# Design of an Epitope-Based Peptide Vaccine Against Dengue Virus Isolate from Eastern Uttar Pradesh, India

Siddharth Gupta[1] · Ajay Kumar[1]

## Abstract

Dengue outbreaks are a serious public health concern that occurs on a regular basis in various locations of India. According to the Government of India's National Center for Vector-Borne Disease Control, a total of 1,23,106 dengue cases were identified in India as of October 2021. The currently available dengue vaccine was found to be ineffective against all serotypes of the virus. Dengue virus serotype 2 was reported to be the sole predominant serotype in Eastern Uttar Pradesh, India. An epitope-based peptide vaccine is believed to be safe and effective against all serotypes of the dengue virus. In this work, an epitope-based peptide vaccine based on envelope protein against the dengue virus was developed using the reverse vaccinology method. T-cell epitopes present in the envelope protein were screened using different immunoinformatic tools. Epitopes predicted by all servers were chosen and additionally picked out on the grounds of their antigenic reactivity, immunogenicity, toxicity, and allergenicity assessment. Three potent T cell epitopes as IVQPENLEY, ILIGVVITW, and DTAWDFGSL were screened, which binds with HLA-B*35:01, HLA-B*58:01, HLA-A*26:01 alleles, respectively. To build a 3D structure model of epitopes and alleles, the PepstrMod and Swiss-Model servers were used. Predicted epitopes and HLA alleles were docked using the HPEPDOCK server to confirm binding ability. These anticipated epitopes were found to cover the greatest number of populations in India and around the world. These identified epitopes have a high potential for eliciting an immune response in the development of a vaccine against the dengue virus, while further experimental validation is required for final confirmation.

**Keywords** Dengue virus · MHC class I · T-cell epitopes · Vaccine

## Introduction

Dengue (DEN) is an important mosquito-borne viral infection of humans that contributes to the considerable worldwide epidemic load of subtropical and tropical nations of the Caribbean, Africa, America, the Pacific, and Asia. Dengue Virus is a flavivirus belonging to the Flaviviridae family. Dengue fever (DF) is characterized clinically as moderate and self-limiting, which may result in potentially lethal consequences (Murphy and Whitehead 2011). Dengue Virus strains are divided into four serotypes based on antigenic characteristics: DENV1, DENV2, DENV3, and DENV4 (Murrell et al. 2011).DENV2 infection can cause more severe disease than other serotypes and it can cause shock and fatal internal bleeding. According to the DG of the Indian Council of Medical Research, the DENV2variant is highly virulent and can lead to more fatal cases. The DENV2could be responsible for the recent deaths in the western UP districts of Aligarh, Mathura, Agra, and Firozabad due to a strange fever. Dengue fever has been detected in more than 15 states throughout the country (Deval et al. 2021). Dengue fever cases have been progressively increasing across the country, posing significant hurdles for medical personnel.

Dengue virus is presently prevalent in one hundred twenty-eight countries, threatening around 3.97 billion individuals each year (Brady et al. 2012; Bhatt et al. 2013).DEN outbreaks were widespread in northern India from 2015 to 2016, affecting Uttar Pradesh (UP), Punjab, Haryana, and New Delhi (Choudhary et al. 2017; National center for vector-borne diseases control 2019). In 2020, Asian countries such as India reported 16,439 cases with 12 deaths, Malaysia documented 78,303 cases and 127 deaths, Cambodia

✉ Ajay Kumar
  ajaymtech@gmail.com

[1] Department of Biotechnology, Faculty of Engineering and Technology, Rama University, G.T. Road, Kanpur 209217, India

declared 9108 cases and 14 deaths, while Bangladesh confirmed 494 cases (CDC 2020).According to the most recent statistics, Delhi is currently the worst-affected state, with a dengue death toll of 6 (NCVBDC 2021). With the inclusion of new cases, the number of cases in Delhi has raised to 2794 (NVBDCP 2021). In West Bengal, Punjab, Telangana, Gujarat, and Madhya Pradesh, dengue fever is on the rise. The disease has become a significant public health issue in Uttar Pradesh, India's most populous state, with over 200 million populations. Deoria, Basti, Siddharth Nagar, Kushinagar, Sant Kabir Nagar, and Gorakhpur are located in the eastern part of the Uttar Pradesh state, India.

Seven nonstructural and three structural proteins are encoded by the dengue viral genome (Proutski et al. 1999; Markoff 2003). Global genetic differences across different genotypes and serotypes of the dengue virus have been studied using the envelope protein analysis (Foster et al. 2004; Twiddy et al. 2002; Rico-Hesse et al. 1997; Wang et al. 2000). Dengvaxia, a dengue virus vaccine, was approved in 2015 to treat individuals aged 9 to 16 years who live in endemic areas. Dengvaxia vaccine renders only partial protection against the DENV2 infection (Liu et al. 2016). Clinical study findings for Dengvaxia show that it can cause miscarriage, elective termination, uterine death, elective termination, and stillbirth in pregnant women (https://www.fda.gov/media/124379/download). This condition necessitates the development of a new vaccine to combat the DENV-2 strain. Epitope-based vaccines outperform traditional vaccines and can also address safety and side effects concerns (Srivastava et al. 2020; Krishnan et al. 2020a, b). This research began with the discovery of a number of epitopes in the dengue virus envelope protein, which were antigenic and consensus in different algorithms used in epitope–allele binding. Following that, PEPstrMOD and Swiss-model were used to generate these peptides and HLA alleles in their 3D models for further structural binding study by docking. Following that, PEPstrMOD and Swiss-model were used to construct 3D models of these peptides and HLA alleles, which were then confirmed using a docking study.

## Methodology

For the development of a candidate vaccine against envelope glycoprotein of dengue virus, an immunoinformatics and reverse vaccinology strategy was adopted.

### Retrieval of the Protein Sequences

Envelope protein (Protein Id: AWI48553.1) of Dengue virus 2 isolate NIV/GFU-DV-1615135 was downloaded from the protein database of NCBI. This strain was discovered in serum samples taken clinically confirmed the dengue fever in patients who visited from 2015 to 2016 in Gorakhnath Multispecialty Hospital, Gorakhpur (Deval et al. 2021). E protein of DENV-2 was amplified from the virus in NIV, Gorakhpur, UP (Deval et al. 2021). E protein gene sequences from internationally scattered DENV-2 serotypes share 96–100% sequence similarity at the nucleotide level. The Vaxign version 2.0 beta server (http://www.violinet.org/vaxign2), a vaccine target prediction and analysis tool based on reverse vaccinology, was used to evaluate envelope glycoprotein as a potential vaccine targetfor developing vaccine candidates.

### Instability and Allergenicity Prediction of the Envelope Protein

The instability index of protein was checked by the ProtParam tool (https://web.expasy.org/protparam/) and allergenicity by Allergen FP (https://ddg-pharmfac.net/AllergenFP/).

