Identification of potential vaccine candidates against *SARS-CoV-2*,

A step forward to fight novel coronavirus 2019-nCoV: A Reverse Vaccinology Approach

Ekta Gupta[1], Rupesh Kumar Mishra[2], Ravi Ranjan Kumar Niraj[2#]

1. Dr. B. Lal Institute of Biotechnology, Jaipur (INDIA) - 302017

2. Amity Institute of Biotechnology, Amity University Rajasthan, Jaipur (INDIA)-303002

 # Correspondence Author:

Dr. Ravi Ranjan Kumar Niraj

Assistant Professor

Amity Institute of Biotechnology,

Amity University Rajasthan, Jaipur (INDIA)-303002

Email: rrkniraj@gmail.com

ORCID: 0000-0002-8547-6037

Mobile: +91-9729559580

Abstract: The recent Coronavirus Disease 2019 (COVID-19) causes an immense health crisis to global public health. The World Health Organization (WHO) declared the COVID-19 as a pandemic. The COVID-19 is the etiologic agent of a recently arose disease caused by the severe acute respiratory syndrome coronavirus 2 (*SARS-CoV-2*). Presently, there is no vaccine available against this emerged viral disease. Therefore, it is indeed a need of the hour to develop an effectual and safe vaccine against this decidedly pandemic disease. In the current study, we collected *SARS-CoV-2* genome from Indian geographical origin against human host, further more using reverse vaccinology and immunoinformatics tools here we claim effective vaccine candidates that can be mile stone in battle against COVID19. This novel study divulged two promising antigenic peptide GVYFASTEK and NFRVQPTESIV from surface glycoproteins (protein accession no. - QIA98583.1 and QHS34546.1) of *SARS-CoV-2*, which were predicated to be interacted with class I and class II MHC alleles and showed up to 90% conservancy and high value of antigenicity. Subsequently, the molecular docking studies were verified molecular interaction of these prime antigenic peptides with the residues of HLA-A*11-01 allele for MHC Class I and HLA DRB1*04-01 allele for MHC class II. After vigorous analysis, these peptides were predicted to be suitable epitopes which are capable to elicit the strong cell-mediated immune response against the *SARS-CoV-2*. Consequences from the present study could facilitate selecting *SARS-CoV-2* epitopes for vaccine production pipelines in the immediate future. This novel research will certainly pave the way for a fast, reliable and virtuous platform to provide timely countermeasure of this dangerous pandemic disease, COVID-19.

Keywords: COVID-19, *SARS-CoV-2*, Immunoinformatics, Reverse vaccinology, Molecular docking, Epitope, Vaccine candidates.

1. Introduction

The Coronavirus Disease 2019 (COVID-19) begin in December 2019, like a viral outbreak in Wuhan city of China[1]. It gained rapid foothold across the world resulting in WHO declared it as a pandemic[2]. As on April 5th, worldwide total 1,205,825 cases and 64,978 deaths were reported by WHO[3]. In this period of time when the continuous transmission of the virus across borders and health burden on the global scale is rapidly increasing, more urgent studies are required and in the absence of effective cures majorly drugs, vaccination or immunization therapy is imperative in order to target whole population. Likewise to move forward vaccine development pipeline, immunoinfomatics tools have been proved crucial[4]. Since the Covid19 has affected almost all of the world's population, promiscuous epitopes binding to a variety of HLA alleles for larger dissemination is vital. For that, *in silico* approaches will be remarkably useful in helping develop a cure in as fast manner as possible[5]. The antibody generation by activation of B-cell as well as acute viral clearance by T-cells along with virus-specific memory generation by CD8+ T-cells are analogously important to develop immunity against the virus[6]. The S protein is considered highly antigenic and thereby can evoke strong immune responses and generate neutralizing antibodies that can block the attachment of virus to the host cells[7].

In reverse vaccinology, various tools of *in silico* biology are used to discover the novel antigens by studying the genetic makeup of a pathogen and the genes that could lead to good epitopes are determined. This method is a quick easy and cost-effective way to design vaccine[8]. It is a process of vaccine development where the novel antigens are identified by analyzing the genomic information of a virus or other organism. Reverse vaccinology approach facilitates an easier and productive process of antigen discovery[9].

Herein, we explored the proteome of SARS-CoV-2 of Indian geographical origin against human host to identify potential antigenic proteins and epitopes that can effectively elicit cellular mediated immune response against COVID-19, To do so, we collected *SARS-CoV2* genome from Indian geographical origin against human host and applied *in-silico* approach. This significant research disclosed two promising antigenic peptide GVYFASTEK and NFRVQPTESIV from surface glycoproteins (protein accession no. - QIA98583.1 and QHS34546.1) of *SARS-CoV-2*, they were predicted to be interacted with class I and class II MHC alleles and displayed up to 90% conservancy and significant antigenicity. Further, molecular docking analysis were confirmed the molecular interaction of such prime antigenic peptides with the residues of HLA-A*11-01 allele for MHC Class I and HLA DRB1*04-01 allele for MHC class II. After careful evaluation, these peptides were foreseen to be appropriate epitopes proficient to

elicit the strong cell-mediated immune response against the *SARS-CoV-2*. The reverse vaccinology approach facilitates an easier and productive process of vaccine candidate's determination. The outcomes from this very significant analysis could help selecting *SARS-CoV-2* epitopes for vaccine production pipelines soon. This novel research will certainly pave the way for a fast, reliable and virtuous platform to provide timely countermeasure of this dangerous pandemic disease, COVID-19.

2. Materials and Methods

2.1 Strain selection

The highly virulent strain *SARS-CoV-2* was chosen for *in-silico* analysis. The genome of viral strain is available on the National Center for Biotechnology Information or NCBI (https://www.ncbi.nlm.nih.gov/) with *RefSeq NC_045512.2*

2.2 Protein Identification and Retrieval

Twelve viral protein sequences of *SARS-CoV2* against (Host: Human, Country: India) were retrieved from ViPR database[10].

2.3 Physicochemical Property Analysis

The various physicochemical properties of the protein sequences were determined by ExPASy's online tool ProtParam[11].

2.4 Protein Antigenicity

VaxiJen v2.0 **[12]** was utilized to predict antigenicity of proteins. This software requires FASTA-submitted amino acid sequences and uses the physicochemical properties of proteins to predict their antigenicity. This feature is denoted according to an antigenic score respectively **[13]**.

