Immunoinformatics and System Biology Approaches for Potential Vaccine Candidates Against Burkholderia Pseudomallei

Asfiha Tarannum  
University of Dhaka Faculty of Biological Sciences

Md. Arif Khan  
University of Development Alternative

Ahmad S. Sirajee  
University of Dhaka Faculty of Biological Sciences

Mahmuda Yasmin  
University of Dhaka Faculty of Biological Sciences

Chowdhury R. Ahsan (✉ crahsan@du.ac.bd)  
Dhaka University  https://orcid.org/0000-0001-5024-6293

Research article

Keywords: Burkholderia pseudomallei, Bioinformatics, Homology modelling, Molecular docking, Molecular dynamics simulation, System biology.

DOI: https://doi.org/10.21203/rs.3.rs-38222/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** *Burkholderia pseudomallei*, an intracellular gram-negative bacterium, is the causative agent of melioidosis. It is a highly infectious disease that usually involves the lungs of humans as well as animals endemic in South-East Asia and Northern Australia. At present, we do not have any vaccine or treatment scheme at our disposal which is absolutely effective against this disease. There have been few advances in the development of vaccines against intracellular bacterial pathogens, making vaccine development against intracellular pathogen a more challenging arena. So, we opted for in-silico methods of drug designing intending to combat *B. pseudomallei*.

**Results:** The whole proteome of *B. pseudomallei* was analyzed for determining immunogenic proteins. Combining B-cell and T-cell epitope prediction studies, the sequence ETAAADALY was considered as the most potential epitope for both T and B cells, followed by molecular docking against an MHC (major histocompatibility complex) Class I molecule, HLA-A*02:01 (human leukocyte antigens encoded by the HLA-A locus). Apart from epitope prediction, hydrogen bond study and target site analysis were carried out which concludes that ARG^{237}, ALA^{234}, ARG^{236}, ARG^{262}, GLN^{226}, ASP^{229}, VAL^{230}, VAL^{239}, GLY^{258}, LEU^{259}, ASN^{225}, and ILE^{254} consist the best active site of the protein molecule. The study of the bacterial antigenic protein model’s secondary structure and stereochemical properties provided an insight into the protein's stability as our epitope of choice. The post docking interactions were further subjected to molecular dynamics simulation and the system biology approach for validation.

**Conclusion:** The goal of our endeavor is to delve into an integrative Immunoinformatics study combined with the system biology logistics of *B. pseudomallei* to detect potential vaccine candidates against this pathogen as well as providing a depiction of the bacterial immunome that could be insightful for vaccine generation in the future.

**Background**

Melioidosis, also recognized as Whitmore's disease, is a respiratory infection caused by the Gram-negative, intracellular bacterium *Burkholderia pseudomallei*. Commonly widespread in tropical climates, specifically in Southeast Asia and northern Australia, the bacteria thrive in polluted water and muddy ground and contract its hosts directly. The primary route of infection is assumed to take place through percutaneous inoculation upon getting exposed to contaminated muddy ground or surface water, as in agricultural fields, in endemic areas [1]. Several environmental conditions are commonly believed to trigger melioidosis which involves inhalation as the key mode of pathogenesis, namely tropical storms, cyclones, typhoons, floods, none of which are uncommon to Bangladesh [2]. This notion is signposted by the fact that during the rainy season melioidosis could turn more serious due to prolonged exposure to the pathogen owing to increased odds of contracting the disease. Since heavy rain and aerosol-producing winds may encourage frequent inhalation of fomites, the dose of infection often rises accordingly [3]. As we couldn’t find a way around this intractable airborne pathogen, it is now deemed a biological weapon in the making. US Centers for Disease Control and Prevention designated it into the list of category B agents.
The treatment scheme for *melioidosis* is co-administering a combination of wide ranges of antibiotics. Unfortunately, the pathogen's inherent resistance pattern to an array of antibiotics has prompted the disease to relapse and reactivate after the termination of antibiotic therapy [4] due to a lack of vaccines available against *B. pseudomallei*.

Intracellular pathogens hide inside specific host cells to [5] escape the immune system. Additionally, a commonly practiced notion is that cellular immunity is indispensable to the defense against intracellular bacterial pathogens [6], in particular those involving CD4 + and CD8 + T cells [7]. A key factor in determining the type of vaccine that should be produced is the inference that specific immune responses are needed for defense against different pathogens. To illustrate, vaccines based on killed whole cells, or isolated protein or polysaccharide fractions of cells, may stimulate strong antibody responses but may prove quite futile in inducing cell-mediated immunity [8]. In comparison, live attenuated mutants of virulent strains invoke a varied extent of immune responses including cytotoxic T cell-mediated immunity. This is deemed the best route for a vaccine to be effectively working against intracellular pathogens.

Here we tried to employ the idea of docking simulation to procure a clear conception about peptide vaccine development against *B. pseudomallei*. As observed in the cases of intracellular Gram-negative Category B pathogen, it is quite difficult to find a bacterial component against which we can develop an effective vaccine. *B. pseudomallei* is composed of a variety of PAMPs (Pathogen Associated Molecular Pattern) including, lipid-A, lipopolysaccharide (LPS), flagellin, peptidoglycan, DNA and type III secretion system (TTSS), which can be possibly recognized well by different TLRs (Toll-like Receptors) [9]. Analyzing thoroughly the putative immunogenic proteins enabling *B. pseudomallei* to possess a wide range of cellular function and invasion strategies, we chose the most virulent proteins from the immune-sera expressed during infection [10]. Peptide vaccine development study was further carried out by conducting a docking and dynamics simulation study between the target epitope derived from the selected immunogenic protein and a class 1 MHC molecule, as we intend to trigger cell-mediated immunity. Docking and Molecular Dynamics simulation study would provide us a 3D mechanism and range of efficacy of the vaccine in terms of system biology. So, the overall objective of our study was to identify a putative vaccine candidate that would pass through various immunoinformatic screening as well as molecular docking and dynamics simulation to justify the potency of the candidates as efficient vaccines against *B. pseudomallei*.

