PLOS ONE

RESEARCH ARTICLE

Epitope-based universal vaccine for Human T-lymphotropic virus-1 (HTLV-1)

Md. Thosif Raza[1], Shagufta Mizan[1], Farhana Yasmin[1], Al-Shahriar Akash[1], Shah Md. Shahik[2¤]*

1 Faculty of Biological Sciences, Department of Genetic Engineering and Biotechnology, University of Chittagong, Chattogram, Bangladesh, **2** Bioinformatics Division, Disease Biology and Molecular Epidemiology Research Group, Chattogram, Bangladesh

¤ Current address: Molecular Biology Department, AFC Agro Biotech Ltd., Dhahka, Bangladesh
* sm.shahik@gmail.com, shah_shahik@iscb.org

Citation: Raza MT, Mizan S, Yasmin F, Akash A-S, Shahik SM. (2021) Epitope-based universal vaccine for Human T-lymphotropic virus-1 (HTLV-1). PLoS ONE 16(4): e0248001. https://doi.org/10.1371/journal.pone.0248001

Editor: William M. Switzer, Centers for Disease Control and Prevention, UNITED STATES

Received: September 11, 2020

Accepted: February 17, 2021

Published: April 2, 2021

Data Availability Statement: All relevant data will be made available upon acceptance of this manuscript for publication.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Human T-cell leukemia virus type 1 (HTLV-1) was the first oncogenic human retrovirus identified in humans which infects at least 10–15 million people worldwide. Large HTLV-1 endemic areas exist in Southern Japan, the Caribbean, Central and South America, the Middle East, Melanesia, and equatorial regions of Africa. HTLV-1 TAX viral protein is thought to play a critical role in HTLV-1 associated diseases. We have used numerous bio-informatics and immuno-informatics implements comprising sequence and construction tools for the construction of a 3D model and epitope prediction for HTLV-1 Tax viral protein. The conformational linear B-cell and T-cell epitopes for HTLV-1 TAX viral protein have been predicted for their possible collective use as vaccine candidates. Based on in silico investigation two B cell epitopes, KEADDNDHEPQISPGGLEPPSEKHFR and DGTPMISGPCPKDGQPS spanning from 324–349 and 252–268 respectively; and T cell epitopes, LLFGYPVYV, ITWPLLPHV and GLLPFHSTL ranging from 11–19, 163–171 and 233–241 were found most antigenic and immunogenic epitopes. Among different vaccine constructs generated by different combinations of these epitopes our predicted vaccine construct was found to be most antigenic with a score of 0.57. T cell epitopes interacted strongly with HLA-A*0201 suggesting a significant immune response evoked by these epitopes. Molecular docking study also showed a high binding affinity of the vaccine construct for TLR4. The study was carried out to predict antigenic determinants of the Tax protein along with the 3D protein modeling. The study revealed a potential multi epitope vaccine that can raise the desired immune response against HTLV-1 and be useful in developing effective vaccines against Human T-lymphotropic virus.

Introduction

Human T Lymphocyte Virus 1 or HTLV-1 is a type C retrovirus that possesses proteins capable of oncogenesis [1]. Belonging to the *Retroviridae* family (sub-family: *Orthoretrovirinae*; genus: Deltaretrovirus), HTLV-1 has been identified to be the first human virus with

oncogenicity [2]. Suspected of zoonotic origin, Human T Lymphocytic Viruses (HTLVs) come from a larger group of Primate T-lymphotropic viruses (PTLVs); their cousins being the Simian T-lymphotropic Viruses (STLVs) found in non-human primates (NHPs) [3–11]. To date, 4 types of HTLV viruses have been identified (HTLV-1, HTLV-2, HTLV-3, and HTLV-4) among which HTLV-1 is the most clinically active member [12]. According to a study in 2010, HTLV-1 infects approximately 20 million people globally [13]. Possible modes of transmission include breastfeeding (predominant), sexual intercourse, transfusion of infected blood components and sharing of needles and syringes [14]. Though the maximum of infected cases are asymptomatic, these individuals can still act as "Carriers" of the virus and are capable of transmitting the virus to a healthy individual [14].

HTLV-1 has etiological manifestation with severe inflammatory diseases like HTLV-1-Associated Myelopathy/Tropical Spastic Paraparesis (HAM/TSP), HTLV-1 associated Uveitis [12,15] and a lymphoproliferative disorder called Adult T Cell Leukemia/Lymphoma (ATL) [16–19]. HTLV-1 infected individuals have been found to be susceptible to other neurological [20], pulmonary [21], ophthalmological [22], rheumatological [23] and urological [24] coinfections because HTLV-1 can cause suppression of the immune system [25].

Approximately 100nm in diameter, HTLV-1 is an enveloped virus. The viral matrix protein (MA) lines the inner membrane of the virion envelope, the structure which encompasses the viral capsid (CA). The tenants of the capsid include two strands of genomic RNA (identical) and three enzymes; Pro (functional protease), IN (integrase) and RT (reverse transcriptase) [2,26]. The viral pathogenesis begins by membrane fusion between HTLV-1 and the target CD4+ T cells [27]. Cellular attachment is facilitated by a type of glycosaminoglycan known as Heparan Sulfate Proteoglycan [28] which is a commonly expressed cell surface molecule in case of mammalian cells [29] Glucose Transporter 1 (GLUT1) and Neuropilin-1 (NRP1—receptor for semaphorin-3A and VEGF-A165) have also been observed to play roles in cellular attachment [30,31]. Once the HTLV-1 virus infects a cell, the progression of the infection occurs by cell-to-cell transmission. This transmission process involves the formation of a viral synapse at the site of contact between an HTLV-1 infected T cell and a healthy T cell by polarization of a microtubule organizing center (MTOC) at a junctional point between the two cells.

The single-stranded, complex RNA genome of an HTLV-1 virus encodes a range of structural proteins including Gag, Pro, Pol (polymerase) which have roles in assembling the virion and maturation of the virus; Env, which assists the entrance of the viral nucleic acid inside the host and host transformation [13,32,33]; along with two important regulatory proteins Rex (encoded from ORFIII) and Tax (encoded from ORFIV). Among the two regulatory proteins, the Rex protein acts as a post-transcriptional regulator, Tax has been found to show oncogenic abilities [34–38] and has crucial roles in regulating viral transcription and transformation of T cells mediated by HTLV-1 [39,40]. Accessory genes p12 and p30 and their protein products p8 and p13 are encoded by ORFI and ORFII function in initiating the viral infection and play roles in creating persistence of the virus in animal models [41–44].

When a Human T-cell leukemia virus type 1 enters the human body through any of the transmission modes including transfusion, breastfeeding and sexual intercourse [14], HTLV-1 viruses by the help of membrane fusion facilitated by the virus's own membrane protein Env and the target cell's receptors insert their nucleic acid inside target CD4+ T cells (these cells are more susceptible as HTLV1 targets) [43]. The genomic RNA is reversely transcribed to DNA by reverse transcriptase enzyme (RT); following steps involving transportation of the viral DNA to the host nucleus and integration of the viral DNA with the host DNA. This amalgamated structure consisting of both the host and the viral DNA is termed as Provirus. The process of integration is worthwhile for the HTLV-1 virus as its copies are being made in spite of the whole viral structure not being present in the host by the help of the host's own

replication mechanism. A protein known as the HBZ protein (HTLV-1 basic leucine zipper factor) is encoded by the antisense transcript of the proviral genome [27] which upon interaction with the Jun family of AP-1 transcription factors JunB, c-Jun, JunD [45–47], cAMP response element binding (CREB) and CREB binding protein (CBP)/p300 regulate both cellular and viral gene transcription [48–50] and also plays a vital role in the proliferation of infected T cells [51–53].

Although the HTLV1 mediated pathogenesis is the cumulative result of multiple types of proteins, Tax protein has been found to be a key trigger element of a nuance of cellular events like resistance to apoptosis, cell signaling, cell cycle regulation and interference with checkpoint control and inhibition of DNA repair through its transactivational properties [26]. In patients with HTLV-1-Associated Myelopathy/Tropical Spastic Paraparesis (HAM/TSP), Tax protein is expressed at an elevated level regardless of the proviral load and therefore can be considered as an overt marker for HAM/TSP prognosis and a target for the development of therapeutics [1]. Tax protein has also been found to play significant role as a leukemogen and has successfully managed to arrest apoptosis in T-lymphocytes *in vitro* and cause cancer in transgenic animals [54,55]. Findings have established the fact that in 60% of the leukemic cases, Tax expression is undetectable [56]. This could be due to the fact the high levels of Tax protein in an infected cell make them susceptible to be attacked by Cytotoxic T Cells (CTL) which could be a reason for depleted levels of Tax expression [57].

For more than two and a half decades, multiple studies have been conducted on understanding the pathogenesis of HTLV-1 in order to develop therapeutics and vaccines against the virus. A favourable target for previous approaches to developing HTLV-1 vaccines and therapeutics has been the envelope glycoprotein (Env) since it is important for creating the baseline for viral gene transmission to a non-infected cell through cell mediated fusion [58]. As for a non-structural protein, Tax can also be a promising target for its importance as a determining factor of viral persistence and pathogenesis [1,39,40,55] for playing role as a key transcriptional activator [39,40]. Asymptomatic carriers, if not at an elevated level, have Tax protein expressions in infected cells (Median probability Tax protein expression of an infected cell in an Asymptomatic Carrier is 28% [1]). The goal of the current study is to map small continuous antigenic amino acid sequences in the Tax protein (epitopes) and selection of the most tangible Tax protein epitope candidate for *in silico* vaccine development against HTLV-1 following the principles of comparative modeling, stereochemical and epitope conservation analysis. *In vivo*, epitopes are presented to T cells through Human Leukocyte Antigen (HLA) molecules and are also detected by B cell receptors. The identification of epitopes by both T cells and B cells is a forerunner for the development of adaptive immunity against a specific antigen which the current study aims to develop against HTLV-1.

It is natural to consider the envelope glycoprotein of HTLV to be a more favourable candidate for vaccines; firstly because the envelope glycoprotein plays an important role in initiating viral pathogenesis by acting as the anchoring protein between the virus and the receptor and secondly, because previous studies have found highly immunodominant epitopes in the glycoprotein that are likely to evoke the immune cells [59–61]. However, factors like conservancy of the glycoprotein as well as nature of the epitopes should also be taken into account. In a previous study conducted on the molecular characterization of HTLV-1 gp46 glycoprotein, minor sequence variations were observed across different geographical regions as well as clinical characterization of the samples [62]. Additionally, the envelope glycoprotein is sensitive to mutations and will render non-expressive and dysfunctional if it undergoes any kind of mutation [63]. In case of epitope mapping, often in protein sequences immunodominant epitopes coincide with neutralizing epitopes, which can lead to faulty and diminished levels of epitope identification by immune cells [64]. A similar scenario was observed for envelope glycoprotein

gp46, when Palker et al.'s study found that the central neutralizing region of gp46 (190–209) and the C-terminal neutralizing epitope (296–312) as Desgranges et al. defined had immunodominant properties [65]. Moreover, a study conducted by Horal et al., highlighted overlapping immunodominant and neutralizing epitopes in the regions including 176–199, 190–212, 224–244, 240–262, and 292–314 [66]. These facts imply the chances of improper identification of epitopes by immune cells and subsequent failure of the vaccine itself to provoke proper immune response. The Tax protein on the other hand is predominantly expressed in infected cells as transcription activators as well as exosomes, even in the early stages, therefore identification of the epitopes by immune cells and building the adaptive immune system against HTLV will be a more agile process [67–69].

The approach of this study to designing a Tax protein epitope based vaccine against HTLV-1 *in silico* might prove to be effective in activating the acquired immune system and stimulating specific humoral and cell mediated immune response against HTLV-1 in healthy individuals; treating both symptomatic and asymptomatic individuals with HTLV-1 infection and in preventing the development of further neurological, pulmonary, ophthalmological, rheumatological and urological co-infections in infected individuals.

Materials and methods

A flow chart describing the overall procedures of construction of a multi-epitope based peptide vaccine for HTLV-1 Tax protein has been illustrated in Fig 1.

Protein sequence retrieval

The amino acid sequence of Trans-activating transcriptional regulatory protein of HTLV-1 (Tax protein) was retrieved from the NCBI Protein database in FASTA format. The NCBI Protein database consists of a collection of sequences retrieved from SwissProt, PIR, PRF, PDB. It also includes translations from elucidated coding regions in GenBank, RefSeq and TPA. Numerous bioinformatics and immunoinformatics tools and databases were used to identify and analyse physicochemical, structural and functional properties of the selected Tax protein sequence and prediction of B-cell and T-cell epitopes.

Primary and secondary structure analysis

The primary structure of the protein was evaluated using ProtParam tool of the Expasy server [70]. ProtParam computes various physical and chemical characteristics of given protein sequence, such as number of amino acid, isoelectric point (pI), molecular weight, instability index, aliphatic index, number of total atoms, grand average of hydropathicity (GRAVY). The secondary structure was analyzed using self-optimized prediction method with alignment (SOPMA) [71] and PSIPRED [72] to access properties like transmembrane helices, globular regions, bend region, random coil and coiled-coil region and to obtain a graphical presentation of the protein sequence of interest (HTLV-1 Tax protein).

