INFECTION AND IMMUNITY, Aug. 1990, p. 2452–2458
0019-9567/90/082452-07$02.00/0
Copyright © 1990, American Society for Microbiology
Vol. 58, No. 8

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

KEETA S. GILMORE,[1] ROY R. B. RUSSELL,[2] AND JOSEPH J. FERRETTI[1]*

Department of Microbiology & Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73190,[1] and Hunterian Dental Research Unit, London Hospital Medical College, London E1 2AD, United Kingdom[2]

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 (*gtfI*) from *S. downei* and GTF-I (*gtfB*) and GTF-SI (*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 mutans 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 mutans 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-Is produced by cariogenic species of mutans 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 mutans 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* GS5 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 mutans 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

* Corresponding author.

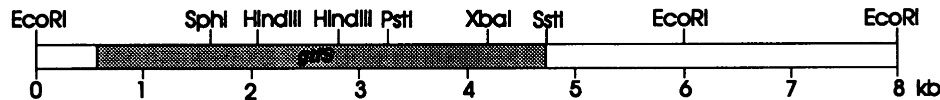

FIG. 1. Physical map of the *S. downei* DNA fragment which encodes GTF-S harbored by pMLG60. Shaded region indicates the location of the *gtfS* structural gene relative to restriction enzyme recognition sites within the 8-kb pMLG60 DNA insert.

by the University of Oklahoma Health Sciences Center Molecular Biology Core Resource Facility. M13 forward primer (17-mer) was purchased from Promega Biotec (Madison, Wis.), and [α-^{35}S]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 *Xba*I-*Sst*I, 0.8-kb *Pst*I-*Xba*I, 0.75-kb *Hin*dIII, 3.0-kb *Sph*I-*Sst*I, and 1.6-kb *Eco*RI-*Sph*I 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 NaH_2PO_4 (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 reprecipitated twice as described above. After the third precipitation, the glucan pellet was air dried.

