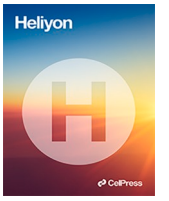
Heliyon 10 (2024) e23616

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Reverse vaccinology and immunoinformatics approach to design a chimeric epitope vaccine against *Orientia tsutsugamushi*

Anutee Dolley [a], Himanshu Ballav Goswami [a], Dikshita Dowerah [b], Upalabdha Dey [a], Aditya Kumar [a], Vanlal Hmuaka [c], Rupak Mukhopadhyay [a], Debasree Kundu [d], George M. Varghese [d], Robin Doley [a], Ramesh Chandra Deka [b], Nima D. Namsa [a,*]

[a] *Department of Molecular Biology and Biotechnology, Tezpur University, Napaam, 784028, Assam, India*
[b] *Department of Chemical Sciences, Tezpur University, Napaam, 784028, Assam, India*
[c] *Entomology and Biothreat Management Division, Defence Research Laboratory, Tezpur, 784001, Assam, India*
[d] *Department of Infectious Diseases, Christian Medical College, Vellore, 632002, Tamil Nadu, India*

ARTICLE INFO

Keywords:
Scrub typhus
O. tsutsugamushi
Immunoinformatics
TSA56
ScaA
Multi-epitope antigen

ABSTRACT

Scrub typhus is a vector-borne infectious disease caused by *Orientia tsutsugamushi* and it is reportedly associated with up to 20 % of hospitalized cases of febrile illnesses. The major challenge of vaccine development is the lack of identified antigens that can induce both heterotypic and homotypic immunity including the production of antibodies, cytotoxic T lymphocyte, and helper T lymphocytes. We employed a comprehensive immunoinformatic prediction algorithm to identify immunogenic epitopes of the 56-kDa type-specific cell membrane surface antigen and surface cell antigen A of *O. tsutsugamushi* to select potential candidates for developing vaccines and diagnostic assays. We identified 35 linear and 29 continuous immunogenic B-cell epitopes and 51 and 27 strong-binding T-cell epitopes of major histocompatibility complex class I and class II molecules, respectively, in the conserved and variable regions of the 56-kDa type-specific surface antigen. The predicted B- and T-cell epitopes were used to develop immunogenic multi-epitope candidate vaccines and showed to elicit a broad-range of immune protection. A stable interactions between the multi-epitope vaccines and the host fibronectin protein were observed using docking and simulation methods. Molecular dynamics simulation studies demonstrated that the multi-epitope vaccine constructs and fibronectin docked models were stable during simulation time. Furthermore, the multi-epitope vaccine exhibited properties such as antigenicity, non-allergenicity and ability to induce interferon gamma production and had strong associations with their respective human leukocyte antigen alleles of world-wide population coverage. A correlation of immune simulations and the in-silico predicted immunogenic potential of multi-epitope vaccines implicate for further investigations to accelerate designing of epitope-based vaccine candidates and chimeric antigens for development of serological diagnostic assays for scrub typhus.

* Corresponding author.
E-mail address: namsa@tezu.ernet.in (N.D. Namsa).

https://doi.org/10.1016/j.heliyon.2023.e23616
Received 30 March 2023; Received in revised form 6 December 2023; Accepted 7 December 2023
Available online 13 December 2023

1. Introduction

Scrub typhus is a re-emerging infectious febrile illness caused by the obligate intracellular bacterium *Orientia tsutsugamushi,* which is transmitted to humans through the bite of larval trombiculid mites [1]. Though highly endemic in the Asia-Pacific region, scrub typhus has recently re-emerged in unexpected geographical locations such as Chile and Kenya, posing significant public health risks due to high rate of mortality and morbidity [2–4]. In the last few decades, several cases of scrub typhus infections across the different states of India have been reported recently [5–10]. The increase in the number of cases can be attributed to multiple factors, including increased anthropogenic activity resulting in changes in vector and rodent reservoirs and the widespread use of β-lactam antibiotics resulting in development of resistance strains of *O. tsutsugamushi* [11,12]. Currently, diagnosing scrub typhus is challenging since eschar formation is not observed in many patients, and clinical symptoms are non-specific and often overlaps with other common febrile illnesses such as dengue, malaria, typhoid, and leptospirosis [13,14]. However, untreated patients may present with severe complications, such as multi-organ failure, with mortality rates of 35–50 %, suggesting proper diagnosis is important for subsequent management of scrub typhus afflicted individuals [15,16]. Diagnostic methods such as immunofluorescence assays, enzyme-linked immunosorbent assay (ELISA), Weil-Felix test, and polymerase chain reaction have certain limitations due to precision accuracy and the requirement of skilled laboratory technician [17]. Hence, a field deployable cost-effective point-of-care test (POCT) for rapid detection of antibodies specific to antigens of *O. tsutsugamushi* in a limited resource settings and endemic regions is required to ascertain the actual prevalence and disease burden of scrub typhus.

The accuracy of diagnostics and the efficiency of vaccine candidates are heavily influenced by antigenic diversity among circulating strains of the causal agent. Of over 20 antigenically distinct *O. tsutsugamushi* strains reported, Karp, Kato, Gilliam, and TA716 are considered prototype [18]. The *O. tsutsugamushi* genome encodes a 56-kDa type-specific antigen (TSA56) that is located on its outer membrane and comprises approximately 20 % of the total bacterial proteome [13]. Interaction of TSA56 with host fibronectin receptors mediates host cell invasion [19,20]. The TSA56 protein has 500–540 amino acids, four variable domains ascribed to antigenic diversity, and conserved domains [21]. TSA56 is considered a major target for diagnostic assay and vaccine development since it is highly immunogenic and reacts to the convalescent serum of scrub typhus-positive patients [22,23]. The weak immune protection against the homologous strains and short-lived heterologous immunity is the major challenge of vaccine development for *O. tsutsugamushi*. Moreover, the *O. tsutsugamushi* genome encodes five distinct autotransporter genes: surface cell antigen A (*ScaA*), *ScaB*, *ScaC*, *ScaD*, and *ScaE*. The passenger domain of ScaA, which is relatively conserved across strains, helps in attachment of bacterium to host cells [24]. Further ScaA antibody blocking experiment showed inhibition of *O. tsutsugamushi* infection in HeLa cells and administration of ScaA and TSA56 in mice provided immune protection against lethal challenges with homologous and heterologous strains. Therefore, these findings suggest that TSA56 and ScaA are potential vaccine candidates for scrub typhus [25].

Recent advances in bioinformatics have facilitated in engineering epitope-based serological assay and vaccines for yellow fever, malaria, leptospirosis, and dengue [26–29]. The outcome of immunoinformatic analyses is fast and cost-effective thereby supplementing the requirements of vaccine development using the biological experiments. Although there are earlier reports that have identified TSA56 as vaccine and diagnostic candidates, a major limitation is the use of a mixture of full-length TSA56 antigens from multiple strains of *O. tsutsugamushi* and a lack of long-term immune protection. This study aims to develop a single chimeric antigen using the in-silico determined immunogenic B-cell epitopes of TSA56 and ScaA and a multi-epitope subunit vaccine candidate with the predicted B-cell and T-cell epitopes of prototype *O. tsutsugamushi* strains. Additionally, our study has identified unique B- and T-cell peptides providing a platform for further experimental validations and development of peptide-based serological assay for rapid diagnosis of scrub typhus.

2. Materials and methods

2.1. Determination of antigenic and bio-chemical properties of TSA56 and ScaA proteins

The amino acid (aa) sequences of TSA56 protein of Gilliam (SPR11258.1), Kato (KJV54848.1), Karp (SPR16111.1), TA716 (KJV77538.1), TA763 (KJV75025.1), and ScaA (SPR03916.1) were retrieved from NCBI GenBank. The antigenic properties of TSA56 and ScaA antigens were determined using the VaxiJen v2.0 server and the biochemical attributes was predicted using the ProtParam tool [30,31].

2.2. Prediction of B-cell epitopes

Bcepred determine linear B-cell epitopes based on hydrophilicity, flexibility, accessibility, polarity, exposed surface, and turns [32]. BepiPred 2.0 employs a random forest algorithm to predict B-cell epitopes derived from antigen-antibody structure. ABCpred was used to predict B-cell epitopes and the prediction method is based on artificial neural network [33]. A window length of 16-mer and threshold of 0.8 were applied to select the predicted epitopes [34]. ElliPro and DiscoTope 2.0 predict discontinuous antibody epitopes based on a protein three-dimensional structure [35,36]. EPSVR and CBTOPE servers were utilized to determine the B-cell epitopes and the prediction is based on support vector regression method [37,38]. Agadir is a prediction algorithm based on the helix/coil transition theory and the helical behaviour of B-cell epitopes can be determined [39].

2.3. T-cell epitope mapping

The protein sequences of TSA56 and ScaA antigens were screened for MHC-I-restricted cytotoxic T lymphocyte ($CD8^+$) epitopes of 9-mer length using TepiTool, NetMHCPan 4.1, and Immune Epitope Database (IEDB) Combined predictor. TepiTool predicts highly immunogenic MHC-restricted T-cell epitopes of 9-mer and 15-mer length, respectively [40]. The NetMHCpan-4.1 server was utilized to predict peptides binding to MHC molecules and the prediction algorithm is based on artificial neural networks. Default parameters for peptides having a threshold of <0.5 % for high-affinity binding peptides and <2 % for low-affinity binding peptides was used during the prediction [41]. MHC class I combined predictor tool is a combination of predictors of proteasomal processing, TAP transport, and MHC binding to predict the properties of T-cell epitope [42]. Twenty-seven HLA-A, HLA-B, HLA-DR, HLA-DQ, and HLA-DP alleles that are frequently observed in the general population were employed in predicting MHC-restricted T-cell epitopes [43]. NetMHCIIpan 4.0 predicts MHC-II-restricted peptides using artificial neural networks and represents nearly 1000 isotypes of human (HLA-DR, HLA-DQ, and HLA-DP) and mouse (H-2) MHC alleles [41,44]. $CD4^+$ T-cell epitopes binding to MHC-class II alleles with an IC50 below 1000 nM are previously reported as immunogenic and strong binders [45,46]. Hence, a threshold of IC50 = 1000 nM was applied to select the strong binders of MHC-class II molecule [40,47]. The TepiTool and NetMHCIIpan 4.0 servers cover a combination of alpha- and beta-protein chains for DP and DQ alleles [40,41,44]. IFNepitope determined the selected T-cell epitopes to stimulate IFN-γ production using motif, SVM, and hybrid approach algorithms [48]. IEDB tool was employed to determine the population coverage of the predicted MHC restricted epitopes based on allelic frequency of 115 nations and 27 distinct ethnic groups assembled into 16 geographical regions [49].

2.4. Determination of antigenic, allergenic, and conservancy of predicted B- and T-cell epitopes

The antigenic and allergenic properties of the identified B- and T-cell epitopes were evaluated using the VaxiJen 2.0 server [30] and AllerTOP v2.0, respectively [50]. The IEDB-based conservancy tool was utilized to compute the conservancy of B- and T-cell epitopes using 160 of TSA56 sequences and 11 full-length amino acid sequences of ScaA [51].

2.5. Molecular docking of T-cell epitopes with MHC alleles

Molecular docking to establish an interaction between predicted T-cell epitopes and MHC alleles is an integral component of reverse vaccinology [52]. Further to facilitate T-cell epitope and MHC allele interaction structures available in the repository of protein data bank were retrieved and the ligand free 3D-structures of MHC molecules were utilized for docking (PDB; http://www.rcsb.org/pdb/). In addition, the top-listed T-cell epitopes with the maximum score of antigenic properties were qualified and included for docking. The 3-D structure of T-cell epitopes was prepared using the PEPFOLD 3 web tool and their interaction with MHC/HLA alleles were analyzed [53]. Modelled structures were capped at their N- and C-terminals using BIOVIA Discovery Studio Visualizer (Dassault Systèmes®) [54]. We utilized a widely used ClusPro server (https://cluspro.bu.edu/login.php) for docking the predicted T-cell epitopes and MHC alleles, fibronectin receptor and vaccine construct [55]. ClusPro server provides a several flexibilities to enable removal of unstructured regions of proteins, application of attraction or repulsion, accounting for pairwise distance restraints, and construction of homo-multimers, and precise mapping of binding sites of interacting protein complexes. The three steps of ClusPro server-based protein-protein docking: (i) PIPER based Fast Fourier transform (FFT) rigid-body docking using sample set of billions protein structures; (ii) root-mean-square deviation (RMSD)-based clustering of the 1000 lowest-energy structures; and (iii) energy minimization to remove steric clashes in docked complexes [55,56]. In general, the ClusPro server performs protein-ligand docking through 70,000 rotations. Then the ligand is translated in x, y, z axis relative to the protein on a grid for each round of the rotation analyzed to select the best scored docked complexes. Of the 70,000 ligand rotations, the 1000 translation combinations with the lowest energy score are selected. The interaction energy of the MHC-peptide complex was analyzed using the PRODIGY web server [57].

2.6. Designing of multi-epitope vaccine constructs based on TSA56 and ScaA proteins

The vaccine constructs were prepared by adding the predicted CTL and HTL epitopes, and linear and conformational B-cell epitopes of TSA56 and ScaA with appropriate linkers. The cleavable AAY and KK linkers were added between the CTL and HTL epitopes and the B-cell epitopes were linked through the GGGGS flexible linker [58]. The T-cell epitopes are processed with the help of proteases and the epitopes are subsequently loaded into the groove of MHC molecules for surface presentation to T-cells [43]. The flexible linkers help to restore protein folding by allowing interaction between different domains [59]. The integrin-binding RGD motif was chosen as a biological adjuvant and added at the N-terminal region of the vaccine constructs and the presence of EAAAK rigid linker provides independent function of bi-functional domains [19]. It was previously shown that the RGD motif could stimulate antibody titers production by more than 10 times upon administration of peptide or antigens conjugated with RGD adjuvant [60]. The 56-kDa type-specific antigen of *O. tsutsugamushi* is known to interact with the RGD motif in the fibronectin to further facilitate the entry of the bacterium into the host cells [19,61]. The antigenic and non-allergenic properties including the other bio-chemical properties of chimeric vaccine constructs were determined as described previously [30,31,49].

