Current Pharmaceutical Biotechnology, 2020, *21*, 325-340

RESEARCH ARTICLE

Prediction of Epitope-based Peptide Vaccine Against the Chikungunya Virus by Immuno-informatics Approach

Saeed Anwar[1,2,†], Jarin T. Mourosi[1,3,†], Md. Fahim Khan[1,†] and Mohammad J. Hosen[1,*]

[1]*Department of Genetic Engineering and Biotechnology, School of Life Sciences, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh;* [2]*Maternal and Child Health Program, Department of Medical Genetics, Faculty of Medicine and Dentistry, University of Alberta, 8440 112 St. NW, Edmonton, AB T6G 2R7, Canada;* [3]*Microbial and Cellular Biology Program, Department of Biology, The Catholic University of America, 620 Michigan Ave. NE, Washington, DC, 20064, USA*

Abstract: ***Background*:** Chikungunya is an arthropod-borne viral disease characterized by abrupt onset of fever frequently accompanied by joint pain, which has been identified in over 60 countries in Africa, the Americas, Asia, and Europe.

***Methods*:** Regardless of the availability of molecular knowledge of this virus, no definite vaccine or other remedial agents have been developed yet. In the present study, a combination of B-cell and T-cell epitope predictions, followed by molecular docking simulation approach has been carried out to design a potential epitope-based peptide vaccine, which can trigger a critical immune response against the viral infections.

***Results*:** A total of 52 sequences of E1 glycoprotein from the previously reported isolates of Chikungunya outbreaks were retrieved and examined through in silico methods to identify a potential B-cell and T-cell epitope. From the two separate epitope prediction servers, five potential B-cell epitopes were selected, among them "NTQLSEAHVEKS" was found highly conserved across strains and manifests high antigenicity with surface accessibility, flexibility, and hydrophilicity. Similarly, two highly conserved, non-allergenic, non-cytotoxic putative T-cell epitopes having maximum population coverage were screened to bind with the HLA-C 12*03 molecule. Molecular docking simulation revealed potential T-cell based epitope "KTEFASAYR" as a vaccine candidate for this virus.

***Conclusion*:** A combination of these B-cell and T-cell epitope-based vaccine can open up a new skyline with broader therapeutic application against Chikungunya virus with further experimental and clinical investigation.

ARTICLE HISTORY

Received: May 05, 2019
Revised: July 16, 2019
Accepted: November 04, 2019

DOI:
10.2174/1389201020666191112161743

Keywords: Chikungunya virus, immunoinformatics, T cell epitopes, B-cell epitopes glycoprotein, vaccine, Zika viruses.

1. INTRODUCTION

Chikungunya is a neglected tropical disease, dispersed by mosquitoes and caused by the Chikungunya Virus (CHIKV), which can be characterized by frequent fever accompanied by arthralgia [1-3]. Other symptoms of this debilitating disease include myalgia, nausea, skin rashes, photophobia, headaches and some neurological and cardiovascular complications [3-6].

CHIKV is an enveloped, positive-sense, and single-stranded RNA virus, which belongs to the alphavirus genus of the Togaviridae family [7]. This virus is generally considered as a non-fatal virus, though, there are reports that it can cause ~10% mortality and ~35% severe morbidity [1, 8]. Furthermore, CHIKV co-infection and overlapping features have recently been reported with Dengue and Zika viruses, which make the diagnosis a challenge. No wonder, the morbidity and mortality associated with CHKV infection has risen with an increase in the number of co-infection cases [9].

Since the first evidence of CHIKV outbreak in 1952 in Tanzania, several major and small outbreaks have been reported in Africa and Asia [10, 11]. A majority of these outbreaks occurred in the 1960s-1980s and then decreased in

*Address correspondence to this author at the Department of Genetic Engineering and Biotechnology, School of Life Sciences, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh; Tel: +8801777195857; E-mail: jakir-gen@sust.edu

[†]*These authors contributed equally.*

1873-4316/20 $65.00+.00

activity until 2004 [10]. However, with a number of major outbreaks in Africa, Asia, and Pacific Oceans in the last era, CHIKV has become a global health threat [10]. As of May 2018, more than one million Chikungunya cases have been reported in more than a hundred countries and territories across the world [12]. The global expansion of this disease is solely due to an increase in international travel, with disease transportation from infected travelers [13, 14]. The mosquito species, *Aedes aegypti* was a principal vector in many outbreaks, while *Aedes albopictus*, the Asian tiger mosquito, has been considered a primary re-emergence factor since 2005 [15].

With the emergence and re-emergence of CHIKV in over 55 countries in the world, the demands for an efficient and cost-effective antiviral strategy to reduce morbidity and mortality are increasing significantly. An ideal vaccine which would initiate both humoral and cell-mediated immune response could help eradication of the virus. Although several endeavors have revealed genetic insights into the genome of the CHIKV and the pathomechanism underlying the infection is thoroughly studied, no suitable antiviral strategies against this aggressive viral infection have been approved to date [16-18]. Thus, failure in developing effective vaccines or specific antivirals has essence in further research [16]. However, unlike most other neglected tropical diseases, pharmaceutical industries remain indifferent about CHIKV vaccine development as it is a costly and time-consuming process [16, 19]. Few researches on vaccines against CHIKV started during the last decade, and several inactivated, attenuated, and recombinant vaccine candidates are currently in the early stages of research and clinical trials [20, 21]. With this end of view, computational algorithms can assist the industries and scientists to cope with the factors associated with time and cost. Notably, in the past two decades, computational algorithms and bioinformatic approaches have shown their success and power in assisting therapeutic discovery and development. Moreover, the recent advancements in genomic and proteomic technologies have opened up a new horizon in the computational immunology.

Conventional vaccination measures usually utilized live attenuated pathogens to boost immunity against particular diseases. However, those immunization approaches were prone to potential threats of re-activation of the vaccine molecule, which could induce infection. Besides, antigenic drifts and shifts are causing the vaccines to lose their efficacy in different regions. But, modern vaccination approaches, *e.g.*, DNA vaccines and epitope-based peptide vaccines, have the promise to defeat these safety concerns. Moreover, these vaccines are believed to produce a more robust, effective, and prolonged immune response with minimal side-effects [22]. Epitope based vaccine design using immunoinformatic approach is a state-of-art method, because of its high level of specificity, capacity to create long-lasting immunity with evading undesirable side-effects, and above all, its ability to reduce time and cost [23]. This method has been used to facilitate several diseases, including coronavirus infection, arboviral infections like Zika and dengue, tuberculosis, Nipah virus infection, rhinovirus infection, and St. Louis Encephalitis virus [24-30].

One of the experimental strategies undertook for the development of an efficient strategy against CHIKV is to inhibit the viral replication within the host cell by interfering with the non-structural proteins. Unfortunately, CHIKV does not offer any facile scope to target its non-structural proteins that make up active enzymatic complexes of its replication machinery [31, 32]. With this end of view, targeting the CHIKV structural proteins is a tactful approach to develop efficient anti-viral strategies against the virus. The E1 glycoprotein, being a subset of CHIKV structural polyprotein (C-E3-E2-6K-E1), is a class II fusion protein, which remains heterodimerically embedded in the CHIKV membrane and is responsible for membrane fusion [32]. However, most neutralizing antibodies target the E2 glycoprotein, a tri-domain cell attachment protein, in response to the infection, and hence, most vaccine efforts target E2 to combat alphavirus infections.

Nevertheless, glycoprotein E1 alone has been successfully used as a vaccine antigen capable of protecting against lethal challenge though neutralizing antibodies against E1 are rare [33, 34]. Moreover, E1 fusion protein remains highly conserved across strains and even the alphavirus family, and conformationally less stable than E2 attachment protein, which makes it a better choice over E2 and essences greater applicability of an E1 glycoprotein derived epitope-based vaccine candidate [35-37]. Besides, most reports to date, including a few computational studies, targeting CHIKV glycoproteins, are either against the E2 or E1-E2 complex [38, 39]. However, specific target selection is of critical importance for efficient epitope mapping [40, 41]. Several studies have inferred that interfering with viral fusion is a novel and attractive immuno-therapeutic strategy to control virus infection, and both T-cell and B-cell epitope-based vaccines eliciting immunogenicity against viral fusion proteins presents significant clinical benefits [39, 42, 43]. We hypothesized immunity against E1 glycoprotein of CHIKV would harness therapeutic strategy development for Chikungunya infection.

In the present study, we adopted an immunoinformatic approach to recognize B-cell and T-cell based epitopes on CHIKV E1 glycoprotein. Our presented vaccine strategy could serve as a potential candidate with the extensive therapeutic application and further future laboratory-based attempts developing efficient management of CHIKV infection.

2. METHODS AND METHODOLOGY

2.1. Retrieval of Protein Sequences

A total of 52 sequences of the E1 glycoprotein of Chikungunya virus were retrieved from the GenBank database of the National Center for Biotechnology Information (NCBI) [44; URL: https://www.ncbi.nlm.nih.gov/genbank]. The dates of when they were isolated were also taken into account to cover the maximum outbreaks from the past. E1 glycoprotein sequences from the most recently sequenced isolates, *e.g.*, from the outbreak in Bangladesh in 2017, were also included (Supplementary dataset **01**) [45].

2.2. Conservancy Analysis

To analyze the level of conservation, we aligned all the retrieved sequences by using the EBI-Clustal Omega program [46] and obtained a multiple sequence alignment (MSA). The MSA was visualized using MView [47] and JalView Version 2 [48] (Supplementary dataset **02**). The absolute site variability in the MSA was calculated using Protein Variability Server (PVS) [49].

2.3. Unraveling the Antigenicity

VaxiJen v2.0 server [50] and Kolaskar & Tongaonkar method [51] were used to predict the antigenicity of the retrieved sequences. The threshold of an overall prediction score for protective viral antigen was set at 0.40 in Vaxijen v2.0, while a window size of 7 amino acids was set to determine the antigenicity of the central amino acid for each residue of E1 protein in the Kolaskar & Tongaonkar antigenicity prediction tool.

2.4 T-cell Epitope Prediction and Conservancy Analysis

To identify potential T cell epitopes, we utilized a non-linear artificial neural networks-based server, namely, NetCTL server [52] at the web-server of IEDB (URL: http://www.iedb.org). The method-specific threshold value for NetCTL was set to 0.5, and the sensitivity and specificity were set to 0.89 and 0.94, respectively. Based on these scores, we selected the top 10 epitopes for further analysis.

MHC-I alleles interacting with each of the ten selected epitopes were determined by MHC-I prediction server [53] at the IEDB interface. Afterward, the stabilized matrix method (SMM) [54] was utilized to predict the half-maximal inhibitory concentration (IC50) of peptide-binding to MHC-I alleles. 200 nM was set at the cut-off value of IC50, and for the analysis of epitope-binding to the allele, all the available MHC-I alleles were selected, and the length of peptides was set to 9 amino acids.

