



Proteomic Characterization and Target Identification Against *Streptococcus mutans* Under Bacitracin Stress Conditions Using LC–MS and Subtractive Proteomics

Sahar Zaidi¹ · Tulika Bhardwaj² · Pallavi Somvanshi^{2,3} · Asad U. Khan¹

Accepted: 25 December 2021 / Published online: 6 January 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The aim of the present study, is to identify potential targets against the highly pathogenic bacteria *Streptococcus mutans* that causes dental caries as well as the deadly infection of endocarditis. The powerful and highly sensitive technique of liquid chromatography-mass spectrometry (LC–MS/MS) identified 321 proteins of *S. mutans* when grown under stressful conditions induced by the antibiotic bacitracin. These 321 proteins were subjected to the *insilico* method of subtractive proteomics to screen out potential targets by utilizing different analyses like CD-HIT, non-homologous sequence screening, KEGG pathway, essentiality screening, gut-flora non-homology, and codon usage analysis. A database of essential proteins was employed to find sequence homology of non-paralogous proteins to determine proteins which are essential for bacterial survival. Cellular localization analysis of the selected proteins was done to localize them inside the cell along with physico-chemical characterization and druggability analysis. Using computational tools, 22 proteins out of 321, that are functionally distinguishable from their human counterparts and passed the criterion of a potential therapeutic candidate were identified. The selected proteins comprise central energy metabolic proteins, virulence factors, proteins of the sortase family, and essentiality factors. The presented analyses identified proteins of the sortase family, which appear as key therapeutic targets against caries infection. These proteins regulate a number of virulence factors, thus can be simultaneously inhibited to obstruct multiple virulence pathways.

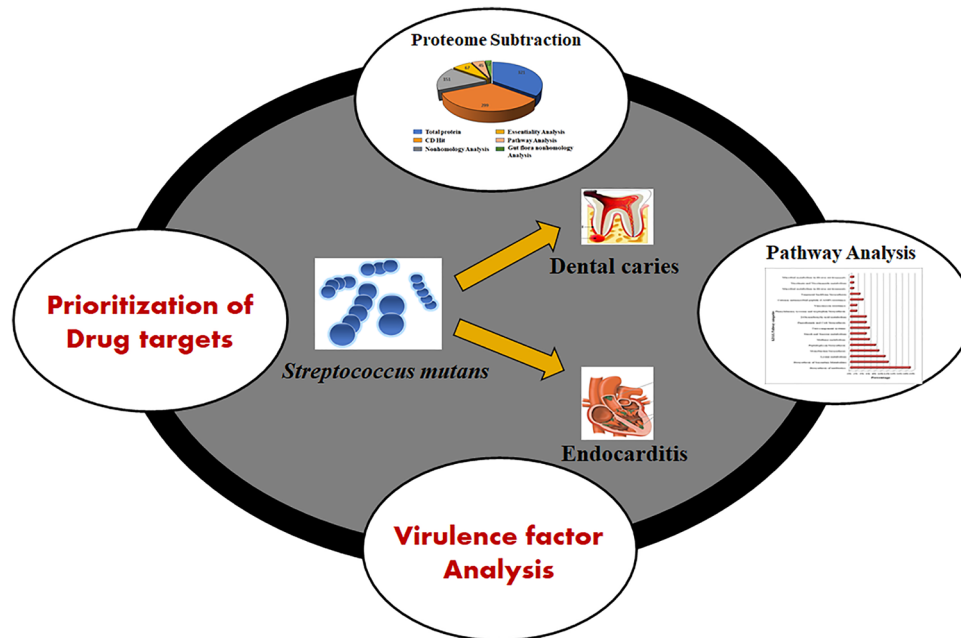
✉ Asad U. Khan
asad.k@rediffmail.com; asadukhan72@gmail.com

¹ Medical Microbiology and Molecular Biology Laboratory, Interdisciplinary Biotechnology Unit, Aligarh Muslim University, Aligarh, UP 202002, India

² School of Computational & Integrative Sciences (SC&IS), Jawaharlal Nehru University, New Delhi 110067, India

³ Special Centre of Systems Medicine (SCSM), Jawaharlal Nehru University, New Delhi 110067, India

Graphical Abstract



Keywords Bacitracin · LC–MS · Essentiality analysis · *S. mutans* · Subtractive proteomics

1 Introduction

Streptococcus mutans was first isolated by J. Clarke in 1924, from carious lesions. Nevertheless, it was around 1950s, when this bacterium grabbed attention of a number of scientific communities as well as clinical and animal-based laboratories, which revealed *S. mutans* as a significant etiologic agent in a number of dental pathologies. *S. mutans* generally resides inside the human mouth or more precisely the dental plaque, which is a biofilm constructed on tooth's hard surface by diverse range of bacteria. Besides, this bacterium is also implicated in bacteremia and infective endocarditis, a deadly heart valves inflammation [1, 2]. Unique feature that makes *S. mutans* one of the most potent cariogenic organisms include the capability of *S. mutans* to metabolize a broad range of carbohydrates into organic acids (acidogenicity) and also it can flourish even under acidic conditions (aciduricity) [3].

While genetic as well as biochemical approaches have been employed for the past few decades to explore the biology of *S. mutans*, the whole sequence of genome of *S. mutans* strain UA159 was only published in 2001, which transformed the landscape dramatically. Currently, *S. mutans* is among the finest characterized Gram-positive bacteria. The genome of *S. mutans* that was first sequenced (serotype c UA159 strain) contained ~2.0 Mb of DNA, encoding around 2000 genes [4, 5]. However, the transcriptome and proteome

of *S. mutans* has got less attention, though they are known to be significant to understand pathogenesis of any bacteria. Proteome is regarded as cell's functional moiety that fluctuates in response to the stimuli imposed externally such as drug stress. Under stressful conditions, harmony inside the bacterial cell is disturbed and alternate cellular functions are adopted by bacteria simultaneously that proteome mediates in order to bypass the effect. Consequently, proteome obtained when bacteria is grown under stressful conditions, could unravel some novel pathways that bacteria adopt to avoid perturbations around. Besides, proteomics along with bioinformatics approach could further assist in resolving such biological problems [6, 7]. The current study intends to understand the pathogenesis of the cariogenic bacteria *S. mutans*. Literature mining was carried out to construct a novel and robust pipeline amalgamating the proteomics analysis employing LC–MS and subtractive proteomics approach to identify potential targets against caries infection. Protein samples of *S. mutans* grown under bacitracin stress condition and without stress, were subjected to LC–MS. Antibiotic bacitracin was selected, since the drug impacts the functioning of the cell by opting multiple routes and disturbs the total cell's physiology. And thereby it elevates the chances of obtaining differential expression of maximum number of virulence proteins [8]. Bacitracin obstructs the process of bacterial cell wall biosynthesis as it hinders the critical step of dephosphorylation of C55-isoprenyl pyrophosphate (IPP)

and certain other cellular processes like action of a number of hydrolytic enzymes, ubiquinone precursors formation, synthesis of membrane derived oligosaccharides as well as role of membranes in cell division process [9]. Proteins identified by LC–MS then underwent subtractive proteomics approach to identify non-homologs against human proteome, followed by essentiality as well as KEGG pathway analysis. Eventually proteins were screened for non-homologs against gut microflora proteome. Once shortlisted, the qualitative characterization of proteins was done. The putative target proteins that have been shortlisted can act as promising candidates to construct therapeutic to control *S. mutans* virulence. The aim of this study is to provide resourceful data to researchers, who are exploring targets or working on drug discovery, so as to decode biological queries in a computational tractable method by exploring, filtering and weighting the huge size of proteome -scale data sets.

