https://doi.org/10.3329/cbmj.v13i2.75317

In silico Identification and Characterization of Novel Drug Targets in *Treponema denticola (strain ATCC 35405 / DSM 14222 / CIP 103919 / JCM 8153 / KCTC 15104)*: A Subtractive Genomics Approach

* Chowdhury ARK[1], Billah SM[2], Meem MB[3], Mahi MH[4], Zaman T[5], Onia SK[6], Ishita AF[7], Hossain MM[8]

Abstract

Treponema denticola is a gram-negative, highly drug resistant bacterium found in primary dentition infections and around teeth. It causes inflammation and tissue homeostasis, linked to periodontal diseases like early-onset periodontitis, necrotizing ulcerative gingivitis, and acute pericoronitis. Research is needed to develop powerful, cost-effective, secure, and environmentally friendly antibiotics and techniques to control and manage infections caused by this bacterium. This study exploits the sophisticated in silico subtractive genomics approach to investigate potential therapeutic targets that are exclusive to the pathogen *T. denticola.* The full sequencing data of the *Treponema denticola (strain ATCC 35405 / DSM 14222 / CIP 103919 / JCM 8153 / KCTC 15104)* proteome has facilitated the computational analysis of its genome. Our analysis found 126 proteins of the pathogen that had no resemblance to the human genome. The Database of Essential Genes (DEG) identified 12 bacterial proteins which were indispensable to the pathogen, while the KEGG Automated Annotation Server (KAAS) found in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways database analysis of the non-homologous proteins revealed 11 *T. denticola* enzymes that can be targeted for drug development. Moreover, based on sub-cellular localization prediction, all selected proteins were cytoplasmic proteins and the no outer membrane proteins were present among the non-homologous proteins, and virulent protein predictions revealed no virulent proteins among the selected proteins. Therefore, eleven proteins were selected based on their involvement in unique metabolic pathways inside the pathogen that were not present in the host. The study further revealed the protein-protein interactions of these eleven essential proteins, and successfully anticipated, assessed, and verified the three-dimensional structures of these proteins. Undertaking a screening process to detect functional inhibitors for these newly revealed targets may lead to the discovery of novel therapeutic drugs that can effectively combat *Treponema denticola.*

CBMJ 2024 July: vol. 13 no. 02 P: 251-264

Keywords: *Treponema denticola*, subtractive genomics, metabolic pathways, novel drug, immunoinformatics

Introduction

To determine potential targets for drugs and vaccines, this current investigation makes use of the subtractive genomics approach, the database of essential genes (DEG), differential pathway analysis, sub-cellular localization prediction, viral protein prediction, protein-protein interactions, and functional annotations to analyze the entire proteome of *Treponema denticola (strain ATCC 35405 / DSM 14222 / CIP 103919 / JCM 8153 / KCTC 15104).*

Treponema denticola is a gram-negative, obligate anaerobic, motile, and highly proteolytic spirochete bacterium with four periplasmic flagella, which facilitate its motility even in a viscous environment.[1,2] It is a highly drug resistant bacteria often seen in primary dentition infections and in pockets around teeth which is intimately linked to periodontal diseases such as early-onset periodontitis, necrotizing ulcerative gingivitis, and acute pericoronitis.[3] Periodontal diseases in the oral cavity are caused by the

1. *Ashiqur Rahman Khan Chowdhury, Graduate Student, Biotechnology and Bioinformatics, Department of Life Sciences, School of Environment and Life Sciences, Independent University Bangladesh, Bashundhara, Dhaka, 1229, Bangladesh.
2. Sajid Mehran Billah, Undergraduate Student, Biotechnology and Genetic Engineering, Department of Mathematics & Natural Sciences, BRAC University, 66 Mohakhali, Dhaka 1212, Bangladesh.
3. Maksuda Begum Meem, Graduate Student, Biotechnology and Bioinformatics, Department of Life Sciences, School of Environment and Life Sciences, Independent University Bangladesh, Bashundhara, Dhaka, 1229, Bangladesh.
4. Mahiea Hossain Mahi, Graduate Student, Biotechnology and Bioinformatics, Department of Life Sciences, School of Environment and Life Sciences, Independent University Bangladesh, Bashundhara, Dhaka, 1229, Bangladesh.
5. Tonima Zaman, Graduate Student, Biotechnology and Bioinformatics, Department of Life Sciences, School of Environment and Life Sciences, Independent University Bangladesh, Bashundhara, Dhaka, 1229, Bangladesh.
6. Sumaiya Khanam Onia, Graduate Student, Biotechnology and Bioinformatics, Department of Life Sciences, School of Environment and Life Sciences, Independent University Bangladesh, Bashundhara, Dhaka, 1229, Bangladesh.
7. Afsana Ferdousi Ishita, Graduate Student, Biotechnology and Bioinformatics, Department of Life Sciences, School of Environment and Life Sciences, Independent University Bangladesh, Bashundhara, Dhaka, 1229, Bangladesh.
8. Dr. M. Mahboob Hossain, Professor, Microbiology, Department of Mathematics & Natural Sciences, BRAC University, 66 Mohakhali, Dhaka 1212, Bangladesh.

