

Design of a multi-epitope vaccine candidate against carrion disease by immunoinformatics approach

Damaris Rivera-Asencios^a, Abraham Espinoza-Culupú^a, Sheyla Carmen-Sifuentes^b, Pablo Ramirez^a, Ruth García-de-la-Guarda^{a,*}

^a Molecular Microbiology and Biotechnology Laboratory, Faculty of Biological Sciences, Universidad Nacional Mayor de San Marcos, Lima, Peru

^b Faculty of Biological Sciences, Universidad Nacional Mayor de San Marcos, Lima, Peru

ARTICLE INFO

Keywords:

Multi-epitope

Vaccine

TLR4

Immunoinformatics

Bartonella bacilliformis

Carrion's disease

ABSTRACT

Carrion's disease, caused by the bacterium *Bartonella bacilliformis*, is a serious public health problem in Peru, Ecuador and Colombia. Currently there is no available vaccine against *B. bacilliformis*. While antibiotics are the standard treatment, resistant strains have been reported, and there is a potential spread of the vector that transmits the bacteria. This study aimed to design a multi-epitope vaccine candidate against the causative agent of Carrion's disease using immunoinformatics tools. Predictions of B-cell epitopes, as well as CD4⁺ and CD8⁺T cell epitopes, were performed from the entire proteome of *B. bacilliformis* KC583 using the most frequent alleles from Peru, Ecuador, Colombia, and worldwide. B-cell epitopes and T-cell nested epitopes from outer membrane and virulence-associated proteins were selected. Epitopes were filtered out based on promiscuity, non-allergenicity, conservation, non-homology and non-toxicity. Two vaccine constructs were assembled using linkers. The tertiary structure of the constructs was predicted, and their stability was evaluated through molecular dynamics simulations. The most stable construct was selected for molecular docking with the TLR4 receptor. This study proposes a vaccine construct evaluated in silico as a potential vaccine candidate against *Bartonella bacilliformis*.

1. Introduction

Bartonella bacilliformis is a gram-negative, mobile pleomorphic and intraerythrocytic bacterium that causes the neglected Carrion's disease, also known as Oroya's fever, in humans [1]. Transmitted by female sand-flies of the genus *Lutzomyia*, primarily *L. verrucarum* and *L. peruensis* [2], the disease is endemic in Peru, Ecuador and Colombia. In Peru it represents a serious public health problem, particularly in inter-Andean valleys and jungle areas [3]. Carrion's disease is characterized by two well-established clinical phases that can occur sequentially or independently [4,5]. During acute phase, *B. bacilliformis* invades human erythrocytes mainly enhanced by flagella, IalB and deformin and hemolysin proteins [1,6,7]. The chronic phase, known as verruga peruana (Peruvian warts) involves abnormal proliferation of endothelial cells in the terminal vasculature of the dermis and subcutaneous tissue, resulting in characteristic dermal eruptions [5]. Although Carrion's disease is currently restricted to certain areas, the expansion of the vector's distribution could lead to its spread to other regions in the

future [8]. Moreover, *B. bacilliformis* has been reported to exhibit constitutive resistance to quinolones [9]. Recent epidemiological reports from Peru, indicate ongoing cases of both acute and chronic Carrion's disease, underscoring the urgent need for strategies to control the disease.

The development of a vaccine could be one of the most feasible and effective strategies for disease control. To date, no vaccine is available, although some vaccine candidates against *B. bacilliformis* have been developed in recent years. One approach has been the design of a multi-epitope protein as a vaccine candidate. In a previous study, *B. bacilliformis* KC583 proteome was analyzed using bioinformatic tools, leading to the development of a multi-epitope vaccine called M1. Cloning, expression and purification of M1 were performed in *E. coli*, followed by mice immunization. In vitro erythrocyte invasion tests demonstrated the potential effectiveness of M1; however, further studies are needed to fully characterize its immunogenic properties and its viability as a vaccine [10]. An immunoinformatics and subtractive proteomic approach was employed to select the best candidates for

* Corresponding author.

E-mail address: rgarciad@unmsm.edu.pe (R. García-de-la-Guarda).

<https://doi.org/10.1016/j.combiomed.2024.109397>

Received 28 June 2024; Received in revised form 9 October 2024; Accepted 7 November 2024

Available online 20 November 2024

0010-4825/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

vaccine design, including the prediction of epitopes and in silico vaccine construction. The final vaccine construct was modeled for docking with Toll-like receptor 4 (TLR-4), and molecular dynamics analysis was conducted. The final vaccine construct was also cloned in silico in *E. coli*. However, future experimental validation is required [11].

In our study, we employed an immunoinformatics approach to design a potential vaccine candidate. Utilizing the *B. bacilliformis* KC583 proteome, we predicted T and B-cell epitopes and applied several filters were employed to select the most effective nested epitopes (linear CD4⁺ T-cell epitopes containing linear CD8⁺ T-cell epitopes within their sequence) capable of stimulating both CD4⁺ and CD8⁺ T-cell responses. This process led to the design of two vaccine constructs. We designed two vaccine constructs based on these nested epitopes and, to assess the potential innate immune response, we conducted Molecular dynamics simulation and docking analysis with TLR-4, which is a key receptor that activates the innate immune system by recognizing pathogens, triggering cytokine production and antigen-presenting cell activation, making it a vital target for enhancing multi-epitope vaccines immune responses [12].

Based on our comprehensive analysis, we present a promising vaccine candidate against *B. bacilliformis*.

2. Materials and methods

The study framework used for designing multi-epitope vaccine constructs against *B. bacilliformis* is illustrated in Fig. 1.

2.1. Retrieval of proteome of *Bartonella bacilliformis*

The complete proteome of *B. bacilliformis* strain KC583 (UP000000643) was obtained from the UniProt database (<https://www.uniprot.org/proteomes/UP000000643>) in FASTA format.

2.2. Evaluation of human non-homologous proteins

The *B. bacilliformis* proteome was evaluated to identify proteins that are non-homologous to human proteins. BLASTp tool was used against the *Homo sapiens* proteome obtained from the UniProt server following the methodology described by Samavia [13]. Proteins that did not show significance were selected for the subsequent analyses.

2.3. T-cell epitope prediction

MHC I and MHC II binding epitopes were predicted using NetMHCpan-4.0 and NetMHCIIpan-4.0, respectively [14,15]. These tools analyze peptides ranging from 8 to 14 amino acids (8-mer peptides were selected for this study) and 13 to 25 amino acids (15-mer peptides were selected for this study), respectively. Both tools provide outputs in the form of binding affinities (IC50 values) and percentile ranks. Peptides with IC50 < 50 nM were categorized as strong binders and were selected.

The most frequent HLA alleles were divided into two groups: group A, representing the world population, (HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-A*26:01, HLA-B*07:02, HLA-B*08:01, HLA-B*15:01, HLA-B*27:05, HLA-B*39:01, HLA-B*40:01 and HLA-B*58:01) and group B, representing Peru, Ecuador and Colombia population, (HLA-A*02:01, HLA-A*24:02, HLA-B*51:01, HLA-C*04:01 and HLA-C*07:02). MHC II binding epitopes were also predicted for the most frequent alleles in the world population (group A), including HLA-DRB1*03:01, HLA-DRB1*07:01, HLA-DRB1*15:01, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB4*01:01 and HLA-DRB5*01:01, as well as for Peru, Ecuador and Colombia population (group B) (HLA-DRB1*04:07, HLA-DRB1*08:02 and HLA-DRB1*14:02).

2.4. Nested epitope identification

As previously described, nested epitopes, were obtained [16]. We selected nested epitopes for outer membrane proteins predicted by PSORT [17]. PSORT predicts subcellular localization analyzing the amino acid sequence. To identify virulence-associated proteins, we employed VFDB [18], as well as those already reported as possible vaccine candidates [1,19].

2.5. B-cell epitope prediction

We used the online server BepiPred-2.0 (<https://services.healthtech.dtu.dk/services/BepiPred-2.0/>) [20] to predict 15-aminoacids B-cell epitopes corresponding to outer membrane and virulence-associated proteins. This tool combines a hidden Markov model with propensity scale methods to identify regions likely to be recognized by B-cell epitopes. Epitopes with scores greater than 0.5 and 0.55, respectively, were selected.