### Prediction of CTL Epitopes

Three servers having a different approach to search such as the IEDB MHC-I prediction tool (Kim et al. 2012), NetCTLpan1.1 (Stranzl at el. 2010) and NetMHC-4.0 (Andreatta and Nielsen 2016) have been used to identify envelope protein epitopes of Cytotoxic T lymphocytes (CTL). IEDB MHC-I method of prediction identifies epitopes that may interact with genes of MHC-I. The NetCTLpan1.1 (https://services.healthtech.dtu.dk/service.php?NetCTLpan-1.1) tool predicts epitopes of MHC class I by the use of weight matrix, the efficiency of TAP transport, and ANN. NetMHC 4.0 (https://services.healthtech.dtu.dk/service.php?NetMHC-4.0) server predicts T cell epitopes using ANN.

### Immunogenicity, Toxicity Antigenicity and Allergenicity Analysis of the Epitopes

The epitopes Identified by all servers were exposed to the IEDB class I immunogenicity tool using usual limitations (Calis et al. 2013). The antigenicity of the epitopes was determined using the VaxiJen v2.0 server, with a threshold of 0.4 used to resolve the prediction's accuracy (Doytchinova and Flower 2007). To generate allergenicity, the server AllerTOP v.2.0 was used. AllerTOP is an alignment-independent online allergenicity prediction tool that produces accurate findings (Dimitrov et al. 2014). Furthermore, ToxinPred was used to predict toxicity using an SVM-based algorithm with default parameters (Gupta 2013).

## Peptide Modeling

Only the selected epitopes were chosen for the Peptides and HLAs interaction pattern study after several bioinformatics investigations. The PEPstrMOD tool was used to build a three-dimensional model of selected T-cell epitopes (Singh et al. 2015a, b). The prediction technique is based on the fact that, in addition to regular structures, the β-turn is an essential and constant property of short peptides. Energy minimization and molecular dynamics simulations are used to refine the structure.

## Allele Modeling

HLA alleles having known crystal structures were retrieved from Protein Data Bank. The sequence of HLA alleles having unknown structures was downloaded from IPD-IMGT/HLA Database (Robinson et al. 2015). The 3D structures of these HLA alleles were modeled by the SWISS-MODEL server (Waterhouse et al. 2018).

## Tertiary Structure Validation

The structure of the refined model was validated using the Swiss-Model Structure Assessment page (https://swiss model.expasy.org/assess). Another tool, ProSA-web (Wiederstein and Sippl, 2007) was used for validating the protein structures. ProSA takes a protein structure (PDB file) as input and calculates an overall model quality score (z-score), and outputs it in a plot. If the z-score is outside a range characteristic for native proteins, the structure probably contains errors.

## Molecular Docking

The molecular docking of epitopes with their respective HLA binding alleles was conducted using an online HPEP-DOCK server (Zhou et al. 2018).HPEPDOCK is a server that uses a hierarchical approach to perform blind peptide-protein docking.

## Conservancy Analysis of Epitopes

Conserved epitopes are thought to give more protection across species than epitopes from highly varied genomic regions. Similar sequences were found by comparing the amino acid sequence of the envelope to the nr sequence database using the BLASTP software (Altschul et al. 1997). The IEDB's conservancy analysis tool (Bui et al. 2007) was utilized to determine the conservancy of CTL epitopes employed in candidate vaccine design among screened homologues.

## Population Coverage Analysis

In order for vaccination to be effective, a vaccine molecule must give broad-spectrum protection against the disease in distinct world populations. The IEDB population coverage tool (http:/tools.iedb.org/population/) was used in population coverage analysis of epitopes.

# Results

## Vaccine Target

Four parameters of the Vaxign version 2.0 beta server (Xiang and He 2009) were used to check envelope protein as a potential vaccine target. Two transmembrane helix was predicted in protein. The predicted adhesion probability of this protein is 0.494. Adhesion probability > 0.51 suggested that protein is an adhesion. Predicted protein does not have a similarity to human proteins.

## Instability and Allergenicity of Envelope Protein

The instability index of envelope protein is 28.91 and protein is stable. The resultof Allergen FP too shows that protein is non-allergen with a similarity index of 0.82.

## CTL Epitopes Analysis

We selected the recommended IEDB 2020.09 prediction method and HLA allele reference set, which cover the largest number of the world population to predict epitopes with a length of 9- mers. The set of expected binders was made for IEDB MHC class I epitope predictionsbased on the percentile range < 0.5 percent to cover the topmost immune responses. In our calculated results the threshold value for NetCTLpan1.1 and NetMHC-4.0 servers was selected as < 0.5. To improve accuracy, we filter out those epitopes, which are commonly predicted by these three servers. Of these three servers, 28 commonly predicted epitopes (Table 1) from a total of 119 predicted epitopes (Supplementary Table ST-1) were selected from antigenic protein.

## Immunogenicity, Toxicity, Antigenicity and Allergenicity Analysis of Epitopes

Immunogenicity analysis of selected 28 T cell epitopes reported a positive immunogenicity value for 15 epitopes. A high score of immunogenicity results in high potency for the stimulation of naive T cells. Toxicity, antigenicity and allergenicity prediction were carried out for all the epitopes. After the assessment, three best MHC class-I epitopes Such

