Research Article

Structural annotation of hypothetical proteins of *Streptococcus mutans*

M. Aparnaa, V. Vishnu Priya*, R. Gayathri

ABSTRACT

Introduction: *Streptococcus mutans* is one of the serious pathogens, facultative Gram-positive, anaerobic cocci predominantly found in the human oral cavity known to be involved in promoting tooth decay. Although the genome of this bacterium is already sequenced, there still remain proteins for which neither the function nor the structures are known yet. Therefore, in this study, we will be performing structural annotation of the hypothetical proteins present in *S. mutans* to find proteins that are similar in fold for further understanding. **Materials and Methods:** The FASTA sequence of the selected hypothetical protein of *S. mutans* was subjected to structural annotation using the online server HHpred. **Results and Conclusion:** It was found that the hypothetical proteins which were selected had glucosyltransferase function and structural annotation supported the finding.

KEY WORDS: Glucosyltransferase, HHpred, Hypothetical protein, *Streptococcus mutans*, Structural annotation

INTRODUCTION

Streptococcus mutans provides its name to a bunch of seven closely connected species collectively stated as the mutans streptococci (MS). The first habitats for *S. mutans* are mouth, pharynx, and gut. Several factors such as adherence to enamel surfaces, production of acidic metabolites, the capability to make up polyose reserves, and also the ability to synthesize extracellular polysaccharides (EPSs) are present in dental caries.[1] *S. mutans* and *Streptococci sobrinus* have a central role within the etiology of tooth decay because these will adhere to the enamel secretion investment and to alternative plaque microorganism. MS and lactobacilli are robust acid producers and, therefore, cause acidic surroundings making the chance for cavities. Usually, the looks of *S. mutans* within the tooth cavities are followed by cavity when 6–24 months.[2] The acidogenic *S. mutans* and *S. sobrinus* are ready to form EPS within the presence of disaccharide, however, additionally from levulose and aldohexose. The EPSs are long chained and high molecular mass polymers. The energy made glycosidic bond between the aldohexose and levulose moieties provides the free energy required for the synthesis of EPS. Aldohexose homopolysaccharides are referred to as glucans, whereas levulose homopolysaccharides are referred to as fructans. Glucans area unit made by glucosyltransferases (GTFs), whereas fructans area unit made by fructosyltransferases. The assembly of enormous quantities of EPSs from disaccharide is a crucial issue of *S. mutans* cariogenicity.[3]

S. mutans may be known by a chalky white spot on the surface of the tooth indicating a district of demineralization of enamel that is usually brought up as an unhealthy lesion.[4] Because the lesion more demineralizes, it will flip brown and can eventually end in a cavity.[5] Before the formation of the cavity, the method is reversible; however, once the streptococci mutans forms the cavity, the tooth structure is lost and cannot be regenerated. A lesion that seems shiny and dark brown suggests that a lesion was once present; however, the demineralization has stopped feat a stain. Because the enamel and dentin are destroyed, the cavity becomes a lot of noticeable.[6] The affected space of the tooth changes color and becomes sensitive. Once the decay passes through the enamel, the dentin tubules permit passages to the nerves creating the tooth exposed, which ends

Access this article online

Website: jprsolutions.info | ISSN: 0975-7619

Department of Biochemistry, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai- 600 077, Tamil Nadu, India.

***Corresponding author:** V. Vishnu Priya, Department of Biochemistry, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai - 600 077, Tamil Nadu, India. E-mail: drvishnupriyav@gmail.com

Received on: 24-12-2018; Revised on: 19-01-2019; Accepted on: 15-02-2019

in pain that quickly worsens with exposure to heat, cold, sweet food, or drinks. If the tooth is weakened by substantial amounts of internal decay, a fracture could occur below traditional change of state forces. Pain can arise a lot of constant once the decay has progressed enough to wherever the microorganism will overwhelm the pulp tissue that is located within the center of the tooth.[7] *S. mutans* and *S. sobrinus* are two microorganisms that reside within the mouth of an individual. *S. mutans* is carried by everyone, whereas *S. sobrinus* is carried between eight and 35 you look after individuals in several countries Table 1. These completely different microorganism strains may be distinguished by laboratory tests though the procedure is dear and time consuming; thus, it is not invariably practicable to spot the species during a

