Analysis of the *Streptococcus downei* gtfS Gene, Which Specifies a Glucosyltransferase That Synthesizes Soluble Glucans

KEETA S. GILMORE, ROY R. B. RUSSELL, AND JOSEPH J. FERRETTI*

Department of Microbiology & Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73190, and Hunterian Dental Research Unit, London Hospital Medical College, London E1 2AD, United Kingdom

Received 9 February 1990/Accepted 24 April 1990

The complete nucleotide sequence was determined for the *Streptococcus downei* (previously *Streptococcus sobrinus*) MFe28 gtfS gene which specifies a glucosyltransferase (GTF-S) producing water-soluble glucan. A single open reading frame which encodes a mature protein with a molecular weight of 147,408 (1,328 amino acids) and a putative signal peptide 36 or 37 amino acids in length was detected. GTF-S shares extensive sequence similarity with GTF-I (gtf) from *S. downei* and GTF-I (gtfB) and GTF-S (gtfC) from *Streptococcus mutans*. GTF-S contains a highly conserved enzymatic domain and C-terminal repeated sequences which appear to be involved in glucan binding. Comparison of the deduced GTF-S protein sequence with other sequenced GTF genes of mutants streptococci revealed that these C-terminal repeats occurred in all cases, although the patterns of repeated sequences varied with respect to each other and to the glucan-binding protein of *S. mutans*. GTF-S contains four C-terminal repeat sequences ranging from 49 to 51 amino acids in length and a partial repeat of 13 amino acids. Nuclear magnetic resonance analysis of the glucan produced by GTF-S revealed that the product consisted of more than 90% α-1,6-linked glucosyl residues.

*Streptococcus mutans* is a principal etiologic agent of dental caries (8, 13). The virulence of this organism is in part due to the production of a group of enzymes called glucosyltransferases (GTFs). These extracellular enzymes cleave dietary sucrose and polymerize the resulting glucose moiety to form water-soluble and/or water-insoluble glucans, which are important components of dental plaque (8).

Since the initial observation that a single *S. mutans* strain can produce a number of electrophoretically distinct GTFs (7), several laboratories have isolated and characterized multiple GTFs from various mutants streptococci (4, 6, 17, 24). These GTFs fall into three categories: (i) GTFs which produce water-soluble, primarily α-1,6-linked glucan (GTF-S); (ii) GTFs which produce water-insoluble, primarily α-1,3-linked glucan (GTF-I); and (iii) GTFs which produce a combination of water-soluble and-insoluble glucans (GTF-SI). The insoluble glucan produced primarily by GTF-I mediates the sucrose-dependent attachment of *S. mutans* to the smooth surfaces (8). The contribution of the water-soluble glucan to the dental caries process is unknown, but it causes aggregation of certain bacteria and can serve as an extracellular energy store. Many GTF-IIs produced by cariogenic species of mutants streptococci require a primer for glucan synthesis (24), whereas certain GTF-Ss have been shown to be primer independent (24). Such GTF-Ss may provide the primer necessary for GTF-I-mediated insoluble-glucan synthesis. Therefore, soluble-glucan-forming GTF-Ss may play a central although indirect role in bacterial adherence and colonization of the oral cavity.

Genes encoding GTF-I and GTF-S (a primer-independent GTF which catalyzes formation of a soluble glucan) from the cariogenic mutants streptococcus *Streptococcus downei* (23) (formerly *Streptococcus sobrinus*) MFe28 have been cloned in *Escherichia coli* (6), and the nucleotide sequence for the gene encoding GTF-I has been determined (3). The nucleotide sequence has also been determined for genes from *S. mutans* GSS which encode GTF-I and GTF-SI (gtfB [19] and gtfC [21], respectively). Repeated amino acid sequence motifs have been observed to occur in the amino acid sequences inferred for all of these proteins. More recently, similar repeats have also been observed to occur in a glucan-binding protein (GBP) derived from *S. mutans* Ingbritt (1). This protein has no GTF activity, thus providing evidence that the repeated sequences may be involved in glucan binding by GTFs.

