



Mapping Immunogenic Regions In SARS-CoV-2 to Understand Vaccine Design Using Bioinformatics

Joanna A. Sanchez Rocha¹, Brady Anderson², Jatniel Morales¹, Christopher Salgado¹, Cidney Smith¹, Mahita Jarjapu³, Marcus Mendes³, Nina Blazeska³, and Sheela Vemu¹

¹*Department of Liberal Arts & Sciences, Waubensee Community College, Sugar Grove, Illinois, 60554 USA*

²*Cleveland Clinic; Cleveland, OH*

³*Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology, La Jolla, California, USA*

* svemu@waubensee.edu

Abstract: Disparities in undergraduate STEM degree completions across the United States are a national concern. Undergraduate-level research opportunities are vital for developing future researchers and building their scientific identity. These experiences can help students in community colleges acquire 21st-century skills and build confidence in their ability to do science [1-3]. The development and implementation of guided research experiences provide users with a topic they are familiar with but not necessarily experts in, like SARS-CoV2 infections. In this particular study, the Immune Epitope Database (IEDB) was used to identify amino acid residues located on the immunogenic regions of the spike glycoprotein of SARS-CoV-2 variants: Alpha, Beta, Gamma, Delta, and Omicron. IEDB is a web-based bioinformatics tool that contains published epitope information and prediction aids that can be used as a research platform for studying infectious diseases. The objective of this study aimed to map the immunogenic regions on the spike glycoproteins of the SARS-CoV-2 variants and predict the immune evasion of these variants [4-6]. Identifying the antigenic determinations that bind to the antibodies is essential for designing future candidates for peptide-based vaccines.

This study aims to map the immunogenic regions on the spike glycoproteins of the SARS-CoV-2 variants and predict the immune evasion of these variants [4-6]. Identifying the antigenic determinations that bind to the antibodies is essential for designing future candidates for peptide-based vaccines. This research identifies regions where mutations have occurred in the virus, which are important to study as they can affect the virus's immune evasion and impact available vaccines. Targeting multiple immunogenic regions unaffected by mutations can serve as potential targets for new vaccines, providing better protection against different variants.

Keywords: Immunology, Bioinformatics, Biotechnology, engineering, IEDB, B-cell, peptide-based vaccines, scientific reasoning

© 2023 under the terms of the J ATE Open Access Publishing Agreement

Introduction

The adaptive immune system in animals can recognize a wide range of antigens or ligands from various pathogens. The B cell and T cell receptors are responsible for identifying these antigens and triggering an



immune response. The specific regions on these antigens are called epitopes, which are recognized by B and T cell receptors. The IEDB database includes validated and benchmarked methods to predict epitope-paratope binding, antigen processing, and T and B cell recognition receptors for infectious diseases, allergens, autoimmune diseases, and transplants.

As part of this research experience, students Brady Anderson, Christopher Salgado, Jatniel Morales Gomez, Brianna Carr, and Joanna Sanchez Rocha were introduced to bioinformatics analytical tools that can help in the development of new vaccines, diagnostics, and therapeutics. They were guided by their mentor, Dr. Sheela Vemu, Associate Professor of Biology at Waubensee Community College. They had an amazing opportunity to interact with the La Jolla Institute of Immunology team, who provided further clarification on the use of the database and how it can be incorporated into their research.

The Immune Epitope Database Analysis Resource (IEDB) IEDB webpage (www.iedb.org) is a tool created by NIAID (National Institute of Allergy and Infectious Diseases). It is freely available to all to provide access to a variety of epitope analysis and prediction tools [7]. The database consists of experimental published data from humans, non-human primates, mice, and other animal species. This database component can also be queried using a flexible user interface that links pathogen-specific and immunological assays into one set of data [8]. It is populated using information captured and curated from peer-reviewed scientific publications worldwide and contains a vast amount of data.

As of August 2022, there were 1,730,172 peptidic epitopes available on the ImmunomeBrowser section in IEDB. Of that number, 1,119,684 epitopes were mapped on structural complexes leaving 610,488 epitopes not mapped. The mapped epitopes are split with 501,491 on a PDB complex and 618,193 mapped on an Alphafold-modeled structure [9]. As of March 2023, over 100,000 unique epitopes have been put into the IEDB database from cutting-edge scientific research. It is continually updated and is the single most powerful repository in the world for a comprehensive collection of experimental data that can be queried for known epitopes and their immunogenic regions.

