JBUON 2018; 23(5): 1514-1527
ISSN: 1107-0625, online ISSN: 2241-6293 • www.jbuon.com
E-mail: editorial_office@jbuon.com

ORIGINAL ARTICLE

In silico CD4+, CD8+ & humoral immunity associated antigenic epitope prediction and HLA distribution analysis of HTLV-I

Tasneem Bano[1], Essam Mohammed Janahi[2], Anupam Dhasmana[3], Mohtashim Lohani[1,4], Shafiul Haque[5], Raju K. Mandal[5], Sajad A. Dar[5], Arshad Jawed[5], Mohd Wahid[5], Naseem Akhter[6], Mohammed Y. Areeshi[5]

[1]Environmental Carcinogenesis and Toxicoinformatics Laboratory, Department of Bioengineering & Biosciences, Integral University, Lucknow, Uttar Pradesh, India; [2]Department of Biology, College of Science, University of Bahrain, Kingdom of Bahrain; [3]Himalayan School of Biosciences & Cancer Research Institute, Swami Rama Himalayan University, Dehradun (Uttarakhand), India; [4]Department of Emergency Medical Sciences, College of Applied Medical Sciences, Jazan University, Jazan-45142, Saudi Arabia; [5]Research and Scientific Studies Unit, College of Nursing and Allied Health Sciences, Jazan University, Jazan-45142, Saudi Arabia; [6]Department of Laboratory Medicine, Faculty of Applied Medical Sciences, Albaha University, Albaha-65431, Saudi Arabia

Summary

Purpose: *The linkage of human T-cell leukemia virus type 1 (HTLV-1) to fatal diseases is a well known fact for many years. However, there has been no significant progress in the field of the treatment that can lead to the development of a successful vaccine. Furthermore, there are no means of assessing the risk of disease and its prognosis in the infected people.*

Methods: *The current study has taken the cognizance of the importance of host's immune response in reducing the risk of infectious diseases to carry out immunoinformatics driven epitope screening strategy of vaccine candidates against HTLV-1. In this study, a genetic variability and HLA distribution analysis among the documented HTLV-1 genotypes I, II, III, IV, V & VI was performed to ensure the coverage of the vast majority of population, where vaccine would be employed. The meticulous screening of effective dominant immunogens was done with the help of ABCPred and Immune Epitope Database.*

Results: *The results showed that the identified epitopes might be protective immunogens with high conservancy and potential of inducing both protective neutralizing antibodies and T-cell responses. The peptides "PSQLPPTAPPLLPHSNLDHI", "PCPNLVAYSSYHATY", and "YHATYSLYLF", were 100% conserved among different isolates from far and wide separated countries, suggesting negligible antigenic drift in HTLV-1.*

Conclusions: *Overall, the mentioned epitopes are soluble, non-toxic suitable candidates for the development of vaccine against HTLV-1 and warrant further investigation and experimental validation.*

Key words: *HTLV-1, immunity, immunoinformatics, oncogenic virus, peptide vaccine*

Introduction

HTLV-1 belongs to delta retrovirus family and other members of this family are simian T-lymphotrophic virus type 1 (STLV-1) and bovine leukemia virus (BLV) [1]. HTLV-1 was first reported in a cell line derived from a patient with cutaneous T-cell lymphoma in 1980 [2]. The symptoms of HTLV-1 infection do not appear in the majority of people, whereas in some people it causes severe problems like neoplastic diseases: Adult T-cell Lymphoma (ATL), Cutaneous T-cell Lymphoma [3], inflammatory syndromes: HAM/TSP [4], uveitis [5] and opportunistic infections: dermatitis, lymphadentis,

Correspondence to: Shafiul Haque, PhD. Research and Scientific Studies Unit, College of Nursing and Allied Health Sciences, Jazan University, Jazan-45142, Saudi Arabia.
Tel & Fax: +966 173174383, E-mail: shafiul.haque@hotmail.com
Received: 11/03/2018; Accepted: 03/04/2018

Sjogren's syndrome [6]. ATL and HAM/TSP are the most predominant among them. ATL has 5 different clinical stages: asymptomatic carrier state, preleukemic state, chronic/smoldering ATL, lymphoma type ATL and acute ATL [7]. HTLV-I infects T-cells, B-cells, dendritic cells, monocytes, and endothelial cells equally, yet the virus has preferential tropism for CD4 T-cells [8]. HTLV-1, a single-stranded diploid RNA virus, is 9,030–9,040 nucleotides long and contains two flanking long terminal repeat (LTR) sequences [9] and is packaged in the viral core with the viral nucleocapsid protein (p15 NC). NC is enclosed in capsid (p24 CA) and matrix (p19 MA) proteins. Similar to the other retroviruses, the HTLV-1 genome encodes the structural proteins 'Gag' (NC, CA, and MA) and 'Env', and the enzymatic 'Pol' proteins RT, Rnase H (RH), integrase (IN), and protease. The *env* gene encodes two subunit proteins: (SU) gp46 and transmembrane (TM) gp21 [10]. The gp46 confers cell receptor-binding activity. The fusion of receptor protein is performed by gp21. HTLV-I 'Env' proteins are genetically highly conserved among isolates, with very low variability (1 to 8%) within their amino acid sequences [11,12]. On the other hand, the cellular entry of HTLV-1 involves three host cell surface proteins: glucose transporter 1 (GLUT1), neuropilin-1 (NRP-1) and heparan sulfate proteoglycans (HSPG) [13], leading to a multi-receptor model host entry for HTLV-1 [14,15]. The CTD of HTLV-1 SU interacts with HSPGs during the initial attachment phase. This improves the possibility and strength of binding of HTLV-1 SU with NRP-1. The binding of HTLV-I SU to NRP-1 leads to a conformational change, resulting in exposure of the GLUT1- binding domain. Finally, the interaction of SU with GLUT1 activates the translocation of TM gp21 that contributes in the fusion mechanism [1]. HTLV-1 virus has six known genotypes (HTLV-1a-g), distributed worldwide [3].

Comprehensive knowledge of the adaptive immune system is required for the development of epitope-based vaccines. T_H cells and T_C cells can recognize antigen when bound with major histocompatibility Complex (MHC) class II and I molecule, respectively [15,16]. MHC is a membrane glycoprotein. In humans, it is also known as the human leukocyte antigen (HLA) and is extremely polymorphic in nature. The HLA molecules may bind to an array of antigenic linear epitopes derived from antigen processing, which initiate an immune response. However, HLA binding alone does not en-