Because of differences in the solubility of the glucan produced and in primer requirements, it was of interest to characterize the gene encoding a primer-independent GTF at the molecular level. In this communication, we report the complete nucleotide sequence of gtfS from *S. downei* MFe28 and compare the deduced amino acid sequence of GTF-S with those of other enzymes involved in sucrose metabolism by mutants streptococci.

**MATERIALS AND METHODS**

**Bacteria and media.** The gtfS gene was cloned from *S. downei* MFe28, a serotype h strain (23). The 8-kilobase (kb) fragment containing the gtfS was cloned into the plasmid vector pACYC184, resulting in construction of pMLG60, which was transformed into *E. coli* JM109 (25). Fragments of the 8-kb pMLG60 insert were cloned into M13 bacteriophage vectors mp18 and mp19 (25) for DNA sequencing. Detection of transfectants was accomplished by using 2× YT broth (15) with 0.75% agar, 0.33 mM isopropyl-β-D-thiogalactopyranoside (Sigma Chemical Co., St. Louis, Mo.), and 0.02% 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Sigma). Recombinant phages were harvested from 2× YT broth cultures of infected JM109.

**Enzymes and chemicals.** Restriction enzymes, exonuclease III, M13 vectors, and T4 DNA ligase were purchased from either Bethesda Research Laboratories, Inc. (Gaithersburg, Md.), or Fisher Scientific Co. (St. Louis, Mo.) and were used in accordance with the specifications of the manufacturer. DNA sequence reactions were performed with a T7 DNA polymerase sequencing kit (Pharmacia, Inc., Piscataway, N.J.). Custom oligonucleotide primers were produced...
by the University of Oklahoma Health Sciences Center Molecular Biology Core Resource Facility. M13 forward primer (17-mer) was purchased from Promega Biotech (Madison, Wis.), and [α-35S]dATP was purchased from Dupont, NEN Research Products (Boston, Mass.).

**Nucleotide sequence determination.** The complete nucleotide sequence for both DNA strands of gtfS was determined by using the dideoxy-chain termination method as modified by Tabor and Richardson (20). Nucleotide sequence determinations were made for cloned fragments and nested deletion derivatives (11). The 0.6-kb XbaI-SsrI, 0.8-kb PstI-XbaI, 0.75-kb HindIII, 3.0-kb SphI-SsrI, and 1.6-kb EcoRI-SphI restriction fragments of pMLG60 were cloned into the vectors M13mp18 and M13mp19. Gaps in the resulting nucleotide sequence were closed by the use of custom synthesized oligonucleotides to prime the chain termination reactions. The sequence information was analyzed by using the James M. Pustell DNA-protein-sequencing program from International Biotechnologies (New Haven, Conn.) and programs from the University of Wisconsin Genetics Computer Group version 6.1. The GenBank/EMBL accession number for gtfS is M30943.

**Glucan isolation.** Cells, centrifugally harvested from an overnight culture of *E. coli* (pMLG60), were disrupted with a French pressure cell (American Instrument Co., Inc., Silver Spring, Md.). Cell lysate (1 ml) was diluted 1:100 in 0.05 M NaH2PO4 (pH 6.5)-5% sucrose-0.01% ethylmercurithiosalicylate. The diluted lysate was incubated at 37°C for 48 h. After incubation, 3 volumes of 95% ethanol were added and allowed to precipitate at −20°C overnight. Precipitated glucan was collected by centrifugation at 16,270 × g for 15 min. The glucan was suspended in 100 ml of water and precipitated twice as described above. After the third precipitation, the glucan pellet was air dried.

**NMR analysis.** The composition of the glucan was analyzed by 13C-nuclear magnetic resonance (13C-NMR) spectrometry. The glucan was assayed in D2O at 20°C. The spectra were obtained at 125 MHz on a VXRS00 spectrometer (Varian, Palo Alto, Calif.) and also at 75 MHz on an XL300 spectrometer (Varian). Chemical shifts were measured (and reported in parts per million) downfield from sodium-3-trimethyl-silyl propionate, the internal standard. Assignment of peaks was based on the report by Colson et al. (2).