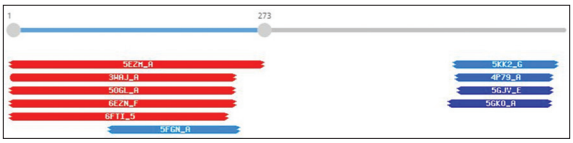

Figure 1: Sequence identity matching

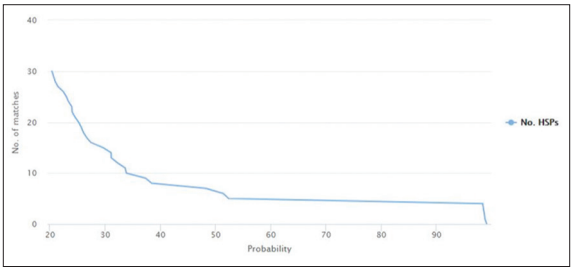

Figure 2: Probability distribution

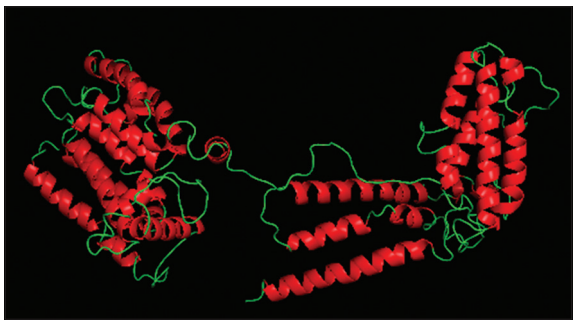

Figure 3: Three-dimensional model of the annotated protein (hypothetical) glucosyltransferase

massive medicine scale.[8] Researchers not fancied a selective media that may permit one to seem at the presence of one species in spit samples. Attributable to this, researchers have combined the two species along as MS conjointly called (MS) Table 2. Since streptococci mutans is a lot of prevailing than *S. sobrinus* most of the isolates can exclusively be *S. mutans*. MS may be refined during a laboratory by taking a sample of the microorganisms within the mouth. The life scientist will then use two varieties of selective media that are wide wont to isolate caries-related streptococci on Mitis-Salivarius agar and tryptone, yeast extract, and cystine agar within which antibiotic drug is added.[9] These varieties of media suppress the expansion of most microorganisms, however, permit the expansion of streptococci mutans and *S. sobrinus*. The inclusion of disaccharide ends up in the formation of glucans and also the look of colony formation to help identification.[10] Diagnostic kits utilized in dental clinics are supported similar selective media; however, it is vital to notice that they are measuring for total MS, not just *S. mutans*.[11] Although the ordering of this microorganism is already sequenced, there still stay proteins that neither the perform nor the structures are glorious yet.[12] Therefore, during this study, we are going to be performing structural annotation of the hypothetic proteins present in *S. mutans* to seek out proteins that are similar in the fold for more understanding.

MATERIALS AND METHODS

Data Collection

The National Center for Biotechnology Information (NCBI) is a leading and most sought database portal across the world by scientist to access different types of biological data. NCBI consists of several databases. Keyword *S. mutans* was used to search the list of entire genome entries of this species. Database search led to nine strains of this species whose whole genome had been sequenced. In this study, we chose the one which was recently updated. The strain that we took in this study was *S. mutans* LAB761 species (table). Further, keyword hypothetical was used filter out non-hypothetical proteins. Moreover, only those

