International Journal of Mosquito Research

ISSN: **2348-5906**
CODEN: **IJMRK2**
IJMR 2019; 6(4): 10-21

Received: 04-05-2019
Accepted: 06-06-2019

Renu Jakhar
Centre for Medical Biotechnology, Maharshi Dayanand University, Rohtak, Haryana, India

Pawan Kumar
Centre for Bioinformatics, Maharshi Dayanand University, Rohtak, Haryana,

Neelam Sehrawat
Department of Genetics, Maharshi Dayanand University, Rohtak, Haryana, India

SK Gakhar
Centre for Medical Biotechnology, Maharshi Dayanand University, Rohtak, Haryana, India

Correspondence
SK Gakhar
Centre for Medical Biotechnology, Maharshi Dayanand University, Rohtak, Haryana, India

Immunoinformatics prediction for designing potential epitope-based malaria vaccine in the Aminopeptidase N1 protein of *Anopheles gambiae* (Diptera: Culicidae)

Renu Jakhar, Pawan Kumar, Neelam Sehrawat and SK Gakhar

Abstract
Vector based Malaria Transmission Blocking Vaccines target antigens present on midgut surface of *Anopheles* mosquitoes. Aminopeptidase N is one of the important immunogenic protein involved in parasitic infection and its development in mosquitoes. In the present study, the epitopes that can encourage cell-mediated as well as humoral immunity against *Anopheles gambiae* Aminopeptidase N1 have been identified using immunoinformatics approach. Potential epitopes for cytotoxic T cells (MHCI) were identified using NetCTL and IEDB server. Selected T cell epitope were further analyzed by using computational docking technique to check interactions against the HLA molecules. Linear and conformational B cell epitope was also predicted and found to satisfy the threshold scores of all the parameters of predictive tools available on IEDB server. Therefore, an effort was done to design potential epitope based malaria vaccine by *in silico* analysis in the Aminopeptidase N1 protein of *Anopheles gambiae*.

Keywords: *Anopheles gambiae*, Aminopeptidase N, malaria, epitope, immunoinformatics

1. Introduction
Malaria continues to remain a life threatening infectious disease throughout the tropical region of the world. The World malaria report 2018 states that there were about 219 million cases in ninety countries in 2017 [1]. The African continent has the maximum malaria cases because of the presence of the *An. gambiae* complex among different *Anopheles* vector.
A midgut specific protein, Aminopeptidase N 1 (APN1) is glycosylphospotidyl inositol anchored protein reported to play an important role in ookinete invasion of *Plasmodium* in the *Anopheles gambiae* [2]. Aminopeptidase N belongs to a group of membrane bound ubiquitous zinc metalloproteases (ZMP). APN1 has been identified and characterized from *An. stephensi* as a candidate for Transmission Blocking Vaccines (TBV) [3]. Because of the lack of any effective and economical control strategy, TBVs, promise a more efficient way to malaria control. Other studies have shown that the APN protein is a candidate antigen for vaccine development [4]. Studies on the APN 1 gene of *Anopheles gambiae* have shown it as a potential candidate to induce specific humoral and cellular immunity in BALB/c mice [5]. Structural analysis of midgut APN1 in *Anopheles gambiae* has revealed B cell epitope based malaria transmission blocking activity [6]. However, T-cell-based epitope mapping is lacking for cellular immunity which is also essential for cleaning parasite infection.
The vaccination aim to induce immunity against specific pathogens. It will be induced by selectively stimulating antigen specific cytotoxic T-cells, helper T-cells and B-cells. Ideally, a vaccine is divided into two classes based on antigenic epitopes, firstly a B-cell epitope and a helper T-cell epitope, secondly a CTL epitope. By combination of these epitopes, the vaccine is capable to induce either specific humoral or cellular immune response against the specific pathogens [7]. For development of an effective transmission blocking vaccine for malaria, an ideal aim should be the conservation of epitopes in various species of *Anopheles* that elicit both neutralizing antibody and cellular immunity against parasite.
In the present study, we have therefore systemically B-cell and CTL epitopes in the APN1 protein of *An. gambiae* by *in silico* analysis to achieve putative epitope vaccine candidates.

The present study should also help identify the epitopes to draw strategy for transmission blocking malaria vaccine development.

2. Materials and Methods

2.1 Retrieval of protein sequence from database

Protein sequences of all Aminopeptidase from *Anopheles gambiae* were retrieved from the NCBI. The antigenicity of this sequence was predicted by VaxiJen v2.0 server [8] with all default parameter. All the antigenic proteins were then filtered in tabular form with their respective score. Further the APN1 protein sequences from different *Anopheles* species (Diptera: Culicidae) were retrieved from the vectorbase (https://www.vector base.org/). Multiple sequence alignment of APN1 protein sequences from these species carried out through Clustal W.

2.2 Secondary structure analysis

Antigenicity depends on the protein secondary structure. Therefore, for prediction of secondary structures ExPASy's server ProtParam [9] was used. Various parameters like the amino acid composition, extinction coefficient, instability index, aliphatic index and molecular weight are included. Self-optimized prediction method (SOPMA) [10] was also used to study transmembrane helices, solvent accessibility, globular and coiled regions for the analysis of secondary structure APN1 protein. These methods provided information about the protein stability and functional role.

2.3 Prediction of B cell epitope

Immune Epitope Database (IEDB) was used to predict B cell epitopes. The tools at the IEDB, Bepipred linear epitope prediction [11], Emini surface accessibility [12], Kolaskar and Tongaonkar antigenicity [13], Parker hydrophobicity [14], Chou and Fasman beta turn prediction [15] and Karplus & Schulz Flexibility Prediction [16] were applied during prediction. The predicted linear epitopes having equal or more values than average default threshold values are surface accessible, antigenic, hydrophilic and flexible and lie in beta turn regions. ElliPro [17] at IEDB was used to predict conformational B-cell epitopes.

2.4 Prediction of cytotoxic T cell epitope

NetCTL server [18] was used to predict T-cell epitopes. The parameter set with 50 to have highest specificity and sensitivity of 0.94 and 0.89, respectively and all the super types were taken during the submission of protein sequence. A combined algorithm of Major Histocompatibility Complex (MHC)-1 binding, transport efficiency, Transporter of Antigenic Peptide (TAP) and proteosomal cleavage efficiency were applied to predict overall scores. Based on combined score selection of best epitopes was done. Putative epitope candidates epitopes further tested for MHC-1 binding tool at IEDB [19]. Stabilized Matrix Base Method (SMM) was used to calculate the threshold values for strong binding peptides (IC50). 9 amino acids length peptide was selected for all the alleles. For further analysis, alleles having IC 50 value less than 200nm were selected [20]. Overall, the higher immunogenicity of peptides shows more expected to be CTL epitopes. Therefore, immunogenicity prediction tool at IEDB was used to predict immunogenicity of the epitopes [21].

2.5 Conservancy and Allergenicity assessment

The epitope conservancy analysis was carried out in the APN1 protein sequences from different species of *Anopheles* by conservation across antigens tool at IEDB [22]. Similarly, the allergenicity of the epitopes was analyzed by the Allertop for evaluation of allergenicity in proteins [23].

2.6 CTL Epitopes three dimensional structures

CTL Epitopes three dimensional structures were designed by the PEP-FOLD [24].

2.7 Population coverage prediction

Human population coverage for selected epitopes was checked by population coverage tool at IEDB [25]. Every epitopes and their HLA alleles were added, also various ethnic group and geographical regions across the world were also selected.