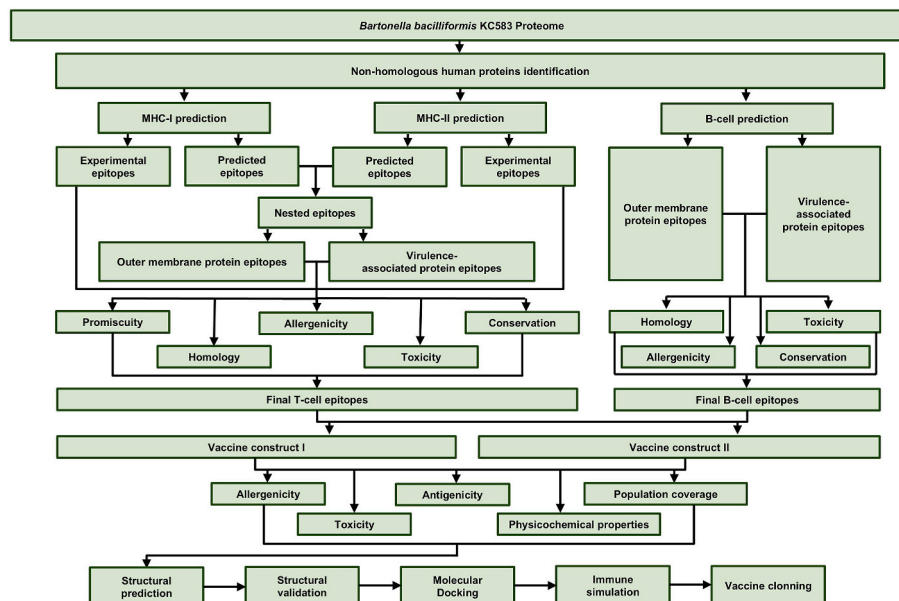


Fig. 1. Flowchart of the methodology that was used for designing multi-epitope vaccine constructs against *B. bacilliformis*.

2.6. Epitope evaluation

All the epitopes identified in T and B-cells were evaluated for promiscuity, human proteome non-homology, allergenicity, toxicity and conservation using online bioinformatics tools.

2.6.1. Promiscuous epitope identification

Promiscuous epitopes are defined as T-cell epitopes that bind to many alleles. To select them, we used a cutoff of number of alleles.

2.6.2. Identification of non-homologous epitopes

Epitopes that showed 100 % identity and sequence coverage with the human proteome (UP000005640) and the mouse proteome (*Mus musculus*) (UP000005589) were excluded to avoid potential autoimmune responses in humans and future clinical trials in mice.

2.6.3. Allergenicity of the epitopes

All selected epitopes were evaluated for potential allergic reactions using the AllerTOP online server (<https://www.ddg-pharmfac.net/AllerTOP/>) [21]. This tool utilizes machine learning algorithm based on auto-cross covariance (ACC) transformation to classify proteins as either "allergen" or "non-allergen". The input sequences are compared against a curated database of known allergens and non-allergens. In this study, only epitopes classified as non-allergenic were selected.

2.6.4. Toxicity of the epitopes

To determine the potential toxicity of all epitopes, the ToxinPred tool (https://webs.iitd.edu.in/raghava/toxinpred/multi_submit.php) was used [22]. This tool utilizes machine learning models based on amino acid composition and known toxic motifs to classify peptides as "toxic" or "non-toxic." Only epitopes classified as non-toxic were selected.

2.6.5. Epitope conservation in *B. bacilliformis* genomes

The conservation of the epitopes selected by the previous filters was analyzed across all reported genomes of *B. bacilliformis* in the NCBI database. This analysis was performed using IEDB epitope conservancy tool (<http://tools.iedb.org/conservancy/>) [23].

2.7. Multi-epitopes vaccine construction

The previously selected epitopes were overlapped and, any repeated epitopes in group A and B, were manually eliminated. The final filtered epitopes along with the adjuvant β -defensine were concatenated using linkers to form two vaccine constructs (vaccine construct I and II). Vaccine construct I was composed of nested epitopes from outer membrane proteins, B-cell epitopes, and experimental epitopes. Vaccine construct II was composed of nested epitopes from virulence-associated proteins, B-cell epitopes and experimental epitopes. The "EAAAK" linker was used to bind the adjuvant and the first nested T-cell epitope. Nested T-cell, experimental and B-cell epitopes were linked via GPGPS, KK, and AAY linkers, respectively (Fig. 2).

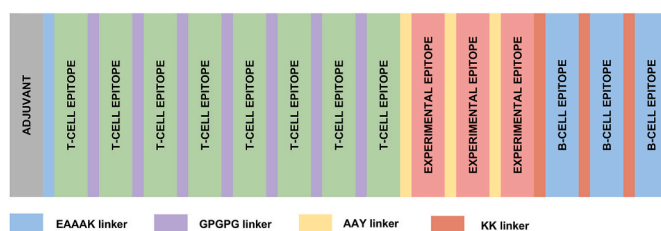


Fig. 2. Schematic representation of the vaccine constructs.

2.8. Allergenicity, toxicity, antigenicity and other physicochemical properties assessment for vaccines constructs

The vaccine constructs were assessed using several bioinformatics tools. AllerTOP [21], ToxDL [24] and VaxiJen v2.0 [25] were utilized. VaxiJen predicts the antigenicity of the selected epitopes based on an alignment-independent method, using physicochemical properties of proteins to classify them as "antigenic" or "non-antigenic". The threshold value of the last server was 0.4. The amino acid composition, molecular weight, half-life, instability index, aliphatic index, theoretical PI and GRAVY (grand average of hydropathicity) were analyzed using ProtParam (<https://web.expasy.org/protparam/>) [26]. Water solubility predictions were performed with ProteinSol (<https://protein-sol.manchester.ac.uk/>) [27]. Additionally, the presence of signal peptides and transmembrane helices was evaluated using SignalP 5.0 (<https://service.s.healthtech.dtu.dk/services/SignalP-5.0/>) [28] and TMHMM v. 2.0 (<https://services.healthtech.dtu.dk/services/TMHMM-2.0/>) [29], respectively.

2.9. Population coverage

The population coverage of the selected epitopes was analyzed using the IEDB population coverage tool (<http://tools.iedb.org/population/>) [30]. This analysis included both group A and group B alleles.

2.10. Secondary structures prediction

The secondary structures of the vaccine constructs were predicted and analyzed PSIPRED 4.0 (<http://bioinf.cs.ucl.ac.uk/psipred/>) [31], to determine their transmembrane topology. This tool employs a neural network-based approach to predict alpha helices, beta strands and coils from amino acid sequences.

2.11. Tertiary structure prediction and refinement

The tertiary structure of the vaccine constructs was predicted using the Robetta server (<https://rosetta.bakerlab.org/>) [32], which employs the Rosetta de novo structure prediction methodology. The refinement of these models was conducted with the ModRefiner server (<https://zhanglab.dcm.med.umich.edu/ModRefiner/>) [33].

2.12. Tertiary structure validation

For structural validation of the models we employed ERRAT [34], Prosa-web (<https://prosa.services.came.sbg.ac.at/prosa.php>) [35] and PROCHECK 3.6.2 [36]. ERRAT assesses the accuracy of protein models by analyzing the statistics of non-bonded atomic interactions. The overall quality factor generated by ERRAT was used to ensure that the predicted structures were reliable. Prosa-web calculates a Z-score that reflects overall quality of the protein model by comparing it with experimentally validated structures. Models with a Z-score within the range of native proteins were considered reliable and selected for further analysis. PROCHECK evaluates the stereochemical quality of the predicted structures by analyzing Ramachandran plots. Models with a high percentage of residues in favored regions were considered reliable.

The best model for each vaccine construct was then selected.

2.13. Energy optimization and simulation of molecular dynamics

Energy minimization of the best models for vaccine constructs I and II was carried out using the steepest descents algorithm of GROMACS (Groningen Machine for Chemical Simulations) [37]. TIP3P was chosen as the water model, and the AMBER99SB-ILDN force field was used to generate the topology file required for energy minimization and equilibration. AMBER99SB-ILDN force field has extensive validation and acceptance in protein studies. The ILDN refinement corrects side chain

interactions and enhances the accuracy of protein dynamics in atomic-level simulations [38]. Following equilibration, a 40 ns molecular dynamics simulation was performed to determine the RMSD (Root Mean Square Deviation) of the vaccine constructs' backbone.

2.14. Molecular docking with TLR4 protein

TLR4 crystal structure was obtained from the PDB (Protein Data Bank) (ID: 3FXI), a single chain was selected and binding sites, covalent linkages as well as metal coordination atoms, were removed. The molecular docking analysis was performed using the HADDOCK 2.4 program [39]. The docking protocol involved rigid-body energy minimization, semi-flexible refinement and solvent refinement. Active residues for TLR4 were selected from a previous study [11]. All other parameter settings were set as default. The docking results were filtered to identify the lowest HADDOCK score, which is based on electrostatic, van der Waals and desolvation energies. The HADDOCK Refinement Interface was then used to refine the selected cluster, improving the precision of the docking results. Finally, the interaction residues between the multi-epitope vaccine and the TLR4 protein were mapped using PDBsum [40].

2.15. Immuno-simulation of constructed vaccine

The immune response to vaccine construct was examined using the simulation server C-ImmSim (<https://kraken.iac.rm.cnr.it/C-IMMSIM/>) [41]. This tool employs a position-specific matrix (PSSM) and models both humoral and cell-mediated immune responses over time. The results provided insights into cytokine production, T-cell responses, and antibody generation, helping to evaluate the potential immunogenicity of the vaccine construct. Default settings were applied. The immune simulation was conducted with homozygous host haplotypes HLA-A*0201, HLA-A*2401, HLA-B*3901, HLA-B*5801, HLA-DRB4*0101 and HLA-DRB5*0101.

