JOURNAL OF BACTERIOLOGY, Feb. 1988, p. 810–816
0021-9193/88/020810-07$02.00/0

Vol. 170, No. 2

Sequence Analysis of the *Streptococcus mutans* Fructosyltransferase Gene and Flanking Regions

TERUAKI SHIROZA AND HOWARD K. KURAMITSU*

Department of Microbiology-Immunology, Northwestern University Medical-Dental Schools, Chicago, Illinois 60611

Received 11 September 1987/Accepted 19 November 1987

The nucleotide sequence of the *ftf* gene from *Streptococcus mutans* GS-5 was determined. The deduced amino acid sequence indicates that the unprocessed fructosyltransferase gene product has a molecular weight of 87,600. A typical streptococcal signal sequence is present at the amino terminus of the protein. The processed enzyme is relatively hydrophilic and has a pI of 5.66. An inverted repeat structure was detected upstream from the *ftf* gene and may function in the regulation of fructosyltransferase expression. Sequencing of the regions flanking the gene revealed the presence of four other putative open reading frames (ORFs). Two of these, ORFs 2 and 3, appear to code for low-molecular-weight proteins containing amino acid sequences sharing homology with several gram-positive bacterial DNA-binding proteins. In addition, ORF 3 is transcribed from the *ftf* DNA coding strand. Partial sequencing of ORF 4 suggests that its gene product may be an extracellular protein.

Streptococcus mutans has been implicated as the principal causative agent in the development of human dental caries (13). Extensive investigations have focused on the synthesis of water-insoluble glucan by these organisms since these polymers play a role in the colonization of tooth surfaces. Many strains of *S. mutans* are also capable of synthesizing extracellular fructan polymers (3). However, the role of these polysaccharides in the cariogenicity of *S. mutans* has not been elucidated. Recent evidence from this laboratory (S. Sato and H. K. Kuramitsu, unpublished results) suggested that mutants of *S. mutans* GS-5 defective in fructosyltransferase (FTF; EC 2.4.1.10) activity still display normal sucrose-dependent colonization of smooth surfaces in vitro. However, since fructans can be degraded by several organisms present in human dental plaque (25), these polymers could function as a reserve source of carbohydrate for plaque microorganisms.

Although the *S. mutans ftf* gene coding for FTF activity has been recently isolated and characterized (18), little information is currently available regarding the regulation of expression of this gene. Growth of *S. mutans* Ingbritt in continuous culture indicated that the enzyme is constitutively expressed (27). In contrast, another enzyme catalyzing the synthesis of fructan from sucrose, levansucrase from *Bacillus subtilis*, was shown to be inducible by sucrose (12). An analysis of various levansucrase mutants has indicated that multiple genes (*sacR*, *sacS*, *sacQ*, and *sacH*) may be involved in the regulation of levansucrase expression (12). In addition, more recent sequence analysis of the levansucrase gene (*sacB*) suggested the presence of a potential termination structure upstream from this gene (23). This site could be involved in the regulation of gene expression (20).

To examine the regulation of the *S. mutans ftf* gene, nucleotide sequencing of the gene was carried out. The present report describes the sequence of the intact gene along with both upstream and downstream flanking sequences. Several potential open reading frames (ORFs) were detected in these regions, and this suggests that the product of one of these genes could be involved in gene regulation.

* Corresponding author.

MATERIALS AND METHODS

Plasmids. The construction of plasmid pSS22 has been recently described (18). The plasmid encoding the intact *ftf* gene, pTS102, was identified in a *Hin*dIII clone bank of *S. mutans* GS-5 chromosomal DNA. Since Southern blot analysis indicated that the *ftf* gene resides on a 4.5-kilobase (kb) *Hin*dIII fragment (data not shown), DNA fragments of 4 to 6 kb were isolated from an agarose gel following *Hin*dIII digestion of chromosomal DNA. The purified fragments were ligated to *Hin*dIII-cleaved vector pUC8. The ligation mixture was transformed into *Escherichia coli* JM83, and transformants harboring chimeric plasmids were selected on Luria-Bertoni (LB) agar plates containing ampicillin (50 μg/ml) and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal). Transformants expressing FTF activity were initially identified following replica plating of the clone bank onto LB agar-ampicillin plates supplemented with 1% sucrose. Colonies growing poorly in the presence of sucrose (synthesizing large amounts of intracellular fructan [2]) were isolated, grown in small samples (2 ml), and assayed for FTF activity. In this manner, one colony expressing FTF activity was identified. The plasmid contained in this transformant, pTS102, was used for sequencing the intact *ftf* gene.

DNA manipulations. DNA isolation, endonuclease restrictions, ligations, transformation of competent *E. coli* cells, and Southern blot analysis were carried out as recently described (2).

