



In silico designing of peptide based vaccine for Hepatitis viruses using reverse vaccinology approach

Dwaipayan Chaudhuri, Joyeeta Datta, Satyabrata Majumder, Kalyan Giri*

Department of Life Sciences, Presidency University, Kolkata, India



ARTICLE INFO

Keywords:
Hepatitis virus
Peptide based vaccine
Reverse vaccinology
MD simulation
Immune-simulation
in silico cloning

ABSTRACT

Five different Hepatitis virus from different viral species cause viral-hepatitis, which is a life threatening disease leading to a high number of loss of lives every year. The mode of infection and transmission is different for each species and mostly spreads by direct contact and body fluids (for HBV and HCV). No such vaccine is available that can cure all types of Hepatitis with cross-protection. Thus our study involves a peptide based vaccine design with the help of Immunoinformatics approach. We focused only on the secretory and extracellular proteins of each types and identified their epitopes. Epitopes were examined for antigenicity, allergenicity, toxicity, anti-inflammatory property and IFN- γ induction. The short-listed peptides were stitched using linkers and TLR4 adjuvant. This final vaccine was proven to have good physico-chemical and structural properties. Simulation study to determine structural stability of the vaccine showed good result. Docking structure of vaccine with TLR4 has high affinity binding. Immune-simulation reveals favourable induction of immune response with high level of interleukins production important for immunity. Periplasmic expression in *E.coli* K12 strain was quite satisfactory. This study of designing recombinant chimeric vaccine using reverse vaccinology method provides some idea about the vaccine production against Hepatitis virus.

1. Introduction

Hepatitis is the inflammation of liver which is caused by illness or due to other conditions. Some viruses like cytomegalovirus and mononucleosis causing virus can also lead to liver inflammation but they do not attack liver primarily. Infection by Hepatitis virus still remains the common cause of liver diseases, because they attack the liver primarily by executing their replication inside the liver cell. From the 5 types of hepatitis virus (A, B, C, D and E), A and E are transmitted by contaminated food and water whereas rest are spread by parental contact and sexual contact that is by direct contact of body fluids. HAV (hepatovirus) is a symmetrical RNA virus comprising of a 7500-bp long genome which encodes 3 polyproteins (P1, P2, P3) in a single open reading frame (ORF). Viral proteases cleave those 3 polyproteins into 4 capsid protein (P1 region) and 7 non-structural protein (P2 and P3). HBV (hepadnavirus) has 3200 bp long dsDNA with four ORFs: S (Virus envelop surface protein, divided into pre S1, pre S2 and S regions), C (Core region; gives rise to viral capsid), P (large polymerase protein; gets divided into three domains) and X (a large protein involved in different functions). HCV (flavivirus) has 9600 bp long positive sense ssRNA with a single ORF and encodes structural proteins (core, E1, E2) and non-structural proteins (p7, NS2, NS3, NS4A, NS4B, NS5A, NS5B).

HDV (deltavirus) has circular 1700 bp (approximately) negative sense ssRNA and it requires HBV to provide HBsAg during infection. HEV (orthohepevirus) has approximately 7200 bp positive sense ssRNA and three open reading frame (ORF): ORF1 (nonstructural proteins), ORF2 (capsid protein) and ORF3 (multifunctional proteins). Recently, another type G was discovered which resembles more with flaviviruses, but, the effect of this virus is still not clear.

All the Hepatitis viruses cause acute hepatitis but type B and C turn the disease to chronic hepatitis. Acute viral hepatitis symptoms are generally like flu, fever, jaundice etc. but those do not cause fulminant hepatic failure. Chronic viral hepatitis operates in a very mild and nonspecific way, thus the diagnosis often gets delayed. This chronic viral hepatitis turns to fatal diseases like liver damage and failure, cirrhosis of liver, liver cancer and thus treatment at proper time is important for chronic viral hepatitis. Infection by hepatitis virus can be avoided by applying vaccines. Vaccines are available for hepatitis types A, B and E but those are not capable of doing cross-protection. So, immunization with one specific hepatitis vaccine can't assure the full protection from viral hepatitis.

Vaccination involves the induced production of antigen specific response in host by the employment of inactive live-attenuated form of virus and bacteria. Few proteins of a pathogen have the capability of

* Corresponding author at: Department of Life Sciences, Presidency University, 86/1 College Street, Kolkata 700073, India.
E-mail address: kalyan.dbs@presiuniv.ac.in (K. Giri).

inducing the immune response. Also some proteins can show allergic response and thus must be excluded from the vaccination process. Later it was proved that only some parts of the sorted proteins have the ability to act as a vaccine against the diseases. Considering the aforementioned properties, the idea of developing peptide driven vaccine become an interesting topic to investigate. Long peptides which constitute the immunodominant epitopes can be used as vaccines only if they have the capability to induce humoral and cellular immunity against a broad range of serological variants. Short peptides need some carrier molecules to be used as vaccine but the chosen carriers should not elicit any allergic reaction(Li et al., 2014a). B cell epitopes are needed to produce monoclonal antibody and T_c cells are induced by intracellular pathogens. Both of them are required for prior induction of T_H cells. Human major histocompatibility complex (MHC), an important component of immune system, can create problems due to the heterogeneous population of itself. Thus, designing of a potential vaccine and the application of that particular vaccine in *in vitro* system is problematic in both time and cost due to this variability in the host MHC proteins. The vaccine should thus be presented in such a way that it can bind to a large population of the versatile MHC family.

Immunoinformatics deals with experimental immunology with the aid of computer knowledge(Schönbach et al., 2008; Tomar and De, 2010). Immunology based databases and software are there for making the experimental immunology problem much easier. Reverse Vaccinology is the most unique one, where genome gets identified as pathogen(Davies and Flower, 2007; Meunier et al., 2016). B cell and T cell epitopes can be identified, refined and validated using web servers. Previously, many *in silico* drug and vaccine designing studies were done for Dengue, Filaria, Pseudomonas, HIV, Malaria(Ali et al., 2017; Shey et al., 2019; Solanki et al., 2019; Yang et al., 2015; Pandey et al., 2018). This kind of study starts with many epitopes to generate one potential vaccine.

This study is aimed at designing a common peptide based vaccine which can be used for all the 5 types of hepatitis. 7 proteins were selected for designing the vaccine. Antigenic, non-allergenic, non-toxic, anti-inflammatory peptides capable of inducing IFN- γ and other cytokines were screened, clustered and stitched together to form the vaccine after addition of adjuvant for TLR4. Physico-chemical and peptidase cleavage properties were checked and Molecular Dynamics simulation study carried out to determine the stable structure of the vaccine. Docking study was done against TLR4 receptor and immune-simulation performed to assess the immunogenicity of the vaccine. *In silico* expression of the codon optimised vaccine in pET28a(+) plasmid of *E. coli* K12 was conducted and periplasmic expression checked to evaluate the predicted expression level of the vaccine. The entire procedure led to a highly stable structure of 601 amino acids which abides by all the immunologic and physico-chemical properties stated above and needed for prompt and effective action. Thus a peptide-based vaccine was designed which in its native state or with some other modifications can be used for further assessments *in vitro* and *in vivo*.

2. Methodology

2.1. Sorting of peptides for epitope enquiry

All the proteins of each type of Hepatitis virus were downloaded from ViralZone (www.expasy.org/viralzone) database(Hulo et al., 2011; Masson et al., 2013). There were 25 proteins and they were proceeded with, for the next procedures.

2.2. Antigenicity, subcellular localization and transmembrane domain analysis

To check antigenicity we have used VaxiJen v.2.0 (www.jenner.ac.uk/VaxiJen) server with a cut-off 0.5 to consider a protein as an antigen (Doytchinova and Flower, 2007a; Doytchinova and Flower, 2007b;

Doytchinova and Flower, 2010). The server uses auto cross covariance (ACC) transformation method on protein sequences to convert them into uniform vectors of principal amino acid properties(Wold et al., 1993).

Localization of those proteins were checked since we only focused on the outer-membrane and secretory proteins. We have used recorded data from UniProt (www.uniprot.org/), and Virus-mPLoc server (www.csbio.sjtu.edu.cn/bioinf/virus-multi/) to get those information (Apweiler, 2009; Shen et al., 2010). Virus-mPLoc server uses either Swiss-Prot followed by the GO descriptor or RPS-BLAST and PSI-BLAST, depending on homologous proteins of the query sequences and finally using the ensemble classifier the result is predicted(Bairoch, 1994; Altschul et al., 1997).

After identifying the proteins that satisfied both the above criteria, TMHMM Server, v. 2.0 (www.cbs.dtu.dk/services/TMHMM/) was used, which follows Hidden Markov Model to find the most probable topology to give a set of exact helix boundaries and uses the posterior probability plot to ascertain whether a residue lies in the cytoplasmic side or transmembrane region or outside the membrane(Krogh et al., 2001; Sonnhammer and Krogh, 2008).

2.3. Primary selection of epitopes

Selected 7 proteins were analyzed using CTLpred (crdd.osdd.net/raghava/ctlpred/algo.html) web-server, which was used to detect the presence of any TCell specific epitope in the protein(Bhasin and Raghava, 2004). CTLpred server works by shortlisting peptides as Epitopes in 5 steps including both ANN and SVM methods with high accuracy(Nelson and Wang, 1992; Vapnik, 1999).

In IEDB database (www.iedb.org) tool for MHC-I binding site prediction, a consensus method was used which consists of Artificial neural network (ANN), stabilized matrix method and scoring matrices derived from combinatorial peptide libraries to predict the epitopes based on percentile rank where a lower rank corresponds to higher affinity (Nelson and Wang, 1992; Vita et al., 2019; Peters and Sette, 2005). For prediction a reference set of antigens, based on most common specificities in general population was taken. The cut-off for percentile ranks was taken at 1 and epitopes of 9 and 10 amino acids were determined corresponding to the reference set.

