

# Subtractive Proteomics Analysis Revealed Lipid A-4'phosphatase (LpxF) as a Potential Candidate for Epitope-based Vaccine Design Against *Helicobacter pylori* Infection

Pavan Gollapalli (✉ [gollapallipavan@nitte.edu.in](mailto:gollapallipavan@nitte.edu.in))

K. S. Hegde Medical Academy, Nitte (Deemed to be University)

Tamizh Selvan Gnanasekaran

K. S. Hegde Medical Academy, Nitte (Deemed to be University)

Aditya SJ Rao

CSIR-Central Food Technological Research Institute

Manjunatha Hanumanthappa

Bangalore University

Koigoora Srikanth

University of Aveiro

Praveenkumar Shetty

K. S. Hegde Medical Academy, Nitte (Deemed to be University)

Suchetha Kumari

K. S. Hegde Medical Academy, Nitte (Deemed to be University)

---

## Research Article

**Keywords:** H. pylori infections, antibacterial agents, computational biology, Lipoprotein A-4'-phosphatase (LpxF)

**Posted Date:** May 27th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-557276/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Amidst the surge in the prevalence of resistant *H. pylori* infections, WHO in 2017 has given a high priority to clarithromycin-resistant *H. pylori* for research and to develop new antibacterial agents. In this study, the *Helicobacter pylori* 26695 strain was investigated with extensive computational biology applications to identify novel therapeutic drug targets or vaccine candidates. During the proteomic functional annotation of an organism, it is crucial to determine the function of proteins. The pathogen-specific pathways were found to include only twelve proteins, paving the way further to determine drug or vaccine targets. Lipoprotein A-4'-phosphatase (LpxF) was found to be a novel vaccine target with the highest antigenicity. Having broad-spectrum conservancy with other *H. pylori* strains. Further, an immunoinformatic approach was used to predict an effective epitope-based vaccine against *H. pylori*. LpxF protein has been predicted to have linear and conformational B-cell epitopes and cytotoxic T-lymphocyte epitopes. Virtual screening of all the predicted 35 peptides against human TLR2 receptors resulted in identifying the top 5 peptides. Subsequent redocking with exhaustive parameters reported two peptides with docking energy of -6.9 kcal/mol with a good interaction pattern between the peptide-TLR2 complexes. Furthermore, a panel of two potent epitopes has been proposed that could be used to immunize populations against multiple *H. pylori* infections.

## Introduction

Despite decades of effort, overcoming *Helicobacter pylori* infections has been an arduous task for the scientific community. More than half of the world's population is infected with *H. pylori*, a leading cause of duodenal and gastric ulcer as well as gastric cancer [1]. Targeting the bacterium has become quite challenging task due to its ability to adapt to adverse environmental conditions, such as pH, temperature, adhesion and phenotypic forms [2]. The previous unsuccessful treatment attempts and growing antibiotic resistance impede the pathogen's eradication leading to the development of multidrug resistance (MDR) strains. Among the high-priority pathogen list of world health organizations (WHO), the clarithromycin-resistant *H. pylori* are of precedence for research, discovery, and development of new antimicrobial agents (3). Several factors and mechanisms contribute to MDR development by *H. pylori*, like point mutations, efflux pumps and biofilm formation. The emergence of MDR has become a most significant challenge in the management of *H. pylori* infection. The alarming *H. pylori* antibiotic resistance levels have been found with a prevalence of 80% and 40% in developing and industrialized countries, respectively (4).

*H. pylori* are associated with causing gastritis, a pre-condition in developing a gastric and duodenal ulcer. It is a risk factor for the gastric cancers adenocarcinoma of the distal stomach and gastric mucosa-associated lymphoid tissue (MALT) lymphoma. The outer membrane of Gram-negative bacteria like *H. pylori* is generally constituted with macromolecules like lipopolysaccharide (LPS), one of the essential virulence factors. This LPS comprises lipid A embedded in the outer membrane known as endotoxin that induces fatal hypersensitive reactions of the human immune system at low concentrations. The latter is a core oligosaccharide followed by O-antigen (5). This bacterium can form chronic colonization in the

human stomach by modifying its surface structure, i.e., lipid A, by removing phosphate groups from the 1- and 4'-positions of the lipid A backbone. The studies of Cullen et al. 2011 demonstrated that the ability of *H. pylori* colonization in a mammalian host is due to the dephosphorylation of the lipid A domain of LPS by lpxE and lpxF (6).

Generally, biological processes like purine and pyrimidine metabolism and oxidative phosphorylation appear as cell functions that act as therapeutic targets. Other therapeutic interventions related to metabolic aspects of flagellar assembly, ABC transporter, chemotaxis, protein folding, regulation, metabolic homeostasis, genetic information processing, and resistance to acid and oxidative stresses are also crucial in the pathogenesis of *H. pylori* (7). Global analysis of gene expression explored four independent studies to identify the essential genes of *H. pylori* for its survival in both *in vitro* and *in vivo* conditions (8–11). In contrast, the advances in whole genome sequencing and computational biology have offered various alternative approaches to identify the drug targets with a worth of experimental follow-up (12). Applications of genomics to analyze both pathogen and host-genome sequences have revolutionized focused identification of drug targets more easily (13–14). Findings of novel and unique drug targets remain a specific step in the drug discovery to overcome the infection caused by several drug-resistant pathogens. In this scenario, the advancement in bioinformatics applications has paved the way towards subtractive genomics/proteomics approaches in identifying novel drug targets. Prior to developing the novel therapeutic agents (either by drug-repurposing or new compound discovery) with a novel mode of action, putative targets must be identified in the form of pathogenetic bacterial properties or mechanisms involved in pathogenicity. The approaches like whole-genome mutagenesis allow identifying the potential molecular targets in a pathogen (15) or by elucidating protein-protein interactions, protein-DNA interactions, and protein-RNA interactions potential molecular components of bacterium can be targeted.

This strategy aims to predict non-homologous genes from pathogens against the human host. Further, these non-homologous genes must be essential for the survival, replication, and sustainability of the pathogen and critical in the bacterium's unique metabolic pathways. The essential proteins are prioritized by using various computational biology tools and databases. The screening of the shortlisted proteins for their subcellular localization and their absence as pre-existing targets have revealed the novelty of the targets identified. This study also emphasizes the discovery of B and T-cell epitopes, which are projected as effective peptide vaccine candidates using the immunoinformatics approach. The potential epitope candidates for the development of peptide vaccines were investigated further by molecular docking studies to understand the interactions in peptide-TLR2 complexes.

## Results

The primary focus of this study is to explore the novel therapeutic targets or vaccine candidates against *H. pylori* by employing a structural genomics approach. We used an unreviewed hypothetical proteome set for analysis through several bioinformatics databases and computational biology tools. The entire findings of subtractive proteomics study were briefly summarized in Figure 1.

# Candidate protein for vaccine design

## Analysis of the *H. pylori* hypothetical proteins

A preliminary prediction for the functional annotation was carried out by using the GO FEAT platform. From the total set of *H. pylori* 26695 reference proteome (1, 115 proteins), a set of 944 unreviewed hypothetical proteins were used for the analysis. Further, the proteins of the known domain and/or families and their GO terms were selected (542 proteins) for further analysis (Supplementary File 1). These functionally annotated proteins may play an important role in the cell and are thus labeled as hypothetical proteins (HPs). The functional annotation of HPs assist in gaining the knowledge of structure, function and pathways abetting in the pathogenesis of bacterium and thus crucial in identifying novel therapeutic targets. The human homologous proteins in the pathogen and proteins in metabolic pathways associated with pathogen and host in common were determined using various web-based bioinformatics resources.