The ImmunomeBrowser tool in the IEDB maps epitope recognition information to an antigen and computes an immunogenicity score for each position. This identifies immunogenic hotspots in a protein. So far, this data has only been plotted based on the linear sequence of the proteins. As a part of their research experience, the students wanted to construct a 3D visualization of the data scores and show how immunogenic regions in a protein are located in a 3D space.

Throughout this research experience, the students implemented a series of action steps:

- Step 1: Idea development, managing teams in the class with faculty mentorship
- Step 2: Learning new resources and databases - gallery walk
- Step 3: Ideation, brainstorming of ideas
- Step 4: Plan a research project
- Step 5: Reaching out to fellow students and regular follow-through

Research experiences such as these students' can provide opportunities for others to participate in consequential research. Which then can increase the science identity of each student and help them feel a sense of belonging to the larger scientific community.

Materials and Methods

ImmunomeBrowser maps and visualizes queries in IEDB linear peptidic epitopes along the length of a protein sequence (a target, or reference, protein). The tool's purpose is to allow users to explore how often each protein



region has been studied in immune assays and how many assays the immune response was positive or negative. ImmunomeBrowser provides summarized data by reference antigen because: (i) epitopes reported in IEDB were identified for different strains and protein isoforms – mapping to the reference protein allows to visualize and study of such epitopes as they would have the same antigen; (ii) different mutant variants of the same epitope were tested and reported; and (iii) immune response varies among studies and assays due to heterogeneity of samples and complexity of immune response [11]. Immunome Browser is accessible in IEDB via two entry points (or tabs): the Antigen or Epitope navigation paths [1]. A prominent feature in the IEDB is the ImmunomeBrowser tool [10], which maps epitope recognition information back to an antigen and computes an immunogenicity score for each position in that antigen. This can identify immunogenic hotspots of epitope recognition in a protein as compared to other areas that are not recognized. These data have so far only been plotted based on the linear sequence of the proteins. The ImmunomeBrowser [10] is a tool that retrieves all epitopes available in the IEDB related to a given parent protein and calculates a score called the response frequency (RF). This score attempts to draw attention to regions of the antigen that are more immunogenic. The RF score is based on the number of positive assays at each protein position and uses the lower bound of the 95% confidence interval to provide a conservative estimate.

The resulting RF scores reveal immunogenic hotspots within the protein, providing valuable insights into regions more prone to epitope recognition. The data generated by the ImmunomeBrowser are presented in graphical form, allowing researchers to visually analyze and interpret the immunogenicity profile along the length of the protein. Through these methods, researchers gain a comprehensive understanding of the immune response to specific antigens and can efficiently explore immunogenic regions within the protein. The IEDB web-based interface and the ImmunomeBrowser tool streamline the data analysis process, facilitating informed decision-making and contributing to advancements in immunological research.

Overview of The Immune Epitope Database

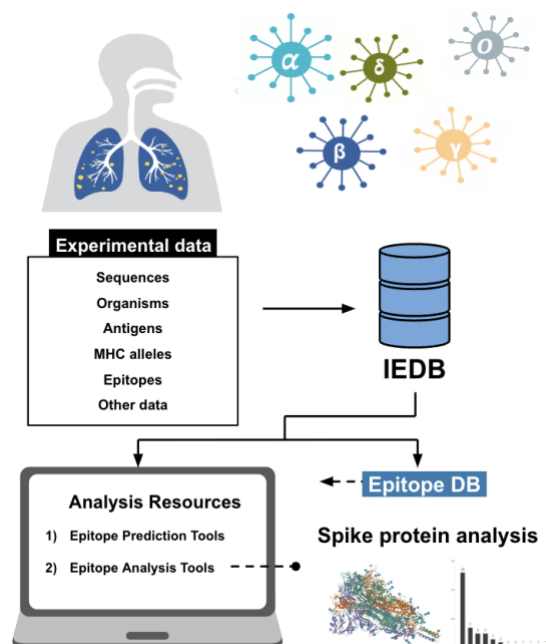


Fig 1. The immune epitope database is composed of real experimental data and provides summarized data by reference antigens. IEDB is composed of the Database and analysis resources.