2.16. In silico codon optimization and cloning of constructed vaccine

The amino acid sequence of vaccine construct II was translated into DNA and optimized for expression in *E.coli* using the JCat (Java Codon Adaptation Tool) [42] and Geneious Prime® 2023.2.1 program (<https://www.geneious.com>). The optimized sequence was then cloned into the pET28a + vector by adding EcoRI and HindIII restriction sites at the C and N terminal, respectively. Parameters such as the codon adaptation index (CAI) and percent GC content for the vaccine construct were evaluated. Additionally, a His 6x tag sequence and a termination codon (TGA) were added.

3. Results

3.1. Non-homologous human proteins

A total of 181 homologous proteins were excluded due to their similarity to the human proteome. As a result, 1074 proteins were selected for further analysis.

3.2. Predicted T-cell epitopes

For group A and B, 131,501 and 68,960 total epitopes were obtained, respectively. After removing repeated epitopes, the nested epitopes were selected (Table 1).

3.3. Identified novel nested epitopes

A total of 38,234 and 11,487 novel nested epitopes were obtained for group A and B, respectively. Epitopes from outer membrane and virulence-associated proteins were selected, as shown in Table 2.

Table 1
Number of total T-cell epitopes obtained for group A and B.

Group	Epitopes	Predicted epitopes		Experimental epitopes		Total
		CD8 ⁺	CD4 ⁺	CD8 ⁺	CD4 ⁺	
A	Total epitopes	88,805	42,686	10	–	131,501
	Non-repeated epitopes	88,174	42,439	9	–	130,633
B	Total epitopes	51,538	17,419	3	–	68,960
	Non-repeated epitopes	51,130	17,309	2	–	68,441

Table 2
Number of total nested epitopes obtained for group A and B.

Group	Total Nested epitopes	Proteins	Nested epitopes
A	38,234	Outer membrane	1826
		Virulence-associated	1373
B	11,487	Outer membrane	542
		Virulence-associated	453

3.4. Predicted B-cell epitopes

We obtained 79 and 24 non-repeated B-cell epitopes from outer membrane and virulence-associated proteins, respectively (Tables S1–S2).

3.5. Epitopes evaluation

3.5.1. Promiscuous epitopes

From group A, nested epitopes corresponding to outer membrane proteins that interacted with 2–13 alleles were identified. We selected those interacting with a total number of alleles (MHC I and II) ≥ 7 , resulting in 63 epitopes. Additionally, epitopes from virulence-associated proteins interacting with 2–11 alleles were identified, and those that interacted with a total number of alleles (MHC I and II) ≥ 9 were selected, resulting in 44 epitopes. From group B, nested epitopes corresponding to outer membrane proteins interacting with 2–7 alleles were identified. We selected those interacting with a total number of alleles (MHC I and II) ≥ 6 , resulting in 31 epitopes. Furthermore, epitopes from virulence-associated proteins interacting with 2–8 alleles were identified and those interacting with a total number of alleles (MHC I and II) ≥ 6 were selected, yielding 30 epitopes.

From the total numbers of experimental epitopes, those binding to a total number of alleles (MHC I and II) ≥ 3 were selected, resulting in 5 epitopes for group A and 2 epitopes for group B meeting this criterion.

3.5.2. Non-homologous epitopes to the human and mouse proteome

None of the previously selected nested epitopes corresponding to outer membrane and virulence-associated proteins from groups A and B showed 100 % identity or coverage with any region of the human or murine proteome. For the experimental epitopes, a total of 4 from group A and 1 from group B were obtained. Similarly, none of the B-cell epitopes corresponding to outer membrane and virulence-associated proteins exhibited 100 % identity or coverage with any region of the human or murine proteome.

3.5.3. Non-allergen epitopes

In group A, 34 nested epitopes from outer membrane epitopes and 22 nested from virulence-associated proteins were identified as non-allergenic. In group B, 21 nested epitopes from outer membrane proteins and 20 from virulence-associated proteins were also identified as non-allergenic. Among the experimental epitopes, 3 from group A and 1

from group B were determined to be non-allergenic. Regarding the B-cell epitopes, 40 and 22 from outer membrane and virulence-associated proteins, respectively, were found to be non-allergenic.

3.5.4. Non-toxic epitopes

None of the previously selected epitopes (nested, experimental and B-cell epitopes) were predicted to be potentially toxic.

3.5.5. Conserved epitopes among all *B. bacilliformis* strains

In group A, 11 nested epitopes from outer membrane proteins and 14 from virulence-associated proteins were conserved among all the strains. In group B, 18 nested epitopes from outer membrane proteins and 6 from virulence-associated proteins, were similarly conserved. Among the experimental epitopes, 3 from group A and 1 from group B were conserved across all the strains. Regarding B-cell epitopes, 9 from outer membrane proteins and 11 from virulence-associated proteins were conserved among all strains.

3.6. Vaccine constructs

We finally identified 14 epitopes (11 from outer membrane proteins and 3 experimental epitopes) that comprised vaccine construct I (Table 3) (Table S3). Additionally, we obtained 15 epitopes (12 from virulence-associated proteins and 3 experimental epitopes) that comprised vaccine construct II (Table 4) (Table S4).

3.7. Allergenicity, toxicity, antigenicity and other physicochemical properties

Neither of the vaccine constructs were predicted to be potentially allergens or toxic. Vaccine construct I comprised 306 amino acids with a molecular weight of 33.18 kDa, while vaccine construct II consisted of 327 amino acids with a molecular weight of 35.46 kDa. Both vaccine constructs had antigenicity scores above the threshold value. The theoretical PI was 9.37 for vaccine construct I and 9.82 for vaccine construct I and II, indicating a slightly basic nature. The instability index suggested the stable state of both vaccine constructs. The half-life was 30 h for construct I and over 10 h for construct II in both mammals and *E. coli*. The GRAVY scores were −0.338 for vaccine construct I and

Table 3

Final epitopes of the vaccine construct I. The epitope and the corresponding outer membrane protein are indicated.

Epitopes	Protein	
	Protein ID	Protein name
FAISLSSYGKNTLSS	A1UTU0	Outer membrane autotransporter
KRAIAAMQLLIERYP	A1UTB7	Outer membrane protein assembly factor BamD
ARSNYREILSALYAD	A1UUD8	Surface antigen/outer membrane protein, OMP85 family
NTQNGFIIPTLALRADT	A1US68	LPS-assembly protein LptD
YAKYLVDASSYNVKA	A1USK0/	Type I secretion outer membrane
	A1USN1	protein, TolC family
EQNLFLDVIKAYANVY	A1USK0/	Type I secretion outer membrane
	A1USN1	protein, TolC family
EGIYNYVNTQHVIEIE	A1URQ3	Putative lipoprotein
FRGAVFADAATLYGNVY	A1USE4	Outer membrane protein assembly factor BamA
ISASLAYSIVQEEYN	A1USE4	Outer membrane protein assembly factor BamA
FLSAQVKMQAQRESFRNAE	A1USK0/	Type I secretion outer membrane
	A1USN1	protein, TolC family
QAQAARSFAVSEYSLARS	A1USK0/	Type I secretion outer membrane
	A1USN1	protein, TolC family
VLDEPSIGL	A1UST0	UvrABC system protein A
WTIGYDTIY	A1URV9	4-hydroxybenzoate octaprenyltransferase
KLYPNVDFY	A1USF4	Citrate synthase

Table 4

Final epitopes of the vaccine construct II. The epitope and the corresponding virulence-associated protein are indicated.

Epitopes	Protein	
	Protein ID	Protein name
PRKVTFESSNSLPAG	A1UTS9	Flagellar hook-associated protein FlgK
KAARVHIVLPDKGSF	A1UTT7	Flagellar M-ring protein
MTRLIEVSRAFERVEAM	A1UTV2	Flagellar basal body protein
NARRIIAQTAIKLSEQ	A1UTV8	Putative flagellar motor switch protein Flg
EYIRQKLVLDLSLFI	A1UTV9	Flagellar biosynthetic protein FlhB
PLDFSAPFTVLLIAT	A1UTW6	Flagellar biosynthesis protein FlhA
QRLLPQQVELANRVA	A1UTW6	Flagellar biosynthesis protein FlhA
MAPQYQKQLEDEMASA	A1UTS9	Flagellar hook-associated protein FlgK
QKPYQAMQNMADALP	A1UTV4	Lateral flagellar motor protein MotA
LMDEEIHITSSDSQK	A1UTV4	Lateral flagellar motor protein MotA
DEYKNLEGNPMVKSRMR	A1UTV9	Flagellar biosynthetic protein FlhB
NPMVKSRMRSLARDRIRKMLIA	A1UTV9	Flagellar biosynthetic protein FlhB
VLDEPSIGL	A1UST0	UvrABC system protein A
WTIGYDTIY	A1URV9	4-hydroxybenzoate octaprenyltransferase
KLYPNVDFY	A1USF4	Citrate synthase

−0.471 for vaccine construct II, indicating the hydrophilic nature of both vaccine constructs. The aliphatic index was 75.39 for vaccine construct I and 72.57 for vaccine construct II, further suggesting their stability. The results are shown in Table 5. Additionally, the calculated solubility was 0.647 for vaccine construct I and 0.638 for vaccine construct II (Supplementary Fig. S1). No signal peptides or trans-membrane helices were identified in either vaccine constructs.