2.7. Structural prediction of vaccine constructs

The PSIPRED online server (http://bioinf.cs.ucl.ac.uk/psipred/) was utilized to determine the secondary structure of vaccine

constructs [62]. It is based on two feed-forward neural networks and uses a stringent cross-validation evaluation method to give out simple and accurate secondary structures. The three-dimensional structures of vaccine constructs were determined through modelling. I-TASSER modelling server was employed to simulate tertiary structures and the algorithm works by identifying the structural templates in the PDB repository by homology-independent multiple threading method [35]. Then the full-length three-dimensional atomic models are constructed through several threading alignments and iterative template-based fragment assembly simulations [55]. A template modeling (TM) score is generated to quantify the modelled protein structure similarity assessment and the TM score of more than 0.5 indicates a model with accurate topology, and the scores below 0.17 shows a a randomly chosen unrelated proteins as reported previously [58]. The structural validation of simulated structures of vaccine constructs were performed using the Alphafoldv2.0, ProCheck, and ProSA web servers [63,64]. The quality of modelled structures was evaluated based on the Ramachandran

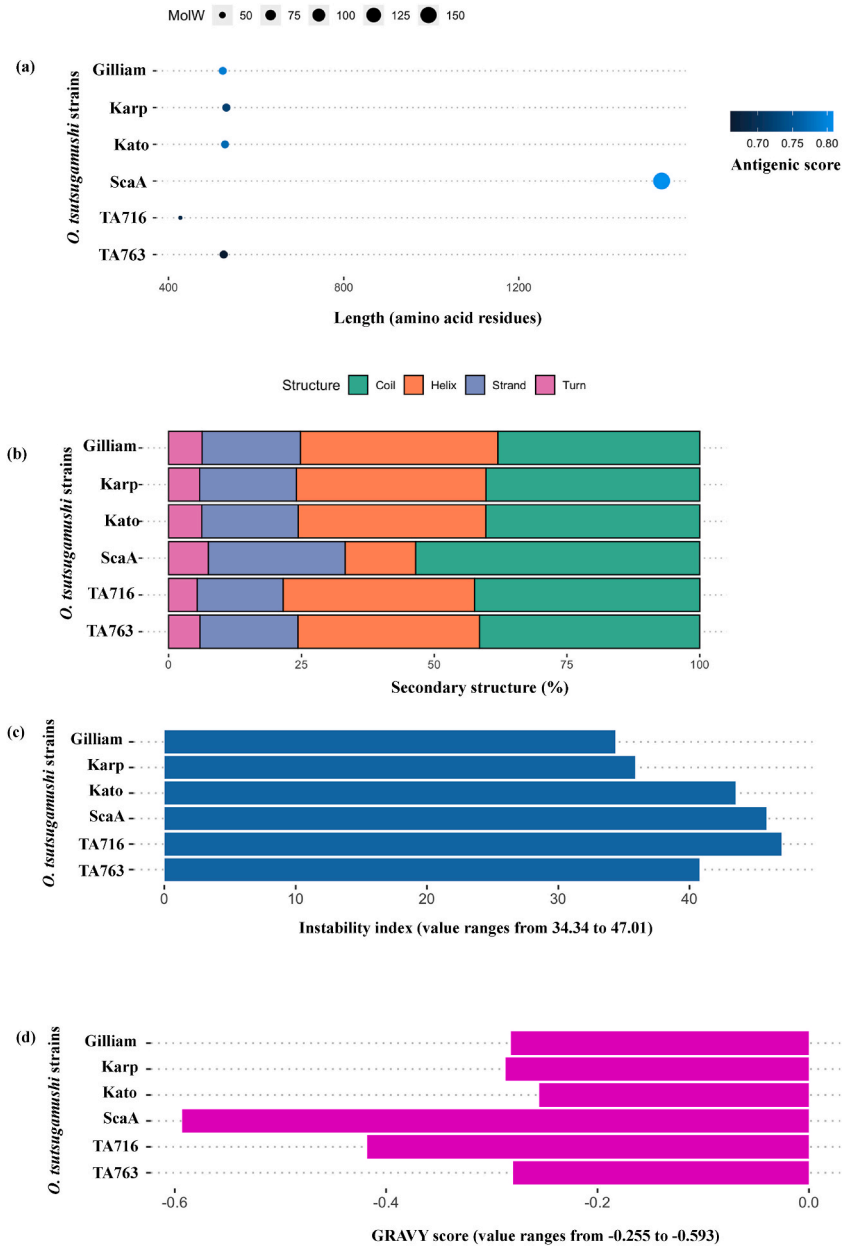


Fig. 1. Bio-physicochemical properties Physicochemical properties of the 56-kDa type-specific antigen (TSA56) and surface cell antigen (ScaA) of Orientia tsutsugamushi. (a) The antigenic score of TA763, Gilliam, Karp, Kato, TA716, and ScaA was observed between 0.6604 and 0.8087. (b) The secondary structures such as helix, strand, turn and coil were determined for all five strains using SOPMA. (c) The instability index and (d) GRAVY score for all five strains were determined. Similarly, the (e) pI value and (f) aliphatic index of Gilliam, Karp, Kato, TA716, TA763, and ScaA have been determined. All five TSA56 strains and the ScaA protein had an antigenic score >0.6604 on the antigenicity scale.

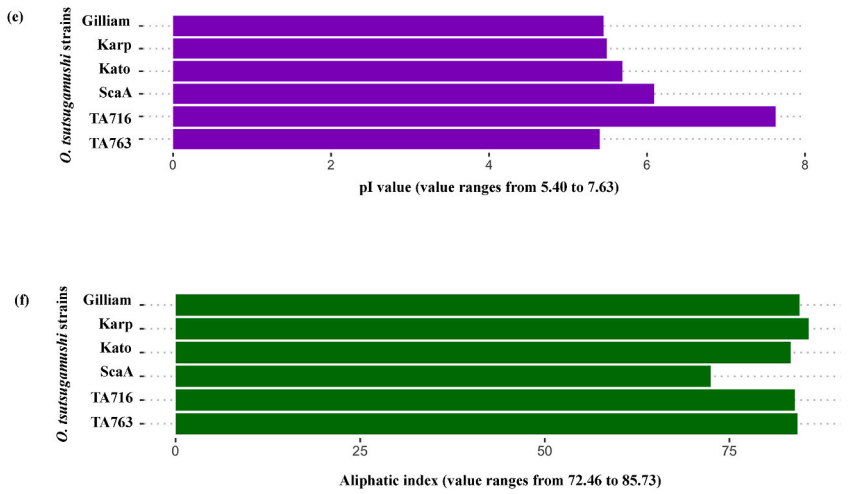


Fig. 1. (*continued*).

plots and Z-scores of ProCheck and ProSA and the model with the highest predicted local difference test (pLDDT) scores was selected as described previously [63–65].

2.8. Fibronectin-epitope vaccine construct docking and simulation

Previously it was reported that the TSA56 antigen of *O. tsutsugamushi* binds to RGD motif in the fibronectin and helps the bacterium to enter the host cell [19,61]. A total of eight vaccine constructs (1–8) were prepared using the rearrangement of predicted $CD4^+$, $CD8^+$ T-cell and B-cell epitopes of TSA56 and ScaA proteins with appropriate linkers. Constructs 1, 2, 5, and 6 comprised only linear and a combination of linear and conformation B-cell epitopes, with the idea to develop chimeric antigens for rapid diagnostic assays of scrub typhus. Vaccine constructs 3, 4, 7, and 8 comprised of both B- and T-cell predicted epitopes. Molecular docking of vaccine constructs with fibronectin (PDB ID: 1FNF) was performed using Cluspro with default settings [55]. The best docked vaccine constructs 3, 4, 7, and 8 with fibronectin (1FNF) were considered for molecular dynamics (MD) simulation using the AMBER18 software [66]. The AMBER ff14SB force field was applied to optimize the backbone dihedral parameters that can potentially change the preference of secondary structure of proteins [67]. The systems were first neutralized using counter ions (Na^+/Cl^-) ions and then solvation was carried out through addition of TIP3P water molecules. The model was constructed using the 'LEAP' module of AMBER. The systems were minimized using the steepest descent method for 5000 steps, followed by a conjugate gradient method for another 5000 steps. The systems were then heated gradually from 0 K to 300 K using a constraint of 100 kcal/mol. Equilibration was conducted for 1 ns by gradually reducing the constraint. Simulations were carried out with periodic boundary conditions using the Particle Mesh Ewald method to compensate for electrostatic interactions and long-range van dar Waals interactions. MD simulations were performed with a 2.0 fs timestep and a non-bonded cutoff of 10.0 Å using the SHAKE algorithm was applied on all hydrogen atoms throughout the simulation trajectories [68]. The Berendsen coupling algorithm was employed throughout the simulation process to maintain a constant pressure of 1 atm and temperature of 300 K and after every 100 steps the list was updated. All MD simulations were performed with a trajectory dataset for 100 ns time and the results of three independent simulations were used for further analysis. The structural stability, compactness and flexibility of docked complexes were performed by calculating the root mean square deviation (RMSD), root mean square fluctuation (RMSF), and radius of gyration (RoG) using the CPPTRAJ module of AMBER tool [69].

2.9. Immune simulation of multi-epitope vaccine constructs

C-IMMSIM server was used for in silico immune stimulation and immune response profile of vaccine constructs 3, 4, 7, and 8. This server employs the Celada-Seiden model to produce the immune reactions in the vaccinated mammalian system [70]. Administration of vaccine in the absence of lipopolysaccharide (LPS) was included and performed the simulations using the default parameters settings (https://kraken.iac.rm.cnr.it/C-IMMSIM/index.php?page=1). The sigma-70 factor of RNA polymerase subunit was included as negative control antigen in the study. The measurement of immune diversity was performed and interpreted from the plots using the Simpson index D [58].

3. Results

3.1. Prediction of antigenicity and physicochemical properties of TSA56 and ScaA

All protein sequences of TSA56 and ScaA of *O. tsutsugamushi* were found antigenic with scores (>0.6604) above the threshold value

(Fig. 1a). All sequences had pI > 5: TA763 and TA716 had the lowest (5.40) and highest (7.63) values, respectively (Fig. 1e). The negatively charged amino acid residues were observed more than the positively charged residues in all strains, except for TA716. The instability index ranged from 34.35 to 47.01 for TSA56 and 45.86 for ScaA (Fig. 1c). The grand average of hydropathicity score was −0.255 to −0.418, suggesting the globularity and hydrophilicity properties of proteins (Fig. 1d). The results of secondary structure revealed that TSA56 and ScaA antigens were rich in random coils (Fig. 1b). Furthermore, the aliphatic index ranged from 72 to 85, indicating that the antigens were thermostable (Fig. 1f).

Table 1

Predicted linear B-cell epitopes of TSA56 and ScaA using Bcepred (1), BepiPred (2) and ABCpred (3). Antigenicity, allergenicity, Agadir score, and conservancy is shown across antigens. Selected epitopes for chimeric construct design are highlighted in bold. The antigenicity of the epitopes was evaluated using antigenic scores ≥0.4, the threshold value. A, Allergen; NA, Non-allergen. Eleven ScaA strains and 160 TSA56 strains were considered for conservancy analysis by the IEDB tool.

TSA56 protein	Name of identified peptide	Linear B-cell epitopes	Length	Server	Allergeicity by AllerTope v. 2.0	Agadir Score	Conservancy (%)
Karp	KP1	48-VESARLDPADAEGKKH-63	16	1,2	NA	2.02	2.50
	KP2	95-LTNITAQVEEGKVKAD-110	16	3	NA	0.67	29.38
	KP3	141-MPISIADRDFGIDIPN-156	16	3	NA	0.45	23.75
	KP4	150-FGIDIPNIPQQQAQAAQPQLNDEQRAAA-177	28	2	NA	0.65	11.25
	KP5	277-PFADIAGIDVPDTSLPNSASVEQIQNKMQE-306	30	2	A	0.92	13.12
	KP6	387-HAGIKKAMEKLAAQQE-402	16	3	A	4.11	33.12
	KP7	398-AAQQEEDAKNQG-409	12	1,2	NA	2.73	47.50
	KP8	402-EEDAKNQGEGDCKQQQ-417	16	2,3	A	0.65	39.38
	KP9	414-KQQQGTSEKSKKGKDK-429	16	1,2,3	A	0.73	3.12
Kato	KaP1	48-GVESTRLDPADAGGKK-63	16	1,2,3	NA	1.01	1.88
	KaP2	134-TIMPISIADRDLGVDI-149	16	3	NA	0.53	5.62
	KaP3	392-DGGCNGGGDNKKKRGAS-408	17	1,2,3	NA	0.83	3.75
	KaP4	425-TKETEFDLSMIVGQVK-440	16	3	A	1.71	3.75
Gilliam	GP1	20-ASAIELGEEGGLECGP-35	16	2,3	A	0.42	11.88
	GP2	48-GAESTRLDSTDSEGKKHL-65	18	1,2,3	NA	0.34	6.88
	GP3	96-LRNISAEVEVGKGKVD-111	16	1,2,3	A	0.71	16.25
	GP4	121-GGGTDTPIRKRFKLTP-136	16	1,2,3	NA	2.59	1.25
	GP5	140-TIMPISIADRDVGVDT-155	16	3	NA	0.35	5.62
	GP6	269-PFADIAGIDVPDTGLPNSASVEQIQSKMQ-297	29	2	NA	0.30	4.38
	GP7	385-AMEKLAAQQEEDAKNQG-401	17	1,2,3	A	5.47	46.25
	GP8	394-EEDAKNQGEGDCKQQQ-409	16	1,2,3	A	0.65	39.38
	GP9	405-CKQQQGASEKSKEGKGKETE-424	20	1,2	NA	0.59	4.38
	GP10	420-GKETEFDLSMIVGQVK-435	16	3	NA	1.81	10.00
TA763	TP1	20-ASAIELGDEGGLECGP-35	16	2,3	A	0.3	63.75
	TP2	84-TPSIRAELGVMYLRNI-99	16	3	NA	0.89	0.62
	TP3	96-LRNISAEVELGKVKAD-111	16	3	A	1.25	6.88
	TP4	104-ELGKVKADSGSKTKADSG-121	17	1,2,3	NA	0.59	0.62
	TP5	122-GETDAPIRKRFKLTPPQP-139	18	1,2,3	A	0.39	6.25
	TP6	142-MPISIADRDFGVDVTN-157	16	2,3	NA	0.46	6.25
	TP7	278-GIDVPDAGLPNSATVEQIQNKMQ-300	23	1,2,3	NA	1.75	1.25
	TP8	393-AAQQEEDAKNQGEGDCKQQQGTSEK-417	25	1,2	A	2.85	11.88
TSA56 & ScaA protein	Name of identified peptide	Linear B-cell epitopes	Length	Server	Allergeicity by AllerTope v. 2.0	Agadir Score	Conservancy (%)
TA716	TaP1	104-ESGKTNSGTDTRADTDSPILQRPKFT-129	26	1,2,3	NA	0.30	2.50
	TaP2	138-ISIADRNLGVDVTNVP-153	16	3	A	0.83	2.50
	TaP3	274-DALPNSASVEQIQNKMQ-294	17	1,2	NA	1.38	2.50
	TaP4	405-GASEDSDSGGGLKKEKAKRQ-424	20	1,2	NA	1.1	0.62
ScaA	SP1	148-GEVTDGKKHKKSPLKKLKDLGSTL-171	24	1,2,3	NA	1.24	9.09
	SP2	225-SDDTKKQVEVKQNEGYGV-242	18	1,2,3	NA	2.86	9.09
	SP3	247-IKQDEVAFDKIEGRSASI-264	18	2,3	NA	0.48	9.09
	SP4	355-FFSSTTHGGKTEGTDRSKFTGI-376	22	1,2,3	NA	0.43	54.55
	SP5	822-QCSISDKLSAKSEMTL-837	16	3	NA	0.84	9.09
	SP6	851-SGRIDAGNAKVNLGLN-866	16	3	NA	0.75	9.09
	SP7	878-SGHKSNKGSSTNPEKFKKKPVTVF-901	24	1,2,3	NA	0.36	9.09
	SP8	1117-QQQQQQQQQQQQQQQQQQQQQQQQQQ-1142	26	1,2	NA	1.17	27.27
	SP9	1438-NSNSTFNPDVITSREQCLHK-1457	20	1,2	NA	0.44	27.27

Table 2
Predicted conformational B-cell epitopes using one primary sequence-dependent server (CBTOPE [1]) and three different structure-dependent tools (Ellipro [2], Discotope [3], and EPSVR [4]). Agadir scores and conservancy of the epitopes are seen across antigens; Selected epitopes for chimeric construct are highlighted in bold.