EpiTOP 3.0 (<http://www.pharmfac.net/EpiTOP>) was used to predict the epitopes for MHC-II (DQ, DP, DR) antigens separately(Dimitrov et al., 2010; Jacobsen and Wiersma, 2010; Wiesen et al., 2016; Hung et al., 2015). The server uses a quantitative matrix derived using iterative self-consistent algorithm from known binders of varying lengths. The initial training set was used to extract the first model which was used to calculate the pIC50s and the best scoring peptides formed the second set. The process was repeated till the number of peptides between 2 consecutive sets was around 99%. Epitopes in the extracellular region and those that bind to at least 1 antigen were selected. The binding level was set at pIC50 value of 6.

B cell epitopes were predicted using BCPRED (<http://ailab.cs.iastate.edu/bcpreds>) and ABCpred (www.imtech.res.in/raghava/abcpred/) servers(Chen et al., 2007; El-Manzalawy et al., 2008a; El-Manzalawy et al., 2008b; Saha and Raghava, 2006). Antigenicity (ANTIGENpro, <http://scratch.proteomics.ics.uci.edu>), Solvent Accessibility and flexibility of linear B-cell epitopes were checked. BepiPred 1.0 (www.cbs.dtu.dk) identifies the location of linear B cell epitopes using a combination of a Hidden Markov Model and propensity scale. BepiPred-2.0 (<http://www.cbs.dtu.dk/services/BepiPred/>) works by Random Forest algorithm and the process was used to further validate epitopes precisely(Sonnhammer and Krogh, 2008; Magnan et al., 2010; Fieser et al., 1987; Larsen et al., 2006; Jespersen et al., 2017; Gordon et al., 1984; Moore, 1987). Discontinuous B cell epitopes were not taken into account as very few of the proteins have solved structure.

2.4. Secondary selection of epitopes and vaccine stitching

Peptides having 75% conservancy with each other were taken. Vaxijen v.2.0 (<http://www.jenner.ac.uk/VaxiJen>) server was used for antigenicity check of those resulted peptides(Doytchinova and Flower, 2007a; Doytchinova and Flower, 2007b; Doytchinova and Flower, 2010). These antigenic Peptides were then proceeded for allergenicity check by 2 servers consecutively; AllerTOP v.2 (<http://www.ddg-pharmfac.net/AllerTOP>) and AllergenFP (www.ddg-pharmfac.net/AllergenFP/)(Dimitrov et al., 2014a; Dimitrov et al., 2014b). Both of them work on the basis of ACC transformation and E-descriptor value and identify the allergenicity of a protein by some mathematical and machine learning methods(Apweiler, 2009; Venkatarajan and Braun, 2001). ToxinPred (crdd.osdd.net/raghava/toxinpred/), which works by machine learning methods and quantitative matrix based models by using different properties of amino acids, was used for identifying non-toxic peptides(Gupta et al., 2013). AIPpred (<http://www.thegleelab.org/AIPpred>) server was used to check anti-inflammatory property which works on the basis of ML algorithm and compositional features and gives data quite accurately(Manavalan et al., 2018; Cao et al., 2019). Though we have checked IFN- γ induction efficiency later, but to design a vaccine precisely, IFN- γ induction efficiency of individual peptides was checked using IFNepitope (crdd.osdd.net/raghava/ifnepitope/) server which is based on both Motif and SVM based models using MERCi software(Vapnik, 1999; Dhanda et al., 2013). These steps ended up with almost 88 peptides which satisfy all the positive criteria to be used as vaccine candidate. Clustering of peptides were done using IEDB clustering (www.iedb.org) and cut-off was set to 70%. These peptides were then stitched together with AAY linker and one TLR4 adjuvant peptide APPHALS added in the front with EAAAK linker(Yang et al., 2015; Dhanda et al., 2018; Shanmugam et al., 2012; Ito et al., 2017; Arai et al., 2001; Kahl, 2015).

2.5. Physicochemical properties prediction

Physicochemical properties of the resulted vaccine was calculated using ProtParam (<http://web.expasy.org/protparam/>) web-server which gives various parameter data like Molecular Weight (MW), theoretical pI, GRandAVerage of hydropathicitY (GRAVY) etc.(Gasteiger et al., 2005). This server has different methods of calculations and respective cut-off values for each parameter. To know about solubility of the vaccine we used Protein-Sol server (<http://protein-sol.manchester.ac.uk>), which compares query sequence solubility and a reference solubility by a graph where the cut-off is set to 0.45(Hebditch et al., 2017; Hebditch and Warwicker, 2019).

2.6. Structure validation and refinement of the vaccine

To know about structural information of vaccine, first protein-protein blast of the vaccine against full human genome was done and there was no such prominent hit. PSIPRED was used (<http://globin.bio.warwick.ac.uk/psipred/>) to get information about the secondary structure of the vaccine(Shey et al., 2019; Vapnik, 1999; Buchan and Jones, 2019; Jones, 1999). This server uses the output obtained from PSI-BLAST, which identifies sequences having homology with the query sequence and PSIPRED uses those sequences to construct a position specific scoring matrix and further validates by two feed-forward neural network(Pandey et al., 2018; Altschul et al., 1997). RaptorX-Property (<http://raptorx2.uchicago.edu/StructurePropertyPred/predict/>) was used to define extra information about secondary structure(Wang et al., 2016).

2.7. Tertiary structure prediction

I-TASSER (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) server was used to compute tertiary structure of the vaccine, which

predicts the 3D-structure of a protein by having the amino-acid sequence(Roy et al., 2010; Yang and Zhang, 2015; Yang et al., 2014). It identifies PDB files using LOMETS based approach with respect to the provided amino acid sequence and creates models by template-based threading mechanism(Wu and Zhang, 2007). It uses state-of-the-art algorithms to predict most accurate structure and functions.

2.8. Molecular dynamics simulation

GROMACS 2018.3 was used to simulate the predicted three dimensional structure of the vaccine (Abraham et al., 2015). CHARMM36 force field with TIP3P water model and dodecahedron simulation box was used in the simulation(Huang and MacKerell, 2013). Distance between protein and the boundary of the box was set 10 Å and periodic boundary condition was used. Protein was solvated with 34,862 water molecules and neutralized by adding 16 chloride ions. Steepest Descent method (maximum no. of steps 50,000) was employed to energy-minimize the neutralized system. Energy minimized system was subjected to 500 ps NVT equilibration at 310 K and successively NVT equilibrated system was subjected to 500 ps NPT equilibration at 1 atm pressure and 310 K temperature. During NVT and NPT equilibration period, the protein position was fully restrained. NPT equilibrated system was then subjected to 15 ns production run with 2 fs time step with no restrain. Smooth Particle-Mesh Ewald method was used to calculate long-range electrostatic interactions and a 12 Å cut-off was used for both PME and van der Waals interactions(Essmann et al., 1995). Average structure of the protein during simulation was calculated using inbuilt GROMACS command line tools.

2.9. Docking with TLR4

TLR4 complex (PDB ID: 3FXI) was taken from PDB and using Discovery Studio, only the TLR4 chains were selected for docking (Pandey et al., 2018). PatchDock (<http://bioinfo3d.cs.tau.ac.il>) server was used to execute docking study using this modified TLR4 and our designed vaccine(Duhovny et al., 2002; Schneidman-Duhovny et al., 2005). PatchDock server uses proteins, peptides, DNA or small molecule as input structure and the output comes as a list of docked structures following shape complementarity and atomic contact energy criteria.

2.10. Peptide cleavage identification

Netchop 3.1 (www.cbs.dtu.dk/services/Netchop) server was used which predicts human specific proteasome cleavage sites, based on novel sequence coding and neural network training strategies(Nielsen et al., 2005). Amino acid sequence in FASTA format is given to get output, which can be in both short and detailed format. Cathepsin is a proteasomal pathway specific enzyme, whose cleavage site is specially examined by Site Prediction (<http://www.dmbr.ugent.be/prx/biot2-public/SitePrediction/>) server(Verspurten et al., 2009).

2.11. Immune-simulation study

C-ImmSim server (<http://150.146.2.1/C-IMMSIM/index.php>) is an agent based model which was used for Immune-simulation study(Rapin et al., 2010). It uses position-specific scoring matrix and machine learning methods to identify the epitope peptides and other immune interactions. Three mammalian anatomical regions get simulated at a time by this server which are Bone Marrow (lymphoid and myeloid cell), thymus (T cell) and a lymphatic organ to exhibit immune response (Rapin et al., 2010). All parameters are set to default at the time of vaccine introduction and only one injection was given.

2.12. Periplasmic expression

SLP-Local (<https://sunflower.kuicr.kyoto-u.ac.jp>) /smatsuda/

slplocal) server was used for the purpose of identifying the expression of the vaccine and it identified periplasmic location for the particular vaccine(Matsuda et al., 2005). Periscope (<http://lightning.med.monash.edu/periscope/>) server was used to predict the periplasmic expression level(Chang et al., 2016). Having the amino-acid sequence as input the server works with two architecture strategies i.e., SVM and SVR respectively and gives an output of location and level of expression with high probability(Vapnik, 1999).

2.13. Codon optimization and *in silico* expression

To execute codon optimization experiment, JCcat (<http://www.prodoric.de/JCat>) server was used and it provides both graphical and Codon Adaptation Index value as output(Grote et al., 2005). *E.coli* K12 strain was chosen for all these predictions. SnapGene software (SnapGene software from GSL Biotech; available at snapgene.com) was used to predict *in silico* expression of the peptide based vaccine by introducing the nucleotide sequence between the site of two restriction enzymes *Xba*I and *Nde*I inside a pET28a(+) vector.

3. Result

3.1. Sub-cellular localization, antigenicity and transmembrane domain analysis for potential proteins

Antigenicity of a protein refers to the potentiality to generate immune response against the organism to which the protein belongs, hence, it is a necessary factor to use the protein as vaccine. Antigenicity of all the proteins of each type of Hepatitis virus were checked using VaxiJen Webserver. The cut-off antigenicity is taken 0.5. Cellular localization was tracked for these proteins. Depending on the subcellular localization data, domain analysis of particular proteins was done using TMHMM server. We have considered the proteins which have extra-cellular domain or are secreted. Depending on the data tabulated in Table 1, 7 proteins: VP3, VP1-2A, C, CORE, S, E2 and ORF2 were taken for further studies.

