



# Study of the Binding Pattern of HLA Class I Alleles of Indian Frequency and cTAP Binding Peptide for Chikungunya Vaccine Development

Prashant Saxena<sup>1,2</sup> · Sanjay Mishra<sup>2</sup>

Accepted: 27 January 2020 / Published online: 3 February 2020  
© Springer Nature B.V. 2020

## Abstract

Chikungunya is a mosquito-borne disease, caused by the member of the *Togaviridae* family belongs to the genus *alphavirus*, making it a major threat in all developing countries as well as some developed countries. The mosquito acts as a vector for the disease and carries the CHIK-Virus. To date there is no direct treatment available and that demands the development of more effective vaccines. In this study author employed Immune Epitope Database and Analysis Resource, a machine learning-based algorithm principally working on the Artificial Neural Network (ANN) algorithm, also known as (IEDB-ANN) for the prediction and analysis of Epitopes. A total of 173 epitopes were identified on the basis of IC50 values, among them 40 epitopes were found, sharing part with the linear B-cell epitopes and exposed to the cTAP1 protein, and out of 40, 6 epitopes were noticed to show interactions with the cTAP with their binding energy ranging from  $-3.61$  to  $-1.22$  kcal/mol. The six epitopes identified were exposed to the HLA class I alleles and from this all revealed interaction with the HLA alleles and minimum binding energy that ranges from  $-4.12$  to  $-5.88$  kcal/mol. Besides, two T cell epitopes i.e. <sup>145</sup>KVFTGVYPE<sup>153</sup> and <sup>395</sup>STVPVAPPR<sup>403</sup> were found most promiscuous candidates. These promiscuous epitopes-HLA complexes were further analyzed by the molecular dynamics simulation to check the stability of the complex. Results obtained from this study suggest that the identified epitopes i.e. and <sup>395</sup>STVPVAPPR<sup>403</sup>, are likely to be capable of passing through the lumen of ER to bind with the HLA class I allele and provide new insights and potential application in the designing and development of peptide-based vaccine candidate for the treatment of chikungunya.

**Keywords** Immunoinformatics · Epitope · HLA · Chikungunya · NAMD · Vaccine · cTAP · Docking

## Introduction

Apart from the variously known arboviruses, CHIKV (Chikungunya Virus) is one among them transmitted by the bite of female *Aedes* mosquito generally from the *Aedes aegypti* and *Aedes albopictus*. The outbreak of chikungunya has re-emerged in various continents including Asia and America and has become a major threat to the human population. The infection caused by the CHIKV results in a broad range of clinical symptoms starts from fever and body aches to joint pains which may lead to polyarthralgia which could persist

from days to months. Unfortunately, there is no direct treatments or vaccinations available for this severe infection.

Till today doctors follow the indirect way to treat this infection such as prescribing analgesic and antipyretic drugs to lower down body pains and fever (An et al. 2017).

Chikungunya infection had started declining in India before the massive out-break in 2016. This massive outbreak that occurred in India in the year of 2016, affecting millions of people. The positive rate of the CHIKV virus was approximately 70% among them 97% of patients were suffering from severe joint pains and restriction in movements (Kaur et al. 2017). The burden of chikungunya infection increases day by day, the reason behind that is, the vector for CHIKV and dengue are similar in some ways, which leads to confusion in the treatment plans. Most of dengue negative patients were analyzed under the chikungunya infection test and were found chikungunya positive (Dutta et al. 2017). Another reason for the increasing burden of this infection load is the unavailability of the treatment or a proper vaccination. Even

✉ Prashant Saxena  
prashant347@gmail.com

<sup>1</sup> Department of Biotechnology, K. S. Vira College of Engineering & Management, Bijnor, UP(W) 246701, India

<sup>2</sup> School of Biotechnology, IFTM University, Delhi Road (NH 24), Moradabad, UP(W) 244102, India

though various scientific groups are working independently in order to design a vaccine for this infectious disease, but still there isn't any treatment or vaccination available (Pow-ers 2017).

The era is moving very fast with the advent of technology, the chemical and enzymatic sequencing methods is becoming history and new technology known as NGS (next-generation sequencing) has taken its place. With this technology, a new ray of hope arises for developing and designing new therapeutics and vaccines for the treatment of infectious diseases (Kchouk et al. 2017).

With the advent of next-generation sequencing technology, the human genome and the genome of other organisms are being studied extensively, thus providing an increase in the amount of immunogenic data. A large volume of immunogenic data has been deposited in the immunological databases which lead the researchers to look into the emerging field of immunoinformatics (Tomar and De 2010). A group of newly developed tools in the field of immunoinformatics brought out the methods for the design of peptide-based vaccines for various microbial and viral diseases. Some of the tools help in deciphering the information from the immunogenic data available in the immunogenic database. (Groot et al. 2002).

In the cell-mediated immune response, the immune system responds only when the antigen is processed by the T cell and bound with the HLA class I and class II molecules. (Choo 2007) The MHC molecule present on the surface of TH & TC (T- Helper [CD4] & T- Cytotoxic [CD8]) cell. The CD4 & CD8 cells help in the processing of the bound antigen to the T cell for generating the effective immunogenic response to kill the antigen and provide the required immunity in the host cell. Major histocompatibility complex encodes the human leukocyte antigens genes and presents the high polymorphism. This polymorphism maintained through the successfully processing of the foreign peptides to the T cell antigen. The different loci of class I and class II allele help in the molecular diagnosis. (Williams 2001).

The challenges in the designing of peptide-based vaccines are the identification of the best methods for the identification and analysis of the short peptides as a promiscuous T cell epitope. (Kangueane and Sakharkar 2005) The cell-mediated and adaptive immunity is responsible for presenting the antigenic peptide to the cell surface of class I MHC molecules.

The small antigenic peptides must reach the lumen of the rough endoplasmic reticulum (RER) where they will bind with the class I MHC molecule and then these bounded small peptides are processed by the T cell. In the facilitation of this process transporter associated protein (TAP) (Procko and Gaudet 2005) uses energy in the form of ATP molecule. The two chains of TAP protein TAP1 and TAP2 form a cavity for the movement of the peptide molecule.

The TAP protein bound at the RER membrane and having a cytosolic domain. Each chain of TAP protein contains an ATP binding site and the transportation of small peptide is mediated by the hydrolysis of ATP molecule into ADP + Pi. The ATP induced conformational changes which close the peptide binding cavity closes to the cytosol and opens up for the endoplasmic reticulum. These peptide molecules translocated into the RER lumen, where they will be associated with class I MHC molecule.

These binding analysis of small peptide to the TAP protein and peptide with the MHC class I allele could be predicted by the in-silico approaches such as docking analysis. Docking analysis had been employed for the identification of various potent candidate as a diagnostic agent for the treatment of mosquito-borne diseases such as malaria and chikungunya (Singh and Mishra 2019).

Therefore, the present studies were undertaken the computational approach to identify and analyze short peptides, likely to act as a promiscuous epitope for the treatment of chikungunya.

