

An Integrated *In-Silico* Approach to Develop Epitope-Based Peptide Vaccine against SARS-CoV-2

Prekshi Garg¹, Neha Srivastava^{1,2}, Prachi Srivastava^{1*}

1. AMITY Institute of Biotechnology, AMITY University Uttar Pradesh Lucknow Campus, UP, India.
2. Bioinformatics Centre, Biotech Park, Sec-G, Jankipuram, Kursi Road, Lucknow, UP, India.
[*psrivastava@amity.edu](mailto:psrivastava@amity.edu); 9453141916, 8601909444.

Abstract

SARS-CoV-2 has been the talk of the town ever since the beginning of 2020. The pandemic has brought the complete world on a halt. Every country is trying all possible steps to combat the disease ranging from shutting the complete economy of the country to repurposing of drugs and vaccine development. The rapid data analysis and widespread tools, software and databases have made bioinformatics capable of giving new insights to the researchers to deal with the current scenario more efficiently. Vaccinomics, the new emerging field of bioinformatics uses concepts of immunogenetics and immunogenomics with *in silico* tools to give promising results for wet lab experiments. This approach is highly validated for the designing and development of potent vaccines. The present *in-silico* study was attempted to identify peptide fragments from spike surface glycoprotein that can be efficiently used for the designing and development of epitope-based vaccine designing approach. Both B-cell and T-cell epitopes are predicted using integrated computational tools. VaxiJen server was used for prediction of protective antigenicity of the protein. NetCTL was studied for analyzing most potent T cell epitopes and its subsequent MHC-I interaction through tools provided by IEDB. 3D structure prediction of peptides and MHC-I alleles (HLA-C*03:03) was further done to carry out docking studies using AutoDock4.0. Various tools from IEDB were used to predict B-cell epitopes on the basis of different essential parameters like surface accessibility, beta turns and many more. Based on results interpretation, the peptide sequence from 1138-1145 amino acid and sequence WTAGAAAYY and YDPLQPEL were obtained as a potential B-cell epitope and T-cell epitope respectively. This *in-silico* study will help us to identify novel epitope-based peptide vaccine target in spike protein of SARS-CoV-2. Further, *in-vitro* and *in-vivo* study needed to validate the findings.

Keywords: SARS-CoV-2, Peptide Vaccine, Spike Protein, Vaccinomics, Epitope Prediction.

1. Introduction

The sudden outbreak of disease COVID-19 has brought the complete world under lockdown. The disease has led to a worldwide concern ever since it has originated [1]. The pandemic has affected lakhs of people all across the globe and thousands of people have lost their lives because of the deadly disease. SARS-CoV-2, the causative agent of COVID-19 is believed to originate from Wuhan, China and spread in all corners of the world [2-4]. Today, repurposing drugs and development of vaccines against SARS-CoV-2 has become an urgent need and lies as a challenge before the researchers of all countries. With the advancement in informatics-based approaches and time-consuming development process of the conventional methods, the field of vaccinomics for the development of vaccines has gained much popularity. Vaccinomics is an emerging field of biotechnology that integrates bioinformatics with immunogenomics and immunogenetics [5]. Peptide vaccines refer to an alternative vaccine strategy, that uses peptide fragments to elicit a strong immune response against a pathogen, instead of using whole organism for the purpose [6]. The use of peptide vaccines is growing ever since with the advances in biotechnology, because of the *in silico* tools and databases that give safe and promising results [7,8]. In such a pandemic situation, when the world is fighting against coronavirus, vaccinomics approach for predicting potent peptide vaccines against SARS-CoV-2 can help.

Human coronavirus is a member of *Coronaviridae* family and is known to contain a large enveloped, single stranded, positive sense RNA of approximate 30kb length. Entry of virus into the human cell is one of the most essential steps for the spread of COVID-19. SARS-CoV-2 is known to bind with angiotensin-converting enzyme 2 (ACE2) of human cell as its receptor in order to gain entry into the cell. This virus encodes a surface glycoprotein known as spike protein [9]. The receptor binding domain (RBD) of the spike protein is responsible for binding with the host cell receptor (ACE2) in order to mediate the entry of the virus into the cell. The entry of virus into the host cell is further facilitated by the release of spike fusion peptide formed due to cleavage of spike protein by the host protease [10, 11]. Therefore, it is suggested the use of potent peptide fragments of spike glycoprotein of SARS-CoV-2 as peptide vaccine against COVID-19. The current *in-silico* study was attempted to identify peptide fragments from spike surface glycoprotein that can be efficiently used for the designing and development of epitope-based vaccine using integrated computational tools for vaccine development.

