https://doi.org/10.48047/AFJBS.6.14.2024.2131-2151

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Identification of CTL epitopes on envelope protein of Zika virus using Immunoinformatics approach

Kirti[1], Inamul Hasan[2]and VivekSrivastava[1*]

[1]Department of Biotechnology, Faculty of Engineering and Technology, Rama University, Kanpur, UP, India

[2]CSIR-CIMAP, Lucknow, Uttar Pradesh, India.

*Corresponding author
viveksrivastavabio@gmail.com

Article History

Volume 6, Issue 14, Aug 2024

Received: 15 July 2024

Accepted: 7 August 2024

Published: 8 August 2024

doi: 10.48047/AFJBS.6.14.2024.2131-2151

Abstract

The spread of *Zika* virus infection in India has become a serious public health concern due to its associationwith birth defects and neurological disorders.The development of an effective vaccine to prevent viral infection is urgently needed.In this study, epitope-based peptide vaccine components based on the *Zika* virus envelope protein were developed using the reverse vaccinology method.The envelope protein was examined for cytotoxic T lymphocyte (CTL) epitopes using immunoinformaticstools.Common epitopes predicted by variousmethods were chosen and evaluated for immunogenicity score, toxicity, antigenicity,and allergenicityassessment as well as human homology.Four potent CTL epitopes were identified as DTAWDFGSV, SIQPENLEY, GTVTVEVQY, and MMLELDPPF.These predicted epitopes were identified to cover the maximum number of populations in India and worldwide as well as conserved across different stains sequences. Further, thePepstrMod server was used to build the 3D structure of predicted epitopes. The sequence of theHLA-A*68:02allele was downloaded from IPD-IMGT/HLA database and model structure with the SWISS-MODELserver.HLA-A*01:01, HLA-A*03:01and HLA-B*35:01alleles had known crystal structures, whichwere retrieved from Protein Data Bank.Epitope SIQPENLEY and GTVTVEVQY were docked with allele HLA-A*01:01using the HPEPDOCK server.DTAWDFGSV and MMLELDPPF epitopes were docked with alleles HLA-A*68:02 and HLA-B*35:01, respectively.Docked epitopes exhibited a good docking score with alleles. After that,the docked epitope-allele complex was subjected to molecular structure dynamics simulation to determine the stability of the complex. GROMACS 5.1.4 package was used to study the structural properties and interaction between epitope and allele at the atomic level. Molecular dynamics simulation confirmed the stable nature of the docked complexes.These discovered epitopes were expected to have a high potential for eliciting an immune response in the development of a *Zika* virus vaccine.While additional experimental validation is needed for definitive confirmation.

Keywords: *Zika* virus, CTL epitopes, MHC Class I, vaccine.

Introduction

The *Flaviviridae* virus family includes the vector-borne *Zika* virus (ZIKV), which is spread by the bite of infected *Aedes* mosquitoes, primarily *Aedesaegypti* and *Aedesalbopictus*(Hamel et al., 2015).During yellow fever (YF) monitoring, ZIKV was initially discovered in 1947 in the *Zika* forest of Africa in rhesus monkey serum(Dick, 1952). The *Zika* virus disease (ZVD) has symptoms that are comparable to those of other flavivirus diseases like Dengue, West Nile, Yellow Fever, and Japanese Encephalitis (Pielnaa et al., 2020). Blood transfusions, sexual contact, and vertical contact can also spread the virus (Rothan et al., 2018). In 1952, the first human ZIKV case was reported, and over the next several decades, sporadic mild cases from Asia and Africa were reported. The current outbreaks of the *Zika* virus in India are extremely worrying. 159 *Zika* virus infections were reported in Rajasthan at the end of 2018 (Yadav et al., 2019), 127 Madhya Pradesh cases (Saxena et al., 2019). It was confirmed that a citizen of Kerala state in southwest India had contracted the *Zika* virus on July 8, 2021 (WHO, 2021). The first laboratory-confirmed *Zika* infection in Maharashtra was reported on July 31, 2021, from Belsar, a village in the Pune district (WHO, 2021).Initially, ZVD causes moderate sickness such as fever, myalgia, headache, rash, arthritis, and conjunctival congestion, with up to $4/5^{th}$ of infected people being asymptomatic (Duffy et al., 2009; Calvet et al., 2016). ZVD has also been implicated in the development of Guillain-Barre Syndrome (GBS)(Cao-Lormeau et al., 2016).On February 1, 2016, the World Health Organization declared a Public Health Emergency of International Concern, at which time autochthonous ZIKV transmissions were reported in 22 countries and territories in Latin America and the Caribbean(Yakob and Walker, 2016).

There are no effective antivirals or vaccines available for *Zika* virus infection at the moment (Saiz and Martín-Acebes, 2017).Identifying CTL epitopes has become a vital part of epitope-based vaccine design. The wet experiment for epitope identification is time-consuming, labor-intensive, and expensive. With the growing availability of experimentally produced epitopes, it is now possible to develop computer algorithms for epitope prediction that are both faster and less expensive (Flower, 2007). We utilized immunoinformatics tools to examine the envelope protein of the *Zika* virus and discovered CTL epitopes that were antigenic and consensus in several epitope–allele binding algorithms.Epitope-based vaccines outperform conventional vaccines and potentially address safety and adverse effect problems (Srivastava et al., 2021; Krishnan et al.,

2020). This study established a theoretical background for *Zika* virus vaccine development and prevention.

2. Materials and methods

An immunoinformatics and reverse vaccinology technique was used to develop potential vaccine componentsagainst the envelope protein of the *Zika* virus.

2.1. Retrieval of the envelopeprotein sequence

The ZIKV envelope protein amino acid sequence was extracted from a polyprotein sequence (GenBank ID: AZS35409.1) that was deposited in the NCBI protein database. The Vaxign version 2.0 beta server (Xiang and He, 2009) was used to investigate envelope protein as a potential vaccine target for the development of vaccine components.

2.2.Instability and allergenicityprediction of the envelope protein

The Protparam software was used to predict the physicochemical properties of the envelope protein, such as the instability index, grand average of hydropathicity (GRAVY), theoretical pI, and molecular weight (Walker, 2005; Wilkins et al., 1999). The allergenicity of the protein was evaluated using AllerTOP v.2.0 (Dimitrov et al., 2014).

2.3 Prediction of CTL epitopes

To identify CTL epitopes of the envelope protein, two separate servers, the IEDB MHC-I prediction tool (Kim et al., 2012) and NetCTLpan1.1 (Stranzl et al., 2010), were employed. The IEDB MHC-I prediction approach identifies epitopes that may interact with MHC-I genes. The NetCTLpan1.1 tool predicts epitopes of MHC class I using the weight matrix, the efficiency of TAP transport, and ANN.

2.4 Immunogenicity,toxicity, antigenicity andallergenicity analysis of the epitopes

The epitopes identified by two servers were subjected to the IEDB class I immunogenicity tool using default parameters (Calis et al., 2013). The antigenicity of the epitopes was calculated using the VaxiJen v2.0 server, with a 0.4 threshold used to determine the accuracy of the prediction (Doytchinova and Flower, 2007). The server AllerTOP v.2.0 was used to compute the allergenicity

of epitopes.AllerTOP v.2.0 is an alignment-independent online allergenicity prediction tool that yields reliable results (Dimitrov et al., 2014). Furthermore, ToxinPred was utilized to predict toxicity using an SVM-based algorithm with default parameters (Gupta et al., 2013).

