J. Oral Biosci. **47**(2)：171-174, 2005

SHORT COMMUNICATION

Single Nucleotide Polymorphisms（SNPs）Detected in the *gbpC* Gene Coding Region of *Streptococcus mutans*

Yutaka Sato[1,2)§], Kazuko Okamoto-Shibayama[1,2)], Akiyoshi Kagami[1,2)], Yasuhito Yamamoto[1,2)] and Harutoshi Kizaki[1,2)]

[1)]*Department of Biochemistry, Tokyo Dental College*
（Chief：Prof. Harutoshi Kizaki）
[2)]*Oral Health Science Center, Tokyo Dental College*
（Chief：Prof. Harutoshi Kizaki）
[1,2)]*2-2, Masago 1-chome, Mihama-ku, Chiba 261-8502, Japan*
〔Received on December 8, 2004；Accepted on February 9, 2005〕

Key words：*Streptococcus mutans*／glucan-binding protein C／the *gbpC* gene／single nucleotide polymorphisms（SNPs）

Abstract：*Streptococcus mutans* glucan-binding protein C encoded by the *gbpC* gene is an important virulent factor in human dental caries. However, a nonsense mutation and several nucleotide substitutions in the *gbpC* gene have been detected in strain GS-5. Therefore, *gbpC* nucleotide sequences from 17 additional *S. mutans* strains were determined and single nucleotide polymorphisms were detected. Both conserved and polymorphic regions of the *gbpC* gene will be useful for estimation of functional domains of GbpC protein and identification of *S. mutans* strains.

Streptococcus mutans is regarded as the principal etiologic agent of human dental caries[1)]. This organism possesses several extracellular proteins or enzymes for adherence and survival on tooth surfaces as a constituent of dental plaque biofilm[2)]. Glucan-binding protein C（GbpC）encoded by the *gbpC* gene identified in our laboratory[3)] is one such extracellular protein classified as a cell wall-anchored surface protein（WASP）in Gram-positive bacteria[4)] and is involved in dextran（α-1,6 glucan）-dependent aggregation（ddag）of this organism. *S. mutans* cells are able to adhere to an immobilized glucan substratum mediated by this protein[5)]. Therefore, this protein is an important virulent factor of this organism[6,7)]. However, we have found that *S. mutans* strain GS-5 did not exhibit the ddag, and detected a nonsense mutation in the *gbpC* gene[8)] in addition to a frame-shift mutation in the *pac* gene encoding the major surface protein antigen of this organism[9)]. Besides this nonsense mutation in the *gbpC* gene, three non-synonymous and 8 synonymous nucleotide substitutions were detected when the entire 1,752 bp *gbpC* coding region of strain GS-5 was compared with that of strain 109cS. We have examined the ddag of hundreds of *S. mutans* strains from our stock culture collection, and found three other ddag-negative strains（Z1, OMZ175, and 98-4）in addition to strain GS-5. This result suggests that the *gbpC* gene may be divergent in *S. mutans* species and that single nucleotide polymorphisms（SNPs）may be detected in the *gbpC* gene. While SNPs are useful for individual identification or

§ Corresponding author
E-mail：yusato@tdc.ac.jp

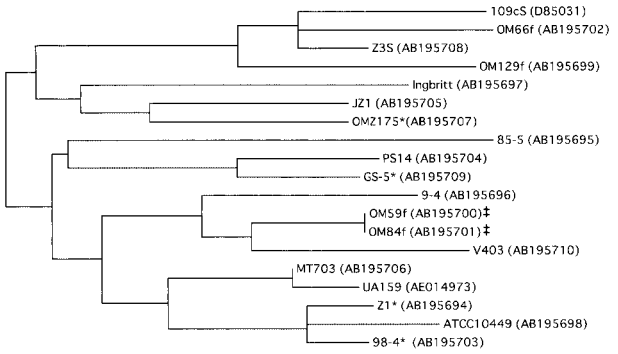

Fig. 1 Phylogenetic tree of the 19 *gbpC* sequences
Multiple alignment and tree analyses were performed with default settings (TOSSGAPS-ON) of the ClustalW program at DDBJ. The *gbpC* sequences from strains OM59f and OM84f (symbolized with ‡) were identical. Asterisks indicate the strains exhibiting no ddag and suffered *gbpC* gene truncation because of nonsense or frame shift mutations. Accesion numbers of the sequences are in parentheses

for potential individualized medical treatments in humans, they are also useful for strain identification in bacterial infections；for example, mother-infant vertical transfer of *S. mutans*[10]. SNPs in the *gbpC* gene, if they are actually observed, may also be helpful in identifying essential regions for glucan-binding activity in the protein encoded by the gene. Therefore, 17 additional *S. mutans* strains from our stock culture collection were selected and determined the *gbpC* nucleotide sequences of these strains.

S. mutans strains were cultured in Todd-Hewitt broth, and isolation of chromosomal DNA was carried out as previously described[3]. A rapid chromosomal DNA isolation method using chloroform extraction[11] originally described by Gillespie, *et al*.[12] was also employed. The region encoded by the *gbpC* gene was amplified by the PCR method and was directly sequenced as previously described with the same primers listed in the previous report[8]. Amplified fragment length polymorphisms (AFLPs) analysis using the Ready-To-Go RAPD Analysis Kit (Amersham Biosciences Corp. Piscataway, NJ, USA) with the chromosomal DNA from the 19 strains was also carried out as previously described[8].

The 17 *gbpC* nucleotide sequence data reported in this paper appear in the DDBJ, EMBL, and GenBank nucleotide sequence databases under the accession numbers AB195694–AB195710. Analysis for multiple alignment and tree-making with the 19 *gbpC* sequences including those previously determined (D85031 and AE014973 from strains 109cS and UA159 respectively) were carried out with the ClustalW program at DDBJ and an unrooted phylogenetic tree was drawn with the DendroMaker for Macintosh Ver 4.1 from the output results by the ClustalW program (Fig. 1.).

All sequences except for that of strain Z1[13] have 1,752 bp length coding regions. The Z1 *gbpC* open reading frame (ORF) was truncated at 615 nt by a flame-shift mutation resulted from one nucleotide deletion at 430 nt. In addition to this mutation and previously determined mutation in the GS-5 gene (G193T), nonsense mutations were detected in the genes from strains OMZ175 and 98-4 at 808 nt (A＞T) and 193 nt (G＞T), respectively. Since 2 (Z1 and OMZ175) out of 4 strains were found to be serotype *f*, we were interested in whether or not serotype *f* strains often suffered nonsense mutations. Therefore, four serotype *f* strains were included in the 19 strains analyzed. However, these four strains possessed no nonsense mutations. The other 11 strains including 9 randomly selected strains exhibited synonymous and nonsynonymous nucleotide substitutions but no nonsense mutations. The 19 strains together exhibited 18 distinct *gbpC* sequences and one sequence from strain OM59f was completely identical to that from strain OM84f. In this respect, we concluded that the *S. mutans gbpC* gene exhibited SNPs. The number of sites at which mutations/substitutions were detected within 1,752 bp was 47 and the sites were relatively diffuse over the entire coding region. However, a relatively high polymorphic region (13 SNPs in a 250 bp sites；nt 101—350；Fig. 2) and several conserved stretches encompassing almost 100 nt or more (i. e. nt 1—114, nt 1162—1298, nt 1300—1435, nt 1437—1535) were detected. The former region from 19 strains contained 15 distinct sequences and may not be essential for GbpC protein function. However, such polymorphic regions were useful and important