## Selection of non-homologous human proteins

All the hypothetical protein sequences analyzed for functional domain/family were then screened only for the non-homologous sequences. BLASTp with an e-value of  $10^{-3}$  threshold against *H. sapiens* from NCBI was performed. This prompted us to acquire a set of 412 non-homologous proteins from *H. pylori* (Supplementary File 1).

## Essential protein analysis

A total of 77 essential genes were identified by performing BLASTp against the DEG database (e-value < 0.0001). These proteins considered crucial for *H. Pylori* survival are unique (Supplementary File 1) and are believed to be in identifying species-specific drug targets/vaccine candidates (44).

## Human Gut microbiota analysis

The inadvertent blockage of the gut floral proteins due to homologous proteins of the pathogen may lead to hostile effects (31). To avoid this, homologous gut microbial proteins which are similar to essential proteins of *H. pylori* were omitted for further analysis. This step is accomplished using BLASTp by choosing the search set against human gut metagenome 16S ribosomal RNA with an expected threshold at 0.05. Found no significant matches suggesting that the entire protein set of essential proteins are unique for the pathogen.

## Analysis of metabolic pathways

We retrieved a total of 342 human metabolic pathways and 95 *H. pylori*-specific metabolic pathways from the KEGG database (Supplementary file 2). For screening novel therapeutic targets, the proteins exclusively involved in the pathways specific for the pathogen were considered. In our study, the set of 77 essential proteins were submitted to KAAS server for assigning KEGG ontology (KO) and specific metabolic pathways. A total of fifty-five proteins were assigned with KO and forty-two proteins were involved in pathways common to *H. sapiens*. These forty-two proteins were omitted for further screening to circumvent cross-reactivity with other human pathogens. Finally, we arrived at conclusion with twelve unique proteins found to be involved in pathogen-specific pathways (Table 1).

## Analysis of subcellular location

The prediction of protein localization serves as a vital parameter in identifying therapeutic targets because many pathogens can span multiple locations (45). Among 12 proteins, only two were identified as inner membrane proteins, 07 as cytoplasmic and the remaining 03 as periplasmic proteins (Table 2).

## Analyzing druggability of hypothetical protein

The novelty of the membrane proteins as a drug target was analyzed using the DrugBank database (Supplementary File 3). The proteins with significant similarity in the database were excluded and the penicillin-binding protein 1A, exhibited similarity above the threshold under consideration. But, the second protein Lipid A-4'-phosphatase (LpxF), reported no similarity with any of the current drug targets in the database.

## 'Anti-target' analysis of the novel drug target

In view of inadvertent side effects, various drug candidates were either withdrawn or reduced their usage except under extreme situation. The cross-reactivity and carcinogenesis check is crucial in selecting an effective drug molecule (46-48). The toxicity caused by the misconstrued binding of drugs to host 'anti-targets' instead of pathogenic targets must be avoided. In this concern, this result revealed no similarity with any of the human 'anti-target' proteins (Supplementary File 4) and thus LpxF is considered as the host 'non-anti-target' protein.

## Antigenicity and allergenicity prediction

The reverse vaccination method is considered one of the powerful approaches in designing a candidate vaccine (49-50). The small antigenic protein sequences were considered for developing a safe recombinant vaccine with the potency to fight against infectious diseases (51). The identified novel drug target of *H. pylori*, LpxF, is subjected to the VaxiJen v2.0 server. The results revealed that the LpxF was the probable antigen protein of *H. pylori* with a score of 0.5232 (threshold:0.4). Further, this membrane

protein can be used to detect the high immunodominant peptide to develop an efficient subunit vaccine against *H. pylori*.

## Conservancy analysis of *H. pylori* 26695lpxF sequence with other strains

A very high conservancy pattern of lpxF was found among various strains of *H. pylori*. This wide range of conservancies found for the predicted lpxF ensures it as a potential drug and/or vaccine target against *H. pylori* (52). The results of pBLAST showing lpxF conservancy from all *H. pylori* strains from diverse geographical locations were tabulated in Supplementary File 5. Hence, lpxF protein might be a potential vaccine candidate/therapeutic target against *H. pylori*.

## Virulence factors of pathogenic *H. pylori*

The virulence mechanism of the non-homologous, essential protein can be explored by submitting it to the virulence factor database. For the query protein lpxF, the virulence factors retrieved were listed in Table 3.

## Peptide vaccine discovery

### B cell epitopes prediction.

Three linear B-cell epitopes were predicted by BCPred with a score value of  $\geq 0.90$  and length of each epitope with 20 amino acids. Another B-cell epitope prediction server from IEDB identified epitopes based on five different methods with all default parameters and were listed in Table 4.

The Surface accessibility of *H. pylori* lpxF is predicted based on the threshold value  $>1$ . The amino acids that fall above this value are probably considered their presence on the protein surface. Here, the maximum surface probability score was found to be 11.053 for FTSRYKPKRWML165-176. Figure 2A depicts the expected surface accessibility of *H. pylori*, while Table S1 (Supplementary Material) lists the maximum and minimum accessibility scores. Karplus and Schulz study on surface flexibility of *H. pylori* lpxF has revealed a highly systematic and disordered structure indicated by the low and high b-factor values. The maximum predicted surface flexibility score is 1.107 for FKGSSRY184-190. The predicted surface flexibility of *H. pylori* lpxF is graphically represented in Figure 2B and the minimum and maximum scores were shown in Table S6 (Supplementary Material).

The Parker approach was used to predict the hydrophobicity of the predicted epitopes from *H. pylori* (53) and was graphically illustrated in Figure 2C. The maximum and minimum hydrophobicity calculated was

5.329 and -7.071, respectively, from all the predicted peptides for IpxF at the amino acid residue positions STAHKDG79-85 and FLSLLLW8-14, and predicted to act as active B-cell epitopes.

## Prediction of Cytotoxic and HTC epitopes

NetCTL 1.2 was used to predict the cytotoxic epitopes for IpxF protein. A total of eight CTL epitopes were predicted based on the defined criteria and specific MHC binding score (Table 5). These potential epitopes for MHC class I molecules against the HLA-A\*24:02 allele was predicted using the SMM method. Further, MHC-I binding and proteasome-dependent C-terminal cleavage of IpxF were also predicted based on weight matrix and artificial neural network. Finally, the prominent epitopes were predicted based on the MHC binding affinity, the TAP score and the C-terminal cleavage score.

## Structural modeling peptide and molecular docking studies

A total of thirty-five epitopes were shortlisted by combining B-cell and T-cell epitopes from the IpxF protein (Table S7, Supplementary Material). The selected top five epitopes from the protein were re-docked with an exhaustive parameter set to 100 for better conformation search. This led to selection of two peptides with the best conformational binding towards the TLR2 target protein and identified the interaction pattern between the peptide-TLR2 complexes. The epitope AGFVYYR (PEP\_19) was able to bind TLR2 with docking energy of -6.9 kcal/mol. The residues Phe3, Tyr5, Tyr6 and Arg7 from the PEP\_19 epitope were found in interaction with and Lue119, Asn137, Phe144 and Asn143 from TLR2. Similarly, the docking energy for binding epitope FAYLFTSRY (PEP\_31) with TLR2 was also calculated as -6.9 kcal/mol. The residues Phe1, Tyr3, Phe5, Thr6 and Ser7 from the PEP\_31 epitope interacted with Asn265, Ser184, His159 and Val154 of TLR2 receptor protein. The hydrogen bonding and hydrophobicity for the docked complexes were analyzed using the Discovery Studio tool. The pattern of interaction was demonstrated in Figures 3 and 4.