## Results

### Retrieval of sequences

Consisting of two chromosomes of 4,074,542 bp and 3,173,005 bp, respectively, the reference genome used in this study for *B. pseudomallei* is 7,247,547 bp in size with a (G+C) content of 68.06% [57], as reported by Sanger Institute. On the other hand, the strain used in this study was K96243 (Taxonomy ID: 272560). The proteome sequence was retrieved from both the NCBI GenBank sequence database [58] and UniProt [59].
T cell epitope prediction

The protein sequence having the UniProtKB ID: Q63NM3, NCBI ID (YP_110296.1) achieved one of the highest scores of 0.5693 in the study of VaxiJen throughout the proteome. We chose it after carrying out subcellular localization by PSORTb. It seemed that our selected sequence is of a transmembrane flagellar protein. T cell epitopes of the protein were predicted by the NetCTL server. Three top-ranked epitopes ETTAAADALY, ESGAQTVDY, FADMWATLK were chosen based on their combinatorial score. Wide coverage of the MHC-I molecules coupled with IC_{50} value (half maximal inhibitory concentration) less than 200 nm (IC_{50}<200), guaranteed the selection criteria of the HLA molecules having a higher affinity to the top three epitopes. The identification of epitopes combined with the most effective HLA molecules governs the strength of the immune response. The higher the number of HLA alleles reacting with a peptide, the greater the likelihood of a strong immune response. The 9-mer epitope ETTAAADALY, amongst the three, showed an affinity for the highest numbers (nine) of MHC-I allele HLA-A*26:01. The scores gleaned from MHC-I binding analysis and processing tools of IEDB are assembled in Table 1. For a vaccine scheme based on an epitope, the use of well-conserved epitopes is supposed to provide comparatively wider surveillance across multitudes of strains, than the ones hailing from the highly variable genomic region. That is why epitope conservancy analysis was carried out [21]. Previously recognized optimal MHC-I molecules were further subjected to population coverage analysis of the epitopes in Table 2. The query sequence of the protein Aerotaxis Receptor was considered non-allergen and non-cross-reactive by both AllerTop 1.0 and AllerHunter.

Molecular docking

The epitope sequence ETTAAADALY was surmised to meet the highest number of prerequisites of an ideal vaccine, clearly predominating over other epitopes in terms of epitope conservancy, population coverage, and affinity towards a greater number of MHC molecules. The conformations for the epitope ETTAAADALY to bind HLA molecules were predicted by AutoDockVina. To illustrate, we retrieved the HLA sequence, HLA-A*26:01, attached with an Azo Benzene Containing Peptide crystal structure. Afterward, docking was simulated with Azo Benzene containing peptide epitope from which the HLA molecule was eliminated (Figure 2.A). The aim was to compare the spontaneous binding energy of the native peptide and the target peptide (ETAAADALY). Docking simulation between HLA-A*26:01 and the epitope of our interest (Figure 2.B) demonstrated that the binding energy of our chosen epitope with the MHC-1 receptor molecule was around -6.5 kcal/mol. The binding energy of Azo Benzene containing peptide epitope with retrieved HLA-A*26:01 (-7.1 kcal/mol) was then compared to the binding energy of the previously docked complex between epitope our choice and MHC receptor, which was found to less spontaneous than the binding energy of the chosen epitope-MHC complex. The 3D structures of HLA and the epitope were visualized using PyMOL molecular graphics system.

Hydrogen bond analysis
Molecular docking interactions of our target complex was studied through hydrogen bond analysis. Hydrogen bonds are a pivotal non-covalent structural force, predominant in molecular systems. They are formed through the association of a single hydrogen atom and the heavy atoms sharing the hydrogen atom. In this hypothesis, we studied the hydrogen bonds between the MHC molecule and the chosen epitope. The amino acid residues ASN 62, ASN 65, ASN 76, ARG 96, ARG 162, LYS 145, THR 142, TRP 146 and TYR 6 demonstrated to be involved in hydrogen bonding via Discovery Studio (Figure 3).

**B cell epitope prediction**

Identification of B cell epitopes was carried out by quite a few online tools available in IEDB (shown in Table 3), VaxiJen and TMHMM prediction servers. Surface accessibility of B cell epitopes is an integral part of vaccine designing because, surface hydrophilic regions, upon exposure, are more likely to elicit a spontaneous immune response from B cells. The Emini surface accessibility prediction (Figure 5) and Parker hydrophilicity prediction tools (Figure 4) were utilized as well. The beta-turn protein motifs, hydrophilic by convention and presumably initiating antigenicity, are usually accessible to the surface. At this point, the Chou and Fasman beta-turn predictions were conducted to screen the beta-turn regions within the query protein. Flexible regions within the query protein were also documented using the Karplus Schulz flexibility prediction tool. The epitopes of linear B cells were determined using ABCpred and Bepipred, machine learning processes based on the hidden Markov model. After review of all the data obtained from the previous B cell epitope prediction tools, the region from 280-344 amino acid sequences, 1-32 amino acid sequences, and 100-128 amino acid sequences have been proved to be the strongest region modulating B cell response.