**Table 1** Identification of MHC-I epitopes using three different servers

| Sl. No | Start | End | Epitope | IEDB MHC I Binding (percentile rank) | NetMHCpan (% rank) | NetCTLpan (% rank) |
|---|---|---|---|---|---|---|
| 1 | 204 | 212 | KAWLVHRQW | HLA-B*53:01(0.27), HLA-A*32:01(0.04), HLA-B*58:01(0.01), HLA-B*57:01(0.01) | HLA-B5801(0.01), HLA-B5701(0.01), HLA-A3201(0.30) | HLA-B*58:01(0.05), HLA-B*57:01(0.05) |
| 2 | 232 | 240 | IQKETLVTF | HLA-A*02:06(0.29), HLA-A*24:02(0.19), HLA-A*32:01(0.08), HLA-A*23:01(0.14), HLA-B*15:01(0.01) | HLA-B1501(0.05), HLA-A3201(0.40) | HLA-B*15:01(0.10) |
| 3 | 353 | 361 | TVNPIVTEK | HLA-A*31:01(0.21), HLA-A*30:01(0.03), HLA-A*03:01(0.01), HLA-A*68:01(0.01), HLA-A*11:01(0.01) | HLA-A6801(0.17), HLA-A1101(0.05) | HLA-A*68:01(0.05), HLA-A*11:01(0.10), HLA-A*03:01(0.30) |
| 4 | 170 | 178 | ITEAELTGY | HLA-A*30:02(0.29), HLA-A*01:01(0.01) | HLA-A0101(0.07) | HLA-A*01:01(0.05) |
| 5 | 238 | 246 | VTFKNPHAK | HLA-A*31:01(0.41), HLA-A*30:01(0.02), HLA-A*68:01(0.16), HLA-A*03:01(0.02), HLA-A*11:01(0.01) | HLA-A1101(0.17), HLA-A0301(0.30) | HLA-A*11:01(0.20), HLA-A*03:01(0.40) |
| 6 | 412 | 420 | MAILGDTAW | HLA-B*35:01(0.13), HLA-B*57:01(0.15), HLA-B*53:01(0.02), HLA-B*58:01(0.04) | HLA-B5801(0.01), HLA-B5701(0.10), HLA-B5301(0.02), HLA-B3501(0.09) | HLA-B*57:01(0.20), HLA-B*53:01(0.05), HLA-B*35:01(0.15) |
| 7 | 49 | 57 | EAKQPATLR | HLA-A*33:01(0.02), HLA-A*68:01(0.05) | HLA-A6801(0.40), HLA-A3301(0.15) | HLA-B*58:01(0.05) |
| 8 | 256 | 264 | QEGAMHTAL | HLA-B*44:02(0.34), HLA-B*44:03(0.34), HLA-B*40:01(0.06) | HLA-B4403(0.50), HLA-B4001(0.15) | HLA-B*40:01(0.30) |
| 9 | 383 | 391 | EPGQLKLSW | HLA-B*53:01(0.02) | HLA-B5301(0.06) | HLA-B*53:01(0.40) |
| 10 | 65 | 73 | LTNTTTASR | HLA-A*11:01(0.48), HLA-A*33:01(0.17), HLA-A*31:01(0.18), HLA-A*68:01(0.06) | HLA-A6801(0.30), HLA-A3301(0.40), HLA-A3101(0.40) | HLA-A*68:01(0.05), HLA-A*33:01(0.20), HLA-A*31:01(0.20) |
| 11 | 445 | 453 | GAAFSGVSW | HLA-A*32:01(0.29), HLA-B*53:01(0.18), HLA-B*57:01(0.15), HLA-B*58:01(0.06) | HLA-B5801(0.30), HLA-B5701(0.25) | HLA-B*58:01(0.30), HLA-B*57:01(0.40) |
| 12 | 213 | 221 | FLDLPLPWL | HLA-A*02:03(0.25), HLA-A*02:06(0.09), HLA-A*02:01(0.04) | HLA-A0206(0.50), HLA-A0201(0.09) | HLA-A*02:01 (0.30) |
| 13 | 202 | 210 | ENKAWLVHR | HLA-A*33:01(0.02) | HLA-A3301(0.01) | HLA-A*33:01(0.40) |
| 14 | 129 | 137 | IVQPENLEY | HLA-A*32:01(0.37), HLA-B*53:01(0.37), HLA-A*26:01(0.3), HLA-A*11:01(0.44), HLA-B*58:01(0.38), HLA-A*30:02(0.03), HLA-B*35:01(0.08), HLA-B*15:01(0.05), HLA-A*01:01(0.05) | HLA-B3501(0.50) | HLA-A*35:01 (0.15), HLA-A*01:01(0.20) |

**Table 1** (continued)

| Sl. No | Start | End | Epitope | IEDB MHC I Binding (percentile rank) | NetMHCpan (% rank) | NetCTLpan (% rank) |
|---|---|---|---|---|---|---|
| 15 | 457 | 465 | ILIGVVITW | HLA-A*24:02(0.49), HLA-A*23:01(0.25), HLA-B*53:01(0.22), HLA-A*32:01(0.04), HLA-B*58:01(0.13), HLA-B*57:01(0.17) | HLA-B5801(0.20), HLA-A3201(0.40) | HLA-B*58:01(0.40) |
| 16 | 438 | 446 | QVFGAIYGA | HLA-A*02:06(0.3), HLA-A*68:02(0.08) | HLA-A6802 (0.04), HLA-A0206 (0.25) | HLA-A*68:02(0.30) |
| 17 | 467 | 475 | GMNSRSTSL | HLA-B*15:01(0.44), HLA-A*02:03(0.1), HLA-B*08:01(0.06) | HLA-B0801(0.17), HLA-A0203(0.40) | HLA-A*02:03(0.30) |
| 18 | 453 | 461 | WTMKILIGV | HLA-A*02:06(0.29), HLA-A*68:02(0.09) | HLA-A6802 (0.01), HLA-A0206(0.07), HLA-A0203(0.50), HLA-A0201(0.30) | HLA-A*68:02(0.30) |
| 19 | 271 | 279 | QMSSGNLLF | HLA-A*23:01(0.32), HLA-A*24:02(0.35), HLA-A*32:01(0.15), HLA-B*15:01(0.16) | HLA-B1501 (0.40) | HLA-B*15:01(0.40) |
| 20 | 417 | 425 | DTAWDFGSL | HLA-A*26:01(0.23), HLA-A*68:02(0.11) | HLA-A6802(0.40), HLA-A2601(0.25) | HLA-A*26:01(0.20) |
| 21 | 463 | 471 | ITWIGMNSR | HLA-A*33:01(0.21), HLA-A*68:01(0.42), HLA-A*31:01(0.19) | HLA-A3301(0.25), HLA-A3101(0.50) | HLA-A*33:01(0.20), HLA-A*31:01(0.20) |
| 22 | 173 | 181 | AELTGYGTV | HLA-B*40:01(0.23) | HLA-B4001(0.15) | HLA-B*40:01( 0.30) |
| 23 | 298 | 306 | SYSMCTGKF | HLA-A*23:01(0.25), HLA-A*24:02(0.16) | HLA-A2402(0.25), HLA-A2301(0.50) | HLA-A*24:02(0.30), HLA-A*23:01(0.30 |
| 24 | 297 | 305 | MSYSMCTGK | HLA-A*11:01(0.33), HLA-A*03:01(0.33) | HLA-A6801(0.12), HLA-1101(0.25), HLA-A0301(0.06) | HLA-A*11:01(0.30), HLA-A*03:01(0.40) |
| 25 | 212 | 220 | WFLDLPLPW | HLA-A*24:02(0.26), HLA-A*23:01(0.16) | HLA-B5301(0.40), HLA-A2301(0.20) | HLA-A*24:02(0.40), HLA-A*23:01(0.30) |
| 26 | 293 | 301 | QLKGMSYSM | HLA-B*15:01(0.34), HLA-B*08:01(0.18) | HLA-B0801(0.20), HLA-A3201(0.50) | HLA-B*08:01(0.40) |
| 27 | 206 | 214 | WLVHRQWFL | HLA-B*08:01(0.31), HLA-A*02:01(0.33) | HLA-B0801(0.17), HLA-A0201(0.17) | HLA-B*08:01(0.30), HLA-A*02:01(0.20) |
| 28 | 210 | 218 | RQWFLDLPL | HLA-A*32:01(0.29) | HLA-B4001(0.40), HLA-B1501(0.50), HLA-A3201(0.01) | HLA-B*40:01(0.40), HLA-A*32:01(0.40) |

as IVQPENLEY, ILIGVVITW and DTAWDFGSL were selected for vaccine candidates (Table 2).