## Discussion

The *Helicobacter pylori* responsible for chronic gastrointestinal infections in more than half of the world's population (54) and a high prevalence is found in the East Asian countries (55–56). The systemic effect of *H. pylori* infection on the entire body has drawn considerable attention in recent years (57–58).

Subtractive genomics-based approaches are used to identify the druggable targets or vaccine candidates for *H. pylori*. To recognize specific features such as non-homologous, important, and antigenic proteins, we used the *H. pylori* 26695 strain. Candidate vaccine proteins were chosen from the known membrane proteins that play a critical role in *H. pylori* virulence and survival, as well as pathogen-specific metabolic pathways.

In the present study, we used the entire set of uncharacterized hypothetical proteins to analyze and predict the drug candidate protein. The primary step employed is to find the functional domains/ family

of the hypothetical proteins. The particular function depends on the domains, which are the structural, functional, and evolutionary protein units. Understanding the function of the protein domain is essential to explore its role at cellular level. Secondly, striking homology exists between bacteria and human protein since these proteins are involved in a typical cellular system (59–60). Hence, the cross-reactivity with host homologous proteins must be avoided during therapeutic development and administration to bind pathogen-specific target proteins. Thirdly, critical genes were defined from the non-homologous set collected, since essential genes are required for the bacterial proteome's cellular processes to continue functioning (61).

Further, a gut microbiota refers to the large population of bacteria that colonize the human intestinal tract(62). Pathogens that cause human inflammatory diseases are closely associated with gut microbiota; these pathogens co-evolve and self-multiply in a symbiotic relationship with gut microbiota (63). The IpxF, which is predicted to be localized in the cell's inner membrane, is hypothesized as a potential vaccine candidate. While cytoplasmic proteins may be potential drug targets, we will focus on inner membrane proteins for this study because membrane proteins account for more than 60% of therapeutic targets (64). Considering membrane proteins as targets for a variety of reasons that include: (1) protein functions will be calculated using machine learning and computer-based methods prior to in vivo or in vitro laboratory trials; (2) A nature of the unique structure of membrane proteins will facilitate to predict and generate their secondary structure (65). The mutational change and exchange of genes among pathogens occur due to the overuse of broad-spectrum therapeutics. The emergence of the antibiotic resistance crisis is mainly because of the misuse of antibiotics and the lack of new therapeutics (64–65).

The IpxF removes the 4'-phosphate group from tetra- and hexa- acylated lipid A species and has no 1-phosphatase or 2-keto-3-deoxy-D-manno-octurosonic acid (Kdo2) hydrolase activity. The absence of the 4'-phosphate group renders the bacteria resistant to host-derived cationic antimicrobial peptides (CAMP), allowing it to camouflage itself from the host's innate immune response, and plays a critical role in the long-term colonization of the host's stomach. This protein is involved in the LPS lipid A biosynthesis pathway, which is part of bacterial outer membrane biogenesis (66). To identify an epitope-based peptide vaccine that can enhance the immunity against *H. pylori* infections, we predicted and validated both B-cell and T-cell epitopes. Upon virtual screening of the peptide library of 35 epitopes based on MHC binding affinity, C-terminal cleavage score, TAP score and molecular docking, we hypothesize that epitopes AGFVYYR (PEP\_19) and FAYLFTSRY (PEP\_31) from IpxF protein as potential epitope-based peptides. These T-cell epitopes could boost the immune responses in the human host cell. For a wide variety of bacterial products like lipoproteins and LPS, the TLR2 plays an important role in initiating the cellular innate immune responses (67). TLR2 is a transmembrane type I glycoprotein composed of three domains: the N-terminal extracellular domain, the transmembrane region, and the cytoplasmic toll/IL1 receptor (TIR) domain. The N-terminal extracellular domain, which consists of leucine-rich repeats (LRR) with the conserved motif "LxxLxLxxN" (20–29 amino acid residues), is known to be involved in ligand recognition(68–70). TLR2 recognizes a wide range of ligands, especially those made up of bacterial cell walls. Lipoproteins, peptidoglycan (PGN), lipoteichoic acid (LTA), lipopolysaccharides (LPS) from certain

bacterial species (e.g., *Porphyromonas gingivalis*), porins from *Neisseria*, lipoarabinomannan from mycobacteria, and zymosan and phospholipomannan from yeast cell walls are only a few examples [69, 71, 72]. The TLR2 has unique properties to be investigated in vaccine development and its ability to covalently bind TLR2 ligands to antigens and the enhancement of antigens coupled to TLR2-targeting lipid moieties for direct and cross-presentation. TLR2 stimulation's ability to induce healthy Th responses and regulatory mechanisms and its mucosal imprinting properties can aid in the resolution of actual vaccine challenges [73].

We hypothesize that the extracellular domain of TLR2 may be involved in this interaction as LRRs of TLR2 carry the specificity allowing binding of both epitope-based peptides (PEP\_19 and PEP\_31) and signaling. These epitopes' allergic and antigenic profiles have also been confirmed, indicating that they are potent candidates for vaccine production. Furthermore, B-cell epitopes have been identified as the preferred method for eliciting a B-cell immune response. Overall study will aid progress in the peptide-based vaccine development against *H. pylori* infections. As a result of our research, we've identified novel peptides that could help to combat the MDR bacteria crisis by creating new *H. pylori* therapeutic.

## Conclusions

In the current study, the computational subtractive genomic method explored the *lpxF* of *H. pylori* as a potential vaccine target candidate. It is found that this protein is found to be involved in the pathogen-specific 'lipopolysaccharide biosynthesis' metabolic pathway. This inner membrane protein *lpxF* does not show any homology with the human proteome to avoid a potential autoimmune response. Moreover, the absence of cross-reactivity with other pathogenic antigens and suitable antigenic and adhesion properties are most essential for the pathogenesis of the microbe and protection against the infectious diseases. The *lpxF* also showed no similarity against the anti-targets in the human. Therefore, this study shows *lpxF* as a prioritized protein of possible vaccine target candidates. Further, the epitopes predicted for *lpxF* were screened for the potential binding against the human TLR2 receptor using virtual screening and molecular docking approaches. The two peptides named PEP\_19 (AGFVYYR) and PEP\_31 (FAYLFTSRY) were reported with strong hydrophobic and hydrogen-bond interaction for TLR2. We conclude by recommending these epitope-based peptides validation further by *in vivo* and *in vitro* studies to prove their effectiveness.

## Methods

### Subtractive proteomics approach for identifying candidate protein for vaccine design

### Retrieval of the proteomic data set

The whole proteome (unreviewed) of *Helicobacter pylori* (strain ATCC700392/26695)/ *Campylobacter pylori* were retrieved from the UniProt database (proteome ID: UP00000429). The reference proteome contains a total of 1,115 proteins and 944 unreviewed proteins.

## Functional annotation of hypothetical proteins

The entire set of 944 unreviewed hypothetical proteins (HPs) were subjected to the GO FEAT 1.0 server as a preliminary prediction for the functional annotation with e-value  $1e-03$  (16). Multiple databases like Uniprot (17), InterPro (18), Pfam (19), NCBI (<https://www.ncbi.nlm.nih.gov/>) and EMBL (<https://www.ebi.ac.uk/>) were used for rapid functional annotation of the HPs set using GO FEAT server. The HPs set describing the functional domain and/or protein family were considered for further analysis.

## Identifying non-homologous bacterial proteins in humans.

All the HP sequences analyzed by the GO FEAT tool were then screened for non-homologous proteins. Identifying the protein sequences that are not analogous to the human proteins is an objective of this step and is achieved by performing BLASTp against the human proteome set with e-value  $<0.0005$  (20).