**RESULTS**

**Subcloning of gtfS into M13.** The gtfS gene was originally cloned into the bacteriophage lambda replacement vector λA71.1 (6, 12). An 8-kb EcoRI fragment containing the gene was then subcloned into pACYC184, and the resulting construct was termed pMLG60 (Fig. 1). The gtfS gene was found to be located within a 4.7-kb EcoRI-SsrI fragment. This 4.7-kb EcoRI-SsrI fragment was further digested into smaller restriction fragments and cloned into M13mp18 and M13mp19 for sequencing in both directions. In some cases, custom oligonucleotide primers were synthesized to extend the sequences and to determine the nucleotide sequence at restriction fragment junctions.

**Nucleotide sequence.** The nucleotide sequence was determined in both directions and is shown in Fig. 2. The gtfS gene is encoded by a 4,094-base-pair open reading frame. This open reading frame is preceded by a putative ribosome-binding-site sequence (GGAGG) located 9 bases upstream from the ATG translation initiation codon. Further upstream is an A-T-rich region containing a potential promoter with a −10 (ATTTAA) at base 149 and a −35 (TGACAA) at base 127.

**Amino acid composition.** The deduced amino acid sequence indicated a highly hydrophilic 151,591-dalton protein containing 1,365 amino acids. The N-terminal portion of GTF-S displayed a hydrophobic region typical of a signal peptide, which, according to the rule proposed by von Heijne (22), would be cleaved after amino acid 36 or 37, leaving a mature protein with a molecular weight of 147,408 and a theoretical isoelectric point of 5.28. These values are close to those determined experimentally (6, 16).

**Comparison of the deduced GTF-S protein sequence with other sequenced genes of mutans streptococci.** To date, three other GTF sequences and a fructosyltransferase sequence have been published (3, 18, 19, 21). Additionally, the DNA sequence of the GBP-encoding gene (gfb) has been recently determined (1). The inferred GTF-S amino acid sequence from *S. downei* possesses 45% overall identity with the amino acid sequences deduced from gtfS from *S. downei* and gtfB and gtfC from *S. mutans* and 24% identity with the GBP from *S. mutans* in the C-terminal portion of the protein. No significant homology was found between GTF-S and fructosyltransferase. A common feature in the inferred amino acid sequences from gtfS, gtfB, gtfC, and gfb is the presence of repeated regions in the C-terminal portion of the gene products (1). As expected, GTF-S also contains repeated sequences.

GTF-S has four repeated sequences ranging from 49 to 51 amino acids in length and a fifth, partial repeat of 13 amino acids (Fig. 3). These repeats share from 59.2 to 80.4% similarity with a consensus sequence (WYFYFxNxDQAAT GLQTIDGQTVYFDDNGxQVKxGxAVTDxxGKLRYFYD ADQG). The repeated sequences begin at amino acid 1082 and continue through to the termination codon. The repeats described by Banas et al. (1) as type A and C repeats are found in GTF-S, alternating (A-C-A-C, etc.). The 48-amino-acid type B repeats found by Ferretti et al. (3) in the GTF-I is not present in the GTF-S.

A portion of the repeat region was also found to have homology with the autolysin gene (lytA) from *Streptococcus pneumoniae* (5) and an α-galactosidase gene (our unpublished data) from *S. mutans*. Partial homology was also found between the N terminus of GTF-S and the Fc receptor protein from group A streptococci (fcrA76) (10).

**Glucan analysis.** The composition of the glucan synthesized by *E. coli* (pMLG60) was estimated by 13C-NMR spectral analysis (on the basis of peak height) to consist of more than 90% α-1,6-d-glucopyranosyl residues (Fig. 4). The NMR spectrum revealed six intense peaks which correspond to the relative peak positions for α-1,6-linked dextran T70 (2).
Repeat regions are underlined.