2.8 Assessment of HLA-Epitope Interaction by Molecular Docking

Molecular docking studies screen out whether or not these epitope will bind with HLA molecules when applied *in vivo* [26]. Autodock Vina [27] was used to analyse the interactions between HLAs and epitopes and Lig Plot $^{+}$ [28]. To carry out the docking simulations, HLA class I, three D structures were first obtained from RCSB [29]. Prior to docking simulation, already bound epitope, with these alleles was removed by using AutoDockTools. For some alleles, three dimensional structures was not available on PDB, so they were modeled by the protein homology modeling tool Swissmodeler [30]. Energy minimization was done with Chimera [31] and structure validation was carried out with saves [32] procheck, qmean [33] and Prosa web server [34].

3. Results

3.1 Retrieval of protein sequence and Antigenicity detemination

A total of eighteen APN protein sequences of *Anopheles gambiae* retrieved from NCBI in FASTA format were screened using the VaxiJen server to predict the most immunogenic polypeptide (Table 1). In the present study, the APN1 (Vectorbase id: AGAP004809) was predicted to be the most immunogenic protein based on overall scores obtained by all specific APN proteins examined.

Table 1: Showing the accession no. of various protein sequences and their predicted score for antigenicity from NCBI.

S. No	Protein	Vectorbase id	Predicted score for Antigenicity	Probable antigen / non-antigen
1	APN1	AGAP004809	0.5664	Antigen
2	APN2	AGAP013393	0.4317	Non-antigen
3	APN3	AGAP013255	0.4301	Non-antigen
4	APN4	AGAP013188	0.4317	Non-antigen
5	APN5	AGAP013146	0.4669	Non-antigen
6	APN	AGAP000885	0.3755	Non-antigen
7	APN	AGAP012757	0.4130	Non-antigen

8	APN	AGAP003692	0.4265	Non-antigen
9	APN	AGAP003695	0.4231	Non-antigen
10	APN	AGAP013001	0.5082	Antigen
11	APN	AGAP003926	0.5326	Antigen
12	APN	AGAP002508	0.5339	Antigen
13	APN	AGAP013155	0.4881	Non-antigen
14	APN	AGAP001881	0.5122	Antigen
15	APN	AGAP012984	0.4423	Non-antigen
16	APN	AGAP013150	0.4828	Non-antigen
17	APN	AGAP004860	0.3885	Non-antigen
18	APN	AGAP004808	0.4730	Non-antigen

3.2 Secondary structure analysis

Secondary structure of most antigenic protein, APN1 was analyzed that included total number of 1020 amino acid, molecular weight of 113kDa, isoelectric point of 5.06, formula of $C_{5058}H_{7824}N_{1342}O_{1545}S_{34}$, 458 alpha helixes, 150 extended strands, 35 beta turns, and 377 random coils are also obtained (Table S2 and Fig. S1).

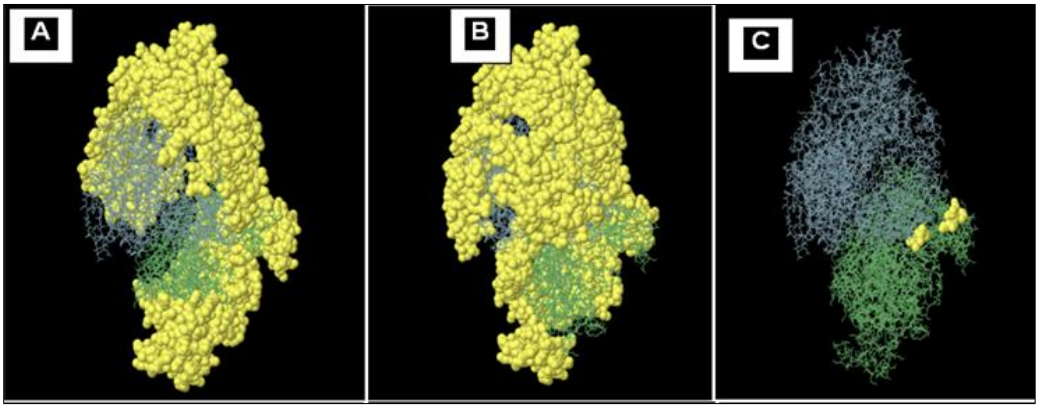
3.3 B-cell epitopes identification.

Kolaskar and Tongaonkar's Antigenicity method indicated that average antigenic prospensity value was 1.025. Potential antigenic determinants are considered on the basis of values greater than the average value and the maximum threshold value of the predicted epitopes was 1.243 with corresponding amino acid position from ^{10}LGLCCLL^{16}.The average value scored as 1.000 for surface accessibility areas of the protein by Emini surface accessibility, all values equal or greater than this value were regarded potentially in the surface and maximum threshold was 6.972 at position ^{665}RNEREY^{670} of protein, where 667E was the surface residue. The predicted peptide ^{992}TTEEDDG^{998} showed the maximum score 7.386 where 995E acts as a hydrophillic residue and the average Parker hydrophilicity Prediction threshold score of the protein was 1.271; all values equal or greater than this were potentially hydrophilic. The Chou and Fasman beta turn prediction method with the default threshold 0.960 was used in this study because the antigenic parts of a protein belong to the beta turn regions. ^{22}PPYKSSG^{28} were shown to have maximum number of beta turn regions. The average Karplus & Schulz Flexibility threshold score of APN1 was 0.989, all values equal or greater than this value were potentially flexible. The maximum flexibility value was 1.095 at amino acid position from ^{452}YDKSGSV^{458}, where 455 S was the flexible residue.

Three epitopes were found to have cutoff prediction scores above threshold scores, namely ^{59}VDERYRL^{65}, ^{137}VQYSTDT^{143} and ^{450} VAYDKSG^{456} (Table 2). Among them VDERYRL and VAYDKSG were found to be nonallergic and conserved in five different types of *Anopheles* species (*An. arabiensis* (AARA016470), *An. merus* (AMEM002547), *An. quadriannulatus* (AQUA016895), *An. farauti* (AFAF015666), *An. culicifacies* (MK033514). The conformational B-cell epitopes were also obtained in four chains of APN1 protein by using ElliPro. ElliPro gives the score to each output epitope, which is Protrusion Index (PI) value averaged over each epitope residue. A number of ellipsoids approximated the tertiary structure of the protein. The highest probability of a conformational epitope was calculated at 75.5% (PI score: 0.755). Residues positions on 3D structures of APN1 protein involved in conformational epitopes shown in Fig. 1A to 1J.

Table 2: Showing B cell epitopes with their allergenicity predicted by IEDB tools.