NMR analysis. The composition of the glucan was analyzed by ^{13}C-nuclear magnetic resonance (^{13}C-NMR) spectrometry. The glucan was assayed in D_2O at 20°C. The spectra were obtained at 125 MHz on a VXR500 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 L47.1 (6, 12). An 8-kb *Eco*RI 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 *Eco*RI-*Sst*I fragment. This 4.7-kb *Eco*RI-*Sst*I 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 (GGAGA) 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 (ATTAAA) 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 (*gbp*) 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 *gtfI* 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 *gtfI*, *gtfB*, *gtfC*, and *gbp* 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 (WYYFNxDGQAAT 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 ^{13}C-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).

```
                         16                  32
                         *                   *
     TGG CCT ATG GAG TAG CGG TCA TTC TCA CCA TCT TGA ACA TCA AGT TAA

                         64                  80
                         *                   *
     TTA TTG ATT TGT TCC AAT AAA TAT GAA TCT CTT TAA AGA ATA AAT TTA

                         112                 128
                         *                   -35
     AAG AGA TTT TTA TGT TGT TTT ACA ATC TAA [TGA CAA] ATT TAG GAT TTC

                 -10     160                 176
                         *                   *
     TGA G[AT TAA A]CT TAA ATA ACA TGT TAA AAT AAA TTG ATA ATT AAA ATT

                         208                 224             RBS
                         *                   *
     ATT TGG TTA TTA AAG TTT TAA TTG TCG AAT TTT TTT ATA A[GG AGA] ATG

                         256                 272
                         *                   *
     ATA TTC ATG GAA AAA AAT TTG CGC TAC AAA TTG CAT AAG GTT AAA AAG
              M   E   K   N   L   R   Y   K   L   H   K   V   K   K

                         304                 320
                         *                   *
     CAA TGG GTA GCC ATT GGT GTG ACG ACC GTA ACG CTT AGC TTT CTA GCT
15    Q   W   V   A   I   G   V   T   T   V   T   L   S   F   L   A

                         352                 368
                         *                   *
     GGA GGC CAA GTC GTT GCA GCC GAC ACA AAC AAT AAT GAC GGG ACA AGT
31    G   G   Q   V   V   A   A   D   T   N   N   N   D   G   T   S

                         400                 416
                         *                   *
     GTT CAG GTC AAC AAA ATG GTG CCA AGT GAT CCC AAG TTT GAC GCT CAA
47    V   Q   V   N   K   M   V   P   S   D   P   K   F   D   A   Q

                         448                 464
                         *                   *
     GCC CAA AAC GGT CAA CTT GCT CAA GCT ATG TTT AAG GCT GCT AAT CAA
63    A   Q   N   G   Q   L   A   Q   A   M   F   K   A   A   N   Q

                         496                 512
                         *                   *
     GCA GAT CAA ACT GCT ACA AGT CAA GTT AGC CCA GCA ACA GAT GGC AGG
79    A   D   Q   T   A   T   S   Q   V   S   P   A   T   D   G   R

                         544                 560
                         *                   *
     GTT GAT AAC CAG GTG ACA CCA GCA GCT AAT CAA CCT GCG GCT AAT GTG
95    V   D   N   Q   V   T   P   A   A   N   Q   P   A   A   N   V

                         592                 608
                         *                   *
     GCC AAT CAA GAT GTG GCC AAT CCA GCG ACT GAT GCC GGA GCA CTT AAT
111   A   N   Q   D   V   A   N   P   A   T   D   A   G   A   L   N

                         640                 656
                         *                   *
     CGC CAA TCG GCC GCT GAT ACG TCA ACA GAT GGT AAG GCT GTA CCT CAG
127   R   Q   S   A   A   D   T   S   T   D   G   K   A   V   P   Q

                         688                 704
                         *                   *
     ACT AGT GAC CAA CCA GGT CAT CTG GAA ACA GTT GAT GGC AAA ACT TAC
143   T   S   D   Q   P   G   H   L   E   T   V   D   G   K   T   Y

                         736                 752
                         *                   *
     TAT GTG GAT GCC AAT GGT CAA CGG TTG AAA AAC TAT TCA ATG GTC ATT
159   Y   V   D   A   N   G   Q   R   L   K   N   Y   S   M   V   I

                         784                 800
                         *                   *
     GAT GGC AAG ACC TAT TAC TTT GAC GGT CAG ACA GGC GAA GCG CAA ACC
175   D   G   K   T   Y   Y   F   D   G   Q   T   G   E   A   Q   T

                         832                 848
                         *                   *
     GAC TTG CCT AAG ACC GGT CAA GCT AAT CAA GAC AAT GTG CCC GAT AGC
191   D   L   P   K   T   G   Q   A   N   Q   D   N   V   P   D   S

                         880                 896
                         *                   *
     TAT CAG GCC AAT AAT CAA GCC TAT AGT AAT GAG GCT TCT AGC TTT GAG
207   Y   Q   A   N   N   Q   A   Y   S   N   E   A   S   S   F   E

                         928                 944
                         *                   *
     ACA GTT GAT AAT TAC CTG ACA GCG GAT TCC TGG TAT CGC CCT CGC AAG
223   T   V   D   N   Y   L   T   A   D   S   W   Y   R   P   R   K

                         976                 992
                         *                   *
     ATT TTG AAA AAT GGC CAG AGC TGG CAG GCT AGT TCA GAA GGT GAC CTG
239   I   L   K   N   G   Q   S   W   Q   A   S   S   E   G   D   L

                         1024                1040
                         *                   *
     CGT CCT ATT CTC ATG ACC TGG TGG CCA GAT GCG GCC ACC AAG GCA GCC
255   R   P   I   L   M   T   W   W   P   D   A   A   T   K   A   A

                         1072                1088
                         *                   *
     TAT GCC AAC TTC TGG GCT AAA GAA GGC CTG ATT TCA GGT TCT TAC AGG
271   Y   A   N   F   W   A   K   E   G   L   I   S   G   S   Y   R

                         1120                1136
                         *                   *
     CAA AAT TCT GCC AAT CTT GAT GCG GCT ACC CAG AAT ATT CAA TCT GCC