Nucleotide sequencing. The two large DNA fragments containing the intact *ftf* gene, i. e., the 1.8-kb *Hin*dIII-*Eco*RI fragment from pSS22 and the 2.0-kb *Xho*I-*Hin*dIII fragment from pTS102 (Fig. 1), were subcloned into M13mp18 or M13mp19. A series of deleted bacteriophage clones were constructed as described by Henikoff (9) following treatment with exonuclease III and nuclease S1. The desired subfragments were cloned into the M13 phage vectors by using the appropriate restriction sites as described below. Nucleotide sequences were determined by the dideoxy chain termination procedure (17) with lambda phage single-stranded DNA, the 17-mer universal primer (Bethesda Research Laboratories, Gaithersburg, Md.), and [α - ^{35}S]dATP (600 Ci/mmol; Amersham Corp., Arlington Heights, Ill.) as recently de-

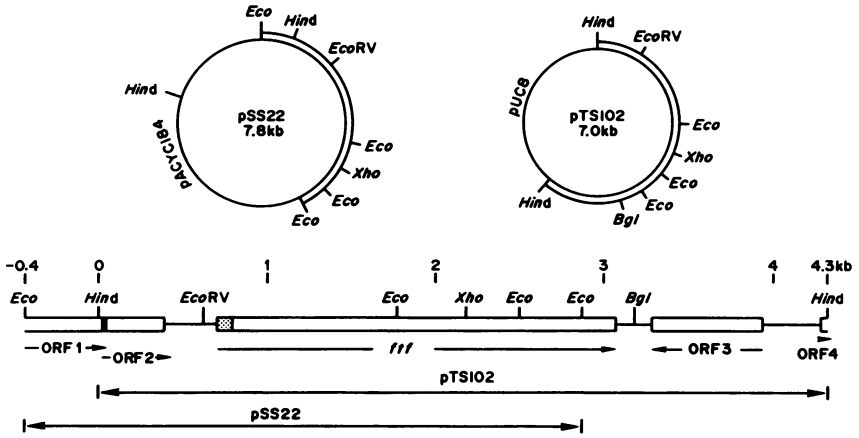

FIG. 1. Restriction maps of pSS22 and pTS102 and the deduced total structure of the 4.7-kb *Eco*RI-*Hin*dIII fragment. The GS-5 chromosomal DNA insert of plasmid pSS22 consists of three *Eco*RI fragments, while pTS102 contains a 4.3-kb *Hin*dIII chromosomal insert. The transcriptional directions of the genes are indicated by the arrows (ORF 3 is transcribed in the opposite direction relative to the other genes). A small segment of ORF 2 overlaps the 3′ end of ORF 1. ■, Putative signal sequence region at the 5′ end of the *ftf* gene. Abbreviations: *Bgl*, *Bgl*II; *Eco*, *Eco*RI; *Hin*d, *Hin*dIII; *Xho*, *Xho*I.

scribed (22). However, putative ORF 4 (Fig. 1) was sequenced by using double-stranded DNA and the 16-mer reverse primer (New England BioLabs, Inc., Beverly, Mass.). Both DNA strands of the 4.7-kb *Eco*RI-*Hin*dIII fragment (Fig. 1) were completely sequenced.

Sequence analysis. The nucleotide and amino acid sequences were analyzed with the Pustell sequence analysis program (International Biotechnologies, Inc., New Haven, Conn.).

Enzyme assays. FTF activity was measured as previously described (18) by using either the sucrase assay or [^{3}H]fructan synthesis from [^{3}H-fructose]sucrose.

RESULTS

Identification of ORFs on the 4.7-kb *Eco*RI-*Hin*dIII fragment of *S. mutans* chromosomal DNA. Our previous studies of the *ftf* gene have shown that *E. coli* transformants harboring plasmid pSS22 (Fig. 1), containing a 3.2-kb *S. mutans* chromosomal DNA insert, exhibit strong FTF activity (18). Nucleotide sequencing of the insert (Fig. 2 and data not shown) revealed that the insert contains three potential ORFs (Fig. 1). Two of these, putative ORFs 1 and 2, encode polypeptides of at least 159 and 122 amino acids, respectively. The other large ORF specifies a protein of 728 amino acids. Therefore, the nucleotide sequence data revealed that only the latter ORF is compatible with the predicted molecular weight of the *ftf* gene (18). However, no termination codon could be found within the longer ORF corresponding to the *ftf* gene in pSS22.

Southern blot analysis of *S. mutans* chromosomal DNA cleaved with *Hin*dIII by using plasmid pSS22 as a probe indicates two positive signals of approximately 4.5 and 1.0 kb (data not shown). Since the predicted molecular size of the *ftf* gene is less than 3.0 kb (18) and one *Hin*dIII site is located 680 base pairs (bp) upstream from the initiator Met codon of the gene (Fig. 2), these results suggested that the entire *ftf* gene should be contained within the 4.5-kb *Hin*dIII fragment. Accordingly, an *S. mutans* GS-5 *Hin*dIII clone bank was constructed in *E. coli* JM83 by using plasmid vector pUC8. One of the clones was demonstrated to express FTF activity following screening of the clone bank for sucrase activity (15). The plasmid from this clone, designated pTS102, was purified from the transformant, and a restriction map was constructed (Fig. 1). Restriction analysis indicates that the plasmid contains a 1.4-kb *Eco*RI-*Hin*dIII fragment which should contain the termination codon for the *ftf* gene. Nucleotide sequencing of this region was carried out by using the 2.1-kb *Xho*I-*Hin*dIII fragment of pTS102. In addition to the expected termination codon for the *ftf* gene (position 3072, Fig. 2), two additional potential ORFs (ORFs 3 and 4) were identified within this fragment (Fig. 1). The coding strand for ORF 3 is opposite to that for the *ftf* gene and ORFs 1 and 2. Even though only a small portion of putative ORF 4 was sequenced (Fig. 2), the data presented below suggests that this sequence codes for an extracellular protein. Therefore, the sequence data revealed the presence of four additional potential ORFs in addition to the *ftf* gene on the 4.7-kb *Eco*RI-*Hin*dIII fragment.