3.8. Population coverage analysis

All epitopes from vaccine construct I had MHC I and MHC II coverage score of almost 90 % (Table 6) (Supplementary Tables S5–S6). The same value was obtained for all epitopes from vaccine construct II (Table 7) (Supplementary Tables S7–S8).

3.9. Secondary structure prediction

Vaccine construct I consists of 43.79 % α -helices, 18.63 % β -sheets and 37.58 % coils (Fig. 3A). The vaccine construct II has 47.09 %, 9.79 % β -sheets and 43.12 % coils (Fig. 3B).

3.10. Tertiary structure, refinement, and validation

A total of five models were generated for each vaccine construct and visually evaluated using PyMOL. Subsequent refinement was performed. Ramachandran plot analysis was conducted on the structure of the five models before and after refinement, and the best-refined model of each construct was selected for further analysis. The PROCHECK server revealed a Ramachandran plot of the best model of the vaccine construct I with 97.6 % residues in the most favored regions, 2.4 % residues in the additional allowed regions, and none in the generously allowed or disallowed regions (Fig. 4A). For vaccine construct II, it the Ramachandran plot showed 97.3 % residues in the most favored regions, 2.7 % residues in the additional allowed regions, and no residues in the generously allowed or disallowed regions (Fig. 4B).

According to ERRAT, the calculated score for the best model of vaccine construct I was 89.9654 and for vaccine construct II, it was 92.0266. The z-score according to Prosa-web was −6.35 (Fig. 5A) and −6.13 (Fig. 5B) for the best model of the vaccine construct I and II, respectively.

Table 5

Physicochemical analysis for vaccine constructs I and II.

Vaccine construct	Allergenicity	Toxicity	Antigenicity	No. of amino acids	Molecular weight	Theoretical pI	Half life in mammals	Half life in <i>E. coli</i>	Instability index	Aliphatic index	GRAVY
I	No	0.001	0.7348	306	33.18 kDa	9.37	30 h	>10 h	31.17	75.39	−0.338
II	No	0.0009	0.6283	327	35.46 kDa	9.82	30 h	>10 h	35.23	72.57	−0.471

Table 6

The calculated MHC I and MHC II population coverage of epitopes from vaccine construct I.

Population/area	Class combined		
	Coverage ^a	Average hit ^b	pc90 ^c
Colombia	66.25 %	3.07	0.3
Ecuador	93.87 %	5.37	1.39
Peru	98.91 %	11.86	5.6
World	97.07 %	17.56	7.03
Average	89.02	9.46	3.58
Standard deviation	13.27	5.68	2.81

^a Projected population coverage.^b Average number of epitope hits/HLA combinations recognized by the population.^c Minimum number of epitope hits/HLA combinations recognized by 90 % the population.**Table 7**

The calculated MHC I and MHC II population coverage of epitopes from vaccine construct II.

Population/area	Class combined		
	Coverage ^a	Average hit ^b	pc90 ^c
Colombia	66.25 %	3.2	0.3
Ecuador	93.87 %	9.05	4.05
Peru	98.91 %	15.43	8.33
World	97.07 %	18.78	5.96
Average	89.02	11.62	4.66
Standard deviation	13.27	5.99	2.94

^a Projected population coverage.^b Average number of epitope hits/HLA combinations recognized by the population.^c Minimum number of epitope hits/HLA combinations recognized by 90 % the population.

The three-dimensional structure of vaccine constructs I and II is shown in Fig. 6.

3.11. Molecular dynamics simulation

The energy minimization of vaccine constructs I and II was achieved in 2291 and 1844 steps, respectively. During the equilibration of the structure of the vaccine constructs I and II, the following values were generated (Table 8):

The RMSD of vaccine constructs I and II is shown in Fig. 7. The vaccine construct II was selected for a longer molecular dynamics simulation of 100 ns and further analysis (Fig. 8).

3.12. Molecular docking with TLR4 protein

HADDOCK clustered 24 structures into 5 clusters, representing 12 % of the water-refined models generated. The top-ranked cluster with the lowest Z-score is the most significant for docking analysis. A representative model from this cluster was further refined using the HADDOCK Refinement Interface, which grouped 20 structures into a single cluster, representing 100 % of the water-refined models. The refined cluster statistics are shown in Table 9. A total of 30 residues from vaccine

construct II interacted with 34 residues from TLR4 (Fig. 9A). The docking complex exhibited 20 hydrogen bonds, as detailed in Table 10. The hydrogen bonds involved the following amino acids: [Chain A (TLR4)- Chain B (vaccine construct II); 42-7, 58–60, 60-58, 60-60, 62-60, 87–92, 87–96, 135-92, 135-92, 135-93, 135-93, 264-203, 264-203, 265-139, 266-139, 339-199, 529-284, 578-287, 603-286, 606-285] (Fig. 9B).

3.13. Immune simulation of constructed vaccine

The predicted immune response profile of vaccine construct II is shown in Fig. 10. The primary response was characterized by high levels of IgM and IgG, along with a decrease in antigenic concentration, which corresponds to an increase in the immunoglobulin expression (Fig. 10A). Additionally, there was an elevated production of B-cells and Th-cells in response to the vaccine construct II (Fig. 10B and C). Furthermore, a high production of IFN- γ was observed (Fig. 10D).

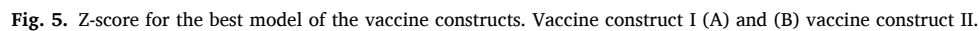
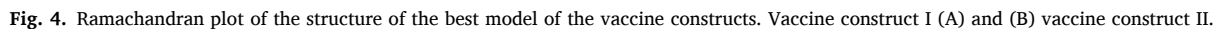
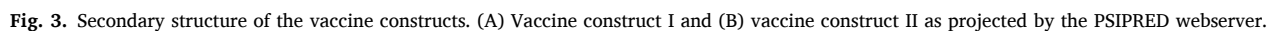
3.14. In silico codon optimization and cloning of constructed vaccine

The reverse translation and codon optimization for vaccine construct II were performed using the JCat server to facilitate the expression in *E. coli*. The average GC content was predicted to be 54 %, and the Codon Optimization Index (CAI) value was 0.98. The restriction sites were identified, and vaccine construct II was successfully cloned as represented in Fig. 11.

4. Discussion

In this study, an immunoinformatics analysis of the *B. bacilliformis* proteome was performed to select the best epitopes for designing a vaccine candidate. This approach not only addresses problems of cost, time and safety compared to traditional methods but also facilitates the creation of multi-epitope vaccines [43]. Recently, numerous researchers have designed novel multi-epitope vaccines using immunoinformatics tools against bacterias [44,45], viruses [46] and parasites [47].

B. bacilliformis is a pathogenic intraerythrocytic bacterium responsible for the neglected Carrion's disease. The erythrocyte invasion process first involved the adhesion to the erythrocyte surface mediated by flagella. The importance of the role of flagella has been previously demonstrated since blocking flagella using antibodies decreases the association with erythrocytes [48]. After adhesion, *B. bacilliformis* enters the erythrocyte through different bacterial factors such as the protein ialB. It has been shown that *E. coli* acquires the ability to invade erythrocytes when transformed with ialB from *B. bacilliformis* [49]. The entry to the erythrocyte is followed by replication and persistence of Bartonella mediated by other virulence factors. Therefore, given the importance of virulence-associated proteins and outer membrane proteins for pathogenesis, these have been used as targets for multi-epitope vaccine development [50]. Our approach incorporated both T and B-cell epitopes predictions because humoral immune response is as important as cellular immune response to improve the vaccine's efficacy. For T-cell epitope prediction, we considered both the most common alleles of the world population (group A) and those prevalent in endemic countries (group B) [51,52]. From the pool of predicted epitopes, experimental and nested epitopes were selected. While many studies have focused solely on overlapping CD4⁺ and CD8⁺ epitopes for multi-epitope



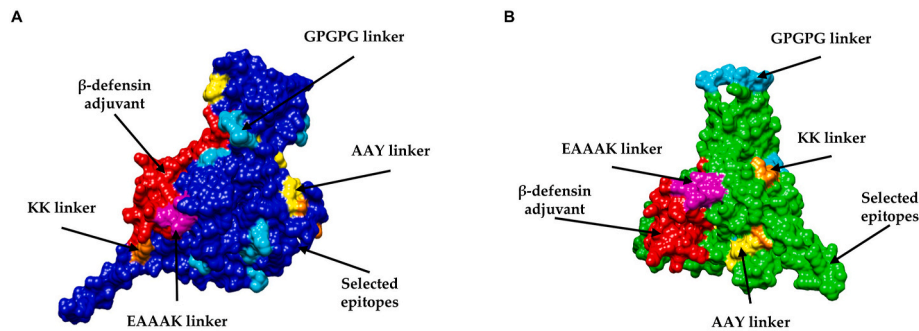


Fig. 6. Three-dimensional structure of the vaccine constructs. Vaccine construct I (A) and (B) vaccine construct II.