## Methodology

Chikungunya virus strain of the Indian outbreak of 2006 that is DRDE-06 complete genome was retrieved from the NCBI (National Center for Biotechnology Information) database (<https://www.ncbi.nlm.nih.gov/>). The complete genome and protein sequences (Structural & Nonstructural) were retrieved (Accession No: EF210157) from the sequence database of NCBI (Santhosh et al. 2008; Sayers et al. 2019).

## Prediction of T Cell Epitopes

All the structural and non-structural proteins of the CHIKV (Chikungunya virus) was analyzed and all possible dominant HLA class I binding T cell epitopes were predicted using the immunoinformatics approach. The IEDB (Immune Epitope Database) (<https://tools.iedb.org/mhci/>) (Vita et al. 2015) was used. The tool is based on the ANN (Artificial Neural Network) (Lancashire et al. 2009) approach which is a supervised learning method. A total of 05 HLA super-type (A2, A3, A24, B7, B15) (Sidney et al. 2008) alleles that are common in the Indian population were used more specifically to predict the epitopes. The IEDB-ANN method used IC50 value for differentiating among the good binder ( $IC_{50} < 50$  nm), binder ( $IC_{50} = 500$  nm), and poor binder ( $IC_{50} > 500$  nm) (Fleri et al. 2017) Predicted epitopes have possessed the length of 9 amino acids because the preferable length of the liner T cell epitope is 8–11 amino acid (Nevagi et al. 2018).

All the predicted T cell epitopes of chikungunya have undergone the conservancy analysis (Nevagi et al. 2018)

among all available genotypes of chikungunya using the IEDB conservancy tool. For the study of the conservancy analysis, protein sequences of all genotypes belonging to the various geographical areas were collected randomly from the NCBI sequence database. The identified epitopes which were nanomeric in size and having IC<sub>50</sub> value less than 50 and the conservancy value in the range of 88–100% were selected for further analysis. The identified conserved epitopes were also checked for their TAP binding activity by the TAPPred, (Bhasin and Raghva 2004) (<https://crdd.osdd.net/raghava/tappred/>) and SVM (Support Vector Machine) (Byvatov and Schneider 2003) based tool finds the TAP affinity of the epitopes and categorized them three categories i.e. High Affinity, average affinity, and low affinity according to their predicted score. This top-down approach of immunoinformatics (Fig. 1) helped in edging a large amount of epitope to the promiscuous T cell epitope. Epitopes are known to be promiscuous if they bind with the HLA alleles of HLA supertypes. In this study, we included a total of 09 most common Indian HLA alleles (Rani et al. 2007; Dedhia et al. 2015; Chowdhryet al. 2016) to cover up the maximum population.

### B Cell Epitope Sequence Contains the T Cell Epitope Sequence

For the prediction of T cell epitopes who are the part of B cell epitopes, all the structural and non-structural proteins were analyzed by the BCPred. (El-Manzalawy et al. 2008) (<https://ailab-projects1.ist.psu.edu:8080/bcpred/predict.html>) BCPred is a machine learning-based tool that uses the concept of SVM and based on the AAP method. This tool predicts the linear B cell epitopes of the length of 20 amino acids. An antibody or MHC molecule approaches the B cell & T cell epitopes, and if T cell epitopes share the part with the linear B cell epitopes it will become more achievable by the antibody and became more capable of neutralizing the effect of antigen. (Castelli et al. 2013) In this research work, we had included only those T cell epitopes who were the part linear B cell epitopes.

### Modeling and Validation of Predicted Epitopes and HLA Alleles

From the identified epitopes those epitopes which were positive in all the checking criteria such as IEDB-ANN score, population allele coverage, TAP binder, high conservancy, and B cell epitope was selected for further study and these structure of epitope were modeled using PepStrMOD (Singh et al. 2015; Kaur et al. 2007) (<https://webs.iitd.edu.in/raghava/pepstrmod/>) and the modelled structure were analyzed by the stability test by running the molecular dynamics and simulation using Amber 6.0. The sequence of HLA alleles was

retrieved from the IMGT-HLA (Robinson et al. 2001) database of that sequence whose structure is not available in the PDB database (Parasuraman 2012). The structure of HLA alleles was modeled using the Modeller 9.21 (Webb and Sali 2016), (<https://salilab.org/modeller/>) the template for building the model was selected by using BLAST (Altschul et al. 1990) and retrieved from the PDB database. The modeled HLA structure was validated using various tools such as ProSA (Wiederstein and Sippl 2007), (<https://prosa.services.came.sbg.ac.at/prosa.php>) saves (Structure Analysis and Verification Server) (<https://servicesn.mbi.ucla.edu/SAVES/>) Verify 3D, ERRAT (Colovos and Yeates 1993), Prove (Pontius et al. 1996), ProCheck (Laskowski et al. 1993), WhatCheck, ProQ (Cristobal et al. 2001) and RAMPAGE (<https://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) to check the stability of the modelled protein structure.

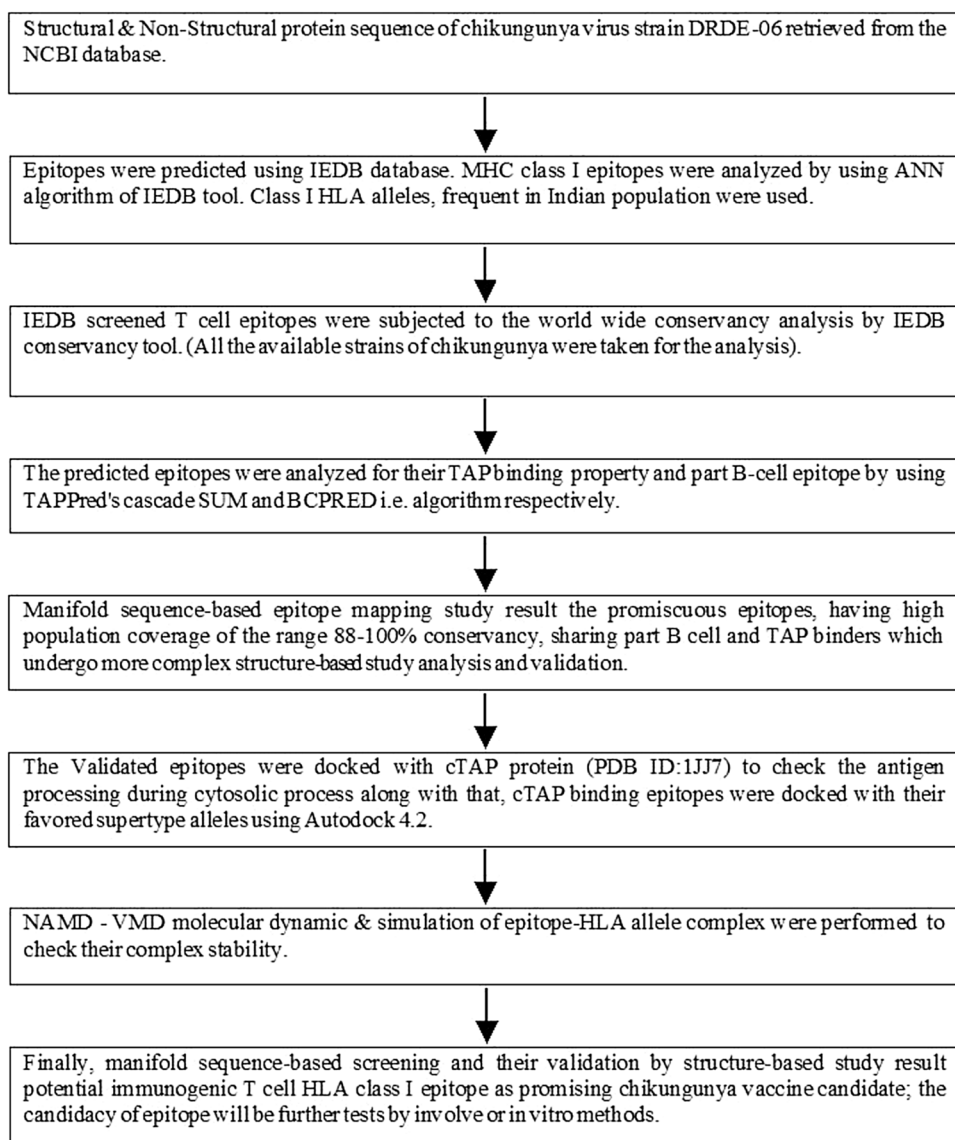
### Molecular Docking Study of Selected Epitopes with cTAP1 and HLA Allele

