Molecular Immunology 93 (2018) 193–205

Contents lists available at ScienceDirect

Molecular Immunology

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

In silico prediction of potential vaccine candidates on capsid protein of human bocavirus 1

Narayanan Kalyanaraman*

Viral Research Diagnostic Laboratory (VRDL), Department of Microbiology, Govt Theni Medical College, Theni, Tamil Nadu, India

ARTICLE INFO

Keywords:
Docking
Epitope prediction
Human bocavirus
Human leukocyte antigen
Vaccine design

ABSTRACT

Human bocavirus 1 (HBoV1) is a newly identified parvovirus that causes serious respiratory infection among children across the globe. Aim of the present study was to predict immunogenic residues located on the VP2 protein of HBoV1 towards development of epitope based vaccines. Several computational tools were employed to predict epitopes (both T and B cell restricted) with stringent regulation for the improvement of confidence. After meticulous analysis, the peptide "TTPWTYFNFNQY" was identified as potential candidate for development of preventive vaccine. Of note, the epitope "TTPWTYFNFNQY" was found to be recognized by fifteen different alleles belonging to seven HLA supertypes (A1, A3, A24, A26, B7, B58 and B62). Further, mutational variability analysis pointed that most of the amino acids were well conserved. Docking scores obtained from ClusPro and Autodock Vina for selected epitopes displayed energetically favorable and stable interaction of peptide-HLA-I complexes. The core peptide "LLYQMPFFL" was found to recognize by wide range of HLA class II allele recognition thereby qualified as candidate for therapeutic vaccine. Five distinct linear peptides (with T cell epitope superimposition) belonging to B cells were identified in the VP2 protein. Further attention on the enlisted epitopes may shed light on the path for development of diagnostic, therapeutic and preventive tools against HBoV1 infection. Additionally, the predicted epitopes may help us to address the original antigenic sin phenomena observed during consecutive HBoV2-4 infection.

1. Introduction

New viruses capable to infect humans continuously emerge around every corner of the world. This imposes us to develop counter measures at the earliest for better clinical management and prevention. Presently, the easiest, safest and effective method for prevention of human diseases mainly infectious origin was vaccination. Conventional methods to design vaccine candidate is attributed as a laborious process with higher consumption of time and economy. Hence, computational approaches were often utilized nowadays for delineation of immunogenic residues that can function as a putative candidate.

Human bocavirus (HBoV), a newly identified virus that shares sequence similarities and genomic organization with bovine parvoviruses and minute virus of canines. HBoV belongs to the genus Bocaparvoviruses within the subfamily of Parvovirinae. So far, four species (HBoV1-4) have been identified with considerable rate of genetic heterogeneity and variability in clinical outcome of infection (Broccolo et al., 2015; Jartti et al., 2012; Kailasan et al., 2015; Schildgen et al., 2012). Epidemiological and observational evidences endorsed that HBoV1 causes serious respiratory infection among children and immunologically challenged individuals (Broccolo et al., 2015; Jartti et al., 2012). Further, observation of antibodies specific for HBoV1 even among healthy donors alarmed us to consider it as a possible concern during transfusion and transplantation procedure (Li et al., 2015a).

Development of vaccines or therapeutic measures often requires prior understanding on the immunological aspects during the natural course of an infection. The observation of strong $CD4^+$ T cell response among recovered individuals provided the first line of evidence for cellular response against HBoV1 infection (Lindner et al., 2008a). Later studies displayed that HBoV1 elicit typical virus-induced immune response involving both T_H1 and T_H2 cells (Deng et al., 2014; Jartti et al., 2012). Presence of long-lasting high-avidity IgG antibodies in human serum pointed the need of B cell mediated immunity against HBoV1 infection (Jartti et al., 2012; Lindner et al., 2008b). This led to understand that the selected vaccine candidate(s) of HBoV1 must possess the ability to elicit both arms of immune system.

Molecular works on HBoV1 disclosed that the viral genome was made of linear single stranded DNA (ssDNA) with a size of ~5.3 kb and consisted of three different Open Reading Frames (ORFs). The first and

* Communication address: Laboratory of Molecular Biology, Meenakshi Mission Hospital and Research Centre, Madurai, Tamil Nadu, India.
E-mail address: n.kalyanaraman@ymail.com.

https://doi.org/10.1016/j.molimm.2017.11.024
Received 1 August 2017; Received in revised form 18 November 2017; Accepted 26 November 2017
Available online 05 December 2017
0161-5890/

second ORF encodes for two non-structural proteins (NS1 and NP1) and third ORF encodes for two (VP1 and VP2) capsid proteins (Broccolo et al., 2015; Schildgen et al., 2012). Available studies concurred that the capsid protein (VP2) grasps the ability to induce both humoral and cellular immune response of the host (Lindner et al., 2008a; Deng et al., 2014). Hence, VP2 protein was often considered as a candidate for development of vaccine.

The concept of rationally designed epitope based vaccine was originated from our understanding on antigen recognition by both T and B lymphocytes (Patronov and Doytchinova, 2013; Testa and Philip, 2012). The Cytotoxic T lymphocytes (CTL) and Helper T lymphocytes (HTL) sees foreign antigen in the form of peptides that were presented with Major Histocompatibility Complex (MHC) which expressed on the surface of all nucleated cells. The MHC molecules are broadly classified in to two major types (MHC-I and MHC-II) based on tissue distribution pattern, source of antigen and responding T cells (Trowsdale, 2011). The influence of human MHC or Human Leukocyte Antigen (HLA) in clinical outcome of disease was well established beyond uncertainty.

The direct relationship between the diversity of HLA and repertoire of peptides presented by T cells that result in varied clinical manifestations and vaccine response (Ovsyannikova et al., 2004; Trowsdale, 2011). Till now, a total of 16,755 HLA alleles (Class I: ~ 12,351 + Class II: ~ 4404) have been identified across the world (IMGT/HLA statistic,[1] as on 04 July 2017). However, this much higher degree of HLA polymorphism also poses a threat to develop a vaccine with single epitope that can cover entire human ethnics and races. Hence, the concept of HLA supertype (similarity in peptide specificity among the HLA alleles) utilized to identify candidates that can confer wide range of population coverage (Sette and Sidney, 1998).

Nevertheless, experimental assays for identification of such promiscuous peptides were tedious process. Hence, *in silico* epitope prediction tools were often employed to determine potential candidates with the advantage of reduction in number of validation experiments and time (Nielsen et al., 2007; Yang and Yu, 2009).Presently, huge numbers of computational tools are available to predict peptides (T and B cell) with necessary properties (Yang and Yu, 2009). Algorithms based on binding motifs, Position Specific Scoring Matrices (PSSM), Artificial Neural Network (ANN) and Support Vector Machine (SVM) were often used to predict potential MHC binders. Of note, each method have its own merit and demerits which was discussed elsewhere (Bhasin and Raghava, 2004; Patronov and Doytchinova, 2013; Yang and Yu, 2009).

Computational approaches for the prediction of highly immunogenic epitope has been employed for viruses such as Chikungunya (Pratheek et al., 2015), Ebola (Srivastava et al., 2016), Influenza A (Staneková and Varečková, 2010) and Zika (Dikhit et al., 2016).Till now, no licensed vaccines or therapeutic peptides is available to prevent or treat HBoV1 infection. The present study aimed to predict immuno-dominant epitopes (T and B cell specific) located on the VP2 protein of HBoV1. In the following section, details of the strategy used for determination of T and B cell sites on VP2 protein were described. Subsequently, docking analysis was conducted to ensure the interaction between appropriate alleles and peptides.

2. Method

2.1. Amino acid sequence retrieval

The VP2 protein sequence of HBoV1 was retrieved from National Centre for Biotechnology Institute (NCBI) protein databank (Accession No.YP_338089.1) and used as an input for bioinformatics analysis.

[1] Website link for IMGT/HLA statistic https://www.ebi.ac.uk/ipd/imgt/hla/stats.html

2.2. Anitgenicity prediction

Antigenic nature of entire VP2 protein of HBoV1 and shortlisted peptides (predicted using various *in silico* tool) were evaluated using VaxiJen (v 2.0). The antigenicity of a given protein was predicted by VaxiJen with an accuracy rate of 70–89% (Doytchinova and Flower, 2007). The threshold value for being a probable antigen was set at 0.4 (ACC output).

2.3. Protein evaluation and modeling

The physiochemical properties such as molecular weight, isoelectric point (pI) value, iteration of amino acids within the protein, instability index, aliphatic index, estimated half-life extinction coefficient and grand average hydropathicity (GRAVY) were analyzed using the online tool Protparam (Gasteiger et al., 2005). Further, properties like solvent accessibility, transmembrane helices, globular region, bend region, random coil and coiled-coil region were determined with improved self-optimized prediction method (SOPMA) with default setting (Geourjon and Deleage, 1995).

2.3.1. Homology modeling and validation

The three dimensional structure of the selected protein was constructed using the Protein Homology/Analogy Recognition Engine, Phyre 2 (Kelley et al., 2015). The protein sequence was submitted in FASTA format and homology modeling was done under normal mode. Next, local structural distortions that generated during homology modeling were reduced using ModRefiner (Xu and Zhang, 2011). Structural refinement was performed several times until a model with minimal Root Mean Square Deviation (RMSD) value and maximal Template Modeling (TM) score was obtained. Accuracy and stereochemical nature of the constructed model was determined based on Ramchandran plot generated by PROCHECK (Ramachandran et al., 1963; Laskowski et al., 1996). The overall quality of the constructed model was evaluated with QMEAN6 (Qualitative Model Energy ANalysis) score which comprised of six different assessment available in Protein structure and model assessment tools of SWISS-MODEL work space (Arnold et al., 2006; Benkert et al., 2011) and Verify 3D (Eisenberg et al., 1997).

2.4. T cell epitope prediction

2.4.1. Preliminary enlistment of HLA class I ($CD8^+$ T cells) epitope

Peptides (nanomers) restricted to CTL was predicted (at supertype and allele specific level) using computational tools such as BIMAS-HLA, CTLpred, NetCTL, Propred-I and Immune Epitope Database Analysis (IEDB). The peptides recognized by HLA supertypes (A1, A3, A24, B7 and B40) and alleles (A*0201, B*3501, B*3701, B*5101, and B*5801) were predicted using ANN algorithm by BIMAS-HLA (Parker et al., 1994). Susequently, CTLpred a combined approach (ANN = 0.51, SVM = 0.36) was employed for peptide prediction and nHLApred was used to determine HLA restriction (Bhasin and Raghava, 2004). Further, NetCTL (version 1.2) was utilized to predict peptides recognized by twelve HLA supertypes (A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58 and B62) based on ANN algorithm. The cut-off for being a probable peptide was set at 0.75. Additionally, Propred-I which covers forty seven HLA Class I alleles was employed to predict peptides (without proteosomal filters) based on PSSM was employed. The threshold for being a probable antigenic peptide was set at 4.0 (Singh and Raghava, 2003). Then, top ten rank peptides predicted using BIMAS-HLA, CTLpred and Propred-I for the defined HLA supertype and allele were compiled to avoid reiteration and filtration. At this junction, peptides (at supertype level) tagged as epitopes by NetCTL was also included in primary analysis. Furthermore, top one percentile results of HLA Class I epitopes predicted using IEDB recommended consensus method with reference allele set (Kim et al., 2012) were included.

