Vaccine 36 (2018) 2262–2272

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Immunoinformatics approaches to design a novel multi-epitope subunit vaccine against HIV infection

Rajan Kumar Pandey [a], Rupal Ojha [a], Veeranarayanan Surya Aathmanathan [b], Muthukalingan Krishnan [a,b], Vijay Kumar Prajapati [a,*]

[a] Department of Biochemistry, School of Life Sciences, Central University of Rajasthan, Bandarsindri, Kishangarh, Ajmer 305817, Rajasthan, India
[b] Department of Environmental Biotechnology, Bharathidasan University, Tiruchirappalli 620 024, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 27 October 2017
Received in revised form 19 February 2018
Accepted 15 March 2018
Available online 20 March 2018

Keywords:
Immunoinformatics
HIV infection
AIDS
Subunit vaccine
Epitopes
Adjuvant

ABSTRACT

The end goal of HIV vaccine designing requires novel strategies to elicit a strong humoral and cell-mediated immune response. The emergence of drug resistance and the requirement of next line treatment necessitate the finding of the potential and immunogenic vaccine candidate. This study employed a novel immunoinformatics approach to design multi-epitope subunit vaccine against HIV infection. Here, we designed the subunit vaccine by the combination of CTL, HTL and BCL epitopes along with suitable adjuvant and linkers. Physiochemical characterization of subunit vaccine was assessed to ensure its thermostability, theoretical PI, and amphipathic behavior. In further assessment, subunit vaccine was found to be immunogenic with the capability to generate humoral and cell-mediated immune response. Further, homology modeling and refinement was performed and the refined modeled structure was used for molecular docking with the immune receptor (TLR-3) present on lymphocyte cells. Consequently, molecular dynamics simulation ensured the molecular interaction between TLR-3 and subunit vaccine candidate. Disulfide engineering was performed by placing the cysteine residues in the region of high mobility to enhance the vaccine stability. At last, *in silico* cloning was performed to warrant the translational efficiency and microbial expression of the designed vaccine.

1. Introduction

The human immunodeficiency virus (HIV) has become a major public health concern for the developing world. Around 36.7 million people throughout the world are currently living with this severe complication, and almost 40% patients are being unaware of their diseased condition. The worst-hit areas are regions of sub-Saharan Africa where one in every 25 individuals is affected by this retroviral syndrome [1]. In spite of the steps taken by the National AIDS Control Organization (NACO) to spread awareness about this syndrome, in the year 2015, around 2.1 million new individuals were enthralled to this infection. Although there are no specific symptoms of HIV infection, some common flu-like symptoms may be experienced [2]. Over the time, HIV infection leads to acquired immunodeficiency syndrome (AIDS) which was common in intravenous drug users, male homosexuals and hemophiliacs in the United States [3]. Broadly, HIV is classified into two categories namely HIV-1 and HIV-2. Among these two strains, formerly is responsible for the 95% of the global spread of AIDS, while the influence of HIV-2 is mainly restricted to Central and Western Africa. The strain of HIV-1 can be additionally ordered into four groups to be specific M, N, O, and P [4]. Of these, M (95% cases) is the major group further classified into nine genetically distinct subtypes namely A, B, C, D, F, G, H, J and K. Of these, B-subtype is dominant in the Australia, America and Western Europe while C-subtype has shown its prevalence in the southern Africa and India. Both these subtype contributes to the major HIV infectious cases, globally.

HIV genome consists of gag, pol, and env genes flanked by long terminal repeats. Apart from this, HIV also has six regulatory genes that are rev, tat, nef, vif, vpr, and vpu [5,6]. Rev is an RNA binding protein that performs a nuclear export of intron-containing HIV-1 RNA [7]. Tat protein binds to trans-activation response RNA, downstream of the transcription start site, hence is crucial for activating transcription. [8]. Nef protein is crucial for viral survival by aiding in evasion of the immune system and anti-apoptosis of the HIV infected cells [9]. Vif protein hinders two human enzymes namely APOBEC3F and APOBEC3G, the cellular cytosine deaminases by forming a complex with these two enzymes. It can also interact

* Corresponding author at: Department of Biochemistry, Central University of Rajasthan, NH-8, Bandarsindri, Ajmer, Rajasthan 305817, India.
E-mail address: vkprajapati@curaj.ac.in (V.K. Prajapati).

https://doi.org/10.1016/j.vaccine.2018.03.042
0264-410X/

and inhibit translation of APOBEC3G [10]. Vir protein causes cell cycle arrest in G2 phase which gives ample amount of time to the virus for the expression of the viral genome [11]. HIV protein Vpu works along with Nef to escape host immune surveillance. Vpu depletes virion tethering bone marrow stromal antigen-2 (BST2). The viral RNA is reverse transcribed into proviral DNA by reverse transcriptase encoded by the viral RNA. Integrase performs the task of integrating proviral DNA into the host genome by performing various cleaving reactions followed by ligation [12]. Newly formed HIV particles are released with a thick layer of radially arranged Gag and Gag-pol precursors. It is the protease that performs the cleaving of these precursors into their mature counterparts in an organized series of steps. These proteases are aspartyl proteases in nature and perform cleaving during or shortly after budding [5].

Currently, available HIV/AIDS treatment consists of antiretroviral therapy (ART). ART has evolved from days of high toxicity and pill burden to a more effective option that has less toxic side effects. Although with current treatment, complete eradication of the disease is not possible, ART helps to increase the life expectancy of AIDS patient. A lot of efforts are still needed to find some drug alternatives like a vaccine. Vaccines symbolize the most cost-effective life-saving device in history that prepares our body beforehand to fight the deadly and contagious disease. Many approaches can be taken for designing of a vaccine which includes its mode of infections and how immune system of the host responds. Types of vaccine that currently exist are live attenuated, inactivated vaccine, subunit vaccine, toxoid vaccine, conjugate vaccine, DNA vaccine and recombinant vector vaccine. All of these have their respective advantages and disadvantages and hence are used according to in different disease conditions. Regardless of many years of efforts, the failure to develop a vaccine against the worldwide pandemics of HIV needs to search for the new vaccine candidate that can confer protective immunity. Subunit vaccines do not consist of any live pathogenic components, therefore, nullify the chances of pathogenicity reversal as in case of live attenuated vaccine [13,14]. It consists of only the antigenic part of the pathogens which may have the capability to elicit a protective immune response within the human body. The major advantage of using subunit vaccine includes its application to a person with the weak immune system, long-lived immunity and low risk of reaction. Therefore, this study made an effort to design a novel subunit vaccine against HIV infection. This study utilizes the immunogenic epitopes viral proteome to be specific tat, rev, vpu, vpr, vif, nef, protease, integrase and reverse transcriptase. Our designed vaccine consisting of humoral as well as cell-mediated immune response specific immunogenic epitopes. These epitopes were further checked for their conservancy among other HIV subtypes and worldwide population coverage. Subunit vaccine was further evaluated on the antigenicity, allergenicity, and physiochemical parameters. Nextly, the 3D model of subunit vaccine was docked against the TLR-3 receptor and the complex stability was determined by molecular dynamics simulation studies. Vaccine stability was enhanced by disulfide engineering, and *in silico* cloning was performed to ensure the high expression in *E. coli* K12 expression system. Overall, this study applied a combinatorial approach to design an immunogenic, thermostable and non-allergic vaccine candidate to tackle the HIV infection.

2. Methodology

2.1. HIV genome analysis and selection of crucial proteins

A thorough literature survey was done to identify the crucial proteins of the HIV metabolic pathways. Total nine proteins encoded by HIV-1 group M subtype B genome were selected consisting of two regulatory proteins namely Tat (UniProt ID: P04608) and Rev (UniProt ID: P04618); four accessories proteins Vpu (UniProt ID: P05919), Vpr (UniProt ID: P69726), Vif (UniProt ID: P69723), Nef (UniProt ID: P04601), and three enzymes encoded by pol (UniProt ID: P04585) namely Protease (UniProt ID: P04585), Integrase (UniProt ID: P04585) and Reverse transcriptase (UniProt ID: P04585), were selected for the purpose of subunit vaccine development. Sequences of all these proteins were retrieved from UniProt database (http://www.uniprot.org/) in FASTA format.