The selected epitopes from the structural analysis were docked against the cTAP1 protein using AutoDock 4.2 (Morris et al. 2009) (<https://autodock.scripps.edu/>) and Cygwin terminal for running the tool in the Windows environment. Those epitopes, which showed the binding with cTAP1 expose to their favored HLA allele for identification of a binding pattern of epitopes on the binding groove of HLA alleles. The binding poses were analyzed by using Discovery Studio Visualizer (BIOVIA 2015). The various parameters were analyzed for identification of binding between epitope and cTAP1 & HLA alleles. The parameters which were taken into considerations are Binding energy, number of H-Bonds, Distance of H-Bonds, K<sub>i</sub> and RMSD value.

### NMAD-VMD Simulation Study of Epitope-HLA Allele Complex

The molecular dynamics & simulation studies were performed on the epitope-HLA complexes using the NAMD (Nanoscale Molecular Dynamics) (Phillips et al. 2005) (<https://www.ks.uiuc.edu/Research/namd/>) and VMD (Visual Molecular Dynamics) (Humphrey et al. 1996) simulation. In the process of running the dynamics and simulation we generated the only ATOM file, then this file (.pdf file) used for building the PSF (Protein complex structure file) by using the auto PSF builder of VMD. To continue the process CHARMM force field (Vanommeslaeghe et al. 2010) was used and molecular dynamics were run on 310° K with 100,000 runs. The RMSD value was calculated using the command line of VMD and then rmsd.dat file is generated and using the Origin tool interactive RMSD plot was generated.

**Fig. 1** Workflow of the top-down approach of immunoinformatics, to employ the prediction of Chikungunya vaccine



The veracity of the procured result by the molecular dynamics simulation, intended on the conformational sampling and the force field used. CHARMM (Chemistry at Harvard Macromolecular Mechanics). CHARMM is an epitome in the field of MD studies, especially for the study of biomolecular systems (Vanommeslaeghe and MacKerell 2015). The CHARMM force field is versatile in nature and deals with the proteins, peptides, lipids, nucleic acids, carbohydrates, small molecules, etc. CHARMM is a dynamic method for the MD studies uses concepts of classical and mechanical energy function for biomolecular system of many contrasting classes and size. CHARMM has been recognized as the main asset for dealing with small-scale molecules. So, in this research work authors used CHARMM for the MD study of peptide-HLA complexes (Brooks et al. 2009).

## Results

### Identification of Epitopes Based on Their Conservancy, HLA Allele & TAP Binding

Epitopes or antigenic determinants of the Chikungunya virus were identified using a statistical approach. The epitopes were identified using the IEDB (Immune Epitope Database) tool. Total of five supertypes of HLA alleles was selected and a total of 9 HLA allele was used for the prediction of the epitopes. The artificial neural network (ANN) approach was of the IEDB tool and was used for the prediction of epitopes based on the IC<sub>50</sub> value. The small IC<sub>50</sub> value refers to the good binders and those were selected from the predicted epitopes. The predicted epitopes who had IC<sub>50</sub> values less than or equal to 50 nM (Nano Molar) were again examined for their conservancy in the proteome of the chikungunya



virus. A total of 11 protein sequences was used to check the conservancy of each epitope. Those epitopes found their conservancy in the range of 88–100 were used. The <sup>71</sup>TVI PNTVGV<sup>79145</sup>KVFTGVYPE<sup>153</sup> (Structural Protein) is the part of envelope glycoprotein and <sup>395</sup>STVPVAPPR<sup>40318</sup>CV-VNAANPR<sup>26</sup> (Non-Structural protein) are the part of nsP3 which are the macrodomain of the non-structural protein. All these epitopes were found conservancy score is 100%. The predicted epitopes were further analyzed for their TAP binding activity. The TAPPred tool was used for finding that predicted epitopes are High Binder, Intermediate binder or Non-Binder. It has been identified that the above-specified epitopes were falling in the category of intermediate binder to High binder with a score value of 4.238 nM, 3.841 nM, 3.662 nM, & 4.720 nM respectively. Then again further the identified epitopes were examined whether the predicted epitopes share the part with B cell epitope or not by using the BCPred tool. It was found that the epitopes shared the part with B cell epitopes (Tables 1, 2).

The predicted T cell epitopes who shares the part with B cell epitopes are <sup>64</sup>VSAYEHVTVIPNTVGVVPYKT<sup>83</sup>, <sup>40</sup>PDYSCKVFTGVYPFMWGGAY<sup>159</sup>, <sup>388</sup>AVSDWVM STVPV APPRRRRGRN<sup>409,12</sup>AKNDEECVVNAANPRGLPGDGV<sup>33</sup> with the predicted score value of 0.912, 0.819, 1.0 & .999 respectively.

### Modeling of Epitopes and HLA Alleles

The predicted epitopes who met with all the designated criteria were modeled using the PepStrMOD, a peptide modeling tool, the modeled epitopes were gone under molecular dynamics simulation using AMBER 6.0 to find out their stability. The structure of HLA class I alleles (Human Leukocyte Antigens) were modeled using the Modeller v.21.0. The protein sequence of HLA-A\*02:06, HLA-A\*11:01, HLA-A\*68:01 & HLA-A\*31:01 were retrieved from the IMGT/HLA database in the FASTA format. The sequence of all the mentioned HLA alleles was run against the PDB (Protein Databank) using BLAST (Basic Local Alignment Tool) for the identification of a template for building the model with the method of homology modeling. The templates were retrieved from PDB with the PDB ID 6APN, 4NQX, 6EI2 respectively for modeling the structure of HLA class I alleles (Fig. 2a, b). The modeled HLA alleles class I structure were analyzed for quality testing (Table 3). The PROSA, Errat, ProQ, PROVE, Verify3D, ProCheck& RAM-PAGE tools were used for the quality testing of the predicted model. The model was evaluated by their Z score, quality factor, LG score, and the Ramachandran plot (Fig. 3a, b).

