Molecular Diversity
https://doi.org/10.1007/s11030-022-10539-w

ORIGINAL ARTICLE

Designing a novel SOX9 based multi-epitope vaccine to combat metastatic triple-negative breast cancer using immunoinformatics approach

HemaNandini Rajendran Krishnamoorthy[1] · Ramanathan Karuppasamy[1]

Received: 29 August 2022 / Accepted: 28 September 2022

Abstract
Immunotherapies are a promising treatment option especially for the management of TNBC owing to its higher levels of tumour-associated antigens together with higher mutational load. Of note, the administration of preventive vaccines in the early stage of the cancer holds promise for effective disease management. Therefore, the present study aimed to develop a novel multi-epitope peptide-based vaccination against TNBC employing SOX9, which has recently been recognized as a key regulator of TNBC metastasis. The immunodominant regions from the SOX9 protein were computed and assessed based on their ability to elicit both T and B lymphocyte mediated responses. The resultant epitopes were fused using appropriate linkers (EAAAK, KK, AAY and GPGPG) and adjuvant (50S ribosomal protein L7/L12) to enhance the vaccine's immunogenicity. The physicochemical properties and population coverage were also anticipated for the constructed vaccine. Adding together, docking and dynamics simulation studies were performed on the modelled vaccine against TLR-4 to provide insight into the stability. Finally, the designed vaccine was cloned into the pET28 (+) vector and immunological simulation studies were carried out. These results demonstrate that our designed vaccine had the potency to trigger humoral and cellular immune responses. Based on these collective evidences, the final proposed vaccine could be an interesting therapeutics for the management of TNBC in the near future.

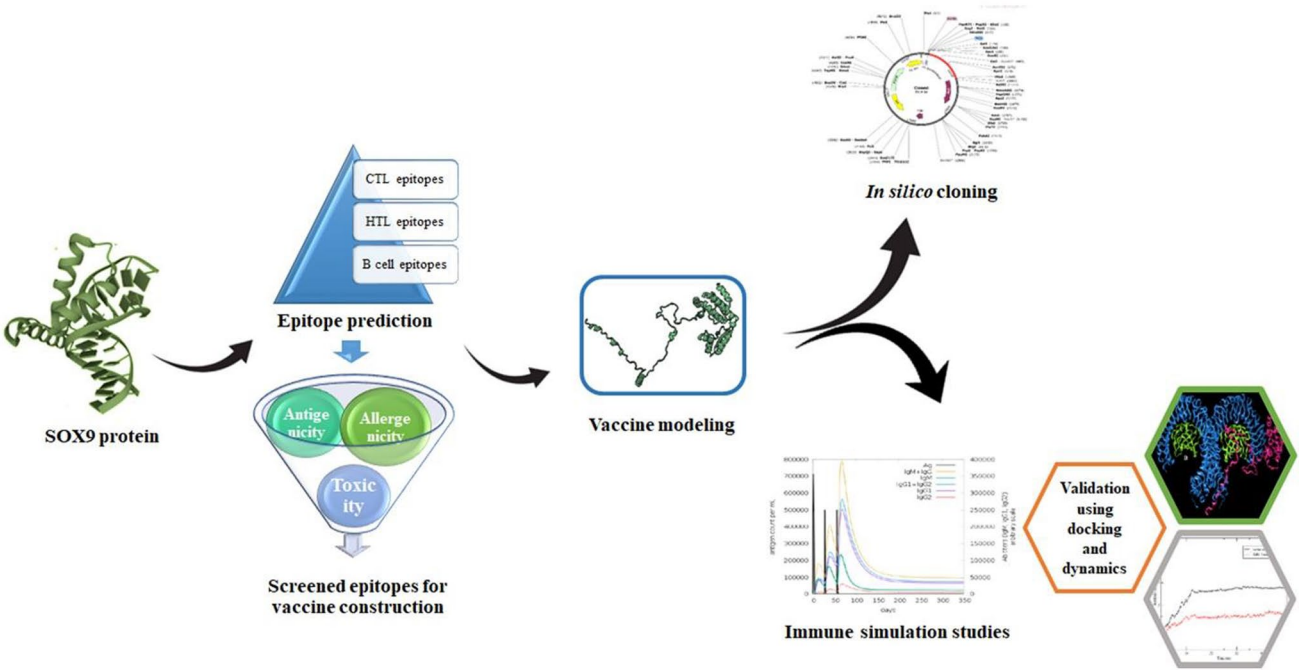
✉ Ramanathan Karuppasamy
kramanathan@vit.ac.in

[1] Department of Biotechnology, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu, India

Published online: 10 October 2022

Graphical abstract

Schematic representation of an efficient vaccine design framework by combining the range of immunoinformatics strategies.



Keywords Triple-negative breast cancer · SOX protein · In silico cloning · Dynamics · Immune simulation

Introduction

Triple-negative breast cancer (TNBC), which represents 10–30% of breast cancer cases, is characterized by the absence of the progesterone receptor, epidermal growth factor receptor 2 and oestrogen receptor [1]. Currently, surgery and chemotherapy are the primary treatment modalities for patients with TNBC [2]. However, the response to these options had limited efficacy and significant toxicity [3]. Furthermore, TNBC has also grown resistance to targeted and endocrine therapy due to molecular heterogeneity [4]. For these reasons, efficient treatment options and potential biomarkers are an unmet need for the management of TNBC.

Recent advances in the immune-landscape of the tumour microenvironment (TME) have gained insight into cancer immunotherapy [3]. Immunotherapy primarily aims to eliminate tumour cells by boosting both the cell and antibody mediated immune response, thereby reducing the probability of metastasis, relapse and resistance [5]. In TNBC, the TME is characterized by higher levels of lymphocyte infiltration and tumour mutational burden, suggesting that it would respond more likely to immunotherapy. Studies in the literature have shown that cancer subunit vaccines are a promising active immunotherapy approach that stimulate the CD 8+, CD 4+ and B cells to eliminate cancer cells. The main advantages of these subunit vaccines include minimal toxicity and specific adaptive immune response [6]. In fact, 45.5% of clinical trials for breast cancer vaccines are peptide-based. PVX-410, P10s-PADRE and TPIV200 are some of the chimeric vaccines for TNBC that are currently in clinical phase trials [7]. Furthermore, a plethora of studies are being conducted to identify potential peptide vaccines to combat TNBC. For instance, Kumar et al. proposed a subunit vaccine by identifying the potential cell surface markers for TNBC management [8]. Similarly, Mahdevar et al. and Parvizpour et al. designed novel peptide-based vaccines for cancer-testis antigens (BORIS, MAGE-A and NY-ESO-1) by fusing the immunodominant CTL and HTL epitopes to prevent TNBC [9, 10]. The most frequently explored tumour antigens to date for vaccine development include HER2 protein, cancer testis-antigen, mucin-1 (MUC-1) and carcino-embryonic antigen (CEA) [11]. However, there are countless overexpressed proteins that might be promising vaccine candidates.

Recently, ten transcription factors have been discovered to play a significant part in the progression of TNBC, including androgen receptor (AR), bromodomain 4 (BRD4), forkhead box C1 (FOXC1) homeobox protein engrailed-1 (EN1), MYC, myeloid zinc finger (MZF1), p53, sex-determining region Y-related HMG-Box (SOX) members, zinc finger

E-box binding homeobox 1 (ZEB1) and high mobility group AT-hook 1 (HMGA1). In particular, the SOX proteins SOX 9, SOX 10, SOX 2, SOX 4, SOX 8 and SOX 11 are reported to be overexpressed in TNBC. However, effective inhibition of these TFs remains elusive because of the absence of binding pockets and intracellular localization [12, 13]. Meanwhile, scientific advancement has drawn attention to the development of vaccination strategy using these transcription factors. For instance, Polakova et al. designed a DNA vaccine against the transcription factor SOX 2 which is essential for stem cell self-renewal and is also overexpressed in numerous cancer [14]. Similarly, Friedman et al. demonstrated the presence of immunodominant B and T lymphocyte epitopes from SOX 4 and suggested that could be used in a peptide-based vaccine against lung carcinoma [15]. Despite the fact that SOX9 has been identified as the key regulator of TNBC progression, no therapeutics approved for the treatment of patients who have SOX9 overexpression [16]. In light of this evidence, we implemented immunoinformatics approaches alongside simulation strategies to develop a novel multi-epitope chimeric vaccine that could be an effective therapy option for the management of TNBC.

