An In-depth Genomic Investigation to Design a Multi-Epitope Based Vaccine against Brucellosis

Alaa R. Hameed  
University of Baghdad

Bassam Qasim Mohammed  
University of Baghdad

Tabarak Sabah Jassim  
Al-Nahrain University

Sajjad Ahmad (✉ sajjad.ahmad@abasyn.edu.pk)  
Abasyn University  https://orcid.org/0000-0003-0754-1742

Research Article

Keywords: Brucella melitensis, Brucellosis, Vaccine, Pan-omics, Molecular dynamics simulation

Posted Date: March 27th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2727459/v1

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

Version of Record: A version of this preprint was published at Informatics in Medicine Unlocked on May 19th, 2023. See the published version at https://doi.org/10.1016/j.imu.2023.101276.
Abstract

*Brucella melitensis* is a gram-negative coccobacillus that causes brucellosis in humans when they come into contact with infected animal meat or consumed raw milk. The lack of effective treatment and increasing antibiotic resistant patterns shown by *B. melitensis* warrant the search for novel therapeutic targets. In this study, comprehensive bioinformatics, reverse vaccinology, and biophysics techniques were employed to design a novel multi-epitopes-based vaccine (MEBV) against *B. melitensis*. Pan-genomics, subtractive proteomics and immunoinformatic studies revealed three core proteins: Flagellar hook protein (FlgE), TonB-dependent receptor, and Porin family protein as promising vaccine targets. The proteins have exposed topology, are antigenic, and are adhesin. Furthermore, B and T cell epitopes were predicted from these target proteins. Highly antigenic, immunogenic, and non-allergic epitopes were shortlisted and used in the MEBV vaccine design. The designed MEBV also showed stable docked conformation with different immune receptors such as MHC-I, MHC-II, and TLR-4. It was found that all three systems showed robust binding energies with net binding energy < -300 kcal/mol. The van der Waals and electrostatic energies were the dominating energies and were found to be the stabilizing factors of complexes. The designed vaccine contains antigenic epitopes that were filtered using stringent criteria. The vaccine was also predicted to generate promising immunological responses and thus could be an attractive candidate for evaluation in experimental studies.

1. Introduction

Brucellosis is a zoonotic infection caused by a gram-negative coccobacillus called *Brucella melitensis* [1]. *Brucella* comprises 12 recognized species, five (*B. melitensis, B. suis, B. abortus, and B. canis*) of which are capable of infecting humans [2]. It spreads to humans by consumption of contaminated dairy products [3]. Butchers, shepherds, milkmen, and veterinary persons are at a higher risk of contracting Brucellosis than the general public [4, 5]. *B. melitensis* is the most pathogenic among *Brucella* species [6]. *B. melitensis* is highly infectious, virulent, and responsible for persistent human infections that result in major health complications [7]. It is an etiological agent of endemic Brucellosis in the Middle East [8]. *Brucella* strains are often considered common laboratory pathogens as they are transmitted via air [9]. It’s capacity to serve as a bio-weapon against humans has earned class B pathogen by the Centers for Disease Control and Prevention [10].

*Brucella* species can successfully infect host mucosal surfaces and remain latent for a longer time [3]. The smooth lipopolysaccharides of the cell wall shell the *B. melitensis* cells during infection, which facilitates their intracellular survival and protection from the immune system [11]. Besides, *Brucella* species can survive for a longer time inside phagocytes, avoid access of the immune system, and cause long-term infection [12]. The persistent infections of *Brucella* are characterized by mononuclear apoptotic cells suppression, inhibition of dendritic cell maturation, reduction of antigens presentation, and inactivation of T-cells [13]. The delays in *Brucella* identification by the host immune system *Brucella* is due to modification or completely shut down pathogen associated molecular patterns [5]. Several studies have been conducted to understand *B. melitensis* resistance to drugs such as rifampicin and trimethoprim/sulfamethoxazole [14, 15]. Therefore, unveiling novel targets against *B. melitensis* still is of high significance [16, 17].

Traditional steps of vaccine development are time-consuming, complex, and costly, and the success rate is low [18, 19]. Development of a subunit failed to prevent brucellosis. Epitope-based vaccines are safe to produce and easy to develop [20]. The peptide vaccines also lack allergenic sequences and can provoke specific immune responses [21]. Additionally, antigenic peptides can be fused to stimulate strong, safe, and specific immune responses [22]. The availability of genetic sequence data has recently drawn the attention of the scientific community to surface previously uncharacterized vaccine targets. Presently, two popular vaccines against brucellosis are used in animals. The live attenuated *B. abortus* strain 19 (S19 vaccine) used for preventing brucellosis in cattle. The Rev1 vaccine is effective against caprine and ovine brucellosis. Both vaccines produced good immunological responses, however, lack of a persistent response is a major limitation of the vaccines [23, 24]. In the current study, comparative proteomic studies of *B. melitensis* strains were done, followed by reverse vaccinology approach to identify potential vaccine candidates. The screened proteins were subjected to T and B cell epitope mapping, and best scored epitopes were used to design a multi-epitopes-based vaccine (MEBV). The designed vaccine was then evaluated for structural and immunological properties to evaluate its biological potency against *B. melitensis*.

2. Materials And Methods

Details of the study methodology are given in Fig. 1. The proteomic data of 57 sequenced strains of *B. melitensis* available at time of this research work was obtained from the genome database of National Center for Biotechnology Information (NCBI) [25]. The proteomic sequence data was then investigated in pan-genome analysis to identify core, unique, and accessory genes [26, 27]. For this purpose, the bacterial pan-genomic analysis (BPGA) program was used [27].