Table 8
Values generated during equilibration phases. The average and drift values are detailed.

Vaccine construct	Vaccine construct I		Vaccine construct II	
System evaluation	Average	Drift	Average	Drift
Potential energy (KJ/mol)	-3.08522e+06	-280343	-1.96767e+06	-197577
Temperature (K)	299.77	1.23254	299.834	1.2522
Pressure (bar)	0.612972	2.05587	1.47076	-6.84067
Density (kg/m ³)	1006.35	0.277002	1012.03	0.275575

vaccine design [53], our approach incorporates linear CD4⁺ T-cell epitopes containing multiple linear CD8⁺ T-cell epitopes within the same sequence. This strategy has been previously used and offers the advantage of dual specificity (binding to both MHC-I and MHC-II), while also reducing the overall length of the vaccine construct, thereby enhancing its immunogenic potential [16,54].

The primary selection criterion for nested T cell epitopes was high promiscuity, as epitopes that bind to many alleles increase the population coverage of the vaccine [55]. Additionally, T and B cell epitopes that did not have homology with the human and murine proteomes were chosen to avoid potential autoimmune responses in humans and in future clinical trials in mice. The epitopes were also evaluated for allergenicity and toxicity to reduce potential adverse effects. The final criterion for selection was conservation among all strains of *B. bacilliformis*. Ideally, only epitopes with 100 % conservation across all strains were selected. However, for outer membrane protein epitopes, those with at least 94.12 % conservation (100 % conservation among 15 out of 16 strains) were also considered because amino acid changes in some peptides were observed with respect to the strain Ver097, which is considered a possible subspecies of *B. bacilliformis* [56]. These criteria have been applied previously to select the best epitopes for constructing a multi-epitope vaccine against *B. bacilliformis* [10,11]. Nonetheless, unlike previous studies, our approach considered non-homology with

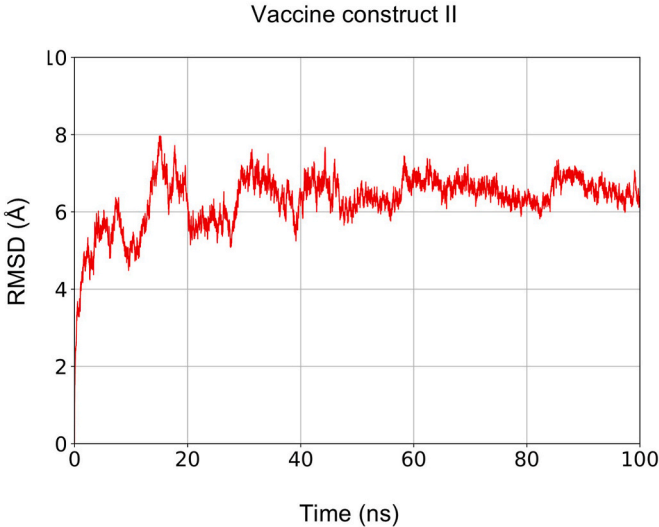


Fig. 8. RMSD plot of the alpha carbon atoms for the 100 ns molecular dynamics simulation of vaccine construct II.

Table 9
Protein–protein docking results between vaccine construct II and the TLR4.

Parameters	Values
HADDOCK score	-224.8 ± 4.9
Cluster size	20
RMSD from the overall lowest-energy structure	0.6 ± 0.4
Van der Waals energy	-99.5 ± 3.0
Electrostatic energy	-486.3 ± 21.3
Solvation energy	-28.1 ± 2.5
Restraint's violation of energy	0.0 ± 0.00
Buried Surface Area	3583.3 ± 55.7
Z-Score	0.0

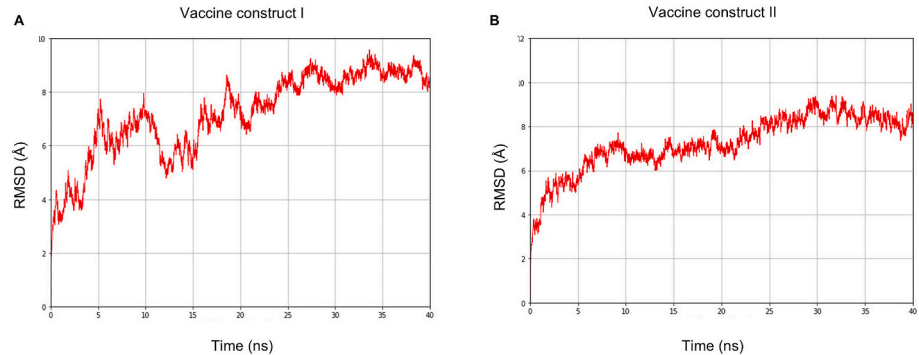


Fig. 7. RMSD plot of the alpha carbon atoms for the 40 ns molecular dynamics simulation. (A) Vaccine construct I and (B) Vaccine construct II.

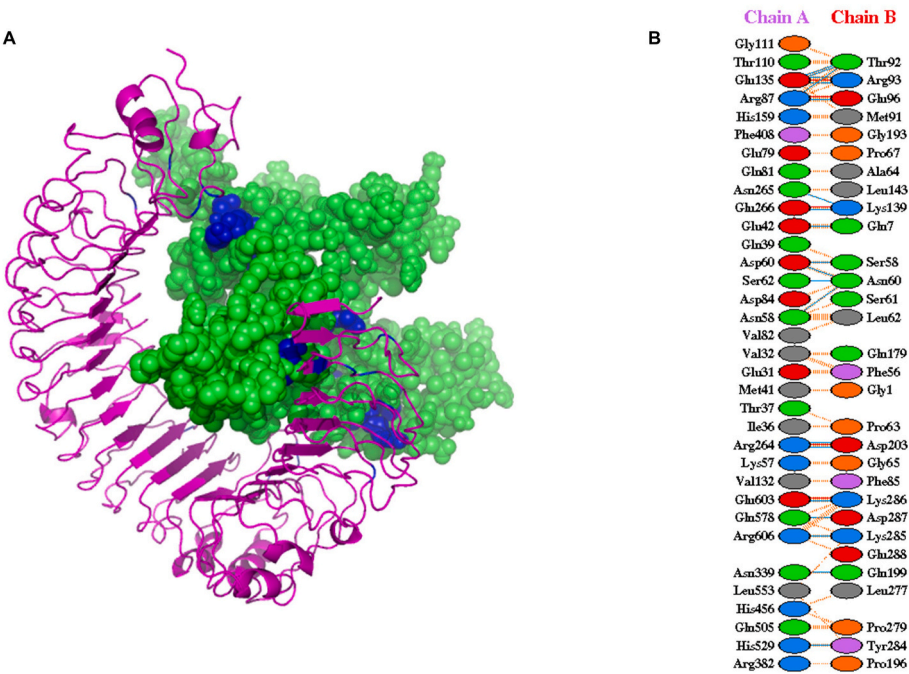


Fig. 9. Docking complex of the vaccine construct II and the TLR4. (A) Vaccine construct II (green) and TLR4 (purple). The residues that mediate interaction between the two proteins are highlighted on blue. (B) Detailed residues involved in the vaccine construct II-TLR4 complex.

Table 10
Interactions between vaccine construct II and the TLR4.

Hydrogen Bond	Salt Bridges	Non-bonded contacts
20	5	144

the murine proteome and high epitope promiscuity. While none of our selected epitopes have yet been reported; most of the epitopes from virulence-associated proteins are related to the flagellar apparatus, aligning with studies proposing flagella proteins as potential vaccine

candidates [11,57]. Moreover, Dichter et al., reported an epitope corresponding to the outer membrane autotransporter protein (A1UTU0) an experimentally validated immunodominant antigen [58].