Methodology

Retrieval of protein sequence

The amino acid sequence of the sex-determining region Y-related HMG-box transcription factor 9 (SOX9) (UniProt ID: P48436) was retrieved in FASTA format from the UniProt database. The protein sequence was submitted to an ensemble of computational tools for the prediction of the immunodominant epitopes.

Prediction of immunogenic epitopes

Prediction of T-cell epitopes

The NetCTL 1.2 online server was employed to anticipate MHC class I restricted (CTL) epitopes, which utilizes artificial neural networks (ANN) to predict MHC-peptide binding. In the current investigation, prediction was carried out using a peptide length of 9 mers and a threshold value of 0.75 against twelve MHC class I super types (A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58 and B62). The threshold is set to achieve highest prediction performance both in terms of sensitivity and specificity.

Similarly, the NetMHC II 2.3 server was used to estimate the HTL epitopes that bind to HLA-DQ, HLA-DP and HLA-DR alleles using artificial neural networks [17]. The epitopes were predicted using a 15-mer peptide length and threshold values of 2% for strong binders and 10% for weak binders, respectively.

Peptides of increased length have an increased affinity for MHC class II relative to shorter peptides for MHC class I. More precisely, MHC class I accommodate peptides of 8–10 residues. On the other hand, the open binding groove of MHC class II proteins typically accommodates peptides of 13–25 residues [18]. Given the importance of peptide-MHC interaction to the cellular immune response, we have utilized the default peptide length to retrieve the epitopes with greater affinity. NetCTL-1.2 also achieves the highest sensitivity of 0.75.

Prediction of continuous and discontinuous B-cell epitopes

The BCPREDS server, a SVM-based classifier with an accuracy of 74.57% was utilized for the prediction of linear B-cell epitopes [19]. In the present study, the SOX9 protein sequence was examined for potent B-cell epitopes with a default threshold value of 0.75 and an epitope length of 20 mers. Parallely, the Ellipro server computes a protrusion index (PI) value for each estimated conformational B-cell epitope by combining the Tornton's method with residue clustering algorithms [20, 21]. The PI value represents the proportion of protein atoms that are responsible for antibody binding and extend beyond the molecular bulk.

Assessment of immunogenic properties of the predicted epitopes

From the array of predicted peptides, potent antigenic regions were shortlisted depending on their allergenicity, toxicity and antigenicity characteristics. Initially, the antigenicity of the epitopes was assessed using the VaxiJen (version 2.0) server, which has an accuracy range of 70 to 89% at a threshold of 0.4 and virus as the target organism [22]. The default parameters were employed for the consistent and reliable prediction of protective antigens.

Similarly, the toxicity of the predicted epitopes was examined using the ToxinPred server with an SVM (Swiss-Prot) based method being employed [23]. Additionally, the AllerTOP v. 2.0 server, which is based on auto cross covariance (ACC) was used to evaluate the allergenicity of the epitopes [24]. It should be noted that not all HTL peptides are capable of inducing cytokine production. Hence IFNepitope a SVM-based server was utilized to estimate IFN-γ inducing HTL epitopes.

Population coverage analysis and multi–epitope vaccine construction

The development of peptide-based vaccines relies greatly on population coverage analysis [25]. The IEDB's population

coverage analysis tool was used with the default options to calculate the population coverage of potential epitopes and their HLA alleles. The final vaccine molecule was constructed using the selected CTL, HTL and B-cell epitopes. To enhance the immunogenicity, the 50S ribosomal protein L7/L12 (Locus RL7_MYCTU; Acc No. P9WHE3) was chosen as an adjuvant.

Assessment of immunological and physicochemical properties of the construct

The overall purpose of vaccination is to observe the post-injection activity of the designed vaccine in the body. Therefore, we utilized AllerTop (version 2.0) and Vaxijen (version 2.0) respectively to assess the antigenic and non-allergenic behaviour of the vaccine. In addition, various physicochemical properties including theoretical isoelectric point (pI), extinction coefficient, molecular weight, instability index, amino acid composition, aliphatic index and grand average hydropathy (GRAVY) were computed for the chimeric vaccine by the Expasy ProtParam server to gain insight into the physicochemical characteristics of final vaccine candidate [26].

Prediction and validation of secondary and tertiary structure

The PSIPRED online server, which works by incorporating neural networks and PSI-BLAST (Position-Specific Iterated-BLAST), was employed for the prediction of the alpha helices, beta sheets and coils in the designed vaccine [27].

The tertiary structure of the final vaccine sequence was anticipated using the ROBETTA server which generates three-dimensional structures based on deep learning methods. The modelled vaccine construct was then validated using the PROCHECK, ERRAT and VERIFY 3D tools from the SAVES v6.0 server. These servers evaluate the tertiary structure with different aspect ratios [28, 29].

Docking studies of designed vaccine and immune receptors (TLR-2 and TLR-4)

Molecular docking is a key tool to predict the predominant binding modes of a ligand with a protein of a known three-dimensional structure [30]. In the present study, the designed vaccine was docked against TLRs such as TLR-2 (PDB ID: 6NIG) and TLR-4 (PDB ID: 4G8A) using the ClusPro 2.0, a fully automated algorithm that performs rigid-body protein–protein docking. It filters docked conformations with good surface complementarity and ranks them based on their clustering properties [31]. The docked complex was then visualized using the PyMOL software and the interaction between the vaccine and the TLRs (TLR-2 and TLR-4) were analysed to gain insight into the binding pattern.

Molecular dynamic simulation

Molecular dynamics simulation was carried out using GROMACS v2020.1 with GROMOS force field to further confirm the stability of the developed vaccine. Initially, the system was solvated using simple point charge (SPC) water molecules in a 10 Å dodecahedron box which was configured using the editconf inbuilt tool of GROMACS. The total charge of the system was -28 which was neutralized by adding 28 sodium counter ions using the genion tool. Subsequently, the energy of the complex was minimized prior to MD run with 10 kJ/mol force constant using the Steepest Descent algorithm. The Particle-Mesh Ewald method was applied to explore the electrostatic interactions. The hydrogen bonds and the Van der Waals interactions are constrained and truncated by implementing LINCS algorithm. The Number of particles, volume and temperature (NVT) and Number of particles, pressure and temperature (NPT) ensembles were run for 50 ps to restrain the position for the heavy atoms. The system was then heated to 300 K, the optimal physiological temperature of the biological host system, with a lapsing time of 0.1 ps and pressure of 1 bar using the Berendsen algorithm. Finally, the vaccine and the vaccine-TLR-4 complex were subjected to 50 ns MD simulations with a 2 fs time step interval. The parameters such as root-mean-square deviation (RMSD), radius of gyration (Rg) and solvent accessible surface area (SASA) were evaluated using GROMACS utilities.

Codon optimization and in silico cloning of the vaccine construct

Back translation and codon optimization were done using the Java Codon Adaptation Tool (JCat) with *Escherichia coli* (K12 strain) as the expression host system for the protein sequence of the designed vaccine. The server computes two parameters important for assessment of protein expression level, namely, the codon adaptive index (CAI) and the GC content. The protein expression will be noble if the CAI value is 1.0 and GC content is in the range between 30 and 70%. Beyond this range an unfavourable effects on translation and transcriptional efficiencies will be observed [32]. The SnapGene software was utilized for in silico cloning of the codon optimized vaccine into the pET28 (+).

Immune simulation studies

C-ImmSim, an online simulation server was used to illustrate the immune response profile between the human immune system and the vaccine construct. The C-ImmSim

Table 1 Predicted CTL epitopes of SOX9 protein and its immunogenic properties

Peptides	Allele	Antigenicity	Allergenicity	Toxicity
GQVTYTGSY	A1, B27, B62	0.6997	Non-allergen	Non-toxic
NLPHYSPSY	A1, A26, B62	0.4475	Non-allergen	Non-toxic
AAGQGTGLY	A1, A26, B62	0.6651	Non-allergen	Non-toxic

portrays the humoral and cellular responses of a mammalian system against the constructed vaccine using position-specific scoring matrix (PSSM) and machine learning techniques, respectively [33]. For the analysis, three injections containing 1000 vaccine proteins with no LPS were given with a 4 week interval of 1, 84 and 168 steps. The remaining parameters, such as the random seed, simulation volume and number of antigens to inject, were left at their default settings with the exception of "Simulation Steps," which was set to 1050.

