

In silico identification of drug targets and vaccine candidates against *Bartonella quintana*: a subtractive proteomics approach

Shabir Ahmad, Hugo Verli/+

Universidade Federal do Rio Grande do Sul, Centro de Biotecnologia, Porto Alegre, RS, Brasil

BACKGROUND The availability of genes and protein sequences for parasites has provided valuable information for drug target identification and vaccine development. One such parasite is *Bartonella quintana*, a Gram-negative, intracellular pathogen that causes bartonellosis in mammalian hosts.

OBJECTIVE Despite progress in understanding its pathogenesis, limited knowledge exists about the virulence factors and regulatory mechanisms specific to *B. quintana*.

METHODS AND FINDINGS To explore these aspects, we have adopted a subtractive proteomics approach to analyse the proteome of *B. quintana*. By subtractive proteins between the host and parasite proteome, a set of proteins that are likely unique to the parasite but absent in the host were identified. This analysis revealed that out of the 1197 protein sequences of the parasite, 660 proteins are non-homologous to the human host. Further analysis using the Database of Essential Genes predicted 159 essential proteins, with 28 of these being unique to the pathogen and predicted as potential putative targets. Subcellular localisation of the predicted targets revealed 13 cytoplasmic, eight membranes, one periplasmic, and multiple location proteins. The three-dimensional structure and B cell epitopes of the six membrane antigenic protein were predicted. Four B cell epitopes in KdtA and mraY proteins, three in lpxB and BQ09550, whereas the ftsI and yidC proteins were located with eleven and six B cell epitopes, respectively.

MAINS CONCLUSIONS This insight prioritises such proteins as novel putative targets for further investigations on their potential as drug and vaccine candidates.

Key words: *Bartonella quintana* - subtractive proteomics - drug targets - vaccine candidate

Bartonella quintana is a bacterial pathogen that can infect humans and cause trench fever, first identified during World War I and spread by human body lice (*Pediculus humanus corporis*).⁽¹⁾ Although other arthropods have been associated with transmitting *Bartonella* species, the human body louse is considered the primary vector for trench fever.^(1,2,3) This bacterial infection is associated with bartonellosis, a condition characterised by prolonged bacteraemia and found in a wide range of host reservoirs such as humans, cats, dogs, rabbits, rodents, horses, cattle, and other wild animals.^(1,3,4,5,6) While several *Bartonella* species can cause fever and culture-negative endocarditis in humans and animals, *B. quintana* and *B. henselae* (a causal agent of cat scratch disease) are most commonly associated with human endocarditis cases.^(7,8,9) The three species *B. quintana*, *Bartonella bacilliformis*, and *B. henselae* are the causative agents for human disease, producing a variety of signs from mild symptoms like fever, headache, and malaise to more serious symptoms such as hallucinations.^(4,6,10,11,12) The severity of clinical manifestations is often correlated with the immune status of the patient, and the large spectrum of animal reservoir hosts and arthropod vectors that can transmit these bacteria among animals and humans are major public health concerns.

Co-infection with *B. henselae* in people suffering from persistent symptoms after borreliosis treatment has been reported, and transmission with other pathogens, such as *Borrelia*, transmitted by ticks, can contribute to atypical disease progression.^(9,13,14) As reviewed by Lantos & Wormser, laboratory diagnostics were not performed in most reported cases of *Bartonella* and *Borrelia* co-infections.⁽¹⁵⁾ With different populations exposed to animals and arthropod vectors in various dimensions, veterinarians, veterinary technicians, and zookeepers are at increased risk of infection with *Bartonella* spp.^(15,16,17) *Bartonella* infections are suspected of having contributed to the deaths of two veterinarians in 2013.⁽¹⁵⁾

Recent progress in the field of bioinformatics has generated various *in silico* strategies and drug-designing approaches that reduce the time and cost associated with trial-and-error experiments for drug development.⁽¹⁸⁾ These methods serve to shortlist potential drug targets that may be used for experimental validation.^(19,20,21,22,23) One such approach is subtractive proteomics, an *in silico* method used to identify essential and non-host homologous proteins within a pathogen proteome.^(18,24,25,26,27) By selecting essential proteins unique to pathogen survival and propagation, the subtractive proteomics approach allows the identification of novel drug targets and vaccine candidates within a pathogen, as shown for life-threatening pathogens such as *Pseudomonas aeruginosa*,⁽²⁸⁾ *Streptococcus pneumoniae*⁽²⁹⁾ and *Mycobacterium tuberculosis*.^(30,31)

Inceptive steps in discovering a novel drug target or vaccine candidate include identifying target proteins.⁽³²⁾ However, to our knowledge, no studies have examined *B. quintana* utilising subtractive proteome analy-

doi: 10.1590/0074-02760230040

+ Corresponding author: hugoverli@gmail.com

https://orcid.org/0000-0002-4796-8620

Received 22 February 2023

Accepted 29 February 2024



sis. Therefore, in this study, we employed an approach to identify potential drug targets and vaccine candidates in *B. quintana* by predicting the subcellular localisation of non-host homologous essential proteins, conducting druggability and antigenicity analyses, and predicting the three-dimensional (3D) structure of membrane antigenic proteins using AlphaFold. This study provides a novel insight into potential drug targets and vaccine candidates in *B. quintana*.

MATERIALS AND METHODS

Retrieval of proteomes of host and pathogen - The selection of *B. quintana* str. Toulouse project accession PRJNA44's proteome was based on the organism's name and the availability of its pathways on the KEGG database. The proteome of *B. quintana* (accession no. GCF_000046685.1) and *Homo sapiens* (accession no. GCF_000001405.40) was obtained from the National Centre for Biotechnology Information (NCBI).^(33,34)

Screening of non-host homologous proteins - To eliminate paralogs or duplicates from the *B. quintana* proteome, a 60% sequence identity threshold was applied using Cluster Database at High Identity with Tolerance (CD-HIT)^(27,35) before further analysis. The resulting protein set was then subjected to BLASTp against the *H. sapiens* proteome, using an E-value cut-off of 10^{-5} , as reported by Altschul et al.⁽³⁶⁾ Only the protein sequences of *B. quintana* that exhibited no similarity with *H. sapiens* proteins were considered for subsequent analysis.