Characterization of the *ftf* gene. The *ftf* structural gene begins with the ATG initiation codon (position 681, Fig. 2) and ends with the TAA termination codon (position 3072). The intact gene codes for a 797-amino-acid protein with a predicted molecular weight of 87,600. The *ftf* gene is preceded by an inverted repeat region (positions 565 to 593) which may function as a regulatory region, as has been postulated for a similar region upstream of the levansucrase gene of *B. subtilis* (23). However, this same sequence could also act as a termination sequence for ORF 2. In addition, a promoterlike sequence ATGATA-N_{17}-TAGGAT, which resembles the *E. coli* consensus sequence TTGACA-N_{17}-TATAAT (8) exists between positions 617 and 645. The nucleotide sequence downstream from the *ftf* gene contains an inverted repeat structure (positions 3150 to 3179) which might act as a transcription terminator (16) for the gene. It is of interest that this potential stem-loop structure is also positioned to play a similar role in regulating the expression of ORF 3, with transcription of this gene occurring in the opposite direction to that of the *ftf* gene.

Characterization of the FTF protein. The amino-terminal region of the FTF protein (first 34 amino acid residues) resembles a typical signal sequence for extracellular proteins observed in gram-positive bacteria (1, 6, 11). The first 14 residues are strongly basic (7 of 14), and they are followed by a highly hydrophobic region (13 of 17). On the basis of the application of the general rules for determining signal pepti-