### Structure Variation in the Modelled HLA Class I Alleles

The predicted structure of HLA class I alleles was analysed to find the available variation in the structure to predict the amount of difference among them. The Chimera 1.13 0.1 tool was used for predicting the RMSD (Root Mean Square Deviation) value among the HLA class I alleles. The predicted RMSD value was 1.374 Å keeping HLA-A\*02:06 as a reference molecule (Fig. 4). The RMSD value defines the present variability in the HLA alleles even though the visually the modeled structure looks similar.

### Molecular Docking Studies Between Structured Epitope and cTAP1 and Epitope and HLA Alleles

The molecular docking studies were performed in two categories. Firstly, the binding patterns of structures epitopes were identified with the cTAP1 (PDB ID: 1JJ7). The binding analysis of epitopes and cTAP1 decipher that weather the epitope is capable of reaching the lumen of ER where it will bind with HLA alleles or not. Among the 40 predicted epitopes, only six epitopes were found interacting with the cTAP1 protein molecule. The detail interaction and binding energy analysis are shown in Table 4. In the second docking category, the identified six epitopes who shown interaction with the cTAP1 protein molecule were docked against their respective HLA alleles. All the 06 epitopes shown the interaction with the HLA molecule and they show good binding energy. The detailed analysis has been shown in Tables 4 and 5.

Tables 4 and 5 include the binding energy, inhibition constant, RMSD value from the reference structure, interactions and the distance of H-bonds. The binding energy of the docking studies includes intermolecular energy, Vander wall's energy, de-solvation energy, and electrostatic energy. All these energies combinedly refer to as Binding energy. Ki refers to as the inhibition constant for the analysis and measured in the form of Micro molar (µM). RMSD stands for the root mean square deviation and describes the deviation in the molecular structure from its reference structure. It is measured in the unit of Angstrom (Å). The distance of H bonds refers to the length of the bond between the donor and the acceptor atom. The shorter the length stronger the bond, while the acceptable length of the bond length is 4 Å (Fig. 5a, b).

### Molecular Dynamics and Simulation Study

Molecular dynamics and simulation studies are the way to understand the actual behavior of the molecule in the computer system with the defined parameters. The molecular

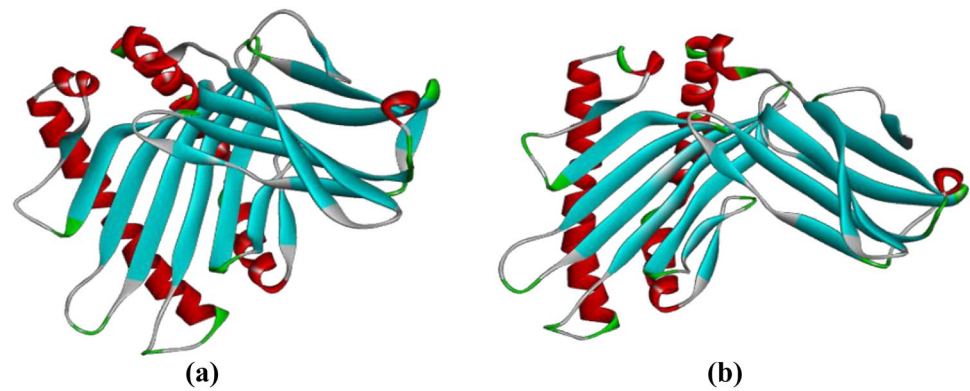
**Table 1** Identified HLA Class I epitopes of CHIKV structural protein and their analysis

S. no	T cell epitope (start–end position)	HLA allele	Protein	IC <sub>50</sub> value (nM)	B cell epitope property	Conservancy (%)	TAP binding score
1	LQISFSTAL	A-0206	E1	6.8	No	100	4.076
2	TVIPNTVGV		E1	11.11	Yes (0.912)	100	4.238
3	FIVGPMSSA		E1	14.19	No	100	7.704
4	SLTDMSCFV		E1	18.99	No	100	5.689
5	YSCKVFTGV		E1	19.3	No	100	8.145
6	KVFTGVYYPF		E1	26.55	Yes (0.819)	100	3.841
7	SQLQISFST		E1	30.02	No	100	5.026
8	KTVIPSPYV		E1	49.82	No	100	3.959
9	TVIPSPYVK	A-1101	E1	10.18	No	100	3.984
10	SAWTPFDNK		E1	18.31	No	100	3.865
11	TVIPSPYVK	A-6801	E1	14.61	No	100	3.984
12	HAAVTNHKK		E2	17.66	No	100	Non-binder
13	ITVNGQTVR		E2	25	No	100	Non-binder
14	FTIPTGAGK		C	8.79	Yes (1.0)	100	Non-binder
15	DASKFTHEK		C	18.27	No	100	Non-binder
16	VYNMDYPPF	A-2402	E1	32.56	No	100	6.568
17	QYNSPLVPR	A-3101	E2	12.18	No	100	Non-binder

**Table 2** Identified HLA Class I epitopes of CHIKV non-structural protein and their analysis

S. no	T cell epitope (start–end position)	HLA allele	Protein	IC <sub>50</sub> value (nM)	B cell epitope property	Conservancy (%)	TAP binding score (nM)
1	FVVPSLWSS	A-0206	nSP-1	4.85	No	100	3.092
2	VQAEFDSFV		nSP-1	5.93	No	100	Non-Binder
3	AIYQDVYAV		nSP-1	12.19	No	100	5.832
4	YQDVYAVHA		nSP-1	13.37	No	100	Non-binder
5	IIVCSSFPL		nSP-3	8.78	No	100	3.249
6	WVMSTVPVA		nSP-3	45.98	No	100	Non-binder
7	TQMRELPTL		nSP-4	13.2	No	100	6.178
8	ITDEYDAYL		nSP-4	41.71	No	100	3.254
9	MTFDTFQNK	A-1101	nSP-2	9.99	No	100	3.133
10	SSFPLPKYK		nSP-3	8.28	No	100	5.117
11	STVPVAPPR		nSP-3	22.58	Yes (1.0)	100	3.662
12	MTFDTFQNK	A-6801	nSP-2	7.06	No	100	3.133
13	YVYDQRR		nSP-2	8.84	No	100	5.690
14	ITTDVMRQR		nSP-2	22.9	No	100	3.384
15	STVPVAPPR		nSP-3	9.18	Yes (1.0)	100	3.662
16	CVVNAANPR		nSP-3	19.13	Yes (0.999)	100	4.720
17	SSFPLPKYK		nSP-3	37.1	No	100	5.117
18	LTTYVTKLK		nSP-4	17.91	No	100	Non-binder
19	VYKKWPESF	A-2402	nSP-3	12.37	No	100	5.399
20	KYHCVCMPR	A-3101	nSP-1	8.84	No	100	Non-binder
21	KQICVTTRR		nSP-2	14.32	No	100	4.165
22	STVPVAPPR		nSP-3	44.48	Yes (1.0)	100	3.662
23	NMKATHIQR		nSP-4	14.38	No	100	5.856