Results and discussion

Prediction and screening of CTL, HTL and B-cell epitopes

Cytotoxic T lymphocytes (CTL) and helper T lymphocytes (HTL) are key components of the adaptive immune system as they play a major role in acquiring long-term immunity [34, 35]. In the present study, a total 115 CTL epitopes and 303 HTL epitopes were predicted by the NetCTL 1.2 and NetMHC II 2.3 servers, respectively. Initially, the predicted epitopes were narrowed down by their ability to bind to at least three MHC I and MHC II alleles. Subsequently, these epitopes were assessed for antigenicity, toxicity, allergenicity and IFN-γ inducing property (only for HTL) using AllerTOP, VaxiJen, ToxinPred and IFNepitope servers, respectively. Finally as shown in Tables 1 and 2, 3 CTL epitopes (GQVTYTGSY, NLPHYSPSY and AAGQGTGLY) and 3 HTL epitopes (NIETFDVNEFDQYLP, GLYSTFTYMNPAQRP and GISSTAATPASAGHV) were selected for further analysis.

B-cell epitopes are antigenic regions to which antibodies bind to trigger an immune response [36]. Therefore, recognizing these epitopes in an antigen is essential for designing a chimeric vaccine. The BCPREDS server predicted 17 B-cell epitopes with scores ranging from 0.8 to 1.0. Only 10 of these epitopes were non-toxic, antigenic and non-allergenic (Table 3). Epitopes with an antigenic score of 1.0 (GEHSGQSQGPPTPPTTPKTD, GKADLKREGRPLPEGGRQPP and SEDSAGSPCPSGSGSDTENT) were considered for further analysis.

Assembling of multi-epitope vaccine construct

The chosen CTL, HTL and linear B-cell epitopes were fused to form a robust vaccine construct using appropriate linkers and adjuvants. As a result, the final vaccine consisted of three CTL, three HTL and three B-cell epitopes with a final length of 295 amino acids (Fig. 1). In order to boost the immunogenicity, 50S ribosomal protein was used as an adjuvant at the N-terminal end of the vaccine using EAAAK linker. EAAAK is a rigid α-helix peptide linker that provides efficient functional domain separation in fusion proteins [37].

The CTL epitopes were joined using the AAY linkers. In mammalian cells, the Ala-Ala-Tyr (AAY) linker acts as the proteasomes' restriction site. As a result, epitopes coupled by an AAY linker are efficiently separated within the cells. Moreover, the vaccine construct's immunogenicity is also enhanced by the AAY

Table 2 Predicted HTL epitopes of SOX9 protein and its immunogenic properties

Peptides	Alleles	Immunogenic properties			
		Antigenicity	Allergenicity	IFN-γ	Toxicity
NIETFDVNEFDQYLP	HLA-DPA10103, HLA-DPB10402, HLA-DQA10101, HLA-DQB10501, HLA-DQA10102, HLA-DQB10502	0.4356	Non-allergen	Inducer	Non-toxic
GLYSTFTYMNPAQRP	HLA-DQA10102, HLA-DQB10602, HLA-DQA10601, HLA-DQB10402, HLA-DRB1_0401, HLA-DRB1_1001, HLA-DRB5_0101	0.5122	Non-allergen	Inducer	Non-toxic
GISSTAATPASAGHV	HLA-DQA10102, HLA-DQB10602, HLA-DQA10103, HLA-DQB10603, HLA-DQA10201, HLA-DQB10301, HLA-DQA10201, HLA-DQB10303, HLA-DQA10201, HLA-DQB10402, HLA-DQA10501, HLA-DQB10301, HLA-DQA10501, HLA-DQB10302, HLA-DQA10501, HLA-DQB10303	0.5486	Non-allergen	Inducer	Non-toxic

Table 3 Predicted linear B-cell epitopes of SOX9 protein

Protein	Epitope	Score	Vaxijen score	Allergenicity	Toxicity
SOX9	*GEHSGQSQGPPTPPTTPKTD	1	0.7068	Non-allergen	Non-toxic
	*GKADLKREGRPLPEGGRQPP	1	0.6941	Non-allergen	Non-toxic
	*SEDSAGSPCPSGSGSDTENT	1	0.8349	Non-allergen	Non-toxic
	QENTFPKGEPDLKKESEEDK	0.999	0.4860	Non-allergen	Non-toxic
	PFMKMTDEQEKGLSGAPSPT	0.991	0.8493	Non-allergen	Non-toxic
	NESEKRPFVEEAERLRVQHK	0.987	0.7546	Non-allergen	Non-toxic
	TLVPMPVRVNGSSKNKPHVK	0.986	0.7353	Non-allergen	Non-toxic
	TTLSSEPGQSQRTHIKTEQL	0.949	0.7346	Non-allergen	Non-toxic
	QPRRRKSVKNGQAEAEEATE	0.933	1.1107	Non-allergen	Non-toxic
	SEQQQHSPQQIAYSPFNLPH	0.889	0.5050	Non-allergen	Non-toxic

*Epitopes considered for vaccine design

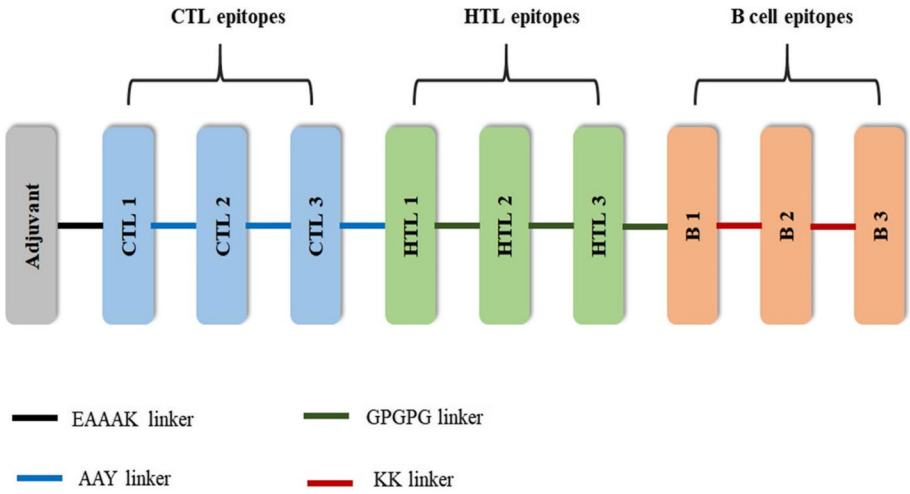


Fig. 1 The structural representation of the final vaccine construct

linker [37]. Similarly, the GPGPG linkers, which have the potential to trigger HTL responses necessary for a synthetic vaccine, were utilized to combine the CD4+ epitopes [38]. Finally, to incorporate the B-cell peptides KK linkers were utilized, which are also capable of enhancing the immunogenicity [39].

Evaluation of population coverage

The peptide binding domains of MHC class I and MHC class II restricted HLA alleles are polymorphic, implying that they possess different binding specificities [40]. As a result, determining population coverage for predicted antigenic regions is an essential part of developing a multi-epitope vaccine. According to Table 4 and Fig. 2a, the estimated global population coverage for the shortlisted epitopes is 99.65% across 109 countries. Similarly, the population coverage across Indian population is found to be 99.83% as shown in Fig. 2b. In a nutshell, the findings indicate that the selected immunodominant peptides could be used to develop a synthetic vaccine that could be administered to a larger population.