Screening of essential proteins in B. quintana - To identify essential proteins in *B. quintana*, a BLASTp analysis was conducted on non-homologous protein sequences using a database of essential proteins downloaded from the Database of Essential Genes (DEG), as reported by Luo et al.⁽³⁷⁾ The analysis utilised an E-value cut-off score of 10^{-5} and $\geq 30\%$ identity, while a minimum bit-score cut-off of 100 was used to include proteins that could potentially represent essential genes. The resulting protein sequences represent the non-homologous essential proteins of *B. quintana*.

Metabolic pathway analysis - The essential proteins of *B. quintana* were subjected to metabolic pathway analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the functional annotation of the genes was identified using the Automatic Annotation Server (KAAS). The KAAS analysis was performed using the BLAST option and the assignment method was used as a bi-directional best hit (BBH), which provided KEGG Orthology (KO) codes indicating the presence of particular proteins in the specific pathways of the pathogen. Comparative analysis of the metabolic pathways of the host and pathogen proteins was carried out using the KEGG pathway database, as reported by.^(38,39) The aim was to identify unique proteins involved in pathogen-specific metabolic pathways, which could serve as potential therapeutic targets.

Subcellular localisation and sequence conservation analysis of unique essential proteins - The analysis of unique essential proteins was performed using two online servers: CELLO v2.5 and QuickGo. CELLO v2.5,

with a reported accuracy of up to 89% based on a training dataset of 1443 proteins,⁽⁴⁰⁾ was used to identify the location of essential proteins in various cellular organelles such as the outer membrane, cytoplasm, and extracellular space. The analysis was performed using the parameters of the organism as gram-negative and the sequence parameters as proteins. Additionally, QuickGo, with an accuracy of up to 65% based on training datasets containing over 1000 proteins,⁽⁴¹⁾ was utilised for the same purpose. These tools provided valuable insights into the subcellular localisation of essential proteins in the study. These essential target proteins were used to map these proteins to the 25 genomes of *B. quintana* strains available on the NCBI website and were analysed using tBLASTn. The tBLASTn is a variant of the BLAST algorithm, which compares a query consisting of an amino acid sequence with the translated nucleotide (coding sequence) database. The unique essential target protein sequences were uploaded as a query using a database search as RefSeq Genome Database (refseq_genomes) against the organism *B. quintana* (taxid:803). Protein sequence searches are performed on all six translated frames in the database to enable protein-to-protein comparisons. To perform multiple sequence alignment (MSA), the amino acid sequences of the target protein were downloaded and then uploaded to CLUSTAL W.⁽⁴²⁾ The sequences that had been aligned were subsequently imported into ESPript 3.0 (esprpt.ibcp.fr/ESPript/ESPript)⁽⁴³⁾ in order to map sequence conservation.

Evaluation of druggability potential of the unique essential proteins - To identify potential drug targets in *B. quintana*, the essential proteins associated with unique pathways were subjected to BLASTp analysis with a bit score > 100 and E value $< 10^{-3}$ against a customised database of Food and Drug Administration (FDA)-approved drug targets, as reported by.⁽⁴⁴⁾ The protein targets that exhibited BLASTp hits against the FDA-approved drug target database were considered druggable targets.

Conversely, protein targets that did not exhibit any matches with the FDA-approved drug target database were considered novel targets for new drug identification. These novel targets offer the potential for the discovery of new drugs that can target unique metabolic pathways in *B. quintana*, thus enhancing the current arsenal of drugs available for the treatment of infections caused by this pathogen.

Analysis of virulence factors (VFs) - The Virulence Factor Database (VFDB) was employed to identify VFs in *B. quintana*. The VFDB is a comprehensive database comprising four categories of VFs, namely offensive, defensive, non-specific, and virulence-associated regulated proteins, from 25 pathogenic bacteria, as reported by.⁽⁴⁵⁾

To identify potential VFs in *B. quintana*, the non-homologous essential proteins were subjected to a BLASTp search against the database of protein sequences from the VFDB core dataset (R1), with a cut-off E-value of 10^{-5} . This approach allowed us to identify *B. quintana* proteins that exhibit significant sequence similarity to known VFs, which can aid in understanding the virulence mechanisms of this pathogen and identify potential targets for the development of new therapeutics.

Antigenicity analysis and 3D structure prediction - We employed the VaxiJen server developed for the antigenicity prediction of the membrane proteins. The VaxiJen server is a computational tool for predicting protein sequences' antigenicity. It uses auto-cross covariance (ACC) transformation followed by a probabilistic neural network (PNN) algorithm to classify proteins as either antigenic or non-antigenic. This server has an accuracy range of 70% to 89% for antigenicity prediction. Proteins with scores > 0.4 were considered antigenic, and those with scores < 0.4 were labelled non-antigenic, leading to further investigation of antigenic proteins.⁽⁴⁶⁾

Due to the absence of crystal structure data for the predicted antigenic membrane target proteins, we utilised AlphaFold⁽⁴⁷⁾ and ColabFold⁽⁴⁸⁾ to predict their 3D models using the parameters number of relax (num_relax = 1), template mode PDB100, number of cycles (num_recycles = 3), and number of seeds (num_seeds = 1). These tools employ a threading technique that considers protein folding rather than similarities to generate models of the proteins.

B cell epitope prediction - To predict B cell linear epitopes, we employed three algorithms: BCPred,⁽⁴⁹⁾ which uses a default epitope length of 14 amino acids and reports only non-overlapping epitopes with over 75% specificity; ABCPred,⁽⁵⁰⁾ which uses an epitope length of 14 amino acids and reports only non-overlapping epitopes above the 0.8 thresholds; and BepiPred,⁽⁵¹⁾ which requires a minimum threshold of 0.5 and at least six amino acids to identify an epitope.

Amino acid sequences identified as probable epitopes by at least two algorithms were then subjected to antigenicity analysis using VaxiJen v.2.0.⁽⁴⁶⁾ VaxiJen classifies antigens based on their physicochemical features, and the default recommendation threshold for analysing the sequence of parasite proteins was 0.5.