Results

Analysis of the SARS-CoV-2 spike glycoprotein was achieved through the accession of the IEDB database, specifically the ImmunomeBrowser. Five variants of interest were identified: Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), and Omicron (B.1.1.529). The IEDB database sourced all scientific literature for positive and negative assay results at the protein residue level [10]. By sorting through the ImmunomeBrowser for data on SARS-CoV-2 spike glycoprotein in human B cells, the students could visualize the frequency of positive assays led by researchers worldwide at specific regions in the spike glycoprotein. Upon ImmunomeBrowser input of various SARS-CoV-2 variants of interest, point mutations from each variant were visualized overlay the region of the spike glycoprotein. In this overlay, the students saw that the variants contained unique mutations in multiple domains of the spike protein but also appeared to share regions of similar point mutation occurrences. Of interest, the Omicron variant appeared to possess the highest frequency of point mutations between 333 to 527 residues, in which the receptor binding domain (RBD) consists.

Overview of The Immune Epitope Database

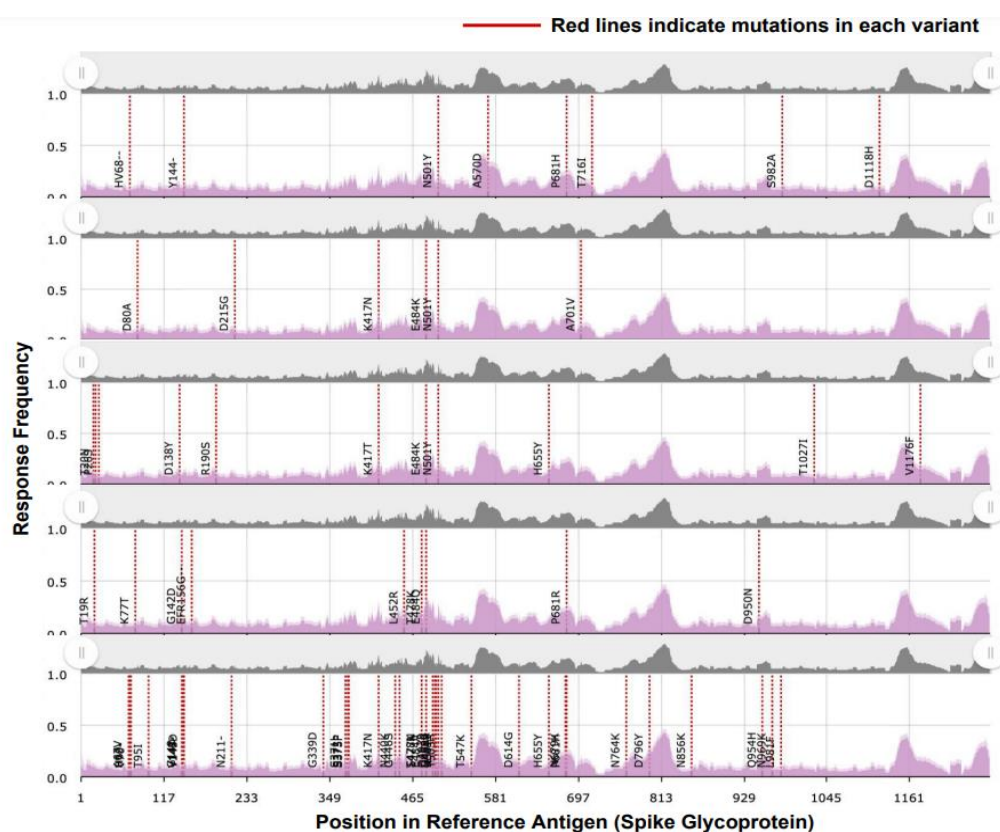


Fig 2: The Immuno Browser graphs above show the map of the response frequency data associated with an amino acid residue. Our results suggest that specific mutations are conserved among these five variants. The immunogenicity hotspots were found in the residue range of 300 - 550, corresponding to the receptor binding domain. The positive epitope assay counts showed activity, while the response frequency exhibited no change when tested against our reference antigen. Omicron displayed the most mutations within the overall range of residues within the hotspot range.



Further analysis of the number of assay counts and associated response frequencies of each residue in the spike protein was plotted and colorimetrically highlighted by each corresponding variant. The students noticed that the Omicron variant had more mutations in residues with increasing response frequencies and increasing assay counts than the other variants of interest.

Next, the students looked at visualizing the spike protein mutations across each of the variants in terms of their unique respective mutations and their shared mutations shown in Fig 2. Using the UpsetR package in RStudio, we saw that Omicron possesses 26 specific mutations on the spike glycoprotein, whereas other variants, such as Gamma, Delta, Alpha, and Beta, have 7, 5, 5, and 3, respectively. Of the residues that shared a mutation with more than one variant, Omicron was common to all. Some residues, such as 417 and 681, had three variants with mutations at these sites. Residues 484 and 501 each contained four of the five variants.