	TSA & ScaA Protein	Name of identified peptide	Conformational B-cell Epitope	No. Of residues	Servers	Agadir Score	Conservancy (%)
I-TASSER Model	Karp		Y94,L95,T96,N97,I98,T99,A100,Q101,V102,E103, E104,G105,K106,V107,K108,A109,D110	17	1,2,3	0.69	29.38
		KC1	D280,I281,A282,G283,I284,D285,V286,P287,D288, T289,S290,L291,P292,N293,S294,A295,S296,V297, E298,Q299,I300,Q301,N302,K303,M304,Q305	26	1,2,3,4	0.91	13.75
		KC2	D319,G320,Y321,L322,G323,G324,N325,A326,F327, A328,N329,Q330,I331,Q332,L333,N334,F335,V336, M337	19	1,4	0.80	16.25
			Q408,G409,E410,G411,D412,C413,K414,Q415,Q416, Q417,G418,T419,S420,E421,K422,S423,K424,K425, G426,K427,D428,K429,E430,A431,E432,F433,D434, L435,S436,M437,I438,V439,G440,Q441	34	2,3	1.52	3.12
			G418,T419,S420,E421,K422,S423,K424,K425,G426, K427,D428,K429,E430,A431,E432,F433,D434,L435, S436,M437,I438,V439,G440,Q441	24	1,2,3,4	1.54	3.12
	Kato		G44,M45,I46,T47,G48,V49,E50,S51,T52,R53,L54,D55, P56,A57,D58,A59,G60,G61,K62,K63,Q64,L65,P66,L67, T68,T69,S70,M71,P72,F73,G74,G75,T75	32	1,2,3,4	0.70	0.00
		KaC1	N98,V99,K100,A101,E102,V103,E104,S105,G106, K107,T108,G109,S110,D111,A112,D113,I114,R115, S116,G117,A118,D119,S120,*P*121,M122,P123,Q124, R125,Y126,K127	30	3,4	2.07	3.12
		KaC2	P131,Q132,P133,T134,I135,M136,P137,I138,S139, I140,A141,D142,R143,D144,L145,G146,V147,D148, I149,P150	20	2,3,4	0.47	3.12
		KaC3	S287,V288,E289,Q290,I291,Q292,N293,K294,M295, Q296,E297,L298,N299	13	1,2,3	0.77	25.00
		KaC4	G314,G315,N316,A317,F318,A319,N320,Q321,I322, Q323,L324,N325,F326,R327	14	1	0.90	4.38
			K421,G422,K423,E424,T425,K426,E427,T428,E429, F430,D431,L432,S433,M434	14	1	0.40	3.75
	Gilliam	GC1	M45,I46,T47,G48,A49,E50,S51,T52,R53,L54,D55,S56, T57,D58,S59,E60,G61,K62,K63,H64,L65,S66,L67,T68, T69,G70,L71,P72,F73,G74	30	1,4	1.07	6.88
			D277,V278,P279,D280,T281,G282,L283,P284,N285, S286,A287,S288,V289,E290,Q291,I292,Q293,S294, K295,M296,Q297	21	1,2,3,4	0.40	4.38
		GC2	E387,K388,L389,A390,A391,Q392,Q393,E394,E395, D396,A397,K398,N399,Q400,G401,E402,G403,D404, C405,K406,Q407,Q408,Q409,G410,A411,S412,E413, K414,S415,K416,E417,G418,K419,G420,K421	35	1,2,3	3.53	5.00
			G401,E402,G403,D404,C405,K406,Q407,Q408,Q409, G410,A411,S412,E413	13	1,2,3	0.66	28.75
			E413,K414,S415,K416,E417,G418,K419,G420,K421, E422,T423,E424,F425,D426,L427,S428,M429	17	1,2,3,4	0.50	5.00
		GC3	E417,G418,K419,G420,K421,E422,T423,E424,F425, D426,L427,S428,M429,I430,V431,G432,Q433,V434, K435,L436,Y437,A438,D439	23	1,4	1.85	7.50
	TA763	TC1	T57,D58,S59,E60,G61,K62,K63,H64,L65,S66,L67,T68, T69,G70	18	2,3	1.43	1.88
		TC2	P72,F73,G74,G75,T76,L77,A78,A79,G80,M81,T82,I83, T84,P85,S86,I87,R88	17	4	0.3	0.62
		TC3	N288,S289,A290,T291,V292,E293,Q294,I295,Q296, N297,K298,M299,Q300,E301,L302,N303,D304	17	1,2,3	2.76	0.62
			R310,E311,S312,F313,D314,G315,Y316,L317,G318, G319,N320,A321,F322,A323,N324,Q325,I326,Q327, L328,N329,F330,V331,M332,P333	24	4	0.68	13.12
		TC4	Q395,Q396,E397,E398,D399,A400,K401,N402,Q403, G404,E405,G406,D407	13	2,4	1.57	39.38
			G413,T414,S415,E416,K417,S418,K419,E420,G421, S422,K423,K424,E425,P426,E427,F428,D429,L430, S431,M432,I433	21	4	0.35	0.62
	TA716		M81,T82,I83,A84,P85,G86,F87,R88,A89,E90,I91,G92, V93,L96,T97,N98,V99,K100,A101,E102,V103	21	1,3	0.3	2.50

(*continued on next page*)

Table 2 (*continued*)

TSA & ScaA Protein	Name of identified peptide	Conformational B-cell Epitope	No. Of residues	Servers	Agadir Score	Conservancy (%)
	TaC1	L96,T97,N98,V99,K100,A101,E102,V103,E104,S105, G106,K107,T108,N109	14	1,3	0.84	2.50
	TaC2	P275,S276,D277,A278,L279,P280,N281,S282,A283, S284,V285,E286,Q287,I288,Q289,N290,K291,M292, Q293,E294	20	2,3	1.41	1.88
		A342,Q343,A344,T345,A346,Q347,E348,R404,G405, A406,S407,E408,D409,S410,D411,S412,G413,G414, G415,L416,K417,K418	22	3	0.44	0.62
	TaC3	R376, H377, A378, G379, I380, K381, K382, A383, M384, K386, L387, Q390, D391, C394, D395, G396, G398, D399, N400, K402, K403	21	3,4	0.56	2.50
	TaC4	S407,E408,D409,S410,D411,S412,G413,G414,G415, L416,K417,K418,E419,K420,A421,K422,R423,Q424, S425,L426,I427	22	1,2,4	1.89	0.62
ScaA	SC1	L224,S225,D226,D227,T228,K229,K230,Q231,V232, E233,V234,K235,Q236,N237,E238,G239,Y240,G241, V242	19	2	5.9	9.09
	SC2	F289,N290,L291,S292,D293,S294,E295,F296,I297, V298,N299,N300,T301,P302,H303,S304,N305,S306, D307,L308,K309	21	2	0.62	9.09
	SC3	V1026,S1027,D1028,D1029,E1030,E1031,G1032, G1033,K1034,G1035,E1036,D1037,I1038,Q1039, A1040,S1041,T1042,A1043,S1044,H1045	20	2,4	0.45	18.18
		Q1109,P1110,I1111,A1112,A1113,G1114,A1115, G1116,Q1117,Q1118,Q1119,Q1120,Q1121,Q1122, Q1123,Q1124,Q1125,Q1126,Q1127,Q1128,Q1129	21	2,4	1.24	18.18
	SC4	I1252,S1253,G1254,I1255,S1256,I1257,G1258,A1259, D1260,K1261,H1262,L1263,K1264,N1265,N1266, A1267,I1268	17	1,4	0.4	100.00
	SC5	K1457,T1458,F1459,Y1460,Q1461,L1462,G1463, V1464,K1465,L1466,A1467,I1468,K1469,R1470, N1471,I1472,M1473,D1474,I1475,G1476,I1477,A1478	22	4	1.54	81.82

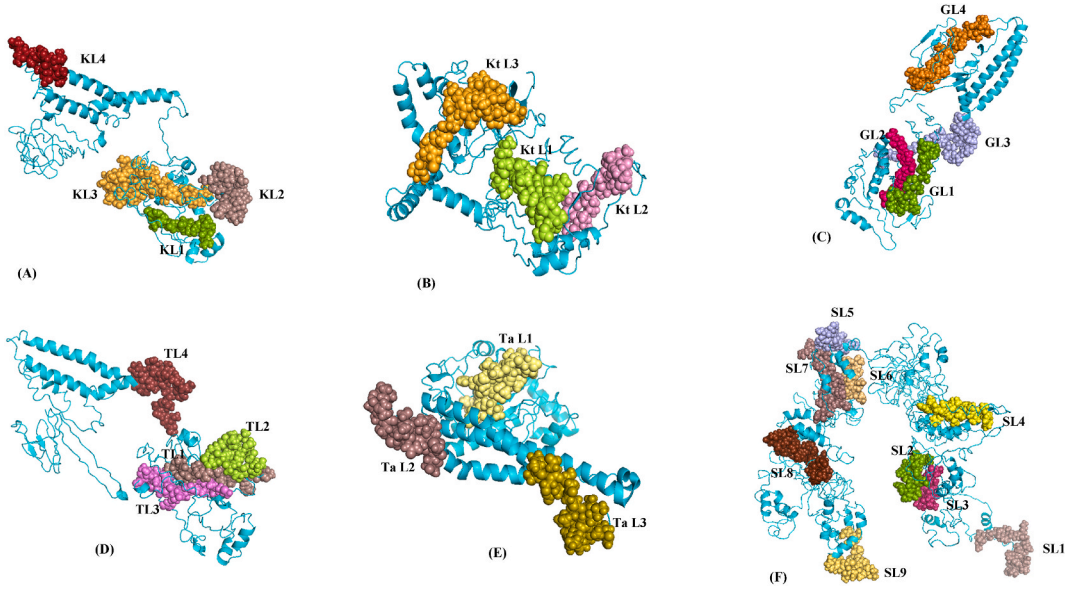


Fig. 2. Structural localization of predicted linear B-cell epitopes of TSA56 and ScaA on the native antigens of five strains of *Orientia tsutsugamushi*. (A) Karp: KL1 (48–63), KL2 (95–110), KL3 (141–177), and KL4 (398–409); **(B) Kato**: KtL1 (48–63), KtL2 (134–149), and KtL3 (392–408); **(C) Gilliam**: GL1 (48–65), GL3 (140–155), GL4 (269–279), and GL5 (405–435); **(D) TA763**: TL1 (84–99), TL2 (104–121), TL3 (142–157), TL4 (278–300); **(E) TA716**: TaL (104–129), TaL (274–294), and TaL (405–424); **(F) ScaA**: SL1 (148–171), SL2 (225–242), SL3 (247–264), SL4 (355–376), SL5 (822–837), SL6 (851–866), SL7 (878–901), SL8 (1117–1142), and SL9 (1438–1457).

3.2. Prediction of B-and T-cell epitopes of TSA56 and ScaA

The linear B-cell epitopes were predicted using three servers namely, Bcepred, BepiPred, and ABCpred and the predicted continuous B-cell epitopes are presented in Fig. S1 and Table S1. The predicted epitopes ranged from 9 to 21 amino acid residues. Using the parameters such as conservancy, helical content, non-allergenicity, and antigenicity the top 35 and 9 linear B-cell epitopes of TSA56 and ScaA, respectively were selected (Table 1). The linear B-cell peptides, P1 was conserved in all five strains (Karp, Kato, Gilliam, TA763, and TA716), P2 was conserved in Karp, Kato, and Gilliam strains, and P3 was conserved in Karp and Gilliam strains (Table S2).

ElliPro, DiscoTope, EPSVR, and CBTOPE servers were used in prediction of discontinuous B-cell epitopes of TSA56 of Karp, Kato, Gilliam, TA763, and TA716 strains (Fig. S1). The top 29 discontinuous B-cell epitopes of TSA56 (5 for Karp and 6 each for Gilliam, Kato, TA763, and TA716) and 6 discontinuous B-cell epitopes of ScaA were selected as the best predicted epitopes (Table 2). The TSA56 peptides, P4 and P5 were conserved in Karp, Gilliam, and TA763; P6 was conserved in Karp, Kato, and TA763; P7 was common in Gilliam and TA763; and P8 was conserved in TA763 and TA716 (Table S2). The proper folding of the in-silico determined immunogenic B-cell epitopes is important in eliciting a long-term antibody response and immune protection. The predicted linear and discontinuous B-cell epitopes are found on the accessible regions of the modelled structures of TSA56 and ScaA as observed using BIOVIA Discovery Studio 2017 R2 (Fig. 2A–F, Table 1, Fig. 3A–F & Table 2). The majority of the computationally determined B-cell epitopes were mapped to the antigenic domain as compared to the variable domain of TSA56. Thus, the findings of this study suggest the immunogenic potential of TSA56 and ScaA antigens of *O. tsutsugamushi*. The peptides, GP2, GP6, GP7, GP9, GP10 of Gilliam, KaP2 and KaP4 of Kato, KP2 of Karp, TaP4 of TA716, and TP8 of TA763 had epitopes common for linear and discontinuous B-cells (Table 1). Similarly, the presence of overlapping epitopes for linear and discontinuous B-cells were observed for the peptides, SP2 and SP8 of ScaA (Table 1). However, future investigations are required to characterize the functionality of B-cell epitopes in designing peptide vaccines and diagnostic assays.