**Target site analysis**

ProtParam generated an instability index (II) of 40.82 and an aliphatic index of 90.66 and negative GRAVY (grand average hydropathy) of -0.115 for the protein. Secondary structural traits of the protein were assessed by SOPMA and the protein was reported to be mostly occupied by alpha-helix (61.09%) entailing random coils and extended strands which constituted 18.29% and 13.04% of the protein, respectively. Finally, it showed beta-turns accounting for 7.59% of the whole protein. Three disease-causing regions were identified by GlobPlot 2.3 (Figure 8). The 3D model created with an aid of both Modeller 9.11 and SwissPDB was validated to prove the point of accuracy in comparison with other high-resolution models. PROCHECK generated a general overview of both the model and the Ramachandran plot (Figure 7). An environmental profile was chalked out and later, verified by QMEAN. The catalytically-active binding sites of the protein were discovered with the help of the CASTp server. The outcome delineated an array of amino acids to be conserved within the active site (Figure 6). They are ARG 237, ALA 234, ARG 236, ARG 262, GLN 226, ASP229, VAL 230, VAL 239, GLY 258, LEU259, ASN 225, and ILE 254. For smoothing out the drug discovery against the bacterium, the hydrophobicity of the protein was measured (Figure 4), which shows that sequence of our desired protein is strongly hydrophobic.

**PKPD modelling for validation**
A portrayal of a biological network has been used to clarify interaction and mechanisms between responding entities in a biochemical reaction. Bacterial chemotaxis pathway (Figure 9) based on pathway database (KEGG) [61] has been used to illustrate the pharmacokinetic mechanisms in this specific living system. To approach mathematical modeling, the biological process was retrieved from a pathway database and the building blocks were incorporated in the simulated model.

In a time-course simulation (Figure 10) with the chosen peptide (obtained from the screening), the initial concentration dose of the peptide was set at 0.2 μm (micro-meter) using MATLAB and Statistics Toolbox version 2012b [60]. The X-axis represented the transition time of entities and the Y-axis represented the concentration of the peptide.

**Molecular dynamics simulation**

To equilibrate interaction regions and obtain the precise structure of the complex close to realistic conditions, the docked peptide-protein complex was subjected to molecular dynamics simulation using GROMACS 2018.1 [55] for 1 nanosecond (Figure 11). We found the deviation levels approximately around 0.1 nm (1Å), which indicates that the structure is very stable.

**Discussion**

Our experiments were conducted to devise a way out that would trigger cell-mediated immunity against the Gram-negative intracellular opportunistic pathogen, *B. pseudomallei*. The task is seemingly easy, but the result is very cumbersome to obtain. Hence, to facilitate vaccine discovery against this opportunistic and occasionally fatal pathogen, we conducted this dry lab experiment which would accelerate further drug discovery. Many recent advances in the field of vaccination have been based on unique diagnostic approaches concerning the T cell epitope. Even though antigenic drift can supposedly circumvent antibody memory response continually over time, cell-mediated immunity recurrently evokes longer-lasting immune response [62]. Traditionally, pragmatic approaches towards vaccination have been extensively investigated. Nevertheless, a large body of evidence suggests that the limiting factors for the success of these approaches are big data on the pathogenesis of the disease, host-pathogen interaction, and mechanisms of virulence. On that account, the execution of successful vaccination is buttressed by decades of underpinning research. Many current vaccination programs lack this background data on the pathogens they are dealing with. Herein come immunoinformatics approaches in our rescue which are remarkably user-oriented, fully utilizing genome sequence data ready at hand. This scheme of vaccine designing based on of genome sequence has been termed reverse vaccinology, a proteomic approach to select the most immunogenic protein to be used as a vaccine candidate. This framework highlights the potential utility of sequence data in the detection of biochemical pathways, which upon inactivation, engender attenuated mutants functional in vaccine logistics. Those data might also be of use for rapid screening of open reading frames (ORF) throughout the whole genome sequence, capable of encoding potent subunits as vaccine candidates. [63, 64, 65].
A sequence should embody specific unique features to be a success as a peptide epitope. Initially, the epitope has to hail from the reasonably conserved regions among the proteome. Besides, it must incorporate qualities that promote immune reactions with HLA alleles. Moreover, the processed peptide has to demonstrate a high-affinity score with MHC alleles. Eventually, a decent population coverage score has to be ensured through the interaction between the MHC allele and the epitope. The T cell epitope we extrapolated meets all of the above requirements. The epitope was also identified as non-allergen. Therefore, an epitope vaccine employing selected peptides has nearly zero to very low probability to elicit allergic reactions. Hydrogen bond analysis along with the determination of the active sites of the protein was performed to facilitate docking. Furthermore, the selected epitope underwent \textit{in silico} docking simulation study. The findings suggested that our chosen epitope can bind as efficiently as the control engineered in this set-up as they have nearly the same binding energy. ProtParam and SOPMA tools helped us understanding the primary and secondary structural configurations of the protein using, which may offer an insight into the post-therapeutic strategy. To further designing drugs that are in the pipeline, ligand binding sites were unveiled and hydrophobicity analysis was conducted. Hydrophobicity study is of high impact in regards to drug design and delivery. It can be said with conviction that knowing and capitalizing the hydrophobicity study of drugs, such as docking and target entity prediction, would be of great value in the long run [66]. Furthermore, Molecular Dynamics simulations and PK/PD modeling approaches were also carried out to validate that our chosen peptide has experimental potential. Both studies corroborated the stability of the engineered vaccine with considerable efficiency. However, we need further wet-lab experiments to authenticate our hypothesis about the vaccine.