On the other hand, the whole sequences were run in the IEDB MHC-II binding prediction interface [55]. IC50 values of the epitopes binding to MHC-II molecules were computed using the Stabilized Matrix Based Method, and the epitopes containing 15-amino acid residues that interacted with the highest number of alleles were selected. The SMM-align method was applied to choose the binders that exhibit at least an IC50 of 100 nm. Among the epitopes which potentially would interact with the highest number of MHC I and MHC II alleles, we selected three overlapping epitopes for epitope conservancy analysis as well as population coverage evaluation [56]. Later on, the predicted T-cell epitopes were put forward to the IEDB conservancy analysis interface with sequence identity threshold set at 80% [56].

2.4.1. Allergenicity and Toxicity Analysis of the T-cell Epitopes

The AllerTOP v. 2.0 [57; URL: http://www.pharmfac.net/allertop] were employed to determine the allergenicity of the selected T cell epitopes. Only the epitopes not predicted as allergens were furthered to toxicity analysis using ToxinPred web-server [58; URL: https://webs.iiitd.edu.in/raghava/toxinpred]. Any peptide predicted as toxic by the ToxinPred web-server was discarded from the next analysis steps.

2.4.2. Population Coverage by the T-cell Epitope-based Peptides

The non-allergenic, as well as non-toxic T-cell epitope-based peptide sequences with their corresponding class of HLA alleles, were run on the population coverage analysis tool at the interface of IEDB with all parameters set as default and (1) projected population coverage, (2) average number of epitope hits or HLA combinations recognized by the populations and (3) minimum number of epitope hits or HLA combinations recognized by 90% of the population (PC_{90}) were calculated. These calculations were accomplished based on the frequency of HLA genotypes considering non-linkage disequilibrium among HLA loci [57]. Only the epitopes having ≥40% population coverage were selected for further analysis.

2.4.3. Prediction of the 3D Structures of Conserved T-cell Epitopes and Selected HLA Molecule

Three dimensional (3D) structures of the predicted and/or selected peptides were constructed using the PEP-FOLD Peptide Structure Prediction server [59, 60].

The MHC-I allele, HLA-C 12*03, was found to interact with the majority of the predicted T cell epitopes. Unfortunately, a lack of a high-resolution X-ray crystallographic structure (http://www.rcsb.org/pdb/home/home.do) of HLA-C 12*03 creates difficulties to screen binding affinity of the epitopes with the MHC class I molecule, however, we utilized a comprehensive *in silico* homology modeling and docking techniques for HLA-C 12*03. At first, the primary sequence of human HLA C-12*03 was retrieved from the NCBI database (Accession: CDK41181.1). To find a suitable template for the HLA C-12*03, we performed a PSI-BLAST against all existing molecules in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) [61]. This blast revealed the best homology of HLA C-12*03 with PDB: *5vgd.1.A* [62]. Consequently, we selected this protein as the template for 3D model generation using the MODWEB server and MODELLER v9.21 [63]. A total of 04 models were generated for further analysis.

Several procedures were used to validate the 3D models [64, 65]. The stereochemical characteristics and energetic parameters were evaluated to determine whether the bond lengths and angles were within normal ranges and whether there were too many bumps in the models (corresponding to high van der Waals forces). The stereochemical characteristics of the models were evaluated using different servers, including ERRAT [66], PROCHECK [67], and Verify3D [68], which were used to determine the statistical significance of a protein 3D model taking into consideration the spatial position of amino acids and overall stability of the structure. Usually, more than 90% accuracy is needed for the validity of a model [69].

*2.4.4. Blind Docking Assay of T-cell Epitopes with the HLA-C 12*03 Allele*

To evaluate the interaction of the T cell epitopes with the corresponding MHC-I molecules, we accomplished a blind

structure-based docking analysis on the PyRx interface [70], which is an integrated platform combining AutoDOCK Vina, AutoDOCK 4.2, Mayavi, Open Babel, etc. Initially, the epitope was separated from the allele and then docked with HLA-A*32:15 (Accession ID: AM422702). The binding energy for control epitope was calculated and compared with the energy found for the predicted T-cell epitopes of note, the 3D structure of HLA-A*32:15 molecule docked with the octapeptide epitope, HKEGAFFLY, was utilized as a positive control for the docking assay [71]. The binding of the epitope to the allele was then visualized using PyMol [72].

2.5. Potential Linear B-cell Epitope Prediction

To predict B cell epitopes without being prone to any false positiveness, we utilized two popular and frequently used tools. The retrieved sequences were run in BepiPred 2.0 [73; URL: http://www.cbs.dtu.dk/services/BepiPred] and ABCpred [74; URL: http://crdd.osdd.net/raghava/abcpred] web servers. Seventy-five percent threshold and a window length of 12 amino acids were set as parameters for the prediction of epitopes in the ABCpred server. All the predicted epitopes were scrutinized using a simple python script, and only the epitopes recognized commonly by both of the servers were selected for further analysis.

2.5.1. Prediction of Surface Accessible Regions

Emini surface accessibility scale of IEDB was utilized to evaluate the surface accessibility of the selected epitopes [75]. Only the epitopes having surface accessibility score of over 1.0 (threshold value) were selected and analyzed for conservancy across multiple strains.

2.5.2. Conservancy Evaluation of B Cell Epitopes

The epitope conservancy tool [76] at the IEDB interface was utilized to assess the conservancy of the selected epitopes based on their surface accessibility. In order to predict the conservancy of the epitopes with strong confidence, the sequence identity threshold was set as 80%, and only the epitopes showing conservancy across multiple strains were further analyzed for their antigenicity, flexibility, and hydrophobicity.

2.5.3. Prediction of Antigenicity, Flexibility, and Hydrophilicity of the Selected B-cell Epitopes

The selected B-cell epitopes were tested for their antigenicity in the VaxiJen v2.0 server [50]. At the same time, these epitopes were also run on the Karplus and Schulz flexibility tool [77] and the Parker hydrophilicity prediction tools [78] at the interface of IEDB. Epitopes, other than those that overrode the threshold values for antigenicity, flexibility, and hydrophilicity, were discarded.

2.5.4. Prediction of Carcinogenicity of the Selected B-cell Epitope

The cancerogenic potential of the selected B-cell based epitope(s) was analyzed using a two-dimensional term frequency (2DTF) formula adopted from frequency-inverse document frequency method at the Tumor Associated Gene (TAG) interface available at the webserver of Bioinformatics Center, National Cheng Kung University [79; URL: http://www.binfo.ncku.edu.tw/TAG/GeneDoc.php]. Only the non-carcinogenic epitopes were chosen as potential B-cell based vaccine candidate designing.

3. RESULTS

3.1. Conservancy and Antigenicity of E1 Glycoprotein

Multiple sequence alignment of the 52 E1 glycoproteins revealed that the protein remained mostly conserved (Fig. **1a**). The absolute variability computed by Protein Variability Server (PVS) [49] revealed 90.6% conserved nucleotides across the strains (Fig. **1b**). Particularly, two regions with amino acid positions 1-17 and 20-76 were conserved in all the strains. Antigenicity assessment using VaxiJen server identified all the proteins as probable antigens (Supplementary dataset **03**). Moreover, the Kolaskar & Tongaonkar antigenicity prediction tool [51] evaluated the proteins for B-cell epitopes investigating the physicochemical properties of amino acids and their affluence in known B cell epitopes which revealed that the average antigenicity propensity value was 1.033, while the minimum and maximum values were 0.952 and 1.153 respectively.

3.2. T-cell Epitope Identification and Conservancy Analysis

At 75% threshold value for epitope identification, NetCTL 1.2 server (available at: http://www.cbs.dtu.dk/services/NetCTL/) [52] recognized 19 potential T cell epitopes using 12 MHC supertypes (Supplementary Dataset 04). However, 10 out of the 19 epitopes, which had above 100% conservancy across all the protein sequences, were selected (Table **1**).

Later, based on IC50 cut off values (200 nM), seven overlapping T-cell based epitopes between MHC I and MHC II were predicted (Table **2**; Supplementary Dataset **05**).

3.2.1. Allergenicity and Toxicity of T-cell Epitopes

Four out of the seven selected T-cell based epitopes were predicted to be non-allergens (NA) (Table **3**). The rest three epitopes, which were predicted as allergens, were discarded from further analysis. Toxicity analysis of the seven epitopes using the ToxinPred server revealed that all of these epitopes were non-toxic (NT) to cell, proving their potential as candidate vaccines (Table **3**).

3.2.2. Population Coverage of T-cell Epitopes

Considering MHC class I and II alleles, population coverage by the four non-allergic and non-toxic epitopes varied between 26-65% (Table **4**). Only two epitopes, SASAKLRVL and KTEFASAYR, exceeded 40% population coverage and hence were furthered for additional analysis.

*3.2.3. Homology Modeling of T-cell Epitopes and HLA-C 12*03 and Validation*

PEPFOLD Peptide Structure Prediction server was utilized for the generation of 3-D structure of the two selected T cell epitopes SASAKLRVL and KTEFASAYR (Fig. **2**). As both epitopes were predicted to interact with HLA-C 12*03

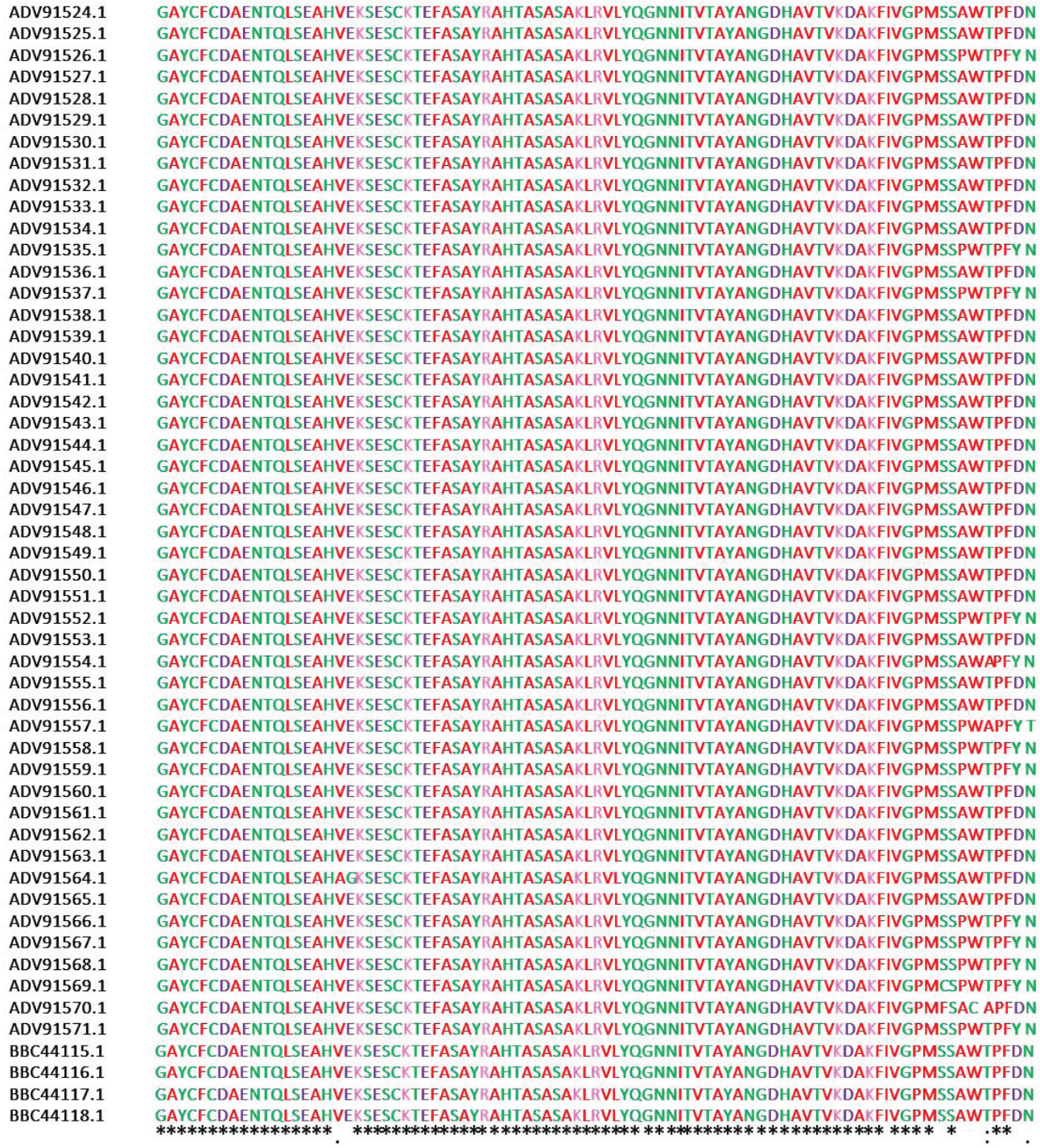


Fig. (1a). Multiple sequence alignment of the 52 E1 glycoproteins.