Table 1: Function prediction using similarity search

#Organism/Name	Strain	Size (Mb)	GC%	Genes	Proteins	Release date	Modify date
S. mutans	UA159	2.03	36.8	2040	1958	October 25, 2002	May 21, 2017
S. mutans	NN2025	2.01	36.8	2024	1876	July 31, 2009	March 22, 2017
S. mutans	GS-5	2.03	36.8	2024	1842	June 25, 2012	March 23, 2017
S. mutans	LJ23	2.02	37.1	2028	1878	March 5, 2012	March 23, 2017
S. mutans	UA159-FR	2.03	36.8	2072	1940	January 9, 2015	March 25, 2017
S. mutans	NG8	1.98	36.9	1988	1844	February 11, 2016	March 27, 2017
S. mutans	LAR01	2.09	37	2123	1961	March 13, 2018	March 16, 2018
S. mutans	NCTC10449	2.02	36.9	2067	1909	June 17, 2018	June 25, 2018
S. mutans	LAB761	2.08	36.7	2117	1841	October 29, 2018	October 29, 2018

S. mutans: *Streptococcus mutans*. *p* value is <0.05 , Hence ,the data is considered to be statistically significant

Table 2: Structural annotation of AYO47875.1

Accesson#	Matching hits	Score	QC	E-value	Identity	Accession# of matched hit
AYO48030.1	MutB (*S. mutans*)	1732	1	0	0.85	AAG48566.1
AYO48451.1	Putative transmembrane protein (*Streptococcus suis*)	1105	0.99	0	0.56	CYV49150.1
AYO48449.1	Putative transmembrane protein (*S. mutans* A19)	1872	0.98	0	0.99	EMB91506.1
AYO48430.1						
AYO47875.1	Glucosyltransferase (*S. mutans*)	1075	0.88	0	0.99	BAC75681.1
AYO48733.1						
AYO48160.1	Signal peptide (*S. mutans*)	1044	0.91	0	0.99	SUN72694.1
AYO48862.1	Holliday junction-specific endonuclease (*S. mutans*)	1049	1	0	0.99	SUN72667.1
AYO47827.1						
AYO47622.1						
AYO48672.1	Radical SAM protein (*Streptobacillus moniliformis*)	95.1	0.71	6E-17	0.31	WP_064615423.1
AYO47878.1	Serotype determinant, transmembrane protein (*Streptococcus lutetiensis*)	387	0.9	2E-128	0.53	SQG57110.1
AYO47878.1	Beta-carotene 15,15'-monooxygenase (*Streptococcus macedonicus*)	258	0.89	6E-78	0.39	WP_099412027.1
AYO48575.1						
AYO47290.1						
AYO48073.1	DUF4071 domain-containing protein (*Dorea formicigenerans*)	336	0.94	3E-109	0.45	WP_117657998.1
AYO48261.1						
AYO48677.1	Renal dipeptidase (*Oceanobacillus* sp. YLB-02)	66.2	0.33	0.00000004	0.33	WP_121524145.1
AYO47584.1						
AYO48084.1	Nitroreductase family protein (*Hungatella* hathewayi)	384	0.99	3E-129	0.46	CUO02896.1
AYO48161.1	DNA primase (*Streptococcus troglodytae*)	646	1	0	0.95	BAQ24875.1
AYO47539.1	GNAT family N-acetyltransferase (*Streptococcus gallolyticus*)	498	1	4E-175	0.73	WP_064592523.1
AYO48675.1						
AYO47752.1						
AYO47145.1	ATP-binding protein (*Staphylococcus aureus*)	110	0.5	7E-23	0.43	WP_089536011.1
AYO47624.1						
AYO48259.1	Peptidase M50 (*Bacillus wiedmannii*)	120	0.89	3E-28	0.31	WP_098221316.1
AYO48259.1	Zn-dependent proteases (*S. pneumoniae*)	113	0.56	1E-26	0.35	CKI43669.1
AYO47186.1						
AYO47237.1						
AYO48008.1	Membrane protein (*S. mutans*)	546	1	0	1	WP_002262983.1
AYO47411.1						
AYO48369.1	Carboxypeptidase regulatory-like domain-containing protein (*Streptococcus orisasini*)	295	0.87	1E-97	0.61	WP_057491681.1
AYO48681.1	Seryl-tRNA synthetase (*S. pneumoniae*)	201	0.64	1E-61	0.57	WP_050269417.1

S. pneumonia: *Streptococcus pneumonia, S. mutans*: *Streptococcus mutans.* BLAST hits table. Blank rows indicates no match found

hypothetical proteins were chosen which above the median/average sequence length. Hence, a total of 32 hypothetical proteins were considered for further study.