2.1. Initial Screening - Subtractive Proteomics

Subtractive proteomics on the core proteome was performed to identify potential vaccine candidates. The initial step of subtractive proteomics was to remove paralogue sequences [28]. CD-HIT (Cluster Database at High Identity with Tolerance) is a well-known and effective tool for comparing and grouping sequences to reduce sequence redundancy [29]. In this case, the core proteome was filtered at a 90% threshold. Subcellular localization prediction of non-redundant proteins was performed with PsortB 3.0 [30]. The subcellular localization results were revalidated by CELLO, which is an SVM-based software [31]. Proteins with membrane localization predicted in the periplasmic space, and extracellular matrix were chosen for further examination. Proteins having multiple and uncertain localization were eliminated.

2.2. Prioritization of Potential Candidates

Virulence factors are crucial as they play a vital role in immunological responses against antigens [32, 33]. VFDB (Virulence Factor Database) was used to identify virulent factors from surface localized proteins [34]. Virulence factors were identified by defining bit score > 100 and sequence identity score > 30% in Blastp search in VFDB. The TMHMM (Transmembrane Helices; Hidden Markov Model) server was utilized to predict transmembrane helices [35]. The Proteins
having multiple transmembrane helices were removed from the study because they are difficult to express, purify, and clone, making them unsuitable for vaccine development [36, 37]. The pathogenic proteins having 0 or 1 transmembrane helices were then analyzed for physicochemical properties by ProtParam tool to select proteins that are very likely to be used in experiments [38]. The instability index was the most important characteristic that was accessed. The instability index calculator predicts the presence of specific dipeptides that are absent in in vivo unstable proteins. The protein instability index had a cut-off of 40, and those with a predicted value of more than 40 were classified as unstable. Molecular weight analysis is considered crucial in terms of proteins purification [26]. Ideally, proteins with a molecular weight of less than 110 kDa are regarded as effective vaccine targets [34, 39, 40]. VaxiJen 2.0 was used to determine antigenicity of proteins, with organism set to bacteria and a threshold of > 0.6 [41]. Antigenicity refers to the ability for provoking adaptive immunity products such as antibodies and T-cell receptors [42]. Adhesins are promising vaccine targets as they facilitate bacterial attachment and adherence to host tissues, which is critical for microbial pathogenicity [43]. Afterward, using AllerTOP 2.0, an in silico allergen prediction approach, was employed to predict allergenic sequences [44]. Non-allergic epitopes were screened using NCBI’s online BLASTp against a reference human proteome (taxonomic id: 9606) to eliminate proteins with homology greater than the input parameters. Homology search parameters used were: E-value 1.0 E^{-5}, bit score > 100, and sequence identity greater than 30%. Non-homologous proteins were then aligned with the proteomes of probiotic bacteria to discard identical proteins and as such to avoid beneficial microorganisms’ accidental inhibition. The input criteria used were the same as discussed above.

2.3. B cell-derived T cell Epitopes Prediction

The screen candidates were then considered in the epitope prediction step, which predicted B-cell and T-cell epitopes. The Bepipred Linear Epitope Prediction 2.0 of Immune Epitope Database (IEDB) was used to predict linear B-cell epitopes with a threshold of 0.5 [45, 46]. The B-cell epitopes were then utilized in T-cell epitopes prediction in IEDB T-cell epitope sever to identify subsequences that bind to MHC class I and II alleles. The IEDB-recommended 2.22 technique was used for T-cell epitope prediction, and the peptides were sorted by percentile score. High-affinity binders were defined as those with a low percentile score. The binding affinity of screened T-cell epitopes for DRB*0101 was predicted using MHCPred 2.0 [47]. Antigenicity, allergenicity, and solubility of predicted B and T cell epitopes were checked through VaxiJen [41], AllerTOP 2.0 [44], and ProSol, respectively. The toxicity of antigenic, non-allergic and soluble filtered epitopes was checked through Toxin-pred [48] and the resultant non-toxic epitopes were selected.

2.4. Population Coverage Analysis

The expression of HLA alleles depends on ethnic populations, which helps to improve the design of an epitope-based vaccine. The role of HLA allele distribution in the global population is critical for MEBV development. The population coverage of the identified epitopes and their unique HLA binding alleles were investigated using IEDB-AR v2.20 population Coverage Tool [49].

2.5. Designing and Validation of MEBV

MEBV with adequate adjuvants and immunodominant epitopes may overcome weak immunogenicity [21]. GPGPG linkers were utilized to join the shortlisted epitopes with each other [35]. The last peptide was then added to the beta-defensins, which was used an adjuvant molecule [50]. The different physicochemical properties of the MEBV were predicted using ProtParam (II) [38]. The VaxiJen v2.0 server was also used to test the antigenicity and immunogenicity of the MEBV construct [41]. AllerTop was used to assess the vaccine’s allergenicity [44]. The MEBV secondary structure was evaluated using the SOPMA [51]. In addition, the SolPro server was used to test the solubility of the MEBV construct [52].

2.6. Tertiary Structure Prediction and Validation

SCRATCH protein predictor’s 3Dpro was used to predict the construct’s tertiary structure [53]. GalaxyLoop was used to model MEBV loops, and GalaxyRefine was used to refine it [54]. The quality of the vaccine 3D model was assessed through the use of online tools such as ERRAT [55] and ProSA-web [56].

2.7. Blind Docking Protocol

The vaccine construct must effectively interact with host immune cells and receptors in order to elicit sufficient and accurate immunological responses. The binding affinity of MEBV with human immunological receptors was investigated using molecular docking approach. The PatchDock server [57] was used to execute blind docking of MEBV with immune receptors (MHC-I, MHC-II, and TLR-4). The receptors 3D structure was obtained from protein data bank (PDB). The PatchDock docking was done based on RMSD value of 4 angstroms. The resulting docked solutions were refined using FireDock [58, 59]. Then using UCSF Chimera 1.16 [60], all complexes with lowest global energy were shortlisted intermolecular conformation and interactions visualization.