Two vaccine constructs were designed from the final epitopes concatenating them using linkers, and human β -defensin at the C-terminal end as an adjuvant. This TLR agonist has previously demonstrated antimicrobial and immunomodulatory activity [59]. Analysis of both vaccine constructs revealed their safety (non-allergenicity and non-toxicity) and potential effectiveness (antigenicity). Physicochemical properties indicated that both constructs are stable, hydrophilic,

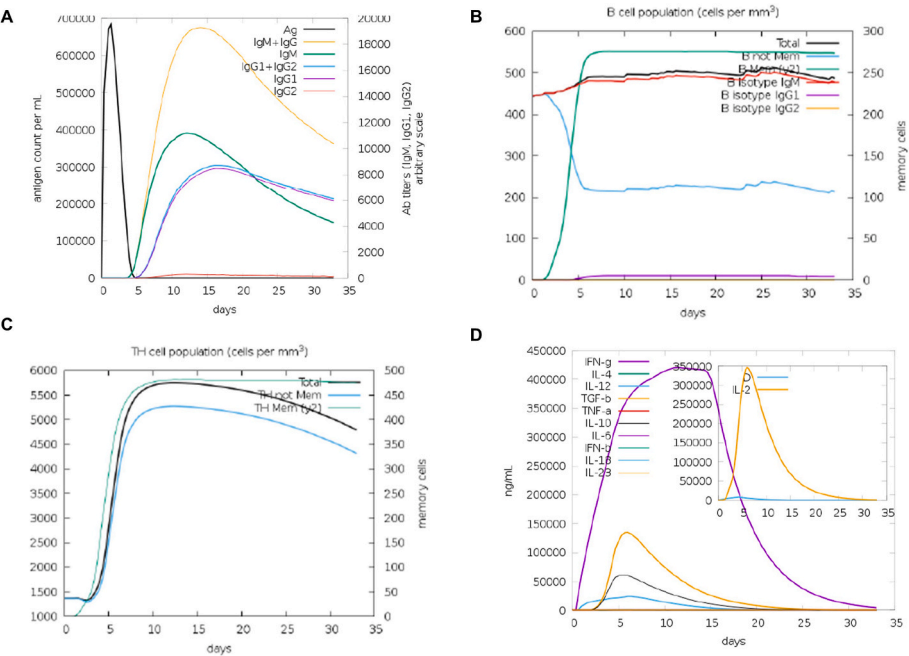


Fig. 10. Immune profile of the vaccine construct II. (A) Antigen count and antibody titer with Ig subclass. (B) B-cell population. (C) Th-cell population. (D) Cytokines and interleukins population.

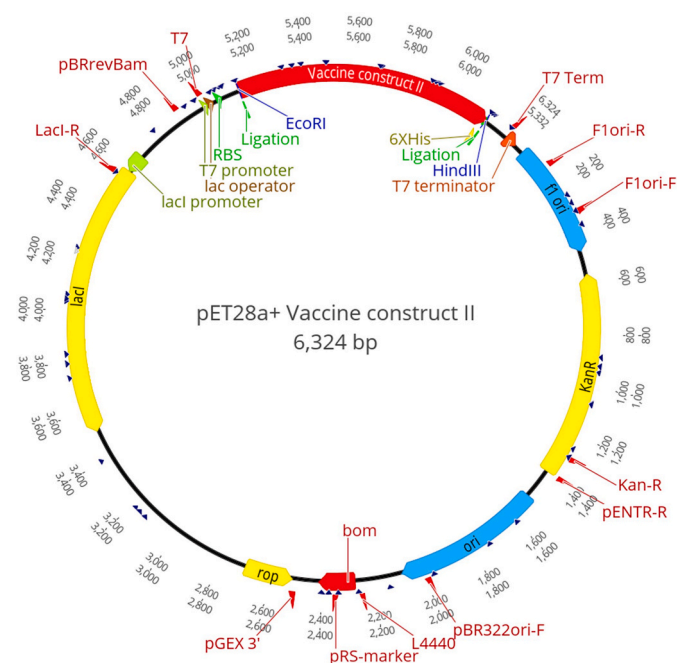


Fig. 11. Codon optimization and in-silico cloning of the vaccine construct II. *In silico* restriction cloning of the vaccine construct II sequence into the pET28a (+) expression vector. Vaccine construct II gene coding (red) and vector backbone (black).

soluble, and basic. Particularly their stability and solubility in physiological simulated conditions support its potential for effective expression and immunogenicity. Similar results have been reported in a study of multi-epitope vaccine against *Klebsiella pneumoniae* which demonstrated robust solubility and thermostability characteristics [60]. Additionally, they do not show transmembrane helices or signal peptides, which are undesirable characteristics in a vaccine [61]. The generated 3D models were further refined and evaluated using Ramachandran plots, showing improved quality of the 3D structures post-refinement. Validation of the best models for both constructs yielded high calculated scores, indicating that both constructs fall within the range of structures resolved by X-ray crystallography and nuclear magnetic resonance (NMR). The 40 ns molecular dynamics simulation results (Fig. 7) indicated stable behavior for vaccine construct II, and the 100 ns simulation further confirmed this stability (Fig. 8), leading to its selection for molecular docking with TLR4. Although these in silico results was promising, an extended simulation length in futures studies may provide additional insights into long-term stability of the vaccine construct II.

The docking analysis (Fig. 9) revealed salt bridges, non-bonded interactions, and hydrogen bonds. The distance of the hydrogen bonds between the vaccine and TLR4 ranged from 2 to 3 Å, suggesting a strong interaction between the designed vaccine construct and TLR4, which plays a crucial role in innate immune activation [12]. These findings imply that the vaccine construct II may effectively trigger downstream signaling pathways essential for a strong immune response.

The immune simulation revealed that vaccine construct II has significant potential to elicit robust immune responses. The results indicated the induction of differentiation of various immune cells, including antibody-producing and memory B-cells as well as various T-cell subsets, alongside elevated cytokines production. In addition, to enhance the precision of immune system simulations in multi-epitope vaccine development, other advanced mathematical models can be used [62]. These findings align with the expected immune response a vaccine would generate in the human immune system.

Cloning vaccine construct II using conventional techniques appeared feasible. The calculated Codon Adaptation Index (CAI) was nearly 1.0,

and the GC content was within the ideal range (30–70 %), suggesting high expression levels in the *E. coli* K12 system [63]. In silico cloning was performed using the pET28a(+) vector as a template for future vaccine synthesis. This plasmid includes a His-tag as a fusion partner, facilitating easier future purification.

Future research will focus on the experimental validation of the designed multi-epitope vaccine, including cloning, expression, and in vitro assays to assess its immunogenic potential. In vivo studies will be crucial to evaluate the immune response, including the production of specific antibodies. These efforts will help establish a solid foundation for further preclinical development and testing of a potential vaccine against *B. bacilliformis*.

5. Conclusions

Designing an effective and affordable vaccine against the neglected Carrion's disease is crucial to addressing a public health issue in Peru, Ecuador and Colombia. This study applied an immunoinformatics approach to identify the best nested epitopes from outer membrane and virulence-associated proteins with the potential to induce both cellular and humoral immune responses, leading to the design of two multi-epitope vaccine constructs. Most of the selected epitopes are novel and have not been previously predicted. Vaccine construct II demonstrated greater stability during molecular dynamics simulations. A robust interaction network with the immune receptor TLR4 ensured that vaccine construct II could elicit a variety of strong immunological responses, as shown by immune simulation. The in-silico characterization of vaccine construct II strongly suggests that it could be an effective potential vaccine candidate and lays the groundwork for future studies; however, in vitro and in vivo evaluation is necessary.

CCRediT authorship contribution statement

Damaris Rivera-Asencios: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Abraham Espinoza-Culupú:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Sheyla Carmen-Sifuentes:** Software. **Pablo Ramirez:** Writing – review & editing, Writing – original draft. **Ruth García-de-la-Guarda:** Writing – review & editing, Writing – original draft, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Funding

This research was funded by Vicerrectorado de Investigación y Posgrado de la Universidad Nacional Mayor de San Marcos, grant number B20100010a, B21100981 and Dr. Abraham Espinoza-Culupú is the recipient of a postdoctoral fellowship from the CONCYTEC/PROCIENCIA program No. PE501084176-2023-PROCIENCIA.

Declaration of competing interest

We, the authors of this article, titled, “Design of a Multi-Epitope Vaccine candidate against Carrion disease by immunoinformatics approach.” by Damaris Rivera-Asencios et al. We hereby declare that no artificial intelligence (AI) tools or methods were used in the creation, research, writing, or analysis presented in this manuscript. All findings, interpretations and conclusions are the result of the effort and experience of our research group.

We affirm that the content herein reflects original research conducted by the authors and that all sources have been duly acknowledged. No part of this work has been generated, manipulated or influenced by artificial intelligence technologies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compbmed.2024.109397>.