2.5 Conservancy analysis of epitopes

Conserved epitopes are expected to provide greater cross-species protection than epitopes from highly diverse genomic regions. Using the BLASTP software, similar sequences were discovered by comparing the amino acid sequence of the envelope to the nr sequence database (Altschul et al., 1997). To determine the conservancy of CTL epitopes used in candidate vaccine design among screened homologs, the IEDB's conservancy analysis tool (Bui et al., 2007) was used.

2.6 Population coverage analysis

A vaccine molecule must give broad-spectrum disease protection in distinct world populations in order to be effective. The IEDB population coverage tool (http://tools.iedb.org/population/) was used to analyze epitope population coverage.

2.7Peptide modeling

After multiple bioinformatics investigations, only the specified epitopes were chosen for the peptides and HLAs interaction pattern analysis. The PEPstrMOD was used to construct a three-dimensional model of chosen CTL epitopes(Singh et al., 2015).The technique is predicated on the concept that, in addition to regular structures, the β-turn is a necessary and consistent attribute of short peptides.The structure is refined using energy minimization and molecular dynamic simulations.

2.8Allele modeling

Protein Data Bank was used to retrieve HLA alleles with known crystal structures. The sequence of HLA alleles with unknown structures was obtained from the IPD-IMGT/HLA database (Robinson et al. 2015). The SWISS-MODEL server was used to model the 3D structures of these HLA alleles (Waterhouse et al., 2018).

2.8.1 Tertiary structure validation

The Swiss-Model Structure Assessment page (https://swissmodel.expasy.org/assess) was used to validate the refined model's structure. For confirming the protein structures, another tool, ProSA-web was utilized (Wiederstein and Sippl, 2007). ProSA takes a protein structure (PDB file) as input and generates a plot of the overall model quality score (z-score). If the z-score falls outside of the normal range for natural proteins, the structure most likely contains errors.

2.9Molecular docking

An online HPEPDOCK server was used to perform molecular docking of epitopes with their corresponding HLA binding alleles (Zhou et al., 2018).HPEPDOCK is a server that performs blind peptide-protein docking using a hierarchical approach.

2.10 Molecular Dynamics simulation

The effective method for the molecular analysis of biological systems is known to be molecular dynamics simulation (Hospital et al., 2015; Lazim and others, 2020; Sinha et al., 2022). It was utilized in previous research to examine the stability of various protein complexes (Hajighahramani et al., 2017; Nezafat and other, 2016; Narula and other, 2018).After performing the epitope-allele docking study, the docked complex was subjected to molecular structure dynamics simulation to determine the stability of the epitope-allele complex. The GROMACS 5.1.4 package was used to study the structural properties and interaction between epitope and allele at the atomic level. All molecular dynamic simulation was performed using the CHARMM27 force field (Abraham et al., 2015; Vanommeslaeghe et al., 2010). To guarantee that the geometry of the system is adequate and no steric clashes are occurring using the steepest descent algorithm approach energy, minimization was performed prior to simulation. During the equilibration phase (100 ps), the temperature was increased up to 300K and the pressure up to 1 bar. Finally, the trajectories generated from the simulation (20000 Ps) were analyzed for the stability of the complex in terms of the root mean square deviation (RMSD), root mean square fluctuation (RMSF), and Rg (Radius of Gyration) of the epitope-allelecomplex.The RMSD, RMSF, and Rg can be used to obtain information about our bio systems. The RMSD includes the calculation concerning the reference structure (backbone) of the average atom location in a molecule. This method is used to calculate the average changes in atom displacement for evaluating conformational shifts and bio systems stability. The average deviation of a particle (for example,

a protein residue) from the reference location (usually the average particle location) over time is calculated with root-mean-square fluctuation (RMSF). RMSF, therefore, analyzes the structural portions which fluctuate more (or less) from their mean structure. The gyration radius (Rg) represents the compactness of the structure. The lower degree of fluctuation with its simulation stability indicates the higher compactness and rigidity of the device.

3. Results

3.1 Vaccine target

Using the Vaxign version 2.0 beta server, the envelope protein was explored as a potential vaccine target (Xiang and He, 2009).In envelope protein, two transmembrane helixes were predicted. The predicted adhesion probability for this protein is 0.713. The protein is an adhesion if the probability of adhesion is greater than 0.51. The predicted protein bears no similarity to proteins found in humans.

3.2Envelope protein sequence analysis

The protein antigenicity score of 0.6268 shows that the vaccine antigenic characteristic is good. The protein was non-allergenic and had a molecular weight of 54410.22 Da.The protein exhibits a negative GRAVY value (-0.087), indicating that it was hydrophilic and interacted extensively with water molecules. The envelope protein has a value of 22.77 for instability, indicating that it is a stable protein. The calculated pI of the protein is 6.51, indicating that it is acidic.

3.3 CTL epitopes analysis

To predict epitopes with a length of 9 mers, we used the recommended IEDB 2020.09 (NetMHCpan EL 4.1) prediction method and HLA allele reference set, which cover the largestnumber of people in the world. The set of predicted binders was made for IEDB MHC class I epitope based on the percentile range < 0.5 to cover the top most immune responses. The threshold value for the NetCTLpan1.1 server was set to 1.0 in our calculations. To improve accuracy, we selected those epitopes, which are commonly predicted by these two servers. Antigenic protein was used to pick 42 commonly predicted epitopes (Table 1) from these two servers.