**Fig. 2** Modelled structure of HLA class I alleles. **a** Molecular structure of HLA-A\*02:06 allele and **b** Molecular structure of HLA-A\*31:01 allele



**Table 3** Comparative quality analysis of modelled structure by different computational tool

S. no	HLA allele HLA-A*	<sup>a</sup> Verify 3D	<sup>b</sup> ERRAT	<sup>c</sup> PROVE	<sup>d</sup> Pro- Check E W P	<sup>e</sup> ProSa	<sup>f</sup> PROQ	<sup>g</sup> RAMPAGE
1	02:06	98.18	94.4223	2.8	0 6 3	-9.14	2.981	~98.9
2	11:01	99.64	91.5709	3.8	4 3 2	-8.87	2.369	~95.6
3	31:01	98.55	98.4848	2.3	1 5 3	-8.92	2.630	~98.2
4	68:01	99.27	98.4906	2.3	0 6 3	-9.32	2.811	~97.4

<sup>a</sup>At least 80% amino acid residue must have 3D score  $\geq 0.2$

<sup>b</sup>It defines the quality factor of the modelled protein structure

<sup>c</sup>On the basis of Voronai radical planes predicts the quality of the structure

<sup>d</sup>Defines the available error warning and pass of the modelled structure

<sup>e</sup>The quality of the modelled structure is predicted in terms of Z score

<sup>f</sup>The quality of the modelled score is predicted in terms of LG score

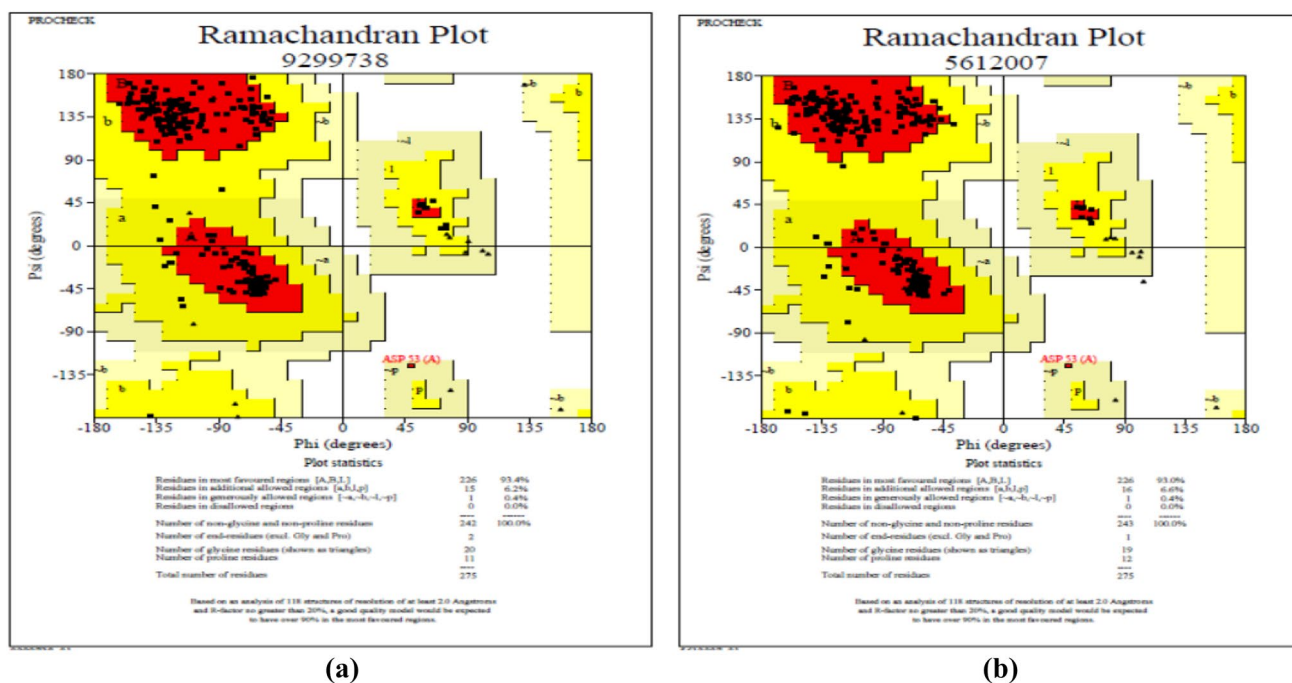
<sup>g</sup>The data represents the number of amino acids available in the favored region in the Ramachandran plot

dynamics and simulation studies were performed using the NAMD-VMD tool for all the predicted six epitopes-HLA complexes using the topology and structural files of the VMD tool. The MD simulation was run for 100,000-time steps and the minimization of energy was run for the 1000 steps by using the default parameters of the program. The simulation was run for the time step of 1 Fs (Femtosecond). After successfully running the algorithm the md.out and md.dcd files were used for the analysis of the MD simulation result. There were two graphs were plotted between the RMSD and Time (PS) and Energy and Timestep (TS). For building both the graph VMD tool was used. The protein.psf file was loaded first and then the protein\_md.dcd file was uploaded for further analysis. The NAMD plot tool of VMD was used to plot the Energy vs Time step graph and the RMSD trajectory tool was used to plot the RMSD vs Time graph. In the Energy vs TS graph protein\_wb\_md.out file of MD simulation was used as a Y-axis and time steps kept on the X-axis. In the RMSD vs Time graph protein\_wb.psf and

protein\_wb\_md.dcd files were used. The RMSD trajectory tool first performs the alignment and then plot the RMSD vs Time graph using the time slot 0.0–1.0. Among the six identified epitopes HLA complex epitopes i.e. <sup>145</sup>KVFTG VYPE<sup>153</sup> & <sup>395</sup>STVPVAPPR<sup>403</sup> (1 from structural & 1 from Non-structural protein) were found stable in the MD simulation analysis.

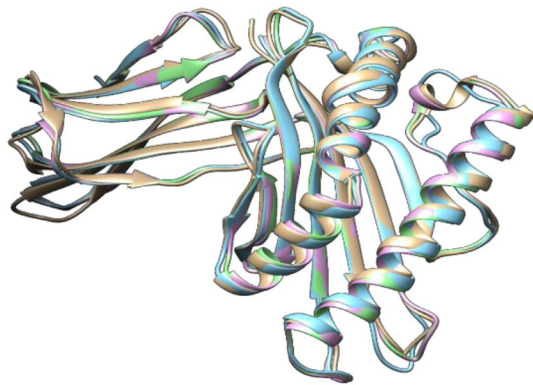
The energy graph explains that at the beginning of the MD simulation, there were lots of fluctuations present and the molecule was looking for the constant energy point, but with the progression of the MD simulations, it was found that energy was continuously increasing and it got stable approx. the 300 kcal/mol and it stops stepping up and goes the steady-state. The graph shown in Figs. 6, 7 depicts that the Epitope-HLA complex was reached to the stable state after the completion of the process of MD simulation.