## Identification of essential non-homologous proteins

Essential genes are those indispensable for the survival of the organism by constituting various cellular processes. Hence, identifying the essential genes is a critical step in the screening of potential drug targets. BLASTp from the database of essential proteins (DEG) is used to predict the essential non-homologous protein from the protein set with a threshold e-value  $<0.0001$  (21).

## Identification of orthologs in the human gut microbiome.

Gastrointestinal digestion in humans mainly depends on the activity of microbes that reside in the gut region as, gut microbiota plays a critical role in keeping the health of gastrointestinal. The gastrointestinal tract of healthy humans houses about  $10^{10}$ - $10^{14}$  microbes and exist in a symbiotic association with the host (22). Adverse pharmacokinetic side effects in the host may be attributed due to interaction and binding of the pathogenic homologous proteins with human gut microbiota proteins. Hence, the above-identified essential proteins were screened for non-homology with human gut attributed using BLASTp at significance of e-value  $<0.0001$  (23-24).

## Metabolic pathway analysis.

The KEGG pathway database was used for metabolic pathway comparison of *H. pylori* 26695 (hpy) and *H. sapiens* (hsa) (25). To identify pathogen-specific unique pathways. The non-homologous essential proteins were further screened by performing BLASTp analysis using KASS server (26) and the only proteins involved in pathogen-specific pathways were sorted for further analysis.

## Prediction of protein subcellular localization

The five major localizations of Gram-negative bacterial proteins include cytoplasm, inner membrane, peripheral membrane, outer membrane and extracellular region (27). The prediction of subcellular localization is essential for categorizing the proteins as a drug target or vaccine candidate. Generally, the proteins in the cytoplasm can be referred to as drug targets and those on the membrane as vaccine targets (28). In this study, subcellular localization of the shortlisted proteins is identified using PSORTb v3.2 server (27), CELLO v2.5 server (27) and PSLPred (29). For protein localization, the exact position predicted by all three servers or at least any two servers. Only membrane proteins were considered as potential drug target and vaccine candidates, used for druggability analysis.

## Druggability analysis

Using the BLASTp software with an e-value of 0.001, the DrugBank 5.0 database (30) was used to classify novel drug targets or vaccine candidates from membrane proteins with default parameters.

## 'Anti-target' analysis

The adverse pharmacological effects could be developed in the host due to the non-specific binding and interaction of the therapeutic compounds instead of acting against pathogenic proteins due to shared homology. Anti-targets are analogous human proteins and a total of 210 proteins were identified via a literature search, as shown in Supplementary Table 4 (31) using BLASTp at NCBI was used by submitting the novel membrane proteins against these 'anti-targets' with e-value<0.005 and identity of < 25%.

## Antigenicity and allergenicity of druggable proteins

The magnitude protein-antigenicity is greatly important for designing a subunit vaccine. VaxiJen v2.0 server (32) was used to screen antigenicity of the novel drug target protein. This step reveals the possible antigenic proteins from the set of druggable membrane proteins, by which further immunodominant peptides can be screened for peptide-based vaccine development. The protein showing hits >0.4 were used for conservative analysis.

## Comparison of predicted sequences with other strains

The spectrum of drugs in the entire homologous bacterial community can be determined by estimating conservancy patterns among predicted sequences with different classically used strains of the same species. BLASTp was used for the analysis by keeping all default parameters except “*Helicobacter pylori*” in the organism option.

## **Virulence factor analysis**

A comprehensive data set of virulence factors defined by sixteen dominant bacterial pathogens were provided by Virulence Factor Database (VFDB) (33). These virulence factors cause the colonization of bacterial pathogens leading to the destruction of host cell.

## **Immunoinformatic approach for epitope-based peptide design and validation**

### **Linear B-Cell Epitopes prediction**

The identification and characterization of B-cell epitopes play an important role in vaccine design. B-cell epitopes were predicted using BCPREDS (<http://ailab-projects1.ist.psu.edu:8080/bcpred/>) using protein sequence information (34).

To predict 20-mer linear B-cell epitopes, the BCPred server uses kernel technique. BCPred predicts B-cell (linear) epitopes using a support vector machine (SVM) algorithm. The Parker hydrophilicity prediction values (35), Emini surface accessibility prediction values (36), Kolaskar and Tongaonkar's antigenicity scale values (37), Karplus and Schulz Flexibility Prediction values, and Chou and Fasman  $\beta$ -turn prediction methods were obtained using IEDB (<http://www.iedb.org/>) online tools. B-cell epitopes have a hydrophilic nature and are accessible for flexible regions (38).

### **Potential Cytotoxic T-lymphocyte (CTL) Epitopes Prediction**

The NetCTL.1.2 server (<http://www.cbs.dtu.dk/services/NetCTL/>) was used to predict CTL epitopes. NetCTL accepts the FASTA sequence format for various analysis, including MHC class I binding affinity prediction, TAP transport performance, and C-terminal cleavage prediction. MHC-I binding and proteasome-dependent C-terminal cleavage were predicted using an artificial neural network and a weight matrix(39).

## **Construction of Peptide library and analysis of peptide-target interactions**

Based on each peptide's sequence information, the 3D structures were generated using discovery studio visualized v.20. The energy minimization was done using Maestro software V12.7. Thus prepared peptide structures were added with Gasteiger charges [40] using Autodock tools and processed.PDBQT file'. Subsequently, the molecular docking study was performed to screen the library of epitopes modeled against the human TLR2 receptor (PDB ID: 6NIG) to predict the best fit peptide and their binding interactions. The target protein structure was retrieved from the protein data bank (PDB) and edited to remove hetero atoms and added Kollmann charges [41]. Grid box of size 74x, 110y and 110z with centers 4.468x, 7.467y and 23.688z was set around the residues forming the binding site of the target protein and the Broyden-Fletcher-Goldfarb-Shanno algorithm implemented in the AutoDock Vina was employed to study appropriate binding modes of each peptide in their most stable conformation [42]. For the peptide molecules, all the torsions were allowed to rotate during docking [43].

## Declarations

## Acknowledgements

The authors are thankful to Nitte (Deemed to be University), Mangalore, India for providing all the facilities to complete this work.

## Author contributions statement

P.G. Supervised and planned the experiments, performed the experiments and wrote the paper draft. P.G., A.R.S.J. and T.S.G. carried out the experiments. A.R.S.J. carried out docking studies. P.G., A.R.S.J. and T.S.G contributed to the interpretation of the results. M.H. contributed in manuscript editing. K. S. assisted in language editing and formatting the manuscript. S.K., and P.K.S. helped shape the research. P.G. took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript.

## Declaration of Competing Interest

The authors declare that they have no conflict of interest.

## References

1. Hathroubi, S., Zerebinski, J. &Ottemann, K. M. Helicobacter pylori Biofilm Involves a Multigene Stress-Biased Response, Including a Structural Role for Flagella. *mBio*. 9(5), e01973-18 (2018).
2. Prasad, A. et al. Phyto anti-biofilm elicitors as potential inhibitors of Helicobacter pylori. *3 Biotech* 9, 53 (2019).
3. Boyanova, L., Markovska, R. &Mitov, I. Multidrug resistance in anaerobes. *Future Microbiol*. 14, 1055–1064 (2019).

