Pertanika J. Trop. Agric. Sci. 46 (1): 277 - 313 (2023)

PERTANIKA JOURNALS

TROPICAL AGRICULTURAL SCIENCE

Journal homepage: http://www.pertanika.upm.edu.my/

Characterisation of the Putative Antigenic Genes of the Outer Membrane Proteins of *Pasteurella multocida* B:2 Strain PMTB2.1 through *in silico* Analysis

Tahera Hashimi[1,4,5], Deborah Joyce[1], Sufia Mohd Nasir[1], Mas Jaffri Masarudin[1,3], Annas Salleh[2] and Sarah Othman[1,3]*

[1]*Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*
[2]*Department of Veterinary Laboratory Diagnostics, Faculty of Veterinary Medicine, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*
[3]*UPM-MAKNA Cancer Research Laboratory, Institute of Bioscience, Universiti Putra Malaysia, 43400 Serdang, Selangor Darul Ehsan, Malaysia*
[4]*Department of Biology, Faculty of Education, Ghor Institute of Higher Education, Chaghcharan 3201, Ghor, Afghanistan*
[5]*Department of Biomedical Science, City University of Hong Kong, Kowloon, Hong Kong, China*

ABSTRACT

Outer membrane proteins (OMPs), usually found in Gram-negative bacteria, have long been shown to elicit immune responses in infected hosts. This tendency of OMPs to generate immune reactions makes them ideal candidates for vaccine development against pathogenic bacteria. *Pasteurella multocida* is a Gram-negative pathogen responsible for the economically significant veterinary disease, hemorrhagic septicemia (HS). HS is an endemic and highly fatal disease affecting buffaloes and cattle. In Malaysia, outbreaks of this disease cost about half a million USD each year. Thus, despite current treatment and prevention measures, HS is a prevalent issue that needs to be overcome. *Pasteurella multocida* subsp. *multocida* PMTB2.1, a Malaysian strain of the pathogen, has recently had its entire genome sequenced after being isolated from HS outbreaks in the region. Antigenic OMPs from this strain have since been identified and published for further characterisation. LptD, Wza, and TbpA are integral membrane proteins, while Pal is a peripheral membrane protein that has not been characterised in-depth. This study, therefore, aims to analyse these OMPs through *in silico* methods. First,

ARTICLE INFO

Article history:
Received: 20 August 2022
Accepted: 16 November 2022
Published: 14 February 2023

DOI: https://doi.org/10.47836/pjtas.46.1.16

E-mail addresses:
tahera.hashimi53@gmail.com (Tahera Hashimi)
deborahjoyce98@gmail.com (Deborah Joyce)
sufia.mohdnasir@gmail.com (Sufia Mohd Nasir)
masjaffri@upm.edu.my (Mas Jaffri Masarudin)
annas@upm.edu.my (Annas Salleh)
sarahothman@upm.edu.my (Sarah Othman)
* Corresponding author

ISSN: 1511-3701
e-ISSN: 2231-8542

protein homology modelling was performed using SWISS-MODEL, whereafter, the structures generated were validated using the SWISS-MODEL structure assessment page, PROCHECK, ERRAT, and PROSA programs. The Pal, Wza, and TbpA structures were good models, while the LptD structure was found to be a near-good model based on the validation performed. Analyses using BCPREDS, NetMHCpan4.1, and NetBoLAIIpan1.0 revealed that these four OMPs could potentially elicit humoral and cellular immune responses.

Keywords: Antigenic OMPs, hemorrhagic septicemia, homology modelling, outer membrane proteins, *Pasteurella multocida*

INTRODUCTION

Pasteurella multocida is a Gram-negative pathogen from the Pasteurerellaceae family responsible for an economically significant veterinary disease, hemorrhagic septicemia (HS). HS is a highly fatal and chronic disease affecting animals, such as buffaloes and cattle in tropical regions. *Pasteurella multocida* strains are categorised into five serogroups (A, B, D, E, and F) and 16 serotypes based on lipopolysaccharide antigens (Peng et al., 2018). In Malaysia, the strain *P. multocida* subsp. *multocida* PMTB2.1 has been isolated from HS outbreaks in the region and had its complete genome sequenced as of late (Jabeen et al., 2017).

Pasteurella multocida expresses a range of virulence factors, including capsule, fimbriae, lipopolysaccharides (LPS), adhesion proteins, toxins, iron control, iron acquisition proteins, sialic acid synthesis, hyaluronidases, and OMPs (Peng et al., 2016).

Many vaccines have been developed and routinely used to prevent this acute disease. For example, bacterin, oil-adjuvant, and alum-precipitated vaccines were used in Asia to control the disease (Muenthaisong et al., 2020).

However, due to a high record of cattle and buffalo mortality resulting in a loss of nearly USD3.9 million, none of the above control techniques or live attenuated vaccines developed so far were completely effective (Rafidah et al., 2010). The failure and low vaccination coverage against the disease originated from several reasons, such as the high viscosity of the oil adjuvant vaccine that causes difficulties in injection and the variety of cattle management methods (Benkirane & De Alwis, 2002; Zamri-Saad & Annas, 2016). In Malaysia, a parental strain was isolated from an outbreak, and the mutant strain derived was labelled as *P. multocida* B:2 GDH7. This mutant was weakened by deleting a housekeeping gene, *gdhA*, from *P. multocida* B:2 wild type (Sarah et al., 2006).

Subunit vaccine has been suggested to be a potential candidate vaccine that is stable and safe compared to the others. Furthermore, a study by Rita et al. (2018) reported that production of a recombinant subunit vaccine had successfully cloned the ABA392 gene isolated from *P. multocida* B:2 into a protein expression vector. Consequently, the vaccine triggered an immune response against HS disease, thus, confirming its potential as a vaccine candidate for further studies.

Outer membrane proteins (OMPs) of *P. multocida* are one of the virulence factors in host animals to allow the bacteria to adapt to host niches. *Pasteurella multocida* OMPs are classified into structural proteins, transport proteins, binding proteins, adhesion, protein-assembly machinery, and membrane-associated enzymes based on their functional characteristics (Peng et al., 2019).

Joshi et al. (2013) demonstrated the immunoprotective efficacy of *P. multocida* OMPs in mouse models by proving the survival of OMP-vaccinated animals when exposed to live pathogenic bacteria. Prasannavadhana et al. (2014) used a proteomics approach to identify OMPs unique to the Indian strain of *P. multocida* serotype B:2, while Azam et al. (2020) used *in silico* methods to identify and assess the antigenicity of antigenic OMPs of the Malaysian isolate of *P. multocida* B:2. Among the 105 OMPs identified by Azam et al. (2020), 53 of them were found to be possibly antigenic. Individual OMPs of *P. multocida* have also been reported to play a function in activating an immunological response in several studies (Hatfaludi et al., 2012).

Bioinformatics tools assist researchers in finding possible proteins or genes associated with a pathogen's antigenic or virulence components (Azam et al., 2020). Bioinformatics analysis of protein localisation in bacteria is also crucial in understanding their structure and function, especially in disease development by the pathogen. Efforts to identify antigenic genes of *P. multocida* OMPs have been seen through *in silico* analysis to select potential vaccine targets for other strains, such as avian influenza (Al-Hasani et al., 2007). However, bioinformatics analysis reveals that the composition of the subproteome of *P. multocida* outer membrane (OM) is far from comprehensive. Further experimental efforts are needed to identify the complete complement of OMPs and to scribe a role to each protein. Several tools are now available that are fast, accurate, and easy to use, particularly for users with no previous experience in bioinformatics research (Hensen et al., 2012).

In the previous study by Azam et al. (2020) and this current study, these methods were used to predict some crucial antigenic genes of OMPs and to further characterise the potential OMPs by identifying subcellular localisation, predicting OM lipoproteins, and predicting barrel integral OMPs.

The availability of genome sequencing provides a new approach to the study of pathogens. More than 100 genome sequences are available for *P. multocida* B:2; however, extensive and complete genome sequence analysis is restricted. Based on this current study and Jabeen et al. (2019), the genome of *P. multocida* subsp. *multocida* strain PMTB2.1 (accession no. CP007205.2) contains 2,315,138 bp DNA, 2,176 genes, and 2,065 SDS coding genes, with more than 40 coding regions for iron homeostasis and 140 virulence genes, such as the complete tad locus. Among them, a few genomes, such as the PM70 strain (serotype F) (May et al., 2001), PM36950 (serotype A) (Peng et al., 2016), and the genomes of *P. multocida* strains harbouring the *P. multocida* toxin (PMT) gene have been studied and analysed (W. Liu et al., 2012).

The previous study by Azam et al. (2020) identified 105 OMPs of *P. multocida*

B:2 PMTB2.1, a Malaysian strain that includes 53 integral membrane proteins (IMPs) and 52 peripheral membrane proteins (PMPs). The finding of Azam et al. (2020) was elaborated in this work, primarily on the potential antigenic OMPs, Pal, Wza, TbpA, and LptD of the strain. The details are summarised in Table 1. The purpose of this study was to characterise the 3D model structure predictions further and to identify the antigenic epitopes of each OMP for future applications.

MATERIALS AND METHODS

Protein Sequence Retrieval

Pasteurella multocida subsp. *multocida* PMTB2.1 is a pathogenic *P. multocida* B:2 isolated from an outbreak in Malaysia from buffaloes, and the genomic DNA of the bacteria has been sequenced in the NCBI database (accession no. CP007205.2) (Jabeen et al., 2017).

The protein sequences of the OMPs from *P. multocida* subsp. *multocida* PMTB2.1 Malaysian strains were retrieved from National Center for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov/). The retrieved protein sequences were used in additional computer analysis to find the antigenic gene sequences for vaccine development or treatment. The nucleotide sequences of antigenic OMP genes, *pal*, *tbpA*, *wza*, and *lptD*, were retrieved and analysed from the genomic DNA of *P. multocida* B:2 subsp. *multocida* PMTB2.1 strain. The protein sequences of the antigenic OMP genes were used as input in the protein-protein basic local alignment search tool (blastp) (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The results of blastp were used to determine the exact location of the antigenic genes.

The complete genome of *P. multocida* subsp. *multocida* strain PMTB2.1 was retrieved and then downloaded from NCBI (https://www.ncbi.nlm.nih.gov/). The complete genomic DNA file of *P. multocida* subsp. *multocida* strain PMTB2.1 downloaded earlier, was opened using SnapGene viewer 5.2.3 software. Then, the sequence IDs obtained earlier were used to obtain the gene sequences of the putative antigenic genes using SnapGene viewer 5.2.3 software.

Table 1
Characterisation of the four antigenic OMPs of the P. multocida *B:2 PMTB2.1 Malaysian strain that were retrieved through* in silico *study*

Protein group	Accession Number	Function	Length (amino acids)	Molecular weight (kDa)	Protein type
Membrane assembly-associated proteins	AMM81068.1	Peptidoglycan associated lipoprotein	150	16.2	Perpheral membrane protein
	AMM81201.1	Thiamine ABC transporter substrate-binding protein	334	37.9	Integral membrane protein
	AMM81019.1	Sugar transporter	387	42.2	Integral membrane protein
	AMM82596.1	LPS assembly protein LptD	782	90.6	Integral membrane protein

Note. The table was adapted and modified from Azam et al. (2020) and the current study

The sequences of the four putative antigenic genes were retrieved from *in silico* study, and the primers were designed for each gene according to their sequences. First, the genes were amplified through polymerase chain reaction (PCR) using the extracted genomic DNA of *P. multocida* B:2. Then, the amplified genes were ligated into pJET1.2 plasmid vector using the *Escherichia coli* TOP10 as host cells.

The *in vitro* sequencing results of the cloned OMP genes (*pal*, *tbpA*, *wza*, and *lptD* into pJET1.2 plasmid vector) and the nucleotide sequences obtained from NCBI GenBank (https://www.ncbi.nlm.nih.gov) were uploaded to Expasy Translate Tool (https://web.expasy.org/translate/) to obtain the open reading frames (ORFs) and corresponding amino acid sequences. In addition, the frame with the longest and continuous amino acid sequence was retrieved for downstream analyses.

3D Homology Structure Modelling

In this study, the SWISS-MODEL server available at https://swissmodel.expasy.org/ was used for the 3D structure modelling of query proteins. The amino acid sequences retrieved from the Expasy Translate Tool were used as input. It was followed by a template search against the SWISS-MODEL template library (SMTL) derived from the Protein Data Bank (PDB) (Biasini et al., 2014) using two database search methods: BLAST (Camacho et al., 2009) and HHblits (Remmert et al., 2012). BLAST is useful for closely related templates, while HHblits is employed for distant homology searches. First, the template with a global model quality estimate (GMQE) score closest to 1, with a sequence identity of more than 30%, and a numerically lower structure resolution was chosen for building optimal models. 'Build models' was selected after that. Next, models with quality model energy analysis (QMEAN) Z-score above -4 and closer to 0 were selected. An updated GMQE score was also weighed, and models with a score closest to 1 were selected. The comparison plot was also evaluated for a visual representation of the models' Z-score before choosing a suitable homology model. Models of choice for Pal, LptD, Wza, and TbpA were saved in PDB format for further analysis. The steps applied in homology modelling using SWISS-MODEL are summarised in Figure 1.