allele, the 3-D structure of the allele was generated in the MODELLER 9.21 server by homology modeling [63].

PSI-BLAST result showed that HLA-C 12*03 shares a 96.03% sequence identity and 100% query coverage with *5vgd.1.A*. Usually, the identity of >30% between the target and template sequence is widely accepted for comparative modeling [80]. However, for membrane transporter proteins, an identity of 20-40% is expected to establish a protein model of sufficient quality [81].

As open conformation or native state is the more stable form of a protein, we used an open conformation state of the protein to investigate potential substrate binding sites. To generate high-quality models, we used the MODWEB server and the Modeller program. We generated 04 open conformation models.

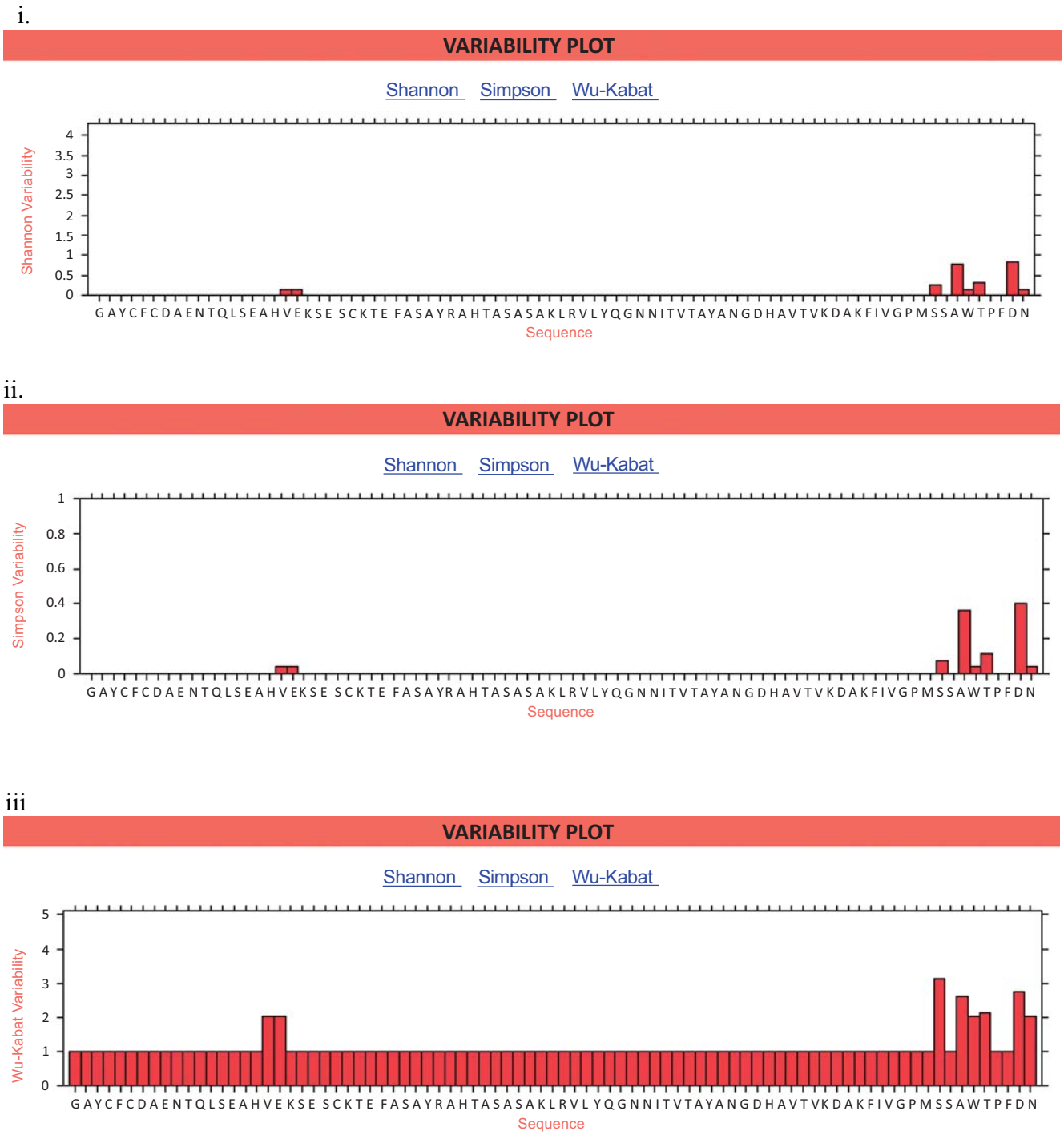


Fig. (1b). The absolute variability computed by Protein Variability Server. i) Shannon variability plot; ii) Simpson variability plot; iii) Wu-Kabat variability plot.

The validation (for quality and reliability) of generated 3D models for HLA-C 12*03 was done through various structure assessment methods, including PROCHECK, ERRAT, and Verify3D. PROCHECK evaluates the stereochemical characteristics of a protein structure by analyzing residue-by-residue geometry and overall structure geometry and utilizes the Ramachandran plotting. On the other hand, ERRAT verifies protein structures by plotting error values as a function of the position of a sliding 9-residue window. It uses an error function based on the statistics of non-bonded interatomic interactions in a given structure, while Verify3D assesses the local quality of a given structure by assigning a structural class based on the location and environment of each residue position and by comparing the results to suitable structures. In our assessment, only one out of the four models that Modeller generated could overcome the inclusion criteria set by each of the three validation procedures (Fig. **2**). PROCHECK revealed that 95.3% residues were plotted in the core region of the Ramachandran plot, while 4.1% residues were plotted in the allowed areas (usually more than

Table 1. 10 epitopes selected on the basis of NetCTL 1.2 combinatorial scores and 100% conservancy across the available strains.

Epitope Sequences that Shows 100% Conservancy	MHC Supertypes
ASAKLRVLY	A1, A3, A26, B58, B62
CKTEFASAY	A1
DAKFIVGPM	A26
FCDAENTQL	B39
GNNITVTAY	B62
HTASASAKL	A1, B39, B58
KSESCKTEF	A3
KTEFASAYR	A24
LYQGNNITV	B7
QLSEAHVEK	A1
SASAKLRVL	A1, A3, A26, B58, B62
VTAYANGDH	A1

Table 2. Seven shortlisted T-cell overlapping epitopes between MHC I and MHC II binding predictions.

Epitope	MHC- I Allele	Peptide Sequence	MHC- II Allele
ASAKLRVLY	HLA- C*12:03 HLA- C*03:03 HLA- A*30:02 HLA- A*29:02	RAHTASASAKLRVLY AHTASASAKLRVLYQ	HLA-DRB5*01:01
		RAHTASASAKLRVLY	HLA-DRB1*07:01
LYQGNNITV	HLA- C*14:02 HLA- C*12:03 HLA- C*03:03	KLRVLYQGNNITVTA SAKLRVLYQGNNITV AKLRVLYQGNNITVT LRVLYQGNNITVTAY	HLA-DRB1*13:02
HTASASAKL	HLA- B*15:02 HLA- C*03:03 HLA- A*68:02 HLA- C*15:02 HLA- C*12:03	ASAYRAHTASASAKL SAYRAHTASASAKLR AYRAHTASASAKLRV YRAHTASASAKLRVL	HLA-DRB1*01:01
		ASAYRAHTASASAKL SAYRAHTASASAKLR	HLA-DRB1*09:01
		YRAHTASASAKLRVL RAHTASASAKLRVLY AHTASASAKLRVLYQ SAYRAHTASASAKLR AYRAHTASASAKLRV	HLA-DRB1*01:01
		YRAHTASASAKLRVL RAHTASASAKLRVLY AYRAHTASASAKLRV	HLA-DRB1*07:01
		ASAYRAHTASASAKL	HLA-DRB1*04:01
GNNITVTAY	HLA- C*12:03 HLA- C*14:02 HLA- B*15:02 HLA- C*03:03	LRVLYQGNNITVTAY	HLA-DRB1*13:02

(Table 2) Contd….

Epitope	MHC- I Allele	Peptide Sequence	MHC- II Allele
SASAKLRVL	HLA- C*03:03 HLA- C*15:02 HLA- C*12:03 HLA- C*07:01 HLA- C*06:02 HLA- B*15:02	YRAHTASASAKLRVL	HLA-DRB1*01:01
		YRAHTASASAKLRVL RAHTASASAKLRVLY AHTASASAKLRVLYQ	HLA-DRB5*01:01
		YRAHTASASAKLRVL RAHTASASAKLRVLY	HLA-DRB1*07:01
KTEFASAYR	HLA- A*31:01 HLA- C*12:03 HLA- C*15:02 HLA- A*68:01 HLA- C*14:02 HLA- C*03:03	KTEFASAYRAHTASA CKTEFASAYRAHTAS SCKTEFASAYRAHTA ESCKTEFASAYRAHT	HLA-DRB1*11:01
DAKFIVGPM	HLA- C*12:03 HLA- C*03:03 HLA- C*14:02 HLA- B*15:02	KDAKFIVGPMSSAWT DAKFIVGPMSSAWTP VKDAKFIVGPMSSAW TVKDAKFIVGPMSSA	HLA-DRB1*01:01

Table 3. Allergenicity and toxicity of the T-cell epitopes.