Table 1: Identification of MHC-I epitopes bytwo servers

S.No	Start	End	Epitope	IEDB MHC I Binding (percentile rank)	NetCTLpanb(% rank)
1	454	462	KSLFGGMSW	HLA-B*57:01(0.01),HLA-B*58:01(0.01), HLA-A*32:01(0.02)	HLA-A*32:01(0.40), HLA-B*57:01 (0.05), HLA-B*58:01(0.05)
2	392	400	GEKKITHHW	HLA-B*44:02(0.01),HLA-B*44:03(0.01)	HLA-B*44:02(0.05),HLA-B*44:03(0.15)
3	320	328	ETLHGTVTV	HLA-A*68:02(0.01),HLA-A*26:01(0.29)	HLA-A*68:02(0.05)
4	261	269	QEGAVHTAL	HLA-B*40:01(0.02),HLA-B*44:03(0.18)	HLA-B*40:01(0.15),HLA-B*44:02(0.80), HLA-B*44:03(0.80)
5	244	252	EFKDAHAKR	HLA-A*33:01(0.01)	HLA-A*33:01(0.05)
6	195	203	GLDFSDLYY	HLA-A*01:01(0.02),HLA-A*30:02(0.3)	HLA-A*01:01(0.05),HLA-A*30:02(0.80)
7	297	305	KLRLKGVSY	HLA-B*15:01(0.01),HLA-A*03:01(0.27), HLA-A*30:02(0.12),HLA-A*30:01(0.1), HLA-A*32:01(0.18)	HLA-A*03:01(0.80),HLA-A*30:01(0.80), HLA-A*30:02 (0.80),HLA-B*15:01(0.80)
8	106	114	GLFGKGSLV	HLA-A*02:03(0.03),HLA-A*02:01(0.12), HLA-A*02:06(0.38)	HLA-A*02:03(0.15)
9	48	56	TTVSNMAEV	HLA-A*68:02(0.02),HLA-A*02:06(0.29), HLA-A*26:01(0.35)	HLA-A*26:01(0.80),HLA-A*68:02(0.05)
10	421	429	MAVLGDTAW	HLA-B*58:01(0.06),HLA-B*53:01(0.03), HLA-B*35:01(0.18)	HLA-B*35:01(0.15),HLA-B*53:01(0.05), HLA-B*57:01(0.20),HLA-B*58:01(0.05)
11	228	236	GADTGTPHW	HLA-B*58:01(0.06),HLA-B*57:01(0.21), HLA-B*53:01(0.16)	HLA-B*58:01(0.80)
12	215	223	KEWFHDIPL	HLA-B*40:01(0.07)	HLA-B*40:01(0.01),HLA-B*44:02(0.30), HLA-B*44:03(0.15)
13	49	57	TVSNMAEVR	HLA-A*68:01(0.11)	HLA-A*68:01(0.40)
14	426	434	DTAWDFGSV	HLA-A*68:02(0.05)	HLA-A*26:01(0.20),HLA-A*68:02(0.05)
15	129	137	SIQPENLEY	HLA-A*01:01(0.06),HLA-B*15:01(0.08), HLA-B*35:01(0.13),HLA-A*11:01(0.35), HLA-A*26:01(0.22)	HLA-A*01:01 (0.30),HLA-A*30:02 (0.30)
16	466	474	ILIGTLLMW	HLA-B*57:01(0.2),HLA-B*58:01(0.14), HLA-A*32:01(0.09),HLA-A*23:01(0.38), HLA-B*53:01(0.44)	HLA-B*58:01 (0.80)
17	299	307	RLKGVSYSL	HLA-A*32:01(0.02),HLA-A*02:01(0.19), HLA-B*15:01(0.24),HLA-B*08:01(0.14), HLA-A*30:01(0.14),HLA-A*02:06(0.43)	HLA-A*02:03(0.80),HLA-A*30:01 (0.80), HLA-A*32:01(0.10),HLA-B*08:01(0.40)
18	341	349	VPAQMAVDM	HLA-B*35:01(0.08),HLA-B*07:02(0.24), HLA-B*53:01(0.28)	HLA-B*07:02(0.80),HLA-B*35:01(0.40), HLA-B*53:01(0.80)
19	308	316	CTAAFTFTK	HLA-A*11:01(0.08),HLA-A*68:01(0.26), HLA-A*03:01(0.39)	HLA-A*03:01(0.30),HLA-A*11:01(0.05), HLA-A*68:01(0.05)
20	490	498	ALGGVLIFL	HLA-A*02:01(0.1),HLA-A*02:06(0.26)	HLA-A*02:01(0.80)
21	159	167	ETDENRAKV	HLA-A*68:02(0.07)	HLA-A*68:02(0.30)
22	53	61	MAEVRSYCY	HLA-A*01:01(0.09),HLA-B*35:01(0.2)	HLA-A*01:01(0.10),HLA-B*35:01 (0.80), HLA-B*53:01(0.80)
23	309	317	TAAFTFTKI	HLA-B*51:01(0.07),HLA-A*68:02(0.19)	HLA-A*68:02(0.30),HLA-B*51:01(0.80)
24	304	312	SYSLCTAAF	HLA-A*24:02(0.11),HLA-A*23:01(0.17)	HLA-A*23:01(0.15),HLA-A*24:02(0.10)
25	324	332	GTVTVEVQY	HLA-A*30:02(0.08),HLA-A*01:01(0.2), HLA-B*15:01(0.33),HLA-B*58:01(0.38), HLA-A*26:01(0.21)	HLA-A*01:01(0.80)

26	412	420	EATVRGAKR	HLA-A*68:01(0.44),HLA-A*33:01(0.18)	HLA-A*33:01(0.80)
27	249	257	HAKRQTVVV	HLA-B*08:01(0.09),HLA-B*51:01(0.31)	HLA-B*08:01 (0.80)
28	489	497	LALGGVLIF	HLA-B*53:01(0.25),HLA-B*58:01(0.24), HLA-B*35:01(0.16),HLA-B*57:01(0.44)	HLA-B*35:01(0.40),HLA-B*51:01(0.80), HLA-B*53:01 (0.30), HLA-B*57:01(0.80), HLA-B*58:01 (0.40)
29	366	374	TESTENSKM	HLA-B*40:01(0.2),HLA-B*44:02(0.25)	HLA-B*44:02(0.80)
30	205	213	TMNNKHWLV	HLA-A*02:01(0.23),HLA-B*08:01(0.28), HLA-A*02:03(0.22)	HLA-A*02:01(0.40,HLA-A*02:03 (0.30), HLA-A*02:06 (0.80)
31	455	463	SLFGGMSWF	HLA-B*15:01(0.24),HLA-A*32:01(0.22), HLA-A*26:01(0.28)	HLA-A*26:01(0.40),HLA-A*32:01 (0.20), HLA-B*15:01(0.40)
32	217	225	WFHDIPLPW	HLA-A*23:01(0.15),HLA-B*53:01(0.41), HLA-A*32:01(0.42)	HLA-A*23:01(0.40),HLA-A*24:02 (0.40)
33	394	402	KKITHHWHR	HLA-A*31:01(0.33)	HLA-A*31:01 (0.80)
34	198	206	FSDLYYLTM	HLA-A*01:01(0.19)	HLA-A*01:01(0.10)
35	25	33	LEHGGCVTV	HLA-B*40:01(0.28)	HLA-B*40:01 (0.30)
36	494	502	VLIFLSTAV	HLA-A*02:03(0.25), HLA-A*02:01(0.36)	HLA-A*02:03 (0.10),HLA-A*02:06 (0.80), HLA-A*02:01 (0.30)
37	472	480	LMWLGLNTK	HLA-A*03:01(0.4)	HLA-A*03:01(0.80)
38	306	314	SLCTAAFTF	HLA-A*32:01(0.16)	HLA-A*32:01 (0.80)
39	357	365	RLITANPVI	HLA-A*32:01(0.21)	HLA-A*32:01(0.80)
40	374	382	MMLELDPPF	HLA-B*35:01(0.42)	HLA-A*02:06 (0.80), HLA-A*32:01(.80), HLA-B*15:01(0.80),HLA-B*35:01(0.15), HLA-B*53:01(0.80)
41	221	229	IPLPWHAGA	HLA-B*51:01(0.48)	HLA-B*51:01(0.80)
42	135	143	LEYRIMLSV	HLA-B*40:01(0.49)	HLA-B*40:01 (0.30),HLA-B*44:02 (0.80), HLA-B*44:03(0.80)

3.4 Immunogenicity, toxicity, antigenicity, allergenicity and human homology of epitopes

Immunogenicity analysis of selected 42CTL epitopes reported a positive immunogenicity value for 24 epitopes (table 2). A high score of immunogenicity results in high potency for the stimulation of naive T cells. Toxicity, antigenicity, and allergenicity predictions were carried out for allthe epitopes. The BLASTP program was used to predict the homology of the epitopes with *Homo sapiens*(Altschul et al., 1997). After the assessment, the four best MHC class-I epitopes,such as DTAWDFGSV, SIQPENLEY, GTVTVEVQY, and MMLELDPPF, wereselected for vaccine componentsand shown as bold text in table 2.