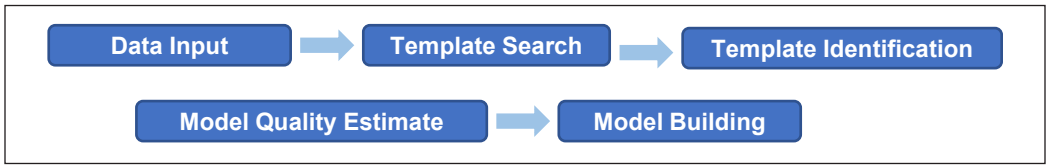

Figure 1. The general steps were taken to build 3D models of the outer membrance proteins in SWISS-MODEL. The target amino acid sequence inputs were provided in FASTA format. The input data served as a query sequence during the template search to identify evolutionary-related protein structures against the SWISS-MODEL template library. In template selection, the high reliability of a certain template is indicated by a high Global model quality estimate score that is expressed as a number between zero and one. Different models are built based on the templates selected. Following this, model building was continued with the generation of 3D models and model quality estimation using the Structure Assessment page of SWISS-MODEL

3D Homology Structure Validation

Evaluation of the 3D models were performed using several tools to determine their quality. Firstly, validation was performed using comparison plots and the local quality estimates available on the Structure Assessment page of the SWISS-MODEL server (https://swissmodel.expasy.org/). Following this, PROCHECK and ERRAT programs substantiated the model validation results. These programs are available in the SAVES 6.0 server, which is accessible at https://saves.mbi.ucla.edu/. The PROCHECK program makes a detailed scan of the stereochemistry of a protein structure, providing an assessment of the overall quality of the structure against highly refined structures of the same resolution (Laskowski et al., 1993). Meanwhile, the ERRAT program validates the model by the characteristic of atomic interactions and identifies the incorrect as well as the correct regions in a protein structure (Colovos & Yeates, 1993).

The ProSA server, which is available at https://prosa.services.came.sbg.ac.at/prosa.php, was then used to check the validity of the 3D structures. PDB files of the 3D structures were used as input data in all the servers mentioned above to verify the generated model. ProSA is a structure analysis program that displays scores and energy plots of local model quality for potential problems detected in protein structures. Global protein quality scores are displayed in the context of all known protein structures (Wiederstein & Sippl, 2007).

Protein Antigenicity Prediction and Comparison

Several parameters are important for predicting the immunogenicity of an antigen, such as flexibility, antigenicity, and solubility (Kim et al., 2013). The antigenicity of all OMPs was analysed by the ANTIGENpro server at http://scratch.proteomics.ics.uci.edu/. It is sequence-based, alignment-free, and pathogen-independent. A protein's antigenicity is predicted using protein antigenicity microarray data (Magnan et al., 2010). The input was provided in plain sequence format. Karplus Schulz's flexibility prediction method was used from the IEDB server (http://tools.iedb.org/bcell/) to determine the average flexibility of peptides within the OMPs. The flexible regions of the proteins are those shown above the threshold value of 1.0, following the method by Hasan et al. (2015). The SOLpro, which is available at http://scratch.proteomics.ics.uci.edu/, was used to predict the solubility of the OMPs and to report the solubility probabilities (above 0.5).

Epitope Prediction

B-cell epitopes were predicted using the B-cell epitope prediction server (BCPREDS), which is available at http://ailab-projects1.ist.psu.edu:8080/bcpred/predict.html. BCPREDS provides a selection of three prediction methods; the amino acid pairs (AAP) method (Chen et al., 2007), BCPred, and FBCPred (EL-Manzalawy et al., 2008b). The AAP method was chosen, and amino acid sequences were provided in plain text format, with epitope length set

to 12 and specificity set at default (75%) based on the method of Baliga et al. (2018). The program was also set to only report non-overlapping epitopes before submitting the query.

Prediction of major histocompatibility complex class I (MHC-I) T-cell epitopes and binding affinities were made using NetMHCpan 4.1 (Fisch et al., 2021), which is available at http://www.cbs.dtu.dk/services/NetMHCpan/. This server predicts the binding of peptides to MHC-I molecules of the known sequence using artificial neural networks (ANNs). The method is trained on quantitative binding affinity (BA) and mass-spectrometry eluted ligands (EL) peptides. The % rank threshold for strong binding peptides was set at 0.5, while the % rank threshold for weak binding peptides was set at 2. The peptide length selected was 9-mer, and the alleles chosen were BoLA-1:00901, BoLA-1:02301, BoLA-2:01201, BoLA-2:03001, BoLA-3:00101, and BoLA-3:00201. Input data was provided in FASTA format. Earlier findings reported that the vast majority (96.5%) of natural ligands are identified at a very high specificity (98.5%) using a percentile rank threshold of 2%. Input amino acid sequences are digested into 9-mer peptides. Research has shown that most presented MHC class I ligands are between 8-10 amino acids in length (Trolle et al., 2016). The BoLA alleles listed were also chosen as these were annotated in the IPD-MHC database (https://www.ebi.ac.uk/ipd/mhc/group/BoLA/).

Prediction of major histocompatibility complex class II (MHC-II) T-cell epitopes and binding affinities were made using the NetBoLAIIpan 1.0 server (http://www.cbs.dtu.dk/services/NetBoLAIIpan/) (Fisch et al., 2021). This server uses EL data derived from cell lines expressing a range of DRB3 alleles for model training. The % rank threshold for strong binding peptides was set at two, while the % rank threshold for weak binding peptides was set at 10. The peptide length selected was 13-mer, and all the alleles listed as BoLA-DRB3 alleles from mass spectrometry (MS) data were selected. Input amino acid sequences were digested into 13-mer peptides as research has shown that most presented MHC class II ligands are between 13–17 amino acids in length (Chang et al., 2006). Input data was provided in FASTA format, and additional configurations to display only the strongest binding core was selected alongside checking the 'Sort' output by prediction score option. The epitopes predicted by the BCPREDS, NetMHCpan 4.1, and NetBoLAIIpan 1.0 servers were manually combined to arrive at a consensus antigenic epitope sequence for each OMP.

RESULTS AND DISCUSSION

3D Homology Structure Modelling

Although X-ray crystallography or protein nuclear magnetic resonance spectroscopy (NMR) can be used to identify protein structures more precisely, these procedures are frequently time-consuming and expensive. In this sense, protein homology modelling using online servers and computer software can provide valuable models that help establish hypotheses about protein

function and drive additional experimental effort in the shortest time attainable while being cost-effective (Haddad et al., 2020). Such downstream applications include designing site-directed mutations through primer design and predictions of receptor-ligand binding for drug targeting (Sailapathi et al., 2020). Freely available sites for molecular docking are LightDock (Jiménez-García et al., 2018), MedusaDock 2.0 (Wang & Dokholyan, 2019), and LeDock (N. Liu & Xu, 2019).

Protein homology modelling generates 3D structures of proteins using a known experimental structure of a homologous protein as a template (Hasani & Barakat, 2016). 'Coverage' in templates quantifies the degree to which the template covers the target sequence. Although in most cases, achieving coverage of 100% is not possible, there is much to be inferred about a protein if the important domains are included within the template. As is observed with all the OMPs' models, none of the templates had 100% coverage. As such, other parameters were also taken into consideration. The number of amino acids perfectly matched between the target and template sequences is referred to as 'Identity.' Because the alignment of amino acids between the target and template sequences can be inferred, the higher the identity, the better the model. Because of the existence of a twilight zone, less than 30% of identities might be challenging to model. A homologous template backbone is no longer sufficient to regulate the exact assembly of protein side chains in this zone (Chung & Subbia, 1996). Thus, by default, the templates listed were filtered for those having an identity of 30% or above.

Models generated with the nearest GMQE score of 1 would indicate reliable templates (Biasini et al., 2014). This score estimates the model's quality given the target-template alignment and the template structure (Waterhouse et al., 2018). The models were also selected based on their QMEANDisCo global score, where a higher score is better as it shows the accuracy of local quality estimates (Waterhouse et al., 2018). High numeric values of 'resolution', such as 4 Å, mean poor resolution, while low numeric values, such as 1.5 Å, mean good resolution (Martz, 2014). For example, 2.05 Å is the PDB's median resolution for X-ray crystallographic results (Martz, 2014). Therefore, it is a crucial factor in modelling, as poor-resolution templates will give rise to poor-quality models. Details about the generated homology models are summarised in Table 2.

The Pal model in Figure 2A has a distinctly simpler structure compared to the LptD model in Figure 2B, owing to the shorter nucleotide sequence of *pal* and subsequently shorter amino acid sequence of Pal compared to that of LptD. The Pal model consists of four α-helices and four β-sheets linked by flexible loop regions, forming an α/β sandwich domain. This domain is characteristic of Pal proteins reported across the literature (Kumar et al., 2015; Mathelié-Guinlet et al., 2020). As for the LptD model, periplasmic and extracellular dimensions are obvious with the membrane annotations

Table 2
Details of P. multocida *PMTB2.1* outer membrance proteins*' homology models*

Outer membrane protein	Template SMTL ID	Coverage	Identity	GMQE	QMEAN DisCo Global	Resolution
Pal	4pwt.1	0.75	71.43	0.65	0.84	1.8 Å
TbpA	2qry.1	0.91	50.99	0.77	0.79	2.25 Å
Wza	2j58.1	0.86	56.46	0.75	0.78	2.25 Å
LptD	4q35.1	0.97	47.93	0.77	0.74	2.4 Å

Note. SMTL = SWISS-MODEL template library. A structural database produced from the Protein Data Bank comprises experimentally established protein structures. Global model quality estimate (GMQE) is a score between 0 and 1. A score closer to 1 is preferable. Qualitative model energy analysis distance constraint (QMEAN DisCo) global is a composite scoring function assessing the major geometrical aspects of protein structures

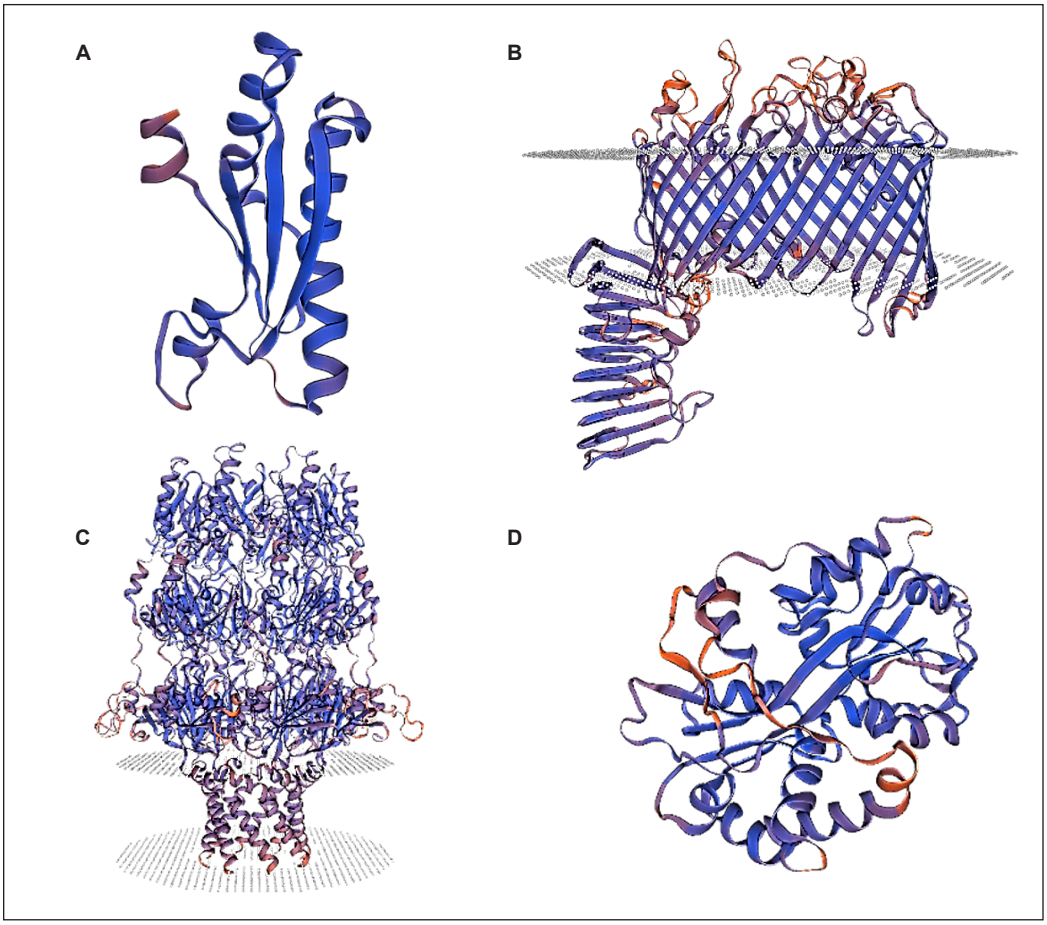

Figure 2. The 3D protein homology models of *Pasteurella multocida* PMTB2.1 outer membrance proteins. (A) 3D structure of Pal, (B) 3D structure of LptD, (C) 3D structure of Wza with membrane annotations transferred from the template, and (D) 3D structure of TbpA. The model colour scheme indicates regions of low confidence (red) and regions of high confidence (blue)

transferred from the template used. The modelled LptD structure was also observed to have a 26-stranded-β-barrel domain and a β-jellyroll domain consistent with LptD structures reported across literature (Qiao et al., 2014; Villa et al., 2013). SWISS-MODEL colour codes the modelled structures by indicating regions of low confidence in red and regions of high confidence in blue. It is visible in Figure 2B that the modelled LptD structure's loop regions on the extracellular dimension are mostly of low confidence and therefore coloured in red. It is because loop regions can be particularly difficult to model as sequences and structures might differ in these areas (Waterhouse et al., 2018). However, loop refinement or remodelling can be employed using other programs like the GalaxyLoop (Park et al., 2014), DaReUS-Loop (Karami et al., 2019), and MODELLER (Fiser & Šali, 2003).