287   Q   N   S   A   N   L   D   A   A   T   Q   N   I   Q   S   A

                         1168                1184
                         *                   *
     ATT GAA AAG AAG ATT GCC AGT GAA GGC AAT ACC AAC TGG TTG CGC GAC
303   I   E   K   K   I   A   S   E   G   N   T   N   W   L   R   D

                         1216                1232
                         *                   *
     AAG ATG TCC CAA TTC GTC AAG TCA CAA AAC CAG TGG AGT ATT GCT TCT
319   K   M   S   Q   F   V   K   S   Q   N   Q   W   S   I   A   S

                         1264                1280
                         *                   *
     GAA AAT GAG ACT GTC TAT CCT AAT CAG GAT CAT ATG CAA GGG GGG GCC
335   E   N   E   T   V   Y   P   N   Q   D   H   M   Q   G   G   A

                         1312                1328
                         *                   *
     CTG CTC TTT TCA AAC AGT AAG GAT ACC GAG CAT GCT AAC TCA GAC TGG
351   L   L   F   S   N   S   K   D   T   E   H   A   N   S   D   W

                         1360                1376
                         *                   *
     CGC CTG CTC AAC CGC AAT CCA ACC TTC CAA ACT GGT AAA CAA AAG TAT
367   R   L   L   N   R   N   P   T   F   Q   T   G   K   Q   K   Y

                         1408                1424
                         *                   *
     TTT ACA ACC AAC TAT GCG GGT TAT GAA TTG CTC TTA GCC AAT GAT GTT
383   F   T   T   N   Y   A   G   Y   E   L   L   L   A   N   D   V

                         1456                1472
                         *                   *
     GAT AAT TCT AAC CCG GTC GTC CAA GCC GAA CAG CTC AAT CAC CTC CAC
399   D   N   S   N   P   V   V   Q   A   E   Q   L   N   H   L   H

                         1504                1520
                         *                   *
     TAC CTT ATG AAC TGG GGT GAC ATT GTT ATG GGG GAT AAG GAT GCT AAC
415   Y   L   M   N   W   G   D   I   V   M   G   D   K   D   A   N

                         1552                1568
                         *                   *
     TTT GAT GGT GTC CGA GTT GAT GCG GTT GAT AAT GTC AAT GCC GAC CTG
431   F   D   G   V   R   V   D   A   V   D   N   V   N   A   D   L

                         1600                1616
                         *                   *
     CTA CAA ATT CAG AGG GAT TAT TAT AAG GCC AAA TAT GGC ACC GAT CAA
447   L   Q   I   Q   R   D   Y   Y   K   A   K   Y   G   T   D   Q

                         1648                1664
                         *                   *
     AAT GAA AAA AAT GCC ATT GAC CAC CTC TCT ATC CTA GAA GCC TGG TCT
463   N   E   K   N   A   I   D   H   L   S   I   L   E   A   W   S

                         1696                1712
                         *                   *
     GGC AAC GAC AAT GAC TAC GTC AAG GAC CAA AAC AAC TTT TCT CTC TCT
479   G   N   D   N   D   Y   V   K   D   Q   N   N   F   S   L   S

                         1744                1760
                         *                   *
     ATT GAC AAT GAT CAA CGC AGT GGT ATG TTG AAA GCT TTC GGT TAT GCT
495   I   D   N   D   Q   R   S   G   M   L   K   A   F   G   Y   A

                         1792                1808
                         *                   *
     TCT GCC TAT CGT GGT AAT TTG AGT AAT CTC GCA ACT GCT GGC TTG AAA
511   S   A   Y   R   G   N   L   S   N   L   A   T   A   G   L   K

                         1840                1856
                         *                   *
     AAC CGT AGT GCA AAT CCT GAT AGT GAC CCA GTT CCA AAC TAT GTC TTT
527   N   R   S   A   N   P   D   S   D   P   V   P   N   Y   V   F

                         1888                1904
                         *                   *
     ATT CGT GCC CAT GAC TCT GAG GTA CAG ACA CGG ATT GCT AAG ATT ATT
543   I   R   A   H   D   S   E   V   Q   T   R   I   A   K   I   I

                         1936                1952
                         *                   *
     CGT GAG AAA TTG GGT AAG ACC AAT GCT GAT GGT TTG ACC AAT CTG ACC
559   R   E   K   L   G   K   T   N   A   D   G   L   T   N   L   T

                         1984                2000
                         *                   *
     TTA GAT GAT TTA AAC AAG GCC TTT GAC ATC TAC AAC CAG GAT ATG AAT
575   L   D   D   L   N   K   A   F   D   I   Y   N   Q   D   M   N

                         2032                2048
                         *                   *
     GCC ACC GAT AAG GTT TAT TAT CCT AAT AAT TTA CCG ATG GCC TAT GCT
591   A   T   D   K   V   Y   Y   P   N   N   L   P   M   A   Y   A

                         2080                2096
                         *                   *
     TGG ATG CTG CAA AAT AAG GAT ACG GTC ACT CGG GTC TAC TAT GGT GAT
607   W   M   L   Q   N   K   D   T   V   T   R   V   Y   Y   G   D

                         2128                2144
                         *                   *
     ATG TAC ACC GAT AAT GGC CAG TAT ATG GCG ACC AAG ACC CCA TTT TAT
623   M   Y   T   D   N   G   Q   Y   M   A   T   K   T   P   F   Y

                         2176                2192
                         *                   *
     AAT GCT ATT GAG ACC CTG CTT AAG GGC CGG ATT AAG TAT GTT GCT GGT
639   N   A   I   E   T   L   L   K   G   R   I   K   Y   V   A   G

                         2224                2240
                         *                   *
     GGT CAA GCG GTC TCC TAC AAG CAA GAT TGG TCT AGT GGA ATT TTG ACC
655   G   Q   A   V   S   Y   K   Q   D   W   S   S   G   I   L   T

                         2272                2288
                         *                   *
     TCG GTT CGC TAT GGT AAG GGA GCT AAT TCG GCT AGT GAT GCT GGG AAC
671   S   V   R   Y   G   K   G   A   N   S   A   S   D   A   G   N

                         2320                2336
                         *                   *
     ACG GAA ACC CGC AAT TCT GGT ATG GCC CTC CTG ATT AAC AAT CGA CCT
687   T   E   T   R   N   S   G   M   A   L   L   I   N   N   R   P

                         2368                2384
                         *                   *
     AAC TTC AGG GCT TAT CGG AAC TTG ACC TTG AAC ATG GGG GCA GCC CAT
703   N   F   R   A   Y   R   N   L   T   L   N   M   G   A   A   H
```