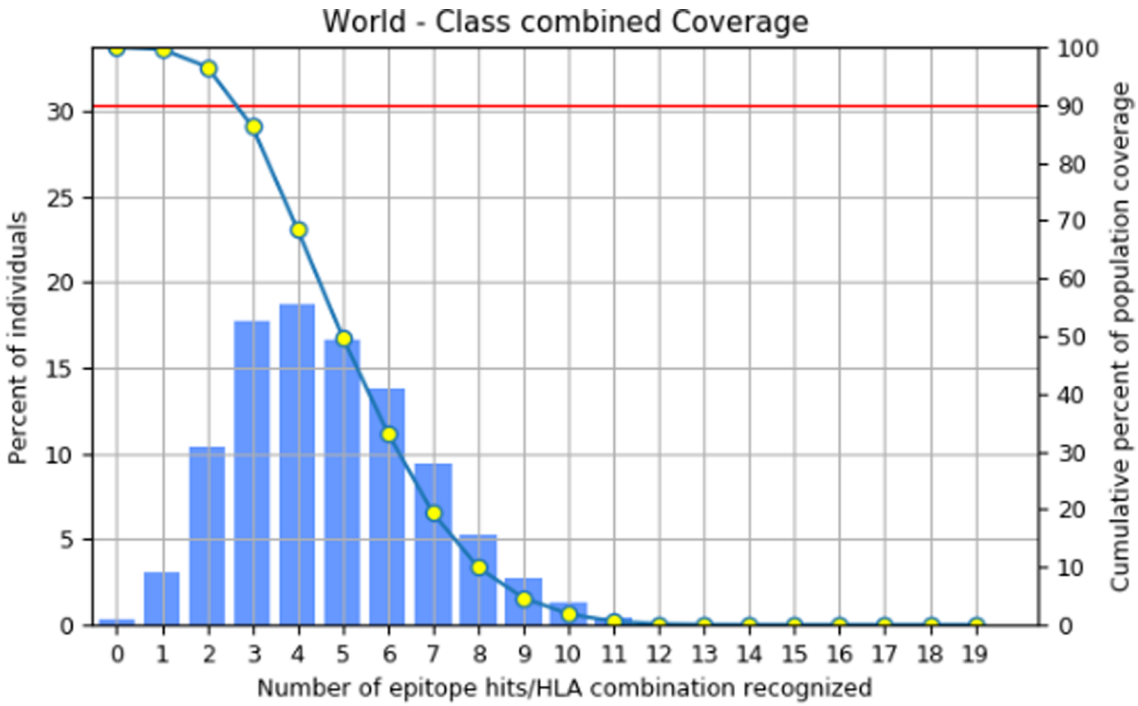
Table 4 Estimated population coverage score for Indian and world population

Population/area	Class combined		
	Coverage[a]	average_hit[b]	pc90[c]
India	99.83%	4.72	2.77
World	99.65%	4.7	2.63
Average	99.74	4.71	2.7
Standard deviation	0.09	0.01	0.07

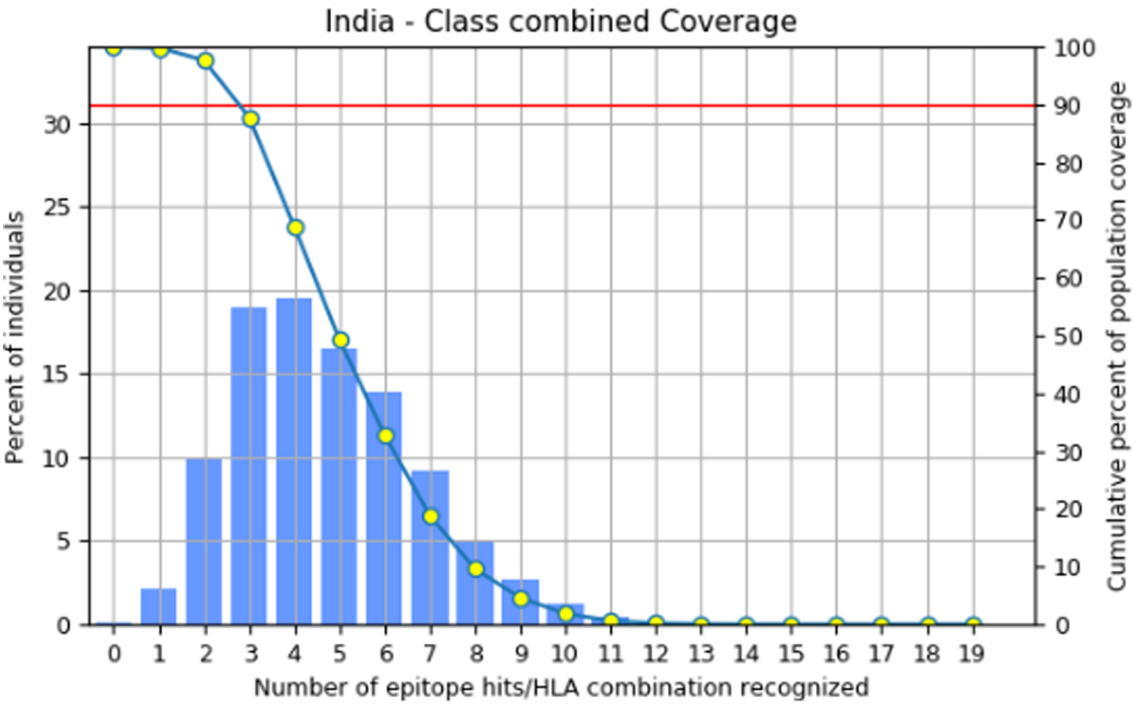
[a]Projected population coverage

[b]Average number of epitope hits/HLA combinations recognized by the population

[c]Minimum number of epitope hits/HLA combinations recognized by 90% of the population



(a)



(b)

Fig. 2 Population coverage analysis **a** Worldwide, **b** Indian population

Table 5 Analysis of physicochemical properties of the vaccine construct

Physicochemical parameters	Value
Molecular weight (kDa)	30.14
No. of amino acids	295
Theoretical pI	4.91
Instability index	39.84
Aliphatic index	65.36
GRAVY	− 0.408
Antigenicity	0.5945 (probable antigen)
Allergenicity	Non-allergen

nature of the vaccine. The values of the physicochemical properties are tabulated in Table 5.

Modelling and validation of secondary and tertiary structure

Secondary structures are important in protein folding and structure [41]. The PSIPRED server was used to calculate the percentage of secondary structure, which included 33.56% (alpha-helix), 12.20% (extended strand) and 57.24% (coils). The result is shown in Fig. S1. The tertiary structure of the designed vaccine was modelled by ROBETTA server. Out of 5 models predicted by the server, model 2 was chosen based on the model evaluation performed using SAVES server. The modelled tertiary structure was assessed by PROCHECK, VERIFY 3D and ERRAT provided by the SAVES servers to determine the possible errors and reliability of the model. According to the Ramachandran plot, 90.5% of the residues are in the favoured region (Fig. S2a). In addition, the overall quality factor of the model is found to be 94.50. The modelled 3D structure is visualized using PyMOL (Fig. S2b).

Table 6 Predicted discontinuous B-cell epitopes

S. no.	Residues	Number of residues	Score
1	SQGPPTPPTTPKTDKKGKADLKREGRPLPEGGRQPPKKSEDSAGSPCPSGSGSDTENT	58	0.87
2	MAKLSTDELLDAEMDFVKKFEETEVTAAAPV	31	0.686
3	QEFILEAAGDKKIGVGAPKPVAKEADEAKAKLEAAGATVTVKEAAAKGQ	49	0.604
4	SSTAATPASAGV	12	0.584
5	IET	3	0.537
6	FDVN	4	0.523

Prediction of immunogenic property, physicochemical property and secondary structure of the designed vaccine

The constructed chimeric vaccine was subjected to allergenicity, toxicity and antigenicity evaluation. The designed vaccine candidate had an antigenicity score of 0.5945, indicating that it is a probable antigen capable of eliciting an immune response. Also, the vaccine construct is found to be non-toxic and non-allergic thus proving its safety. Adding together, the Expasy Protparam online server was utilized to explore the vaccine's other physical and chemical properties. It is evident that the final synthetic vaccine is composed of 295 amino acids and has a molecular weight of 30.14 kDa which is ideal for vaccine construction and purification during experimental evaluation. The instability index of the constructed vaccine is found to be 39.84, implying its high stability. Further, the aliphatic index and the theoretical pI are observed to be 65.36 and 4.91, respectively, reflecting its high thermostability and acidic nature. The designed vaccine has an estimated half-life of 30 h in mammalian reticulocytes (in vitro), > 20 h in yeast (in vivo) and > 10 h in *E. coli* (in vivo). The GRAVY (grand average of hydropathicity) score is found to be − 0.408 denoting the hydrophilic

Prediction of conformational B-cell peptides

The discontinuous B-cell epitopes are regions in the antigen that are commonly recognized by the antibodies, which play a substantial role in antibody-mediated immunity. The ElliPro server was employed to discover the non-sequential B-cell epitopes from the constructed vaccine. As illustrate in Fig. S3, the algorithm estimated 6 conformational B-cell epitopes in our proposed vaccine with scores ranging from 0.87 to 0.54 (Table 6) indicating that these regions will be distinguished by antibodies.