2 Materials and Methods

2.1 Bacteria, Media, Culturing Conditions and Drug Susceptibility Testing

Institute of Microbial Technology, Chandigarh, India provided the *Streptococcus mutans* MTCC 497 strain employed in this study. Brain heart infusion broth (BHI) with 1% sucrose (Himedia Labs, Mumbai, India) was used to grow bacteria. Antibiotic bacitracin (Sigma Chemical Company) was utilized at sub-MIC concentrations. Sub-MIC concentration of bacitracin was determined by diluting the stock of 10 mg/ml of antibiotic serially in 96 well microtiter plate by means of microdilution method [10].

2.2 Culture and Drug Induction

S. mutans was grown in Brain heart fusion broth (BHI) for overnight at 37 °C and 220 rpm. Secondary culture was grown in two flasks using overnight culture (primary culture), and bacitracin at half of the MIC (39 µg/ml) value was added to one of the flasks labelled as treated while no drug was added to control. Cells in both flasks were grown up to mid log phase (OD₆₀₀ = 0.5–0.6) and were harvested by the technique of centrifugation operated at 10,000 rpm for 10 min at 4 °C. The precipitated cell pellet obtained was subsequently kept at –80 °C till needed. All the experiments that have been performed in this study were replicated thrice biologically.

2.3 Protein Extraction

Washing of bacterial cells was done with normal saline, and then lysis buffer was used to re-suspend the washed cells.

Composition of lysis buffer used was 50 mM Tris–HCl containing 0.1% sodium azide, 10 mM MgCl₂, 1 mM ethylene glycol tetra acetic acid (EGTA; pH 7.4) and 1 mM phenylmethylsulfonyl fluoride (PMSF) at the concentration of 5 mg per 1 g wet weight. Cells were broken down by sonicating, intermittently, at 40% amplitude (Sonics & Materials Inc., Newtown, CT, USA). Homogenate obtained was then subjected to centrifugation operated at the speed of 12,000×g, for the duration of 20 min, at 4 °C. Precipitation of supernatant was done by mixing cold acetone to supernatant in ratio 1: 4 and storing at –20 °C for overnight. After incubation proteins present in the supernatant were precipitated and obtained by centrifuging at 12,000 rpm, for 30 min approximately and were allowed to dry in the air. Finally, in the appropriate volume of dissolving buffer, proteins were solubilized and concentration of protein was calculated with the help of Bradford assay [11].

2.4 Separation as Well as Identification of Proteome by nanoLC-TripleTOF5600MS

Trypsin was used to digest protein samples taken in equal amounts that were further studied employing Triple TOF 5600 mass-spectrometer (AB-Sciex, USA) which is provided with Eksigent MicroLC 200 system (Eksigent, Dublin, CA) as well as possesses an Eksigent C18 reverse phase column (150×0.3 mm, 3 µm, 120 Å). For each sample, data dependent analysis (DDA) was carried out with the purpose of generating spectral library that can be used for protein identification experiment, as well as to produce spectral ion libraries of superior quality to be used for SWATH analysis, with specific parameters to operate the mass spectrometer. Information dependent acquisition (IDA) mode was employed to generate spectral library after inserting tryptic digest of 2 µg on column by means of Eksigent nano LC-Ultra™2D plus system that possesses SCIEX Triple TOF @5600 system equipped with nano sprayIII source. Sample afterwards were loaded on trap (Eksigent Chrom XP 350 µm×0.5 mm, 3 µm 120 Å) and washing was done for 30 min at the rate of 3 µl/min. A gradient in multiple steps (range varying between 5 and 50% acetonitrile in water, with 0.1% formic acid) of 120 min was set up so that peptides from Chrom XP3-C18, 0.075×150 mm, 3 µm, 120 Å analytical column can be eluted [11].

2.5 Information Dependent Acquisition (IDA) Parameters

IDA is the ion library generation procedure was employed in the study. In this up to 20 most strong and multiple charged ions per MS cycle were chosen to achieve MS/MS fragmentation. Each of the ions, were subjected to a dynamic

exclusion criterion for 10 s. The accumulation time was adjusted to 70 ms for each MS/MS experiment.

2.6 SWATH Parameters for Label Free Quantification

In SWATH acquisition method, adjustment of Q1 transmission window to 12 Da from the mass range for 350–1250 Da. With an accumulation duration of 62 ms, sum of 75 windows was attained independently, with each set having three technical replicates. Total cycle time i.e. < 5 s was kept constant. Spectral library was generated by employing Protein Pilot™ v.5.0. Spectral alignment and the peak extraction were carried out for label free quantification with the help of Peak View® 2.2 software with parameters that have been defined example, transitions number as 5, 95% as peptide confidence, peptide number as 2, 30 ppm as XIC width, 3 min as XIC extraction window. Data was then, processed by Marker View software V1.3 (AB Sciex) in order to obtain interpretation of statistical data. In Marker View, normalization of peak area underneath the curve for the peptides that were selected was done with internal standard protein (beta galactosidase) spike throughout the SWATH accumulation. Results displayed as AUC of the ions; ions total intensity for the peptide as well as peptide total intensity for protein in the form of three output files. SWATH Acquisition Micro App 2.0 in Peak View® Software. 2.7 was used to process all SWATH Acquisition data.

2.7 Data Analysis

Protein Pilot Software v. 5.0 (AB SCIEX, Foster City, CA) was employed to process Data analysis by using Paragon and Pro group Algorithm. Analysis was also completed by means of integrated tools in Protein Pilot at the 1% false discovery rate (FDR).

2.8 Subtractive Proteome Screening

2.8.1 CD-Hit Analysis

Identified proteome dataset was subjected to CD-HIT analysis to remove duplicity in dataset based on clustering algorithm. An identity cut-off of 0.6 (60% identity) was selected to screen proteomes as proteins sharing 60% identity with relatively similar structure and functional domains. A bandwidth of 20 amino acids was selected to perform sequence based global alignment [12].

2.8.2 Non-homologous Sequence Analysis

Pathogen specific screening of drug targets was supported by performing non-homology search against human proteome.

This prevents the cross reactivity and binding of therapeutic compounds with host proteome. Therefore, non-redundant protein sequence (nr) database of host *Homo sapiens* (taxID: 9606) was subjected to similarity search analysis against bacterial proteome dataset (321) employing BLASTp at threshold cut-off e-value of 0.5 with identity of 60% [13, 14].

2.8.3 Essentiality Analysis

Non-homologous dataset underwent similarity screening using Database of Essential Genes (DEG) version 15.2. The complete screening was performed using BLASTp against experimentally validated 53,885 essential proteins along with 786 essential non-coding sequences of pathogens. Proteins hit with expectation value $\leq 10^{-100}$, identity $\geq 25\%$ and minimum bit score 100 as parameters were considered as essential proteins of the pathogen [12–14].