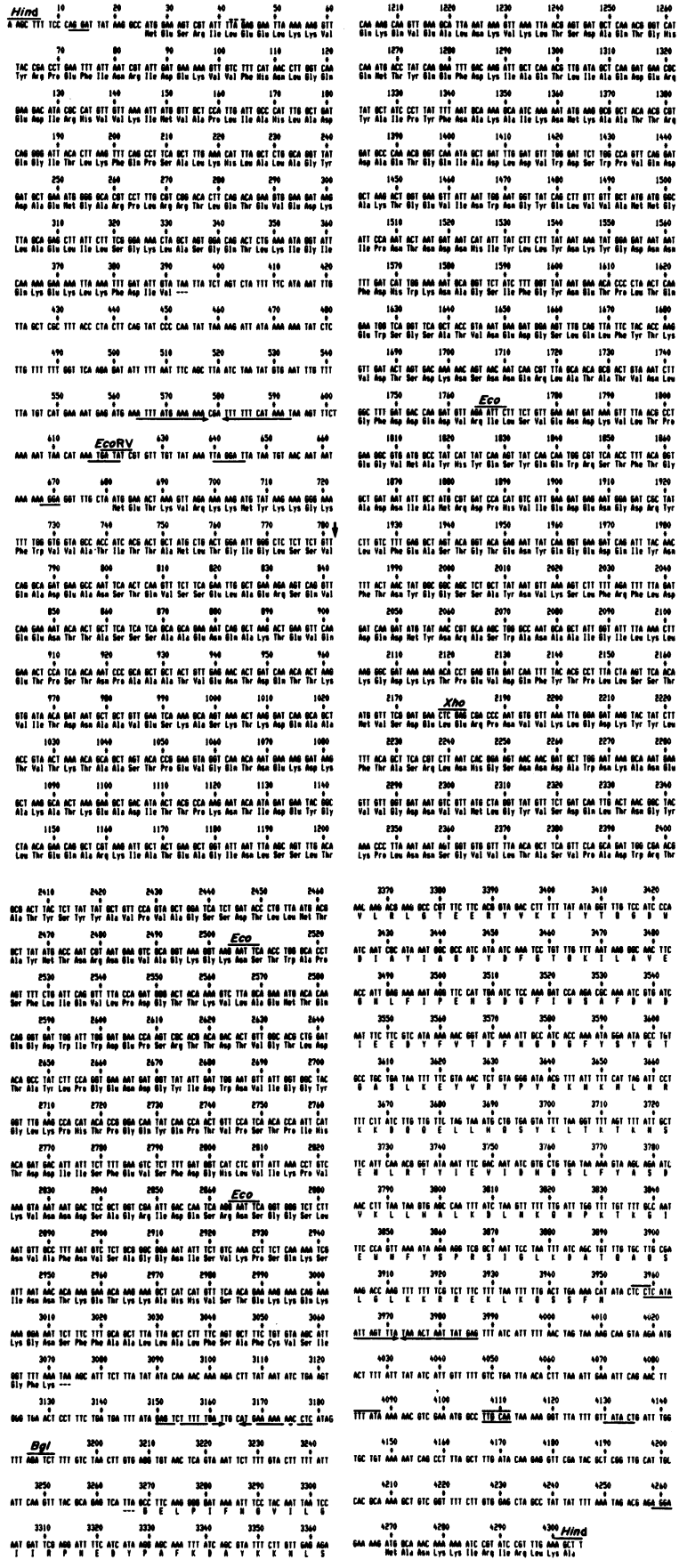
HindIII
EcoRV
Eco
Xho
Bgl
HindIII

TABLE 1. Codon usage of the *ftf* gene

Codon	Amino acid	No. of residues	Codon	Amino acid	No. of residues	Codon	Amino acid	No. of residues	Codon	Amino acid	No. of residues
TTT	Phe	19	TCT	Ser	17	TAT	Tyr	25	TGT	Cys	1
TTC	Phe	4	TCC	Ser	1	TAC	Tyr	8	TGC	Cys	0
TTA	Leu	17	TCA	Ser	19	TAA		1	TGA		0
TTG	Leu	7	TCG	Ser	2	TAG		0	TGG	Trp	14
CTT	Leu	13	CCT	Pro	7	CAT	His	10	CGT	Arg	8
CTC	Leu	3	CCC	Pro	4	CAC	His	1	CGC	Arg	3
CTA	Leu	4	CCA	Pro	14	CAA	Gln	30	CGA	Arg	3
CTG	Leu	4	CCG	Pro	2	CAG	Gln	11	CGG	Arg	0
ATT	Ile	25	ACT	Thr	26	AAT	Asn	50	AGT	Ser	17
ATC	Ile	4	ACC	Thr	9	AAC	Asn	10	AGC	Ser	4
ATA	Ile	5	ACA	Thr	30	AAA	Lys	35	AGA	Arg	4
ATG	Met	15	ACG	Thr	13	AAG	Lys	21	AGG	Arg	1
GTT	Val	35	GCT	Ala	40	GAT	Asp	46	GGT	Gly	27
GTC	Val	9	GCC	Ala	7	GAC	Asp	8	GGC	Gly	11
GTA	Val	9	GCA	Ala	21	GAA	Glu	35	GGA	Gly	10
GTG	Val	5	GCG	Ala	4	GAG	Glu	6	GGG	Gly	4

dase cleavage sites (22), a putative cleavage site following Val-34 was assigned (Fig. 2).

Examination of a hydropathy plot of the deduced FTF amino acid sequence revealed that the enzyme is highly hydrophilic with only two significant hydrophobic regions (data not shown). One of these corresponds to a portion of the signal sequence, while the other was found at the carboxy terminus of the protein (nine consecutive hydrophobic amino acids from base positions 3015 to 3040; Fig. 2). On the basis of the amino acid composition of the putative processed protein (Table 1), a pI of 5.66 was determined.

Codon utilization of the *ftf* gene. Examination of the codon usage for the *ftf* gene (Table 1) revealed a high frequency (78%) of A+T in the third positions of the codons. This frequency is higher than that observed for the same position in the codons for the strain GS-5 *gtfB* gene (63%) and reflects the relatively low G+C content (36 to 38%) of the *S. mutans* chromosomal DNA (7). Only one residue of cysteine was detected in the entire protein (Table 1). This is consistent with recent observations that extracellular proteins from streptococci appear to contain little or no cysteine (22).

Comparison of FTF with the levansucrase of *B. subtilis*. The conversion of sucrose to a fructan polymer is catalyzed not only by the FTF of *S. mutans* but also by the levansucrase of *B. subtilis* (4). The former enzyme produces primarily an inulinlike polymer (3), while the latter enzyme synthesizes a levan product (23). Since Steinmetz et al. (23) have recently determined the nucleotide sequence of the latter gene, it was of interest to compare the sequence with that of the *S. mutans* FTF. A comparison of the two genes by a homology matrix indicates significant homologies at both the amino acid and nucleotide sequence levels (data not shown). When the two amino acid sequences are aligned (including gaps to account for the differences in molecular weights), regions of extensive homology can be detected (Fig. 3). High degrees of amino acid sequence homologies are especially apparent at the amino termini of both proteins (signal sequence regions) and in the central region of the FTF protein. Contained in these latter regions are sequences with as many as nine consecutive amino acids in common between the two proteins.

Characterization of ORFs 1 and 2. Putative ORF 1, which ends at the termination codon TGA, 42 bp downstream from a *Hin*dIII site, appears to encode a 159-amino-acid-residue carboxy terminus of an unknown protein, since no other termination codon is present in the 0.4-kb *Eco*RI-*Hin*dIII fragment (data not shown). ORF 2, encoding a 122-residue polypeptide, begins with the ATG initiation codon (position 26, Fig. 2), is preceded by a potential Shine-Dalgarno (SD) sequence (21), and ends at the termination codon TAA (position 392). Therefore, the coding sequences for the amino-terminal region of ORF 2 appear to overlap with the carboxy-terminal region of ORF 1.

Characterization of ORF 3. The putative ORF 3, transcribed from the opposite DNA strand relative to the *ftf* gene, begins with the ATG initiation codon (position 3951, Fig. 2) and terminates at the TAA codon (position 3267). This putative gene would code for a 228-amino-acid polypeptide with a molecular weight of 26,500. The putative gene is preceded by potential −35 and −10 promoter regions (TTGCAA and TATAAA), although the distance between the two sequences is 18 bp rather than the 17 bp difference found in most *E. coli* promoter sequences (8). Furthermore, a potential SD sequence was identified at position 3956 (Fig. 2). Interestingly, an inverted repeat sequence (positions 3958 to 3978) partially overlaps the SD sequence of the gene,