3.2. Epitope selection for cytotoxic T cell (T_c cell)

Conservancy is an important criteria of epitope-driven vaccine development. So, we executed a conservancy test with 75% cut-off conservation of epitopes with pathogen variant of different infection stages. An epitope needs to bind with different MHC alleles with high affinity to introduce immune response. Epitope should work on vast range of population. Identification of epitopes were carried out using CTLpred server having ANN cut-off 0.8. 37 peptides were sorted and validated using different servers. VaxiJen server was used to check antigenicity taking a cut-off 0.6. 20 out of 37 predicted to be antigenic. Allergenicity check was performed using Allertop v.2 and AllergenFP servers and predicted 14 peptides to be allergen out of 20. Finally, 2 peptides were predicted non-toxic, anti-inflammatory and capable of IFN- γ induction with the help of ToxinPred, AIPpred and IFNepitope servers respectively. Those 2 peptides were identified as potential epitopes for VP3 protein of HAV. Using same procedure we found 1 epitope of HAV VP1-2A, 3 epitopes of HBV CORE Antigen and 4 epitopes of HEV ORF2 (Supplementary 1).

3.3. MHC-I epitope selection for helper T cell

Identification of epitopes were carried out using IEDB tool specific for 27 MHC-I HLA. Conservancy test result sorted with 101 peptides of both 9mer and 10mer and those were validated using different servers. VaxiJen server was used to check antigenicity taking a cut-off 0.6 and 59 out of 101 came out to be antigenic. Allergenicity check with Allertop and AllergenFP servers came up with 15non-allergen peptides. This processes continued with those 15 peptides which were gone

through toxin test gave 13 non-toxin peptides. Anti-Inflammatory and IFN- γ induction tests were executed with the help of respective servers. Finally, results ended up with 2 potential epitopes. The above example was about VP3 protein of HAV. Other proteins were also gone through with same procedure resulting 3⁺ epitopes of HAV VP1-2A, 2⁺ epitopes of each HBV C and CORE Antigen proteins, 1 of HBV S protein and 3⁺ epitopes of HEV ORF2 (Supplementary 1).

3.4. MHC-II epitope selection for helper T cell

Epitopes were identified for each of DR, DQ, DP class and separately examined. pIC50 values were recorded for each of the 3 class epitopes with a cut-off values 6(Supplementary 2). Antigenicity positive result came for 268 out of 448. Allergenicity result opted for 66 epitopes out of 268. Toxin check, Anti-Inflammatory check and IFN- γ induction test results finally ended up with 6⁺ epitopes. This result is for VP3 protein of HAV and same protocol was followed for each proteins and 6⁺ epitopes of HAV VP1-2A protein, 2 epitopes of each HBV C protein and CORE Antigen protein, 2⁺ epitopes for S protein, 1 of HEV E2 protein and 23⁺ epitopes of HEV ORF2 protein (Supplementary 1).

3.5. Epitope selection for B cell

BCpred and ABCpred were used for initial Epitope selection. Antigenicity Check, Solvent Accessibility, Flexibility, BepiPred1.0 and BepiPred2.0 were carried out with 8 epitopes. These 8 epitopes were validated further to test antigenicity, allergenicity, toxicity and anti-Inflammation property. Though no epitopes were identified having all satisfactory characteristics for VP3 of HAV but there were 2 epitopes of each HAV VP1-2A, HBV C, CORE Antigen proteins and HEV ORF2 (Supplementary 1).

3.6. Vaccine preparation

73 Peptides (Table 2) were taken and BLAST was performed against human proteome. No such specific protein hits were found. So, clustering of the peptides were carried out using IEDB cluster programme with a cut-off of 70% identity, resulting in 45 peptide sequences of which 10 are consensus sequences and the rest singletons. These peptides were then stitched together using specific linker AAY tripeptide between every pair of epitopes. Adjuvant of TLR4; APPHALS was added with a different linker EAAAK. This process results in a 601 amino acid peptide and this peptide was further validated (Fig. 1).

3.7. Physicochemical properties of the designed vaccine

The physicochemical properties of the vaccine was calculated using Expasy Protparam Server. Molecular weight of the vaccine is 65,087.41 g/mol and pI is 9.24, which is quite acceptable as it is close to histidine pI. Aliphatic index is 77.0, thus can be said as thermostable. The vaccine is moderately nonpolar in nature with a GRAVY value –0.078. Half-life of that vaccine has satisfactory data for all the system. Solubility of the vaccine has a value of 0.288 (scale 0–1).

3.8. Structure prediction

Secondary structure of the protein was predicted using RaptorX-Property (Figure: 2A) and PSIPRED (Figure: 2C) servers. The predicted secondary structure results indicate 29% helix, 17% beta-sheet and 52% coil structure. 6% of the protein consisting of 41 residues is disordered structure. Solvent accessibility of the vaccine was concluded as 41% exposed, 17% moderately exposed and 41% fully buried (Fig. 2B).

Tertiary structure (Fig. 2D), obtained from Threading based server I-tasser with respect to 10 templates having Z-score value ranging from 0.51–2.87. Result provided 5 models from which MODEL-1 with a C-score of –0.90 was taken. This particular model has an estimated TM-

Table 1

All the proteins of 5 different types of Hepatitis viruses. 7 proteins from all 5 types of Hepatitis virus are taken for antigenicity checking, location in cell and transmembrane domain analysis.

Hepatitis virus types	Name of proteins (uniprot id)	Antigenicity	Location in cell	Transmembrane domain analysis	
				Domain	Amino-acid span
Hepatitis A virus (HAV)	Viral protease 0 (P08617)	0.4994	Virion	Outside	1–245
	Viral protease 3 (P08617)	0.6265	Host multivesicular body	Outside	1–246
	Viral protease 1-2A (P08617)	0.5332	Host multivesicular body	Outside	1–345
	2 BC (P08617)	0.5517	Host membrane (2B) and Host membrane (2C)	Outside	1–586
	3ABCD (P08617)	0.5258	Host membrane	Inside	1–39
Hepatitis B virus (HBV)				TM Helix	40–59
				Outside	60–805
	C (Q76R61)	0.5333	Host cytoplasm	Outside	1–183
	S (Q76R62)	0.5371	Virion membrane	Outside	1–253
				TM helix	254–272
				Inside	273–343
				TM helix	344–366
				Outside	367–375
				TM helix	376–398
				Inside	399–400
Hepatitis C virus (HCV)	Core (P0C767)	0.5328	Host nucleus, secreted	Outside	1–212
	P (Q69028)	0.5809	Host cytoplasm, nucleus, mitochondria		
	X (Q69027)	0.4637	Host nucleus (virus MPLOC server)		
	P21 (P27958)	0.3824	Host ER membrane, host lipid droplet		
	E1 (P27958)	0.4302	Virion membrane, host ER membrane		
Hepatitis D virus (HDV)				Inside	1–73
				TM helix	74–96
				Outside	97–127
				TM helix	128–147
				Inside	148–167
				TM helix	168–190
				Outside	191–192
				Outside	1–363
	E2 (P27958)	0.5643	Virion membrane, host ER membrane		
	P7 (P27958)	0.6073	Host ER membrane		
Hepatitis E virus (HEV)	NS2–3 (P27958)	0.4437	Host ER membrane		
	NS3 (P27958)	0.4868	Host ER membrane		
	NS4A (P27958)	0.5546	Host ER membrane		
	NS4B (P27958)	0.5106	Host ER membrane		
	NS5A (P27958)	0.4428	Host ER membrane, host perinuclear region, host mitochondrion		
	RDRP (P27958)	0.4536	Host ER membrane		
	SMALL (P0C6L3)	0.4496	Virion, host nucleus		
	LARGE (P29996)	0.4194	VIRION, host nucleolus		
	ORF1 (Q81862)	0.4410	Host cytoplasm		
	ORF2 (Q81871)	0.6108	Secreted, virion, host cytoplasm, ER, golgi apparatus, cell surface	Outside	1–660
	ORF3 (Q81870)	0.5718	Host ER, cytoplasm		

score of 0.60 ± 0.14 and estimated RMSD of $9.8 \pm 4.6 \text{ \AA}$. C-score was considered to choose a structure and which was further validated using Ramachandran plot.

3.9. Simulation study

From the RMSD plot of the protein backbone, it can be concluded that the structure takes very short time to reach equilibrium. It takes about initial 4 ns for equilibration and the protein then reaches a stable plateau with average RMSD of $1.113 \pm 0.07 \text{ nm}$. RMSF values indicate that the core amino acid residues of the protein have lower fluctuations than the terminal residues. Relatively the N-terminal adjuvant part has high RMSF values at about 1.60 nm and visual inspection of the trajectory shows that it moves heavily in the solvated box. Apart from the terminal residues, Met492 has high RMSF value of 1.32 nm . Ramachandran plot of this average simulated structure is provided, which shows 88.6% residues in favoured region, 8.7% residues in allowed region and remaining 2.7% residues in outlier region (Fig. 3). Thus the structure has a high percentage of residues where there are no steric clashes present, which indicates high stability.

3.10. Docking study

The simulated structure was docked with the TLR4 which plays various role in Hepatitis virus infection (Tomar and De, 2010; Davies

and Flower, 2007; Meunier et al., 2016). The vaccine interacts with both the chains of TLR4. Adjuvant portion predominantly made H-bond and Hydrophobic interaction similar to the whole vaccine (Fig. 4). The decrease in energy due to complex formation is 396.92 kcal and the interface has an area of 2697.7 \AA^2 .

3.11. Immunological study

C-ImmSim server provides different Immune response results for the vaccine sequence. Looking at the different graphs obtained from the server it was observed that the induced B cell as well as T cell population were satisfactory after the vaccine injection. Active T cell populations increase for almost all types of T cell between 5 and 10 days. Interleukins are one of the important factor for Immune response. Result showed that the vaccine can induce both IFN- γ and IL-2 with a satisfactory D value. By using IL-10Pred server it was checked if that vaccine can induce IL-10 and the result came with a value of 0.84 which is a good induction level interpretation (Fig. 5).