Table 2: Predicted epitope immunogenicity, toxicity, antigenicity, allergenicity and human homology

Sl. No	Epitopes	Immunogenic-ity score	Toxicity	Antigenici-ty score	Antigenicity	Allergenicity	Human homology
1	KSLFGGMSW	-0.06576	Non-toxin	0.7431	Antigen	Non-allergen	Non-homologous
2	GEKKITHHW	-0.07466	Non-toxin	0.2060	Non-antigen	Allergen	Non-homologous
3	ETLHGTVTV	0.15601	Non-toxin	0.5422	Antigen	Allergen	Non-homologous
4	QEGAVHTAL	0.17664	Non-toxin	0.0330	Non-antigen	Non-allergen	Non-homologous
5	EFKDAHAKR	-0.07211	Non-toxin	0.6898	Antigen	Non-allergen	Non-homologous
6	GLDFSDLYY	-0.02674	Non-toxin	1.7878	Antigen	Non-allergen	Non-homologous
7	KLRLKGVSY	-0.23428	Non-toxin	1.8471	Antigen	Allergen	Non-homologous
8	GLFGKGSLV	-0.2521	Non-toxin	0.8202	Antigen	Allergen	Non-homologous
9	TTVSNMAEV	-0.23315	Non-toxin	0.6972	Antigen	Allergen	Non-homologous
10	MAVLGDTAW	0.11174	Non-toxin	0.8599	Antigen	Allergen	Non-homologous
11	GADTGTPHW	0.12534	Non-toxin	0.3712	Non-antigen	Non-allergen	Non-homologous
12	KEWFHDIPL	0.34792	Non-toxin	0.0523	Non-antigen	Allergen	Non-homologous
13	TVSNMAEVR	-0.08576	Non-toxin	0.7175	Antigen	Allergen	Non-homologous
14	**DTAWDFGSV**	0.29933	Non-toxin	1.9031	Antigen	Non-allergen	Non-homologous
15	**SIQPENLEY**	0.09179	Non-toxin	2.1083	Antigen	Non-allergen	Non-homologous
16	ILIGTLLMW	-0.0073	Non-toxin	0.4446	Antigen	Non-allergen	Non-homologous
17	RLKGVSYSL	-0.25121	Non-toxin	1.1682	Antigen	Allergen	Non-homologous
18	VPAQMAVDM	-0.19023	Non-toxin	0.4764	Antigen	Non-allergen	Non-homologous
19	CTAAFTFTK	0.32409	Non-toxin	0.3747	Non-antigen	Non-allergen	Non-homologous
20	ALGGVLIFL	0.25558	Non-toxin	0.2254	Non-antigen	Non-allergen	Non-homologous
21	ETDENRAKV	0.05739	Non-toxin	0.0354	Non-antigen	Allergen	Non-homologous
22	MAEVRSYCY	-0.06591	Non-toxin	1.2831	Antigen	Non-allergen	Non-homologous
23	TAAFTFTKI	0.18526	Non-toxin	0.8551	Antigen	Non-allergen	Non-homologous
24	SYSLCTAAF	-0.02494	Non-toxin	-0.1391	Non-antigen	Non-allergen	Non-homologous
25	**GTVTVEVQY**	0.15407	Non-toxin	1.4859	Antigen	Non-allergen	Non-homologous
26	EATVRGAKR	0.04346	Non-toxin	0.6828	Antigen	Allergen	Non-homologous
27	HAKRQTVVV	-0.03522	Non-toxin	0.7216	Antigen	Non-allergen	Non-homologous
28	LALGGVLIF	0.17076	Non-toxin	0.6495	Antigen	Non-allergen	Non-homologous
29	TESTENSKM	-0.18885	Non-toxin	0.8013	Antigen	Allergen	Non-homologous
30	TMNNKHWLV	-0.0077	Non-toxin	1.0538	Antigen	Allergen	Non-homologous
31	SLFGGMSWF	-0.0704	Non-toxin	0.3681	Non-antigen	Non-allergen	Non-homologous
32	WFHDIPLPW	0.13614	Non-toxin	0.7204	Antigen	Non-allergen	Non-homologous
33	KKITHHWHR	0.35005	Non-toxin	-0.3415	Non-antigen	Non-allergen	Non-homologous
34	FSDLYYLTM	0.00228	Non-toxin	0.7397	Antigen	Allergen	Non-homologous
35	LEHGGCVTV	0.08437	Non-toxin	0.5027	Antigen	Allergen	Non-homologous
36	VLIFLSTAV	0.05009	Non-toxin	0.2934	Non-antigen	Non-allergen	Non-homologous
37	LMWLGLNTK	0.10052	Non-toxin	1.8391	Antigen	Allergen	Non-homologous
38	SLCTAAFTF	0.21797	Non-toxin	0.1343	Non-antigen	Non-allergen	Non-homologous
39	RLITANPVI	0.12903	Non-toxin	-0.2339	Non-antigen	Allergen	Non-homologous
40	**MMLELDPPF**	0.09139	Non-toxin	1.0318	Antigen	Non-allergen	Non-homologous
41	IPLPWHAGA	0.28421	Non-toxin	0.8344	Antigen	Non-allergen	Non-homologous
42	LEYRIMLSV	-0.09084	Non-toxin	0.8181	Antigen	Allergen	Non-homologous

3.5. Conservancy across the sequences

The envelope protein produced 98 homologous sequences when searched against the nr database using a similarity search tool (BLASTP) (Altschul et al., 1997).Protein sequences with greater than 99 percent sequence identity and 100 percent query coverage were considered homologous in our study.The conservation value of chosen CTL epitopes among examined Envelope protein homologous sequences is shown in Table 3. In this analysis, we discovered that DTAWDFGSV, SIQPENLEY, GTVTVEVQY, and MMLELDPPF have 100% conservancy.

Table 3: The potential CTL epitopes conservancy

Protein Name	Epitope sequence	Epitope conservancy
Envelope protein	DTAWDFGSV	100%
	SIQPENLEY	100%
	GTVTVEVQY	100%
	MMLELDPPF	100%

3.6. Worldwide population coverage analysis

To analyze global population coverage, the selected MHC class-I epitopes used for vaccine production were collected, as well as their associated HLA binding alleles, as anticipated in table1. MHC class-I epitopes covered a large portion of the global population (figures1 to 4). This research reveals that the suggested vaccine could be an effective option for the majority of the world's population.