Molecular docking

Understanding the molecular details of antigen detection is a crucial and effective step for the successful development of

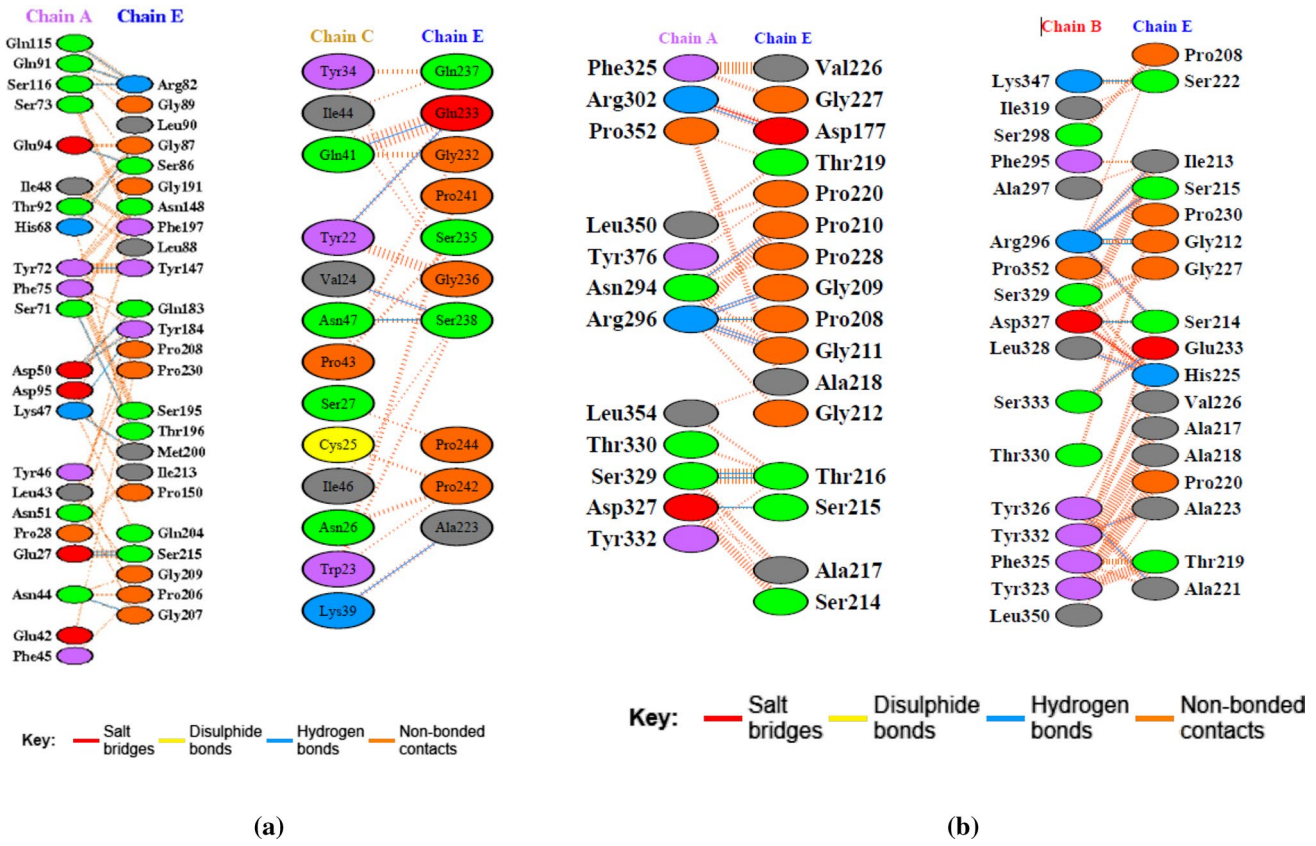


Fig. 3 **a** Interacting residues between chain A and chain C of TLR-4 with vaccine (chain E). The hydrogen bonds are represented in blue colour. **b** Interacting residues between chain A and chain B of TLR-2 with vaccine (chain E). The hydrogen bonds are represented in blue colour

peptide-based vaccinations. Toll-like receptors (TLRs) are essential for identifying particular patterns in host molecules and also play a significant role in innate immunity during cancer. According to reports, TNBC cells exhibits nearly tenfold less TLR-2 expression than ER/PR+ cells [42]. Similarly, TLR-4 was also reported to be less expressed in TNBC patient compared to luminal and HER2+ patients [43]. These data suggest that the low expression of TLR-2 and TLR-4 may aid in immune evasion and tumour growth in TNBC. Therefore, keeping these facts in mind, we carried out a molecular docking study to assess how the designed vaccine will interact with these immune molecules. The docking results showed that the proposed vaccine significantly binds to the TLR-4 (− 1074.3 kcal/mol) than the TLR-2 receptor (− 981.2 kcal/mol). Thus, we selected TLR-4 for further analysis.

The PyMOL software was used to view the bound complex of TLR-2 and TLR-4 with vaccine (Figs. S4 and S5) and the binding residues between the vaccine-TLR-2 and vaccine-TLR-4 were visualized using the PDBSum. Based on the interaction analysis report, it is observed that the designed subunit vaccine formed 15 hydrogen bonds with TLR-4 chain A and 5 hydrogen bonds with TLR-4 chain C (Fig. 3a). On the contrary, it exhibits 10 hydrogen bonds with TLR-2 chain A and 12 hydrogen bonds with TLR-2 chain B (Fig. 3b). Of note, the binding of vaccine with TLR-4 protein was facilitated by highest number of non-bonded contacts than with TLR-2 protein. Although these non-bonded contacts are weaker, collectively it could support the binding of vaccine construct with the immune molecule. These interacting patterns implicate that the designed vaccine has a strong affinity and structural stability against TLR-4 than TLR-2 protein. It is worth noting that docking score correlates favourably with the available literature information and thus validates the docking analysis. For instance, the literature highlights TLR-4 recognition by 50S ribosomal protein L7/L12 and causes DC maturation which further results in the release of pro-inflammatory cytokines (TNF-alpha, IL-1beta and IL-6). The evidences also suggest that naïve T cells could also be stimulated by 50S ribosomal protein L7/L12-activated DCs, which could also efficiently polarize CD4+ and CD8+ T cells to produce IFN-gamma and trigger T cell-mediated cytotoxicity [44].