Peptide Sequence	SVM Score	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt	Allergenicity Prediction	Toxicity Prediction
ASAKLRVLY	-0.91	-0.11	0.54	-0.23	2	1020.36	NA	NT
LYQGNNITV	-0.77	-0.02	-0.04	-0.8	0	1021.28	NA	NT
HTASASAKL	-0.53	-0.1	-0.02	-0.07	1.5	885.1	Allergen	NT
GNNITVTAY	-0.83	0.01	0.04	-0.72	0	952.16	Allergen	NT
SASAKLRVL	-0.86	-0.14	0.6	0.06	2	944.26	NA	NT
KTEFASAYR	-0.68	-0.31	-0.92	0.34	1	1072.29	NA	NT
DAKFIVGPM	-0.15	0.07	0.64	-0.18	0	977.31	Allergen	NT

SVM = Support Vector Methods; Mol wt = Molecular weight; NA = Non-Allergen; NT = Non-toxic.

Table 4. World population coverages by the selected epitopes.

Epitope Sequence	Population Coverage (%)
ASAKLRVLY	37.15
LYQGNNITV	26.05
SASAKLRVL	65.89
KTEFASAYR	40.4

90% accuracy is needed for the validity of a model [69]) (Supplementary dataset **06**). Analysis by Verifi3D revealed that above 75% (77.05%) of the residues had averaged 3D-1D score ≥ 0.2, which indicated that the constructed model was fairly reliable (Supplementary dataset **07**). Model validation by the ERRAT (ERRAT2) showed (Supplementary dataset **08**) an overall quality factor of 78.2007.

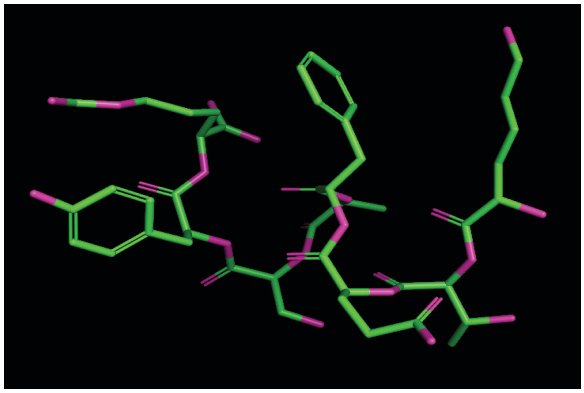
*3.2.4. Blind Docking of T-cell Epitope against HLA-C 12*03*

At first, binding simulations for both of the conserved T cell epitopes with the HLA-C 12*03 were generated using the AutoDock Vina tool in PyRx [70]. Analysis of the simulations revealed that the energy values calculated for binding both of the epitopes, SASAKLRVL and KTEFASAYR, to the binding groove of the HLA-C 12*03 were -28.0328 kJ/mol and -58.1576 kJ/mol, respectively.

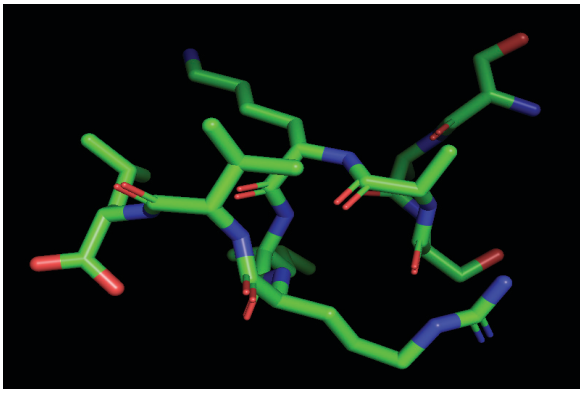
Control peptide "HKEGAFFLY" bound to the HLA-A*32:15 allele with the binding energy of -57.1825 kJ/mol. As lower binding energy favors the formation of stable interaction, we can expect that KTEFASAYR will interact with MHC-I molecules *in vivo* readily. The binding of the epitope KTEFASAYR to the HLA-C 12*03 allele was then visualized using PyMol (Fig. **3**).

3.3. Identification of B-cell Epitopes

Bepipred 2.0 and ABCpred identified 3 and 8 B-cell based potential epitopes, respectively. By comparing between the



Epitope sequence: SASAKLRVL



Epitope sequence: KTEFASAYR

Fig. (2). 3D structure of the T-cell epitopes as generated by the PEPFOLD Peptide Structure Prediction Server.

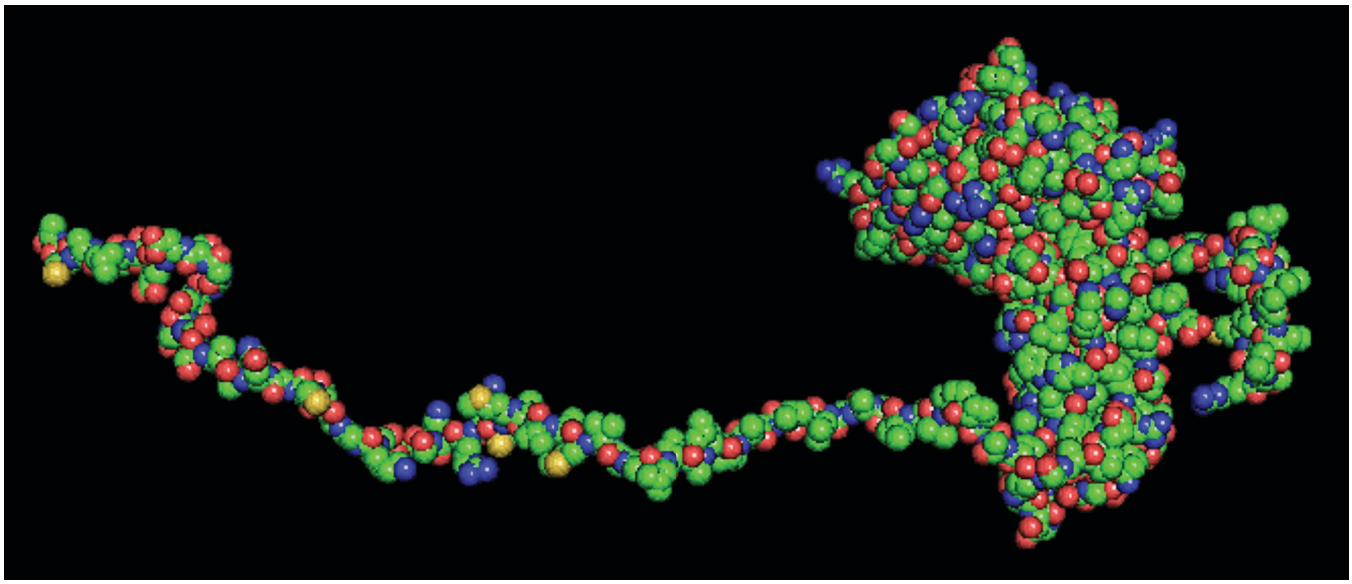


Fig. (3). Ball and surface view of Modeller 9.21 generated 3D homology model of HLA-C 12*03 molecule. PDB: 5vgd1.A was used as a template for this model.

epitopes predicted by these tools, we chose five potential B-cell epitopes for further analysis (Table **5**). These epitopes were overlapping between the two tools.

3.3.1. Surface Accessibility Properties and Conservancy Analysis

At threshold cutoff 1.0, only two peptide fragments of E1 glycoproteins were obtained to have satisfactory surface accessibility (Table **6**). Four among the five selected epitopes were found to have consensus sequences with the two peptides having surface accessibility (Tables **5**, **6**). These four epitopes, including NTQLSEAHVEKS, KSESCKT, AHVEKSESC, and DAENTQLS, were found highly conserved across all the proteins (Table **7**).

3.3.2. Antigenicity of Epitopes

Analysis of Overall Protective Antigen Prediction scores generated by the VaxiJen v2.0 server identified NTQLSEAHVEKS, AHVEKSESC, and DAENTQLS as probable antigens with values of 0.4654, 0.4108 and 1.1035 respectively. Epitope KSESCKT produced a score of -1.1135, which designated it as a probable non-antigen (Table **8**).

3.3.3. Flexibility and Hydrophilicity Analysis

Among the three antigenic epitopes, including two highly similar ones (NTQLSEAHVEKS and DAENTQLS), NTQLSEAHVEKS and AHVEKSESC were found to be highly flexible in Karplus and Schulz flexibility prediction analysis (Supplementary dataset **09**). With a window size set at seven amino acids and center position as 4, two peptide fragments of NTQLSEAHVEKS, NTQLSEA, and TQLSEAH were found to have flexibility above the threshold of 1.0 (Table **9**). At the same time, AHVEKSE and HVEKSES peptide fragments of AHVEKSESC had similar flexibility predicted.

Analysis of hydrophilicity by the Parker Hydrophilicity Analysis module at the interface of IEDB revealed that both of the epitopes were significantly hydrophilic. At a window

Table 5. Potential B-cell based epitopes identified.

Epitope Sequence	Start	End	Length (aa)
DAENTQLS	7	14	8
NTQLSEAHVEKS	10	21	12
AHVEKSESCKT	16	26	11
KSESCKT	20	26	7
NGDHAVT	59	65	7

aa = amino acids.

Table 6. Surface accessibility properties of the B-cell epitopes.

Epitope Sequence	Start	End	Length (aa)	Average Surface Accessibility Score
ENTQLS	9	14	6	1.8775
AHVEKSESC	16	24	9	1.613

aa = amino acids.

Table 7. Conservancy analysis of the B-cell epitopes.

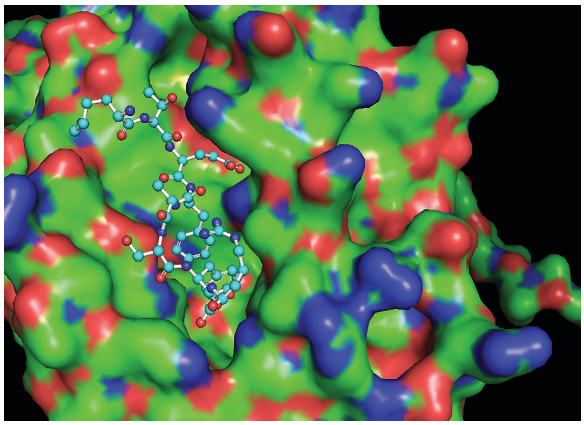
Epitope Sequence	% Matches at Identity = 100%	Minimum Identity	Maximum Identity
NTQLSEAHVEKS	98.08% (51/52)	83.33%	100.00%
KSESCKT	100.00% (52/52)	100.00%	100.00%
AHVEKSESC	98.08% (51/52)	77.78%	100.00%
DAENTQLS	100.00% (52/52)	100.00%	100.00%

Table 8. Antigenicity prediction of the B-cell epitopes by VaxiJen v2.0.

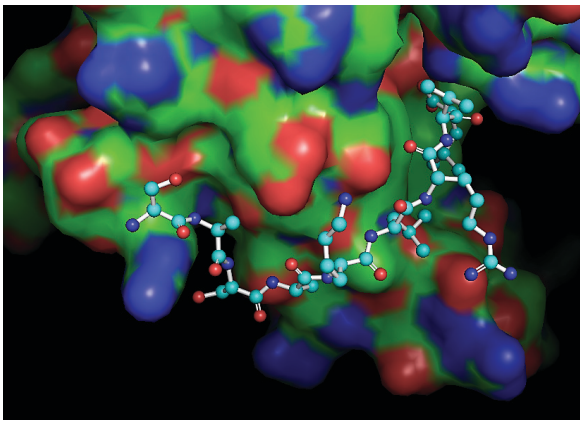
Epitope Sequence	VaxiJen v2.0 Score	Comment
NTQLSEAHVEKS	0.4654	Probable antigen
KSESCKT	-1.1135	Probable non-antigen
AHVEKSESC	0.4108	Probable antigen
DAENTQLS	1.0293	Probable antigen

Table 9. Karplus and Schulz flexibility prediction analysis of the B-cell epitopes.