The RMSD time graph depicts the variations that occurred in the Epitope-HLA complex with the change



**Fig. 3** The Ramachandran plot of the modelled structure generated by the SAVES server to predict the stability of the modelled structure of protein molecules. **a** The modelled protein structure of HLA-A\*02:06 has 93.4% amino acid residue in the favored region while 0.0% amino

acid residue falls in the disallowed region. **b** The modelled protein structure of HLA-A\*31:01 has 93.0% amino acid residue in the favored region while 0.0% amino acid residue falls in the disallowed region



**Fig. 4** The Superimposition model of HLA class I alleles. Each color represents a different HLA allele. The green color shows HLA-A\*68:01, Violet color shows HLA-A\*31:01, Blue color shows HLA-A\*11:01 and Brown color shows the structure of HLA-A\*02:06 respectively

in the time. As long as the process of MD simulation proceeds the variations in the molecule kept fluctuating and it found maximum at time 1709 with RMSD 1.766 Å (Fig. 2b) and 1547 with the RMSD of 1.74 Å (Fig. 2b) respectively.

## Discussion

In the present study, authors have used the immunoinformatics top-down approach for the prediction of the promiscuous T cell epitope for designing vaccine for the treatment of chikungunya. This research work started with the prediction of the nanomeric T cell epitopes by using the immune epitope database. The ANN (Artificial Neural Network) algorithm of the IEDB database was used. Epitopes with IC<sub>50</sub> values below 50 nM were selected. Only two of the T cell epitopes among all identified epitopes were found highly conserved, they also share the part with the B cell linear epitope and they show good binding energy with the cTAP1 and HLA class I allele. The predicted epitopes show the good binding with the TAP molecule which infers the movement of the peptides through the lumen of RER (Rough endoplasmic reticulum) with help of TAP protein to bind with the MHC class I allele. To understand the structural analysis, the modelled epitopes were docked with cTAP1 protein and HLA class I alleles using the AutoDock 4.2. In this study author had performed docking studies with the downloaded structure of HLA-A\*02:06 (PDB ID:3OXR) from PDB database and the obtained energy value are not good in comparison to the modelled structure. The obtained binding energy with the PDB structure was 3.24 kcal/mol while the obtained binding energy with the modelled structure was 4.53 kcal/



**Table 4** Molecular docking analysis of identified epitope & cTAP1 protein and their interaction analysis

S. no	Epitope & cTAP1	E <sub>Binding</sub> (Kcal/Mol)	K <sub>i</sub> (mM)	RMSD (Å)	H-bonds	H-bond distance (Å)
1	KVFTGVYPE	− 1.41	92.34	139.738	GLY585–TYR7	2.62
					GLN586–THR4	3.18
					GLU587–VAL2	2.55
					TYR555–PHE9	3.31
					GLY585–VAL5	2.66
					GLN552–ARG9	3.07
2	STVPVAPPR	− 3.99	1.18	144.162	GLN586–SER1	3.35
					ALA583–PRO8	2.80
					HIS574–ARG9	2.85

**Table 5** Molecular docking analysis of identified epitope & HLA class I allele and their interaction analysis

S. no	Epitope & HLA	E <sub>binding</sub> (Kcal/Mol)	K <sub>i</sub> (μM)	RMSD (Å)	H-bonds	H-bond distance (Å)
1	TVIPNTVGV & HLA-A*02:06	− 4.58	442.19	18.547	THR34–VAL9	2.9
					GLN120–VAL7	3.3
					THR257–THR1	2.8
					GLU256–THR1	2.7
					TYR51–TLE3	3.9
2	KVFTGVYPE & HLA-A*02:06	− 5.88	48.89	17.411	PHE32–ILE3	3.3
					SER28–PHE3	3.0
					ARG30–THR4	2.8
					THR34–PHE9	2.3
					GLN120–PHE9	2.8
3	STVPVAPPR & HLA-A*11:01	− 4.64	395.95	100.402	ASP54–VAL6	2.7
					ARG30–PRO4	2.5
					TYR51–THR2	3.21
					TYR137–ARG9	3.27
					PHE32–VAL3	3.98
4	STVPVAPPR & HLA-A*31:01	− 5.65	72.42	75.956	ASP146–ARG9	3.57
					THR34–PRO7	2.66
					TYR51–THR2	2.39
					GLN120–PRO7	3.24
					GLY261–SER1	2.69
5	CVVNAANPR & HLA-A*68:01	− 5.39	112.20	70.891	ASP54–VAL3	2.88
					GLU256–ARG9	2.67
					SER28–VAL5	2.70
					ASP54–SER1	3.17
					PHE265–PRO8	3.61
6	STVPVAPPR & HLA-A*68:01	− 4.12	950.29	64.03	ASP54–ARG9	3.19
					TYR51–ASN7	2.93
					GLN121–VAL3	2.57

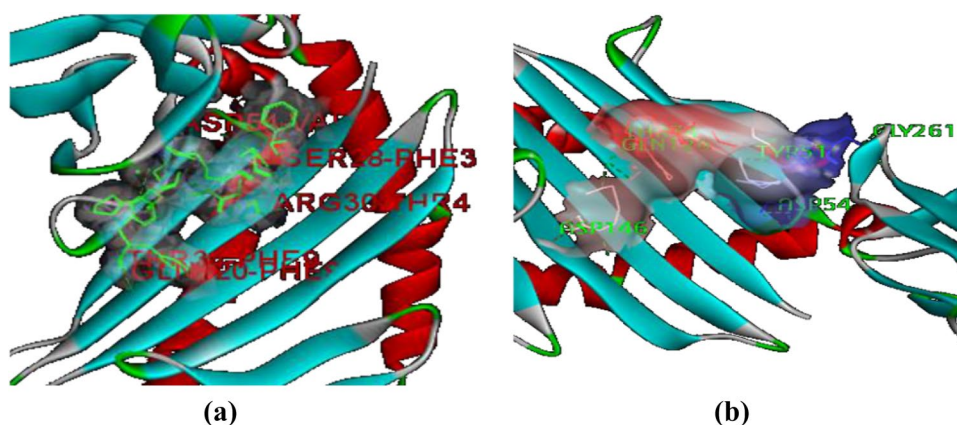
Binding energy = intermolecular energy + Vander wall's energy + De-solvation energy + electrostatic energy

K<sub>i</sub> = inhibition constant (μM)

