# Sequence Analysis of the Glucosyltransferase A Gene (*gtfA*) from Streptococcus mutans Ingbritt

JOSEPH J. FERRETTI.<sup>1\*</sup> TING-TING HUANG,<sup>1</sup> AND ROY R. B. RUSSELL<sup>2</sup>

Department of Microbiology and Immunology, University of Oklahoma Health Science Center, Oklahoma City, Oklahoma  $73190$ ,<sup>1</sup> and Dental Research Unit, Royal College of Surgeons of England, Downe, Kent BR6 7JJ, United Kingdom<sup>2</sup>

Received 16 December 1987/Accepted <sup>2</sup> March 1988

The complete nucleotide sequence of a 2.4-kilobase fragment containing the glucosyltransferase A gene  $(gt/4)$ of Streptococcus mutans Ingbritt (serotype c) was determined. The gtfA gene contains 481 codons and specifies a protein of molecular weight 55,665. There is no evidence of a signal peptide in the protein or that the glucosyltransferase A enzyme is secreted. No sequence homologies were observed between the gtfA gene or protein and the  $gtfI$  or  $gtfB$  gene and its protein.

Streptococcus mutans has been implicated as the major causative agent of dental caries (10, 11), and hence an appreciation of the ways in which it metabolizes sucrose is essential to our understanding of the disease process. S. mutans has numerous enzymes which utilize sucrose as a substrate, and there has been particular interest in the extracellular glucosyltransferases (GTF; EC 2.4.1.5) which form glucans containing various proportions of  $\alpha(1-3)$ - and  $\alpha(1-6)$ -linked glucose units (3). Although uncertainty remains about the precise number and nature of these enzymes, it is known that  $S$ . *mutans* and other oral streptococci commonly contain more than one GTF, and these have generally been found to have molecular weights of the order of 150,000 (reviewed in references 4 and 14). An exception is the GTF-A of S. mutans, with a molecular weight of 55,000. This enzyme was first described by Robeson et al. (19) after its expression from a recombinant plasmid in Escherichia coli. Subsequently, the  $gtA$  gene was cloned and expressed from a variety of other host-vector systems (1, 2, 17, 18, 20). The function of GTF-A in  $S$ . mutans, and its relationship with the other high-molecular-weight GTFs is unknown. In this study, we determined the nucleotide sequence of the  $gtf A$ gene from S. mutans Ingbritt, originally isolated from a recombinant gene bank constructed in bacteriophage lambda (20).

# MATERIALS AND METHODS

Bacteria and media.  $E.$  coli JM109 was the recipient used in all transfection experiments with bacteriophage M13 derivatives and transformation experiments with plasmid pUC18 derivatives (24). The growth medium used in all experiments was  $2 \times$  YT broth (15). Soft agar overlays consisted of  $2 \times$  YT broth supplemented with final concentrations of  $0.75\%$  agar,  $0.33$  mM isopropyl- $\beta$ -D-thiogalactopyranoside, and  $0.22\%$  5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside when differentiating between recombinant and nonrecombinant phage.

Enzymes and chemicals. Restriction enzymes, T4 DNA ligase, the Klenow fragment of DNA polymerase, and the M13 15-base-pair primer were purchased from Bethesda Research Laboratories, Inc., Gaithersburg, Md., and used as specified by the manufacturer. Deoxy- and dideoxynucleotide triphosphates were purchased from P-L Biochemicals, Inc., Milwaukee, Wis., and  $[\alpha^{-32}P]dATP$  was purchased from New England Nuclear Corp., Boston, Mass. Isopropyl-3-D-thiogalactopyranoside and 5-bromo-4-chloro- $3$ -indolyl- $\beta$ -D-galactopyranoside were purchased from Sigma Chemical Co., St Louis, Mo.