Epitope	Start	End	Length	Residue	Emini Surface Accessibility Prediction score / Threshold 1.000	Karplus & Schulz Flexibility Prediction score / Threshold 0.989	Kolaskar & Tongaonkar Antigenicity Prediction score / Threshold 1.025	Parker Hydrophiliciy Prediction score / Threshold 1.271	Chou & Fasman Beta-Turn Prediction score / Threshold 0.960	Allergenicity
VDERYRL	59	65	7	R@ 62	3.448	1.009	1.037	1.629	0.904	Non allergen
VQYSTDT	137	143	7	S@ 140	3.042	1.012	1.036	3.9	1.061	Allergen
VAYDKSG	450	456	7	D@ 453	1.686	1.038	1.041	3.486	1.109	Non allergen



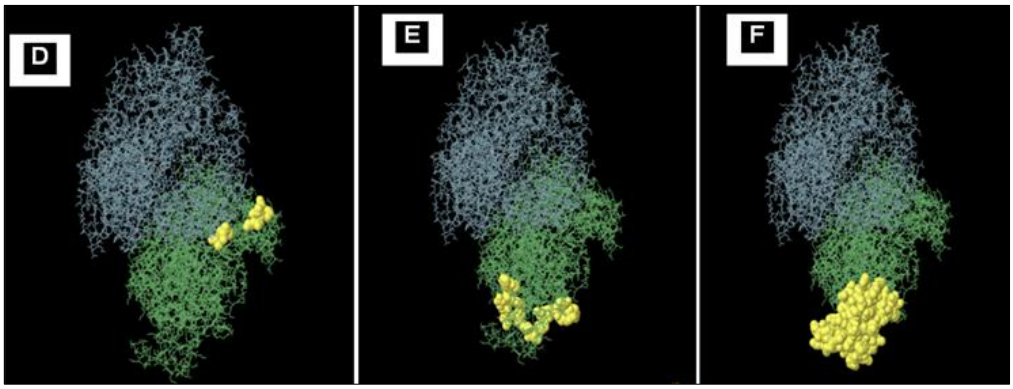


Fig 1: 2D view of conformational epitopes on four chains (A to D) of APN 1 protein. The epitopes showed by yellow surface, and the bulk of the grey and green sticks represents chain A and chain B, respectively of APN1 protein. E and F represent the surface Accessibility of VDERYRL and VQYSTDT, B cell epitopes respectively on B chain.

3.4 Identification of cytotoxic T-cell epitopes. Eight epitopes having high combinatorial scores were considered as most potential epitopes as predicted by NetCTL prediction tool. MHC-I allele interactions with these epitopes were carried out on SMM-based IEDB MHC-I binding prediction tool. The epitopes with higher affinity (IC50 less than 200) with MHC-I alleles were selected for further analysis (Table 3). The affinity for binding of the epitopes with the MHC-I alleles were inversely proportional with the IC50 values. The predicted total score of proteasome score, tap score, MHC score, processing score and MHC-I binding are summarized as total score in Table 5. These epitopes are antigenic and nonallergic in nature. Among these 8 T-cell epitopes, 9-mer epitope, LVSPEWWEY was found to have the highest immunogenicity and combined score. The conservancy analysis of these epitopes indicated that LVSPEWWEY was found to be 45 % conserve. However, another epitope RRYLATTQF was found to be 100 % conserve, which was highest among all epitope. This epitope found to be conserve in *An. arabiensis* (AARA016470), *An. merus* (AMEM002547), *An. quadriannulatus* (AQUA016895), *An. sinensis* (ASIC009153), *An. atroparvus* (AATE011993), *An. darlingi* (ADAC006959), *An. maculatus* (AMAM007684), *An. albimanus* (AALB015678), *An. culicifacies* (MK033514), *An. stephensi* (KJ573522).

This epitope found to have maximum number of allelic interactions with good population coverage than other epitopes.

Table 3: The percentage conservancy, immunogenicity score, Population coverage and total processing score of putative candidate's CTL epitopes on interaction with MHC-I alleles.

S. No.	Epitopes	Position in sequence	Combined score	Interaction of MHC-1 allele with an affinity <200 IC50 (total score of proteasome score, TAP score, MHC score, processing score and MHC 1 binding)	Conservancy (%)	Immunogenicity	Antigenicity	Allergenicity	Population coverage (%)
1.	LVSPEWWEY	376	2.8	HLA-C*03:03(1.31), HLA-C*12:03(1.25), HLA-B*35:01(1.02), HLA-A*29:02(0.87)	45	0.4	Antigenic	Non allergen	23
2.	AEFAAETLL	941	2.1	HLA-B*40:01(0.65), HLA-C*12:03(0.33), HLA-C*03:03(0.32), HLA-B*15:02(0.13), HLA-B*40:02(0.07), HLA-C*07:01(-0.36)	18	0.2	Antigenic	Non allergen	34
3.	RRYLATTQF	190	2.1	HLA-B*27:05(1.06), HLA-C*03:03(0.85), HLA-C*12:03(0.8), HLA-C*14:02(0.65), HLA-C*07:02(0.64), HLA-A*32:01(0.5)	100	0.1	Antigenic	Non allergen	33
4.	AEYHNFLIF	691	1.99	HLA-B*40:02(1.69), HLA-B*40:01(1.42), HLA-B*44:03(1.37), HLA-B*18:01(1.13), HLA-B*44:02(1.02), HLA-B*50:01(0.85), HLA-B*49:01(0.71), HLA-B*15:25(0.69), HLA-B*13:01(0.68),	27	0.2	Antigenic	Non allergen	34

				HLA-B*15:01(0.44)					
5.	GRLDLRVAL	650	1.82	HLA-C*03:03(0.52), HLA-C*14:02(0.52), HLA-B*39:01(0.19), HLA-B*27:05(0.13), HLA-C*12:03(0.11)	55	0.1	Antigenic	Non allergen	17
6.	TYREQALLF	339	1.73	HLA-C*14:02(1.27), HLA-A*23:01(0.76), HLA-C*12:03(0.51), HLA-A*24:02(0.45), HLA-B*15:02(0.25)	55	0.02	Antigenic	Non allergen	34
7.	RVALRFMTY	655	1.68	HLA-C*03:03(1.66), HLA-C*12:03(1.22), HLA-A*32:01(1.08), HLA-A*29:02(0.81), HLA-B*15:01(0.74), HLA-A*03:01(0.68), HLA-A*30:02(0.58)	36	0.03	Antigenic	Non allergen	33
8.	AMSTYLLAF	257	1.32	HLA-B*15:01(1.1), HLA-C*14:02(0.64), HLA-A*32:01(0.45), HLA-B*15:02(0.37), HLA-C*12:03(0.32)	45	0.14	Antigenic	Non allergen	18

3.5 Population Coverage

The population coverage of all predicted epitopes has been analyzed based on their binding with alleles in sixteen identified ethnic group and geographical regions across the world (Table 4). The percentage of population coverage rate for selected MHC I epitope 'RRYLATTQF' of APN1 protein was found to be highest among all epitopes.

Table 4: Showing percentage of Population coverage rate for all selectable putative cytotoxic T-cell and helper T-cell epitopes of APN1 protein.

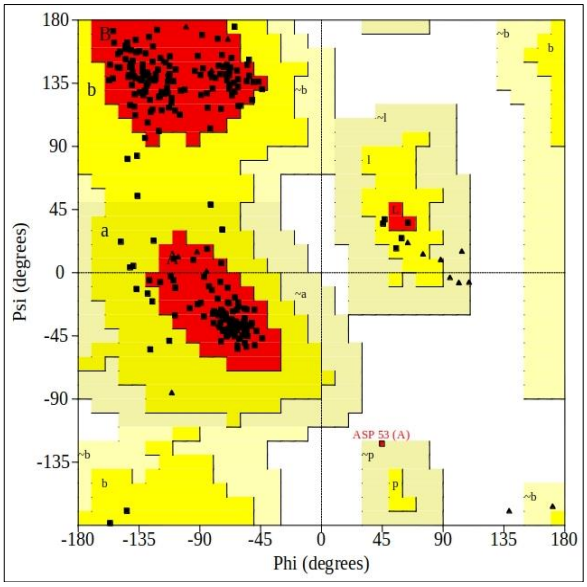
Population/area	Class I HLAs		
	Coverage	Average_hit	pc90
Central Africa	68.52%	1.18	0.32
Central America	1.99%	0.02	0.1
East Africa	57.31%	0.87	0.23
East Asia	83.88%	1.99	0.62
Europe	89.43%	2.34	0.95
North Africa	74.73%	1.51	0.4
North America	81.85%	1.71	0.55
Northeast Asia	76.96%	1.54	0.43
Oceania	76.63%	1.74	0.43
South Africa	1.14%	1.33	-209.43
South America	69.12%	1.29	0.32
South Asia	65.33%	1.3	0.29
Southeast Asia	79.83%	1.5	0.5
Southwest Asia	64.44%	1.23	0.28
West Africa	62.14%	0.98	0.26
West Indies	63.54%	0.84	0.27
Average	63.55	1.34	-12.72