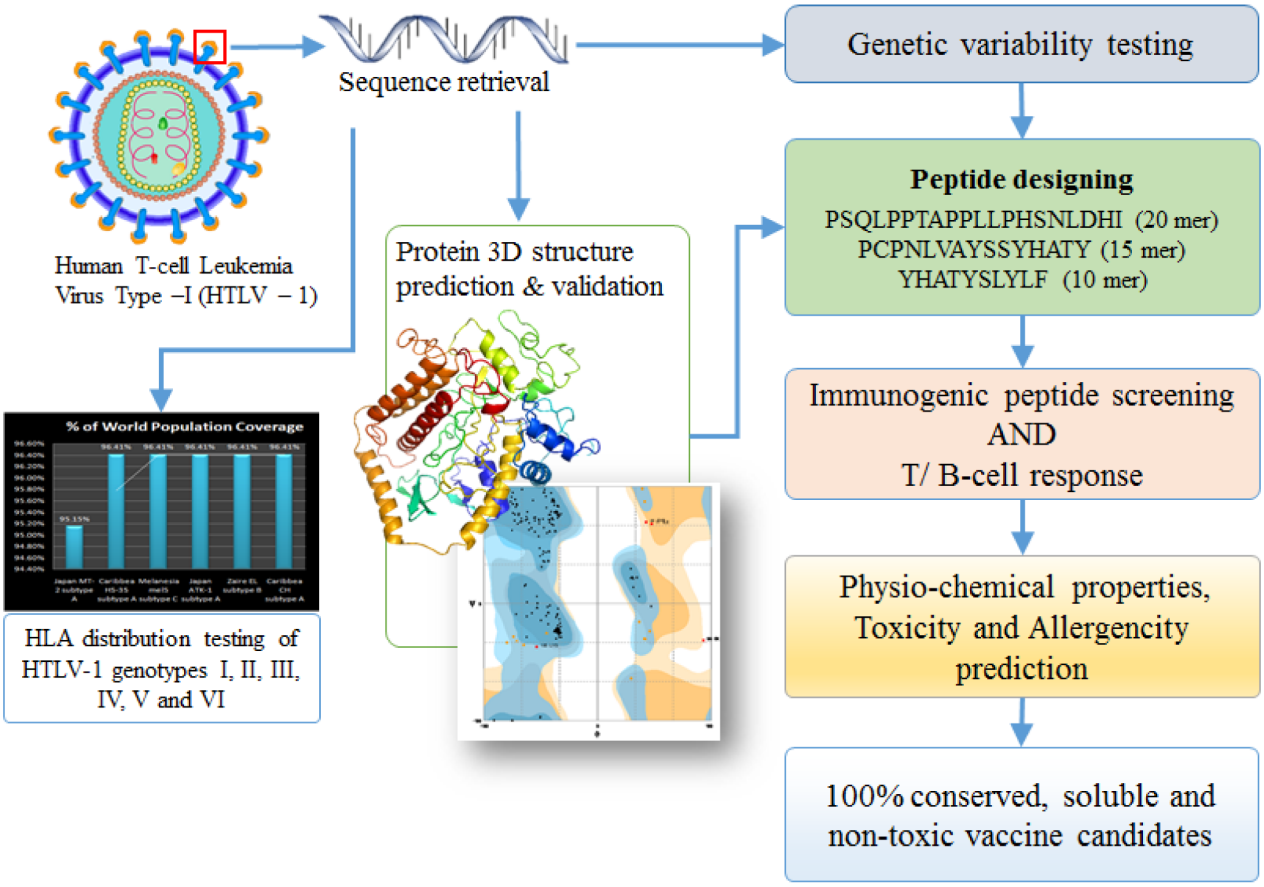

Figure 1. Schematic representation of the entire methodology adopted in this work.

sure the generation of T-cell immune response. The different HLA alleles have varying peptide binding specificity among ethnic populations. It has been reported that only few HLA supertypes can cover the majority of alleles, where diverse members of a supertype bind similar peptides; these similar peptides are called super antigens. Recently, eighteen major HLA class I supertypes (HLA-HLA- A*0101, A*03:01, A*26:01, A*02:01, A*11:01, A*68:01, B*08:01, B*37:01, B*07:02, B*14:02, B*27:05, B*38:01, B*39:01, B*40:02, B*44:02, B*48:01, B*49:01, B*51:01,) and six HLA class II supertypes (DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*07:07, DRB1*15:01) have been determined by comparing peptide-binding data [17]. The problem of HLA allele's population coverage and the chance of antigen escape due to antigenic drift or shift may be overcome by the binding of peptides exhibiting super antigenic property to a maximum number of HLA alleles with their conserved nature.

The present study has been designed for a meticulous analysis of predicted HLA binding peptides of HTLV-1 viral proteins of all the isolates. The schematic diagram of the entire methodology adopted in this work has been given in Figure 1.

Methods

Retrieval of protein sequences

The protein sequence ODS of Envelop glycoprotein of six strains of Human T Lymphotropic Virus-I: Japan MT-2 subtype A (ENV_HTL1M Envelope glycoprotein gp62, P23064), Carribbea HS-35 subtype A (ENV_HTL1C Envelope glycoprotein gp62, P14075), Melanesia mel5 subtype C (ENV_HTL1L Envelope glycoprotein gp62, P0C212), Japan ATK-1 subtype A (ENV_HTL1A Envelope glycoprotein gp62, P03381), Zaire EL subtype B (ENV_HTL1F Envelope glycoprotein gp62, Q03817) and Caribbea CH subtype A (ENV_HTL1N Envelope glycoprotein gp62, Q03816) were retrieved from UniProtKB Database (www.uniprot.org). The retrieved sequences were used for various bioinformatics approaches such as 3D-structure prediction, antigenicity assessment, and epitope prediction.

Immunogenicity prediction of viral protein and structural analysis

An antigen, which can be recognized by the immune system is called Immunogenic. VaxiJen v2.0 server [18] (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) was used for analyzing 'Env' protein immunogenicity of HTLV-1 of all the genotypes from different geographical areas by finding the best potential candidate immunogen epitope. VaxiJen is a web server to predict the protective antigens, tumor antigens and subunit vaccines, with an accuracy of approximately 70 to 89% for the discrimination between antigens and non-antigens [18]. The server is based on Auto Cross Covariance (ACC) conversion of protein sequences into uniform vectors of principal amino acid properties. It is being considered as the first server predicting the protective antigens in an alignment-independent manner.

The crystal or NMR structure of viral envelop protein was not available in Protein Data Bank (PDB), thus selected sequences (from UniProtKB Database) were used to develop the three-dimensional (3D) structure. In order to build the 3D model, I-TASSER online server [19] was used, which employs a fragment based method for the modeling of the protein structure. The server predicted five models along with their C-score. The models with highest C-score value, which is directly proportional to the confidence level of the model were selected. The modeled structures were validated by the approach of Ramachandran Plot using RAMPAGE [20]. Furthermore, the generated model was submitted to ModRefiner for atomic level high resolution protein structure refinement.

Multiple sequence alignment

Using ClustalW (www.ebi.ac.uk/clustalw) multiple sequence alignment program, best match (conserved regions) for the selected sequences were calculated. ClustalW is a multiple sequence alignment program for DNA or Proteins. It calculates the best matches for the selected sequences & arranges them in order of the identities, similarities & differences seen.

Epitope prediction

Humoral immunity associated epitope prediction

For all of the six strains of HTLV-I, B-cell epitopes were predicted using ABCPred online server (www.imtech.res.in/abcpred), selecting the threshold 1.0 and window length 20. The server is based on information-processing paradigm "neural network" inspired by the mammalian brain. All the predicted epitopes were ranked as per their respective scores. Higher score of the peptides depicted the greater probability to be the most suitable (immunogenic) epitope for the human B-Lymphocytes.

CD4+ and CD8+ epitope prediction

The peptides capable of stimulating T-cell epitopes bind to MHC molecules. Most important attribute of MHC molecules is their allelic polymorphism. The ImMunoGeneTics/HLA (IMGT) database listed 3,411 human leukocyte antigens (HLA) class-I and 1,222 HLA class II molecules in July 2010 [21].