**Conclusions**

It is about time we took a colossal leap of faith for making headway in therapeutics for melioidosis and we have seen that significant changes have been introduced over the last few years. That being said, the path towards fruition is incomplete without the knowledge of system biology and the pathogenesis of the disease. Throughout this study, we focused on investigating a novel putative vaccine using computational biology. Also, we explored possible target sites for our proposed drug as well as conventional treatment schemes. A repertoire of B cell and T cell epitopes has been reported which might contribute to the innovation of novel therapeutics. As such, it is proposed that this vaccine should be further tested in vivo and in vitro to validate the findings of the study.

**Methods**

This study aims to start from the very beginning of proteome retrieval to using in silico approaches for drug discovery. The overall study design is outlined in a flow diagram (Figure 1).

**Epitope based vaccine design**

Screening the complete proteome of \textit{B. pseudomallei}, the reference strain we used was K96253, followed by the retrieval of the amino acid sequences in FASTA format using the UniProt Knowledgebase.
Filtering out the non-structural proteins (NSP) from this pool, we analyzed the sequences of the structural proteins to study several key properties of prospective vaccine candidates, which include antigenicity, solvent accessibility, surface accessibility, flexibility, and MHC class-I binding affinity [11,12].

To analyze the degree of digression of the protein sequences over time, phylogenetic trees were constructed by MEGA software with a bootstrapping value of about 1000 [13]. In order to predict effective antigens and subunit vaccines out of the protein sequences, VaxiJen v2.0 server [14] was used while maintaining its default parameters. Only one antigenic surface protein was screened out based on antigenicity score for further evaluation.

**Cellular localization of proteins**

PSORTb (http://www.psort.org/psortb/) [15] was used to predict the cellular location and function of proteins. Among these proteins, we selected one transmembrane protein, Aerotaxix Receptor, flagellar by nature as a potential vaccine target. The flagella and motility, along with the bacterial resistance to the bactericidal action of human immunity, is equally important to showcase significant roles in mobility, chemotaxis, and dissemination of *B. pseudomallei* from primary sites of infection, namely lungs or skin, to a distant part of the body through systemic circulation [16,17,18].

**T cell epitope prediction**

Systematic projections of CTL (cytotoxic T lymphocytes) epitopes are integral to the effective design of vaccines. NetCTL-1.2, a web-based device for screening human CTL epitopes in a target sequence, calculated the total score of CTL epitopes by adding up the values of TAP (transporter associated with antigen processing protein complex) transport efficiency, proteasomal cleavage, and MHC-I molecules binding affinity. We have set the parameter at 0.5 as per our study requirements with a sensitivity and specificity of 0.89 and 0.94, respectively. The epitope embodying the highest score was carefully chosen for further *in silico* investigation. To assess the binding affinity of peptide epitope to MHC-I molecules, IC\textsubscript{50} values that serve as a definitive measure of drug potency were evaluated [19, 20]. Prior to computation, peptide length was defaulted to 9 amino acids. To validate further, we carefully picked the alleles with a binding affinity (IC\textsubscript{50}) of less than 200 nm. Another IEDB (Immune Epitope Database) analytics software tool was used as well to measure processing score, TAP score, proteasomal cleavage score, and MHC-I binding affinity. Epitope conservancy, population coverage, and coverage area were also calculated [21]. The candidacy of the peptide epitopes were further scrutinized by IEDB MHC class I binding prediction tool [22] for all the available class I alleles [23].

**Allergenicity testing**

For assessing the allergenicity of the selected epitope, AllerHunter, [24] an online tool, was implemented. To verify the results of this tool, allergenicity was also evaluated by an additional online server called AllerTop 2.0 [25].
3D structure prediction

The highest conserve ETAAADALY epitope was submitted to PEP-FOLD [26] (peptide structure prediction server) to launch a molecular docking simulation study which gave us 5 rational 3D structures. The most reliable model for studying the interactions with HLAs was chosen. UCSF Chimera, a high throughput program for interactive visualization has also been used [27] for molecular modeling and visualization of the structure.

Docking simulation

AutoDockVina verified the interaction between HLA molecules and our target (peptide epitope) through a docking simulation study [28]. We retrieved 4uq2 (PDB id of a crystal structure of HLA-A*26:01) from the RCSB (Research Collaboratory for Structural Bioinformatics) database [29], followed by elimination of the Azo Benzene Containing Peptide by Discovery Studio, which forms a complex with the HLA-A*26:01 binding site [30]. The removed Azo Benzene Containing Peptide was also allowed to dock with the HLA molecule which is regarded as control docked complex in this construct. The binding energy of our predicted epitope (ETAAADALY) to the MHC-1 allele was compared to the one of the docked structure set as control.

B cell epitope prediction

Linear B cell epitopes were predicted based on the immunogenicity of the protein sequence via B cell epitope prediction tools of IEDB [31]. Emini surface accessibility prediction tool [32] (default threshold level 1.0) predicted the likelihoods of exposure of surface amino acids within the proteins constituting B cell epitope, while hydrophilicity was probed by Parker hydrophilicity prediction tool [33] (default set-up). The IEDB analysis resource also provided the Karplus and Schulz flexibility scale [34] for determining the flexibility of the potential B cell epitope [35]. Chou & Fasman Beta Turn Prediction Tool checked the frequency of beta-turns in the prospective antigenic region among the protein [36]. Beta turns are the most common protein motifs that help the polypeptide chain change its direction. Additionally, the Bepipred tool was used to predict linear B cell epitopes [37]. ABCpred server was employed for further corroboration [38]. To make a precise distinction between intracellular and surface proteins, transmembrane topology prediction was carried out using TMHMM (Transmembrane protein topology prediction method based on hidden markov model) v0.2 server [39]. Eventually, the epitopes that have been found to completely overlap between Bepipred and ABCpred prediction tools are selected as the optimal B cell epitope(s).