3.12. Peptidase activity on the vaccine

The epitopes should be presented by APCs to satisfy the main motive of vaccine. Proteasomal cleavage is important for TCell epitope presentation. Thus we have investigated both proteasomal and Cathepsin specific peptidase activity on the vaccine. NetChop 3.1 Server provided

Table 2

The selected proteins of Hepatitis viruses and their respective Peptides which are used as Epitopes for vaccine design (before clustering through IEDB Cluster Program).

Protein Name	Hepatitis Virus	T Cell										B Cell					
		T _C Cell			T _H Cell												
		Initial	Final	Peptides	MHC-I			MHC-II			Initial	final	Peptides	Initial	Final	Peptides	
					Initial	Final	Peptides	Initial	Final	Peptides							
VP3	HAV	37	2	ARAKMSFAL EDARAKMSF	101	2	GVQSTLRF VQSTLRFV	448	6 [†]	ITLKQATT VQSTLRFV [†] FAPLYHAMD [‡] FALDQEDWK GVQSTLRFV [†] LLFCFVPGN	8	-	-				
VP1-2A	HAV	54	1	KTALGAVRF	158	3 [†]	KTALGAVRF [†] EVGKQLKY TTVSTEQNV	585	6 [†]	RSHFLCTFT HFCLCTFTFN EYTFPITLS [†] TTVSTEQNV EDPVLAKKV KTALGAVRF	22	2	KEYTFPITLSSTSNNPPHGLP NNKEYTFPITLSSTSNNPPHGLP				
C	HBV	22	-	-	94	2	SPRRRRSQS [†] RRRSQSPRR	322	2 [†]	SFLPSDFP SPRRRRSQS [†]	13	2	STLPETTVVRRRGR STLPETTVVRRRGRSP				
CORE	HBV	32	3	VRRRGRSPR PRRRRSQSP SPRRRRSQS	113	2 [†]	SPRRRRSQS [†] RRRSQSPRR	382	2 [†]	SFLPSDFP SPRRRRSQS [†]	6	2	STLPETTVVRRRGR STLPETTVVRRRGRSP				
S	HBV	34	-	-	69	1	CPGYRWMCL	396	2 [†]	WTSLNFLGG CPGYRWMCL [†]	42	-	-				
E2 ORF2	HCV HEV	51 102	- 4	YAISISFWP YNYNTTASD PNAVGYYAI LRANDVLWL	155 273	- 3 [†]	NLADTLLGGL LPLQDGTNTH TEASNYAQY [†]	878 1175	1 23 [†]	LSTGLIHLH TAAAGAGPR AGPRVRQPA SPLPLQDG TIRYRPLVP [†] EATGLVML [†] PYTGALGLL [†] TPGNTNTRV SANGEPTVK VATGAQAVA TLDGRPLST PLRGKLSFW CPECRPLGL LHYRNQGWR [†] LELEFRNLIT LDWTKVTLD AAGAGPRVR PLGSAWRDQ YAISISFWP AISISFWPQ LRANDVLWL PLRGKLSFW KAGYPYN GRRSGGSGG	31 62	- 2	-	SERLHYRNQGWRSV EAGTTKAGYPYNNTTASDQLL			

101 proteasomal sites and Site prediction Server provided 1 peptidase link for Cathepsin L, K and D (human Specific) with 99.9% specificity and 42 peptidase link for other Cathepsin with 99% specificity.

3.13. Expression in *E. coli*

Expression data in *E. coli* were predicted by the SLP-Local server, which provided the subcellular localization probability of protein for Cytoplasmic, periplasmic or Extracellular chambers. The result shows SVM score of 0.94 for periplasmic chamber and SVM score of -0.43 for extra-cellular chamber. Periscope server provides periplasmic expression of 13.103 mg/l in *E. coli* for the vaccine.

3.14. Codon optimization and *in silico* expression

Codon optimization using Jcat server shows CAI (codon adaptation index) value of 0.9954, which is an acceptable value for good expression in *E. coli* K12 system. Codon usage bias while forming cDNA is also predicted (Fig. 2). The GC content of the particular vaccine is 56.53%

(range 30–70%), which implies a good genomic expression. Lastly, Using SnapGene software we predicted the *in silico* expression of the vaccine in pET28a(+) vector. The vaccine nucleotide sequence was inserted between the cut site of two restriction enzymes; *Xba*I and *Nde*I (Fig. 6).

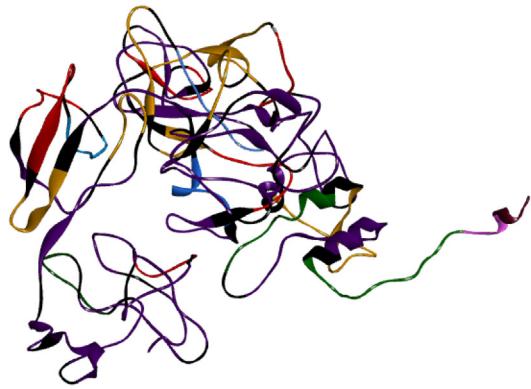
4. Discussion

Peptide based vaccines have emerged in the last few years and thus study is unexplored in many aspects. Till 2014, 452 clinical studies of peptide vaccines have been performed. 270 peptide based vaccines are in phase 1 of clinical trials, 224 in Phase 2 while 12 in phase 3. The phase 3 vaccines all were against various types of cancer (Li et al., 2014b). This rise of number of studies in peptide based vaccines as a possible future alternative to traditional vaccines is due to various features of these vaccines which include complete control of chemical nature of vaccine, ability to be chemically synthesized, cost-effectiveness, easy reproducibility, stability upon storing, solubility, specificity and less number of side effects. But the major hurdles here are the poor

A VP3 VP1-2A C CORE S E2 ORF2 Linker Adjuvant-linker Adjuvant

APPHALSEAAAKSTLPETTVVRRGRSPRAAYEAGTTKAGYPNYNTTAS
DQLLAAYSPPRRRSQSPRRAAYNNKEYTFPITSSTSNNPPHGLPAAYSERLH
YRNQGWRSVAAAYAISIFWPQAAYGVQSTLRFRVAAYRSHFLCTFTNA
AYTAAAGAGPRVRAYEDARAKMSFAAAYAGPRVRQPAAYNLADTL
GGLAAVIAINYNHDSDEAAYTIVSTEQNVAAWTSLNFLGGAAYTEASNYA
QYAAYDSTFGLVSIAYEVGKQLRKAAYLPLQDGNTHAAYGRRSGGSG
GAAYTLDRPLSTAAYPLRGKLSFWAAAYSANGEPTVKAAYLELEFRNLTAA
YFAPLYHAMDAAYSPLPLQDGAAAYCPGYRWMLAAYTIRYRPLVPAAYE
DPVLAKVAAAYLLFCFPVPGNAAYLSTGLIHLHAAYTILQATTAAYVATGA
QAVAAAYPVKELRLEVAAYLRLNDVWLAAYEATSGLVMMLAAYCPECRPL
GLAAVKTALGAVRFAAYSFLPSDFPPAAYPYTGALGLLAAYPLGSAWRDQ
AAVFALDQEDWKAAYPNAVGGYIAAYTPGNTNTRVAAYLDWTKVLD

C



B

GCTCCGCCGACGCTCTGCTGAAGCTGCTGCTAAATCTACCGTCCGGAAACAC
CGTTGTCGTCGTCGGTCGTTCTCGCGCTGCTGCTACGAAGCTGGTACCA
AAGCTGGTACCCGTACAACACTACAACACCAACCGCTCTGACCAGCTGCTGGCTGC
TACTCTCCGCGTCGTCGTCGTTCTGAGTCCGCGTCGCTGCTTACAACAAACAAA
GAATACACCTTCCGATCACCTGCTTCTACCTCTAACCCGCCGACGGTCTGCC
GCTGCTTACTCTGAACGCTGCACTACCGTAACCAAGGGTTGGCTTCTGTTCTGC
TTACTACGCTATCTCTCTTCTGCCGCGCAGGCTGCTTACGGTGTTCAGTCTAC
CCTGCCCTCCGCTGCTGCTTACCGCTTACTCTCTGTCGACCTTCACCTTCAAC
GCTGCTTACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAAG
CGCTGCTGCTAAATGCTTCTGCCGCTGCTTACGCTGGTCCGCTGTTCTCA
GCCGGCTGCTGCTTACACCTGCTGACACCCCTGCTGGGTGGTCTGGCTGCTTACA
TCGCTAACTACAACCAACTCTGACGAAGCTGCTTACACCACCGTTTACCGAACAG
AACGTTGCTGCTTACTGGACCTCTGAACTCTGGGTGGTGTGCTTACACCGA
AGCTTCTAACTACGCTCAGTACGCTGCTTACGACTCTACCTCGGCTGCTGGTTCTAT
CGCTGCTTACGAAGTTGTAACACGGCTGCTAACATGCTGCTTACCTCGGCTGCTG
AGGACGGTACCAACACCCACGCTGCTTACGGCTGCTGCTTCTGGTCTGGTGGT
GCTGCTTACCCCTGGACGGCTGCTGCTGCTTACCGCTGCTTACCGCTGCTGCTG
TAAACTGCTTCTGGGCTGCTTACTCTGTAACCGTGAACCGACCGTTAAAGCTG
CTTACCTGGAACCTGAAATTCCGTAACCTGACCGCTGCTTACTCTCCGCTGCTG
ACGCTATGGACGCTGCTTACTCTCCGCTGCTGCCGCTGCTGAGGACGGTGTGCTTAC
TGGCCGGGTTACCGTGGATGTCCTGGCTGCTTACCATCGGTTACCGTCCGCT
GGTTCGGGCTGCTTACGAAGACCCGTTCTGGCTAAAAAAAGTTGCTGCTTACCTG
TGTTCGCTTCTCGTCCGGTAACGCTGCTTACCTGCTTACCGGCTGATCACCTG
ACGCTGCTTACATCACCCGTAACACAGGCTACCACCGCTGCTGCTTACGTTGCTACC
GGTGCCTAGGCTGTTGCTGCTTACCCGTAACAAAGAAACTGCGCTGGAAGTTG
TGCTTACCTGCGTCAACGACGTTGCTGGCTGGCTGCTTACGAAGCTACCTG
GTCCTGGTATGTCGGCTGCTTACTGCCGGAATGCCGCTGCCGCTGGTCTGGTGT
TACAAAACCGCTCTGGGTGCTTACCCGTAACCCGGCTGCTGGTCTGCTGCTG
TCTTCCCGGCTGCTTACCCGTAACCCGGCTGCTGGTCTGCTGGTCTGCTG
CTGGGTTCTGCTGGGTGCTGCTGGGCTGCTTACTCGCTCTGGACAGGAAAGACTG
GAAAGCTGCTTACCCGAAACGCTGTTGGTGGTACCGCTATCGCTGCTTACCCCCG
GTAACACCAACACCCGTTGCTTACCTGGACTGGACCAAAGTTACCGTGGAC