Epitope	Position	Residue	Start	End	Peptide	Score
NTQLSEAHVEKS	4	L	1	7	NTQLSEA	1.024
	5	S	2	8	TQLSEAH	1.007
AHVEKSESC	4	E	1	7	AHVEKSE	1.029
	5	K	2	8	HVEKSES	1.061



KTEFASAYR docked at Grid: x – 40, y – 26, z – 34 with HLA-C 12*3; Affinity: -58.1576 kJ/mol



SASAKLRVL docked at Grid: x – 40, y – 28, z – 34 with HLA-C 12*3; Affinity: -28.0328 kJ/mol

Fig. (4). Graphical representation of the T-cell epitopes docked with HLA-C 12*03. Docking simulation was visualized with PyMol. Docking simulation by AutoDock Vina revealed that KTEFASAYR had higher binding affinity to the binding groove of HLA-C 12*03.

size of 7 amino acids and center position 4, both the epitopes resulted in hydrophilicity score above the threshold of 3.036 (Fig. **4**).

3.3.4. Cancerogenicity Analysis

Non-tumor associated gene and oncogenic domain search revealed that NTQLSEAHVEKS had no match with any oncogenic domains or genes. Hence, this epitope was listed as non-carcinogenic.

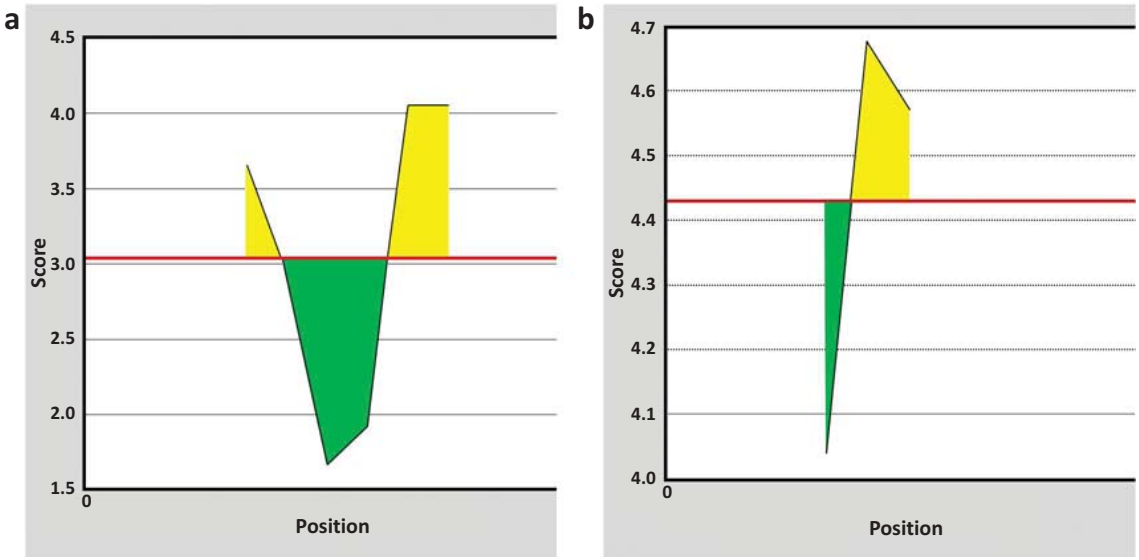
4. DISCUSSION

The present study utilized a pipeline *in silico* tool to identify major immunogenic epitopes on the E1 glycoprotein of CHIKV. In this study, we have analyzed all the available sequences of a specific structural protein of CHIKV, glycoprotein E1. Previous studies have concerted on the whole proteome or the whole sequence of the structural polyprotein (C-E3-E2-6K-E1). In fact, glycoprotein E1 and E2 carry the main antigenic determinants [18, 20]. Among these two, E2 is the most commonly utilized antigen used for the development of vaccines against alphavirus infections as natural antibodies targeting glycoprotein E1 are scanty [82, 83]. However, epitopes from glycoprotein E1 can elicit significant immunity [52, 53]. Nevertheless, there is a dearth of reports on vaccines specifically targeting the glycoprotein E1 of CHIKV. Two recent studies by Jadoon *et al.* (2019) and Narula *et al.* (2018) have also analyzed the whole proteome of CHIKV using in silico techniques [84, 85]. Another study held by ul Qamar *et al.* (2018) investigated the structural polyproteins instead of the whole proteome; however, no specific data was provided for the glycoprotein E1 [16]. Moreover, the selection of target protein is critical in any *in silico* epitope-based endeavor [59, 60]. Current computational methods for epitope mapping are underestimated because these methods often produce false-positive results when the target protein is not selected efficiently [86, 87]. However, faithful selection of target protein enables the mapping of most efficacious epitopes [59]. So, this manuscript describes one of the first hand studies on the *in silico* vaccine candidate identification specifically against glycoprotein E1 of CHIKV.

As an RNA virus, CHIKV is submissive to frequent mutations since RNA polymerase lacks proof-reading capacity [88, 89]. A putative epitope that would induce active immunity against CHIKV infection should, therefore, be derived from the portion of the target protein showing significant conservancy to will ensure efficacy across strains. We initially built a local library consisting of 52 E1 glycoprotein sequences from different strains of CHIKV, which included all the sequences from past outbreaks to most recently revealed genomic data [45]. Each of the sequences in the constructed library was then subjected to sequence alignment, which uncovered a high degree of sequence similarity, and thereby, it is expected that targeting E1 glycoprotein would produce robust and cross-reactive immunity against the infection.

Any protein targeted for epitope mapping is expected to be highly antigenic to provoke sufficient immune response. Hence, the antigenicity of the E1 glycoprotein was determined from the library we prepared by utilizing the Vaxijen server and Kolaskar & Tongaonkar antigenicity prediction tool. Analysis by both of the tools exhibited the E1 glycoprotein as an antigenic protein, and hence, it was subjected to various immunoinformatic analysis through a long pipeline of computational approaches.

Seven out of nineteen T-cell based epitopes predicted by the NetCTL 1.2 server [52] could meet the IC_{50} criterion and showed 100% conservancy across the strains (Table **2**). None of these epitopes were described in previously reported similar studies where either the whole proteome or the total structural polyprotein was examined [16, 84, 85]. For MHC-I binding prediction, peptides with IC_{50} value <50 nm were considered to have a high affinity, while <500 nm and <5000 nm were classified as intermediate and low affinity, respectively. In this study, we chose maximum alleles having binding affinity <200 nm [90].



Epitope sequence: NTQLSEAHVEKS
Average score: 3.036 (range: 1.657 - 4.043)

Epitope sequence: AHVEKSESC
Average score: 4.429 (range: 4.043 - 4.671)

Fig. (5). Hydrophilicity of the epitopes. Most of the amino acid residues of both of the epitope were hydrophilic in nature. Residues above the cut-off (horizontal red line, denoting the threshold value) are in the yellow region.

An effective T-cell based epitope vaccine is expected not to induce any allergenicity or toxicity [91]. To design allergenicity and toxicity reduced T-cell based epitope, we analyzed the allergenicity and toxicity of the predicted epitopes. Our analysis uncovered that three out of the seven predicted epitopes were non-toxic as well as non-allergenic.

It is generally accepted that promiscuous T-cell based epitopes can substantially increase the coverage of a higher proportion of populations [91, 92]. From a molecular epidemiological point of view, at least 40% of world population coverage is expected for each of the epitope-based vaccines [76]. Besides, only a highly conserved epitope would ensure the boosting of robust and long-lasting immunity. In our analysis, only two epitopes, SASAKLRVL and KTEFASAYR, exceeded 40% population coverage and hence considered as an epitope of choice.

We additionally validated each epitope by molecular docking simulation against human HLA-C 12*03 proteins, as it was found frequently in the results from MHC-I binding interaction analysis. However, before docking simulations, a favorable 3D structure of the HLA-C 12*03 molecule was generated using the technique called homology modeling by the Modeller 9.21 tool [63]. ERRAT [66], PROCHECK [67], and Verify3D [68] tools were utilized to endorse the reliability of the generated model.

Receptor-ligand molecule docking is a fast and robust technique to examine the relative binding affinity of the ligand towards its receptor. In this study, docking simulation was performed using AutoDOCK Vina in the PyRx interface [70] to evaluate the intermolecular interaction between the epitopes and the MHC class I molecule (HLA-C 12*03 allele). Since data regarding the binding pocket of HLA-C 12*03 is still scarce, we performed molecular blind docking. In the blind docking simulation, grid volume covers the whole protein for both ligand compounds, and two different conformers were generated for each ligand. Between the two conformers, conformer with the lowest affinity was considered. It was evident that KTEFASAYR obtained minimum binding affinity, and hence, this peptide was selected as the best T-cell based epitope. The binding affinity of this peptide with the HLA molecule was considerably higher than previously described T-cell based epitopes [16, 84].

B-cell epitopes are designed to mimic cognate antigens to stimulate specific B-cells to undergo proliferation and eventually elicit strong immunity, with no noticeable side effects on the human body. Attempts of predicting and producing B-cell epitopes were initiated in the early 70s, and to date, most B-cell epitope prediction methods solely depend on structural data, however, sequence data alone can be used to produce effective predictions [93-96]. *In silico* B-cell epitope prediction tools operate without extensive structural information regarding the antibody, rather intents to predict a set of residues on a given antigen capable of binding an antibody [96]. In this study, several B-cell epitope prediction software tools were utilized to find out a potent B-cell based vaccine candidate. Basically, we focused on amino acid properties within the E1 glycoprotein of CHIKV, which included hydrophobicity, hydrophilicity, antigenicity, propensities, and specific patterns [97]. In the beginning, two commonly used B-cell based epitope prediction tools, namely BepiPred 2.0 [73] and ABCpred [74], were utilized to find five potent B-cell epitopes (Table **5**). The utilization of two epitope prediction tools would reduce the chance of obtaining false-positive results. Later on, we evaluated the surface accessibility, hydrophilicity, antigenicity, and conservancy of the five predicted B-cell epitopes (Fig. **5**). Indeed, a B-cell based epitope should have the characteristics of high antigenicity and enough surface accessibility to facilitate binding with antibodies to generate immune response [98]. Hence, along with

the whole E1 protein manifesting antigenicity, it is also desirable that the individual epitopes shall rake out antigenicity separately.

Being conserved across all strains of a pathogen, an ideal epitope should also have high level of flexibility, hydrophilicity as well as non-carcinogenicty as fundamental features [77, 78, 99]. We analyzed five potential B-cell based epitopes for these fundamental properties, however, only NTQLSEAHVEKS could meet all these requirements and thus enabling itself as a B-cell epitope against CHIKV.