Tertiary structure analysis

The three dimensional structure from the amino acid sequence of the selected HTLV-1 Tax protein was predicted and analyzed using Phyre2 and $(PS)^2$ Servers [73–75]. Phyre2 server allows 3D modelling of a provided amino acid sequence using the alignment of hidden Markov models by the help of HHsearch [76], improving accuracy of alignment and detection rate in a pronounced way. Phyre2 also models 3D structures for regions in a query sequence that

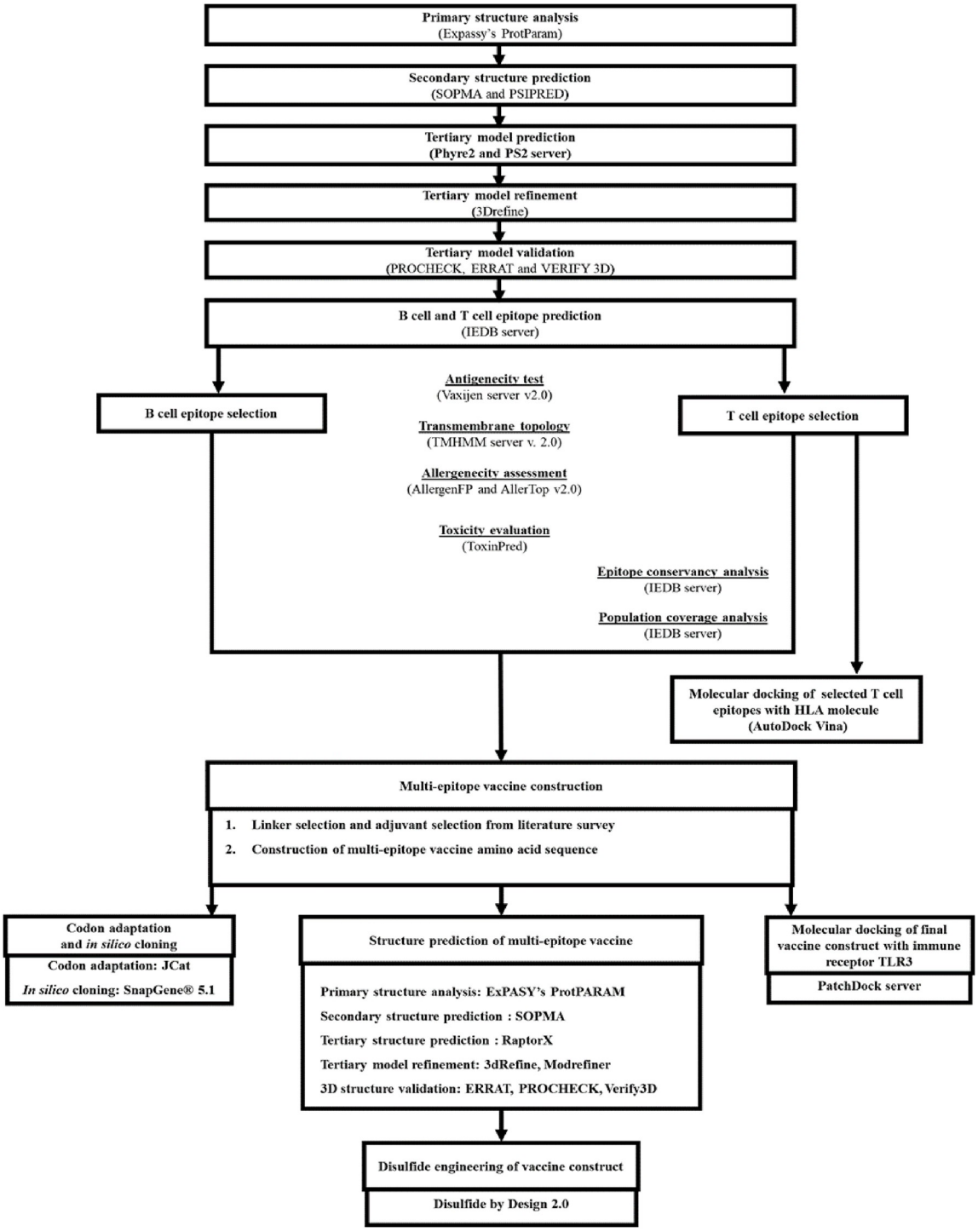

Fig 1. Working-flowchart of the multi-epitope vaccine construct design against HTLV-1.

https://doi.org/10.1371/journal.pone.0248001.g001

does not have any detectable homology with recognized structures by incorporating an ab initio folding simulation called Poing [77]. The $(PS)^2$ homology modelling server operates through a substitution matrix for detecting homologous proteins by combining both sequence and secondary structure information [74,75].

Model refinement

The generated 3D model of HTLV-1 Tax protein was refined at an atomic level based on a reference 3D protein model using high resolution protein structure refinement tool ModRefiner [78] and 3DRefine [79].

Validation of protein structure

Additional evaluation of the 3D structure was done by the help of multiple web tools. Ramachandran plot was generated using PROCHECK [80] to further assess the Phi/Psi angles for better understanding of the protein backbone confirmation. PROCHECK [80] was also used for stereochemical analysis of the predicted protein using a residue-by-residue approach. An additional evaluation of the generated 3D structure was carried out using ERRAT [81] and Verify 3D [82].

B-cell epitope prediction

Linear B cell epitopes were predicted using sequential B-cell epitope predictor BepiPred-2.0 under IEDB (Immune Epitope Database) server [83]. Selection of the predicted B cell epitopes was done based on transmembrane topology (to discriminate between soluble and membrane proteins) and antigenicity for which TMHMM server 2.0 [84] and Vaxijen 2.0 web server [85] were used respectively. Allergenicity and toxigenicity of the selected epitopes were checked using AllerTOP v. 2.0 [86] and ToxinPred [87] respectively.

T-cell epitope prediction

NetCTL 1.2 server was used to predict CTL epitopes for the selected HTLV-1 Tax protein sequence [88]. Binding affinity towards multiple HLA class molecules was assessed using MHC I binding prediction tools of IEDB server [89]. In case of predicting MHC class I binding epitopes, the Stabilized Matrix Base Method (SMM) was used in order to calculate the IC50 values [90]. Epitopes having an IC50 value less than 200nm were selected for the steps following.

Population coverage calculation

Population coverage for the selected epitopes was observed using the IEDB resource for population coverage [91] considering denominated MHC restriction of T cell responses and polymorphic HLA combinations for different regions of the world. The selected epitopes were then scrutinized for their antigenic and immunogenic potentials using Vaxijen 2.0 [83] and immunogenicity analysis resources of the IEDB server [92]. The IEDB analysis resource was also used for assessing epitope conservancy. AllerTOP v. 2.0 [86] and ToxinPred [87] were used for assessing allergenic and toxigenic aptitude of the selected epitopes.

3D epitope structure prediction and molecular docking analysis

To substantiate the interaction between the selected T cell epitopes and HLA molecules, *in silico* docking was carried out between the molecules using AutoDock Vina [93]. For the process, HLA-A*0201 was selected due to having a high genetic frequency. In order to stimulate the docking, a crystal structure of HLA-A*0201 was retrieved from RCSB protein database (Research Collaboratory for Structural Bioinformatics) [94] having the PDB ID 1B0R. PEPFOLD [95] server was used to predict peptide structures of the selected T cell epitopes based on antigenicity. To carry out docking simulation first of all the peptide ligand (Influenza Matrix Peptide) bound to HLA-A*0201 was removed by PyMol software package. This

removed peptide ligand was then docked with the binding groove of HLA-A*0201. The binding groove had center box coordinates X: 29.900, Y: 0.614, Z: 49.512 and the dimensions of the grid box were X: 40, Y: 40 and Z: 40 (unit of the dimensions, Å) targeting the active site of the protein.

Construction of multi epitope vaccine, modeling, and validation

The selected B cell and T cell epitopes based on antigenicity and allergenicity were linked in order to create a fusion peptide using GPGPG and AAY peptide linker [96] in multiple combinations. VaxinPad [97] was used to predict a peptide based adjuvant for the amino acid sequences generated using the selected epitopes. Selection of the adjuvant was based on immunomodulatory potential accompanied with properties like hydrophobicity, hydrophilicity, steric hindrance, solvation, hydropathicity and isoelectric point (pI). EAAK rigid peptide linker was used to link the selected adjuvant in the N-terminal end of the multi epitope combinations. The multi epitopes were then assessed for their antigenic and allergenic potencies using Vaxijen 2.0 web server [85] and AllerTOP v. 2.0 [86]. Multi epitope sequences exhibiting the maximum antigenic potential and non allergenicity were selected for tertiary structure evaluation. I-TASSER (Iterative Threading ASSEmbly Refinement) homology modelling tool was used to predict 3D structures [98–100] for the selected multi epitope sequences. The final 3D structures were further validated using PROCHECK [80], ERRAT [81] and Verify3D [82]. Upon completion of evaluation, the 3D structures were visualized using Pymol [101].

Disulfide engineering of the 3D multi epitope vaccine constructs

To check if the structurally validated 3D vaccine constructs were accessible to the addition of novel disulfide bonds to provide the protein with increased stability and decreased conformational entropy [102], Disulfide by Design 2 (DbD2) was used for disulfide engineering of the 3D multi epitope vaccine constructs [102].

Codon adaptation and *in silico* cloning

In order to express multi epitope vaccine construct selected based on antigenicity, peptide validation parameters and binding affinity towards HLA molecules in an *Escherichia coli* K12 strain codon optimization was done using Java Codon Adaptation Tool (JCat) [103]. The optimized codon sequence was further screened for expression parameters, codon adaptation index (CAI) and percentage of GC-content. For *in silico* cloning simulation, a bacterial, kanamycin resistant expression vector pETSUMO which consists of a 6x polyhistidine tag was selected [104]. The *in silico* restriction cloning simulation between the adapted codon sequence and pETSUMO expression vector was carried out using Snapgene software. The use of a pETSUMO expression vector will facilitate TA cloning as well as provide an easier screening modus operandi for transformed cells. The hexa histidine tag present in pETSUMO expression vector will facilitate swift detection of the recombinant vaccine construct in immunochromatographic assays [105].

Molecular docking of vaccine constructs with TLR4

In order to predict the binding affinity between the Multi Epitope Vaccine Construct and Toll like receptor 4 [TLR-4 (PDB: 4G8A)], a member of the toll like receptor family of protein involved in triggering the innate immunity system [106], molecular docking approaches were implemented computationally. PatchDock (Beta 1.3 Version) docking server [107,108] was adopted to receptor-ligand docking and the generated Protein Data Bank (PDB) file of the

protein-peptide docking complex was visualized in BIOVIA Discovery Studio Visualizer v12.1.0.15350.

Results

Sequence retrieval, transmembrane topology and antigenicity features analysis of the HTLV-1 Tax protein sequences

The complete amino acid sequences of HTLV-1 Tax protein were retrieved from NCBI database. Seven amino acid sequences of the protein were retrieved in FASTA format with accession number AYN25353.1, BBA30574.1, BBD74587.1, AYN25375.1, AYN25364.1, AYN25342.1, AYN25331.1 and BAX77785.1. Antigenicity of the protein was confirmed through assessment in VaxiJen v2.0 server and the Tax protein amino acid sequence with accession number BAX77785.1 was found to be most antigenic with a score of 0.4610 at threshold 0.4. Antigenicity is a prerequisite for a protein or amino acid sequence to be a vaccine candidate. The transmembrane topology of the protein was determined by TMHMM Server v. 2.0 and it was found that the protein fulfills the criteria of exo-membrane protein.

Primary and secondary structure analysis

A physico-chemical analysis of the protein primary structure was performed by the Expasy server's ProtParam tool. From the result generated by ProtParam it was found that the 353 amino acids long HTLV-1 tax protein has a molecular weight of 39470.53Da with a high aliphatic index of 87.56%. Having pI of 6.45, less than 7, the protein belongs to negatively charged proteins. Instability index of 48.9 indicates that the protein is unstable *in vitro*. Interestingly it has a negative grand average hydropathy score accompanied with a high extinction coefficient as summarized in Table 1. Secondary structural features analysed by SOPMA and PSIPRED reveals the abundance of random coil (54.96%) followed by extended strand (21.53%), alpha helix (17.85%) and 5.67% of beta turn. in HTLV-1 Tax protein. The secondary structure plot is presented in Fig 2a while Fig 2b illustrates the distribution of different forms of secondary structure in HTLV-1 Tax protein.

Tertiary structure prediction, refinement and validation

The tertiary structure of a protein is critical to its function and stability. 3D structure of the protein was predicted using the Phyre2 server and PS^2 server. PDB ID 2I46 was selected as a template. The structure was predicted with 75% reliability. A pymol generated 3D structure of

Table 1. Different physio-chemical properties of HTLV-1 Tax protein.

Parameter	Value
Amino acids	353
Molecular weight	39470.53 Da
Theoretical isoelectric point (pI)	6.45
The instability index (II)	48.96
Grand average of hydropathicity (GRAVY)	-0.054
Total number of atoms	5551
Aliphatic index	87.56
No. of negatively charged residues (asp+glu)	25
No. of positively charged residues (arg+lys)	21

https://doi.org/10.1371/journal.pone.0248001.t001

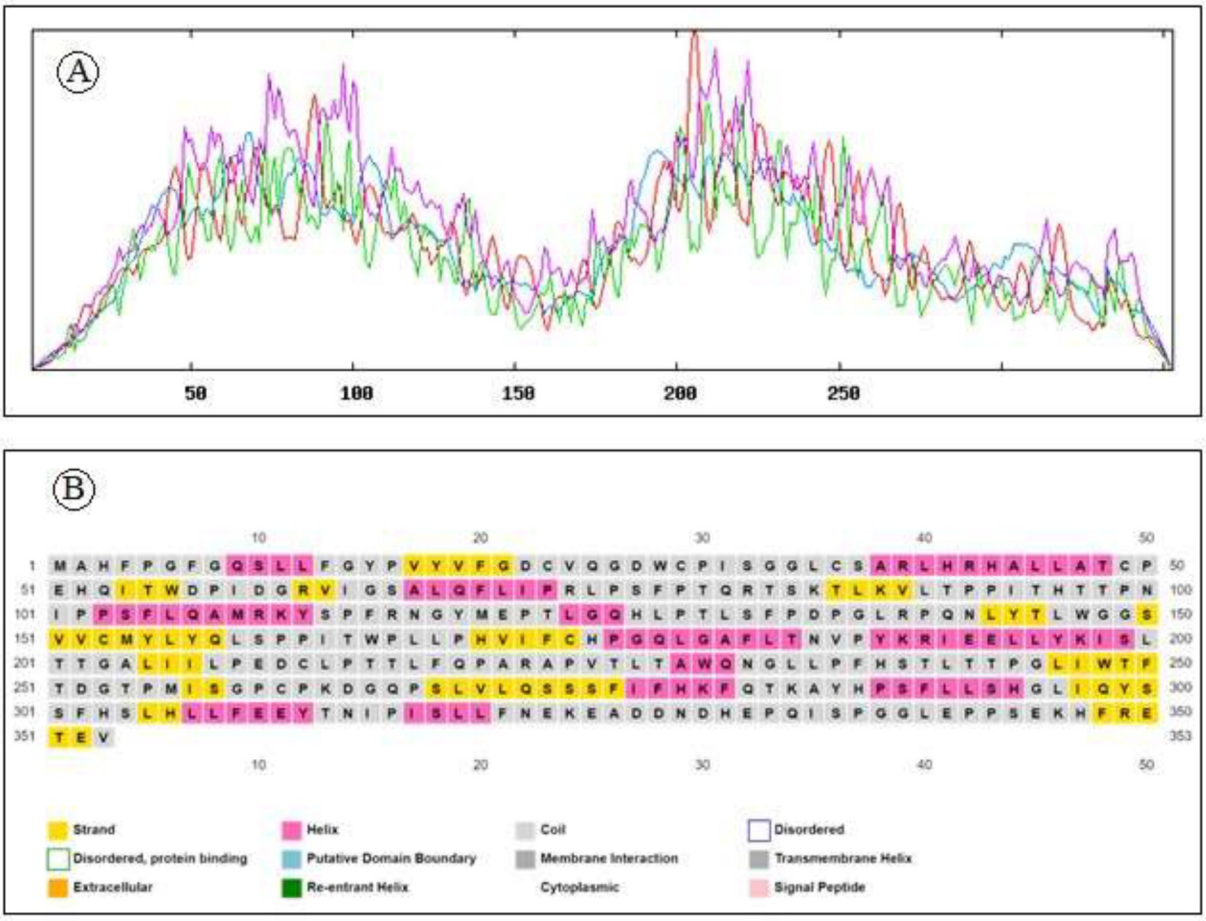

Fig 2. (a) Secondary structure plot of HTLV-1 Tax protein. Here, helix is indicated by blue, while extended strands and beta turns are indicated by red and green, respectively. (b) Distribution of different forms of secondary structure in HTLV-1 Tax protein.

https://doi.org/10.1371/journal.pone.0248001.g002

HTLV-1 Tax protein is displayed in Fig 3a. Major local distortions contained by homology based modeling include irregular H-hydrogen bonding networks, steric clashes and unphysical phi/psi angles which reduce the structure models usefulness for high-resolution functional analysis. Protein model refinement is required to overcome these distortions.