Target site analysis

To evaluate the physicochemical profile of the designated protein, ProtParam [40] and, the self-optimized prediction method with alignment (SOPMA) of the Expasy server [40, 41] were utilized. GlobPlot 2.3 predicted disease-causing regions among the protein sequence [42]. The three-dimensional (3D) structure of the selected protein was visualized via Modeller 9.14 through HHpred [43,44]. The protein sequence
underwent rigorous modeling followed by frequent energy minimization refinement program of Swiss-
PDB Viewer. Both Swiss-PDB Viewer [45] and PyMOL [46] were used as visualization tools. PROCHECK [47] checked the stereochemistry of the protein structure which is executed by Ramachandran Plot analysis [48] found in the ‘protein structure and model assessment tools’ section of SWISS-MODEL workspace [49]. Upon submission of the refined PDB format in the server, 2.5 Å resolution value was picked for a better protein structure. The protein structure and 3D profile were assessed using ERRAT [50] and model quality was estimated by QMEAN [51]. The catalytic binding sites of the protein were scanned as per the structural correlation between the template and the model construct with CASTp (computed atlas of surface topography of proteins) [52] server. Discovery Studio 4.0 client [53] has been employed in the measurement of the hydroobicity.

**Pharmacokinetic/Pharmacodynamic modelling for validation**

Pharmacokinetic/pharmacodynamic (PK/PD) modeling and simulation can be viewed as an apt, economical approach to work on issues like efficacy and safety of novel high throughput drugs. Ranging from the preclinical phase to phase III of drug development, all phases might benefit from this simulation tactic. The proposed peptide was confirmed by this simulation too for the testing of the dosage. The classic and most commonly used pharmacodynamic model is the nonlinear kinetics scheme that follows typically, Michaelis–Menten equation [54] as shown below:

\[
V = \frac{V_{\text{max}} \times s}{K_m + s}; \text{ it can be transformed into } C = \frac{C_{\text{max}} \times D}{K_m + D}, \text{ or, } V = \frac{dp}{dt} = \frac{A_{\text{max}} \times D}{K_m + D};
\]

Here C represents the steady-state concentration, \( C_{\text{max}} \) the theoretical maximum for C, A the amount absorbed, \( A_{\text{max}} \) the theoretical maximum for A, and D the dose.

**Molecular dynamics simulation**

The magnitude of affinity between the peptide and the HLA molecule was ascertained by Molecular Dynamics (MD) simulation. MD simulation was performed using Gromacs 2018.1 program [55]. OPLS-AA/L all-atom force field (2001 amino acid dihedrals) was used for preparing topology of the structure, and water molecules were added using the SPC/E water model. The structure was placed at the center of a cubic box, where the distance between the structure and the cubic box was one nanometre. To make the structure well solvated with an explicit solvent, the SPC216 model was used, an all-inclusive equilibrated three-point solvent model. Using that model, 15033 water molecules were added inside the box. The long-range electrostatic interactions were estimated using Particle-Mesh Ewald (PME) method with an interpolation order of four and maximum Fast Fourier Transforms (FFT) grid spacing of 0.16 nanometre. The coulomb radius was set to one nanometre and the short-range interactions also known as van der waals was set to a threshold value of one nanometre [56]. In this construct, periodic boundary conditions were applied, and the bonds were constrained using the LINCS algorithm to their equilibrium position. Three Sodium (Na\(^{+}\)) ions were added into the box to neutralize the net charge of the structure.
Subsequently, the structure was minimized for 50000 steps using the steepest descent minimization algorithm and subjected to 100 picosecond equilibrations at constant temperature and pressure of 300K and 1 bar, respectively. After the structure was well-equilibrated at the desired temperature and pressure, the production MD was performed in isothermal-isobaric conditions for one nanosecond, and the coordinates were saved every ten picoseconds. The following equation was solved to calculate the stability of the complexes after one nanosecond.

\[ \text{RMSD} = \sqrt{\frac{\sum_{i=0}^{N} (M_i \times (X_i \times Y_i)^2)}{M}} \]

Where \( N \) is the number of atoms, \( M_i \) is the mass of atom \( i \), \( X_i \) is the coordinate vector for target atom \( i \), \( Y_i \) is the coordinate vector for reference atom \( i \), and \( M \) is the total mass. If the RMSD is not mass-weighted, all \( M_i = 1 \) and \( M = N \).

**Abbreviations**

3D: 3 Dimensional

CASTp: Computed Atlas of Surface Topography of proteins

CTL: Cytotoxic T-Lymphocyte

FFT: Fast Fourier Transforms

GRAVY: Grand Average of Hydropathy

GROMACS: GROningen MAchine for Chemical Simulations

HLA: Human Leukocyte Antigen

IC\(_{50}\): Half maximal Inhibitory Concentration

IEDB: Immune Epitope Database

KEGG: Kyoto Encyclopedia of Genes and Genomes

MEGA: Molecular Evolutionary Genetics Analysis

MD: Molecular Dynamics

MHC: Major Histocompatibility Complex

OPLS-AA/L: Optimized potentials for liquid simulations-all atoms/ Local second order Moller Plesset Perturbation Theory
PAMPS: Pathogen-associated Molecular Patterns

PDB: The Protein Data Bank

PK/PD: Pharmacokinetic/pharmacodynamic modeling

QMEAN: Qualitative Model Energy ANalysis

RCSB: The Research Collaboratory for Structural Bioinformatics

SOPMA: Self-optimized Prediction Method with Alignment

SPC216: Simple Point Charge water model of a small periodic box containing 216 water molecules

SPC/E: Extended Simple Point Charge model

TAP: Transporter Associated with Antigen Processing Protein Complex

TLR: Toll-Like Receptors

UniProtKB: UniProt Knowledgebase

**Declarations**

**Ethics approval and consent to participate:** Not applicable

**Consent for publication:** Not applicable

**Availability of data and materials:** All data generated or analyzed during this study are included in this published article (and its supplementary information files).