2.4.2. Immunogenicity and stability prediction

The antigenic and immunogenic propensity of the shortlisted peptides was predicted using VaxiJen and HLA Class I Immunogenicity prediction tool, respectively. Then, key parameters such as stability and affinity of the peptide-MHC-I (*p*-MHC-I) complexes were assessed. NetMHCstabpan (version 1.0) predicts epitopes for twelve HLA supertypes (A1, A2, A3, A24, A26, B7, B8. B27, B39, B44, B58 and B62) and binding affinity for appropriate *p*-MHC-I complexes using NetMHCpan (version 2.8).The cut-off value for being a potential and weak *p*-MHC-I complexes was set at 0.5 and 2.0, respectively (Nielsen et al., 2007). Further, the binding affinity of *p*-MHC-I complexes calculated in the subsequent segment (2.4.3) was also taken into account.

2.4.3. Promiscuity assessment

The ability of a peptide to bind with multiple alleles was often referred as promiscuity. Peptides that holds immunogenic and antigenic propensity were shortlisted for the evaluation of promiscuity. A total of 64 HLA alleles (17 HLA −A; 27 HLA-B and 20 HLA-C) were selected and peptide prediction was performed (for each loci separately) with IEDB Class I prediction tool using consensus (recommended) method. The list of selected 64 alleles was provided in the Supplementary workbook 1. Then, top four percentile results of IEDB Class I prediction was considered and alleles were enlisted based on the peptide binding affinity value. Here, binding affinity calculations depend on the Inhibitory Concentration 50 (IC50) value using ANN and Stabilized Matrix Method (SMM) algorithm. Allele specific cut-off value (Paul et al., 2013) was employed as threshold for the selection of HLA-A and HLA-B alleles. Further, a generalized threshold setting of IC50 value ≥ 250 nM was used for alleles that were not reported in the literature (Paul et al., 2013) and alleles belonging to HLA-C. Preference for MHC binding affinity values were given according to certain hierarchy (ANN + SMM > ANN > SMM). Moreover, overlapping peptide was clumped into single peptide to obtain better results while epitope conservation, population coverage and allergenicity evaluation.

The probability for being T cell epitope was assessed with MHC-I processing prediction tool an integrated approach which consist of proteasomal cleavage score, Transporter of Antigen Presentation (TAP) efficiency value and MHC Class I binding affinity (Kim et al., 2012). For consistency, sixty four alleles (previously selected) were considered for evaluation with NetMHCpan. The immuno-proteasomal filter was selected as cleavage prediction type. The maximum precursor extension was set at 1 with an alpha factor of 0.2 for TAP transport prediction. The peptide length was set at 9 and immunogenic parameters (Proteasome, TAP, MHC processing, and total score) were normalized on a scale of 0–1. Here, IC50 values for the MHC binding peptides were calculated based on SMM algorithm. Of note, the IC50 value calculated by MHC-I processing prediction tool was considered as threshold for alleles belonging to HLA-C.

2.4.4. HLA class II ($CD4^+$ T cells) specific peptide prediction

NetCTLIIpan (version 3.1) and MHC Class II processing tool (consensus method) was employed to predict peptides with strong binding affinity recognized by HTLs. Peptides with strong binding affinity (IC50 ≤ 50 nM) were selected and subjected to interleukin 4 (IL-4) and interferon γ (IFNγ) induction assessment (Dhanda et al., 2013a; Dhanda et al., 2013b). Peptides that displayed positivity for both parameters (IL-4 and IFNγ) were examined for the propensity of antigenicity and allergenicity. The promiscuous nature of the enlisted peptides was compiled based on top four percentile results of IEDB MHC Class II prediction. Finally, alleles that holds peptides with binding affinity value < 250 nM (according to the order ANN + SMM > ANN > SMM IC50 values) were considered for next level of evaluation. Additionally, peptides (nanomers) restricted to HLA Class II (20 alleles) predicted using Propred-II with a threshold setting of 3.0 were taken into consideration (top 10 peptides for each HLA allele).

2.4.5. Overlapping of HTL and CTL epitopes

The core peptides belonging to HTL were listed out to determine the residues that overlap with CTL restricted peptides. Preferences were given to core peptides predicted by combinatorial library method and compared with results generated from Propred-II. At this junction, peptides that possess overlapping amino acid sequences were merged to generate a single peptide fragment. Finally, the shortlisted peptides were subjected to epitope conservancy and population coverage.

2.5. B cell epitope prediction

2.5.1. Linear B cell epitope

Linear epitopes (20mers) limited by B cells were predicted using online tools such as BCpred (EL-Manzalawy et al., 2008), BCpred AAP (Chen et al., 2007), ABCpred (Saha and Raghava, 2006) and FBCpred (Yasser et al., 2008). Of note, filter for removal of overlapping peptides was applied and peptides with threshold value (≥90.0) were considered for subsequent analysis. Meanwhile, shortlisted peptides were manually verified to avoid reiteration and subjected to the evaluation of antigenicity and allergenicity. Additionally, linear epitopes of variable length predicted by Bepipred with a window size of seven at the threshold setting of 0.35 (Larsen et al., 2006) and ElliPro (Ponomarenko et al., 2008) were taken into account.

2.5.2. Conformational B cell epitope

Conformational (discontinuous) B cell epitopes were predicted using ElliPro and Discotope (version 2.0). ElliPro, uses a combination of geometric features and single amino acid propensity of an antigen for conformational epitope prediction (Ponomarenko et al., 2008). Discotope (v 2.0) predicts conformational epitopes on the basis of novel definition for the spatial neighborhood used to sum propensity scores and half-sphere exposure as a surface sphere (Kringelum et al., 2012).

2.5.3. B cell epitope properties evaluation

Fundamental properties such as antigenicity (Kolaskar and Tongaonkar, 1990), β-turn (Chou and Fasman, 1978), flexibility (Karplus and Schulz, 1985), hydrophilicity (Parker et al., 1986) and surface accessibility (Emini et al., 1985) of epitopes specific for B cells were evaluated. Since the window scale was seven for all parameters, a single peptide of 20 amino acids length could possess 14 possible fragments or positions. Therefore, consideration of an average for these 14 positions may collectively interpret the overall quality for being most probable B cell epitope. While scoring, positions within a peptide that failed to possess the basic property above appropriate threshold value were counted as zero. Thus, the rank of the basic attribute for each peptide varies from 0 to 14.

2.6. Epitope conservation analysis

Conservancy was often defined as the portion of protein sequences that restrain the epitope measured at or exceeding a specific level of identity (Bui et al., 2007). The conservancy of shortlisted T and B cell epitopes was assessed using the epitope conservancy tool (IEDB) with default threshold values. A total of 88 sequences matching with VP2 protein belonging to HBoV1 was retrieved from UniProt database. The filter for removal of duplicate sequences was applied and incomplete sequences were manually omitted to prevent bias when identifying conserved regions. Then, mutation variability of the peptides of VP2 was analyzed with freely available online tool, ConSurf (Ashkenazy et al., 2010). This server works on the basis of multiple sequence alignment with three dimension structural co-ordinate of protein and calculates bayesian conservation score (range 1–9), where lower score corresponds to extreme mutational variability and higher score correspond to maximum conservation.

2.7. Population coverage analysis

Population coverage analysis was performed using an online tool available in the IEDB analysis database. The primary objective was to determine the fraction of individuals respond to a given set of peptides with known MHC restrictions identity (Bui et al., 2007). The shortlisted epitopes with appropriate alleles (HLA Class I and core peptides of Class II) were used as input and analysis was performed for total world as well as individual major populations.

2.8. Allergenicity evaluation

AllerTop (version 2.0) predicts allergenicity on the basis of amino acid properties such as hydrophobicity, size, helix-forming propensity, relative abundance of amino acids and β-strand forming propensity. The proteins were classified according to k-nearest neighbor algorithm (kNN, k = 1) based on training set containing 2427 known allergens from different species and 2427 non-allergens(Dimitrov et al., 2014).

2.9. Peptide-HLA docking analysis

2.9.1. Preparation of peptide and HLA structure for molecular docking

The three dimensional structure of the finalized epitopes were constructed using PEP-FOLD3. It works on the basis of the structural alphabets (SA) letters to explain the structural conformation of four consecutive amino acid residues coupled with a series of SA greedy algorithm and coarse-grained force field (Lamiable et al., 2016). The best model predicted by PEP-FOLD3 was downloaded in PDB file format. The protein sequences of the shortlisted HLA alleles were retrieved from the IMGT/HLA database. Then, three dimensional structure of these alleles were constructed using the webserver Phyre2; structural refinement was performed using ModRefiner (refer section 2.3.1) and subsequently subjected to PROCHECK, Verify 3D and QMEAN6 assessments.

2.9.2. Molecular docking using ClusPro

ClusPro (v.2), a web server performs blind molecular docking through a rigid PIPER docking program based on Fast Fourier Transform (FFT) followed by filtering and clustering of docked confrontation using pairwise RMSD subsequently stabilization using Monte Carlo simulations(Kozakov et al., 2010).The PIPER docking program incorporated within the ClusPro calculates the binding energy score based on attributes such as shape complementarity, electrostatics, desolvation contribution and Deocys as reference states (DARS). Docking was performed under hydrophobic environment and optimal model was selected based on RMSD value, peptide location on the binding groove and size of the cluster. The optimal model was visualized using UCSF Chimera version 1.11 (Pettersen et al., 2004).

2.9.3. Molecular docking using AutoDockVina (v 1.1.2)

AutoDockVina (v.1.2), an open-source program often used to study molecular interactions between two protein molecules (Trott and Olson, 2010). The input files were prepared with dockprep option available in UCSF Chimera (version 1.11). Addition of hydrogen, water molecules and removal of incomplete side chain on the constructed peptide and allele model were performed. The grid (centre) and size dimensions set was adjusted in a manner to cover the peptide binding grooves of the HLA moiety with accommodation for free movement. The number of binding modes, exhaustiveness parameter, energy (kcal/mol) was kept at the maximum limit 10, 8 and 3, respectively. The best output was selected on the basis of binding energy and visualized using UCSF Chimera version 1.11 (Pettersen et al., 2004).

2.10. Active site analysis

The active binding sites of the pVP2 protein was evaluated using the web server CASTp (Computed Atlas of Surface Topography of protein).The refined structure of the VP2 protein (PDB format) was provided as input file in CATSp server. The active binding sites, surface structural pockets, area, shape and volume of every pocket and internal cavities of the proteins was assessed. The active site that displayed large volume and area was considered for further comparison (Dundas et al., 2006).