Refinement of the predicted 3D structure is required for bringing it closer to its native structure. 3Drefine was used for the structure refinement. The 3Drefine tool utilizes repetitive optimization of hydrogen bonding networks along with energy minimization at atomic-level on the optimized model by utilizing knowledge-based force fields and composite physics for efficient protein structure refinement. It generated 5 refined models with RMSD ranging from 0186 to 0.357. The best refined structure had the highest 3D score of 34961.1 accompanied by least RMSD 0.186.

After refinement the refined model was validated using PROCHECK, ERRAT and VERIFY 3D. Ramachandran plot generated by PROCHECK showed that 77.7 7% of the residues fall within the most favoured region (Fig 3b). Results generated from ERRAT showed an overall quality factor of 39.86 for the best refined structure (Fig 3c) and Verify3D depicts that only 27.20% of the residues in the protein had 3D-1D score equal or above 0.2.

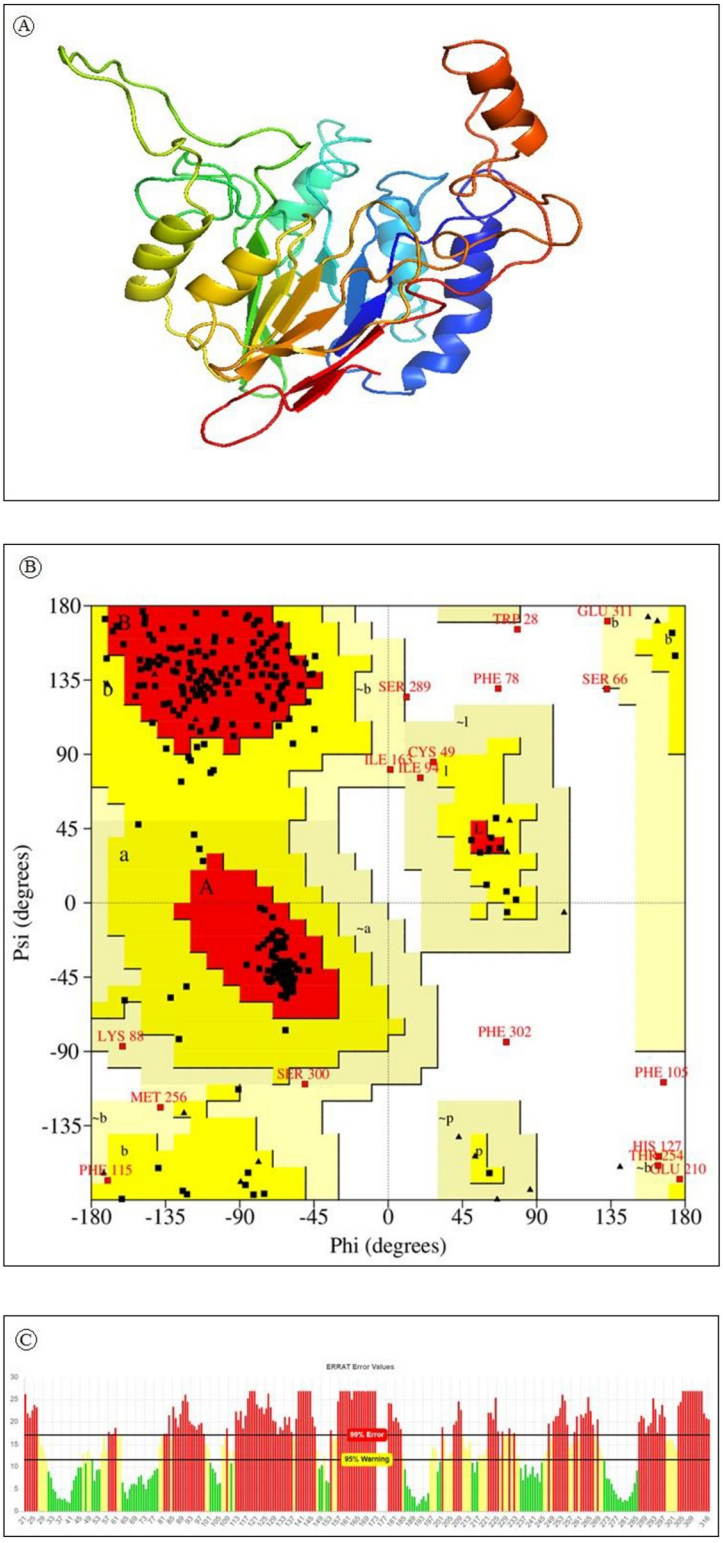

Fig 3. **a)** Predicted 3-dimensional structure of HTLV-1 Tax protein using comparative modeling. **b)** Analysis of Ramachandran plot of HLTV-1 Tax protein. Here, yellow indicates allowed region, red for favored region, light yellow shows generously allowed region while white for disallowed region. Torsion angels are determined by phi and psi angels. **c)** ERRAT generated result of Tax protein where 95% indicates rejection limit.

https://doi.org/10.1371/journal.pone.0248001.g003

B cell epitope prediction

Overall antigenicity assessment of HTLV-1 tax protein by VaxiJen server regarded protein as a promising antigen at threshold 0.4. Transmembrane topology analysed by TMHMM predicted that all the residues of the protein are exposed outside the cell membrane. Epitope was predicted by the IEDB server using Kolaskar and Tongaonkar antigenicity and Bepipred linear epitope prediction method. Based on antigenicity test performed by VaxiJen server 7 epitopes (^{324}KEADDNDHEPQISPGGLEPPSEKHFR^{349}, ^{252}DGTPMISGPCPKDGQPS^{268}, ^{114}PFRNGYMEPTLGQ^{126}, ^{51}EHQITWDPIDGR^{62}, ^{93}PITHTTPNIPPS^{104}, ^{131}LSFPDPGLRPQ^{141}, ^{77}SFPTQRTS^{84}) were found to be antigenic (Table 2). TMHMM is used to evaluate the transmembrane topology of the antigenic B cell epitopes. EHQITWDPIDGR and SFPTQRTS with antigenicity score 1.3783 and 0.7813 respectively were exposed inside while epitope LSFPDPGLRPQ, PITHTTPNIPPS, PFRNGYMEPTLGQ, DGTPMISGPCPKDGQPS and KEADDNDHEPQISPGGLEPPSEKHFR having antigenicity score of 1.3581, 0.9582, 0.9083, 0.7207 and 0.7170 were exposed outside.

Toxicity of the antigenic epitopes was accomplished by the ToxinPred tool. All the examined epitopes reported to be non-toxic. Hydrophilic epitopes were subjected to assessment for electricity using AllerTopv2.0. KEADDNDHEPQISPGGLEPPSEKHFR and DGTPMISGPCPKDGQPS demonstrated no allergenicity while PFRNGYMEPTLGQ, PITHTTPNIPPS and LSFPDPGLRPQ were predicted as potential allergens as summarized in Table 4. Based on the antigenicity, surface accessibility and allergenicity 2 epitopes KEADDNDHEPQISPGGLEPPSEKHFR and DGTPMISGPCPKDGQPS spaning region 324–349 and 252–268 respectively were finally selected for vaccine model construction.

T cell epitope prediction

NetCTL server and IEDB T cell epitope prediction tools were used to predict T cell epitopes of HTLV-1Tax protein. Among the 94 MHC class I epitopes predicted based on combined score 64 epitopes wear selected with IC50 value less than 200nm. 15 epitopes among them we are found to interact with multiple which were selected for antigenicity assessment at threshold level of 0.4. Epitopes with position 11–19, 151–159, 163–171, 178–186, 233–241, 297–305 and 307–315 were found antigenic (Table 3). LLFEEYTNI, QLGAFLTNV, LLFGYPVYV, ITWPLLPHV, GLLPFHSTL peptide sequence were found immunogenic when analysed with IEDB class I immunogenicity prediction tool. Immunogenic epitopes are listed in Table 4. From the transmembrane topology determination of these immunogenic epitopes by TMHMM server v. 2.0 it was found that all of these epitopes fulfills criteria of exomembrane protein.

Table 2. Assessment of antigenicity score, transmembrane topology and toxicity of the antigenic B cell epitopes.

Start	End	Peptide	Length	Antigenecity	TMHMM	Toxicity
324	349	KEADDNDHEPQISPGGLEPPSEKHFR	26	0.7170	Outside	No
252	268	DGTPMISGPCPKDGQPS	17	0.7207	Outside	No
114	126	PFRNGYMEPTLGQ	13	0.9083	Outside	No
51	62	EHQITWDPIDGR	12	1.3783	Inside	No
93	104	PITHTTPNIPPS	12	0.9582	Outside	No
131	141	LSFPDPGLRPQ	11	1.3581	Outside	No
77	84	SFPTQRTS	8	0.7813	Inside	No

https://doi.org/10.1371/journal.pone.0248001.t002

Table 3. MHC class I epitopes based on their antigenicity, immunogenicity and epitope conservancy.

Epitopes	Antigenicity score	Immunogenicity score	Epitope conservancy (%)
LLFEEYTNI	0.4534	0.26	100
QLGAFLTNV	0.4664	0.18	100
LLFGYPVYV	0.4126	0.09	100
ITWPLLPHV	0.6704	0.05	100
GLLPFHSTL	0.9387	0.01	100
VVCMYLYQL	0.6350	-0.28	100
IQYSSFHSL	0.8634	-0.29	100

https://doi.org/10.1371/journal.pone.0248001.t003

Epitope conservancy, toxicity and allergenicity assessment

Epitope conservancy of the expected epitopes were tested by examining and matching all the epitopes obtained from HTLV-1 Tax protein. Epitope sequence LLFEEYTNI, QLGAFLTNV, LLFGYPVYV, ITWPLLPHV, GLLPFHSTL spaning 307–315,178–186, 11–19, 163–171, 233–241 position of HLTV-1 Tax were found to show 100% conservancy as predicted by the IEDB epitope conservancy prediction tool.

Toxicity was analysed by ToxinPred which indicated all 5 immunogenic epitopes as non toxic. The Allergenicity test carried out by AllerTop v2.0 depicted LLFGYPVYV, ITWPLLPHV and GLLPFHSTL as non allergen. These epitopes were also reported non toxic by ToxinPred (Table 4).

Population coverage

As the different HLA alleles have different rates of occurrence in different ethnicities in the world, the analysis of individuals coverage by the respective HLA alleles of the predicted T-cell epitopes is an important part of an effective vaccine design. The population coverage analysis showed maximum coverage in Mexico (90.21%) followed by England (89.88%), South Africa (81.56%), North America (76.82%), South America (74.95%) and Japan (70.92%). Very low population coverage of 1.20% was observed in the United Arab of Emirates while it was more than 70% in three major parts of Asia, 70.48%, 71.78% and 74.53% respectively in Southeast Asia, Northeast Asia and East Asia. Cumulative population distribution of HLA alleles for the selected T cell epitopes is pictured by Tableau Public online public software in Fig 4.

Table 4. Allergenicity, epitope conservancy and interacting MHC class I alleles for the immunogenic MHC class I epitopes.

Epitopes	Interacting Class I alleles (IC50<200)	Allergenicity	Epitope conservancy (%)
LLFEEYTNI	HLA-A*02:02, HLA-A*02:01, HLA-A*02:06, HLA-A*02:11, HLA-A*02:16, HLA-A*02:12, HLA-A*02:03, HLA-A*02:50, HLA-C*12:03, HLA-A*02:19, HLA-C*14:02, HLA-A*02:17.	Allergen	100
QLGAFLTNV	HLA-A*02:02, HLA-A*02:01, HLA-A*02:06, HLA-A*02:03, HLA-A*02:11, HLA-C*12:03, HLA-A*02:19, HLA-A*02:12, HLA-A*02:16.	Allergen	100
LLFGYPVYV	HLA-A*02:02, HLA-A*02:01, HLA-A*02:06, HLA-A*02:11, HLA-A*02:16, HLA-A*02:12, HLA-A*02:19, HLA-A*02:03, HLA-A*02:50, HLA-C*12:03, HLA-C*14:02, HLA-A*02:17, HLA-C*06:02, HLA-C*07:01.	Non-allergen	100
ITWPLLPHV	HLA-A*02:06, HLA-A*02:01, HLA-A*02:16, HLA-A*02:11, HLA-A*02:50, HLA-C*12:03, HLA-C*14:02.	Non-allergen	100
GLLPFHSTL	HLA-A*02:02, HLA-A*02:01, HLA-A*02:16, HLA-A*02:11, HLA-A*02:50, HLA-C*03:03, HLA-A*02:12, HLA-C*07:02, HLA-B*15:02, HLA-A*02:19, HLA-A*02:17, HLA-C*12:03, HLA-A*02:03.	Non-allergen	100

https://doi.org/10.1371/journal.pone.0248001.t004

Fig 4. Worldwide population coverage of the selected for the multi-epitope vaccine. Global distributions of the selected epitopes are indicated by circles. Circle of different colors represents different ranges of population coverage.

https://doi.org/10.1371/journal.pone.0248001.g004

Molecular docking simulation of the HLA allele-peptide interaction

Binding energy achieved for the peptide ligand (Influenza Matrix Peptide) was -5.7 Kcal/mol. A higher binding energy was acquired for our predicted epitopes ITWPLLPHV, LLFGYPVYV and GLLPFHSTL as follows -8.8, -8.6 and -8.5 Kcal/mol which suggest profound interaction between the predicted epitopes and HLA molecules. Analysis of the nonbonding interactions of the Influenza Matrix Peptide interacted with the active site A:TYR27, A:GLN32, A:PRO235, B:TYR426, B:SER452, B:ASP453, B:TYR463 and B:LEU465. T cell epitopes with the HLA-A*0201 reveals that the selected compounds interacted mostly with A:Try27, A:Pro235 and B:Tyr463 catalytic residues detected by Autodock Vina, as shown in Table 5 and Fig 5.