2.2. Helper T-lymphocytes epitope prediction

Helper T-lymphocyte cell (HTLs) receptor specific epitopes were screened among the above-selected nine HIV-1 group M subtype B proteins by utilizing Immune Epitope Database (IEDB) (http://tools.iedb.org/mhcii/) MHC-II epitope prediction tool [15,16]. In this server, different prediction methods were available for the epitope prediction purpose but we selected the IEDB recommended option to use the best-suited prediction method for the epitope selection. During species selection, the human was selected as target species. In the MHC allele selection panel, IEDB recommended allele option was selected, consisting of a reference panel of 27 alleles with a coverage of >99% population. The selected MHC-II alleles were HLA-DRB1∗01:01, HLA-DRB1∗03:01, HLA-DRB1∗04:01, HLA-DRB1∗04:05, HLA-DRB1∗07:01, HLA-DRB1∗08:02, HLA-DRB1∗09:01, HLA-DRB1∗11:01, HLA-DRB1∗12:01, HLA-DRB1∗13:02, HLA-DRB1∗15:01, HLA-DRB3∗01:01, HLA-DRB3∗02:02, HLA-DRB4∗01:01, HLA-DRB5∗01:01, HLA-DQA1∗05:01/DQB1∗02:01, HLA-DQA1∗05:01/DQB1∗03:01, HLA-DQA1∗03:01/DQB1∗03:02, HLA-DQA1∗04:01/DQB1∗04:02, HLA-DQA1∗01:01/DQB1∗05:01, HLA-DQA1∗01:02/DQB1∗06:02, HLA-DPA1∗02:01/DPB1∗01:01, HLA-DPA1∗01:03/DPB1∗02:01, HLA-DPA1∗01/DPB1∗04:01, HLA-DPA1∗03:01/DPB1∗04:02, HLA-DPA1∗02:01/DPB1∗05:01, and HLA-DPA1∗02:01/DPB1∗14:01. As the majorly HIV-1 affected countries are Sub-Saharan Africa followed by Eastern Europe and Central Asia, North America, Latin America, south and south-East Asia, and Oceania. For the HTL epitope shorting, allelic population coverage was taken into consideration and only one epitope per protein was selected that belongs to the alleles of the aforementioned region which is severely affected by the HIV-1 infection.

2.3. Cytotoxic T lymphocyte epitope prediction

Cytotoxic T lymphocyte cells (CTL) are the key player in MHC-I mediated cellular immune response. They perform their function by killing the cancerous cells, virus-infected cells or other damaged cells by recognizing the epitope presented by MHC-I molecule on the cell surface. CTL epitopes were predicted for the same HIV proteins by utilizing the NetCTL 1.2 server [17] at the sensitivity of 0.80. Among the different parameters of prediction, A2, A3, and B7 were selected as a supertype to cover 83–88.8% person in different ethnic groups [18], weight on C terminal cleavage and tap transport efficiency was kept default at 0.15 and 0.05, respectively. While the threshold score set for the prediction was 0.75.

2.4. T-cell (HTL & CTL) epitope conservancy analysis

The only objective of this study of subunit vaccine development was to provide broader protection against multiple HIV-1 strains using conserved epitopes. Therefore, it was necessary to check the conservancy of CTL and HTL epitopes among different strains of HIV-1 group-M strains. Aforementioned, T-cell epitopes were predicted for the HIV-1 group M subtype B, therefore IEDB Epitope Conservancy Analysis tool [19] was used to calculate the degree of

conservancy of predicted CTL and HTL epitopes among the protein sequences of HIV-1 group M subtype C. Here, we selected B and C subtype because of their prevalence throughout the world. The reference protein set of HIV-1 group M subtype C was obtained from the NCBI (http://tools.iedb.org/ncbi_seq_browser/) consisting of 587 proteins.

2.5. B-cell epitope prediction

The humoral immune response specific epitopes mainly linear B-cell epitopes were predicted for the aforementioned HIV proteins by utilizing the ABCpred server [20]. This server predicts B-cell epitopes using the artificial neural network. The algorithm of this server is based on the recurrent neural network consisting of 700 B-cell epitopes and 700 non B-cell epitopes of maximum 20 residues long for the training and testing purposes. This server has shown an accuracy of 65.93% using the recurrent neural network. Therefore, the FASTA sequence of all HIV proteins was used to predict the linear B-cell epitopes at the threshold of 0.51 and window length of 10.

2.6. Vaccine construction

Subunit vaccine consists of an antigenic component of a pathogenic organism to elicit an immunological response. In order to achieve the same, predicted HTL, CTL and BCL epitopes were combined together in a sequential manner to obtain the final vaccine construct. Subunit vaccine sequence starts with the Adjuvant (β-Defensin peptide) of 45mer length followed by top one HTL epitopes for each HIV protein. Next, top one CTL epitopes corresponding each supertype and BCL epitopes were added in a sequential manner for each HIV protein. EAAAK linkers were used to join the adjuvant and HTL epitopes. GPGPG, AAY, and KK linkers were used to join the HTL and CTL and BCL epitopes, respectively. Aforementioned linkers were used to achieve the effective separation of individual epitopes *in vivo* [21,22].

2.7. Antigenicity and allergenicity prediction

The antigenicity evaluation of final vaccine construct was performed by VaxiJen v2.0 server [23] to predict the probable antigenicity based on an alignment-independent algorithm. It predicts the antigenicity solely based on the physiochemical properties of the input protein sequence without recourse to sequence alignment. The accuracy of this server varies from organism to organism and ranges from 70% to 89%.

Allergenicity of the final vaccine protein was calculated by using AllerTOP v.2.0 [24] server at http://www.pharmfac.net/allertop/. This server is based on the auto cross-covariance (ACC) transformation of protein sequences into uniform equal-length vectors. K-nearest neighbor algorithm was used to classify the protein based on a training set of 2210 known allergen from different species and 2210 non-allergens from the same species.

2.8. Physicochemical characterization of vaccine protein

ProtParam server [25] was used to predict the physicochemical properties of the vaccine protein for the characterization purposes. Predicted properties include a number of amino acids, molecular weight, theoretical pI, extinction coefficient, estimated half-life *in vitro* and *in vivo*, instability index, aliphatic index and Grand average of hydropathicity. ProtPram doesn't require additional information for the prediction purpose except for the primary amino acids sequence.

2.9. Secondary and tertiary structure prediction

The secondary structure of the predicted vaccine was carried out by using PSIPRED v3.3 analysis workbench [26] and the percentage of the helix, strand and coils were determined. Similarly, the tertiary structure of vaccine protein was predicted using the RaptorX structure prediction module [27,28]. RaptorX is a template based, fully automated tertiary structure prediction server. The protein sequence of the subunit vaccine was used in the FASTA format to perform the same. RaptorX having the capability to predict high-quality tertiary structure even remotely related template sequence has experimentally solved the structure. It uses a novel nonlinear context-specific alignment potential and probabilistic consistency algorithm.

2.10. Tertiary structure refinement and validation

The 3D modeled structure of final vaccine construct obtained from RaptorX server depends on the degree of similarity between the target protein and available template structure options. Therefore, the model structure was improved beyond the accuracy by using the GalaxyRefine server [29,30]. The refinement method used by the GalaxyRefine server has been successfully tested in CASP10 server. As per the CASP10 report, it is one of the best-performing methods to improve the local structure quality. GalaxyRefine first rebuilds the side chain followed by side chain packaging and overall relaxation by molecular dynamics simulation. It can have the ability to improve both local and global quality of model structure predicted by state-of-the-art protein tertiary structure prediction servers.

2.11. Vaccine protein disulfide engineering

Enhancing the stability of model vaccine was an essential objective that was completed by applying a logical approach to disulfide bond formation. Disulfide engineering was performed to design novel disulfide bonds in the vaccine protein construct. Disulfide bonds are known to provide significant stability to protein structure along with well-defined geometric conformation. Disulfide by Design-2 (DbD2) is an online program, which was used to design the same for the model vaccine [31]. DbD2 algorithm is based on the identification of residues pairs that are probably going to form a disulfide bond if the individual amino acids are mutated to cysteine. It results in the list of residue pairs which have appropriate geometry and capability to form disulfide bonds.