2.5 B-cell and T-cell Epitope Prediction

The B-cell and T-cell epitopes of the selected two surface glycoprotein sequences were predicted via IEDB (The Immune Epitope Database). The IEDB database holds large amount of experimental data on epitopes and antibodies. It allows robust analysis on many epitopes in the context of some tools like: conservation across antigens, population coverage, and clusters with similar sequences [14]. In order to obtain MHC class-I restricted CD8+ cytotoxic T-lymphocyte (CTL) epitopes of the selected surface glycoprotein sequences, NetMHCpan EL 4.0 prediction method was applied for HLA-A*11-01 allele. For MHC class-II restricted CD4+ helper T-lymphocyte (HTL) epitopes were obtained for HLA DRB1*04-01 allele using Sturniolo prediction method. Top ten MHC class-I and MHC class-II epitopes were randomly selected on the basis of their percentile scores and antigenicity scores (AS). Five random B-cell lymphocyte epitopes (BCL) were selected based on of their higher length using Bipipered linear epitope prediction method[8].

2.6 Antigenicity, Allergenicity of the predicted epitopes

VaxiJen v2.0 was utilized to predict protein antigenicity. This software requires FASTA-submitted amino acid sequences and uses the physicochemical properties of proteins to predict their antigenicity. This feature is denoted according to an antigenic score respectively[12]. The allergenicity of the selected epitopes was predicted via AllerTOP v2 (https://www.ddg-pharmfac.net/AllerTOP/).

2.7 Transmembrane Topology and Toxicity Prediction of the predicated epitopes

The transmembrane topology of the selected epitopes was identified using the transmembrane topology of protein helices determinant, TMHMM v2.0 server (http://www.cbs.dtu.dk/services/TMHMM/). The server predicts whether the epitope would be transmembrane, remain inside or outside of the membrane. The toxicity prediction of the selected epitopes was carried out via ToxinPred server https://webs.iiitd.edu.in/raghava/toxinpred/protein.php.

2.8 Prediction of Conservancy for the Selected Epitopes

The conservancy analysis of the earlier selected epitopes was performed via the epitope conservancy analysis tool of IEDB server [14]. During analysis, the sequence identity threshold was kept at '>=50'.

2.9 The MHC Alleles; Cluster analysis

Cluster analysis was carried out by MHCcluster 2.0 [15]. During cluster analysis, the number of peptides to be included was kept at 50,000, the number of bootstrap calculations were kept at 100. For cluster analysis, the NetMHCpan-2.8 prediction method was used.

2.9. Generation of the 3D Structures of the Selected Epitopes

The 3D structures of the selected best epitopes were generated using online 3D generating tool PEP-FOLD3 (http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/). The server is a tool for generating de novo peptide 3-dimensional structure [16-18].

2.10. Molecular Docking of the Selected Epitopes

Pre-docking was carried out by UCSF Chimera [19]. The peptide-protein docking of the selected epitopes was performed by online docking tool PatchDock (https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php) results of PatchDock were refined and re-scored by FireDock server (http://bioinfo3d.cs.tau.ac.il/FireDock/php.php). Later on, docking was performed by HPEPDOCK server[20]. Docking pose analysis was done by using Ligplot[21].

3. Results

3.1 Selection and Retrieval of Viral Protein Sequences

The *SARS-CoV-2*, strain was identified. Twelve viral protein sequences of *SARS-CoV2* against (Host: Human, Country: India) were retrieved from ViPR database and selected for the possible vaccine candidate identification **(Table-1)**. These proteins were: Orf10 protein **QIA98591.1**, Orf8 protein **QIA98589.1**, Orf7a protein **QIA98588.1**, Orf6 protein **QIA98587.1**, Orf3a protein **QIA98584.1**, Membrane glycoprotein **QIA98586.1**, Envelope protein **QIA98585.1**, Surface glycoprotein **QIA98583.1**, Surface glycoprotein **QHS34546.1**, Nucleocapsid protein **QII87776.1**, Nucleocapsid protein **QII87775.1** and Nucleocapsid phosphoprotein **QIA98590.1**. The FASTA sequence of proteins mentioned in **(Additional File:1)**

3.2 Physicochemical Property Analysis and Protein Antigenicity

In the physicochemical property analysis, the number of amino acids, the molecular weights, theoretical pI, extinction coefficients (in M-1 cm-1), Est. half-life (in mammalian cell), instability indexes, aliphatic indexes and

grand average of hydropathicity (GRAVY) of the twelve proteins were predicted **(Table-2)**. For antigenicity prediction threshold value kept at 0.4, all proteins were found to be antigenic **(Table-3)**. The physicochemical study revealed that the surface glycoproteins **QIA98583.1** and **QHS34546.1** had the highest extinction co-efficient of 148960M-1 cm-1 and 147470 M-1 cm-1 and lowest GRAVY value of -0.077 and -0.071. In addition, both surface glycoproteins were stable and antigenic. We selected both surface glycoproteins for further analysis.

3.3 T-cell and B-cell Epitope Prediction

The T-cell epitopes of MHC class-I for both of the proteins were determined by NetMHCpan EL 4.0 prediction method of the IEDB server keeping the sequence length 9. The server generated epitopes further analyzed on the basis of the antigenicity scores (AS) and percentile scores, top ten potential epitopes were selected randomly for antigenicity, allergenicity, toxicity and conservancy tests. The server ranks the predicted epitopes based on the ascending order of percentile scores **(Table-4a & 4b)**. The T-cell epitopes of MHC class-II (HLA DRB1*04-01 allele) of the proteins were also determined by IEDB server **(Table-5a & 5b)**, where the Sturniolo prediction methods was used. For each protein, ten of the top epitopes were selected randomly for further analysis. Moreover, the B-cell epitopes of the proteins were selected using Bipipered linear epitope prediction method of the IEDB server and epitopes were selected based on their higher lengths **(Fig-1).**

3.4 Topology Identification of the Epitopes

The topology of the selected epitopes was determined by TMHMM v2.0 server (http://www.cbs.dtu.dk/services/TMHMM/). **Table 4a** and **Table 4b** list the potential T-cell epitopes of surface glycoproteins **QIA98583.1**, **Table 5a** and **Table 5b** list the potential T-cell epitopes of surface glycoproteins **QHS34546.1**, and **Table 6** list the potential B-cell epitopes with their respective topologies.