Envelop protein sequences of all the six strains were analyzed for the screening of possible dominant T cell epitopes using immunoinformatics tool IEDB online server (www.iedb.org). MHC-I binding and MHC-II binding resources available in IEDB were used for the binding analysis of all the possible peptides of 18 class I HLA alleles: A*0101, HLA-A*03:01, A*26:01, A*02:01, A*11:01, A*68:01, B*08:01, B*37:01, B*07:02, B*14:02, B*27:05, B*38:01, B*39:01, B*40:02, B*44:02, B*48:01, B*49:01, B*51:01, and 6 class II HLA alleles: DRB1*01:01,

Table 1. Overall immunogenecity prediction of all isolates of HTLV-I

Strain	*UniProt ID*	*Threshold*	*Overall antigen prediction*	*Antigen*
Japan MT-2 subtype A	P23064	0.4	0.5004	Probable-Antigen
Caribbea HS-35 subtype A	P14075	0.4	0.5161	Probable-Antigen
Melanesia mel5 subtype C	P0C212	0.4	0.5015	Probable-Antigen
Japan ATK-1 subtype A	P03381	0.4	0.4928	Probable-Antigen
Zaire EL subtype B	Q03817	0.4	0.4830	Probable-Antigen
Caribbea CH subtype A	Q03816	0.4	0.4978	Probable-Antigen

DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*07:07, DRB1*15:01 described earlier [17].

The peptide length was set 10 and 15 for MHC-I and MHC-II, respectively. Default IEDB recommended prediction method was selected as this selection tries to use the best possible method that combines the predictions from ANN, SMM, and CombLib. The epitopes are predicted on the basis of percentile rank; the low percentile rank, the high binding affinity.

HLA distribution analysis

The analysis of population coverage in the context of MHC polymorphism becomes typical as different HLA types are expressed at considerably different frequencies in different ethnicities. Due to the MHC dependence of T-cell response, the peptides with more different HLA binding specificities result in more population coverage in the geographical regions, where the peptide-based vaccine might be employed. The population coverage rate of the epitopes was calculated using the IEDB population coverage tool (http://tools.immuneepitope.org/tools/ population/iedb_input) [22]. Each epitope with its binding HLA alleles was added, and different geographic areas were also included in the analysis.

Peptide modelling

B-cell and T-cell epitopes were designed using the program Discovery Studio 2.5. It was developed and distributed by Accelrys, USA. Discovery Studio is a software suite that has been frequently used to analyze and model molecular structures, sequences, and other information relevant to life science researchers. It includes utility for viewing and editing the data along with the tools for conducting the basic data analysis.

Protein-peptide interaction

Docking analyses between receptor (human HSPG) and T-cell epitopes (predicted through IEDB) were performed using ZDOCK online server having great prediction abilities in Critical Assessment of Predicted Interactions (CAPRI). ZDOCK is authored & maintained by ZLAB (Zhipping Weng's Lab). It is a rigid-body molecular docking program based on Fast Fourier Transform algorithm. It analyses all the possible binding modes in the translational and rotational space between the two proteins and evaluates each pose on the basis of an energy-based scoring function. The resulting z-scores with highest value were used as appropriate conformational pose.

Table 2. Score obtained through alignment of sequences

Sequences	*Aligned score*
1:2	98.7705
1:3	99.3852
1:4	98.5656
1:5	99.3852
1:6	99.1803
2:3	98.1557
2:4	97.3361
2:5	98.1557
2:6	97.9508
3:4	97.9508
3:5	100
3:6	98.5656
4:5	97.9508
4:6	97.7459
5:6	98.5656

Antigenicity of identified epitopes

Both B-cell and T-cell predicted epitopes were checked for their immunogenicity using VexiJen v2.0 online server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html).

Epitope physio-chemical property calculations

Peptide Property Calculator (https://www.genscript.com) server was used to determine the best solvent for the predicted peptide. The working methodology of this server is based on its (peptide's) amino acid sequence.

Virtual scanning of toxic/non-toxic peptides and allergenicity assessment

The final epitope set was submitted to ToxinPred server to identify the toxic/non-toxic peptides. This server predicts the toxicity along with all the important physio-chemical properties. A quantitative matrix was generated on the basis of probability or frequency of amino acid at a particular position. The performance was assessed by using 5-folds cross validation technique. Further, the AllerHunter server was employed to analyze the allergenicity of the epitopes (http://tiger.dbs.nus.edu.sg/AllerHunter). AllerHunter is a support vector machine (SVM)-pair wise system that assesses the al-

Table 3. Ramachandran Plot calculation of 3D model of Human receptor and viral protein sequences

Protein	*Ramachandran plot statistics*		
	% residues in most favored regions	*% residues in allowed regions*	*% residues in outlier regions*
Human HSPG	91.5	6.5	2.0
Japan MT-2	94.2	4.9	0.8
Caribbea HS-35	92.2	6.2	1.6
Malensia Mel5	92.8	5.8	1.4
Japan ATK-1	91.2	7.8	1.0
Zaire EL	93.2	5.4	1.4
Caribbea CH	93.6	4.7	1.6