3.6 Homology Modeling, Refinement and Validation of HLA class I alleles.

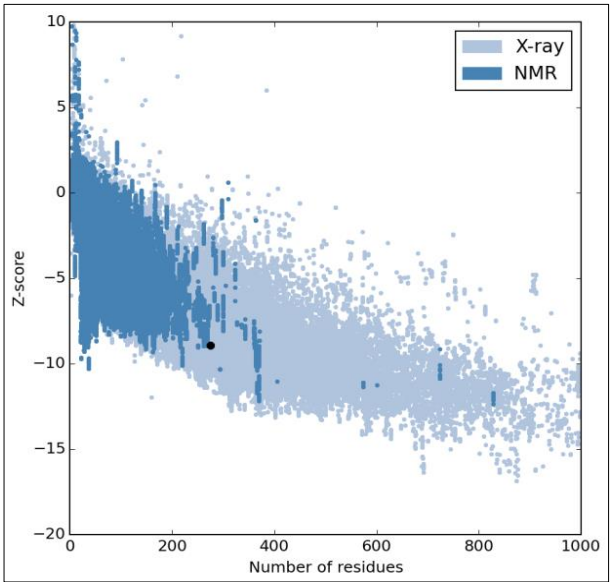
In this study binding of epitope RRYLATTQF were shown with six HLA class I alleles (HLA-A*32: 01, HLA-C*12:03, HLA-C*03:03, HLA-C*14:02, HLA-C*07:02 and HLA-B*27:05). Three dimensional structures of HLA-C*07:02, HLA-C*03:03 and HLA-B*27:05 (PDB ID: 5VGE, 1EFX and 1HSA respectively) were obtained from RCSB. The structures of HLA-C*12:03, HLA-C*14:02 and HLA-A*32:01 was not available on RCSB. So, these were modeled using homology structure modeling tool Swissmodeler. HLA-C*12:03, HLA-C*14:02 and HLA-A*32:01 showed good model with Swissmodeler by using PDB ID: 5VGD.1A, 5VGE.1A and 6EL2.1A respectively as template having more than 90% identity and more than 60% similarity with the query structure. Energy minimization of these models was done by YASARA server (Table 5**)**. All three dimensional structures were evaluated on the basis of scores using Qmean, Ramachandran Z-score, Verify 3D, Prosa Z-score and validation by Ramachandran Plot. The Ramachandran plot and Prosa Z-score validation (Fig. 2) indicated that >92-96% residues in favored region for all of these modeled HLA class I alleles

Table 5: Evaluation scores of modeled structures of HLA-C*12:03, HLA-C*14:02 and HLA-A*32:01.

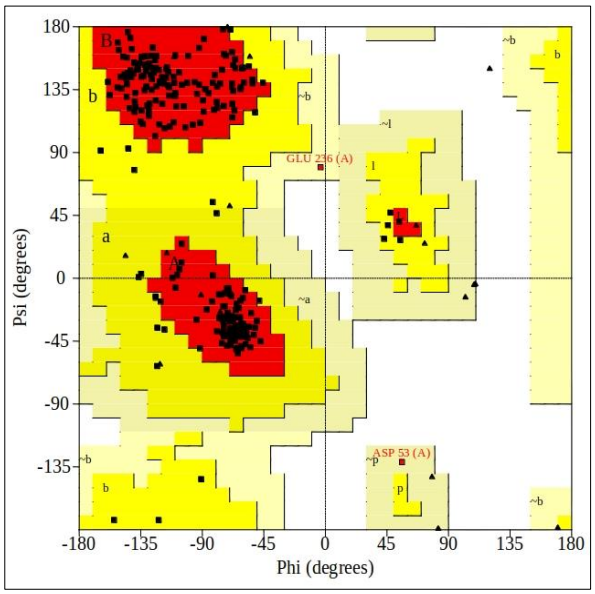
Quality parameters	HLA-C*12:03	HLA-C*14:02	HLA-A*32:01
Template (PDB ID)	5VGE.1A	5VGD.1A	6EL2.1A
Similarity	61%	61%	60%
Identity	95.71%	95.31%	92%
Ramachandran Plot • Most Favoured region	90.0	88.6	90.6
• Additional Allowed Region	9.6	10.5	9.0
• Generously Allowed Region	0.4	0.8	0.4
Disallowed Region	0.0	0.0	0.0
Ramachandran Z-score	-1.710	-2.632	-0.618
Q Mean	-0.11	-0.38	0.42
Verify 3D	96.73	99.27	99.64
Prosa Z-score	-8.92	-8.58	-8.94



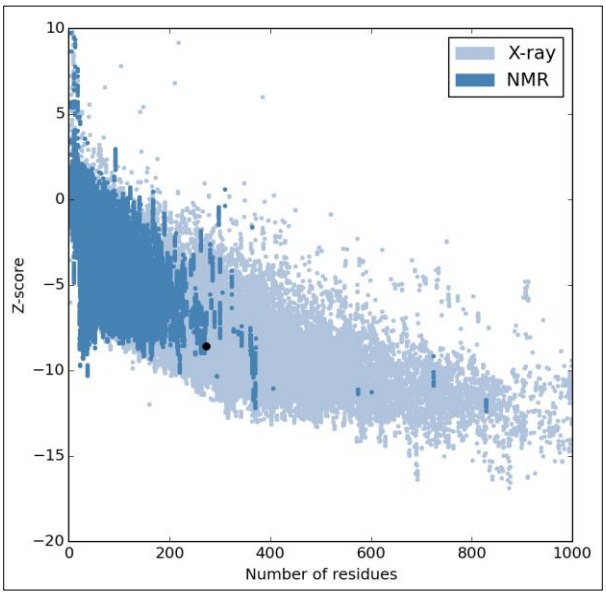
HLA-C*12:03



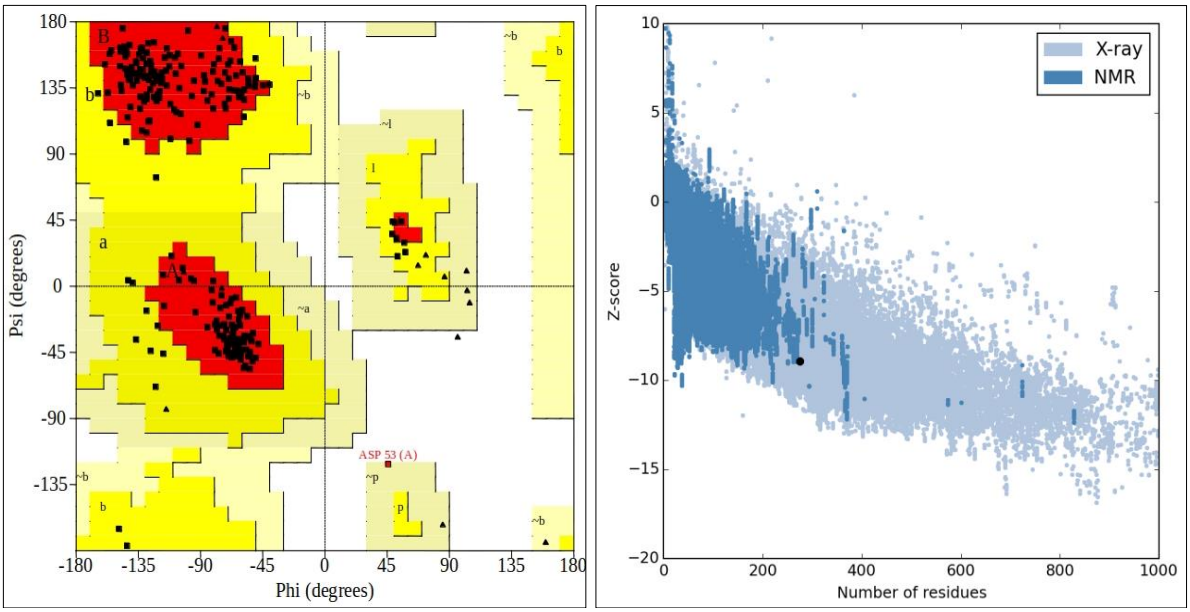
HLA-C*12:03 (Score: -8.92)