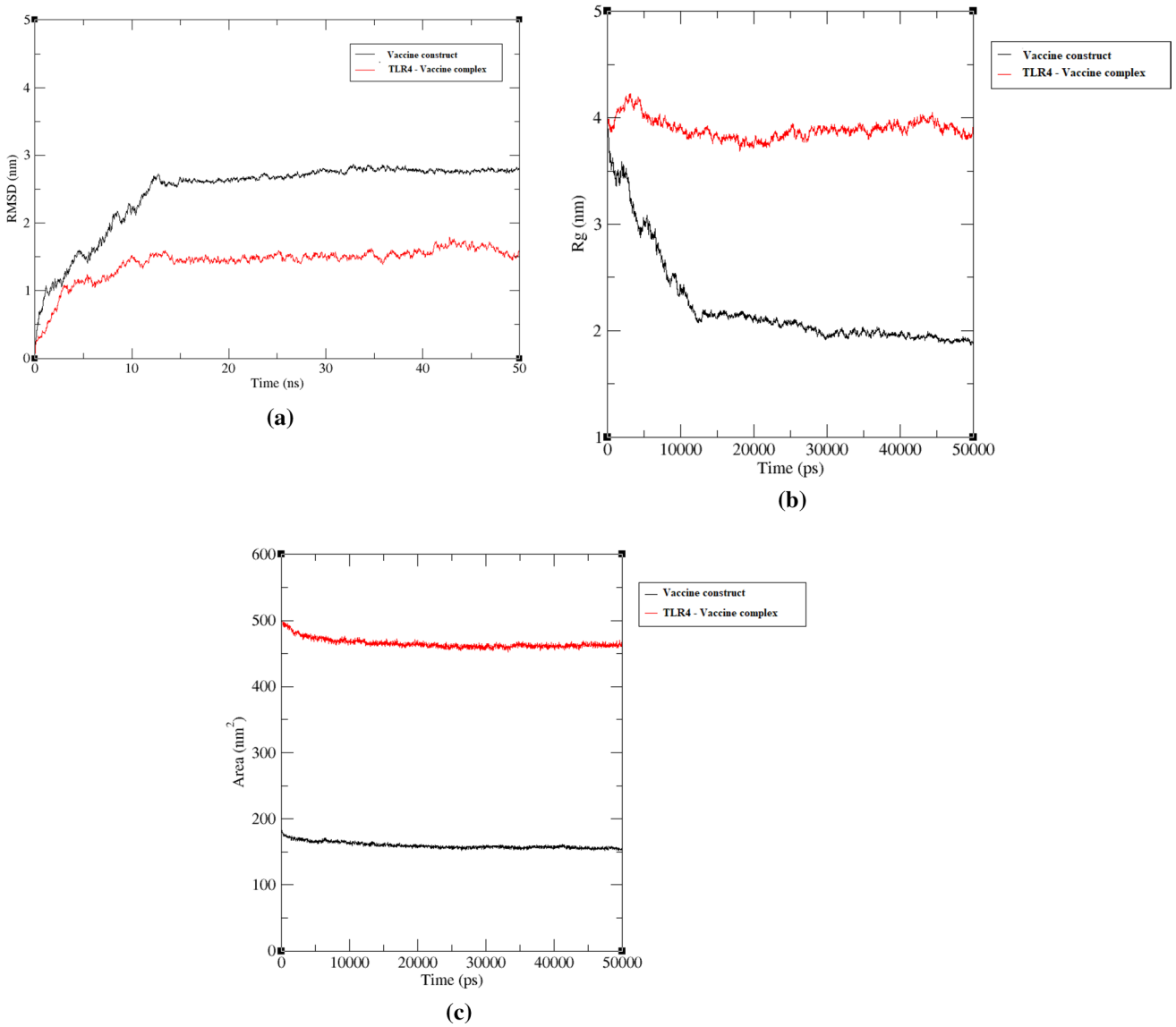


Fig. 4 Molecular dynamics simulation analysis **a** RMSD graph of TLR-4—vaccine complex. **b** Radius of gyration of TLR-4—vaccine complex and **c** solvent accessible surface graph of TLR-4—vaccine complex

Molecular dynamics simulation

A molecular dynamics simulation was executed for 50 ns using both vaccine-TLR-4 complex and vaccine in order to validate the binding mode and stability of the complex. The results were analysed using parameters such as RMSD, Rg and SASA to gain insights into the stability at molecular level. Initially, the vaccine's RMSD value increased sharply, reaching nearly 2.78 nm in 12 ns. Between 12 and 15 ns, minimal fluctuations were observed. Around 50 ns, the RMSD value reached to 2.5 nm. Similarly, the complex's RMSD value increased in a positive trend, reaching 1.5 nm at 12 ns. Afterwards, the RMSD value dropped to 1.4 nm around 14 ns and continued till the end of the simulation. Therefore, from Fig. 4a it can be observed that the vaccine-TLR-4 complex exhibited a lower RMSD value, which could be beneficial for structural stability.

Radius of gyration (Rg) is used to characterize the flexibility of the receptor-ligand complex. A small Rg value illustrates that the proteins are in folded state during the simulation. As depicted in Fig. 4b, the average Rg score for the construct was found to be 1.90 nm and the average Rg score for the vaccine-TLR-4 complex was 3.8 nm. These results indicate that the designed vaccine remained compact during the simulation.

Solvent accessible surface area (SASA) is used to determine the surface area that can be accessed by solvent molecules. The average SASA values of our vaccine and

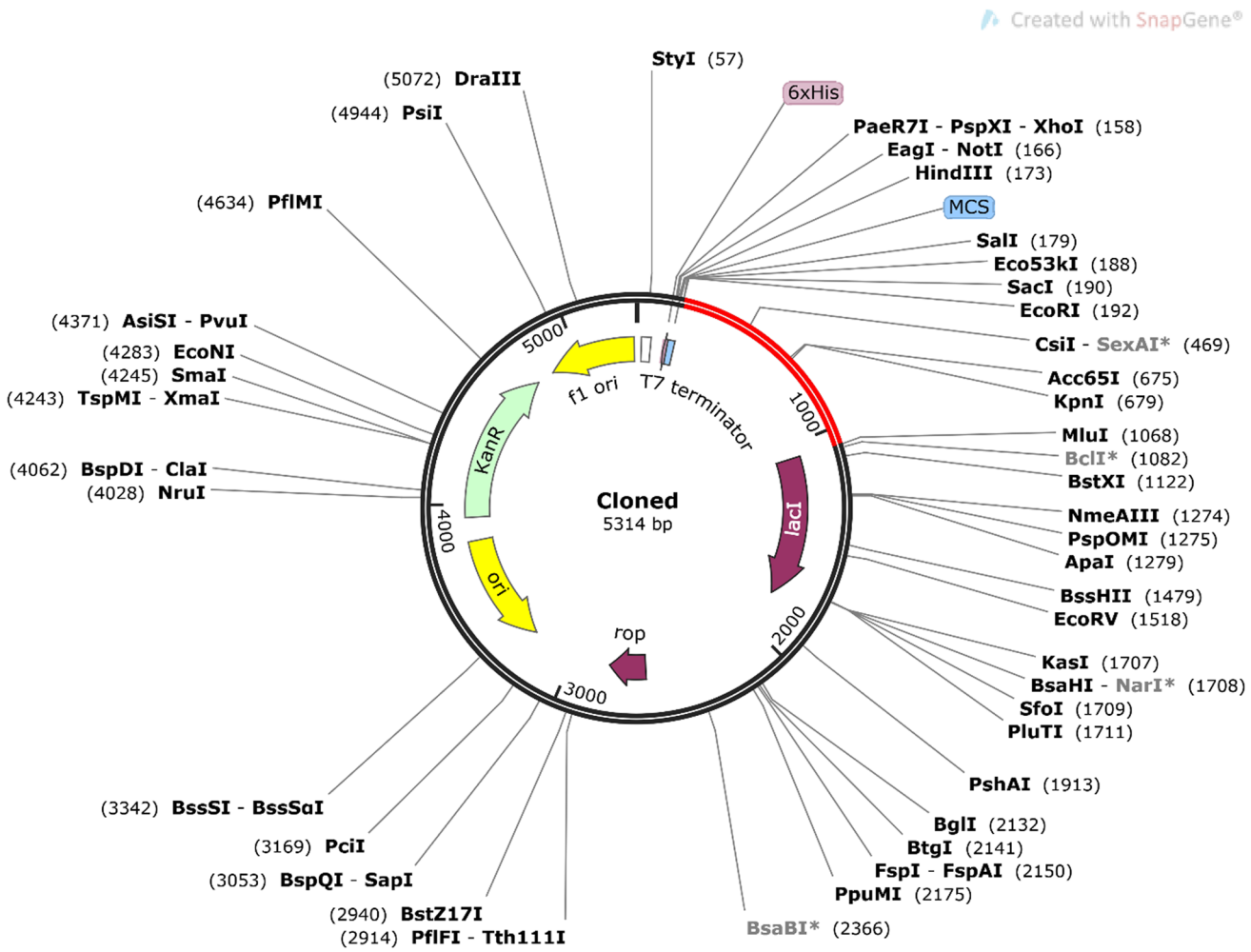


Fig. 5 In silico cloning of vaccine construct into the pET28 (+) vector using SnapGene software

vaccine—TLR-4 complex were monitored during the 50 ns MD simulation and are found to be 156.501 nm^2 and 480 nm^2 respectively. The complex's higher SASA value implies that it is stable in the solvent environment (Fig. 4c).