FIG. 2. Nucleotide sequence of *gtfS* 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.

```
                         2416                2432
                         *                   *
     AAG AGC CAA GCC TAC CGT CCT CTC TTG CTG TCT ACC AAG GAT GGC ATT
719  K   S   Q   A   Y   R   P   L   L   L   S   T   K   D   G   I

                         2464                2480
                         *                   *
     GCC ACC TAC CTT AAT GAT AGC GAT GTG GAT TCT CGC CAA TAC AAG TAC
735  A   T   Y   L   N   D   S   D   V   D   S   R   Q   Y   K   Y

                         2512                2528
                         *                   *
     ACC GAT AGT CAG GGG AAC TTA AGC TTT AGT GCC TCT GAA CTG CAA AGT
751  T   D   S   Q   G   N   L   S   F   S   A   S   E   L   Q   S

                         2560                2576
                         *                   *
     GTT GCT AAC GCC CAA GTC TCA GGT ATG ATT CAG GTT TGG GTA CCC GTC
767  V   A   N   A   Q   V   S   G   M   I   Q   V   W   V   P   V

                         2608                2624
                         *                   *
     GGT GCG GCA GAT AAC CAA GAT GTC CGT ACT TCT CCA AGT ACA CAA GCG
783  G   A   A   D   N   Q   D   V   R   T   S   P   S   T   Q   A

                         2656                2672
                         *                   *
     ACC AAG GAT GGC AAT ATC TAT CAT CAA AGT GAC GCC CTA GAT TCC CAA
799  T   K   D   G   N   I   Y   H   Q   S   D   A   L   D   S   Q

                         2704                2720
                         *                   *
     GTC ATC TAT GAA GGT TTC TCT AAT TTC CAA GCC TTC GCC CAA AGT CCT
815  V   I   Y   E   G   F   S   N   F   Q   A   F   A   Q   S   P

                         2752                2768
                         *                   *
     GAC CAA TAT ACC AAT GCT GTC ATT GCT AAA AAT GGC GAC CTC TTT AAG
831  D   Q   Y   T   N   A   V   I   A   K   N   G   D   L   F   K

                         2800                2816
                         *                   *
     TCT TGG GGC ATT ACC CAA TTT GAA ATG GCA CCG CAG TAC GTG TCA TCA
847  S   W   G   I   T   Q   F   E   M   A   P   Q   Y   V   S   S

                         2848                2864
                         *                   *
     GAA GAT GGA ACT TTC CTT GAT TCC GTT ATT TTA AAT GGT TAT GCC TTC
863  E   D   G   T   F   L   D   S   V   I   L   N   G   Y   A   F

                         2896                2912
                         *                   *
     TCA GAC CGC TAT GAC CTG GCT ATG AGT AAG AAT AAT AAA TAT GGC TCC
879  S   D   R   Y   D   L   A   M   S   K   N   N   K   Y   G   S

                         2944                2960
                         *                   *
     AAG CAG GAT TTG GCC AAT GCC ATT AAA GGA CTT CAG TCT GCA GGT ATC
895  K   Q   D   L   A   N   A   I   K   G   L   Q   S   A   G   I

                         2992                3008
                         *                   *
     AAG GTC TTG TCA GAC TTG GTA CCG AAC CAA CTT TAT AAT CTG CCT GGA
911  K   V   L   S   D   L   V   P   N   Q   L   Y   N   L   P   G

                         3040                3056
                         *                   *
     AAA GAA GTG GTA ACG GCA ACC CGG GTC AAT CAA TAT GGT CAA GCC AAA
927  K   E   V   V   T   A   T   R   V   N   Q   Y   G   Q   A   K

                         3088                3104
                         *                   *
     TCA GGG GCG ACC ATC AAT AAG ACT CCT TAT GTG GCC AAT ACC CGT TCC
943  S   G   A   T   I   N   K   T   P   Y   V   A   N   T   R   S

                         3136                3152
                         *                   *
     TAT GGT GAC TAC CAA GAA CAG TAC GGT GGA AAA TTC TTG GAT GAC TTG
959  Y   G   D   Y   Q   E   Q   Y   G   G   K   F   L   D   D   L

                         3184                3200
                         *                   *
     CAA AAG CTC TAC CCA AGA CTC TTT AGC ACC AAG CAA ATT TCA ACG GGC
975  Q   K   L   Y   P   R   L   F   S   T   K   Q   I   S   T   G

                         3232                3248
                         *                   *
     AAA CCA ATT GAT CCA TCT GTT AAG ATT ACC AAT TGG TCG GCT AAG TAT
991  K   P   I   D   P   S   V   K   I   T   N   W   S   A   K   Y

                         3280                3296
                         *                   *
     TTC AAT GGC TCC AAT ATT TTG GGG CGT GGT GCC AAG TAT GTT CTG AGT
1007 F   N   G   S   N   I   L   G   R   G   A   K   Y   V   L   S

                         3328                3344
                         *                   *
     GAG GGC AAT AAG TAT CTC AAC TTG GCT GAT GGC AAG CTC TTC TTG CCA
1023 E   G   N   K   Y   L   N   L   A   D   G   K   L   F   L   P

                         3376                3392
                         *                   *