Address of Correspondence:

Email: khan.chowdhury21@gmail.com

growth and buildup of bacteria resulting from inadequate oral hygiene.[4] Both basic studies and clinical data indicate that the abundance of *T. denticola,* along with other proteolytic gram-negative bacteria, in significant quantities inside periodontal pockets, likely has a substantial impact on the development of periodontal disease.[5] However, the precise impact of the pathogen in causing disease still requires more investigation and evidence.[6,7] Periodontal diseases affect about 5 to 20% of the worldwide population and are amongst the most common chronic infections.[8,9]

The vast amount of data from over 1000 bacterial genomes sequenced so far, primarily from harmful bacteria, can be used to discover potential drug targets and understand virulent factors in these pathogens.[10] Bioinformatics-based techniques offer a cost-effective alternative to standard laboratory-based investigations, reducing costs and time spent on finding candidate molecules as drug targets. A subtractive genomics approach can be applied to discover proteins that are exclusively present in the pathogen by deducing the homologous proteins from the whole proteome of both the host and the pathogen. This efficient technique has been effectively used to uncover new therapeutic targets in pathogens such as *Mycobacterium tuberculosis F11*[11], *Edwardsiella tarda*[12], and others pathogenic bacteria.[13–15] In this study, we utilized a combination of advanced tools and databases such as Universal Protein Resource Database (UniProt), Basic Local Alignment Search Tool for Proteins (BLASTP), Database of Essential Genes (DEG), KEGG Automatic Annotation Server (KAAS) of Kyoto Encyclopedia of Genes and Genomes (KEGG),

pSORTb, CELLO, VirulentPred 2.0, STRING, and SWISS-MODEL. Our aim was to identify, characterize, analyze, predict, and validate the structures of the essential proteins of the human pathogen *Treponema denticola* that could have the potential to serve as promising candidates for vaccines and drugs that specifically target this pathogen.

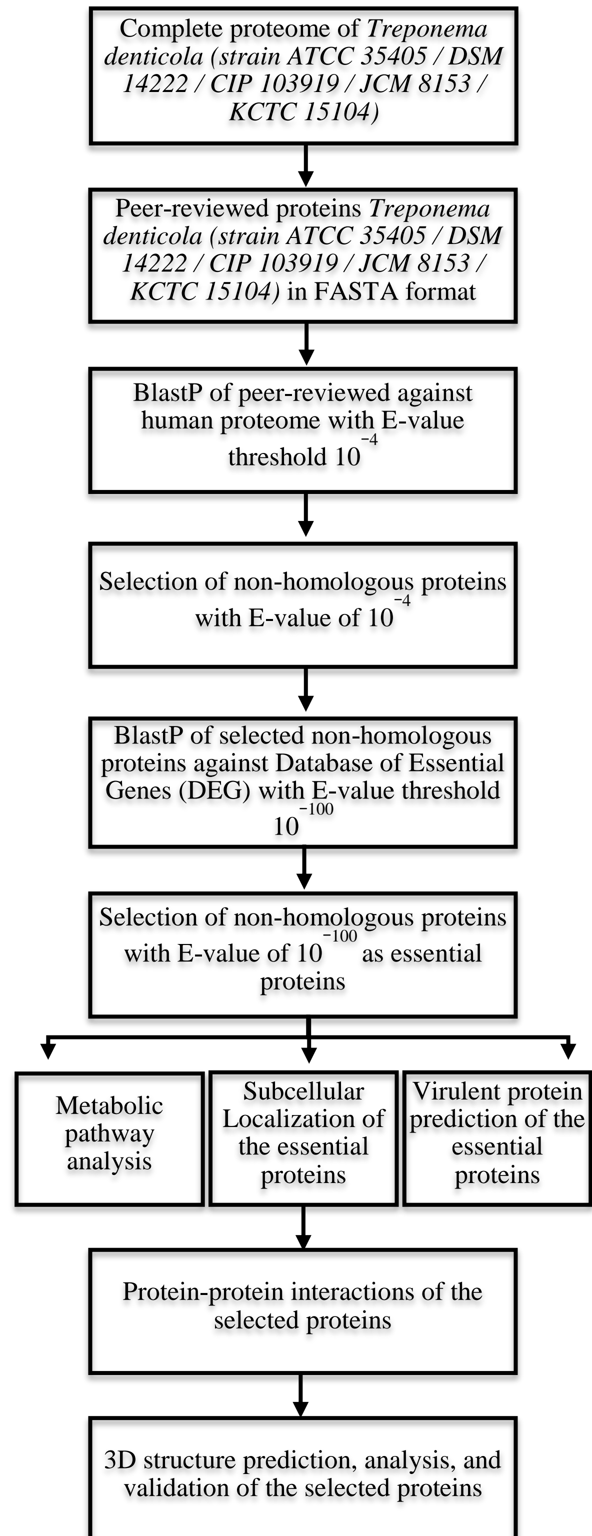
Remarkable capabilities of computational databases were demonstrated and the use of subtractive genomics approach in identifying potential targets for antibacterial drugs, particularly in the completely sequenced *Treponema denticola* genome. Additionally, our approach effectively identified several protein targets, paving the way for the development of novel drug targets, showcasing the potential of computational databases in drug discovery.

Methods

The flow chart describing the detailed methodology for identifying pathogen specific essential proteins in *T. denticola* for identifying potential drug targets is shown in Fig. 1.