4. Thung, I. et al. Review article: the global emergence of *Helicobacter pylori* antibiotic resistance. *Aliment Pharmacol Ther.* 43(4), 514–533 (2016).
5. Raetz, C. R. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu Rev Biochem.* 71, 635–700 (2002).
6. Cullen, T. W. et al. *Helicobacter pylori* versus the host: remodeling of the bacterial outer membrane is required for survival in the gastric mucosa. *PLoS Pathog.* 7(12), e1002454 (2011).
7. Duckworth, M. J., Okoli, A. S. & Mendz, G. L. Novel *Helicobacter pylori* therapeutic targets: the unusual suspects. *Expert Rev Anti Infect Ther.* 7(7), 835–867 (2009).
8. Salama, N. R., Shepherd, B. & Falkow, S. Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*. *Journal Bacteriol.* 186(23), 7926–7935 (2004).
9. Chalker, A. F. et al. Systematic identification of selective essential genes in *Helicobacter pylori* by genome prioritization and allelic replacement mutagenesis. *J Bacteriol.* 183(4):1259-1268 (2001).
10. Baldwin, D. N. et al. Identification of *Helicobacter pylori* genes that contribute to stomach colonization. *Infect Immun.* 75(2), 1005–1016 (2007).
11. Kavermann, H. et al. Identification and characterization of *Helicobacter pylori* genes essential for gastric colonization. *J Exp Med.* 197(7), 813-822 (2003).
12. Mondal, S. I. et al. Identification of potential drug targets by subtractive genome analysis of *Escherichia coli* O157:H7: an in silico approach. *Adv Appl Bioinform Chem.* 8, 49–63 (2015).
13. Allsop, A. E. Bacterial genome sequencing and drug discovery. *Curr Opin Biotechnol.* 9(6), 637–642 (1998).
14. Stumm, G., Russ, A. & Nehls, M. Deductive genomics: a functional approach to identify innovative drug targets in the post-genome era. *Am J Pharmacogenomics.* 2(4):263-271 (2002).
15. Fang, G., Rocha, E. & Danchin, A. How essential are nonessential genes? *Mol Biol Evol.* 22(11):2147-2156 (2005).
16. Araujo, F. A. et al. GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. *Sci Rep.* 8(1), 1794 (2018).
17. Chen, C., Huang, H. & Wu, C. H. Protein Bioinformatics Databases and Resources. *Methods Mol Biol.* 1558, 3–39 (2017).
18. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* 47(D1), D427–D432 (2019).
19. Finn, R. D. et al. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* 45(D1), D190–D199 (2017).
20. Anishetty, S., Pulimi, M. & Pennathur, G. Potential drug targets in *Mycobacterium tuberculosis* through metabolic pathway analysis. *Comput Biol Chem.* 29(5), 368–378 (2005).
21. Zhang, R., Ou, H. Y. & Zhang, C. T. DEG: a database of essential genes. *Nucleic Acids Res.* D271–D272 (2004).
22. Fujimura, K. E. et al. Role of the gut microbiota in defining human health. *Expert Rev Anti Infect Ther.* 8(4), 435–454 (2010).

23. Altschul, S. F. et al. Basic local alignment search tool. *J Mol Biol.* 215(3), 403–410 (1990).
24. Raman, K., Yeturu, K. & Chandra, N. targetTB: a target identification pipeline for *Mycobacterium tuberculosis* through an interactome, reactome and genome-scale structural analysis. *BMC Syst Biol.* 2, 109 (2008).
25. Paduano, F. & Forbes, A. G. Extended LineSets: a visualization technique for the interactive inspection of biological pathways. *BMC Proc.* 9(Suppl 6 Proceedings of the 5th Symposium on Biological Data), S4 (2015).
26. Moriya, Y. et al. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35(Web Server issue), W182–W185 (2007).
27. Yu, N. Y. et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics.* 26(13), 1608–1615 (2010).
28. Barh, D. et al. In silico subtractive genomics for target identification in human bacterial pathogens. *Drug Development Research.* 72(2):162-77 (2011).
29. Bhasin, M., Garg, A. & Raghava, G. P. PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinformatics.* 21(10), 2522–2524 (2005).
30. Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46(D1), D1074–D1082 (2018).
31. Shanmugham, B. & Pan, A. Identification and characterization of potential therapeutic candidates in emerging human pathogen *Mycobacterium abscessus*: a novel hierarchical in silico approach. *PLoS One.* 8(3), e59126 (2013).
32. Doytchinova, I. A. & Flower, D. R. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine.* 25(5), 856–866 (2007).
33. Chen, L. et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* D325–D328 (2005).
34. El-Manzalawy, Y., Dobbs, D. & Honavar, V. Predicting linear B-cell epitopes using string kernels. *J Mol Recognit.* 21(4), 243–255 (2008).
35. Parker, J., Guo, D. & Hodges, R. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry.* 25, 5425–5432 (1986).
36. Emini, E. A. et al. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol.* 55, 836–839 (1985).
37. Kolaskar, A. & Tongaonkar, P. C. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* 276, 172–174 (1990).
38. Fieser, T. M. et al. Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein alpha-helix. *Proc Natl Acad Sci U S A.* 84, 8568–8572 (1987).

39. Larsen, M. V. et al. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*. 8, 424 (2007).
40. Gasteiger, J. & Marsili, M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* [Internet]. Elsevier {BV}; 36:3219–3228 (1980).
41. Janakirama, A., R., et al. Characterization of isolated compounds from *Morus* spp. and their biological activity as anticancer molecules. *BiolImpacts*. (2020)
42. Trott, O, & Olson, A.J. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 31:455–461 (2010).
43. Raghavendra, S, et al. Multiple ligand simultaneous docking (MLSD): A novel approach to study the effect of inhibitors on substrate binding to PPO. *Comput Biol Chem*. 59 (2015).
44. Judson, N. & Mekalanos, J. TnAraOut, a transposon-based approach to identify and characterize essential bacterial genes. *Nat Biotechnol*. 18(7), 740–745 (2000).
45. Golkar, Z., Bagasra, O. & Pace, D. G. Bacteriophage therapy: a potential solution for the antibiotic resistance crisis. *J Infect Dev Ctries*. 8(2), 129–136 (2014).
46. Fenner, H. Evaluation of the efficacy and safety of NSAIDs. A new methodological approach. *Scand J Rheumatol Suppl*. 80, 32–39 (1989).
47. Bakke, O. M., Wardell, W. M. & Lasagna, L. Drug discontinuations in the United Kingdom and the United States, 1964 to 1983: issues of safety. *Clin Pharmacol Ther*. 35(5), 559–567 (1984).
48. Moses, P.L. et al. Severe hepatotoxicity associated with bromfenac sodium. *Am J Gastroenterol*. 94(5):1393-1396 (1999).
49. Galan, M. V. et al. The burden of acute nonfulminant drug-induced hepatitis in a United States tertiary referral center [corrected]. *J Clin Gastroenterol*. 39(1), 64–67 (2005).
50. Rappuoli, R. Reverse vaccinology. *Curr Opin Microbiol*. 3(5), 445–450 (2000).
51. Rappuoli, R. Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine*. 19(17-19), 2688–2691 (2001).
52. Khan, M. T. et al. Subtractive genomics approach towards the identification of novel therapeutic targets against human *Bartonella bacilliformis*. *Informatics in Medicine Unlocked*. Volume 20, 100385 (2020).
53. Parker, J., Guo, D. & Hodges, R. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*. 25, 5425–5432 (1986).
54. Hooi, J. et al. Global Prevalence of *Helicobacter pylori* Infection: Systematic Review and Meta-Analysis. *Gastroenterology*. 153(2), 420–429 (2017).
55. Hunt, R. H. et al. *Helicobacter pylori* in developing countries. World Gastroenterology Organisation Global Guideline. *J Gastrointestin Liver Dis* 20(3), 299–304 (2011).