3.4. Antigenicity, allergenicity, toxicity and conservancy analysis of the epitopes

In the antigenicity, allergenicity, toxicity and conservancy analysis, the T-cell epitopes that were found to be highly antigenic as well as non-allergenic, non-toxic, and had conservancy of over 90% were selected. Among the ten selected MHC class-I epitopes and ten selected MHC class-II epitopes of both of the proteins, total four epitopes were selected based on the mentioned criteria: GVYFASTEK, TLADAGFIK, NFRVQPTESI and LLIVNNATNV.

3.5. Cluster Analysis of the MHC Alleles

The cluster analysis of the possible MHC class-I alleles that may interact with the predicted epitopes were performed by online tool MHCcluster 2.0 (http://www.cbs.dtu.dk/services/MHCcluster/). The tool generates the clusters of the alleles in phylogenetic manner. Results illustrate the outcome of the experiment where the red zone indicates strong interaction and the yellow zone corresponds to weaker interaction **(Fig-2)**.

3.6. Generation of the 3D Structures of the Epitopes

All the T-cell epitopes were subjected to 3D structure generation by the PEP-FOLD3 server. The 3D structures were generated for peptide-protein docking **(Fig-3)**.

3.7 Peptide-Protein Docking & Vaccine candidate's prioritization

The docking was performed to find out, whether all the epitopes had the ability to bind with the MHC class-I and MHC class-II molecule. The selected epitopes were docked against the HLA-A*11-01 allele (PDB ID: 5WJL) and HLA DRB1*04-01 (PDB ID: 5JLZ). The docking was performed using PatchDock online docking tool and then the results were refined by FireDock online server. Results were also analysed by HPEPDOCK server **(Additional File: Fig-1)**. Among the Four epitopes of two selected glycoproteins QIA98583.1and QHS34546.1, GVYFASTEK (MHC class I epitope) and NFRVQPTESI (MHC class II epitope) showed the best result with the lowest global energy of -52.82 and -42.00. Further, docking pose was analyzed via Ligplot (**Fig 4).**

We identified highly antigenic and non-allergenic B-cell vaccine candidates LTPGDSSSGWTAG and VRQIAPGQTGKIAD from Surface glycoprotein (QIA98583.1) and QIAPGQTGKIAD and ILPDPSKPSKRS from Surface glycoprotein (QHS34546.1).

4. Discussion

Vaccine is one of the most important and extensively produced pharmaceutical products. Millions of infants and people are getting vaccinated every year. However, the development and research processes of vaccines are costly and sometimes, it takes many years to develop an appropriate vaccine candidate against a pathogen. In modern

times, various tools and methods of immune-informatics, bioinformatics, and reverse vaccinology are extensively used for vaccine development, which save time and cost of the vaccine development process[8, 22].

In current study, the physicochemical study revealed that surface glycoproteins **QIA98583.1** and **QHS34546.1** had the highest extinction co-efficient of 148960M-1 cm-1 and 147470 M-1 cm-1and lowest GRAVY value of -0.077 and -0.071. In addition, both surface glycoproteins were highly stable (instability index of less than 40) and antigenic. Antigenicity of the proteins was determined by VaxiJen V2.0 server. If a compound had instability index of more than 40, then the compound is referred to be unstable [23]. The extinction coefficient means the amount of light, that is absorbed by a compound at a certain wavelength[24, 25]. The various physicochemical properties like number of amino acids, molecular weight, theoretical pI, extinction co-efficient, instability index, aliphatic index, GRAVY was determined by ProtParam (https://web.expasy.org/protparam/) server. The two selected proteins performed quite similarly in the physicochemical property assessment.

In immune system, two of the major cells that functions are the T lymphocytic cell and B lymphocytic cell. After recognized by an antigen presenting cell or APC (like macrophage, dendritic cell etc.), the antigen is presented by the MHC class-II molecule present on the surface of these standard APCs, to the helper T cell. Since, the helper T cell contains CD4+ molecule on its surface, it is also called as CD4+ T cell. After activated by APC, the T-helper cell then activates the B cell and causes the production of antibody producing plasma B cell along with memory B cell. The plasma B cell produces a large number of antibodies and the memory B cell functions as the immunological, long term memory. However, macrophage and CD8+ cytotoxic T cell are also activated by the T-helper cell that destroys the target antigen [26-30].

The possible B cell and T cell epitopes of the selected *SARS-CoV-2* proteins were identified by the IEDB (https://www.iedb.org/) server. The IEDB server generates and ranks the epitopes on the basis of their antigenicity scores (AS) and percentile scores. The top ten MHC class-I and MHC class-II epitopes were taken for analysis. The topology of the selected epitopes was determined by TMHMM v2.0 server (http://www.cbs.dtu.dk/services/TMHMM/). In the antigenicity, allergenicity, toxicity and conservancy analysis, the T-cell epitopes that were found to be highly antigenic as well as non-allergenic, non-toxic, and had conservancy of over 90% . Among the ten selected MHC class-I epitopes and ten selected MHC class-II epitopes of both proteins, total four epitopes were selected based on the mentioned criteria: GVYFASTEK, TLADAGFIK, NFRVQPTESI and LLIVNNATNVV as well as antigenic and non-allergenic B-cell epitopes were selected for further vaccine candidate

analysis. The cluster analysis of the possible MHC class-I and MHC class-II alleles that may interact with the predicted epitopes were performed by online tool MHC cluster 2.0 (http://www.cbs.dtu.dk/services/MHCcluster/). Antigenicity, defined as the ability of a foreign substance to act as antigen and activate the B cell and T cell responses, through their epitope also called antigenic determinant portion [31]. The allergenicity is defined as the ability of that substance to act as allergen and induce potential allergic reactions within the host body [32].