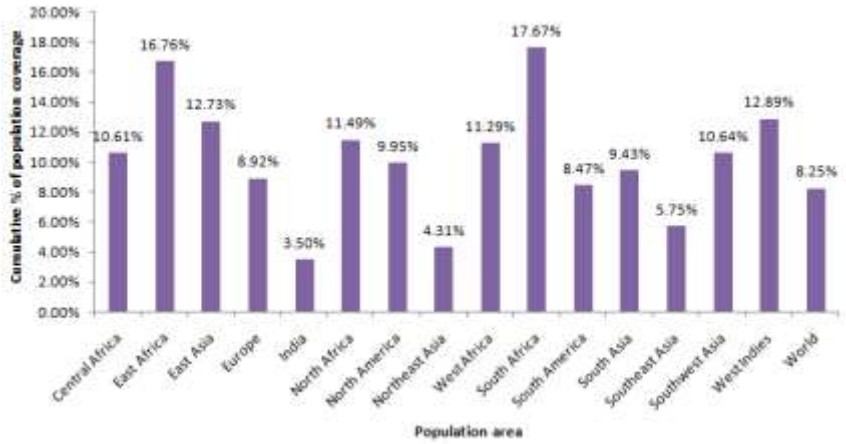

Figure 1: Population coverage analysis predicted based on epitope DTAWDFGSV along with their respective alleles

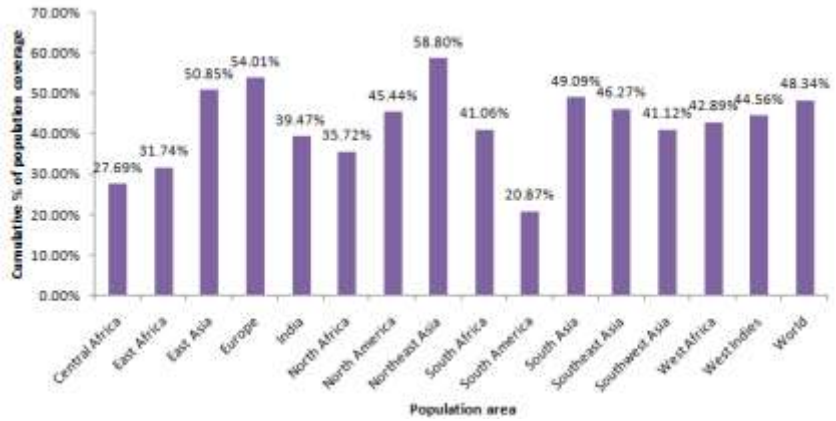

Figure 2: Population coverage analysis predicted based on epitope SIQPENLEY along with their respective alleles.

Figure 3: Population coverage analysis predicted based on epitope GTVTVEVQY along with their respective alleles.

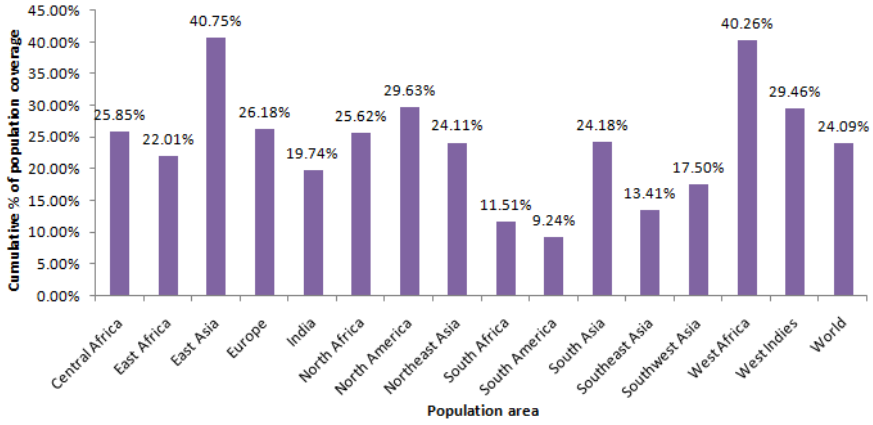

Figure 4: Population coverage analysis predicted based on epitope MMLELDPPF along with their respective alleles.

3.7 Structure of the alleles

Crystal structure of HLA-A*01:01(6MPP), and HLA-B*35:01 (PDB ID: 4LNR) were retrieved from PDB database.The HLA-A*68:02 allele sequence was obtained from the IPD-IMGT/HLA database (Robinson et al. 2015) and used as the target protein sequence for homology modeling.The modeling of the target protein structure was done in a phased manner, beginning with a template structure search, sequence alignment,and model building by the Swiss-Model server. From a substantial number of results, a suitable template structure with PDB-Id: 7RTD was picked as the basis for model building.The query coverage for the target sequence is 100%, and the sequence identity with the template sequence is 96.71%. Swiss-Model server generated a homology model of the target sequence based on the template and target alignment.

3.7.1 Tertiary structure validation

The Swiss model/Structure evaluation server was utilized to validate the refinement results with Ramachandran plot analysis. Figure 5a shows a Ramachandran plot of the predicted model with 98.53 percent of the residues in the favored region and 0.74 percent in the outlier region. The ProSA-web server computes and displays the total quality score for a specific input structure in the context of all known protein structures. The Z score of the model protein was -9.49by using the ProSA web server as shown in figure 5b, in the broad blackdot.

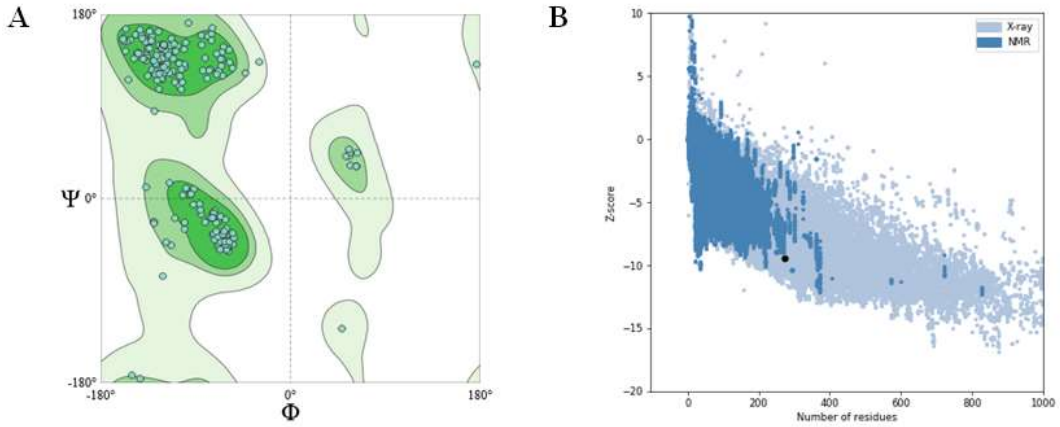

Figure 5: The Ramachandran plot and the ProSA-web server were used to validate the tertiary structure.(A) Ramachandran plot shows that the amount of amino acid residues in favorable region is 98.53%.(B) The ProSA-web result provides a Z score of -9.49.

3.8. Docking analysis

The HPEPDOCK server generated 10 conformations and the one with the lowest binding energy score was picked. The lower the binding energy, the greater will be the binding interaction between the HLA allele and the epitope. The docking scores of peptide-allele complexes are shown in Table 4.

Table4: Docking results of epitopes with HLA alleles

Epitope sequence	**Allele**	**Docking score**
DTAWDFGSV	HLA-A*68:02	-242.951
SIQPENLEY	HLA-A*01:01	-210.425
GTVTVEVQY	HLA-A*01:01	-216.146
MMLELDPPF	HLA-B*35:01	-212.653

Interactions between the epitopesand their respective allele were shown figure 6a, b, c& d.