HLA-C*14:02



HLA-C*14:02(Score: -8.58)



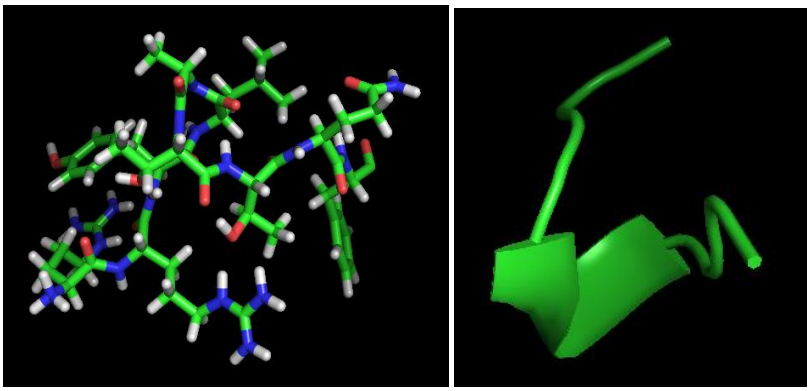
HLA-A*32:01 HLA-A*32:01 (Score: -8.94)

Fig 2: Ramachandran Plot and Prosa Z-score validation of HLA-C*12:03, HLA-C*14:02 and HLA-A*32:01.

3.7 Docking simulation. Binding interactions between epitopes and HLA alleles were assessed using Autodock Vina. The 3D structure of epitopes was predicted using PEP-FOLD and energy minimization was carried out by using Yasara (Fig. 3). The receptors used for docking studies included three reported HLAs (HLA-C*07:02, HLA-C*03:03 and HLA-B*27:05) and three modeled HLAs (HLA-C*14:02, HLA-C*12:03 and HLA-A*32:01). However epitope (RRYLATTQF) was used as ligand for MHC class I. The grid coordinates from selected receptor molecules for docking with their epitope are summarized in Table 6. 1Å spacing was used to select the binding site. The grid box was positioned carefully to make the docking of ligands at the binding groove of the receptors. The binding energies of predicted epitope with their respective allele's receptor were as shown in table 6. HLA-C*14:02 was observed to have the best interaction with the RRYLATTQF epitope with lower binding energy (-8.8 kCal/mol). The predicted peptides showed significant binding affinities with all HLAs (Fig. 4). The more negative ΔGbinding value, stronger is the interaction between the epitope and HLA. Also, the binding energy of the predicted epitopes were compared with the binding energy of the already experimentally verified peptides and found to be negative.

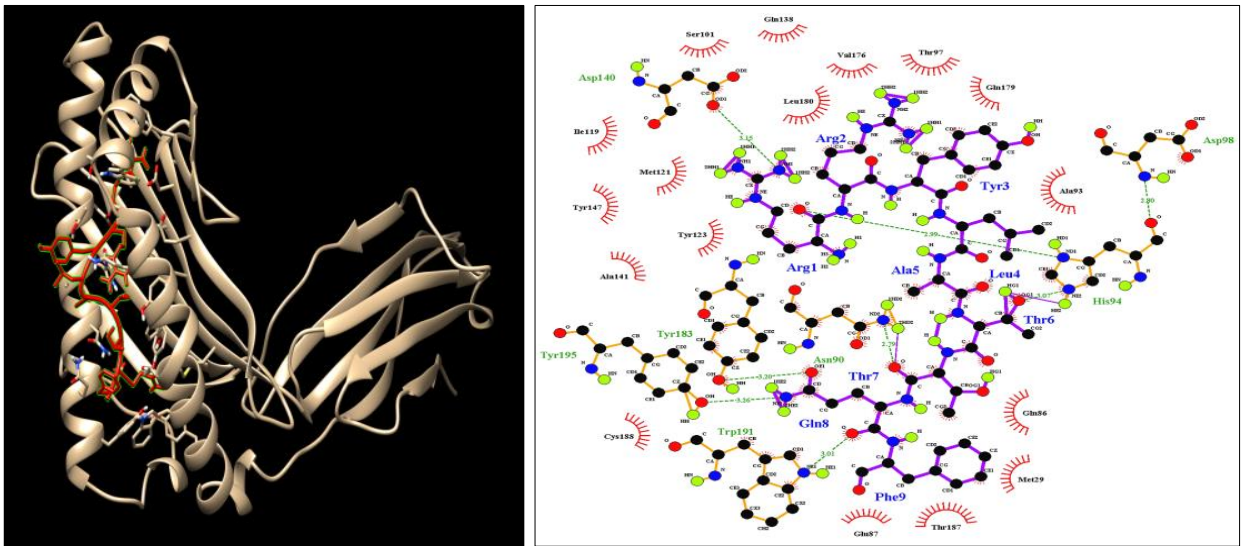
Table 6: Binding site coordinates for protein-ligand docking between MHC 1 molecules and peptides prepared by autodock tools. Also, docking simulation scores for MHCI Binding prepared by autodock vina for predicted epitopes.

Protein Name	Axis	Center Box	Size	Binding Affinity (KCal/mol)
HLA-A*32:01	X	48.328	52	-8.3
	Y	43.377	68	
	Z	24.999	68	
HLA-B*27:05	X	1.027	64	-8.4
	Y	10.411	102	
	Z	18.775	94	
HLA-C*03:03	X	18.268	80	-7.3
	Y	16.942	64	
	Z	63.061	66	
HLA-C*07:02	X	18.108	64	-7.1
	Y	-37.573	72	
	Z	4.034	64	
HLA-C*12:03	X	16.951	64	-6.5
	Y	-38.62	72	
	Z	3.709	60	
HLA-C*14:02	X	11.127	50	-8.8
	Y	86.972	66	
	Z	17.826	70	