2.11. Peptide match for prediction of autoimmunity

The possibility to elicit autoimmune reaction has to be addressed before selection of peptide as a vaccine candidate. For this, Peptide Match Service an online tool available in Protein Information Resource (PIR) was employed to predict the presence of the selected candidate peptides in human proteome (Chen et al., 2013). Homo sapiens were selected as the target organism and the limitation for UniRef100 representative sequences within the UniProtKB was applied.

3. Results and discussion

Our understanding on immunodominant epitopes located on the VP2 protein of HBoV1 remains elusive. Hence the present study was undertaken to delineate immunogenic residues that can function as candidate(s) for development of epitope based vaccine. Several computational tools were employed in the present investigation to narrow down the putative list of T and B cell epitopes. Further, preferences were given to peptides predicted by consensus method for the enhancement of sensitivity and specificity.

3.1. Anitgenicity prediction

The overall anitgenicity (0.505) suggested VP2 protein as a probable antigen of HBoV1. Further, Kolaskar and Tongaonkar anitgenicity reported that nearly fifty percent of residues present in the VP2 protein hold the property to elicit immune response of the host. These results provide a partial explanation for the observation of copious amount of VP2 specific antibodies in the peripheral blood stream of infected individuals (Jartti et al., 2012; Lindner et al., 2008a).

3.2. Primary and secondary structure analysis

The basic properties of a protein often play a major role in development of host immune response. Protoparam analysis displayed that the VP2 protein had a molecular weight of 60539.6 Da and theoretical isoelectric point (pI) of 5.97. An isoelectric point below 7 suggested that the VP2 protein was negatively charged. The total number of negatively charged residues (Asp + Glu) and positively charged residues (Arg + Lys) were 51 and 44 respectively. The aliphatic index (60.65), instability index (34.19) and negative GRAVY value (-0.595) indicated that the VP2 protein was stable, thermostable and hydrophobic in nature, respectively. The estimated half-life of VP2 protein in mammalian reticulocytes, yeast and *E.coli* was 30 (*in-vitro*), > 20 (*in-vivo*) and > 10 h (*in-vivo*), respectively.

Analysis of secondary structures with SOPMA revealed that the target protein was dominated by random coil (51.11%), followed by extended strand (27.86%), beta turn (10.89%) and alpha helix (10.15%). The occurrence of higher proportion of random coil and extended strand suggested that the residues located in these region increases the likelihood of forming an antigenic epitope. Further, the richness of coiled region indicated that the VP2 protein was well conserved evolutionarily. These findings collectively inferred that VP2 protein of HBoV1 could be most probable candidate to design and develop vaccine.

The structure of VP2 protein was constructed based on the VP2 protein of pathogenic bovine parvoviruses (PDB ID: C4QC8) which displayed 49% of identity with 100% confidence. Structural refinement

Table 1
Results of Ramchandran plot for VP2 protein and selected HLA alleles.

Protein	Accession	Ramachandran plot	ModRefinement		QMEAN6		Verify 3D
		MFR (%)	RMSD	TM	raw	Z	%
VP2	YP_338089.1	86.0	0.247	0.999	0.421	−4.22	72.24
A*24:02	AAB87053.1	93.1	0.136	0.999	0.635	−1.34	95.56
A*29:02	CAQ16345.1	92.7	0.149	0.9994	0.689	−0.91	98.92
B*07:02	AAA59622.1	93.7	0.163	0.9989	0.654	−1.14	100
B*08:01	AAA52662.1	93.0	0.116	0.9996	0.781	0.09	98.92
B*15:01	AAL30409.1	93.7	0.183	0.9987	0.614	−1.56	100
B*35:01	AAL30411.1	94.3	0.147	0.999	0.773	0.08	100
B*58:01	AAB82306.1	91.8	0.095	0.9996	0.612	−1.57	100
C*03:02	AAA59686.1	93.8	0.12	0.9996	0.787	0.16	96.4

Note: VP2–Viral protein 2; MFR − Most Favored Region; % − percentage; Root Mean Square Deviation (RMSD) value; TM − maximal Template Modeling; QMEAN − Qualitative Model Energy ANalysis.

(three times) was performed using ModRefiner and a final model of VP2 protein was obtained with minimal RMSD (0.247) value and maximal TM (0.999) score (Table 1). The quality and reliability of the constructed VP2 model was evaluated with PROCHECK, Verify3D and QMEAN6 scores. PROCHECK measures the stereochemical quality of a protein by analyzing residue-by-residue and structural geometry features. The congregation of majority of restudies (86.7%) under favorable region in the Ramachandran plot pointed that the constructed model was good for further analysis (Table 1, Supplementary workbook 1). Verify 3D analysis displayed that 72.24% of residues possess an average of 3D-1D score ≥ 0.2 and QMEAN6 score of VP2 protein was 0.421 (Table 1). Results obtained from Ramachandran plot, Verify 3D and QMEAN6 score (Table 1, Supplementary workbook 1) collectively indicated that the constructed model can be used as a template to improve our understanding in terms of structure and function of the VP2 protein, drug designing and protein–protein interactions.

3.3. T cell epitope prediction

The rationality of T cell epitope based vaccine majorly depends on the consistency of peptide prediction. Hence, confidence in present study was substantiated by employment of several computational tools with stringent regulatory parameters. Literature evidence suggested that allele specific affinity rather than generalized threshold value (IC50) of 500 nM provided better prediction for peptide-allele binding affinity (Paul et al., 2013). Nevertheless, integration of both peptide binding affinity and stability of *p*-MHC-I complex conferred superior result than any of the single method (Jørgensen et al., 2014). Therefore, allele specific affinity and stability of the appropriate pMHC-I complex was combined to obtain appropriate peptide-HLA combinations. Despite of peptides with stronger binding values, peptides with strong immunogenicity presumably functions as CTL epitopes than those with weaker immunogenicity. Hence, immunogenicity of the selected peptides was also considered as a key parameter for peptide enlistment.

3.3.1. HLA class I epitope prediction

As per the criteria (section 2.4.1), a total of 76, 10, 102, 170 and 57 distinct peptides was predicted by BIMAS-HLA, CTLpred, NetCTL, Propred I and IEDB (top 1 percentile), respectively. Here, preference was given to peptides predicted by at-least three or more different tools. A total of 59 out of 194 distinct peptides were scrutinized for the assessment of immunogenic and antigenic propensity (Supplementary workbook 2). Finally, a total of 16 peptides were found to be potential candidate for downstream analysis. The consolidated results of the enlisted peptides with appropriate supertypes and relevant parameters were provided in Table 2.

The complete results of MHC class I binding and MHC-I processing prediction tools were provided in the Supplementary workbook 2. Potential MHC binders were identified with strict adherence to threshold setting of IC cut-off value as explained previously in methodology (section 2.4.3). Peptides with overlapping sequences were congregate in to single epitope for the improvement of maximum population coverage across the globe. Subsequently, eight epitopes were identified as most probable candidates without allergenicity (Table 3).

Table 2
List of supertype specific peptides predicted from VP2 protein of HBoV1 using different approaches.

S.No	Peptide	Methods	Supertypes	Immunogenicity	VaxiJen
1	VPTRRVQYI	1,2,3,4,5	A11, B7, B8, B14	0.05	1.31
2	WTYFNFNQY	1,2,3,4,5	A1, A3, A26, B7, B58, B62	0.14	0.8
3	CLAPTNLEY	1,3,4,5	A1, A3,A26,B58, B62	0.08	2.31
4	DPFDGSIAM	1,3,4,5	B7, B62	0.07	0.47
5	FPITRENPI	1,3,4,5	B7, B8, B14,	0.21	0.84
6	FTFNFDCEW	1,3,4,5	B7, B58	0.17	1.96
7	GYIPIENEL	1,3,4,5	A24, B39	0.3	0.6
8	RPERRHTAL	1,3,4,5	A24, B7, B8, B27, B39	0.22	1.29
9	SSDTAPFMV	1,3,4,5	A1	0.07	0.46
10	TTPWTYFNF	1,3,4,5	A1, A24,A26, B58	0.34	1.6
11	ASTGRIQPY	1,3,4	A1, B62	0.11	1.1702
12	ESTEFTFNF	1,3,4	A26, B58	0.35	1.6
13	HVDPTGAYI	1,3,4	A1	0.09	0.53
14	LSNGADTTY	3,4,5	A1, B58, B62	0.14	0.67
15	TEFTFNFDC	1,3,4	B44	0.29	1.75
16	YKRFRPKAM	3,4,5	B7, B8	0.01	0.83

Note: 1 − BIMAS-HLA; 2–CTLpred; 3 − Net CTL; 4 − Propred I; 5 − IEDB consensus. For additional information on immunogenicity and VaxiJen readers are requested to refer appropriate reference provided in the manuscript.

Results generated from NetMHCstabpan demonstrated that few peptides displayed stronger interaction with several HLA Class I alleles (Table 3). In general, lower IC50 values (< 50 nM) HC-I complex inferred the ability of the peptide to induce effective immune response even at lower dosage. Further, peptides with higher stability or longer half-lives hold better chances to encounter appropriate CTL clone when compared to pMHC-I complexes with shorter half-lives (Jørgensen et al., 2014). For instance, the peptide “WRPERRTHAL” displayed more stable and longer duration (11.57 h) of interaction with HLA-B*07:02 (Table 3).

Of note, magnitude of immune response mostly depends on a cluster of tightly overlapping epitopes (polyclonal T cell response) rather than single epitope (Berzofsky et al., 1987). Consolidate results of *in silico* analysis depicted that the peptide “TTPWTYFNFQY” grasped the ability to induce polyclonal T cell response through epitope overlapping phenomena. The combination of wide recognition by seven major HLA supertypes (Table 2), strong binding affinity and higher stability with several alleles (Table 3) provide substantial evidence to the peptide “TTPWTYFNFNQY” as potential candidate. Further, observation of

Table 3
List of putative epitopes along with interacting HLA class I alleles, IC50 values and relevant scores.

