N., S. Lofdahl, M. J. Betley, M. O'Reilly, P. M. Schlievert, M. S. Bergdoll, and R. P. Novick. 1983. The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. Nature (London) 305:709-712.
- Larson, T. J., S. Ye, D. L. Weissenborn, H. J. Hoffmann, and H. Schweizer. 1987. Purification and characterization of the repressor for the *sn*-glycerol 3-phosphate regulon of *Escherichia coli* K12. J. Biol. Chem. 262:15869–15874.
- Lee, C. Y., and J. J. Iandolo. 1986. Lysogenic conversion of staphylococcal lipase is caused by insertion of the bacteriophage L54a genome into the lipase structural gene. J. Bacteriol. 166:385-391.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Maxwell, E. S., K. Kurahashi, and H. M. Kalckar. 1962. Enzymes of the Leloir pathway. Methods Enzymol. 5:174–189.
- Moran, C. P., Jr., N. Lang, S. F. J. LeGrice, G. Lee, M. Stephens, A. L. Sonenshein, J. Pero, and R. Losick. 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol. Gen. Genet. 186:339–346.
- Orchard, L. M. D., and H. L. Kornberg. 1990. Sequence similarities between the gene specifying 1-phosphofructokinase (*fruK*), genes specifying other kinases in *Escherichia coli* K12, and *lacC* of *Staphylococcus aureus*. Proc. R. Soc. London B 242:87-90.
- 32. Oskouian, B., E. L. Rosey, F. Breidt, Jr., and G. C. Stewart. 1990. The lactose operon of *Staphylococcus aureus*, p. 99–112. *In R. P. Novick and R. Skurray (ed.), Molecular biology of the* staphylococci. VCH Publishers, New York.
- Oskouian, B., and G. C. Stewart. 1987. Cloning and characterization of the repressor gene of the *Staphylococcus aureus* lactose operon. J. Bacteriol. 169:5459-5465.
- Oskouian, B., and G. C. Stewart. 1990. Repression and catabolite repression of the lactose operon of *Staphylococcus aureus*. J. Bacteriol. 172:3804–3812.
- Rosey, E. L., B. Oskouian, and G. C. Stewart. 1991. Lactose metabolism by *Staphylococcus aureus*: characterization of *lacABCD*, the structural genes of the tagatose 6-phosphate pathway. J. Bacteriol. 173:5992–5998.
- Rosey, E. L., and G. C. Stewart. 1989. The nucleotide sequence of the *lacC* and *lacD* genes of *Staphylococcus aureus*. Nucleic Acids Res. 17:3980.
- 37. Saier, M. H., Jr., and J. Reizer. 1992. Proposed uniform nomenclature for the proteins and protein domains of the

bacterial phosphoenolpyruvate:sugar phosphotransferase system. J. Bacteriol. 174:1433-1438.

- 38. Saier, M. H., Jr., M. Yamada, B. Erni, K. Suda, J. Lengeler, R. Ebner, P. Argos, B. Rak, K. Schnetz, C. A. Lee, G. C. Stewart, F. Breidt, Jr., E. B. Waygood, K. G. Peri, and R. F. Doolittle. 1988. Sugar permeases of the bacterial phosphoenolpyruvate-dependent phosphotransferase system: sequence comparisons. FASEB J. 2:199-208.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- Shortle, D. 1983. A genetic system for analysis of staphylococcal nuclease. Gene 22:181–189.
- Simoni, R. D., and S. Roseman. 1973. Sugar transport. VII. Lactose transport in *Staphylococcus aureus*. J. Biol. Chem. 248:966-976.
- 42. Smith, G. E., and M. D. Summers. 1980. The bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzy-loxymethyl-paper. Anal. Biochem. 109:123–129.
- 43. Smorawinska, M., J. C. Hsu, J. B. Hansen, E. Katarzyna, J. Krynicka, Y. Abiko, and R. Curtiss III. 1983. Clustered genes for galactose metabolism from *Streptococcus mutans* cloned in *Escherichia coli*. J. Bacteriol. 153:1095–1097.
- 44. Sullivan, M. A., R. E. Yasbin, and F. E. Young. 1984. New shuttle vectors for *Bacillus subtilis* and *Escherichia coli* which allow rapid detection of inserted fragments. Gene 29:21–26.
- 45. Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla. 1973. Improved estimation of secondary structure in ribonucleic acids. Nature

(London) New. Biol. 246:40-41.

- 46. Uhlen, M., B. Guss, B. Nilsson, S. Gatenbeck, L. Philipson, and M. Lindberg. 1984. Complete sequence of the staphylococcal gene encoding protein A: a gene evolved through multiple duplications. J. Biol. Chem. 259:1695–1702.
- Valentin-Hansen, P., P. Hojrup, and S. Short. 1985. The primary structure of the DeoR repressor from *Escherichia coli* K12. Nucleic Acids Res. 13:5927-5936.
- 48. van Rooijen, R. J., and W. M. de Vos. 1990. Molecular cloning, transcriptional analysis, and nucleotide sequence of *lacR*, a gene encoding the repressor of the lactose phosphotransferase system of *Lactococcus lactis*. J. Biol. Chem. 256:18499–18503.
- van Rooijen, R. J., M. J. Gasson, and W. M. de Vos. 1992. Characterization of the *Lactococcus lactis* lactose operon promoter: contribution of flanking sequences and LacR repressor to promoter activity. J. Bacteriol. 174:2273-2280.
- van Rooijen, R. J., S. van Schalkwijk, and W. M. de Vos. 1991. Molecular cloning, characterization, and nucleotide sequence of the tagatose 6-phosphate pathway gene cluster of the lactose operon of *Lactococcus lactis*. J. Biol. Chem. 266:7176-7181.
- 51. von Bodman, S. B., G. T. Hayman, and S. K. Farrand. 1992. Opine catabolism and conjugal transfer of the nopaline Ti plasmid pTiC58 are coordinately regulated by a single repressor. Proc. Natl. Acad. Sci. USA 89:643–647.
- 52. Wu, L. F., A. Reizer, J. Reizer, B. Cai, J. M. Tomich, and M. H. Saier, Jr. 1991. Nucleotide sequence of the *Rhodobacter capsulatus fruK* gene, which encodes fructose-1-phosphate kinase: evidence for a kinase superfamily including both phosphofructokinases of *Escherichia coli*. J. Bacteriol. 173:3117–3127.