Multi-epitope vaccine construction

Three CTL epitopes (LLFGYPVYV, ITWPLLPHV and GLLPFHSTL) and two B cell epitopes (KEADDNDHEPQISPGGLEPPSEKHFR, DGTPMISGPCPKDGQPS). B cell epitopes were linked using GPGPG linkers while CTL epitopes by AAY linkers. A peptide adjuvant PMISWPCPKD was selected based on immunogenicity and high molecular weight using VaxinPred server. The Adjuvant was added to N termini of the vaccine construct using EAAAK linker. EAAAK linkers were used on both the terminals of the vaccine construct. Antigenicity score of the Multi-epitope vaccine construct was 0.57 as predicted by VaxiJen v2.0 server. A schematic diagram of 109 aa long multi-epitope vaccine is shown in Fig 6.

Structure prediction of multi-epitope vaccine

Primary and secondary structure analysis provides insight about stability of a peptide. Various physicochemical parameters were determined by ExPASY's ProtParam server. A high molecular weight of 11.523kDa is an indication of the antigenic nature of the vaccine construct. Theoretical pH is valuable in determining the buffer system for the vaccine purification process

Table 5. Non-bonding interactions of three T-cell epitope ligands and an Influenza Matrix Peptide with HLA-A*0201.

Legends	Bonds [Donor. (Distance, Å). Acceptor] (Bond type)		
	Hydrogen Bond	**Electrostatic Bond**	**Hydrophobic Bond**
LLFGYPVYV	A:ARG6:HH11 (2.300):LEU2:O (HB) A:ARG6:HH12 (2.378):TYR5:OH (HB) A:TYR27:HH (2.187):VAL7:O (HB) A:GLN32:HN (2.147):VAL9:O1 (HB) A:GLY237:HN (2.983):TYR8:O (HB) A:GLY239:HN (2.319):TYR8:O (HB) B:LYS458:HN (2.966):LEU1:O (HB) B:TYR467:HH (2.328):TYR8:OH (HB) :VAL9:O1 (3.381) A:ASP30:O (HB) :TYR5:OH (3.319) A:ASP29:O (HB) A:ARG6:CD (3.297):TYR5:OH (CHB) A:THR31:CA (3.301):VAL9:O1 (CHB)	A:ASP30:OD2 (3.789):PHE3 (Pi-Anion)	:LEU1:CD2 (3.812) A:TYR113 (Pi-Sigma) B:TYR467 (5.311):TYR8 (Pi-Pi Stacked) A:ALA211 (4.904):PRO6 (A) A:PRO235 (4.589):PRO6 (A) A:PRO235 (4.365):VAL7 (A) B:LYS458 (4.867):LEU1 (A) B:LYS458 (4.180):LEU2 (A) :VAL7 (4.979) B:LEU465 (A) A:PHE241 (5.039):PRO6 (Pi-A) B:TYR426 (5.412):PRO6 (Pi-A) B:TYR463 (5.200):VAL7 (Pi-A)
ITWPLLPHV	A:ARG6:HH11 (2.792):HIS8:O (HB) A:TYR27:HH (2.216):TRP3:O (HB) A:GLN32:HN (2.281):THR2:OG1 (HB) A:GLU212:HN (2.545):VAL9:O2 (HB) B:TYR463:HH (2.132):LEU6:O (HB) :THR2:OG1 (3.371) A:ASP30:O (HB) :VAL9:N (3.296) A:ASP30:OD1 (HB) :VAL9:O2 (3.342) A:ASP30:OD1 (HB) A:ASP30:CA (3.338):LEU5:O (CHB) A:THR31:CA (3.542):THR2:OG1 (CHB)		B:TYR467 (5.495):TRP3 (Pi-Pi Stacked) :TRP3 (4.502) B:TYR467 (Pi-Pi Stacked) A:ALA49 (5.108):ILE1 (A) A:PRO50 (5.198):ILE1 (A) A:ALA211 (5.155):LEU5 (A) A:PRO235 (4.422):PRO4 (A) A:PRO235 (5.145):LEU5 (A) B:LYS458 (5.007):PRO7 (A) :PRO4 (5.006) B:LEU465 (A) A:TYR27 (5.433):LEU6 (Pi-A) A:PHE241 (5.371):LEU5 (Pi-A) B:TYR426 (5.493):PRO4 (Pi-A) B:TYR463 (4.805):PRO4 (Pi-A) B:TYR463 (5.333):LEU5 (Pi-A) :HIS8 (4.403) A:ALA211 (Pi-A)
GLLPFHSTL	A:GLY1:HT3 (2.059):LEU9:O2 (HB) A:SER2:HN (1.961):LEU9:O2 (HB) A:ARG6:HH12 (2.532):SER7:OG (HB) A:THR233:HN (2.685):LEU2:O (HB) A:THR233:HG1 (2.336):GLY1:O (HB) A:LYS243:HZ3 (2.532):GLY1:O (HB) B:SER457:HG (2.571):LEU3:O (HB) B:LYS458:HN (2.406):HIS6:NE2 (HB) :LEU3:N (3.215) A:GLU232:OE1 (HB) :LEU9:O1 (3.365) A:ASP29:OD1 (HB) :THR8:OG1 (3.190) A:ASP30:OD1 (HB) A:TYR27:HH (3.000):PHE5 (Pi-Donor HB) :SER7:OG (3.513):HIS6 (Pi-Donor HB)	:GLY1:N (3.609) A:GLU212:OE2 (Salt Bridge)	B:TYR463 (5.312):PHE5 (Pi-Pi Stacked) A:PRO210 (4.398):LEU9 (A) A:ALA211 (4.254):PRO4 (A) :PHE5 (4.024) A:PRO235 (Pi-A)
Influenza Matrix Peptide	A:TYR27:HH (1.971):ILE702:O (HB) A:GLN32:HN (2.086):GLY701:O (HB) B:SER452:HG (2.305):LEU703:O (HB)	:GLY701:N (5.208) B:ASP453:OD1 (Attractive Charge)	A:PRO235 (4.034):LEU703 (A) B:LEU465 (4.753):ILE702 (A) B:TYR426 (5.443):LEU703 (Pi-A) B:TYR463 (4.861):LEU703 (Pi-A)

(Pose predicted by AutoDock Vina where, HB = Conventional Hydrogen Bond, CHB = Carbon hydrogen bond, A = alkyl, Pi-A = Pi-Alkyl).

https://doi.org/10.1371/journal.pone.0248001.t005

which was computed 5.18 representing the slightly acidic nature of the vaccine. Number of positively and negatively charged residues were found to be 8 and 13 respectively.

Assuming all cysteine residues are reduced, the extinction coefficient was 18450M-1cm−1 at 280 nm measured in water. The calculated half life was found 1 h in mammalian reticulocytes (*in vitro*), while >10 h in *E. coli* (*in vivo*) and 30 min in yeast (*in vivo*). The instability index was 38.16. Having an instability index below 40 classified the vaccine as stable. The aliphatic index was high and assessed to be 65.60 whereas the grand average of hydropathicity (GRAVY) was -0.367. A high aliphatic index supports the thermostability of the designed

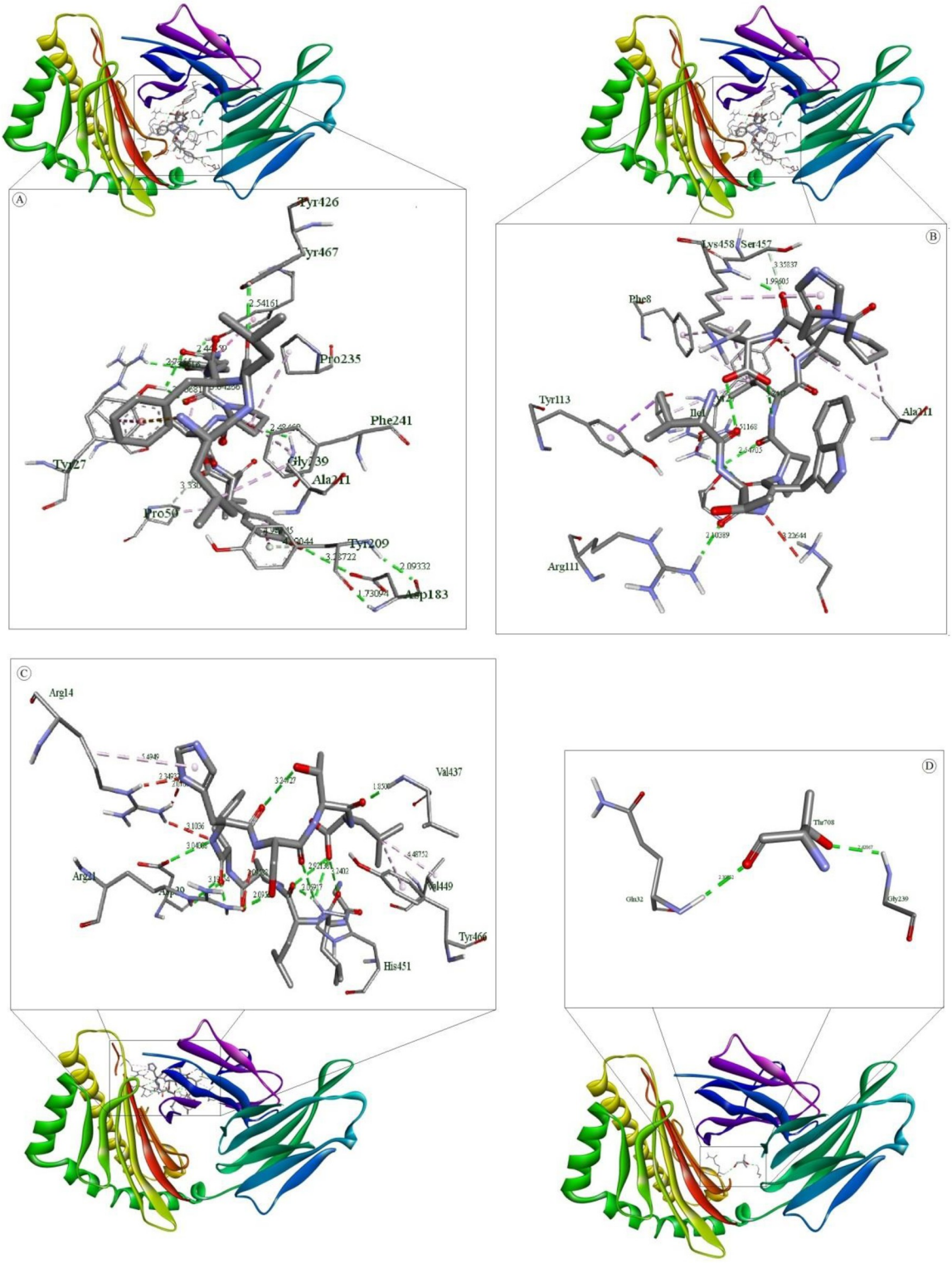

Fig 5. Molecular docking of selected T cell epitopes with HLA-A*0201. (A) shows interaction of LLFGYPVYV and HLA-A*0201 (B) shows interaction of ITWPLLPHV and HLA-A*0201 (C) shows interaction of GLLPFHSTL and HLA-A*0201 and (D) shows interaction of Influenza Matrix Peptide and HLA-A*0201.

https://doi.org/10.1371/journal.pone.0248001.g005

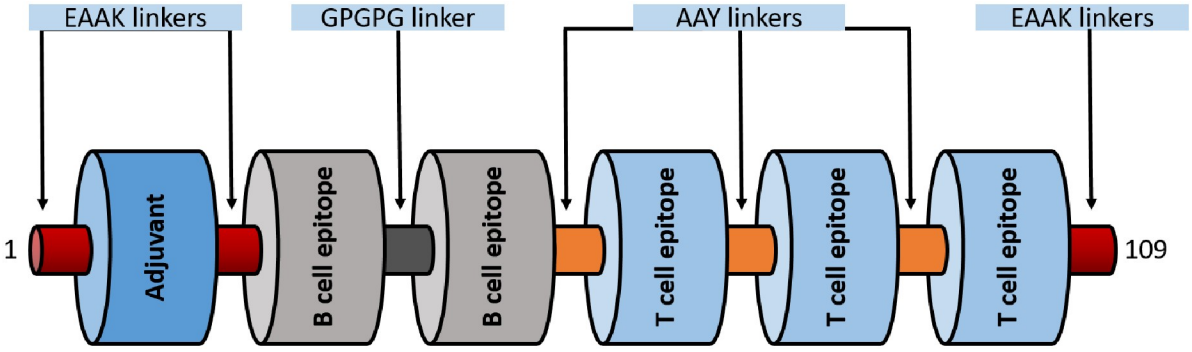

Fig 6. A diagrammatic representation of the final multi-epitope vaccine peptide. The 109-amino acid long peptide sequence containing an adjuvant (blue) at the N-terminal end linked with the multi-epitope sequence through an EAAAK linker (garnet). T cell epitopes are linked with the help of AAY linkers (orange) while the B cell epitopes are linked using GPGPG linkers (dark grey).

https://doi.org/10.1371/journal.pone.0248001.g006

vaccine while a negative GRAVY score represents its hydrophilic nature. Secondary structure of the vaccine was assessed by SOPMA server using 109 aa long sequence. It predicted that 25.69%, 4.59%, 4.59% and 65.14% amino acids are involved in α-helix, extended strand, β-turn, and random coil, respectively. A probability score graph of frequency of helix, strand, turn, and coil at each amino acid position in the secondary structure of the final vaccine construct. Tertiary structure of the multi-epitope vaccine prediction by I-tasser which is an ordered approach for the prediction of protein structure and function. It identifies structural templates from the PDB by multiple threading approach LOMETS. Structure generated by the I-tasser is shown in Fig 7.