S.No	Peptide	Allele	Method of prediction	MHC I binding and processing prediction scores					NetMHC stab
				ANN IC50	SMM IC50	Proteasome	TAP	MHC IC50	Half life (hrs)
1	VPTRRVQYI	B*08:01	a	801	613.1	1.21	0.15	380.36	0.25
		B*51:01	a	791	936.7	1.21	0.15	1234.1	
		B*07:02	a	631	551.9	1.21	0.15	530.18	0.79
2	TTPWTYFNFNQY	A*26:01	b	153	474.9	1.41	1.27	58.66	0.79
		A*29:02	b	4	32.2	1.41	1.27	5.8	
		A*01:01	b	64	356.9	1.41	1.27	234.5	0.67
		A*25:01	b	150	927.0	1.41	1.27	288.7	
		A*68:01	b	50	220.5	1.41	1.27	77.3	
		A*30:02	b	42	92.3	1.41	1.27	46.0	
		B*35:01	a	172	187.7	1.41	1.27	58.6	
		B*15:01	a	113	222.8	1.41	1.27	116.4	2.62
		C*16:01	c	–	–	1.41	1.27	127.1	
		C*03:02	c	–	–	1.41	1.27	29.67	
		C*12:02	c	–	–	1.41	1.27	200.35	
		C*14:02	c	–	–	1.41	1.27	229.0	
		C*12:03	c	–	–	1.41	1.27	28.6	
		A*23:01	b	68	243.6	1.39	1.09	23.3	
		A*24:02	b	23	216.5	1.39	1.09	29.6	1.5
3	DPFDGSIAM	B*35:01	a	2	1.5	1.12	−0.07	3.9	
		B*35:03	b	1067	356.6	1.12	−0.07	48.0	
		B*18:01	b	313	402.7	1.12	−0.07	1338.3	
		B*39:01	b	120	969.1	1.12	−0.07	370.2	0.35
		B*15:02	b	3811	45.9	1.12	−0.07	675.9	
4	FTFNFDCEWV	A*32:01	a	194	95.7	1.45	0.31	70.5	
		B*58:01	a	11	26.6	1.45	0.31	3.4	5.53
		B*53:01	a	273	216.7	1.45	0.31	38.8	
		B*57:01	b	20	45.0	1.45	0.31	12.4	
		C*03:02	c	–	–	1.45	0.31	43.1	
		C*12:02	c	–	–	1.45	0.31	230.2	
5	WRPERRHTAL	B*07:02	a	7	4.3	1.44	0.3	2.5	11.57
		B*08:01	a	11	21.1	1.44	0.3	24.3	0.33
		B*14:02	b	659	39.6	1.44	0.3	626.7	
		C*07:01	b	375	29.0	1.02	−0.1	1028.2	
6	ASTGRIQPY	A*30:02	b	812	191.1	0.94	1.4	84.9	
		B*15:01	a	214	291.0	0.94	1.4	201.7	5.31
		B*40:02	b	2367	373.9	1.13	−0.7	5192.8	
		C*03:02	c	–	–	0.94	1.4	139.2	
		C*12:03	c	–	–	0.94	1.4	209.4	
7	ESTEFTFNFDC	B*40:02	b	40	246.5	0.73	0.01	1292.0	
		B*40:01	b	291	172.1	0.73	0.01	4033.9	
		B*44:03	b	1610	699.0	0.73	0.01	3043.0	
8	YKRFRPKAM	B*08:01	a	21	57.48	0.96	0.09	33.3	0.84
		B*14:02	b	649	168.2	0.96	0.09	747.31	

Note: a − Consensus (ANN/Comblib_sidney2008/SMM); b − Consensus (ANN/SMM); c − netMHCpan; ANN − Artificial Neural Network; SMM − Stabilized Matrix Method; IC50–Inhibitory concentration 50; TAP − Transporter of Antigen Presentation; MHC − Major Histocompatibility Complex; Hrs − hours.

positive scores (proteasomal cleavage and TAP efficiency) with several alleles also shed light on the immunogenic potential of the peptide "TTPWTYFNFNQY".

3.3.2. HLA class II (CD4) restricted eptiopes

NetCTLIIpan (version 3.1) was used to predict peptides recognized by HLA Class II alleles. A total of 80 peptides were listed out based on calculated MHC binding affinity value (IC50 ≤ 50 nM). Further, analyses (IL-4 and IFNγ inducing propensity) tagged 12 peptides (Table 4) as candidates for subsequent comparison with peptides predicted by IEDB HLA Class II tool. At this junction, top four percentile ranks of peptides predicted using IEDB MHC II tool were consolidated to improve the consistency, accuracy and allele coverage of putative candidates (Supplementary workbook 3). A total of 27 unique peptides were found to consist of both IL-4 and IFNγ inducing propensity (Table 4). Moreover, a total of 59 different peptides (nanomers) predicted using Propred-II (Supplementary workbook 3) was compared for core peptide identification.

After synchronization, peptides were manually verified to avoid reiteration and the results of each tool were not discussed in detail due to space constraint. A total of 34 discrete peptides (combined results of NetCTLpanII and IEDB) were subjected to filtration based on allergenicity and antigenicity parameter. Ten of the 34 peptides were identified as probable candidates with exceptional qualities for being HLA Class II epitope (Table 4). The combination of lower binding affinity (IC50 ≤ 50 nM) and intricate propensity (IL-4 and IFNγ) enhances the confidence level on the predicted top ten peptides (Table 4). However, functional experiments are mandatory to reassure the quality of the predicted $CD4^+$T cell restricted epitopes. Inclusion of HLA class II restricted epitopes was recommended in vaccine formulation due to the cognate support provided by HTLs which leads to mount vigorous immune response with optimal CTL response and production of neutralizing antibodies (Rosa et al., 2010). Therefore, HLA class II epitope was considered in the present investigation to improve the quality of vaccine candidate.

Table 4
List of putative HLA class II peptides along with relevant scores.

S.No	Peptide	Method of prediction	Core peptide	IL-4 score	IFNγ Score	Conservancy (%)	VaxiJen
1*	ALLYQMPFFLLENSD	1	LLYQMPFFL	0.27	1.0[a]	81.82	0.47
2*	EFTFNFDCEWVNNER	2	EFTFNFDCE	0.34	6.0[a]	80.68	1.46
3*	LLYQMPFFLLENSDH	1, 2	LLYQMPFFL	0.28	1.0[a]	81.82	0.45
4*	LYQMPFFLLENSDHQ	2	LYQMPFFLL	0.28	1.0[a]	81.82	0.61
5*	NFDCEWVNNERAYIP	2	WVNNERAYI	0.31	6.0[a]	81.82	0.94
6*	PTYHVDPTGAYIQPT	2	YHVDPTGAY	0.26	3.0[a]	72.73	0.96
7*	SNYTPTYHVDPTGAY	2	YHVDPTGAY	0.24	0.32[b]	73.86	0.73
8*	TPTYHVDPTGAYIQP	2	YHVDPTGAY	0.27	5.0[a]	73.86	0.68
9*	TYHVDPTGAYIQPTS	2	YHVDPTGAY	0.27	0.35[b]	72.73	1.009
10*	YHVDPTGAYIQPTSY	2	YHVDPTGAY	0.26	0.34[b]	72.73	1.088
11*	ATEKALLYQMPFFLL	1	LLYQMPFFL	0.27	0.79[b]	ND	0.31
12*	DGSIAMDHPPGTIFI	2	IAMDHPPGT	0.27	1.0[a]	ND	0.17
13*	DPFDGSIAMDHPPGT	2	FDGSIAMDH	0.28	1.0[a]	ND	0.25
14*	IWMFPNQVWDRFPIT	2	QVWDRFPIT	0.27	0.015	ND	0.13
15*	NATEKALLYQMPFFL	1	LLYQMPFFL	0.25	0.49	ND	0.25
16*	PFDGSIAMDHPPGTI	2	IAMDHPPGT	0.27	1.0[a]	ND	0.19
17*	TEKALLYQMPFFLLE	1	LLYQMPFFL	0.27	1.0[a]	ND	0.29
18*	RRVQYIRQNGSTAAS	1	IRQNGSTAA	0.27	2.0[a]	ND	0.07
19	EKALLYQMPFFLLEN	1	LLYQMPFFL	0.27	1.0[a]	ND	0.21
20	ESTEFTFNFDCEWVN	2	EFTFNFDCE	0.34	6.0[a]	ND	1.64
21	FDGSIAMDHPPGTIF	2	IAMDHPPGT	0.27	1.0[a]	ND	0.04
22	FFLLENSDHQVLRTG	2	FFLLENSDH	0.27	1.0[a]	ND	0.21
23	FNFDCEWVNNERAYI	2	FDCEWVNNE	0.3	6.0[a]	ND	1.21
24	GESTEFTFNFDCEWV	2	FTFNFDCEW	0.34	6.0[a]	ND	1.62
25	KALLYQMPFFLLENS	1, 2	LLYQMPFFL	0.27	1.0[a]	ND	0.38
26	MPFFLLENSDHQVLR	1, 2	FFLLENSDH	0.27	1.0[a]	ND	0.17
27	NYTPTYHVDPTGAYI	2	YHVDPTGAY	0.27	0.25[b]	ND	0.43
28	PFFLLENSDHQVLRT	1, 2	FFLLENSDH	0.27	1.0[a]	ND	0.05
29	QMPFFLLENSDHQVL	1, 2	FFLLENSDH	0.28	1.0[a]	ND	0.43
30	QVWDRFPITRENPIW	2	FPITRENPI	0.27	0.56[b]	ND	0.102
31	STEFTFNFDCEWVNN	2	FTFNFDCEW	0.35	6.0[a]	ND	1.499
32	TEFTFNFDCEWVNNE	2	FTFNFDCEW	0.34	6.0[a]	ND	1.509
33	YQMPFFLLENSDHQV	1, 2	FFLLENSDH	0.28	1.0[a]	ND	0.54
34	YTPTYHVDPTGAYIQ	2	YHVDPTGAY	0.26	5.0[a]	ND	0.53

Note: * identified as non-allergen; 1 – NetMHCIIpan; 2 – IEDB consensus; IL – Interleukin; all IL scores are derived from SVM; IFNγ – Interferon gamma; IFN score – [a] MERCI, [b] support vector machine; ND – not done; % – percentage; Amino acids in Bold (ConSurf score – 8), Bold and Underlined (ConSurf score – 9).

Table 5
Conservancy and population coverage conferred by potential T cell epitopes.

		Population*											
Epitope	Conservancy*	World#	Amerindian	Arab	Asian	Black	Caucasoid	Hispanic	Jews	Oriental	Persian	Polynesian	Siberian
VPTRRVQYI	83.6.	29.13	8.59	31.3	16.08	20.79	45.52	28.12	4.72	11.86	31.44	13	26
TTPWTYFNFNQY	63.6	72.11	98.58	76.8	95.97	87.56	99.27	75.2	76.95	78.29	61.4	95.2	100
DPFDGSIAM	87.5	23.16	22.17	15.6	20.18	17.73	28.93	23.71	24.74	18.83	29.78	29.2	16
FTFNFDCEWV	81.8	19.41	15.4	49.5	55.58	66.07	49.13	70.33	61.98	51.51	73.22	28.7	56
WRPERRHTAL	82.9	39.7	7.76	33.3	14.76	37.9	56.41	39.89	23.39	8.41	32.06	4.28	19
ASTGRIQPY	79.5	25.21	17.47	22.9	25.09	19.9	27.69	27.2	9.28	30.82	21.87	28.2	36
ESTEFTFNFDC	90.9	17.03	16.02	11	20.16	10.64	19.67	23.02	13.09	24.94	4.94	29	31
YKRFRPKAM	69.3	13.28	2.67	12.3	3.53	10.62	21.43	15.73	6.47	0.71	17.37	0	1.6
Epitope set	NA	90.13*	99.24	95.58	99.07	98.34	99.95	98.18	95.06	94.39	96.25	98.49	99.93

Note: Amino acids in Bold (ConSurf score – 8), Bold and Underlined (ConSurf score – 9); * – values were expressed in percentage (%); # – 20 different populations.