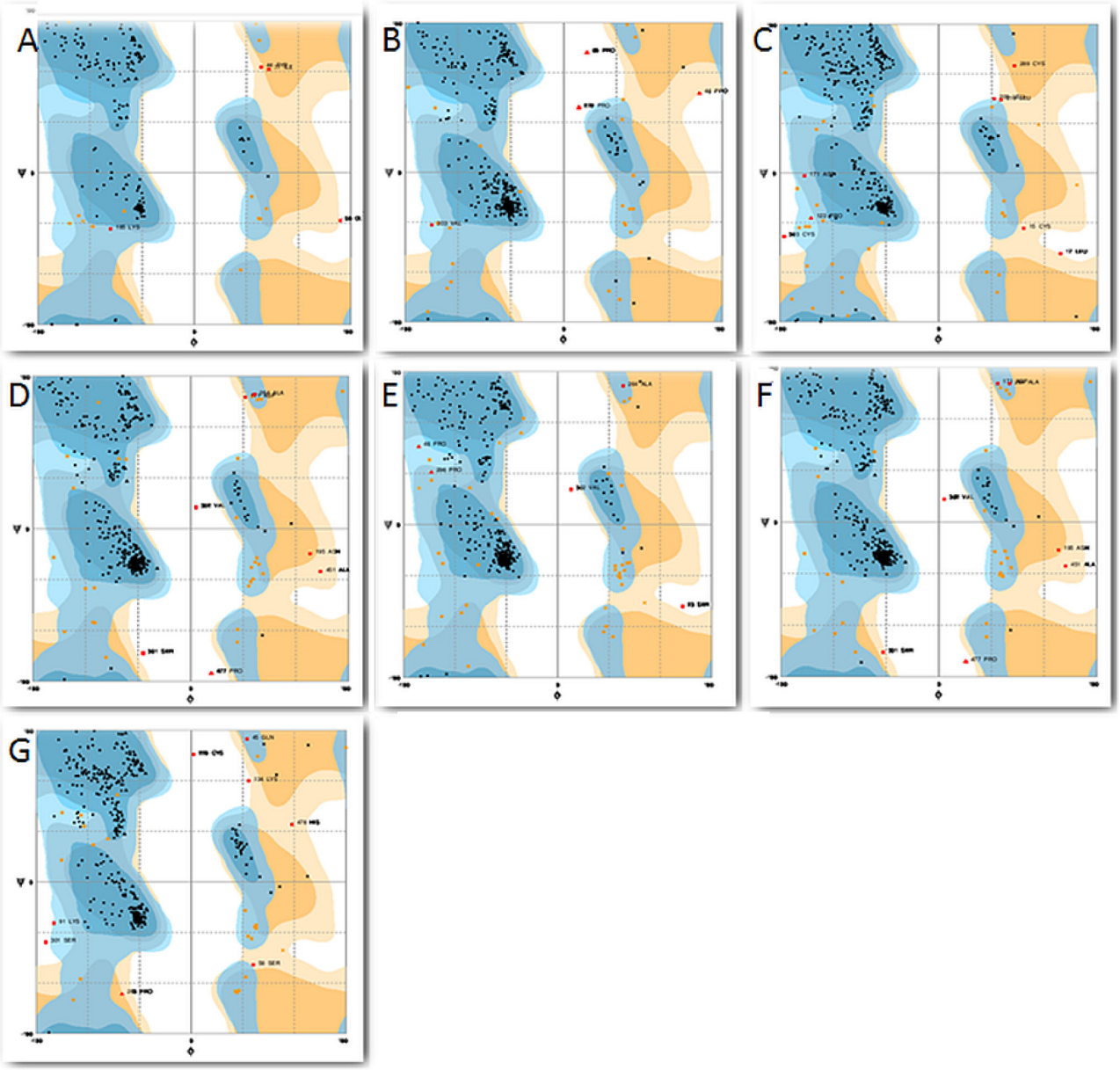

Figure 2. Structural validation of modeled protein using RAMPAGE: **(A)** Human HSPG, **(B)** Japan MT-2, **(C)** Caribbea HS-35, **(D)** Malensia Mel5, **(E)** Japan ATK-1, **(F)** Zaire EL, **(G)** Caribbea CH.

lergenicity and allergic cross-reactivity in the proteins. AllerHunter predicts allergens and non-allergens with high sensitivity and specificity, without compromising the efficiency at the classification of proteins with similar sequences to known allergens [23].

Results

Immunogenic protein identification and genetic variability analysis

The selection of the immunogen is considered as the preliminary step of vaccine development. The said selection was performed on the sequences of all the isolates retrieved for 'Env' protein. The results revealed all the isolates to be immunogenic (Table 1) and genetically highly conserved (Table 2). The VaxiJen v2.0 online server was used to determine the immunogenicity of protein on the basis of overall score produced by the specific protein sequence.

Structure validation

Once the sequences were established as immunogenic, the 3D-structures of 'Env' proteins were predicted using I-TASSER. The structures were further validated with Ramachandran plot analysis using 'RAMPAGE' as shown in Figure 2 and Table 3. The I-TASSER modelled structure is depicted in Figure 3.

Identification of humoral immunity associated epitope

One of the key steps in epitope-based vaccine designing is the prediction and identification of the B-cell epitopes in the target antigens. The ABCPred

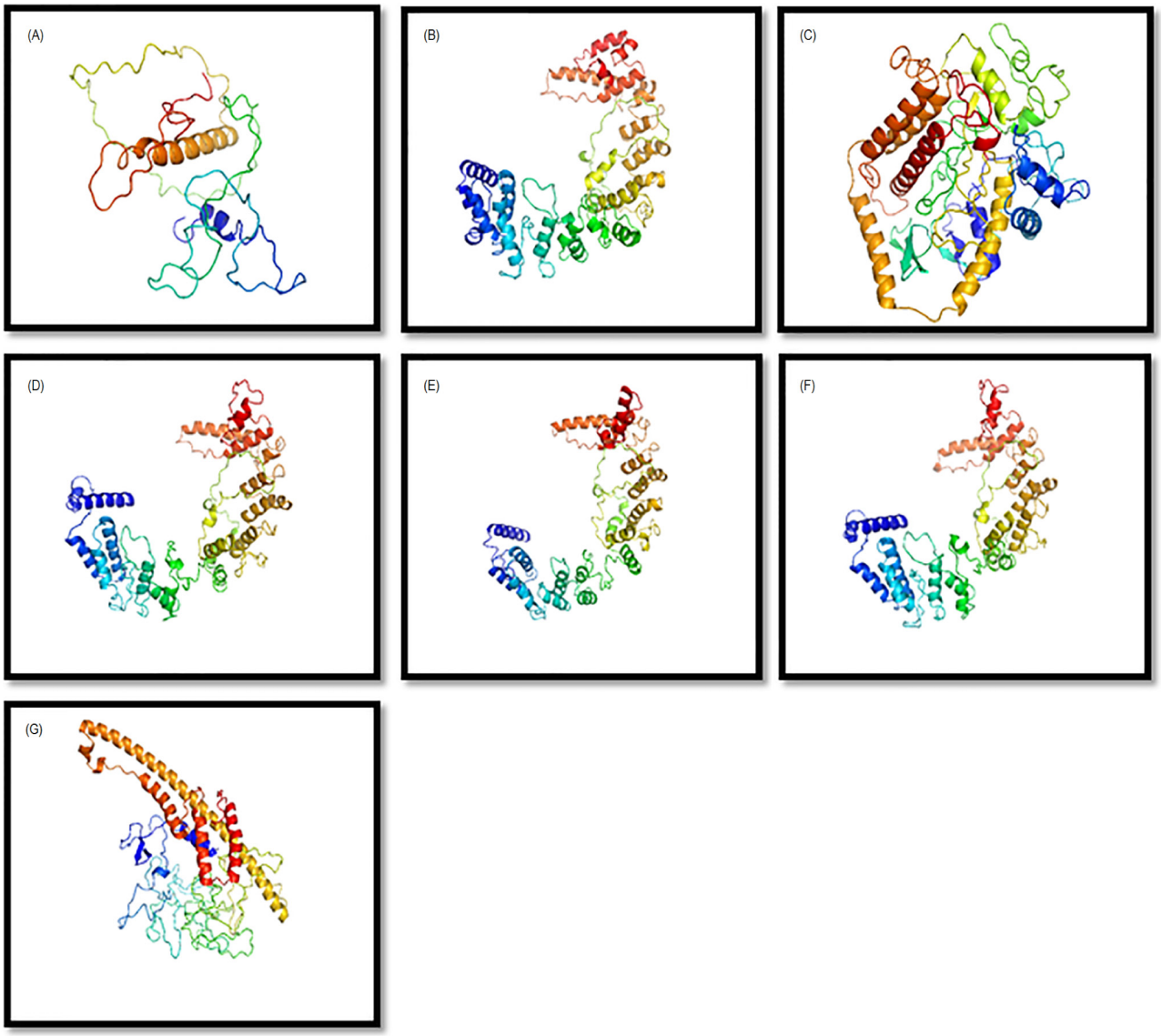

Figure 3. 3D structure modeled using I-TASSER and refined by ModRefiner. **(A)** Human HSPG, **(B)** Japan MT-2, **(C)** Caribbea HS-35, **(D)** Malensia Mel5, **(E)** Japan ATK-1, **(F)** Zaire EL, **(G)** Caribbea CH.

database was used to obtain the B-cell epitope candidates in the 'Env' protein of HTLV-1. Based on Artificial Neural Network method of the ABCPred server, the number of epitopes identified for different isolates were as follows:

11 potential epitopes for Japan MT-2; 10 potential epitopes for Caribbea HS-35; 11 potential epitopes for Malesia mel5; 11 potential epitopes for Japan ATK-1; 11 potential epitopes for Zaire EL; 12 potential epitopes for Caribbea CH.

Humoral immunity associated epitopes predicted by ABCpred were in descending order of their score, depicting the top scorer at the highest point having the best binding affinity. Selecting the first (top) epitope for each strain has given a total of six epitopes. Out of the 6 B-cell epitopes, 5 had 100% sequence conservation. Therefore, a total of two B-cell epitopes *"PSQLPPTAPPLLPHSNLDHI"* and *"EPSQLPPTTPPLLPHSNLDH"* with score 1 at position 180 and 179, respectively, showing highest probability to be as antigenic determinants, were identified.