     ACG GTC CTC AAT AAT ACT TAT GGT CAG CCG CAA GTA TCG GCT AAT GGT
1039 T   V   L   N   N   T   Y   G   Q   P   Q   V   S   A   N   G

                         3424                3440
                         *                   *
     TTT ATT TCT AAG AAT GGC GGT ATT CAT TAT CTT GAT AAA AAT GGT CAG
1055 F   I   S   K   N   G   G   I   H   Y   L   D   K   N   G   Q

                         3472                3488
                         *                   *       Repeat #1
     GAA GTC AAG AAT CGG TTC AAG GAA ATT TCT GGC AGC TGG TAT TAC TTC
1071 E   V   K   N   R   F   K   E   I   S   G   S   W   Y   Y   F

                         3520                3536
                         *                   *
     GAT TCT GAC GGT AAG ATG GCT ACT GGA AAA ACG AAG ATT GGC AAT GAT
1087 D   S   D   G   K   M   A   T   G   K   T   K   I   G   N   D

                         3568                3584
                         *                   *
     ACC TAC CTC TTT ATG CCT AAT GGT AAA CAA CTT AAG GAA GGT GTC TGG
1103 T   Y   L   F   M   P   N   G   K   Q   L   K   E   G   V   W

                         3616                3632
                         *                   *
     TAT GAT GGT AAG AAG GCC TAC TAT TAT GAT GAT AAT GGT AGG ACG TGG
1119 Y   D   G   K   K   A   Y   Y   Y   D   D   N   G   R   T   W

                         3664                3680
                         *                   *
     ACC AAT AAG GGC TTC GTT GAA TTT AGA GTT GAC GGT CAA GAC AAG TGG
1135 T   N   K   G   F   V   E   F   R   V   D   G   Q   D   K   W

                         3712                3728
     Repeat #2           *                   *
     CGT TAC TTC AAT GGT GAT GGC ACT ATC GCC ATT GGA CTA GTT AGT CTA
1151 R   Y   F   N   G   D   G   T   I   A   I   G   L   V   S   L

                         3760                3776
                         *                   *
     GAT AAT CGC ACC CTC TAC TTT GAT GCC TAT GGC TAT CAA GTC AAG GGA
1167 D   N   R   T   L   Y   F   D   A   Y   G   Y   Q   V   K   G

                         3808                3824
                         *                   *
     CAA ACC GTG ACG ATT AAT GGT AAG TCC TAT ACC TTT GAT GCC GAT CAG
1183 Q   T   V   T   I   N   G   K   S   Y   T   F   D   A   D   Q

                         3856                3872
                         *                   *
     GGT GAC TTG GTC CAA ACA GAC AAT GCC AAT CCA GCT CCT CAG GGT CAA
1199 G   D   L   V   Q   T   D   N   A   N   P   A   P   Q   G   Q

                         3904                3920
                         *                   Repeat #3
     GCA GGC TGG AAA CTC CTA GGA GAT AAC CAG TGG GGC TAC CGC AAG GAC
1215 A   G   W   K   L   L   G   D   N   Q   W   G   Y   R   K   D

                         3952                3968
                         *                   *
     GGT CAA CTC TTG ACG GGT GAG CAA ACT ATT GAT GGT CAA AAG GTC TTC
1231 G   Q   L   L   T   G   E   Q   T   I   D   G   Q   K   V   F

                         4000                4016
                         *                   *
     TTC CAA GAT AAT GGC GTC CAA GTC AAA GGT GGA ACT GCG ACA GAT GCT
1247 F   Q   D   N   G   V   Q   V   K   G   G   T   A   T   D   A

                         4048                4064
                         *                   *
     TCA GGT GTC TTG CGT TTC TAC GAC CGT GAC CAG GGC CAC CAA GTT GGC
1263 S   G   V   L   R   F   Y   D   R   D   Q   G   H   Q   V   G

                         4096                4112
                         *                   Repeat #4
     AAG GGC TGG TAC TCA ACC TCC GAC GAT AAT TGG GTC TAT GTC AAT GAA
1279 K   G   W   Y   S   T   S   D   D   N   W   V   Y   V   N   E

                         4144                4160
                         *                   *
     TCC GGT CAA GTT CTG ACA GGC TTA CAA ACC ATT GAT GGT CAG ACG GTC
1295 S   G   Q   V   L   T   G   L   Q   T   I   D   G   Q   T   V

                         4192                4208
                         *                   *
     TAT TTC GAT GAC AAG GGT ATC CAG GCC AAG GGC AAG GCT GTC TGG GAT
1311 Y   F   D   D   K   G   I   Q   A   K   G   K   A   V   W   D

                         4240                4256
                         *                   *
     GAA AAT GGC AAC CTG CGT TAC TTT GAT GCC GAT TCA GGT AAT ATG CTT
1327 E   N   G   N   L   R   Y   F   D   A   D   S   G   N   M   L

                         4288                4304
                         *                   Repeat #5
     CGG GAC CGT TGG AAG AAC GTT GAC GGA AAC TGG TAT TAC TTC AAC CGC
1343 R   D   R   W   K   N   V   D   G   N   W   Y   Y   F   N   R

                         4336                4352
                         *                   *
     AAT GGC CTA GCC ACG AGG TGG TAA GCC TAA TAC TCG CTA AAT ATT AAA
1359 N   G   L   A   T   R   W   *

     AGA GCT CG
```