The leveraging of more than one server to predict potential T-cell epitopes might contribute in reducing the false positive targets for vaccine preparation. Three prediction algorithms were compared and employed in identifying HLA-class I epitope and two servers were used in prediction of HLA-class II epitope (Fig. S2 and Table S1). The predicted peptides were found antigenic, non-allergenic, and can induce IFN-γ production. We found 51 (fifty-one) non-overlapping strong binders of MHC-I for TSA56 (13 CTL epitopes for Karp, 11 for Kato, 12 for Gilliam, nine for TA763, and six for TA716), and 9 (nine) for ScaA protein (Table 3). Notably, we observed 14 $CD8^+$ T-cell epitopes common to all five prototype strains (Table S3). CTL and HTL epitopes were found antigenic, non-allergenic, and IFN-γ-inducing, with binding-free energy $\geq$ -6.3 kJ/mol. Molecular docking confirmed the interaction between the T-cell epitopes and MHC alleles which could further increase the reliability of predicted T-cell epitopes for future laboratory investigations. The lack of structures of MHC alleles limited the scope of docking analysis for predicted T-cell epitopes in the present study. We found 27 non-overlapping strong binders of MHC-II for TSA56 (eight $CD4^+$ T-cell epitopes each for Karp and Kato, five for Gilliam, and three each for TA763 and TA716), and three for ScaA (Table 4). Notably, we found nine $CD4^+$ T-cell epitopes that were conserved in all prototype strains (Table S4).

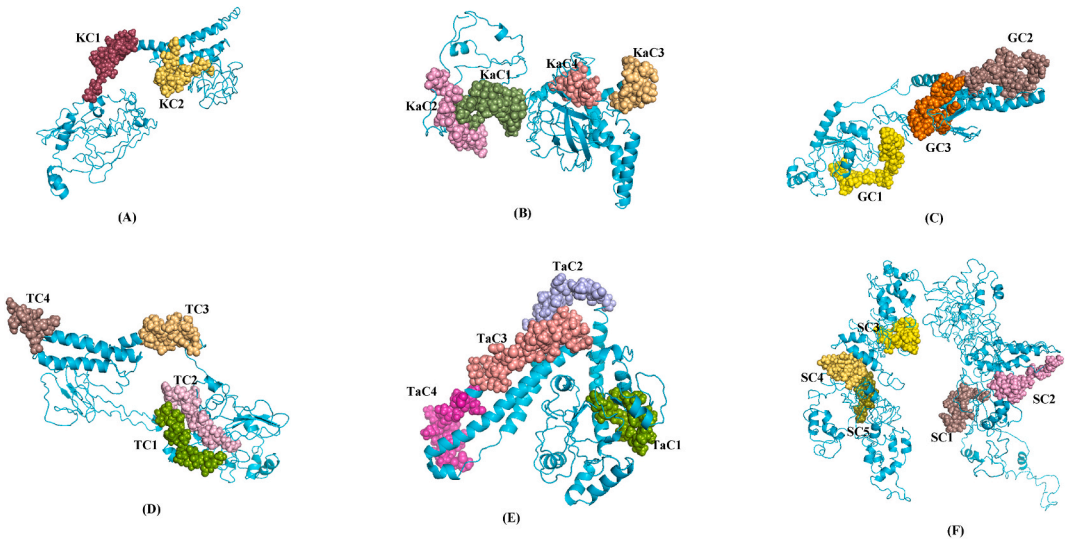


Fig. 3. Structural localization of predicted conformational B-cell epitopes of TSA56 and ScaA on the native antigens of five strains of *Orientia tsutsugamushi*. (A) Karp: KC 1 and KC2. **(B) Kato**: KaC1, KaC2, KaC3, and KaC4. **(C) Gilliam**: GC1, GC2, and GC3. **(D) TA763**: TC1, TC2, TC3, and TC4. **(E) TA716**: TaC1, TaC2, TaC3, and TaC4. **(F) ScaA**: SC1, SC2, SC3, SC4, and SC5. KC, KaC, GC, TC, and TaC refer to conformational B-cell epitopes predicted in the 56-kDa type-specific antigen (TSA56) of Karp, Kato, Gilliam, TA763, and TA716 strains. SC refers to conformational B-cell epitopes predicted in the surface cell antigen A (ScaA). The full-list of predicted conformational B-cell epitopes is provided in Table 2.

Table 3

Binding energies for docked MHC-I-epitope complexes. The 27 HLA-I supertypes given below have a cumulative population coverage of ≥97 % for the five major ethnicities, namely African Americans, Caucasians, Hispanics, Asians, and North American Natives. The epitopes selected for the chimeric construct had good antigenic values and were positive for IFN-γ. Sixteen ScaA strains and 162 TSA56 strains were considered for conservancy analysis using the IEDB tool.

HLA I alleles	PDB ID
HLA-A*01:01	4nqv
HLA-A*02:01	1duz
HLA-A*02:03	3oxr
HLA-A*02:06	3oxr
HLA-A*03:01	2xpg
HLA-A*24:02	3i6l
HLA-A*68:01	4hwz
HLA-A*68:02	4i48
HLA-B*07:02	3vcl
HLA-B*08:01	1m05
HLA-B*35:01	4prn
HLA-B*51:01	1e27
HLA-B*53:01	1a1m
HLA-B*57:01	3x12

HLA I alleles (used I-TASSER model)
HLA-A*11:01
HLA-A*23:01
HLA-A*26:01
HLA-A*30:01
HLA-A*30:02
HLA-A*31:01
HLA-A*32:01
HLA-A*33:01
HLA-B*15:01
HLA-B*40:01
HLA-B*44:02
HLA-B*44:03
HLA-B*58:01

Protein	MHC I Epitope (9 mer)	Allele	Binding energy in ΔG (kcal mol^{-1})	Server	Conservancy (%)
Karp	80-MTIAPGFRA-88	HLA-A*6801	-10.2	1,2,3	45.00
	98-ITAQVEEGK-106	HLA-A*68:01	-7.9	1,2,3	37.50
	121-GGKDAPIRK-129	HLA-A*30:01	-10.8	2	3.75
	137-QPTIMPISI-145	HLA-B*07:02 HLA-B*51:01 HLA-B*53:01	-9.8 -10 -10.5	1,2,3	80.62
	157-IPQQQAQAA-165	HLA-B*07:02	-9.8	3	15.62
	174-RAAARIAWL-182	HLA-B*07:02 HLA-A*30:01 HLA-B*58:01 HLA-A*02:03 HLA-A*02:06 HLA-B*08:01	-10.1 -10.3 -8.7 -11.1 -9.2 -8.3	1	37.50
	203-VINPILLNI-211	HLA-A*02:03 HLA-A*02:06 HLA-A*32:01 HLA-A*02:01	-7.8 -9.7 -10.3 -9.6	1,2	33.75
	319-DGYLGGNAF-327	HLA-B*35:01	-7.8	1,2	16.25
	325-NAFANQIQL-333	HLA-A*6802 HLA-B*3501 HLA-B*5101	-10.4 -9.4 -9.6	1,2	53.75
	358-EAVAAAAVR-366	HLA-A*3301 HLA-A*6801	-9.4 -10.9	1,2,3	55.62
	383-KLQRHAGIK-391	HLA-A*0301 HLA-A*3101	-11.1 -8	1,2,3	35.00
	458-AGVGAGLAY-466	HLA-B*15:01 HLA-B*35:01 HLA-A*30:02	-8.6 -9.5 -8.2	1,2	31.25
	522-LMASVSVRY-530	HLA-A*30:02 HLA-B*35:01 HLA-A*03:01 HLA-A*32:01 HLA-A*11:01 HLA-A*01:01	-11 -11.4 -10.2 -10.2 -9.5 -10.5	1,2,3	1.88

Protein	MHC I Epitope (9 mer)	Allele	Binding energy in ΔG (kcal mol^{-1})	Server	Conservancy (%)
Kato	81-MTIAPGFRA-89	HLA-A*6801	-10.2	1,2,3	45.00% (72/160)
	132-QPTIMPISI-140	HLA-B*07:02 HLA-B*51:01 HLA-B*53:01	-8.9 -9.6 -9.4	2,3	80.62% (129/160)
	316-NAFANQIQL-324	HLA-A*68:02, HLA-B*35:01 HLA-B*51:01	-9.8 -10.3 -7.5	1,2	53.75% (86/160)
	348-QEAAAAAAV-356	HLA-B*44:02 HLA-B*44:03 HLA-B*40:01	-10.2 -10.3 -10.5	1,2,3	20.00% (32/160)
	374-KLKRHAGIK-382	HLA-A*0301 HLA-A*3001	-10.1 -10.1	1,2,3	15.00% (24/160)
	385-MEELAAQDG-393	HLA-B*44:02	-9.4	3	4.38% (7/160)
	446-FSIYAGVGA-454	HLA-A*68:02	-10.4	1,3	40.00% (64/160)
	458- HTYGKIDDK-466	HLA-A*03:01 HLA-A*11:01 HLA-A*68:01	-9.5 -8.4 -10.2	1,2,3	16.25% (26/160)
	496-YVDIEGSYM-504	HLA-A*01:01	-10	2	28.12% (45/160)
	516-ASVGVRYNF-524	HLA-A*32:01 HLA-B*58:01 HLA-B*57:01 HLA-B*15:01	-9.6 -10.7 -10.3 -12.1	1,2,3	24.38% (39/160)
	514-LMASVGVRY-522	HLA-A*0301 HLA-A*3002 HLA-B*3501 HLA-B*1501	-10.8 -8.3 -12.3 -8.4	1,2,3	23.75% (38/160)
Gilliam	81-MTIAPGFRA-89	HLA-A*6801	-10.2	1,2,3	45.00% (72/160)
	95-YLRNISAEV-103	HLA-A*02:03 HLA-A*02:01 HLA-A*02:06	-10.4 -10.4 -10.1	1,2,3	36.88% (59/160)
	138-QPTIMPISI-146	HLA-B*07:02 HLA-B*51:01 HLA-B*53:01 HLA-B*57:01	-8.9 -9.6 -9.4 -10.3	1,2,3	80.62% (129/160)
	219-RPRQNLDIL-227	HLA-B*0702	-9.7	1,2,3	5.62% (9/160)
	316-NAFANQIQL-324	HLA-A*68:02, HLA-B*35:01 HLA-B*51:01	-9.8 -10.3 -7.5	1,2	53.75% (86/160)
	349-QEAVAAAAV-357	HLA-B*4001	-10	1,2,3	55.00% (88/160)
	406-KQQQGASEK-414	HLA-A*0301 HLA-A*3001	-8.3 -8.7	2	31.25% (50/160)
	448-IYAGVGAGL-456	HLA-A*2301 HLA-A*2402	-8.9 -9.2	1,2,3	38.12% (61/160)

Protein	MHC I Epitope (9 mer)	No. of alleles	Allele	Binding energy in ΔG (kcal mol^{-1})	Server	Conservancy (%)
Gilliam	458-HTYGKIDDK-466	3	HLA-A*0301 HLA-A*1101 HLA-A*6801	-9.7 -9.7 -10.5	1,2,3	16.25
	494-LEGSYMHSF-502	2	HLA-B*4402 HLA-B*4403	-9.3 -9.6	1,2,3	15.62
	514-LMASVGVRY-522	5	HLA-A*0301 HLA-A*3002 HLA-B*3501 HLA-A*0101	-10.8 -8.3 -12.3 -10.3	1,2,3	23.75
	316-NAFANQIQL-324	2	HLA-A*6802 HLA-B*3501 HLA-B*5101	-10.4 -9.4 -7.5	1,2,3	53.75
TA763	95-YLRNISAEV-103	4	HLA-A*02:03 HLA-A*02:01 HLA-A*02:06 HLA-B*08:01	-10.4 -10.4 -10.1 -8.1	1,2,3	36.88
	138-QPTIMPISI-146	3	HLA-B*07:02 HLA-B*51:01 HLA-B*53:01	-9.8 -10 -10.5	1,2,3	80.62
	320-NAFANQIQL-328	2	HLA-B*3501 HLA-A*6802	-9.4 -10.4	1,2,3	53.75
	352-QEAVAAAAV-360	3	HLA-B*4001 HLA-B*4403 HLA-B*4402	-10.9 -10.1 -10.6	1,2,3	55.00
	378-KLQRHAGIK-386	2	HLA-A*0301 HLA-A*3101	-11.1 -8	1,2,3	35.00
	451-IYAGVGAGL-459	2	HLA-A*2301 HLA-A*2402	-8.9 -9.2	1,2,3	38.12
	453-AGVGAGLAY-461	3	HLA-A*3002 HLA-B*1501 HLA-B*3501	-8.2 -8.7 -8.5	1,2,3	31.25
	493-YVDIEGSYM-501	3	HLA-A*0101 HLA-A*0206 HLA-B*3501	-10.1 -10.6 -8.1	2	28.12
	516-LMASVGVRY-524	6	HLA-A*0301 HLA-A*3002 HLA-B*3501	-10.8 -8.3 -12.3	1,2,3	23.75
TA716	81-MTIAPGFRA-89	4	HLA-A*0206 HLA-A*3001 HLA-A*6801 HLA-A*6802	-11 -11.2 -10.2 -11.1	1,2,3	45.00
	123-LQRPKFTPP-131	1	HLA-A*3001	-11.6	1,3	2.50
	132-QPTIMPISI-140	4	HLA-B*0702 HLA-B*5101 HLA-B*5301 HLA-B*5701	-9.8 -10 -10.5 -10.3	1,2,3	80.62

Protein	MHC I Epitope (9 mer)	Allele	Binding energy in ΔG (kcal mol^{-1})	Server	Conservancy (%)
	139-SIADRNLGV-147	HLA-A*0201 HLA-A*0203 HLA-A*0206 HLA-A*6802	-10 -10.2 -10.8 -11	1,2,3	2.50
	348-EAAAAAAVR-356	HLA-A*3301 HLA-A*6801 HLA-A*3301	-9.4 -10.9 -9.4	1,2,3	20.00
	373-KLKRHAGIK-381	HLA-A*0301 HLA-A*3001 HLA-A*3101	-10.1 -10.1 -9.9	1,2,3	15.00
ScaA	260-RSASIEMSI-268	HLA-A*3001 HLA-A*3201 HLA-B*5701 HLA-B*5801	-11.3 -8.1 -10 -10.3	1,2	9.09
	434-QSALTANIK-342	HLA-A*1101 HLA-A*6801	-8.2 -12.3	1,3	9.09
	527-GTPDSVMEK-535	HLA-A*1101 HLA-A*6801	-8.4 -7.2	1,2,3	9.09
	528-TPDSVMEKI-	HLA-B*5101 HLA-B*5301	-6.3 -10.4	1,2	9.09
	1008-SSSSGSSSV-1016	HLA-A*0203 HLA-A*6802	-8 -8.7	1	36.36
	1169-EPAPALQPI-1177	HLA-B*0702 HLA-B*5101 HLA-B*5301	-8.4 -9.2 -9.2	1,2,3	18.18
	1342-TAGHYFYPL-1350	HLA-A*0206 HLA-A*6802 HLA-B*0801	-10.3 -11.7 -11	1,3	63.64
	1360-PTVGIRHSY-1368	HLA-A*2601	-10.9	1,2	18.18
	1480-DVYLAEKYI-1488	HLA-A*6802	-10.8	1,2,3	100.00

Table 4
Binding energies for docked MHC-II-epitope complexes. The 27 HLA-II supertypes given below are known to have cumulative population coverages of ≥97 % for the five major ethnicities. The epitopes selected for the chimeric construct had good antigenic values and were positive for IFN-γ. Eleven strains of ScaA full length protein and 160 TSA strains were considered for conservancy analysis using the IEDB tool.