2.12. Protein-protein docking

Molecular docking is a computational approach to check the binding affinity between the ligand and receptor molecule [32,33]. Immune receptors mainly TLR-3 present over the immune cells has been implicated in the HIV-1 induced inflammation and AIDS. Literature survey reveals that the HIV-1 mediated activation of dendritic cells is TLR-3 dependent. In order to evaluate the binding affinity of designed subunit vaccine with the TLR-3 immune receptor, protein-protein docking was performed by using ClusPro 2.0 [34] and PatchDock server [35]. Both these servers are fully automated online web server commonly used for the protein-protein docking. Only the PDB file of vaccine protein and TLR-3 immune receptor was uploaded to the ClusPro and PatchDock web interface to obtain those complexes having good electrostatic interaction and desolvation free energy. The output solutions of PatchDock were subjected to the FireDock server for the refinement purposes and global energy of top 10 solutions was determined.

Table 1
Predicted HTL epitopes among HIV proteins to be a part of multi-epitope subunit vaccine.

Protein name	Allele	Epitope	Method	Percentile rank	IC_{50}
TAT	HLA-DRB5*01:01	CFITKALGISYGRKK	Consensus (smm/nn)	0.55	46
NEF	HLA-DRB3*01:01	REVLEWRFDSRLAFH	Consensus (smm/nn)	0.28	20
VIF	HLA-DRB1*01:01	KVGSLQYLALAALIT	Consensus (smm/nn)	0.04	2
REV	HLA-DRB1*11:01	EELIRTVRLIKLLYQ	Consensus (smm/nn)	0.02	37
VPR	HLA-DRB4*01:01	AIIRILQQLLFIHFR	Consensus (smm/nn)	0.01	21
VPU	HLA-DRB5*01:01	IIEYRKILRQRKIDR	Consensus (smm/nn)	0.09	25
Protease	HLA-DPA1*02:01/ DPB1*01:01	FIKVRQYDQILIEIC	Consensus (smm/nn)	0.74	242
Integrase	HLA-DRB1*11:01	NSDIKVVPRRKAKII	Consensus (smm/nn)	0.15	5
Reverse Transcriptase	HLA-DRB4*01:01	SESELVNQIIEQLIK	Consensus (smm/nn)	0.04	15

Table 2
Region wise population coverage and presence of selected alleles corresponding to top scoring HTL-epitopes.

S. No.	Region	Alleles consider for HTL epitope screening
1	Australia	HLA-DRB1*01:01 DPB1*01:01
2	Europe	HLA-DRB3*01:01 HLA-DRB1*11:01 HLA-DPA1*02:01 DPB1*01:01 HLA-DRB1*01:01
3	North Africa	HLA-DRB1*01:01 HLA-DRB1*11:01 DPB1*01:01
4	North America	HLA-DRB5*01:01 HLA-DRB3*01:01 HLA-DRB4*01:01 HLA-DRB1*11:01 HLA-DPA1*02:01 DPB1*01:01 HLA-DRB1*01:01
5	N. E. Asia	HLA-DRB1*01:01 HLA-DRB1*11:01 HLA-DPA1*02:01 DPB1*01:01
6	Oceania	HLA-DRB1*01:01 HLA-DRB1*11:01 HLA-DPA1*02:01
7	S. C. America	HLA-DRB1*01:01 HLA-DRB1*11:01 HLA-DPA1*02:01
8	South Asia	HLA-DRB1*01:01 HLA-DRB1*11:01 HLA-DPA1*02:01 DPB1*01:01
9	S. E. Asia	HLA-DRB1*01:01 HLA-DRB1*11:01 HLA-DPA1*02:01 DPB1*01:01
10	S.S. Africa	HLA-DRB1*01:01 HLA-DRB1*11:01
11		HLA-DPA1*02:01 DPB1*01:01
12	W. Asia	HLA-DRB1*01:01 HLA-DRB1*11:01 HLA-DPA1*02:01 DPB1*01:01

2.13. Molecular dynamics simulation

Molecular dynamics simulation is a unique approach commonly used to check the stability of protein-ligand complex [30,33]. Here the complex was formed of vaccine protein and TLR-3 immune receptor, therefore, their complex stability was performed by using GROningen MAchine for Chemical Simulations (GROMACS) molecular dynamics package. The gromos96 43a1 force field was used to generate the topology and determine the intermolecular interactions for the whole duration of molecular dynamics simulation process. Rhombic dodecahedron boundary box was used to covers the protein complex followed by the addition of water and ion using SPC water model. Energy minimization of the complex was also performed to overcome the stearic clashes and bad contacts. Further, the system moves through NVT and NPT ensemble simulation for the time period of 100 ps. Finally, molecular dynamics simulation was performed for a time duration of 10 ns and respective RMSD and RMSF was determined for the protein backbone and side chains, respectively.

2.14. Codon adaptation and in silico cloning

Codon adaptation is a unique algorithm to adapt the codon usage to the most sequenced prokaryotic organism which may lead to the higher expression rate. This approach was applied to achieve the higher expression rate of the vaccine protein in the *E. coli* K12 host because the codon usage of human and selected host differs from each other. In the options section, we checked to avoid rho-independent transcription termination, prokaryote ribosome binding site and cleavage site of restriction enzymes. The optimized codon sequence was further checked for the presence of commercially available restriction enzyme cutting site and found that none of the restriction sites (mainly XhoI and BamH1) are present. Therefore, the optimized sequence of vaccine protein was reversed and added with XhoI and BamHI restriction site at the N- and C-terminal sites, respectively, to ensure the restriction cloning. SnapGene restriction cloning module was used to insert the adapted sequence into pET28a(+) vector between the XhoI (158) and BamHI (198).

3. Results and discussion

3.1. Sequence retrieval for vaccine designing

Amino acid sequences of 9 HIV proteins namely Tat, Rev, Vpu, VPR, Vif, Nef, Protease, Integrase and reverse transcriptase, were retrieved from the UniProt Database. All these proteins were retrieved in the FASTA format and basically, they belong to the three protein categories namely regulatory proteins, accessories proteins and pol encoded enzymes. All these proteins have shown to play a crucial role in the HIV metabolic pathways (details available in Section 1). These sequences were further utilized for the HTL, CTL, and BCL epitope prediction.

3.2. HTL epitope prediction

Helper T-lymphocyte cells are the key player of an adaptive limb of the immune response [36]. It leads to the whole body massive immune response by coordination signaling to activate cytotoxic T-cell, macrophages, and B-cell mediated immune response. A variable number of epitopes were predicted using the reference allele sets covering >99% of the population for all input protein sequences as described elsewhere [14,37]. The epitope length was found to be 15mer with variable percentile rank and IC50 values [37]. Top one epitope per input sequence (9 proteins) was

Table 3
Predicted CTL epitopes among HIV proteins to construct multi-epitope subunit vaccine.

S. No.	HIV-1 protein	Accession No	CTL epitopes predicted using NetCTL server		
			A2 supertype	A3 supertype	B7 Supertype
1	Protein TAT	sp_P04608_1	–	QTHQASLSK	–
2	Protein NEF	sp_P04601_2	AAVDLSHFL	SVIGWPTVR	RVGAASRDL
3	Protein VIF	sp_P69723_1	SLVKHHMYV	ALAALITPK	TPKKIKPPL
4	Protein REV	sp_P04618_1	YLGRSAEPV	LIRTVRLIK	LPPLERLTL
5	Protein VPR	sp_P69726_1	QLLFIHFRI	FIHFRIGCR	FPRIWLHGL
6	Protein VPU	sp_P05919_1	ALVEMGVEM	RLIDRLIER	IPIVAIVAL
7	PROTEASE	sp_P04585_48	LLTQIGCTL	ILIEICGHK	KVRQYDQIL
8	INTEGRASE	sp_P04585_11	KQMAGDDCV	KLAGRWPVK	VPRRKAKII
9	Reverse Transcriptase	Sp_P04585	YQYMDDLYV	AIFQSSMTK	SPAIFQSSM

Table 4
Epitope conservancy analysis of HIV-1 group M subtype B protein based CTL and HTL epitopes against HIV-1 group M subtype C.