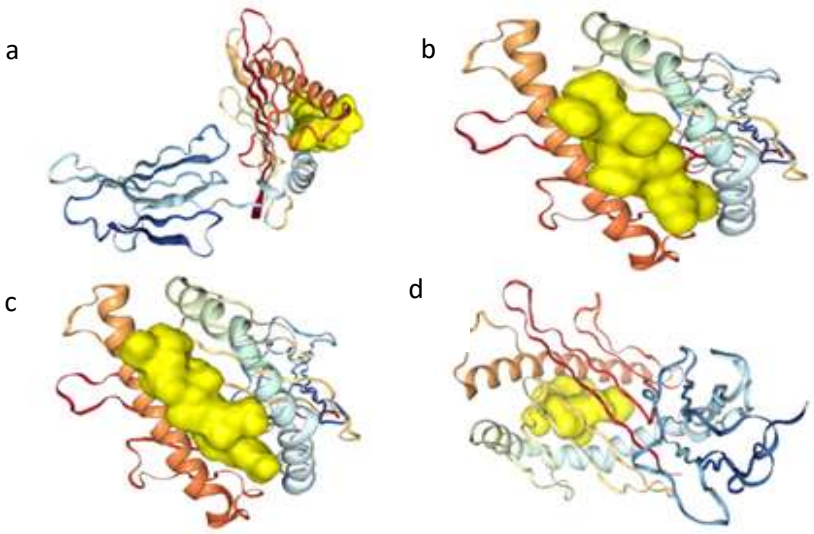

Figure 6: The interactions between epitopes and alleles are depicted in this diagram. Here, (a)is the interaction between epitope DTAWDFGSV and HLA class I alleleHLA-A*68:02, (b) is the interaction between SIQPENLEY and HLA-A*01:01 allele, (c)the interaction between GTVTVEVQY and HLA-A*01:01 allele, and (d) is the interaction between MMLELDPPF and HLA-B*35:01allele.

3.9 Molecular Dynamics simulation

To confirm the proper engagement of the peptide sequence and with the HLA alleles, MD simulation was conducted. Subsequently, the final trajectory was used to analyze some essential parameters; the first one was the RMSD. This parameter measures the stability between the allele and the epitope where mild fluctuations point to a stable interaction. Figure 7 shows stable RMSD of all the complexes throughout the 20000ps, which is considerable stability of the complexes.

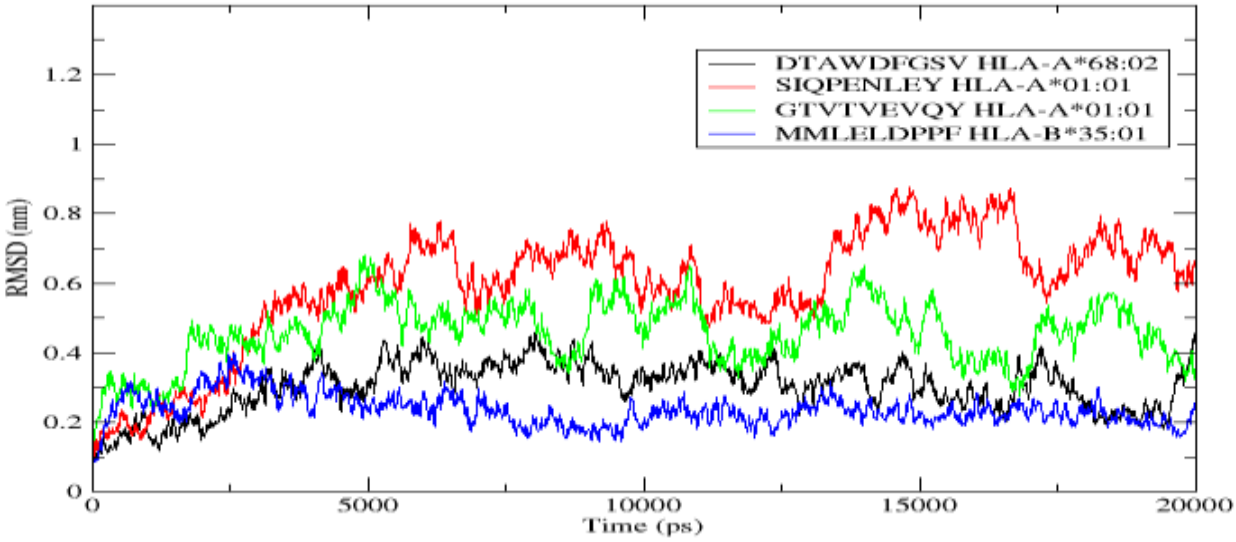

Figure 7: Root Mean Square Deviation (RMSD) plots for all four docked complexesversus the MD simulation time.

The second parameter, RMSF, represents amino acids side chains fluctuations. RMSF-Root Mean Square Fluctuation plot of docked complexes remained fairly flat reflecting the flexibility of the side chain of docked protein complexes (Figure 8).

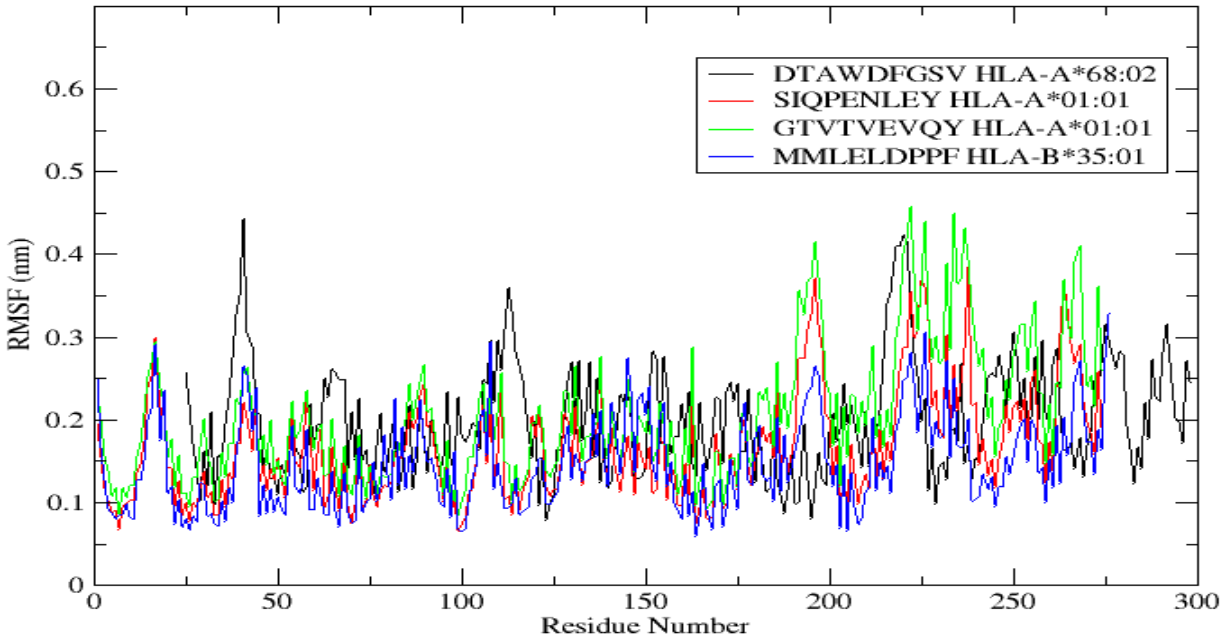

Figure 8: RMSF-Root Mean Square Fluctuation plots for all four docked complexes.