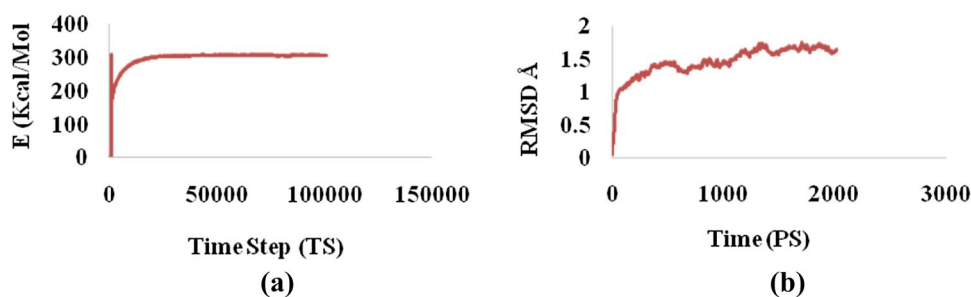
RMSD deviation in the molecular structure from its reference structure (Å)

Distance of H bonds length of the bond between the donor and acceptor atom

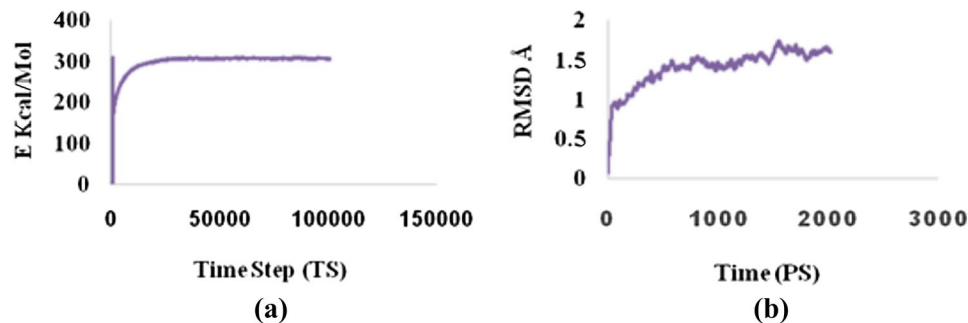
**Fig. 5** Interaction complex of epitope & HLA. **a** Depicts the interaction analysis of E1 envelope glycoprotein epitope KVFTGVYPE-HLA-A\*02:06 complex. Showing the epitope-interacting with HLA molecule in the binding groove, clearly showing the 5 H-bonds in the HLA pocket. **b** It depicts the interaction analysis of the nsp3 epitope-HLA complex. Showing the STVPVAPPR-HLA-A\*31:01 complex with 6 H-bonds



**Fig. 6** The Energy vs TS and RMSD vs Time graph for the <sup>145</sup>KVFTGVYPE<sup>153</sup>-HLA-A\*02:06 Complex obtained by the simulation study by NAMD-VMD tool



**Fig. 7** The Energy vs TS and RMSD vs Time graph for the <sup>395</sup>STVPVAPPR<sup>403</sup>-HLA-A\*31:01 Complex obtained by the simulation study by NAMD-VMD tool



mol. Due to good binding energy the modelled structure was preferred over the PDB structure. The scoring function (Rizvi et al. 2013) of the AutoDock 4.2 makes it more reliable for studying the molecular behavior and its interaction with the ligand molecule and it found that KVFTGVYPE and STVPVAPPR were highly conserved.

The KVFTGVYPE is part of the E1 envelope glycoprotein of the structural protein and STVPVAPPR is part of nsp3 of the Non-structural protein. The E1 envelope glycoprotein helps in the which facilitate the access of the viral protein to the host cell (Singh et al. 2018). And the nsp3 protein of non-structural protein of chikungunya functions as Macrodomein, functions as a regulator of viral RNA transcription (Bakar and Ng 2010). After the successful

docking and interaction analysis, the molecular dynamics simulation studies were performed by using the NAMD and VMD tool to predict the stable complex of T cell epitope and HLA class I molecules by means of their energy and RMSD values.

The present study had identified that KVFTGVYPE and STVPVAPPR as the most promiscuous ones for chikungunya treatment. The identified epitopes are nanomeric in their size and have the capacity to become the most promiscuous vaccine candidate for chikungunya. As reported previously that structural and non-structural proteins help in designing the peptide-based vaccine for Dengue and Hepatitis (Kausik et al. 2014; Henriques et al. 2013). Research had been documented, supports the immunoinformatics approach

for predicting the best and most promiscuous epitopes or antigenic determinant for the purpose of designing vaccine candidate in-silico to overcome the burden of viral infections (Sharma et al. 2016). Though in-silico studies are playing major role in the recent days' research. The cost effectiveness and rapidity in obtaining the result make in-silico studies an important asset in the today's research. Even though these advantages are there with the in-silico studies, there are certain limitations. The result obtained from the in-silico studies could not be directly employed in the real world, the obtained results must be tested in the laboratory before implementing to the real world. Another disadvantage of in-silico studies is that the result is based on the tools/methods used, so there is possibility that the obtained result might get vary and produced ambiguity among the obtained results. The reliability of the tools is the main concern in the field of in-silico studies, if tools are not statistically tested, the obtained result could not be believed to be reliable. So, in in-silico studies selection of tools with the reliable algorithm is very much important. Similarly, this study identified the promiscuous T cell epitopes which could be analyzed and tested in the laboratory as vaccine candidates for designing the vaccine for chikungunya.

## Conclusion

With the advancement of the computational power and the bioinformatics & immunoinformatics tools and techniques, it became possible to overcome the burden of healthcare to predict the most promiscuous vaccine candidates for the treatment of deadly diseases like chikungunya and dengue. In the present study, the authors reported the two T cell epitopes who found highly interacting with the HLA class I allele common in Indian populations. The identified nanomeric epitopes <sup>145</sup>KVFTGVYPE<sup>153</sup> envelope glycoprotein and <sup>395</sup>STVPVAPPR<sup>403</sup> macrodomain of the viral proteome bind with TAP protein which helps in the smooth conduction in the lumen of ER facilitate the binding with HLA class I allele. Hence, we can be stated that these identified epitopes function as the most promiscuous vaccine candidates for the treatment of chikungunya. These identified results provide new insights into the development of a platform for chikungunya vaccine research.

## Compliance of Ethical Standards

## Conflict of interest

The authors declare that they have no conflict of interest.

## Ethical Approval

The authors declare that there were no animal or human objects involved in this present study.

**Acknowledgements** The authors are thankful to the Management of IFTM University and KS Vira College of Engineering and Management for providing all the necessary facilities and support in order to carry out this research work. Besides, the authors are highly grateful to Dr. Madhvi Saxena and Jyothi Jose for support during the preparation of this manuscript.