Epitope	Epitope name	Epitope sequence	Epitope length	Maximum identity
1	Epitope_HTL1	CFITKALGISYGRKK	15	86.67%
2	Epitope_HTL2	REVLEWRFDSRLAFH	15	66.67%
3	Epitope_HTL3	KVGSLQYLALAALIT	15	86.67%
4	Epitope_HTL4	EELIRTVRLIKLLYQ	15	66.67%
5	Epitope_HTL5	AIIRILQQLLFIHFR	15	86.67%
6	Epitope_HTL6	IIEYRKILRQRKIDR	15	80.00%
7	Epitope_HTL7	FIKVRQYDQILIEIC	15	100.00%
8	Epitope_HTL8	NSDIKVVPRRKAKII	15	100.00%
9	Epitope_HTL9	SESELVNQIIEQLIK	15	100.00%
10	Epitope_CTL1	QTHQASLSK	9	44.44%
11	Epitope_CTL2	AAVDLSHFL	9	77.78%
12	Epitope_CTL3	SVIGWPTVR	9	77.78%
13	Epitope_CTL4	RVGAASRDL	9	88.89%
14	Epitope_CTL5	SLVKHHMYV	9	100.00%
15	Epitope_CTL6	ALAALITPK	9	77.78%
16	Epitope_CTL7	TPKKIKPPL	9	88.89%
17	Epitope_CTL8	YLGRSAEPV	9	77.78%
18	Epitope_CTL9	LIRTVRLIK	9	55.56%
19	Epitope_CTL10	LPPLERLTL	9	88.89%
20	Epitope_CTL11	QLLFIHFRI	9	88.89%
21	Epitope_CTL12	FIHFRIGCR	9	88.89%
22	Epitope_CTL13	FPRIWLHGL	9	88.89%
23	Epitope_CTL14	ALVEMGVEM	9	66.67%
24	Epitope_CTL15	RLIDRLIER	9	77.78%
25	Epitope_CTL16	IPIVAIVAL	9	44.44%
26	Epitope_CTL17	LLTQIGCTL	9	100.00%
27	Epitope_CTL18	ILIEICGHK	9	88.89%
28	Epitope_CTL19	KVRQYDQIL	9	100.00%
29	Epitope_CTL20	KQMAGDDCV	9	88.89%
30	Epitope_CTL21	KLAGRWPVK	9	100.00%
31	Epitope_CTL22	AIFQSSMTK	9	100.00%
32	Epitope_CTL23	YQYMDDLYV	9	100.00%
33	Epitope_CTL24	AIFQSSMTK	9	100.00%
34	Epitope_CTL25	SPAIFQSSM	9	100.00%

Table 5
Predicted BCL epitopes among HIV proteins to make multi-epitope subunit vaccine.

S. No.	Protein	Sequence	Position	Score
1	TAT	SYGRKKRRQR	49	0.8
2	NEF	QVPLRPMTYK	75	0.85
3	VIF	YRHHYESPHP	43	0.77
4	REV	LYQSNPPPNP	25	0.83
5	VPR	MEQAPEDQGP	4	0.79
6	VPU	LRQRKIDRLI	36	0.77
7	Protease	LEEMSLPGRW	36	0.71
8	Integrase	VVAKEIVASC	34	0.83
9	Reverse transcriptase	LDVGDAYFSV	112	0.84

shorted to be a part of subunit vaccine construct (Table 1). During epitope shorting, the population coverage of alleles was defined on the basis of top one epitope for each antigenic protein considered for vaccine design. Therefore, total nine HTL epitopes were selected based on their allelic population coverage of the Sub Saharan Africa, North Africa, Europe, North America, South Central America, Oceania, Asia and Australia. As per the epitope selection data, the allelic distribution and population coverage of selected alleles namely HLA-DRB5∗01:01, HLA-DRB3∗01:01, HLA-DRB1∗01:01, HLA-DRB1∗11:01, HLA-DRB4∗01:01, HLA-DRB5∗01:01, HLA-

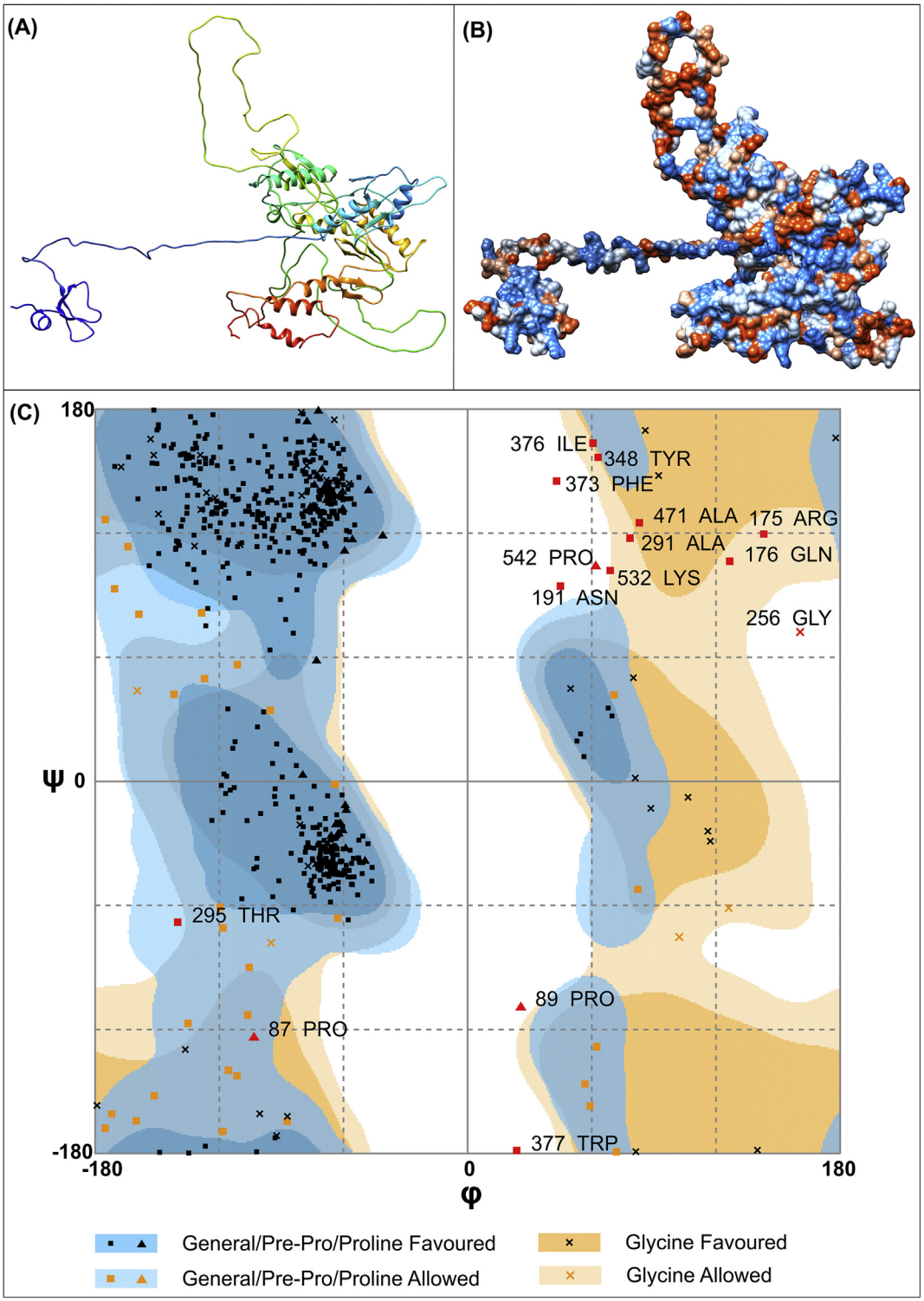

Fig. 1. Tertiary structure prediction, refinement and validation (A-B) Showing 3D structure of refined vaccine model (C) which was validated using RamPage and found that 92.2% residues are in the favored region, 5.4% residues in the allowed region and only 2.4% residues in an outlier.