Protein	MHC II Epitope (15 mer)	Allele	Binding energy in ΔG (kcal mol-1)	Server	Conservancy (%)
Karp	36-AKVGVVGGMITGVES-50	HLA-DQA1*05:01/DQB1*03:01	−7.9	1,2	46.88
	84-PGFRAEIGVMYLTNI-98	HLA-DPA1*01:03/DPB1*04:01		2	2.27
		HLA-DPA1*02:01/DPB1*14:01	−10.2		
		HLA-DRB3*01:01	−7.9		
		HLA-DRB3*02:02	−12.6		
			−9.7		
	169-LNDEQRAAARIAWLK-183	HLA-DPA1*02:01/DPB1*14:01	−11.7	1	21.25
	205-NPILLNIPQGNPNPV-219	HLA-DRB3*02:02	−7	1	11.25
	346-QGQQQQAQATAQEAV-360	HLA-DQA1*03:01/DQB1*03:02	−9.2	1,2	43.75
		HLA-DQA1*01:02/DQB1*06:02	−9.2		
		HLA-DQA1*04:01/DQB1*04:02	−12.9		
	356-AQEAVAAAAVRLLNG-370	HLA-DPA1*02:01/DPB1*14:01	−9	1,2	45.62
		HLA-DQA1*01:02/DQB1*06:02	−7.8		
		HLA-DQA1*04:01/DQB1*04:02	−9.5		
		HLA-DQA1*05:01/DQB1*03:01	−10.1		
	491-AINAAEGVYVDIEGS-505	HLA-DQA1*05:01/DQB1*02:01	−9.5	1	26.88
		HLA-DQA1*03:01/DQB1*03:02	−12.5		
	518-SINPLMASVSVRYNF-532	HLA-DPA1*02:01/DPB1*14:01	−8	2	2.27
		HLA-DQA1*01:02/DQB1*06:02	−10.5		
		HLA-DRB1*03:01	−8.6		
		HLA-DRB1*07:01	−6.7		
		HLA-DRB1*09:01	−9.7		
		HLA-DRB1*12:01	−10.8		
		HLA-DRB1*13:02	−11.6		
		HLA-DRB3*02:02	−10		
Kato	37-AKVGVVGGMITGVES-51	HLA-DQA1*05:01/DQB1*03:01	7.9	1	46.88
	135-IMPISIADRDLGVDI-149	HLA-DRB4*01:01	−9.1	1,2	5.62
		HLA-DRB1*03:01	−8.9		
	341-QQAQATAQEAAAAAA-355	HLA-DQA1*0102/DQB1*0602	−9.3	1,2	10.62
	347-AQEAAAAAAVRVLNN-361	HLA-DQA1*05:01/DQB1*02:01	−8.6	1,2	10.62
		HLA-DQA1*05:01/DQB1*03:01	−9.6		
		HLA-DQA1*04:01/DQB1*04:02	−10.2		
		HLA-DQA1*01:02/DQB1*06:02	−9.2		
		HLA-DPA1*02:01/DPB1*14:01	−11.4		
	SFSIYAGLGAGLAYT	HLA-DQA1*05:01/DQB1*03:01	−11.9	2	27.27 % (24/88)
		HLA-DRB1*01:01	−12.7		
		HLA-DRB1*09:01	−14.9		
		HLA-DRB1*15:01	−13.8		
	SGALGVAINAAEGVY	HLA-DRB1*13:02	−11.3	1	70.00 % (112/160)
		HLA-DRB3*02:02	−9.5		
		HLA-DQA1*01:02/DQB1*06:02	−7.2		
Kato	488-AINAAEGVYVDIEGS-502	HLA-DQA1*05:01/DQB1*02:01	−9.5	1	26.88 % (43/160)
		HLA-DQA1*03:01/DQB1*03:02	−12.5		
		HLA-DQA1*04:01/DQB1*04:02	−11.9		
	513-KYSINPLMASFGVRY-527	HLA-DPA1*02:01/DPB1*14:01	−13.2	2	13.64 % (12/88)
		HLA-DRB1*09:01	−13.5		
		HLA-DRB1*12:01	−9.9		
		HLA-DRB1*13:02	−8		
		HLA-DRB1*15:01	−11.4		
		HLA-DRB3*02:02	−9.5		
Gilliam	152-GVDTDILAQAAAGQP-166	HLA-DQA10301DQB10302	−13.2	1	17.50 % (28/160)
		HLA-DRB40101	−7.8		
	338-QGQQQQAQATAQEAV-352	HLA-DQA10102DQB10602	−9.8	1	43.75 % (70/160)
		HLA-DQA10301DQB10302	−13		
	348-AQEAVAAAAVRLLNG-362	HLA-DQA10501-DQB10301	−10.1	1,2	45.62 % (73/160)
		DRB10901	−10.6		
		HLA-DPA10201DPB11401	−9		
		HLA-DQA10102DQB10602	−8.5		
		HLA-DQA10401DQB10402	−9.5		
		HLA-DR1	−11.6		
	444-ESFSIYAGVGAGLAH-458	HLA-DRB10901	−15.1	1	9.38 % (15/160)
		HLA-DRB11501	−11.6		
	477-SGALGVAINAAEGVY-491	HLA-DRB1*13:02	−11.3	1	70.00 % (112/160)
		HLA-DRB3*02:02	−9.5		
		HLA-DQA1*01:02/DQB1*06:02	−7.2		

(*continued on next page*)

Table 4 (*continued*)

Protein	MHC II Epitope (15 mer)	Allele	Binding energy in ΔG (kcal mol-1)	Server	Conservancy (%)
TA763	323-ANQIQLNFVMPQQAQ-337	DQA10501-DQB10301	−10.4	1,2	26.88 % (43/160)
		DRB1-0802	−10.9		
		DRB10901	−17.8		
		DRB11101	−10.4		
		DRB11201	−10.1		
		DRB11302	−11.3		
		DRB11501	−12.3		
		DRB30202	−12.3		
		DRB40101	−10.2		
		DRB50101	−11.2		
		HLA-DPA10201DPB10101	−12.6		
		HLA-DPA10301DPB10402	−13.3		
		HLA-DQA10101DQB10501	−10.2		
		HLA-DQA10102DQB10602	−7.9		
		HLA-DQA10501DQB10201	−11.7		
		HLA-DRB10101	−10.9		
		HLA-DRB10401	−7.5		
		HLA-DRB10405	−10.9		
		HLA-DRB10701	−16		
		HLA-DPA10103DPB10201	−13.5		
	342-GQQQQAQATAQEAVA-356	HLA-DQA10102DQB10602	−10.7	1,2	44.38 % (71/160)
		HLA-DQA10401DQB10402	−10.9		
		HLA-DQA10501DQB10301	−11.1		
	448-SVSIYAGVGAGLAYT-462	DQA10501-DQB10301	−11.8	1,2	5.62 % (9/160)
		DRB10901	−15.9		
		DRB11501	−10.7		
		HLA-DRB10101	−9.6		
TA716	113-DTRADTDSPILQRPK-127	HLA-DQA10501DQB10201	−9.1	1	2.50 % (4/160)
	340-QQAQATAQEAAAAAA-354	HLA-DQA1*0102/DQB1*0602	−9.3	1	10.62 % (17/160)
	176-LKDYAGIEYMVKDPN-200	HLA-DPA1*0201/DPB1*0501	−6.6	1	1.25 % (2/160)
ScaA	236-QNEGYGVQKVIIKQD-250	DQA10501-DQB10301	−12.8	1	9.09 % (1/11)
		HLA-DQA10102DQB10602	−9.2		
		HLA-DQA10301DQB10302	−12.6		
		HLA-DQA10401DQB10402	−10.8		
		HLA-DQA10501DQB10201	−11.1		
		HLA-DR1	−10		
	255-DKIEGRSASIEMSIE–269	DQA10501-DQB10301	−12.8	2	9.09 % (1/11)
		HLA-DQA10102DQB10602	−9.2		
		HLA-DQA10301DQB10302	−12.6		
		HLA-DQA10401DQB10402	−10.8		
		HLA-DQA10501DQB10201	−11.1		
		HLA-DR1	−10		
	1194-VAAISSIVQSRASNS-1208	DQA10501-DQB1030	−15.6	2	9.09 % (1/11)
		DRB1-0802	−11.4		
		DRB10901	−14.6		
		DRB11101	−10.3		
		DRB11201	−12		
		DRB11302	−11.3		
		DRB11501	−9		
		DRB30202	−9.8		
		DRB40101	−10.4		
		DRB50101	−10		
		HLA-DQA10102DQB10602	−8.3		
		HLA-DR1	−10.8		
		HLA-DRB10401	−9.9		
		HLA-DRB10405	−9.8		
		HLA-DRB10701	−7.5		

3.3. Designing multi-epitope chimeric vaccine constructs and molecular dynamics simulation

Using bioinformatically predicted epitopes of TSA56 and ScaA antigens, eight chimeric vaccine constructs were designed (Fig. S3). Briefly, construct 1 has linear B-cell epitope and construct 2 has both linear and discontinuous B-cell epitope of TSA56. Construct 5 has linear B-cell epitope of ScaA while construct 6 has both linear and discontinuous B-cell epitopes of ScaA (Fig. S3). These constructs were prepared with an idea to develop chimeric antigens for rapid diagnostic assays of scrub typhus. The multi-epitope vaccine construct 3 has linear B-cell epitope, CTL and HTL T-cell epitopes, while construct 4 has both linear, discontinuous B-cell epitope, MHC predicted epitopes of TSA56. Similarly, construct 7 has linear B-cell epitope, MHC-I, and MHC-II predicted epitopes, while construct 8 has both linear, conformational B-cell epitope, MHC-I, and MHC-II predicted epitopes of ScaA. The B- and T-cell epitopes joined with the appropriate linkers and adjuvants were included in the vaccine constructs to enhance immunogenicity. The modelled structure of

vaccine constructs was validated through a Ramachandran plot as described previously [71,72]. For the vaccine construct 4, 60 % of amino acid residues were observed in the favorable regions; the additional allowed regions comprised of 32 % residues, the generously allowed regions has a distribution of 5.8 % residues and 2.4 % in the disallowed regions (Figs. S4A–D). Similarly, for the vaccine construct 8, the favorable and additional allowed regions consisted of 67 % and 25 % amino acid residues, respectively, 4.7 % residues were clustered in the generously allowed regions and 3.3 % in the disallowed regions (Fig. S4). The structural validation of modelled structures of vaccine constructs were calculated using the Z-scores of ProSA web server [63–65].

The ClusPro server (https://cluspro.org) is widely used for studying the protein-protein interactions and this server was applied to analyze the interaction of predicted T-cell epitopes with the MHC alleles. The T-cell epitopes included in the vaccine constructs were selected based on the lowest binding energy scores. Similarly, we utilized the ClusPro server to perform the docking of fibronectin and vaccine constructs and the structures of the best docked conformations of simulated fibronectin-vaccine recombinant constructs obtained after 100 ns of MD simulation trajectory (Fig. 4A–D). MD trajectories were analyzed by observing the root mean square deviation (RMSD), root mean square fluctuation (RMSF) and radius of gyration (RoG) plots for three independent sets of simulation (Set-I, Set-II & Set-III) each simulated for 100 ns simulation time. RMSD plots of unligated fibronectin (1FNF) and docked fibronectin complexes 1FNF-Construct 3, 1FNF-Construct 4, 1FNF-Construct 7, 1FNF-Construct 8 are provided in Fig. 5. The average RMSD values were also calculated for all the complexes (Table 5). The mobility of the residues was analyzed by calculating the residue-wise RMSFs of unligated fibronectin (1FNF) and docked fibronectin complexes 1FNF-Construct 3, 1FNF-Construct 4, 1FNF-Construct 7, 1FNF-Construct 8 plotted in Fig. 6 and RMSFs of the individual constructs are graphed in Fig. S5. For Set-I, analysing the RMSD plots it was observed that fibronectin in presence of construct 3, 4, 7 exhibited lower deviations compared to construct 8. From the RMSF plots it was observed that construct 8 exhibited higher fluctuations from residue 500 onwards which might have attributed to the higher deviation of the complex 1FNF-Construct 8. Accordingly, fluctuations of construct-3, construct-4 and construct-7 were lower compared to construct-8 hence RMSDs of 1FNF in complex with construct-3, construct-4 and construct-7 were lower compared to 1FNF in complex with construct-8. However, for Set-II & Set-III, RMSDs of all the docked 1FNF complexes were fairly stable for both the sets of simulations. Compared to the unligated fibronectin there is not much difference in the fluctuations of individual residues of the docked fibronectins thereby indicating that the protein structure was intact throughout the simulation even after being bound to the constructs which has also been observed from the RoG plots. RoG plots for the complexes are presented in Fig. 7 and averaged RoG values are listed in Table 5. The average RoG values clearly depicted the overall stability of the complexes for both 1FNF as well as the constructs throughout the simulation time for all the three sets of simulation.