Fig. 1. Pictorial representation of epitope based vaccine. (A) Amino acid sequence of the vaccine with respective colour code of proteins of all Hepatitis viruses (B) Reverse translated nucleic acid sequence of vaccine (C) Tertiary structure of the vaccine with respective colour code. Respective colour-code of proteins of all Hepatitis viruses is mentioned above.

immunogenicity of the proteins and their degradation by the various enzymes. To overcome the first one, specific immunogenic epitopes are selected which have the capacity to generate an immune response and to overcome the latter one, multiple epitopes are used so that even if one gets cut other epitopes remain. This multiple epitope strategy has reached advanced clinical trials for HPV (Skwarczynski and Toth, 2016) and is again a new emerging field as well.

Our current study is focused on the 5 different types of Hepatitis viruses (A, B, C, D and E). Most of the Hepatitis viruses spread through body fluid except A and E types. Hepatitis B, C and E are the most fatal types as they cause chronic and acute liver disease. Hepatitis B has more percentage of death than Hepatitis C and E(Hepatitis, 2019; Hepatitis, 2019; Hepatitis, 2019). Type A Hepatitis is not fatal but it contributes 0.5% of total death caused by hepatitis(Hepatitis, 2019). Hepatitis D can't survive alone so it needs co-infection of Hepatitis B, which provides the envelope (surface) protein to Hepatitis D to cause infection(Hepatitis, 2019). At present vaccine is available for Hepatitis A which consists of inactivated attenuated virus which in normal circumstances has lost the power to cause infection while can still generate an immune response(Jacobsen and Wiersma, 2010). But still there is a small probabilistic chance that the virus can switch to the infectious cycle. Apart from this as a side effect allergic response can be seen in certain individuals due to the allergenicity of the vaccine. For Hepatitis B, the vaccine consists of parts of the virus specifically viral envelope and surface antigen(Jacobsen and Wiersma, 2010; Wiesen et al., 2016; Hepatitis, 2019). In case of Hepatitis C virus, no vaccine is available till date mainly due to the huge level of variation which reaches a highest of 70% between the various genotypes which would make the vaccine useless against a different genotype when raised against a particular one. In case of Hepatitis D virus, no vaccine is there but vaccine for Hepatitis B virus works in this case as the Hepatitis D being incomplete itself needs HBV to cause infection. Finally in case of Hepatitis E virus,

vaccine is available which is licensed only in China and is of subunit type raised against the capsid protein. Though the diseases are pretty much similar but as the genomic nature of those viruses are different, it is hard to develop one common vaccine for all Hepatitis viruses. There may be a chance of getting infected by another type of virus in spite of having vaccination against one specific type of Hepatitis virus. In this study we have attempted to construct a vaccine which can be used against all the four types (A, B, C and E) of Hepatitis using *in silico* methods.

We have considered only the extra-cellular and secretory proteins from the total 25 proteins present in all the Hepatitis viruses. Sub-cellular localization of these proteins were checked and transmembrane domain analysis were performed for the extracellular proteins, thus resulting in 7 proteins which were selected to identify the peptides to be used as epitopes. Antigenicity, non-allergenicity, non-toxicity, anti-inflammatory properties are needed for a vaccine to elicit a specific immune response and not a generalized inflammatory one upon vaccination, apart from also generating the IFN- γ response. The epitopes were clustered and stitched with AAY linkers and EAAAK linker was used for joining the TLR4 adjuvant(Yang et al., 2015; Arai et al., 2001; Kahl, 2015). Final form of this peptide derived vaccine was then subjected to antigenicity, allergenicity, toxicity and solubility validation. PSI-BLAST and HHPRED against human proteome provides no specific hit which indicates that the vaccine may not create an auto-immune response and also may not overlap with some other human protein with a different function thus incurring its specificity(Pandey et al., 2018; Altschul et al., 1997; Söding et al., 2005). This vaccine is thermostable with a good GRAVY score and has a moderate basic pI and molecular weight. Secondary structure of the vaccine showed a good proportion of helical and sheet elements with some random coil. Tertiary structure using threading based software gave 5 models which were further validated by Ramachandran plot and as per server property, model with

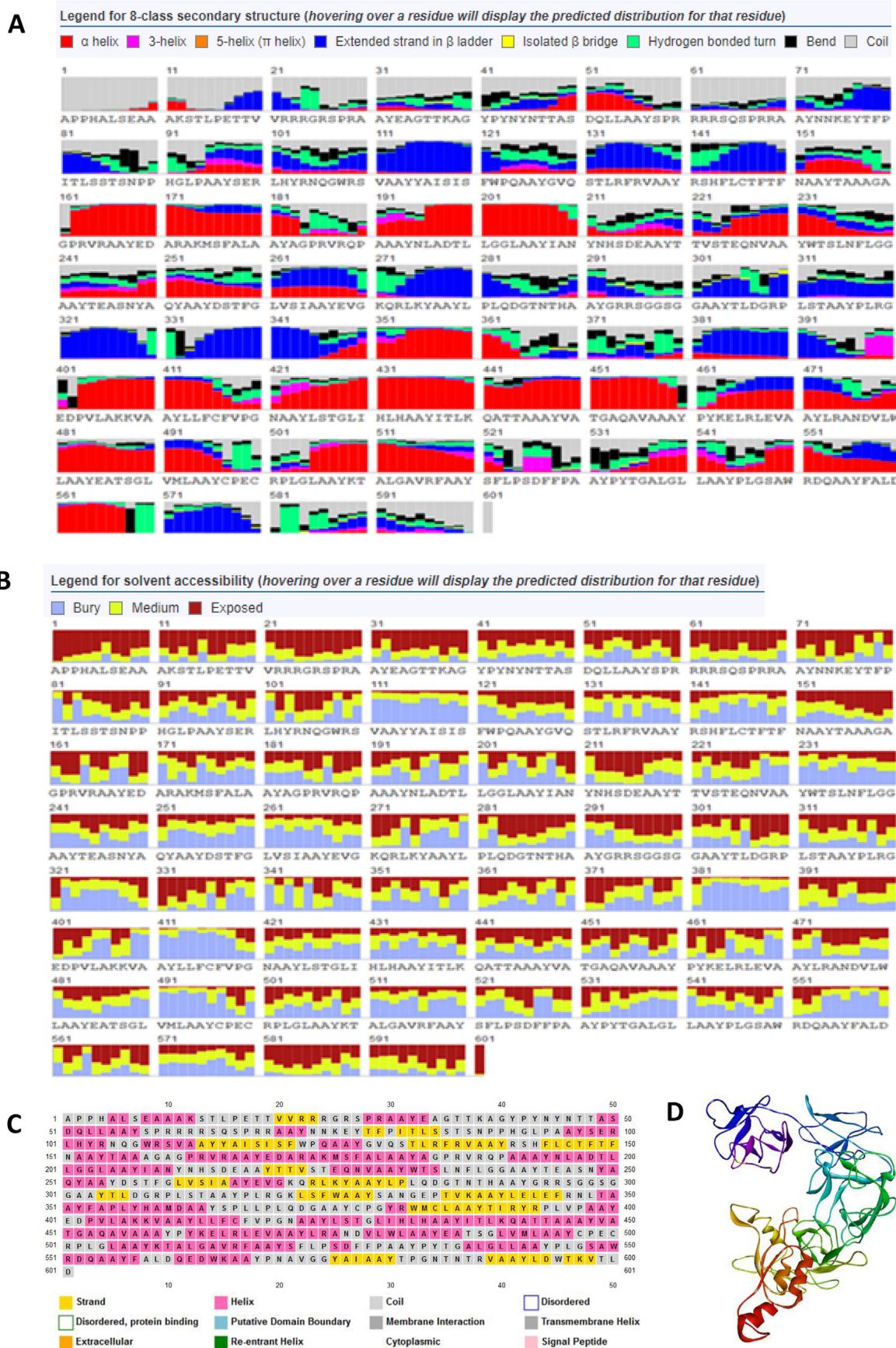


Fig. 2. Representation of structural properties. (A) 8 classes of secondary structure, (B) Solvent accessibility using RAPTORX-Property server, (C) Representation of secondary structure by PSIPRED, (D) Tertiary structure by I-TASSER server of designed vaccine.