Moreover, the cluster analysis of the MHC class-I alleles and MHC class-II alleles were also carried out to identify their relationship with each other and cluster them functionally on the basis of their predicted binding specificity[15]. In the next step, the peptide-protein docking was carried out between the selected epitopes and the MHC alleles. The MHC class-I epitopes were docked with the MHC class-I molecule (PDB ID: 5WJL) and the MHC class-II epitopes were docked with the MHC class-II molecule (PDB ID: 5JLZ) respectively. The peptide-protein docking was performed to analyze the ability of the epitopes to bind with their respective MHC molecule. Pre-docking was performed by UCSF-chimera and later on, we performed 3D structure generation of the epitopes. The docking was carried out by PatchDock and FireDock servers also analysed by HPEPDOCK server based on global energy. GVYFASTEK, and NFRVQPTESI generated the best scores in the peptide-protein docking respectively. All the vaccine candidates were proved to be potentially antigenic and non-allergenic, for this reason they should not cause any allergenic reaction within the host body. However, more *in vitro* and *in vivo* researches should do to finally confirm the safety, efficacy and potentiality of the predicted vaccines candidates.

5. Conclusion

In the face of the enormous tragedy of suffering, death and social disaster caused by COVID-19 pandemic. It is of utmost importance to develop an effectual and safe vaccine against this highly pandemic disease. Bioinformatics, Reverse vaccinology and related technologies are widely used in vaccine design and development since these technologies reduce the cost and time. In this study, first the potential proteins belong to *SARS-CoV-2*against (host: human, country: India) are identified. Further, the potential B cell and T cell epitopes that can effectively elicit cellular mediated immune response related to these selected proteins were determined through robust processes. These potential T-cell epitopes GVYFASTEK, NFRVQPTESI and B-cell epitopes LTPGDSSSGWTAG, VRQIAPGQTGKIAD, QIAPGQTGKIAD and ILPDPSKPSKRS play vital role in subunit and multi-epitope vaccine construction in the near future. To conclude, reverse vaccinology justified as a powerful tool for identifying

new vaccine candidates and their consequent precise application. This study will lead the research in a new and effectual direction and the outcome of our study will provide a fast, reliable and significant platform in search of effective and timely cure of this dangerous pandemic disease, COVID-19 caused by *SARS-CoV2*.

Acknowledgement: RKM acknowledges the financial support and award of Ramalingaswami fellowship from Department of Biotechnology, New Delhi, India. Authors highly acknowledge the Amity Institute of Biotechnology, Amity University Rajasthan, Jaipur and Dr. B. Lal Institute of Biotechnology, Jaipur.

Funding:

This research did not receive any specific grant from funding agencies.

Compliance with Ethical Standards:

Conflict of interest

The authors declare there is no conflict of interest.

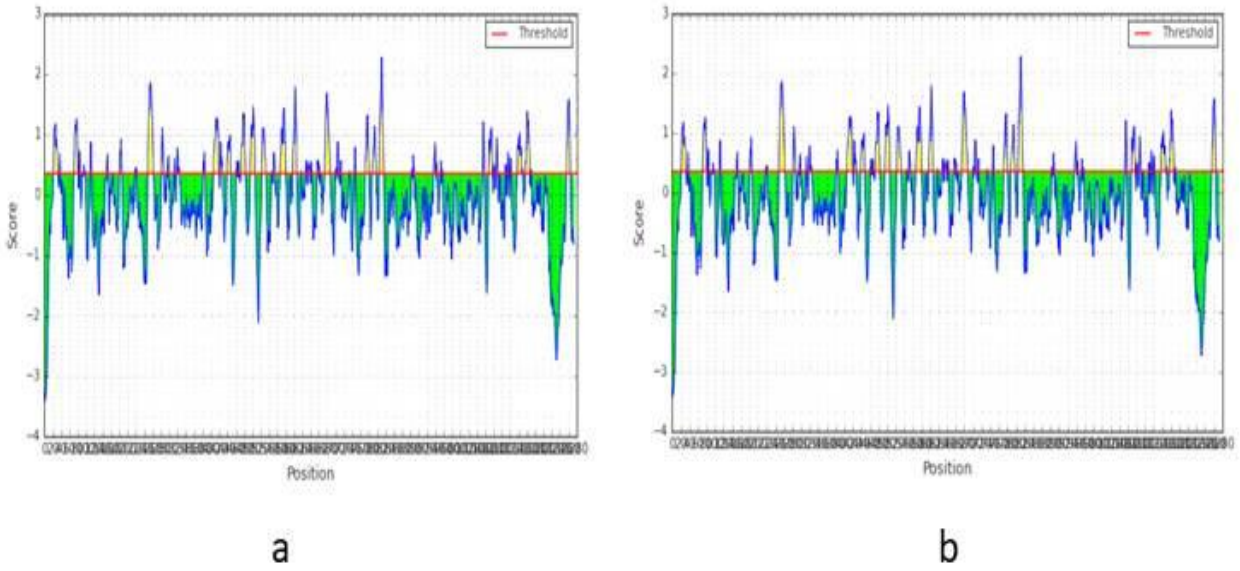
Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