56. Parsonnet, J. et al. Helicobacter pylori infection and the risk of gastric carcinoma. *N Engl J Med.* 325(16), 1127–1131 (1991).
57. Goni, E. & Franceschi, F. Helicobacter pylori and extragastric diseases. *Helicobacter.* 21 Suppl 1, 45–48 (2016).
58. Kyburz, A. & Müller, A. Helicobacter pylori and Extragastric Diseases. *Curr Top Microbiol Immunol.* 400, 325–347 (2017).
59. Gupta, S.K. et al. Definition of Potential Targets in Mycoplasma Pneumoniae Through Subtractive Genome Analysis. *J AntivirAntiretrovir* 2: 038-041 (2010).
60. Hediger, M.A., Turk, E. & Wright, E.M. Homology of the human intestinal Na<sup>+</sup>/glucose and Escherichia coli Na<sup>+</sup>/proline cotransporters. *Proc Natl Acad Sci U S A.* 86(15):5748-5752 (1989).
61. Zhang, Z. & Ren Q. Why are essential genes essential? - The essentiality of Saccharomyces genes. *Microb Cell* 2(8): 280-287 (2015).
62. Jandhyala, S. M. et al. Role of the normal gut microbiota. *World J Gastroenterol.* 21(29), 8787–8803 (2015).
63. Pickard, J. M. et al. Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. *Immunol Rev.* 279(1), 70–89 (2017).
64. Michael, C.A., Dominey-Howes, D. & Labbate, M. The antimicrobial resistance crisis: causes, consequences, and management. *Front Public Health.* 2:145 (2014).
65. Bartlett, J.G., Gilbert, D.N. & Spellberg, B. Seven ways to preserve the miracle of antibiotics. *Clin Infect Dis.* 56(10):1445-50 (2013).
66. Cullen, T.W. et al. Helicobacter pylori versus the host: remodeling of the bacterial outer membrane is required for survival in the gastric mucosa. *PLoS Pathog.* 7(12):e1002454 (2011).
67. Vasselon, T. et al. TLR2 recognizes a bacterial lipopeptide through direct binding. *J Immunol.* 173(12):7401-7405 (2004).
68. Kawai, T. & Akira, S. The roles of TLRs, RLRs and NLRs in pathogen recognition. *Int Immunol.* vol. 21, no. 4, 317–337 (2009).
69. Kumar, H., Kawai, T. & Akira, S. Pathogen recognition in the innate immune response. *Biochem J.* 420(1); 1–16 (2009).
70. Kumar, H., Kawai, T. & Akira, S. Toll-like receptors and innate immunity. *Biochemical and Biophysical Research Communications.* 388(4), 621–625 (2009).
71. Warshakoon, H. J. et al. Potential adjuvant properties of innate immune stimuli. *Hum Vaccin.* 5(6); 381–394 (2009).
72. Hirschfeld, M. et al. Signaling by Toll-like receptor 2 and 4 agonists results in differential gene expression in murine macrophages. *Infect Immun.* 69(3); 1477–1482 (2001).
73. Basto, A.P. & Leitão, A. Targeting TLR2 for vaccine development. *J Immunol Res.* 2014:619410 (2014).

## Tables

Table 1				
Proteins involved only in <i>H. pylori</i> specific unique pathways				
Sl. No	KO Assignment	Protein Name	Gene Symbol	Pathway
01	K00172	pyruvate ferredoxin oxidoreductase gamma subunit	porC	Biosynthesis of secondary metabolites; Microbial metabolism in diverse environments; Methane metabolism
02	K00175	2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase subunit beta	korB	Biosynthesis of secondary metabolites; Microbial metabolism in diverse environments
03	K00177	2-oxoglutarate ferredoxin oxidoreductase subunit gamma	korC	Biosynthesis of secondary metabolites; Microbial metabolism in diverse environments
04	K01808	ribose 5-phosphate isomerase B	rpiB	Biosynthesis of secondary metabolites; Microbial metabolism in diverse environments
05	K11753	riboflavin kinase / FMN adenylyltransferase	ribF	Biosynthesis of secondary metabolites
06	K11784	cyclic dehypoxanthinylfutalosine synthase	mqnC	Biosynthesis of secondary metabolites
07	K02527	3-deoxy-D-manno-octulosonic-acid transferase	kdtA	Lipopolysaccharide biosynthesis
08	K23159	lipid A 4'-phosphatase	lpxF	Lipopolysaccharide biosynthesis
09	K05366	penicillin-binding protein 1A	mrcA	Peptidoglycan biosynthesis; beta-Lactam resistance
10	K03203	type IV secretion system protein VirB8	virB8	Bacterial secretion system
11	K00407	cytochrome c oxidase cbb3-type subunit IV	ccoQ	Two-component system
12	K12368	dipeptide transport system substrate-binding protein	dppA	Bacterial chemotaxis

**Table 2**

Subcellular location of twelve non-homologous, essential proteins only involved in the unique metabolic pathways of the pathogen.

Sl. No	Protein	PSORTb	CELLO v.2.5	PSLpred
1	tr O30648	Cytoplasm	Cytoplasm	Cytoplasm
2	tr O25312	Cytoplasm	Cytoplasm	Cytoplasm
3	tr D7FDT9	Cytoplasm	Cytoplasm	Cytoplasm
4	tr AOA024C7N3	Cytoplasm	Cytoplasm	Cytoplasm
5	tr AOA3Q9Y0Z5	Cytoplasm	Cytoplasm	Cytoplasm
6	tr O25370	Cytoplasm	Cytoplasm	Cytoplasm
7	tr AOA7G1HQN3	Cytoplasm	Cytoplasm	Cytoplasm
8	tr Q9ZJ31	Cytoplasmic membrane	Inner membrane	Inner membrane
9	tr AOA3Q9Y2E2	Cytoplasmic membrane	Outer membrane	Inner membrane
10	tr AOA086RT63	Cytoplasm	Outer membrane	Extracellular protein
11	tr O24953	unknown	cytoplasm	Periplasmic membrane
12	tr AOA1Q2RQ85	Periplasma	Periplasma	Periplasmic membrane

**Table 3**

Virulence factor of the hypothetical protein of lpfX.

<b>Protein</b>	<b>Virulence Factors</b>				
Lipid A 4'-phosphatase (lpxF)	<b>VFDB ID</b>	<b>Gene symbol</b>	<b>Description</b>	<b>Score (Bits)</b>	<b>E-value</b>
	VFG002340	fliR	Flagellar biosynthetic protein	28	0.66
	VFG011439	lpxE	Phosphatidyl glycerophosphates	27	1.5
	VFG010882	ccmF	Cytochrome c heme lyase subunit	26	1.9
	VFG045531	lpg1661	Dot/Icm Type IV secretion synthase effect	25	4.3
	VFG000495	ssaC	Type III secretion system synthetin	24	7.3
	VFG020191	yadA	Trimeric autotransporter adhesin	24	9.5
	VFG002437	boaB	Autotransporter protein	24	9.5

**Table 4**

B-Cell epitopes predicted by BCPRe tool.