The presence of discontinuous epitopes in the 3D models of the predicted membrane proteins was evaluated using two different prediction tools. ElliPro⁽⁵²⁾ employs Thornton's method along with residue clustering algorithms to predict epitopes with a minimum score of 0.5 and a maximum distance of six amino acids. The second tool used was DiscoTope,⁽⁵³⁾ which calculates the surface accessibility of epitope fragments and uses the standard threshold of -3.7 in this analysis.

RESULTS

Identification of non-host homologous and essential proteins analysis - The proteome of *B. quintana* was analysed in this study. Initially, 1197 proteins were retrieved from NCBI, and those with an amino acid sequence of fewer than 100 residues were removed, resulting in 1,026 non-redundant proteins according to the stepwise flowchart as shown in (Fig. 1). BLASTp analysis was conducted against the human proteome and found 660 proteins that were non-homologous to *H. sapiens*. The remaining 336 proteins were excluded. The 660 proteins were then subjected to DEG analysis to find the essential genes needed for the survival of the pathogen. This yielded 159 hits that were involved in structural organisation, nutrient uptake, pathogenesis, and other essential processes for *B. quintana* survival (Table I).

Metabolic pathway, subcellular localisation, and sequence conservation analysis - 159 essential proteins were analysed to determine their involvement in metabolic pathways. Only 25 pathways specific to *B. quintana* and thus not present in humans were found, possessing 28 unique proteins non-homologous to humans. The pathways that are unique to *B. quintana* with the names of essential genes and their KO, have been presented in (Table II). Segregation of these 28 proteins into their metabolic pathways revealed that four proteins of them belong to the two-component system: 10 proteins in the peptidoglycan biosynthesis pathway, three proteins in the quorum-sensing pathway, five proteins in the lipopolysaccharide biosynthesis pathway, and six in the lysine biosynthesis pathway. Similarly, there are five proteins in the vancomycin pathway, one protein in the phosphotransferase system (PTS), one protein in the bacterial secretion system, one in beta-lactam resistance, and one in monobactam biosynthesis. Since these target proteins are essential and unique to *B. quintana*, their suppression will render bacteria more vulnerable to various medications, and proteins implicated in these pathways make better therapeutic targets (Table II, Fig. 2).

The subcellular localisation analysis of the 28 therapeutic targets revealed that 13 proteins were cytoplasmic, eight were membrane, one was periplasmic, and six had multiple locations. Some of the many functions of bacterial secreted proteins include acting as immunogens, toxins, or virulence factors.⁽⁵⁴⁾ Of the 28 final targets, not a single one was predicted as a secretory protein (Fig. 2). These protein targets were mapped across the 25 available strains of *B. quintana* (taxid:803) by analysing the 28 therapeutic targets of *B. quintana* str. Toulouse with NCBI tBLASTn. We analysed the tBLASTn strain data, and the amino acid sequences were downloaded and imported into CLUSTAL W to find sequence conversations across all the strains. To map the similarity information of the therapeutic targets across the *B. quintana* strains, the multiple sequence aligned file was subsequently uploaded to ESPript 3.0. The target proteins such as gene lpxD (accession no: WP_011179438.1), gene mraY (accession no: WP_011179606.1), and gene murB (accession no: WP_034449436.1) in all of the strains share the gene name and the accession number, showing 100% sequence conservation in the multiple sequence alignment. Residues that are conserved between the strain sequences and the template proteins (*B. quintana* str. Toulouse) are shaded red. While the unshaded residue is not conserved across the template and strain sequences [see Supplementary data (Fig. 1)].

Druggability of unique essential proteins - The cytoplasmic, periplasmic, and multiple localisation essential proteins involved in a distinctive metabolic pathway in *B. quintana* were subjected to further investigation to assess their druggability. The analysis showed that seven proteins of the shortlisted 20 unique, essential cytoplasmic, periplasmic, and multiple localisation proteins are druggable according to the FDA-approved DrugBank databases, while the remaining 13 proteins were predicted to be novel druggable targets (Table II).

Virulence factor and antigenicity analysis of the vaccine candidate - Further, the VFs for the 28 essential unique proteins of *B. quintana* were analysed, and 10 potential VFs were identified including the cytoplasmic and membrane proteins (Table II). Pathogen virulence-related membrane proteins have been shown to play an important role in adhesion and survival in host cells.⁽⁵⁵⁾ Potential antigen candidates are membrane proteins, which are typically immunogenic and exposed on the cell surface. VaxiJen was used to as-

sess the antigenicity of eight membrane proteins, and six proteins with an expected threshold value greater than 0.4 were considered antigenic. The remaining two proteins were deemed non-antigenic due to their values falling below the cut-off threshold.

B cell epitope and 3D structure prediction - To predict B cell epitopes, *in silico* algorithms were used for the six membrane antigenic proteins. Linear epitope prediction algorithms (ABCPred, BCPred, and BepiPred) were employed to generate consensus sequences across