Subcloning of the *gtfA* gene and nucleotide sequencing. Bacteriophage lambda L47.1, containing the  $g$ tfA gene as previously described (20), was propagated, and DNA was isolated as described by Maniatis et al. (12). The lambda DNA containing  $g\,t\,f\,A$  was digested with  $E\,co\,RI$ , ligated to similarly cut pUC19, and transformed into E. coli JM109. The transformation mixture was spread onto M9 minimal medium (12) agar plates containing 4% sucrose. Only cells containing gtfA were capable of growing on these plates. A 2.4-kilobase DNA fragment containing the  $g$ tfA gene was obtained from plasmid pSF90, and it was unidirectionally digested with Bal31 and subcloned into mpl8 and mpl9 vectors as previously described (8, 9). The Sanger et al. dideoxy-chain termination method (21) was used to determine the nucleotide sequence. All sequences were confirmed from at least two overlapping clones, and the entire sequence was determined on both strands. The sequence information was analyzed by the James M. Pustell DNAprotein sequencing program obtained from International Biotechnologies, Inc.

Purification of GTF-A. GTF-A was purified from a sonic extract of E. coli RX4 grown in the presence of isopropyl-3-D-thiogalactopyranoside by a series of steps involving ammonium sulfate precipitation, gel filtration on Ultrogel AcA34, ion-exchange chromatography on DEAE-Trisacryl, chromatofocusing (the pl of GTF-A was found to be 4.1), and hydrophobic interaction chromatography in Phenyl-Sepharose (Pharmacia, Inc., Piscataway, N.J.).

**GTF-A assay.** Incorporation of radiolabel from  $[^{14}C]$ sucrose into an insoluble product was assayed by the thin-layer chromatography procedure described previously (20). When required, the inhibitor 5,5-dithiobis(2-nitrobenzoic acid) or parahydroxymercuribenzoate was added at a concentration of up to  $2 \times 10^{-3}$  M.

N-terminal sequence analysis. N-terminal sequence analysis of purified GTF was done with an Applied Biosystems 470A protein sequencer with an on-line 120A phenylthiohydantoin (PTH) analyzer in accordance with the instructions of the manufacturer.

<sup>\*</sup> Corresponding author.



FIG. 1. Partial restriction map of the 2,379-base-pair fragment containing the  $g t f A$  gene. The 5' end of the gene is at the left, and the <sup>3</sup>' end is at the right.

### RESULTS

Subcloning of the *gtfA* gene. The *gtfA* gene of  $S$ . *mutans* was previously cloned in bacteriophage lambda L47.1 (20) and was subcloned as a 2.4-kilobase EcoRI fragment into plasmid pUC19. Detection of  $E$ . coli containing the  $gtf A$ subclone was made possible by growth on minimal medium with sucrose as the sole carbon source. Since the E. coli strains containing the *gtfA* gene in either orientation in pUC19 grew in this medium, the fragment containing  $g\mathit{tfA}$ most likely contained its own transcription and translation signals. A physical map of the 2.4-kilobase EcoRI fragment containing *gtfA* is presented in Fig. 1.

Nucleotide sequence. The nucleotide sequence of the 2,379 base-pair fragment containing the  $g\,t\,f\,A$  gene was determined by the dideoxy-chain termination method of Sanger et al. (21) and is shown in Fig. 2. An open reading frame which starts at the ATG codon at nucleotide <sup>821</sup> and extends to the TAA termination codon at nucleotide <sup>2264</sup> specifies <sup>a</sup> protein of <sup>481</sup> amino acids with <sup>a</sup> molecular weight of 55,665. A putative ribosome-binding site sequence (AGAGGA) is located 6 nucleotides upstream of the translation initiation codon. Further upstream is a probable  $-10$  region (TA-TAAA) separated by 16 nucleotides from a possible  $-35$ region (CTGTTA).

Immediately following the *gtfA* termination codon are another possible ribosome-binding site sequence and a translation initiation codon for another protein. This sequence appears to be under transcriptional control of the  $g t f A$  gene and may be the beginning of the sequence coding for a 38-kilodalton protein identified by Burne et al. (2) to follow the  $gtfA$  gene immediately. Upstream of the  $gtfA$  gene is a portion of another reading frame which specifies a partial protein of 207 amino acids.