Retrieval of protein sequence: The whole proteome of *Treponema denticola* was retrieved from the UniProt (Universal Protein Resource) (https://www.uniprot.org/) database in FASTA format. UniProt contains over 227 million sequence records in UniProtKB.[16] The UniProtKB Proteomes portal (https://www.uniprot.org/proteomes/) provides access to more than 451,000 proteomes, which are sets of protein sequences originating from completely sequenced viral, bacterial, archaeal and eukaryotic genomes.[16,17]

Fig. 1: Flowchart for identification of potential drug targets in *Treponema denticola (strain ATCC 35405 / DSM 14222 / CIP 103919 / JCM 8153 / KCTC 15104).*



Selection of non-homologous proteins to host: The set of proteins obtained after excluding the unreviewed protein sequences were subject to BlastP tool of NCBI database (National Centre for Biotechnology Information) (https://blast.ncbi.nlm .nih.gov/Blast.cgi) against Homo sapiens proteome with the expectation value (E-value) cut-off of 10^{-4}. The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences.[18] The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.[18] The resultant dataset was with no homologs in *Homo sapiens.*

Screening of Essential Protein Sequences: The Database of Essential Genes or DEG (https://tubic.org/deg/public/index.php/index) was used to identify the essential proteins involved in *Treponema denticola (strain ATCC 35405 / DSM 14222 / CIP 103919 / JCM 8153 / KCTC 15104).* BlastP analysis was performed for the non-homologous protein sequences of the pathogen against DEG with e-value cut-off score of 10^{-100} to screen out genes that appeared to represent essential genes. The resulted protein sequences with significant homogeneity with the DEG database are delineated as sequences of proteins that are essential for the bacterial survival.[19–22]

Analysis of metabolic pathways: Metabolic pathway analysis of the essential proteins of Treponema denticola was done by KAAS (KEGG Automatic Annotation Server) (https :// www. genome.jp/tools/kaas/) at KEGG for the identification of potential targets. KAAS provides functional annotation of genes by BLAST comparisons against the manually curated KEGG GENES database. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG pathways.[23]

Prediction of Sub-Cellular Localization: Protein sub-cellular localization prediction involves the computational prediction of where a protein resides in a cell. Prediction of protein sub-cellular localization is an important component as it predicts the protein function and genome annotation, and it can aid in the identification of targets. Cell membrane proteins are often used as targets for vaccines, while cytoplasmic proteins are often targeted for therapeutic uses. The essential proteins involved in the unique metabolic pathways of *Treponema denticola* were therefore analyzed for outer membrane proteins. Initially, the program PSORTb v.3.03 (https://www.psort.org/psortb/) was used for sub-cellular localization prediction.[24,25] Afterwards, the essential proteins were crosschecked using CELLO v2.5 subcellular localization predictor (http://cello.life.nctu.edu.tw/).[26,27]

Identification of Virulent Proteins: VirulentPred2.0 algorithm (https://bioinfo.icgeb. res.in/virulent2/ predict.html) was employed for the prediction of virulent protein sequences.[28] This algorithm is based on bilayer cascade support vector machine (SVM). Different prediction approaches were available for prediction of virulent protein sequences, i.e. amino acid composition-based, dipeptide composition based, similarity search based, higher order dipeptide composition based, PSSM-based and cascaded SVM module.[28]

Protein Interaction Network Analysis: To identify whether these proteins can be a hub protein and validate their functional interactions, the PPIs network of pathogen proteins as potential drug targets were generated through the STRING database (https://string-db.org/).[29] The STRING is a database of experimentally known and predicted PPIs, including direct (physical) and indirect (functional) network. Node degrees and clustering coefficients are used to classify these PPIs as hub proteins. STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 59,309,604 proteins from 12,535 organisms.[30] This tool was used to study the protein-protein interactions of the selected essential proteins.

Prediction and Validation of 3D Structures: The 3D structure prediction of the selected outer membrane proteins, was done by SWISS-MODEL (https://swissmodel.expasy.org/) server. The SWISS-MODEL is a homology-based structure prediction tool used to predict 3D structures of protein from its amino acid sequences.[31] The amino acid sequences of the selected proteins were inserted into the online web server SWISS-MODEL, and the 3D structure of the selected proteins were modelled from a predicted set of templates. The tool then built a structure based on the homology on the alignment between the target and template.[31] The template with the highest identity percentage and GMQE score was selected for the building of the 3D model.[32–34] Afterwards, the predicted 3D model was checked for structure validation within the SWISS-MODEL server.[35,36]

Results

The primary objective of this study was to identify possible drug targets that can effectively combat *T. denticola.* The drug targets that are being considered must satisfy to the drug efficacy criteria, which require that the targets be dissimilar to the human host, essential for the pathogen, and play an important part in the primary bacterial metabolic system. The overall results of the subtractive genomic analysis approach are shown in Table-I.

Table-I: Subtractive genomic and metabolic pathway analysis result for *Treponema denticola (strain ATCC 35405 / DSM 14222 / CIP 103919 / JCM 8153 / KCTC 15104)*

Sl. No.	Steps	*Treponema denticola*
1	Total number of proteins	2753
2	Peer-reviewed proteins	259
3	No. of sequences non-homologous to *H. sapiens* applying BLASTp (e-value 10^{-4})	126
4	Screening of essential protein sequences in DEG (e-value 10^{-100})	12
5	Essential non-homologous proteins involved in unique metabolic pathways (KEGG)	11
6	Pathways unique to the pathogen (KEGG)	22
7	Common pathways with the host (KEGG)	2
8	Prediction of subcellular localization of outer membrane proteins (CELLO v2.5)	No outer membrane proteins
9	Prediction of subcellular localization of outer membrane proteins (pSORTb v3.03)	No outer membrane proteins
10	Prediction of Virulent Proteins (VirulentPred2.0)	No virulent proteins

Retrieval of protein sequence: The whole proteome of Treponema denticola was retrieved from the UniProt (Universal Protein Resource) (https://www.uniprot.org/) database in FASTA format. The reference proteome Treponema denticola (strain ATCC 35405 / DSM 14222 / CIP 103919 / JCM 8153 / KCTC 15104) has 2753

proteins in total, among which 259 were peer-reviewed (Swiss-Prot) proteins. The peer-reviewed proteins were selected for the next step and the rest were discarded.