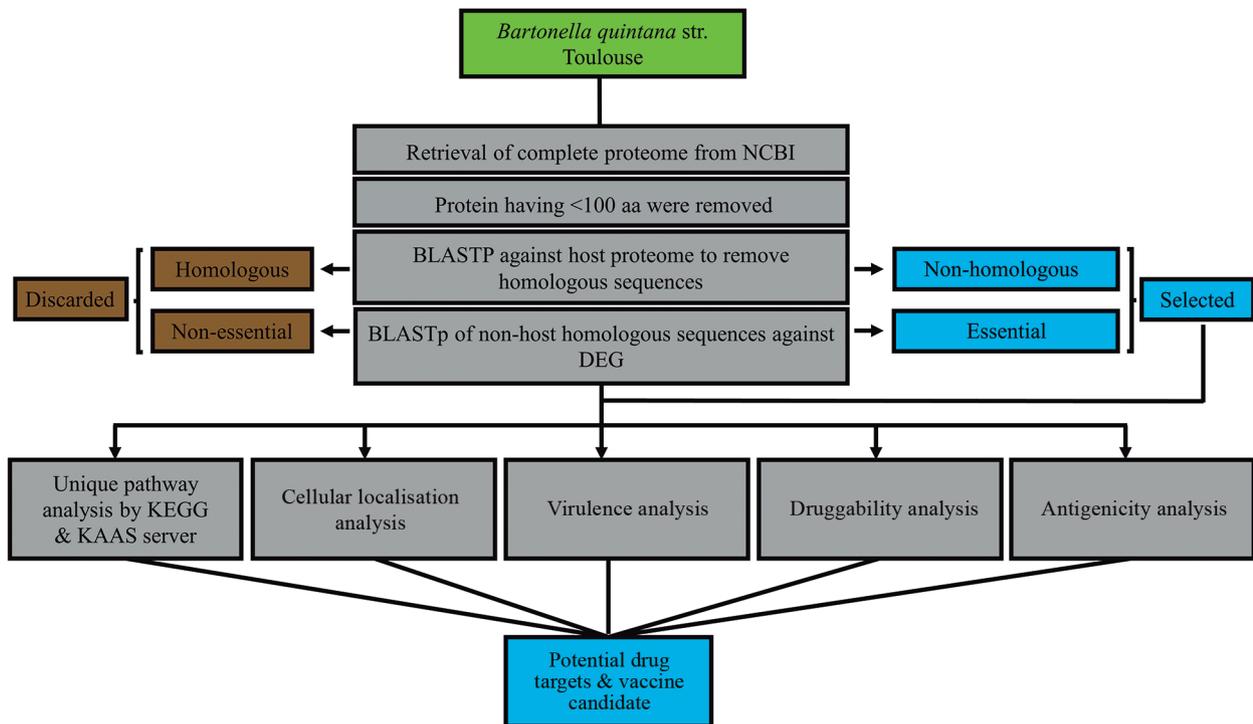


Fig. 1: stepwise analysis flowchart for the subtractive proteomics analysis of *Bartonella quintana*. The green colour box consists of the pathogen name (*B. quintana* str. Toulouse), which is the starting step of the analysis. The grey colour box represents the methodological aspects at each analysis step. The gold colour box represents the homologous and non-essential data and is discarded. The light blue colour box represents the essential non-homologous and therapeutic targets.

TABLE I
Classification of *Bartonella quintana* proteins using subtractive proteome analysis

SN	Step	No. of proteins in <i>B. quintana</i> str. Toulouse
1	Total proteome	1197
2	Non-paralogs (> 60% identical) in CD-HIT	1026
3	Non-homologs	660
4	Essential proteins in DEG	159
5	Unique metabolic pathway KEGG	25
6	Essential proteins involved in unique pathways	28
7	Druggable targets with a cut-off E-value of 10^{-5}	7
8	Novel targets with a cut-off E-value of 10^{-5}	13
9	Antigenic vaccine candidate	6

CD-HIT: Cluster Database at High Identity with Tolerance; DEG: Database of Essential Genes; KEGG: Kyoto Encyclopedia of Genes and Genomes; SN: serial number.

TABLE II
Non-homologous, essential, and unique proteins involved in twenty-five unique metabolic pathways and their druggability

SN	Accession no.	Gene name	KO list	Pathway's name	DrugBank	Subcellular localisation	QuickGo	Virulence factor	Antigenicity
1	WP_011178881.1	<i>dapE</i>	K01439	Lysine biosynthesis	Druggable	Cytoplasmic	N/A	Non-virulent	-
2	WP_011178891.1	<i>BQ00560</i>	K07658	Two-component system	Novel	Cytoplasmic	N/A	Virulent	-
3	WP_011178913.1	<i>yfeA</i>	K11601	Two-component system	Novel	Multiple	N/A	Virulent	-
4	WP_011178939.1	<i>dnaA</i>	K02313	Two-component system	Novel	Cytoplasmic	Cytoplasm	Non-virulent	-
5	WP_011178959.1	<i>fruB</i>	K08483	Phosphotransferase system (PTS)	Novel	Multiple	Cytoplasm	Non-virulent	-
6	WP_011179066.1	<i>kdtA</i>	K02527	Lipopolysaccharide biosynthesis	-	Membrane	Membrane	Virulent	Antigenic
7	WP_011179067.1	<i>lpxK</i>	K00912	Lipopolysaccharide biosynthesis	Novel	Multiple	N/A	Virulent	Non-antigenic
8	WP_011179434.1	<i>lpxB</i>	K00748	Lipopolysaccharide biosynthesis	-	Membrane	N/A	Virulent	Antigenic
9	WP_011179438.1	<i>lpxD</i>	K02536	Lipopolysaccharide biosynthesis	Druggable	Cytoplasmic	N/A	Virulent	-
10	WP_011179451.1	<i>dacA2</i>	K07258	Peptidoglycan biosynthesis	Druggable	Periplasmic	N/A	Non-virulent	-
11	WP_011179558.1	<i>ybeJ</i>	K10001	Two-component system	Novel	Multiple	N/A	Virulent	-
12	WP_011179596.1	<i>lpxC</i>	K02535	Lipopolysaccharide biosynthesis	Novel	Cytoplasmic	N/A	Virulent	-
13	WP_011179600.1	<i>ddl</i>	K01921	Peptidoglycan biosynthesis, vancomycin resistance	Druggable	Cytoplasmic	Cytoplasm	Non-virulent	-
14	WP_011179602.1	<i>murC</i>	K01924	Peptidoglycan biosynthesis, vancomycin resistance	Novel	Cytoplasmic	Cytoplasm	Non-virulent	-
15	WP_011179603.1	<i>murG</i>	K02563	Peptidoglycan biosynthesis, vancomycin resistance	-	Membrane	Membrane	Non-virulent	Non-antigenic
16	WP_011179606.1	<i>mraY</i>	K01000	Peptidoglycan biosynthesis, vancomycin resistance	-	Membrane	Membrane	Virulent	Antigenic
17	WP_011179607.1	<i>murF</i>	K01929	Lysine biosynthesis, peptidoglycan biosynthesis, vancomycin resistance	Novel	Multiple	Cytoplasm	Non-virulent	-
18	WP_011179608.1	<i>murE</i>	K01928	Lysine biosynthesis, peptidoglycan biosynthesis	Novel	Cytoplasmic	Cytoplasm	Non-virulent	-
19	WP_011179609.1	<i>fisI</i>	K03587	Peptidoglycan biosynthesis, beta-lactam resistance	-	Membrane	Membrane	Non-virulent	Antigenic
20	WP_011179659.1	<i>BQ09550</i>	K02034	Quorum sensing	-	Membrane	Membrane	Non-virulent	Antigenic
21	WP_011179660.1	<i>BQ09560</i>	K02033	Quorum sensing	-	Membrane	Membrane	Non-virulent	Antigenic
22	WP_011179677.1	<i>yidC</i>	K03217	Quorum sensing, bacterial secretion system	-	Membrane	Membrane	Non-virulent	Antigenic
23	WP_011179680.1	<i>dapB</i>	K00215	Lysine biosynthesis, monobactam biosynthesis	Novel	Cytoplasmic	Cytoplasm	Non-virulent	-
24	WP_011179695.1	<i>alr</i>	K01775	Vancomycin resistance	Druggable	Multiple	N/A	Non-virulent	-
25	WP_011179784.1	<i>murA</i>	K00790	Peptidoglycan biosynthesis	Druggable	Cytoplasmic	Cytoplasm	Non-virulent	-
26	WP_011179951.1	<i>BQ13390</i>	K00290	Lysine biosynthesis	Novel	Cytoplasmic	N/A	Non-virulent	-
27	WP_011179960.1	<i>dapF</i>	K01778	Lysine biosynthesis	Novel	Cytoplasmic	Cytoplasm	Non-virulent	-
28	WP_034449436.1	<i>murB</i>	K00075	Peptidoglycan biosynthesis	Druggable	Cytoplasmic	Cytoplasm	Virulent	-