3.3.3. Overlapping $CD4^+$ and $CD8^+$ t cell epitope

Peptides restricted by CTLs often fall in the range of 9–12 amino acids (predominantly nanomer) whereas no such limitations were observed for HTLs. However, a core peptide (nanomer) often accounted for most of the recognition of HTL restricted peptides (Rosa et al., 2010). Therefore, core peptide sequences defined while HTL peptide prediction was utilized to determine the superimposition with CTL epitopes. This resulted in identification of two peptides "TTPWYFNFNQY" and "FTFNFDCEWV" that possess the core peptides restricted by $CD4^+$T cells and $CD8^+$T cells (Supplementary workbook 3). These findings clearly indicated the superiority of the peptides ("TTPWYFNFNQY" and "FTFNFDCEWV") vaccine candidacy over other peptides provided in Table 3.

Of note, the core peptide "LLYQMPFFL" was consistently identified as a part of several peptides restricted to HTLs (Table 4). However, it failed to qualify as putative candidate of CTL epitope due to poor immunogenicity (Supplementary Work book 2). At the outset, it led to the assumption that the core peptide "LLYQMPFFL" could be considered as a vaccine candidate for therapeutic purpose only. But, under the influence of HTL mediated response it may aid CTL to mount sustained immune response.

3.4. Epitope conservancy and population coverage analysis

The effectiveness of a vaccine program solely depends upon the level of conservation and population coverage conferred by the selected epitopes. Epitopes and HLA alleles (Class I and core peptides of Class II) selected for population coverage were provided in the Supplementary

workbook 1. Conservation analysis (IEDB) displayed that most of the epitopes were well conserved among available VP2 protein sequences of HBoV1 (Table 5). Even though the peptide "ESTEFTFNFDC" displayed maximum conservation (90.9%), lack of promiscuity (Table 3) resulted in very poor rate (17.03%) of global population coverage (Table 5).

Astonishingly, entire epitope set activate T cell response of major populations that resulted in 90.1% coverage across the globe (Table 5). It indicated that development of a muti-epitopic vaccine may provide better protection *via* the concept of immunodominance in cellular immunity. The broad allele specificity (fifteen alleles of HLA Class I) of the epitope "TTPWTYFNFNQY" contributed for the observation of excellent worldwide coverage (Tables 3 and 5). Intriguingly, mutation variability analysis (using ConSurf) of the peptide "TTPWTYFNFNQY" displayed that most of the amino acids (underlined) were highly conserved (Table 5).

Of note, the epitope "TTPWTYFNFNQY" provided maximum coverage for major populations such as Asian, Blacks and Caucasoid (Table 5). Several African and Asian countries (Bangladesh, China, India and Pakistan) have the possibility to hold higher prevalence and incidence of HBoV1 infection due to limited medical diagnostic facilities and over-crowding, respectively. This implies that there is a possibility to develop regional specific vaccine for containment of HBoV1 infection. In nutshell, on the basis of promiscuity (Table 3), conservancy and population coverage (Table 5) the epitope "TTPWTYFNFNQY" was considered as potential candidate for experimental investigations.

Additionally, certain HLA class II epitopes displayed excellent conservancy rate and restricted by several common alleles. Among them, the peptide "LLYMPFFLLENSDH" provided promising result (81.8%) to consider as a candidate for the development of therapeutic vaccine against HBoV1 infection (Table 4). Remarkably, this particular peptide was restricted to thirteen HLA class alleles that are often seen in almost all population (Supplementary workbook 3).

All these findings collectively indicated that the magnitude of the T cell mediated immune response mounted by the host might be partly attributed by enlisted epitopes derived from VP2 protein of HBoV1 (Tables 3 and 4). Nevertheless, functional studies based on the proposed epitopes will aid us to understand the cell mediated immunity of HBoV1 infection thereby towards development of preventive and therapeutic peptide vaccine.

3.5. B cell epitope

Knowledge on conserved immunodominant B cell epitopes provide better options for the development of sensitive diagnostic kits, immunotherapeutic tools and vaccine. The consolidated results of shortlisted peptides with sequences, method of prediction, basic properties and conservancy rate were provided in the table (Tables 6–8). The number of linear peptides (20mers) predicted by BCpred, BCpred AAP, FBCpred and ABCpred were 16, 16, 18 and 9 respectively. The details of the shortlisted forty five epitopes with appropriate software score along other parameters were shown in the Supplementary workbook 4. A total of 27 out of 45 peptides were found to be antigenic (VaxiJen cutoff $\geq$ 0.4) in nature subsequently subjected to epitope conservancy and allergenicity analysis. Finally, a total of fourteen peptides were identified as non-allergen with greater conservancy among VP2 protein sequences (Table 6).

At this junction, physio-chemical properties of the predicted peptides were considered to pinpoint potential candidates for further consideration. The threshold, maximal and minimal scores for each parameter with appropriate position and range were provided in Table 7. The graphical representation generated by the IEDB tool for each intrinsic property was provided in Fig. 2. In general, peptides with hydrophilic nature and higher probability for being found on the surface of a protein with flexibility were often reported as antigenic residues of B cells (Emini et al., 1985; Karplus and Schulz, 1985; Kolaskar and Tongaonkar, 1990; Parker et al., 1986). Two epitopes (numbered 2 and 9) failed to qualify due to lack of intrinsic properties and peptide numbered 10 was eliminated owing to extremely poor conservancy rate (Table 6). Few epitopes were found to possess excellent propensity scales above the threshold value with good conservancy rate. For instance, "CEIVWEVERYATKNWRPERR" and "KQILSNGADTTYNNDLTAGV" epitopes displayed more promising results as a candidate for being potential B cell sites (Tables 6 and 8). The complete results of intrinsic properties for each peptide were provided in the Supplementary work book 4.

SOPMA analysis of secondary structures revealed that majority of amino acids present on the predicted B cell epitopes was located in the coiled coil strand and helical region thus providing access for sufficient surface interaction (Table 8). In general, peptides with good anitgenicity, surface accessibility and flexibility offer higher success rate in terms of therapeutic and diagnostic tools development. The observation of threonine residues across the peptide sequence conferred indirect evidence for the hydrophilic and antigenic nature of the enlisted epitopes (Table 8). Experiments based on antibodies derived from patient sera are required to address the potentiality of these peptides as B cell linear epitopes. Earlier, Zhou et al. (2014) reported two B cell epitopes (position 1–20 and 162–180) as probable site for the development of a common diagnostic tool for HBoV1-4. But, present study did not find those residues as potential targets of B cells.

3.5.1. Conformational epitope

The finalized linear peptides (Table 8) shortlisted were cross examined with the results generated from Bepipred and ElliPro conformational linear epitope prediction. A total of twenty seven and twelve conformational peptides with different length were predicted by Bepipred and ElliPro, respectively (Supplementary workbook 4). The threshold, minimal and maximal value with appropriate position predicted by Bepipred was provided in Table 8. Further, maximum (position: 1–53; score: 0.815) and minimum (position: 304–316; score: 0.533) value predicted by ElliPro was provided in Supplementary work book 4. The graphical representation generated by the IEDB tool for linear epitope prediction was provided along epitope intrinsic properties (Fig. 1).

Then, conformational discontinuous epitopes were compared with the listed linear (20mer) and linear conformational epitope. ElliPro analysis reported nine distinct discontinuous epitope with a maximum and minimum value of 0.993 and 0.536, respectively. Discotope (version 2) analysis identified 265 distinct residues with a window size of 1 at the threshold setting of -3.700. Majority of the peptides reported in Table 8 were reiterated as B cell epitopes in Bepipred and ElliPro conformational epitope analysis. These results suggested that the shortlisted epitopes (Tables 6 and 8) play a crucial role in development of humoral response against HBoV1 infection. As a whole, further attention on the shortlisted epitopes (Tables 6 and 8) may illuminate the path for development of diagnostic and immunotherapeutic tools.

3.5.2. T and B cell epioptes superimposition

Generally, peptide that holds the capability to induce both cellular and humoral response was often suggested as most probable candidate. The B cell specific epitopes numbered 4–8 (Table 6) seems to overlap with predicted CTL epitopes (Table 3) thereby increasing the likelihood for being more effective candidate. Remarkably, epitopes "SGKSQRCVTTPWTYFNFNQY" and "LRTGESTEFTFTFNFDCEWVNN" found to grasp the capability to induce strong $CD8^+$ T cell response *via* CTL restricted major epitopes (underlined). Likewise, B cell epitopes numbered 13–14 (Table 6) was found to overlap with core peptides restricted for CD4 + T cells (Table 4). Therefore, *in-vivo* functional proficiency of these epitopes has to be evaluated through antibodies derived from human sera and animal model studies.

The epitope "SNYTPTYHVDPTGAYIQPTS" located at position

Table 6
Most probable linear epitopes predicted from VP2 protein of HBoV1.

S.No	Method	Sequence (position)	Cons	Intrinsic propensities				
			(%)	ES	KA	KF	pH	BT
1	Bcpred AAP	VDAPQNTSGGGTGSIGGGKG (14–33)	57.95	4	1	14	14	14
2	Bcpred/ FBCpred	APQNTSGGGTGSIGGGKGSG (16–35)	59.09	2	0	14	14	14
3	ABCpred	GHLYKTEAIETTNQSGKSQR (69–88)	63.64	12	3	10	12	5
4	ABCpred	SGKSQRCVTTPWTYFNFNQY (83–102)	63.64	3	9	5	5	6
5	FBCpred	KQILSNGADTTYNNDLTAGV (136–155)	77.27	8	2	12	13	12
6	FBCpred	LRTGESTEFTFNFDCEWVNN (232–252)	81.82	4	3	5	6	1
7	Bcpred AAP	AASTGRIQPYSKPTSWMTGP (282–301)	78.41	10	6	8	6	12
8	FBCpred	ASTGRIQPYSKPTSWMTGPG (283–302)	79.55	10	6	12	7	13
9	Bcpred AAP	VCTNPEGTHINTGAAGFGSG (321–340)	82.95	1	1	5	14	11
10	Bcpred/ FBCpred	GTHINTGAAGFGSGFDPPNG (327–346)	2.27	1	0	6	12	12
11	Bcpred	CEIVWEVERYATKNWRPERR (477–496)	86.36	10	7	7	7	4
12	ABCpred	HTALGMSLGGESNYTPTYHV (497–516)	73.86	6	4	8	9	10
13	Bcpred/Bcpred AAP	SNYTPTYHVDPTGAYIQPTS (508–527)	72.73	6	13	5	9	9
14	ABCpred	TPTYHVDPTGAYIQPTSYDQ (511–530)	72.73	6	14	6	9	10

Note: Peptides underlined overlaps with predicted T cell peptides; Cons − conservancy; % − percentage; ES − Emini's surface accessibility, KA − Kolaskar and Tongankar antigenicity, KF − Karplus flexibility, pH − Parker's hydrophilicity, BT − Chou and Fasman beta turn.