2.8.4 KEGG Pathway Analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) organism code was utilized for the retrieval of metabolic pathways of disease-causing bacteria and host *Homo sapiens*. KAAS server at KEGG database was utilized to identify essential proteins participating in pathways specific for pathogen. KAAS offers genes functional annotation by means of BLAST comparison compared with the manually curated database of KEGG GENES. KO (KEGG Orthology) assignments define metabolic proteins. KAAS also produces KEGG pathways which display these metabolic proteins [12].

2.8.5 Non-homology Analysis Against Gut-Flora

This analysis was performed to filter out essential proteome content sharing structural similarity with gut microbe proteome as it prevents toxicity caused by gut microbe proteome interaction with drugs. *In-house* temporary database containing proteome dataset of gut microbes was developed using MySQL and PHP (Hypertext Preprocessor). Pathogen specific essential protein dataset then underwent non-homology sequence similarity screening against gut microbe proteome using BLASTp at bit score of 10 and threshold e-value of 10^{-10} and identity of 0.6 (60% similarity) [13, 14].

2.8.6 Codon Usage Analysis

Codon adaptation Index (CAI) defines the geometric mean of relative fitness values for each synonymous codon which ranges between 0 and 1. Higher the expression values of essential genes higher is the suitability for consideration as drug target [15].

2.9 Qualitative Characterization

2.9.1 Sub Cellular Localization Analysis

This analysis categorizes essential proteins significance of serving as either drug or vaccine target. CELLO 2.0 [16], a two-level support vector machine tool comprising 1444 and 7589 bacterial and eukaryotic proteins, respectively, were taken as training dataset was utilized to identify the protein subcellular localization of prioritized proteins after gut-flora non-homology analysis [13, 14, 16].

2.9.2 Physico-Chemical Characterization

ProtParam tool (<https://web.expasy.org/protparam/>) at ExPASy webserver enables physico-chemical characterization of essential proteome datasets. It is used to determine isoelectric point, molecular weight, nature of charge (positive or negative), GRAVY, extinction coefficient of prioritized proteomes, aliphatic as well as instability index [13].

2.9.3 Druggability Analysis

DrugBank 3.0 [17], including therapeutics that have been validated experimentally and approved by FDA along with target details was employed and probable (E-value) cut off of 10^{-5} was included to characterize potential targets capable of binding to a number of therapeutics with high affinity. Additionally, Therapeutic Targets Database (TTD) was investigated and studied to explore targets for nucleic acid, infections or diseases that can be targeted, pathway details as well as the equivalent drugs molecules aiming specifically the targets possessing significant E-value i.e. < 1 [12, 16].

3 Results

3.1 Protein Identification Using LC–MS/MS

S. mutans was cultured under bacitracin drug stress (half-MIC i.e., 39 µg/ml) [18]. The proteome of *S. mutans* was analyzed by LC–MS/MS, by means of SWATH workflow. Proteins were enumerated at 1% FDR. 321 proteins were identified and the scale of differential expression was characterized using log folds change vs. p-value ratio, tabulated in Supplementary file 1 (S1). These 321 proteins underwent subtractive proteome scrutiny to further cut down the list.

3.2 Subtractive Proteome Screening

3.2.1 Non-homology Analysis

Proteins with human homologs were eliminated and only non-human homologs were screened out employing

comparative non-homology proteome analysis. This analysis identified approximately 67% evolutionary relatedness among pathogen and host [15]. Among 321 pathogen proteome dataset, 151 protein sequences were identified non-homologous (S2; Sheet 1) to host *Homo sapiens* after BLASTp at NCBI server with 0.005 (E = 0.005) as E-value cutoff and also a match of greater than 50% of query length.

3.2.2 Essentiality Analysis

The core proteome dataset constituting essential proteins, which are necessary for bacterial existence as well as persistence. Thus, essential proteins are regarded as potential targets for identification of novel and innovative chemical scaffolds for the purpose of drug development. Several studies have demonstrated the potential of essential proteins towards synthetic biology [19]. Functional codes were utilized to define such system of classification in accordance with COG (Cluster of Orthologous Groups of proteins) classification [13, 15]. The collection of data containing pathogen sequences that are non-homologous to humans were then employed to predict their essentiality. It utilized BLASTp, with the dataset of essential proteins of bacteria employing DEG version 15.2 with an E-value cutoff of 0.0001. In the list of essential proteins (S2 Sheet 2), that were identified, the greatest value of significant similarities with *S. sanguinis* (35 hits), *Francisella novicida* U122 (16 hits), *M. tuberculosis* H37Rv (8 hits), *S. pneumoniae* R6 (5 hits), *Helicobacter pylori* 26,695 (2 hits) along with *Mycoplasma pulmonis* UAB CTIP (1 hit).

3.2.3 KEGG Pathway Analysis

Pathways regulating metabolism of the host *Homo sapiens* as well as the pathogen were retrieved from the KEGG server. These pathways were grouped into different categories, namely: metabolism, genetic information processing, and environmental information processing. Manual comparative analysis among host and pathogen-related pathways indicates the presence of 21 pathogen-related unique pathways. The KAAS webserver helps in the identification of participation of the essential proteins identified (67) towards major metabolic pathways of the pathogen in terms of percentage. After obtaining the score and related metabolic pathways from KAAS for each essential protein, a percentage calculation of participation was performed using *in-house* pipeline and cross-validated manually [15] (S2 Sheet 3). The majority of essential proteins participate in biosynthesis of antibiotics (19%) [20, 21], biosynthesis of secondary metabolites (12%) [22], lysine metabolism (11%) [23], monobactam biosynthesis (9%) [24], peptidoglycan biosynthesis and degradation proteins (8%) [25, 26], methane metabolism (6%) [27], starch and sucrose metabolism (5%) [28], two component

system (TCS) (6%) [29, 30], pantothenate and CoA biosynthesis (5%) [31, 32], 2-Oxocarboxylic acid metabolism (5%) [33], phenylalanine, tyrosine and tryptophan biosynthesis (2%) [34, 35], vancomycin resistance (2%) [36], cationic antimicrobial peptide (CAMP) resistance (4%) [37], Jacobs terpenoid backbone biosynthesis (3%) [38], metabolism of bacteria in diverse environments (1%) [39] and nicotinate and nicotinamide metabolism (1%) [40] as shown in Fig. 1.

3.2.4 Non-homology Analysis Against Gut-Flora

Vitamin synthesis in human gut is assisted by microflora of the gut [41, 42] along with other processes that gut microflora performs like degradation of xenobiotics [43] and assimilation of incompletely digested components particularly dietary compounds [44]. Furthermore, they avert pathogenic as well as opportunistic bacteria present in the gut from establishing a foothold, hence are crucial for human wellbeing. For the development of a temporary database, FASTA sequences of the Human Microbiome Project database Reference proteome were mined [45]. The essential proteome dataset was subjected to sequence-based similarity screening at default parameters. Also, non-homology screening against gut-flora prioritizes sequences showing fewer than the identity of 30% and identified 22 proteins as potential drug targets (S2 Sheet 4),