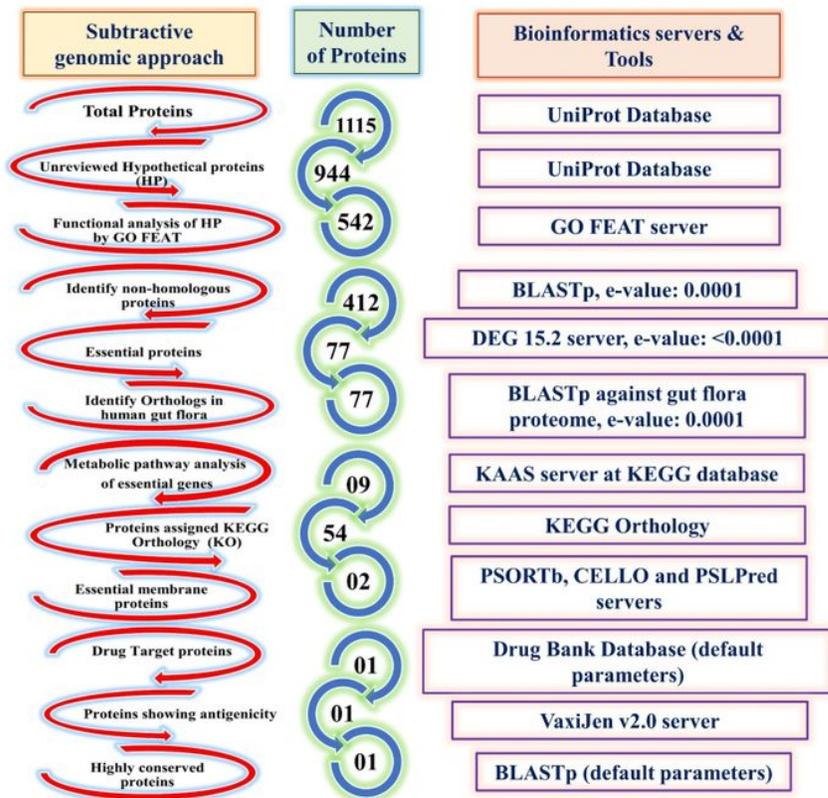
<b>Position</b>	<b>Epitope</b>	<b>Score</b>
93	RPCCNSWRGMPSGHAGGAFS	0.991
13	LWVYPLKSEPINEGAYILEE	0.963
169	YKPKRWMLYPEISSDFKGSS	0.951

**Table 5**

List of the total T-cell epitopes vaccine candidate peptides predicted by NetCTL tool.

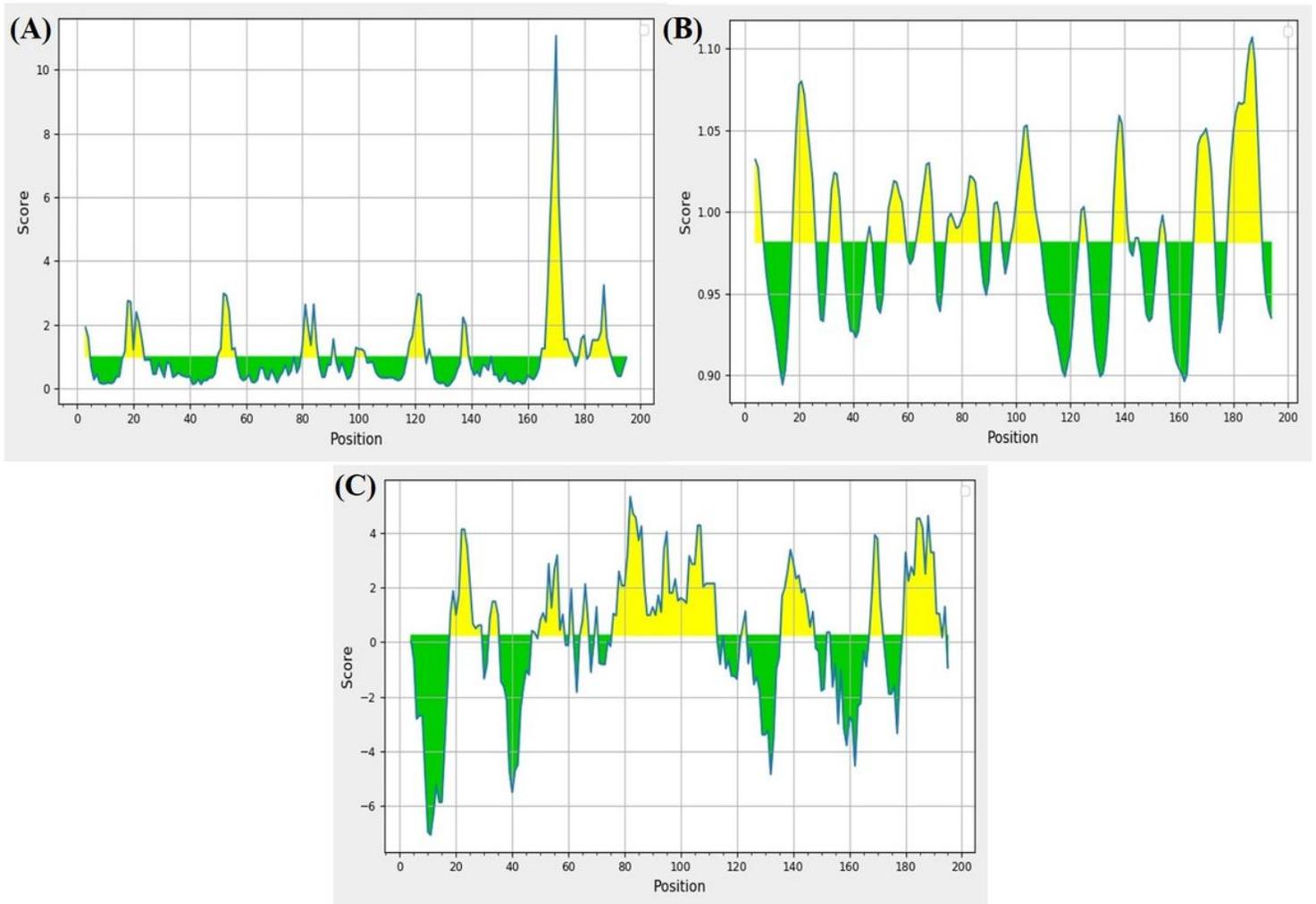
S. No	Peptide Sequence	MHC Binding Affinity	Rescale Binding Affinity	C-terminal Cleavage Affinity	Transport Affinity	Prediction Score	MHC-I Binding
111	FSAAGFVYY	0.5983	2.5403	0.9459	3.0410	2.8343	Yes
8	FLSLLLWVY	0.5043	2.1413	0.9658	2.8930	2.4309	Yes
46	TVSLAMRDY	0.2432	1.0326	0.8803	3.0080	1.3150	Yes
161	FAYLFTSRY	0.2235	0.9489	0.9668	3.0390	1.2459	Yes
155	SLIAWGFAY	0.2058	0.8737	0.9717	3.0360	1.1712	Yes
64	TLVTQGVYI	0.1671	0.7093	0.9445	2.9710	0.9996	Yes
169	YKPKRWMLY	0.1587	0.6740	0.6809	3.0710	0.9297	Yes
113	AAGFVYYRY	0.1562	0.6633	0.9697	3.0120	0.9594	Yes