Identification of CD4+ and CD8+ epitope

MHC-II restricted CD4+ T-cells activation plays an important role for initiating and upholding a proficient antibody response or Cytotoxic T Lymphocytes (CTL) response. On the other hand, MHC-I restricted CD8+ CTLs has a vital role in combating the viral infection. Thus, comprehending the principles of T cell activation and epitope-based vaccine design directs the identification of helper T-cell epitopes & CTL epitopes. Hence, the helper T-cell epitopes & CTL epitopes in the 'Env' protein of HTLV-1 were identified using *in silico* approach. Here, the immunogenic peptides, obtained with recommended methods were ranked according to their consensus percentile rank from SWISS-PROT proteins. The lowest percentile rank demonstrated the best binders. Selecting the first (lowest percentile rank) epitope for each of the 6 alleles of MHC CLASS II and 18 alleles of MHC CLASS I has given 6 helper T-cell epitopes and 18 CTL epitopes for each of the six strains of HTLV-I using IEDB.

Table 4. HLA distribution and population coverage analysis results

Strain	*Coverage, %*	*Average hit*	*PC90*
Japan MT-2 subtype A	95.15	2.21	1.23
Caribbea HS-35 subtype A	96.41	2.39	1.35
Melanesia mel5 subtype C	96.41	2.39	1.35
Japan ATK-1 subtype A	96.41	2.39	1.35
Zaire EL subtype B	96.41	2.39	1.35
Caribbea CH subtype A	96.41	2.39	1.35

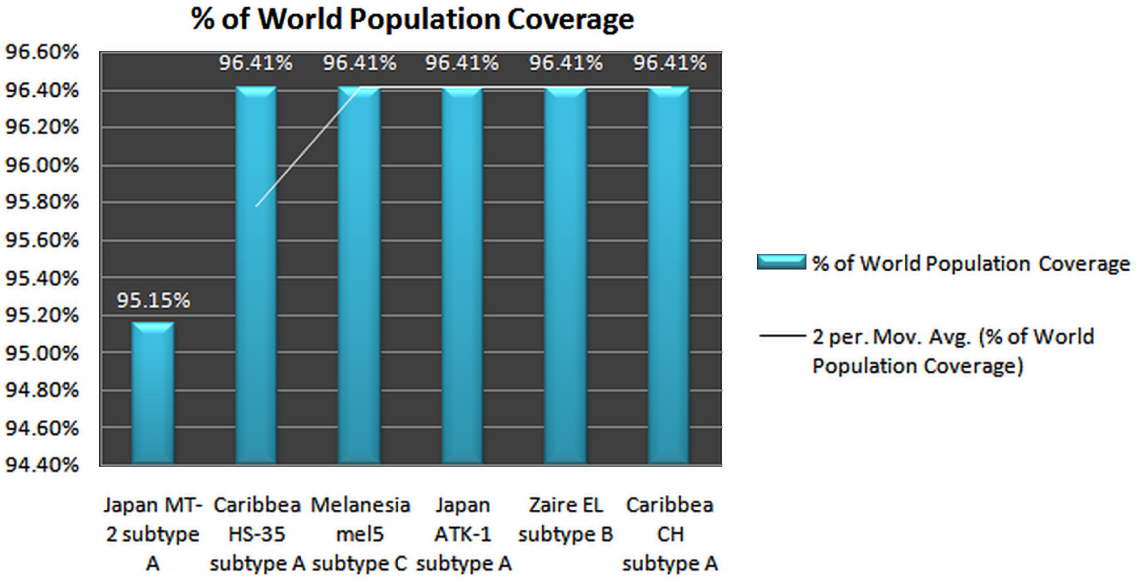

Figure 4. Graphical representation of predicted epitope / HLA distribution among different isolates of HTLV-I.

Table 5. HSPG-peptide docking results

S. No.	*Epitope*	*Z dock-Score*	*Intermolecular Hydrogen Bond*	*Hydrogen Bond Distance*
CLASS II (CD4+)				
1	LVALLLLVILAGPCI	1579.921	:LEU79:N - A:GLY12:O	3.16958
2	SSPYWKFQHDVNFTQ	1662.858	:ALA4:N - A:SER2:OG	2.15667
			:TRP5:N - A:SER2:OG	2.72544
			A:SER1:N - :ARG3:O	3.09156
			A:SER2:N - :ARG2:O	2.7973
			A:SER2:N - :ARG3:O	2.05351
			A:SER2:OG - :ARG2:O	3.05556
			A:SER2:OG - :ARG3:O	2.37172
			A:ASN12:N - :GLU126:OE1	2.85617
			A:ASN12:N - :GLU126:OE2	2.95054
3	YWKFQQDVNFTQEVS	1612.931	:TYR44:OH - A:ASP7:O	2.35794
			:THR80:N - A:THR11:OG1	1.87384
			:GLY182:N - A:TRP2:O	3.05172
			:SER188:N - A:PHE4:O	2.76898
			:SER188:OG - A:PHE4:O	2.88083
			A:VAL1:N - :ASP180:OD1	1.17792
			A:TRP2:N - :ASP180:OD2	2.92515
			A:TRP2:NE1 - :MET171:O	2.1366
			A:TRP2:NE1 - :ARG172:O	2.99427
			A:TRP2:NE1 - :LYS173:O	3.124
			A:LYS3:NZ - :GLU183:OE2	1.74793
			A:GLN5:N - :SER188:OG	2.71964
			A:VAL8:N - :TYR44:OH	3.07846
			A:THR11:OG1 - :LEU79:O	2.5652
			A:THR11:OG1 - :THR80:O	2.94651
4	YSSYHATYSLYLFPH	1874.750	A:THR7:N - :ASP48:OD1	2.26765
			A:TYR8:N - :ASP48:OD1	2.16293
			A:TYR8:OH - :ARG72:O	2.75544
			A:SER9:N - :ASP47:OD2	2.46973
			A:SER9:OG - :ASP47:OD1	2.3081
			A:SER9:OG - :ASP47:OD2	1.82501
			A:TYR11:OH - :TYR191:O	3.02424
5	PCPNLVAYSSYHATY	1607.413	:MET1:N - A:LEU5:O	3.18292
			:TRP5:NE1 - A:SER7:OG	1.98532
			:TRP5:NE1 - A:SER7:O	2.92805
			:TYR44:OH - A:HIS12:O	3.13358
			:GLY156:N - A:SER10:OG	2.83934
			A:LEU5:N - :SER103:O	2.91771
			A:SER10:OG - :ILE155:O	2.01553
			A:HIS12:N - :TYR44:OH	3.06947
			A:ALA13:N - :TYR44:OH	2.5348