2. Methodology

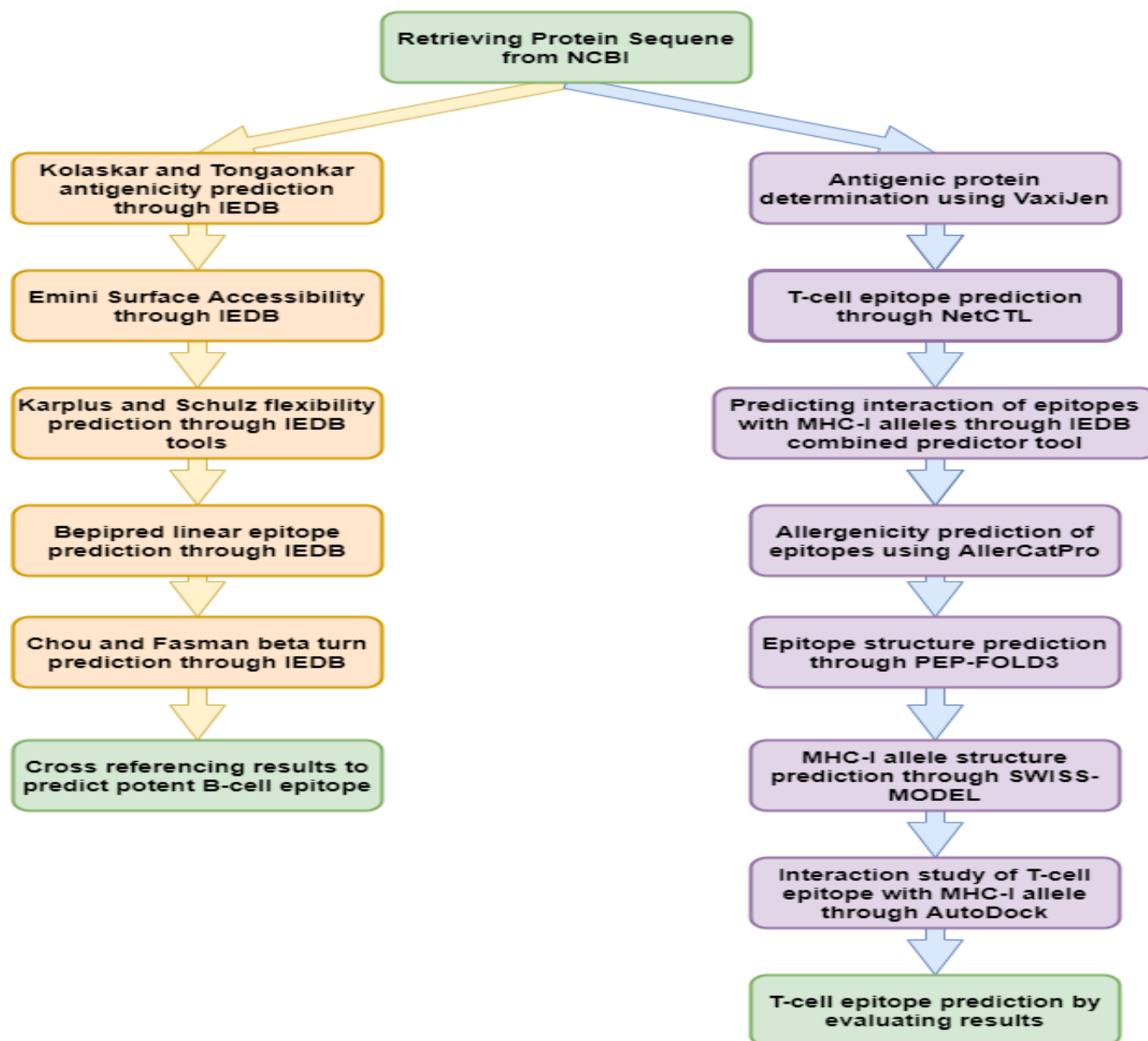


Figure 1: Complete methodology adopted for the study

2.2. Data collection and identification:

The National Centre for Biotechnology Information (NCBI) is a meta-database comprising of a series of databases that act as an important tool and resource in the field of biotechnology and biomedicine.

On the basis of literature survey, spike glycoprotein present on the surface of SARS-CoV-2 was selected as the target protein. Spike protein is known to attach to the receptor protein of host, ACE2. The sequence of spike protein of SARS-CoV-2 was retrieved from NCBI having Accession ID QIA20044. The protein was identified to be 1273 amino acids long.

2.3. Antigenic protein identification

VaxiJen v2.0 server [12] is an *in silico* tool for prediction of antigens and subunit vaccines. Default parameters of this server were used to identify most potent antigenic peptides from the protein sequence.

2.4. T-cell epitope identification and conservancy analysis

NetCTL 1.2 server [13] is an *in silico* tool used for identification of T-cell epitopes. The prediction method adopted for identifying T-cell epitopes comprised of peptide major histocompatibility complex class I (MHC-I) binding, proteasomal C terminal cleavage and TAP transport efficiency. The threshold parameter for prediction was set to 0.75 to obtain 0.80 sensitivity and 0.97 specificity. IEDB combined predictor tool was used to analyze MHC class I interaction with T-cell epitopes. IEDB analysis resource used stabilized matrix method (SMM) for the prediction [14]. The half maximal inhibitory concentration (IC_{50}) was set as a parameter for selection. Those epitopes were chosen which had $IC_{50} < 200nM$ for their binding to class I molecules. Prior to the run, the length of peptides was set at 9.0.

2.5. Allergenicity assessment

Web based AllerCatPro tool [15] was used to determine the allergenicity of the predicted antigens. Since, peptide antigens act as a foreign substance for human body, it is a very vital step to predict the allergenicity of the epitopes to ensure that the epitopes do not cause any inflammation or allergic reactions in an individual.

2.6. Designing three-dimensional (3D) structure of epitope and MHC-I alleles

PEP-FOLD3 [16] is a web-based server which was used to predict the 3D structure of the most potent epitope in order to carry out docking studies in future. PEP-FOLD3 predicted five most provable structures. The model which had the lowest energy, was selected for further analysis.

The structure of MHC-I allele was unavailable in PDB; therefore, it was necessary to design the structure of the allele. UniProt was used to determine the protein sequence of each MHC-I allele. Further, the three-dimensional structure of the allele was designed using an *in silico* tool,

SWISS-MODEL [17]. The validation of the predicted structure was done using Procheck and ERRAT.

2.7. Docking studies

The T-cell epitope which showed affinity with maximum number of MHC alleles was selected and docked using AutoDock4.0 [18] with specific MHC class I alleles in order to visualize the bonding of the antigen with the MHC so that the antigen can be presented to T-cell in human body and elicit immune response against it. The docked model was visualized using ICM Browser [19].

2.8. B cell epitope identification

B cell epitopes are identified to predict potential antigens that can interact with B lymphocytes and generate an immune response. IEDB tools were utilized to predict B cell epitopes were identified on the basis of parameters like antigenicity, flexibility, hydrophilicity, prediction of linear epitope and accessibility of surface [20]. BepiPred linear epitope prediction analysis [21], Kolaskar and Tongaonkar antigenicity scale, Emini surface accessibility prediction, Chou and Fasman prediction tool and Karplus and Schulz flexibility prediction tools provided by IEDB were used to predict flexibility, surface accessibility, linear epitopes and hydrophilicity [22].