Selection of non-homologous proteins to host: To identify non-homologous sequences of all the selected 259 peer-reviewed proteins, the proteins were screened using BLASTp of the NCBI database against Homo sapiens, with a threshold e-value of 10^{-4}. All 259 selected sequences of *Treponema denticola* were screened against the host organism, Homo sapiens. Scoring parameters of the BLASTp program were set as the default values and no filters or masking were applied. A total of 126 proteins out of 259 given protein sequences resulted in being non-homologous proteins, and the rest having similarities with the host organism. The 126 non-homologous proteins were selected, as they did not have any similarities with the host proteins, and the remaining homologous proteins were discarded for further analysis.

Screening of Essential Protein Sequences: The non-homologous proteins comprising essential genes responsible for the biological activities of the bacteria are considered to be potential druggable targets. With an e-value cutoff of 10^{-100}, and a minimum percentage identity of >25%, the selected non-homologous proteins were then subjected to identify the essential proteins to the bacteria using BLASTp against the Database of Essential Genes (DEG) (https://tubic.org/deg/public/index.php/index). Out of 126 non-homologous protein sequences, 12 proteins were identified and selected as essential to *T. denticola*.

Analysis of metabolic pathways: Metabolic pathway analysis of the non-homologous essential proteins of *T. denticola* was done by KEGG Automatic Annotation Server (KAAS). Using the three letter KEGG organism codes “hsa” and “tde” for human and *T. denticola* respectively, all of the metabolic pathways present in the host (*H. sapiens*) and the pathogen were collected separately. A comparison was made manually to identify the metabolic pathways only present in the pathogen using the KEGG pathway database. Metabolic pathway analysis of these 12 non-homologous essential proteins showed the involvement of these proteins in 24 pathways of the pathogen, and 3 pathways of the host. However, there were 2 common pathways that were found between the pathogen and host. As a result, 22 metabolic pathways were unique only to pathogen. These 22 unique pathways involved the selected essential, non-homologous, and non-redundant proteins of *T. denticola*. The result from the KAAS server at KEGG revealed that all essential proteins had assigned KEGG Orthology (KO) identifiers. The protein sequences related to these pathways, as shown in Table-II, are the most potent targets for drug development because there are no rival pathways in humans so there is no chance of bad effects.

Prediction of Sub-Cellular Localization:
Protein localization is important to understand throughout the drug development process because it influences the design of novel drugs and vaccines, and the best method that is used for subcellular localization prediction is pSORTb24. Therefore, for the initial prediction of subcellular localization of the 12 shortlisted

essential proteins, pSORTb v3.0.3 (https://www.psort.org/psortb/) server was utilized. Among these essential proteins, all 12 proteins were found to be cytoplasmic proteins and no outer membrane proteins were predicted. Afterwards, the selected essential proteins were cross-checked using the CELLO v2.5 server. The server also predicted all 12 proteins as cytoplasmic proteins and no proteins were predicted in the outer membrane region.

Table-II: Unique metabolic pathways *Treponema denticola (strain ATCC 35405 / DSM 14222 / CIP 103919 / JCM 8153 / KCTC 15104)*

Sl. No	KO ID	Unique Pathways	Proteins
1	00010	Glycolysis / Gluconeogenesis	1
2	00051	Fructose and mannose metabolism	1
3	00500	Starch and sucrose metabolism	1
4	00520	Amino sugar and nucleotide sugar metabolism	1
5	00620	Pyruvate metabolism	2
6	00640	Propanoate metabolism	2
7	00650	Butanoate metabolism	1
8	00710	Carbon fixation by Calvin cycle	1
9	00720	Other carbon fixation pathways	2
10	00680	Methane metabolism	2
11	00061	Fatty acid biosynthesis	2
12	00430	Taurine and hypotaurine metabolism	1
13	00550	Peptidoglycan biosynthesis	1
14	00750	Vitamin B6 metabolism	1
15	00900	Terpenoid backbone biosynthesis	1
16	03018	RNA degradation	1
17	03420	Nucleotide excision repair	1
18	03440	Homologous recombination	1
19	02020	Two-component system	1
20	04066	HIF-1 signalling pathway	1
21	04112	Cell cycle - Caulobacter	1
22	02026	Biofilm formation - *Escherichia coli*	1

Identification of Virulent Proteins: Identification of virulent/non-virulent properties of these essential non-homologous proteins was then carried out through VirulentPred2.0 having leading to the hypothesis that these essential proteins could have significant role in the normal functionality of the pathogen inside the host. However, all 12 selected essential proteins of *Treponema denticola* were predicted to be non-virulent.

Protein Interaction Network Analysis: Among the 12 essential non-homologous proteins identified, one protein, 1-deoxy-D-xylulose-5-phosphate synthase (Q73LF4), was excluded from further analysis due to its involvement in both the host and pathogen pathways. The remaining eleven proteins were chosen based on their involvement in vital and unique metabolic pathways of the pathogen, as there were no outer membrane proteins or virulent proteins predicted from previous steps. The PPI and functional annotation of the selected proteins were determined using the STRING server. The results showed different nodes and edges of each protein and demonstrated that the prioritized target proteins may act as hub protein inter-acting with more than ten proteins. Therefore, targeting such proteins can affect the activity of all interacting proteins. Among the 11 essential proteins, the STRING analysis revealed that the cytoplasmic protein Trans-2-enoyl-CoA reductase (NADH) interacted with proteins involved in fatty acid synthesis pathways. It resulted in 4 PPI networks (Fig. 2A) represented as fabV in red node with its neighbouring proteins, whereas the cytoplasmic protein RecA, interacted with various proteins involved in cell division pathways. It resulted in 27 PPI networks (Fig. 2B) represented