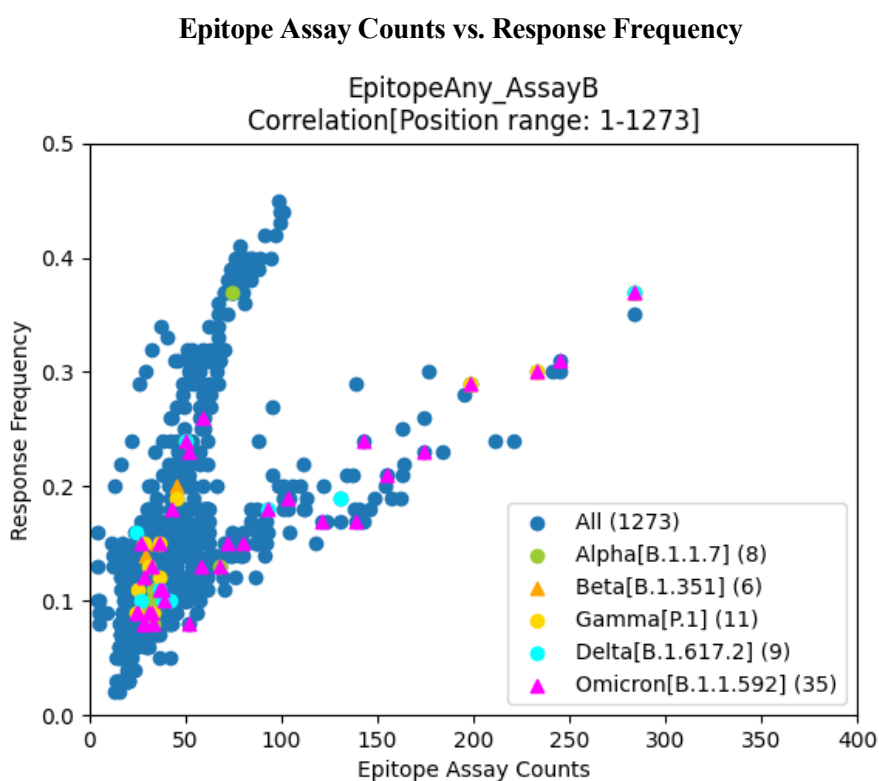


Fig.3: Data from IDEB was analyzed to study correlations between epitope assay counts and response frequency among all five variants. Omicron has the most immunogenic regions and follows the larger count trend concerning higher response frequency. We see two trends: Smaller counts with respect to responses more frequently responded. Larger count with respect to responding less frequently responded. Many mutations in Omicron follow the larger count trend with respect to higher response frequency.



SPIKE MUTATION INTERSECTIONS

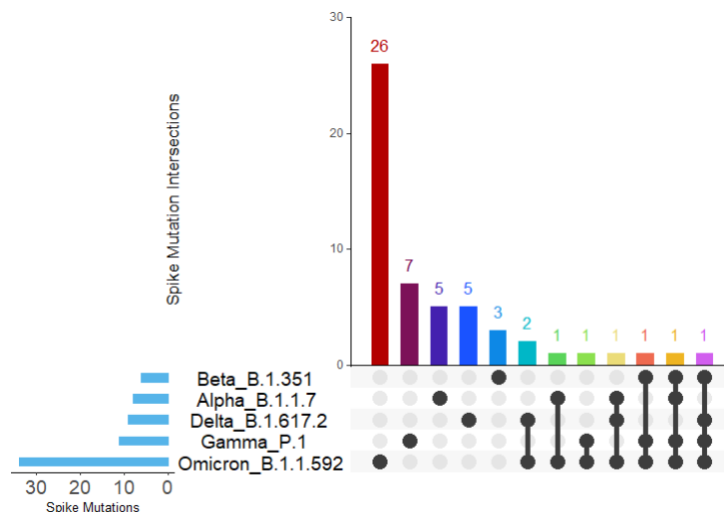


Fig. 4: Shows the shared mutations (amino acid residues) on the spike glycoprotein among the various SARS-CoV-2 variants: Beta, Alpha, Delta, Gamma, and Omicron. Variants with shared mutation residues are depicted by a linking line between each variant (represented as dots). Omicron variants possess 26 unique mutations on the spike glycoprotein and over 30 total spike mutations. Amino acid residues 404 (magenta), 501 (yellow), 417 (orange), and 681 (gold).