References

- [1] C. Henriquez-Camacho, P. Ventosilla, M.F. Minnick, J. Ruiz, C. Maguina, Proteins of *Bartonella bacilliformis*: candidates for vaccine development, *Int. J. Pept.* 2015 (2015) 702784, <https://doi.org/10.1155/2015/702784>.
- [2] V. Zorrilla, G. Vasquez, L. Espada, P. Ramirez, [Update on tegumentary leishmaniasis and carrion's disease vectors in Peru], *Rev. Peru. Med. Exp. Salud Pública* 34 (2017) 485–496, <https://doi.org/10.17843/rpmesp.2017.343.2398>.
- [3] N. Sanchez Clemente, C.A. Ugarte-Gil, N. Solorzano, C. Maguina, P. Pachas, D. Blazes, R. Bailey, D. Mabey, D. Moore, *Bartonella bacilliformis*: a systematic review of the literature to guide the research agenda for elimination, *PLoS Neglected Trop. Dis.* 6 (2012) e1819, <https://doi.org/10.1371/journal.pntd.0001819>.
- [4] B. Alexander, A review of bartonellosis in Ecuador and Colombia, *Am. J. Trop. Med. Hyg.* 52 (1995) 354–359, <https://doi.org/10.4269/ajtmh.1995.52.354>.
- [5] M.F. Minnick, B.E. Anderson, A. Lima, J.M. Battisti, P.G. Lawyer, R.J. Birtles, Oroya fever and verruga peruana: bartonelloses unique to South America, *PLoS Neglected Trop. Dis.* 8 (2014) e2919, <https://doi.org/10.1371/journal.pntd.0002919>.
- [6] M. Garcia-Quintanilla, A.A. Dichter, H. Guerra, V.A.J. Kempf, Carrion's disease: more than a neglected disease, *Parasites Vectors* 12 (2019) 141, <https://doi.org/10.1186/s13071-019-3390-2>.
- [7] H. Deng, Q. Pang, B. Zhao, M. Vayssier-Taussat, Molecular mechanisms of *Bartonella* and mammalian erythrocyte interactions: a review, *Front. Cell. Infect. Microbiol.* 8 (2018) 431, <https://doi.org/10.3389/fcimb.2018.00431>.
- [8] C. Gomes, M.J. Pons, J. Del Valle Mendoza, J. Ruiz, Carrion's disease: an eradicable illness? *Infect. Dis. Poverty* 5 (2016) 105, <https://doi.org/10.1186/s40249-016-0197-7>.
- [9] L.J. del Valle, L. Flores, M. Vargas, R. Garcia-de-la-Guarda, R.L. Quispe, Z. B. Ibanez, D. Alvarado, P. Ramirez, J. Ruiz, *Bartonella bacilliformis*, endemic pathogen of the Andean region, is intrinsically resistant to quinolones, *Int. J. Infect. Dis.* 14 (2010) e506–e510, <https://doi.org/10.1016/j.ijid.2009.07.025>.
- [10] C.P.P. Rojas, P.N.L. Pari, L.S. Solis, C.O. Mogollon, L.I. Angulo, H.B. Calderon, A. M. Nique, J. Morales, G.E.V. Egusquiza, [Design and evaluation of a multi-epitope protein as a candidate for a carrion disease vaccine], *Rev. Peru. Med. Exp. Salud Pública* 36 (2019) 414–422, <https://doi.org/10.17843/rpmesp.2019.363.4430>.
- [11] H. Gul, S.S. Ali, S. Saleem, S. Khan, J. Khan, A. Wadood, A.U. Rehman, Z. Ullah, S. Ali, H. Khan, et al., Subtractive proteomics and immunoinformatics approaches to explore *Bartonella bacilliformis* proteome (virulence factors) to design B and T cell multi-epitope subunit vaccine, *Infect. Genet. Evol.* 85 (2020) 104551, <https://doi.org/10.1016/j.meegid.2020.104551>.
- [12] T. Kawai, S. Akira, The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors, *Nat. Immunol.* 11 (2010) 373–384, <https://doi.org/10.1038/ni.1863>.
- [13] N. Samavia, P. Fahed, W. Yasir, A. Tasneem, N. Syeda, Prediction of promiscuous epitopes in ORF2 of Hepatitis E virus: an In-Silico approach, *Afr. Health Sci.* 22 (2022) 626–639, <https://doi.org/10.4314/ahs.v22i3.67>.
- [14] V. Jurtz, S. Paul, M. Andreatta, P. Marcatili, B. Peters, M. Nielsen, NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data, *J. Immunol.* 199 (2017) 3360–3368, <https://doi.org/10.4049/jimmunol.1700893>.
- [15] B. Reynisson, C. Barra, S. Kaabinejad, W.H. Hildebrand, B. Peters, M. Nielsen, Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data, *J. Proteome Res.* 19 (2020) 2304–2315, <https://doi.org/10.1021/acs.jproteome.9b00874>.
- [16] L.F. Soto, A.C. Romani, G. Jimenez-Avalos, Y. Silva, C.M. Ordinola-Ramirez, R. M. Lopez Lapa, D. Requena, Immunoinformatic analysis of the whole proteome for vaccine design: an application to *Clostridium perfringens*, *Front. Immunol.* 13 (2022) 942907, <https://doi.org/10.3389/fimmu.2022.942907>.
- [17] N.Y. Yu, J.R. Wagner, M.R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S.C. Sahinalp, M. Ester, L.J. Foster, et al., PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, *Bioinformatics* 26 (2010) 1608–1615, <https://doi.org/10.1093/bioinformatics/btq249>.
- [18] B. Liu, D. Zheng, Q. Jin, L. Chen, J. Yang, VfdB 2019: a comparative pathogenomic platform with an interactive web interface, *Nucleic Acids Res.* 47 (2019) D687–D692, <https://doi.org/10.1093/nar/gky1080>.
- [19] A. Thibau, A.A. Dichter, D.J. Vaca, D. Linke, A. Goldman, V.A.J. Kempf, Immunogenicity of trimeric autotransporter adhesins and their potential as vaccine targets, *Med. Microbiol. Immunol.* 209 (2020) 243–263, <https://doi.org/10.1007/s00430-019-00649-y>.
- [20] M.C. Jespersen, B. Peters, M. Nielsen, P. Marcatili, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, *Nucleic Acids Res.* 45 (2017) W24–W29, <https://doi.org/10.1093/nar/gkx346>.
- [21] I. Dimitrov, I. Bangov, D.R. Flower, I. Doytchinova, AllerTOP v.2—a server for in silico prediction of allergens, *J. Mol. Model.* 20 (2014) 2278, <https://doi.org/10.1007/s00894-014-2278-5>.
- [22] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, C. Open Source Drug Discovery, G.P. Raghava, In silico approach for predicting toxicity of peptides and proteins, *PLoS One* 8 (2013) e73957, <https://doi.org/10.1371/journal.pone.0073957>.
- [23] H.H. Bui, J. Sidney, W. Li, N. Fusseder, A. Sette, Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines, *BMC Bioinf.* 8 (2007) 361, <https://doi.org/10.1186/1471-2105-8-361>.
- [24] X. Pan, J. Zuaalart, X. Wang, H.B. Shen, E.P. Campos, D.O. Marushchak, W. De Neve, ToxDL: deep learning using primary structure and domain embeddings for assessing protein toxicity, *Bioinformatics* 36 (2021) 5159–5168, <https://doi.org/10.1093/bioinformatics/btaa656>.
- [25] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, *BMC Bioinf.* 8 (2007) 4, <https://doi.org/10.1186/1471-2105-8-4>.
- [26] M.R. Wilkins, E. Gasteiger, A. Bairoch, J.C. Sanchez, K.L. Williams, R.D. Appel, D. F. Hochstrasser, Protein identification and analysis tools in the ExPASy server, *Methods Mol. Biol.* 112 (1999) 531–552, <https://doi.org/10.1385/1-59259-584-7:531>.
- [27] M. Hebditch, M.A. Carballo-Amador, S. Charonis, R. Curtis, J. Warwicker, Protein-Sol: a web tool for predicting protein solubility from sequence, *Bioinformatics* 33 (2017) 3098–3100, <https://doi.org/10.1093/bioinformatics/btx345>.
- [28] J.J. Almagro Armenteros, K.D. Tsirigos, C.K. Sonderby, T.N. Petersen, O. Winther, S. Brunak, G. von Heijne, H. Nielsen, SignalP 5.0 improves signal peptide predictions using deep neural networks, *Nat. Biotechnol.* 37 (2019) 420–423, <https://doi.org/10.1038/s41587-019-0036-z>.
- [29] A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, *J. Mol. Biol.* 305 (2001) 567–580, <https://doi.org/10.1006/jmbi.2000.4315>.
- [30] H.H. Bui, J. Sidney, K. Dinh, S. Southwood, M.J. Newman, A. Sette, Predicting population coverage of T-cell epitope-based diagnostics and vaccines, *BMC Bioinf.* 7 (2006) 153, <https://doi.org/10.1186/1471-2105-7-153>.
- [31] D.T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, *J. Mol. Biol.* 292 (1999) 195–202, <https://doi.org/10.1006/jmbi.1999.3091>.
- [32] S. Raman, R. Vernon, J. Thompson, M. Tyka, R. Sadreyev, J. Pei, D. Kim, E. Kellogg, F. DiMaio, O. Lange, et al., Structure prediction for CASP8 with all-atom refinement using Rosetta, *Proteins* 77 (Suppl 9) (2009) 89–99, <https://doi.org/10.1002/prot.22540>.
- [33] D. Xu, Y. Zhang, Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization, *Biophys. J.* 101 (2011) 2525–2534, <https://doi.org/10.1016/j.bpj.2011.10.024>.
- [34] C. Colovos, T.O. Yeates, Verification of protein structures: patterns of nonbonded atomic interactions, *Protein Sci.* 2 (1993) 1511–1519, <https://doi.org/10.1002/pro.5560020916>.
- [35] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Res.* 35 (2007) W407–W410, <https://doi.org/10.1093/nar/gkm290>.
- [36] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* 26 (1993) 283–291, <https://doi.org/10.1107/S0021889892009944>.
- [37] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1 (2015) 19–25, <https://doi.org/10.1016/j.softx.2015.06.001>.
- [38] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, D. E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field, *Proteins* 78 (2010) 1950–1958, <https://doi.org/10.1002/prot.22711>.
- [39] R.V. Honarato, M.E. Trellet, B. Jimenez-Garcia, J.J. Schaarschmidt, M. Giulini, V. Reys, P.I. Koukos, J. Rodrigues, E. Karaca, G.C.P. van Zundert, et al., The HADDOCK2.4 web server for integrative modeling of biomolecular complexes, *Nat. Protoc.* (2024), <https://doi.org/10.1038/s41596-024-01011-0>.
- [40] R.A. Laskowski, J. Jablonska, L. Pravda, R.S. Varkova, J.M. Thornton, PDBsum: structural summaries of PDB entries, *Protein Sci.* 27 (2018) 129–134, <https://doi.org/10.1002/pro.3289>.
- [41] N. Rapin, O. Lund, M. Bernaschi, F. Castiglione, Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system, *PLoS One* 5 (2010) e9862, <https://doi.org/10.1371/journal.pone.0009862>.
- [42] A. Grote, K. Hiller, M. Scheer, R. Munch, B. Nortemann, D.C. Hempel, D. Jahn, JCat: a novel tool to adapt codon usage of a target gene to its potential expression host, *Nucleic Acids Res.* 33 (2005) W526–W531, <https://doi.org/10.1093/nar/gki376>.
- [43] S. Parvizpour, M.M. Pourseif, J. Razmara, M.A. Rafi, Y. Omid, Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches, *Drug Discov. Today* 25 (2020) 1034–1042, <https://doi.org/10.1016/j.drudis.2020.03.006>.
- [44] M. Shafaghi, Z. Bahadori, H. Madanchi, M.M. Ranjbar, A.A. Shabani, S.F. Mousavi, Immunoinformatics-aided design of a new multi-epitope vaccine adjuvanted with domain 4 of pneumolysin against *Streptococcus pneumoniae* strains, *BMC Bioinf.* 24 (2023) 67, <https://doi.org/10.1186/s12859-023-05175-6>.
- [45] M. Tahir Ul Qamar, S. Ahmad, I. Fatima, F. Ahmad, F. Shahid, A. Naz, S.W. Abbasi, A. Khan, M.U. Mirza, U.A. Ashfaq, et al., Designing multi-epitope vaccine against *Staphylococcus aureus* by employing subtractive proteomics, reverse vaccinology and immuno-informatics approaches, *Comput. Biol. Med.* 132 (2021) 104389, <https://doi.org/10.1016/j.compbmed.2021.104389>.
- [46] A.J. Obaidullah, M.M. Alanazi, N.A. Alsaif, H. Albassam, A.A. Almezahia, A. M. Alqahtani, S. Mahmud, S.A. Sami, T.B. Emran, Immunoinformatics-guided design of a multi-epitope vaccine based on the structural proteins of severe acute