as recA in red node with its neighbouring proteins such as DNA Polymerases, helicases, topoisomerases, etc. Another protein, Glycogen synthase, comprised of 491 amino acids, interacted with proteins involved in glycogen biosynthesis pathways and resulted in 7 PPI networks (Fig. 2C) represented as glgA in red node with its neighbouring proteins. Moreover, Acetate kinase, interacted with proteins involved in pyruvate metabolism. It resulted in 6 PPI networks (Fig. 2D) represented as ackA in red node with its neighbouring proteins. To add, UDP-N-acetylglucosamine 1-carboxyvinyl transferase, comprised of 426 amino acids, interacted with proteins involved in cell wall formation. It resulted in 9 PPI networks (Fig. 2E) represented as murA in red node with its neighbouring proteins. Furthermore, Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha is a component of the acetyl coenzyme A carboxylase (ACC) complex. It resulted in 12 PPI networks (Fig. 2F) represented as accA in red node with its neighbouring proteins. Significantly, Pyridoxal 5'-phosphate synthase subunit PdxS, interacted with proteins involved in vitamin B6 biosynthesis pathways and resulted in 8 PPI networks (Fig. 2G) represented as pdxS in red node with its neighbouring proteins. The Chromosomal replication initiator protein DnaA interacted with proteins involved in cell division pathways. It resulted in 8 PPI networks (Fig. 2H) represented as dnaA in red node with its neighbouring proteins such as DNA polymerase III, DNA gyrase, DNA helicase, etc. In addition, UvrABC system protein C interacted with proteins involved in DNA repairing. It resulted in 7 PPI networks(Fig. 2I) represented as uvrC in red node with its neighbouring proteins. Likewise, Fructose - bisphosphate aldolase class-I, comprised of 295 amino acids, interacted with various proteins involved in glycolysis pathways. It resulted in 20 PPI networks (Fig. 2J) represented as fda in red node with its neighbouring proteins. Lastly, Ribonuclease Y, interacted with proteins involved in RNA degradation and resulted in 3 PPI networks (Fig. 2K) represented as rny in red node with its neighbouring proteins. It has been thus observed that these 11 essential proteins are involved in a variety of critical functions in the pathogen *T. denticola.* According to the STRING results, targeting these essential proteins can result in the diminution of function of the other interconnected proteins and can prove to be invaluable when utilized as therapeutic targets.

Prediction and Validation of 3D Structures: The 3D structures of the 11 selected essential proteins were predicted using online web server SWISS-MODEL of ExPaSy. The amino acid sequences (FASTA sequences) of the selected proteins were inserted into the online web server SWISS-MODEL and the 3D structures of the selected proteins were modelled from a predicted set of templates. For the selected proteins, the SWISS-MODEL server predicted fifty templates for each protein, and the templates having the highest GMQE score and identity percentage were used to model the 3D structures of the selected proteins. Furthermore, the SWISS-MODEL structure validation analysis of the predicted 3D structures of the eleven selected proteins satisfied all the criteria of structure validation, with maximum residues in Ramachandran favoured regions and minimum Ramachandran outliers. The values and calculations showed that the predicted 3D structures were of good quality, as shown in Fig. 3.