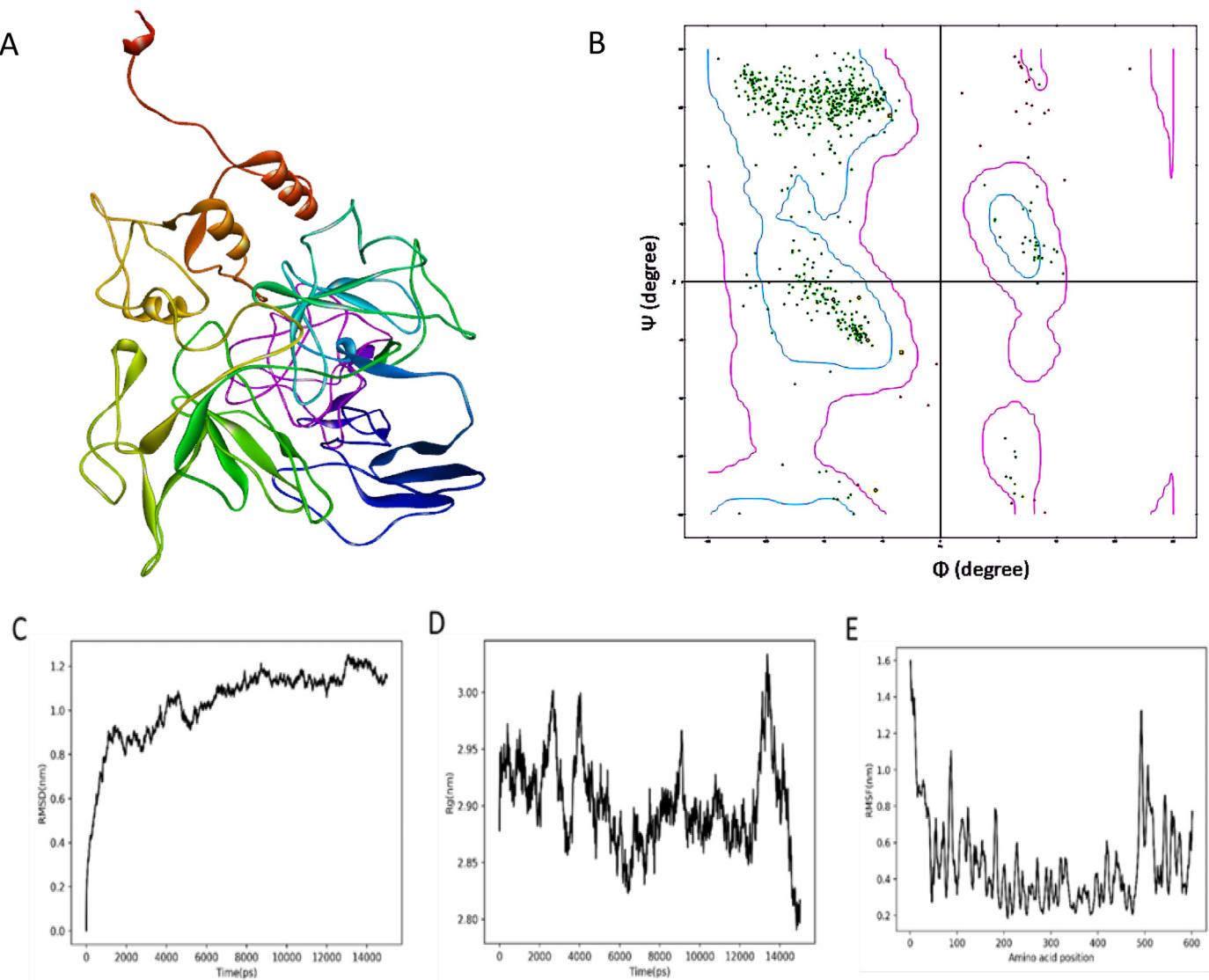


Fig. 3. Molecular Dynamics Simulation study of the vaccine. (A) Simulated average structure of the vaccine using GROMACS. (B) Ramachandran plot of the vaccine structure (residues in favoured region: 88.6%, residues in allowed region: 8.7%, residues in outlier region: 2.7%). (C) RMSD, (D) Radius of gyration (R_g) and (E) RMSF graph of vaccine during simulation.

the highest confidence score (C score) was selected and subjected to simulation study. RMSD plot of the protein backbone illustrated that the equilibrium for the vaccine structure was achieved within 4 ns, after which a plateau remained throughout rest of the time scale. From RMSF plot we can conclude that the terminal portion of the vaccine showed higher fluctuation than the core thus emphasizing the absence of any type of stable influence on this part which has the highest RMSF value. During simulation all the secondary structures except coils and turns are preserved indicating the protein's structural stability. From the simulation it can be concluded that the designed vaccine is very much stable at physiological temperature.

Docking of the vaccine with TLR4 peptide was carried out using PatchDock server and model 2 was chosen for refinement and investigation based on atomic contact energy and interfacial surface area (Wu and Zhang, 2007; Schneidman-Duhovny et al., 2005). Docking analysis showed the presence of different interactions like Hbond, Hydrophobic and Electrostatic interactions with TLR4 for the adjuvant as well as for whole vaccine. They showed negative interaction energy and a decrease in energy during the complex formation.

These lead us to the fact that the resulting structure is a stable complex. It is seen that the adjuvant forms bonds with both the chains

of the TLR4. The adjuvant specific for the aforementioned Toll-like receptor (TLR) is used to increase the immunogenicity of the vaccine. Large difference between the monomer area and complex area indicates that, while in the complex the solvent accessible area of all the residues present in the adjuvant and the linker is reduced. Thus it can be said that the complex formation leads to burying of those residues.

Proteasomal cleavage is important for presenting epitopes on the surface of antigen presenting cells to elicit an immune response. T cell epitopes express via proteasomal cleavage by specific proteases. Peptide cleavage server NetChop 3.1 results reported that the vaccine contains many proteasomal cleavage sites (Nielsen et al., 2005). Proteases like different classes of Cathepsin enzyme have some cleavage sites, which suggests that the designed vaccine is capable of producing different epitopes through the antigen presenting cell (APC) which serves the motive of vaccination perfectly.

Immunesimulation study showed that induction of B cell for both plasma and memory cell generation as well as induction of T cell was very prompt by the vaccine. Interleukin induction is necessary for any kind of cellular immunity and the vaccine satisfies the criteria having good induction potentiality. Antibody induction remains constant for a long time after the vaccine has been introduced. Memory B cell and

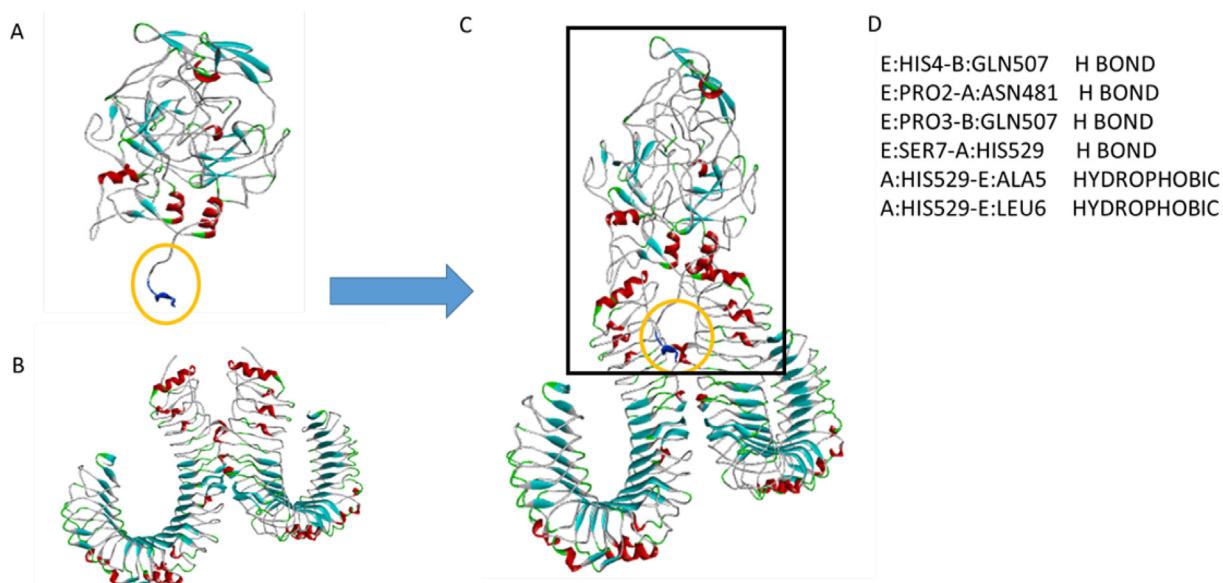


Fig. 4. Docking study analysis of vaccine with TLR4. (A) Simulated structure of the vaccine. (B) Structure of TLR4 peptide. (C) Docking structure of the vaccine with TLR4. Black bordered box assigned the vaccine and yellow circle assigned the adjuvant. (D) Type of bonds between TLR4 (Chain A, B) and adjuvant of the designed vaccine (Chain E). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

antigen presenting cells both show a rise in expression level, similarly, T_H and Tc cells increase in number which is required for immunity. It has been shown that both IFN- γ and IL-2 got induced within the tenure of 5–20 days which indicates prompt action of the vaccine. Simpson index (D) value helped to assume that it can stimulate a large and

diverse immune response. CAI value represents the codon usage in the formation of cDNA and it is 1 for this particular vaccine which is acceptable (> 0.8) (Morla et al., 2016). Predicted GC content suggests higher melting temperature and high quality PCR amplification. GC content outside the range of 30–70% can affect the expression both at