6. References

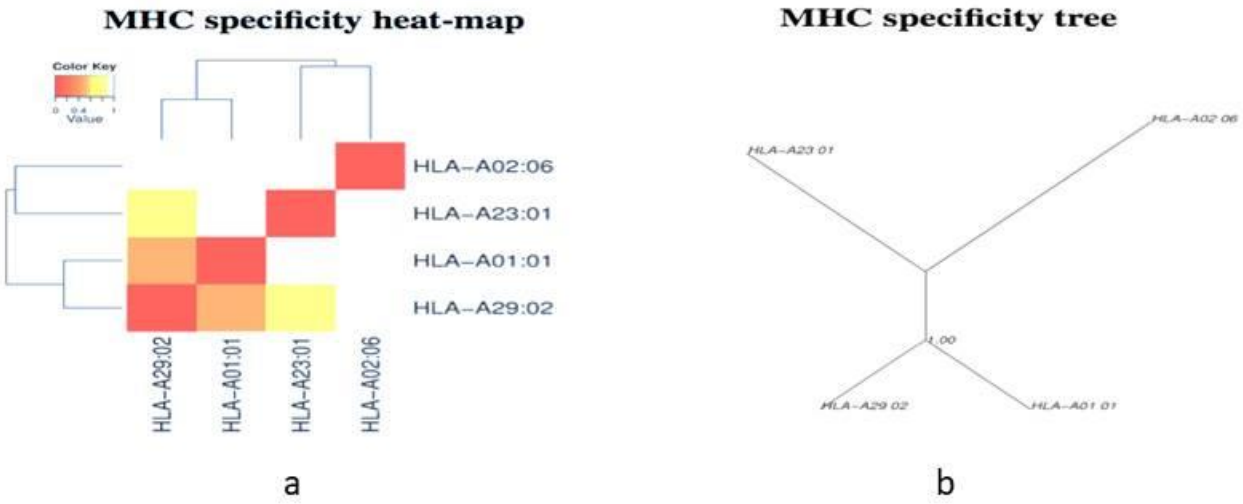
1. Holshue, M.L., et al., *First case of 2019 novel coronavirus in the United States.* New England Journal of Medicine, 2020.
2. https://www.cnbc.com/2020/03/11/who-declares-the-coronavirus-outbreak-a-global-pandemic.html.
3. (https://www.worldometers.info/coronavirus/).
4. Mishra, S. and S. Sinha, *Immunoinformatics and modeling perspective of T cell epitope-based cancer immunotherapy: a holistic picture.* Journal of Biomolecular Structure and Dynamics, 2009. **27**(3): p. 293-305.
5. Mishra, D., *T Cell Epitope-Based Vaccine Design for Pandemic Novel Coronavirus 2019-nCoV*. 2020.
6. Enjuanes, L., et al., *Molecular basis of coronavirus virulence and vaccine development*, in *Advances in virus research*. 2016, Elsevier. p. 245-286.
7. Du, L., et al., *The spike protein of SARS-CoV—a target for vaccine and therapeutic development.* Nature Reviews Microbiology, 2009. **7**(3): p. 226-236.
8. Ullah, M.A., B. Sarkar, and S.S. Islam, *Exploiting the Reverse Vaccinology Approach to Design Novel Subunit Vaccine against Ebola Virus.* medRxiv, 2020.
9. Gupta, E., S.R. Gupta, and R.R.K. Niraj, *Identification of drug and vaccine target in mycobacterium leprae: a reverse vaccinology approach.* International Journal of Peptide Research and Therapeutics, 2019: p. 1-14.
10. Pickett, B.E., et al., *ViPR: an open bioinformatics database and analysis resource for virology research.* Nucleic acids research, 2012. **40**(D1): p. D593-D598.
11. Walker, J.M., *The proteomics protocols handbook*. 2005: Springer.
12. Doytchinova, I.A. and D.R. Flower, *VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines.* BMC bioinformatics, 2007. **8**(1): p. 4.

13. Meunier, M., et al., *Identification of novel vaccine candidates against Campylobacter through reverse vaccinology.* Journal of immunology research, 2016. **2016**.
14. Vita, R., et al., *The immune epitope database (IEDB): 2018 update.* Nucleic acids research, 2019. **47**(D1): p. D339-D343.
15. Thomsen, M., et al., *MHCcluster, a method for functional clustering of MHC molecules.* Immunogenetics, 2013. **65**(9): p. 655-665.
16. Thévenet, P., et al., *PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides.* Nucleic acids research, 2012. **40**(W1): p. W288-W293.
17. Shen, Y., et al., *Improved PEP-FOLD approach for peptide and miniprotein structure prediction.* Journal of chemical theory and computation, 2014. **10**(10): p. 4745-4758.
18. Lamiable, A., et al., *PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex.* Nucleic acids research, 2016. **44**(W1): p. W449-W454.
19. Pettersen, E.F., et al., *UCSF Chimera—a visualization system for exploratory research and analysis.* Journal of computational chemistry, 2004. **25**(13): p. 1605-1612.
20. Zhou, P., et al., *HPEPDOCK: a web server for blind peptide–protein docking based on a hierarchical algorithm.* Nucleic acids research, 2018. **46**(W1): p. W443-W450.
21. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.* Protein engineering, design and selection, 1995. **8**(2): p. 127-134.
22. María, R., et al., *The impact of bioinformatics on vaccine design and development*. 2017: InTech, Rijeka, Croatia.
23. Guruprasad, K., B.B. Reddy, and M.W. Pandit, *Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence.* Protein Engineering, Design and Selection, 1990. **4**(2): p. 155-161.
24. Ikai, A., *Thermostability and aliphatic index of globular proteins.* The Journal of Biochemistry, 1980. **88**(6): p. 1895-1898.
25. Pace, C.N., et al., *How to measure and predict the molar absorption coefficient of a protein.* Protein science, 1995. **4**(11): p. 2411-2423.
26. Goerdt, S. and C.E. Orfanos, *Other functions, other genes: alternative activation of antigen-presenting cells.* Immunity, 1999. **10**(2): p. 137-142.
27. Tanchot, C. and B. Rocha, *CD8 and B cell memory: same strategy, same signals.* nature immunology, 2003. **4**(5): p. 431-432.
28. Pavli, P., et al., *Dendritic cells, the major antigen-presenting cells of the human colonic lamina propria.* Immunology, 1993. **78**(1): p. 132.
29. Arpin, C., et al., *Generation of memory B cells and plasma cells in vitro.* Science, 1995. **268**(5211): p. 720-722.
30. Cano, R.L.E. and H.D.E. Lopera, *Introduction to T and B lymphocytes*, in *Autoimmunity: From Bench to Bedside [Internet]*. 2013, El Rosario University Press.
31. Fishman, J.M., K. Wiles, and K.J. Wood, *The acquired immune system response to biomaterials, including both naturally occurring and synthetic biomaterials*, in *Host Response to Biomaterials*. 2015, Elsevier. p. 151-187.
32. Andreae, D. and A. Nowak-Węgrzyn, *The Effect of Infant Allergen/Immunogen Exposure on Long-Term Health*, in *Early Nutrition and Long-Term Health*. 2017, Elsevier. p. 131-173.



a b

Fig 1: B cell Epitopes prediction (a) epitope prediction for surface glycoproteins QIA98583.1 (b) The graph of epitope prediction for surface glycoproteins QHS34546.1



a b

Fig 2: MHC Class Cluster Analysis (a) Heat map (b) Specificity tree
Note: Here, red zone indicates strong interaction and the yellow zone corresponds to weaker interaction.