DISCUSSION

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

```
Amino Acid                                                      Percent Identity

1083 WYYFDSDGKMATGKTKIGNDTYLFMPNGKQLKEG--VWYD-GKKAYYYD-DNG  59.2%
1150 WRYFNGDGTIAIGLVSLDNRTLYFDAYGYQVK-GQTVTIN-GKS-YTFDADQG  64%
1225 WGYR-KDGQLLTGEQTIDGQKVFFQDNGVQVKGGTA-TDASGVLR-FYDRDQG  80%
1289 WVYVNESGQVLTGLQTIDGQTVYFDDKGIQAK-GKAVWDENGNLRYF-DADSG  80.4%
1353 WYYFNRNGLA-T-RW                                        69%

Consensus:
     ---------------- A ----------------    -------- C -------
     WYYFNxDGQAATGLQTIDGQTVYFDDNGxQVKxGxAVTDxxGKLRYFYDADQG
     * *    *                *   * * * *       *       * * *
```

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.

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-S 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* Ingbritt 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

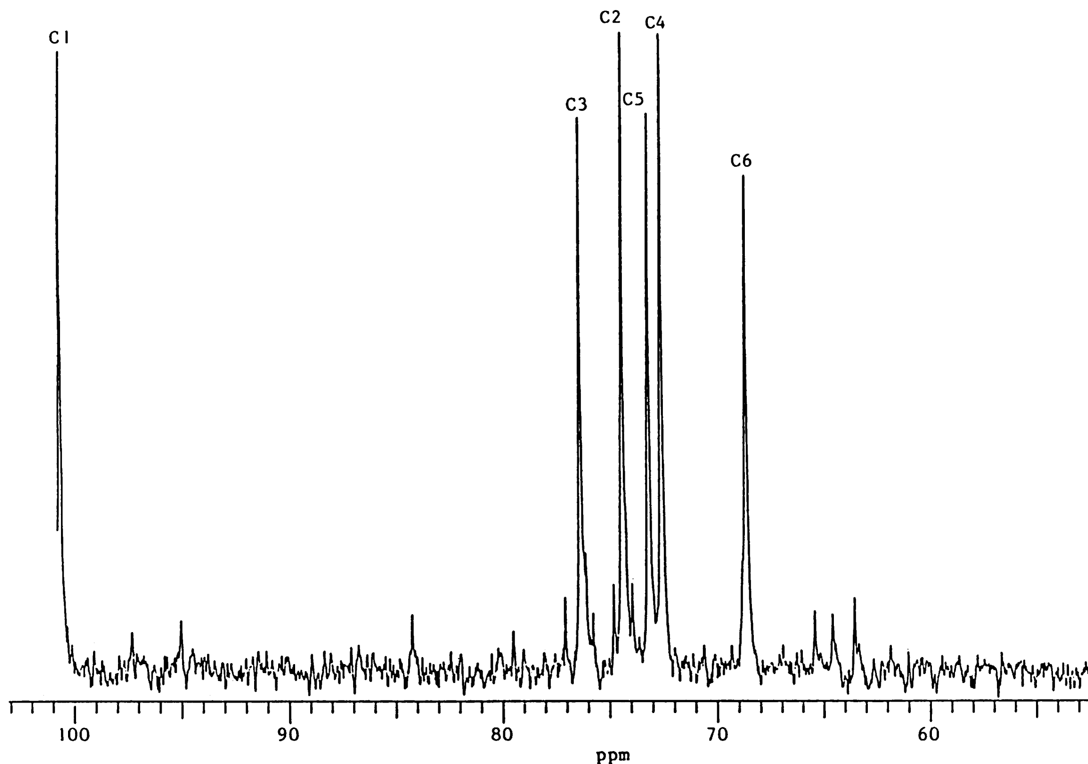

FIG. 4. Proton-decoupled carbon spectrum of the glucan produced by the *gtfS* cloned into *E. coli*. Chemical shifts are measured in parts per million (ppm) downfield from a sodium-3-trimethyl-silyl propionate internal standard. The assignment of carbon positions is shown.