3. Results

3.1. Antigenic protein prediction

VaxiJen server was used to predict antigenic properties of the protein. The server predicted spike protein of SARS-CoV-2 as antigenic having an overall prediction score of **0.4671**.

3.2. T cell epitope identification

The NetCTL server identifies the potent T cell epitopes from the given protein sequence in a preselected environment. Based on high combinatorial score, the best five epitopes were selected for the further study. Table 1 list down the selected epitopes along with their score.

Table 1: List of T-cell epitopes obtained from NetCTL Server

S.No.	Epitope sequence	Overall score
1.	LTDEMIAQY	3.6616
2.	WTAGAAAYY	3.1128
3.	TSNQVAVLY	3.0758
4.	CVADYSVLY	2.5759
5.	KTSVDCTMY	2.3795

3.3. MHC class I interaction analysis

The interaction of MHC-I allele with the five best epitopes was predicted through stabilized matrix base method (SMM). Those MHC-I alleles were selected for which the epitopes showed higher affinity ($IC_{50} < 200nM$). Table 2 lists down the five best epitopes along with their interacting MHC-I alleles, their total processing score and IC_{50} value.

The epitopes predicted retained the highest conservancy of 100.00% as observed from IEDB conservancy analysis tool. The results obtained are tabulated in table 3. The overall score predicted by IEDB defines the intrinsic potential of each peptide to be a T cell epitope. Higher overall score denotes higher processing capabilities of each peptide. Among the five T cell epitopes, 9mer WTAGAAAYY epitope was found to interact with maximum number of class I MHC alleles with high affinity.

Table 2: List of T-cell epitopes along with their MHC-I alleles overall score and IC_{50}

S.No	Epitope Sequence	Interacting MHC-I allele	Total Score (proteasome score, TAP score, MHC-I score, processing score)	IC_{50}
1.	LTDEMIAQY	HLA-C*12:03	1.67	6.5
		HLA-C*05:01	1.46	10.5
		HLA-A*01:01	0.66	67.1
		HLA-C*14:02	0.24	178.1
2.	WTAGAAAYY	HLA-C*03:03	1.66	6.8
		HLA-A*29:02	1.09	25.3
		HLA-C*12:03	0.77	52.7
		HLA-A*30:02	0.72	58.8

		HLA-B*15:02	0.64	71.8
		HLA-A*26:01	0.6	78.5
		HLA-B*35:01	0.57	83.1
		HLA-A*68:01	0.34	141.4
		HLA-C*14:02	0.2	195.3
3.	TSNQVAVLY	HLA-C*12:03	1.15	28.3
		HLA-C*07:01	1.11	30.9
		HLA-C*05:01	1.09	32
		HLA-B*58:01	0.57	105.5
		HLA-A*29:02	0.46	136.7
		HLA-A*30:02	0.45	141.7
		HLA-C*06:02	0.33	185.7
4.	CVADYSVLY	HLA-A*29:02	1.47	19.2
		HLA-B*35:01	1.1	45
		HLA-C*12:03	1.02	54.2
		HLA-C*03:03	1	57.6
		HLA-C*07:01	0.62	138.7
5.	KTSVDCTMY	HLA-C*12:03	1.15	31.5
		HLA-A*30:02	0.87	59.9
		HLA-C*03:03	0.84	64.8
		HLA-A*29:02	0.35	199

Table 3: Conservancy analysis results obtained from IEDB tool

Epitope	Epitope sequence	Epitope length	Percent of protein sequence matches at identity <= 100%	Minimum identity	Maximum identity
1	LTDEMIAQY	9	100.00% (1/1)	100.00%	100.00%
2	WTAGAAAYY	9	100.00% (1/1)	100.00%	100.00%
3	TSNQVAVLY	9	100.00% (1/1)	100.00%	100.00%
4	CVADYSVLY	9	100.00% (1/1)	100.00%	100.00%
5	KTSVDCTMY	9	100.00% (1/1)	100.00%	100.00%

3.4. The 3D Structure Prediction and Validation of HLA

The 3D structure of HLA-C*03:03 protein was modeled using online server SWISS-MODEL (Figure2A). Procheck and Errat server was used to validate the modeled structure of HLA. The Ramachandran plot showed 91.3% residues of the protein in the favored region, reflects the good stability and reliability of predicted structure (Figure2B). The ERRAT plot error frequency rate of 89.5911 indicates good quality rate of structure (Figure2C).