Thirty-two hypothetical proteins were subjected to BLAST to search for similar sequences, thereby giving us some idea about the potential function of the sequences. BLAST is a similarity search tool which works on the principle of sequence alignment by search against the known sequence databases. The search was carried with default parameters.

From the BLAST search, we chose one protein for which there were good query coverage and sequence identity and also which has an important function. Therefore, we chose, AYO47875.1 found to be a putative GTF. The FASTA sequence of this protein was subjected to structural annotation using the online server HHpred. All the parameters at the HHpred were set to default.

RESULTS AND DISCUSSION

S. mutans is a type of acid-producing bacteria that have been reported to be involved in causing dental caries and plaque. This is caused by a mixture of microorganisms and food debris. *S. mutans* affects the hard tooth structure in the favorable environment

```
Q ss_pred        CCccchhhHHHHHHHHHHHHHHHHHHHHHHhhCCccccCCCCCCHHHHHHHHHHHHHcCCCCCccCCc--HHhhccccHHHHH
Q AYO47875.1   1 MNKNVTIKKSTLFYIILCFIGIFRIALLRSAPWELDANTGYDDLLQLKNAISIASGNWLGKTYSY--ISMTKNIGYPLFL  78 (603)
Q Consensus    1 m~~~~~~~~~~~~l~~il~l~lilrl~l~~~~~p~~~~~~~~~~~D~~~yi~~A~~l~~g~~~~~~~y~~~~~~~~~~~P~Ypl~L  78 (603)

T Consensus   22 ~~~~~~~~~~~~~~~~l~~~~~~~~~~~~~~~~~~~~~~~De~~~~~~a~~~~~~~~~~~~~~~~~~~~~ppl~~~l~  96 (578)
T 5EZM_A      22 QGAVGWSAATGWVVLFVAVALVVWFVSLDMRHL-----VGPDEGRYAEISREMFASGDWVTIRYNALKYFEKPPFHMWVT  96 (578)
T ss_dssp        ------CCTHHHHHHHHHHHHHHHHHHHGGGSSCC-----CTTHHHHHHHHHHHHHHHHCCSSSCEETTEECCSSCSHHHHHH
T ss_pred        chhcchHHHHHHHHHHHHHHHHHHHHHHcccccc-----cCCCHHHHHHHHHHHHHHHHCCCeeEEECCEeccCCCHHHHHHH

Q ss_pred        HHHHHC-CCCH---HHHHHHHHHHHHHHHHHHHHHHHcCCHHHHHHHHHHHHHhCccchhhHHHHHHHhcHHHHHHHHHHHHH
Q AYO47875.1  79 ALTQLL-NIPY---SVLYGLLISLSSFSFIKAIQPVVKSKKLLLLIFFVIIFTPINHGAFYRIYRNALVPWVLLLIISSY  154 (603)
Q Consensus   79 a~~~~l-g~~~~~~l~~~ll~~la~~ll~~~~~~~l~~~~~~~~~~~~~~ll~l~~P~~~~~~~~~~~i~~e~l~~~~l~ll~l~~l  154 (603)

T Consensus   97 ~~~~~l~g~~~~~~r~~~~l~~~~~~~~~~~y~l~~~~~~~~~~a~~a~~l~~~~p~~~~~~~~~~~~~~~~~~~~~~~~l~~~  175 (578)
T 5EZM_A      97 VVGYELFGLGEWQARLAVALSGLLGIGVSMMAARRWFGAR-AAAFTGLALLAAPMWSVAAHFNTLDMTLAGVMSCVLAFM  175 (578)
T ss_dssp        HHHHHHHCSSHHHHTHHHHHHHHHHHHHHHHHHHHHHHHHCHH-HHHHHHHHHHHCHHHHHHHHTSCCHHHHHHHHHHHHHHHHH
T ss_pred        HHHHHHHCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhChh-HHHHHHHHHHHhHHHHHHHHhHHhhHHHHHHHHHHHHHHH

Q ss_pred        HHHHHhc-----ccccccHHHHHHHHHHHHHHHHHHhcCcchhhHHHHHHHHHHHHHHHHHHHHHHHcCCCHHHHHHHHHHHHHHHH
Q AYO47875.1 155 IAIFIRR-----RDRLSLFLPWTILAFFSIMYFWTLREDSIWILPFILVAAITIIITIVILYKKDLHEILLRSVLVLLPL  229 (603)
Q Consensus  155 l~~~~~~~~~~~~~~~~~~~~~~~~~~~~l~Gl~lgl~~ltR~~~i~llp~l~l~ll~~~~~~~~~~~~~~~~~~~~~~~~~~l~~  229 (603)

T Consensus  176 ~~~~~~~~~~~~~~~~~~~~~~~~~~g~~~gl~~~~k~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  240 (578)
T 5EZM_A     176 LMGQHPDASVAAR------RGWMVACWAAMGVAILTKGLVGIALPGLVLVVY------TLVTRDWGLWRRLH---LALGV  240 (578)
T ss_dssp        HHHTCTTSCHHHH------HHHHHHHHHHHHHHHHHHHTTHHHHHHHHHHHHHHH------HHHSCCTTHHHHTC---HHHHH
T ss_pred        HHHhCCCcchhcc------cHHHHHHHHHHHHHHHHCccHHHHHHHHHHHHHH------HHHcCCchHHHHhh---HHHHH

Q ss_pred        HHHHHHHHHHHHHHHHHHHHCCceeecCCCHHHHHHHHHhhhcccCC
Q AYO47875.1 230 IGIVVSNVWVSAINYSYYGIWGVNDRSDTAAAKAMSLIYKIEDN  273 (603)
Q Consensus  230 l~~~~~~~~~~~~N~~~~G~~~~~~~~~~~~~~~f~~~~~~~l~~i~~~  273 (603)
                  |
T Consensus  241 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  284 (578)
T 5EZM_A     241 VVMLVITVPWFYLVSVRNPEFPNFFFIHEHWQRYTSNIHSRSGS  284 (578)
T ss_dssp        HHHHHHHHHHHHHHHHHHHCTTHHHHHHHHCCCCCC-------CCCC
T ss_pred        HHHHHHHHHHHHHHHHHHHCccchHHHhhhhhHHHHhcccccCCCC
```