Another region of homology models that can be hard to model is N-terminal tails (Waterhouse et al., 2018). Due to insufficient coverage between the template and target amino acid sequences (75% coverage, from residue 38 to 150), the N-terminal tail in the modelled Pal protein is found to be absent when compared to the template that was used to model the Pal protein (SMTL ID: 4pwt.1) and the structure of Pal protein reported by Mathelié-Guinlet et al. (2020). This N-terminal tail is, in fact, central to the role of the Pal protein. It functions to bind the Pal protein to the inner leaflet of the outer membrane (Cascales & Lloubès, 2004), while the C-terminal region of Pal interacts with the cell wall peptidoglycan (Lazzaroni & Portalier, 1992). Similarly, the LptD model's structure determines the protein's function. The β-barrel domain, along with the N-terminal β-jellyroll domain, forms a hydrophobic portal for the transport of lipopolysaccharide from the periplasm to the extracellular surface of the outer membrane (Botos et al., 2017). In brief, protein homology modelling can aid in characterising protein functions concerning their structures. It may direct predictions on how to regulate or modify proteins. However, limitations in this approach remain. These include the inability to predict sections of the sequence that do not align with the template sequence, as well as inaccuracies in modelling loop regions.

Wza is an inner membrane-lipoprotein (IM-lipoprotein) with the crystal structure of an octameric structure that is shaped like an amphora without its handles (Figure 2C) (Reid & Whitfield, 2005). Its transmembrane region is a novel α-helical barrel. The bulk of its structure is in the periplasm, which is also made up of three novel domains that create a large central cavity. It opens at a narrow 'neck' in the outer membrane towards the extracellular environment but is closed at its base at the periplasm (Dong et al., 2006). Wza's monomer structure comprises four domains. However, in ribbon format, Wza is an octameric structure made of four rings annotated as R1-R4 (Ford et al., 2009). Dong et al. (2006) mentioned the characteristics of each domain in Wza. Domain 1 comprises an anti-parallel β-sandwich with an α-helix at one edge representing a novel fold. Eight copies of Domain 1 would make up the Ring 1 structure at the bottom of Wza. Thus, Ring 1 has a centre filled with eight loops with a

concave surface at the base. The conserved polysaccharide biosynthesis/export (PES) motif is in Domain 1. Domain 2 shares a similar dimension to Domain 1 despite having a novel structure. An eightfold symmetric ring structure, Ring 2, is formed by eight copies of Domain 2. The domain comprises of central five-stranded mixed β-sheet with three α-helices. Domain 3 has a larger diameter than Domain 2 despite having the same structural duplication as Domain 2. N-terminus is located at the top of Ring 3. Hence, this indicates its location on Domain 3. Domain 4 is an amphipathic helix with a carboxy terminus of the monomers (Dong et al., 2006). An α-helical barrel was produced at the 'neck' of the structure due to the rotational symmetry of helices from eight monomers (Dong et al., 2006). Rings 1–3 are located inside the periplasmic (Kong et al., 2013), while Ring 4 has a C-terminus outside the cell (Dong et al., 2006). Ring 4 is a hydrophobic helical barrel, while Rings 1–3 consist of hydrophilic domains (Collins & Derrick, 2007; Ford et al., 2009).

According to a study by Soriano et al. (2008), the crystal structure of the thiamine-binding protein (TbpA) from *E. coli* in a complex with thiamine monophosphate (ThMP) was identified. Hence, it was referred to understand the crystal structure of TbpA from *P. multocida* in this study (Figure 2D). An adenosine triphosphate (ATP)-binding cassette (ABC) transporter in Gram-negative bacteria consists of periplasmic binding protein membrane permease and ATPase (Sippel et al., 2009). ABC transporters function by transporting ions and molecules into cells by utilising cellular energy (Jones & George, 2004), and in this case, it was used to take up the thiamine. Organisms used ABC transporters to take up thiamine. A study by Harper et al. (2006) mentioned that ABC transport protein is a predicted protein from the outer membrane of *P. multocida* A:3 strain P1059 that can stimulate cross-serotype protection (Rimler, 2001). TbpA structure consists of two domains with a flexible hinge region and a binding cleft between two domains (Soriano et al., 2008). Domain 1 consists of a three-layer αβα sandwich. The β layer has a five-stranded mix of β-sheet with a strand of the β-strand topology of ↑β2↑β1 ↑β3 ↓β9↑β4a. Meanwhile, for the α layer, five α-helices flanked the centre of the β-sheet with three helices below the β-sheet and two helices above it. Domain 2 also consists of a five-stranded mix of β-sheet with a strand of β-strand with a different topology compared to Domain 1, which is ↑β6↑β5 ↑β7↓β4b↑β8. Six and three α-helices flanked the above and below of the central β-sheet, respectively (Soriano et al., 2008). Jones and George (2004) stated that clarifying the mechanism and structure of ABC transporters is crucial to making pathways for rational protein design. Furthermore, it provides a guiding point for the proteins to be validated experimentally.

3D Homology Structure Validation

Determining the quality of a 3D protein homology model is a crucial step in understanding a model's utility and its potential downstream applications (Benkert et al., 2009).

Quality estimation was first done by analysing the local quality estimates of the generated 3D models. The local quality estimates in Figure 3 display sections of the models that are reliable and those that are not. The x-axis represents the residue number, while the y-axis represents predicted local similarity to the target. Red bars indicate residues of low confidence and low similarity between target and template residues, while blue bars are residues of high confidence that indicate high similarity between target and template residues, and therefore, high quality (Waterhouse et al., 2018) stretches below 0.6 hint towards regions of very low quality. Between the Pal (Figure 3A) and LptD models (Figure 3B), it can be observed that the LptD model has many more residues with stretches below the 0.6 threshold. It indicates that there are many erroneous regions in the LptD model. On closer inspection, it was noted that the stretches below 0.6 correspond to the loop regions of the LptD model, specifically from residues 260 to 737. This finding substantiates the observation made in the 3D protein homology model of LptD (Figure 2B), where the loop regions were also indicated in red. It was expected as loops are often the most vulnerable to model inaccuracies due to their increased variability in terms of sequence and structure when contrasted with the rest of the protein (Waterhouse et al., 2018). As for Wza and TbpA, the local quality estimates of the models are good, as most residues had scores close to 1. Wza (Figure 3C) and TbpA (Figure 3D) appear, with most residues between 0.8 and 0.9. Regions with low quality are indicated with residues that fall below 0.6. Wza shows a great number of residues in position 280-320 that falls below 0.6, while TbpA shows low-quality residues in regions 30-60, 180-210, and 210-250.

Further validation was performed with QMEAN, which relies on the database's context of all known experimental 3D structures by the scores representing energies (Haddad et al., 2020). These are shown in Figure 4. The comparison plots provide comparisons with a non-redundant set of PDB structures, a comparison to observe how well properties of known structures are represented in the model of interest. It consists of the x-axis representing the size of protein structures, the y-axis representing the normalised QMEAN4 scores of properties, and numerous dots, each representing one structure from the non-redundant set of PDB database structures. QMEAN4 refers to a version of the scoring system based solely on the statistical potentials of the four geometric aspects of the protein structure.

The red star represents the model of interest. The dark grey zone represents structures with a Z-score ranging from within one standard deviation of the mean in a normal distribution. Ideally, the star should be in the dark grey zone, indicating that the model behaves similarly to the experimentally determined PDB structure (Agrawal et al., 2013). Based on the comparison plots, the modelled structure of Pal (Figure 4A) ideally behaves as the star representing it is in the dark grey zone. It means that the modelled Pal structure has a QMEAN Z-score within one standard deviation of the mean in a

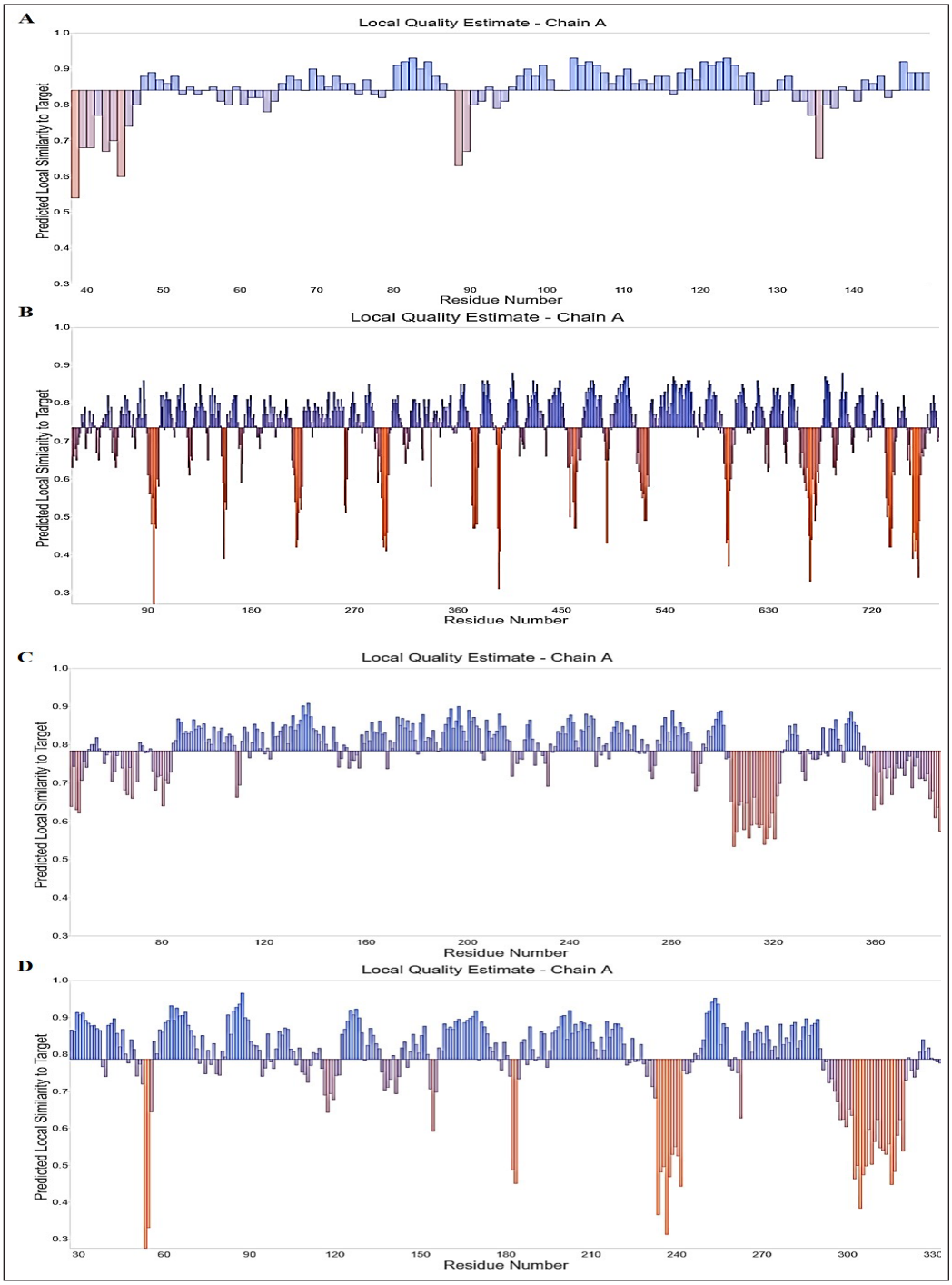

Figure 3. Local quality estimate of residues on modelled (A) Pal, (B) LptD, (C) Wza, and (D) TbpA structures. These were obtained from the Structure Assessment page of the SWISS-MODEL program. Low confidence residues are shown in red, while high confidence regions are displayed in blue. The threshold for the determination of very low-quality residues lies below 0.6 in predicted local similarity to target protein residues

normal distribution and, therefore, a higher "degree of nativeness" (Benkert et al., 2011) comparable to experimentally determined Pal structures of similar size in the PDB with a QMEAN Z-score of 0.15. In contrast, the LptD model (Figure 4B) is relatively not-ideal as the red star representing the model is located on the far right, with a QMEAN Z-score of -2.22, outside of the range of experimentally determined PDB structures. However, the LptD model is still acceptable as the QMEAN Z-score value is above -4.0 (Benkert et al., 2011). Meanwhile, Figure 4C shows that the Wza protein structure is reliable as it is within the range of other protein structures in PDB with a Z-score less than 1, -0.44. However, like LptD, the modelled TbpA protein structure lies outside the range of other structures in PDB (Figure 4D) with a Z-score of more than -2.63, indicating that the TbpA model is not ideal as well.