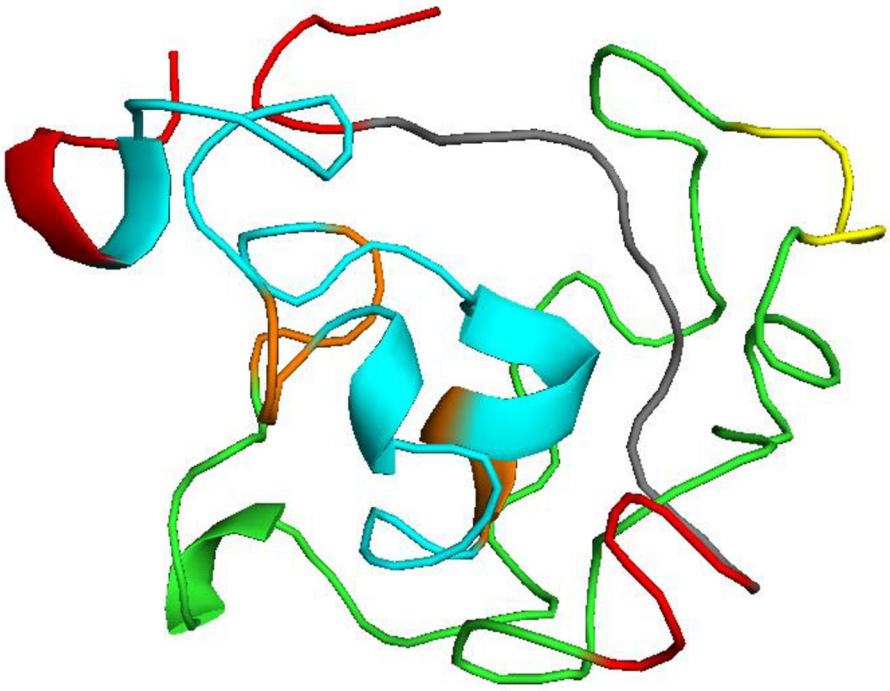

Fig 7. Tertiary structure of the multi-epitope vaccine. Red, orange and yellow color indicate EAAAK, AAY and GPGPG linkers respectively. B cell epitopes are shown in green while T epitopes in blue. A 10 mer adjuvant is visualized by gray color.

https://doi.org/10.1371/journal.pone.0248001.g007

Disulfide engineering, codon adaptation, and *in silico* cloning of the vaccine construct

Disulfide engineering was done through Disulfide by Design 2 (DbD2) which predicted a total of 20 pairs of residues for the probable disulfide bonds formation (Table 6) But, after considering the Chi3 value between −87 and +97 and the energy score less than 4 kcal/mol only 4 pairs of residue pairs were selected for making disulfide bonds.

JCat server was used for the adaptation of codon usage of the designed vaccine constructs for *E. coli* K12. An optimized codon sequence of 329 nucleotides was provided by JCat. The percentage of the GC-content of 50.78% and CAI of 1.0 of the optimized codon sequence ensured that the vaccine construct was highly expressed in *E. coli* K12. The improved DNA sequence was translated into our vaccine protein appropriately and GC-content increased to 58.10% for *E. coli* K12.

Later on, the adapted codon sequence was cloned into the *E. coli* pET SUMO vector and *E. coli* strain DH5 alpha was selected as host. A 5970 bp recombinant vector was obtained after TA cloning of vaccine DNA sequence (Fig 8).

Molecular docking of the multi-epitope vaccine with immune receptor *TLR4*

The docking of vaccine-receptor was performed using Patchdock server for evaluating the complex formation of the our vaccine construct with an immune receptor such as TLR4 and their binding affinity. The PatchDock server provided 20 docking complexes. Among the complexes, we selected only the docking complex with the highest negative Atomic Contact Energy (ACE) value for analysis. The ACE value of the selected docking complex was -247.59 which

Table 6. Possible disulfide bond between the residues of the vaccine construct.

Residue 1			Residue 2			Bond		
Chain	Seq	AA	Chain	Seq	AA	X_3	kcal/mol	Σ B-factor
A	2	ALA	A	99	PRO	-67.47	5.50	5.17
A	2	ALA	A	101	LEU	+114.49	3.48	4.10
A	3	ALA	A	102	PRO	-73.27	3.51	4.13
A	4	ALA	A	99	PRO	-76.90	3.79	6.36
A	5	LYS	A	103	HIS	+81.03	3.37	6.01
A	13	PRO	A	16	GLU	+83.62	6.42	4.34
A	13	PRO	A	74	PHE	+122.24	1.95	4.61
A	24	PRO	A	26	ILE	+95.00	4.47	5.63
A	24	PRO	A	27	SER	+119.72	3.94	5.31
A	26	ILE	A	31	PRO	+118.95	4.74	4.57
A	27	SER	A	86	LEU	-61.55	3.42	4.60
A	28	GLY	A	51	GLU	-85.03	2.64	5.08
A	32	LYS	A	38	GLY	+114.69	1.93	6.06
A	52	PRO	A	55	SER	+123.75	6.02	5.12
A	70	ALA	A	74	PHE	+107.47	3.56	5.03
A	73	LEU	A	108	ALA	+78.29	5.00	4.98
A	77	PRO	A	104	VAL	+104.13	5.34	5.83
A	80	VAL	A	93	ALA	+87.23	6.55	5.90
A	94	ALA	A	98	TRP	+84.33	5.61	5.64
A	98	TRP	A	100	LEU	+120.55	7.58	4.13

https://doi.org/10.1371/journal.pone.0248001.t006

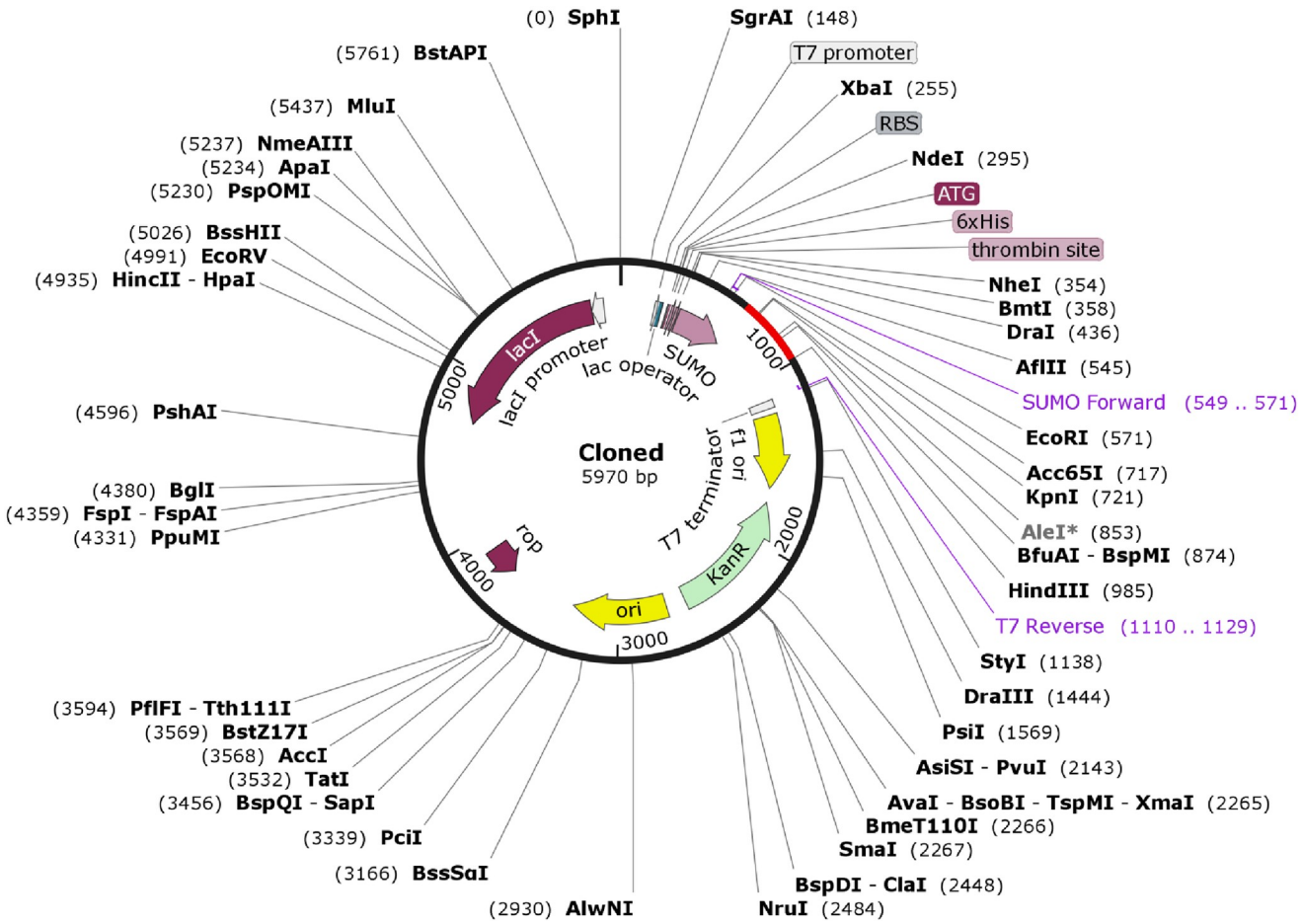

Fig 8. Cloning of the multi-epitope vaccine coding sequence in to pET-SUMO vector. Here the multi-epitope vaccine coding sequence of 329 bp is indicated by red color and pET-SUMO vector by black. A recombinant vector of 5970 bp is generated after insertion of the target nucleotide sequence.

https://doi.org/10.1371/journal.pone.0248001.g008

indicates spontaneous binding between the vaccine component and TLR-4. The selected docking complex is illustrated in Fig 9 along with molecular surface interaction as well as some bonding interactions. The elaborate interface residues between the vaccine component and TLR-4, bonding interactions and their distance are in the supplementary table (Table 1). The proper protein-protein docking between the vaccine component and TLR-4 will activate immune cascades for destroying the viral antigens 26.

Discussion

HTLV-I is the prime causative agent leading to a disabling inflammatory disease HAM/TSP and an aggressive malignancy named Adult T cell Leukemia. To date no effective vaccine and treatment is available for HTLV-1 infection. Our study is focused on development of an effective vaccine against HTLV-1 Tax protein which plays a major role in viral pathogenesis. While the CTL response unique to Tax is a common occurrence in many HTLV-1-carriers, Tax protein is a major antigenic target for HTLV-1-specific CTLs [57]. Tax protein is the first HTLV-1 protein to be expressed in an infected cell [1]. It was reported as a potential immunotherapeutic target against HAM/TSP and ATL [109]. HBZ is another HTLV-1 protein which also plays an important role in viral infectivity surged with survival and growth of leukemic cells [110].

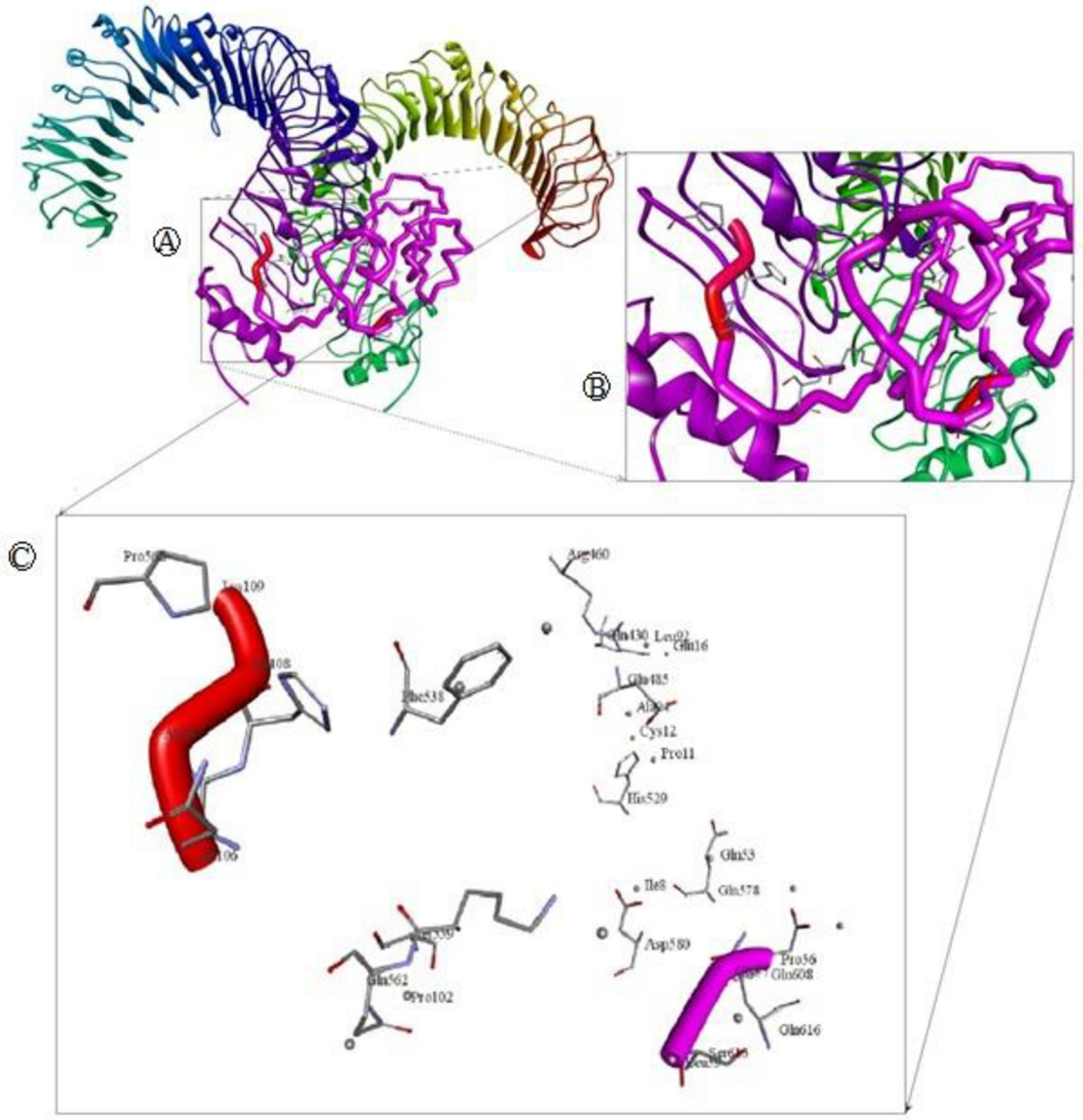

Fig 9. Molecular docking of multi epitope vaccine (MEV) with immune receptor (TLR4). (A) Whole Docked of MEV-TLR4 complex where MEV is purple and adjuvant colored red, chain A and B of TLR4 colored green and blue respectively. (B) Closer look of the main docking site. (C) Main interacting racedues and there interactions.

https://doi.org/10.1371/journal.pone.0248001.g009

Despite the fact HBZ performs an essential role in the proliferation of HTLV-1-infected cells, it could additionally provide a unique mechanism that lets them evade immune recognition [111]. Antigenicity prediction of Tax and HBZ using VaxiJen server at threshold 0.4 regarded Tax protein as probable antigen with score of 0.461 while HBZ as non antigen with score of 0.3063. These contrasting facts show that Tax is a potential candidate for vaccine design albeit HBZ is a potential drug target.

A physico-chemical analysis of the protein sequence was done by the Expasy server's Prot-Param tool. It revealed an instability index of 48.9, which denotes, this protein will be unstable *in vitro* because a value over 40 is considered unstable [70]. Interestingly this protein was also predicted to have a high aliphatic index; it is the total volume occupied by aliphatic side chains

and higher value is considered a positive factor for increased thermostability. Along with high extinction coefficient and negative GRAVY, the extents of other parameters imply the stability of the protein.