DPA1∗02:01, DPB1∗01:01, HLA-DRB1∗11:01, HLA-DRB4∗01:01 is reported in Table 2. All nine epitopes having the lowest percentile rank ranging from 0.02 to 0.74 reflecting their good binding affinity for the MHC-II receptor. Out of these nine epitopes, eight epitopes having an IC50 value below 50 showing high affinity while the IC50 value obtained for the protease HTL epitope (FIKVRQYDQILIEIC) was 242 showing an intermediate affinity for the respective epitopes [38]. Therefore, total 9 HTL epitopes with good binding affinity and epidemic population coverage were used for vaccine designing.

3.3. CTL epitope prediction

HIV-specific cytotoxic T-lymphocytes have shown to play a crucial role in controlling the viral infection during asymptomatic HIV infection phase [39]. Total 43, 57 and 31 number of CTL epitope

was predicted respectively to A2-, A3- and B7-supertype for all 9 HIV proteins using NetCTL server [37,38]. In case of TAT protein, none of the CTL epitopes were determined for A2- and B7-supertype while epitopes for only A3-supertype was obtained. All output epitope sequences were obtained with a combined score of greater than 0.75, where a higher score represents good binder [40,41]. Top one high scoring CTL epitopes for each HIV protein respective to each MHC-I supertype were picked and utilized for the vaccine designing. Therefore, total 25 CTL epitopes for 9 HIV proteins were selected to become a part of subunit vaccine (Table 3).

3.4. T-cell (HTL & CTL) epitope conservancy analysis

This study was designed to develop a conserved multi-epitopes subunit vaccine that has the capability to elicit immunogenicity against HIV-1 group M subtype B as well as HIV-1 group M subtype C. The epitope conservancy analysis shows that most of the CTL and HTL epitopes belongs to subtype-B having 80–100% identity with the HIV-1 group M subtype-C, While few epitopes have shown $\leq$80% identity. This result concludes that the designed vaccine is immunogenic against the subtype B as well as subtype C (Table 4).

3.5. B-cell epitope prediction

B-cells contributes towards the humoral limb of immune response and responsible for the clearance of blood pathogen in an antibody-dependent manner [42]. Therefore, it was necessary for us to determine the presence of BCL epitopes among the HIV protein sequences. BCPRED server helped us to determine the presence of humoral immune response specific B-cell epitopes in the input viral sequences [14,36]. Total 160 BCL epitopes were obtained having a variable score, representing their binding affinity to the B-cell receptor. An epitope with high score represents the good binder and vice versa. Therefore, total 9 BCL epitopes, one epitope each viral protein with a high score, were utilized for the vaccine designing (Table 5).

3.6. Vaccine construction

Recently, many research works have been reported to be based on subunit vaccine design using multiple, HTL, CTL and B-cell epitopes [14,36,37,40,41]. This study also develops a multi-epitope subunit vaccine of 635 amino acid residues which was constructed to make a contribution towards the path of HIV-1 vaccine development. Out of 635 residues, 47 residues were occupied by the adjuvant sequence, while remaining sequence was occupied by the HTL, CTK, and BCL epitopes and their respective linkers (EAAK, AAY, GPGPG, and KK). The designed vaccine consists of 09 HTL epitope sequences, 25 CTL epitope and 09 BCL epitopes where HTL, CTL, and BCL epitopes were joined together by GPGPG, AAY and KK linkers. The specialty of the designed vaccine is its component; it starts with the adjuvant (β-Defensin peptide) which is necessary for the enhanced immune response followed by HTL CTL and BCL epitopes separated by the suitable linker to ensure their effective separation within the body for maximal immunity.

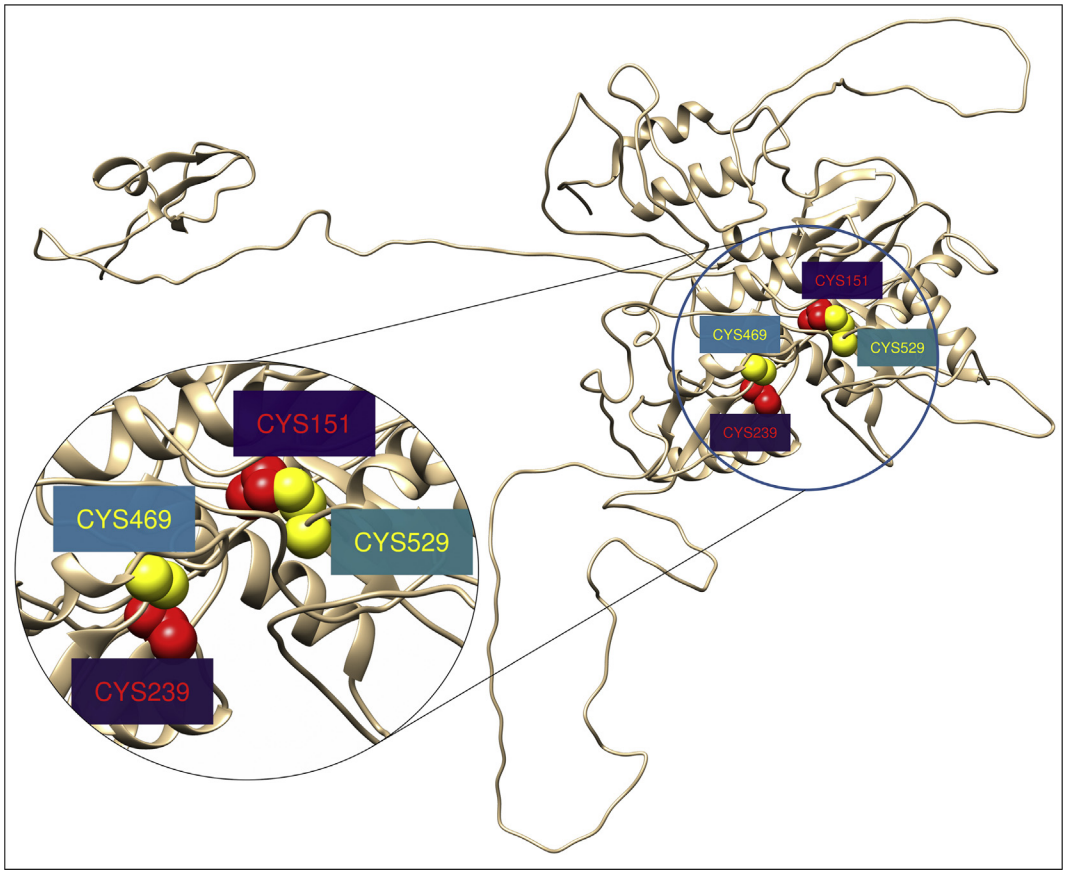

Fig. 2. Disulfide engineering was performed and to determine amino acid residues whose replacement with cysteine will lead to the di-sulfide bond formation and ultimately increasing the protein stability. Total 4 pairs of residues were identified to have the ability to form disulfide bond.

3.7. Antigenicity and allergenicity prediction

Antigenicity is the major parameter which determines the success of any vaccine candidate. A vaccine candidate must be immunogenic to activate the humoral and cell-mediated immune cells which will result in the memory cell formation against the target organism. As per the result of VaxiJen server, designed vaccine candidate was found to be antigenic with the probable antigenicity score of 0.45 at the threshold value capping of 0.4. The resulting output of AllerTOP for the allergenicity determination explored the non-allergenic behavior of the vaccine protein. Both these result output has shown the immunogenic and non-allergenic behavior of the vaccine protein.