Validation was then continued with the PROCHECK program (Laskowski et al., 1993). PROCHECK validates the stereochemical quality and accuracy of the models and reports results as Ramachandran plots (Figure 5) that indicate residues in the favoured regions in red, allowed regions in yellow, generously allowed regions in light

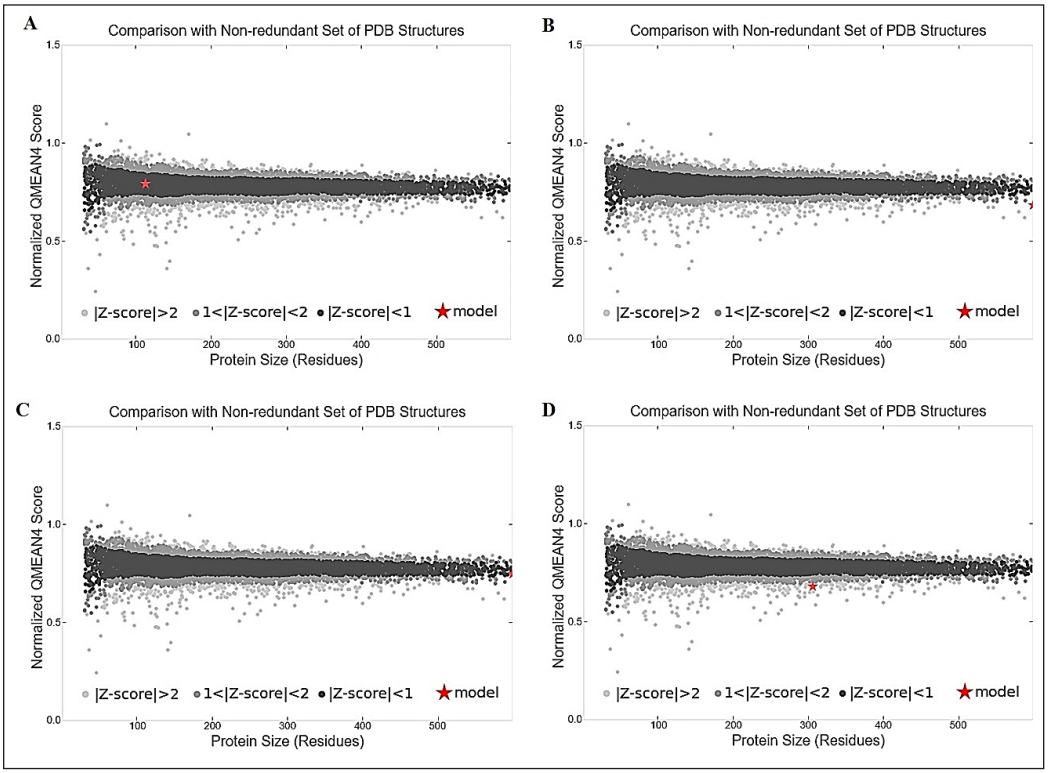

Figure 4. Comparison plot of the models represented by a red star (A) Pal, (B) LptD, (C) Wza, and (D) TbpA against a non-redundant set of Protein Data Bank structures (PDB). These were obtained from the Structure Assessment page of the SWISS-MODEL program. Each dot represents an individual protein crystal structure from the PDB database

yellow, and disallowed regions in white. In Figure 5, the red regions that determine the most favoured regions are marked as {A, B, L}. The additionally allowed regions are marked as {a, b, l, p} in yellow regions. Small black squares represent proline and non-glycine residues, while black triangles are glycine- residues (non-end). Disallowed residues are shown as red squares. In Ramachandran plot analyses, the percentage of residues in the most favoured regions must be more than 90% to ensure the reliability of the predicted structure. Meanwhile, G-factors measure how unusual the property is, with values below -0.5 and below -1.0 highly unusual.

The Ramachandran plot analyses indicate that the Pal model (Figure 5A) is a good model with more than 90% of its residues in the favoured region, while the LptD model (Figure 5B) is a near-good model with 88.7% residues in the favoured region (Prajapat et al., 2014). Both models are also observed to have an overall G-factor

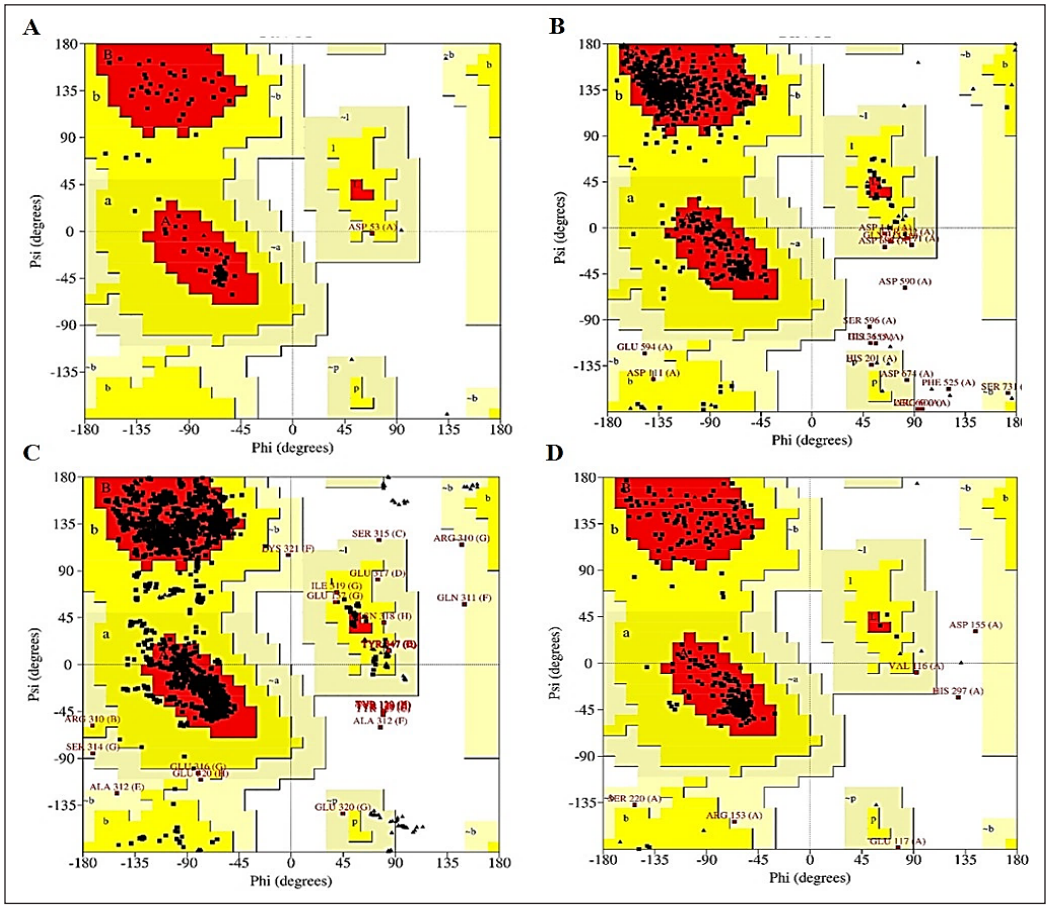

Figure 5. Main Ramachandran plot of the PROCHECK program for modelled (A) Pal, (B) LptD, (C) Wza, and (D) TbpA structures. Residues are indicated in favoured regions which are shown in red (represented by A, B, L), allowed regions in yellow (represented by a, b, l, p), generously allowed regions in light yellow (represented by ~a, ~b, ~l, ~p) and disallowed regions in white

of more than -0.5, indicating that they are reliable (Ramachandran et al., 1963). Similar to the LptD model, the Ramachandran plots for modelled Wza (Figure 5C) and TbpA (Figure 5D) did not record more than 90% of residues in the most favoured regions. Instead, the percentage of residues in the most favoured regions for Wza and TbpA is 89.5% and 89.9%, respectively. Thus, they are near-good quality models. Although both models did not exceed the 90% score in most favoured regions, they are still valid as the residues in disallowed regions are merely 0.5% and 0.7% (Prajapat et al., 2014). Hence, this represents only a slight steric hindrance between the model's side-chain C-beta methylene group and main-chain atoms. The overall G-factor for the Wza and TbpA structures are -0.08 and -0.17, respectively. Therefore, each modelled structure is acceptable as both scores are greater than the G-factor threshold value of -0.5. Data regarding the percentage residues in each region and overall G-factors have been summarised in (Table 3).

Quality estimation was done using the ERRAT server, where the overall quality factor for good high-resolution structures lies above 95%, while low resolutions have values of around 91% (Colovos & Yeates, 1993). Overall ERRAT quality factor values are expressed as the percentage of the proteins for which the calculated value falls below the 95% rejection limit. The quality factor values for Pal and LptD models were 100% and 91.3%, respectively (Figures 6A and 6B). It was established earlier that high-resolution structures generally produce values around 95% or higher (Colovos & Yeates,1993). Hence, the Pal model gets a maximum score as it was modelled after a high-resolution template (1.8 Å). On the other hand, the low ERRAT quality factor value for the LptD model is expected owing to its lower resolution template (2.4 Å). The Wza and TbpA were modelled after templates with a resolution of 2.25 Å. The overall quality factor of the Wza and TbpA models showed a score of 94.537 and 85.619, respectively, as shown in Figures 6C and 6D, respectively. It indicates that Wza has a good quality factor while TbpA might be considered near good as it is near 91%. Furthermore, the OMPs only had several regions that exceeded the error values of 95 and 99%. Thus, indicating that most residues are acceptable.

Lastly, ProSA was used to provide energy plots that can highlight potential errors in protein structures. The plots in

Table 3
Result summary of the Ramachandran plots from PROCHECK

Outer membrane protein	Residues in the favoured region (%)	Residues in the additionally allowed region (%)	Residues in the generously allowed region (%)	Residues in the disallowed region (%)	Overall G-factor
Pal	91.9	7.1	1.0	0.0	-0.02
LptD	88.7	8.7	1.5	1.0	-0.16
Wza	89.5	9.2	0.8	0.5	-0.08
TbpA	87.5	10.8	1.1	0.7	-0.17

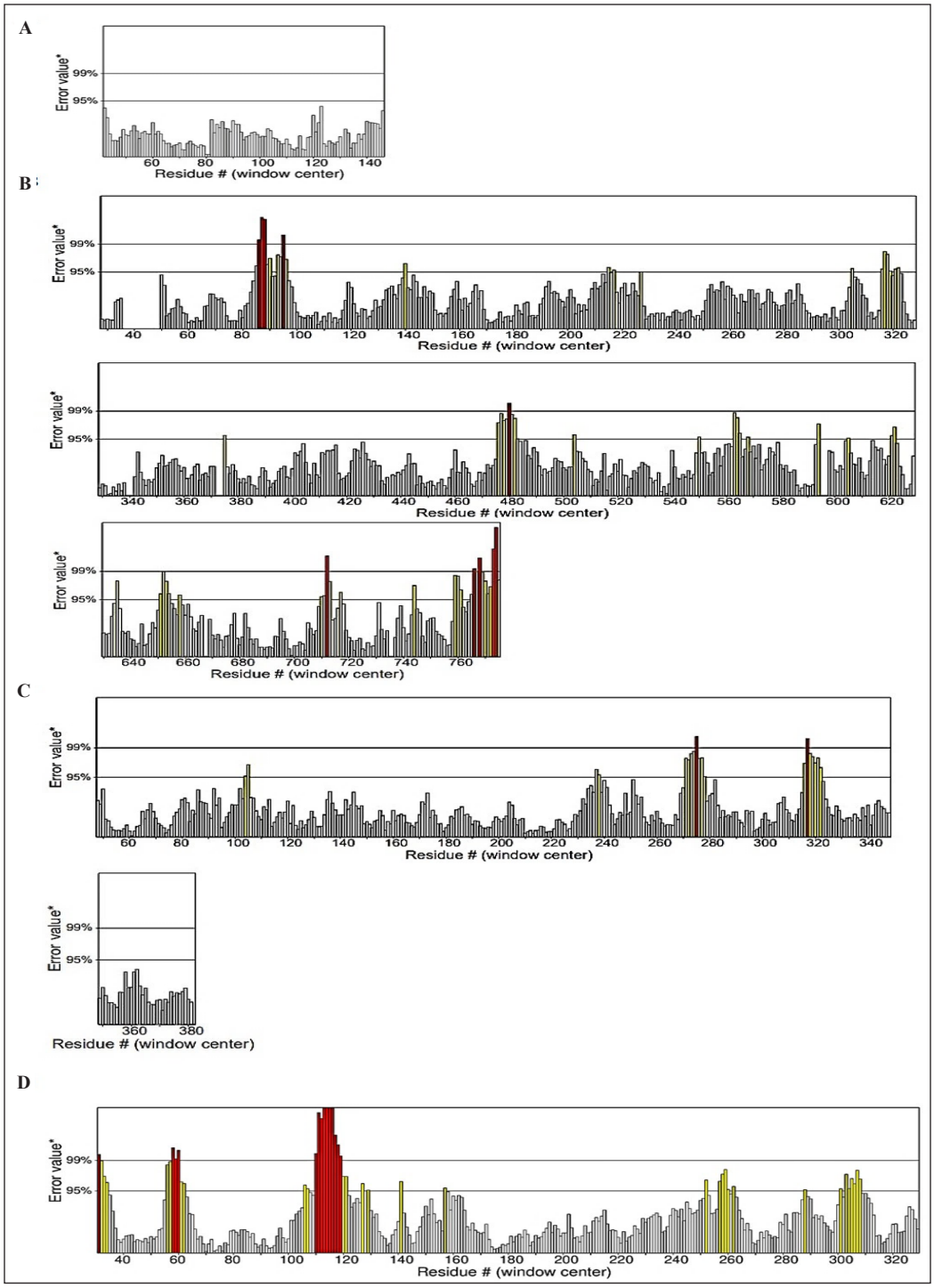

Figure 6. The ERRAT program evaluated the overall quality of the model. (A) Pal model, (B) LptD model, (C) Wza model, and (D) TbpA model. On the error axis, the two lines (95% and 99%) indicate the confidence with which it is possible to reject regions that exceed that error value

Figure 7 represent data obtained from the PDB regarding protein chains' scores discovered experimentally (Berman et al., 2002). All the plots were determined by X-ray crystallography (light blue) or NMR spectrophotometry (dark blue) about individual protein lengths (Agrawal et al., 2013). The inaccuracies of the models generated can be determined by the analysis of Z-scores being outside of the range of scores characteristic of native proteins. Analysing the position of the target protein (represented as black dots) in the Z-score plot could indicate whether the protein is in the typical range of scores as other proteins of similar length in PDB. The Z-score of the model presents the overall model quality.