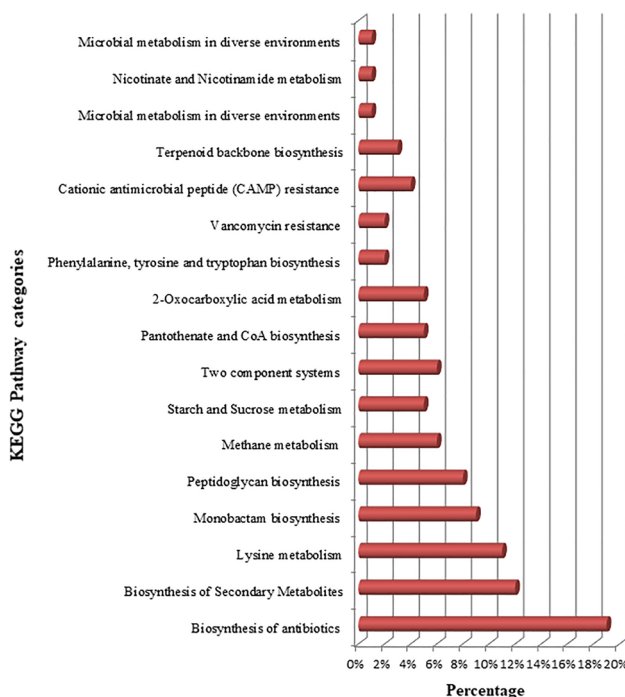


Fig. 1 KEGG Pathway analysis identifies 21 pathways unique to bacteria only. KAAS webservice encodes majority of the selected proteins involved in the biosynthesis of antibiotics (19%) followed by secondary metabolites biosynthesis (12%) and lysine metabolism (11%)

3.2.5 Codon Usage Analysis

The computation of CAI values assists in scrutinizing non-homologous essential proteins, expressing their translational efficiency ranging from 0 to 1. The prioritized protein sequences support major criteria of essentiality and selectivity of proteins to serve as targets (Table 1). The higher the translational efficiency of the protein, the higher the potential to serve as a drug target [46]. Therefore, we further subjected the dataset to qualitative characterization, including sub-cellular localization, physico-chemical characterization, and druggability analysis. (An illustration of all proteome subtractive methods is given in Table 2.

3.3 Qualitative Characterization

3.3.1 Sub Cellular Localization Analysis

CELLO version 2 assists in the selective distribution of prioritized drug targets within a bacterial cell. The sorting of protein sequences was mainly in the cell membrane, the extracellular matrix, and cytoplasmic domains (Table 1).

3.3.2 Physico-Chemical Characterization

The physico-chemical properties of essential protein sequences were computed using ProtParam based on amino acid residue composition. Instability index, aliphatic index, and extinction coefficient predict the *in-vivo* lifespan of protein sequences which are considered as targets. Such analysis supports quality control and assurance procedure during drug development [47, 48]. Instability index > 40 predicts stability of proteins *in-vivo* experiments. The space that aliphatic side chains occupy represents aliphatic index while GRAVY is an indicator of sum of hydrophathy values all amino acids divided to the total number of residues of amino acids in a protein. These parameters of selected drug targets are listed in Table 3

3.3.3 Druggability Analysis

Similarity search analysis against the therapeutics approved by the FDA, nutraceutical as well as experimental small-molecule compounds using BLASTp rendered 22 potential drug targets. Shortlisted and selected targets can be exploited for drug designing as they are harmless to the host (Table 4).

4 Discussion

Highly pathogenic Gram-positive bacteria *Streptococcus mutans*, is the chief causal organism in the pathogenesis of dental caries in humans as well as it may also be the source

Table 1 List of 22 potential target, their function, location and codon adaptation index

S.No.	Potential targets	Function	Location	Codon adaptation index
1	Glucose-6-phosphate isomerase	Glycolysis Biosynthesis of antibiotics Biosynthesis of secondary metabolites	Cytoplasmic	0.76
2	4-Hydroxy-tetrahydrodipicolinate reductase	Lysine biosynthesis Monobactam biosynthesis	Cytoplasmic	0.80
3	Glycogen synthase	Biosynthesis of secondary metabolites	Cytoplasmic	0.71
4	Glucan-binding protein GbpC	Bacterial adherence and extracellular polymeric substance (EPS) synthesis	Extracellular	0.75
5	Cell surface antigen SpaP	Adherence	Extracellular	0.91
6	Thiol peroxidase	H2O2 Receptor and Redox-Transducer	Cytoplasmic	0.87
7	Levansucrase	Two-component system Starch and sucrose metabolism	Extracellular	0.89
8	Phospho-2-dehydro-3-deoxyheptonate aldolase	Phenylalanine, tyrosine and tryptophan biosynthesis Biosynthesis of secondary metabolites	Cytoplasmic	0.79
9	ATP synthase F0F1 subunit alpha	Oxidative phosphorylation	Cytoplasmic	0.92
10	ATP synthase F0F1 subunit beta	Oxidative phosphorylation	Cytoplasmic	0.87
11	Glucosyltransferase-SI (GtfC)	Two-component system	Extracellular	0.88
12	Glucosyltransferase-I (GtfB)	Two-component system	Extracellular	0.89
13	secreted antigen GbpB/SagA	Peptidoglycan biosynthesis and degradation proteins	Extracellular	0.92
14	Formate acetyltransferase/ Pyruvate formate-lyase	Pyruvate and butanoate metabolism	Cytoplasmic	0.93
15	Fructose-1,6-biphosphate aldolase	Glycolysis	Cytoplasmic	0.80
16	Endolytic murein transglycosylase	Terminates nascent peptidoglycan synthesis	Extracellular	0.88
17	L-lactate dehydrogenase	Glycolysis / Gluconeogenesis Cysteine and methionine metabolism	Cytoplasmic	0.82
18	Ketol-acid reductoisomerase	Pantothenate and CoA biosynthesis 2-Oxocarboxylic acid metabolism	Cytoplasmic	0.92
19	Aspartate-semialdehyde dehydrogenase	Lysine biosynthesis Monobactam biosynthesis	Cytoplasmic	0.97
20	Enolase	Glycolysis Methane metabolism	Cytoplasmic	0.85
21	UDP-N-acetylglucosamine 1-carboxyvinyltransferase I	Peptidoglycan biosynthesis	Cytoplasmic	0.95
22	ATP-dependent protease ClpE	Genetic information processing	Cytoplasmic	0.79

Table 2 Details of the proteome subtractive and qualitative characterization methods

<i>Streptococcus mutans UA159</i>						
Module	Proteome subtraction methods	Softwares	Total dataset	Selected	Excluded	
1	Total proteome content	CD-Hit	321	299	22	
	Non-homology analysis	BLASTp	299	151	148	
2	Essentiality analysis	DEG	151	67	84	
	Pathway analysis	KEGG	67	45	22	
	Gut-flora non-homology analysis	BLASTp	45	22	23	
3	Codon usage analysis	CAI	22	ALL	–	
	Subcellular localization analysis	CELLO	22	22	–	
	Physicochemical characterization	ProtParam	22	22	–	
	Druggability analysis	DrugBank	22	22	–	