All positions in the Spike-Glycoprotein, where Omicron has shared mutations with other variants

Position	142	144	417	478	484	501	655	681
Alpha		o				o		o
Beta					o	o		
Gamma			o		o	o	o	
Delta	o		o	o	o			o
Omicron	o	o	o	o	o	o	o	o

Table 1: Indicates a high degree of overlap (green) as emerging potential hotspots for immune evasion and target for new peptide-based vaccines.

Discussion

The emergence of the Omicron variant has raised numerous concerns, including but not limited to the origin of exposure, the impact of mutations on vaccine efficacy, modifications to host immunity in response to mutations, and the level of lethality and transmissibility of the variant. The student findings suggest that Omicron and Delta have more affinity towards immune escape compared to the other variants due to the receptor binding motif 437-508. This was shown between the amino acid 300 -550 residues. Mutations in this immunogenic region may pose challenges for adaptive immune responses. This could be from the specific protein conformation, which is possibly elusive to antibody and/or vaccine therapies. In the mutation, L452R (Leucine), which is hydrophobic, is bound to a charged arginine in the receptor binding motif of the spike glycoprotein receptor binding domain (RBD). Mutations in this immunogenic region might pose challenges for adaptive immune responses as shown in this data. The increased mutations between Omicron and the other



variants may highlight some of the selective pressures on the virus to avoid natural and therapeutically immune detection via vaccines.

This research helps identify specific regions where mutations have occurred. These regions are essential to study because mutations can potentially alter the virus's behavior, including its infectivity and immune system evasion. Knowing which regions have undergone mutations is fundamental to understanding how the virus evolves and how it might impact available vaccines. Also, the approach applied in this study investigates other immunogenic regions that remain unchanged across different variants. By targeting multiple immunogenic regions unaffected by mutations, these regions can serve as potential targets for new vaccines, eliciting a more robust and broad immune response, making it harder for the virus to evade immune detection, and providing better protection against different variants.

Designing broadly applicable epitope-based vaccines against highly variable pathogens, like SARS-CoV-2, often require epitopes to be conserved across variants [12]. Our results can be used to filter possible candidate epitopes by removing residues that are recurrently hypermutated on the spike glycoprotein across globally important variants. Removing such residues may reduce the footprint of vaccine development in silico, in vitro, and in vivo. Other studies utilizing IEDB have successfully collated datasets on multiple epitopes and predicted possible vaccine constructs [13], binding affinity toward T and B cell motifs [14], and immunogenicity effects [15].

Various bioinformatics approaches have been used to expedite the discovery of potential drugs, design vaccines, and understand the concept behind COVID-19 pathogenesis. Bioinformatics has been instrumental in identifying interactions with SARS-CoV-2 proteins, predicting immunogen and antigen epitopes from SARS-CoV-2 proteins, and identifying new potential pathways in COVID-19 progression and pathogenicity. By leveraging these bioinformatics tools, researchers can model and predict complex biochemical processes related to the virus, accelerating the evaluation of existing vaccines against emerging variants. This will aid in timely and effective decision-making to maintain good immunity and combat the disease.

Therefore, this study demonstrated how immune-dominant parts of SARS-CoV-2 proteins have significantly mutated. Due to the need to identify the correct target for vaccines, we were curious about how an algorithm could suggest an epitope that is both highly antigenic and sufficiently conserved. As we've seen over the pandemic period, SARS-CoV-2 strain nomenclature has evolved considerably and continues to do so. Likewise, new mutations continue to be identified for existing or new strains. The IEDB ImmunomeBrowser utilizes SARS-CoV-2 strain and mutation data from NCBI's ACTIV-TRACE program. This information is updated weekly; therefore, a key limitation of the data is the availability of updated SARS-CoV-2 strains and mutations. Therefore, analysis of the ImmunomeBrowser information must be periodically reviewed as SARS-CoV-2 data is updated.

This research study is an excellent example of how exposure to research can create a catalyst for future generations to explore science. Repurposing IEDB can be a valuable learning tool for students interested in immunology. In addition to the research findings, the students gained the confidence & skills like communication & presentation to apply for regional and national internships & fellowships. By participating in this research experience, the students expressed confidence in building a repertoire of skills that are specific to the industry and, in turn, make them more competitive candidates for future jobs.