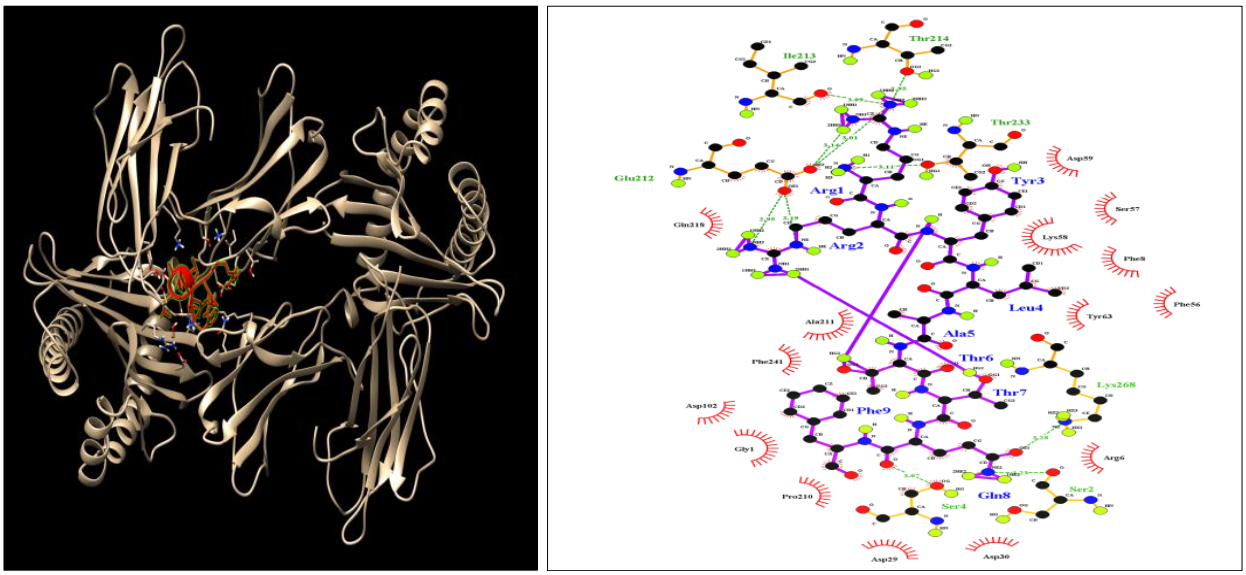


RRYLATTQF

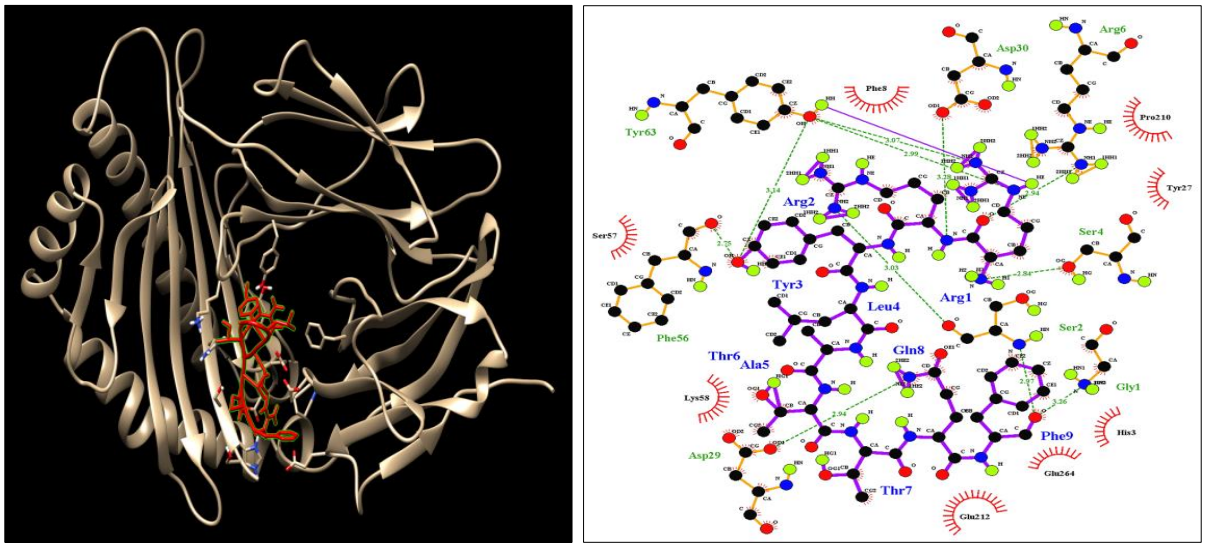
Fig 3: Peptide Structure of predicted epitope “RRYLATTQF” for MHC I prepared by PEP-FOLD.