Table 3 List of the physico-chemical properties of selected potential targets

Potential drug targets	Number of Amino acids	Molecular weight	Isoelectric Point	negatively charged residues	positively charged residues	aliphatic index	Instability index	Extinction coefficient ($M^{-1} \text{ cm}^{-1}$)	GRAVY	Nature
Glucose-6-phosphate isomerase	449	49,421.66	4.84	63	45	88.69	27.26	54,780	-0.232	Stable
4-Hydroxy-tetrahydrodipicolinate reductase	255	27,804.82	5.07	39	27	69.51	20.48	239,060	-0.063	Stable
Glycogen synthase	476	54,311.80	5.69	45	36	87.01	32.11	7572	-0.287	Stable
Glucan-binding protein GbpC	583	63,349.90	6.75	20	20	104.93	25.67	4576	0.079	Stable
Cell surface antigen SpaP	1562	169,971.93	5.3	35	31	104.52	23.72	16,008	-0.223	Stable
Thiol peroxidase	161	17,552.71	5.08	36	31	122.83	27.03	24,535	-0.061	Stable
Levanucrase	795	87,384.68	5.8	27	22	93.44	21.92	13,076	-0.114	Stable
Phospho-2-dehydro-3-deoxyheptonate aldolase	343	38,718.65	5.53	43	42	77.33	39.84	22,256	-0.455	Stable
ATP synthase F0F1 subunit alpha	501	54,372.84	5.42	28	15	7450	47.87	19,878	-0.094	Unstable
ATP synthase F0F1 subunit beta	689	75,218.11	4.88	45	30	101.02	40.2	22,045	-0.473	Unstable
Glucosyltransferase-SI	1455	162,966.24	5.28	41	36	100.1	33.5	14,900	-0.151	Stable
Glucosyltransferase-I	1476	165,846.81	5.09	68	53	103.68	30.32	25,245	-0.16	Stable
secreted antigen GbpB/SagA	431	44,620.42	9.46	66	64	94.65	27.18	17,880	-0.219	Stable
Formate acetyltransferase	775	87,605.80	5.27	49	38	92.7	21.62	26,860	-0.321	Stable
Fructose-1,6-biphosphate aldolase	293	31,425.80	5.46	32	26	99.96	31.78	13,535	0.007	Stable
Endolytic murein transglycosylase	614	68,543.36	6.61	37	36	102.59	20.19	16,180	-0.226	Stable
L-lactate dehydrogenase	328	35,245.09	5.85	33	31	96.28	25.97	14,440	-0.089	Stable
Ketol-acid reducto isomerase	340	37,285.33	5.53	68	56	92.79	27.1	37,650	-0.277	Stable
Aspartate-semialdehyde dehydrogenase	358	38,903.31	6.21	45	41	89.54	29.85	23,452	-0.365	Stable
Enolase	432	46,857.61	9.32	56	92	87.62	38.9	47,655	-0.555	Stable
UDP-N-acetylglucosamine 1-carboxyvinyl-transferase 1	423	45,616.65	4.95	45	33	92.46	34.55	11,564	-0.51	Stable
ATP-dependent protease ClpE	753	83,733.90	5.59	31	55	100.34	22.87	17,886	-0.195	Stable

Table 4 Listing of FDA approved drugs (Drug names, DrugBank ID) for the 22 screened potential drug targets

S.No.	Drug target	Drug name	DrugBank ID
1	Glucose-6-phosphate isomerase	Vitafof-one Artemimol	DB09130 DB11638
2	4-Hydroxy-tetrahydrodipicolinate reductase	Dipicolinic acid	DB04267
3	Glycogen synthase	Valproate Tideglusib	DB00313 DB12129
4	Glucan-binding protein GbpC	Eraxis Vfend	DB00362 DB00582
5	cell surface antigen SpaP	Gleevec	DB00619
6	Thiol peroxidase	S-oxy-L-cysteine	DB03382
7	Levansucrase	Sucrose	DB02772
8	Phospho-2-dehydro-3-deoxyheptonate aldolase	Phosphoenolpyruvate D-erythrose 4-phosphate	DB01819 DB03937
9	ATP synthase F0F1 subunit alpha	Aurovertin B Piceatannol Quercetin Artemimol	DB07394 DB08399 DB04216 DB11638
10	ATP synthase F0F1 subunit beta	Aurovertin B Piceatannol Quercetin Artemimol	DB07394 DB08399 DB04216 DB11638
11	Glucosyltransferase-SI	Bezlotoxumab	DB13140
12	Glucosyltransferase-I	Bezlotoxumab	DB13140
13	Secreted antigen GbpB/SagA	Inebilizumab Blinatumomab Tisagenlecleucel	DB12530 DB09052 DB13881
14	Formate acetyltransferase	OxamicAcid D-Treitol	DB03940 DB03278
15	Fructose-1,6-biphosphate aldolase	Artemimol	DB11638
16	Endolytic murein transglycosylase	Bicine	DB03709
17	L-lactate dehydrogenase	Oxamic Acid Artemimol	DB03940 DB11638
18	Ketol-acid reductoisomerase	Cocarboxylase	DB01987
19	Aspartate-semialdehyde dehydrogenase	Nicotinamide adenine dinucleotide phosphate Aspartate Semialdehyde	DB03461 DB04498
20	Enolase	Sodium fluoride	DB09325
21	UDP-N-acetylglucosamine 1-carboxyvinyl-transferase 1	Aminomethylcyclohexane 8-Anilino-naphthalene-1-sulfonic acid	DB02435 DB04474
22	ATP-dependent protease ClpE	Bismuth subcitrate potassium	DB09275

of fetal infection of endocarditis [49]. The objective of the present study was to find out new and more effective drug targets against this cariogenic bacteria *S. mutans*, so as to propose the alternative potential therapeutics to combat oral infections. In this study, we employed LC–MS based quantitative proteomics, which is a powerful analytical technique, that is being increasingly utilized for a varied range of biological applications owing to its increasing capabilities for broad proteome coverage, high sensitivity, specificity as well as precision in quantification [50]. LC–MS along with subtractive proteomics approach which has been documented to be a powerful method to identify unique yet uncharacterized sequences as possible therapeutic targets. Besides, the *in*

silico subtractive proteomics is the most efficient, time saving and economical technique, on condition that both proteome of pathogen and host are accessible. This approach involves numerous analyses at multiple stages that generally employ BLAST [15, 51].

The *S. mutans* grown in the presence of bacitracin. Bacitracin is an antibiotic that blocks the biogenesis of cell-wall polysaccharides such as peptidoglycan. The process is facilitated through an efficient carrier for example undecaprenylphosphate (C55–P). Sugar linking to C55–P occurs at the membrane's inner face then its translocation takes place towards periplasm, where glycan moiety is transferred to the growing polymer, thereby it anchors peptidoglycan

units in the membrane. C55P is derived when the precursor undecaprenyl pyrophosphate (C55PP), is dephosphorylated. This dephosphorylation step is obstructed by bacitracin by sequestering C55PP, thereby reducing the pool of C55P available to transport glycan units to the cell wall thereby disturbing cell integrity and causing lysis [18]. When the whole proteome extract of bacteria grown under bacitracin stress subjected to LC–MS, a panel of 321 proteins was identified. At the outset, the duplicity in data set of 321 proteins was removed. For this, the protein panel was subjected to CD-Hit, which clusters those proteins that cross the similarity threshold. In our data, proteins that were found to have 60% sequence or functional similarity with other proteins of the data set were eliminated. The remaining proteins were then subjected to non-homologous analysis, which involves pathogen-specific screening of drug targets against human proteome [52]. In this way cross reactivity as well as binding of therapeutic compounds with the host proteome can be averted. The Non-homology analysis is particularly the foremost step in any *in silico* drug target identification method to avoid any off-target side-effects. Subsequently, those proteins that are necessary for bacterial survival as well as persistence were identified through essentiality analysis by employing Database of Essential Genes (DEG). These essential proteins are regarded as potential targets for identification of novel and innovative chemical scaffolds for the purpose of drug development. These proteins comprise sequences grouped according to their biological functions, that is, (a) information storage and processing (DNA and RNA metabolism), (b) protein processing, folding and secretion, (c) cellular processes (division of cell and transport), (d) energetic and intermediary metabolism (glycolysis, pentose phosphate pathway, lipid metabolism, cofactors as well as nucleotides biosynthesis, production of proton-motive force), (e) and poorly characterized [12, 19].