Fig. 2. Nucleotide sequence of gsf and the inferred amino acid sequence. Boxed regions indicate a putative promoter and ribosome-binding site (RBS). Nucleotide positions (+) and amino acid positions (in the left margin) are indicated. The five C-terminal repeat regions are underlined.
The relationship of GTF-S to GTF-I and to other known GTFs is of considerable interest. The sticky, insoluble glucan produced by the combined activities of GTF-S and GTF-I is thought to be of primary importance in the development of dental plaque. In *S. downei*, GTF-I-mediated glucan synthesis requires the presence of a primer (6). Glucan synthesis by GTF-S is independent of primer glucan, and the soluble glucan produced by GTF-S may act as an intrinsic primer for the synthesis of glucan by GTF-I. However, we have evidence (unpublished) that *S. downei* produces four distinct GTFs, so further work is necessary to elucidate the patterns of interaction among different enzymes.

Nucleotide sequence analysis of the DNA fragment encoding GTF-S from *S. downei* MFe28 revealed a single open reading frame of 4,094 bp. The deduced amino acid sequence of the processed GTF-S was determined to have a molecular weight of 147,408. Consistent with the extracellular location of the protein, the sequence encodes a typical signal peptide thought to be either 36 or 37 amino acids in length. Except for the putative signal peptide, GTF-S is a very hydrophilic protein. As with the other GTFs sequenced thus far, GTF-S...
The presence of by

Similarly, the boundaries was further substantiated to localized. Amino-terminal residues that an initiating residue corresponding to the C-terminal functional. The A repeat is somewhat shorter than the C repeat found in other GTFs. The amino acid residue initiating each repeat (left margin), percentages of amino acid identities with the consensus sequence for each repeat (right margin), and amino acids which remain identical in every repeat of GTF-S (*) are indicated.

contains no cysteine residues. The open reading frame is preceded by a putative ribosome-binding site and an upstream A-T-rich region containing a putative promoter site. The linear sequence of GTF-S is consistent with the presence of two functional domains. Ferretti et al. (3) demonstrated that an N-terminal deletion mutant of GTF-I (pSF86) could bind glucan but was not enzymatically functional. Similarly, Mooser and Wong (14) demonstrated that trypsin digestion of a GTF-S isolated from S. sobrinus yields a domain (fragment) which retains glucan-binding activity but loses its catalytic activity. As discussed below, the C-terminal localization of the GTF glucan-binding domain was further substantiated by sequence identity with a GBP of S. mutans. On the basis of these observations, the boundaries of the two functional domains of GTF-S can be localized to approximately the amino-terminal 1,050 amino acids (catalytic activity) and carboxy-terminal residues 1100 through 1365 (glucan-binding activity).

GTF-I of S. downei was reported to contain two types of amino acid repeats localized to the C terminus: a series of six type A repeats (35 amino acids in length) and two type B repeats (48 amino acids in length) (3). Banas et al. (1) observed the A repeat to be present within the GBP of S. mutans. Ingbrit five times with an additional, partial A repeat truncated by the termination codon. The B repeat was not observed in the GBP, but a different repeat termed the C repeat occurred four times (1). Similar type A and C repeats occur within other GTF enzymes (1). As in the GBP, no sequence corresponding to the B repeat of GTF-I was found in GTF-S. However, four sequences corresponding to A and C repeats were found in GTF-S, alternating A-C-A-C, with an additional, partial A repeat occurring at the C terminus of

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Percent Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1083 WYYFSDGKMATGKTKIGNDTYLMPNGKQLKEG-...-VWYD-GKAYYYYD-DNG</td>
<td>59.2%</td>
</tr>
<tr>
<td>1150 WRYFNGGTIAIGLVLSDNRTLYFDAYGYQVK-GQTIVIN-GKS-YTFDADQG</td>
<td>64%</td>
</tr>
<tr>
<td>1225 WQYR-KDGQILITGQTIDQKVFQDQNVWQGKT-A-DASGVL-R-FYDRDDG</td>
<td>80%</td>
</tr>
<tr>
<td>1289 WYYNESQVLGQLTIQDQTVYFDDKGIQAK-GKAVNENGLRYP-DADSG</td>
<td>80.4%</td>
</tr>
<tr>
<td>1353 WYFNLGGLAA-T-RW</td>
<td>69%</td>
</tr>
</tbody>
</table>

**Consensus:**

WYYFNxGQAATGLQTVYFDDNGxQVxGAVTDxxGKLRYFYDADQG

* * * * * * * * *

**FIG. 3.** Comparison of AC repeats contained within the inferred amino acid sequence of GTF-S with a consensus sequence. Regions corresponding to the A and C repeats are indicated. The C repeat is somewhat shorter than the C repeat found in other GTFs. The amino acid residue initiating each repeat (left margin), percentages of amino acid identities with the consensus sequence for each repeat (right margin), and amino acids which remain identical in every repeat of GTF-S (*) are indicated.