The Z-score for the Pal model (−5.43) (Figure 7A) was within the range of scores typically found for the native proteins of similar size, while this was hardly the case for the LptD model (- 4.27) (Figure 7B). Figures 7C and 7D show the positioning according to individual Z- scores for Wza

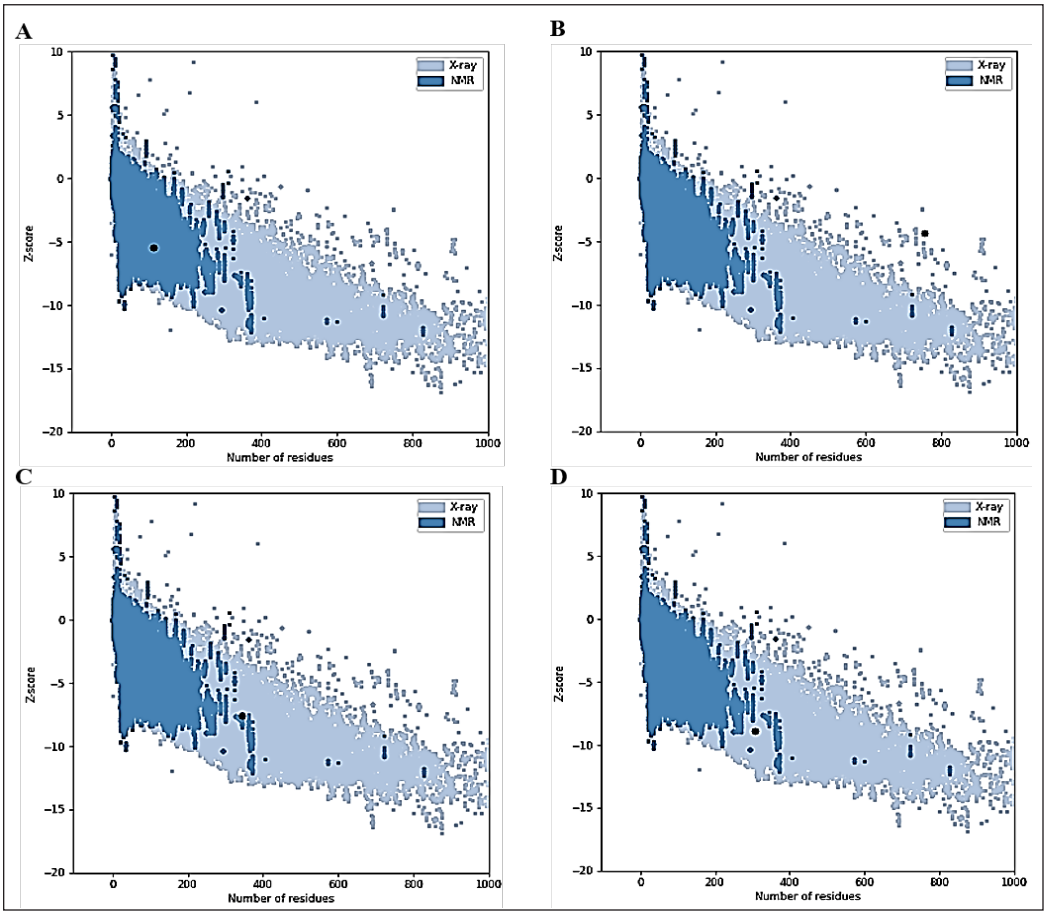

Figure 7. ProSA output of Z-scores for homology models using the ProSA-web server. (A) Pal model, (B) LptD model, (C) Wza model, and (D) TbpA model. The models are represented as black dots in the plot. These plots are used to identify whether the Z-score of protein is within the range with the typical range of scores for experimentally determined proteins of similar sizes

and TbpA with Z-score values of -7.57 and -8.85, respectively. The ProSA outputs indicate that the homology models for Pal, Wza, and TbpA are within the range of native protein conformations. It means that the 3D homology structures of Pal, Wza, and TbpA are highly reliable structures that are well within the range of scores typically found for proteins of similar size.

The plot of residue scores or energy plots provides local model quality by using knowledge-based energy to plot amino acid sequence position (Agrawal et al., 2013), where positive values correspond to problematic or erroneous parts of the input structure. This plot revealed that most of the calculated values were negative for Pal (Figure 8A), while a fair number of LptD residues had negative and positive scores (Figure 8B). It implies that there are many erroneous regions in the LptD model. The Wza model in the energy plot (Figure

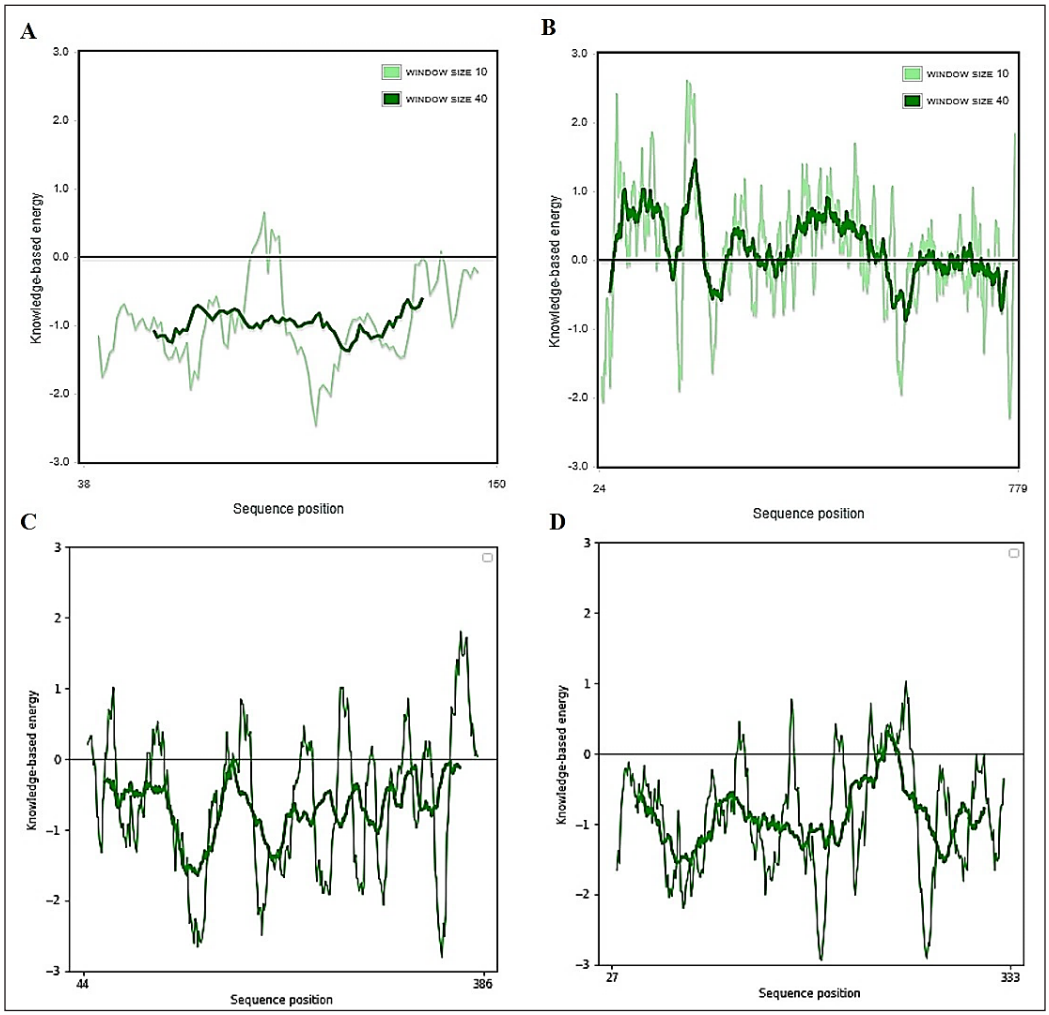

Figure 8. Plot of residue energies for (A) Pal model, (B) LptD model, (C) Wza model, and (D) TbpA model. Dark green and light green lines distinguish window sizes of 40 and 10 residues, respectively

8C) shows that all residues are positioned within the negative quadrant. It implies that the model has good energy stability. Meanwhile, the TbpA model (Figure 8D) shows that some residues are in the positive quadrant but still near zero, indicating that it has average energy stability (Bhattacharya et al., 2020).

Protein Antigenicity Prediction and Comparison

ANTIGENPro provides an accuracy of up to 82%, which surpasses other antigenicity predictors (Magnan et al., 2010). In addition, antigenic proteins in other pathogens were detected by this program (Chin et al., 2014; Magnan et al., 2010). ANTIGENpro server scores 0.816380 for Pal, 0.900650 for LptD, 0.779916 for Wza, and 0.804137 for TbpA. It shows that all proteins scored well above the minimum threshold of 0.7, indicating that they are all antigenic (Magnan et al., 2010) (Table 4). It substantiates the findings from Azam et al. (2020), whose study classified these four OMPs as antigenic following the results obtained from the Vaxijen v2.0 server (Doytchinova & Flower, 2007).

Solubility is crucial for maintaining the homeostasis of proteins. Identifying the solubility of a protein could give access to its conformation, concentration, and quaternary structure (Balch et al., 2008). Furthermore, maintaining a protein's function and preventing its aggregation requires solubility (Ciryam et al., 2013). Studies on structural and functional proteomics also require protein to be soluble (Chan et al., 2010). Therefore, identifying a protein's solubility is important for immunogenicity prediction. In Table 5, it was identified that Wza and LptD proteins are insoluble. Hence, they are not appropriate for the downstream analysis of the immunogenicity effect (Rahman et al., 2020). Meanwhile, Pal and TbpA are considered soluble. Hence, proving it to be suitable as a vaccine candidate for further analysis in the future. On the other hand, the area of a peptide that interacts with the antibody tends to be flexible, according to experimental findings (Hasan et al., 2015). The Karplus Schulz flexibility prediction program found the query protein's flexible regions in the IEDB server. As reported in Table 5, the OMP Pal recorded a flexibility score of 0.997, while the OMP TbpA scored 0.992. Between the two soluble OMPs,

Table 4
Antigenicity scores of outer membrane proteins derived from the ANTIGENpro server

Outer membrane proteins	Probable antigenicity	Prediction score
Pal	Antigen	0.816380
LptD	Antigen	0.900650
Wza	Antigen	0.779916
TbpA	Antigen	0.804137

Table 5
Flexibility and solubility of the OMPs, Pal, LptD, Wza, and TbpA

Outer membrane proteins	Flexibility (IEDB)	Solubility (SOLpro)
Pal	0.997	SOLUBLE with 0.762434
LptD	1.001	INSOLUBLE with 0.745939
Wza	0.996	INSOLUBLE with 0.700107
TbpA	0.992	SOLUBLE with 0.934680

Pal showed a higher flexibility score and is therefore considered relatively more immunogenic than the other OMPs reported in this study.

Epitope Prediction

T-cell or B-cell, epitope-based therapeutic and diagnostic rational design for pathogens, has long since been reported (Arévalo-Herrera et al., 2002; Sollner et al., 2008). Regarding T-cell epitopes, activation of both $CD4^+$ (MHC-II) and $CD8^+$ (MHC-I) T-cells are vital in many cases of infections (Jenkins et al., 2003). Therefore, in developing peptide-based vaccines, OMPs epitopes that elicit both the B-cell and T-cell (MHC-I and MHC-II) mediated immunities are the most advantageous (Bhattacharya et al., 2020).

B-cells are critical in the immune system's battle against invading pathogenic organisms because they secrete antibodies against antigens. A linear or continuous B-cell epitope is a sequence segment of a protein. Presently, the bulk of existing epitope prediction algorithms focuses on continuous epitopes due to the ease with which the amino acid sequence of a protein may be used as an input. These prediction approaches are based on the hydrophilicity, solvent accessibility, secondary structure, flexibility, and antigenicity of amino acids (Shen et al., 2015).

BCPred online server was used in predicting linear B-cell epitopes of the OMPs from their respective amino acid sequences. BCPred provides two algorithms for identifying an epitope's fixed lengths, which are BCPred and amino acid pair (AAP) (EL-Manzalawy et al., 2008a). The AAP technique is based on discovering that B-cell epitopes prefer some AAPs. The AAP technique outperformed conventional scales based on single amino acid propensity (Chen et al., 2007). The epitope length chosen for our study is 12 amino acids, which were motivated by previous work (Baliga et al., 2018). B-cell epitopes of lengths between 12 to 15 were shown to be potent inducers of the production of antibodies (EL-Manzalawy et al., 2008a). B-cell epitope prediction using the AAP method within the amino acid sequences of individual OMPs is summarised in (Supplementary material 1).

Identifying the shortest peptide in an antigen that triggers $CD4^+$ or $CD8^+$ T-cell mediated immunity is the objective in T-cell epitope prediction. In assays, immunogenicity was achieved using synthetic peptides derived from antigen that triggers T-cells (Ahmad et al., 2016). MHC aids in activating the adaptive immune response. MHC genes present peptide fragments from pathogens to T- cells to elicit an immune response. Therefore, the ability of the T-cell epitope region in antigenic proteins to bind to MHCs is essential in vaccine development (Patronov & Doytchinova, 2013). Epitope binding to MHC-I and MHC-II differs in length. For MHC-I binding prediction, input amino acid sequences were digested into 9-mer peptides, as research has shown that most presented MHC class I ligands are between 8 to 10 amino acids in length (Trolle et al., 2016). Meanwhile, the MHC-II binding epitope would be in the range of 13 to 17

amino acids in length (Chang et al., 2006). Considering the optimal epitope lengths for recognition and binding with B- cells and T-cells, antigenic sequences found to be recognised by B-cells and those with high binding affinities for MHC-I and MHC-II make for an effective epitope vaccine (Naz et al., 2015).