- respiratory syndrome coronavirus 2, RSC Adv. 11 (2021) 18103–18121, <https://doi.org/10.1039/d1ra02885e>.
- [47] S. Naz, A. Aroosh, A. Caner, E.A. Sahar, S. Toz, Y. Ozbel, S.W. Abbasi, Immunoinformatics approach to design a multi-epitope vaccine against cutaneous leishmaniasis, *Vaccines (Basel)* 11 (2023), <https://doi.org/10.3390/vaccines11020339>.
- [48] D.C. Scherer, I. DeBuron-Connors, M.F. Minnick, Characterization of Bartonella bacilliformis flagella and effect of anti-flagellin antibodies on invasion of human erythrocytes, *Infect. Immun.* 61 (1993) 4962–4971, <https://doi.org/10.1128/iai.61.12.4962-4971.1993>.
- [49] S.J. Mitchell, M.F. Minnick, Characterization of a two-gene locus from Bartonella bacilliformis associated with the ability to invade human erythrocytes, *Infect. Immun.* 63 (1995) 1552–1562, <https://doi.org/10.1128/iai.63.4.1552-1562.1995>.
- [50] H. Arya, Chapter 7 - key steps in the selection of vaccine targets, in: T.K. Bhatt, S. Nimesh (Eds.), *The Design & Development of Novel Drugs and Vaccines*, Academic Press, 2021, pp. 93–95.
- [51] M. Wang, M.H. Claesson, Classification of human leukocyte antigen (HLA) supertypes, *Methods Mol. Biol.* 1184 (2014) 309–317, https://doi.org/10.1007/978-1-4939-1115-8_17.
- [52] S. Paul, C.S. Lindestam Arlehamn, T.J. Scriba, M.B. Dillon, C. Oseroff, D. Hinz, D. M. McKinney, S. Carrasco Pro, J. Sidney, B. Peters, et al., Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes, *J. Immunol. Methods* 422 (2015) 28–34, <https://doi.org/10.1016/j.jim.2015.03.022>.
- [53] S.H. Apte, P.L. Groves, M. Skwarczynski, Y. Fujita, C. Chang, I. Toth, D.L. Doolan, Vaccination with lipid core peptides fails to induce epitope-specific T cell responses but confers non-specific protective immunity in a malaria model, *PLoS One* 7 (2012) e40928, <https://doi.org/10.1371/journal.pone.0040928>.
- [54] M. Hakimian, A. Doosti, A. Sharifzadeh, A novel chimeric vaccine containing multiple epitopes for simulating robust immune activation against Klebsiella pneumoniae, *BMC Immunol.* 25 (2024) 27, <https://doi.org/10.1186/s12865-024-00617-z>.
- [55] N. Nezafat, Z. Karimi, M. Eslami, M. Mohkam, S. Zandian, Y. Ghasemi, Designing an efficient multi-epitope peptide vaccine against Vibrio cholerae via combined immunoinformatics and protein interaction based approaches, *Comput. Biol. Chem.* 62 (2016) 82–95, <https://doi.org/10.1016/j.compbiolchem.2016.04.006>.
- [56] S. Paul, M.F. Minnick, S. Chattopadhyay, Mutation-driven divergence and convergence indicate adaptive evolution of the intracellular human-restricted pathogen, Bartonella bacilliformis, *PLoS Neglected Trop. Dis.* 10 (2016) e0004712, <https://doi.org/10.1371/journal.pntd.0004712>.
- [57] M.T. Khan, A. Mahmud, M.A. Iqbal, M. Hasan, Subtractive genomic analysis for identification of novel drug targets and vaccine candidates against Bartonella bacilliformis subsp. Ver097, *bioRxiv* (2019) 731570, <https://doi.org/10.1101/731570>.
- [58] A.A. Dichter, T.G. Schultze, A. Wenigmann, W. Ballhorn, A. Latz, E. Schlufte, P. Ventosilla, H. Guerra Allison, C. Ugarte-Gil, P. Tsukayama, et al., Identification of immunodominant Bartonella bacilliformis proteins: a combined in-silico and serology approach, *Lancet Microbe* 2 (2021) e685–e694, [https://doi.org/10.1016/S2666-5247\(21\)00184-1](https://doi.org/10.1016/S2666-5247(21)00184-1).
- [59] D.M. Hoover, Z. Wu, K. Tucker, W. Lu, J. Lubkowski, Antimicrobial characterization of human beta-defensin 3 derivatives, *Antimicrob. Agents Chemother.* 47 (2003) 2804–2809, <https://doi.org/10.1128/AAC.47.9.2804-2809.2003>.
- [60] N. Cuscino, A. Fatima, V. Di Pilato, M. Bulati, C. Alfano, E. Monaca, G. Di Mento, D. Di Carlo, F. Cardinale, F. Monaco, et al., Computational design and characterization of a multi-epitope vaccine against carbapenemase-producing Klebsiella pneumoniae strains, derived from antigens identified through reverse vaccinology, *Comput. Struct. Biotechnol. J.* 20 (2022) 4446–4463, <https://doi.org/10.1016/j.csbj.2022.08.035>.
- [61] T. Kar, U. Narsaria, S. Basak, D. Deb, F. Castiglione, D.M. Mueller, A.P. Srivastava, A candidate multi-epitope vaccine against SARS-CoV-2, *Sci. Rep.* 10 (2020) 10895, <https://doi.org/10.1038/s41598-020-67749-1>.
- [62] B. Maayah, O.A. Arqub, S. Alnabulsi, H. Alsulami, Numerical solutions and geometric attractors of a fractional model of the cancer-immune based on the Atangana-Baleanu-Caputo derivative and the reproducing kernel scheme, *Chin. J. Phys.* 80 (2022) 463–483, <https://doi.org/10.1016/j.cjph.2022.10.002>.
- [63] P. Hengen, Purification of His-Tag fusion proteins from Escherichia coli, *Trends Biochem. Sci.* 20 (1995) 285–286, [https://doi.org/10.1016/s0968-0004\(00\)89045-3](https://doi.org/10.1016/s0968-0004(00)89045-3).