Amino acid sequence. The deduced amino acid sequence of GTF-A in the N-terminal region did not contain a sequence comparable to the signal peptides found in other secreted proteins (16). N-terminal amino acid analysis of purified GTF-A revealed the first 19 amino acids to be Pro-Ile-Thr-Asn-Lys-Thr-Met-Leu-Ile-Thr-Tyr-Ala-Asp-Ser-Leu-Gly-Lys-Asn-Leu. This sequence is identical to that deduced from the nucleotide sequence, except for the N-terminal methionine, which was apparently removed by an aminopeptidase. The deduced amino acid composition of GTF-A consisted of 39% nonpolar, 31% polar, 16% acidic, and 13% basic amino acids (Table 1). A single cysteine residue is present in GTF-A, though there is no evidence that it is important for its activity, since it was not inhibited by the sulfhydryl inhibitors 5,5-dithiobis(2-nitrobenzoic acid) and parahydroxymercaribenzoate. A comparison of the GTF-A sequence with previously described GTF-I (9) and GTF-B (22) sequences by the PRTALN program of Wilbur and Lipman (23) showed no homology with either protein.

## DISCUSSION

The *gtfA* gene and product described in this study appear to be identical to those previously described by Robeson et

TABLE 1. Amino acid composition of GTF-A based on the nucleotide sequence of *gtfA* 

Amino acid	No. of residues
	33
	19
	28
	43
	1
	36
	16
	23
	11
	33
	41
	33
	Q <sub>0</sub>
	24
	15
	22
	29
	6
	31
	28

" This number includes the N-terminal methionine.

al. (19), Pucci and Macrina (17, 18), and Burne et al. (1, 2) on the basis of similarities in the restriction endonuclease maps and the molecular weight of the product. The  $gtf A$  gene fragment contains probable transcription and translation initiation sequences similar to those described for other streptococcal genes (7). The deduced amino acid sequence of GTF-A has a molecular weight of 55,665, in close agreement with previous estimates (1, 17, 19, 20).

In S. mutans, GTF-A has been reported to be primarily cell associated (19), though it can also be detected in culture supernatants by rocket immunoelectrophoresis or Western blotting (immunoblotting) (19, 20). Robeson et al. estimated that 15% of the total GTF-A activity was extracellular, whereas a further 5% could be recovered from the cell surface by washing with <sup>1</sup> M NaCl. Pucci and Macrina (17) introduced  $g\{fA}$  on a shuttle vector into S. sanguis and found a change in extracellular polymer synthesis, suggesting that the enzyme was also secreted in S. sanguis. No such secretion, however, was found when *gtfA* was expressed in Bacillus subtilis (1). When expressed in E. coli, a proportion of the enzyme is secreted through the cytoplasmic membrane into the periplasmic space (5, 19), but the mechanism by which this takes place is unknown-no difference in molecular weight between the intracellular and periplasmic forms was observed (6). This observation is consistent with our failure to detect a region of the  $g$ tfA gene which might code for a sequence of amino acids resembling that found in typical procaryotic signal peptides. Other streptococcal extracellular proteins, including GTF (9, 22), have been found to be synthesized with signal peptides which are cleaved during the secretion process.

Previous investigators have established that regions of DNA adjacent to *gtfA* encode proteins (2, 18). Burne et al. (2) demonstrated the existence of a dextranase gene lying downstream of *gtfA* and transcribed in the same direction, though apparently under the control of its own promoter and ribosome-binding site. They also reported finding an invertase activity in the same region and speculated that there might exist an operonlike arrangement of genes for enzymes involved in sucrose metabolism. Our own previous observations also indicated the linkage of  $g\mathit{tfA}$  with invertase, since