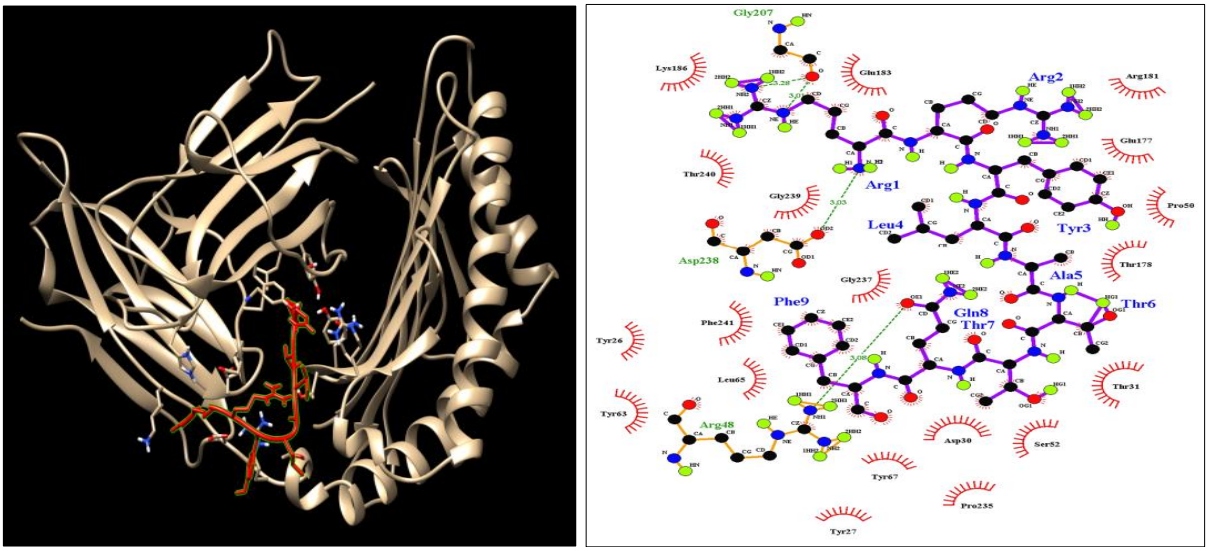
HLA-A*32:01



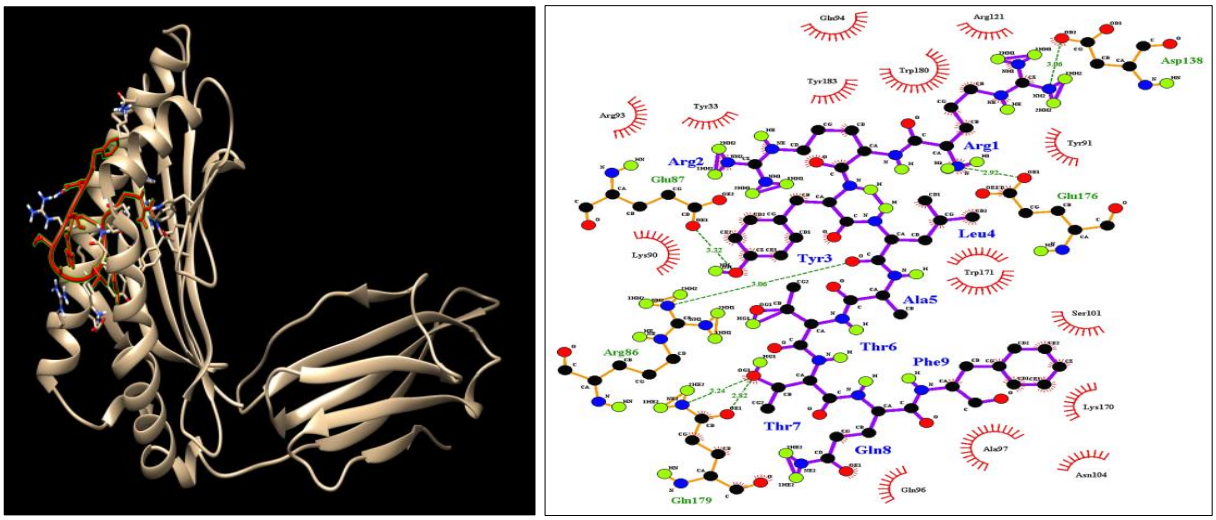
HLA-B*27:05



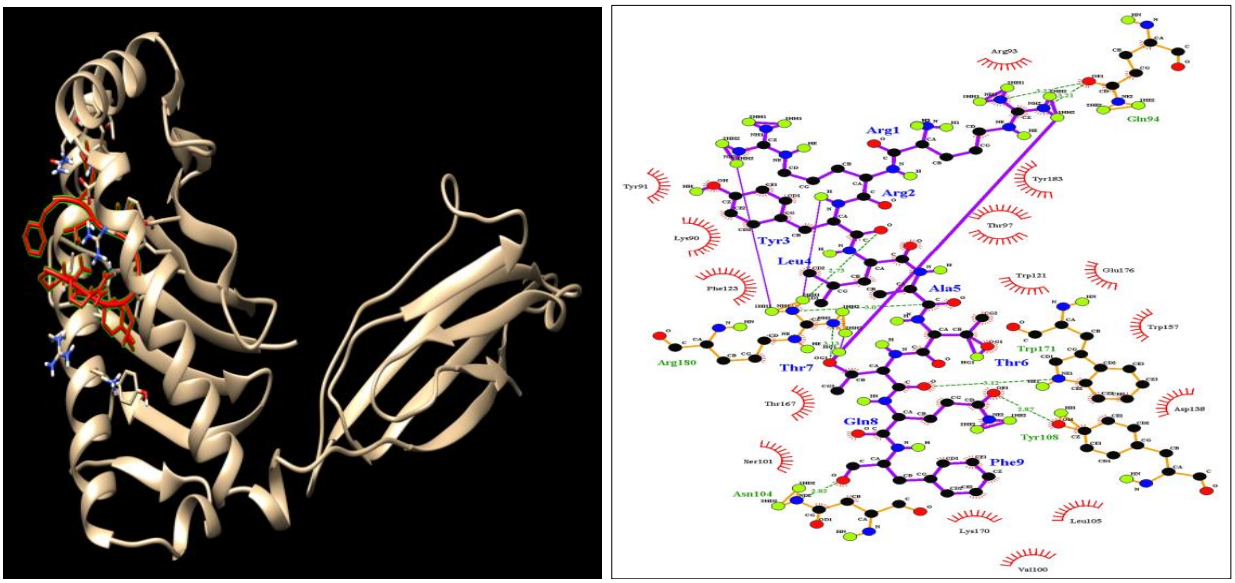
HLA-C*03:03



HLA-C*07:02



HLA-C*12:03



HLA-C*14:02

Fig 4: Docking sites of predicted peptide (RRYLATTQF) against selected MHCI receptors.

4. Discussion

Malaria transmission blocking vaccine helps control malaria without causing ecological imbalance. During the present study, the most potent B and T cell epitopes for transmission blocking vaccine in APN1 protein of *Anopheles gambiae* based on computational techniques.

APN1 was found to be the most immunogenic protein by Vaxijen server and this has also been indicated as a lead TBV candidate [5]. The analysis of secondary structure of APN1 revealed that its antigenic part is more likely to be the beta sheet region as also reported in other experiment [40]. The presence of threonine residues (10.2%), which predominately in the beta sh*eet al*so indicate about the protein's antigenicity. The predicted negative value (-0.129) of grand average of the hydrophobicity rule (GRAVY) of this linear sequence protein not only indicates its hydrophilic nature but also indicates the presence of residues mostly on the surface. In addition, this protein is stable and aliphatic in nature because its Instability Index (32) is smaller than 40 and Aliphatic Index (86.31) has higher value. High aliphatic index seems to be responsible for increasing the thermo stability of globular proteins and higher proportions of coiled region provide more stability.

B cell and T cell epitopes are the key molecules of humoral and cell mediated immunity. Two types of B cell epitopes are linear epitopes and conformational epitopes. We predicted three linear (continuous) epitopes based on scores which were above threshold values of five algorithms- Parker hydrophilicity, Emini surface accessibility, Chou and Fasman beta turn prediction, Kolaskar and Tongaonkar antigenicity and Bepipred linear epitope prediction available on IEDB. The more value of B cell epitope scores then the threshold level in five algorithms indicates that these candidate epitopes ($_{59}$VDERYRL$_{65}$ and $_{137}$VQYSTDT$_{143}$ and $_{450}$VAYDKSG$_{456}$) could be an effective antigenic peptide in response to B cells. First and third epitopes were found to be nonallergic in nature. The localization of conformational (discontinuous) epitopes on four chains (A to D) of the APN1 protein using 3D representation of residues revealed that the presumptive antigenic epitopes sequence that is placed in such a way which enables it to have direct interactions with immune receptor. The B-cell epitopes residues, $_{59}$VDERYRL$_{65}$ and $_{137}$VQYSTDT$_{143}$ situated on the surface of B chain of APN1 protein had good Protrusion Index (PI) score (0.755) were indicative of high accessibility. Ellipsoid value of PI 0.75 indicates that 75% protein residues lie within ellipsoid and the remaining 25% residues lie outside. PI score and solvent accessibility are directly proportional to each other, if PI score is higher; maximum is the solvent accessibility of the residues. Thus, these could be the putative vaccine candidates. T-cell based development of vaccines seems to have potential because of antigenic drift as the foreign particles can easily engineer the escape from antibody memory response. In addition T-cell mediated immunity tends to be a long-lasting. The peptide that passes several criteria has been considered to be a good epitope candidate such as possessing antigenicity, non allergen, highly immunogenic, good conservancy, good interaction with HLA molecules and enough population coverage. During the present study, it was found that the LVSPEWWEY epitope could be used as a potential candidate because it had maximum combined score and immunogenic score, however, it possessed the minimal number of HLA binding alleles amongst other CTL epitopes. Also, the epitope AEYHNFLIF and AEFAAETLL had the maximum number of HLA binding alleles, but having less immunogenic and combined score. This inconsistency of immunological features of epitopes indicates that some other parameters also needed for screening. An ideal epitope, should be highly conserved among different species of *Anopheles*. The conservancy analysis of these epitopes indicated that RRYLATTQF was found to be conserved in ten *Anopheles* species. Previous studies reported a potential 9 mer epitope in *Anopheles gambiae* APN1 at the N terminal region and proved that the amino acids from 60 to195 to be use as TBV [5], this region is safe and highly immunogenic, even in the absence of an adjuvant, in murine models [4]. Interestingly, CTL epitope (RRYLATTQF) and B cell epitope (VDERYRL) predicted during the present *in silico* study coincides with this location.