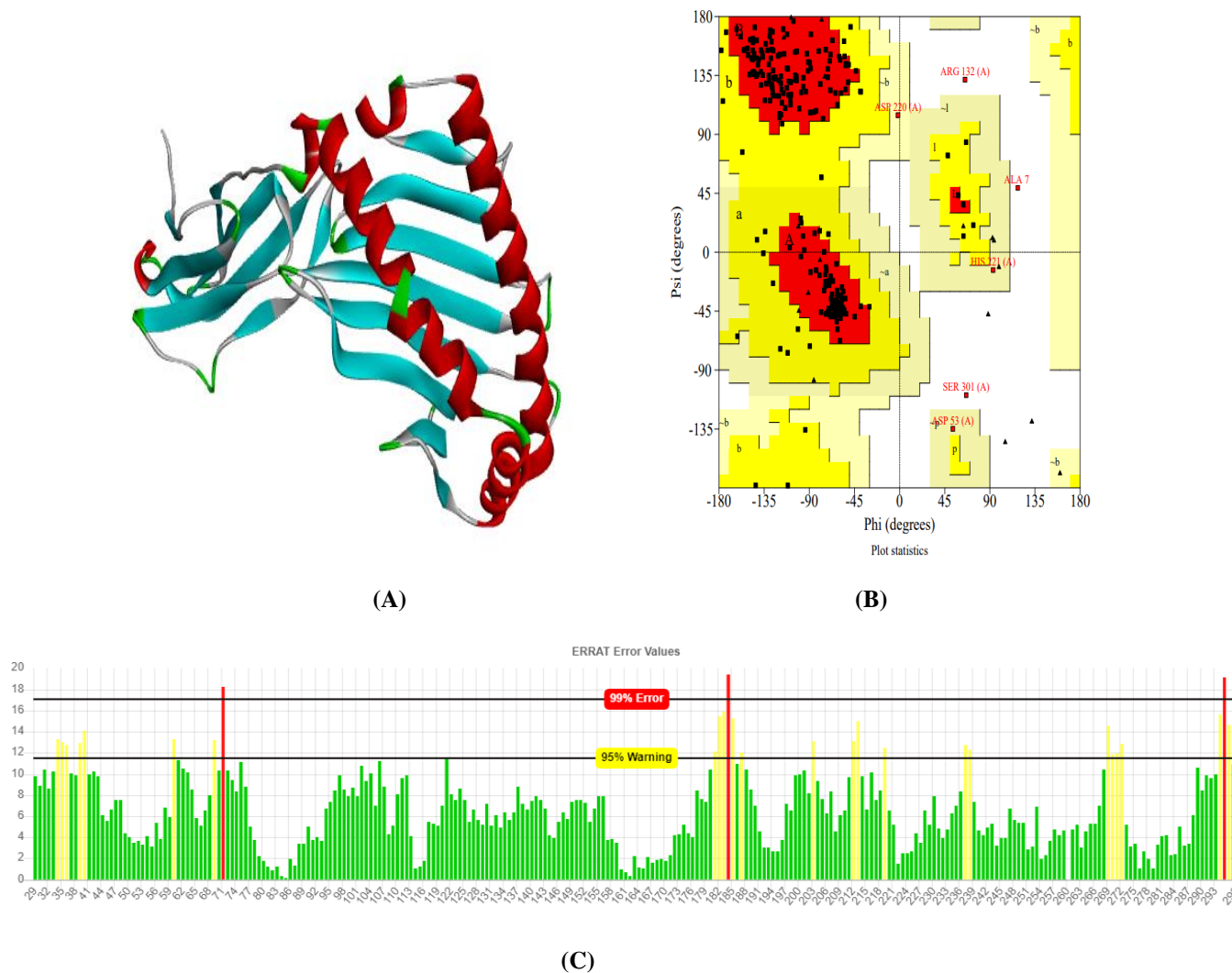


Figure2 (A) The 3D Structure of HLA-C*03:03 (B) Ramachandran plot of HLA-C*03:03 obtained from Procheck server (C) ERRAT Score of HLA-C*03:03 obtained from server ERRAT

3.5. Allergenicity prediction of epitopes

Since, the T-cell epitopes are foreign to the host body, it is very essential to predict if the selected vaccine fragments are susceptible to cause certain allergic reactions in the host body. Therefore, allergenicity prediction of all 5 T-cell epitopes was done using AllerCatPro and it was predicted that there were no evidences about the specific peptide sequences arousing any allergic reactions into the host body.

3.6. Prediction of 3D structure of Epitope

The 3D structure of the epitope WTAGAAAYY was modelled using *in silico* tool, PEP-FOLD3. Figure 3 shows the 3D structure of the epitope obtained.

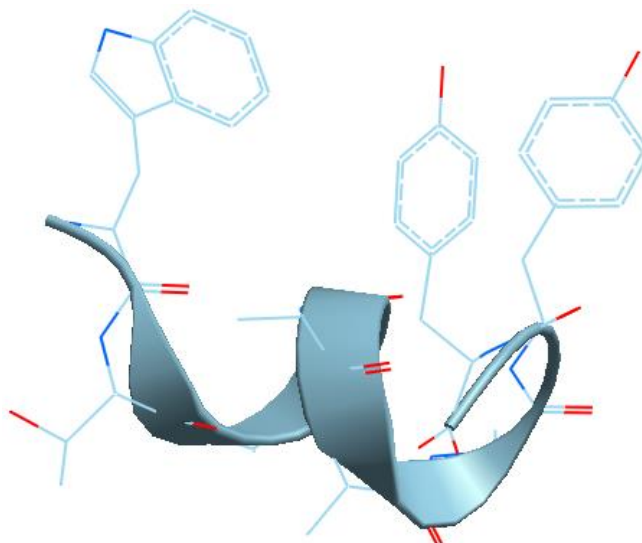


Figure 3: Structure of T-cell epitope (WTAGAAAYY) as obtained from PEP-FOLD3

3.7. Molecular docking analysis

AutoDock4.0 was used to analyze the binding of epitope WTAGAAAYY with each HLA-C*03:03 since the epitope showed highest affinity with the allele. The sequence of HLA-C*03:03 was obtained from UniProt (UniProt ID: B8YBG3) and the structure of the allele was modeled using SWISS-MODEL. The epitope bounded with HLA-C*03:03 with an energy of -5.47 kcal/mol. Figure4 depicts the binding of allele with the epitope.

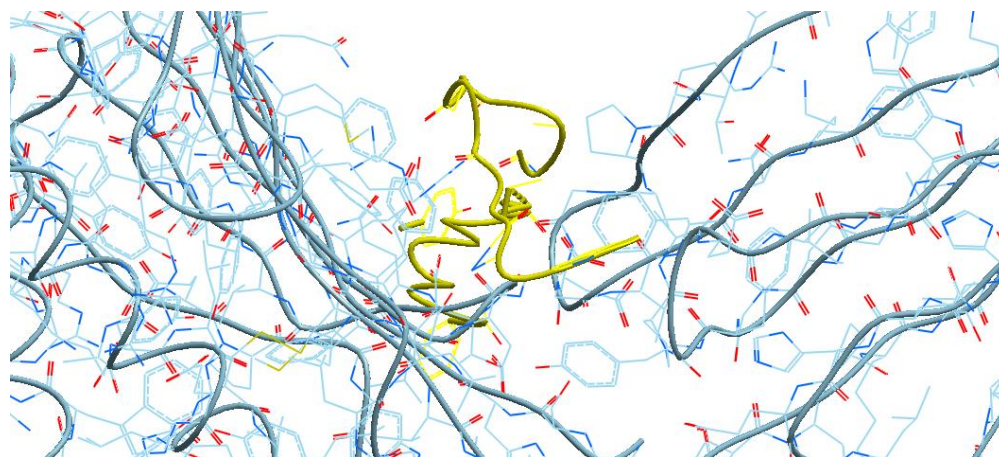


Figure 4: Epitope-HLA-C*03:03 interaction of WTAGAAAYY with MHC-I allele (HLA-C*03:03)

3.7. B cell epitope identification

Different analysis methods were used for the prediction of continuous B cell epitopes on the basis of different parameters.