KO list: Kyoto Encyclopedia of Genes and Genomes Orthology list; N/A: not available; SN: serial number.

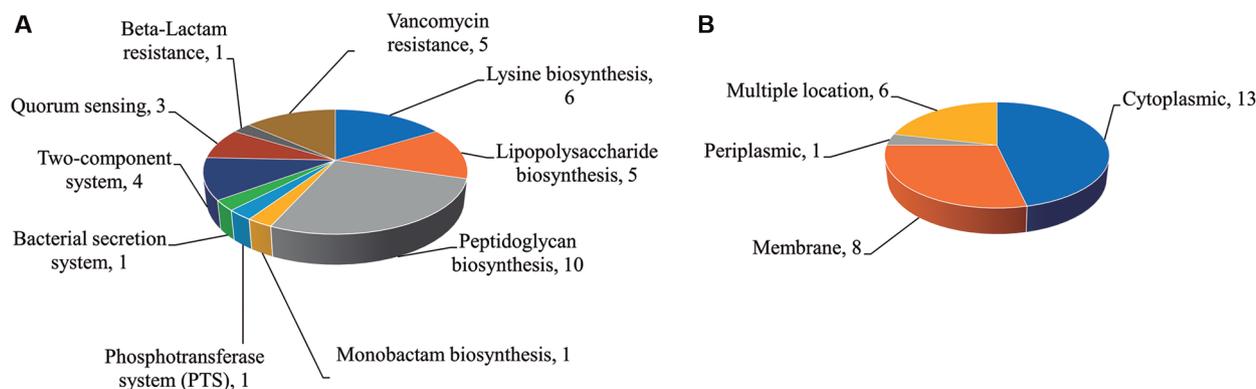


Fig. 2: distribution of non-homologous essential therapeutic targets. (A) Unique essential proteins are present in each unique metabolic pathway; (B) Subcellular localisation of therapeutic target proteins show the majority are cytoplasmic.

all evaluated proteins. Four B cell epitopes were identified within the KdtA (IRGKEEWNRKKE, TIIVPRH-PERSED, RRSNAIPARD, EALRQEMVDK) and mraY (ROGKGQPIRDGPQ, KVTKQTEKG, FVKDY-FINLS, DYLQIHVYVSG) proteins. LpxB (RPERAR, PIVPEYFNE, QKMKTEVPP) and BQ09550 (AHLQPPSM, IFLTTI, LGAQLPSPE) were found to have three B cell epitopes each, whereas the ftsI (GQIEE-AKGGPVL, GIPGIGF, QASSGYII, DFHGKNRPL, PLQTAVGAAA, RTKEQALKQAKQV, MRYLYKL, GRNAKVE, KVENGKYSKT, PKPEDGKYAA, IKP-DFKKEYDSIL) and yidC (LKKYRLTVDKKS, EG-NNTTLTPSTPVT, DSLKTEKYKTLA, EKYPDDRT-KQQQA, MRMNPAPQDQT) proteins were mapped with eleven and six B cell epitopes, respectively [see Supplementary data (Fig. 2)]. The interaction of these epitopes with B cells is heavily influenced by their three-dimensional configuration. This interaction is heavily influenced by the spatial arrangement of amino acids in the epitope region. A linear or two-dimensional representation of the antigenic proteins cannot adequately capture the complexity of these interactions. A comprehensive 3D model is necessary to capture the accurate conformation of these epitopes. Given this requirement, to map B cell epitopes, the 3D structures of the final six selected antigenic proteins were predicted using AlphaFold. This approach was chosen due to the unavailability of crystal structures for these proteins. These models provided a clear understanding of the spatial orientation and configuration of potential B cell epitopes. Discontinuous epitope analyses (ElliPro and DiscoTope) revealed the presence of sequences of interest in all targets (Fig. 3). The results indicated that continuous epitopes are related to discontinuous epitopes and that multiple prediction tools should be utilised, as combining linear and conformational predictions improves the results.⁽⁵⁶⁾