Results generated by secondary structure analysis tool PSIPRED and SOPMA showed the HTLV-1 Tax protein is adorned with 17.85% alpha helix and 54.96% random coils along with 5.26% extended strands and 21.53% beta turns. The abundance of coiled regions is an indication of higher conservation and stability of the model.

Tertiary structure of a protein determines its function and stability. To date no tertiary structure for HTLV-1 Tax is available in Protein Data Bank. We have generated a 3D structure of Tax protein using Phyre2 server and PS^2 server with 75% reliability. Often 3D structures generated by bioinformatics tools contain significant local distortions, including unphysical phi/psi angles, and steric clashes irregular H-hydrogen bonding networks, which make the structure models unfit for high-resolution analysis of functions. Refinement of the modeled structures can bring up a solution of this problem [78,112]. Best model predicted by 3Drefine was selected after refinement based on having the highest 3D score and least RMSD. After refinement the protein model was validated by ERRAT, Verify3D, PROCHECK. ERRAT interpreted the overall quality of the model with the quality factor 78.313; this score represents the proportion of the protein that falls below the rejection limit of 95%. Verify 3D shows 63.49% of the residues have 3D-1D score above 0.2 whereas a good quality model required 80% of the residues to have 3D-1D> = 0.2. Ramachandran plot, a plot of the dihedral angles—phi (φ) and psi (ψ)—of the amino acids contained in a peptide, generated by PROCHECK illustrates that 75.7% of the amino acid fall within the most favored region and 17.4% in additional allowed region. Analysis based on 118 structures of resolution of minimum 2.0 Angstroms and R-factor less or equal to 20%, a good quality model should have over 90% of its residues in the most favoured regions.

Most B cell epitopes are discontinuous epitopes, epitopes composed of amino acid residues residing in different regions of the protein, which are arranged together by the folding of the protein chain [113,114]. These residue groups cannot be isolated in the same conformation from the antigen [114]. As per concern linear B cell epitopes for HTLV-1 tax are curated for IEDB. Bepipred linear epitope prediction method was used to predict 13 B cell epitopes. Among the predicted epitopes 7 of them exhibited antigenicity. From our data, we see that only 2 epitopes were present inside while other 5 epitopes were found outside. These 5 epitopes, KEADDNDHEPQISPGGLEPPSEKHFR, DGTPMISGPCPKDGQPS, PFRNGY-MEPTLGQ, EHQITWDPIDGR, PITHTTPNIPPS, LSFPDPGLRPQ were assessed for toxicity and allergenicity which led as to select 2 non-allergenic and non-toxic B cell epitopes as potential vaccine candidate.

Once the proteins enter the host antigen-presenting cells (APC) of the, they are processed and then the T cell epitopes are proteolytically cleaved from the protein, and represented by the MHC molecules on the surface of APCs, exposing them to the receptors of T cells [114]. MHC class I molecules represent endogenous antigens such as intracellular bacterial, viral and tumor inducing proteins while MHC class II represent epitopes from the exogenous proteins. HTLV-1 Tax is preferentially produced endogenously in the cytoplasm of a host cell and localized in the cytoplasm with fewer speckle-like dots in the nucleus [115]. Induction of virus-specific CD8+ cytotoxic T lymphocytes (CTL) by MHC class I presented peptide is required for effective viral clearance [116]. MHC classI binding T cell epitopes were by NetCTL server and IEDB server. 94 MHC class I T cell epitopes were selected based on combined score from which 48 epitopes were extracted with IC50 less than 200nm. The lower the IC50 value the higher their binding affinity with the HLA molecules. About 16 epitopes interacted with more than 5 HLA alleles. Epitopes those which interacted with $\geq$5 MHC HLA-alleles are most likely

to be potential vaccine candidates [117,118]. Antigenicity assessment with VaxiJen server at threshold 0.4 mined 7 epitopes to be antigenic and among them 5 were immunogenic. Epitope conservancy was found 100% for all 5 epitopes as predicted by IEDB server. Toxicity and allergenicity evaluation of the epitopes rendered all selected epitopes to be non-toxic but only 3 of them are non allergenic. Efficiency of a multiepitope vaccine greatly relies on precise interaction between epitopes and HLA alleles. MHC class I alleles having interaction with LLFGYPVYV, ITWPLLPHV and GLLPFHSTL were searched for population coverage. From our study the highest population coverage was recorded for Mexico (90.21%) followed by England (89.88%) and South Africa which is located in Sub Saharan Africa. South America and North America showed population coverage of 74.95% and 76.82% respectively while in Japan it was slightly lower (70.92%). East Asia, North East Asia and South Asia had population coverage just above 70%. Cameroon, a country of central Africa exhibited 71.39% population coverage while UNited Arab of Emirates was found with the lowest percentage (1.2%) of coverage. The major HTLV-1 highly endemic regions are the sub-Saharan Africa, South America Southwestern part of Japan, the Caribbean area, Australo-Melanesia and foci in the Middle East [119]. Age, ethnicity, mode of transmission are important factors influencing variation in population coverage [119].

Previous study of vaccine designing on HTLV-1 partially aligns with our study in that they have selected epitopes LLFGYPVYV, QLGAFLTNV and GLLPFHSTL as potential candidates for multivalent vaccine [l] but differ in that QLGAFLTNV was excluded from our vaccine as it was found allergenic. Instead, a non-allergenic T cell epitope ITWPLLPHV was assigned to be a part of our designed vaccine.

Docking simulation was carried out on AutoDock to assess the binding efficiency of the selected T cell epitopes to HLA molecules. The epitopes were docked with HLA-A*0201 complexed with a peptide with the carboxyl-terminal group substituted by a methyl group. The peptide had binding energy of -5.0Kcal/mol with the HLA molecules whereas our selected T cell epitopes ITWPLLPHV, LLFGYPVYV and GLLPFHSTL showed higher binding energy, -8.8 Kcal/mol, -8.6 Kcal/mol, -8.5 Kcal/mol respectively suggesting satisfactory binding accuracy of the predicted epitopes. The ITWPLLPHV is stabilized by twelve hydrogen bonds, eleven hydrophobic bonds and one electrostatic bond while interacting with the receptor protein. It also forms hydrogen bonds with catalytic residue A:Tyr27 and hydrophobic interaction with A:Pro235 and B:Tyr463. ITWPLLPHV interacts through ten hydrogen bonds with the catalytic residue A:Tyr27 and B:Tyr463 and fifteen hydrophobic interactions with the catalytic residue A:Pro235 and B:Tyr463. GLLPFHSTL forms thirteen hydrogen bonding interactions with catalytic residue A:Tyr27, one electrostatic and four hydrophobic interactions observed with the catalytic residue A:Pro235 and B:Tyr463. Two potential B cell epitopes and three T cell epitopes were selected for multi-epitope vaccine construction. A suitable adjuvant was selected using the VaxinPad tool based on antigenicity. B cell epitopes were linked using GPGPG linker while MHC class I epitopes were joined by AAY linker. EAAK linkers were added with N -termini and C -termini of vaccine construct. These linkers are widely used in multi-epitope vaccine construction [120–122]. Vaccine construct with highest antigenicity among multiple combinations of these epitopes and linkers was selected for further investigation. The vaccine construct was tested for allergenicity by AllerTop 2.0 and AllergenFP which declared the vaccine construct as non-allergen. An important insight was gained from the physicochemical properties and secondary structure analysis of the protein performed by ProtParam and SOPMA tools. It was found stable with instability index 30.57, negative GRAVY score (-0.367) and high aliphatic index are indications of the stability of the vaccine construct. Results generated by secondary structure prediction tool SOPMA showed the vaccine structure is dominated by 25.69% alpha helix and 65.14% random coils. The abundance of coiled

regions indicates higher conservation and stability of the model [70,71,123]. As no suitable homologous template for homology modelling was not available for the vaccine construct peptide sequence ab initio modelling was executed using the I-Tasser tool to predict its 3D structure. Disulfide engineering was done at several positions on the vaccine tertiary structure. Disulfide bonds increase the stability of the protein [124].

Codon optimization was carried out in order to achieve high-level expression of our recombinant vaccine protein in *E. coli* (strain K12). Both the GC content (58.10%) and the codon adaptability index (1.0) were favourable for increased expression of the protein in bacteria. After codon adaptation the protein nucleic acid sequence was cloned in pET SUMO by TA cloning and expressed in *E. coli* strain DH5 alpha. DH5 alpha is a *E. coli* strain which allows easy selection of transformed colonies by blue-white screening. Again pET SUMO contains His Tag sequence at the 5' end of the vaccine sequence which allows easy purification of the desired protein.

Moreover, it is necessary to know the immune response of TLR4 against the vaccine protein. TLR4 is known to activate innate immunity against HTLV-1 as reported in several studies [125,126]. It has been widely demonstrated that TLR4 plays an important role in the recognition of endogenous molecules which are released by necrotic cells and injured tissues [127]. These molecules activate an intense proinflammatory response through interaction with TLR4 [128]. To assess the binding affinity between the vaccine and TLR4 a molecular docking was performed and analysis of the result showed good binding affinity between them. A high binding affinity to TLR4 supports the acceptability of our predicted multi-epitope vaccine.

Conclusion

Although the approaches in our study to predicting HTLV-1 TAX protein based epitopes and construction of a multi epitope vaccine were all in all *in silico*, the insights and outcome gained from the study can provide an elementary ground for expediting investigations related to designing and constructing epitope based vaccine against HTLV-1 in a wet lab and take it to *in vitro* studies from there. Further studies involving extensive laboratory assays and techniques might also render a higher a higher frequency and array of HLA molecules in terms of detecting the epitopes within the Tax protein.

Supporting information

S1 Table. Docking with TLR-4.
(DOCX)

S2 Table. T cell epitope.
(PDF)

S3 Table. B cell epitope.
(PDF)

Acknowledgments

We cordially thank Dr. Adnan Mannan, Associate Professor of the Department of Genetic Engineering and Biotechnology, University of Chittagong, for his suggestions and inspiration during our research proceedings.

Author Contributions

Conceptualization: Shah Md. Shahik.

Data curation: Md. Thosif Raza, Shah Md. Shahik.

Formal analysis: Md. Thosif Raza, Shah Md. Shahik.

Investigation: Md. Thosif Raza, Shagufta Mizan, Shah Md. Shahik.

Methodology: Md. Thosif Raza, Shagufta Mizan, Farhana Yasmin, Al-Shahriar Akash, Shah Md. Shahik.

Project administration: Md. Thosif Raza, Shah Md. Shahik.

Resources: Shagufta Mizan, Shah Md. Shahik.

Software: Md. Thosif Raza, Shagufta Mizan, Shah Md. Shahik.

Supervision: Shah Md. Shahik.

Validation: Md. Thosif Raza, Shah Md. Shahik.

Visualization: Md. Thosif Raza, Shah Md. Shahik.

Writing – original draft: Md. Thosif Raza, Shagufta Mizan, Farhana Yasmin, Al-Shahriar Akash, Shah Md. Shahik.

Writing – review & editing: Md. Thosif Raza, Shagufta Mizan, Shah Md. Shahik.

References

1. Asquith B, Mosley AJ, Heaps A, Tanaka Y, Taylor GP, McLean AR, et al. Quantification of the virus-host interaction in human T lymphotropic virus I infection. Retrovirology. 2005; 2(1):1–9. https://doi.org/10.1186/1742-4690-2-75 PMID: 16336683
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. Proceedings of the National Academy of Sciences. 1980; 77(12):7415–74159.
3. Salemi M, Desmyter J, Vandamme AM. Tempo and mode of human and simian T-lymphotropic virus (HTLV/STLV) evolution revealed by analyses of full-genome sequences. Molecular Biology and Evolution. 2000; 17(3):374–386. https://doi.org/10.1093/oxfordjournals.molbev.a026317 PMID: 10723738
4. Salemi M, Van Dooren S, Vandamme AM. Origin and evolution of human and simian T-cell lymphotropic viruses. AIDS Rev. 1999; 1(3):131–139.
5. Slattery JP, Franchini G, Gessain A. Genomic evolution, patterns of global dissemination, and interspecies transmission of human and simian T-cell leukemia/lymphotropic viruses. Genome research. 1999; 9(6):525–540. PMID: 10400920
6. Mahieux R, Chappey C, Georges-Courbot MC, Dubreuil G, Mauclere P, Georges A, et al. Simian T-cell lymphotropic virus type 1 from Mandrillus sphinx as a simian counterpart of human T-cell lymphotropic virus type 1 subtype D. Journal of virology. 1998; 72(12):10316–10322. https://doi.org/10.1128/JVI.72.12.10316-10322.1998 PMID: 9811783
7. Van Dooren S, Shanmugam V, Bhullar V, Parekh B, Vandamme AM, Heneine W, et al. Identification in gelada baboons (Theropithecus gelada) of a distinct simian T-cell lymphotropic virus type 3 with a broad range of Western blot reactivity. Journal of general virology. 2004; 85(2):507–519. https://doi.org/10.1099/vir.0.19630-0 PMID: 14769908
8. Gessain A, Mahieux R. Epidemiology, origin and genetic diversity of HTLV-1 retrovirus and STLV-1 simian affiliated retrovirus. Bulletin de la Societe de Pathologie Exotique (1990). 2000; 93(3):163–171.
9. Courgnaud V, Van Dooren S, Liegeois F, Pourrut X, Abela B, Loul S, et al. Simian T-cell leukemia virus (STLV) infection in wild primate populations in Cameroon: evidence for dual STLV type 1 and type 3 infection in agile mangabeys (Cercocebus agilis). Journal of virology. 2004; 78(9):4700–4709. https://doi.org/10.1128/jvi.78.9.4700-4709.2004 PMID: 15078952
10. Meertens L, Mahieux R, Mauclère P, Lewis J, Gessain A. Complete sequence of a novel highly divergent simian T-cell lymphotropic virus from wild-caught red-capped mangabeys (Cercocebus torquatus) from Cameroon: a new primate T-lymphotropic virus type 3 subtype. Journal of virology. 2002; 76(1):259–268. https://doi.org/10.1128/jvi.76.1.259-268.2002 PMID: 11739691

11. Goubau P, Van Brussel M, Vandamme AM, Liu HF, Desmyter J. A primate T-lymphotropic virus, PTLV-L, different from human T-lymphotropic viruses types I and II, in a wild-caught baboon (Papio hamadryas). Proceedings of the National Academy of Sciences. 1994; 91(7), 2848–2852.

12. Calattini S, Chevalier SA, Duprez R, Bassot S, Froment A, Mahieux R, et al. Discovery of a new human T-cell lymphotropic virus (HTLV-3) in Central Africa. Retrovirology. 2005; 2(30), 4–7.