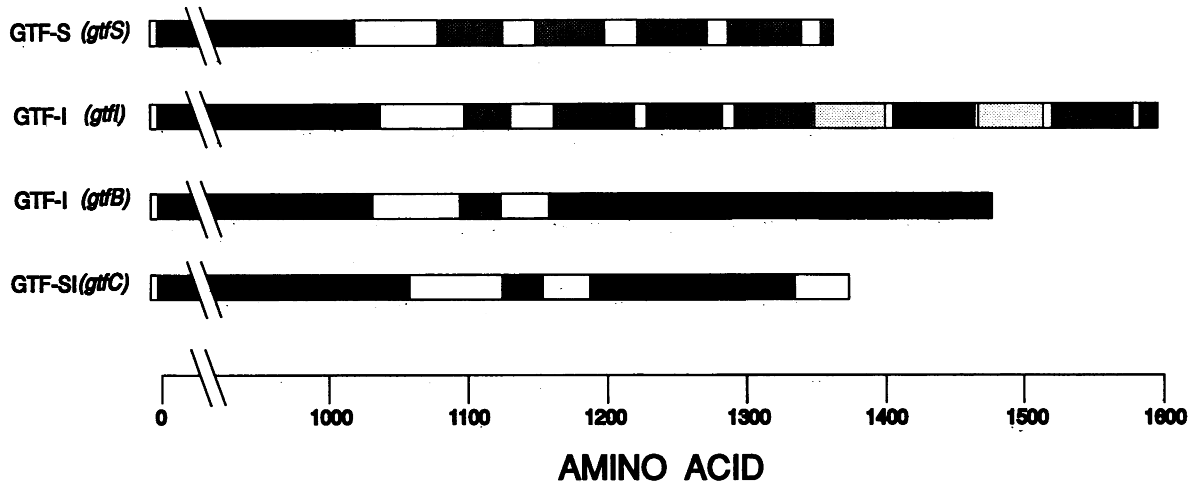

FIG. 5. Organization of conserved amino acid sequence motifs among GTFs. Symbols: ■, highly conserved amino-terminal domains for GTF-S from *S. downei* (this study), GTF-I from *S. downei* (3), and GTF-I (*gtfB*) (19) and GTF-SI (*gtfC*) (21) from *S. mutans* (see Fig. 6); ▩, AC repeated regions (except where the letter A designates a partial repeat); ▒, locations of B repeats, thus far reported only to occur in GTF-I (3); □, nonconserved regions between repeats.

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* GS5 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

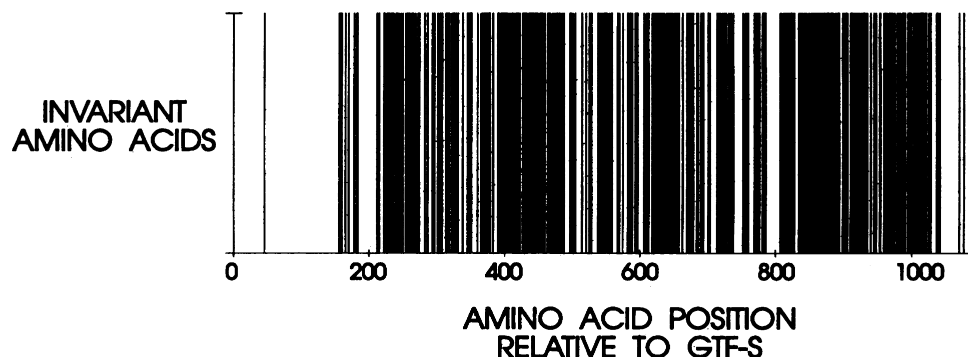

FIG. 6. Graphical depiction of N-terminal amino acid residues within GTF-S which are identical in all GTFs for which the amino acid sequence has been determined. Each vertical line represents an amino acid that is identical in all four GTFs. Regions of highest conservation occur between residues 390 to 460, 625 to 680, 820 to 900, and 960 to 1050. The region immediately following amino acid 1050 contains relatively few invariant amino acids and appears to form the link between the amino-terminal enzymatic domain and the carboxy-terminal glucan-binding domain.