The results of DEG were subjected to KEGG Automated Annotation Server (KAAS), an online server, to investigate the involvement of proteins in different essential metabolic pathways in KEGG. Metabolic pathways shared among the pathogen and host are categorized as common pathways while unique pathways are those which are present in pathogen only. Human metabolic pathways are not included in the bacterial metabolism [53]. This analysis was performed on non-homologous essential genes of *S. mutans*. At first, metabolic pathway analysis is performed on the non-homologous essential proteins by KAAS to identify the metabolic pathway of the target. Further in this study, comparative pathway method is performed among human metabolic pathways and pathogen pathway to identify “unique pathways”. Non-homologous essential proteins among following unique metabolic pathways can be mapped, and these proteins can be key targets for the treatment of diseases and avoiding all possibilities of side effects [12, 52].

Out of the total sequences of proteins selected by the KAAS server, maximum sequences were found to be involved in the biosynthesis of antibiotics, followed by biosynthesis of secondary metabolites, lysine metabolism, monobactam and peptidoglycan biosynthesis etc. as shown in Fig. 1.

In order to screen out essential proteins having structural similarity with the proteome of Gut microbiota and subsequently to prevent off-target effects, non-homology analysis against gut microflora was performed (Table 2).

Suitability of essential proteins to be considered as potential targets depends on two criteria: (a) essentiality and (b) selectivity. Essential proteins constitute the foundation of organism whereas selectivity depends on expression levels of genes. The efficiency of translating mRNA to protein depends partially on the coding strategy of an mRNA and is reflected in codon usage bias which is often measured by the codon adaptation index. CAI gives the geometric mean of relative fitness values for each synonymous codon which lies between 0 and 1 [15, 54]. CAI of the proteins shortlisted was performed to find out their suitability (Fig. 2).

22 proteins obtained after series of subtractive proteomics methods were considered as therapeutic targets against caries infection. These proteins were further subjected to qualitative characterization to localize them within the cell and to know their physio-chemical properties (Table 3). A significant aspect of possible drug target is its localization. Protein targets should be at appropriate cell's compartment so they can optimally perform their functions. Besides, drug needs to bind with the target to act, therefore the protein targets whereabouts in cell should be known to design appropriate drug compound. Subcellular localization of non-homologous protein sequences was performed by CELLO

Proteome Subtraction Methods

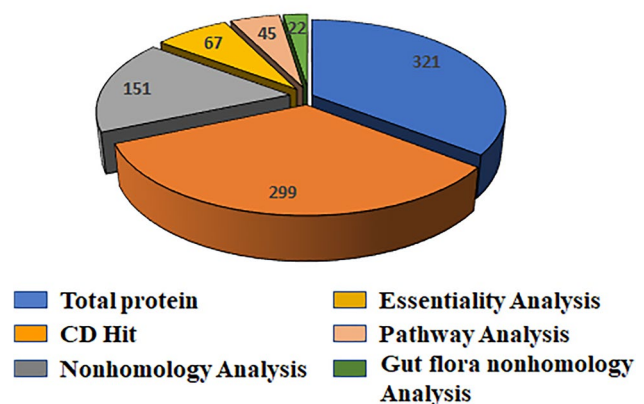


Fig. 2 Protein count for *S. mutans* MTCC 497 strain employing subtractive proteome analysis

2.0 [55, 56]. The sorting of protein sequences was mainly found in cell membrane, extracellular matrix and cytoplasmic domains. Physicochemical characterization which seeks to define physical as well as chemical properties based on amino acid residue composition was computed using ProtParam [57]. Druggability analysis of 22 shortlisted proteins was also performed to predict if the protein can bind to drug-like molecules (Table 4). Identified 22 therapeutic candidates have been illustrated in file S3. (Function as well as the physicochemical properties of 22 shortlisted proteins is listed in Table 1 and 3).

Out of total 22 possible targets identified by this study, sortase family of proteins, which represent highly conserved transpeptidase family have been documented to be critical for bacterial virulence. In *S. mutans* sortases help in anchoring a number of virulence surface proteins such as FruA, GbpC, Pac, WapA and Dex to the cell wall. These proteins mediate bacterial adherence with the tooth surface, that results in the formation of biofilm. Further, sortases are present at the center of a pathway that controls numerous virulence factors. Consequently, if they are targeted, it may result in hindering multiple virulence pathways simultaneously [58]

5 Conclusion

Despite enormous efforts, dental caries is still considered as one of the highly prevalent, though preventable disease around the globe. Efforts for an effective drug and vaccine development to combat infection have been made in many countries for decades but have not gained much success. In this regard, proteomics along with subtractive in silico methods is the most suitable approach to find out the proteins involved in virulence of *S. mutans*. With the help of hierarchical in silico novel approach, this study has presented a list of 22 potential anti-caries drug candidates along with their qualitative characterization. It has been explored that a category of essential non-human homologous, central metabolic proteins fit in the scope of potential drug targets due to two reasons: (1) Either the targets identified don't have any human or gut microbiome counterparts or they appear to use some alternative catalytic mechanisms. Accordingly, an inhibitor can be designed that would not obstruct its human cousin. (2) secondly, targeting energy-making machinery of bacteria circumvents the usual pathogenesis and thereby paving the path to multi-faceted strategies to combat the infection of dental caries. Shortlisted proteins were qualitatively characterized to find if they can gain access into pipelines of drugs as well as vaccine development. It is being anticipated that results obtained in the present study can provide platform to design vaccine or targets against caries infection.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10930-021-10038-1>.

Acknowledgements Sahar Zaidi is thankful to Department of biotechnology for the support of fellowship. PS is thankful to Jawaharlal Nehru University, New Delhi for providing necessary support for Bioinformatics analysis. Authors also acknowledge the SCIE X Pvt. Ltd. Gurgaon, India for proteomics facility support.

Funding The study was supported by the internal funds of Biotechnology Unit, AMU Aligarh.

Data Availability Not applicable.

Code Availability Not applicable.

Declarations

Conflict of interest No potential conflict of interest was reported by the authors.