13. Proietti FA, Carneiro-Proietti ABF, Catalan-Soares BC, Murphy EL. Global epidemiology of HTLV-I infection and associated diseases. Oncogene. 2005; 24, 6058–6068. https://doi.org/10.1038/sj.onc.1208968 PMID: 16155612

14. Gonçalves DU. Proietti FA, Ribas JGR, Araújo MG, Pinheiro SR, et al. Epidemiology, treatment, and prevention of human T-cell leukemia virus type 1-associated diseases. Clin. Microbiol. Rev. 2010; 23, 577–589. https://doi.org/10.1128/CMR.00063-09 PMID: 20610824

15. Mochizuki M, Yamaguchi K, Takatsuki K, Watanabe T, Mori S, Tajima H. HTLV-I and uveitis. The Lancet. 1992; 339(8801):1110.

16. Nakao K, Matsumoto M, Ohba N. Seroprevalence of antibodies to HTLV-I in patients with ocular disorders. British journal of ophthalmology. 1991; 75(2):76–78. https://doi.org/10.1136/bjo.75.2.76 PMID: 1995047

17. Kinoshita K, Amagasaki T, Ikeda S, Suzuyama J, Toriya K, Nishino K, et al. Preleukemic state of adult T cell leukemia: abnormal T lymphocytosis induced by human adult T cell leukemia-lymphoma virus. Blood. 1985; 66:120–127. PMID: 2988665

18. Kinoshita K, Hino S, Amagasaki T, Yamada Y, Kamihira S, Ichimaru M et al. Development of adult T-cell leukemia-lymphoma (ATL) in two anti-ATL-associated antigen-positive healthy adults. Gann. 1982; 73(5): 684–685. PMID: 6984681

19. Matsumoto M. Adult T-cell leukemia-lymphoma in Kagoshima district, southwestern Japan: clinical and hematological characteristics. Jpn J Clin Oncol. 1979; 9:325–336.

20. Araujo AQ, Silva MT. The HTLV-1 neurological complex. The Lancet Neurology. 2006; 5(12):1068–1076. https://doi.org/10.1016/S1474-4422(06)70628-7 PMID: 17110288

21. Kimura I. HABA (HTLV-I associated bronchiolo-alveolar disorder). Nihon Kyobu Shikkan Gakkai Zasshi. 1992; 30(5):787. PMID: 1630042

22. Mochizuki M, Watanabe T, Yamaguchi K, Yoshimura K, Nakashima S, Shirao M, et al. Uveitis associated with human T-cell lymphotropic virus type I. American journal of ophthalmology. 1992; 114 (2):123–129. https://doi.org/10.1016/s0002-9394(14)73974-1 PMID: 1642286

23. Cruz BA, Catalan-Soares B, Proietti F. Higher prevalence of fibromyalgia in patients infected with human T cell lymphotropic virus type I. The Journal of rheumatology. 2006; 33(11):2300–2303. PMID: 17086610

24. Oliveira P, Castro NM, Carvalho EM. Urinary and sexual manifestations of patients infected by HTLV-I. Clinics. 2007; 62(2):191–196. https://doi.org/10.1590/s1807-59322007000200015 PMID: 17505705

25. Goon PK, Bangham CR. Interference with immune function by HTLV-1. Clinical and experimental immunology. 2004; 137(2):234. https://doi.org/10.1111/j.1365-2249.2004.02524.x PMID: 15270838

26. Kannian P, Green PL. Human T lymphotropic virus type 1 (HTLV-1): molecular biology and oncogenesis. Viruses. 2010; 2(9):2037–2077. https://doi.org/10.3390/v2092037 PMID: 21994719

27. Richardson JH, Edwards AJ, Cruickshank JK, Rudge P, Dalgleish AG. In vivo cellular tropism of human T-cell leukemia virus type 1. Journal of Virology. 1990; 64:5682–5687. https://doi.org/10.1128/JVI.64.11.5682-5687.1990 PMID: 1976827

28. Piñon JD, Klasse PJ, Jassal SR, Welson S, Weber J, Brighty DW, et al. Human T-Cell Leukemia Virus Type 1 Envelope Glycoprotein gp46 Interacts with Cell Surface Heparan Sulfate Proteoglycans. Journal of Virology. 2003; 77: 9922–9930. https://doi.org/10.1128/jvi.77.18.9922-9930.2003 PMID: 12941902

29. Spillmann D. Heparan sulfate: Anchor for viral intruders? Biochimie. 2001; 83:811–817. https://doi.org/10.1016/s0300-9084(01)01290-1 PMID: 11530214

30. Jin Q, Agrawal L, VanHorn-Ali Z. Alkhatib G. Infection of CD4+ T lymphocytes by the human T cell leukemia virus type 1 is mediated by the glucose transporter GLUT-1: Evidence using antibodies specific to the receptor's large extracellular domain. Virology. 2006; 349:184–196. https://doi.org/10.1016/j.virol.2006.01.045 PMID: 16519917

31. Ghez D, Lepelletier Y, Lambert S, Fourneau GM, Blot V, Janvier S et al. Neuropilin-1 Is Involved in Human T-Cell Lymphotropic Virus Type 1 Entry. Journal of Virology. 2006; 80: 6844–6854. https://doi.org/10.1128/JVI.02719-05 PMID: 16809290

32. Lee TH, Coligan JE, Homma T, Mclana, Tachibana N, Essex M. Human T-cell leukemia virus-associated membrane antigens: identity of the major antigens recognized after virus infection. Proceedings

of the National Academy of Sciences. 1984; 81(12), 3856–3860. https://doi.org/10.1073/pnas.81.12.3856 PMID: 6328528

33. Nam SH, Kidokoro M, Shida H, Hatanaka M. Processing of gag precursor polyprotein of human T-cell leukemia virus type I by virus-encoded protease. Journal of Virology. 1988; 62: 3718–3728. https://doi.org/10.1128/JVI.62.10.3718-3728.1988 PMID: 2843670

34. Paine E, Gu R, Ratner L. Structure and expression of the human T-cell leukemia virus type 1 envelope protein. Virology. 1994; 199: 331–338. https://doi.org/10.1006/viro.1994.1131 PMID: 8122365

35. Grassmann R, Berchtold S, Radant I, Alt M, Fleckenstein B, Sodroski JG, et al. Role of human T-cell leukemia virus type 1 X region proteins in immortalization of primary human lymphocytes in culture. Journal of Virology. 1992; 66: 4570–4575. https://doi.org/10.1128/JVI.66.7.4570-4575.1992 PMID: 1351105

36. Grassmann R, Dengler C, Müller-Fleckenstein I, Fleckenstein B, McGuire K, Dokhelar MC, et al. Transformation to continuous growth of primary human T lymphocytes by human T-cell leukemia virus type I X-region genes transduced by a Herpesvirus saimiri vector. Proc. Natl. Acad. Sci. U. S. A. 1989; 86(9):3351–3355. https://doi.org/10.1073/pnas.86.9.3351 PMID: 2541443

37. Nerenberg M, Hinrichs SM, Reynolds RK, Khoury G, Jay G. The tat gene of human Lymphotropic virus type I induces mesenchymal tumors in transgenic mice. Science. 1987; 237:1324–1329. https://doi.org/10.1126/science.2888190 PMID: 2888190

38. Tanaka A, Takahashi C, Yamaoka S, Nosaka T, Maki M, Hatanaka M. Oncogenic transformation by the tax gene of human T-cell leukemia virus type I in vitro. Proc. Natl. Acad. Sci. U. S. A. 1990; 87(3):1071–1075. https://doi.org/10.1073/pnas.87.3.1071 PMID: 2300570

39. Yamaoka S, Tobe T, Hatanaka M. Tax protein of human T-cell leukemia virus type I is required for maintenance of the transformed phenotype. Oncogene. 1992; 7:433–437. PMID: 1549359

40. Robek MD, Ratner L. Immortalization of CD4+ and CD8+ T Lymphocytes by Human T-Cell Leukemia Virus Type 1 Tax Mutants Expressed in a Functional Molecular Clone. Journal of Virology. 1999; 73:4856–4865. https://doi.org/10.1128/JVI.73.6.4856-4865.1999 PMID: 10233947

41. Ross TM, Pettiford SM, Green PL. The tax gene of human T-cell leukemia virus type 2 is essential for transformation of human T lymphocytes. Journal of Virology. 1996; 70:5194–5202. https://doi.org/10.1128/JVI.70.8.5194-5202.1996 PMID: 8764028

42. Bartoe JT, Albrecht B, Collins ND, Robek MD, Ratner L, Green PL, et al. Functional role of pX open reading frame II of human T-lymphotropic virus type 1 in maintenance of viral loads in vivo. Journal of virology. 2000; 74(3):1094–1100. https://doi.org/10.1128/jvi.74.3.1094-1100.2000 PMID: 10627519

43. Derse D, Mikovits J, Ruscetti F. X-I and X-II open reading frames of HTLV-I are not required for virus replication or for immortalization of primary T-cells in vitro. Virology. 1997; 237:123–128. https://doi.org/10.1006/viro.1997.8781 PMID: 9344914

44. Collins ND, Newbound GC, Albrecht B, Beard JL, Ratner L, Lairmore MD. Selective ablation of human T-cell lymphotropic virus type 1 p12I reduces viral infectivity in vivo. Blood, The Journal of the American Society of Hematology. 1998; 91(12):4701–4707. PMID: 9616168

45. Cavanagh MH, Landry S, Audet B, Arpin-André C, Hivin P, Paré MÈ, et al. HTLV-I antisense transcripts initiating in the 3'LTR are alternatively spliced and polyadenylated. Retrovirology. 2006; 3(1):15. https://doi.org/10.1186/1742-4690-3-15 PMID: 16512901

46. Eferl R, Wagner EF. AP-1: a double-edged sword in tumorigenesis. Nature reviews. Cancer. 2003; 3(11):859–868. https://doi.org/10.1038/nrc1209 PMID: 14668816

47. Milde-Langosch K. The Fos family of transcription factors and their role in tumourigenesis. European journal of cancer. 2005; 41(16):2449–2461. https://doi.org/10.1016/j.ejca.2005.08.008 PMID: 16199154

48. Hernandez JM, Floyd DH, Weilbaecher KN, Green PL, Boris-Lawrie K. Multiple facets of junD gene expression are atypical among AP-1 family members. Oncogene. 2008; 27(35):4757–4767. https://doi.org/10.1038/onc.2008.120 PMID: 18427548

49. Basbous J, Arpin C, Gaudray G, Piechaczyk M, Devaux C, Mesnard JM. The HBZ factor of human T-cell leukemia virus type I dimerizes with transcription factors JunB and c-Jun and modulates their transcriptional activity. Journal of Biological Chemistry. 2003; 278(44):43620–43627. https://doi.org/10.1074/jbc.M307275200 PMID: 12937177

50. Clerc I, Polakowski N, André-Arpin C, Cook P, Barbeau B, Mesnard JM, et al. An interaction between the human T cell leukemia virus type 1 basic leucine zipper factor (HBZ) and the KIX domain of p300/CBP contributes to the down-regulation of tax-dependent viral transcription by HBZ. Journal of Biological Chemistry. 2008; 283(35):23903–23913.

51. Thébault S, Basbous J, Hivin P, Devaux C, Mesnard JM. HBZ interacts with JunD and stimulates its transcriptional activity. FEBS letters. 2004; 562(1–3):165–170. https://doi.org/10.1016/S0014-5793(04)00225-X PMID: 15044019

52. Lemasson I, Lewis MR, Polakowski N, Hivin P, Cavanagh MH, Thébault S, et al. Human T-cell leukemia virus type 1 (HTLV-1) bZIP protein interacts with the cellular transcription factor CREB to inhibit HTLV-1 transcription. Journal of virology. 2007; 81(4):1543–1553. https://doi.org/10.1128/JVI.00480-06 PMID: 17151132

53. Satou Y, Yasunaga JI, Yoshida M, Matsuoka M. HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells. Proceedings of the National Academy of Sciences. 2006; 103(3):720–715. https://doi.org/10.1073/pnas.0507631103 PMID: 16407133

54. Yoshida M. Multiple viral strategies of HTLV-1 for dysregulation of cell growth control. Annual review of immunology. 2001; 19(1):475–496. https://doi.org/10.1146/annurev.immunol.19.1.475 PMID: 11244044

55. Boxus M, Twizere JC, Legros S, Dewulf JF, Kettmann R, Willems L. The HTLV-1 tax interactome. Retrovirology. 2008; 5(1):76. https://doi.org/10.1186/1742-4690-5-76 PMID: 18702816

56. Takeda S, Maeda M, Morikawa S, Taniguchi Y, Yasunaga JI, Nosaka K, et al. Genetic and epigenetic inactivation of tax gene in adult T-cell leukemia cells. International journal of cancer. 2004; 109(4):559–567. https://doi.org/10.1002/ijc.20007 PMID: 14991578

57. Kannagi M, Harada S, Maruyama I, Inoko H, Igarashi H, Kuwashima G, et al. Predominant recognition of human T cell leukemia virus type I (HTLV-I) pX gene products by human CD8+ cytotoxic T cells directed against HTLV-I-infected cells. International immunology. 1991; 3(8):761–767. https://doi.org/10.1093/intimm/3.8.761 PMID: 1911545

58. Pique C, Tursz T, Dokhelar MC. Mutations introduced along the HTLV-I envelope gene result in a non-functional protein: a basis for envelope conservation?. The EMBO Journal. 1990; 9(13):4243–4248. PMID: 2124968

59. Seki M, Sashiyama H, Hayami M, Shida H. Intracellular processing and immunogenicity of the envelope proteins of human T-cell leukemia virus type I that are expressed from recombinant vaccinia viruses. Virus genes. 1990; 3(3):235–249. https://doi.org/10.1007/BF00393183 PMID: 2189258

60. Sagara Y, Inoue Y, Shiraki H, Jinno A, Hoshino H. Maeda Y. Identification and mapping of functional domains on human T-cell lymphotropic virus type 1 envelope proteins by using synthetic peptides. Journal of virology. 1996; 70(3):1564–1569. https://doi.org/10.1128/JVI.70.3.1564-1569.1996 PMID: 8627675

61. Rosenberg AR., Delamarre L, Preira A, Dokhélar MC. Analysis of functional conservation in the surface and transmembrane glycoprotein subunits of human T-cell leukemia virus type 1 (HTLV-1) and HTLV-2. Journal of virology. 1998; 72(9):7609–7614. https://doi.org/10.1128/JVI.72.9.7609-7614.1998 PMID: 9696862

62. Miranda ACAM, Barreto FK, Amarante MFDC, Batista EDS, Cunha JPM, Vallve MDLF, et al. Molecular characterization of HTLV-1 gp46 glycoprotein from health carriers and HAM/TSP infected individuals. Virology Journal. 2013; 10(75):1–10.