2.8. Binding Confirmation validation of MEBV Using MD Simulations

Molecular dynamics (MD) simulation assay was performed in three steps: 1) system preparation, 2) pre-processing, and 3) production phase [61]. AMBER20 was utilized to execute a 100-ns simulation production. The first phase was completed using an antechamber module of AMBER20 [62, 63]. The complexes of MHC-I MHC-II, TLR-4 with MEBV were solvated into TIP3P water box and Na+ ions were used to neutralize them. During the MD simulation assay, force field of ff14SB was used to investigate intermolecular interactions [64]. The steepest descent approach was utilized for 6000 cycles while conjugate gradient algorithm was used for 3000 cycles to accomplish energy minimization process. Equilibration was carried out at 1 atm and temperature was maintained at 310 K. After that, the MDS was performed for each complex on a time scale of 200 ns using PME (Particle mesh Ewald) [65]. The SHAKE algorithm was employed to constrain hydrogen bonds [66]. The CPTRAJ module was used for trajectory analysis [67].

2.9. MMGBSA Binding Energy Analysis

The binding energies of MEBV-MHC-I, MEBV-MHC-II, and MEBV-TLR-4 were predicted using MM-GBSA and MM-PBSA techniques [68]. Different studies have found that the MM-PB/GBSA is the best method for estimating biological complexes, which include protein-protein/DNA/RNA systems. Using the script MMPBSA.py, the total free energy was estimated [69]. For energy estimation, 1000 frames were selected from simulation trajectories.

2.10. Immune Simulation
C-ImmSim version 10.1 is an immune simulation tool that can be used to assess the MEBV's immunological response [70]. It simulates three primary organs: bone marrow, thymus, and lymph node. HLA (B0702, A0101, B0702, DRB1 0101, A0101, and DRB1 0101), number of steps (100), number of injections set to 1, volume (10), and random seed were among the parameters used in immunological simulation.

2.11. MEBV's In Silico Cloning and optimization

Codon optimization is required for good expression of a foreign gene in a host organism. The host system was set to E. coli K12. For codon adaptation, the MEBV sequence was uploaded to JCat [71]. With the help of SnapGene v4.3 server, in silico cloning was performed in pET28a (+) vector.

3. Results And Discussion

3.1. Core Proteomics Analysis

A total of 57 complete sequenced strains of B. melitensis proteome were collected from the NCBI. The proteome of the strain can be found in its entirety at the following link: https://www.ncbi.nlm.nih.gov/genome/943. The list of strains is also given in S-Table 1. The core proteome of B. melitensis strains contains 153966 proteins. A subtractive proteomics approach was used to identify surface proteins from the core proteome. The full core proteome of the aforementioned pathogen was compared to one another to screen the sequences that shared sequence identity. This was achieved by CD-Hit analysis to accelerate clustering and comparison of input protein sequences. To eliminate redundant proteins, several runs of CD-HIT analysis were performed. Protein sequences with sequence identities of more than 90% were excluded. This redundancy filter revealed 2336 non-redundant proteins. Out of these 2336 non-redundant proteins, 10 proteins were predicted as extracellular, 69 as periplasmic, 23 as outer membrane, 399 as inner membrane, 1050 as cytoplasmic, and 781 have unknown localization. The surface proteins have antigenic epitopes that are easily recognized by host immune system [34]. Further, only 21 proteins were found as virulence factors. The virulence factors initiate infectious pathways and are regarded as good vaccine targets [72–76].

3.2. Physicochemical Properties Investigation

It was necessary to evaluate physicochemical properties, antigenicity, and number of transmembrane helices of the shortlisted virulence factors to guide the selection of potent vaccine targets for experimental assessment. Out of 21 virulence factors, 20 were less than 110 kDa. Out of those 20, 19 proteins were stable and determined to contain one helix and were considered for further investigation. The low molecular weight proteins allow easy purification in experimental studies [77]. Furthermore, a smaller number of transmembrane helices proteins can be easily cloned and expressed [78]. The VaxiJen server was then used to evaluate the selected 19 proteins, and eight of them were found to be antigenic. Allergic sequences were removed further to discard vaccine-related allergies [79]. There were seven proteins identified as non-allergic, suggesting that they could be vaccine targets. In order to prevent bacterial colonization, subsequent pathogenesis and infection cycle, it is necessary to prevent the bacteria from attaching to the host cells. Adhesion is mediated by contacts between adhesions on bacteria surface and the receptor on host cells [80]. Adhesions are virulence factors found on the surface of bacteria's cells, and they could be targeted for the treatment of bacterial infections. Adhesin-based vaccines can be helpful in preventing brucellosis as explained by intranasal immunization with BtaF in mice [81]. Preclinical investigations with E. coli/adenosin FimH demonstrated the formation of anti-FimH antibodies [82]. Identification and characterization of adhesions that could be used in prophylactic/therapeutic vaccination is a high priority in current research. Only three proteins were identified as adhesin having scores greater than 0.7. The remaining four proteins were discovered non-adhesin and were discarded. Further, a comparative homology check was performed on the filtered proteins against Human proteome as well as against probiotic bacteria. Probiotic bacteria are beneficial in maintaining the natural balance of the gut microbiota, which is sometimes altered during illness and the treatment process [83]. All three proteins were discovered to have no hits against human and probiotic proteomes. Over the past century, the mouse has emerged as the main mammalian model for experimental studies [70]. The B and T cells are major lymphocytes of the adaptive immune system that play a crucial role in producing humoral and cell-mediated immunity, respectively against invading pathogens [84]. The B and T cells are major lymphocytes of the adaptive immune system that play a crucial role in producing humoral and cell-mediated immunity, respectively against invading pathogens [85]. The B and T cells are major lymphocytes of the adaptive immune system that play a crucial role in producing humoral and cell-mediated immunity, respectively against invading pathogens [86]. B and T cell epitopes were predicted from three target proteins. A cut-off score of 0.5 was