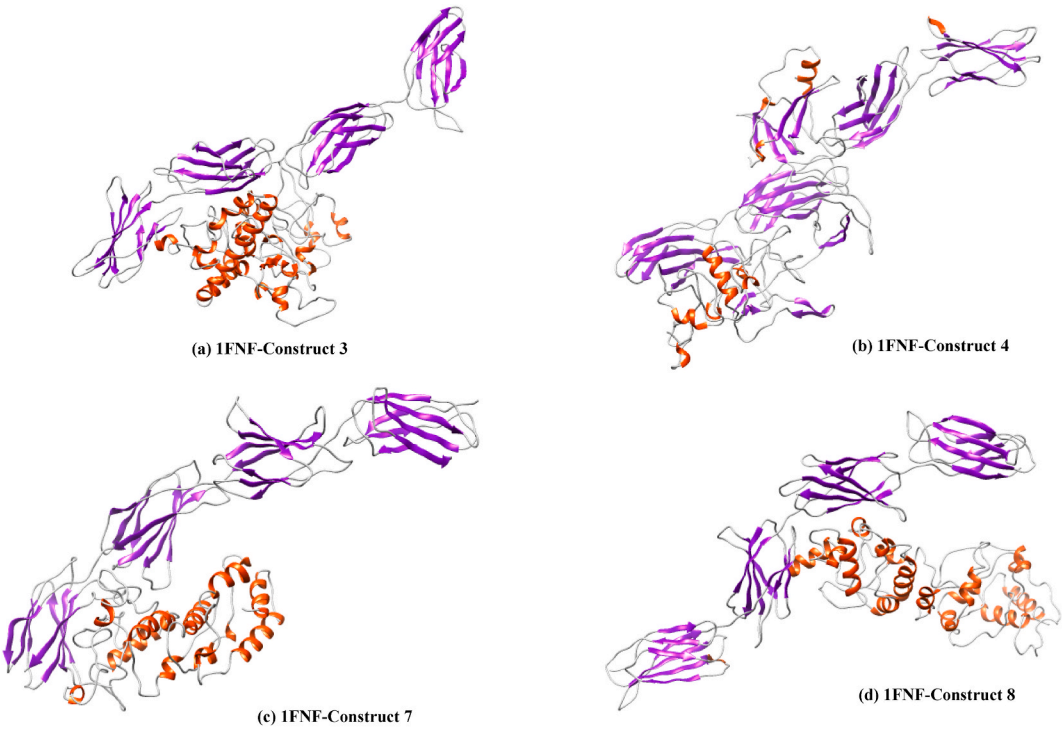


Fig. 4. Structures of the fibronectin (1FNF)-chimeric vaccine constructs (a) Docked structure complexes of 1FNF-Construct 3, **(b)** 1FNF-Construct 4, **(c)** 1FNF-Construct 7, and **(d)** 1FNF-Construct 8 obtained at the end of 100 ns simulation trajectory. Construct 3 has linear B-cell epitope, MHC-I, and MHC-II predicted epitopes of TSA56; Construct 4 has both linear, conformation B-cell epitope, MHC-I, and MHC-II predicted epitopes of TSA56. Construct 7 has linear B-cell epitope, MHC-I, and MHC-II predicted epitopes of ScaA; construct 8 has both linear, conformation B-cell epitope, MHC-I, and MHC-II predicted epitopes of ScaA. 1FNF: fibronectin (PDB ID:1FNF).

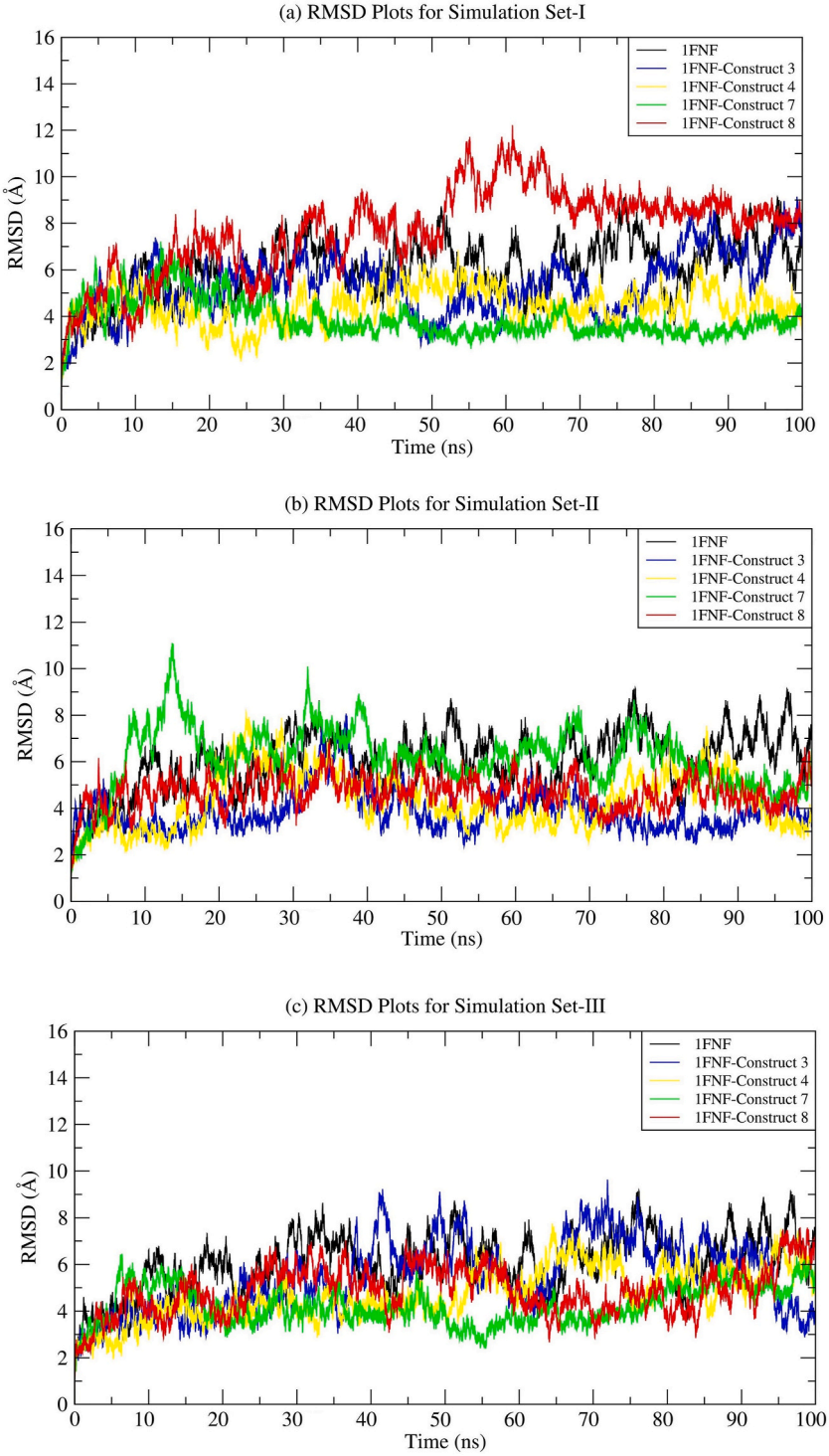


Fig. 5. Root mean square deviation (RMSD) plots of fibronectin (FNF) and multi-epitope vaccine constructs. RMSD plots of 1FNF (black), docked 1FNF complexes 1FNF-Construct 3 (blue), 1FNF-Construct 4 (yellow), 1FNF-Construct 7 (green) and 1FNF-Construct 8 (red). Simulations for the complexes (FNF-construct 3-blue colour, FNF-construct 4-yellow, FNF-construct 7-green, FNF-construct 8-red) were performed for three independent replicates for 100 ns simulation trajectory (a) Set-I, (b) Set-II, and (c) Set-III.

Table 5
The average root mean square deviation (RMSD) and radius of gyration (RoG) values of the unligated fibronectin and multi-epitope vaccine constructs. 1FNF, 1FNF-Construct 3, 1FNF-Construct 4, 1FNF-Construct 7, 1FNF-Construct 8 and the constructs 3, 4, 7, 8 for 3 sets of simulation each simulated for 100 ns simulation trajectory. 1FNF: fibronectin (PDB ID:1FNF).

	1FNF	1FNF-Construct 3		1FNF-Construct 4		1FNF-Construct 7		1FNF-Construct 8	
RMSD (in Å)		1FNF	Construct 3	1FNF	Construct 4	1FNF	Construct 7	1FNF	Construct 8
Set-I	6.03	5.40	4.27	4.43	7.68	3.90	5.30	7.71	5.89
Set-II		3.87	5.21	4.37	7.07	6.17	5.31	4.68	3.41
Set-III		5.48	5.41	4.75	8.22	4.25	4.41	4.89	4.01
	1FNF	**1FNF-Construct 3**		**1FNF-Construct 4**		**1FNF-Construct 7**		**1FNF-Construct 8**	
RoG (in Å)		1FNF	Construct 3	1FNF	Construct 4	1FNF	Construct 7	1FNF	Construct 8
Set-I	43.94	43.94	22.34	43.87	33.69	43.51	20.87	43.30	25.61
Set-II		43.48	22.51	43.37	34.10	42.49	21.15	43.65	25.16
Set-III		44.00	23.25	44.02	36.55	43.08	21.05	43.98	25.37

3.4. Immune simulation

The immune simulation of vaccine constructs was predicted to induce cytokine secretion including IFN-γ upon administration of antigens (Fig. S6). The antibody titer specific to the immunized antigen was much higher when compared to control antigen. Further the antigen specific immune response was followed by peak appearance of memory and plasma B-cell (Fig. S6). In silico immune simulation of the multi-epitope constructs of TSA56 and ScaA revealed high production of IgM antibodies thereby correlating with the observed immunity. The secondary and tertiary immune responses was evident by the increased B-cell sub-populations, IgG1 + IgG2, IgM, and IgG + IgM antibodies (Fig. S6).

4. Discussion

There is a need for developing a new and effective vaccine against scrub typhus due to its high incidence (over 1 million annual cases), wide geographic distribution, diagnostic challenges, lack of affordable point-of-care diagnostic methods, high case fatality rate, and resistance of some bacterium to commonly used antibiotics. Computational methods of identifying the immunogenic antigens can help in accelerating the design and preparation of recombinant vaccines prior to biological experiments. Using in-silico method, this study aimed to identify immunogenic antigens and epitopes of *O. tsutsugamushi* that can elicit humoral and T-cell immune reaction for developing an effective vaccine and sero-diagnostic assay as reported previously [52,73]. The knowledge of B-cell and T-cell epitopes binding regions in the antigen's sequence or structure is important in rational designing of vaccine [26,27,29], diagnostic tests [74], and immuno-therapy [75]. Therefore, identification of putative immunogenic epitopes with the help of in-silico methods can help in minimizing the time, cost, and resources associated with the experimental mapping of epitopes. However, with the advancement of machine learning (ML) has improved performance of in-silico prediction methods for mapping of B- and T-cell epitopes [32–38]. These prediction methods are commonly used for identifying viral, bacterial, and tumor-specific B-cell epitopes [76,77]. Therefore, the state-of-the-art B-cell epitope prediction methods with well-established acceptance and independent performance evaluations implicating the reliability of in-silico methods in identifying potential epitopes in rational design of vaccine [32–38,77]. TepiTool and NetMHCpan are commonly employed in predicting peptides that are specific in recognizing the MHC molecules required for T-cell mediated immune protection [78]. These methods are found useful in determining T-cell epitopes with optimum accuracy [79]. The integration of artificial intelligence in the MHC binding peptides prediction methods has overcome the challenges of T-cell epitope prediction due to existence of MHC polymorphism in the populations [80]. The deep learning integrated epitope prediction methods such as NetMHCpan-4.0 [81] and NetMHCIIpan-4.0 [44] have been successfully utilized in determining HLA-I and HLA-II epitopes of SARS-CoV-2. The algorithm of NetMHCpan-4.0 and NetMHCIIpan-4.0 prediction methods are developed using the dataset of HLA ligand elution assay coupled with mass spectrometry in isolating peptides pool that are processed and presented to MHC-restricted T-cells. The TepiTool algorithm has integrated the hundreds of alleles of humans, chimpanzees, bovines, gorillas, macaques, mice, and pigs in predicting high affinity MHC binding peptides [40]. The ClusPro server was utilized to select the top predicted T-cell

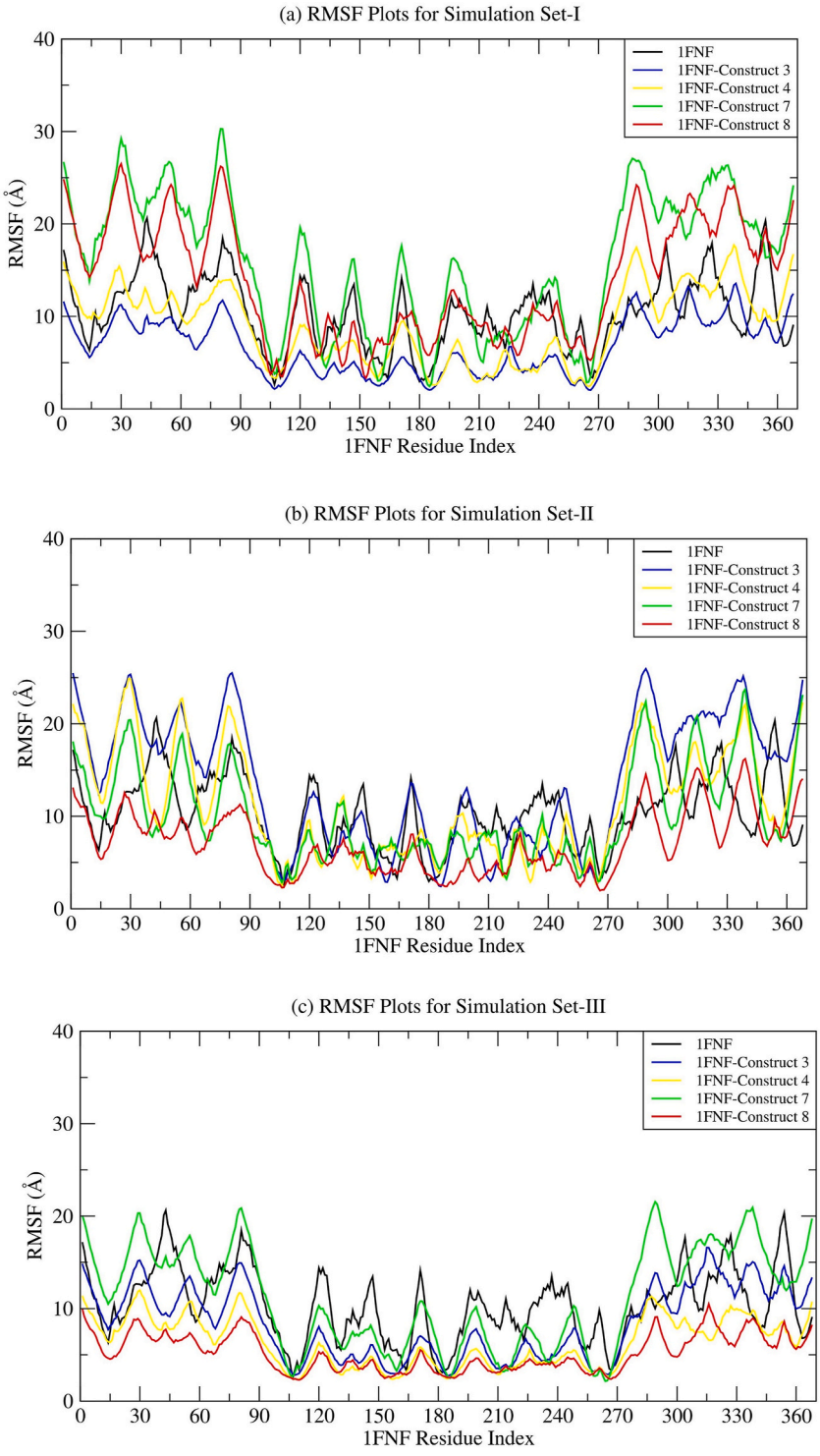


Fig. 6. Root mean square fluctuation (RMSF) plots of fibronectin (FNF) and multi-epitope vaccine constructs. RMSF plots of 1FNF (black), docked 1FNF complexes 1FNF-Construct 3 (blue), 1FNF-Construct 4 (yellow), 1FNF-Construct 7 (green) and 1FNF-Construct 8 (red). Simulations for the complexes (FNF-construct 3-blue, FNF-construct 4-yellow, FNF-construct 7-green, FNF-construct 8-red) were performed for three independent replicates for 100 ns simulation trajectory (a) Set-I, (b) Set-II, and (c) Set-III.