63. Delamarre L, Rosenberg AR, Pique C, Pham D, Callebaut I, Dokhélar MC. The HTLV-I envelope glycoproteins: structure and functions. JAIDS Journal of Acquired Immune Deficiency Syndromes. 1996; 13:85–91. https://doi.org/10.1097/00042560-199600001-00015 PMID: 8797709

64. Lairmore MD, Rudolph DL, Roberts BD, Dezzutti CS, Lal RB. Characterization of a B-cell immunodominant epitope of human T-lymphotropic virus type 1 (HTLV-I) envelope gp46. Cancer letters. 1992; 66(1):11–20. https://doi.org/10.1016/0304-3835(92)90274-y PMID: 1360328

65. Palker TJ, Tanner ME, Scearce RM, Streilein RD, Clark ME, Haynes BF. Mapping of immunogenic regions of human T cell leukemia virus type I (HTLV-I) gp46 and gp21 envelope glycoproteins with env-encoded synthetic peptides and a monoclonal antibody to gp46. The Journal of Immunology. 1989; 142(3):971–978. PMID: 2563272

66. Horal P, Hall WW, Svennerholm B, Lycke J, Jeansson S, Rymo L, et al. Identification of type-specific linear epitopes in the glycoproteins gp46 and gp21 of human T-cell leukemia viruses type I and type II using synthetic peptides. Proceedings of the National Academy of Sciences. 1991; 88(13):5754–5758. https://doi.org/10.1073/pnas.88.13.5754 PMID: 1712105

67. Gessain ANTOINE, Louie A, Gout O, Gallo RC, Franchini G. Human T-cell leukemia-lymphoma virus type I (HTLV-I) expression in fresh peripheral blood mononuclear cells from patients with tropical spastic paraparesis/HTLV-I-associated myelopathy. Journal of virology. 1991; 65(3):1628–1633. https://doi.org/10.1128/JVI.65.3.1628-1633.1991 PMID: 1995955

68. Kinoshita T, Shimoyama M, Tobinai K, Ito M, Ito SI, Ikeda S, et al. Detection of mRNA for the tax1/rex1 gene of human T-cell leukemia virus type I in fresh peripheral blood mononuclear cells of adult T-cell leukemia patients and viral carriers by using the polymerase chain reaction. Proceedings of the

National Academy of Sciences. 1989; 86(14): 5620–5624. https://doi.org/10.1073/pnas.86.14.5620 PMID: 2787512

69. Jaworski E, Narayanan A, Van Duyne R, Shabbeer-Meyering S, Iordanskiy S, Saifuddin M, et al. Human T-lymphotropic virus type 1-infected cells secrete exosomes that contain Tax protein. J Biol Chem. 2014; 289(32): 22284–22305. https://doi.org/10.1074/jbc.M114.549659 PMID: 24939845

70. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. InThe proteomics protocols handbook 2005 (pp. 571–607). Humana press.

71. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics. 1995; 11(6):681–684. https://doi.org/10.1093/bioinformatics/11.6.681 PMID: 8808585

72. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics. 2000; 16(4):404–405. https://doi.org/10.1093/bioinformatics/16.4.404 PMID: 10869041

73. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nature protocols. 2015; 10(6):845–58. https://doi.org/10.1038/nprot.2015.053 PMID: 25950237

74. Chen CC, Hwang JK, Yang JM. (PS)2: protein structure prediction server. Nucleic Acids Res. 2006; 34:152–157.

75. Chen CC, Hwang JK, Yang JM. (PS)2-v2: template-based protein structure prediction server. Bmc Bioinformatics. 2009; 10(1):366. https://doi.org/10.1186/1471-2105-10-366 PMID: 19878598

76. Söding J. Protein homology detection by HMM–HMM comparison. Bioinformatics. 2005; 21(7):951–960. https://doi.org/10.1093/bioinformatics/bti125 PMID: 15531603

77. Jefferys BR, Kelley LA, Sternberg MJ. Protein folding requires crowd control in a simulated cell. Journal of molecular biology. 2010; 397(5):1329–1338. https://doi.org/10.1016/j.jmb.2010.01.074 PMID: 20149797

78. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophysical journal. 2011; 101(10):2525–2534. https://doi.org/10.1016/j.bpj.2011.10.024 PMID: 22098752

79. Bhattacharya D, Nowotny J, Cao R, Cheng J. 3Drefine: an interactive web server for efficient protein structure refinement. Nucleic acids research. 2016; 44(W1):406–409. https://doi.org/10.1093/nar/gkw336 PMID: 27131371

80. Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. Journal of molecular biology. 1993; 231(4):1049–1067. https://doi.org/10.1006/jmbi.1993.1351 PMID: 8515464

81. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein science. 1993; 2(9):1511–1519. https://doi.org/10.1002/pro.5560020916 PMID: 8401235

82. Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. InMethods in enzymology 1997 Jan 1 (Vol. 277, pp. 396–404). Academic Press.

83. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic acids research. 2017; 45(W1):24–29. https://doi.org/10.1093/nar/gkx346 PMID: 28472356

84. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology. 2001; 305(3):567–580. https://doi.org/10.1006/jmbi.2000.4315 PMID: 11152613

85. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC bioinformatics. 2007; 8(1):4. https://doi.org/10.1186/1471-2105-8-4 PMID: 17207271

86. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v. 2—a server for in silico prediction of allergens. Journal of molecular modeling. 2014; 20(6):2278. https://doi.org/10.1007/s00894-014-2278-5 PMID: 24878803

87. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP. Open Source Drug Discovery Consortium. In silico approach for predicting toxicity of peptides and proteins. PloS one. 2013; 8(9):73957.

88. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC bioinformatics. 2007; 8(1):424. https://doi.org/10.1186/1471-2105-8-424 PMID: 17973982

89. Buus S, Lauemøller SL, Worning P, Kesmir C, Frimurer T, et al. Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee'artificial neural network approach. Tissue antigens. 2003; 62(5):378–384. https://doi.org/10.1034/j.1399-0039.2003.00112.x PMID: 14617044

90. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC bioinformatics. 2005; 6(1):132. https://doi.org/10.1186/1471-2105-6-132 PMID: 15927070

91. 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(1):1–5. https://doi.org/10.1186/1471-2105-7-153 PMID: 16545123

92. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. PLoS Comput Biol. 2013; 9(10):e1003266. https://doi.org/10.1371/journal.pcbi.1003266 PMID: 24204222

93. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry. 2010; 31(2):455–461. https://doi.org/10.1002/jcc.21334 PMID: 19499576

94. Wheeler DL, Chappey C, Lash AE, Leipe DD, Madden TL, Schuler GD, et al. Database resources of the national center for biotechnology information. Nucleic acids research. 2000; 28(1):10–14. https://doi.org/10.1093/nar/28.1.10 PMID: 10592169

95. Maupetit J, Derreumaux P, Tuffery P. PEP-FOLD: an online resource for de novo peptide structure prediction. Nucleic acids research. 2009; 37(suppl_2):498–503.

96. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. Advanced drug delivery reviews. 2013; 65(10):1357–1369. https://doi.org/10.1016/j.addr.2012.09.039 PMID: 23026637

97. Nagpal G, Chaudhary K, Agrawal P, Raghava GP. Computer-aided prediction of antigen presenting cell modulators for designing peptide-based vaccine adjuvants. Journal of translational medicine. 2018; 16(1):181. https://doi.org/10.1186/s12967-018-1560-1 PMID: 29970096

98. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols. 2010; 5(4):725–738. https://doi.org/10.1038/nprot.2010.5 PMID: 20360767

99. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nature methods. 2015; 12(1):7–8. https://doi.org/10.1038/nmeth.3213 PMID: 25549265

100. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. Nucleic acids research. 2015; 43(W1):174–181. https://doi.org/10.1093/nar/gkv342 PMID: 25883148

101. Janson G, Zhang C, Prado MG, Paiardini A. PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL. Bioinformatics. 2017; 33(3):444–446. https://doi.org/10.1093/bioinformatics/btw638 PMID: 28158668

102. Craig DB, Dombkowski AA. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. BMC bioinformatics. 2013; 14(1):1–7. https://doi.org/10.1186/1471-2105-14-346 PMID: 24289175

103. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic acids research. 2005; 33(suppl_2):526–531. https://doi.org/10.1093/nar/gki376 PMID: 15980527

104. Kambara H, Liu F, Zhang X, Liu P, Bajrami B, Teng Y, et al. Gasdermin D exerts anti-inflammatory effects by promoting neutrophil death. Cell reports. 2018; 22(11):2924–2936. https://doi.org/10.1016/j.celrep.2018.02.067 PMID: 29539421

105. Choi ES, Lee SG, Lee SJ, Kim E. Rapid detection of 6×-histidine-labeled recombinant proteins by immunochromatography using dye-labeled cellulose nanobeads. Biotechnology letters. 2015; 37(3):627–632. https://doi.org/10.1007/s10529-014-1731-y PMID: 25388454

106. Vaure C, Liu Y. A comparative review of toll-like receptor 4 expression and functionality in different animal species. Frontiers in immunology. 2014; 5:316. https://doi.org/10.3389/fimmu.2014.00316 PMID: 25071777

107. Duhovny D, Nussinov R, Wolfson HJ. Efficient unbound docking of rigid molecules. InInternational workshop on algorithms in bioinformatics 2002 (pp. 185–200). Springer, Berlin, Heidelberg.

108. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic acids research. 2005; 33(suppl_2):363–367. https://doi.org/10.1093/nar/gki481 PMID: 15980490

109. Kannagi M, Hasegawa A, Nagano Y, Kimpara S, Suehiro Y. Impact of host immunity on HTLV-1 pathogenesis: potential of Tax-targeted immunotherapy against ATL. Retrovirology. 2019; 16(1):23. https://doi.org/10.1186/s12977-019-0484-z PMID: 31438973

110. Enose-Akahata Y, Vellucci A, Jacobson S. Role of HTLV-1 Tax and HBZ in the pathogenesis of HAM/TSP. Frontiers in microbiology. 2017; 8:2563. https://doi.org/10.3389/fmicb.2017.02563 PMID: 29312243

111. Suemori K, Fujiwara H, Ochi T, Ogawa T, Matsuoka M, Matsumoto T, et al. HBZ is an immunogenic protein, but not a target antigen for human T-cell leukemia virus type 1-specific cytotoxic T lymphocytes. Journal of General Virology. 2009; 90(8):1806–1811.

112. Patwary NIA, Islam MS, Sohel M, Ara I, Sikder MOF, Shahik SM. In silico structure analysis and epitope prediction of E3 CR1-beta protein of Human Adenovirus E for vaccine design. Biomedical Journal. 2016; 39:382–390. https://doi.org/10.1016/j.bj.2016.11.004 PMID: 28043417

113. Van Regenmortel MH. Synthetic peptide vaccines and the search for neutralization B cell epitopes. InHIV/AIDS: Immunochemistry, Reductionism and Vaccine Design 2019 (pp. 25–37). Springer, Cham.

114. Palatnik-de-Sousa CB, Soares ID, Rosa DS. Epitope discovery and Synthetic Vaccine design. Frontiers in immunology. 2018; 9:826. https://doi.org/10.3389/fimmu.2018.00826 PMID: 29720983

115. Forlani G, Baratella M, Tedeschi A, Pique C, Jacobson S, Accolla RS. HTLV-1 HBZ protein resides exclusively in the cytoplasm of infected cells in asymptomatic carriers and HAM/TSP patients. Frontiers in Microbiology. 2019; 10:819. https://doi.org/10.3389/fmicb.2019.00819 PMID: 31080441

116. Van Montfoort N, van der Aa E, Woltman AM. Understanding MHC class I presentation of viral antigens by human dendritic cells as a basis for rational design of therapeutic vaccines. Frontiers in immunology. 2014; 5:182. https://doi.org/10.3389/fimmu.2014.00182 PMID: 24795724

117. Adhikari UK, Rahman MM. Overlapping CD8+ and CD4+ T-cell epitopes identification for the progression of epitope-based peptide vaccine from nucleocapsid and glycoprotein of emerging Rift Valley fever virus using immunoinformatics approach. Infection, Genetics and Evolution. 2017; 56:75–91. https://doi.org/10.1016/j.meegid.2017.10.022 PMID: 29107145

118. Hossain MU, Keya CA, Das KC, Hashem A, Omar TM, Khan M, et al. An immunopharmacoinformatics approach in development of vaccine and drug candidates for West Nile virus. Frontiers in Chemistry. 2018; 6:246. https://doi.org/10.3389/fchem.2018.00246 PMID: 30035107

119. Gessain A, Cassar O. Epidemiological aspects and world distribution of HTLV-1 infection. Frontiers in microbiology. 2012; 3:388. https://doi.org/10.3389/fmicb.2012.00388 PMID: 23162541

120. Shey RA, Ghogomu SM, Esoh KK, Nebangwa ND, Shintouo CM, Nongley NF, et al. In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. Scientific reports. 2019; 9(1):1–8.

121. Ali M, Pandey RK, Khatoon N, Narula A, Mishra A, Prajapati VK. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. Scientific reports. 2017; 7(1):1–3. https://doi.org/10.1038/s41598-016-0028-x PMID: 28127051

122. Khatoon N, Pandey RK, Prajapati VK. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. Scientific reports. 2017; 7(1):1–2. https://doi.org/10.1038/s41598-016-0028-x PMID: 28127051

123. Guruprasad K, Reddy BB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Engineering, Design and Selection. 1990; 4(2):155–161. https://doi.org/10.1093/protein/4.2.155 PMID: 2075190

124. Zavodszky M, Chen CW, Huang JK, Zolkiewski M, Wen L, Krishnamoorthi R. Disulfide bond effects on protein stability: Designed variants of Cucurbita maxima trypsin inhibitor-V. Protein Science. 2001; 10(1):149–160. https://doi.org/10.1110/ps.26801 PMID: 11266603

125. Hyun J, Ramos JC, Toomey N, Balachandran S, Lavorgna A, Harhaj E, et al. Oncogenic human T-cell lymphotropic virus type 1 tax suppression of primary innate immune signaling pathways. Journal of virology. 2015; 89(9):4880–4893. https://doi.org/10.1128/JVI.02493-14 PMID: 25694597

126. Yokota SI, Okabayashi T, Fujii N. The battle between virus and host: modulation of Toll-like receptor signaling pathways by virus infection. Mediators of inflammation. 2010; 2010. https://doi.org/10.1155/2010/184328 PMID: 20672047

127. Yu L, Wang L, Chen S. Endogenous toll-like receptor ligands and their biological significance. Journal of cellular and molecular medicine. 2010; 14(11):2592–2603. https://doi.org/10.1111/j.1582-4934.2010.01127.x PMID: 20629986

128. Molteni M, Gemma S, Rossetti C. The role of toll-like receptor 4 in infectious and noninfectious inflammation. Mediators of inflammation. 2016; 2016. https://doi.org/10.1155/2016/6978936 PMID: 27293318