FIG. 2. Nucleotide sequence of the 4.3-kb *Hin*dIII fragment. The sequence shown is that for the noncoding strand of the *ftf* gene. The coding strand of ORF 3 is transcribed in the direction opposite to that of the other ORFs. To indicate this, the deduced amino acid sequence of ORF 3 is presented by using the single-letter amino acid code. The termination codon (TGA at position 42) for ORF 1 is indicated by the broken overline. Underlined are: position 12, SD sequence for ORF 2; positions 617, 640, and 668, respectively, the −35, −10, and SD sequences for the *ftf* gene; and positions 4107, 4130, and 4159, respectively, the −35, −10, and SD sequences for ORF 4. Overlined are positions 4112, 4088, and 3959, respectively, the −35, −10, and SD sequences for ORF 3. The three opposing arrows (positions 565 to 593, 3150 to 3179, and 3958 to 3978) denote inverted repeat structures. The putative signal sequence cleavage site for the FTF protein is indicated by the vertical arrow (position 782). The base positions are numbered beginning at the *Hin*dIII site.

```
          *          *          *          *          *       50*
METKVRKKMY-KK-GKFWVVATITTAMLTGIGLSSVQADEANSTQVSSELAERSQVQENT
        MNIKKFAKQATVLTFTTALLAG-GATQAFAKETNQKPYKETYGISHITRHDM
                   *          *          *          *       50*
 *          *          *          *        100*          *
TASSSAAENQAKTEVQETPSTNPAAATVENTDQTTKVITDNAAVESKASKTKDQAATVTK
LQI---

 *          *          *       150*          *          *
TAASTPEVGQTNEKDKAKATKEADITTPKNTIDEYGLTEQARKIATEAGINLSSLTQKQV
  ---PE-QQKNEK---
          60*
 *          *       200*          *          *          *
EALNKVKLTSDAQTGHQMTYQEFDKIAQTLIAQDERYAIPVFNAKAIKNMKAATTRDAQT
                                  ---YQVPEFDSSTIKNISSAKG---
                                                70*          *
 *       250*          *          *          *          *
GQIADLDVWDSWPVQDAKTGEVINWNGYQLVVAMMGIP-NTNDNHIYLLYNKYGDNNFDH
  ---LDVWYSWPLQNAD-GTVANYHGYHIVFALAGDPKNADDTSIYMFYQKVGETSIDS
                   *       100*          *          *          *
  *300          *          *          *          *          *
WKNAGSIFGYNETPL---------TQEWSGSATVNEDGSLQLFYTKVDTSDKNSNNQRLA
WKNAGRVFKDSDKFDANDSILKDQTQEWSGSATFTSDGKIRLFYT--DFSGKHYGKQTLT
     *       150*          *          *          *          *
  *350          *          *          *          *       400*
TATVNLGFDDQDVRILSVENDKVLTPEGVMAYHYQSYQQWRSTFTGA--DNIAMRDPHVI
TAQVNVSASDSSLNINGVEDYKSIFDGDGKT--YQNVQQFIDEGNYSSGDNHTLRDPHYV
   200*          *          *          *          *       250*
  *          *          *          *       450*          *
EDENGDRYLVFEASTGTEN-YQGEDQIYNFTNYGGSSAYNVKSLFRFLDDQDMYNRASWA
EDK-GHKYLVFEANTGTEDGYQGEESLFNKAYYGKSTSFFRQESQK-LLQSDKKRTAELA
          *          *          *          *       300*
    *          *          *       500*          *          *
NAAIGILKLKGDKKTPEVDQFYTPLLSSTMVSDELERPNVVKLGDKYYLFTASRLNHGSN
NGALGMIELNDDYTLKKVMK---PLIASNTVTDEIERANVFKMNGKWYLFTDSR---GSK
*          *          *          *       350*          *
    *          *       550*          *          *          *
NDAWNKANEVVGDNVVMLGYVSDQLTNGYKPLNNSGVVLTASVPADWRTATYSYYAVPVA
MTIDGITSNDIY----MLGYVSNSLTGPYKPLNKTGLVLKMDLDPNDVTFTYSHFAVPQA
     *          *          *       400*          *
    *       600*          *          *          *          *
GSSDTLLMTAYMTNRNEVAGKGKNSTWAPSFLIQVLPDGTTKVLAEMTQQGDWIWDEPSR
KGNNVVI-TSYMTNRGFYADKQ--STFAPSFLLNIKGKKTSVV---
*          *          *       450*
 650*          *          *          *          *       700*
TTDTVGTLDTAYLPGENDGYIDWNVIGGYGLKPHTPGQVQPTVPSTPIHTDDIISFEVSF
                                  ---KDDILEQGQLTVNK.
                                             *       470*
    *          *          *          *       750*          *
DGHLVIKPVKVNNDSAGRIDQSRNSGGSLNVAFNVSAGGNISVKPSQKSINNTKETKKAH

    *          *       790*
HVSTEKKQKKGNSFFAALLALFSAFCVSIGFK.
```

FIG. 3. Alignment and comparison of the amino acid sequences of the *S. mutans* GS-5 FTF (upper sequence) and *B. subtilis* levansucrase (lower sequence). The two amino acid sequences are aligned based on a homology matrix (data not shown), and the conserved residues between the two proteins are underlined. To maximize the homology, gaps in the sequence alignment have been introduced. The downward and upward arrows indicate the predicted and actual signal sequence cleavage sites for the FTF and levansucrase proteins, respectively.

suggesting that the former sequence may be involved in the regulation of expression of ORF 3. The possibility that ORFs 1, 2, and 3 are actually translated in the *S. mutans* cells is suggested by similar codon utilization of the ORFs relative to the *gtfB* and *ftf* genes of *S. mutans* (data not shown).

Comparison of ORFs 2 and 3 with gram-positive regulatory proteins. Recent genetic investigations of the levansucrase from *B. subtilis* (20) have suggested that the expression of the enzyme is positively regulated by the *sacU* gene product. Since such regulatory proteins bind to DNA and exhibit relatively low molecular weights, it is possible that the products of ORFs 2 or 3 also participate in the regulation of gene expression. Therefore, the predicted amino acid sequences of the two ORF products were compared with the sequences of several DNA-binding proteins: sigma 37 subunit (5), *spoOF* (24), *phoP* (19), and *penI* (10) from gram-positive bacteria. No extensive homologies were observed between the two sets of proteins (data not shown). However, both ORFs 2 and 3 displayed several regions of amino acid homology (involving up to eight consecutive amino acids) with these selected DNA-binding proteins. In addition, one sequence (Lys-Lys-Val-Tyr-Arg) was observed in both ORF 2 (positions 54 to 67, Fig. 2) and ORF 3 (positions 3401 to 3388).