10 20 30 40 50 60  $ATT$   $ACT$   $GTT$   $TTA$   $TCG$   $GTT$ <sup>N</sup> <sup>S</sup> <sup>T</sup> <sup>V</sup> <sup>I</sup> <sup>T</sup> <sup>V</sup> <sup>L</sup> <sup>S</sup> <sup>V</sup> <sup>L</sup> <sup>V</sup> <sup>V</sup> <sup>L</sup> <sup>F</sup> <sup>I</sup> <sup>P</sup> A <sup>A</sup> 70 80 90 100 110 120 TCC ATT GOG CGT AAC ATG TCT AGA AGA AGA AAA GCT TTC AAT ATC ATG TAT AGC CTG<br>SIARN MSRRKAFNIMSL A Y S <sup>I</sup> A R N M S R R K A F N <sup>I</sup> Y S L 130 140 150 160 170 180 TTG AT! CTG 0GA ATT TIC GOT CCT TTC CAG GTT ATC ATG ATT COT ATT ACG GTT ATG ATG L <sup>I</sup> L G <sup>I</sup> F V P Q V <sup>I</sup> M <sup>I</sup> P <sup>I</sup> T V H 190 200 210 220 230 240 AGT MA TrA G0C TTG GCT MT ATG TOG GGA TTA AT! ATT CTG TAT TTA ACT TAT GCT ATT S K L G L A N M W G L I I L Y L T Y A I 250 260 270 280 290 300 CCA CAG ACA CTC TTC CTT TAC GTG GGT TAT ATC AAA CTA AGT GTA CCT GAT AGT TTA GAT P Q T L F L Y V G Y <sup>I</sup> K L S V P D S L D 310 320 330 340 350 360 GCT GAA ATT GAT GOT GOG GAT AAA TTG ACA ACT TAT OGT AAA ATC ATT TTC CCT ATT GAT GGT GOG GAT<br>I D G A D 370 380 390 400 410 420 ATG TTA AAA CCA ATG CAC GCA ACA ACT CTG ATT ATT AAT GCA CTT TGG TTC TGG AAC GAT <sup>M</sup> <sup>L</sup> <sup>K</sup> <sup>P</sup> M <sup>H</sup> <sup>A</sup> <sup>T</sup> <sup>T</sup> <sup>L</sup> <sup>I</sup> <sup>I</sup> <sup>N</sup> A <sup>L</sup> <sup>W</sup> <sup>F</sup> <sup>W</sup> <sup>N</sup> <sup>D</sup> 430 440 450 460 470 480 TTr ATG TTG CCA TTG CEG AT! CTT MT MG GAT TCA AG6 ATG T!G ACG CT! CCT CTT TTC F <sup>L</sup> P L L <sup>I</sup> <sup>L</sup> N K D S S W T <sup>L</sup> P L F 490 500 S10 520 530 540 CM TAC MT TAT AGC G0A CM TAT TTC MT GAT TAC GOT COT AGT OTT GCT TCT TAC ATT Q Y N Y S G Q Y <sup>F</sup> N D Y G P S F A S Y <sup>I</sup> 550 560 570 580 590 600 GTT GGT AT! ATT ACC AT! ACA ATT GTT TAT CTT ATT TOC CM A CAC ATT ATT GOCT GOT TT GGT ATT ATT ACC ATT ACA ATT GTT TAT CTT ATT TTC CAA AAA CAC ATT ATT GCT GGT<br>VGIIT II VYLIFQ KHIIA G 610 620 630 640 650 660 GTG AAG TGA GAG CTA TAG ATA ATG ACC GCT TAA GCA GCT CTT CAT ATG AGC AAT GGA 670 680 690 700 710 720 CAA CTA AGG CCA GCC GCT ATG GOG GTC TGC CTA AGC TTT CCA CTA GCT CGC CAC CCG CAA 730 740 750 760 770 780 TAA GAA CGA TTA CTT CTG GCT CGC GTC GTA GGA GCT AAA GAT GTG GCT GTT ATT TTT AGG 790 800 810 820 830 840 TTG AAC TGG TAT AAA CCA AAA TTA ATT AGA GGA GAT AAA ATG