The Kolaskar and Tongaonkar method is an antigenicity prediction approach for the analysis of antigenicity of protein on the basis of physicochemical properties of amino acids and their abundance in already known epitopes. The average antigenic propensity of the protein was 1.041, with maximum of 1.256 and minimum of 0.881 antigenic propensities. The antigenic determination threshold for the protein was 1.00; therefore, all the values greater than 1.00 were determined as potential antigens. Out of 49 predicted antigens, 14 antigens were found to have score above the threshold. Figure5 depicts the graph for antigenic propensity vs sequence position where X-axis represents sequence position and Y-axis represents antigenic propensity score. Table 4 lists down the predicted B cell epitope peptides obtained using Kolaskar and Tongaonkar antigenicity prediction method.

Table 4: List of predicted peptides obtained through Kolaskar and Tongaonkar method

No.	Start	End	Peptide	Length
1	113	147	KTQSLIVNNA TNVVIK VCEFQFCNDPFLGVYYHK	35
2	322	341	PTESIVRFPNITNLCPFGEV	20
3	427	437	DDFTGCVIAWN	11
4	475	496	AGSTPCNGVEGFNCYFPLQSYG	22
5	534	542	VKNKCVNFN	9
6	578	599	DPQTL EILDITPCSF GGVS VIT	22
7	643	654	FQTRAGCLIGAE	12
8	684	698	ARSVASQSIIAYTMS	15
9	911	920	VTQNVLYENQ	10
10	933	956	KIQDSLSTASALGKLQDVVNQNA	24
11	983	995	RLDKVEAEVQIDR	13
12	997	1018	ITGRLQSLQTYVTQQLIRAAEI	22
13	1090	1100	PREGVFVSNGT	11
14	1122	1145	VSGNCDVVIGIVNNTVYDPLQPEL	24

Surface accessibility is another important parameter for a potent B cell epitope. The prediction of surface accessibility of protein was done using Emini surface accessibility prediction method. The threshold value for the prediction was set to 1.000. The regions from 807 to 817 amino acid residues were found to be more accessible. Out of 24 predicted peptides, 17 peptides had score above the threshold value. Figure 6 depicts the graph of surface probability (Y axis) vs sequence position (X axis). Table 5 lists down peptide fragments with most accessibility.

Table 5: List of predicted peptides through Emini surface accessibility prediction method

No.	Start	End	Peptide	Length
1	20	38	TRTQLPPANTNSFTRGVYY	19
2	73	81	TNGTKRFDN	9
3	145	154	YHKNNKSWME	10
4	181	189	GKQGNFKNL	9
5	202	208	KIYSKHT	7
6	247	252	SYLTPG	6
7	352	357	AWNRKR	6
8	528	533	KKSTNL	6
9	553	561	TESNKKFLP	9
10	574	582	DAVRDPQTL	9
11	675	687	QTQTNSPRRARSV	13
12	772	780	VEQDKNTQE	9
13	807	817	PDPSKPSKRSF	11
14	1068	1076	VPAQEKNT	9
15	1138	1162	YDPLQPELDSFKEELDKYFKNHTSP	25
16	1203	1211	LGKYEQYIK	9
17	1258	1263	EDDSEP	6

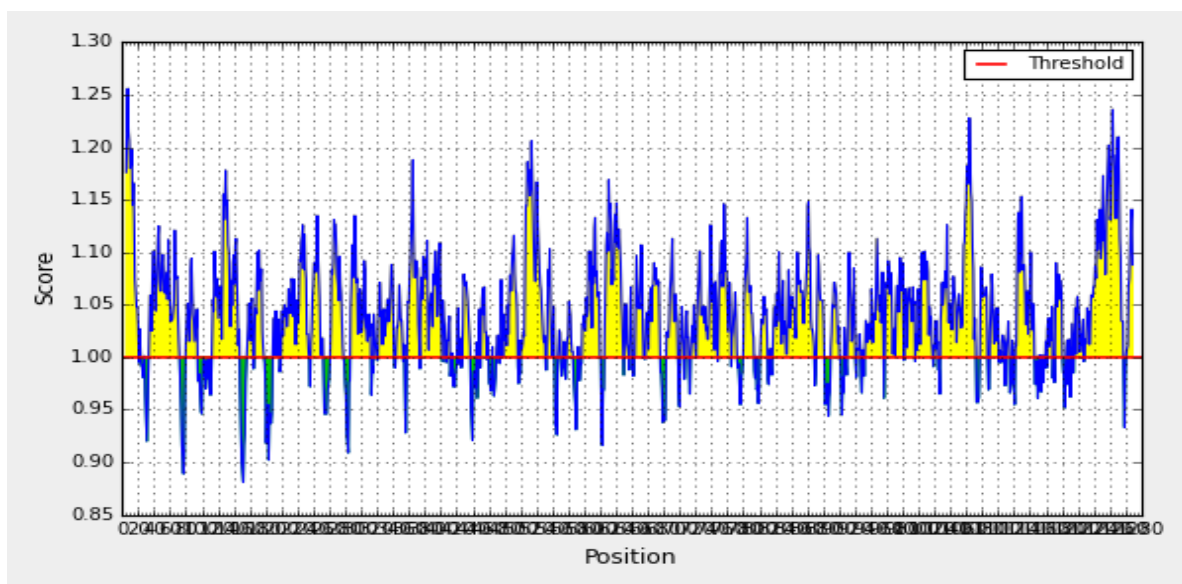


Figure 5: Graph of Kolaskar and Tongaonkar antigenicity prediction method.