Continued on the next page

S. No.	*Epitope*	*Z dock-Score*	*Intermolecular Hydrogen Bond*	*Hydrogen Bond Distance*
CLASS I (CD8+)				
6	SLYLFPHWIK	1666.732	:LEU79:N - A:LYS10:O	2.39997
			:TYR131:OH - A:LEU2:O	2.40817
			:THR196:OG1 - A:LEU4:O	2.22273
			:LYS197:N - A:TYR3:OH	2.85881
			:LYS197:N - A:LEU4:O	2.95916
			:GLU198:N - A:LEU4:O	2.28237
			A:TYR3:N - :TYR131:OH	1.21732
			A:TYR3:OH - :LYS197:O	2.9753
			A:PHE5:N - :GLU198:OE2	2.60355
			A:HIS7:N - :GLY153:O	2.41912
			A:TRP8:NE1 - :GLY156:O	3.12138
7	LHFSKCGFPF	1671.469	:TYR44:OH - A:PHE3:O	2.54674
			:SER178:OG - A:PHE10:O	2.77812
			:SER178:OG - A:PHE10:OXT	1.86774
			:ASP180:N - A:PHE10:O	2.83619
			:SER188:OG - A:CYS6:SG	2.50789
			A:SER4:N - :TYR44:OH	2.76142
			A:CYS6:SG - :SER188:OG	2.50789
			A:PHE10:N - :MET171:O	2.943
			A:PHE10:N - :TYR179:O	2.46562
			A:PHE10:N - :ASP180:OD2	1.70985
8	YHATYSLYLF	1715.898	:MET1:N - A:HIS2:O	2.22991
			:TRP5:NE1 - A:THR4:O	1.93176
			:TYR44:OH - A:LEU9:O	3.09167
			:THR101:OG1 - A:TYR1:OH	2.18107
			A:TYR1:OH - :THR101:OG1	2.18107
			A:LEU9:N - :TYR44:OH	2.64379
			A:PHE10:N - :TYR44:OH	2.19778

Population coverage of epitopes

HLA allocation varies among the diverse ethnic groups and geographic regions around the world. Thus, while designing an effective vaccine, population coverage must be taken into account to cover the maximum possible populations [24]. All the alleles in the supplementary data were identified as the best binders with the predicted epitopes and were used to establish the population coverage for these epitopes (Figure 4). Almost complete population coverage was found for all the putative helper T-cell epitopes and CTL epitopes in six specified geographic regions of the world (Table 4).

Protein peptide docking and CD4+ and CD8+ epitope screening

When selected CD4+ and CD8+ epitope (best five based on best binding affinity; high energy score, between receptor and epitope) for all the six strains of HTLV-I, were analyzed, we found that several of these epitopes had 100% sequence identity. Hence, based on the sequence conservation, a total of eight domains (five CD4+ & three CD8+) were screened out to be used as helper T-cell epitopes: *"YSSYHATYSLYLFPH", "PCPNLVAYSSYHATY", "YWKFQQDVNFTQEVS", "SSPYWKFQHDVNFTQ", "LVALLLLVILAGPCI"*, and CTL epitopes: *"SLYLFPHWIK",*

"LHFSKCGFPF", and "YHATYSLYLF", which may serve as useful diagnostic reagents for evaluating immune response against all the strains of HTLV-I (Figure 5 and Table 5).

Further, the peptides with strong antigenicity are more likely to be B-cell & T-cell epitopes than those with weak antigenicity. Hence, VaxiJen v 2.0 server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) was used to predict the antigenicity of the selected epitopes. This tool predicts the antigenicity of a peptide on the basis of amino acid properties and their positions in the peptide [18].

The obtained results showed that both the predicted B-cell epitopes had antigenic properties and they can be crucial in eliciting the desired immune response, whereas, out of the predicted 8 T-cell epitopes, 5 (four for CD4+ and one for CD8+) were found to be probable antigen (Tables 6 and 7).

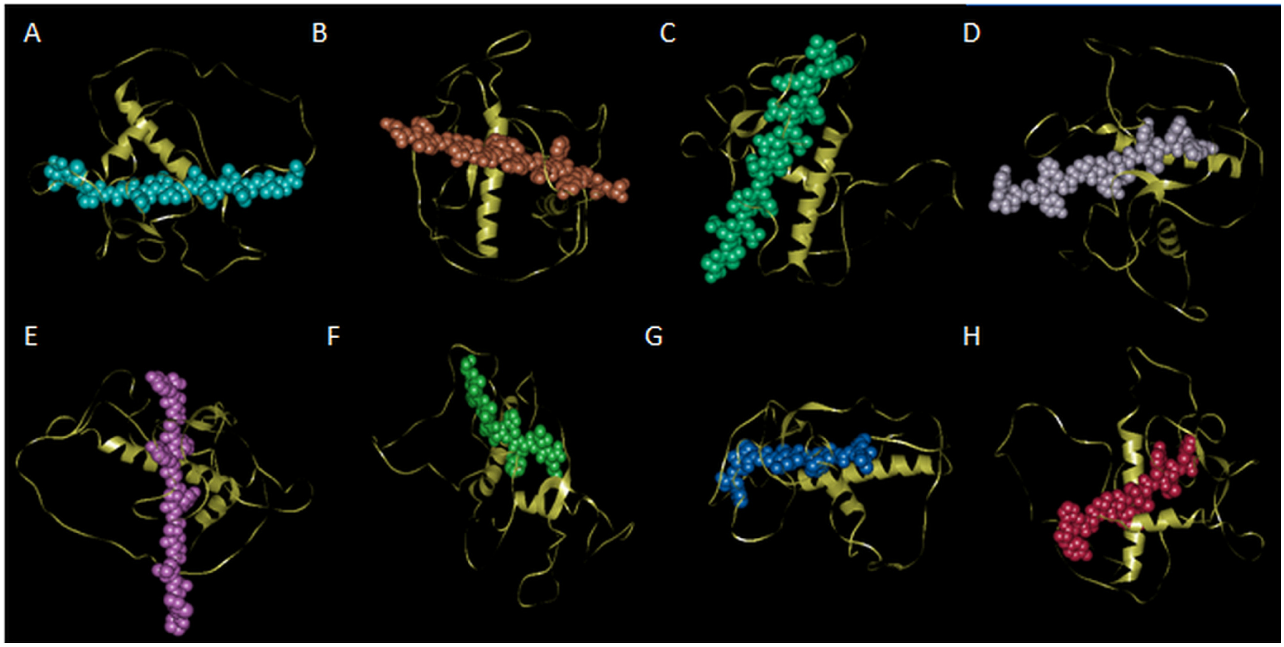

Figure 5. Docked (Peptide-HSPG) structures generated through Discovery Studio 2.5. Epitopes having best binding affinity (highest Z-dock-score) docked with HSPG. **(A)** LVALLLLVILAGPCI-HSPG, **(B)** SSPYWKFQHDVNFTQ-HSPG, **(C)** YWKFQQDVNFTQEVS-HSPG, **(D)** YSSYHATYSLYLFPH-HSPG, **(E)** PCPNLVAYSSYHATY-HSPG, **(F)** SLYLFPHWIK-HSPG, **(G)** LHFSKCGFPF-HSPG, **(H)** YHATYSLYLF-HSPG

Table 6. Humoral immunity associated epitope with their antigenic score

Position	*Humoral associated epitope*	*Antigenic score*	*Antigen/Non-antigen*
180	PSQLPPTAPPLLPHSNLDHI	0.5190	Antigen
179	EPSQLPPTTPPLLPHSNLDH	0.4374	Antigen

Table 7. CD4+ and CD8+ epitope with their antigenic score

Position	*Allele*	*CD4+ and CD8+ epitope*	*Antigenic score*	*Antigen/ Non-antigen*
CD4+ (CLASS II)				
449-463	HLA-DRB1*01:01	LVALLLLVILAGPCI	0.3352	Non-Antigen
129-143	HLA-DRB1*04:01	SSPYWKFQHDVNFTQ	0.8936	Antigen
132-146	HLA-DRB1*04:01	YWKFQQDVNFTQEVS	0.7403	Antigen
73-87	HLA-DRB1*07:07	YSSYHATYSLYLFPH	0.5073	Antigen
66-80	HLA-DRB1*15:01	PCPNLVAYSSYHATY	0.9356	Antigen
CD8+ (CLASS I)				
81-90	HLA-A*03:01	SLYLFPHWIK	-0.0424	Non-Antigen
152-161	HLA-B*14:02	LHFSKCGFPF	0.1898	Non-Antigen
76-85	HLA-B*38:01	YHATYSLYLF	0.5075	Antigen

Among them, the highest scored antigenic epitopes for humoral immunity associated and CD4+ and CD8+ (MHC-II & MHC-I) were *PSQLPPTAPPLLPHSNLDHI*, *PCPNLVAYSSYHATY* and *YHATYSLYLF*, respectively.