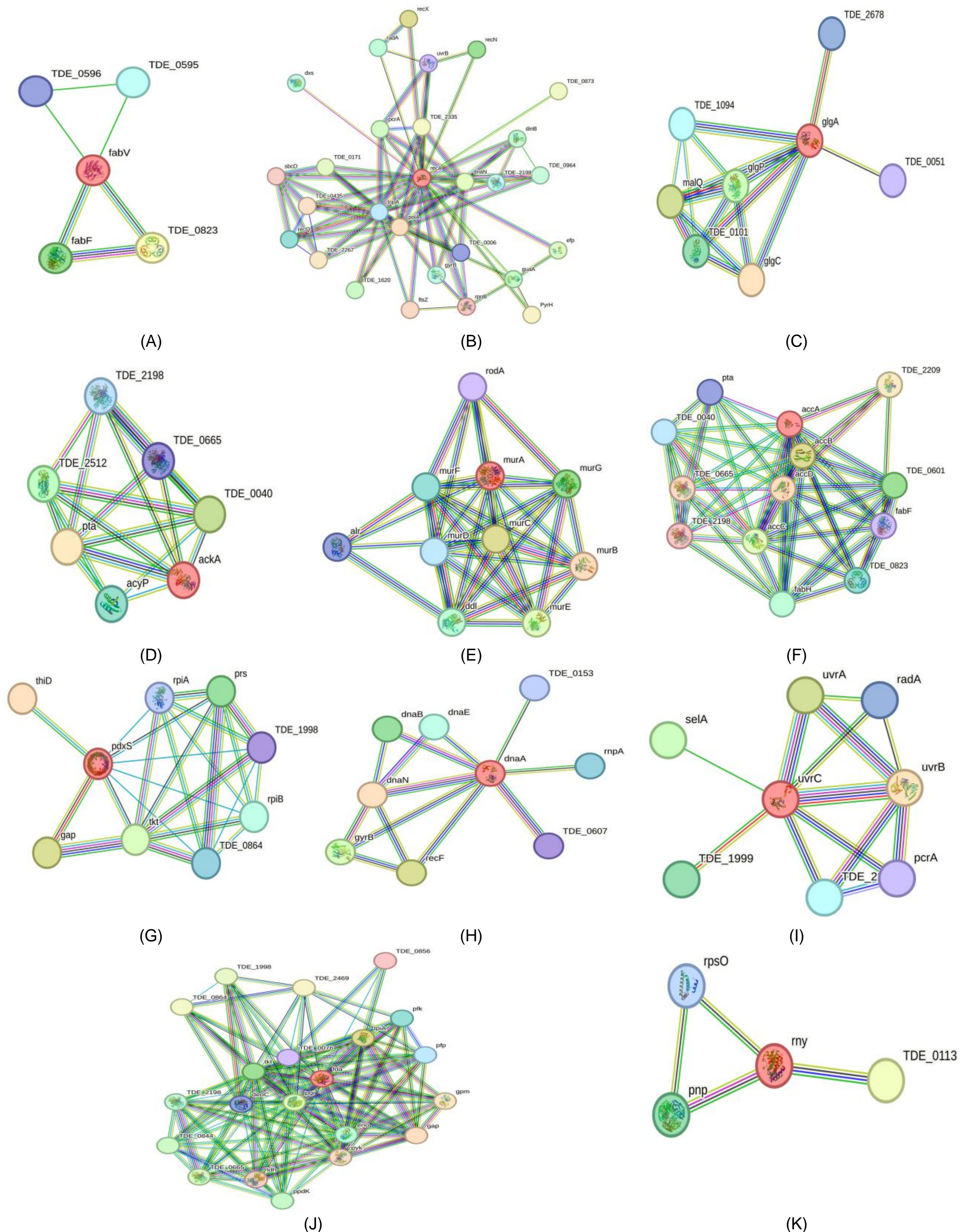


Fig-2: Protein-protein interactions: Schematic PPI network generated through the STRING database for Trans-2-enoyl-CoA reductase (NADH) (A), Protein RecA (B), Glycogen synthase (C), Acetate kinase (D), UDP-N-acetylglucosamine 1-carboxyvinyltransferase (E), Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha (F), Pyridoxal 5'-phosphate synthase subunit PdxS (G), Chromosomal replication initiator protein DnaA (H), UvrABC system protein C (I), Fructose-bisphosphate aldolase class-I (J), and RibonucleaseY(K)

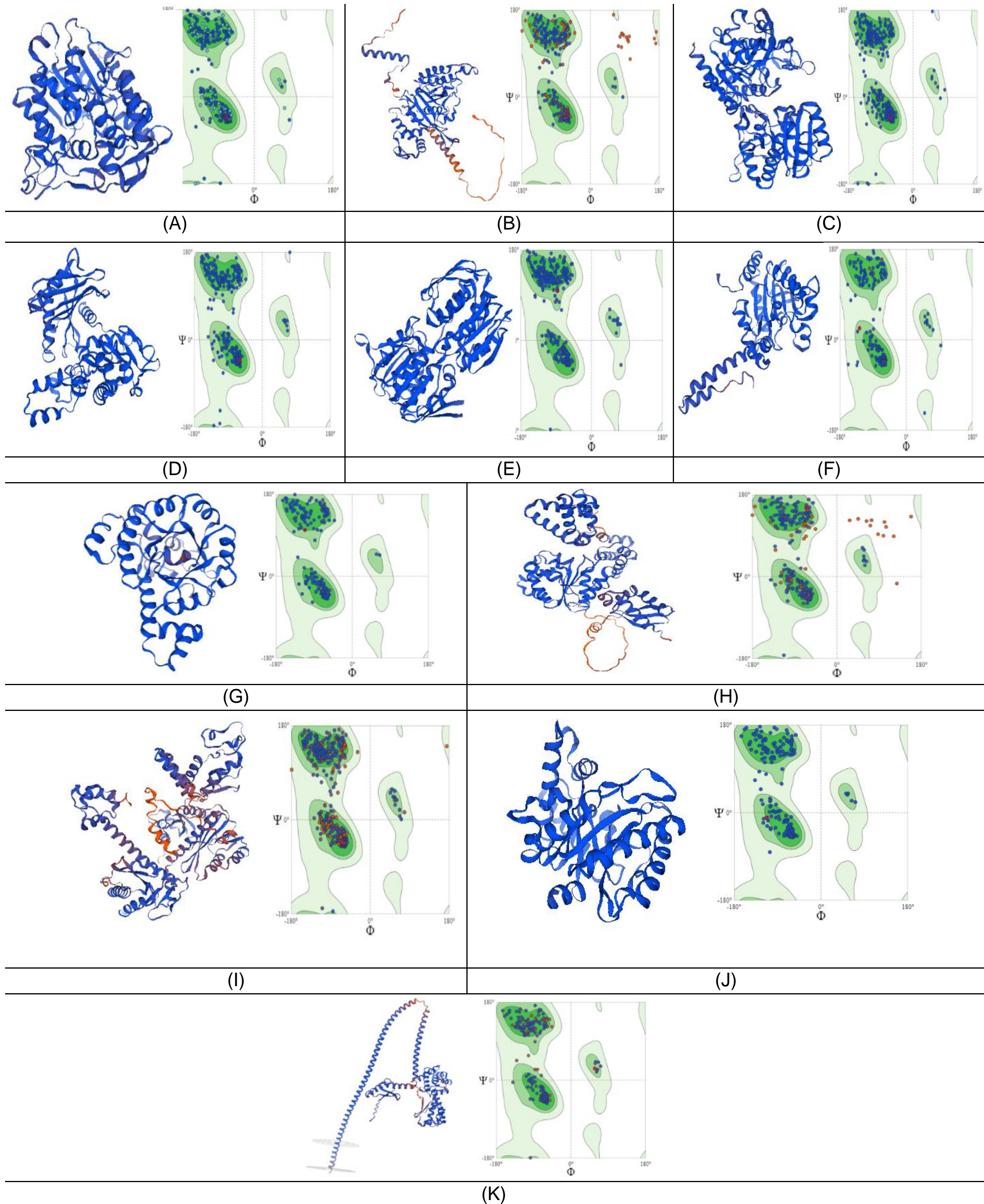


Figure 3: The 3D structures predicted using SWISS MODEL web server and Ramachandran Plot analysis through SWISS MODEL structure validation of the selected proteins Trans-2-enoyl-CoA reductase (NADH) (A), Protein RecA (B), Glycogen synthase (C), Acetate kinase (D), UDP-N-acetylglucosamine 1-carboxyvinyltransferase (E), Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha (F), Pyridoxal 5'-phosphate synthase subunit PdxS (G), Chromosomal replication initiator protein DnaA (H), UvrABC system protein C (I), Fructose-bisphosphate aldolase class-I (J), and Ribonuclease Y (K).