Based on data summarised in (Supplementary material 2), some peptide sequences of MHC-I and MHC-II can bind to more than one allele and have a part of a similar sequence to B-cell epitope prediction. Thus, a consensus antigenic sequence was obtained for each OMP by combining similar sequences found to be recognised by B-cells and those with high binding affinities for MHC-I and MHC-II. These sequences have been summarised in Table 6.

Table 6
B-cell, MHC-I, and MHC-II recognition sites within the amino acid sequences of Pal, LptD, Wza, and TbpA

OMP	Antigenic sequence	Amino acid position	Recognised by
Pal	LQQRYNTVY	41-49	B-cell
			MHC-I: BoLA-1:00901
			MHC-II: BoLA-DRB3_1001
LptD	ALFDSPLNF	393-401	B-cell
			MHC-I: BoLA-1:00901
			MHC-II: BoLA-DRB3_1001
			MHC-II: BoLA-DRB3_1201
Wza	LLVDGDIVH	631-657	B-cell
			MHC-I: BoLA-1:00901
TbpA	EPHFDALKATQM	287-322	B-cell
			MHC-II: BoLA-DRB3_2002
	KTKVQQPPK	770-796	B-cell
			MHC-I: BoLA-2:01201
	TQAVNVYTY	1,109-1,135	B-cell
			MHC-I: BoLA-1:00901
			MHC-I: BoLA-1:02301

CONCLUSION

The *in silico* generated 3D structures of Pal, Wza, and TbpA are considered good models, while the 3D structure of LptD is a near-good model. Good quality models are imperative as these are required for downstream applications, including protein function prediction, site-directed mutations, and ligand docking modelling in drug discovery. This study has also laid out the consensus epitope sequences within the OMPs that could elicit both humoral and cellular immune responses using bioinformatics approaches. Further *in silico* work can be directed towards molecular docking analysis between the identified epitopes and immunological elements to study the specific interactions involved.

Future investigations in validating B-cell and T-cell immunogenicity of the identified antigenic epitopes can be attempted through *in vitro* and *in vivo* means to prepare epitope or peptide-based vaccines against *P. multocida* infections.

ACKNOWLEDGMENTS

Universiti Putra Malaysia supported this research through the UPM Putra Research Grants, GP/2017/9560000 and GP/2017/9559200. The Higher Education Development Program (HEDP)-Afghanistan supported Tahera Hashimi.

REFERENCES

Agrawal, P., Thakur, Z., & Kulharia, M. (2013). Homology modelling and structural validation of tissue factor pathway inhibitor. *Bioinformation*, *9*(16), 808-812. https://doi.org/10.6026/97320630009808

Ahmad, T. A., Eweida, A. E., & El-Sayed, L. H. (2016). T-cell epitope mapping for the design of powerful vaccines. *Vaccine Reports*, *6*, 13-22. https://doi.org/10.1016/j.vacrep.2016.07.002

Al-Hasani, K., Boyce, J., McCarl, V. P., Bottomley, S., Wilkie, I., & Adler, B. (2007). Identification of novel immunogens in *Pasteurella multocida*. *Microbial Cell Factories*, *6*, 3. https://doi.org/10.1186/1475-2859-6-3

Arévalo-Herrera, M., Valencia, A. Z., Vergara, J., Bonelo, A., Fleischhauer, K., González, J. M., Restrepo, J. C., López, J. A., Valmori, D., Corradin, G., & Herrera, S. (2002). Identification of HLA-A2 restricted $CD8^+$ T-lymphocyte responses to *Plasmodium vivax* circumsporozoite protein in individuals naturally exposed to malaria. *Parasite Immunology*, *24*(3), 161-169. https://doi.org/10.1046/j.1365-3024.2002.00449.x

Azam, F. M., Zamri-Saad, M., Rahim, R. A., Chumnanpoend, P., E-kobon, T., & Othman, S. (2020). Antigenic outer membrane proteins prediction of *Pasteurella multocida* serotype B:2. *Asia-Pacific Journal of Molecular Biology and Biotechnology*, *28*(4), 102-116. https://doi.org/10.35118/apjmbb.2020.028.4.09

Balch, W. E., Morimoto, R. I., Dillin, A., & Kelly, J. W. (2008). Adapting proteostasis for disease intervention. *Science*, *319*(5865), 916–919. https://doi.org/10.1126/science.1141448

Baliga, P., Shekar, M., & Venugopal, M. N. (2018). Potential outer membrane protein candidates for vaccine development against the pathogen *Vibrio anguillarum*: A reverse vaccinology based identification. *Current Microbiology*, *75*(3), 368-377. https://doi.org/10.1007/s00284-017-1390-z

Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, *27*(3), 343-350. https://doi.org/10.1093/bioinformatics/btq662

Benkert, P., Künzli, M., & Schwede, T. (2009). QMEAN server for protein model quality estimation. *Nucleic Acids Research*, *37*(suppl_2), W510-W514. https://doi.org/10.1093/nar/gkp322

Benkirane, A., & De Alwis, M. C. L. (2002). Haemorrhagic septicaemia, its significance, prevention and control in Asia. *Veterinary Medicine Journal*, *47*(8), 234-240. https://doi.org/10.17221/5830-VETMED

Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D., & Zardecki, C. (2002). The protein data bank. *Acta Crystallographica Section D: Biological Crystallography*, *58*(6), 899-907. https://doi.org/10.1107/S0907444902003451

Bhattacharya, M., Sharma, A. R., Sharma, G., Patra, P., Mondal, N., Patra, B. C., Lee, S.-S., & Chakraborty, C. (2020). Computer aided novel antigenic epitopes selection from the outer membrane protein sequences of *Aeromonas hydrophila* and its analyses. *Infection, Genetics and Evolution*, *82*, 104320. https://doi.org/10.1016/j.meegid.2020.104320

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., & Schwede, T. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, *42*(W1), W252-W258. https://doi.org/10.1093/nar/gku340

Botos, I., Noinaj, N., & Buchanan, S. K. (2017). Insertion of proteins and lipopolysaccharide into the bacterial outer membrane. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1726), 20160224. https://doi.org/10.1098/rstb.2016.0224

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST^{+}: Architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Cascales, E., & Lloubès, R. (2004). Deletion analyses of the peptidoglycan-associated lipoprotein Pal reveals three independent binding sequences including a TolA box. *Molecular Microbiology*, *51*(3), 873-885. https://doi.org/10.1046/j.1365-2958.2003.03881.x

Chan, W. C., Liang, P. H., Shih, Y. P., Yang, U. C., Lin, W. C., & Hsu, C. N. (2010). Learning to predict expression efficacy of vectors in recombinant protein production. *BMC Bioinformatics*, *11*(Suppl 1), S21. https://doi.org/10.1186/1471-2105-11-S1-S21

Chang, S. T., Ghosh, D., Kirschner, D. E., & Linderman, J. J. (2006). Peptide length-based prediction of peptide–MHC class II binding. *Bioinformatics*, *22*(22), 2761-2767. https://doi.org/10.1093/bioinformatics/btl479

Chen, J., Liu, H., Yang, J., & Chou, K. C. (2007). Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids*, *33*(3), 423-428. https://doi.org/10.1007/s00726-006-0485-9

Chin, C. F., Teh, B. A., Anthony, A. A., Aziah, I., Ismail, A., Ong, E. B. B., & Lim, T. S. (2014). Overexpression, purification and validation of antigenic *Salmonella enterica* serovar Typhi proteins identified from LC-MS/MS. *Applied Biochemistry and Biotechnology*, *174*, 1897-1906. https://doi.org/10.1007/s12010-014-1173-y

Chung, S. Y., & Subbiah, S. (1996). A structural explanation for the twilight zone of protein sequence homology. *Structure*, *4*(10), 1123-1127. https://doi.org/10.1016/S0969-2126(96)00119-0

Ciryam, P., Tartaglia, G. G., Morimoto, R. I., Dobson, C. M., & Vendruscolo, M. (2013). Neurodegenerative diseases and widespread aggregation are associated with supersaturated proteins. *Cell Reports*, *5*(3), 781-790. https://doi.org/10.1016/j.celrep.2013.09.043 PMID: 24183671

Collins, R. F., & Derrick, J. P. (2007). Wza: A new structural paradigm for outer membrane secretory proteins?. *Trends in Microbiology*, *15*(3), 96-100. https://doi.org/10.1016/j.tim.2007.01.002

Colovos, C., & Yeates, T. (1993). ERRAT: An empirical atom-based method for validating protein structures. *Protein Science*, *2*(9), 1511-1519. https://doi.org/10.1002/pro.5560020916

Dong, C., Beis, K., Nesper, J., Brunkan-LaMontagne, A. L., Clarke, B. R., Whitfield, C., & Naismith, J. H. (2006). Wza the translocon for *E. coli* capsular polysaccharides defines a new class of membrane protein. *Nature*, *444*(7116), 226-229. https://doi.org/10.1038/nature05267

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*, 4. https://doi.org/10.1186/1471-2105-8-4

EL-Manzalawy, Y., Dobbs, D., & Honavar, V. (2008a). Predicting linear B-cell epitopes using string kernels. *Journal of Molecular Recognition*, *21*(4), 243-255. https://doi.org/10.1002/jmr.893

EL-Manzalawy, Dobbs, D., & Honavar, V. (2008b). Predicting flexible length linear B-cell epitopes. *Computational Systems Bioinformatics*, *7*, 121-132. https://doi.org/10.1142/9781848162648_0011

Fisch, A., Reynisson, B., Benedictus, L., Nicastri, A., Vasoya, D., Morrison, I., Buus, S., Ferreira, B. R., de Miranda Santos, I. K. F., Ternette, N., Connelley, T., & Nielsen, M. (2021). Integral use of immunopeptidomics and immunoinformatics for the characterization of antigen presentation and rational identification of BoLA-DR–presented peptides and epitopes. *Journal of Immunology*, *206*(10), 2489-2497. https://doi.org/10.4049/jimmunol.2001409

Fiser, A., & Šali, A. (2003). Modeller: Generation and refinement of homology-based protein structure models. *Methods in Enzymology*, *374*, 461-491. https://doi.org/10.1016/S0076-6879(03)74020-8

Ford, R. C., Brunkan-LaMontagne, A. L., Collins, R. F., Clarke, B. R., Harris, R., Naismith, J. H., & Whitfield, C. (2009). Structure–function relationships of the outer membrane translocon Wza investigated by cryo-electron microscopy and mutagenesis. *Journal of Structural Biology*, *166*(2), 172-182. https://doi.org/10.1016/j.jsb.2009.02.005

Haddad, Y., Adam, V., & Heger, Z. (2020). Ten quick tips for homology modeling of high- resolution protein 3D structures. *PLOS Computational Biology*, *16*(4), e1007449. https://doi.org/10.1371/JOURNAL.PCBI.1007449

Harper, M., Boyce, J. D., & Adler, B. (2006). *Pasteurella multocida* pathogenesis: 125 years after pasteur. *FEMS Microbiology Letters*, *265*(1), 1–10. https://doi.org/10.1111/j.1574-6968.2006.00442.x

Hasan, M. A., Khan, M. A., Datta, A., Mazumder, M. H. H., & Hossain, M. U. (2015). A comprehensive immunoinformatics and target site study revealed the corner-stone toward Chikungunya virus treatment. *Molecular Immunology*, *65*(1), 189-204. https://doi.org/10.1016/j.molimm.2014.12.013

Hasani, H. J., & Barakat, K. H. (2016). Protein-protein docking: Are we there yet?. In *Methods and algorithms for molecular docking-based drug design and discovery* (pp. 173-195). IGI Global. https://doi.org/10.4018/978-1-5225-0115-2.ch007

Hatfaludi, T., Al-Hasani, K., Gong, L., Boyce, J. D., Ford, M., Wilkie, I. W., Quinsey, N., Dunstone, M. A. Hoke, D. E., & Adler, B. (2012). Screening of 71 *P. multocida* proteins for protective efficacy in a fowl cholera infection model and characterization of the protective antigen PlpE. *PLOS One*, *7*(7), e39973. https://doi.org/10.1371/journal.pone.0039973

Hensen, U., Meyer, T., Haas, J., Rex, R., Vriend, G., & Grubmüller, H. (2012). Exploring protein dynamics space: The dynasome as the missing link between protein structure and function. *PLOS One*, *7*(5), e33931. https://doi.org/10.1371/journal.pone.0033931

Jabeen, S., Yap, H. Y., Abdullah, F. F. J., Zakaria, Z., Isa, N. M., Tan, Y. C., Joo, Y. S., Satharasinghe, D. A., & Omar, A. R. (2019). Complete genome sequence analysis and characterization of selected iron regulation genes of *Pasteurella multocida* serotype A strain *PMTB2.1*. *Genes*, *10*(2), 81. https://doi.org/10.3390/genes10020081

Jabeen, S., Yong, Y. H., Abdullah, F. J. F., Zakaria, Z., Mat Isa, N., Tan, Y. C., Yee, W. Y., &

Omar, A. R. (2017). Complete genome sequence of *Pasteurella multocida* serotype A strain PMTB2.1 isolated from buffaloes that died of septicemia in Malaysia. *Genome Announcements*, *5*(44), e01190-17. https://doi.org/10.1128/genomeA.01190-17

Jenkins, M. K., Khoruts, A., Ingulli, E., Mueller, D. L., McSorley, S. J., Reinhardt, R. L., Itano, A., & Pape, K. A. (2003). *In vivo* activation of antigen-specific CD4 T cells. *Annual Reviews of Immunology*, *19*, 23–45. https://doi.org/10.1146/annurev.immunol.19.1.23

Jiménez-García, B., Roel-Touris, J., Romero-Durana, M., Vidal, M., Jiménez-González, D., & Fernández-Recio, J. (2018). LightDock: A new multi-scale approach to protein–protein docking. *Bioinformatics*, *34*(1), 49–55. https://doi.org/10.1093/bioinformatics/btx555

Jones, P. M., & George, A. M. (2004). The ABC transporter structure and mechanism: Perspectives on recent research. *Cellular and Molecular Life Sciences*, *61*(6), 682-699. https://doi.org/10.1007/s00018-003-3336-9

Joshi, S., Tewari, K., & Singh, R. (2013). Comparative immunogenicity and protective efficacy of different preparations of outer membrane proteins of *Pasteurella multocida* (B:2) in a mouse model. *Veterinarski Arhiv*, *83*(6), 665-676.