3.8. Physicochemical characterization of vaccine protein

Vaccine protein was characterized by the physical and chemical parameters by using the ProtParam server. The molecular weight of vaccine protein was found to be 70.8 kDa which is good to support the antigenic nature of vaccine construct. The theoretical pI was 10.01 showing that the protein will have net positive charge below the pI and vice versa. It will be helpful in the separation of designed vaccine using isoelectric focusing or chromatographic techniques. The extinction coefficient value was found to be 93,060 at 0.1% absorption and assuming that all cysteine residues are reduced. The estimated half-life of designed vaccine construct was predicted to be 30 h in mammalian reticulocytes, *in vitro* while more than 20 and 10 h in yeast and *E. coli*, *in vivo*, respectively. Vaccine protein was found to be stable at the instability score of 39.96. The aliphatic index score was 94.93 showing the thermostable nature of vaccine protein while Grand average of hydropathicity (GRAVY) value (−0.277) representing the hydrophilic nature.

3.9. Secondary and tertiary structure prediction

The secondary structure of the vaccine protein was obtained from the PSIPRED server and confirms the presence of 15.43% helix, 75.43% coil and only 9.14% of strands. The tertiary structure was predicted as 3 domain structure using the RaptorX web server. The best-suited template used for the homology modeling was *Novel indazole nnrtis* (PDB id: 2jle). All 635 amino acid residues were modeled with the 5% of the residues in the disordered region. P-value parameter defines the quality of the model. Its low value represents good model quality and the P-value of our model was 4.18e−03, showing the stable nature and good quality of modeled structure (Fig. 1A-B).

3.10. Tertiary structure refinement and validation

To improve the quality of predicted 3D modeled structure beyond the accuracy, refinement was performed using GalaxyRefine server. When the refined model was evaluated on the Ramachandran plot, we found that before refinement 88.2% of residues were in the favored region while after refinement 92.2% residues were in the favored region, 5.4% residues in the allowed region while only 2.4% residues were in the outlier region (Fig. 1C).

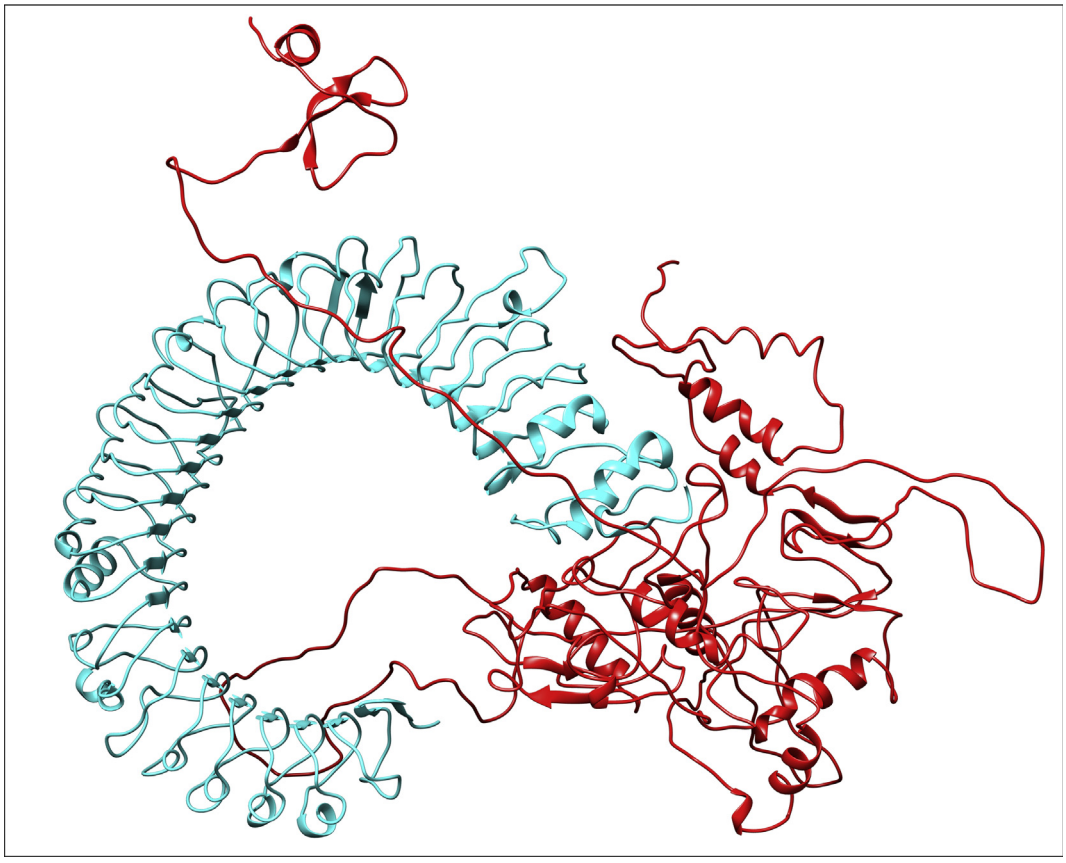

Fig. 3. Protein-protein docking between vaccine protein and TLR-3 using ClusPro where vaccine protein has shown in red color while TLR-3 has shown in the Cyan color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.11. Vaccine protein disulfide engineering

Disulfide engineering was performed to mutate the residues present in high mobility region with cysteine leads to the increase in vaccine stability. Total 69 residue pairs were determined to have the ability of disulfide bond formation. But when all these residue pairs were evaluated on the energy, chi3 and B-factor parameter only 2 pairs of residue satisfied the disulfide bond formation. Those residue pairs were Ile151-Tyr529, and Ala239-Lys496 (Fig. 2). All these 4 residues were mutated to the cysteine residue. The energy value considered for the residue screening was less than 2.2 while the value of chi3 was in between −87 to +97.

3.12. Protein-protein docking

Protein-protein docking was performed using ClusPro and PatchDock web servers to check the binding affinity of designed vaccine construct with the TLR3 receptor. ClusPro results in total 30 protein-ligand complexes as output along with respective lowest energy score (Fig. 3). The lowest energy score of −1553.7 was obtained for the complex namely model.000.01. While, the FireDock output refinement of PatchDock solutions has shown the lowest global energy of −14.7 for the solution number 7. The lowest energy score of these complexes showing the highest binding affinity between TLR-3 and vaccine protein structure.

3.13. Molecular dynamics simulation

The stability of the ClusPro output complex (model.000.01) was determined by performing molecular dynamics simulation study. The whole process of dynamics was performed at ∼300 K temperature and ∼1 bar pressure (Fig. 4A-B). The RMSD value predicted for complex protein backbone was found to be 1.0 nm, while the RMSF value obtained for the side chain was 0.3 nm. Both these value showing the stable nature of the protein-TLR3 complex. The RMSD result showing that initial deviation starts with 0.5 nm and gradually increases up to 1.0 nm where an average deviation was found to be 1.0 nm while the RMSF obtained for the protein side chain was ∼0.3 nm (Fig. 4C-D). Both these results grant the stability of the protein-ligand complex.