DISCUSSION

In this study, the *B. quintana* proteome was analysed to identify its distinctive essential and virulent proteins. A comparison of *B. quintana*'s pathways with those of its host (human) revealed the presence of 28 unique metabolic pathways. These pathways included peptidogly-

can biosynthesis, lipopolysaccharide biosynthesis, two-component system, lysine biosynthesis, Vancomycin resistance pathway, and the phosphotransferase system (PTS). Among these pathways, the peptidoglycan pathway is unique to bacteria, and inhibitors of this pathway are currently employed as antibiotics. Drugs target that peptidoglycan biosynthesis may reduce the pathogenicity of microorganisms.⁽⁵⁷⁾ Additionally, the two-component system according to,⁽⁵⁸⁾ plays a crucial role in the adaptation of bacteria to environmental and intracellular changes. Simultaneously, the lipopolysaccharide biosynthesis pathway is an essential metabolic pathway in *B. quintana*, providing stability to the outer membrane of Gram-negative bacteria by linking a conserved core oligosaccharide, lipid A, to a variable O-antigen.⁽⁵⁹⁾ We found 13 non-homologous, unique essential proteins that can be used as drug targets and the virtual screening against these novel targets might be useful in the discovery of novel therapeutic compounds against *B. quintana*. Additionally, we found six non-homologous, unique, and essential proteins that can be used as vaccine candidates. We located linear B cell epitope in the membrane antigenic proteins and revealed four B cell epitopes in KdtA and mraY proteins, three in lpxB and BQ09550, whereas the ftsI and yidC proteins were mapped with eleven and six B cell epitopes, respectively. A comparative analysis of the predicted antigenic epitopes region in the six vaccine candidate of *B. quintana* str. Toulouse with 24 genomes of *B. quintana* str. shows similarity, except for one epitopes in the ftsI proteins (RTKEQALKQAKQV), where the underlined Q amino acid is replaced with K in only one of the *B. quintana* str. MF1.1 (RTKEQALKQA-KK~~V~~). Formulations based on epitopes are currently considered one of the main strategies in vaccine development due to their cost and time benefit in process optimisation.⁽⁶⁰⁾ One of the main advantages of this method is the ability to focus the immune response on crucial epitopes, avoiding the production of antibodies against regions of no interest.⁽⁶¹⁾ Additionally, discontinuous B cell epitopes were identified in the antigenic membrane proteins and mapped using the 3D structure of the antigenic proteins. We used AlphaFold to predict the 3D structure of these vaccine candidates, as there is no crys-

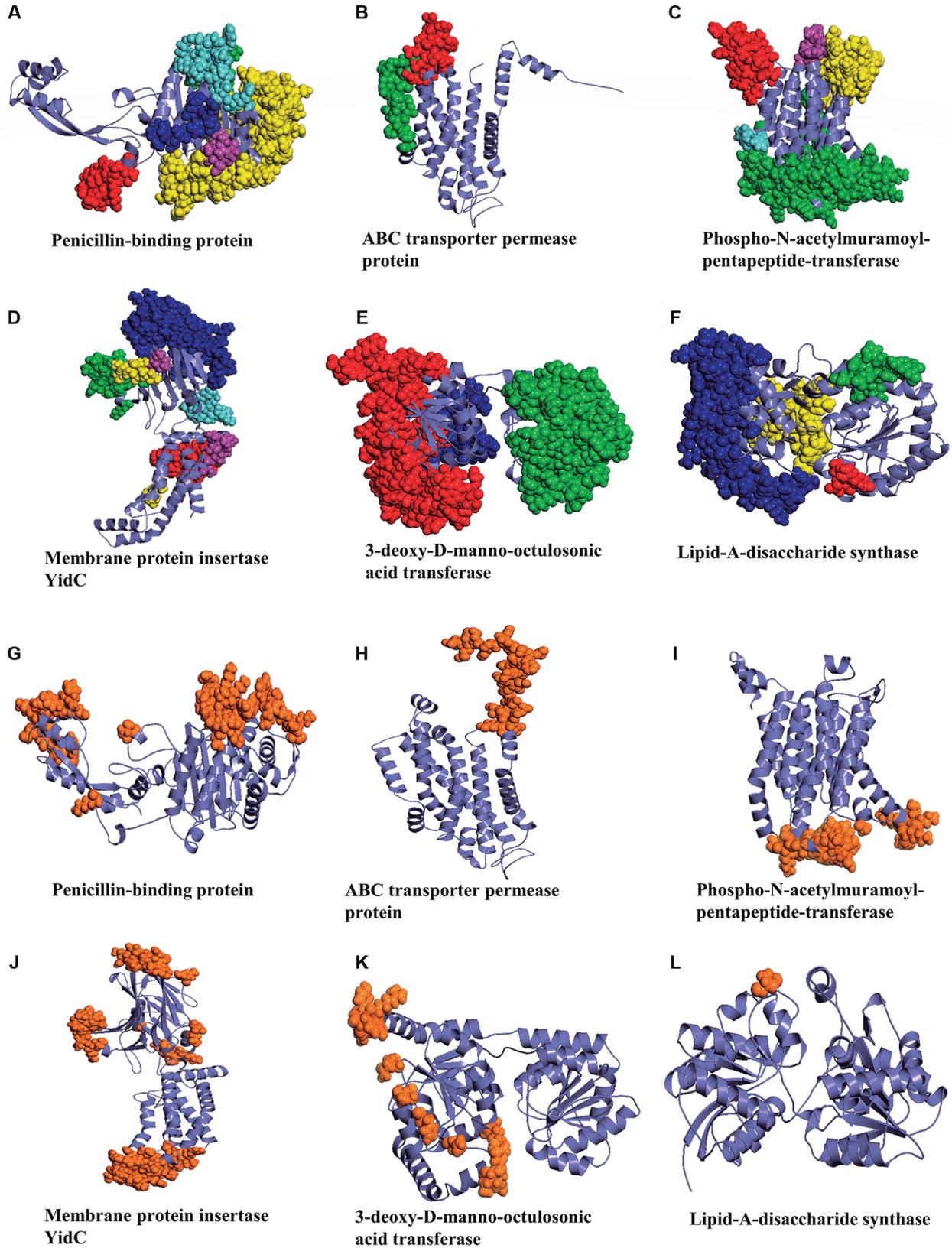


Fig. 3: the discontinuous B cell epitope predicted using ElliPro. The coloured sphere shows the B cell epitope. The sphere is coloured according to the order of the PyMOL colour code Red > Green > Blue > Yellow > Magenta > Cyan. Red colours show the highest score of epitopes from A to F at the top of the figure. Similarity: discontinuous B cell epitopes predicted using DiscoTope are shown as brown from G to L at the bottom of the figure.

tal structure available for these vaccine candidates. The predicted 3D structures provided a great aid in studying protein functions, dynamics, ligand interactions, and other components.⁽⁶²⁾ By obtaining the structural information of these antigenic proteins, we have provided insights into their 3D organisation, the understanding of their functional properties, and the design of targeted interventions. Altogether, the present study has identified 28 essential non-homologous proteins in *B. quintana* that can be targeted for effective therapeutics. These proteins hold promise for further experimental validation to assess their suitability as therapeutic targets.