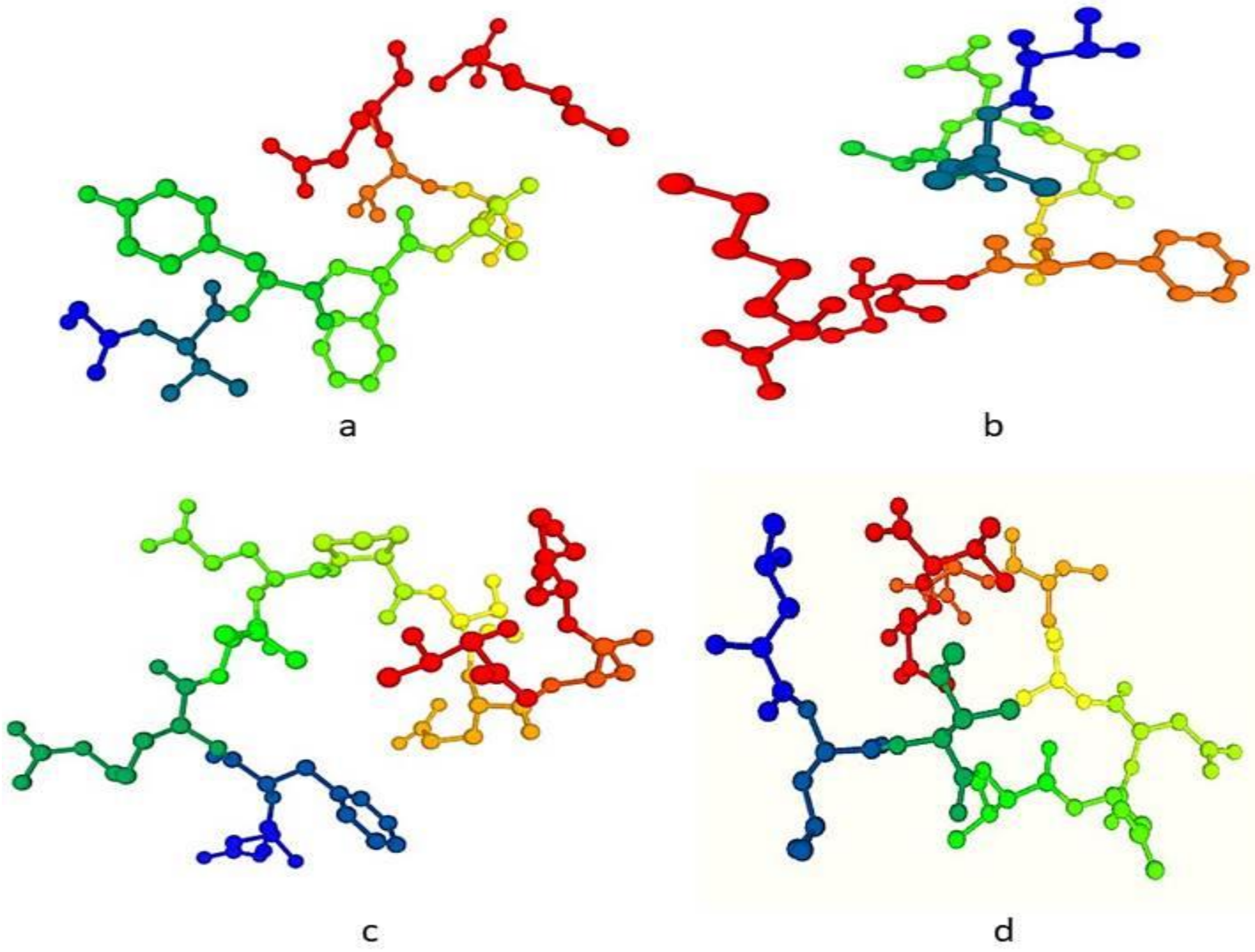


Fig 3: 3D structure generation of T-cell epitopes by the PEP-FOLD3 server. Epitope representation: (a) GVYFASTEK (b) TLADAGFIK (c) NFRVQPTESI and (d) LLIVNNATNV.