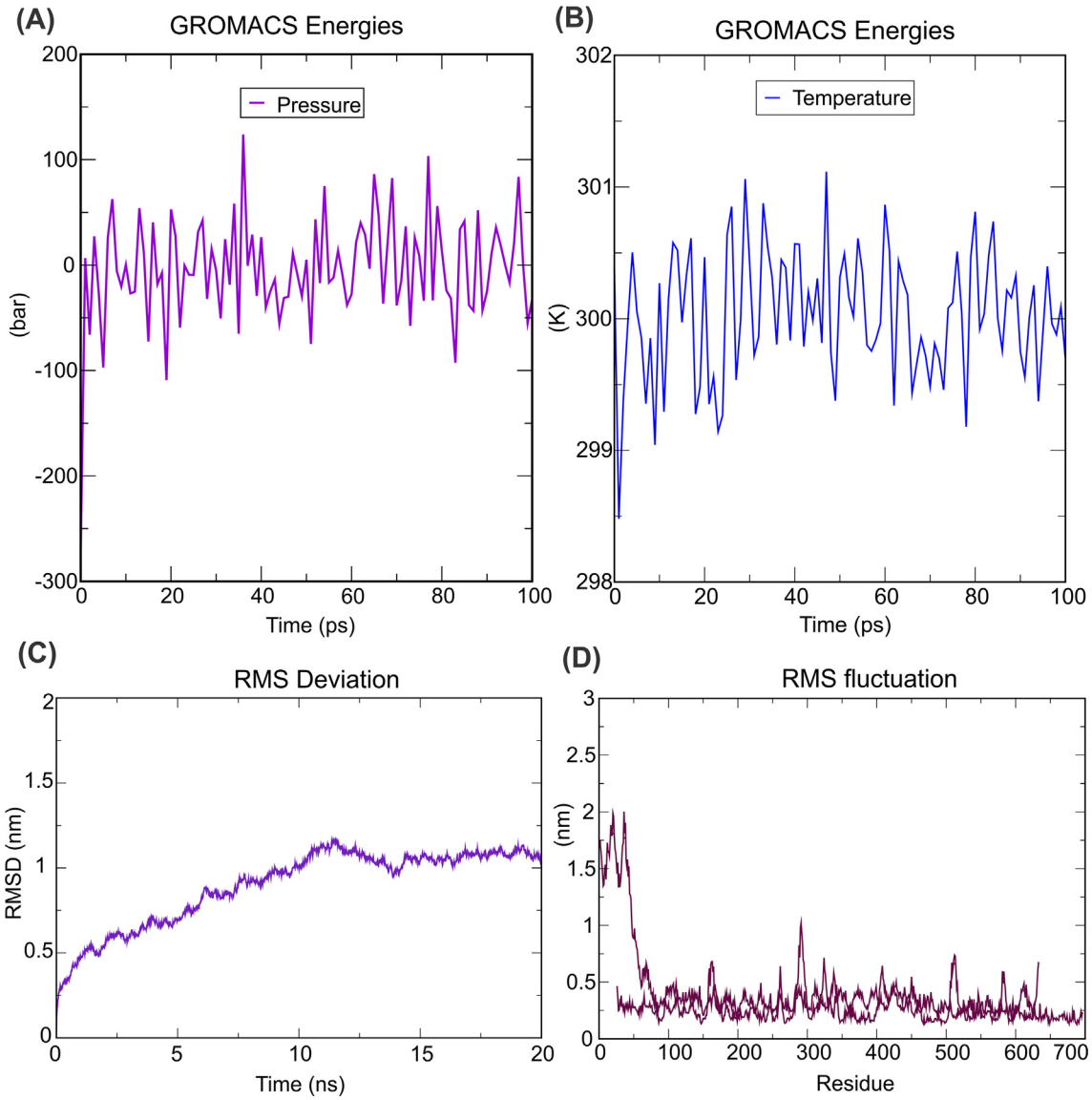

Fig. 4. Molecular dynamics simulation (A-B) showing the temperature and pressure fluctuation for the time duration of 100Ps (C) RMSD plot obtained for the vaccine protein backbone after making complex with the TLR-3 as receptor (B) showing the fluctuation of side chain residue of vaccine protein in complex form with TLR-3.

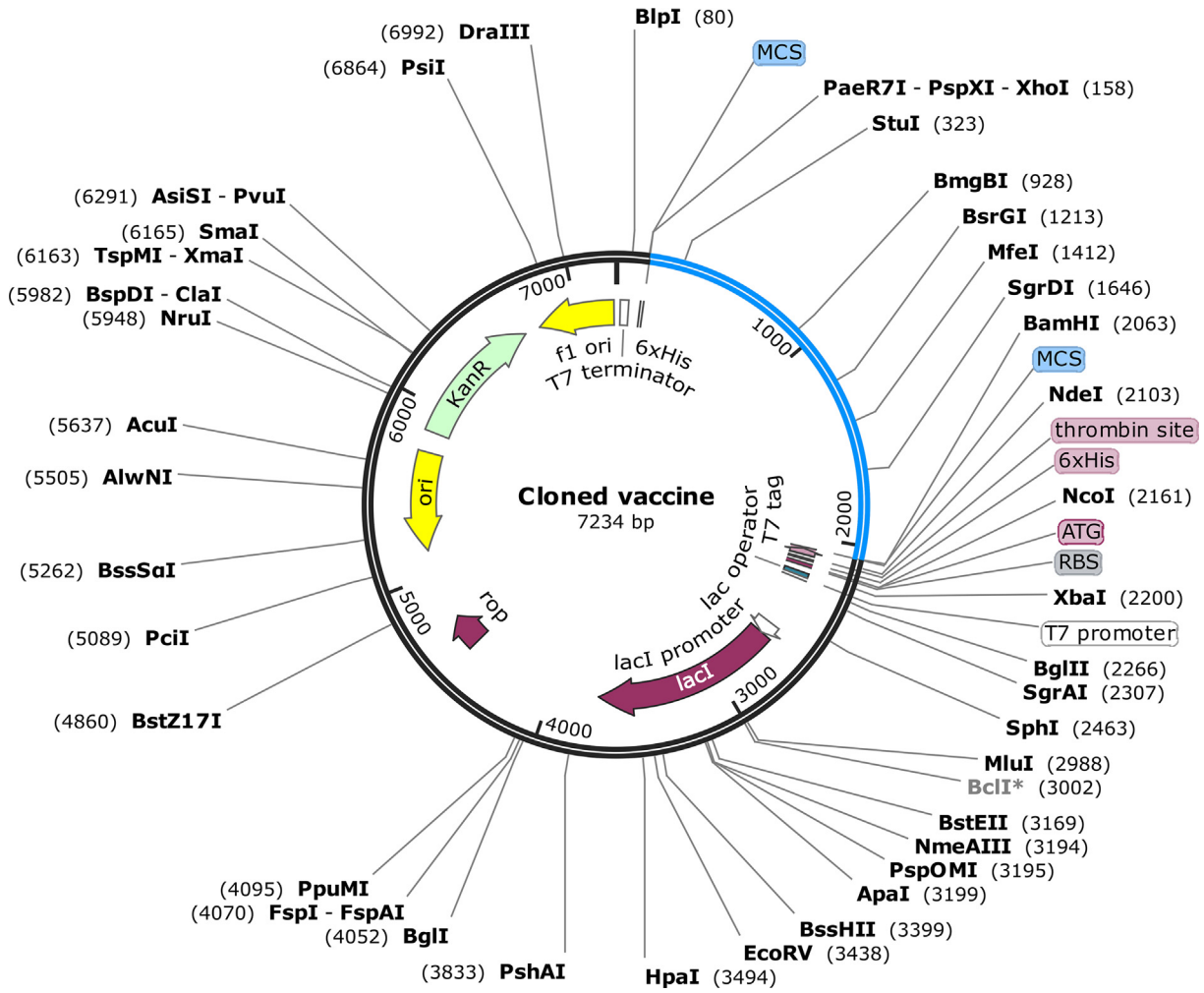

Fig. 5. *In silico* cloning of optimized codons encoding vaccine protein into pET28a(+) vector to ensure the vaccine protein expression in microbial system.

3.14. Codon adaptation and in silico cloning

The expression system of the human and *E. coli* K12 strains differ from each other, needs the adaptation of codon as per the host expression system. Therefore, the vaccine protein was reverse transcribed and adapted as per the *E. coli* K12. The codon adaptation index of optimized codon was found to be 0.986 showing the higher proportion of most abundant codon. The GC content of adapted codons was 51.97 which is also satisfactory. Also, there was an absence of XhoI and BamHI restriction site in the same. So, it was safe for the cloning purposes. Later on, the adapted codon along with XhoI and BamHI restriction sites were inserted into pET28a(+) vector and a clone of 7234 base pair was obtained. The length of the desired sequence was 1911 base pairs while the remaining belongs to the vector. The desired region has shown in purple color in between the pET28a(+) vector sequence (Fig. 5).