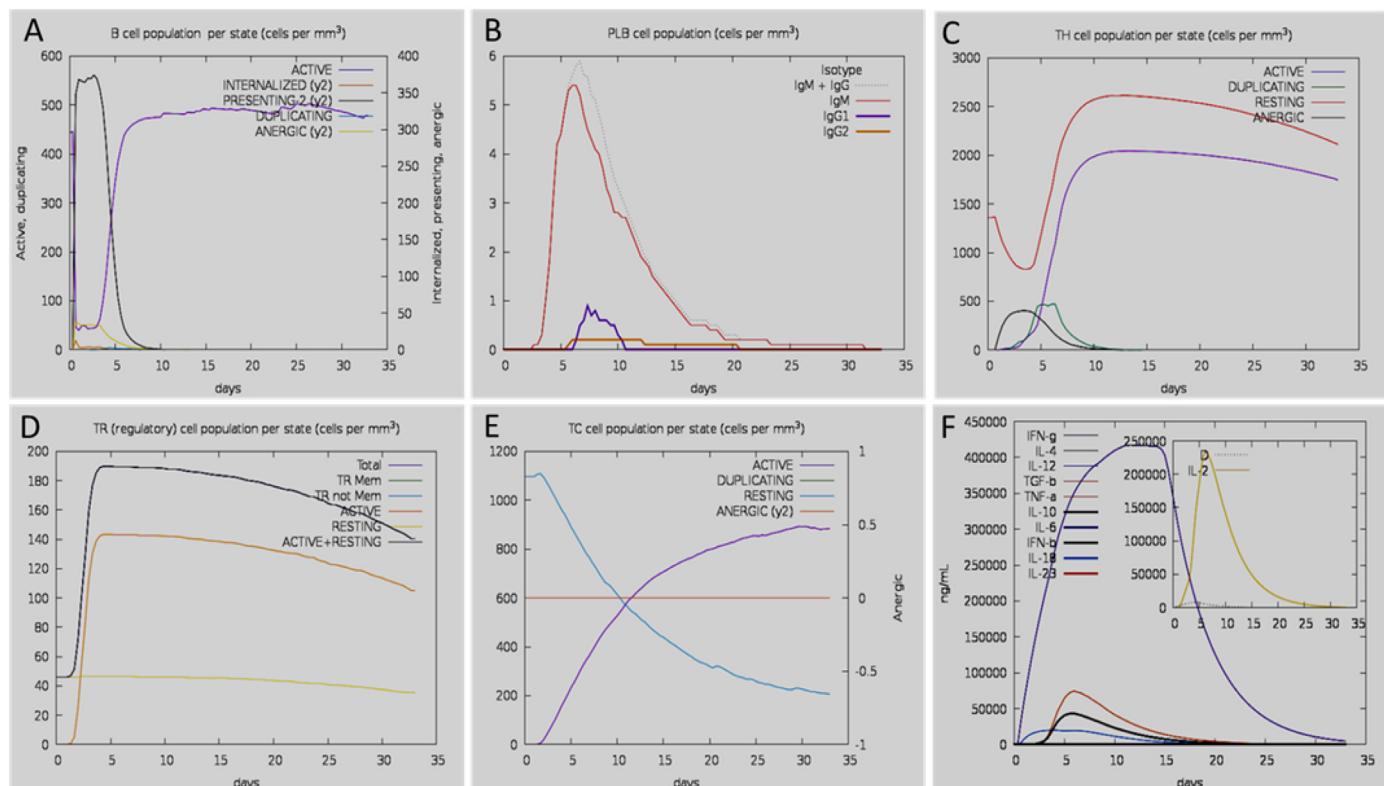


Fig. 5. Immune-simulation using C-ImmSim server. (A) B cell population after the induction of vaccine. Presenting cell count reached a maximum and then fell down from where the active cell count got a hike and saturated. (B) Both IgG and IgM reached their maxima at the stretch of 6–8 days after vaccine induction. (C) and (D) showed the population of TH and TR cell counts, represented almost same type of graphs having high count for active and resting cell population. (E) TC active cell show a different pattern of graph and that is because high active cell count is necessary for antigen presentation. (F) This graph showed the induction of interleukins. Both IFN- γ and IL-2 had an increase.

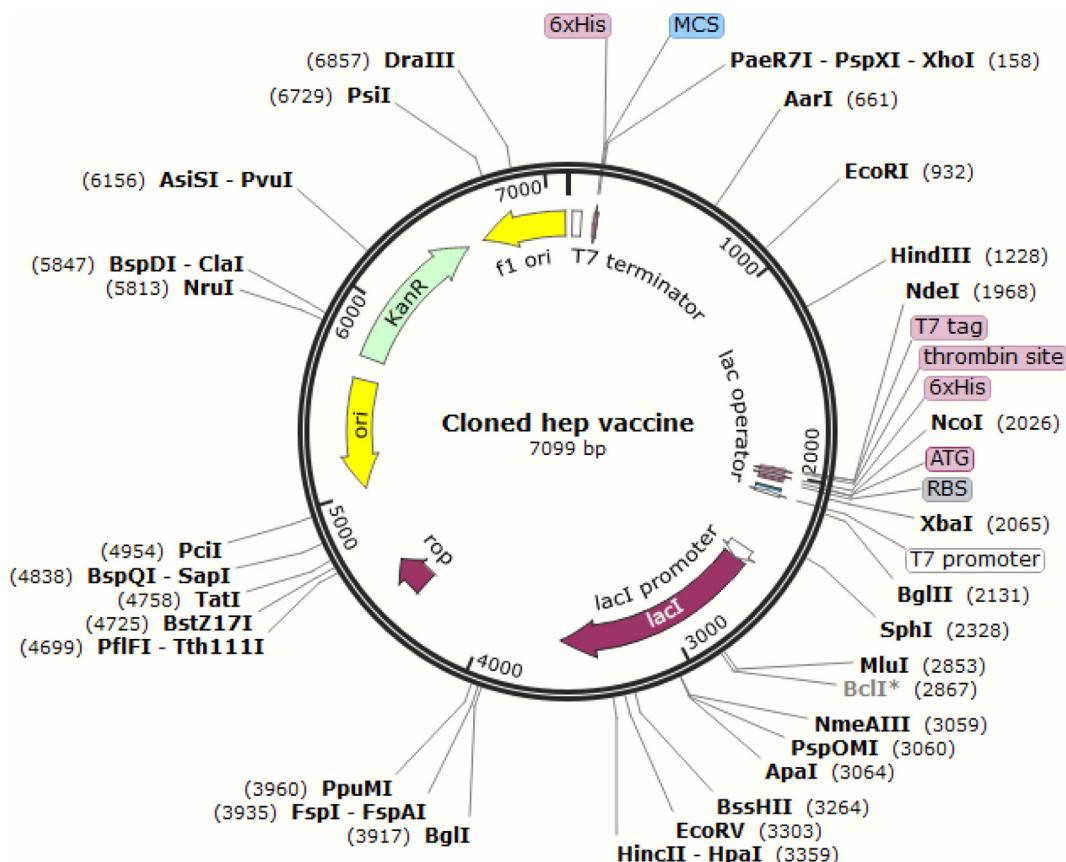


Fig. 6. *In silico* cloning of peptide based Hepatitis vaccine in pET28a + vector in between *Xba*I (N terminal residue 158) and *Nde*I (C terminal residue 1968) with N terminal His-tag.

transcriptional and translational levels(Ali et al., 2017; Shey et al., 2019).

This vaccine is based on peptides which are non-allergenic, have a conservancy of 75% across all sequences of the virus and raised in *E. coli* while other organisms can also be used for cloning by only adapting the codons for that particular organism. In case of *E. coli* production cost is less and expression is high. The presence of cleavage site after His tag also helps to get the vaccine in the native conformation. These properties would help to reduce the allergic response as shown by HAV vaccine, has no chance of going into the infectious cycle as there is no virus particle present, it will be effective for almost all, atleast 3 out of 4 HCV due to its high conservancy and the lower production cost may help the vaccine be accessible for all countries.

This study has been done with the help of servers where data of existing research have been taken into account while forming the database. Thus these predictions have some experimental basis though direct experimental evidences have not been evaluated till now. Thus this vaccine in spite of favoring all details may not be completely suitable for use in the current form and few modifications may be needed. The vaccine structure also has been determined and validated by MD simulation studies but the actual structure cannot be deduced until the vaccine is expressed and purified from an appropriate host for which we have selected *E. coli* and have deduced the gene sequence keeping the codon bias of that particular organism in mind and cloned it in the plasmid which would help in the expression of the protein. The vaccine was predicted to be localized in the periplasm where due to presence of less number of proteins, the purification would be easier and all bonds can be formed favorably due to absence of several enzymes. *In-vitro* validation and clinical trials will be needed in future for further characterization and modification of this epitope based vaccine.

5. Conclusion

Hepatitis has remained a severe problem for a long time. Though some medication has already been discovered, common vaccine for prevention of all types of Hepatitis is still not available. Our study is based on the *in-silico* designing of a peptide derived vaccine, covering all the epitopes of each Hepatitis virus. Both physico-chemical and structural studies supported a stable vaccine with the help of simulation and docking study. Immune-simulation study also suggested that the vaccine can induce good humoral and cellular immunity. Our study helps to identify the potential epitopes which could be used to construct a peptide based vaccine applicable for all 5 types of Hepatitis viruses.

Author contributions

Protocol designed and conceptualized by D.C., manuscript preparation and immune response related analysis done by J.D. and structural analysis done by S.M. Project was done under the supervision of K.G. The manuscript was reviewed and approved by all authors.

Data availability

The protein sequences can be retrieved from ViralZone (expasy.org/viralzone) and UniProt (uniprot.org) databases.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This study was supported by FRPDF grant of Presidency University from Government of West Bengal.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.meegid.2020.104388>.