Codon optimization and in silico cloning

Codon optimization is frequently done to improve the translation efficiency of a target gene by considering the codon bias of the host organism. The Java Codon Adaptation tool was utilized to optimize the codon usage of the vaccine construct in *E. coli* (K12 strain). The Codon Adaptation Index (CAI) was estimated to be 0.98, signifying the success of target gene expression. The GC content of the vaccine construct was found to be 53.67% and it was in the optimal range of 30–70%. Further, the DNA sequence of the vaccine construct was cloned into pET-28 (+) vector between the EcoR1 and MluI restriction sites. These restriction sites were not present in the DNA sequence of the vaccine construct. As a result, they were added at the N-terminal and C-terminal ends of the vaccine, respectively. Finally, the in silico cloning was successfully done using the SnapGene software. The total length of the final vaccine clone was observed to be 5314 bp as depicted in Fig. 5.

Immune simulation studies

According to immune simulation results from C-ImmSim, secondary and tertiary immune responses had higher amounts of IgM, IgM + IgG, IgG1 + IgG2 and IgG1 antibodies than the primary immune response, suggesting that our proposed vaccine molecule has the ability to activate the complement system, which is important in controlling infections. Moreover, it can be observed that the concentration of the antigen also significantly decreased with an increase in the concentration of the antibody during each phase of the immune response (Fig. 6a). In addition to the increase

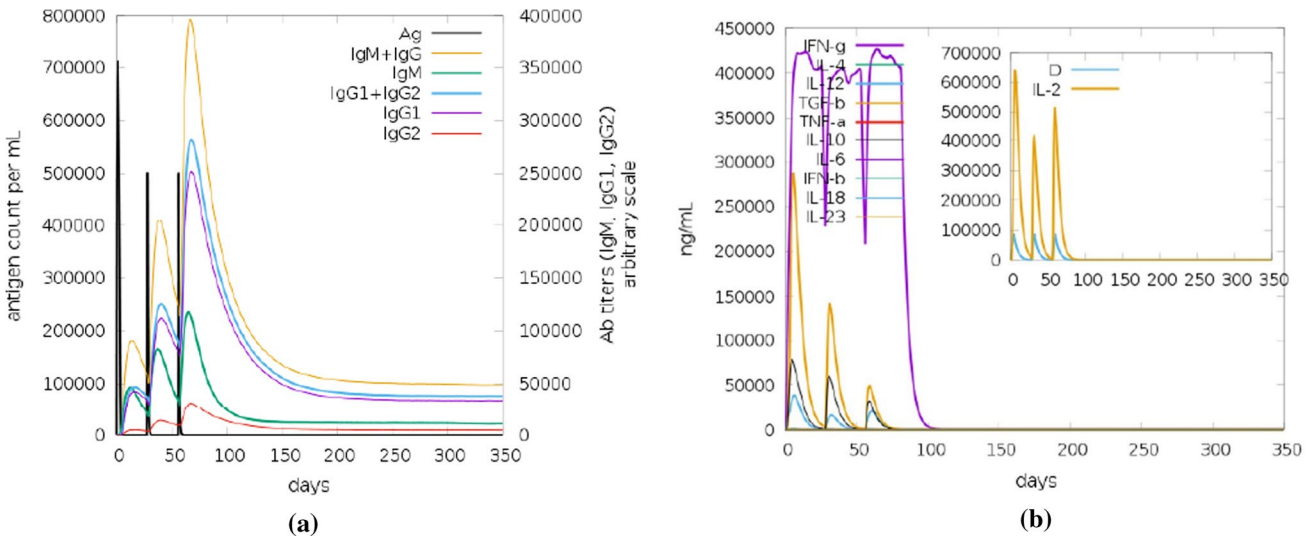


Fig. 6 Immune response profile of the vaccine construct **a** concentration of immunoglobulins, **b** concentration of cytokines and interleukin

in antibodies, other cytokines, including IFN-γ, were also found to be produced at higher levels during immunological reactions (Fig. 6b). Taken together, these immune response results confirm that designed vaccine is able to trigger a robust cellular and humoral immune response.

Conclusion

The present study aimed to identify the antigenic, non-allergic and non-toxic T- and B-cell epitopes of SOX9 with the aid of immunoinformatics strategies. These epitopes, which could elicit cell-mediated as well as humoral immunity, were fused together with an adjuvant (50 s ribosomal L7/L12 protein) to design an epitope-based peptide vaccine against TNBC. The designed vaccine exhibits an instability index of 39.84, which highlights its stability in a test tube. The molecular docking confirmed that the developed vaccine can improve natural defences by binding to immune cell receptor TLR-4. The MD simulation analysis demonstrated that the TLR-4-vaccine complex had a smaller RMSD (1.4 nm) and a higher SASA (480 nm^2), indicating its structural integrity and stability in the solvent environment. Additionally, the Rg value (3.8 nm) of the proposed vaccine was lower, reflecting the molecule's compact nature. Similarly, the immunological simulation studies suggested the developed vaccine could also elicit a potent cellular immune response. The collective evidences from our study highlight that the developed chimeric vaccine candidate could be utilized in the treatment regimen for the management of patients with metastatic triple-negative breast cancer. However, further studies are needed to demonstrate that our vaccine is an effective TNBC preventive.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11030-022-10539-w.

Acknowledgements The authors thank the management of Vellore Institute of Technology for providing the facilities to carry out this research work and acknowledge support from Bioinformatics Resources and Applications Facility (BRAF), C-DAC, Pune.

Declarations

Conflict of interest We wish to confirm that there are no known conflicts of interest associated with this publication.

References

1. Nandini D, Jennifer A, Pradip D (2021) Therapeutic strategies for metastatic triple-negative breast cancers: from negative to positive. J Pharm 14(5):455. https://doi.org/10.3390/ph14050455
2. Luo C, Wang P, He S, Zhu J, Shi Y, Wang J (2022) Progress and prospect of immunotherapy for triple-negative breast cancer. Front Oncol 12:919072. https://doi.org/10.3389/fonc.2022.919072
3. Bai X, Ni J, Beretov J, Graham P, Li Y (2021) Immunotherapy for triple-negative breast cancer: a molecular insight into the microenvironment, treatment, and resistance. J Natl Cancer Inst 1(3):75–87. https://doi.org/10.3390/pharmaceutics13122003
4. Yang R, Li Y, Wang H, Qin T, Yin X, Ma X (2022) Therapeutic progress and challenges for triple negative breast cancer: targeted therapy and immunotherapy. Mol Biol 3(1):1–18. https://doi.org/10.1186/s43556-022-00071-6
5. Benedetti R, Dell'Aversana C, Giorgio C, Astorri R, Altucci L (2017) Breast cancer vaccines: new insights. Front Endocrinol 8:270. https://doi.org/10.3389/fendo.2017.00270
6. Solinas C, Aiello M, Migliori E, Willard-Gallo K, Emens LA (2020) Breast cancer vaccines: heeding the lessons of the past to guide a path forward. Cancer Treat Rev 84:101947. https://doi.org/10.1016/j.ctrv.2019.101947
7. Corti C, Giachetti PP, Eggermont AM, Delaloge S, Curigliano G (2022) Therapeutic vaccines for breast cancer: has the time finally