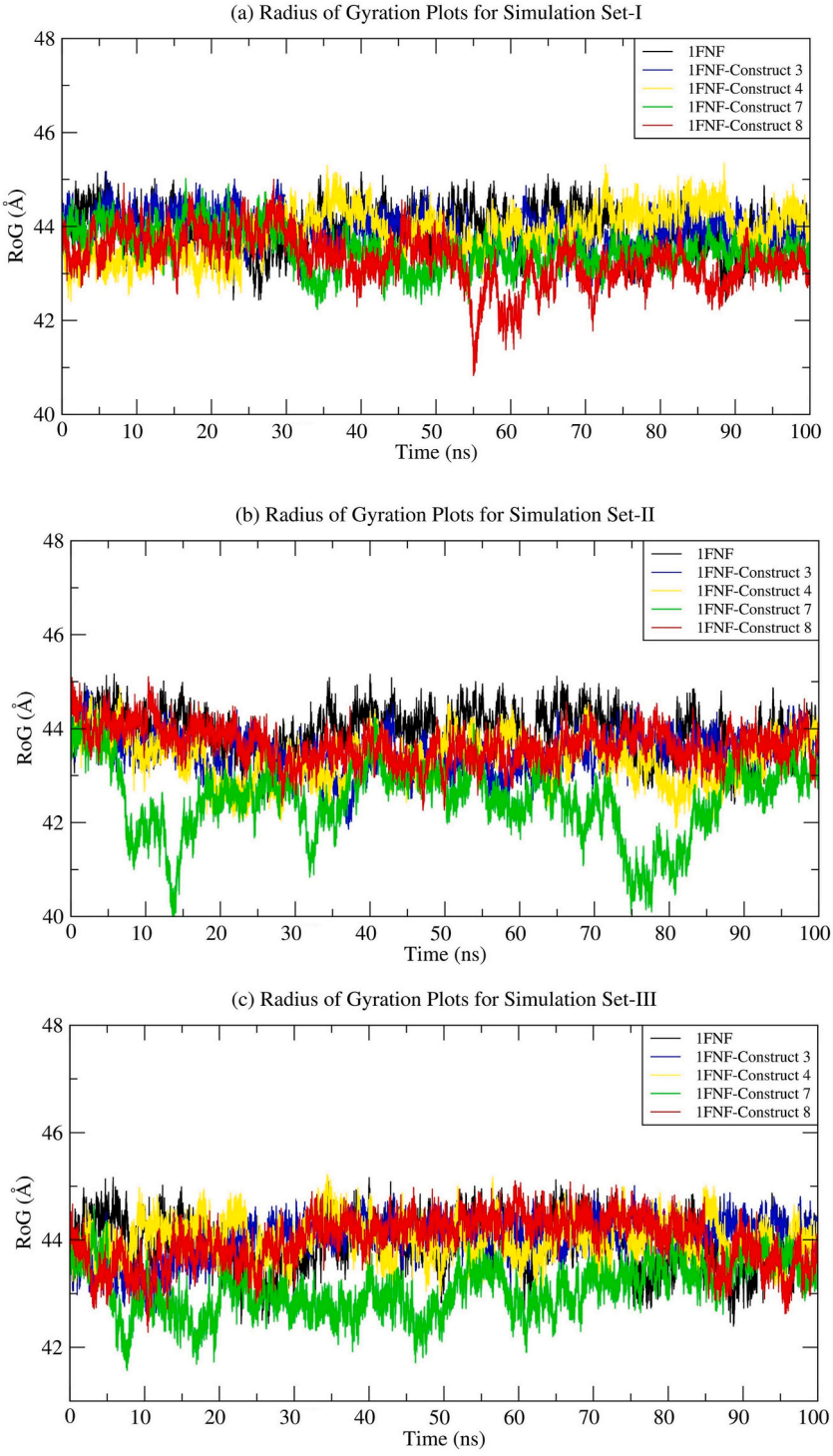


Fig. 7. Radius of gyration (RoG) plots of fibronectin (FNF) and multi-epitope vaccine constructs. Radius of Gyration plots of 1FNF (black), docked 1FNF complexes 1FNF-Construct 3 (blue), 1FNF-Construct 4 (yellow), 1FNF-Construct 7 (green) and 1FNF-Construct 8 (red). Simulations for the complexes (FNF-construct 3-blue colour, FNF-construct 4-yellow, FNF-construct 7-green, FNF-construct 8-red) were performed for three independent replicates for 100 ns simulation trajectory (a) Set-I, (b) Set-II, and (c) Set-III.

epitopes and the best docked conformations of simulated fibronectin-vaccine recombinant constructs in the MD simulation trajectory. The user-friendly ClusPro server is commonly applied in identifying protein-protein or ligand complexes based on the structures of the proteins and thermodynamic considerations to isolate reliable and specific protein–protein interactions to overcome the false positive interacting proteins [55,56,82].

TSA56, an immunodominant antigen, is considered a target for vaccines and serological assays; however, its genetic heterogeneity across *O. tsutsugamushi* strains limits cross-protective immunity [18]. Co-immunization of TSA56 and ScaA in mice has been reported to confer cross-protection, an observation yet to be replicated in humans [20,25]. To overcome scientific obstacles associated with the development of a vaccine and point-of-care test, a deep learning integrated B- and T-cell epitope mapping methods have been implemented and identified immunogenic epitopes in TSA56 and ScaA proteins of *O. tsutsugamushi*. The linear B-cell epitope of the Karp strain (112-VGETKADSVGGKDAPI-127) was previously shown to react with a Karp strain-specific monoclonal antibody [66]. Furthermore, Choi et al. [83] reported that VD I, ADs I, II, III, and IV were associated with homotypic and heterotypic antibody responses [67]. The VD within the antigenic region of TSA56 contributes to poor efficacy of vaccines and diagnostic tests [18]. Seong et al. [84] reported the development of humoral immunity correlating to AD of TSA56. The authors have found that antigenic fragments, ADs I and III, induced the production of IgG and IgM antibodies [83,84]. The immunogenicity of these epitopes may be experimentally verified using the convalescent serum of patients with scrub typhus and may be used for designing peptide-based rapid diagnostic tests. The re-emergence of scrub typhus in recent years has sparked global interest in identifying effective preventative and treatment solutions. The development of an effective vaccine is limited by several factors, including the antigenic diversity of TSA56, the short-lived heterologous strains immunity and the limited knowledge of immune protection against *O. tsutsugamushi*. The role of T-cell mediated immune protection including IFN-γ production in response to *O. tsutsugamushi* infection is reported previously [85,86]. The proliferation of $CD4^+$ and $CD8^+$ T-cells have been observed during the convalescent phase of *O. tsutsugamushi* infection [83,84]. Hence, we screened TSA56 and ScaA to find immunogenic HTL and CTL epitopes and found 51 and 27 high affinity binding peptides of MHC-I and MHC-II for TSA56, respectively. A positive correlation was seen between MHC binding peptides and the coverage of HLA alleles in the population. Antigenic, non-allergenic, IFN-γ-inducing T-cell epitopes with binding-free energy $\geq$ -6.3 kJ/mol were selected. The reliability of predicted T-cell epitopes was further confirmed by studying the interaction between the selected epitopes and MHC molecules. The lack of structures for all MHC alleles in the public repository of PDB limits the docking of predicted epitopes and MHC molecules. In a previous study, TSA56 cloning revealed that convalescent sera from scrub typhus patients contained antibodies against TSA56, and monoclonal antibodies were used to identify strain-specific epitopes [66,86]. The presence of unique and common T-cell epitopes in TSA56 and ScaA suggests that T-cell-based immunity is important in *O. tsutsugamushi* infection [85–90].

The mapping of epitope can help to fasten the cost-effective process of vaccine designing thereby complementing the various steps of biological experiments [91]. The mapping of epitopes is carried out based on the structure and amino acid sequence of vaccine candidate proteins. Hence, immunoinformatic approaches are utilized and credible in predicting immunogenic antigens and peptides that are subsequently validated in biological experiments. Despite all the advantages associated with in-silico methods, some of the disadvantages but not limited to non-availability of information, sub-optimal accuracy of algorithms, and inappropriate tools must be taken into consideration at the time of epitope mapping. The selection and application of methods appropriate for epitope mapping is important in identifying potential immunogens. For example, the B-cell epitopes predicted have 60–70 % accuracy due to the fact that the majority of the epitopes recognized by B-cells exist in the form of specific conformations [92,93].

Selecting an antigen that is conserved across strains is critical in designing an effective vaccine that can stimulate humoral and T-cell mediated immune responses against different strains of *O. tsutsugamushi*. Immunoinformatics tools have improvised the rational design, development and the quality of empirical investigation of constructed vaccine including determination of the conserved sequences. The conserved sequences of immunogenic antigens are known to induce a broad-spectrum of immune protection against the diverse strains of pathogens [94]. In-silico approaches were employed previously to identify immunogenic epitopes that are predicted to possess less toxic, allergenic, and cross-reactive epitopes to host antigens [95]. We designed four multi-epitope vaccine constructs comprising predicted B-cell and T-cell epitopes joined by appropriate linkers. The EAAAK rigid linker is reported to enhance the immunogenicity of the antigen due to helix inducing function [96]. AAY, KK, and GGGGS linkers are included in engineering a recombinant vaccine to provide flexibility to antigen domain function [97]. The characterization of bio-chemical parameters revealed immunogenicity, stability, and solubility of the vaccine constructs in an *E. coli* expression system. The RGD motif in the fibronectin and integrin α5β1 subunits is involved in bacterial uptake and invasion [19]. The findings of molecular docking showed specific and high affinity binding of designed vaccine to the human fibronectin receptor. MD simulation results suggested the stability of fibronectin-bound vaccine constructs 3, 4, and 7 in the cellular environment. The chimeric antigens may be used to develop rapid diagnostic tests for scrub typhus. Although the multi-epitope vaccine and predicted epitopes exhibited promising performance solely based on computational simulations. Therefore, to develop an effective vaccine against *O. tsutsugamushi*, necessary experiments are required to carry out using the predicted epitopes and vaccine constructs. We preferred and utilized the state-of-the-art pipeline ClusPro web server to validate the protein-protein interactions thereby limiting the scope for validation of protein-receptor complexes using other multiple modelling servers [55,56,82].

In conclusion, using *in-silico* approaches we identified several B- and T-cell epitopes of TSA56 and ScaA antigens of *O. tsutsugamushi* prototype strains. Our computational analysis suggests that the designed vaccine components and predicted epitopes are highly immunogenic, safe, non-allergenic, and stable. The experimental validation of epitopes reported in the present study may provide a better understanding of the immune response to *O. tsutsugamushi* infection.

Data availability statement

All the data used from different databases have been provided in the manuscript. Data will be made available on request.

CRediT authorship contribution statement

Anutee Dolley: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Himanshu Ballav Goswami:** Writing – review & editing, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Dikshita Dowerah:** Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Upalabdha Dey:** Software, Methodology, Data curation. **Aditya Kumar:** Software, Resources, Methodology. **Vanlal Hmuaka:** Resources, Methodology. **Rupak Mukhopadhyay:** Resources, Methodology. **Debasree Kundu:** Data curation. **George M. Varghese:** Writing – review & editing, Resources, Investigation. **Robin Doley:** Resources. **Ramesh Chandra Deka:** Software, Formal analysis. **Nima D. Namsa:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare no competing interests. AD and NDN have specific interest to experimentally validate some of the in silico predicted B- and T-cell epitopes reported in this study.

Acknowledgments

AD is a recipient of UGC-CSIR Senior Research Fellowship for PhD. This work was partly supported by the DBT-BIRAC, Government of India (Grant No. BIRAC/KIIT01858/BIG-SP/02/22).

Abbreviations

ELISA	enzyme-linked immunosorbent assay
POCT	point-of-care test
TSA56	56-kDa type-specific antigen
ScaA	surface cell antigen A
IEDB	Immune Epitope Database
CTL	cytotoxic T lymphocyte
HTL	helper T lymphocytes
MHC	major histocompatibility complex
RMSD	root-mean-square deviation
MD	molecular dynamics simulation
RMSF	root mean square fluctuation
RoG	radius of gyration
LPS	lipopolysaccharide
TAP	transporter associated with antigen processing
HLA	human leukocyte antigen
ML	machine learning
1FNF	fibronectin (PDB ID:1FNF)
I-TASSER	Iterative Threading ASSEmbly Refinement
AMBER	Assisted Model Building with Energy Refinement

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e23616.

References

[1] R. Tilak, R. Kunte, Scrub typhus strikes back: are we ready? Med. J. Armed Forces India 75 (2019) 8–17.
[2] T. Weitzel, et al., Endemic scrub typhus in south America, N. Engl. J. Med. 375 (2016) 954–961.
[3] D.H. Paris, et al., Real-time multiplex PCR assay for detection and differentiation of rickettsiae and orientiae, Trans. R. Soc. Trop. Med. Hyg. 102 (2008) 186–193.
[4] A.N. Maina, et al., Q fever, scrub typhus, and rickettsial diseases in children, Kenya, 2011–2012, Emerg. Infect. Dis. 22 (2016) 883–886.
[5] S.K. Chunchanur, Scrub typhus in India-an impending threat, Ann. Clin. Immunol. Microbiol. 1 (2018) 17–19.