The radius of gyration can explain the compactness of the protein structure and stability of the protein. The Rg plot showed less deviation throughout the simulation process and the relatively flat curve suggests the compactness and stability of the complexes (Figure 9).

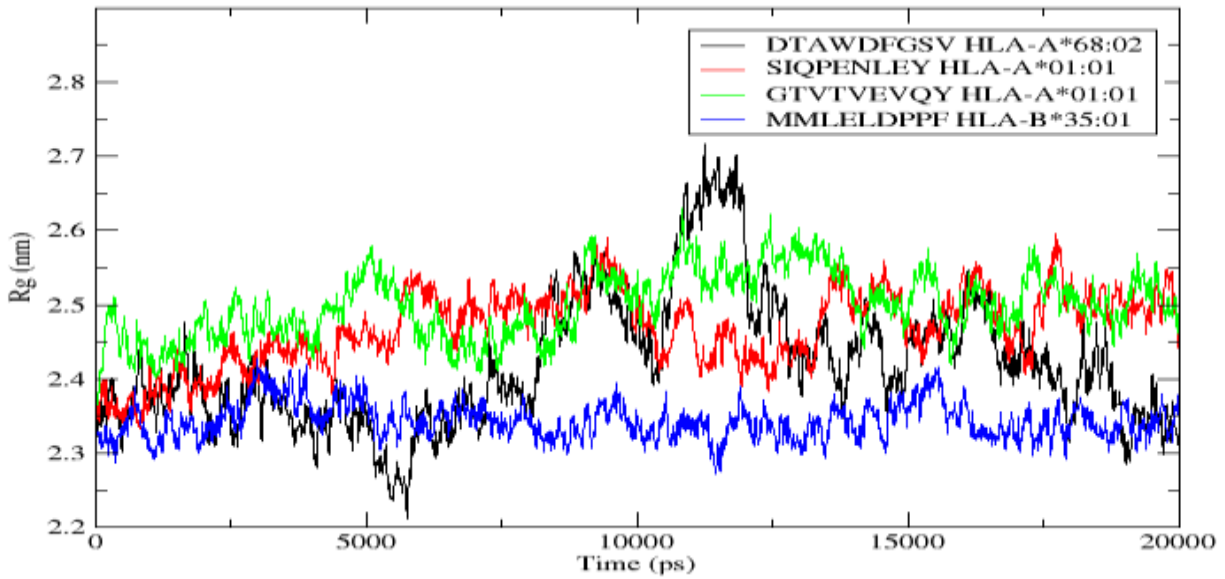

Figure 9: Radius of gyration (Rg) plots of the four docked complexes generated over 20000 ps using GROMACS.

4. Discussion

Immunoinformatics techniques are progressively becoming regarded as the first line of vaccine development in the development of effective vaccines against viruses. Immunoinformatics methods have recently been applied in the development of an epitope-based vaccine for *C.auris* (Akhtar et al, 2021), Dengue virus (Krishnan G et al., 2020),*Orthohantavirus* (Joshi et al, 2022) andSARS-CoV-2 (Sarkar et al, 2020; Rahman et al, 2020). Zika virus infection is a leading cause of morbidity and mortality worldwide.Unfortunately, the lack of effective ZIKV vaccines has resulted in the deaths of many innocent people around the world.As a result, preventive steps to address the global threat of a ZIKV outbreak are urgently needed.Many countries' researchers have quickly employed immunoinformaticsapproaches to predict potential antigenic epitopes on the ZIKV envelope protein for peptide vaccine production (Srivastava and Srivastava, 2023;Ashfaq and Ahmed,2016; Dikhit et al.,2016; Dar et al., 2016; Weltman,2016; Gupta,2014; Badawi et al.,2016). The goal of this work was to develop epitope-based peptide vaccine componentsagainst the *Zika*virus using immunoinformatics techniques.Epitope-based vaccines have a positive effect when compared to traditional vaccine design (Reginald et al., 2018).Several immunoinformaticstools were utilized to screen CTL epitopes in the envelope protein sequence.Regions of protein antigens that bind to immunologic receptors are known as immunodominant epitopes (Ayub et al., 2016). As a result, they were frequently utilized as therapeutic compounds and vaccines. (Ninomiya et al., 2002; Adame-Gallegos et al., 2012; McComb et al., 2015).42CTL epitopes predicted by two servers were chosen and additionally picked out on the grounds of their immunogenicity, toxicity, antigenicity, allergenicity, and human homologyassessment.Using a 0.4 threshold, the VaxiJen tool was used to classify viral components as antigens or non-antigens.According to the immunogenicity analysis results, 24 epitopes exhibited a positive immunogenicity value.A high immunogenicity score suggests a strong ability to stimulate naive T cells. Four potent CTL epitopes such as DTAWDFGSV, SIQPENLEY, GTVTVEVQY,and MMLELDPPFhad antigenicity scores greater than one, and non-toxic and non-allergenic were screened.Peptide–protein docking is widely used to test the stability of vaccines by examining their binding energies.The best predicted epitopes against MHC class I alleles (HLA-A*68:02, HLA-A*01:01, HLA-B*35:01) were examined using this method.The docked complex was also examined for molecular dynamics simulation research, and it was discovered that, according to the RMSD and RMSF plots, they exhibit a stable interaction

pattern.Immunoinformatics approaches are very useful for in-silico studies and can direct laboratory experiments, saving time and money. Nonetheless, the next step is to conduct in vitro immunological tests to validate the predicted epitopes.

Conclusion

Four potent CTL epitopes DTAWDFGSV, SIQPENLEY, GTVTVEVQY and MMLELDPPF in the envelope protein of *Zika* virus were identified as having highimmunogenicity and antigenicity scores as well as non-toxic, non-allergen andnon-homologous with the human protein. These epitopes had the lowest docking score with their associated MHC class I alleles. The stable nature of the docked complexes was confirmed by Molecular Dynamics simulation studies. The projected epitopes cover the maximum number of national and international populations. These predicted epitopes may have strong ability to stimulate naive T cells.

Acknowledgement

The authors would like to express their gratitude to the Department of Biotechnology, Faculty of Engineering and Technology, Rama University, Kanpur, Uttar Pradesh (India).

Ethical approval

I confirm that authors did not perform any experiments on humans or animals.

Declaration of competing interest

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

References

1. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl, E (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, 1:19-25.
2. 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 yoeliimerozoite surface protein 1-19 (MSP119). ExpParasitol., 130:384–393.