### Table 1

<table>
<thead>
<tr>
<th>Accession No</th>
<th>Protein Name</th>
<th>No of A.A</th>
<th>M.W</th>
<th>T.PI</th>
<th>Instability Index</th>
<th>Gravy</th>
<th>Antigenicity</th>
<th>Allergenicity</th>
<th>Adhesion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extracellular</td>
<td>Flagellar hook protein FlgE</td>
<td>396</td>
<td>40.8623</td>
<td>4.49</td>
<td>21.24</td>
<td>Stable</td>
<td>-0.112</td>
<td>0.6889</td>
<td>Non-Allergen</td>
</tr>
<tr>
<td></td>
<td>TonB-dependent receptor</td>
<td>661</td>
<td>72.9328</td>
<td>5.71</td>
<td>31.12</td>
<td>Stable</td>
<td>-0.486</td>
<td>0.6018</td>
<td>Non-Allergen</td>
</tr>
<tr>
<td></td>
<td>Porin family protein</td>
<td>236</td>
<td>25.24</td>
<td>4.72</td>
<td>27.58</td>
<td>Stable</td>
<td>-0.071</td>
<td>0.7617</td>
<td>Non-Allergen</td>
</tr>
</tbody>
</table>

3.3. Identification of Immune Epitopes

The immune system's acquired/specific/adaptive immune responses are highly specialized and systemic in their approach to eliminate a pathogen or inhibiting its proliferation in the body [84]. Acquired immunity results in the formation of memory cells, which can recognize the pathogen on future contacts after initial detection [85]. The B and T cells are major lymphocytes of the adaptive immune system that play a crucial role in producing humoral and cell-mediated immunity, respectively against invading pathogens [86]. B and T cell epitopes were predicted from three target proteins. A cut-off score of 0.5 was
utilized to select epitopes for antibodies, and epitopes with a length of the 9-mer sequence were selected. Based on this, it was discovered that FlgE, TonB-dependent receptor and Porin family protein had 8, 14, and 5 B-cell epitopes, respectively. B-cell epitope prediction is critical because immune system protection is dependent on antibodies that trigger multiple protective mechanisms which include (a) agglutination, (b) complement activation, (c) immunoprosoponization (d) neutralization and (e) cell-mediated cytotoxicity [87]. Each B-cell epitope was subsequently subjected to T-cell epitope mapping, which revealed the presence of significant binding sites for MHC-I and MHC-II. MHC-I molecules are found on the surfaces of nucleated cells and function to present peptides from both internal and external proteins to cytotoxic T cells to trigger an immediate immune response that kills the antigen [88]. MHC-II binding sequences in B cell epitopes were predicted at the same time. MHC-II molecules are expressed by molecular antigen presentation cells (APCs) such as B cells, dendritic cells, monocytes, and thymic epithelial cells [89]. Antigenic peptides derived from extracellular proteins are generated and loaded onto the cell surface, leading to activation of helper T cells, localized inflammation, phagocyte recruitment, and development of antibody response [90]. Based on the lowest percentile score, only common MHC-I and MHC-II binding peptide for the proteins was chosen. The epitopes are B-cell-derived T-cell epitopes that simultaneously activate both arms of adaptive immunity. The MHCpred analysis was used to identify the epitopes with the greatest binding affinity to the DRB1*0101 allele, which is the most prevalent and extensively distributed allele in the human population [91]. Only epitopes with the lowest IC50 (< 100 nM) were selected to guarantee great prediction quality. Finally, to avoid vaccine-related allergies, allergenic sequences were removed. The predicted epitopes are tabulated in Table 2. The population coverage of the final set of epitopes is illustrated in Fig. 2.

**Table 2**

<table>
<thead>
<tr>
<th>Protein</th>
<th>B Cell</th>
<th>MHC-I Score</th>
<th>MHC-I Score</th>
<th>MHCpred Score</th>
<th>Allergenicity</th>
<th>Antigenic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flagellar hook protein</td>
<td>FlgE</td>
<td>YLLGAGADEAAGGL</td>
<td>2.9</td>
<td>GADEAAGGL</td>
<td>30</td>
<td>Non-Allergen</td>
</tr>
<tr>
<td></td>
<td>FVPDKNGDLVNSAGGYL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LGAGADEAAGGLTGAV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LNIVNNAAALPAEGST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGDFTVNLPSDQAPAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GYNHTKSLISNYDKGEK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFLPSAEASGDVRLGFPQG</td>
<td>IFLPSAEASGDVRL</td>
<td>31</td>
<td>1.3</td>
<td>FLPSAEASGD</td>
<td>17.8</td>
<td>Non-Allergen</td>
</tr>
<tr>
<td>Outer membrane</td>
<td>KTYRTRDATDMWDPEV</td>
<td>2.4</td>
<td>0.06</td>
<td>TTRTDATDMW</td>
<td>19.2</td>
<td>Non-Allergen</td>
</tr>
<tr>
<td>TonB-dependent receptor</td>
<td>KHDYNTDNYGHNDRRRYKFN</td>
<td>20</td>
<td>60.3</td>
<td>HNDRRYRKFN</td>
<td>0.708</td>
<td>Non-Allergen</td>
</tr>
<tr>
<td></td>
<td>GRISDEAATAANFYDQ</td>
<td>9.9</td>
<td>0.85</td>
<td>RISDEAATA</td>
<td>4.36</td>
<td>Non-Allergen</td>
</tr>
<tr>
<td>Porin family protein</td>
<td>EYRRTDYGHKDFSVDVG</td>
<td>3.1</td>
<td>0.06</td>
<td>SVVDGLSV</td>
<td>19.7</td>
<td>Non-Allergen</td>
</tr>
<tr>
<td></td>
<td>DLSVEARNKFKTHDRLGV</td>
<td>1.6</td>
<td>99.5</td>
<td>DGDLSEAR</td>
<td>1.498</td>
<td>Non-Allergen</td>
</tr>
</tbody>
</table>