Characterization of putative ORF 4. In the course of sequencing the 1.4-kb *Eco*RI-*Hin*dIII fragment of pTS102 (Fig. 4), it was observed that only one of the two DNA strands (*ftf* antisense strand) of this fragment could be isolated in M13mp18 or M13mp19. However, both strands of two deletion derivatives of this fragment (clones 6 and 12) could be readily isolated in the M13 phages. These results are similar to our recent results in sequencing the *gtfB* gene from strain GS-5 (22). In the latter case, it was also not possible to isolate both DNA strands containing the promoter region of this gene in the M13 phages. By analogy, it is likely that a strong promoter sequence exists on the terminal 0.3-kb region of the *Eco*RI-*Hin*dIII fragment. However, as with the promoter region of the *gtfB* gene (22), it was possible to determine the nucleotide sequence of both DNA strands in the terminal region by using a double-stranded DNA template. Nucleotide sequencing also revealed the presence of the beginning of another potential ORF, termed ORF 4, whose deduced amino acid sequence is very similar to those of the highly basic terminal regions of the signal sequences from both the *ftf* (Fig. 2) and *gtfB* genes (22). Although only a small part of ORF 4 has been sequenced, these results suggest that ORF 4 may code for an extracel-

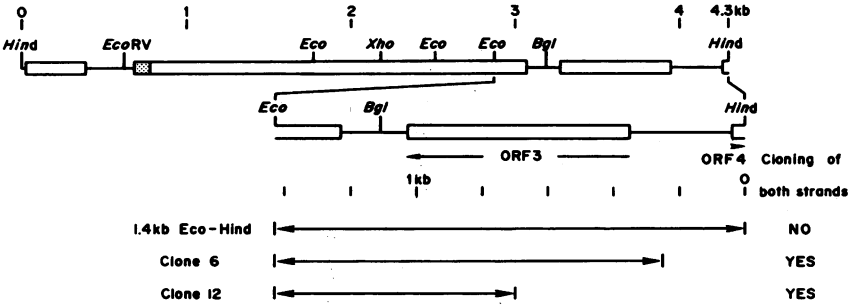

FIG. 4. Isolation of DNA fragments containing ORF 4 and its regulatory regions. Phage clones 6 and 12 are two deletion derivatives of the 1.4-kb *Eco*RI-*Hin*dIII fragment in which approximately 250- and 700-bp sequences from the right end of the *Hin*dIII site are removed, respectively. In the parental 1.4-kb *Eco*RI-*Hin*dIII fragment, only one of the two strands that hybridized with the RNA polymerase sense strand of the *ftf* gene was cloned, while both DNA strands from clones 6 and 12 were isolated in phage M13 vectors. The restriction endonuclease abbreviations and the shaded box are the same as for Fig. 1.

lular protein. It is also of interest that the putative −35 region of ORF 4 could also serve a similar function for ORF 3 (Fig. 2). However, in the latter case, the opposite DNA strand read in the reverse direction would serve as part of the recognition site for RNA polymerase. Thus, the terminal 0.3-kb fragment appears to contain two promoters (one for ORF 3 and the other for ORF 4) (Fig. 2).

DISCUSSION

Nucleotide sequencing of the *ftf* gene from plasmid pSS22 revealed that the cloned gene lacks a termination codon. Therefore, the FTF activity expressed by the plasmid corresponds to a fusion protein (*ftf* gene fused to a pACYC184 vector sequence). Nevertheless, since FTF activity is readily expressed from pSS22 (18), the missing carboxy-terminal portion of the *ftf* gene is not required for activity. Following the isolation of the intact *ftf* gene on plasmid pTS102, the termination codon for the gene was identified 207 bp downstream from the *Eco*RI site in the 1.4-kb *Eco*RI-*Hin*dIII fragment of pTS102. The deduced amino acid sequence of the FTF protein revealed the presence of only a single cysteine residue (Table 1). Likewise, the glucosyltransferase I (GTF-I) protein from strain GS-5 lacks cysteine residues (22). Therefore, streptococcal extracellular proteins appear to contain few cysteine residues (6). However, the significance of these observations has not yet been determined.

Like other streptococcal extracellular proteins (6), FTF contains a relatively long (34 amino acids) signal sequence. This sequence is also very similar (10 of 34 identical amino acids) to that of the GTF-I protein from strain GS-5 (22). A putative signal sequence cleavage site (position 782, Fig. 2) could be predicted based on the general rules of von Heijne (26). However, amino acid sequencing of the secreted FTF will be necessary to confirm this prediction. The predicted molecular size of the mature FTF, 83.8 kilodaltons, is compatible with that of the FTF activity observed in culture fluids of strain GS-5, 83 kilodaltons (data not shown). The previously observed 91-kilodalton FTF from lambda clones (18) apparently results from fusion of the *ftf* gene with a lambda DNA sequence.

A comparison of the FTF protein with the levansucrase from *B. subtilis* indicates that the two proteins contain regions of homology (Fig. 3). Since both enzymes catalyze the same basic reaction, it is likely that these homologous regions are required for enzymatic activity (one of these regions may be involved in binding the substrates). It will be of interest to determine the structural specificity which has resulted in the FTF protein synthesizing a β-1,2-linked fructan, while the levansucrase produces a β-2,6-linked polymer.