Distinguishing epitopes from the background is a reduction in both computational and experimental bias that bears central importance in achieving greater success [100]. This study undertook every measure to reduce computational and experimental biases in the analysis by using multiple approaches from a single evaluation and following a high threshold criterion. Moreover, the stability and efficacy of both T-cell and B-cell based epitopes could be harnessed by the utilization of adjuvants. However, it is emphasized that the stability, efficacy and, most importantly the actual immunogenic potential, as well as probable delivery methods, should be tested *in vitro* and *in vivo*.

CONCLUSION

This study determines potential epitopes for designing an epitope-based peptide universal vaccine for all pathogenic strains of the Chikungunya Virus. NTQLSEAHVEKS is predicted to be a B-cell epitope, and SASAKLRVL is predicted as a T-cell epitope. These epitopes may undergo further in-vivo and human studies to develop a vaccine against this neglected tropical disease.

LIST OF ABBREVIATIONS

CHIKV	=	Chikungunya Virus
HLA	=	Human Leukocyte Antigen
MHC	=	Major Histocompatibility Complex

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are base of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

All relevant data are included in the article and its supplementary materials section.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. Md. Shakhinur Islam Mondal (APEX Marie-Curie Postdoctoral Fellow at APC Microbiome Institute, University College Cork, Ireland) and Mr. Mahmudul Hasan (Assistant Professor and Chairman, Department of Pharmaceuticals and Industrial Biotechnology, Faculty of Biotechnology and Genetic Engineering, Sylhet Agricultural University, Sylhet, Bangladesh) for their wholehearted assistance during this study. Also, the authors take this opportunity to thank Ms. Naushin Sharmin Aurchey (Undergraduate students at the Shahjalal University of Science and Technology, Sylhet, Bangladesh) and Mr. Md. Kamrul Hasan (Faculty, Department of Biochemistry and Molecular Biology, Tejgaon College of National University Bangladesh, Dhaka, Bangladesh) for the help in completing this work.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

REFERENCES

[1] Soumahoro, M.K.; Gérardin, P.; Boëlle, P.Y.; Perrau, J.; Fianu, A.; Pouchot, J.; Malvy, D.; Flahault, A.; Favier, F.; Hanslik, T. Impact of Chikungunya virus infection on health status and quality of life: a retrospective cohort study. *PLoS One,* **2009**, *4*(11), e7800. [http://dx.doi.org/10.1371/journal.pone.0007800] [PMID: 19911058]

[2] Burt, F.J.; Chen, W.; Miner, J.J.; Lenschow, D.J.; Merits, A.; Schnettler, E.; Kohl, A.; Rudd, P.A.; Taylor, A.; Herrero, L.J.; Zaid, A.; Ng, L.F.P.; Mahalingam, S. Chikungunya virus: An update on the biology and pathogenesis of this emerging pathogen. *Lancet Infect. Dis.,* **2017**, *17*(4), e107-e117. [http://dx.doi.org/10.1016/S1473-3099(16)30385-1] [PMID: 28159534]

[3] Runowska, M.; Majewski, D.; Niklas, K.; Puszczewicz, M. Chikungunya virus: A rheumatologist's perspective. *Clin. Exp. Rheumatol.,* **2018**, *36*(3), 494-501. [PMID: 29533749]

[4] Mehta, R.; Soares, C.N.; Medialdea-Carrera, R.; Ellul, M.; da Silva, M.T.T.; Rosala-Hallas, A.; Jardim, M.R.; Burnside, G.; Pamplona, L.; Bhojak, M.; Manohar, R.; da Silva, G.A.M.; Adriano, M.V.; Brasil, P.; Nogueira, R.M.R.; Dos Santos, C.C.; Turtle, L.; de Sequeira, P.C.; Brown, D.W.; Griffiths, M.J.; de Filippis, A.M.B.; Solomon, T. The spectrum of neurological disease associated with Zika and chikungunya viruses in adults in Rio de Janeiro, Brazil: A case series. *PLoS Negl. Trop. Dis.,* **2018**, *12*(2), e0006212. [http://dx.doi.org/10.1371/journal.pntd.0006212] [PMID: 29432457]

[5] Mahendradas, P.; Avadhani, K.; Shetty, R. Chikungunya and the eye: A review. *J. Ophthalmic Inflamm. Infect.,* **2013**, *3*(1), 35. [http://dx.doi.org/10.1186/1869-5760-3-35] [PMID: 23514031]

[6] Jain, M.; Rai, S.; Chakravarti, A. Chikungunya: A review. *Trop. Doct.,* **2008**, *38*(2), 70-72. [http://dx.doi.org/10.1258/td.2007.070019] [PMID: 18453487]

[7] Li, X.; Jiang, T.; Deng, Y.; Zhao, H.; Yu, X.; Ye, Q.; Wang, H.; Zhu, S.; Zhang, F.; Qin, E. No title. In: *Complete genome sequence of a chikungunya virus isolated in Guangdong*; China, **2012**.

[8] Thiberville, S.D.; Moyen, N.; Dupuis-Maguiraga, L.; Nougairede, A.; Gould, E.A.; Roques, P.; de Lamballerie, X. Chikungunya fever: Epidemiology, clinical syndrome, pathogenesis and therapy. *Antiviral Res.,* **2013**, *99*(3), 345-370.

[http://dx.doi.org/10.1016/j.antiviral.2013.06.009] [PMID: 23811281]

[9] Vogels, C.B.F.; Rückert, C.; Cavany, S.M.; Perkins, T.A.; Ebel, G.D.; Grubaugh, N.D. Arbovirus coinfection and co-transmission: A neglected public health concern? *PLoS Biol.,* **2019**, *17*(1), e3000130. [http://dx.doi.org/10.1371/journal.pbio.3000130] [PMID: 30668574]

[10] Zeller, H.; Van Bortel, W.; Sudre, B. Chikungunya: Its history in Africa and Asia and its spread to new regions in 2013-2014. *J. Infect. Dis.,* **2016**, *214*(Suppl. 5), S436-S440. [http://dx.doi.org/10.1093/infdis/jiw391] [PMID: 27920169]

[11] Robinson, M.C. An epidemic of virus disease in Southern Province, Tanganyika territory, in 1952–1953. *Trans. R. Soc. Trop. Med. Hyg.,* **1955**, *49*, 28-32. [http://dx.doi.org/10.1016/0035-9203(55)90080-8] [PMID: 14373834]

[12] Centers for Disease Control and Prevention; National Center for Emerging and Zoonotic Infectious Diseases. https://www.cdc.gov/chikungunya/geo/index.htmlaccessed Oct 09, 2019

[13] Staples, J.E.; Fischer, M. Chikungunya virus in the Americas--what a vectorborne pathogen can do. *N. Engl. J. Med.,* **2014**, *371*(10), 887-889. [http://dx.doi.org/10.1056/NEJMp1407698] [PMID: 25184860]

[14] Mourya, D.T.; Mishra, A.C. Chikungunya fever. *Lancet,* **2006**, *368*(9531), 186-187. [http://dx.doi.org/10.1016/S0140-6736(06)69017-X] [PMID: 16844472]

[15] Coffey, L.L.; Failloux, A.B.; Weaver, S.C. Chikungunya virus-vector interactions. *Viruses,* **2014**, *6*(11), 4628-4663. [http://dx.doi.org/10.3390/v6114628] [PMID: 25421891]

[16] ul Qamar, M. T.; Bari, A.; Adeel, M. M.; Maryam, A.; Ashfaq, U. A.; Du, X.; Muneer, I.; Ahmad, H. I.; Wang, J. Peptide vaccine against chikungunya virus: Immuno-informatics combined with molecular docking approach. *J. Transl. Med.,* **2018**, *16*, 298. [http://dx.doi.org/10.1186/s12967-018-1672-7]

[17] Khan, A.H.; Morita, K.; Parquet Md, Mdel.C.; Hasebe, F.; Mathenge, E.G.; Igarashi, A. Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation site. *J. Gen. Virol.,* **2002**, *83*(Pt 12), 3075-3084. [http://dx.doi.org/10.1099/0022-1317-83-12-3075] [PMID: 12466484]

[18] Sahadeo, N.S.D.; Allicock, O.M.; De Salazar, P.M.; Auguste, A.J.; Widen, S.; Olowokure, B.; Gutierrez, C.; Valadere, A.M.; Polson-Edwards, K.; Weaver, S.C.; Carrington, C.V.F. Understanding the evolution and spread of chikungunya virus in the Americas using complete genome sequences. *Virus Evol.,* **2017**, *3*(1), vex010. [http://dx.doi.org/10.1093/ve/vex010] [PMID: 28480053]

[19] Fenwick, A. The global burden of neglected tropical diseases. *Public Health,* **2012**, *126*(3), 233-236. [http://dx.doi.org/10.1016/j.puhe.2011.11.015] [PMID: 22325616]

[20] Subudhi, B.B.; Chattopadhyay, S.; Mishra, P.; Kumar, A. Current strategies for inhibition of chikungunya infection. *Viruses,* **2018**, *10*(5), 235. [http://dx.doi.org/10.3390/v10050235] [PMID: 29751486]

[21] Rezza, G.; Weaver, S.C. Chikungunya as a paradigm for emerging viral diseases: Evaluating disease impact and hurdles to vaccine development. *PLoS Negl. Trop. Dis.,* **2019**, *13*(1), e0006919. [http://dx.doi.org/10.1371/journal.pntd.0006919] [PMID: 30653504]

[22] Ben-Yedidia, T.; Arnon, R. Epitope-based vaccine against influenza. *Expert Rev. Vaccines,* **2007**, *6*(6), 939-948. [http://dx.doi.org/10.1586/14760584.6.6.939] [PMID: 18034655]

[23] Basu, A. Immunoinformatics based study of T cell epitopes in Zea m 1 pollen allergen. *Medicina (Kaunas),* **2019**, *55*(6), 236. [http://dx.doi.org/10.3390/medicina55060236] [PMID: 31159395]

[24] Ravichandran, L.; Venkatesan, A.; Febin Prabhu Dass, J. Epitope-based immunoinformatics approach on RNA-dependent RNA Polymerase (RdRp) protein complex of Nipah Virus (NiV). *J. Cell. Biochem.,* **2018**, *120*, 7082-7095. [http://dx.doi.org/10.1002/jcb.27979] [PMID: 30417438]

[25] Shah, P.; Mistry, J.; Reche, P.A.; Gatherer, D.; Flower, D.R. *In silico* design of *Mycobacterium tuberculosis* epitope ensemble vaccines. *Mol. Immunol.,* **2018**, *97*, 56-62. [http://dx.doi.org/10.1016/j.molimm.2018.03.007] [PMID: 29567319]

[26] Sabetian, S.; Nezafat, N.; Dorosti, H.; Zarei, M.; Ghasemi, Y. Exploring dengue proteome to design an effective epitope-based vaccine against dengue virus. *J. Biomol. Struct. Dyn.,* **2019**, *37*(10), 2546-2563. [http://dx.doi.org/10.1080/07391102.2018.1491890] [PMID: 30035699]