## Structure of the Alleles

Crystal structure of HLA-B*35:01 (PDB ID: 4LNR) and HLA-B*58:01 (PDB ID: 5VWH) were retrieved from the PDB database. The HLA-A*26:01 allele's sequence was obtained from the IPD-IMGT/HLA Database (Robinson et al. 2015) and used as the target protein sequence. The modeling of the target protein's 3D structure was done in a phased manner, beginning with an automated template structure search on the Swiss-Model server. A probable template structure with PDB-ID: 7RTD was chosen from a large number of hits as the basis for model construction. The query coverage for the target sequence is 98.16% and the sequence identity with the template sequence is 93.42%. Swiss-Model server generated a homology model of the target sequence based on the template and target alignment.

## Tertiary Structure Validation

Ramachandran plot analysis was used to validate the refinement results using the Swiss model/Structure assessment server. Ramachandran plot of the predicted model had 98.16% of the residues in the favored region

**Table 2** Immunogenicity, toxicity, antigenicity and allergenicity of best predicted epitopes

| Sl. No | Epitope | Immunogenicity | Toxicity | Antigenicity | Antigen/ Non- Antigen | Allergenicity |
|---|---|---|---|---|---|---|
| 1 | KAWLVHRQW | 0.10739 | Non-Toxin | 0.3107 | Non-antigen | Allergen |
| 2 | IQKETLVTF | 0.11563 | Non-Toxin | 0.1946 | Non-antigen | Allergen |
| 3 | TVNPIVTEK | 0.24646 | Non-Toxin | 1.0192 | Antigen | Allergen |
| 4 | ITEAELTGY | 0.21149 | Non-Toxin | 0.8778 | Antigen | Non-allergen |
| 5 | VTFKNPHAK | − 0.14558 | Non-Toxin | 0.5252 | Antigen | Allergen |
| 6 | MAILGDTAW | 0.14154 | Non-Toxin | 0.7313 | Antigen | Allergen |
| 7 | EAKQPATLR | − 0.13425 | Non-Toxin | 0.0901 | Non-antigen | Non-allergen |
| 8 | QEGAMHTAL | − 0.03456 | Non-Toxin | 0.0986 | Non-antigen | Non-allergen |
| 9 | EPGQLKLSW | − 0.42538 | Non-Toxin | 1.4199 | Antigen | Non-allergen |
| 10 | LTNTTTASR | 0.04766 | Non-Toxin | 0.7865 | Antigen | Non-allergen |
| 11 | GAAFSGVSW | − 0.06052 | Non-Toxin | 0.6497 | Antigen | Non-allergen |
| 12 | FLDLPLPWL | 0.09486 | Non-Toxin | 0.6739 | Antigen | Allergen |
| 13 | ENKAWLVHR | 0.22837 | Non-Toxin | 0.4696 | Antigen | Non-allergen |
| **14** | **IVQPENLEY** | **0.09179** | **Non-Toxin** | **1.2442** | **Antigen** | **Non-allergen** |
| **15** | **ILIGVVITW** | **0.29136** | **Non-Toxin** | **1.1507** | **Antigen** | **Non-allergen** |
| 16 | QVFGAIYGA | 0.25216 | Non-Toxin | − 0.4694 | Non-antigen | Allergen |
| 17 | GMNSRSTSL | − 0.3378 | Non-Toxin | 1.6491 | Antigen | Non-allergen |
| 18 | WTMKILIGV | − 0.02272 | Non-Toxin | 0.6229 | Antigen | Non-allergen |
| 19 | QMSSGNLLF | − 0.2091 | Non-Toxin | 0.5204 | Antigen | Allergen |
| **20** | **DTAWDFGSL** | **0.29933** | **Non-Toxin** | **1.8229** | **Antigen** | **Non-allergen** |
| 21 | ITWIGMNSR | − 0.0286 | Non-Toxin | 2.1549 | Antigen | Non-allergen |
| 22 | AELTGYGTV | 0.11626 | Non-Toxin | 0.6621 | Antigen | Allergen |
| 23 | SYSMCTGKF | − 0.34376 | Non-Toxin | − 0.1358 | Non-antigen | Non-allergen |
| 24 | MSYSMCTGK | − 0.33686 | Non-Toxin | 0.3117 | Non-antigen | Non-allergen |
| 25 | WFLDLPLPW | − 0.01836 | Non-Toxin | 0.3261 | Non-antigen | Non-allergen |
| 26 | QLKGMSYSM | − 0.46241 | Non-Toxin | 0.8523 | Antigen | Non-allergen |
| 27 | WLVHRQWFL | 0.24265 | Non-Toxin | 0.3188 | Non-antigen | Allergen |
| 28 | RQWFLDLPL | 0.18394 | Non-Toxin | − 0.1195 | Non-antigen | Allergen |

Selected epitopes are highlighted in bold

and 0.74% in the outlier region as shown in Fig. 1a. The total quality score for a given input structure is calculated by the ProSA-web server and displayed in the context of all known protein structures. The Model protein Z score was -9.05 using the ProSA web server as shown in Fig. 1b, in the broad black dot.

## Docking Analysis

The Hpepdock server generated ten conformations, and the best one was chosen based on the binding energy score. The stronger the binding contact between the HLA allele and the epitope, the lower the binding energy. Table 3 displays the docking scores of peptide-allele complexes.

Interactions between the peptides from their respective allele were shown in Figs. 2 a, b & c.

## Conservancy Across the Sequences

In a similarity search tool (BLASTP) against the nr database, envelope protein produced 80 homologous sequences. In our investigation, protein sequences with better than 99 percent sequence identity and 100 percent query coverage were considered homologous. Table 2 displays the conservation value of selected CTL epitopes among tested Envelope protein homologous sequences. In this analysis, we discovered that IVQPENLEY and DTAWDFGSL have 100% conservancy, while ILIGVVITW has 98.75% conservancy (Table 4).
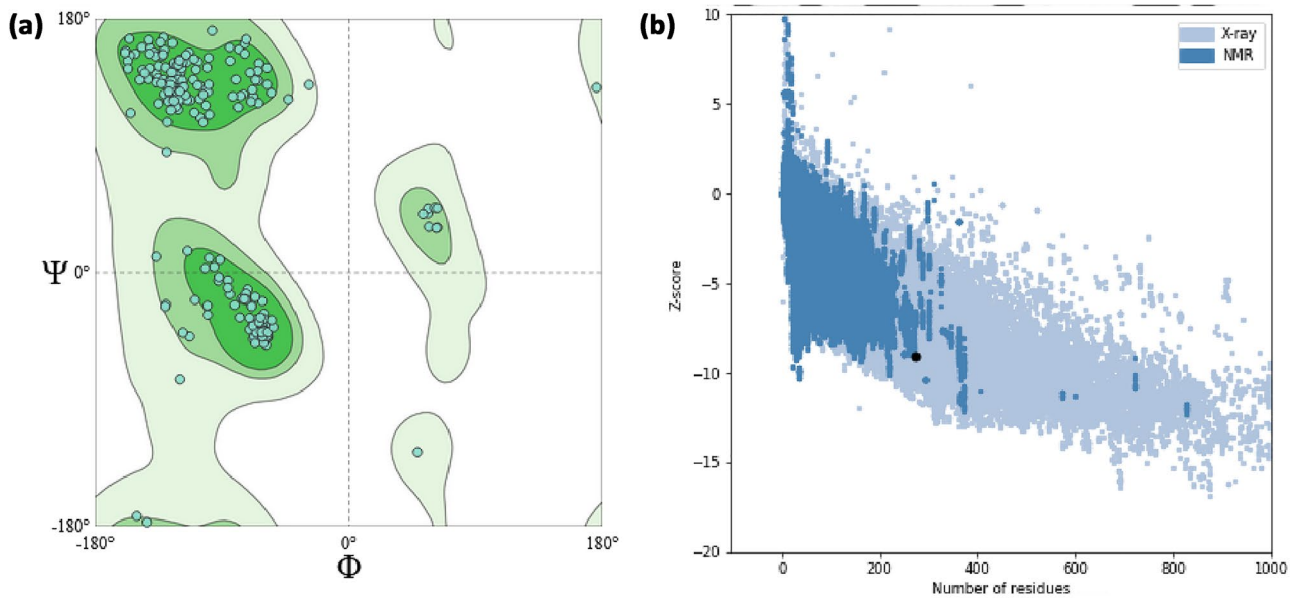
**Fig. 1** The Ramachandran plot and the ProSA-web server were used to validate the tertiary structure. **a** The ProSA-web result provides a Z score of − 9.05; **b** Ramachandran plot shows that the amount of amino acid residues in the favorable region is 98.16%