Figure 4: Secondary structure prediction of the annotated protein. H – Helix; C – Coils; e – strand

such as the presence of sugars. Several strains of this species have been isolated and researched upon. So far, at the time of writing this manuscript, there were totally nine different strains for which the whole was completed sequenced.[13] However, still, there remains a lot of proteins categorized as hypothetical.

In this study, we chose LAB761 strain with total genome size as 2.08 Mb having a GC% of 36.7 part of 2117 genes, leading to 1841 proteins. Of these, 223 were hypothetical proteins. We took median and average into consideration and decided to choose only those sequences which are above the median value in sequence length.[14] Like this, 32 hypothetical proteins were shortlisted and the BLAST results are shown in the table above. Fourteen sequences did not find any similar sequences, and therefore, they were not considered any further. Furthermore, it can be seen from the table that the sequence identity with the matching hits ranged from 31% to 100% and the query coverage 33% to 100%.[15]

From this table, we decided to take up putative GTF (AYO47875.1) protein. The structural annotation for this protein was performed using the HHpred server.[16] Several entries were reported from the HHpred server and it can be seen from the Figure 1 that the best matching one was holding the protein data bank (PDB) ID 5EZM_A which was matching with the query sequence from residue 1 to 273.[17-20]

From the probability distribution chart for the structural annotation performed protein, 31 high scoring hits were found with a probability score ranging from 20 to 95 Figure 2. Based on these annotations, a tertiary structure was modeled and its secondary structural predictions were made. The model structure is shown in Figure 3 and 4. From this figure, it can be seen that it completely consists of alpha helices and coils. Moreover, the protein model has three structural domains. All the structural matches were found to be having oligosaccharide-binding function, peptide-binding protein function, and transmembrane oligosaccharide-binding domain among others. From the secondary structure prediction, it is clear that predominantly the modeled structure consists of most of the residues to be involved and confined in helical structure.