508–527 possess two distinct core peptides ("SNYTPTYHV" and "YHVDPTGAY") for the induction of T cell response. Notably, the core peptide "YHVDPTGAY" was qualified as putative candidates for the induction of $CD4^+$T cell mediated immune response (Table 4). These findings inferred the capability of the peptide to induce both B cell and $CD4^+$T cell response simultaneously thereby qualified as potential candidate for therapeutic vaccine. However, the peptide "YHVDPTGAYI" restricted to $CD8^+$ T cell was labeled as probable allergen. Hence, experimental assays were required to confirm the possibility for utilization of the peptide as a candidate for preventive vaccine without allergenic reactions.

Interestingly, certain degree of serum cross-reactivity was observed between HBoV-1 and HBoV2-4 owing to VP2 gene conservation (Broccolo et al., 2015; Li et al., 2015b). This resulted in the observation of Original Antigenic Sin (OAS) phenomena during consecutive HBoV infection in human and rabbit models (Li et al., 2015b). Yet, mechanism underlying the phenomena of OAS during HBoV infection remains elusive. Theoretically, weaker T cell immune response and antibodies with lower avidity often involved in OAS reaction. Thus, degree of cross-reactivity for the priming and subsequent challenging virus depends on selection of virus specific B and T cells. Comparison of the shortlisted peptides derived from VP2 protein of HBoV1 with VP2 protein of HBoV2-4 showed that several regions were partially conserved (data not shown). Hence, the potential sites reported in the present study may aid us to address the immunological standpoint of OAS reaction during consecutive HBoV infection.

It was suggested that the Virus Like Particles (VLPs) constructed using VP2 of HBoV1 displayed good immunogenicity in mice (Deng et al., 2014). However, VLPs has to be studied extensively for the possibility of infectivity while administration in humans, Further, the possibility to elicit an inflammatory response also has to be addressed. The major advantage of peptide candidates were higher stability, easier in terms of synthesis and characterization rather than the VLPs.

3.6. Active site prediction

A total of 96 different pockets with variable molecular surface area, volume and mouth openings were predicted by CASTp. Among them, pocket ID 96 with an area of 4601.1 made of 158 amino acid residues with 12 openings was identified as the best active site. The residues within the pocket 96 and the surface area covered in the VP2 protein were graphically represented in Fig. 3. Notably, residues (underlined) located on the predicted epitopes "SGKSQRCVTTPWTYFNFNQY and LYQMPFFL" was also a part of the best active site (pocket ID 96).

3.7. Peptide match

A total of 33 predicted peptides (T cell restricted: all 8 epitopes of Table 3 and top 10 conserved epitopes of Table 4; B cell restricted: 15 unique epitopes consolidated from Table 6 and Table 8) were submitted in the FASTA format as mentioned in the Peptide Match Service for batch peptide analysis. No match was obtained with the peptides present in human proteome (Supplementary 5). This clearly indicated that the selected peptides possess the ability to elicit immune response against the viral peptides and indirectly suggest that the selected peptides did not elicit autoimmune reactions on the host.

3.8. Molecular docking analysis

Molecular docking was performed to assess the quality of molecular interaction between the peptide and HLA moiety.The three dimensional structure of the peptide and HLA was constructed using PEP-FOLD3 and Phyre2 web servers, respectively. Alleles were selected on the basis of strong binding affinity with corresponding peptide for docking purpose. The complete result of structural homology modeling, refinement, quality assessment and docking was provided in the Supplementary work book 1.

Table 7
Key values of various intrinsic properties for being B cell epitope calculated using IEDB tool.

Parameter	Threshold value	Minimum value	Position (range)	Maximum value	Position (range)
Bepipred	0.449	−1.963	221 (218–224)	2.322	365 (362–368)
Chou-Fosmann Beta-turn	1.053	0.664	481 (478–484)	1.474	368 (365–371)
Emini surface accessibility	1.000	0.049	155 (153–159)	6.102	424 (422–428)
Karplus flexibility	1.008	0.900	320 (317–323)	1.144	84 (81–87)
Kolaskar antigenicity	1.006	0.837	368 (365–371)	1.173	477 (474–480)
Parker hydrophilicity	1.920	−4.700	222 (219–225)	6.471	10 (7–13)

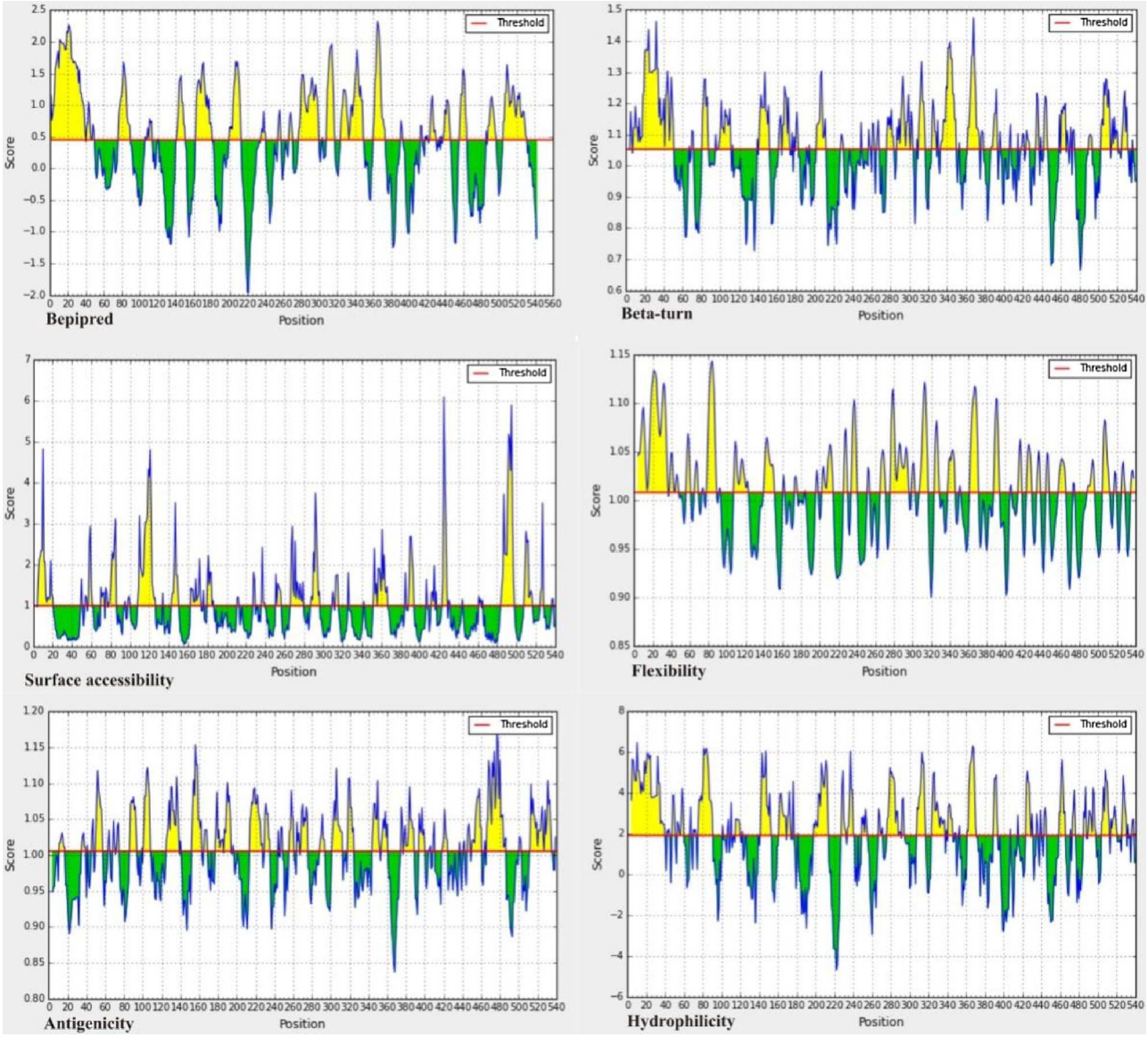

Fig. 2. Prediction of B cell linear epitope and intrinsic properties using different scales.
For each graph: x-axis and y-axis represent the position and score; residues that fall above the threshold value are shown in yellow color; the highest peak in yellow color identifies most favored position.

Table 8
Most probable conformational epitopes derived from VP2 protein of HBoV1.

S.No	Sequence	Conformational epitope (position)	SOPMA
1	VDAPQNTSGGGTGSIGGGKG (14–33)	Bepipred*/ElliPro*	ccccccccccccceeccccc
2	APQNTSGGGTGSIGGGKGSG (16–35)	Bepipred*/ElliPro*	cccccccccccceeccccce
3	GHLYKTEAIETTNQSGKSQR (69–88)	Bepipred (76–88)/Discotope (79–84)	tceeeeeeeeccccccccc
4	SGKSQRCVTTPWTYFNFNQY (83–102)	Bepipred (83–90)/Discotope (100- 103)	ccccceeeccceeeetttt
5	KQILSNGADTTYNNDLTAGV (136–155)	Bepipred (140–151)/Discotope (139–145)	hheehttcceeecttcctte
6	LRTGESTEFTFNFDCEWVNN (232–252)	Bepipred (233–240,250)/Discotope (246–247, 249–252)	eettcccceeeeecttecct
7	AASTGRIQPYSKPTSWMTGP (282–301)	Bepipred*/ Disoctope (282–298)	ccccccccccccccceecct
8	ASTGRIQPYSKPTSWMTGPG (283–302)	Bepipred (283–299)/Disoctope (282–298)	ccccccccccccceeccctt
9	VCTNPEGTHINTGAAGFGSG (321–340)	Bepipred (324–340)/Discotope (331, 333,334, 339)	eeccttcceeeccccccccc
10	GTHINTGAAGFGSGFDPPNG (327–346)	Bepipred*/ Discotope (331, 333,334, 339, 342)	cceeeccccccccccccttc
11	CEIVWEVERYATKNWRPERR (477–496)	Bepipred (486–496)/Discotope (482–496)	hhhhhhhhhhcctttccccc
12	HTALGMSLGGESNYTPTYHV (497–516)	Bepipred (497,505–516)/Discotope*	cheeeeettcccccceeeee
13	SNYTPTYHVDPTGAYIQPTS (508–527)	Bepipred*/ElliPro*/Discotope*	ccccceeeecttcceecccc
14	TPTYHVDPTGAYIQPTSYDQ (511–530)	Bepipred*/ElliPro*/Discotope*	cceeeecttcceeccccccc

Note: *Whole peptide was reported as a part of conformational epitopes; residues underlined overlaps with t cell epitopes; Alphabets under SOPMA column: c − random coil, e- extended strand, h − helix, t − beta-turn.