CCA ATT ACA H P <sup>I</sup> T N K T 850 860 870 880 890 900 ATG TTG ATT ACT TAC GCA GAC AGT TTG GGT AAA AAT TTG AAA GAA TTG AAT GAA AAT ATT M L <sup>I</sup> T <sup>Y</sup> <sup>A</sup> D S <sup>L</sup> G <sup>N</sup> <sup>L</sup> <sup>K</sup> E <sup>L</sup> <sup>N</sup> E <sup>N</sup> <sup>I</sup> 910 920 930 940 950 960 GAG MT TAT TOT G0A GAT GCT OTT G0C G0T GTC CAT TTG CEG CCA TTC TTT CCT TCC ACA E N Y F G D A V G G V H L L P <sup>F</sup> P S T 970 980 990 1000 1010 1020 GGT GAT CGT GGC TIT GCA CCG ATT GAT TAC CAT GAA GTT GAC TCT GCT TTT GCC GAT GAT TAC CAT GAA GTT GAC TCT GCT TTT GCC G D R G F A P I D Y H E V D S A F G D W 1030 1040 1050 1060 1070 1080 O,AT GAT 0TC AM CGC TOG GOT C-A MA TAT TAC CTC ATG TOT OAT TTC AT! AT! AAT CAT D D V K R <sup>L</sup> G E K Y <sup>Y</sup> L F D F <sup>I</sup> N H 1210 1220 1230 1240 1250 1260 GTG GAC CTG ATT TAT AAG CGT AAG GAT 0GA GCA CCT MA CAG GM ATC CM mTT GCA OAT V D L <sup>I</sup> YO R <sup>K</sup> D R A P K Q E <sup>I</sup> Q F A D 1270 1280 1290 1300 1310 1320 GGC AGT GTT GAA CAT CTC TOG AAC ACT TTT GOG GAG GAA CAG ATT GAT CTT GAC GTG ACT<br>G S V E H L W N T F G E E Q I D L D V T G <sup>S</sup> <sup>V</sup> E R L W N T F G E E Q <sup>I</sup> D L D <sup>V</sup> T 1330 1340 1350 1360 1370 1380 CGC TCT ACC ATT GAA AAT TTA GCA GCC AAC GGC TGT<br>R S T I E N L A A N G C <sup>K</sup> <sup>E</sup> <sup>V</sup> <sup>T</sup> N <sup>D</sup> <sup>F</sup> <sup>I</sup> <sup>R</sup> <sup>S</sup> <sup>T</sup> <sup>I</sup> <sup>E</sup> <sup>N</sup> <sup>L</sup> <sup>A</sup> A <sup>N</sup> <sup>G</sup> <sup>C</sup> 1390 1400 1410 1420 1430 1440 GAT CTC ATT CGT TTG GAT GCC TTT GCT TAT GCT GTT AAA AAG CTA GAT ACG AAT GAT TTC D L <sup>I</sup> R L D A F A Y A V K K L D T N D F 1450 1460 1470 1480 1490 1500 TOT GTr GM CCE GM ATC TOG ACE CTG CTA GAT MA GOT CGT OAT ATA GCT GCT GTA TOG <sup>F</sup> <sup>V</sup> <sup>E</sup> <sup>P</sup> <sup>E</sup> <sup>I</sup> <sup>W</sup> <sup>T</sup> <sup>L</sup> <sup>L</sup> <sup>D</sup> <sup>K</sup> <sup>V</sup> <sup>R</sup> <sup>N</sup> <sup>I</sup> <sup>A</sup> A <sup>V</sup> <sup>S</sup> 1510 1520 1530 1540 1550 1560 GGT GCG GAA ATC TTG CCG GAA ATT CAT GAA CAC TAT ACT ATT CAA TTT AAA ATT GCA GAC GAC AACT ATT CAA TE AMA ATT CACA GAC GAC AMA ATT CAA THE HE HE Y TE I Q F K I AD G A E <sup>I</sup> L P E <sup>I</sup> H E H Y T <sup>I</sup> Q F K <sup>I</sup> A D 1570 1580 1590 