Physio-chemical properties of final epitope set

The overall interactive property calculation (Table 8) of most antigenic epitopes revealed that highest scored antigenic epitopes *PSQLPPTAPLLPHSNLDHI*, *PCPNLVAYSSYHATY* and *YHATYSLYLF*, were all suitable to dispense in the basic medium, thus, predicted antigenic epitopes (BASIC), solubilized in 10% or higher acetic acid solution or by adding <50 µl trifluoroacetic acid (TFA) and diluted to 1 ml with deionized water. This information can be used to make suitable diluent for their efficient use as peptide vaccine.

Toxicity and allergenicity screening of the identified epitopes

The final epitope set (CD4+, CD8+ and humoral associated) with best antigenic score were further confirmed for their toxicity. The obtained results are shown in Table 9. Analysis revealed that among the best antigenic peptides none (epitopes) were found allergic against humans.

Discussion

HTLV-I is an infectious virus and causes several pathogenic conditions including fatal T-cell lymphoma. Recent studies have reported that the entry of HTLV-I in humans is mediated by the attachment of the viral envelope (gp46) to the human receptor HSPG. Latest advancements in the field of bioinformatics have allowed the application of immunomics as an alternative approach to the traditional vaccine development strategy. The online accessibility of the complete genome sequences has transformed the approach to *in silico* vaccine development. T lymphocytes are crucial in eliciting a protective immune response against various microbial infections, thus the identification and selection of T-cell epitopes as vaccine targets is a prerequisite task for designing the epitope-based vaccines. The determination of B-cell epitopes is a crucial task for the vaccine development as well as for better diagnosis. Recently, synthetic peptides have been shown to induce antibodies reactive with their associated sequences in the native proteins.

As a first step in defining the B- and T-cell epitopes in HTLV-I, the search of RCSB Protein Data Bank (http://www.rcsb.org) verified that the required tertiary structures of human protein HSPG

Table 8. Analysis of Epitope Physio-Chemical property calculation

Epitopes	*Chemical formula*	*M. Wt*	*Iso-Electric point*	*Hydrophilicity analysis*	
				Charge	*Attribute*
Humoral immunity associated					
PSQLPPTAPPLLPHSNLDHI	$C_{98}H_{154}N_{26}O_{28}$	2144.44	6.50	1	Basic
EPSQLPPTTPPLLPHSNLDH	$C_{98}H_{152}N_{26}O_{31}$	2190.42	5.30	0	Neutral
CD8+					
YHATYSLYLF	$C_{64}H_{84}N_{12}O_{16}$	1277.43	7.53	1	Basic
CD4+					
SSPYWKFQHDVNFTQ	$C_{88}H_{118}N_{22}O_{25}$	1884.02	7.54	1	Basic
YWKFQQDVNFTQEVS	$C_{89}H_{123}N_{21}O_{27}$	1919.06	4.19	-1	Acidic
YSSYHATYSLYLFPH	$C_{90}H_{117}N_{19}O_{24}$	1849.01	7.68	2	Basic
PCPNLVAYSSYHATY	$C_{77}H_{108}N_{18}O_{24}S_1$	1701.86	7.34	1	Basic

Table 9. Toxicity analysis of predicted best antigenic epitope set

Peptide sequence	*SVM score*	*Toxicity prediction*	*Hydrophobicity*	*Hydropathicity*
PCPNLVAYSSYHATY (CD4+ epitope)	-0.64	Non-Toxin	-0.01	-0.13
YHATYSLYLF (CD8+ epitope)	-1.03	Non-Toxin	0.10	0.33
PSQLPPTAPPLLPHSNLDHI (Humoral associated epitope)	-1.36	Non-Toxin	-0.04	-0.37

and 'Env' protein of HTLV-I was not publicly available. The FASTA format amino acid sequences of the above mentioned proteins were retrieved from UniProt (http://www.uniprot.org).

The immunogenicity of each strain of HTLV-I was confirmed with the help of Vaxijen online server. The alignment of the amino acid sequences among six strains of HTLV-I revealed the overall identity of the structural proteins (among the isolates, 'Env' proteins are known to be genetically highly conserved). Following the antigenicity confirmation, I-TASSER was employed to predict the 3D-structure. The refinement of the generated model was performed by using ModRefiner followed by the structure validation by using RAMPAGE. The predicted 3D-structures provided more insight in comprehending the structure and functions of the proteins and were found fit for the interaction analysis.

The candidates for peptide synthesis were derived from the protein sequence with the help of predictive computational algorithms and HLA supertypes. Among all the predicted B-cell epitopes, selecting first (top) from each strain criterion was followed for generating 6 epitopes. Five out of 10 epitopes were conserved, and finally we had 2 B-cell epitopes.

Helper T-cell epitopes are crucial for effective humoral and CTL responses. Although, the response to T-cell epitope is dependent upon HLA proteins, yet the fact is that HLA is highly polymorphic in various ethnic populations. Hence, the HLA specificity of T-cell epitopes must be the principal criterion for screening of the epitopes as vaccine candidates to cover the genetically diverse human population [25,26]. The HLA supertype concept has a substantial role in the understanding of CD4+ and CD8+ epitope's selection, discrimination during T-cell mediated immune response and degeneration. In the HLA supertype analysis, IEDB web server was used to check the binding of the best epitopes with those of HLA alleles; and 18 MHC-I and 6 MHC-II peptides were found in all the known genotypes, as it is established that an epitope elicits a response only if it is capable of binding a MHC molecule in an individual. This phenomenon is called as denominated MHC restriction of T-cell responses. Including multiple peptides with different HLA binding specificities increases the coverage of the population targeted by the peptide-based vaccines or diagnostics. We have performed HLA distribution and population coverage analysis for the predicted epitopes that showed 96.41% of the projected population coverage with 2.39 average numbers of epitope hits/HLA combinations recognized by the population.

It was found that HLA specific predicted docked peptides were highly conserved among the six genotypes, hence, resulted a total of 8 (5 CD4+ and 3 CD8+) epitopes for all the selected genotypes. Among all the predicted humoral associated epitopes, 2 epitopes *"PSQLPPTAPPLLPHSNLDHI"* and *"EPSQLPPTTPPLLPHSNLDH"* were found to be antigenically effective and among 8 T-cell epitopes, 5 epitopes i.e., 1 CD8+ *"YHATYSLYLF"*, and 4 CD4+ *"SSPYWKFQHDVNFTQ"*, *"YWKFQQDVNFTQEVS"*, *"YSSYHATYSLYLFPH"* and *"PCPNLVAYSSYHATY"* were found to be antigenically effective. Further, among these antigenic epitopes, 1 epitope *"PCPNLVAYSSYHATY"* at position 66-80 had the highest antigenic score (0.9356) ensuring maximum immune response.

Additionally, to check the biocompatibility of the best (1 each top scorer from B-cell, MHC-II and MHC-I) antigenic epitopes *"PSQLPPTAPPLLPHSNLDHI"*, *"PCPNLVAYSSYHATY"* and *"YHATYSLYLF"* were tested by Epitope Physio-Chemical property calculator. The analysis revealed that these epitopes have basic attributes to dissolve in water and if not then in 10% acetic acid; otherwise TFA (<50 µl). These properties can be used to make a suitable diluent to dispense the peptide vaccine. In order to evaluate the toxicity of the final epitope set, we employed virtual screening of the peptides using ToxinPred server (that uses 5-folds cross validation technique). The obtained results strongly predicted that all the three epitopes were non-toxic, non-allergic and safe for use.