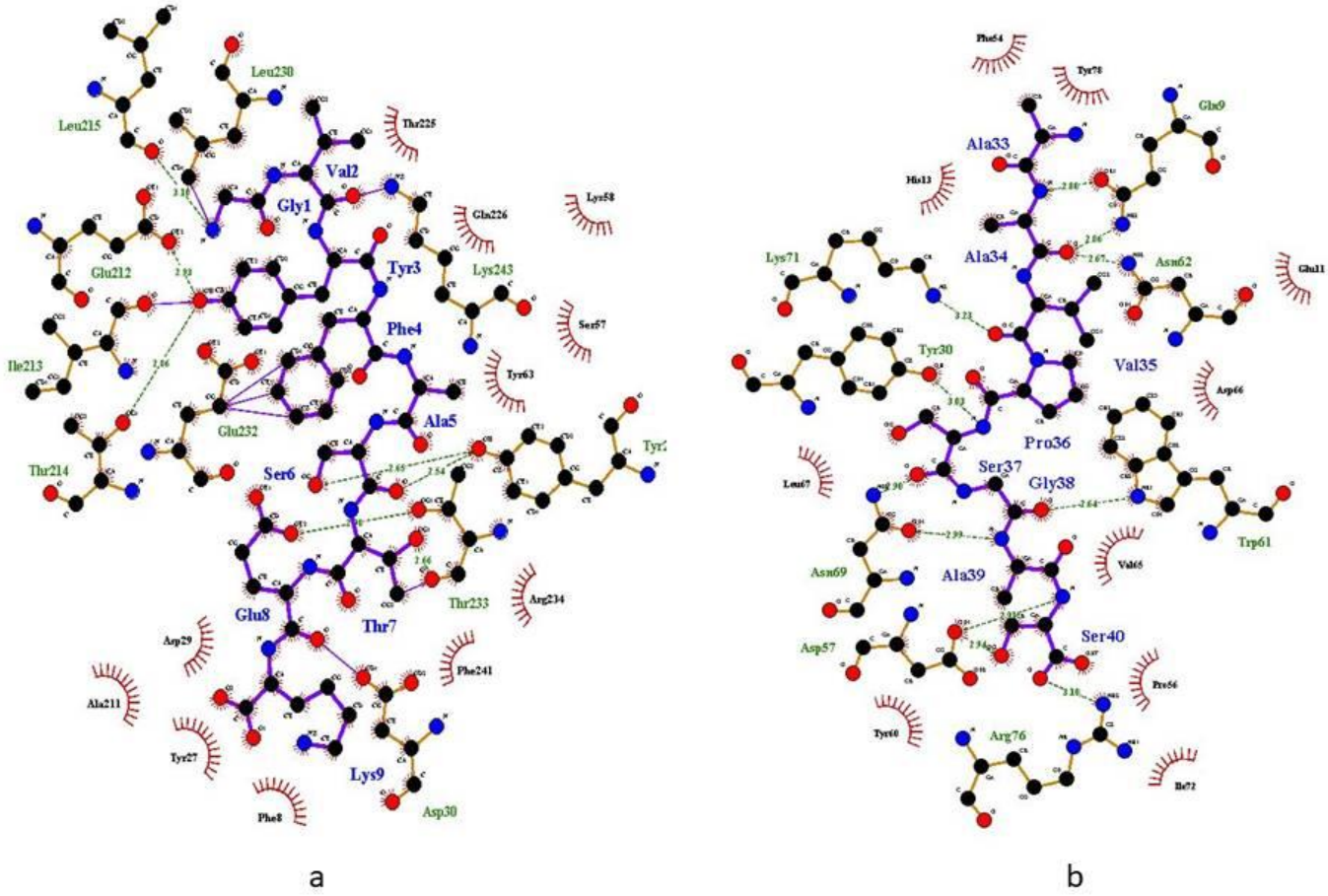


Fig 4: Docking pose analysis via LigPlot [(a: GVYFASTEK epitope docking against the HLA-A*11-01 allele (PDB ID: 5WJL) b: NFRVQPTESI epitope docking against HLA DRB1*04-01 (PDB ID: 5JLZ)]. Note: Molecular Docking result showing protein -ligand interaction; where Oxygen (O), Nitrogen (N) and Carbon (C) atoms are represented in red, blue and black circles

S.No.	*Gene Symbol*	*Protein Name*	*GenBank Accession*	*GenBank Protein Accession*
1	orf10	Orf10 protein	MT050493	QIA98591.1
2	orf8	Orf8 protein	MT050493	QIA98589.1
3	orf7a	Orf7a protein	MT050493	QIA98588.1
4	orf6	Orf6 protein	MT050493	QIA98587.1
5	orf3a	Orf3a protein	MT050493	QIA98584.1
6	M	Membrane glycoprotein	MT050493	QIA98586.1
7	E	Envelope protein	MT050493	QIA98585.1
8	S	Surface glycoprotein	MT050493	QIA98583.1
9	S	Surface glycoprotein	MT012098	QHS34546.1
10	N	Nucleocapsid protein	MT163715	QII87776.1
11	N	Nucleocapsid protein	MT163714	QII87775.1
12	N	Nucleocapsid phosphoprotein	MT050493	QIA98590.1

Table 1: SARS-CoV-2 (Host: Human, Country: India) viral protein sequence identification and retrieval via ViPR database (Viral pathogen database and analysis resource).

	Properties							
Gene Symbol	**No. of amino acids**	**Molecular weight**	**Theoretical pI**	**Ext. coefficient (in M-1 cm-1)**	**Est. half-life (in mammalian cell)**	**Instability index**	**Aliphatic index**	**GRAVY (grand average of hydropathicity)**
orf10	38	4449.23	7.93	4470	30 hours	16.06 (stable)	107.63	0.637
orf8	121	13804.93	5.42	16305	30 hours	46.24 (unstable)	94.13	0.181
orf7a	121	13744.17	8.23	7825	30 hours	48.66 (unstable)	100.74	0.318
orf6	61	7272.54	4.60	8480	30 hours	31.16 (stable)	130.98	0.233
orf3a	275	31122.94	5.55	58705	30 hours	32.96 (stable)	103.42	0.275
M	222	25146.62	9.51	52160	30 hours	39.14 (stable)	120.86	0.446
E	75	8365.04	8.57	6085	30 hours	38.68 (stable)	144.00	1.128
S	1273	141206.52	6.24	148960	30 hours	33.01 (stable)	84.82	-0.077
S	1272	140972.27	6.16	147470	30 hours	32.78 (stable)	85.05	-0.071
N	88	9827.08	10.23	8480	4.4 hours	36.54 (stable)	61.14	-1.067
N	133	14363.88	11.37	8480	1 hours	58.97 (unstable)	44.21	-1.170
N	419	45625.70	10.07	43890	30 hours	55.09 (unstable)	52.53	-0.971

Table 2: Physiochemical property analysis of SARS-CoV-2 against (Host: Human, Country: India) viral proteins.

S. No.	*Protein Name*	*Antigenicity Score*	*Antigenicity; Threshold=0.4*
1	Orf10 protein	0.7185	Antigenic
2	Orf8 protein	0.6063	Antigenic
3	Orf7a protein	0.6441	Antigenic
4	Orf6 protein	0.6131	Antigenic
5	Orf3a protein	0.4945	Antigenic
6	Membrane glycoprotein	0.5102	Antigenic
7	Envelope protein	0.6025	Antigenic
8	Surface glycoprotein	0.4654	Antigenic
9	Surface glycoprotein	0.4687	Antigenic
10	Nucleocapsid protein	0.5767	Antigenic
11	Nucleocapsid protein	0.6235	Antigenic
12	Nucleocapsid phosphoprotein	0.5059	Antigenic

Table 3: Antigenicity predication of SARS-CoV-2 viral proteins (Threshold value: 0.4)

Surface glycoprotein (QIA98583.1)

Epitope	Start	End	Topology	Antigenicity	Score	Allergenicity	Toxicity	Minimum Identity	Conservancy
GVYFASTEK	19	27	Inside	Antigen	0.7112	Non-allergen	Non-toxic	11.11%	100%
VTYVPAQEK	15	23	Inside	Antigen	0.8132	Allergen	Non-toxic	22.22%	100%
ASANLAATK	40	48	Inside	Antigen	0.7041	Allergen	Non-toxic	22.22%	100%
TLADAGFIK	57	65	Inside	Antigen	0.5781	Non-allergen	Non-toxic	22.22%	100%
TLKSFTVEK	22	30	Inside	Non-antigen	0.0809	Allergen	Non-toxic	11.11%	100%
NSASFSTFK	20	28	Inside	Non-antigen	0.1232	Allergen	Non-toxic	11.11%	100%
TEILPVSMTK	24	33	Inside	Antigen	1.4160	Allergen	Non-toxic	10.00%	100%
SSTASALGK	29	37	Outside	Antigen	0.6215	Allergen	Non-toxic	22.22%	100%
GTHWFVTQR	49	57	Inside	Non-antigen	0.0723	Allergen	Non-toxic	11.11%	100%
EILPVSMTK	25	33	Inside	Antigen	1.6842	Allergen	Non-toxic	11.11%	100%

Table 4a: MHC Class-I epitopes of Surface glycoprotein (QIA98583.1): Table represents topology, antigenicity, allergenicity, toxicity and conservancy analysis of protein.