**Table 3** Docking results between selected epitopes and HLA alleles

| Peptide sequence | Allele | Docking score |
| --- | --- | --- |
| IVQPENLEY | HLA-B*35:01 | − 192.383 |
| ILIGVVITW | HLA-B*58:01 | − 274.972 |
| DTAWDFGSL | HLA-A*26:01 | − 262.574 |

**Table 4** The potential T -cell epitopes conservancy

| Protein Name | Epitope sequence | Epitope conservancy (%) |
| --- | --- | --- |
| Envelope protein | IVQPENLEY | 100.00% |
|  | ILIGVVITW | 98.75% |
|  | DTAWDFGSL | 100.00% |

## Worldwide Population Coverage Analysis

The selected MHC class-I epitopes used for vaccine construction and their respective HLA binding alleles as predicted in Table 1, were obtained to assess worldwide population coverage. MHC class-I epitopes offered a high percentage of global population coverage (Figs. 3, 4, and 5). The selected epitopes exhibited interactions with various other HLA alleles from different countries. This analysis suggests that the designed vaccine could be an efficient candidate for most of the population across the world.

## Discussion

Immunoinformatics approaches are gradually becoming accepted as the first line of vaccine development in the production of effective vaccines against viruses. In recent years, immunoinformatics methods were used in the development of an epitope-based vaccine for *C.auris* (Akhtar et al. 2021a), Dengue virus (Krishnan et al., 2020), Orthohantavirus (Joshi et al. 2022), Marburg virus (Kumar et al. 2013), ebola virus (Saraswat et al. 2012),Japanese encephilitis virus (Sharma et al. 2014), Zika virus (Sharma et al. 2021) and SARS-CoV-2 (Sarkar et al. 2020; Rahman et al. 2020). On a global scale, approximately three billion people are at risk of DENV infection (Thomas and Rothman 2015). There is currently no specific therapy for dengue fever, and the major preventive method is the vaccine to limit the disease's burden. Hence, the goal of this study was to use immunoinformatics approaches to build an epitope-based peptide vaccine against the dengue virus. In comparison to traditional vaccine design, the epitope-based vaccine has a positive effect (Reginald et al. 2018). Different immunoinformatics tools were used to screen T-cell epitopes found in the envelope protein. Immunodominant epitopes are regions of protein antigens bound to immunologic receptors (Ayub et al. 2016; Singh et al. 2015a, b).Therefore, they were widely used as vaccine and therapeutic composition (Ninomiya et al. 2002; Adame-Gallegos et al. 2012; McComb et al. 2015). Twenty eight Tcell epitopes predicted by all servers were chosen and additionally picked out on the grounds of their antigenic reactivity, immunogenicity, toxicity and
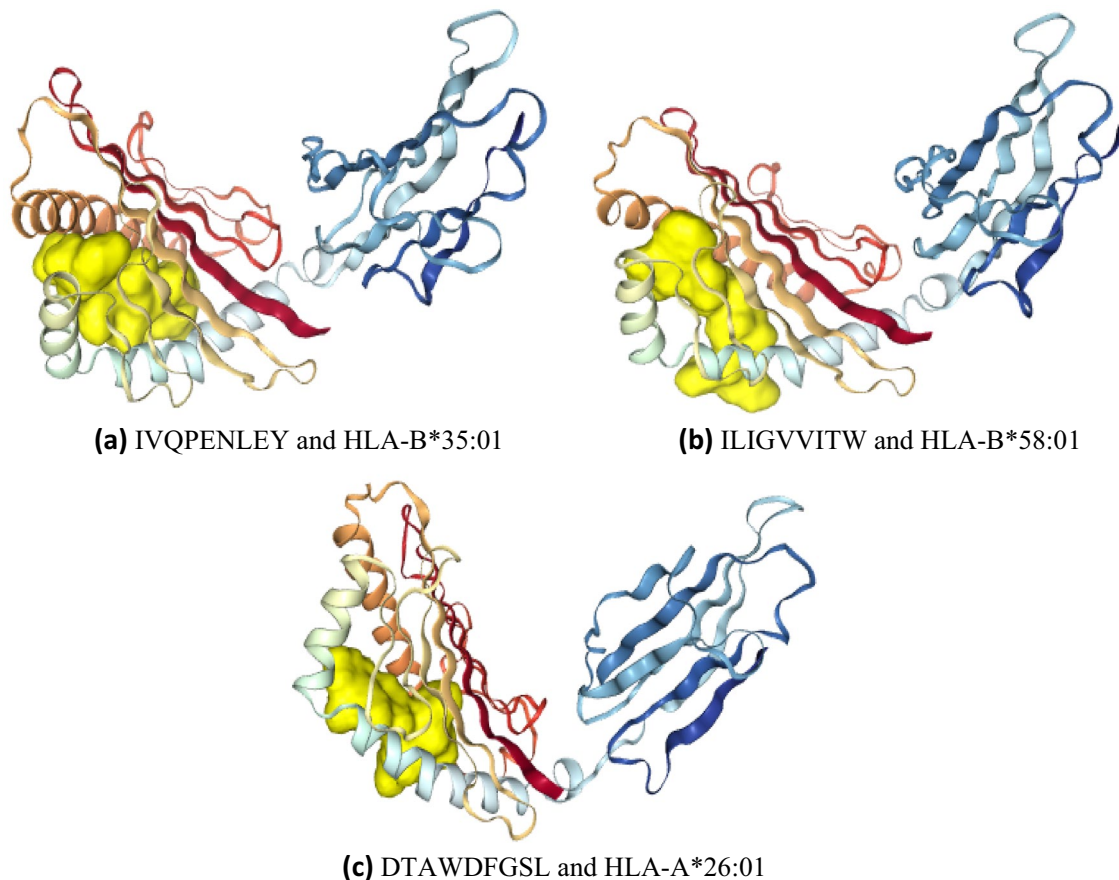
**(a)** IVQPENLEY and HLA-B*35:01