References

- Abraham, M.J., et al., 2015. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25.
- Ali, M., et al., 2017. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci. Rep.* 7, 1–13.
- Altschul, S.F., et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Apweiler, R., 2009. The universal protein resource (UniProt) in 2010. *Nucleic Acids Res.* 38, 190–195.
- Arai, R., Ueda, H., Kitayama, A., Kamiya, N., Nagamune, T., 2001. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng. Des. Sel.* 14, 529–532.
- Bairoch, A., 1994. The SWISS-PROT Protein. 22, pp. 3578–3580.
- Bhasin, M., Raghava, G.P.S., 2004. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine* 22, 3195–3204.
- Buchan, D.W.A., Jones, D.T., 2019. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* 47, W402–W407.
- Cao, Y., Fang, X., Ottosson, J., Näslund, E., Stenberg, E., 2019. A comparative study of machine learning algorithms in predicting severe complications after bariatric surgery. *J. Clin. Med.* 8, 668.
- Chang, C.C.H., et al., 2016. Periscope: quantitative prediction of soluble protein expression in the periplasm of Escherichia coli. *Sci. Rep.* 6, 1–11.
- Chen, J., Liu, H., Yang, J., Chou, K.C., 2007. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids* 33, 423–428.
- Davies, M.N., Flower, D.R., 2007. Harnessing bioinformatics to discover new vaccines. *Drug Discov. Today* 12, 389–395.
- Dhanda, S.K., Vir, P., Raghava, G.P.S., 2013. Designing of interferon-gamma inducing MHC class-II binders. *Biol. Direct* 8, 1–15.
- Dhanda, S.K., et al., 2018. Development of a novel clustering tool for linear peptide sequences. *Immunology* 155, 331–345.
- Dimitrov, I., Garnev, P., Flower, D.R., Doytchinova, I., 2010. EpiTOP-a proteochemical tool for MHC class II binding prediction. *Bioinformatics* 26, 2066–2068.
- Dimitrov, I., Bangov, I., Flower, D.R., Doytchinova, I., 2014a. AllerTOP v.2 - a server for in silico prediction of allergens. *J. Mol. Model.* 20.
- Dimitrov, I., Naneva, L., Doytchinova, I., Bangov, I., 2014b. AllergenFP: Allergenicity prediction by descriptor fingerprints. *Bioinformatics* 30, 846–851.
- Doytchinova, I.A., Flower, D.R., 2007a. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 8, 1–7.
- Doytchinova, I.A., Flower, D.R., 2007b. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine* 25, 856–866.
- Doytchinova, I.A., Flower, D.R., 2010. Bioinformatic approach for identifying parasite and fungal candidate subunit vaccines. *Open Vaccine* 3, 22–26.
- Duhovny, D., Nussinov, R., Wolfson, H.J., 2002. Efficient Unbound Docking of Rigid Molecules. Springer, Berlin, Heidelberg, pp. 185–200. https://doi.org/10.1007/3-540-45784-4_14.
- El-Manzalawy, Y., Dobbs, D., Honavar, V., 2008a. Predicting linear B-cell epitopes using string kernels. *J. Mol. Recognit.* 21, 243–255.
- El-Manzalawy, Y., Dobbs, D., Honavar, V., 2008b. Predicting flexible length linear B-cell epitopes. *Comput. Syst. Bioinformatics Conf.* 7, 121–132.
- Essmann, U., et al., 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8593.
- Fieser, T.M., Tainer, J.A., Geysen, H.M., Houghten, R.A., Lerner, R.A., 1987. Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein alpha-helix. *Proc. Natl. Acad. Sci. U. S. A.* 84, 8568–8572.
- Gasteiger, E., et al., 2005. Protein analysis tools on the ExPASy server. *Proteomics Protoc. Handb. Protein Identif. Anal. Tools ExPASy Serv.* 571–607. <https://doi.org/10.1385/1592598900>.
- Gordon, A.D., Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and regression trees. *Biometrics* 40, 874.
- Grote, A., et al., 2005. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 33, 526–531.
- Gupta, S., et al., 2013. In Silico approach for predicting toxicity of peptides and proteins. *PLoS One* 8.
- Hebditch, M., Warwicker, J., 2019. Web-based display of protein surface and pH-dependent properties for assessing the developability of biotherapeutics. *Sci. Rep.* 9, 1–9.
- Hebditch, M., Carballo-Amador, M.A., Charonis, S., Curtis, R., Warwicker, J., 2017. Protein-sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* 33, 3098–3100.
- Hepatitis B. Available at: <https://www.who.int/news-room/fact-sheets/detail/hepatitis-b>. Accessed: 12th September 2019.
- Hepatitis C. Available at: <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c>. Accessed: 12th September 2019.
- Hepatitis E. Available at: <https://www.who.int/news-room/fact-sheets/detail/hepatitis-e>. Accessed: 12th September 2019.
- Hepatitis A. Available at: <https://www.who.int/news-room/fact-sheets/detail/hepatitis-a>. Accessed: 12th September 2019.
- Hepatitis D. Available at: <https://www.who.int/news-room/fact-sheets/detail/hepatitis-d>. Accessed: 12th September 2019.
- Huang, J., MacKerell, A.D., 2013. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* 34, 2135–2145.
- Hulo, C., et al., 2011. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.* 39, 576–582.
- Hung, G.-Y., Horng, J.-L., Yen, H.-J., Lee, C.-Y., Lin, L.-Y., 2015. Changing incidence patterns of hepatocellular carcinoma among age groups in Taiwan. *J. Hepatol.* 63, 1390–1396.
- Ito, M., et al., 2017. Encryption of agonistic motifs for TLR4 into artificial antigens augmented the maturation of antigen-presenting cells. *PLoS One* 12, 1–13.
- Jacobsen, K.H., Wiersma, S.T., 2010. Hepatitis a virus seroprevalence by age and world region, 1990 and 2005. *Vaccine* 28, 6653–6657.
- Jespersen, M.C., Peters, B., Nielsen, M., Marcatili, P., 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* 45, W24–W29.
- Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. 1 Edited by G. Von Heijne. *J. Mol. Biol.* 292, 195–202.
- Kahl, G., Linker, 2015. Dict. Genomics, Transcr. Proteomics 65, 25.
- Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580.
- Larsen, J.E.P., Lund, O., Nielsen, M., 2006. Improved method for predicting linear B-cell epitopes. *Immune Res.* 2 (2).
- Li, W., Joshi, M., Singhania, S., Ramsey, K., Murthy, A., 2014a. Peptide vaccine: progress and challenges. *Vaccines* 2, 515–536.
- Li, W., Joshi, M.D., Singhania, S., Ramsey, K.H., Murthy, A.K., 2014b. Peptide vaccine: progress and challenges. *Vaccines* 2, 515–536.
- Magnan, C.N., et al., 2010. High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics* 26, 2936–2943.
- Manavalan, B., Shin, T.H., Kim, M.O., Lee, G., 2018. AIPpred: sequence-based prediction of anti-inflammatory peptides using random forest. *Front. Pharmacol.* 9, 1–12.
- Masson, P., et al., 2013. ViralZone: recent updates to the virus knowledge resource. *Nucleic Acids Res.* 41, 579–583.
- Matsuda, S., et al., 2005. A novel representation of protein sequences for prediction of subcellular location using support vector machines. *Protein Sci.* 14, 2804–2813.
- Meunier, M., et al., 2016. Identification of novel vaccine candidates against campylobacter through reverse vaccinology. *J Immunol Res.* 2016.
- Moore, D. H. Classification and regression trees, by Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. Brooks/Cole publishing, Monterey, 1984, 358 pages, \$27.95. *Cytometry* 8, 534–535 (1987).
- Morla, S., Makhija, A., Kumar, S., 2016. Synonymous codon usage pattern in glycoprotein gene of rabies virus. *Gene* 584, 1–6.
- Nelson, D., Wang, J., 1992. Introduction to artificial neural systems. *Neurocomputing* 4, 328–330.
- Nielsen, M., Lundsgaard, C., Lund, O., Keşmir, C., 2005. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 57, 33–41.
- Pandey, R.K., Bhatt, T.K., Prajapati, V.K., 2018. Novel Immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating Anopheles Salivary Protein. *Sci. Rep.* 8, 1–11.
- Peters, B., Sette, A., 2005. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* 6, 1–9.
- Rapin, N., Lund, O., Bernaschi, M., Castiglione, F., 2010. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One* 5.
- Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5, 725–738.
- Saha, S., Raghava, G.P.S., 2006. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins Struct. Funct. Bioinforma.* 65, 40–48.
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., Wolfson, H.J., 2005. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 33, 363–367.
- Schönbach, C., Ranganathan, S., & Brusic, V. *Immunoinformatics*. Immunoinformatics9780387729, (2008).
- Shamugam, A., et al., 2012. Synthetic toll like receptor-4 (TLR-4) agonist peptides as a novel class of adjuvants. *PLoS One* 7.
- Shen, Bin, H., Chou, K.C., 2010. Virus-mploc: a fusion classifier for viral protein subcellular location prediction by incorporating multiple sites. *J. Biomol. Struct. Dyn.* 28, 175–186.
- Shey, R.A., et al., 2019. In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci. Rep.* 9, 1–18.
- Skwarczynski, M., Toth, I., 2016. Peptide-based synthetic vaccines. *Chem. Sci.* 7, 842–845.
- Söding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, 244–248.
- Solanki, V., Tiwari, M., Tiwari, V., 2019. Prioritization of potential vaccine targets using comparative proteomics and designing of the chimeric multi-epitope vaccine against

- pseudomonas aeruginosa. *Sci. Rep.* 9, 1–19.
- Sonnhammer, E.L.L., Krogh, A., 2008. A hidden Markov model for predicting trans-membrane helices in protein sequence. *Sixth Int. Conf. Intell. Syst. Mol. Biol.* 8.
- Tomar, N., De, R.K., 2010. Immunoinformatics: an integrated scenario. *Immunology* 131, 153–168.
- Vapnik, V.N., 1999. An overview of statistical learning theory. *IEEE Trans. Neural Networks* 10, 988–999.
- Venkatarajan, M., Braun, W., 2001. New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties. *J. Mol. Model.* 7, 445–453.
- Versputen, J., Gevaert, K., Declercq, W., Vandeneebele, P., 2009. SitePredicting the cleavage of proteinase substrates. *Trends Biochem. Sci.* 34, 319–323.
- Vita, R., et al., 2019. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.* 47, D339–D343.
- Wang, S., Li, W., Liu, S., Xu, J., 2016. RaptorX-property: a web server for protein structure property prediction. *Nucleic Acids Res.* 44, W430–W435.
- Wiesen, E., Diorditsa, S., Li, X., 2016. Progress towards hepatitis B prevention through vaccination in the Western Pacific, 1990–2014. *Vaccine* 34, 2855–2862.
- Wold, S., Jonsson, J., Sjöström, M., Sandberg, M., Ränna, S., 1993. DNA and peptide sequences and chemical processes multivariately modelled by principal component analysis and partial least-squares projections to latent structures. *Anal. Chim. Acta* 277, 239–253.
- Wu, S., Zhang, Y., 2007. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* 35, 3375–3382.
- Yang, J., Zhang, Y., 2015. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* 43, W174–W181.
- Yang, J., et al., 2014. The I-TASSER suite: protein structure and function prediction. *Nat. Methods* 12, 7–8.
- Yang, Y., et al., 2015. In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. *Hum. Vaccines Immunother.* 11, 795–805.