Surface glycoprotein (QIA98583.1)

Epitope	Start	End	Topology	Antigenicity	Score	Allergenicity	Toxicity	Minimum Identity	Conservancy
SNFRVQPTESI	36	46	Inside	Antigen	0.9897	Allergen	Non-toxic	11.11%	100%
NFRVQPTESIV	37	47	Inside	Antigen	1.0669	Non-Allergen	Non-toxic	22.22%	100%
FRVQPTESIVR	38	48	Inside	Non-antigen	0.3493	Allergen	Non-toxic	9.09%	100%
VYYHKNNKSWM	3	13	Inside	Non-antigen	0.3726	Allergen	Non-toxic	18.18%	100%
LGVYYHKNNKS	1	11	Inside	Antigen	0.8696	Allergen	Non-toxic	9.09%	100%
GVYYHKNNKSW	2	12	Inside	Antigen	0.6685	Allergen	Non-toxic	9.09%	100%
LLIVNNATNVV	47	57	Inside	Antigen	0.4166	Non-Allergen	Non-toxic	9.09%	100%
LIVNNATNVVI	48	58	Inside	Non-antigen	0.2045	Non-Allergen	Non-toxic	9.09%	100%
IVNNATNVVIK	49	59	Inside	Non-antigen	0.2274	Allergen	Non-toxic	9.09%	100%
VFVSNGTHWFV	44	54	Outside	Non-antigen	0.0957	Allergen	Non-toxic	18.18%	100%

Table 4b: MHC Class-II epitopes of Surface glycoprotein (QIA98583.1): Table represents topology, antigenicity, allergenicity, toxicity and conservancy analysis of protein.

Surface glycoprotein (QHS34546.1)

Epitope	Start	End	Topology	Antigenicity	Score	Allergenicity	Toxicity	Minimum Identity	Conservancy
GVYFASTEK	19	27	Inside	Antigen	0.7112	Non-allergen	Non-toxic	11.11%	100%
VTYVPAQEK	14	22	Inside	Antigen	0.8132	Allergen	Non-toxic	22.22%	100%
ASANLAATK	39	47	Inside	Antigen	0.7014	Allergen	Non-toxic	22.22%	100%
TLADAGFIK	56	64	Inside	Antigen	0.5781	Non-allergen	Non-toxic	22.22%	100%
TLKSFTVEK	21	29	Inside	Non-antigen	0.0809	Allergen	Non-toxic	11.11%	100%
NSASFSTFK	19	27	Inside	Non-antigen	0.1232	Allergen	Non-toxic	11.11%	100%
TEILPVSMTK	23	32	Inside	Antigen	1.4160	Allergen	Non-toxic	10.00%	100%
SSTASALGK	28	36	Outside	Antigen	0.6215	Allergen	Non-toxic	22.22%	100%
GTHWFVTQR	48	56	Inside	Non-antigen	0.0723	Allergen	Non-toxic	11.11%	100%
EILPVSMTK	24	32	Inside	Antigen	1.6842	Allergen	Non-toxic	11.11%	100%

Table 5a: MHC Class-I epitopes of Surface glycoprotein (QHS34546.1): Table represents topology, antigenicity, allergenicity, toxicity and conservancy analysis of protein.

Surface glycoprotein (QHS34546.1)

Epitope	Start	End	Topology	Antigenicity	Score	Allergenicity	Toxicity	Minimum Identity	Conservancy
NFRVQPTESIV	35	45	Inside	Antigen	1.0669	Non-Allergen	Non-toxic	0.00%	100%
FRVQPTESIVR	36	46	Inside	Non-antigen	0.3493	Allergen	Non-toxic	9.09%	100%
LLIVNNATNVV	37	47	Inside	Antigen	0.4166	Non-Allergen	Non-toxic	0.00%	100%
LIVNNATNVVI	47	57	Inside	Non-antigen	0.2045	Non-Allergen	Non-toxic	9.09%	100%
IVNNATNVVIK	48	58	Inside	Non-antigen	0.2274	Allergen	Non-toxic	9.09%	100%
VFVSNGTHWFV	49	59	Outside	Non-antigen	0.0957	Allergen	Non-toxic	18.18%	100%
GVFVSNGTHWF	43	53	Outside	Non-antigen	0.2539	Allergen	Non-toxic	18.18%	100%
FVSNGTHWFVT	42	52	Outside	Non-antigen	0.0331	Non-Allergen	Non-toxic	18.18%	100%
IRASANLAATK	44	54	Inside	Antigen	0.6339	Allergen	Non-toxic	0.00%	100%
EIRASANLAAT	37	47	Inside	Antigen	0.9115	Allergen	Non-toxic	18.18%	100%

Table 5b: MHC Class-II epitopes of Surface glycoprotein (QHS34546.1): Table represents topology, antigenicity, allergenicity, toxicity and conservancy analysis of protein.

Surface glycoprotein (QIA98583.1)				*Surface glycoprotein (QHS34546.1)*			
Epitope	**Topology**	**Antigenicity**	**Allergenicity**	**Epitope**	**Topology**	**Antigenicity**	**Allergenicity**
RTQLPPAYTNS	Inside	Antigen	Allergen	QIAPGQTGKIAD	Inside	Antigen	Non-Allergen
SGTNGTKRFDN	Inside	Antigen	Allergen	YGFQPTNGVGYQ	Outside	Antigen	Allergen
LTPGDSSSGWTAG	Outside	Antigen	Non-Allergen	RDIADTTDAVRDPQ	Inside	Antigen	Allergen
VRQIAPGQTGKIAD	Inside	Antigen	Non-Allergen	QTQTNSPRRARSV	Inside	Non-antigen	Non-Allergen
YQAGSTPCNGV	Inside	Non-antigen	Non-Allergen	ILPDPSKPSKRS	Outside	Antigen	Non-Allergen

Table 6: B cell epitopes of Surface glycoprotein (QIA98583.1) and (QHS34546.1): Table represents topology, antigenicity and allergenicity analysis of proteins.