**Competing interests:** The authors declare that they have no competing interests.

**Funding:** Not applicable

**Authors' contributions:** AT and MAK were involved in study design and data analysis. ASS contributed to data analysis. AT, MAK and CRA contributed to discussions and preparation of the manuscript. The work was done under the supervision of CRA and MY. All authors read and approved the final manuscript.

**Acknowledgment:** We express our gratitude to Bio Bio-1 for facilitating this scientific study and Tabassum Tasnim Auroni and Sadia Ahmed for their precious judgment.

**Authors' information:**

**Affiliations:**
References

1. Patel N, Conejero L, Reynal MD, Easton A, Bancroft GJ, Titball RW. Development of vaccines against Burkholderia pseudomallei. Front Microbiol. 2011;2.

2. Currie BJ. Advances and remaining uncertainties in the epidemiology of Burkholderia pseudomallei and melioidosis. Trans R Soc Trop Med Hyg. 2008;102(3):225-227.

3. Wiersinga WJ, Poll T, White NJ, Day NP, Peacock SJ. Melioidosis: insights into the pathogenicity of Burkholderia pseudomallei. Nat Rev Microbiol. 2006;4(4):272.

4. Currie BJ, Fisher DA, Anstey NM, Jacups SP. Melioidosis: acute and chronic disease, relapse and reactivation. Trans R Soc Trop Med Hyg. 2000;94(3):301-304.

5. Moore, T., Ekworomadu CO, Eko FO, MacMillan L, Ramey K, Ananaba GA, et al. Fc receptor–mediated antibody regulation of T cell immunity against intracellular pathogens. J Infect Dis. 2003;188(4):617-624.

6. Casadevall A, Pirofski LA. A reappraisal of humoral immunity based on mechanisms of antibody-mediated protection against intracellular pathogens. Adv Immunol. 2006;91:1-44.

7. Casadevall A. Antibody-mediated protection against intracellular pathogens. Trends Microbiol. 1998;6(3):102-107.

8. Igietseme JU, Eko FO, He Q, Black CM. Antibody regulation of T-cell immunity: implications for vaccine strategies against intracellular pathogens. Expert Rev Vaccines. 2004;3(1):23-34.

9. Wiersinga WJ, Van der Poll T. Immunity to Burkholderia pseudomallei. Curr Opin Infect Dis. 2009;22(2):102-108.

10. Su Y.-C, Wan K-L, Mohamed R, Nathan S. A genome level survey of Burkholderia pseudomallei immunome expressed during human infection. Microbes Infect. 2008;10(12):1335-1345.

11. Consortium U. Activities at the universal protein resource (UniProt). Nucleic Acids Res. 2014;42(D1):D191-D198.

12. Magrane M. UniProt Knowledgebase: a hub of integrated protein data. Database. 2011.

13. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870-4.

14. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007;8(1):4.
15. Brinkman FSL, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010;26(13):1608-1615.

16. DeShazer D, Brett PJ, Carlyon R, Woods DE. Mutagenesis of Burkholderia pseudomallei with Tn5-OT182: isolation of motility mutants and molecular characterization of the flagellin structural gene. J Bacteriol. 1997;179(7):2116-2125.

17. Inglis TJ, Robertson T, Woods DE, Dutton N, Chang BJ. Flagellum-mediated adhesion by Burkholderia pseudomallei precedes invasion of Acanthamoeba astronyxis. Infect Immun. 2003;71(4):2280-2282.

18. Chua K, Chan Y, Gan Y. Flagella are virulence determinants of Burkholderia pseudomallei. Infect Immun. 2003;71(4):1622-1629.

19. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. 2003;12(5):1007-17.

20. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics. 2007;8:424.

21. Bui H-H, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics. 2007;8(1):361.

22. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics. 2005;6:132.

23. Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, et al., Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. Cell Mol Life Sci. 2005;62(9):1025-37.

24. Cheng Muh H, Tong J, Tammi MT. AllerHunter: A SVM-Pairwise System for Assessment of Allergenicity and Allergic Cross-Reactivity in Proteins. PLoS One. 2009;4(6):e5861.

25. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v.2–a Server for in Silico Prediction of Allergens.J Mol Model. 2014;20(6):2278.

26. Thevenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tufféry P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. Nucleic Acids Res. 2012;40(Web Server issue):W288-93.

27. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605-12.

28. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455-461.

29. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, H Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-242.

30. Gras S, Kedzierski L, Valkenburg SA, Laurie K, Liu YC, Denholm JT, et al. Cross-reactive CD8+ T-cell immunity between the pandemic H1N1-2009 and H1N1-1918 influenza A viruses. Proc Natl Acad Sci
31. Ponomarenko J, Bui H-H, 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):514.

32. Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol. 1985;55(3):836-9.

33. Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry. 1986; 25(19):5425-32.

34. Karplus PA, Schulz GE. 1985. Prediction of chain flexibility in proteins. Naturwissenschaften72:212-213.

35. Hasan MA, Khan MA, Datta A, Mazumder MHH, Hossain MU. A comprehensive immunoinformatics and target site study revealed the corner-stone toward Chikungunya virus treatment. Mol Immunol. 2015;65(1):189-204.

36. Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol. 1978;47:45-148.

37. Larsen JEP, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res. 2006;2:2-2.

38. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins. 2006;65(1):40-8.

39. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567-80.

40. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31(13):3784-8.

41. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci. 1995;11(6):681-4.

42. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res. 2003;31(13):3701-8.

43. Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. Evaluation of comparative protein modeling by MODELLER. Proteins. 1995;23(3):318-26.

44. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005;33:W244-8.

45. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. 1997;18(15):2714-23.

46. The PyMOL Molecular Graphics System; DeLano Scientific: San Carlos, CA, 2002.

47. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr. 1993;26(2):283-291.
48. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J Mol Biol. 1963;7:95-9.

49. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al., SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-W303.

50. Colovos C, and Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993;2(9):1511-9.

51. Benkert P, Tosatto SC, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. Proteins. 2008;71(1):261-77.

52. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. 2006;34:W116-8.

53. Dassault Systèmes BIOVIA, Discovery StudioModeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016.

54. Wagner JG, Szpunar GJ, Ferry JJ. A nonlinear physiologic pharmacokinetic model: I. Steady-state. J Pharmacokinet Biopharm. 19985; 13(1):73-92.

55. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1:19-25.

56. Tao Y, Rao ZH, Liu SQ. Insight derived from molecular dynamics simulation into substrate-induced changes in protein motions of proteinase K. J Biomol Struct Dyn. 2010;28(2):143-157.

57. Holden MTG, Titball RW, Peacock SJ, Cerdeño-Tárraga AM, Atkins T, Crossman LC, et al. Genomic plasticity of the causative agent of melioidosis, Burkholderia pseudomallei. Proc Natl Acad Sci U S A. 2004;101(39):14240-14245.

58. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2007; 35: D61–D65

59. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019; 47: D506-515

60. Thompson CM, Shure L. Image Processing Toolbox: For Use with MATLAB. MathWorks. 1995.

61. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids Res. 2000;28(1):27-30.

62. Thompson DB, Crandall K, Harding SV, Smither SJ, Kitto GB, Titball RW, et al., In silico analysis of potential diagnostic targets from Burkholderia pseudomallei. Trans R Soc Trop Med Hyg. 2008;102:S61-S65.

63. Rappuoli R. Reverse vaccinology. Curr Opin Microbiol. 2000;3(5):445-450.

64. Masignani V, Rappuoli R, Pizza M. Reverse vaccinology: a genome-based approach for vaccine development. Expert Opin Biol Ther. 2002;2(8):895-905.
65. Adu-Bobie J, Capecchi B, Serruto D, Rappuoli R, Pizza M. Two years into reverse vaccinology. Vaccine. 2003;21(7):605-610.
66. Sarkar A, Kellogg GE. Hydrophobicity-shake flasks, protein folding and drug discovery. Curr Top Med Chem. 2010;10(1):67-83.

Tables

Table 1
Most potential T cell epitope with interacting MHC-I alleles

| Peptide       | Allele      | IC50  | Conservancy |
|---------------|-------------|-------|-------------|
|               | HLA-A*26:01 | 12.81 | 11.11%      |
|               | HLA-C*03:03 | 17.89 |             |
|               | HLA-A*68:01 | 41.83 |             |
|               | HLA-C*12:03 | 49.91 |             |
| ETAAADALY     | HLA-A*25:01 | 54.46 |             |
|               | HLA-A*29:02 | 74.26 |             |
|               | HLA-B*15:02 | 96.21 |             |
|               | HLA-B*35:01 | 127.23|             |
|               | HLA-A*68:02 | 150.87|             |

The most immunogenic epitope (ETAAADALY), screened by NetCTL-1.2, is shown to be interacting with MHC-I alleles with a definite IC50 score. IC50 measure indicates how much of a particular drug or other substance (inhibitor) is needed to inhibit a given biological process (or component of a process, i.e. an enzyme, cell, cell receptor or microorganism) by half. According to the FDA, IC50 represents the concentration of a drug that is required for 50% inhibition in vitro. Here, we see, the most potent combination of peptide and allele (with the least IC50 measure) contains HLA-A*26:01. In an epitope-based vaccine setting, the use of conserved epitopes would be expected to provide broader protection across multiple strains, or even species, than epitopes derived from highly variable genome regions. Hence, epitope conservancy analysis is done.
Table 2
Population coverage calculated by epitopes

| Population/Area | Class 1 |       |       |
|-----------------|---------|-------|-------|
|                 | Coverage (%)<sup>a</sup> | Average hit<sup>b</sup> | PC90<sup>c</sup> |
| Australia       | 1.74%   | 0.02  | 0.1   |
| Central Africa  | 2.20%   | 0.02  | 0.1   |
| China           | 4.27%   | 0.04  | 0.1   |
| East Africa     | 1.91%   | 0.02  | 0.1   |
| East Asia       | 12.73%  | 0.13  | 0.11  |
| Europe          | 7.73%   | 0.08  | 0.11  |
| North America   | 5.35%   | 0.05  | 0.11  |
| Russia          | 9.74%   | 0.1   | 0.11  |
| South America   | 1.95%   | 0.02  | 0.1   |
| South Asia      | 8.55%   | 0.09  | 0.11  |
| Southeast Asia  | 5.68%   | 0.06  | 0.11  |
| Southwest Asia  | 8.61%   | 0.09  | 0.11  |
| United States   | 5.36%   | 0.05  | 0.11  |
| World           | 5.82%   | 0.06  | 0.11  |
| Average         | 5.23%   | 0.05  | 0.108 |
| Standard Deviation | 3.22% | 0.03  | 13.29 |

<sup>a</sup>Projected population coverage.

<sup>b</sup>Average number of epitope hits/HLA combinations recognized by the population.

<sup>c</sup>Minimum number of epitope hits/HLA combinations recognized by 90% of the population.