1600 1610 1620 CAT GAT TAC TAT GTT TAT GAT TTT GCC CTG CCT ATG GTG AGG CTC TAC AGC CTA TAT TOG<br>H D Y Y Y Y D F A L P N V T L Y S L Y S 1630 1640 1650 1660 1670 1680 GGC AAG GTT GAC CGT CTT GCC AAA TGG CTG AAA ATG AGT CCG ATG AAA CAG TTC ACC ACC G <sup>K</sup> <sup>V</sup> <sup>D</sup> <sup>R</sup> L <sup>A</sup> <sup>K</sup> W <sup>L</sup> <sup>K</sup> H <sup>S</sup> <sup>P</sup> 8 <sup>K</sup> <sup>Q</sup> <sup>F</sup> <sup>T</sup> <sup>T</sup> 1690 1700 1710 1720 1730 1740 CTT GAT ACA CAT GAC GGT AT! GGT GTG GTT GAT GTT MG GAT ATC TTG ACE GAC GAA GAA L D T H D G <sup>I</sup> G V V D V K D <sup>I</sup> L T D E E 1750 1760 1770 1780 1790 1800 ATT ACE TAT ACT TCT MT GAG COT TAT MG GTC GGT GCC MT GTC AAT CGT MG TA.T TOA I T Y T S N E L Y K V G A N V N R K Y S 1810 1820 1830 1840 1850 1860 ACT GCC GAA TAT AAT AAC TTG GAT ATC TAT CAA ATT AAI TCA ACT TAC TAT TCA GCA CTT<br>T A E Y N N L D I Y Q I N S T Y Y S A L 1870 1880 1890 1900 1910 1920 GGT GAT GAT GAT CAA AAA TAC TTT TTG GCC CGC TTG ATA CAA GCT TTT GCT CCA GGT ATT G <sup>D</sup> I <sup>D</sup> Q K <sup>Y</sup> F L A <sup>R</sup> L i Q A <sup>F</sup> A <sup>P</sup> G <sup>I</sup> 1930 1940 1950 1960 1970 1980 CCA CAG GTr TAT TAC GTr GGC Tmr TTA GCT 00GGC G MT GAT CTT GM TTA CTG GM AGC P Q V Y Y V G F L A G K N D L E L L E <sup>S</sup> 1990 2000 2010 2020 2030 2040 ACE MA GAA G0C 0C0 MT ATC M COGT CAT TAT TAT AGT AGT GAA GM AT! GCT AAG GM <sup>T</sup> <sup>K</sup> <sup>E</sup> <sup>G</sup> <sup>R</sup> <sup>N</sup> <sup>I</sup> <sup>N</sup> <sup>R</sup> <sup>H</sup> <sup>Y</sup> <sup>Y</sup> <sup>S</sup> <sup>S</sup> <sup>E</sup> <sup>E</sup> <sup>I</sup> A <sup>K</sup> <sup>E</sup> 2050 2060 2070 2080 2090 2100 GTG AAG OGA OCA GTT GTC AAG GCA CTT TTA AAT CTC TTT ACT TAC OGC AAT CAG TCA GCA V K R P V V K A L L N L F T Y R N Q S A 2110 2120 2130 2140 2150 2160 GCT TTT GAT TTG GAT GGC CGT ATT GAA GTG GAA ACG CCA AAT GAA GCG ACC ATT GTC ATA A F D L D G R <sup>I</sup> E V E T P N E A T <sup>I</sup> V <sup>I</sup> 2170 2180 2190 2200 2210 2220 GM OTT CAA MT MA GAT GGC AGT CAT ATC GCA ACA GCA GAG ATT MT CTC CM GAT ATG E <sup>R</sup> <sup>Q</sup> <sup>N</sup>R <sup>D</sup> <sup>G</sup> <sup>S</sup> <sup>H</sup> <sup>I</sup> <sup>A</sup> T A <sup>E</sup> <sup>I</sup> <sup>N</sup> <sup>L</sup> <sup>Q</sup> ID 2230 2240 2250 2260 2270 2280  $A$ CA TACA AGA AGA ATA AGC TTIGAA TAA TAA GGA GAG AGA CE