Karami, Y., Rey, J., Postic, G., Murail, S., Tufféry, P., & De Vries, S. J. (2019). DaReUS- Loop: A web server to model multiple loops in homology models. *Nucleic Acids Research*, *47*(W1), W423-W428. https://doi.org/10.1093/nar/gkz403

Kim, M., Song, L., Moon, J., Sun, Z. Y., Bershteyn, A., Hanson, M., Cain, D., Goka, S., Kelsoe, G., Wagner, G., Irvine, D., & Reinherz, E. L. (2013). Immunogenicity of membrane-bound HIV-1 gp41 membrane-proximal external region (MPER) segments is dominated by residue accessibility and modulated by stereochemistry. *Journal of Biological Chemistry*, *288*(44), 31888–31901. https://doi.org/10.1074/jbc.M113.494609

Kong, L., Harrington, L., Li, Q., Cheley, S., Davis, B. G., & Bayley, H. (2013). Single-molecule interrogation of a bacterial sugar transporter allows the discovery of an extracellular inhibitor. *Nature Chemistry*, *5*(8), 651–659. https://doi.org/10.1038/nchem.1695

Kumar, A., Mohanty, N. N., Chacko, N., Yogisharadhya, R., & Shivachandra, S. B. (2015). Structural features of a highly conserved Omp16 protein of *Pasteurella multocida* strains and comparison with related peptidoglycan-associated lipoproteins (PAL). *Indian Journal of Microbiology*, *55*(1), 50-56. https://doi.org/10.1007/s12088-014-0489-1

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283-291. https://doi.org/10.1107/S0021889892009944

Lazzaroni, J. C., & Portalier, R. (1992). The excC gene of *Escherichia coli* K-12 required for cell envelope integrity encodes the peptidoglycan-associated lipoprotein (PAL). *Molecular Microbiology*, *6*(6), 735-742. https://doi.org/10.1111/j.1365-2958.1992.tb01523.x

Liu, N., & Xu, Z. (2019). Using LeDock as a docking tool for computational drug design. In *IOP Conference Series: Earth and Environmental Science* (Vol. 218, No. 1, p. 012143). IOP Publishing. https://doi.org/10.1088/1755-1315/218/1/012143

Liu, W., Yang, M., Xu, Z., Zheng, H., Liang, W., Zhou, R., Wu, B., & Chen, H. (2012). Complete genome sequence of *Pasteurella multocida* HN06, a toxigenic strain of serogroup D. *Journal of Bacteriology*, *194*(12), 3292-3293. https://doi.org/10.1128/JB.00215-12

Magnan, C. N., Zeller, M., Kayala, M. A., Vigil, A., Randall, A., Felgner, P. L., & Baldi, P. (2010). High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics*, *26*(23), 2936-2943. https://doi.org/10.1093/bioinformatics/btq551

Martz, F. (2014). *Développement d'une nouvelle méthode de docking basée sur les mécanismes enzymatiques et guidée par des groupes prosthétiques* [Development of a new mechanism-based molecular docking method guided by prosthetic groups] [Doctoral dissertation, Université Paris Sud-Paris XI]. HAL Portal Theses. https://tel.archives-ouvertes.fr/tel-01168482

Mathelié-Guinlet, M., Asmar, A. T., Collet, J. F., & Dufrêne, Y. F. (2020). Bacterial cell mechanics beyond peptidoglycan. *Trends in Microbiology*, *28*(9), 706-708. https://doi.org/10.1016/j.tim.2020.04.013

May, B. J., Zhang, Q., Li, L. L., Paustian, M. L., Whittam, T. S., & Kapur, V. (2001). Complete genomic sequence of *Pasteurella multocida*, Pm70. *Proceedings of the National Academy of Sciences*, *98*(6), 3460-3465. https://doi.org/10.1073/pnas.051634598

Muenthaisong, A., Nambooppha, B., Rittipornlertrak, A., Tankaew, P., Varinrak, T., Muangthai, K., Atthikanyaphak, K., Sawada, T., & Sthitmatee, N. (2020). An intranasal vaccination with a recombinant outer membrane protein H against haemorrhagic septicemia in swamp buffaloes. *Veterinary Medicine International*, *2020*, 3548973. https://doi.org/10.1155/2020/3548973

Naz, A., Awan, F. M., Obaid, A., Muhammad, S. A., Paracha, R. Z., Ahmad, J., & Ali, A. (2015). Identification of putative vaccine candidates against *Helicobacter pylori* exploiting exoproteome and secretome: A reverse vaccinology based approach. *Infection, Genetics and Evolution*, *32*, 280-291. https://doi.org/10.1016/j.meegid.2015.03.027

Park, H., Lee, G. R., Heo, L., & Seok, C. (2014). Protein loop modeling using a new hybrid energy function and its application to modeling in inaccurate structural environments. *PLOS One*, *9*(11), e113811. https://doi.org/10.1371/journal.pone.0113811

Patronov, A., & Doytchinova, I. (2013). T-cell epitope vaccine design by immunoinformatics. *Open Biology*, *3*(1), 120139. https://doi.org/10.1098/rsob.120139

Peng, Z., Liang, W., Liu, W., Wu, B., Tang, B., Tan, C., Zhou, R., & Chen, H. (2016). Genomic characterization of *Pasteurella multocida* HB01, a serotype A bovine isolate from China. *Gene*, *581*(1), 85-93. https://doi.org/10.1016/j.gene.2016.01.041

Peng, Z., Liang, W., Wang, F., Xu, Z., Xie, Z., Lian, Z., Hua, L., Zhou, Z., Chen, H., & Wu, B. (2018). Genetic and phylogenetic characteristics of *Pasteurella multocida* isolates from different host species. *Frontiers in Microbiology*, *9*, 1408. https://doi.org/10.3389/fmicb.2018.01408

Peng, Z., Wang, X., Zhou, R., Chen, H., Wilson, B. A., & Wu, B. (2019). *Pasteurella multocida*: Genotypes and genomics. *Microbiology and Molecular Biology Reviews*, *83*(4), e00014-19. https://doi.org/10.1128/MMBR.00014-19

Prajapat, R., Marwal, A., & Gaur, R. K. (2014). Recognition of errors in the refinement and validation of three-dimensional structures of AC1 proteins of *Begomovirus* strains by using ProSA-web. *Journal of Viruses*, *2014*, 752656. https://doi.org/10.1155/2014/752656

Prasannavadhana, A., Kumar, S., Thomas, P., Sarangi, L. N., Gupta, S. K., Priyadarshini, A., Nagaleekar, V. K., & Singh, V. P. (2014). Outer membrane proteome analysis of Indian strain of *Pasteurella multocida* serotype B:2 by MALDI-TOF/MS analysis. *The Scientific World Journal*, *2014*, 617034. https://doi.org/10.1155/2014/617034

Qiao, S., Luo, Q., Zhao, Y., Zhang, X. C., & Huang, Y. (2014). Structural basis for lipopolysaccharide insertion in the bacterial outer membrane. *Nature*, *511*(7507), 108-111. https://doi.org/10.1038/nature13484

Rafidah, O., Zamri-Saad, M., Nasip, E., Shahiruddin, S., & Saharee, A. A. (2010). Analysis of haemorrhagic septicaemia outbreaks in cattle and buffalo in Malaysia. *Online Journal of Veterinary Research*, *14*(2), 325-333.

Rahman, M. S., Biswas, C., Biswas, P. K., Kader, M. A., Alam, S. N., Sonne, C., & Kim, K. H. (2020). *In silico* analysis of the antigenic properties of iron-regulated proteins against *Neisseria meningitidis*. *Applied Sciences*, *10*(17), 6113. https://doi.org/10.3390/app10176113

Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, *7*(1), 95–99. https://doi.org/10.1016/S0022-2836(63)80023-6

Reid, A. N., & Whitfield, C. (2005). Functional analysis of conserved gene products involved in assembly of *Escherichia coli* capsules and exopolysaccharides: Evidence for molecular recognition between Wza and Wzc for colanic acid biosynthesis. *Journal of Bacteriology*, *187*(15), 5470-5481. https://doi.org/10.1128/JB.187.15.5470-5481.2005

Remmert, M., Biegert, A., Hauser, A., & Söding, J. (2012). HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods*, *9*, 173-175. https://doi.org/10.1038/nmeth.1818

Rimler, R. B. (2001). Purification of a cross-protective antigen from *Pasteurella multocida* grown *in vitro* and *in vivo*. *Avian Diseases*, *45*(3), 572-580. https://doi.org/10.2307/1592897

Rita, D. V., Swee, K. C. W., Shamini, C., Kang, T. L., Nurshamimi, N. R., Hussin, A. R., Nurul, K., & Salmah, I. (2018). A recombinant subunit HSABA392 as a potential vaccine for haemorrhagic septicaemia disease in livestock. *Tropical Biomedicine*, *35*(4), 1075-1086.

Sailapathi, A., Murugan, G., Somarathinam, K., Gunalan, S., Jagadeesan, R., Yoosuf, N., Kanagaraj, S., & Kothandan, G. (2020). Proposing the promiscuous protein structures in JNK1 and JNK3 for virtual screening in pursuit of potential leads. *ACS Omega*, *5*(8), 3969–3978. https://doi.org/10.1021/acsomega.9b03458

Sarah, S. O., Zamri-Saad, M., Zunita, Z., & Raha, A. R. (2006). Molecular cloning and sequence analysis of *gdhA* gene of *Pasteurella multocida* B:2. *Journal of Animal and Veterinary Advances*, *5*(12), 1146- 1149.

Shen, W., Cao, Y., Cha, L., Zhang, X., Ying, X., Zhang, W., Ge, K., Li, W., & Zhong, L. (2015). Predicting linear B-cell epitopes using amino acid anchoring pair composition. *BioData Mining*, *8*, 14. https://doi.org/10.1186/s13040-015-0047-3

Sippel, K. H., Robbins, A. H., Reutzel, R., Boehlein, S. K., Namiki, K., Goodison, S., Agbandje, M. M., Rosser, C. J., & McKenna, R. (2009). Structural insights into the extracytoplasmic thiamine-binding lipoprotein p37 of *Mycoplasma hyorhinis*. *Journal of Bacteriology*, *191*(8), 2585-2592. https://doi.org/10.1128/JB.01680-08

Sollner, J., Grohmann, R., Rapberger, R., Perco, P., Lukas, A., & Mayer, B. (2008). Analysis and prediction of protective continuous B-cell epitopes on pathogen proteins. *Immunome Research*, *4*, 1. https://doi.org/10.1186/1745-7580-4-1

Soriano, E. V., Rajashankar, K. R., Hanes, J. W., Bale, S., Begley, T. P., & Ealick, S. E. (2008). Structural similarities between thiamin-binding protein and thiaminase-I suggest a common ancestor. *Biochemistry*, *47*(5), 1346-1357. https://doi.org/10.1021/bi7018282

Trolle, T., McMurtrey, C. P., Sidney, J., Bardet, W., Osborn, S. C., Kaever, T., Sette, A., Hildebrand, W. H., Nielsen, M., & Peters, B. (2016). The length distribution of class I–restricted T cell epitopes is determined by both peptide supply and MHC allele–specific binding preference. *The Journal of Immunology*, *196*(4), 1480-1487. https://doi.org/10.4049/jimmunol.1501721

Villa, R., Martorana, A. M., Okuda, S., Gourlay, L. J., Nardini, M., Sperandeo, P., Dehò, G., Bolognesi, M., Kahne, D., & Polissi, A. (2013). The *Escherichia coli* Lpt transenvelope protein complex for lipopolysaccharide export is assembled via conserved structurally homologous domains. *Journal of Bacteriology*, *195*(5), 1100-1108. https://doi.org/10.1128/JB.02057-12

Wang, J., & Dokholyan, N. V. (2019). MedusaDock 2.0: Efficient and accurate protein–ligand docking with constraints. *Journal of Chemical Information and Modeling*, *59*(6), 2509–2515. https://doi.org/10.1021/acs.jcim.8b00905

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore. R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296-W303. https://doi.org/10.1093/nar/gky427

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(suppl_2), W407-W410. https://doi.org/10.1093/nar/gkm290

Zamri-Saad, M., & Annas, S. (2016). Vaccination against hemorrhagic septicemia of bovines: A review. *Pakistan Veterinary Journal*, *36*(1), 1-5.