[27] Pandey, K.R.; Ojha, R.; Mishra, A.; Prajapati, K.V. Designing B- and T-cell multi-epitope based subunit vaccine using immunoinformatics approach to control Zika virus infection. *J. Cell. Biochem.,* **2018**, *119*(9), 7631-7642. [http://dx.doi.org/10.1002/jcb.27110] [PMID: 29900580]

[28] Ingale, A.G. Epitopes identification for vaccine design and structural aspects of dengue virus 3 envelope protein *Biochem. Physiol. Open Access.,* **2014**, *3*, 1-7. [http://dx.doi.org/10.4172/2168-9652.1000134]

[29] Hasan, A.; Hossain, M.; Alam, J. A computational assay to design an epitope-based Peptide vaccine against Saint Louis encephalitis virus. *Bioinformat. Biol. Insights,* **2013**, *7*(BBI), S13402. [http://dx.doi.org/10.4137/BBI.S13402]

[30] Lapelosa, M.; Gallicchio, E.; Arnold, G.F.; Arnold, E.; Levy, R.M. In silico vaccine design based on molecular simulations of rhinovirus chimeras presenting HIV-1 gp41 epitopes. *J. Mol. Biol.,* **2009**, *385*(2), 675-691. [http://dx.doi.org/10.1016/j.jmb.2008.10.089] [PMID: 19026659]

[31] Li, L.; Jose, J.; Xiang, Y.; Kuhn, R.J.; Rossmann, M.G. Structural changes of envelope proteins during alphavirus fusion. *Nature,* **2010**, *468*(7324), 705-708. [http://dx.doi.org/10.1038/nature09546] [PMID: 21124457]

[32] Voss, J.E.; Vaney, M.C.; Duquerroy, S.; Vonrhein, C.; Girard-Blanc, C.; Crublet, E.; Thompson, A.; Bricogne, G.; Rey, F.A. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature,* **2010**, *468*(7324), 709-712. [http://dx.doi.org/10.1038/nature09555] [PMID: 21124458]

[33] Gauci, P.J.; Wu, J.Q.; Rayner, G.A.; Barabé, N.D.; Nagata, L.P.; Proll, D.F. Identification of Western equine encephalitis virus structural proteins that confer protection after DNA vaccination. *Clin. Vaccine Immunol.,* **2010**, *17*(1), 176-179. [http://dx.doi.org/10.1128/CVI.00377-09] [PMID: 19923571]

[34] Swayze, R.D.; Bhogal, H.S.; Barabé, N.D.; McLaws, L.J.; Wu, J.Q. Envelope protein E1 as vaccine target for western equine encephalitis virus. *Vaccine,* **2011**, *29*(4), 813-820. [http://dx.doi.org/10.1016/j.vaccine.2010.11.009] [PMID: 21084062]

[35] Weger-Lucarelli, J.; Aliota, M.T.; Kamlangdee, A.; Osorio, J.E. Identifying the role of E2 domains on alphavirus neutralization and protective immune responses. *PLoS Negl. Trop. Dis.,* **2015**, *9*(10), e0004163. [http://dx.doi.org/10.1371/journal.pntd.0004163] [PMID: 26473963]

[36] Weaver, S.C.; Smith, D.W. *Alphavirus Infections. Tropical Infectious Diseases: Principles, Pathogens and Practice: Principles, Pathogens and Practice;* Guerrant, R. L*;* Walker, D.H.; Weller, P.F., Eds.; Elsevier Health Sciences, **2011**.

[37] Kielian, M. Class II virus membrane fusion proteins. *Virology,* **2006**, *344*(1), 38-47. [http://dx.doi.org/10.1016/j.virol.2005.09.036] [PMID: 16364734]

[38] Islam, R.; Sakib, M.S.; Zaman, A. A computational assay to design an epitope-based peptide vaccine against chikungunya virus. *Future Virol.,* **2012**, *7*, 1029-1042. [http://dx.doi.org/10.2217/fvl.12.95]

[39] Erasmus, J.H.; Rossi, S.L.; Weaver, S.C. Development of vaccines for chikungunya fever. *J. Infect. Dis.,* **2016**, *214*(Suppl. 5), S488-S496. [http://dx.doi.org/10.1093/infdis/jiw271] [PMID: 27920179]

[40] Adhikari, U. K.; Tayebi, M.; Rahman, M. M. Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging oropouche virus. *J. Immunol. Res.,* **2018**, *2018* [http://dx.doi.org/10.1155/2018/6718083]

[41] Khan, M.A.; Hossain, M.U.; Rakib-Uz-Zaman, S.M.; Morshed, M.N. Epitope-based peptide vaccine design and target site depiction against Ebola viruses: an immunoinformatics study. *Scand. J. Immunol.,* **2015**, *82*(1), 25-34. [http://dx.doi.org/10.1111/sji.12302] [PMID: 25857850]

[42] Plemper, R.K.; Doyle, J.; Sun, A.; Prussia, A.; Cheng, L.T.; Rota, P.A.; Liotta, D.C.; Snyder, J.P.; Compans, R.W. Design of a small-

molecule entry inhibitor with activity against primary measles virus strains. *Antimicrob. Agents Chemother.,* **2005**, *49*(9), 3755-3761. [http://dx.doi.org/10.1128/AAC.49.9.3755-3761.2005] [PMID: 16127050]

[43] Starr-Spires, L.D.; Collman, R.G. HIV-1 entry and entry inhibitors as therapeutic agents. *Clin. Lab. Med.,* **2002**, *22*(3), 681-701. [http://dx.doi.org/10.1016/S0272-2712(02)00011-2] [PMID: 12244592]

[44] Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.,* **2013**, *41*(Database issue), D36-D42. [PMID: 23193287]

[45] Rahman, M.; Yamagishi, J.; Rahim, R.; Hasan, A.; Sobhan, A. East/central/south african genotype in a chikungunya outbreak, Dhaka, Bangladesh, 2017. *Emerg. Infect. Dis.,* **2019**, *25*(2), 370-372. [http://dx.doi.org/10.3201/eid2502.180188] [PMID: 30666947]

[46] Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J.D.; Higgins, D.G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.,* **2011**, *7*, 539. [http://dx.doi.org/10.1038/msb.2011.75] [PMID: 21988835]

[47] Brown, N.P.; Leroy, C.; Sander, C. MView: A web-compatible database search or multiple alignment viewer. *Bioinformatics,* **1998**, *14*(4), 380-381. [http://dx.doi.org/10.1093/bioinformatics/14.4.380] [PMID: 9632837]

[48] Waterhouse, A.M.; Procter, J.B.; Martin, D.M.; Clamp, M.; Barton, G.J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics,* **2009**, *25*(9), 1189-1191. [http://dx.doi.org/10.1093/bioinformatics/btp033] [PMID: 19151095]

[49] Garcia-Boronat, M.; Diez-Rivero, C.M.; Reinherz, E.L.; Reche, P.A. PVS: A web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res.,* **2008**, *36*(Web Server issue), W35-41 [http://dx.doi.org/10.1093/nar/gkn211] [PMID: 18442995]

[50] Doytchinova, I.A.; Flower, D.R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics,* **2007**, *8*, 4. [http://dx.doi.org/10.1186/1471-2105-8-4] [PMID: 17207271]

[51] Kolaskar, A.S.; Tongaonkar, P.C. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.,* **1990**, *276*(1-2), 172-174. [http://dx.doi.org/10.1016/0014-5793(90)80535-Q] [PMID: 1702393]

[52] Larsen, M.V.; Lundegaard, C.; Lamberth, K.; Buus, S.; Lund, O.; Nielsen, M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics,* **2007**, *8*, 424. [http://dx.doi.org/10.1186/1471-2105-8-424] [PMID: 17973982]

[53] Trolle, T.; Metushi, I.G.; Greenbaum, J.A.; Kim, Y.; Sidney, J.; Lund, O.; Sette, A.; Peters, B.; Nielsen, M. Automated benchmarking of peptide-MHC class I binding predictions. *Bioinformatics,* **2015**, *31*(13), 2174-2181. [http://dx.doi.org/10.1093/bioinformatics/btv123] [PMID: 25717196]

[54] Peters, B.; Sette, A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics,* **2005**, *6*, 132. [http://dx.doi.org/10.1186/1471-2105-6-132] [PMID: 15927070]

[55] Andreatta, M.; Trolle, T.; Yan, Z.; Greenbaum, J.A.; Peters, B.; Nielsen, M. An automated benchmarking platform for MHC class II binding prediction methods. *Bioinformatics,* **2018**, *34*(9), 1522-1528. [http://dx.doi.org/10.1093/bioinformatics/btx820] [PMID: 29281002]

[56] Bui, H.H.; Sidney, J.; Dinh, K.; Southwood, S.; Newman, M.J.; Sette, A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics,* **2006**, *7*, 153. [http://dx.doi.org/10.1186/1471-2105-7-153] [PMID: 16545123]

[57] Dimitrov, I.; Flower, D.R.; Doytchinova, I. In in AllerTOP-a server for *in silico* prediction of allergens; BMC bioinformatics. *BioMed. Central,* **2013**, *14*, S4. [PMID: 23735058]

[58] Gupta, S.; Kapoor, P.; Chaudhary, K.; Gautam, A.; Kumar, R.; Raghava, G.P. Open source drug discovery consortium. *In silico* approach for predicting toxicity of peptides and proteins. *PLoS One,* **2013**, *8*(9), e73957. [http://dx.doi.org/10.1371/journal.pone.0073957] [PMID: 24058508]

[59] Lamiable, A.; Thévenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tufféry, P. PEP-FOLD3: Faster *de novo* structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.,* **2016**, *44*(W1), W449-54. [http://dx.doi.org/10.1093/nar/gkw329] [PMID: 27131374]

[60] Thévenet, P.; Shen, Y.; Maupetit, J.; Guyon, F.; Derreumaux, P.; Tufféry, P. PEP-FOLD: An updated *de novo* structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res.,* **2012**, *40*(Web Server issue), W288-93. [http://dx.doi.org/10.1093/nar/gks419] [PMID: 22581768]

[61] Berman, H.M.; Battistuz, T.; Bhat, T.N.; Bluhm, W.F.; Bourne, P.E.; Burkhardt, K.; Feng, Z.; Gilliland, G.L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J.D.; Zardecki, C. The protein data bank. *Acta Crystallogr. D Biol. Crystallogr.,* **2002**, *58*(Pt 6 No 1), 899-907. [http://dx.doi.org/10.1107/S0907444902003451] [PMID: 12037327]

[62] Mobbs, J.I.; Illing, P.T.; Dudek, N.L.; Brooks, A.G.; Baker, D.G.; Purcell, A.W.; Rossjohn, J.; Vivian, J.P. The molecular basis for peptide repertoire selection in the Human Leucocyte Antigen (HLA) C*06:02 molecule. *J. Biol. Chem.,* **2017**, *292*(42), 17203-17215. [http://dx.doi.org/10.1074/jbc.M117.806976] [PMID: 28855257]

[63] Webb, B. Sali, A. *Protein structure modeling with MODELLER; Protein Structure Prediction*; Springer, **2014**, pp. 1-15.