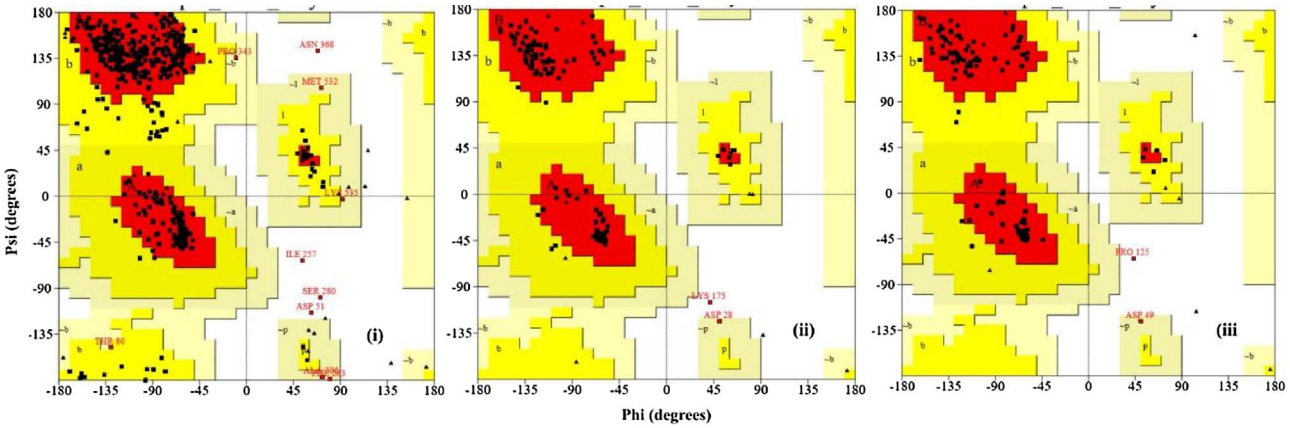

Fig. 1. Ramachandran plot generated by PROCHECK for VP2 protein and HLA alleles.
Note: Red color indicates residues the most favored region (A,B,L), bright yellow indicates additional allowed regions (a,b,l,p), light yellow indicates generously allowed region (~a, ~b, ~l, ~p) and white indicates disallowed region. The torsion angles were determined by Phi and Psi. Figure i, ii and iii represent VP2, HLA A*24:02 and HLA B-35:01, respectively.

3.8.1. Model refinement and validation

Structural refinement was performed for HLA alleles (three times) with ModRefiner. The minimal RMSD value and maximal TM score obtained for the selected alleles were provided in Table 1. Results obtained from Ramachandran plot and QMEAN6 score for the refined models were provided in Tables 1 and 2 along the VP2 protein. As a representation, Ramachandran plot generated by PROCHECK for HLA molecules were provided in Fig. 1(ii) and (iii). The stereochemical qualities of the constructed models were assured by observation of majority of residual cluster at favorable region. Verify 3D analysis reported that majority of residues (> 95%) had an averaged 3D-1D score ≥ 0.2 for generated HLA models. Finally, consolidated results of QMEAN6 scores (Tables 1 and 2) reached the inference that the constructed refined models were qualified as template for docking analysis.

3.8.2. Docking

Of note, structural flexibility of the interacting protein molecules were not considered by ClusPro engine while docking thereby Autodock Vina was employed. The grid (centre) and size dimensions set that covered beta folded chains and alpha helical chains which targets the binding site of the peptide (for Autodock Vina) was provided in the Supplementary work book 1. The binding energy calculated by ClusPro and Autodock Vina for each HLA-peptide complex was provided in the Table 9. In general, observation of strong binding energy (in negative value) indicated higher binding affinity of the peptide with stable intermolecular interaction. The promising results of docking analysis informed that the putative epitopes displayed strong and stable interaction with selected HLA alleles. Consolidated results clearly indicated that the epitope "TTPWTYFNFNQY" grasped the ability to elicit effective immune response and qualified for experimental assays (*in vivo*).As a representation, peptides docked with appropriate allele using ClusPro and AutoDock Vina was displayed in Fig. 4.

4. Conclusion

Several peptides were identified as most probable candidates for development of multi-epitopic vaccine against human bocavirus 1. The capability to induce polyclonal T cell by the epitope "TTPWTYFNFNQY" ensured as a candidate suitable for experimental investigations. Several B cell epitopes with basic intrinsic properties were identified as possible targets for development of therapeutic and diagnostic tools. Additionally, potential sites of VP2 protein that may participate in active conversation between HBoV1 and immune system of human were highlighted. Further, epitopes reported here may aid to address the original antigenic sin reaction while consecutive human bocavirus 2–4 infection. Furthermore, comparison with population based epidemiological studies on natural outcome of HBoV1 infection will open the gates of epitope based vaccine for HBoV1.

Disclosure and funding

Self-supported, no funds received from any funding agencies to perform this investigation

Ethical clearance

Ethic issues are not involved in this study

Disclosure

Hereby I declare no competing interests.

Fig. 3. Results of CASTP most active site and visualization using Chimera. (a) Green color represents active site positions of Pocket ID 96 (b) three dimensional structures of the active site covered on VP2 protein. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 9
Binding energy values calculated by ClusPro and Auto-dock vina.

		CLUSPRO		AUTO DOCK VINA			
		Weighed Score		Energy calculations			
HLA	Peptide	Centre	Lower	Kcal/mol	RMSD I.b	RMSD u.b	H bonds (all)
A*24:02	TTPWTYFNF	−1034.2	−1147.5	−10.8	0	0	10
A*29:02	WTYFNFNQY	−1085.7	−1283	−9.6	0	0	5
B*07:02	VPTRRVQYI	−898.3	−1083.8	−9.6	0	0	6
B*07:02	WRPERRHTAL	−804.7	−926.7	−5.4	0	0	7
B*08:01	VPTRRVQYI	−1005.2	−1256.5	−7.9	0	0	4
B*08:01	WRPERRHTAL	−962.3	−1053.7	−8.7	0	0	3
B*08:01	YKRFRPKAM	−960.5	−1009.8	−7	3.412	14.848	4
B*15:01	ASTGRIQPY	−874.5	−874.5	−8.2	0	0	5
B*35:01	DPFDGSIAM	−879.2	−987.3	−6.6	0	0	2
B*35:01	WTYFNFNQY	−1112.4	−1346.2	−7.6	3.845	14.59	4
B*15:01	WTYFNFNQY	−1334.3	−1334.3	−6.1	3.268	8.459	7
B*58:01	FTFNFDCEWV	−1208.3	−1404.8	−5.7	4.535	7.723	1
C*03:02	ASTGRIQPY	−725.8	−764.6	2.3	4.728	16.178	2
C*03:02	FTFNFDCEWV	−1134.8	−1407.4	20.7	2.453	5.188	2

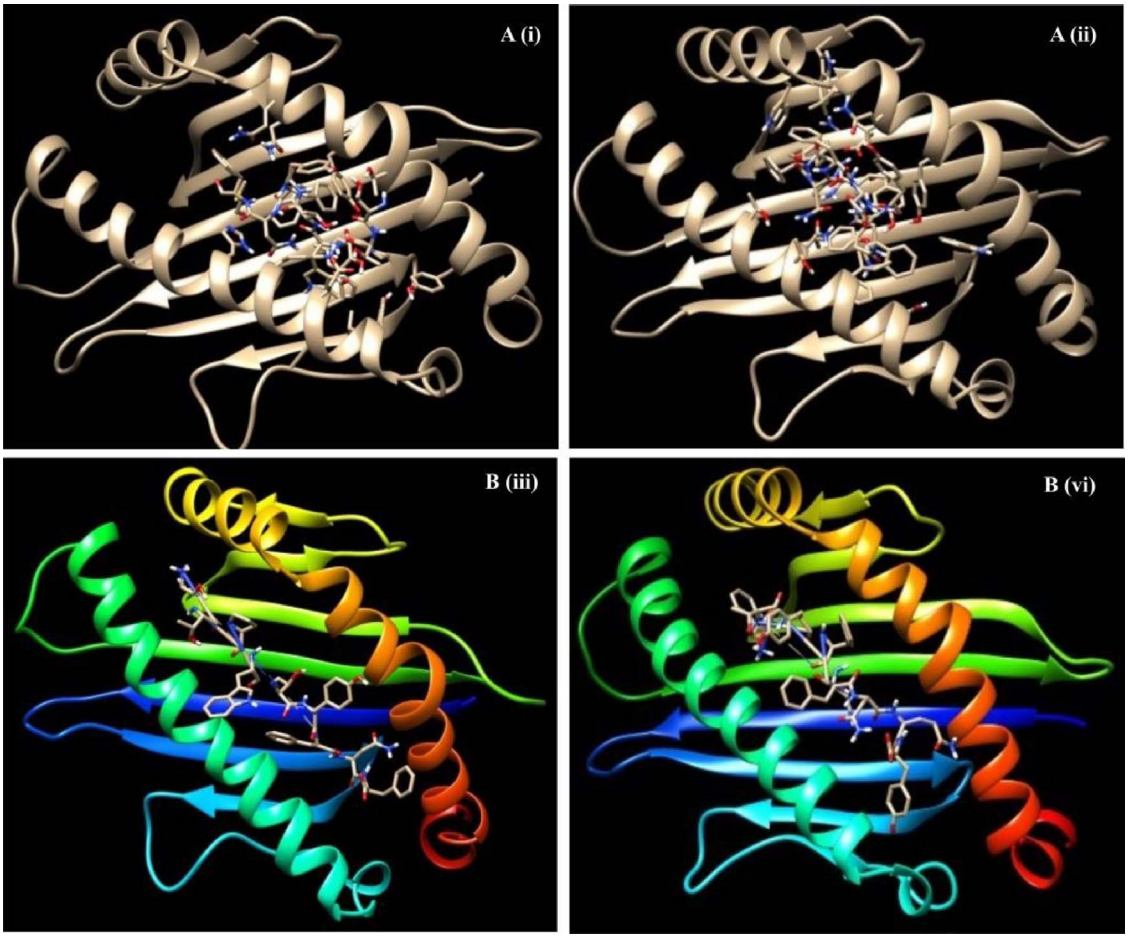

Fig. 4. Docking simulation results generated by ClusPro and AutodockVina. (A) Cluspro docking of Peptide (i) “TTPWTYFNF” with HLA A*24:02 and (ii) “WTYFNFNQY” with HLA B*35:01. (B) AutodockVina docking of Peptide (iii) “TTPWTYFNF” with HLA A*24:02 and (iv) “WTYFNFNQY” with HLA B*35:01.