**FIG. 4.** Proton-decoupled carbon spectrum of the glucan produced by the gtfS cloned into E. coli. Chemical shifts are measured in parts per million (ppm) downfield from a sodium-3-trimethyl-silyl propionate internal standard. The assignment of carbon positions is shown.
GTF-S. Unlike the gbp-encoded protein, in which A and C sequences were observed to occur with various spacing intervals between each other and between A-C repeat cycles, only three to five amino acids separate the A repeat from the C repeat in each A-C cycle of GTF-S (thus forming four type AC repeated regions). The A and C repeat regions which occur in GTF-I from S. downei and in GTF-I and GTF-SI from S. mutans G55 are also found together as an AC repeat (Fig. 5). The GTF-S AC repeat cycles are separated by 13 to 25 nonconserved amino acids, and in GTF-I the AC repeats are separated by 6 nonconserved amino acids and in two cases by the 48-amino-acid B repeat. In contrast to the variable spacing between AC repeats observed for GTFs derived from S. downei, the regularity of the AC repeats in GTFs derived from S. mutans is remarkable in that the spacer amino acids between AC repeat regions are also very highly conserved. However, in the GTF-SI encoded by gtfC, the regularity of the AC repeats abruptly breaks down at amino acid 1338, near the C terminus (21). Interestingly, translation of the reading frame staggered by 1 base from the point of breakdown of the AC repeat mode in the inferred amino acid sequence of gtfC results in continuation of the ongoing AC repeat and the addition of another AC repeat in its entirety. This observation suggests that either the gtfC gene has recently lost about one-and-one-half cycles of the AC repeat, perhaps to modulate the affinity of glucan binding by the enzyme, or a frameshift was inadvertently introduced during nucleotide sequence determination. The sequence of gtfD, which encodes a GTF-S from S. mutans (9), is unknown, but because of their ubiquity, similar AC motifs would be predicted. The B repeats, thus far only observed to occur in the inferred amino acid sequence of gtfI of S. downei, appear to be unique to this GTF species, and the contribution of B repeats to protein function is unknown.

Because of the association of the C termini of GTFs with glucan binding, catalytic activity appears to reside in the N-terminal three-fourths of the protein. Alignment of the first 1,100 amino acids of all GTFs for which the inferred sequences have been reported reveals that this region is generally conserved and that three extensive and nearly invariant regions can be identified (Fig. 6). The degrees of identity which occur among the catalytic domains of GTFs were of interest, since NMR data indicated that GTF-S...
produces primarily α-1,6-linked glucans, as opposed to the α-1,3-linkage synthesized by the other enzymes. Additionally, all GTFs have the ability to hydrolyze sucrose and condense the activated glucose monomer onto the nascent glucan. Thus, subtle alterations in protein structure must account for the differences in GTF enzymatic activities.

The biological role for each GTF species is presently unknown, although collectively they clearly contribute to the deposition of dental plaque on the tooth surface. *S. mutans*, like other mutants streptococci, has been shown to produce more than one discrete but similar GTF (4, 6, 17, 24). One role for the GTF-S studied here may be to provide a primer for the synthesis of glucan by GTF-I. An alternative role may be to introduce α-1,6-linked branch points into the otherwise largely α-1,3-linked glucan, thereby facilitating polymer growth or enhancing adherent properties. The functional role of each GTF species synthesized by mutants streptococci is the subject of ongoing studies.

ACKNOWLEDGMENTS

This research was supported by Public Health Service grant DE 08191 from the National Institutes of Health. K.S.G. was supported by a fellowship from the Phillips Petroleum Foundation, Bartlesville, Okla.

We thank Glen Dreyhurst and Eric Enwall for performing the NMR analysis.

LITERATURE CITED