**(b)** ILIGVVITW and HLA-B*58:01



**(c)** DTAWDFGSL and HLA-A*26:01

**Fig. 2** The interactions between epitopes and alleles are depicted in this diagram. Here, **a** is the interaction between IVQPENLEY and HLA-B*35:01 allele, **b** is the interaction between ILIGVVITW and HLA-B*58:01 allele, and **c** is the interaction between DTAWDFGSL and HLA-A*26:01 allele

allergenicity assessment. VaxiJen tool was used to classify viral components into antigens and non-antigens using a 0.4 threshold. Fifteen epitopes had a positive immunogenicity value, according to the results of the immunogenicity analysis. A high immunogenicity score indicates a strong ability to stimulate naive T cells. Three potent T cell epitopes as IVQPENLEY, ILIGVVITW had antigenicity scores greater than one, non-toxic and non-allergenic were screened. The Ramachandran plot shows that 98.16% of the residues are in the favored regions and 0.74% are in the outlier region, indicating that the overall model quality is excellent. The model structure was further checked by the ProSA-Web server for any potential errors. Based on the input and Z-score of the predicted model, this server generates a graphical output of overall and local model quality. The Z-score of the modeled CP40 protein was − 5.26 in peptide vaccine design using immunoinformatic approach (Droppa-Almeida et al. 2018), whereas the Z-score of the modeled MRE11 protein was − 9.5 (Rekik et al. 2015). This concludes that the anticipated 3D structures were reliable and of high quality. Overall, the validation tool results demonstrate that the predicted model

structure is appropriate and of high quality. Protein–protein docking is commonly used in immunoinformatics to examine the stability of vaccines by looking at their binding energies. The best predicted epitopes against MHC alleles (HLA-B*35:01, HLA-B*58:01 and HLA-A*26:01) were examined using this method. The conservancy analysis, we discovered that IVQPENLEY and DTAWDFGSL have 100% conservancy, while ILIGVVITW has 98.75% conservancy. Previously this of conservancy analysis conducted for dengue virus (Akhtar et al 2021b) Immunoinformatics methods are particularly helpful for doing *In-silico* studies and can guide laboratory experiments, saving time and money. Nonetheless, the next step is to perform in vitro immunological experiments to validate the predicted epitopes and establish their immunogenicity.
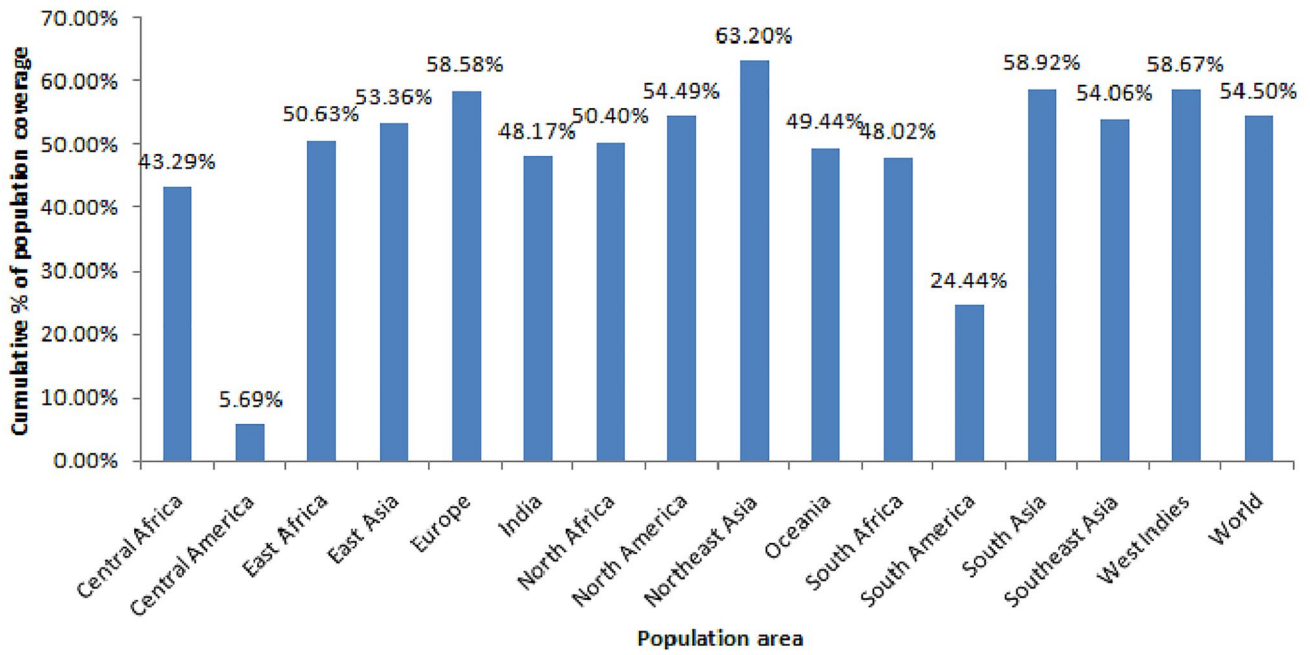
**Fig. 3** Population coverage analysis predicted based on epitope IVQPENLEY along with their respective alleles
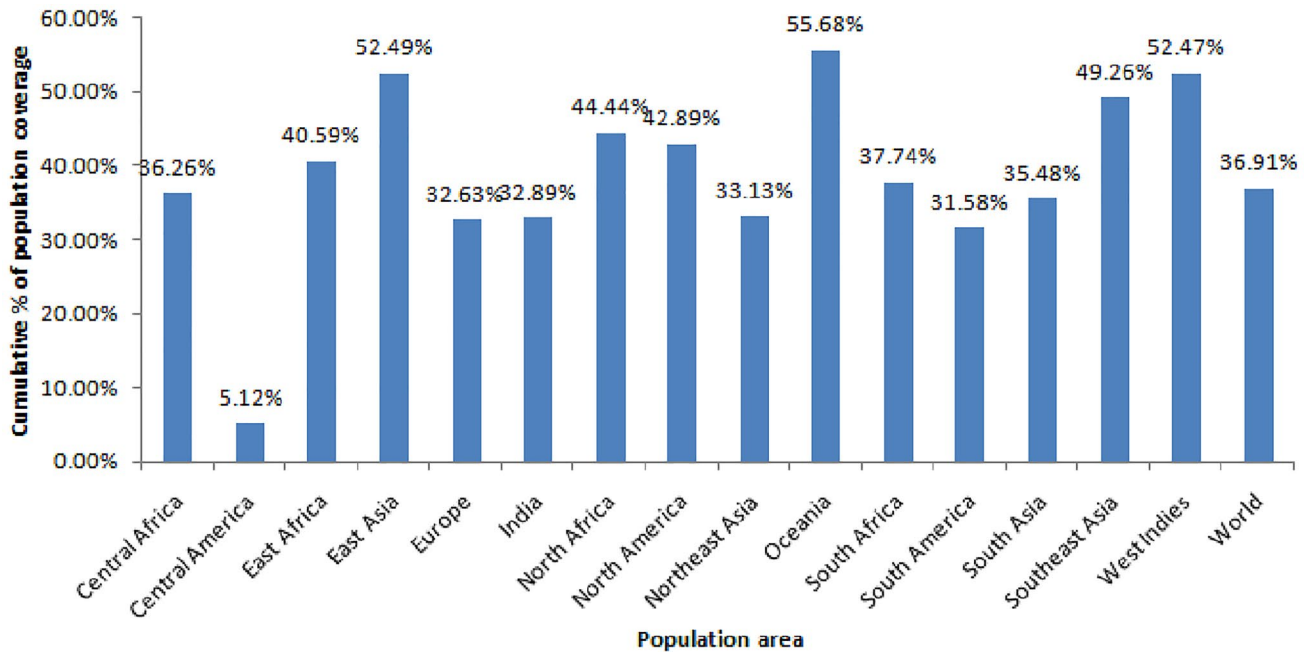