3. Akhtar N, Joshi A, Kaushik V, Kumar M, Amin-ulMannan M (2021) In-silico design of a multivalent epitope-based vaccine against Candida auris. Microbial Pathogenesis, 155:104879.
4. 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.
5. Ashfaq UA, Ahmed B (2016) De novo Structural Modeling and Conserved Epitopes Prediction of Zika Virus Envelope Protein for Vaccine Development. Viral Immunol., 29: 436–443.
6. 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.
7. Badawi MM et al. (2016) Highly Conserved Epitopes of ZIKA Envelopee Glycoprotein May Act as a Novel Peptide Vaccine with High Coverage: Immunoinformatics Approach. American Journal of Biomedical Research, 4(3):46–60.
8. 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 Bioinformatics 8(1):361.
9. 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 computational biology. 9, e1003266.
10. Calvet GA, Santos FB, Sequeira PC (2016) Zika virus infection: epidemiology, clinical manifestations and diagnosis, Current Opinion in Infectious Diseases, 29:459–466.
11. Cao-Lormeau VM, Blake A, Mons S, Lastère S, Roche C, Vanhomwegen J (2016) Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. Lancet, 387:1531–1539.
12. Dar Het al. (2016) Prediction of promiscuous T-cell epitopes in the Zika virus polyprotein: An in silico approach. Asian Pac. J. Trop. Med.,9(9): 844–850.
13. Dick GWA (1952) Zika virus (II). Pathogenicity and physical properties. Transactions of the Royal Society of Tropical Medicine and Hygiene, 5:521–534.
14. Dikhit MR et al. (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.
15. 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):2278.
16. Doytchinova IA, Flower DR (2007)VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines.BMC Bioinformatics, 8:4.
17. Duffy MR, Chen TH, Hancock WT, Powers AM, Kool JL, Lanciotti RS et al. (2009) Zika virus outbreak on Yap Island, federated states of Micronesia. New England Journal of Medicine, 360:2536–2543.
18. Flower DR (2007) Immunoinformatics. Predicting immunogenicity in silico. Preface. Methods Mol Biol., 409: v-vi.
19. Gupta P (2014) Computer Aided Drug Design and Discovery–An Economical Approach to Drug Discovery Industry. Austin J BiotechnolBioeng., 1(4): 2.

20. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R; Open Source Drug Discovery Consortium, Raghava GP (2013). In silico approach for predicting toxicity of peptides and proteins. PLoS One., 8(9):e73957.
21. Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatabadi SS, Ghasemi Y (2017). Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against Staphylococcus aureus. Infection, Genetics and Evolution, 48:83–94.
22. Hamel R, Dejarnac O, Wichit S et al. (2015) Biology of Zika virus infection in human skin cells. *J Virol.*, 89(17):8880-96.
23. Hospital A, Goñi JR, Orozco M, Gelpí JL (2015) Molecular dynamics simulations: advances and applications. Adv. Appl. Bioinform Chem., 8:37–47.
24. 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 Bioinforma11,Article number: 2.
25. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, et al. (2012) Immune epitope database analysis resource. Nucleic acids research, 40: W525–530.
26. Krishnan GS, Joshi A, Kaushik V(2020) T cell epitope designing for dengue peptide vaccine using docking and molecular simulation studies. Molecular Simulation,46(10):787-795.
27. Lazim R, Suh D, Choi S (2020) Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. Int J Mol Sci., 21(17):6339.
28. 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.
29. Narula A, Pandey RK, Khatoon N, Mishra A, Prajapati VK (2018) Excavating chikungunya genome to design B and T cell multi-epitope subunit vaccine using comprehensive immunoinformatics approach to control chikungunya infection. Infection, Genetics and Evolution, 61:4–15.
30. Nezafat N, Karimi Z, Eslami M, Mohkam M, Zandian S, Ghasemi Y (2016) Designing an efficient multi-epitope peptide vaccine against Vibrio cholerae via combined immunoinformatics and protein interaction based approaches. Computational biology and chemistry, 62:82–95.
31. 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.
32. Pielnaa P, Al-Saadawe M, Saro A, Dama MF, Zhou M, Huang Y et al. (2020) Zika virus-spread, epidemiology, genome, transmission cycle, clinical manifestation, associated challenges, vaccine and antiviral drug development. Virology, 543:34–42.
33. 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.
34. Reginald K, Chan Y, Plebanski M, Poh CL (2018) Development of peptide vaccines in dengue. Curr. Pharm. Des., 24:1157-1173.
35. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG (2015) The IPD and IMGT/HLA database: allele variant databases. Nucleic acids research, 43(Database issue), D423–D431.
36. Rothan HA, Bidokhti MRM, Byrareddy SN (2018) Current concerns and perspectives on Zika virus co-infection with arboviruses and HIV. J Autoimmun, 89:11–20.

37. Saiz JC, Martín-Acebes, MA (2017) The race to find antivirals for Zika virus. Antimicrob Agents Chemother, 61(6):e00411-17.
38. Sarkar B, Ullah MA, Johora FT, Taniya MA, Araf Y(2020) Immunoinformatics-guided designing of epitope-based subunit vaccines against the SARS Coronavirus-2 (SARS-CoV-2). Immunobiology, 225:151955.
39. Saxena SK, Kumar S, Sharma R, Maurya VK, Dandu HR, Bhatt MLB (2019)Zika virus disease in India—updates October 2018. Travel Med Infect Dis., 27: 121-122.
40. Singh S, Singh H, Tuknait A, Chaudhary K, Singh B, Kumaran S. Raghava GPS (2015) PEPstrMOD: structure prediction of peptides containing natural, non-natural and modified residues. Biology Direct, 10:73.
41. Sinha S, Tam B, Wang SM (2022) Applications of Molecular Dynamics Simulation in Protein Study. Membranes (Basel), 12(9):844.
42. 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. 2021; 27(1):779-799.
43. Srivastava K, Srivastava V (2023) Prediction of Conformational and Linear B-Cell Epitopes on Envelop Protein of Zika Virus Using Immunoinformatics Approach. Int J Pept Res Ther.,29(1):17.
44. Stranzl T, Larsen MV, Lundegaard C, Nielsen M (2010) NetCTLpan: pan-specific MHC class I pathway epitope predictions. Immunogenetics 62:357–368.
45. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MackerellJr AD (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of computational chemistry, 31(4):671- 690.
46. Walker, JM (2005). The Proteomics Protocols Handbook. https://doi.org/10.1385/1592598900.
47. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gummienny 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.
48. Weltman J (2016) Computer-Assisted Vaccine Design by Analysis of Zika Virus E Proteins Obtained either from Humans or from Aedes Mosquitos. J Med MicrobDiagn 5(3):1000235.
49. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research, 35(Web Server issue):W407–W410.
50. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF (1999) Protein identification and analysis tools in the ExPASy server. Methods in molecular biology (Clifton, N.J.), 112:531–552.
51. World Health Organization (2021) Zika Virus Disease – India. Available online: https://www.who.int/emergencies/disease-outbreak-news/item/zika-virus-disease-india (accessed on 14 October 2021).
52. Xiang Z, He Y (2009) Vaxign: a web-based vaccine target design program for reverse vaccinology. ProcediaVaccinol., 1(1):23–29.
53. Yadav PD, Malhotra B, Sapkal G et al. (2019) Zika virus outbreak in Rajasthan, India in 2018 was caused by a virus endemic to Asia. Infect Genet Evol., 69:199-202.
54. 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 Research, 46(W1):W443–W450,