INFECTION AND IMMUNITY, June 1988, p. 1585–1588
0019-9567/88/061585-04$02.00/0

Vol. 56, No. 6

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

JOSEPH J. FERRETTI,[1]* TING-TING HUANG,[1] AND ROY R. B. RUSSELL[2]

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

Received 16 December 1987/Accepted 2 March 1988

The complete nucleotide sequence of a 2.4-kilobase fragment containing the glucosyltransferase A gene (*gtfA*) 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 α(1-3)- and α(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 *gtfA* 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 *gtfA* 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× YT broth (15). Soft agar overlays consisted of 2× YT broth supplemented with final concentrations of 0.75% agar, 0.33 mM isopropyl-β-D-thiogalactopyranoside, and 0.22% 5-bromo-4-chloro-3-indolyl-β-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 [α-^{32}P]dATP was purchased from New England Nuclear Corp., Boston, Mass. Isopropyl-β-D-thiogalactopyranoside and 5-bromo-4-chloro-3-indolyl-β-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 *gtfA* gene as previously described (20), was propagated, and DNA was isolated as described by Maniatis et al. (12). The lambda DNA containing *gtfA* was digested with *Eco*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 *gtfA* gene was obtained from plasmid pSF90, and it was unidirectionally digested with *Bal*31 and subcloned into mp18 and mp19 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 DNA-protein 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-β-D-thiogalactopyranoside by a series of steps involving ammonium sulfate precipitation, gel filtration on Ultrogel AcA34, ion-exchange chromatography on DEAE-Trisacryl, chromatofocusing (the pI 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 2×10^{-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.

* Corresponding author.

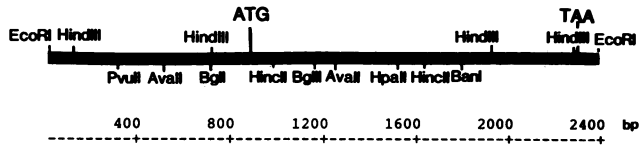

FIG. 1. Partial restriction map of the 2,379-base-pair fragment containing the *gtfA* gene. The 5′ end of the gene is at the left, and the 3′ end is at the right.

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 *Eco*RI fragment into plasmid pUC19. Detection of *E. coli* containing the *gtfA* 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 *gtfA* most likely contained its own transcription and translation signals. A physical map of the 2.4-kilobase *Eco*RI fragment containing *gtfA* is presented in Fig. 1.

Nucleotide sequence. The nucleotide sequence of the 2,379-base-pair fragment containing the *gtfA* 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 821 and extends to the TAA termination codon at nucleotide 2264 specifies a protein of 481 amino acids with a 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 (TATAAA) 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 *gtfA* 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.

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

Amino acid	No. of residues
Alanine	33
Arginine	19
Asparagine	28
Aspartic acid	43
Cysteine	1
Glutamic acid	36
Glutamine	16
Glycine	23
Histidine	11
Isoleucine	33
Leucine	41
Lysine	33
Methionine	9[a]
Phenylalanine	24
Proline	15
Serine	22
Threonine	29
Tryptophan	6
Tyrosine	31
Valine	28

[a] This number includes the N-terminal methionine.

DISCUSSION

The *gtfA* gene and product described in this study appear to be identical to those previously described by Robeson et 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 *gtfA* 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 1 M NaCl. Pucci and Macrina (17) introduced *gtfA* 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 *gtfA* 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 *gtfA* with invertase, since

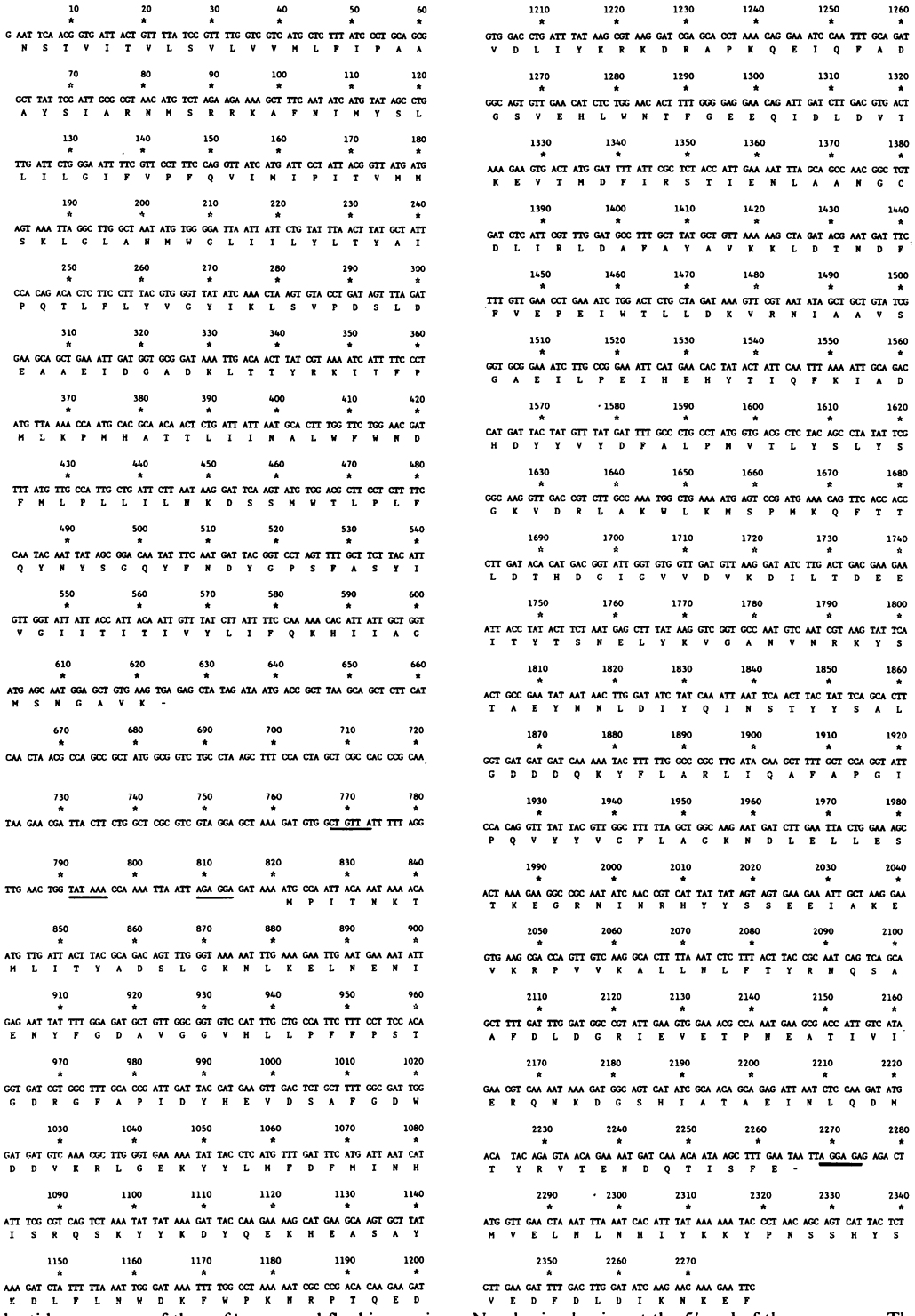

FIG. 2. Nucleotide sequence of the *gtfA* gene and flanking regions. Numbering begins at the 5′ end of the sequence. The deduced amino acid sequence of GTF-A and the flanking proteins is given below the nucleotide sequence. Putative promoter and ribosome-binding site sequences are underlined.

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

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 *Bal*31 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 GTF-A 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 nucleotide 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.