[64] Salas-Burgos, A.; Iserovich, P.; Zuniga, F.; Vera, J.C.; Fischbarg, J. Predicting the three-dimensional structure of the human facilitative glucose transporter glut1 by a novel evolutionary homology strategy: Insights on the molecular mechanism of substrate migration, and binding sites for glucose and inhibitory molecules. *Biophys. J.,* **2004**, *87*(5), 2990-2999. [http://dx.doi.org/10.1529/biophysj.104.047886] [PMID: 15326030]

[65] Sippl, M.J. Recognition of errors in three-dimensional structures of proteins. *Proteins,* **1993**, *17*(4), 355-362. [http://dx.doi.org/10.1002/prot.340170404] [PMID: 8108378]

[66] Colovos, C.; Yeates, T.O. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.,* **1993**, *2*(9), 1511-1519. [http://dx.doi.org/10.1002/pro.5560020916] [PMID: 8401235]

[67] Laskowski, R.A.; MacArthur, M.W.; Thornton, J.M. PROCHECK: Validation of protein-structure coordinates. **2006**. [https://doi.org/10.1107/97809553602060000882]

[68] Eisenberg, D.; Lüthy, R.; Bowie, J.U. In [20] VERIFY3D: assessment of protein models with three-dimensional profiles *Methods in enzymology; Elsevier,* **1997**, *277*, 396-404.

[69] Gao, C. PhD. Thesis, University of Rochester, Rochester, **2010**. *Computational studies on membrane protein structures and protein-ligand binding affinities* **2009**.

[70] Dallakyan, S.; Olson, A.J. Olson, A.J. In Small-molecule library screening by docking with PyRx In: *Chemical biology*; Springer, **2015**; pp. 243-250.

[71] Lovell, S.C.; Davis, I.W.; Arendall, W.B., III; de Bakker, P.I.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure validation by Calpha geometry: φ,ψ and Cbeta deviation. *Proteins,* **2003**, *50*(3), 437-450. [http://dx.doi.org/10.1002/prot.10286] [PMID: 12557186]

[72] Seeliger, D.; de Groot, B.L. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided Mol. Des.,* **2010**, *24*(5), 417-422. [http://dx.doi.org/10.1007/s10822-010-9352-6] [PMID: 20401516]

[73] Jespersen, M.C.; Peters, B.; Nielsen, M.; Marcatili, P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.,* **2017**, *45*(W1), W24-W29. [http://dx.doi.org/10.1093/nar/gkx346] [PMID: 28472356]

[74] Saha, S.; Raghava, G.P. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins,* **2006**, *65*(1), 40-48.

[http://dx.doi.org/10.1002/prot.21078] [PMID: 16894596]
[75] Emini, E.A.; Hughes, J.V.; Perlow, D.S.; Boger, J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol.,* **1985**, *55*(3), 836-839. [PMID: 2991600]
[76] Bui, H.H.; Sidney, J.; Li, W.; Fusseder, N.; Sette, A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics,* **2007**, *8*, 361. [http://dx.doi.org/10.1186/1471-2105-8-361] [PMID: 17897458]
[77] Karplus, P.A.; Schulz, G.E. Prediction of chain flexibility in proteins. *Naturwissenschaften,* **1985**, *72*, 212-213. [http://dx.doi.org/10.1007/BF01195768]
[78] Parker, J.M.; Guo, D.; Hodges, R.S. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry,* **1986**, *25*(19), 5425-5432. [http://dx.doi.org/10.1021/bi00367a013] [PMID: 2430611]
[79] Chen, J.S.; Hung, W.S.; Chan, H.H.; Tsai, S.J.; Sun, H.S. In silico identification of oncogenic potential of fyn-related kinase in hepatocellular carcinoma. *Bioinformatics,* **2013**, *29*(4), 420-427. [http://dx.doi.org/10.1093/bioinformatics/bts715] [PMID: 23267173]
[80] Forrest, L.R.; Tang, C.L.; Honig, B. On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. *Biophys. J.,* **2006**, *91*(2), 508-517. [http://dx.doi.org/10.1529/biophysj.106.082313] [PMID: 16648166]
[81] Reddy, ChS.; Vijayasarathy, K.; Srinivas, E.; Sastry, G.M.; Sastry, G.N. Homology modeling of membrane proteins: A critical assessment. *Comput. Biol. Chem.,* **2006**, *30*(2), 120-126. [http://dx.doi.org/10.1016/j.compbiolchem.2005.12.002] [PMID: 16540373]
[82] Spurgers, K.B.; Glass, P.J. Vaccine development for biothreat alpha viruses. *J. Bioterr. Biodef.,* **2011**. [https://www.researchgate.net/publication/270014073_Vaccine_Development_for_Biothreat_Alpha_Viruses]
[83] Smith, J.L.; Pugh, C.L.; Cisney, E.D.; Keasey, S.L.; Guevara, C.; Ampuero, J.S.; Comach, G.; Gomez, D.; Ochoa-Diaz, M.; Hontz, R.D.; Ulrich, R.G. Human antibody responses to emerging mayaro virus and cocirculating alphavirus infections examined by using structural proteins from nine new and old world lineages. *MSphere,* **2018**, *3*(2), 3. [http://dx.doi.org/10.1128/mSphere.00003-18] [PMID: 29577083]
[84] Jadoon, M.H.; Rehman, Z.; Khan, A.; Rizwan, M.; Khan, S.; Mehmood, A.; Munir, A. *In silico* T-cell and B-cell epitope based vaccine design against alphavirus strain of chikungunya. *Infect. Disord. Drug Targets,* **2019**. [http://dx.doi.org/10.2174/1871526519666190521100521] [PMID: 31109281]
[85] Narula, A.; Pandey, R.K.; Khatoon, N.; Mishra, A.; Prajapati, V.K. Excavating chikungunya genome to design B and T cell multi-epitope subunit vaccine using comprehensive immunoinformatics approach to control chikungunya infection. *Infect. Genet. Evol.,* **2018**, *61*, 4-15. [http://dx.doi.org/10.1016/j.meegid.2018.03.007] [PMID: 29535024]
[86] Kozlova, E.E.G.; Cerf, L.; Schneider, F.S.; Viart, B.T.; NGuyen, C.; Steiner, B.T.; de Almeida Lima, S.; Molina, F.; Duarte, C.G.; Felicori, L.; Chávez-Olórtegui, C.; Machado-de-Ávila, R.A. Computational B-cell epitope identification and production of neutralizing murine antibodies against Atroxlysin-I. *Sci. Rep.,* **2018**, *8*(1), 14904. [http://dx.doi.org/10.1038/s41598-018-33298-x] [PMID: 30297733]
[87] Schneider, F.S.; Nguyen, D.L.; Castro, K.L.; Cobo, S.; Machado de Avila, R.A.; Ferreira, Nde.A.; Sanchez, E.F.; Nguyen, C.; Granier, C.; Galéa, P.; Chávez-Olortegui, C.; Molina, F. Use of a synthetic biosensor for neutralizing activity-biased selection of monoclonal antibodies against atroxlysin-I, a hemorrhagic metalloproteinase from *Bothrops atrox* snake venom. *PLoS Negl. Trop. Dis.,* **2014**, *8*(4), e2826. [http://dx.doi.org/10.1371/journal.pntd.0002826] [PMID: 24762927]
[88] Steinhauer, D.A.; Domingo, E.; Holland, J.J. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene,* **1992**, *122*(2), 281-288. [http://dx.doi.org/10.1016/0378-1119(92)90216-C] [PMID: 1336756]
[89] Sala, M.; Wain-Hobson, S. Are RNA viruses adapting or merely changing? *J. Mol. Evol.,* **2000**, *51*(1), 12-20. [http://dx.doi.org/10.1007/s002390010062] [PMID: 10903368]
[90] Dash, R.; Das, R.; Junaid, M.; Akash, M.F.C.; Islam, A.; Hosen, S.Z. *In silico*-based vaccine design against Ebola virus glycoprotein. *Adv. Appl. Bioinform. Chem.,* **2017**, *10*, 11-28. [http://dx.doi.org/10.2147/AABC.S115859] [PMID: 28356762]
[91] Vrtala, S.; Focke-Tejkl, M.; Swoboda, I.; Kraft, D.; Valenta, R. Strategies for converting allergens into hypoallergenic vaccine candidates. *Methods,* **2004**, *32*(3), 313-320. [http://dx.doi.org/10.1016/j.ymeth.2003.08.016] [PMID: 14962766]
[92] Vani, J.; Shaila, M.S.; Chandra, N.R.; Nayak, R. A combined immuno-informatics and structure-based modeling approach for prediction of T cell epitopes of secretory proteins of *Mycobacterium tuberculosis*. *Microbes Infect.,* **2006**, *8*(3), 738-746. [http://dx.doi.org/10.1016/j.micinf.2005.09.012] [PMID: 16476561]
[93] Meng, X.Y.; Zhang, H.X.; Mezei, M.; Cui, M. Molecular docking: A powerful approach for structure-based drug discovery. *Curr Comput. Aided Drug Des.,* **2011**, *7*(2), 146-157. [http://dx.doi.org/10.2174/157340911795677602] [PMID: 21534921]
[94] Hopp, T.P.; Woods, K.R. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA,* **1981**, *78*(6), 3824-3828. [http://dx.doi.org/10.1073/pnas.78.6.3824] [PMID: 6167991]
[95] Lin, S.Y.; Cheng, C.W.; Su, E.C. Prediction of B-cell epitopes using evolutionary information and propensity scales. *BMC Bioinformat.,* **2013**, *14*(Suppl. 2), S10. [http://dx.doi.org/10.1186/1471-2105-14-S2-S10] [PMID: 23484214]
[96] Kringelum, J.V.; Lundegaard, C.; Lund, O.; Nielsen, M. Reliable B cell epitope predictions: Impacts of method development and improved benchmarking. *PLOS Comput. Biol.,* **2012**, *8*(12), e1002829. [http://dx.doi.org/10.1371/journal.pcbi.1002829] [PMID: 23300419]
[97] Moreau, V.; Fleury, C.; Piquer, D.; Nguyen, C.; Novali, N.; Villard, S.; Laune, D.; Granier, C.; Molina, F. PEPOP: Computational design of immunogenic peptides. *BMC Bioinformatics,* **2008**, *9*, 71. [http://dx.doi.org/10.1186/1471-2105-9-71] [PMID: 18234071]
[98] Caoili, S.E.C. Benchmarking B-cell epitope prediction for the design of peptide-based vaccines: Problems and prospects. *BioMed. Res. Int.,* **2010**, *2010*
[99] Novotný, J.; Handschumacher, M.; Haber, E.; Bruccoleri, R.E.; Carlson, W.B.; Fanning, D.W.; Smith, J.A.; Rose, G.D. Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). *Proc. Natl. Acad. Sci. USA,* **1986**, *83*(2), 226-230. [http://dx.doi.org/10.1073/pnas.83.2.226] [PMID: 2417241]
[100] Bussiere, J.L. McCormick, G.C.; Green, J.D. In: *Preclinical safety assessment considerations in vaccine development; Vaccine Design*; Springer, **1995**; pp. 61-79.