come? Eur J Cancer 160:150–174. https://doi.org/10.1016/j.ejca.2021.10.027
8. Kumar S, Shuaib M, Prajapati KS, Singh AK, Choudhary P, Singh S, Gupta SA (2022) candidate triple-negative breast cancer vaccine design by targeting clinically relevant cell surface markers: an integrated immuno and bio-informatics approach. 3 Biotech 12(3):1–20. https://doi.org/10.1007/s13205-022-03140-3
9. Mahdevar E, Safavi A, Abiri A, Kefayat A, Hejazi SH, Miresmaeili SM, Iranpur Mobarakeh V (2021) Exploring the cancer-testis antigen BORIS to design a novel multi-epitope vaccine against breast cancer based on immunoinformatics approaches. J Biomol Struct Dyn. https://doi.org/10.1080/07391102.2021.1883111
10. Parvizpour S, Razmara J, Pourseif MM, Omidi Y (2019) In silico design of a triple-negative breast cancer vaccine by targeting cancer testis antigens. Bioimpacts 9(1):45. https://doi.org/10.15171/bi.2019.06
11. Zhu SY, Yu KD (2022) Breast cancer vaccines: disappointing or promising? Front Immunol. https://doi.org/10.3389/fimmu.2022.828386
12. Marqués M, Sorolla MA, Urdanibia I, Parisi E, Hidalgo I, Morales S, Salud A, Sorolla A (2022) Are transcription factors plausible oncotargets for triple negative breast cancers? Cancers 14(5):1101. https://doi.org/10.3390/cancers14051101
13. Grimm D, Bauer J, Wise P, Krüger M, Simonsen U, Wehland M, Infanger M, Corydon TJ (2020) The role of SOX family members in solid tumours and metastasis. Semin Cancer Biol 67:122–153. https://doi.org/10.1016/j.semcancer.2019.03.004
14. Polakova I, Duskova M, Smahel M (2014) Antitumor DNA vaccination against the Sox_2 transcription factor. Int J Oncol 45(1):139–146. https://doi.org/10.3892/ijo.2014.2402
15. Friedman RS, Bangur CS, Zasloff EJ, Fan L, Wang T, Watanabe Y, Kalos M (2004) Molecular and immunological evaluation of the transcription factor SOX-4 as a lung tumor vaccine antigen. J Immunol 172(5):3319–3327. https://doi.org/10.4049/jimmunol.172.5.3319
16. Ma Y, Shepherd J, Zhao D, Bollu LR, Tahaney WM, Hill J, Zhang Y, Mazumdar A, Brown PH (2020) SOX9 is essential for triple-negative breast cancer cell survival and metastasisSOX9 regulates TNBC survival and metastasis. Mol Cancer Res 18(12):1825–1838. https://doi.org/10.1158/1541-7786.MCR-19-0311
17. Sanami S, Azadegan-Dehkordi F, Rafieian-Kopaei M, Salehi M, Ghasemi-Dehnoo M, Mahooti M, Alizadeh M, Bagheri N (2021) Design of a multi-epitope vaccine against cervical cancer using immunoinformatics approaches. Sci Rep 11(1):1–15. https://doi.org/10.1038/s41598-021-91997-4
18. Wieczorek M, Abualrous ET, Sticht J, Álvaro-Benito M, Stolzenberg S, Noé F, Freund C (2017) Major histocompatibility complex (MHC) class I and MHC class II proteins: conformational plasticity in antigen presentation. Front Immunol 17(8):292. https://doi.org/10.3389/fimmu.2017.00292
19. Saha S, Raghava GP (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins: Struct Funct Genet 65(1):40–48. https://doi.org/10.1002/prot.21078
20. Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, Peters B (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinform 9(1):1–8. https://doi.org/10.1186/1471-2105-9-514
21. Dey J, Mahapatra SR, Singh P, Patro S, Kushwaha GS, Misra N, Suar MB (2021) T cell epitope-based peptides predicted from clumping factor protein of *Staphylococcus aureus* as vaccine targets. Microb Pathog 160:105171. https://doi.org/10.3390/vaccines9091038
22. Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinform 8(1):1–7. https://doi.org/10.1186/1471-2105-8-4
23. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Open Source Drug Discovery Consortium, Raghava GP (2013) In silico approach for predicting toxicity of peptides and proteins. PLoS ONE 8(9):e73957. https://doi.org/10.1371/journal.pone.0073957
24. Dimitrov I, Bangov I, Flower DR, Doytchinova I (2014) AllerTOP v. 2—a server for in silico prediction of allergens. J Mol Model 20(6):1–6. https://doi.org/10.1186/1471-2105-14-S6-S4
25. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A (2006) Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinform 7(1):1–5. https://doi.org/10.1186/1471-2105-7-153
26. Atapour A, Negahdaripour M, Ghasemi Y, Razmjuee D, Savardashtaki A, Mousavi SM, Hashemi SA, Aliabadi A, Nezafat N (2020) In silico designing a candidate vaccine against breast cancer. Int J Pept Res Ther 26(1):369–380. https://doi.org/10.7324/JAPS.2021.110604
27. Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED protein analysis workbench. Nucleic Acids Res 41(W1):W349–W357. https://doi.org/10.1093/nar/gkt381
28. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26(2):283–291. https://doi.org/10.1107/S0021889892009944
29. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2(9):1511–1519. https://doi.org/10.1002/pro.5560020916
30. Morris GM, Lim-Wilby M (2008) Molecular docking. J Mol Model. https://doi.org/10.1007/978-1-59745-177-2_19
31. Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 20(1):45–50. https://doi.org/10.1093/bioinformatics/btg371
32. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 33(2):526–531. https://doi.org/10.1093/nar/gki376
33. Castiglione F, Bernaschi M (2004) In C-immsim: playing with the immune response. In: Proceedings of the sixteenth international symposium on mathematical theory of networks and systems (MTNS2004)
34. Maher J, Davies ET (2004) Targeting cytotoxic T lymphocytes for cancer immunotherapy. Br J Cancer 91(5):817–821. https://doi.org/10.1038/sj.bjc.6602022
35. Pandey RK, Ojha R, Aathmanathan VS, Krishnan M, Prajapati VK (2018) Immunoinformatics approaches to design a novel multi-epitope subunit vaccine against HIV infection. Vaccine 36(17):2262–2272. https://doi.org/10.1016/j.vaccine.2018.03.042
36. Galanis KA, Nastou KC, Papandreou NC, Petichakis GN, Pigis DG, Iconomidou VA (2021) 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(6):3210. https://doi.org/10.3390/ijms22063210
37. Ayyagari VS, Venkateswarulu TC, Abraham Peele K, Srirama K (2022) Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach. J Biomol Struct Dyn 40(7):2963–2977. https://doi.org/10.1080/07391102.2020.1850357
38. Livingston B, Crimi C, Newman M, Higashimoto Y, Appella E, Sidney J, Sette A (2002) A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. J Immunol 168(11):5499–5506. https://doi.org/10.4049/jimmunol.168.11.5499
39. Li X, Guo L, Kong M, Su X, Yang D, Zou M, Liu Y, Lu L (2015) Design and evaluation of a multi-epitope peptide of

human metapneumovirus. Intervirology 58(6):403–412. https://doi.org/10.1159/000445059
40. Mahapatra SR, Dey J, Kushwaha GS, Puhan P, Mohakud NK, Panda SK, Lata S, Misra N, Suar M (2021) Immunoinformatic approach employing modeling and simulation to design a novel vaccine construct targeting MDR efflux pumps to confer wide protection against typhoidal Salmonella serovars. J Biomol Struct Dyn. https://doi.org/10.1080/07391102.2021.1964600
41. Ji YY, Li YQ (2010) The role of secondary structure in protein structure selection. Eur Phys J E 32(1):103–107. https://doi.org/10.1140/epje/i2010-10591-5
42. Kumar S, Shuaib M, Prajapati KS, Singh AK, Choudhary P, Singh S, Gupta S (2022) A candidate triple-negative breast cancer vaccine design by targeting clinically relevant cell surface markers: an integrated immuno and bio-informatics approach. 3 Biotech 12(3):1–20. https://doi.org/10.1007/s13205-022-03140-3
43. Shi S, Xu C, Fang X, Zhang Y, Li H, Wen W, Yang G (2020) Expression profile of Toll-like receptors in human breast cancer. Mol Med Rep 21:786–794. https://doi.org/10.3892/mmr.2019.10853
44. Lee SJ, Shin SJ, Lee MH, Lee MG, Kang TH, Park WS, Soh BY, Park JH, Shin YK, Kim HW, Yun CH (2014) A potential protein adjuvant derived from *Mycobacterium tuberculosis* Rv0652 enhances dendritic cells-based tumor immunotherapy. PLoS ONE 9(8):e104351. https://doi.org/10.1371/journal.pone.0104351