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

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.