Discussion

The advancements in sequence-based technologies, along with computational studies and the wide range of genomes and proteomics data on different pathogens, enabled the identification of potential therapeutic drug targets and the development of peptide-based vaccines. The in silico subtractive genomic approach has been shown to be an exceptionally promising method widely used against various pathogens such as *Meningococcus B*[37], *Streptococcus pneumonia*[38], and *Salmonella typhii*[14]. In the current study, we have formulated subtractive genomics-based computational approach to screen whole proteome of *Treponema denticola,* comprising of 2753 proteins, for drug target identification. The complete proteome was retrieved from UniProt database and 259 peer-reviewed sequences were selected for further analysis. Afterwards, the selected sequences were then subjected to BLASTp against the human genome. The results showed that 126 proteins of the pathogen were non-homologous with the human host (Homo sapiens) genome. Sequences which are homologous to host genome when treated as drug targets might lead to certain undesirable side effects and cytotoxic reactions in the host. Later on, the identified non-homologous proteins were subjected to the DEG database to get a set of 12 essential proteins. Proteins essential for the survival of pathogen along with the fact that those proteins should also be non-homologous with the host genome; hold great promise to become species-specific drug targets. Following that, pathogen specific metabolic pathways were identified and protein sequences involved in these pathways were analyzed. The common metabolic pathways found in both host and pathogen were excluded while pathways uniquely present in *T. denticola* were considered (Table 2). The proteins involved in these unique pathways were retrieved and proceeded to further downstream analysis to avoid any potential side effects. To narrow down the selection of essential proteins, subcellular localization prediction as well as virulent protein prediction tools were utilized. However, all selected proteins were predicted to be cytoplasmic proteins and no outer membrane proteins were identified and all the proteins were predicted to be non-virulent, when the set of essential proteins were run through pSORTb v3.03, CELLOv2.5 and VirulentPred2.0 respectively. As these eleven proteins were involved in unique metabolic pathways of the pathogen, they are excellent targets for potential drug targets, as reported in different studies.[14,15,39–42] Furthermore, the protein-protein interactions and functional annotation of selected proteins were determined using the STRING server. The analysis successfully revealed that the selected target proteins had the potential to function as central proteins, inter-acting with ten or more other proteins involved in the functioning and survival *T. denticola* and targeting such proteins may affect the activity of all interactor proteins, as reported in different investigations.[11,43] In addition, the 3D structures of the selected proteins were predicted by SWISS-MODEL server and were checked for structure validation within the SWISS-MODEL server. The predicted 3D structures of the eleven selected proteins satisfied all the criteria of structure validation by the use of the aforementioned tools. Finally, we ended up with 11 proteins that are potential target proteins for

drug designing against *T. denticola* based on their substantial role in the pathogen's survival. These proteins need to be characterized for their drug target properties in vitro in the future to achieve newer generation antibiotic drugs. In the future, we are also planning to use these targets to carry out laboratory-based experiments to uncover novel therapeutic compounds. To our knowledge, the current methodology is the first subtractive genomics study against *T. denticola (strain ATCC 35405 / DSM 14222 / CIP 103919 / JCM 8153 / KCTC 15104),* and we believe it will offer an attractive alternative approach to combat the spread of resistant *Treponema denticola* strains.

Conclusion

Bioinformatics tools have revolutionized drug development by analyzing pathogen genomes and proteome sequences across databases. The rise of drug resistance necessitates in silico techniques to find novel drug targets with minimal homology to the host proteome. The subtractive genomics approach can overcome toxicity and dead ends in conventional drug discovery methods, accelerating drug development in unprecedented dimensions. The candidate drug targets which we have sorted out from the whole proteome of *Treponema denticola (strain ATCC 35405 / DSM 14222 / CIP 103919 / JCM 8153 / KCTC 15104)* could accelerate the discovery of novel therapeutic agents against its related infections.

References

1. *Amano A, Takeuchi H, Furuta N. Outer membrane vesicles function as offensive weapons in host–parasite interactions. Microbes Infect. 2010;12(11):791-8.*
2. *Uitto VJ. Dentilisin. In: Handbook of Proteolytic Enzymes. Elsevier; 2013:3217-20.*
3. *Öğrendik M. Periodontal Pathogens in the Etiology of Pancreatic Cancer. Gastrointest Tumors. 2016;3(3-4):125-7.*
4. *Deinzer R, Granrath N, Spahl M, Linz S, Waschul B, Herforth A. Stress, oral health behaviour and clinical outcome. Br J Health Psychol. 2005;10(2):269-83.*
5. *Sela MN. Role of Treponema Denticola in Periodontal Diseases. Crit Rev Oral Biol Med. 2001;12(5):399-413.*
6. *Kaushik SN, Scoffield J, Andukuri A, et al. Evaluation of ciprofloxacin and metronidazole encapsulated biomimetic nanomatrix gel on Enterococcus faecalis and Treponema denticola. Biomater Res. 2015;19:9.*
7. *Loyola-Rodriguez J, Garcia-Cortes J, Martinez-Martinez R, et al. Molecular identification and antibiotic resistant bacteria isolated from primary dentition infections. Aust Dent J. 2014;59(4):497-503.*
8. *Mohanty R, Asopa S, Joseph Md, et al. Red complex: Polymicrobial conglomerate in oral flora: A review. J Family Med Prim Care. 2019;8(11):3480.*
9. *Dentino A, Lee S, Mailhot J, Hefti AF. Principles of periodontology. Periodontol 2000. 2013;61(1):16-53.*
10. *Miesel L, Greene J, Black TA. Genetic strategies for antibacterial drug discovery. Nat Rev Genet. 2003;4(6):442-56.*
11. *Hosen MI, Tanmoy AM, Mahbuba D Al, et al. Application of a subtractive genomics approach for in silico identification and characterization of novel drug targets in Mycobacterium tuberculosis F11. Interdiscip Sci. 2014;6(1):48-56.*