[6] S. Stephen, et al., Scrub typhus in South India: a re-emerging infectious disease, J. Infect. Dis. 66 (2013) 552–554.
[7] K.P. Devi, M. Mae, S. Ningombam, Short Report an Outbreak of Scrub Typhus in Bishnupur, 2010, pp. 69–170.
[8] S.A. Khan, et al., Re-emergence of scrub typhus in northeast India, Int. J. Infect. Dis. 16 (2012) e889–e890.
[9] S.K. Mahajan, et al., Scrub typhus in himalayas, Emerg. Infect. Dis. 12 (2006) 1590–1592.
[10] K.S. Roopa, et al., Serodiagnosis of scrub typhus at a tertiary care hospital from southern India, J. Clin. Diagn. Res. 9 (2015). DC05–DC07.
[11] A. Luce-Fedrow, et al., A review of scrub typhus (orientia tsutsugamushi and related organisms): then, now, and tomorrow, Trav. Med. Infect. Dis. 3 (2018).
[12] E. Devasagayam, et al., The burden of scrub typhus in India: a systematic review, PLoS Neglected Trop. Dis. 15 (2021), e0009619.
[13] Y.J. Kim, et al., Improvement of the diagnostic sensitivity of scrub typhus using a mixture of recombinant antigens derived from Orientia tsutsugamushi serotypes, J. Kor. Med. Sci. 28 (2013) 672–679.
[14] K. Kannan, et al., Performance of molecular and serologic tests for the diagnosis of scrub typhus, PLoS Neglected Trop. Dis. 14 (2020), e0008747.
[15] A.Y. Bonell, et al., Estimating the burden of scrub typhus: a systematic review, PLoS Neglected Trop. Dis. 11 (2017) 1–17.
[16] G.D. Xu, et al., A review of the global epidemiology of scrub typhus, PLoS Neglected Trop. Dis. 11 (2017) 1–27.
[17] K. Pote, R. Narang, P. Deshmukh, Diagnostic performance of serological tests to detect antibodies against acute scrub typhus infection in central India, Indian J. Med. Microbiol. 36 (2018) 108–112.
[18] D.J. Kelly, et al., Scrub typhus: the geographic distribution of phenotypic and genotypic variants of orientia tsutsugamushi, Clin. Infect. Dis. 48 (2009). Suppl. 3.
[19] J.H. Lee, et al., Fibronectin facilitates the invasion of *Orientia tsutsugamushi* into host cells through interaction with a 56-kDa type-specific antigen, J. Infect. Dis. 198 (2008) 250–257.
[20] A. Ramaiah, M.C. Koralur, G.A. Dasch, Complexity of type-specific 56 kDa antigen CD4 T-cell epitopes of *Orientia tsutsugamushi* strains causing scrub typhus in India, PLoS One 13 (2018) 1–14.
[21] N. Ohashi, et al., Diversity of immunodominant 56-kDa type-specific antigen (TSA) of *Rickettsia tsutsugamushi*. Sequence and comparative analyses of the genes encoding TSA homologues from four antigenic variants, J. Biol. Chem. 267 (1992) 12728–12735.
[22] Cao, et al., Preparation of recombinant antigen of *O. tsutsugamushi* Ptan strain and development of rapid diagnostic reagent for scrub typhus, Am. J. Trop. Med. Hyg. 76 (2007) 553–558.
[23] G.M. Varghese, et al., Molecular epidemiology and genetic diversity of *Orientia tsutsugamushi* from patients with scrub typhus in 3 regions of India, Emerg. Infect. Dis. 21 (2015) 64–69.
[24] N.Y. Ha, et al., An autotransporter protein from *Orientia tsutsugamushi* mediates adherence to nonphagocytic host cells, Infect. Immun. 79 (2011) 1718–1727.
[25] N.Y. Ha, et al., Immunization with an autotransporter protein of *Orientia tsutsugamushi* provides protective immunity against scrub typhus, PLoS Neglected Trop. Dis. 9 (2015) 1–18.
[26] J.E. Epstein, et al., Live attenuated malaria vaccine designed to protect through hepatic CD8+ T cell immunity, Science 334 (2011) 475–480.
[27] M. Validi, et al., Immuno-informatics based approaches to design a novel multi epitope-based vaccine for immune response reinforcement against Leptospirosis, Mol. Immunol. 104 (2018) 128–138.
[28] K. Suzuki, et al., Evaluation of novel rapid detection kits for dengue virus NS1 antigen in Dhaka, Bangladesh, in 2017, Virol. J. 16 (2019) 102.
[29] S.F. Tosta, et al., Multi-epitope based vaccine against yellow fever virus applying immunoinformatics approaches, J. Biomol. Struct. Dyn. (2020) 1–17.
[30] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC Bioinf. 8 (2007) 1–7.
[31] E. Gasteiger, et al., The Proteomics Protocols Handbook, Proteomics Protoc, Handb., 2005, pp. 571–608.
[32] S. Saha, G.P.S. Raghava, BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 3239 (2004) 197–204.
[33] M.C. Jespersen, et al., BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, Nucleic Acids Res. 45 (2017) W24–W29.
[34] S. Saha, G.P.S. Raghava, Prediction of continuous B-cell epitopes in an antigen using recurrent neural network, Proteins 65 (2006) 40–48.
[35] J. Yang, Y. Zhang, I-TASSER server: new development for protein structure and function predictions, Nucleic Acids Res. 43 (2015) W174–W181.
[36] J. Ponomarenko, et al., ElliPro: a new structure-based tool for the prediction of antibody epitopes, BMC Bioinf. 9 (2008) 1–8.
[37] J.V. Kringelum, et al., Reliable B cell epitope predictions: impacts of method development and improved benchmarking, PLoS Comput. Biol. 8 (2012), e1002829.
[38] S. Liang, et al., EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results, BMC Bioinf. 11 (2010).
[39] V. Munoz, L. Serrano, Elucidating the folding problem of helical peptides using empirical parameters: III. Temperature and pH dependence, J. Mol. Biol. 245 (1995) 297–308.
[40] S. Paul, et al., TepiTool: a pipeline for computational prediction of T cell epitope candidates, Curr. Protoc. Im. 114 (2016) 18, 19.1-18.19.24.
[41] B. Reynisson, et al., NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data, Nucleic Acids Res. 48 (2020) W449–W454.
[42] B. Peters, et al., Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors, J. Immunol. 171 (2003) 1741–1749.
[43] S. Tenzer, et al., Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding, Cell. Mol. Life Sci. 62 (2005) 1025–1037.
[44] B. Reynisson, et al., 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.
[45] S. Southwood, et al., Several common HLA-DR types share largely overlapping peptide binding repertoires, J. Immunol. 160 (1998) 3363–3373.
[46] W. Fleri, et al., The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design, Front. Immunol. 8 (2017), https://doi.org/10.3389/fimmu.2017.00278.
[47] S. Sankaranarayanan, M. Mohkhedkar, V. Janakiraman, Mutations in spike protein T cell epitopes of SARS-COV-2 variants: plausible influence on vaccine efficacy, Biochim. Biophys. Acta, Mol. Basis Dis. 1868 (2022), 166432.
[48] S.K. Dhanda, P. Vir, G.P.S. Raghava, Designing of interferon-gamma inducing MHC class-II binders, Biol. Direct 8 (2013) 1–15.
[49] H.H. Bui, et al., Predicting population coverage of T-cell epitope-based diagnostics and vaccines, BMC Bioinf. 7 (2006) 1–5.
[50] I. Dimitrov, et al., AllerTOP v.2 - a server for in silico prediction of allergens, J. Mol. Model. 20 (2014) 6.
[51] R. Vita, et al., The immune epitope database (IEDB): 2018 update, Nucleic Acids Res. 47 (2019) D339–D343.
[52] M. Dangi, et al., Advanced in silico tools for designing of antigenic epitope as potential vaccine candidates against coronavirus, in: A. Shanker (Ed.), Bioinformatics: Sequences, Structures, Phylogeny, Springer, Singapore, 2018, pp. 329–357.
[53] Y. Shen, et al., Improved PEP-FOLD approach for peptide and miniprotein structure prediction, J. Chem. Theor. Comput. 10 (2014) 4745–4758.
[54] P. Thévenet, et al., PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides, Nucleic Acids Res. 40 (2012) W288–W293.
[55] D. Kozakov, et al., The ClusPro web server for protein-protein docking, Nat. Protoc. 12 (2017) 255–278.
[56] S. Vajda, et al., New additions to the ClusPro server motivated by CAPRI, Proteins: Struct., Funct., Bioinf. 85 (2017) 435–444.
[57] L.C. Xue, et al., PRODIGY: a web server for predicting the binding affinity of protein–protein complexes, Bioinf 32 (2016) 3676–3678.
[58] Y.D. Devi, et al., Exploring rotavirus proteome to identify potential B- and T-cell epitope using computational immunoinformatics, Heliyon 6 (2020), e05760.
[59] S. Srivastava, et al., Design of novel multi-epitope vaccines against severe acute respiratory syndrome validated through multistage molecular interaction and dynamics, J. Biomol. Struct. Dyn. 37 (2019) 4345–4360.
[60] A. Yano, et al., RGD motif enhances immunogenicity and adjuvanicity of peptide antigens following intranasal immunization, Vaccine 12 (2003) 237–243.
[61] B.A. Cho, et al., Intracellular invasion by *Orientia tsutsugamushi* is mediated by integrin signaling and actin cytoskeleton rearrangements, Infect. Immun. 78 (2010) 1915–1923.
[62] D.W.A. Buchan, et al., Scalable web services for the PSIPRED protein analysis workbench, Nucleic Acids Res. 41 (2013).

[63] 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.
[64] N. Akhtar, et al., Immunoinformatics-aided design of a peptide based multiepitope vaccine targeting glycoproteins and membrane proteins against monkeypox virus, Viruses 14 (2022) 2374.
[65] N. Akhtar, et al., Secreted aspartyl proteinases targeted multi-epitope vaccine design for *Candida dubliniensis* using immunoinformatics, Vaccines 11 (2023) 364.
[66] D.A. Case, et al., AMBER, University of California, San Francisco, 2018.
[67] J.A. Maier, et al., ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB, J. Chem. Theory & Comput. 11 (2015) 3696–3713.
[68] S. Miyamoto, P.A. Kollman, Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models, J. Comp. Chem. 13 (1992) 952–962.
[69] D.R. Roe, T.E. Cheatham III, PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data, J. Chem. Theory & Comput. 9 (2013) 3084–3095.
[70] N. Rapin, et al., Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system, PLoS One 5 (2010), e9862.
[71] Y.D. Devi, et al., Immunoinformatics mapping of potential epitopes in SARS-CoV-2 structural proteins, PLoS One 16 (2021), e0258645.
[72] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Stereochemistry of polypeptide chain configurations, J. Mol. Biol. 7 (1963) 95–99.
[73] S.J. Goodswen, P.J. Kennedy, J.T. Ellis, A guide to in silico vaccine discovery for eukaryotic pathogens, Brief Bioinf 14 (2013) 753–774.
[74] G.A. Schellekens, et al., The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide, Arthritis Rheum. 43 (2000) 155–163.
[75] H. Shirai, et al., Antibody informatics for drug discovery, Biochim. Biophys. Acta Proteins Proteom. 1844 (2014) 2002–2015.
[76] E. Raoufi, et al., Epitope prediction by novel immunoinformatics approach: a state-of-the-art review, Int. J. Pept. Res. Ther. 26 (2020) 1155–1163.
[77] K.A. Galanis, et al., Linear B-cell epitope prediction for in silico vaccine design: a performance review of methods available via command-line interface, Int. J. Mol. Sci. 22 (2021) 3210.
[78] A. Motmaen, et al., Peptide-binding specificity prediction using fine-tuned protein structure prediction networks, Proc. Natl. Acad. Sci. USA 120 (2023), e2216697120.
[79] W. Martin, H. Sbai, A.S. De Groot, Bioinformatics tools for identifying class I-restricted epitopes, Methods 29 (2003) 289–298.
[80] D.V. Desai, U. Kulkarni-Kale, T-cell epitope prediction methods: an overview, Methods Mol. Biol. 1184 (2014) 333–364.
[81] V. Jurtz, et al., NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data, J. Immunol. 199 (2017) 3360–3368.
[82] S.R. Comeau, S. Vajda, C.J. Camacho, Performance of the first protein docking server ClusPro in CAPRI rounds 3-5, Proteins, Structure, Function, and Bioinformatics 60 (2005) 239–244.
[83] M.S. Choi, et al., Homotypic and heterotypic antibody responses to a 56-kilodalton protein of Orientia tsutsugamushi, Infect. Immun. 67 (1999) 6194–6197.
[84] S.Y. Seong, et al., Mapping of antigenic determinant regions of the Bor56 protein of Orientia tsutsugamushi, Infect. Immun. 65 (1997) 5250–5256.
[85] W. Chierakul, et al., Differential expression of interferon-gamma and interferon-gamma-inducing cytokines in Thai patients with scrub typhus or leptospirosis, Clin. Immunol. 113 (2004) 140–144.
[86] C.K. Stover, et al., The 56-kilodalton major protein antigen of Rickettsia tsutsugamushi: molecular cloning and sequence analysis of the STA56 gene and precise identification of a strain-specific epitope, Infect. Immun. 58 (1990) 2076–2084.
[87] H. Iwasaki, et al., Correlation between the concentrations of tumor necrosis factor-alpha and the severity of disease in patients infected with *Orientia tsutsugamushi*, Int. J. Infect. Dis. 14 (2010) e328.
[88] M. de Fost, et al., Activation of cytotoxic lymphocytes in patients with scrub typhus, Am. J. Trop. Med. Hyg. 72 (2005) 465.
[89] H. Chu, et al., *Orientia tsutsugamushi* infection induces CD4+ T cell activation via human dendritic cell activity, J. Microbiol. Biotechnol. 23 (2013) 1159–1166.
[90] B.A. Cho, et al., Phenotypic characterization of peripheral T cells and their dynamics in scrub typhus patients, PLoS Neglected Trop. Dis. 6 (2012) e1789.
[91] H. Can, et al., In silico discovery of antigenic proteins and epitopes of SARS-CoV-2 for the development of a vaccine or a diagnostic approach for COVID-19, Sci. Rep. 10 (2020), 22387.
[92] Y. El-Manzalawy, V. Honavar, Recent advances in B-cell epitope prediction methods, Immunome Res. 3 (Suppl 2) (2010) S2.
[93] V. Yurina, O.R. Adianingsih, Predicting epitopes for vaccine development using bioinformatics tools, Ther. Adv. Vaccines Immunother. 21 (2022), 25151355221100218.
[94] A.T. Heiny, et al., Evolutionarily conserved protein sequences of influenza a viruses, avian and human, as vaccine targets, PLoS One 2 (2007), e1190.
[95] D. Jafari, et al., Designing a chimeric subunit vaccine for influenza virus, based on HA2, M2e and CTxB: a bioinformatics study, BMC Mol. Cell Biol. 21 (2020) 1–13.
[96] N. Hajighahramani, et al., Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*, Infect. Genet. Evol. 48 (2017) 83–94.
[97] S. Aslam, et al., Designing a multi-epitope vaccine against Chlamydia trachomatis by employing integrated core proteomics, immuno-informatics and in silico approaches, Biol. (Basel). 10 (2021) 997.