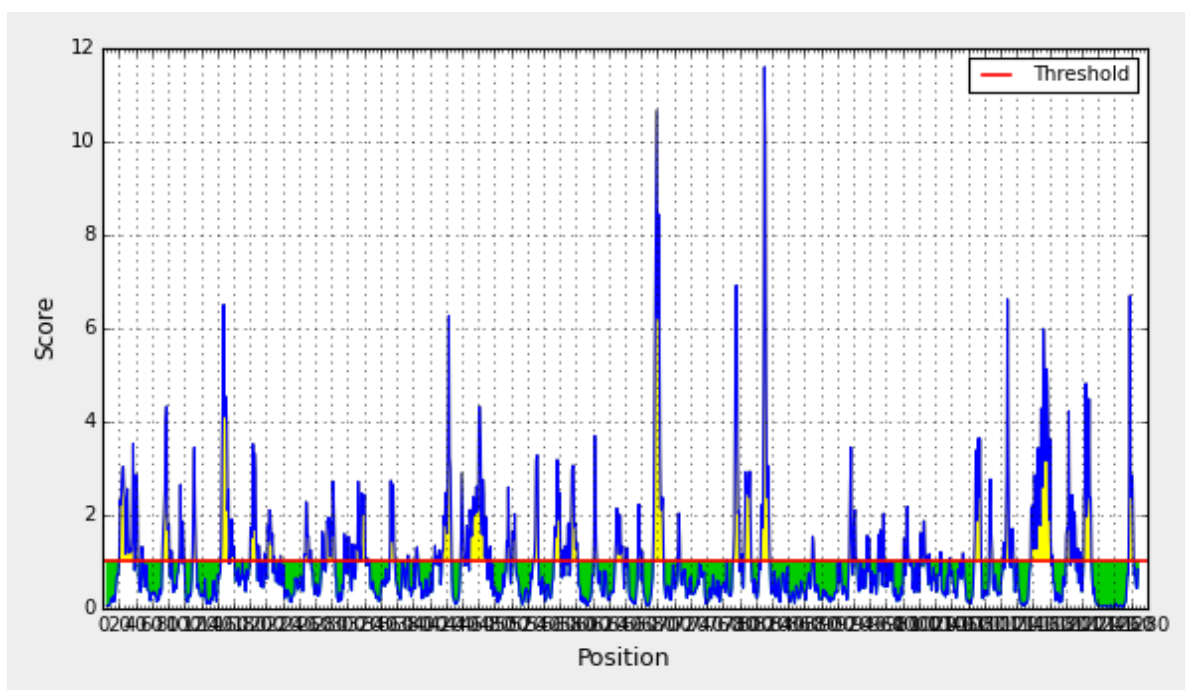


Figure 6: Graph obtained through Emini surface accessibility prediction method

Chou and Fasman beta turn prediction was done to analyze beta regions of the protein because beta regions are often accessible and considerably hydrophilic in nature. Both these properties are antigenic regions of a protein. The region from 249 to 259 was predicted to be beta-turn regions. Figure7 depicts the graph of score (Y-axis) vs position (X-axis). Table 6 lists down the top 5 epitopes on the basis of beta turn prediction.

Table 6: list of top predicted peptides through Chou and Fasman beta turn prediction

Position	Residue	Start	End	Peptide	Score
254	S	250	258	TPGDSSSGW	1.368
255	S	251	259	PGDSSSGWT	1.368
253	D	249	257	LTPGDSSSG	1.327
1250	C	1246	1254	GCCSCGCC	1.326
811	K	807	815	PDPSKPSKR	1.317

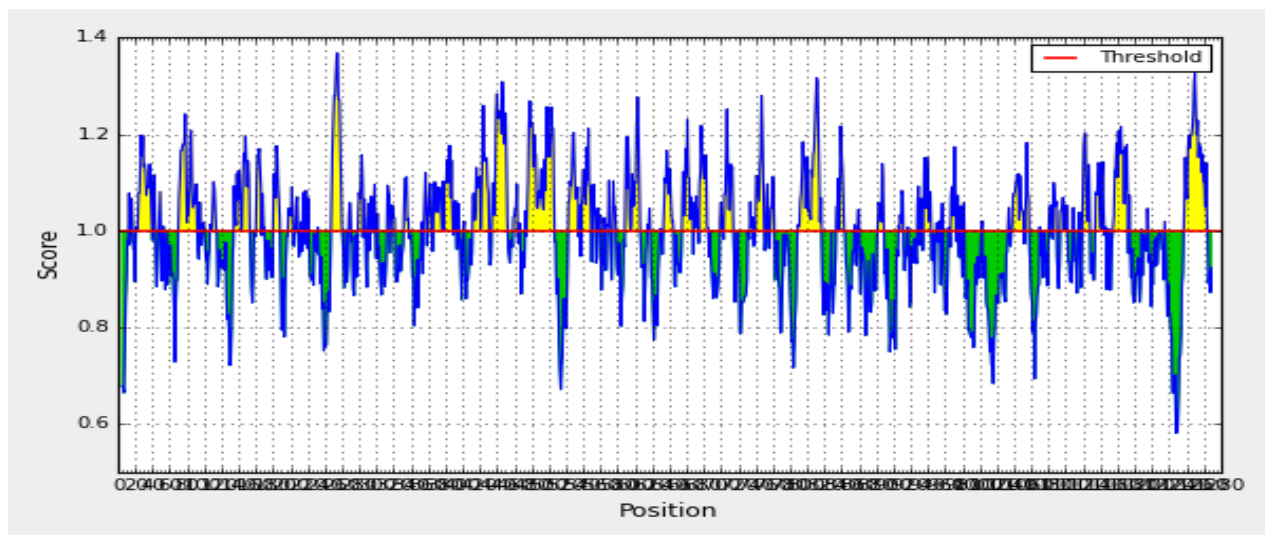


Figure7: Graph obtained from Chou and Fasman beta turn prediction method

Since, flexibility of the peptide is correlated to antigenicity, therefore, Karplus and Schulz flexibility prediction method was used to determine the flexibility of the peptide. The region from 250 to 258 was found to be most flexible, threshold being set at 1.000. Figure8 represents the graph obtained from Karplus and Schulz prediction method wherein Y-axis represents the score and X-axis represents sequence length. Table 7 lists down top five epitopes which were predicted to be more flexible.