12. *Karunasagar, Mohammed Neema, Karunasagar. In silico identification and characterization of novel drug targets and outer membrane proteins in the fish pathogen Edwardsiella tarda. Open Access Bioinformatics. Published online January 2011:37.*

13. *Rathi B, Sarangi AN, Trivedi N. Bioinformation open access Genome subtraction for novel target definition in Salmonella typhi. print) Bioinformation. 2009;4(4):143-50.*

14. *Khan K, Jalal K, Uddin R. An integrated in silico based subtractive genomics and reverse vaccinology approach for the identification of novel vaccine candidate and chimeric vaccine against XDR Salmonella typhi H58. Genomics. 2022;114(2):110301.*

15. *Khan K, Uddin R. Integrated bioinformatics based subtractive genomics approach to decipher the therapeutic function of hypothetical proteins from Salmonella typhi XDR H-58 strain. Biotechnol Lett. 2022;44(2):279-98.*

16. *Bateman A, Martin MJ, Orchard S, et al. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51(D1):D523-31.*

17. *Apweiler R. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. 2004;32(90001):D115-9.*

18. *Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.*

19. *Zhang R. DEG: a database of essential genes. Nucleic Acids Res. 2004;32(90001):D271-2.*

20. *Zhang R, Lin Y. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. Nucleic Acids Res. 2009;37(Database):D455-8.*

21. *Luo H, Lin Y, Gao F, Zhang CT, Zhang R. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements: Table 1. Nucleic Acids Res. 2014;42(D1):D574-80.*

22. *Luo H, Lin Y, Liu T, et al. DEG 15, an update of the Database of Essential Genes that includes built-in analysis tools. Nucleic Acids Res. 2021;49(D1):D677-86.*

23. *Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35(Web Server):W182-5.*

24. *Yu NY, Wagner JR, Laird MR, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010;26(13):1608-15.*

25. *Gardy JL, Laird MR, Chen F, et al. PSORTb v.2.0: Expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. Bioinformatics. 2005;21(5):617-23.*

26. *Yu C, Lin C, Hwang J. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Science. 2004;13(5):1402-6.*

27. *Yu C, Chen Y, Lu C, Hwang J. Prediction of protein subcellular localization. Proteins: Structure, Function, and Bioinformatics. 2006;64(3):643-51.*

28. *Garg A, Gupta D. VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. BMC Bioinformatics. 2008;9(1):62.*

29. *Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-13.*

30. *Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51(D1):D638-46.*

31. *Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006;22(2):195-201.*

32. *Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-303.*

33. *Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis. 2009;30(Suppl 1):S162-73.*

34. *Bienert S, Waterhouse A, de Beer TAP, et al. The SWISS-MODEL Repository—new features and functionality. Nucleic Acids Res. 2017;45(D1):D313-9.*

35. *Waterhouse AM, Studer G, Robin X, Bienert S, Tauriello G, Schwede T. The structure assessment web server: for proteins, complexes and more. Nucleic Acids Res. 2024;52(W1):W318-23.*

36. *Chen VB, Arendall WB, Headd JJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010;66(1):12-21.*

37. *Mora M, Veggi D, Santini L, Pizza M, Rappuoli R. Reverse vaccinology. Drug Discov Today. 2003;8(10):459-64.*

38. *Talukdar S, Zutshi S, Prashanth KS, Saikia KK, Kumar P. Identification of Potential Vaccine Candidates Against Streptococcus pneumoniae by Reverse Vaccinology Approach. Appl Biochem Biotechnol. 2014;172(6):3026-41.*

39. *Sudha R, Katiyar A, Katiyar P, Singh H, Prasad P. Identification of potential drug targets and vaccine candidates in Clostridium botulinum using subtractive genomics approach. Bioinformation. 2019;15(1):18-25.*

40. *Kumar A, Thotakura PL, Tiwary BK, Krishna R. Target identification in Fusobacterium nucleatum by subtractive genomics approach and enrichment analysis of host-pathogen protein-protein interactions. BMC Microbiol. 2016;16:84.*

41. *Prabha R, Singh DP, Ahmad K, Kumar SPJ, Kumar P. Subtractive genomics approach for identification of putative antimicrobial targets in Xanthomonas oryzae pv. oryzae KACC10331. Archives of Phytopathology and Plant Protection. 2019;52(7-8):863-72.*

42. *Omeershffudin UNM, Kumar S. Antibiotic resistance in Neisseria gonorrhoeae: broad-spectrum drug target identification using subtractive genomics. Genomics Inform. 2023;21(1):e5.*

43. *D'Souza SE, Khan K, Uddin R. Proteogenomic analysis of Serratia marcescens using computational subtractive genomics approach. PLoS One. 2023;18(4):e0283993.*