screening of our bank of S. mutans chromosomal DNA binding site and an ATG translation initiation codon are cloned into bacteriophage lambda L47.1 revealed that ap-<br>found within 17 nucleotides of the C-terminal codon of gt cloned into bacteriophage lambda L47.1 revealed that ap-<br>proximately 50% of recombinants carrying *gtfA* could cleave The second protein appears to be under transcriptional proximately 50% of recombinants carrying *gtfA* could cleave The second protein appears to be under transcriptional raffinose as well as sucrose (M. L. Gilpin and R. R. B. control of *gtfA*, since promoterlike sequences a raffinose as well as sucrose (M. L. Gilpin and R. R. B. control of  $g(t)A$ , since promoterlike sequences are lacking in Russell, unpublished data). The results from this study the noncoding region between genes. The identity and role of support the notion of a  $gtfA$  operon, since a ribosome-<br>this second gene product are presently under investigation.

sequences are underlined.

1090 1100 1110 1120 1130 1140 AT! TCG CGT CAG TCT MA TAT TAT MA GAT TAC CM GM MAG CAT GM 0CA A60 GCT TAT <sup>I</sup> <sup>S</sup> <sup>R</sup> Q <sup>S</sup> K Y Y K D Y Q E K H E A <sup>S</sup> A Y 1150 1160 1170 1180 1190 1200 AM GAT CTA TOT TrA MT TGG GAT MMA TO TOG CCT MA MT CGC CCG ACA CM GM GAT GAT AAA TTT TGG CCT AAA DE K F W P K

T Y R V T E N D Q T <sup>I</sup> S F E -

2350 2260 2270 GTT GM GAT T!T GAC TTG GAT ATC MG MC MA GAA TTC V E D F D L D I K N K E F

FIG. 2. Nucleotide sequence of the gtfA gene and flanking regions. Numbering begins at the 5' end of the sequence. The deduced amino acid sequence of GTF-A and the flanking proteins is given below the nucleotide sequence. Putative promoter and ribosome-binding site

2290 2300 2310 2320 2330 2340 ATG GTT GMA CTA MT TTA MT CAC ATT TAT MA MA TAC CCT MC AGC AGT CAT TAC TCT H V E L N L N H <sup>I</sup> Y K K Y P N S S H Y S

The overall amino acid composition and sequence of GTF-A are quite different from those of either GTF-B or GTF-I, GTFs that produce insoluble extracellular glucans, and preclude the possibility that GTF-A is a partially duplicated or truncated form of either of these GTFs. Interestingly, GTF-A contains a single cysteine residue, though its role, if any, in enzyme activity has not been defined. Martin et al. (13) have noted that there is a highly conserved amino acid sequence around a single cysteine residue in three other enzymes concerned with the metabolism of sucrose or sucrose-derived polymers (B. subtilis levanase and sucrase and Saccharomyces cerevisiae invertase), but there is no similarity between this sequence and that of GTF-A. A truncated derivative of gtfA (obtained by Bal31 digestion) missing the coding region for the C-terminal 20 amino acids produces a protein unable to synthesize glucan; thus, the missing region appears to be essential for glucan synthesis activity.

The role of GTF-A in cell physiology remains unknown. However, in view of the data on the intracellular location of GTFA and its lack of similarity to other known GTFs, it seems likely that it is involved in some metabolic pathway distinct from extracellular polysaccharide synthesis.

#### ACKNOWLEDGMENTS

We thank David R. Lorenz for his technical expertise and helpful discussions throughout this investigation.

This research was supported by Public Health Service grant DE 08191 from the National Institutes of Health and by the Medical Research Council.