The application of the predicted potential epitopes instead of whole protein as a vaccine has already been established and currently it is in enormous use as the predicted epitopes are fully capable of representing the complete antigenicity of any protein [27]. Epitope-based vaccines have given promising results against the highly infectious diseases such as H1N1, HIV, and Tuberculosis [28]. These vaccines have low chance of eliciting immune response against the self-antigens, thus evade autoimmune response [29]. The conventional molecular immunology techniques for the search of immunogenic peptide to be used as vaccine candidates are exorbitantly time-taking and expensive. Therefore, the present *in silico* study may help out the experimental biologists by providing assistance in the rapid screening process to facilitate the identification of the probable vaccine candidates. Lately, these computational approaches are recently being used by many researchers across the globe [30]. *In silico* approach reported in the current study for T-cell epitope prediction and population coverage analysis not only reduces the cost in an efficient manner, but also the time and

efforts invested in the planning and execution of the experiments and subsequent development of epitope-based vaccines. The application of *in silico* approaches by immunologists has helped to speed-up the *in vitro*, *in vivo* and *ex vivo* studies by cutting down the large epitope dataset to a smaller dataset that warrants their experimental validation.

In conclusion, genetic variability and HLA distribution analysis of HTLV-1 genotypes I, II, III, IV, V and VI was done for the broader population coverage of the proposed vaccine. The screening of the effective dominant immunogens revealed that the identified epitopes were protective immunogens with high conservancy and bears potential of inducing both protective neutralizing antibodies and T-cell responses. The obtained peptides *"PSQLPPTAPPLLPHSNLDHI"*, *"PCPNLVAYSSYHATY"*, and *"YHATYSLYLF"* were fully conserved among different isolates from far and wide separated countries suggesting negligible antigenic drift in HTLV-1. In conclusion, the predicted epitopes are soluble, non-toxic suitable candidates for the vaccine development against HTLV-1 and warrants for further investigation and experimental validation. The prophecies made in the current study inferred from the primary amino acid sequence from various HTLV-I isolates would tremendously help the researchers involved in the experimental verification of the immunodominant epitopes that are being considered as the fundamental to the understanding of the pathobiology of HTLV-I and its effective ATL vaccine development.

Acknowledgements

The authors are grateful to the Integral University, Lucknow (UP), India and the Deanship of Scientific Research, Jazan University, Jazan, Saudi Arabia for providing the necessary dry-laboratory facilities for this research study.

Financial Statement

No direct/indirect financial support was available for this study.

Conflict of interests

The authors declare no conflict of interests.

References

1. Hoshino H. Cellular factors involved in HTLV-1 entry and pathogenecity. Frontiers Microbiol 2012;10:3389.
2. Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc Natl Acad Sci USA 1980;77:7415-9.
3. Verdonck K, González E, Van Dooren S, Vandamme AM, Vanham G, Gotuzzo E. Human T-lymphotropic virus 1: Recent knowledge about an ancient infection. Lancet Infect Dis 2007;7:266-81.
4. Osame M, Usuku K, Izumo S, Ijichi N, Amitani H . HTLV-I associated myelopathy, a new clinical entity. Lancet 1986;3:1031-43.
5. Mochizuki M, Yamaguchi K, Takatsuki K, Watanabe T, Mori S, Tajima K. HTLV-I and uveitis. Lancet 1992;339-52.
6. Proietti FA, Carneiro-Proietti AB, Catalan-Soares BC, Murphy EL. Global epidemiology of HTLV-I infection and associated diseases. Oncogene 2005;24:6058-68.
7. Shimoyama M. Diagnostic criteria and classification of clinical subtypes of adult T-cell leukemia-lymphoma. A report from the lymphoma study group (1984-1987). Br J Haematol 1991;79: 482-37.
8. Kannian P, Green PL. Human T-cell Lymphotropic Viruses Type-I (HTLV-I): Molecular Biology and Oncogenesis. Viruses 2010;2:2037-77.
9. Seiki M, Hattori S, Hirayama Y, Yoshida M. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc Natl Acad Sci USA 1983;80:3618-22.
10. Hattori S, Kiyokawa, T, Imagawa K et al. Identification of gag and env gene products of human T-cell leukemia virus (HTLV). Virology 1984;136:338-47.
11. Gessain A, Mahieux R, De TG. Genetic variability and molecular epidemiology of human and simian T-cell leukemia/lymphoma virus type-I. J Acquir Immune Defic Syndr Hum Retrovirol 1996;13:132-45.
12. Feuer G, Green PL. Comparative biology of human T-cell lymphotropic virus type-1 (HTLV-1) and HTLV-2. Oncogene 2005;24:5996-6004.
13. Ghez D, Lepelletier Y, Jones KS, Pique C, Hermine O. Current concepts regarding the HTLV-1 receptor complex. Retrovirology 2010;7:99.
14. Jones KS, Lambert S, Bouttier M et al. Molecular aspects of HTLV-1 entry: functional domains of the HTLV-1 surface subunit (SU) and their relationships to the entry receptors. Viruses 2011;3:794-810.
15. Reche PA, Reinherz EL. PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands. Nucleic Acids Res 2005;33:W138 [PMID: 15980443]
16. Zhang GL. PREDBALB/c: a system for the prediction of peptide binding to H2d molecules, a haplotype of

the BALB/c mouse. Nucleic Acids Res 2005;33:W172 [PMID: 15980449]

17. Mota-Miranda AC, Barreto FK, Amarante MF et al. Molecular characterization of HTLV-1 gp46 glycoprotein from health carriers and HAM/TSP infected individuals. Virology J 2013;10:75.
18. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 2007;8:4. [PMID: 17207271]
19. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 2008;9:40.
20. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins Structure Function Genetics 2002;50:437-50.
21. Robinson J, Mistry K, McWilliam H, Lopez R, Parham P. The IMGT/HLA database. Nucleic Acids Res 2011;39:1171-6.
22. 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. [PMID: 16545123].
23. Muh HC, Tong JC, Tammi MT. AllerHunter: a SVM-pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins. PLoS One 2009; 4(6):e5861. doi: 10.1371/journal.pone.0005861 PMID: 19516900
24. Shi J, Zhang J, Li S 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. doi:10.1371.
25. Maenaka K, Jones EY. MHC superfamily structure and the immune system. Curr Opin Struct Biol 1999;9:745-53.
26. Stern LJ, Wiley DC. Antigenic peptide binding by class I and class II histocompatibility proteins. Structure 1994;2:245-51.
27. De Groot AS, Sbai H, Aubin CS, McMurry J, Martin W. Immuno-informatics: Mining genomes for vaccine components. Immunol Cell Biol 2002;80:255-69.
28. Correia BE, Bates JT, Loomis RJ et al. Proof of principle for epitope-focused vaccine design. Nature 2014;507:201-6. doi: 10.1038/nature12966.
29. Azizi A, Anderson DE, Torres JV et al. Induction of broad cross-subtype-specific HIV-1 immune responses by a novel multivalent HIV-1 peptide vaccine in cynomolgus macaques. J Immunol 2008;180:2174-86.
30. Davies MN, Flower DR. Harnessing bioinformatics to discover new vaccines. Drug Discov Today 2007;12:389-95.