In conclusion - The subtractive proteomics approach employed in this study has provided a promising strategy for identifying potential drug targets and vaccine candidates against *B. quintana*. By scrutinising essential and unique proteins within the *B. quintana* proteome, we have identified 28 therapeutic targets. The prioritised list of targets generated by this method can serve as a valuable resource for the development of effective therapeutics against *B. quintana* and potentially other pathogens. Moreover, the targets identified in this study have not been previously characterised as drug targets against *B. quintana*, and therefore hold great potential for the discovery of new and innovative drug compounds. Overall, our findings highlight the utility of subtractive proteomics as a rapid and effective approach for identifying potential drug targets and vaccine candidates against pathogenic bacteria.

AUTHORS' CONTRIBUTION

SA designed the study and performed various *in silico* analyses; SA and HV wrote and revised the manuscript. All the authors performed critical revisions and approved the final version. Both authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declare that they have no conflicts/competing interests.

REFERENCES

- Angelakis E, Raoult D. Pathogenicity and treatment of *Bartonella* infections. *Int J Antimicrob Agents*. 2014; 44(1): 16-25.
- Chomel BB, Boulouis HJ, Breitschwerdt EB, Kasten RW, Vaysier-Taussat M, Birtles RJ, et al. Ecological fitness and strategies of adaptation of *Bartonella* species to their hosts and vectors. *Vet Res*. 2009; 40(2): 1-22.
- Ben-Tekaya H, Gorvel JP, Dehio C. *Bartonella* and *Brucella* — weapons and strategies for stealth attack. *Cold Spring Harb Perspect Med*. 2013; 3(8): a010231.
- Iannino F, Salucci S, Di Provvido A, Paolini A, Ruggieri E. *Bartonella* infections in humans dogs and cats. *Vet Ital*. 2018; 54(1): 63-72.
- Álvarez-Fernández A, Breitschwerdt EB, Solano-Gallego L. *Bartonella* infections in cats and dogs including zoonotic aspects. *Parasit Vectors*. 2018; 11(1): 1-21.
- Cheslock MA, Embers ME. Human bartonellosis: an underappreciated public health problem? *Trop Med Infect Dis*. 2019; 4(2): 69.
- da Cruz Lamas C, Ramos RG, Lopes GQ, Santos MS, Golebiovski WF, Weksler C, et al. *Bartonella* and *Coxiella* infective endocarditis in Brazil: molecular evidence from excised valves from a cardiac surgery referral center in Rio de Janeiro, Brazil, 1998 to 2009. *Int J Infect Dis*. 2013; 17(1): e65-6.
- Guiyedi V, Haddad H, Okome-Nkoumou M, Gire F, Ongali B, Lore P, et al. Cat-scratch disease in adult hospitalized for prolonged-fever associated with multiple lymphadenopathies and weight loss. *Open Microbiol J*. 2013; 7: 152.
- Regier Y, O'Rourke F, Kempf VA. *Bartonella* spp. — a chance to establish One Health concepts in veterinary and human medicine. *Parasit Vectors*. 2016; 9(1): 1-12.
- Breitschwerdt EB. Bartonellosis, One Health and all creatures great and small. *Adv Vet Dermatol*. 2017; 8: 111-21.
- Okaro U, Addisu A, Casanas B, Anderson B. *Bartonella* species, an emerging cause of blood-culture-negative endocarditis. *Clin Microbiol Rev*. 2017; 30(3): 709-46.
- Kalogeropoulos D, Asproudis I, Stefanidou M, Moschos MM, Mentis A, Malamos K, et al. *Bartonella henselae*-and *quintana*-associated uveitis: a case series and approach of a potentially severe disease with a broad spectrum of ocular manifestations. *Int Ophthalmol*. 2019; 39(11): 2505-15.
- Belongia EA. Epidemiology and impact of coinfections acquired from *Ixodes* ticks. *Vector-Borne Zoonotic Dis*. 2002; 2(4): 265-73.
- Holden K, Boothby JT, Kasten RW, Chomel BB. Co-detection of *Bartonella henselae*, *Borrelia burgdorferi*, and *Anaplasma phagocytophilum* in *Ixodes pacificus* ticks from California, USA. *Vector-Borne Zoonotic Dis*. 2006; 6(1): 99-102.
- Lantos PM, Maggi RG, Ferguson B, Varkey J, Park LP, Breitschwerdt EB, et al. Detection of *Bartonella* species in the blood of veterinarians and veterinary technicians: a newly recognized occupational hazard? *Vector-Borne Zoonotic Dis*. 2014; 14(8): 563-70.
- Maggi RG, Mascarelli PE, Pultorak EL, Hegarty BC, Bradley JM, Mozayani BR, et al. *Bartonella* spp. bacteremia in high-risk immunocompetent patients. *Diagn Microbiol Infect Dis*. 2011; 71(4): 430-7.
- Oteo JA, Maggi R, Portillo A, Bradley J, García-Álvarez L, San-Martín M, et al. Prevalence of *Bartonella* spp. by culture, PCR and serology, in veterinary personnel from Spain. *Parasit Vectors*. 2017; 10: 1-9.
- Barh D, Tiwari S, Jain N, Ali A, Santos AR, Misra AN, et al. *In silico* subtractive genomics for target identification in human bacterial pathogens. *Drug Dev Res*. 2011; 72(2): 162-77.
- Schirmer EC, Florens L, Guan T, Yates III JR, Gerace L. Nuclear membrane proteins with potential disease links found by subtractive proteomics. *Science*. 2003; 301(5638): 1380-2.
- Bergmann-Leitner ES, Chaudhury S, Steers NJ, Sabato M, Delvecchio V, Wallqvist AS, et al. Computational and experimental validation of B and T-cell epitopes of the *in vivo* immune response to a novel malarial antigen. *PLoS One*. 2013; 8(8): e71610.
- Yahya MFZR, Alias Z, Karsani SA. Subtractive protein profiling of *Salmonella typhimurium* biofilm treated with DMSO. *Protein J*. 2017; 36: 286-98.
- Kumar S, Mandal RS, Bulone V, Srivastava V. Identification of growth inhibitors of the fish pathogen *Saprolegnia parasitica* using *in silico* subtractive proteomics, computational modeling, and biochemical validation. *Front Microbiol*. 2020; 11: 571093.
- Li J, Qiu J, Huang Z, Liu T, Pan J, Zhang Q, et al. Reverse vaccinology approach for the identifications of potential vaccine candidates against *Salmonella*. *Int J Med Microbiol*. 2021; 311(5): 151508.
- Hosen MI, Tanmoy AM, Mahbuba DA, Salma U, Nazim M, Islam MT, et al. Application of a subtractive genomics approach for *in silico* identification and characterization of novel drug targets in *Mycobacterium tuberculosis* F11. *Interdiscip Sci Comput Life Sci*. 2014; 6(1): 48-56.