## Figures



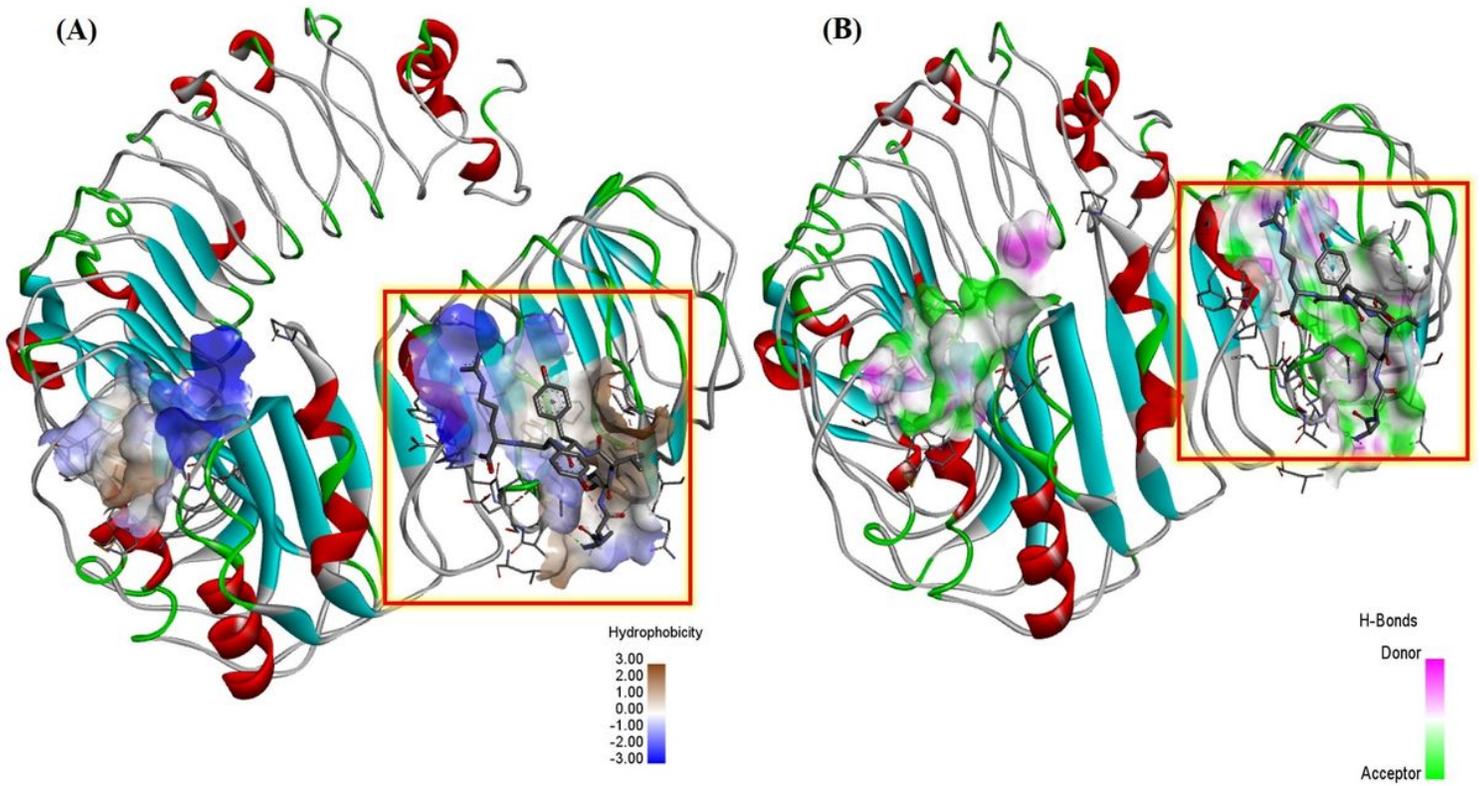
**Figure 1**

Schematic representation of overall work flow employed in the subtractive proteomics approach to identify vaccine candidate protein.



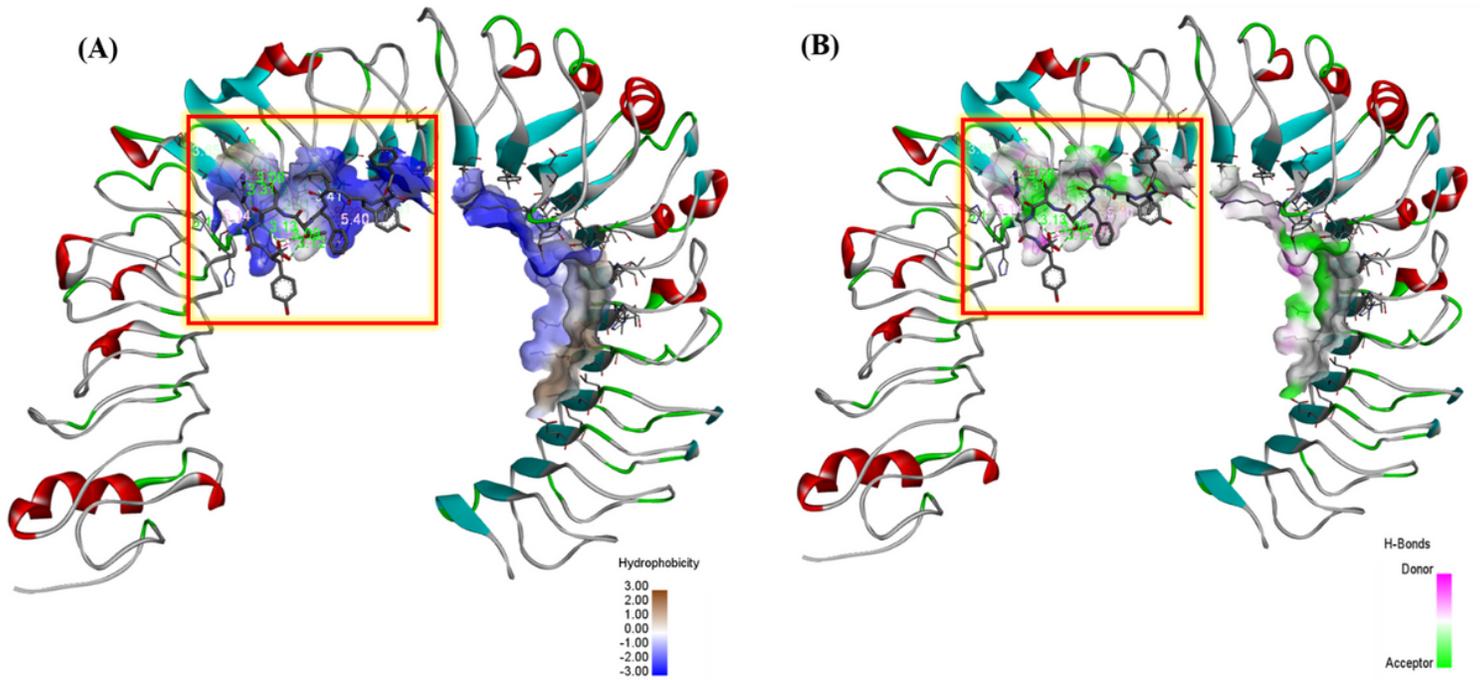
**Figure 2**

Graphical representation of epitopes bases on predicted (A) Surface accessibility, (B) Karplus and Schulz surface flexibility (C) Parker Hydrophilicity Prediction for *H. pylori* IpxF protein.



**Figure 3**

Molecular interaction pattern of the docked PEP\_19 (AGFVYYR) against TLR2 molecule. (A) Hydrophobic interaction. (B) Hydrogen bond formation between peptide and TLR2.



**Figure 4**

Molecular interaction pattern of the docked PEP\_31 (FAYLFTSRY) against TLR2 molecule. (A) Hydrophobic interaction. (B) Hydrogen bond formation between peptide and TLR2.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFileS5.xlsx](#)
- [SupplementaryFileS6.docx](#)
- [SupplementaryFileS7.docx](#)
- [SupplementaryFileS1.xlsx](#)
- [SupplementaryFileS2.xlsx](#)
- [SupplementaryFileS3.xlsx](#)
- [SupplementaryFileS4.xlsx](#)