### 3.4. Construction of MEBV and Physicochemical Properties Evaluation

A single peptide-based vaccine often fails to trigger proper immune reactions and this can be overcome by developing a MEBV, where multiple antigenic peptides are linked via GPGPG linkers [20]. Additionally, the epitope peptide was linked to the Beta defensin adjuvant with the use of an EAAAK linker to create a more effective combination. Both of the used linkers are rigid and enable effective epitope separation, enabling the host immune system to recognize them efficiently. The primary sequence of the designed MEBV is depicted in Fig. 3a. The MEBV comprises 141 amino acids, which have a molecular weight of 14.91 kDa, an aliphatic index of 46.16, an instability score of 20.92, and a theoretical pl of 8.56. In the vaccine, 17 residues are negatively charged, and 24 residues are positively charged. The vaccine GRAVY score is -0.671 suggesting it is hydrophilic. The MEBV is highly antigenic (score: 1.1261), immunogenic (score: 60.3), soluble (score: 0.66), and non-allergenic. Different physicochemical properties of the MEBV are given in Fig. 3b. The solubility analysis and 3D structure are provided in Fig. 3d and Fig. 3e, respectively.

### 3.5. Secondary and Tertiary Structure prediction

In the secondary structure investigations, it was discovered that 17.81% of the residues form α-helix, whereas 16.44%, 6.85%, and 58.90% of the residues form strand, β- turn, or random coils, respectively. The 3D structure of the designed MEBV was modelled ab initio as no suitable template structure was available. The loop modeling for the following residues was completed; Cys7-Glu28, Gly53-Glu65, Gly70-Ala80, Gly85-pro103, Ala113-Gly118, and Asp122-Arg141 are the Cys33-Gly37, Cys41-Ala47, Gly53-Glu65, Gly70-Ala80, Gly85-pro103, Ala113-Gly118, and Asp122-Arg141. The MEBV structure was refined to eliminate any high-energy contacts (Table 3). Model 1 was chosen because it had better Rama-favored residue mapping (96.5%), and improved rotamer score (0.00%), a clash score of 18.6%, and a molProbity score of 1.992%. The refined MEBV 3D obtained 93.1% of residues in the favored regions, 6.9% in additional allowed regions, 0% in generally allowed regions, and 0% percent residues in disallowed regions.

Page 5/19
Table 3
Refinement analysis of MEBV.

<table>
<thead>
<tr>
<th>Model</th>
<th>GDT-HA</th>
<th>RMSD</th>
<th>MolProbity</th>
<th>Clash score</th>
<th>Poor rotamers</th>
<th>Rama favored</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>1.00</td>
<td>0.00</td>
<td>2.915</td>
<td>57.2</td>
<td>1.9</td>
<td>92.4</td>
</tr>
<tr>
<td>MODEL 1</td>
<td>0.98</td>
<td>0.30</td>
<td>1.99</td>
<td>18.6</td>
<td>0</td>
<td>96.5</td>
</tr>
<tr>
<td>MODEL 2</td>
<td>0.99</td>
<td>0.28</td>
<td>2.14</td>
<td>20.5</td>
<td>0</td>
<td>95.1</td>
</tr>
<tr>
<td>MODEL 3</td>
<td>0.98</td>
<td>0.29</td>
<td>2.04</td>
<td>14.3</td>
<td>0</td>
<td>94.4</td>
</tr>
<tr>
<td>MODEL 4</td>
<td>0.98</td>
<td>0.27</td>
<td>2.10</td>
<td>18.6</td>
<td>0</td>
<td>95.1</td>
</tr>
<tr>
<td>MODEL 5</td>
<td>0.98</td>
<td>0.31</td>
<td>2.07</td>
<td>19.5</td>
<td>0</td>
<td>95.8</td>
</tr>
</tbody>
</table>

3.6. MEBV Docking with Immune Receptors

The vaccine construct was docked to TLR-4, MHC-I, and MHC-II innate immune response receptors to understand MEBV binding and interactions. The docking evaluation predicted top 20 complexes, which were ranked on global binding energy, area size of interacting molecules, and desolvation energy. After that, the complexes underwent a successful refinement test on the FireDock web server. According to the global energy, solution 4 of MEBV-TLR-4, solution 5 of MEBV-MHC-I, and solution 1 of MEBV-MHC-II were selected with net global energy of -45.73 kJ/mol, -20.94 kJ/mol and -3.45 kJ/mol, respectively (Table 4). Figure 4 depicts the docked conformation of the MEBV with TLR4, MHC-I, and MHC-II. Visual examination of the complex reveals that the MEBV binds deeply to the ligand binding site of TLR-4 and MHC molecules. The binding is dominated by hydrogen bonds and weak van der Waals interactions. The PDBsum server was used to acquire detailed insights about MEBV-receptors binding interactions. It revealed the H-bond and non-bond interactions. The interaction details are as under; 4, 0 and 4 hydrogen bonds (HB) between MEBV and TLR-4, MHC-I, and MHC-II, respectively. The non-bonded contacts were 206 (TLR-4), 34 (MHC-I), and 137 (MHC-II). MEBV. Figure 4 shows H-bond interaction maps for each complex.
### Table 4
Docked solutions of MEBV with immune receptors.