Worldwide population coverage and prevalence area of the epitope-HLA combination was determined by IEDB analysis resource tool. For every epitope, the MHC-I binders that were recognized as optimum are selected for the population coverage analysis. Identified optimum MHC-I binders for each epitope were considered for the population coverage analysis of the epitopes.
Table 3  
B cell epitopes predicted by IEDB

| Start | End  | Peptide                                      | Length |
|-------|------|----------------------------------------------|--------|
| 1     | 32   | MRNNQPVTQHEFELPDDATLMSTTDPHGRITY            | 32     |
| 42    | 54   | GFSSDEIVGAPHN                                 | 13     |
| 73    | 82   | LKRGEPWTL                                    | 10     |
| 84    | 95   | KNRRKNGDHYWV                                 | 12     |
| 100   | 128  | VPVIRGGQTQGYMSVRTKPARAETAAADA                | 29     |
| 133   | 142  | FREGRAGSRR                                   | 10     |
| 184   | 197  | AWAAGVAGAPLAQL                               | 14     |
| 228   | 242  | LDVATGASRRGVMN                               | 15     |
| 280   | 344  | NEIAQGNHDSARTEQAATSVQQTAASMAQMTATVSSNAQT    | 65     |
|       |      | ATQANRLSESASHAAERGGQAVRE                    |        |
| 349   | 358  | MGEITESSRR                                   | 10     |
| 382   | 394  | VEAARAGEQGRGF                                | 13     |
| 403   | 413  | ALAQRSANAAK                                  | 11     |
| 423   | 442  | VERVESGAQTVDYAGRGMGE                        | 20     |
| 458   | 473  | SASTSEQRAQTQVDD                             | 16     |
| 486   | 500  | AALVEQSAASESLR                               | 15     |

B cell epitope prediction tools of IEDB, Bepipred tool and ABCpred server were used to predict linear B cell epitopes based on the highest immunogenic protein sequence. After cross-checking all the data obtained from the previous B cell epitope prediction tools, the region from 280–344 amino acid sequences, 1–32 amino acid sequences and 100–128 amino acid sequences were found to be the best capable region for inducing B cell response.
Figure 1

Graphical representation of peptide vaccine design and target site analysis against B. pseudomallei
Figure 2

Molecular docking The 3D structures of HLA-A*26:01 and Epitopes are visualized and captured with PyMOL molecular graphics system. AutoDockVina predicted the docked structures with respective docking scores. The aim is to compare the spontaneous binding energy between the HLA-native (control peptide which contains Azobenzene) peptide complex (2.A) and the HLA-target peptide (desired epitope) (ETAAADALY) complex (2.B).

Figure 3
Molecular docking interactions Here, we analyzed the hydrogen bonds between the MHC molecule and the chosen epitope. The amino acid residues ASN 62, ASN 65, ASN 76, ARG 96, ARG 162, LYS 145, THR 142, TRP 146, TYR 6 are demonstrated to be involved in hydrogen bonding via Discovery Studio.

**Figure 4**

Hydrophobicity of the protein for B cell epitope analyses Parker hydrophilicity prediction tools depict the hydrophilic region of the B cell epitopes contained in the transmembrane protein. Blue regions are the most hydrophilic, white regions signify neutrality whereas brown ones are the most hydrophobic amongst the entire flagellar protein.

**Figure 5**

Solvent accessibility surface of the proteins for B cell epitope analyses Emini surface accessibility prediction tool was applied to predict the surface exposure probabilities of amino acids within the proteins selected for designing B cell epitopes in the transmembrane protein. Blue region denotes
exposed amino acid residues having greater than 25% of maximum surface accessibility score while green one denotes hidden residues having less than 10% of maximum surface accessibility score.

Figure 6

Active site residues in the 3D structure The active sites of the protein are identified from CASTp server. The measured outcome indicated that amino acids ARG 237, ALA 234, ARG 236, ARG 262, GLN 226, ASP229, VAL 230, VAL 239, GLY 258, LEU259, ASN 225, ILE 254 are predicted to be conserved with the active site. The server predicted the best active site with these amino acid residues.
Figure 7

Ramachandran Plot showing stability of amino acid residues in the protein model For the assessment of accuracy and stereochemical properties of the predicted protein model, Ramachandran Plot analysis was done. Ramachandran plot is used to assess the chemical correctness of backbone ($\phi$) and $\psi$ angles where the percentage of at least 90% in the most favored regions are preferred. Here we can see that our target protein has most of its region favored for proline residues colored in blue.
Figure 8

GlobPlot disorder prediction The result from GlobPlot 2.3 was used to identify three disease-causing regions in entire transmembrane protein. The regions are shown by green, yellow and brown bars respectively.

Figure 9
Role of Aerotaxis Receptor protein in biochemical pathway derived from KEGG A representation of biological network based on pathway database (KEGG) (https://www.genome.jp/dbget-bin/www_bget?pathway+ko02030) has been used to elucidate mechanisms of interaction among responding entities. This signaling pathway depicts the role of Aerotaxis Receptor protein in bacterial chemotaxis. Appropriate copyright for the use and modification of this pathway has been obtained.

**Figure 10**

Time-Course Simulation against the dosage of the peptide In a time-course simulation with the chosen peptide (obtained from the screening), the initial concentration dose of peptide was set at 0.2 μm (MATLAB and Statistics Toolbox Release 2012b). The X axis represents the transition time of entities and Y-axis represent the concentration of the peptide.
Figure 11

RMSD of crystal structure (red) and energy minimized structure (black) of the peptide vaccine. OPLS-AA/L all-atom force field was used in order to estimate the topology of the structure, and water molecules were added using SPC/E water model. Both the crystal structure and the energy minimized structure were simulated for 100 picoseconds.