It is also of interest that the putative signal sequence of the FTF protein is somewhat longer than that of the levansucrase (Fig. 3). The results from several laboratories (6, 11) appear to suggest that the signal sequences of streptococcal extracellular proteins generally appear to be longer than those of other bacteria. However, the molecular basis for such a difference has not yet been determined. Previous localization studies (2) have indicated that most of the GTF-I activity expressed in *E. coli* clones is associated with the cytoplasmic membrane (2). In contrast, the majority of the cloned FTF activity is secreted into the periplasmic space (18). Since the signal sequences of the two enzymes appear to be quite similar, it may be possible that structural differences in other regions of the two proteins are responsible for the difference in localization. However, the results of the sequence analysis suggest an alternate hypothesis. Translation of the *ftf* gene could actually begin at an internal ATG codon (position 705, Fig. 2), since this codon is preceded by a potential SD sequence (AGAAAAAAG). In this case, the resultant signal sequence would still retain its required characteristics (initial basic region followed by a hydrophobic region). The resultant signal sequence would be similar in size to that of the levansucrase of *B. subtilis*, which is secreted into the periplasmic space of *E. coli* clones (23). Thus, it is possible that signal sequences with relatively short basic regions are able to pass through the *E. coli* cytoplasmic membrane, while streptococcal proteins with longer signal sequences (GTF-I protein) may not be able to penetrate this structure. Examination of more cloned streptococcal extracellular proteins will be required to test this hypothesis.

Nucleotide sequencing of regions flanking the *ftf* structural gene indicate the presence of potential regulatory sequences. The *ftf* gene is preceded by an inverted repeat sequence (positions 565 to 593, Fig. 2), which by analogy to a similar structure upstream from the *B. subtilis* levansucrase gene (23) may be involved in the regulation of FTF expression. It has been proposed that the repeat sequence upstream from the levansucrase gene may act as a recognition site for a regulatory protein (20). However, this latter sequence is positioned between the promoter region and the levansucrase structural gene. In contrast, the present results indicate that the inverted repeat sequence preceding the *ftf* gene lies upstream from the promoter sequence for this gene (Fig. 2). Additional mutagenesis experiments designed to alter this sequence will be required to demonstrate its possible role in regulating FTF expression.

Since the regulatory proteins involved in controlling bacterial gene expression are of relatively low molecular weights (14), it was of interest that two ORFs (ORFs 2 and 3) coding for low-molecular-weight proteins could be identified flanking the *ftf* gene. A comparison of the deduced amino acid sequences of these two genes with those of selected gram-positive bacterial DNA-binding proteins revealed short sequences of homology (data not shown). ORF 3 contains a 20-amino-acid sequence (Gln-18 to Ile-27, Fig. 2) which shares partial homology with amino acids 223 to 238 of the sigma 37 subunit (5). The latter sequence has been postulated to play a direct role in binding to DNA (5). The DNA-binding sites of several regulatory proteins share the common structure Ala-N_3-Gly-N_5-Val/Ile/Leu (14). The homologous region from ORF 3 displays the sequence Ala-N_3-Gly-N_5-Tyr beginning with amino acid 22 (Fig. 2), which might also function in the same capacity. However, no comparable sequence was detected for ORF 2, although this protein also shares several short regions of homology with the regulatory proteins (data not shown). In addition, an extensive amino acid sequence comparison of ORFs 2 and 3 with a wide variety of different DNA-binding proteins was not carried out in the present investigation. Since ORF 3 is preceded by an inverted repeat structure (positions 3958 to 3987, Fig. 2), it is possible that this gene is also subject to regulation by another regulatory protein (cascade regulation). To establish the possible roles of ORFs 2 or 3 in the regulation of FTF expression, it will be necessary to isolate and characterize both genes and their protein products. In addition, the cloned genes will be insertionally inactivated and transformed into strain GS-5 to produce mutants which are defective in the expression of ORFs 2 and 3 (2). It is clear that these additional approaches will be required to confirm

the regulatory roles for these ORFs suggested by the sequence data and their possible relationships, if any, to FTF synthesis.

The inability to isolate one of the two DNA strands from the 1.4-kb *Eco*RI-*Hin*dIII fragment in the M13 phages (Fig. 4) is similar to our experience in attempting to isolate both single-stranded DNA fragments containing the promoter region of the *gtfB* gene (22). In the latter case, only the RNA polymerase antisense strand containing this region could be isolated in the M13 phages. This result suggested that insertion of a strong promoter (sense strand) into the multicloning site of M13 phages such that the direction of transcription from the heterologous promoter is in the opposite direction relative to the phage genes results in the inability to recover phage particles. The results of the present investigation also indicate that only the RNA polymerase antisense strand (relative to the *ftf* gene) of the 1.4-kb *Eco*RI-*Hin*dIII fragment could be isolated in M13mp18 or M13mp19. This is indicated by the observation that the single-stranded DNA isolated in the M13 phages from the intact fragment hybridized with the sense strand of the *ftf* gene (data not shown). The inability to isolate phage particles containing the sense strand of the fragment suggests the presence of a strong promoter on this strand.