References

- Lemos JA, Palmer SR, Zeng L, Wen ZT, Kajfasz JK, Freires IA (2019) The biology of *Streptococcus mutans*. Gram-Positive Pathogens. <https://doi.org/10.1128/9781683670131.ch27>
- Lemos JA, Quivey RG Jr, Koo H, Abranches J (2013) *Streptococcus mutans*: a new Gram-positive paradigm? Microbiology 159:436. <https://doi.org/10.1099/mic.0.066134-0>
- Lemos JA, Burne RA (2008) A model of efficiency: stress tolerance by *Streptococcus mutans*. Microbiology 154:3247. <https://doi.org/10.1099/mic.0.2008/023770-0>
- Ajdić D, McShan WM, McLaughlin RE, Savić G, Chang J, Carson MB et al (2002) Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc Natl Acad Sci USA 99:14434–14439. <https://doi.org/10.1073/pnas.172501299>
- Nakano K, Ooshima T (2009) Serotype classification of *Streptococcus mutans* and its detection outside the oral cavity. Future Microbiol 4:891–902. <https://doi.org/10.2217/fmb.09.64>
- Garcia-Seco D, Chiapello M, Bracale M, Pesce C, Bagnaresi P, Dubois E et al (2017) Transcriptome and proteome analysis reveal new insight into proximal and distal responses of wheat to foliar infection by *Xanthomonas translucens*. Sci Rep 31:1–3. <https://doi.org/10.1038/s41598-017-10568-8>
- Guo MS, Gross CA (2014) Stress-induced remodeling of the bacterial proteome. Curr Biol 19:R424–R434. <https://doi.org/10.1016/j.cub.2014.03.023>
- Ming LJ, Epperson JD (2002) Metal binding and structure–activity relationship of the metalloantibiotic peptide bacitracin. J Inorg Biochem 91:46–58. [https://doi.org/10.1016/S0162-0134\(02\)00464-6](https://doi.org/10.1016/S0162-0134(02)00464-6)
- Pollock TJ, Thorne L, Yamazaki M, Mikolajczak MJ, Armentrout RW (1994) Mechanism of bacitracin resistance in gram-negative bacteria that synthesize exopolysaccharides. J Bacteriol 176:6229–6237. <https://doi.org/10.1128/jb.176.20.6229-6237.1994>
- Zhu M, Ajdić D, Liu Y, Lynch D, Merritt J, Banas JA (2009) Role of the *Streptococcus mutans* sirvA gene in GbpC-independent, dextran-dependent aggregation and biofilm formation. Appl Environ Microbiol 75:7037–7043. <https://doi.org/10.1128/aem.01015-09>
- Sharma D, Garg A, Kumar M, Khan AU (2019) Proteome profiling of carbapenem-resistant *K. pneumoniae* clinical isolate

- (NDM-4): exploring the mechanism of resistance and potential drug targets. *J Proteom* 200:102–110. <https://doi.org/10.1016/j.jprot.2019.04.003>
12. Solanki V, Tiwari V (2018) Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against *Acinetobacter baumannii*. *Sci Rep* 8:9044. <https://doi.org/10.1038/s41598-018-26689-7>
 13. Satyam R, Bhardwaj T, Jha NK, Jha SK, Nand P (2020) Toward a chimeric vaccine against multiple isolates of Mycobacteroides—an integrative approach. *Life Sci*. <https://doi.org/10.1016/j.lfs.2020.117541>
 14. Shanmugham B, Pan A (2013) Identification and characterization of potential therapeutic candidates in emerging human pathogen *Mycobacterium abscessus*: a novel hierarchical *in silico* approach. *PLoS ONE* 8:e59126. <https://doi.org/10.1371/journal.pone.0059126>
 15. Bhardwaj T, Haque S, Somvanshi P (2019) Comparative assessment of the therapeutic drug targets of *C. botulinum* ATCC 3502 and *C. difficile* str. 630 using *in silico* subtractive proteomics approach. *J Cell Biochem* 120:16160–16184. <https://doi.org/10.1002/jcb.28897>
 16. Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, Lu CH (2014) CELLO2GO: a web server for protein subCELLular-Localization prediction with functional gene ontology annotation. *PLoS ONE* 9:e99368. <https://doi.org/10.1371/journal.pone.0099368>
 17. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A et al (2011) DrugBank 3.0: a comprehensive resource for ‘omics’ research on drugs. *Nucleic Acids Res* 39:D1035–D1041. <https://doi.org/10.1093/nar/gkq1126>
 18. Zaidi S, Singh SL, Khan AU (2020) Exploring antibiofilm potential of bacitracin against streptococcus mutans. *Microb Pathog* 1(149):104279. <https://doi.org/10.1016/j.micpath.2020.104279>
 19. Chen X, Ji ZL, Chen YZ (2002) TTD: therapeutic target database. *Nucleic Acids Res* 30:412–415. <https://doi.org/10.1093/nar/30.1.412>
 20. Gil R, Silva FJ, Peretó J, Moya A (2004) Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev* 68:518–537. <https://doi.org/10.1128/MMBR.68.3.518-537.2004>
 21. Escano J, Stauffer B, Brennan J, Bullock M, Smith L (2015) Biosynthesis and transport of the lantibiotic mutacin 1140 produced by *Streptococcus mutans*. *J Bacteriol* 197:1173–1184. <https://doi.org/10.1128/JB.02531-14>
 22. Chen P, Qi F, Novak J, Caufield PW (1999) The specific genes for lantibiotic mutacin II biosynthesis in *Streptococcus mutans* T8 are clustered and can be transferred en bloc. *Appl Environ Microbiol* 65:1356–1360. <https://doi.org/10.1128/AEM.65.3.1356-1360.1999>
 23. Liu L, Hao T, Xie Z, Horsman GP, Chen Y (2016) Genome mining unveils widespread natural product biosynthetic capacity in human oral microbe *Streptococcus mutans*. *Sci Rep* 6:37479. <https://doi.org/10.1038/srep37479>
 24. Cardineau GA, Curtiss R (1987) Nucleotide sequence of the *asd* gene of *Streptococcus mutans*. Identification of the promoter region and evidence for attenuator-like sequences preceding the structural gene. *J Biol Chem* 262:3344–3353
 25. Tang X, Kudo Y, Baker JL, LaBonte S, Jordan PA, McKinnie SM, Edlund A (2020) Cariogenic *Streptococcus mutans* produces tetramic acid strain-specific antibiotics that impair commensal colonization. *ACS Infect Dis* 6:563–571. <https://doi.org/10.1021/acscinfdis.9b00365>
 26. Park SN, Ahn SJ, Kook JK (2015) Oleanolic acid and ursolic acid inhibit peptidoglycan biosynthesis in *Streptococcus mutans* UA159. *Braz J Microbiol* 46:613–617. <https://doi.org/10.1590/S1517-838246246220130209>
 27. Chatfield CH, Koo H Jr, Quivey RG (2005) The putative autolysin regulator LytR in *Streptococcus mutans* plays a role in cell division and is growth-phase regulated. *Microbiology* 151:625–631. <https://doi.org/10.1099/mic.0.27604-0>
 28. Hu X, Wang Y, Gao L, Jiang W, Lin W, Niu C, Huang Z (2018) The impairment of methyl metabolism from luxS mutation of *Streptococcus mutans*. *Front Microbiol* 9:404. <https://doi.org/10.3389/fmicb.2018.00404>
 29. Klein MI, Duarte S, Xiao J, Mitra S, Foster TH, Koo H (2009) Structural and molecular basis of the role of starch and sucrose in *Streptococcus mutans* biofilm development. *Appl Environ Microbiol* 75:837–841. <https://doi.org/10.1128/aem.01299-08>
 30. Alves LA, Harth-Chu EN, Palma TH, Stipp RN, Mariano FS, Hoffling JF et al (2017) The two-component system VicRK regulates functions associated with *Streptococcus mutans* resistance to complement immunity. *Mol Microbiol* 32:419–431. <https://doi.org/10.1111/omi.12183>
 31. Li YH, Lau PC, Tang N, Svensäter G, Ellen RP, Cvitkovitch DG (2002) Novel two-component regulatory system involved in biofilm formation and acid resistance in *Streptococcus mutans*. *J Bacteriol* 184:6333–6342. <https://doi.org/10.1128/jb.184.22.6333-6342.2002>
 32. Carlsson J (1970) Nutritional requirements of *Streptococcus mutans*. *Caries Res* 4:305–320. <https://doi.org/10.1159/000259653>
 33. Leonardi R, Jackowski S (2007) Biosynthesis of pantothenic acid and coenzyme A. *EcoSalPlus*. <https://doi.org/10.1128/ecosa1plus.3.6.3.4>
 34. O’kane DJ, Gunsalus IC (1948) Pyruvic acid metabolism: a factor required for oxidation by *Streptococcus faecalis*. *J Bacteriol* 56:499. <https://doi.org/10.1128/jb.56.4.499-506.1948>
 35. Sparnins VL, Chapman PJ (1976) Catabolism of L-tyrosine by the homoprotocatechuate pathway in gram-positive bacteria. *J Bacteriol* 127:362–366. <https://doi.org/10.1128/aac.25.4.527>
 36. Holden JT (1956) Inhibition of phenylalanine and tyrosine synthesis in *Streptococcus faecalis* and *Lactobacillus arabinosus* by α -keto acids. *Arch Biochem Biophys* 61:128–136. [https://doi.org/10.1016/0003-9861\(56\)90323-X](https://doi.org/10.1016/0003-9861(56)90323-X)
 37. Shlaes DM, Marino JOHN, Jacobs MR (1984) Infection caused by vancomycin-resistant *Streptococcus sanguis* II. *Antimicrob Agents Chemother* 25:527–528. <https://doi.org/10.1128/AAC.25.4.527>
 38. LaRock CN, Nizet V (2015) Cationic antimicrobial peptide resistance mechanisms of streptococcal pathogens. *Biochim Biophys Acta* 1848:3047–3054. <https://doi.org/10.1016/j.bbamem.2015.02.010>
 39. Xu P, Ge X, Chen L, Wang X, Dou Y, Xu JZ (2011) Genome-wide essential gene identification in *Streptococcus sanguinis*. *Sci Rep* 1:125. <https://doi.org/10.1038/srep00125>
 40. Klein MI, DeBaz L, Agidi S, Lee H, Xie G, Lin AHM (2010) Dynamics of *Streptococcus mutans* transcriptome in response to starch and sucrose during biofilm development. *PLoS ONE* 5:e13478. <https://doi.org/10.1371/journal.pone.0013478>
 41. Lemos JA, Nascimento MM, Lin VK, Abranches J, Burne RA (2008) Global regulation by (p) ppGpp and CodY in *Streptococcus mutans*. *J Bacteriol* 190:5291–5299. <https://doi.org/10.1128/jb.00288-08>
 42. Kovatcheva-Datchary P, Zoetendal EG, Venema K, de Vos WM, Smidt H (2009) Tools for the tract: understanding the functionality of the gastrointestinal tract. *Therap Adv Gastroenterol* 2:S9–S22. <https://doi.org/10.1177/1756283X09337646>
 43. Hooper LV, Bry L, Falk PG, Gordon JI (1998) Host–microbial symbiosis in the mammalian intestine: exploring an internal ecosystem. *BioEssays* 20:336–343. [https://doi.org/10.1002/\(SICI\)1521-1878\(199804\)20:4%3c336](https://doi.org/10.1002/(SICI)1521-1878(199804)20:4%3c336)