25. Solanki V, Tiwari V. Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against *Acinetobacter baumannii*. *Sci Rep*. 2018; 8(1): 1-19.
26. Ali A, Ahmad S, Wadood A, Rehman AU, Zahid H, Khan MQ, et al. Modeling novel putative drugs and vaccine candidates against tick-borne pathogens: a subtractive proteomics approach. *Vet Sci*. 2020; 7(3): 129.
27. Ali A, Ahmad S, de Albuquerque PMM, Kamil A, Alshammari FA, Alouffi A, et al. Prediction of novel drug targets and vaccine candidates against human lice (Insecta), acari (Arachnida), and their associated pathogens. *Vaccines*. 2021; 10(1): 8.
28. Sakharkar KR, Sakharkar MK, Chow VT. A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa*. *In Silico Biol*. 2004; 4(3): 355-60.
29. Singh S, Malik BK, Sharma DK. Metabolic pathway analysis of *S. pneumoniae*: an in silico approach towards drug-design. *J Bioinform Comput Biol*. 2007; 5(01): 135-53.
30. Anishetty S, Pulimi M, Pennathur G. Potential drug targets in *Mycobacterium tuberculosis* through metabolic pathway analysis. *Comput Biol Chem*. 2005; 29(5): 368-78.
31. Asif SM, Asad A, Faizan A, Anjali MS, Arvind A, Neelesh K, et al. Dataset of potential targets for *Mycobacterium tuberculosis* H37Rv through comparative genome analysis. *Bioinformation*. 2009; 4(6): 245.
32. Chan JN, Nislow C, Emili A. Recent advances and method development for drug target identification. *Trends Pharmacol Sci*. 2010; 31(2): 82-8.
33. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez gene: gene-centered information at NCBI. *Nucleic Acids Res*. 2007; 35(Suppl. 1): D26-31.
34. Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, et al. The NCBI biosystems database. *Nucleic Acids Res*. 2010; 38(Suppl. 1): D492-6.
35. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006; 22(13): 1658-9.
36. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215(3): 403-10.
37. 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. *Nucleic Acids Res*. 2014; 42(D1): D574-80.
38. Fukuda M, Asano S, Nakamura T, Adachi M, Yoshida M, Yanagida M, et al. CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature*. 1997; 390(6657): 308-11.
39. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016; 44(D1): D457-62.
40. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. *Proteins*. 2006; 64(3): 643-51.
41. Binns D, Dimmer E, Huntley R, Barrell D, O'donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*. 2009; 25(22): 3045-6.
42. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994; 22(22): 4673-80.
43. Gouet P, Courcelle E, Stuart DI, Métoz F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*. 1999; 15(4): 305-8.
44. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. Drug-Bank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res*. 2010; 39(Suppl. 1): D1035-41.
45. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res*. 2019; 47(D1): D687-92.
46. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007; 8(1): 4.
47. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021; 596(7873): 583-9.
48. Mirdita M, Schütze K, Moriawaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods*. 2022; 19(6): 679-82.
49. EL-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. *J Mol Recognit Interdiscip J*. 2008; 21(4): 243-55.
50. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins Struct Funct Bioinforma*. 2006; 65(1): 40-8.
51. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res*. 2017; 45(W1): W24-9.
52. Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*. 2008; 9(1): 1-8.
53. Andersen PH, Nielsen M, Lund OLE. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci*. 2006; 15(11): 2558-67.
54. Mukherjee P, Mani S. Methodologies to decipher the cell secretome. *Biochim Biophys Acta*. 2013; 1834(11): 2226-32.
55. Xing X, Bi S, Fan X, Jin M, Liu W, Wang B. Intranasal vaccination with multiple virulence factors promotes mucosal clearance of *Streptococcus suis* across serotypes and protects against meningitis in mice. *J Infect Dis*. 2019; 220(10): 1679-87.
56. Assis LM de, Sousa JR, Pinto NFS, Silva AA, Vaz AFM, Andrade PP, et al. B-cell epitopes of antigenic proteins in *Leishmania infantum*: an *in silico* analysis. *Parasite Immunol*. 2014; 36(7): 313-23.
57. Belete TM. Novel targets to develop new antibacterial agents and novel alternatives to antibacterial agents. *Hum Microbiome J*. 2019; 11: 100052.
58. Liu C, Sun D, Zhu J, Liu W. Two-component signal transduction systems: a major strategy for connecting input stimuli to biofilm formation. *Front Microbiol*. 2019; 9: 3279.
59. Deutscher J, Francke C, Postma PW. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev*. 2006; 70(4): 939-1031.
60. Parvizpour S, Pourseif MM, Razmara J, Rafi MA, Omid Y. Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches. *Drug Discov Today*. 2020; 25(6): 1034-42.
61. Oscherwitz J. The promise and challenge of epitope-focused vaccines. *Hum Vaccines Immunother*. 2016; 12(8): 2113-6.
62. Kryshchafyovych A, Schwede T, Topf M, Fidelis K, Moulton J. Critical assessment of methods of protein structure prediction (CASP) — Round XIII. *Proteins Struct Funct Bioinforma*. 2019; 87(12): 1011-20.