## References

- Abu Bakar F, Ng LFP (2018) Nonstructural proteins of alphavirus potential targets for drug development. *Viruses* 10(2):71
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410
- An W, Ge N, Cao Y, Sun J, Jin X (2017) Recent progress on chikungunya virus research. *Virology* 32(6):441–453
- Bhasin M, Raghava GP (2004) Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci* 13(3):596–607
- BIOVIA DS, (2015) Discovery studio modelling environment. San Diego, Dassault Systems, Release, p 4
- Brooks BR, Brooks CL, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Cafilisch A, Cavas L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoseck M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30(10):1545–1614
- Byvatov E, Schneider G (2003) Support vector machine applications in bioinformatics. *Appl Bioinform* 2(2):67–77
- Castelli M, Cappelletti F, Diotti RA, Sautto G, Criscuolo E, Dal Peraro M, Clementi N (2013) Peptide-based vaccinology: experimental and computational approaches to target hypervariable viruses through the fine characterization of protective epitopes recognized by monoclonal antibodies and the identification of T-cell-activating peptides. *Clin Dev Immunol*. <https://doi.org/10.1155/2013/521231>
- Choo SY (2007) The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei Med J* 48(1):11–23
- Chowdhry M, Makroo RN, Singh M, Agrawal S, Kumar M, Thakur Y (2016) Human leucocyte antigen Class I and II alleles associated with anti-hepatitis C virus-positive patients of North India. *Indian J Med Microbiol* 34(3):299–302
- Colovos C, Yeates T (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 2(9):1511–1519
- Cristobal S, Zemla A, Fischer D, Rychlewski L, Elofsson A (2001) A study of quality measures for protein threading models. *BMC Bioinform* 2(1):5
- De Groot AS, Sbai H, Aubin CS, McMurry J, Martin W (2002) Immuno-informatics: mining genomes for vaccine components. *Immunol Cell Biol* 80(3):255–269
- Dedhia L, Gadekar S, Mehta P, Parekh S (2015) HLA haplotype diversity in the South Indian population and its relevance. *Indian J Transpl* 9:138–143
- Dutta P, Khan SA, Hazarika NK, Chetry S (2017) Molecular and phylogenetic evidence of chikungunya virus circulating in Assam, India. *Indian J Med Microbiol* 35:389–393

- El-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting linear B-cell epitopes using string kernels. *Journal of molecular recognition* 21(4):243–255
- Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, Sette A (2017) The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Frontiers Immunol* 8:278
- Henriques HR, Rampazo EV, Gonçalves AJS, Vicentin ECM, Amorim JH et al (2013) Targeting the non-structural protein 1 from dengue virus to a dendritic cell population confers protective immunity to lethal virus challenge. *PLOS Negl Trop Dis* 7(7):e2330
- Humphrey W, Dalke A, Schulten K (1996) VMD—visual molecular dynamics. *J Mol Graph* 14:33–38
- Kangueane P, Sakharkar MK (2005) T-epitope designer: a HLA-peptide binding prediction server. *Bioinformatics* 1(1):21–24
- Kaur H, Garg A, Raghava GPS (2007) PEPstr: a de novo method for tertiary structure prediction of small bioactive peptides. *Protein Pept Lett* 14:626–630
- Kaur N, Jain J, Kumar A et al (2017) Chikungunya outbreak in Delhi, India, 2016: report on coinfection status and comorbid conditions in patients. *New Microbes New Infect* 20:39–42
- Kaushik V, Chauhan G, Singh J (2014) In silico peptide-based vaccine design against non-structural protein 5 of hepatitis c virus. *Int J Pharm Pharm Sci* 6(9):80–82
- Kchouk M, Gibrat JF, Elloumi M (2017) Generations of sequencing technologies: from first to next generation. *Biol Med (Aligarh)* 9:395
- Lancashire LJ, Lemetre C, Ball GR (2009) An introduction to artificial neural networks in bioinformatics application to complex microarray and mass spectrometry datasets in cancer studies. *Brief Bioinform* 10(3):315–329
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK—a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283–291
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 16:2785–2791
- Nevagi RJ, Toth I, Skwarczynski M (2018) Peptide applications in biomedicine, biotechnology and bioengineering. Woodhead Publishing, Cambridge, MA, pp 327–358
- Parasuraman S (2012) Protein data bank *J Pharmacol Pharmacother* 3(4):351–352
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802
- Pontius J, Richelle J, Wodak SJ (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol* 264(1):121–136
- Powers AM (2017) Vaccine and therapeutic options to control chikungunya virus. *Clin Microbiol Rev* 31(1):104–116
- Procko E, Gaudet R (2005) Antigen processing and presentation: TAP-ping into ABC transporters. *Curr Opin Immunol* 21(1):84–91
- Rani R, Marcos C, Lazaro AM, Zhang Y, Stastny P (2007) Molecular diversity of HLA-A, -B and -C alleles in a North Indian population as determined by PCR-SSOP. *Int J Immunogenet* 34(3):201–208
- Rizvi SM, Shakil S, Haneef M (2013) A simple click by click protocol to perform docking: AutoDock 4.2 made easy for non-bioinformaticians. *EXCLI J* 12:831–857
- Robinson J, Waller MJ, Parham P, Bodmer JG, Marsh SG (2001) IMGT/HLA Database—a sequence database for the human major histocompatibility complex. *Nucleic Acids Res* 29(1):210–213
- Santhosh SR, Dash PK, Parida MM, Khan M, Tiwari M, Lakshmana Rao PV (2008) Comparative full genome analysis revealed E1: A226V shift in 2007 Indian Chikungunya virus isolates. *Virus Res* 135(1):36–41
- Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, Connor R, Fiorini N, Funk K, Hefferon T, Holmes JB, Kim S, Kimchi A, Kitts PA, Lathrop S, Lu Z, Madden TL, Marchler-Bauer A, Phan L, Schneider VA, Schoch CL, Pruitt KD, Ostell J (2019) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 47(D1):D23–D28
- Sharma PK, Srivastav S, Mishra SK, Kumar A (2016) Sequence and structure based binding prediction study of HLA Class I and cTAP binding peptides for Japanese encephalitis vaccine development. *Int J Pept Res Ther* 23:269–279
- Sidney J, Peters B, Frahm N, Brander C, Sette L (2008) HLA class I supertypes: a revised and updated classification. *BMC Immunol*. <https://doi.org/10.1186/1471-2172-9-1>
- Singh IV, Mishra S (2019) Molecular docking studies of benzamide derivatives for PfDHODH inhibitor as potent antimalarial agent. *Am J Biochem Mol Biol* 9:1–6
- 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. *Biol Direct* 10:73
- Singh A, Kumar A, Uversky VN, Giri R (2018) Understanding the interactability of chikungunya virus proteins via molecular recognition feature analysis. *R Soc Chem* 8:27293–27303
- Tomar N, De RK (2010) Immunoinformatics: an integrated scenario. *Immunology* 131(2):153–168
- Vanommeslaeghe K, MacKerell AD Jr (2015) CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim Biophys Acta* 5:861–871
- Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD Jr (2010) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 31(4):671–690
- Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK, Gabbard JL, Hix D, Sette A, Peters B (2015) The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* 43(Database issue), D405–D412
- Webb B, Sali A (2016) Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinform* 54, 5.6.1–5.6.37. <https://doi.org/10.1002/cpbi.3>
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35(Web Server issue): W407–W410
- Williams TM (2001) Human leukocyte antigen gene polymorphism and the histocompatibility laboratory. *J Mol Diagn JMD* 3(3):98–104

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.