Table 7: List of top epitopes predicted through Karplus and Schulz flexibility prediction method

Position	Residue	Start	End	Peptide	Score
254	S	251	257	PGDSSSG	1.125
255	S	252	258	GDSSSGW	1.119
414	Q	411	417	APGQTGK	1.112
679	N	676	682	TQTNSPR	1.109
253	D	250	256	TPGDSSS	1.108

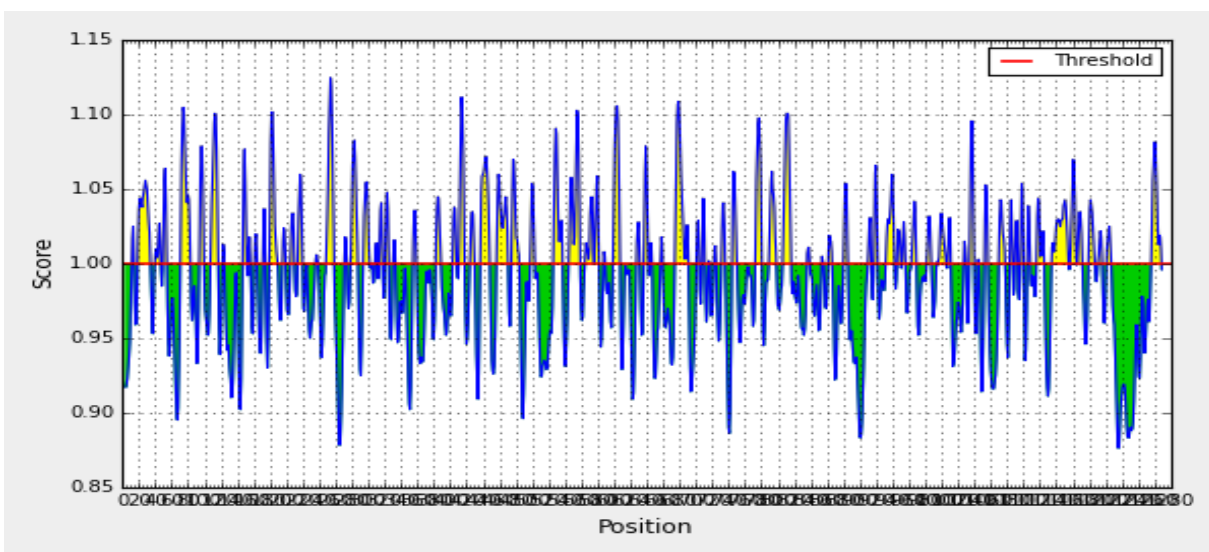


Figure 8: Graph obtained from Karplus and Schulz flexibility prediction method

Finally, the last parameter taken into consideration for prediction of B-cell epitope was Bepipred linear epitope prediction tool 2.0. The tool is based on Hidden Markov Model and is one of the best methods for prediction of linear B-cell epitopes. Method predicted total 33 peptides that had score above the threshold value. Table 8 lists down the B-cell epitopes predicted by Bepipred linear epitope prediction tool 2.0. Figure 10 depicts the graph obtained by the prediction with Score at Y-axis and sequence position at X-axis, the threshold being set at 0.500.

Table 8: List of B-cell epitopes obtained from Bepipred linear epitope prediction tool 2.0

No.	Start	End	Peptide	Length
1	13	36	SQCVNLTTRTQLPPANTNSFTRGV	24
2	59	81	FSNVTWFHAIHVSGTNGTKRFDN	23
3	93	99	ASTEKSN	7
4	138	155	DPFLGVYYHKNNKSWMES	18
5	177	189	MDLEGKQGNFKNL	13
6	206	221	KHTPINLVRDLPQGFS	16
7	250	260	TPGDSSSGWTA	11
8	293	296	LDPL	4
9	304	322	KSFTVEKGIYQTSNFRVQP	19
10	329	363	FPNITNLCPFGEVFNATRFASVYAWNKRISNCVA	35
11	369	393	YNSASFSTFKCYGVSPTKLNDLCFT	25
12	404	426	GDEVQRQIAPGQTGKIADYNYKLP	23
13	440	500	NLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIQAGST PCNGVEGFNCYFPLQSYGFQPT	61
14	516	536	ELLHAPATVCGPKKSTNLVKN	21
15	555	562	SNKKFLPF	8
16	580	583	QTLE	4
17	602	606	TNTSN	5
18	616	632	NCTEVPVAIHADQLTPT	17

19	634	644	RVYSTGSNVFQ	11
20	656	666	VNNSYECDIPI	11
21	672	690	ASYQTQTNSPRRARSVASQ	19
22	694	710	AYTMSLGAENSVAYSNN	17
23	748	748	E	1
24	773	779	EQDKNTQ	7
25	786	800	KQIYKTPPIKDFGGF	15
26	807	815	PDPSKPSKR	9
27	828	842	LADAGFIKQYGDCLG	15
28	987	993	VEAEVQI	7
29	1035	1043	GQSKRVDFC	9
30	1107	1118	RNFYEPQIITTD	12
31	1133	1172	VNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDIG I	40
32	1203	1206	LGKY	4
33	1252	1269	SCCKFDEDDSEPVLKGVK	18