Although only a small portion of ORF 4 has been sequenced in the present investigation, the deduced amino acid sequence suggests the presence of a signal sequence typical of GS-5 extracellular proteins. In addition, both potential SD and promoter sequences were identified upstream from this putative structural gene. The ORF 4 promoter is apparently responsible for the strong promoter activity preventing the isolation of the sense strand of the 1.4-kb *Eco*RI-*Hin*dIII fragment in the M13 phages. It will be of interest to isolate a DNA fragment containing intact ORF 4 in order to characterize this potential extracellular protein. It also may be possible that ORF 3 is involved in regulating the expression of this protein.

ACKNOWLEDGMENT

This investigation was supported in part by Public Health Service grant DE-03258 from the National Institutes of Health.

LITERATURE CITED

1. **Abrahmsen, L., T. Moks, B. Nilsson, U. Hellman, and M. Uhlen.** 1985. Analysis of signals for secretion in staphylococcal protein A gene. EMBO J. **4**:3901–3906.
2. **Aoki, H., T. Shiroza, M. Hayakawa, S. Sato, and H. K. Kuramitsu.** 1986. Cloning of a *Streptococcus mutans* glucosyltransferase gene coding for insoluble glucan synthesis. Infect. Immun. **53**:587–594.
3. **Carlsson, J.** 1970. A levansucrase from *Streptococcus mutans*. Caries Res. **4**:97–113.
4. **Dedonder, R.** 1966. Levansucrase from *Bacillus subtilis*. Methods Enzymol. **8**:500–505.
5. **Duncan, M. L., S. S. Kalman, S. M. Thomas, and C. W. Price.** 1987. Gene encoding the 37,000-dalton minor sigma factor of *Bacillus subtilis* RNA polymerase: isolation, nucleotide sequence, chromosomal locus, and cryptic function. J. Bacteriol. **169**:771–778.
6. **Fahnestock, S. R., P. Alexander, J. Nagle, and D. Filpula.** 1986. Gene for an immunoglobulin-binding protein from a group G streptococcus. J. Bacteriol. **167**:870–880.
7. **Hamada, S., and H. D. Slade.** 1980. Biology, immunology, and cariogenicity of *Streptococcus mutans*. Microbiol. Rev. **44**: 331–384.
8. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
9. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**: 351–359.
10. **Himeno, T., T. Imanaka, and S. Aiba.** 1986. Nucleotide sequence of the penicillinase repressor gene *penI* of *Bacillus licheniformis* and regulation of *penP* and *penI* by the repressor. J. Bacteriol. **168**:1128–1132.
11. **Hollingshead, S. K., V. A. Fischetti, and J. R. Scott.** 1986. Complete nucleotide sequence of type 6M protein of the group A *Streptococcus*. J. Biol. Chem. **261**:1677–1686.
12. **Lepesant, J.-A., F. Kunst, M. Pascal, J. Lepesant-Kejzlarova, M. Steinmetz, and R. Dedonder.** 1976. Specific and pleiotropic regulatory mechanisms in the sucrose system of *Bacillus subtilis* 168, p. 58–69. *In* D. Schlessinger (ed.), Microbiology—1976. American Society for Microbiology, Washington, D.C.
13. **Loesche, W. J.** 1986. Role of *Streptococcus mutans* in human dental decay. Microbiol. Rev. **50**:353–380.
14. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53**:293–321.
15. **Reeves, R. E., and A. Sols.** 1973. Regulation of *Escherichia coli* phosphofructokinase *in situ*. Biochem. Biophys. Res. Commun. **50**:459–466.
16. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
17. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
18. **Sato, S., and H. K. Kuramitsu.** 1986. Isolation and characterization of a fructosyltransferase gene from *Streptococcus mutans* GS-5. Infect. Immun. **52**:166–170.
19. **Seki, T., H. Yoshikawa, H. Takahashi, and H. Saito.** 1987. Cloning and nucleotide sequence of *phoP*, the regulatory gene for alkaline phosphatase and phosphodiesterase in *Bacillus subtilis*. J. Bacteriol. **169**:2913–2916.
20. **Shimotsu, H., and D. J. Henner.** 1986. Modulation of *Bacillus subtilis* levansucrase gene expression by sucrose and regulation of the steady-state mRNA level by *sacU* and *sacQ* genes. J. Bacteriol. **168**:380–388.
21. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequence of *Escherichia coli* 16S ribosomal RNA; complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
22. **Shiroza, T., S. Ueda, and H. K. Kuramitsu.** 1987. Sequence analysis of the *gtfB* gene from *Streptococcus mutans*. J. Bacteriol. **169**:4263–4270.
23. **Steinmetz, M., D. Le Coq, S. Aymerich, G. Gonzy-Treboul, and P. Gay.** 1985. The DNA sequence of the gene for the secreted *Bacillus subtilis* enzyme levansucrase and its genetic control sites. Mol. Gen. Genet. **200**:220–228.
24. **Track, K. A., J. W. Chapman, P. J. Piggot, and J. A. Hoch.** 1985. Deduced product of the stage 0 sporulation gene *spoOF* shares homology with the *spoOA*, *ompR*, and *sfrA* proteins. Proc. Natl. Acad. Sci. USA **82**:7260–7264.
25. **Van Houte, J., and H. M. Jansen.** 1968. Levan degradation by streptococci isolated from human dental plaque. Arch. Oral Biol. **13**:827–830.
26. **von Heijne, G.** 1983. Patterns of amino acids near signal sequence cleavage sites. Eur. J. Biochem. **133**:17–21.
27. **Wenham, D. G., T. D. Hennessey, and J. A. Cole.** 1979. Regulation of glucosyl- and fructosyltransferase synthesis by continuous culture of *Streptococcus mutans*. J. Gen. Microbiol. **114**:117–124.