CONCLUSION

This study was undertaken to annotate hypothetical protein of *S. mutans* structurally and functionally. It was found that the hypothetical proteins which were selected had GTF function and structural annotation supported the finding.

REFERENCES

1. Ajdić D, McShan WM, McLaughlin RE, Savić G, Chang J, Carson MB, *et al.* Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc Natl Acad Sci U S A 2002;99:14434-9.
2. Bailey TL, Gribskov M. Combining evidence using p-values: Application to sequence homology searches. Bioinformatics 1998;14:48-54.
3. Banas JA. Virulence properties of *Streptococcus mutans.* Front Biosci 2004;9:1267-77.
4. Banas JA, Vickerman MM. Glucan-binding proteins of the oral streptococci. Crit Rev Oral Biol Med 2003;14:89-99.

5. Bhagwat SP, Nary J, Burne RA. Effects of mutating putative two-component systems on biofilm formation by *Streptococcus mutans* UA159. FEMS Microbiol Lett 2001;205:225-30.
6. Burne RA, Schilling K, Bowen WH, Yasbin RE. Expression, purification, and characterization of an exo-β-d-fructosidase of *Streptococcus mutans*. J Bacteriol 1987;169:4507-17.
7. Peterson JD, Umayam LA, Dickinson T, Hickey EK, White O. The comprehensive microbial resource. Nucleic Acids Res 2001;29:123-5.
8. Lawrence JG. Selfish operons and speciation by gene transfer. Trends Microbiol 1997;5:355-9.
9. Lawrence JG. Gene transfer, speciation, and the evolution of bacterial genomes. Curr Opin Microbiol 1999;2:519-23.
10. Ochman H, Lerat E, Daubin V. Examining bacterial species under the specter of gene transfer and exchange. Proc Natl Acad Sci U S A 2005;102 Suppl 1:6595-9.
11. Kado CI. Horizontal gene transfer: Sustaining pathogenicity and optimizing host-pathogen interactions. Mol Plant Pathol 2009;10:143-50.
12. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature 2000;405:299-304.
13. Ragan MA, Harlow TJ, Beiko RG. Do different surrogate methods detect lateral genetic transfer events of different relative ages? Trends Microbiol 2006;14:4-8.
14. Beiko RG, Ragan MA. Detecting lateral genetic transfer: A phylogenetic approach. In: Keith JM, editor. Bioinformatics: Data, Sequence Analysis, and Evolution. Totowa, NJ: Humana Press; 2008. p. 457-69.
15. Su XD, Liang Y, Li L, Nan J, Brostromer E, Liu P, *et al*. A large-scale, high-efficiency and low-cost platform for structural genomics studies. Acta Crystallogr D Biol Crystallogr 2006;62:843-51.
16. Vonrhein C, Blanc E, Roversi P, Bricogne G. Automated structure solution with auto SHARP. Methods Mol Biol 2007;364:215-30.
17. Sabri M, Häuser R, Ouellette M, Liu J, Dehbi M, Moeck G, *et al*. Genome annotation and intraviral interactome for the streptococcus pneumoniae virulent phage dp-1. J Bacteriol 2011;193:551-62.
18. Afonine PV, Grosse-Kunstleve RW, Adams PD. The Phenix Refinement Framework. Berkeley: CCP4 Newsletter; 2005.
19. Auswin MK, Ramesh S, Gayathri R, Priya VV. Inhibition of superoxide dismutase using herbal compounds to treat oral diseases caused by *Streptococcus mutans* an *in silico* study. J Adv Pharm Educ Res 2017;7:148.
20. Pillay SR, Geetha RV. *In vitro* evaluation anti bacterial activity of *Coriandrum sativum* extract in reducing *Streptococcus mutans* count. Drug Invent Today 2018;2:10.

Source of support: Nil; Conflict of interest: None Declared