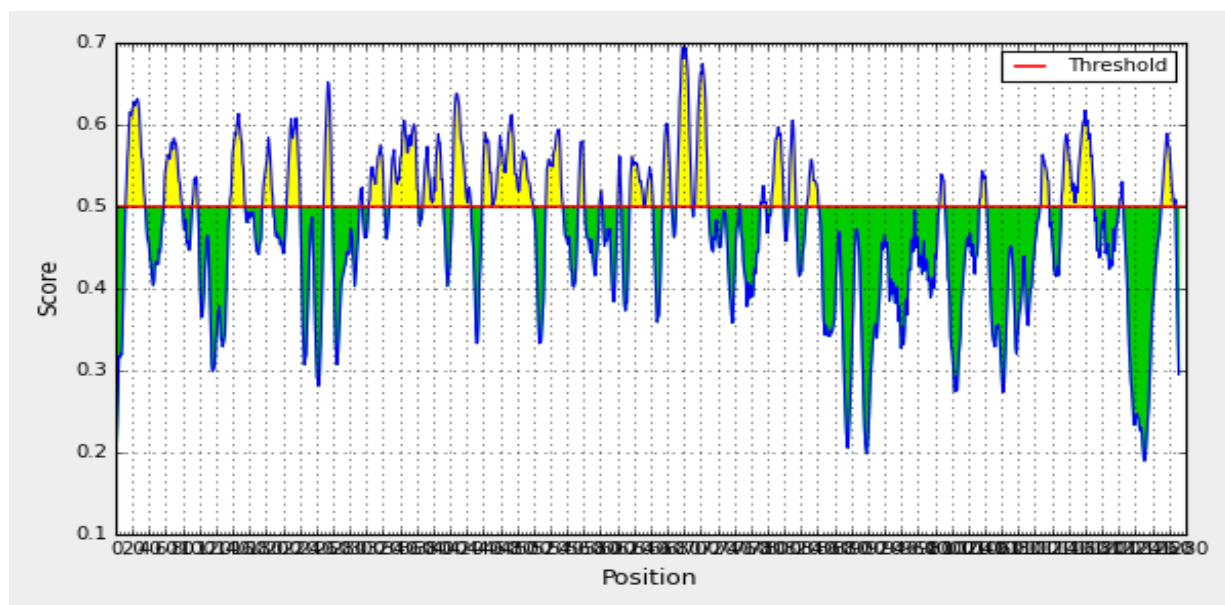


Figure 10: Graph obtained from BepiPred linear epitope prediction tool 2.0

4. Discussion

Covid 19 emerged out to be severe health threat to world population. Currently there is no effective cure measurements are available. There is no widely used vaccines are available for the deadly virus despite so many attempt by scientific community around the world. Hence, there is a need of more innovative approaches for better diagnosis and treatment options to stop the development and progression of the disease. The current *in-silico* study was attempted to identify peptide fragments from spike surface glycoprotein that can be efficiently used for the designing and development of epitope-based vaccine using integrated computational tools for vaccine development. Interpreting the results obtained using different *in silico* tools and databases demonstrate that the epitope sequence WTAGAAYY binds to more number of MHC alleles and is not allergic to human beings. Thus, WTAGAAYY epitope sequence can be used as a T cell epitope for designing of peptide vaccine against SARS-CoV-2. Moreover, the results obtained from IEDB suggest that the protein sequence from 1138-1145, that is, YDPLQPEL had allergenicity, surface accessibility and linear epitope prediction above the threshold value. Also, beta turn prediction score of the peptide sequence was accessed to be 1.136 and the flexibility of the peptide was obtained to be 1.029. The peptide sequence, YDPLQPEL, follows all the essential parameters of a B-cell epitope. There were no clear evidences found for the sequence to exhibit allergenicity effects on the host body. Therefore, the present study concludes that the epitope sequences WTAGAAYY and YDPLQPEL can be used as possible candidates for epitope-based peptide vaccine designing for COVID-19.

References

1. Dahai Zhao, F. Y. (2020). A comparative study on the clinical features of COVID-19 pneumonia to other pneumonias. *Clinical Infectious Diseases*.
2. Zhu N, Z. D. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*.
3. Li Q, G. X. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*.
4. Wu JT KL, L. G. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*, 395.
5. Arafat Rahman Oany, A.-A.-E. T. (2014). Design of an epitope based peptide vaccine against spike protein of human coronavirus: an in silico approach. *Drug Des Devel Ther*.
6. Weidang Li, M. D. (2014). Peptide Vaccine: Progress and Challenges. *Vaccines* 2 (3).
7. McComb RC, H. C. (2015). Presentation of peptides from *Bacillus anthracis* protective antigen on tobacco mosaic virus as an epitope targeted anthrax vaccine. *Vaccine*.
8. Garg N, S. R. (2016). Immunoprotective potential of in silico predicted *Acinetobacter baumannii* outer membrane nuclease, NucAb. *Int J Med Microbiol*.
9. Li, F. (2016). Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* 3.
10. Simmons, G. Z. (2013). Proteolytic activation of the SARS-coronavirus spike protein: cutting enzymes at the cutting edge of antiviral research. *Antiviral Res.* 100.
11. Bertram, S. e. (2011). Cleavage and activation of the severe acute respiratory syndrome coronavirus spike protein by human airway trypsin-like protease. *J. Virol.* 85.
12. Irini A Doytchinova, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 8.
13. Larsen MV, L. C. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*. .
14. Peters B, S. A. (2005). Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* .
15. Sebastian Maurer-Stroh, N. L. (2019). AllerCatPro—prediction of protein allergenicity potential from the protein sequence . *Bioinformatics* 35.
16. Lamiabile A, T. P. (2016). PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res*.

17. Andrew Waterhouse, M. B. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*.
18. Garrett M. Morris, R. H. (2010). AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem*.
19. Marco A. C. Neves, M. T. (2012). Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J Comput Aided Mol Des*.
20. Muhammad Tahir ul Qamar, A. B. (2018). Peptide vaccine against chikungunya virus: immuno-informatics combined with molecular docking approach. *J Transl Med*.
21. Jespersen MC, P. B. (2017). BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res*.
22. Oktavia Rahayu Adianingsih, V. D. (2019). Study of B cell epitope conserved region of the Zika virus envelope glycoprotein to develop multi-strain vaccine. *Journal of Applied Pharmaceutical Science* Vol. 9(01).