#### LITERATURE CITED

- 1. Burne, R. A., B. Rubinfeld, W. H. Bowen, and R. E. Yasbin. 1986. Cloning and expression of a Streptococcus mutans glucosyltransferase gene in Bacillus subtilis. Gene 47:201-209.
- 2. Burne, R. A., B. Rubinfeld, W. H. Bowen, and R. E. Yasbin. 1986. Tight genetic linkage of a glucosyltransferase and dextranase of Streptococcus mutans GS-5. J. Dent. Res. 65: 1392-1401.
- 3. Chassy, B. M. 1983. Sucrose metabolism and glucosyltransferase activity in oral streptococci, p. 3-10. In R. J. Doyle and J. E. Ciardi (ed.), Glucosyltransferases, glucans, sucrose and dental caries. Spec. Suppl. Chemical Senses. IRL Press, Washington, D.C.
- 4. Ciardi, J. E. 1983. Purification and properties of glucosyltransferases of Streptococcus mutans: a review, p. 51-64. In R. J. Doyle and J. E. Ciardi (ed.), Glucosyltransferases, glucans, sucrose and dental caries. Spec. Suppl. Chemical Senses. IRL Press, Washington, D.C.
- 5. Curtiss, R. 1985. Genetic analysis of Streptococcus mutans virulence. Curr. Top. Microbiol. Immunol. 118:253-275.
- 6. Curtiss, R., III, J. P. Robeson, R. Barletta, Y. Abiko, and M. Smorawinska. 1982. Synthesis and function of Streptococcus mutans cell surface proteins in Escherichia coli, p. 253-277. In D. Schlessinger (ed.), Microbiology-1982. American Society for Microbiology, Washington, D.C.
- 7. Ferretti, J. J., and R. Curtiss III. 1987. Compilation of nucleo-

tide sequences that signal the initiation of transcription and translation in streptococci, p. 293. In J. J. Ferretti and R. Curtiss III (ed.), Streptococcal genetics. American Society for Microbiology, Washington, D.C.

- 8. Ferretti, J. J., K. S. Gilmore, and P. Courvalin. 1986. Nucleotide sequence analysis of the gene specifying the bifunctional 6'-aminoglycoside acetyltransferase 2"-aminoglycoside phosphotransferase enzyme in Streptococcus faecalis and identification and cloning of gene regions specifying the two activities. J. Bacteriol. 167:631-638.
- 9. Ferretti, J. J., M. L. Gilpin, and R. R. B. Russell. 1987. Nucleotide sequence of the glucosyltransferase gene from Streptococcus sobrinus MFe28. J. Bacteriol. 169:4271-4278.
- 10. Hamada, S., and H. D. Slade. 1980. Biology, immunology, and cariogenicity of Streptococcus mutans. Microbiol. Rev. 44:331-384.
- 11. Loesche, W. J. 1987. Role of Streptococcus mutans in human dental decay. Microbiol. Rev. 50:353-380.
- 12. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 13. Martin, J. E., M. Debarbouille, E. Ferrari, A. Klier, and G. Rappaport. 1987. Characterization of the levanase gene of Bacillus subtilis which shows homology to yeast invertase. Mol. Gen. Genet. 208:177-184.
- 14. Mukasa, H. 1986. Properties of Streptococcus mutans glucosyltransferases, p. 121-132. In S. Hamada, S. Michalek, H. Kiyono, L. Menaker, and J. R. McGhee (ed.), Molecular microbiology and immunobiology of Streptococcus mutans. Elsevier Science Publishing, Inc., New York.
- 15. Muller-Hill, B., L. Crapo, and W. Gilbert. 1968. Mutants that make more *lac* repressor. Proc. Natl. Acad. Sci. USA 59:1259-1264.
- 16. Oliver, D. 1985. Protein secretion in Escherichia coli. Annu. Rev. Microbiol. 39:615-648.
- 17. Pucci, M., and F. L. Macrina. 1985. Cloned gtfA gene of Streptococcus mutans LM7 alters glucan synthesis in Streptococcus sanguis. Infect. Immun. 48:704-712.
- 18. Pucci, M. J., and F. L. Macrina. 1986. Molecular organization of the gtfA gene of Streptococcus mutans LM7. Infect. Immun. 54:77-84.
- 19. Robeson, J. P., R. G. Barletta, and R. Curtiss III. 1983. Expression of a Streptococcus mutans glucosyltransferase gene in Escherichia coli. J. Bacteriol. 153:211-221.
- 20. Russell, R. R. B., P. Morrissey, and G. Dougan. 1985. Cloning of sucrase genes from Streptococcus mutans in bacteriophage lambda. FEMS Microbiol. Lett. 30:37-41.
- 21. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- 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. Wilbur, W. J., and D. J. Lipman. 1982. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA 80:726-730.
- 24. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13 mp18 and pUC19 vectors. Gene 33:103-119.