APPENDICES

Supplementary Material 1

B-cell Epitope Predictions for Pal, LptD, Wza, and TbpA

Table I
B-cell epitope prediction within the amino acid sequence of Pal using the AAP prediction module in the BCPREDS server

Start position	Sequence
36	YSVQDLQQRYNT
75	TPATKVVVEGNT
108	HYLSAKGVQAGQ
21	GSSKKDESAGQM

Note. AAP = Amino acid pair prediction module in the BCPREDS server

Table II
B-cell epitope prediction within the amino acid sequence of LptD using the AAP prediction module in the BCPREDS server

Start position	Sequence
296	QGRAAEAEQVER
247	QAVHFDNKSPLM
360	NNAYLGFGYDSS
402	GTTRIYDKKGNE
375	QDYFGLFRDRRY
173	SDFDSAYGSSTD
231	YQNALFDSPLNF
159	LNLDYTRVSDQR
576	QTANQIFYDKSI
11	PQDNAWSIEAKE
418	SAGQIYYLQDSR
191	FRVAYYQPHYNI

Note. AAP = Amino acid pair prediction module in the BCPREDS server

Table III
B-cell epitope prediction within the amino acid sequence of Wza using the AAP prediction module in the BCPREDS server

Start position	Sequence
33	GVKVYAQGGPLD
238	RLLVDGDIVHVP
78	KIKAYQYRVGPG
95	TVWDHPELTTPA
339	GTEFYLKPYDVV
172	AYITGEVNRPGQ
369	PTLSGFDSITES
298	FVIRGQRSPSTT
259	IGEVAQPQMLKI

Note. AAP = Amino acid pair prediction module in the BCPREDS server

Table IV
B-cell epitope prediction within the amino acid sequence of TbpA using the AAP prediction module in the BCPREDS server

Start position	Sequence
178	HVYTENEVAQAW
295	EPHFDALKATQM
192	LAKHTVTVGKGW
134	IYDKTKVQQPPK
89	DNYNLEEAEKSG
24	TQAVNVYTYDSF
161	QDPRTSSVGRGL
76	EGKKTKADVVVG

Note. AAP = Amino acid pair prediction module in the BCPREDS server

Supplementary Material 2

MHC-I and MHC-II Binding Predictions for Pal, LptD, Wza, and TbpA

Table V
MHC-I and MHC-II binding peptide sequence prediction within the amino acid sequence of Pal through NetMHCpan 4.1 and NetBoLAIIpan 1.0, respectively

MHC	Start position	Peptide sequence
MHC-I: BoLA-1:00901	41	LQQRYNTVY
MHC-1: BoLA-1:02301	42	QQRYNTVYF
	1	MKKLTKVLL
	41	LQQRYNTVY
MHC-1: BoLA-2:01201	121	STVSYGEEK
MHC-1: BoLA-2:03001	102	RADAVKHYL
	125	YGEEKPAVL
	140	YSKNRRAVL
MHC-1: BoLA-3:00101	125	YGEEKPAVL
	90	GTPEYNIAL
	102	RADAVKHYL
MHC-1: BoLA-3:00201	90	GTPEYNIAL
MHC-II: BoLA-DRB3_1001	46	NTVYFGFDKYNIE
	137	EAAYSKNRRAVLA
MHC-II: BoLA-DRB3_1201	58	EGEYVQILDAHAA
MHC-II: BoLA-DRB3_1601	58	EGEYVQILDAHAA
MHC-II: BoLA-DRB3_2002	59	GEYVQILDAHAAF
	58	EGEYVQILDAHAA

Note. BoLA = Bovine leukocyte antigen

Table VI
MHC-I and MHC-II binding peptide sequence prediction within the amino acid sequence of LptD through NetMHCpan 4.1 and NetBoLAIIpan 1.0, respectively

MHC	Start position	Peptide sequence
MHC-I: BoLA-1:00901	326	SQQTFIEGY
	529	DQWAVVARH
	542	ALRKPVEQY
	234	ALFDSPLNF
	40	KVHGVPIFY
	105	WQLNGEFRY
MHC-1: BoLA-1:02301	612	GKLPYLQAF
	316	VKVDLQTVL
	234	ALFDSPLNF
	609	LQRGKLPYL
	39	FKVHGVPIF

Table VI *(continue)*

MHC	Start position	Peptide sequence
	326	SQQTFIEGY
	207	KQFQIFNEV
	542	ALRKPVEQY
	517	KQLGLTVAW
MHC-I: BoLA-2:01201	463	YQYDTRLNK
MHC-I: BoLA-2:03001	614	LPYLQAFSL
	44	VPIFYTPYL
	347	RPYKDQSNI
	545	KPVEQYLGL
MHC-I: BoLA-3:00101	87	IAPNLDATM
	48	YTPYLQLPI
	264	FHAEPSINL
	585	KSIGINIEL
	289	VHFDNKSPL
MHC-I: BoLA-3:00201	87	IAPNLDATM
	585	KSIGINIEL
	48	YTPYLQLPI
MHC-II: BoLA-DRB3_0101	339	EPHVQYLYRPYKD
	307	RKINRVLPQVKVD
MHC-II: BoLA-DRB3_1001	207	KQFQIFNEVDIGP
	225	QVDFNYYQNALFD
	217	IGPYRALPQVDFN
	45	PIFYTPYLQLPIG
	193	VAYYQPHYNIAIS
	28	QEEYAEMWHARFK
	36	HARFKVHGVPIFY
	309	INRVLPQVKVDLQ
	239	PLNFKLFSQAVHF
	528	TDQWAVVARHYQD
	448	ASNWKINDQWRWQ
	108	NGEFRYLSPIGEG
MHC-II: BoLA-DRB3_1101	108	NGEFRYLSPIGEG
	119	EGKIAGEYLKQDR
	241	NFKLFSQAVHFDN
	321	QTVLASQQTFIEG
MHC-II: BoLA-DRB3_1201	225	QVDFNYYQNALFD
	343	QYLYRPYKDQSNI
	419	AGQIYYLQDSRID
	207	KQFQIFNEVDIGP

Table VI *(continue)*

MHC	Start position	Peptide sequence
	375	QDYFGLFRDRRYS
	160	NLDYTRVSDQRYF
	45	PIFYTPYLQLPIG
	217	IGPYRALPQVDFN
MHC-II: BoLA-DRB3_1601	419	AGQIYYLQDSRID
	160	NLDYTRVSDQRYF
	261	AWRFHAEPSINLP
	168	DQRYFSDFDSAYG
	82	PFYWNIAPNLDAT
	578	ANQIFYDKSIGIN
	197	QPHYNIAISAKQF
MHC-II: BoLA-DRB3_2002	108	NGEFRYLSPIGEG
	217	IGPYRALPQVDFN

Note. BoLA = Bovine leukocyte antigen

Table VII
MHC-I and MHC-II binding peptide sequence prediction within the amino acid sequence of Wza through NetMHCpan 4.1 and NetBoLAIIpan 1.0, respectively

MHC-class: BoLA	Position	Peptide sequence
MHC-I: BoLA-1:00901	60	KQLAPSIPF
	112	AAESGSQVH
	165	ATYQSKKAY
	202	GLSEHADWH
	239	LLVDGDIVH
MHC-I: BoLA-1:02301	2	YKLSRLLIL
	45	AQKAVDAYL
	60	KQLAPSIPF
	76	DQKIKAYQY
	165	ATYQSKKAY
	183	QQYLTNVPL
	263	AQPQMLKIM
	343	YLKPYDVVY
MHC-I: BoLA-2:01201	20	YLKPYDVVY
	70	RANPPLDQK
	138	LTVSQIRNK
	141	SQIRNKLTK
	162	VSVATYQSK
	163	SVATYQSKK
	223	ISVEALIQR
	313	TSSEENIEK

Table VII *(continue)*

MHC-class: BoLA	Position	Peptide sequence
MHC-I: BoLA-2:03001	29	SPVSGVKVY
	50	DAYLITPSL
	111	SAAESGSQV
	127	YPYVGSIHV
	206	HADWHNVTL
	363	VISQIVPTL
MHC-I: BoLA-3:00101	28	KSPVSGVKV
	54	ITPSLVKQL
	71	ANPPLDQKI
	86	VGPGDVLNI
	130	VGSIHVSGL
	178	VNRPGQQYL
	206	HADWHNVTL
	231	RGDLSQNRL
	263	RGDLSQNRL
	273	YGMTLTEAI
MHC-I: BoLA-3:00201	28	KSPVSGVKV
	54	ITPSLVKQL
	71	ANPPLDQKI
	86	VGPGDVLNI
	130	VGSIHVSGL
	263	AQPQMLKIM
	273	YGMTLTEAI
MHC-II:BoLA-DRB3_0101	59	VKQLAPSIPFARA
	60	KQLAPSIPFARAN
	61	QLAPSIPFARANP
	134	HVSGLTVSQIRNK
	135	VSGLTVSQIRNKL
	136	SGLTVSQIRNKLT
	137	GLTVSQIRNKLTK
	159	QIEVSVATYQSKK
	374	FDSITESMLRIRN
	375	DSITESMLRIRNW
MHC-II:BoLA-DRB3_1001	90	DVLNITVWDHPEL
	91	VLNITVWDHPELT
	92	LNITVWDHPELTT
	93	NITVWDHPELTTP
	151	LANYISEPQIEVS
	152	ANYISEPQIEVSV
	340	TEFYLKPYDVVYV

Table VII *(continue)*

MHC-class: BoLA	Position	Peptide sequence
	341	EFYLKPYDVVYVT
	342	FYLKPYDVVYVTT
MHC-II: BoLA-DRB3_1101	322	IADIYQLDVTDAT
MHC-II: BoLA-DRB3_1201	152	LANYISEPQIEVS
	153	ANYISEPQIEVSV
MHC-II:BoLA-DRB3_1501	20	CSVMLPSTKSPVS
	21	SVMLPSTKSPVSG
	345	KPYDVVYVTTAPV
	346	PYDVVYVTTAPVA
	347	YDVVYVTTAPVAR
MHC-II: BoLA-DRB3_1601	151	LANYISEPQIEVS
	152	ANYISEPQIEVSV
	153	NYISEPQIEVSVA
MHC-II: BoLA-DRB3_2002	294	ATGIFVIRGQRSP

Note. BoLA = Bovine leukocyte antigen

Table VIII
MHC-I and MHC-II binding peptide sequence prediction within the amino acid sequence of TbpA through NetMHCpan 4.1 and NetBoLAIIpan 1.0, respectively

MHC	Start position	Peptide sequence
MHC-I: BoLA-1:00901	22	AQTQAVNVY
	24	TQAVNVYTY
	44	KVKKAFETH
	97	EKSGLFVQH
	187	QAWQKLAKH
	213	GEADVVLSY
	219	LSYNTSPLY
	229	MVFEQKDQY
	251	ETAARVAQH
	271	LIHPEAQGH
	289	VINTNIEPH
MHC-I: BoLA-1:02301	22	AQTQAVNVY
	24	TQAVNVYTY
	45	VKKAFETHF
	80	TKADVVVGL
	118	WKNQTFLPY
	160	YQDPRTSSV
	213	GEADVVLSY
MHC-I: BoLA-2:01201	98	KSGLFVQHK
	137	KTKVQQPPK

Table VIII *(continue)*

MHC	Start position	Peptide sequence
	140	VQQPPKSLK
	145	KSLKELVER
	219	LSYNTSPLY
	315	KVNAEQVKK
MHC-I: BoLA-3:00101	5	KTSFFFTAL
	41	AGPKVKKAF
	99	SGLFVQHKV
	142	QPPKSLKEL
	223	TSPLYHMVF
	272	IHPEAQGHL
MHC-I:BoLA-2:03001	106	KVDLTPLSL
	110	TPLSLPVEW
	139	KVQQPPKSL
	142	QPPKSLKEL
	160	YQDPRTSSV
	223	TSPLYHMVF
MHC-I:BoLA-3:00201	5	KTSFFFTAL
	41	AGPKVKKAF
	223	TSPLYHMVF
MHC-II: BoLA-DRB3_0101	190	QKLAKHTVTVGKG
	291	NTNIEPHFDALKA
	292	TNIEPHFDALKAT
	293	NIEPHFDALKATQ
	298	FDALKATQMNTKV
	299	DALKATQMNTKVL
MHC-II: BoLA-DRB3_1001	97	EKSGLFVQHKVDL
	98	KSGLFVQHKVDLT
	99	SGLFVQHKVDLTP
	120	NQTFLPYDFGQFA
	203	WSDTYGAFLKGEA
	204	SDTYGAFLKGEAD
	205	DTYGAFLKGEADV
	322	KKWIAVWQTILTQ
MHC-II: BoLA-DRB3_1101	105	HKVDLTPLSLPVE
	106	KVDLTPLSLPVEW
	107	VDLTPLSLPVEWK
	108	DLTPLSLPVEWKN
	263	CADHFLAFLIHPE
	264	ADHFLAFLIHPEA
	265	DHFLAFLIHPEAQ

Table VIII *(continue)*

MHC	Start position	Peptide sequence
MHC-II: BoLA-DRB3_1201	57	QVNFTAFGDSGTM
	58	VNFTAFGDSGTMF
	99	SGLFVQHKVDLTP
	100	GLFVQHKVDLTPL
MHC-II: BoLA-DRB3_1501	90	NYNLEEAEKSGLF
	91	YNLEEAEKSGLFV
	247	VLQIETAARVAQH
	248	LQIETAARVAQHD
MHC-II: BoLA-DRB3_1601	130	QFAFIYDKTKVQQ
	131	FAFIYDKTKVQQP
	208	GAFLKGEADVVLS
MHC-II: BoLA-DRB3_2002	7	SFFFTALSTLSLS
	294	IEPHFDALKATQM
	295	EPHFDALKATQMN
	296	PHFDALKATQMNT

Note. BoLA = Bovine leukocyte antigen