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. downei*, like other mutans 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 mutans 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

1. **Banas, J. A., R. R. B. Russell, and J. J. Ferretti.** 1990. Sequence analysis of the gene for the glucan-binding protein of *Streptococcus mutans* Ingbritt. Infect. Immun. **58:**667–673.
2. **Colson, P., H. H. Jennings, and I. C. P. Smith.** 1974. Composition, sequence, and conformation of polymers and oligomers of glucose as revealed by carbon 13 nuclear magnetic resonance. J. Am. Chem. Soc. **96:**8081–8087.
3. **Ferretti, J. J., M. L. Gilpin, and R. R. B. Russell.** 1987. Nucleotide sequence of a glucosyltransferase gene from *Streptococcus sobrinus* MFe28. J. Bacteriol. **169:**4271–4278.
4. **Furuta, T., T. Koga, T. Nisizawa, N. Okahashi, and S. Hamada.** 1985. Purification and characterization of glucosyltransferases from *Streptococcus mutans* 6715. J. Gen. Microbiol. **131:**285–293.
5. **Garcia, E., J. L. Garcia, P. Garcia, A. Arraras, J. M. Sanchez-Puelles, and R. Lopez.** 1986. Molecular evolution of lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. Proc. Natl. Acad. Sci. USA **85:**914–918.
6. **Gilpin, M. L., R. R. B. Russell, and P. Morrissey.** 1985. Cloning and expression of two *Streptococcus mutans* glucosyltransferases in *Escherichia coli* K-12. Infect. Immun. **49:**414–416.
7. **Guggenheim, B., and E. Newbrun.** 1969. Extracellular glucosyltransferase activity of an HS strain of *Streptococcus mutans*. Helv. Odontol. Acta **16:**637–648.
8. **Hamada, S., and H. D. Slade.** 1980. Biology, immunology, and cariogenicity of *Streptococcus mutans*. Microbiol. Rev. **44:**331–384.
9. **Hanada, N., and H. K. Kuramitsu.** 1989. Isolation and characterization of the *Streptococcus mutans gtfD* gene, coding for primer-dependent soluble glucan synthesis. Infect. Immun. **57:**2079–2085.
10. **Heath, D. G., and P. O. Cleary.** 1989. Fc-receptor and M protein genes of group A streptococci are products of gene duplication. Proc. Natl. Acad. Sci. USA **86:**4741–4745.
11. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:**351–359.
12. **Loenen, W. A. M., and W. J. Brammar.** 1980. A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. Gene **20:**249–259.
13. **Loesche, W. J.** 1986. Role of *Streptococcus mutans* in human dental decay. Microbiol. Rev. **50:**353–380.
14. **Mooser, G., and C. Wong.** 1988. Isolation of a glucan-binding domain of glucosyltransferase (1,6-α-glucan synthase) from *Streptococcus sobrinus*. Infect. Immun. **56:**880–884.
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. **Russell, R. R. B., M. L. Gilpin, H. Mukasa, and G. Dougan.** 1987. Characterization of glucosyltransferase expressed from a *Streptococcus sobrinus* gene cloned in *Escherichia coli*. J. Gen. Microbiol. **133:**935–944.
17. **Shimamura, A., H. Tsumori, and H. Mukasa.** 1983. Three kinds of extracellular glucosyltransferases from *Streptococcus mutans* 6715 (serotype g). FEBS Lett. **157:**79–84.
18. **Shiroza, T., and H. K. Kuramitsu.** 1988. Sequence analysis of the *Streptococcus mutans* fructosyltransferase gene and flanking regions. J. Bacteriol. **170:**810–816.
19. **Shiroza, T., S. Ueda, and H. K. Kuramitsu.** 1987. Sequence analysis of the *gtfB* gene from *Streptococcus mutans*. J. Bacteriol. **169:**4263–4270.
20. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84:**4767–4771.
21. **Ueda, S., T. Shiroza, and H. K. Kuramitsu.** 1988. Sequence of the *gtfC* gene from *Streptococcus mutans* GS-5. Gene **69:**101–109.
22. **von Heijne, G.** 1983. Patterns of amino acids near signal sequence cleavage sites. Eur. J. Biochem. **133:**17–21.
23. **Whiley, R. A., R. R. B. Russell, J. M. Hardie, and D. Beighton.** 1988. *Streptococcus downei* sp. nov. for strains previously described as *Streptococcus mutans* serotype h. Int. J. Syst. Bacteriol. **38:**25–29.
24. **Yamashita, Y., N. Hanada, and T. Takehara.** 1989. Purification of a fourth glucosyltransferase from *Streptococcus sobrinus*. J. Bacteriol. **171:**6265–6270.
25. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19. Gene **33:**103–119.