Acknowledgement

I express my gratitude to the reviewers for their critical comments and fruitful suggestions for the improvement of this manuscript. Further, I express my sincere thanks to the scientific community for developing appropriate *in silico* tools.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.molimm.2017.11.024.

References

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a

web-based environment for protein structure homology modelling. Bioinformatics 22 (2), 195–201.
Ashkenazy, H., Erez, E., Martz, E., Pupko, T., Ben-Tal, N., 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 38 (suppl_2), W529–W533.
Benkert, P., Biasini, M., Schwede, T., 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27 (3), 343–350.
Berzofsky, J.A., Cease, K.B., Cornette, J.L., Spouge, J.L., Margalit, H., Berkower, I.J., Good, M.F., Miller, L.H., Delisi, C., 1987. Protein antigenic structures recognized by t cells; potential applications to vaccine design. Immunol. Rev. 98 (1), 9–52.
Bhasin, M., Raghava, G.P.S., 2004. Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine 22 (23), 3195–3204.
Broccolo, F., Falcone, V., Esposito, S., Toniolo, A., 2015. Human bocaviruses: possible etiologic role in respiratory infection. J. Clin. Virol. 72, 75–81.
Bui, H.H., Sidney, J., Li, W., Fusseder, N., Sette, A., 2007. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinf. 8 (1), 1.
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 (3), 423–428.
Chen, C., Li, Z., Huang, J., Suzek, B.E., Wu, C.H., 2013. A fast peptide match service for UniProt knowledgebase. Bioinformatics 29 (21), 2808–2809.
Chou, P.Y., Fasman, G.D., 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. Relat. Areas Mol. Biol. 47, 45–148.
Deng, Z.H., Hao, Y.X., Yao, L.H., Xie, Z.P., Gao, H.C., Xie, L.Y., Zhong, L.L., Zhang, B., Cao, Y.D., Duan, Z.J., 2014. Immunogenicity of recombinant human bocavirus-1, 2 VP2 gene virus-like particles in mice. Immunology 142 (1), 58–66.
Dhanda, S.K., Gupta, S., Vir, P., Raghava, G.P.S., 2013a. Prediction of IL4 inducing peptides. Clin. Dev. Immunol (2013).
Dhanda, S.K., Vir, P., Raghava, G.P., 2013b. Designing of interferon-gamma inducing MHC class-II binders. Biol. Direct 8 (1), 1.
Dikhit, M.R., Ansari, M.Y., Mansuri, R., Sahoo, B.R., Dehury, B., Amit, A., Topno, R.K., Sahoo, G.C., Ali, V., Bimal, S., Das, P., 2016. Computational prediction and analysis of potential antigenic CTL epitopes in Zika virus: a first step towards vaccine development. Infect. Genet. Evol. 45, 187–197.
Dimitrov, I., Bangov, I., Flower, D.R., Doytchinova, I., 2014. AllerTOP v: 2–a server for in silico prediction of allergens. J. Mol. Model. 20 (6), 1–6.
Doytchinova, I.A., Flower, D.R., 2007. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinf. 8 (1), 1.
Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., Liang, J., 2006. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. 34 (suppl 2), W116–W118.
Eisenberg, D., Lüthy, R., Bowie, J.U., 1997. [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods Enzymol. 277, 396–404.
Emini, E.A., Hughes, J.V., Perlow, D., Boger, J., 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J. Virol. 55 (3), 836–839.
Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.E., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server. Humana Press, pp. 571–607.
Geourjon, C., Deleage, G., 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput. Appl. Biosci.: CABIOS 11 (6), 681–684.
Jørgensen, K.W., Rasmussen, M., Buus, S., Nielsen, M., 2014. NetMHCstab–predicting stability of peptide–MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. Immunology 141 (1), 18–26.
Jartti, T., Hedman, K., Jartti, L., Ruuskanen, O., Allander, T., Söderlund-Venermo, M., 2012. Human bocavirus–the first 5 years. Rev. Med. Virol. 22 (1), 46–64.
Kailasan, S., Agbandje-McKenna, M., Parrish, C.R., 2015. Parvovirus family conundrum: what makes a killer? Annu. Rev. Virol. 2, 425–450.
Karplus, P.A., Schulz, G.E., 1985. Prediction of chain flexibility in proteins. Naturwissenschaften 72 (4), 212–213.
Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10 (6), 845–858.
Kim, Y., Ponomarenko, J., Zhu, Z., Tamang, D., Wang, P., Greenbaum, J., Lundegaard, C., Sette, A., Lund, O., Bourne, P.E., Nielsen, M., 2012. Immune epitope database analysis resource. Nucleic Acids Res (p.gks438).
Kolaskar, A.S., Tongaonkar, P.C., 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 276 (1–2), 172–174.
Kozakov, D., Hall, D.R., Beglov, D., Brenke, R., Comeau, S.R., Shen, Y., Li, K., Zheng, J., Vakili, P., Paschalidis, I.C., Vajda, S., 2010. Achieving reliability and high accuracy in automated protein docking: clusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13–19. Proteins: Struct. Funct. Bioinf. 78 (15), 3124–3130.
Kringelum, J.V., Lundegaard, C., Lund, O., Nielsen, M., 2012. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. PLoS Comput. Biol. 8 (12) (p.e1002829).
Lamiable, A., Thévenet, P., Rey, J., Vavrusa, M., Derreumaux, P., Tufféry, P., 2016. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res (p.gkw329).
Larsen, J.E.P., Lund, O., Nielsen, M., 2006. Improved method for predicting linear B-cell epitopes. Immunome Res. 2 (1), 1.
Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., Thornton, J.M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8 (4), 477–486.
Li, H., He, M., Zeng, P., Gao, Z., Bian, G., Yang, C., Li, W., 2015a. The genomic and seroprevalence of human bocavirus in healthy Chinese plasma donors and plasma derivatives. Transfusion (Paris) 55 (1), 154–163.
Li, X., Kantola, K., Hedman, L., Arku, B., Hedman, K., Söderlund-Venermo, M., 2015b. Original antigenic sin with human bocaviruses 1–4. J. Gen. Virol. 96 (10), 3099–3108.
Lindner, J., Zehentmeier, S., Franssila, R., Barabas, S., Schroeder, J., Deml, L., Modrow, S., 2008a. CD4+ T helper cell responses against human bocavirus viral protein 2 viruslike particles in healthy adults. J. Infect. Dis. 198 (11), 1677–1684.
Lindner, J., Karalar, L., Zehentmeier, S., Plentz, A., Pfister, H., Struff, W., Kertai, M., Segerer, H., Modrow, S., 2008b. Humoral immune response against human bocavirus VP2 virus-like particles. Viral Immunol. 21 (4), 443–450.
EL-Manzalawy, Y., Dobbs, D., Honavar, V., 2008. Predicting linear -cell epitopes using string kernels. J. Mol. Recogn. 21 (4), 243–255.
Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S., Røder, G., Peters, B., Sette, A., Lund, O., Buus, S., 2007. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and-B locus protein of known sequence. PloS one 2 (8) (p.e796).
Ovsyannikova, I.G., Jacobson, R.M., Poland, G.A., 2004. Variation in vaccine response in normal populations. Pharmacogenomics 5 (4), 417–427.
Parker, J.M.R., Guo, D., Hodges, R.S., 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry 25 (19), 5425–5432.
Parker, K.C., Bednarek, M.A., Coligan, J.E., 1994. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J. Immunol. 152 (1), 163–175.
Patronov, A., Doytchinova, I., 2013. T-cell epitope vaccine design by immunoinformatics. Open Biol. 3 (1), 120139.
Paul, S., Weiskopf, D., Angelo, M.A., Sidney, J., Peters, B., Sette, A., 2013. HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. J. Immunol. 191 (12), 5831–5839.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. 25 (13), 1605–1612.
Ponomarenko, J., Bui, H.H., Li, W., Fusseder, N., Bourne, P.E., Sette, A., Peters, B., 2008. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinfor. 9 (1), 1.
Pratheek, B.M., Suryawanshi, A.R., Chattopadhyay, S., Chattopadhyay, S., 2015. In silico analysis of MHC-I restricted epitopes of Chikungunya virus proteins: implication in understanding anti-CHIKV CD8+ T cell response and advancement of epitope based immunotherapy for CHIKV infection. Infect. Genet. Evol. 31, 118–126.
Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V., 1963. Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7 (1), 95–99.
Rosa, D.S., Ribeiro, S.P., Cunha-Neto, E., 2010. CD4+ T cell epitope discovery and rational vaccine design. Archivum Immunologiae Et Therapiae Experimentalis 58 (2), 121–130.
Saha, S., Raghava, G.P.S., 2006. Prediction of continuous Bâ€?cell epitopes in an antigen using recurrent neural network. Proteins: Struct. Funct. Bioinf. 65 (1), 40–48.
Schildgen, O., Qiu, J., Söderlund-Venermo, M., 2012. Genomic features of the human bocaviruses. Future Virol. 7 (1), 31–39.
Sette, A., Sidney, J., 1998. HLA supertypes and supermotifs: a functional perspective on HLA polymorphism. Curr. Opin. Immunol. 10 (4), 478–482.
Singh, H., Raghava, G.P.S., 2003. ProPred1: prediction of promiscuous MHC Class-I binding sites. Bioinformatics 19 (8), 1009–1014.
Srivastava, P.N., Jain, R., Dubey, S.D., Bhatnagar, S., Ahmad, N., 2016. Prediction of epitope-based peptides for vaccine development from coat proteins GP2 and VP24 of Ebola virus using immunoinformatics. Int. J. Pept. Res. Ther. 22 (1), 119–133.
Staneková, Z., Varečková, E., 2010. Conserved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development. Virol. J. 7 (1), 1.
Testa, J.S., Philip, R., 2012. Role of T-cell epitope-based vaccine in prophylactic and therapeutic applications. Future Virol. 7 (11), 1077–1088.
Trott, O., Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31 (2), 455–461.
Trowsdale, J., 2011. The MHC, disease and selection. Immunol. Lett. 137 (1), 1–8.
Xu, D., Zhang, Y., 2011. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys. J . 101 (10), 2525–2534.
Yang, X., Yu, X., 2009. An introduction to epitope prediction methods and software. Rev. Med. Virol. 19 (2), 77–96.
Yasser, E.M., Dobbs, D., Honavar, V., 2008. Predicting flexible length linear B-cell epitopes. Comput. Syst. Bioinf. 7, 121 (NIH Public Access).
Zhou, Z., Gao, X., Wang, Y., Zhou, H., Wu, C., Paranhos-Baccalà, G., Vernet, G., Guo, L., Wang, J., 2014. Conserved B-cell epitopes among human bocavirus species indicate potential diagnostic targets. PLoS One 9 (1), e86960.