#### MEBV - TLR-4

<table>
<thead>
<tr>
<th>Rank</th>
<th>Solution Number</th>
<th>Global Energy</th>
<th>Attractive VdW</th>
<th>Repulsive VdW</th>
<th>ACE</th>
<th>HB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>-45.73</td>
<td>-34.31</td>
<td>24.07</td>
<td>-1.67</td>
<td>-5.83</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>14.71</td>
<td>-6.86</td>
<td>3.00</td>
<td>4.87</td>
<td>-0.28</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>24.57</td>
<td>-29.43</td>
<td>23.80</td>
<td>7.03</td>
<td>-2.77</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>40.87</td>
<td>-14.90</td>
<td>80.06</td>
<td>-0.16</td>
<td>-3.28</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>66.02</td>
<td>-25.44</td>
<td>58.93</td>
<td>9.33</td>
<td>-2.14</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>114.12</td>
<td>-23.60</td>
<td>166.90</td>
<td>6.43</td>
<td>-1.89</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>336.12</td>
<td>-36.37</td>
<td>436.88</td>
<td>14.51</td>
<td>-3.22</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1104.56</td>
<td>-46.80</td>
<td>1393.96</td>
<td>16.98</td>
<td>-6.57</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>1273.37</td>
<td>-43.16</td>
<td>1613.50</td>
<td>20.30</td>
<td>-8.46</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>2073.50</td>
<td>-83.59</td>
<td>2695.10</td>
<td>12.08</td>
<td>-6.95</td>
</tr>
</tbody>
</table>

#### MEBV - MHC-I

<table>
<thead>
<tr>
<th>Rank</th>
<th>Solution Number</th>
<th>Global Energy</th>
<th>Attractive VdW</th>
<th>Repulsive VdW</th>
<th>ACE</th>
<th>HB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>-20.94</td>
<td>-29.87</td>
<td>7.26</td>
<td>15.58</td>
<td>-4.32</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>-5.49</td>
<td>-33.16</td>
<td>24.90</td>
<td>11.10</td>
<td>-3.40</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>-3.25</td>
<td>-8.55</td>
<td>2.55</td>
<td>2.26</td>
<td>-0.42</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>-1.38</td>
<td>-34.29</td>
<td>27.14</td>
<td>19.71</td>
<td>-7.36</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>7.29</td>
<td>-6.88</td>
<td>1.27</td>
<td>5.49</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>18.53</td>
<td>-9.07</td>
<td>1.61</td>
<td>7.81</td>
<td>-0.98</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>19.85</td>
<td>-10.78</td>
<td>1.96</td>
<td>8.14</td>
<td>-0.98</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>23.18</td>
<td>-35.83</td>
<td>43.87</td>
<td>12.24</td>
<td>-3.53</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>52.46</td>
<td>-35.22</td>
<td>79.07</td>
<td>15.70</td>
<td>-3.20</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>231.36</td>
<td>-39.56</td>
<td>297.56</td>
<td>20.37</td>
<td>-6.93</td>
</tr>
</tbody>
</table>

#### MEBV - MHC-II

<table>
<thead>
<tr>
<th>Rank</th>
<th>Solution Number</th>
<th>Global Energy</th>
<th>Attractive VdW</th>
<th>Repulsive VdW</th>
<th>ACE</th>
<th>HB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-3.45</td>
<td>-5.03</td>
<td>1.17</td>
<td>2.24</td>
<td>-2.42</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-0.39</td>
<td>-24.97</td>
<td>17.93</td>
<td>5.90</td>
<td>-0.63</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>1.32</td>
<td>-34.85</td>
<td>30.21</td>
<td>5.56</td>
<td>-1.93</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>4.09</td>
<td>-2.82</td>
<td>0.00</td>
<td>1.36</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>12.22</td>
<td>-1.70</td>
<td>0.00</td>
<td>0.59</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>12.87</td>
<td>-0.14</td>
<td>0.00</td>
<td>0.30</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>15.48</td>
<td>-34.43</td>
<td>24.33</td>
<td>1.12</td>
<td>-5.06</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>23.13</td>
<td>-2.82</td>
<td>0.00</td>
<td>0.68</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>43.22</td>
<td>-30.67</td>
<td>20.13</td>
<td>16.20</td>
<td>-2.75</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>603.77</td>
<td>-47.78</td>
<td>859.33</td>
<td>5.22</td>
<td>-9.93</td>
</tr>
</tbody>
</table>

### 3.7. MD Simulations

The stability of docked complexes was investigated in MD simulation. This was done to check whether the MEBV epitopes are exposed to the host immune system for detection and processing [92]. The dynamics stability of complexes was disclosed using root mean square deviation (RMSD) which was calculated based on the carbon alpha atoms of the systems [93] (Fig. 5A). All three systems revealed a steady increase in the RMSD. However, all of them are stable in
terms of intermolecular binding conformation and chemical interactions. In case of MEBV-TLR4 complex, the maximum RMSD reaches 11 Å, while for MEBV-MHC-I and MEBV-MHC-II, the highest RMSD is 9 and 8 Å, respectively. Secondly, the atomic level fluctuations were noticed via root mean square fluctuation (RMSF) [94] (Fig. 8B). The RMSF reported that the systems are quite stable with respect to residue level deviations. The maximum RMSF deviation reported for the MEBV-TLR4 complex is 8 Å. In the case of MEBV-MHC-I, the maximum RMSF value determined is 4 Å while the MEBV-MHC-II complex is 3 Å.

### 3.8. Hydrogen Bonds Analysis

Hydrogen bonds play a critical role in ensuring the structural integrity of biological systems [95]. Hydrogen bond analysis indicated that the MEBV-MHC-I complex formed more than 350 hydrogen bonds, while the MEBV-MHC-II and MEBV-TLR4 complexes resulted in 200 and 250 hydrogen bonds, respectively. The hydrogen bonds analysis is presented in Fig. 6. These findings suggest that the intermolecular binding between and receptors is very strong and there are high chance that the epitopes are easily recognized by the host immune system for subsequent activation of the immune cells.