**Fig. 4** Population coverage analysis prediction on the basis of epitope ILIGVVITW along with their respective alleles
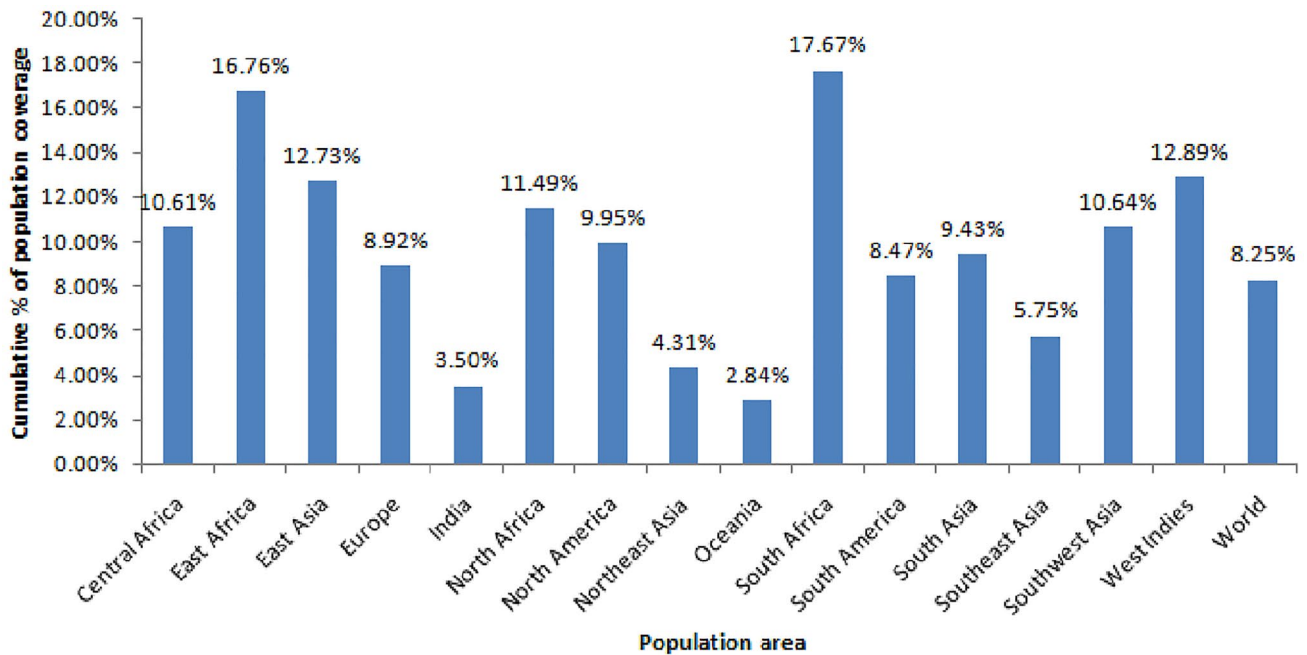
**Fig. 5** Population coverage analysis prediction based on epitope DTAWDFGSL along with their respective alleles

## Conclusion

Three potent T cell epitopes IVQPENLEY, ILIGVVITW and DTAWDFGSL were identified as having high antigenicity scores. These epitopes had the lowest docking score with their associated alleles. The predicted epitopes cover the maximum number of national and international populations. As a result, the current research aims to identify the most promising vaccine candidates from envelope protein in the shortest amount of time and with the least amount of experimental effort, with the goal of reducing the worldwide burden of dengue infection.

## Declarations

**Conflict of interest** I confirm that the authors hereby declare they that have no conflict of interest.

**Ethical approval** I confirm that the authors did not perform any experiments on humans or animals.

## References

Adame-Gallegos JR, Shi J, McIntosh RS, Pleass RJ (2012) The generation and evaluation of two panels of epitope-matched mouse IgG1, IgG2a, IgG2b and IgG3 antibodies specific for Plasmodium falciparum and Plasmodium yoelii merozoite surface protein 1–19 (MSP119). Exp Parasitol 130:384–393

Akhtar N, Joshi A, Kaushik V, Kumar M, Amin-ul Mannan M (2021a) In-silico design of a multivalent epitope-based vaccine against Candida auris. Microb Pathog 155:104879

Akhtar N, Joshi A, Singh J, Kaushik V (2021b) Design of a novel and potent multivalent epitope based human cytomegalovirus peptide vaccine: an immunoinformatics approach. J Mol Liq 335:116586

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Andreatta M, Nielsen M (2016) Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics 32(4):511–517

Ayub G, Waheed Y, Najmi MH (2016) Prediction and conservancy analysis of promiscuous T-cell binding epitopes of Ebola virus L protein: an in silico approach. Asian Pac J Trop Dis 6:169–173

Bhatt S, Gething PW, Brady OJ et al (2013) The global distribution and burden of dengue. Nature 496(7446):504–507

Brady OJ, Gething PW, Bhatt S et al (2012) Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLOS Negl Trop Dis 6(8):e1760

Bui HH, 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 Bioinform 8(1):361

Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD et al (2013) Properties of MHC class I presented peptides that enhance immunogenicity. PLoS Computat Biol 9:e1003266

Centers for Disease Control and Prevention. CDC (2020) https://www.ecdc.europa.eu/en/dengue-monthly

Choudhary MC, Gupta E, Sharma S, Hasnain N, Agarwala P (2017) Genetic signatures coupled with lineage shift characterise endemic evolution of Dengue virus serotype 2 during 2015 outbreak in Delhi, India. Trop Med Int Health 22(7):871–880

Deval H, Behera SP, Agrawal A, Singh R, Misra B, Janardhan V, Patil G, Sah K, Kumar N, Singh R, Bondre VP (2021) Genetic characterization of dengue virus serotype 2 isolated from dengue fever outbreaks in eastern Uttar Pradesh and western Bihar India. J Med Virol 93(6):3322–3329

Dimitrov I, Bangov I, Flower DR, Doytchinova I (2014) "AllerTOP v. 2 - a server for in silico prediction of allergens. J Mol Model 20(6).

Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinform 8:4

Droppa-Almeida D, Franceschi E, Padilha FF (2018) Immune-informatic analysis and design of peptide vaccine from multi-epitopes against corynebacterium pseudotuberculosis. Bioinform Biol Insights 12:1177932218755337

Foster JE, Bennett SN, Carrington CV, Vaughan H, McMillan WO (2004) Phylogeography and molecular evolution of dengue 2 in the Caribbean basin, 1981–2000. Virology 24(1):48–59

Gupta S et al (2013) In silico approach for predicting toxicity of peptides andproteins. PLoS ONE 8:e73957

Joshi A, Ray NM, Singh J et al (2022) T-cell epitope-based vaccine designing against Orthohantavirus: a causative agent of deadly cardio-pulmonary disease. Netw Model Anal Health Inform Bioinform 11

Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P et al (2012) Immune epitope database analysis resource. Nucleic Acids Res 40:W525-530

Krishnan GS, Joshi A, Kaushik V (2020a) T cell epitope designing for dengue peptide vaccine using docking and molecular simulation studies. Mol Simul 46(10):787–795

Krishnan GS, Joshi A, Kaushik V (2020b) T cell epitope designing for dengue peptide vaccine using docking and molecular simulation studies. Mol Simul. https://doi.org/10.1080/08927022.2020.1772970

Kumar A, Jain A, Verma SK (2013) Screening and structure-based modeling of T-cell epitopes of Marburg virus NP, GP and VP40: an immunoinformatic approach for designing peptide-based vaccine. Trends Bioinform 6(1):10

Liu Y, Liu J, Cheng G (2016) Vaccines and immunization strategies for dengue prevention. Emerg Microbes Infect 5(7):e77

Markoff L (2003) 5′- and 3′-noncoding regions in flavivirus RNA. Adv Virus Res 59:177–228

McComb RC, Ho CL, Bradley KA, Grill LK, Martchenko M (2015) Presentation of peptides from Bacillus anthracis protective antigen on tobacco mosaic virus as an epitope targeted anthrax vaccine. Vaccine 33:6745–6751

Murphy BR, Whitehead SS (2011) Immune response to dengue virus and prospects for a vaccine. Annu Rev Immunol 29:587–619

Murrell S, Wu SC, Butler M (2011) Review of dengue virus and the development of a vaccine. Biotechnol Adv 29(2):239–247

National center for vector borne diseases control. NVBDCP (2019) Dengue cases and deaths in the country since (2015) https://nvbdcp.gov.in/index4.php?lang=1&level=0&linkid=431&lid=3715). Accessed 28 Aug 2019.

National center for vector borne diseases control. NVBDCP (2021) https://nvbdcp.gov.in/index4.php?lang=1&level=0&linkid=431&lid=3715. Accessed Oct 2021.

Ninomiya A, Ogasawara K, Kajino K, Takada A, Kida H (2002) Intranasal administration of a synthetic peptide vaccine encapsulated in liposome together with an anti-CD40 antibody induces protective immunity against influenza A virus in mice. Vaccine 20:3123–3129

Proutski V, Gritsun TS, Gould EA, Holmes EC (1999) Biological consequences of deletions within the 3′-untranslated region of flavi-viruses may be due to rearrangements of RNA secondary structure. Virus Res 64(2):107–123

Rahman N, Ali F, Basharat Z, Shehroz M, Khan MK, Jeandet P, Nepovimova E, Kuca K, Khan H (2020) Vaccine design from the ensemble of surface glycoprotein epitopes of SARS-CoV-2: an immunoinformatics approach. Vaccines 8(3):423

Reginald K, Chan Y, Plebanski M, Poh CL (2018) Development of peptide vaccines in dengue. Curr Pharm Des 24:1157–1173

Rekik I, Chaabene Z, Grubb CD et al (2015) In silico characterization and Molecular modeling of double-strand break repair protein MRE11 from Phoenix dactylifera v deglet nour. Theor Biol Med Model 12:23

Rico-Hesse R, Harrison LM, Salas RA et al (1997) Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. Virology 230:244–251

Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG (2015) The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res 43(Database Issue) D423–D431.

Saraswat A, Shraddha JA, Pathak A, Verma SK, Kumar A (2012) Immuno-informatic speculation and computational modeling of novel MHC-II human leukocyte antigenic alleles to elicit vaccine for ebola virus. J Vaccines Vaccin 3(141):2

Sarkar B, Ullah MA, Johora FT, Taniya MA, Araf Y (2020) Immuno-informatics-guided designing of epitope-based subunit vaccines against the SARS Coronavirus-2 (SARS-CoV-2). Immunobiology 225:151955

Sharma P, Saxena K, Mishra S, Kumar A (2014) A comprehensive analysis of predicted HLA binding peptides of JE viral proteins specific to north Indian isolates. Bioinformation 10(6):334

Sharma P, Sharma P, Kumar A (2021) Top down computational approach: a vaccine development step to find novel superantigenic HLA binding epitopes from dengue virus proteome. Int J Pept Res Ther 27(2):1469–1480

Singh S, Singh H, Tuknait A, Chaudhary K, Singh B, Raghava KS, GPS (2015a) PEPstrMOD: structure prediction of peptides containing natural, non-natural and modified residues. Biol Direct 10:73

Singh SK, Roeffen W, Andersen G et al (2015b) A Plasmodium falciparum 48/45 single epitope R0.6C subunit protein elicits high levels of transmission blocking antibodies. Vaccine 33:1981–1986

Srivastava S, Sharma SK, Srivastava V, Kumar A (2020) Proteomic exploration of Listeria monocytogenes for the purpose of vaccine designing using a reverse vaccinology approach. Int J Pept Res Ther. https://doi.org/10.1007/s10989-020-10128-1

Stranzl T, Larsen MV, Lundegaard C, Nielsen M (2010) NetCTLpan: pan-specific MHC class I pathway epitope predictions. Immunogenetics 62:357–368

Thomas SJ, Rothman AL (2015) Trials and tribulations on the path to developing a dengue vaccine. Vaccine 33:D24–D31

Twiddy SS, Farrar JJ, Vinh Chau N et al (2002) Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. Virology 298:63–72

Wang E, Ni H, Xu R et al (2000) Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. J Virol 74(7):3227–3234

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, De Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303

Xiang Z, He Y (2009) Vaxign: a web-based vaccine target design program for reverse vaccinology. Procedia Vaccinol 1(1):23–29

Zhou P, Jin B, Li H, Huang S-Y (2018) HPEPDOCK: a web server for blind peptide-protein docking based on a hierarchical algorithm. Nucleic Acids Res. https://doi.org/10.1093/nar/gky357