4. Conclusion

The human immunodeficiency virus has become a major health concern in the developing countries that attack the body immune system leads to the reduced number of immune cell (CD4 T-cell). Such reduction in CD4 T-cell population leads to the immunocompromised condition and make the body prone to the infectious agents. Among the two HIV strains, HIV-1 is the most severe leads to the global spread of AIDS. Along with this severe situation, treatment only relies on the antiretroviral therapy and none of the vaccines has been reported. Therefore, this study provides an add-on to the path of vaccine development against HIV-1. We retrieved the major proteins of HIV metabolic pathway in FASTA format and most antigenic epitopes suitable for the HTL, CTL and BCL receptor were screened and along with suitable adjuvants and linkers used for the subunit vaccine construction. Further, the designed vaccine go through the characterization on the allergenicity, antigenicity, physiochemical characterization, 2D and 3D structure prediction, di-sulfide engineering and *in silico* cloning parameters. At last, molecular docking and dynamics study was performed to check the binding affinity and complex stability of vaccine construct and TLR-3 receptor. Moreover, this study needs the experimental validation of the vaccine protein to confirm this study.

Author contributions

Protocol designed by RKP, VKP.
Methodology performed by RKP, RO, VKP, VSA.
Manuscript was written by RKP, VKP, MK.

Acknowledgment

RKP is thankful to Department of Science and Technology for providing INSPIRE fellowship. VKP is thankful to the Central University

of Rajasthan for providing the computational facility. VSA is thankful to Lady Tata Memorial Trust for providing fellowship. MK is thankful to Department of Biotechnology, Department of Science and Technology – PURSE scheme and University Information Centre, Bharathidasan University for providing the high-performance computational facility.

Competing financial interests

The authors have declared no competing interest.

References

[1] Cluck DB, Underwood RF. A therapeutic perspective of living with human immunodeficiency virus/AIDS in 2017. Nurs Clin North Am 2017. https://doi.org/10.1016/j.cnur.2017.10.009.
[2] Miedzinski LJ. Early clinical signs and symptoms of HIV infection: delaying progression to AIDS. Can Fam Physician 1992;38:1401–10.
[3] Janeway Jr CA, Travers P, Walport Mea. Immunobiology: the immune system in health and disease. 5th ed. New York: Garland Science; 2001 [Acquired immune deficiency syndrome]. <https://www.ncbi.nlm.nih.gov/books/NBK27126/>.
[4] Hemelaar J. The origin and diversity of the HIV-1 pandemic. Trends Mol Med 2012;18:182–92.
[5] Kirchhoff F. HIV life cycle: overview. Encyclopedia of AIDS. New York, NY: Springer; 2013. p. 1–9.
[6] Chander S, Pandey R, Penta A, Choudhary B, Sharma M, Malik R, et al. Molecular docking and molecular dynamics simulation based approach to explore the dual inhibitor against HIV-1 reverse transcriptase and Integrase. Comb Chem High Throughput Screen 2017;20:734–46.
[7] Taniguchi I, Mabuchi N, Ohno M. HIV-1 Rev protein specifies the viral RNA export pathway by suppressing TAP/NXF1 recruitment. Nucl Acids Res 2014;42:6645–58.
[8] Veschambre P, Simard P, Jalinot P. Evidence for functional interaction between the HIV-1 Tat transactivator and the TATA box binding protein in vivo. J Mol Biol 1995;250:169–80.
[9] Fackler OT, Baur AS. Live and let die: Nef functions beyond HIV replication. Immunity 2002;16:493–7.
[10] Mercenne G, Bernacchi S, Richer D, Bec G, Henriet S, Paillart J-C, et al. HIV-1 Vif binds to APOBEC3G mRNA and inhibits its translation. Nucl Acids Res 2009;38:633–46.
[11] Goh WC, Rogel ME, Kinsey CM, Michael SF, Fultz PN, Nowak MA, et al. HIV-1 Vpr increases viral expression by manipulation of the cell cycle: a mechanism for selection of Vpr in vivo. Nat Med 1998;4:65–71.
[12] Farnet CM, Haseltine WA. Integration of human immunodeficiency virus type 1 DNA in vitro. Proc Natl Acad Sci USA 1990;87:4164–8.
[13] Ahluwalia PK, Pandey RK, Sehajpal PK, Prajapati VK. Perturbed microRNA expression by mycobacterium tuberculosis promotes macrophage polarization leading to pro-survival foam cell. Front Immunol 2017;8:107.
[14] Pandey RK, Bhatt TK, Prajapati VK. Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating anopheles salivary protein. Sci Rep 2018;8:1125.
[15] Wang P, Sidney J, Dow C, Mothe B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 2008;4:e1000048.
[16] Pandey RK, Prajapati P, Goyal S, Grover A, Prajapati VK. Molecular modeling and virtual screening approach to discover potential antileishmanial inhibitors against ornithine decarboxylase. Comb Chem High Throughput Screen 2016;19:813–23.
[17] Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinform 2007;8:424.
[18] Lund O, Nielsen M, Kesmir C, Petersen AG, Lundegaard C, Worning P, et al. Definition of supertypes for HLA molecules using clustering of specificity matrices. Immunogenetics 2004;55:797–810.
[19] Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinform 2007;8:361.
[20] Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 2006;65:40–8.
[21] Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatabadi SS, Ghasemi Y. Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against Staphylococcus aureus. Infect Genet Evol 2017;48:83–94.
[22] Pandey RK, Sundar S, Prajapati VK. Differential expression of miRNA regulates T cell differentiation and plasticity during visceral leishmaniasis infection. Front Microbiol 2016;7:206.
[23] Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinform 2007;8:4.
[24] Dimitrov I, Flower DR, Doytchinova I. AllerTOP – a server for in silico prediction of allergens. BMC Bioinform 2013;14:S4.
[25] Gasteiger E, Hoogland C, Gattiker A, Duvaud SE, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy server. Springer; 2005.
[26] McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics 2000;16:404–5.
[27] Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, et al. Template-based protein structure modeling using the RaptorX web server. Nat Protoc 2012;7:1511–22.
[28] Pandey RK, Verma P, Sharma D, Bhatt TK, Sundar S, Prajapati VK. High-throughput virtual screening and quantum mechanics approach to develop imipramine analogues as leads against trypanothione reductase of leishmania. Biomed Pharmacother 2016;83:141–52.
[29] Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. Nucl Acids Res 2013;41:W384–8.
[30] Pandey RK, Narula A, Naskar M, Srivastava S, Verma P, Malik R, et al. Exploring dual inhibitory role of febrifugine analogues against Plasmodium utilizing structure-based virtual screening and molecular dynamic simulation. J Biomol Struct Dyn 2017;35:791–804.
[31] Craig DB, Dombkowski AA. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. BMC Bioinform 2013;14:346.
[32] Pandey RK, Kumbhar BV, Srivastava S, Malik R, Sundar S, Kunwar A, et al. Febrifugine analogues as *Leishmania donovani* trypanothione reductase inhibitors: binding energy analysis assisted by molecular docking, ADMET and molecular dynamics simulation. J Biomol Struct Dyn 2017;35:141–58.
[33] Pandey RK, Kumbhar BV, Sundar S, Kunwar A, Prajapati VK. Structure-based virtual screening, molecular docking, ADMET and molecular simulations to develop benzoxaborole analogs as potential inhibitor against *Leishmania donovani* trypanothione reductase. J Recept Signal Transduct 2017;37:60–70.
[34] Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for protein-protein docking. Nucl Acids Res 2004;32:W96–9.
[35] Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucl Acids Res 2005;33:W363–7.
[36] Pandey RK, Prajapati VK. Exploring sand fly salivary proteins to design multi-epitope subunit vaccine to fight against visceral leishmaniasis. J Cell Biochem 2018. https://doi.org/10.1002/jcb.26719.
[37] Mehla K, Ramana J. Identification of epitope-based peptide vaccine candidates against enterotoxigenic *Escherichia coli*: a comparative genomics and immunoinformatics approach. Mol BioSyst 2016;12:890–901.
[38] 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. Infect Genet Evol 2017;56:75–91.
[39] Walker B, McMichael A. The T-cell response to HIV. Cold Spring Harb Perspect Med 2012;2:a007054.
[40] Ali M, Pandey RK, Khatoon N, Narula A, Mishra A. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. Sci Rep 2017;7:9232.
[41] Khatoon N, Pandey RK, Prajapati VK. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. Sci Rep 2017;7:8285.
[42] Moir S, Fauci AS. B cells in HIV infection and disease. Nat Rev Immunol 2009;9:235–45.