### 3.9. MMGB/PBSA Binding Free Energies

To better reveal the atomic level binding energies, MMGB/PBSA analysis was performed. These analyses are vital in deciphering the stronger intermolecular stability of vaccine-receptor complexes [96, 97]. It was found that all three systems showed tremendous stability with net binding energy < -300 kcal/mol. The van der Waals and electrostatic energies were the dominating energies and were found to be the stabilizing factors of complexes. The net MMGBSA binding free energy of MEBV-TLR-4, MEBV-MHC-I and MEBV-MHC-II complex is -452.85 kcal/mol, -300.93 kcal/mol and -326.23 kcal/mol, respectively. In case of MMPBSA, the net binding energy of MEBV-TLR4 is -448.54 kcal/mol, MEBV-MHC-I is -305.33 kcal/mol and MEBV-MHC-II -324.26 kcal/mol. The binding free energies calculated through MMGB/PBSA are given in Fig. 7.

### 3.10. Disulde Engineering and In silico cloning of MEBV

The refined structure of the vaccine was subjected to disulde engineering to increase the stability of the folded conformation, which was achieved by decreasing conformational entropy [98]. Structure was screened for in both the inter- and intra-chains. With 250 residues, the best Chi3 angles were +97/-87 with a tolerance of 30, the best Ca-Cb-S angles were 114.60 with a tolerance of 10, and the best Ca-Cb angles were +97/-87 with a tolerance of 30. A total of 13 pairs of residues were identified to be modified; however, only 10 pairs of residues were selected for a mutant generation. The pair residues are as follows: TYR9-CYS33, TYR10-LYS32, CYS11-CYS19, CYS23-ALA48, CYS24-LYS44, ALA47-GLY51, LEU59-SER68, PRO67-GLU70, GLY79-HIS95, PRO108-PRO134, GLY119-VAL125 having χ3 angle value of -100.15, -111.35, +98.57, +84.84, +113.8, -109.67, -83.66, +125, +119.15, +84.03, +119.15 and energy value (kcal/mol) of 4.55, 5.64, 4.39, 6.48, 2.43, 7.69, 5.45, 5.41, 8.15, 5.09, 4.43, 3.92, respectively. Figure 8A depicts the structure of the original and altered vaccine constructs with disulde links while Fig. 8B illustrates highlighted disulde residues. Codon optimization is a technique that allows improvement of gene expression as well as an increase in translation efficiency. The codon adaptation index (CAI) for the vaccine was 0.93, and the GC content was 58.67 percent. In the context of the expressible process, both values are regarded as optimal. Finally, to maximize the expression of the vaccine candidate, the pET-28a (+) expression vector was used for in silico cloning. Figure 8C depicts the in silico cloned construct.

### 3.11. Immune System Simulation by MEBV

The C-ImmSim server, which uses a position-specific score matrix (PSSM) and other machine learning approaches to predict and analyze epitope and immunological interactions, was used to execute a computational immune simulation to test the vaccine's immunogenic potency. The MEBV was seen in immunological interactions, was used to execute a computational immune simulation to test the vaccine's immunogenic potency. The MEBV was seen in

### 3.12. MMGB/PBSA Binding Free Energies

To better reveal the atomic level binding energies, MMGB/PBSA analysis was performed. These analyses are vital in deciphering the stronger intermolecular stability of vaccine-receptor complexes [96, 97]. It was found that all three systems showed tremendous stability with net binding energy < -300 kcal/mol. The van der Waals and electrostatic energies were the dominating energies and were found to be the stabilizing factors of complexes. The net MMGBSA binding free energy of MEBV-TLR-4, MEBV-MHC-I and MEBV-MHC-II complex is -452.85 kcal/mol, -300.93 kcal/mol and -326.23 kcal/mol, respectively. In case of MMPBSA, the net binding energy of MEBV-TLR4 is -448.54 kcal/mol, MEBV-MHC-I is -305.33 kcal/mol and MEBV-MHC-II -324.26 kcal/mol. The binding free energies calculated through MMGB/PBSA are given in Fig. 7.

### 3.13. Disulfide Engineering and In silico cloning of MEBV

The refined structure of the vaccine was subjected to disulfide engineering to increase the stability of the folded conformation, which was achieved by decreasing conformational entropy [98]. Structure was screened for in both the inter- and intra-chains. With 250 residues, the best Chi3 angles were +97/-87 with a tolerance of 30, the best Ca-Cb-S angles were 114.60 with a tolerance of 10, and the best Ca-Cb angles were +97/-87 with a tolerance of 30. A total of 13 pairs of residues were identified to be modified; however, only 10 pairs of residues were selected for a mutant generation. The pair residues are as follows: TYR9-CYS33, TYR10-LYS32, CYS11-CYS19, CYS23-ALA48, CYS24-LYS44, ALA47-GLY51, LEU59-SER68, PRO67-GLU70, GLY79-HIS95, PRO108-PRO134, GLY119-VAL125 having χ3 angle value of -100.15, -111.35, +98.57, +84.84, +113.8, -109.67, -83.66, +125, +119.15, +84.03, +119.15 and energy value (kcal/mol) of 4.55, 5.64, 4.39, 6.48, 2.43, 7.69, 5.45, 5.41, 8.15, 5.09, 4.43, 3.92, respectively. Figure 8A depicts the structure of the original and altered vaccine constructs with disulfide links while Fig. 8B illustrates highlighted disulfide residues. Codon optimization is a technique that allows improvement of gene expression as well as an increase in translation efficiency. The codon adaptation index (CAI) for the vaccine was 0.93, and the GC content was 58.67 percent. In the context of the expressible process, both values are regarded as optimal. Finally, to maximize the expression of the vaccine candidate, the pET-28a (+) expression vector was used for in silico cloning. Figure 8C depicts the in silico cloned construct.