The peptide that fulfills the above parameters, RRYLATTQF epitope for MHC class I, were further chosen for docking

studies. Docking simulation study of the predicted MHC peptides with HLA molecules was performed to find out that whether the designed epitope will elicit the sufficient immunological responses *in vivo*. The binding energy of predicted MHCI epitope with HLA-B*27:05 receptor was found to be -8.4 kcal/mol as compared to the binding energy of Nipah virus Vprotein predicted epitope (NPTAVPFTL) with HLA-B*27:05 (-3.13 kcal/mol) and was observed to be lower in the predicted epitope [36]. The interaction between the epitope and HLA are stronger if ΔG binding value is more negative

The strong binding affinity showed that peptide vaccine designed by using these selected epitopes may be well work *in vivo* to elicit cell mediated immunity.

Different ethnic populations have high polymorphism in HLA and its reaction to T-cell epitopes is being restricted by HLA proteins. Therefore, to stimulate immune responses in human populations among world, the HLA specificity of T-cell epitopes has to be measured as main criteria for selection of the epitopes. On the basis of above study, the epitope candidates should bind maximum HLA alleles to get better population coverage. In this study, the eight HTL epitopes have shown good population coverage (63% for MHC I) in average) and reached above average values in Europe, North America and East Asia population. Further analysis has shown that helper T-cell epitopes RRYLATTQF (32.97%) for MHC class I (that bind the maximum number of HLA alleles) were chosen as putative vaccine candidates. This epitopes have good coverage of population and it may provide a broad immune protection to human beings from different regions of the world.

The predicted CTL epitope RRYLATTQF for cellular immunity, and B cell epitope VDERYRL for humeral immunity may be synthesized for further *in vivo* and *in vitro* assays. These results are based on an analysis of available data on various immune databases. The results of the present study suggest that the predicted epitopes are good candidates for making a peptide vaccine which may initiate an effective immune response *in vivo*.

5. Conclusion

During the present study, the immunoinformatics driven tools were used for selection of vaccine targets to step up for the development of malaria Transmission Blocking Vaccine. Based on this approach, the CTL epitope (RRYLATTQF) and B-cell epitope (VDERYRL) of the APN1 protein of *Anopheles gambiae* were selected as vaccine candidates. Docking simulation energy scores of the predicted epitopes RRYLATTQF for MHC I were found significantly low. Therefore, lower energy scores represents better binding between ligand and receptor. Overall, T and B cell stimulation potentials of the screened epitopes are required to be tested by *in vitro* and *in vivo* experiments along with this *in silico* study for their proficient use as vaccines against malaria. The present study provide us a new and significant epitope candidates which may be helpful for the development of Transmission Blocking Vaccine of malaria and other infectious diseases of same kind.

6. Acknowledgements

The author Renu Jakhar highly acknowledges the financial assistance and support provided by DBT-IPLS, Govt. of India

8. Author contribution: Renu Jakhar conducted the study, performed *in silico* analysis and wrote the manuscript. S.K. Gakhar plans the study and revises the manuscript. Neelam Sehrawat analyzed the data. Pawan Kumar performed the *in silico* analysis.

9. References

1. World malaria report. 2018.
2. Rosenfeld A, Vanderberg JP. Plasmodium berghei: induction of aminopeptidase in malaria-resistant strain of Anopheles gambiae. Experimental parasitology. 1999; 93(2):101-4.
3. Shabani S, Karimi A, Mahboudi F, Poursharif A, Djadid ND, Motalleb G *et al*. Molecular identification and characterization of aminopeptidase N1 from *Anopheles stephensi*: A candidate for transmission blocking vaccines. Gene Reports. 2016; 5:157-62.
4. Mathias DK, Plieskatt JL, Armistead JS, Bethony JM, Abdul-Majid KB, McMillan A *et al*. Expression, immunogenicity, histopathology, and potency of a mosquito-based malaria transmission-blocking recombinant vaccine. Infection and Immunity. 2012; 80(4):1606-14.
5. Armistead JS, Morlais I, Mathias DK, Jardim JG, Joy J, Fridman A *et al*. Antibodies to a single, conserved epitope in Anopheles APN1 inhibit universal transmission of falciparum and vivax malaria. Infection and Immunity. 2014; 82(2):818-29.
6. Atkinson SC, Armistead JS, Mathias DK, Sandeu MM, Tao D, Borhani-Dizaji N *et al*. Structural analysis of Anopheles midgut aminopeptidase N reveals a novel malaria transmission-blocking vaccine B-cell epitope. Nature Structural & Molecular Biology. 2015; 22(7):532.
7. Shi J, Zhang J, Li S, Sun J, Teng Y, Wu M *et al*. Epitope-based vaccine target screening against highly pathogenic MERS-CoV: an in silico approach applied to emerging infectious diseases. PloS one. 2015; 10(12):e0144475.
8. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007; 8(1):4.
9. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD *et al*. Protein identification and analysis tools in the ExPASy server. Methods Molecular Biology. 1999; 112:531-52.
10. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*. 1995; 11(6):681-684.
11. Andersen PH, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Science. 2006; 15(11):2558-67.
12. Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. Journal of virology. 1985; 55(3):836-9.
13. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 1990; 276(1-2):172-4.
14. Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X ray-derived accessible sites. Biochemistry. 1986; 25(19):5425-32.
15. Chou PY, Fasman GD. Prediction of the secondary

structure of proteins from their amino acid sequence. Advances in enzymology and related areas of molecular biology. 1978; 47:45.
16. Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. Naturwissenschaften. 1985; 72(4):212-3.
17. Wilson JA, Hart MK. Protection from Ebola virus mediated by cytotoxic T lymphocytes specific for the viral nucleoprotein. Journal of virology. 2001; 75(6):2660-4.
18. Larsen MV, Lundegaard C, Lamberth K. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics. 2007; 8:424.
19. Buus S, Lauemøller SL, Worning P. Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. Tissue Antigens. 2003; 62(5):378-84.
20. Vita R, Overton JA, Greenbaum JA. The immune epitope database (IEDB) 3.0. Nucleic Acids Research. 2015; 43(D1):D405-412.
21. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A *et al*. Properties of MHC classI presented peptides that enhance immunogenicity. PLoS Computational Biology. 2013; 9(10):e1003266.
22. Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics. 2007; 8:361.
23. Ivan Dimitrov, Lyudmila Naneva, Irini Doytchinova, Ivan Bangov, Allergen FP. allergenicity prediction by descriptor fingerprints. Bioinformatics. 2014; 30(6):846-851.
24. Thévenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tufféry P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. Nucleic Acids Research. 2012; 40:W288-W293.
25. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics. 2006; 7:153.
26. Trott O, Olson AJ. Auto Dock Vina. Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. Journal of Computational Chemistry. 2010; 31(2):455-61.
27. Tambunan USF, Zahroh H, Utomo BB, Parikesit AA. Screening of commercial cyclic peptide as inhibitor NS5 methyltransferase of dengue virus through molecular docking and molecular dynamics simulation. Bioinformation. 2014; 10(1):23-27.
28. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. Protein Engineering. 1996; 8:127-134.
29. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H *et al*. The protein data Bank. Nucleic Acids Research. 2000; 28:235-242.
30. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R *et al*. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research. 2018; 46:296-303.
31. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC *et al*. UCSF Chimera-a visualization system for exploratory research and analysis. Journal of Computational Chemistry. 2004; 25(13):1605-12.
32. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. Procheck - a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography. 1993; 26:283-291.
33. Benkert P, Tosatto SC, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. Proteins: Structure, Function, and Bioinformatics. 2008; 71(1):261-77.
34. Wiederstein, Sippl. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research. 2007; 35:407-410.
35. Kamthania M, Sharma DK. Screening and structure-based modeling of T-cell epitopes of Nipah virus proteome: an immunoinformatic approach for designing peptide-based vaccine. 3 Biotech. 2015; 5(6):877-82.