44. Neish AS (2009) Microbes in gastrointestinal health and disease. *Gastroenterology* 136:65–80. <https://doi.org/10.1053/j.gastro.2008.10.080>
45. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI (2007) The human microbiome project. *Nature* 449:804–810
46. Vetrivel U, Subramanian G, Dorairaj S (2011) A novel in silico approach to identify potential therapeutic targets in human bacterial pathogens. *HUGO J* 5:25–34. <https://doi.org/10.1007/s11568-011-9152-7>
47. Bhardwaj T, Somvanshi P (2018) A computational approach using mathematical modeling to assess the peptidoglycan biosynthesis of *Clostridium botulinum* ATCC 3502 for potential drug targets. *Gene Rep* 12:179–186. <https://doi.org/10.1016/j.genrep.2018.07.002>
48. Bhardwaj T, Haque S, Somvanshi P (2018) Identification of potential molecular mimics involved in *Clostridium botulinum* ATCC 3502 pathogenesis. *Microb Pathog* 121:238–244. <https://doi.org/10.1016/j.micpath.2018.05.017>
49. Nomura R, Matayoshi S, Otsugu M, Kitamura T, Teramoto N, Nakano K (2020) Contribution of severe dental caries induced by *Streptococcus mutans* to the pathogenicity of infective endocarditis. *Infect Immun* 88:e00897–e919. <https://doi.org/10.1128/IAI.00897-19>
50. Xie F, Liu T, Qian WJ, Petyuk VA, Smith RD (2011) Liquid chromatography-mass spectrometry-based quantitative proteomics. *J Biol Chem* 286:25443–25449. <https://doi.org/10.1074/jbc.R110.199703>
51. Solanki V, Tiwari V (2018) Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against *Acinetobacter baumannii*. *Sci Rep* 13:1–9. <https://doi.org/10.1038/s41598-018-26689-7>
52. Satyam R, Bhardwaj T, Goel S, Jha NK, Jha SK, Nand P et al (2021) miRNAs in SARS-CoV 2: a spoke in the wheel of pathogenesis. *Curr Pharm Des* 27:1628–1641. <https://doi.org/10.2174/1381612826999201001200529>
53. Shanmugham B, Pan A (2013) Identification and characterization of potential therapeutic candidates in emerging human pathogen *Mycobacterium abscessus*: a novel hierarchical in silico approach. *PLoS ONE* 19:e59126. <https://doi.org/10.1371/journal.pone.0059126>
54. Vetrivel U, Subramanian G, Dorairaj S (2011) A novel in silico approach to identify potential therapeutic targets in human bacterial pathogens. *HUGO J* 1:25–34. <https://doi.org/10.1007/s11568-011-9152-7>
55. Singh S, Malik BK, Sharma DK (2006) Molecular drug targets and structure based drug design: a holistic approach. *Bioinforma-tion* 1(8):314
56. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R et al (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26:1608–1615. <https://doi.org/10.1093/bioinformatics/btq249>
57. Gupta MK, Vadde R (2019) In silico identification of natural product inhibitors for γ -secretase activating protein, a therapeutic target for Alzheimer's disease. *J Cell Biochem* 120:10323–10336. <https://doi.org/10.1002/jcb.28316> (**Epub 2018 Dec 19**)
58. Kato C, Nakano Y, Lis M, Kuramitsu HK (1992) Molecular genetic analysis of the catalytic site of *Streptococcus mutans* glucosyltransferases. *Biochem Biophys Res Commun* 15:1184–1118. [https://doi.org/10.1016/0006-291x\(92\)92329-v](https://doi.org/10.1016/0006-291x(92)92329-v)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.