### 3.14. Immune System Simulation by MEBV

The C-ImmSim server, which uses a position-specific score matrix (PSSM) and other machine learning approaches to predict and analyze epitope and immunological interactions, was used to execute a computational immune simulation to test the vaccine's immunogenic potency. The MEBV was seen in increasing generation of adaptive immune responses in the form of IgG and IgM antibodies. The IgM antibody was also detected at a significant concentration. According to Fig. 9A, secondary immune responses were followed by tertiary immune responses, resulting in the highest level of B-cell generation as well as the highest levels of "IgM + IgG, IgM, and IgM, IgG1 + IgG2, IgG1 and IgG2" production. Similarly, as shown in Fig. 9B, the generation of interferon-gamma was greater than 400,000 ng/ml for over three weeks.

### 4. Conclusions

In this study, three proteins (Flagellar hook protein (FlgE), TonB-dependent receptor, and Porin family protein) are prioritized as potential subunit targets against B. melitensis. The proteins are part of the pathogen core proteome and are found in all bacterial sequencing strains. These proteins met eligibility checklist such as being antigenic, non-homologous to the human host, and highly antigenic to trigger the host immune system. Further, these proteins contain potential B and T-cell epitopes that were used in multi-epitope vaccine designing. The designed molecule showed robust interactions with MHC-I, MHC-II, and TLR-4 immune receptors and was found to have stable time dependent dynamics. Previous literature found several productive epitope based vaccine design efforts against intracellular pathogens like B. melitensis. For example, Fang et al., reported EscE protein harbors potential epitopes to activate protective CD4 + and CD8 + T-cell responses against Edwardsiella tarda in an animal model [99]. In another work, outer membrane protein (Omp) of E. tarda interacts with MHC-I and MHC-II alleles [100]. Several subunit vaccine candidates have also been tested against B. melitensis. For example, Omp31 [101], Omp16 [102], rE2o [103], Dnak [104], OMP25-BLS fusion protein [105], Ndk, Dps, InpB and Aspc [106], etc [107] but none of them can clear the pathogen. Similarly, some studies reported nanoparticles containing Brucella vaccine candidates to induce IgM, IgG and mucosal IgA, together with IL-6, IFN-γ, and IL-12 in animal models [108]. Several Brucella DNA vaccines are also proposed which are safe, elicit strong cellular immune responses and simultaneous expression of antigens, and requirement of simple storage conditions. Some of the DNA vaccines are Cu-Zn superoxide dismutase [109], Brucella lumazine synthase [110], antigenic surface protein [111] and ribosomal protein L9 [112]. Hence, the designed vaccine might be used in in vivo and in vitro studies to determine its real immune protection efficacy. Furthermore, because the study is computationally based, experimental validation is highly encouraged.

### Declarations


**Author Contributions:** Conceptualization, Alaa R. Hameed, Bassam Qasim Mohammed, Tabarak Sabah Jassim and Sajjad Ahmad; Methodology, Alaa R. Hameed, Bassam Qasim Mohammed, Tabarak Sabah Jassim and Sajjad Ahmad; Software, Alaa R. Hameed, Bassam Qasim Mohammed, Tabarak Sabah Jassim and Sajjad Ahmad; Validation, Alaa R. Hameed, Bassam Qasim Mohammed, Tabarak Sabah Jassim and Sajjad Ahmad; Formal analysis, Alaa R. Hameed, Bassam Qasim Mohammed, Tabarak Sabah Jassim and Sajjad Ahmad; Investigation, Alaa R. Hameed and Sajjad Ahmad; Resources, Alaa R. Hameed, Bassam Qasim Mohammed, Tabarak Sabah Jassim and Sajjad Ahmad; Writing – original draft, Alaa R. Hameed, Bassam Qasim Mohammed, Tabarak Sabah Jassim and Sajjad Ahmad; Data curation, Alaa R. Hameed, Bassam Qasim Mohammed and Tabarak Sabah Jassim; Writing – review & editing, Alaa R. Hameed and Sajjad Ahmad; Visualization, Bassam Qasim Mohammed and Sajjad Ahmad; Supervision, Alaa R. Hameed and Sajjad Ahmad; Project administration, Alaa R. Hameed and Sajjad Ahmad; Funding acquisition, Sajjad Ahmad.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available within the article.

**Acknowledgments:** The authors extend appreciation to Dijlah University College for funding this work.

**Conflicts of Interest:** The author declares no conflict of interest.

**References**


Figures

Figure 1

The flow chart of the entire project. Antigenic factors were evaluated in order to predict B-cell derived T-cell epitopes. The final design was obtained and subjected to a series of tests to determine its biological potency as a good candidate.
Figure 2

Population coverage analysis of selected antigenic epitopes.
Figure 3

Different sequence analysis of MEBV. A. primary sequence of the vaccine, B. Physicochemical analysis, C, D. Water solubility and E. 3D structure.

Figure 4
Docked pose and interactions of MEBV with immune receptors.

**Figure 5**
Molecular dynamics simulation analysis of vaccine-receptor complexes. RMSD (A) and RMSF (B).

**Figure 6**
Time dependent hydrogen bonds formed between vaccine and immune receptors.
Figure 7

Intermolecular vaccine candidate with receptors binding free energies in kcal/mol.
Figure 8

A. Disulfide engineering of designed vaccine candidate, B. structure with disulfide bonds, C. cloned pET-28a (+) expression vector.
Figure 9

C-Immune simulation analysis. Production of different immunoglobulins (A) and cytokines and interferon (B) to the antigen.

Supplementary Files

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

- SupplementaryFiles.docx
- GraphicalAbstract.jpg