Acknowledgments. The students would like to thank their faculty advisor, Sheela Vemu, Ph.D., Associate Professor of Biology at Waubensee Community College. They would like to express their appreciation for the input from Sandy Porter, Ph.D. during the Antibody Hackathons. President Digital World Biology. They would also like to thank Sharon Garcia, Executive Dean for Liberal Arts and Sciences, for advocating the National Institute for Innovation in Manufacturing Biopharmaceuticals (NIIMBL) institutional membership to support student research fellowships and Waubensee Community College for extending support to showcase student



research on campus. This project was supported by the National Science Foundation (NSF) under the award DUE 2055036.

Disclosures. The authors declare no conflicts of interest.

References

- [1] Bangera G, Brownell SE. Course-based undergraduate research experiences can make scientific research more inclusive. *CBE Life Sci Educ.* 2014 Winter;13(4):602-6. doi: 10.1187/cbe.14-06-0099. PMID: 25452483; PMCID: PMC4255347.
- [2] Goodwin EC, Cary JR, Shortlidge EE (2022) Not the same CURE: Student experiences in course-based undergraduate research experiences vary by graduate teaching assistant. *PLoS ONE* 17(9): e0275313.
- [3] Broadening Participation in Undergraduate Research Experiences (URES): The Expanding Role of the Community College James A. Hewlett *CBE—Life Sciences Education* 2018 17:3
- [4] Motozono, C., Toyoda, M., Zahradnik, J. et al. (2021). SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell Host & Microbe*, 29(7), 1124-1136.e11. <https://doi.org/10.1016/j.chom.2021.06.006>
- [5] Taka, E., Yilmaz, S. Z., Golcuk, M. et al. (2021). Critical Interactions Between the SARS-CoV-2 Spike Glycoprotein and the Human ACE2 Receptor. *The Journal of Physical Chemistry B* 2021 125 (21), 5537-5548 DOI: 10.1021/acs.jpcc.1c02048
- [6] Lan, J., Ge, J., Yu, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220 (2020). <https://doi.org/10.1038/s41586-020-2180-5>
- [7] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Research*, 2018; 47(D1):D339–D343. <https://doi.org/10.1093/nar/gky1006>
- [8] Ponomarenko J, Papangelopoulos N, Zajonc DM, Peters B, Sette A, Bourne PE. IEDB-3D: structural data within the immune epitope database. *Nucleic Acids Res.* 2011; 39(Database issue): D1164– 70
- [9] Mendes, M, Mahita, J, Blazeska, N, Greenbaum, J, Ha, B, Wheeler, K, et al. IEDB-3D 2.0: Structural data analysis within the Immune Epitope Database. *Protein Science*. 2023; 32(4):e4605. <https://doi.org/10.1002/pro.4605>
- [10] Dhanda SK, Vita R, Ha B, Grifoni A, Peters B, Sette A. ImmunomeBrowser: a tool to aggregate and visualize complex and heterogeneous epitopes in reference proteins. *Bioinformatics*. 2018; 34(22): 3931–3.
- [11] Beaver JE, Bourne PE, Ponomarenko JV. EpitopeViewer: a java application for the visualization and analysis of immune epitopes in the Immune Epitope Database and Analysis Resource (IEDB). *Immunome Res.* 2007; 3: 3.
- [12] De Groot AS, Moise L, McMurry JA, Martin W. Epitope-Based Immunome-Derived Vaccines: A Strategy for Improved Design and Safety. *Clinical Applications of Immunomics*. 2009; 2: 39-69. https://doi.org/10.1007%2F978-0-387-79208-8_3



- [13] Sarkar B, Ullah MA, Johora FT, Taniya MA, Araf Y. Immunoinformatics-guided designing of epitope-based subunit vaccines against the SARS Coronavirus-2 (SARS-CoV-2). *Immunobiology*. 2020; 225 (3):151955. <https://doi.org/10.1016/j.imbio.2020.151955>
- [14] Rakib A, Sami SA, Mimi NJ, et. al. Immunoinformatics-guided design of an epitope-based vaccine against severe acute respiratory syndrome coronavirus 2 spike glycoprotein. *Computers in Biology and Medicine*. 2020;124:103967. <https://doi.org/10.1016/j.combiomed.2020.103967>
- [15] Jamil FANAS, Auliyana N, Nur M, Nabilah RK. Developing an Epitope-Based Peptide Vaccine for the Hepatitis C Virus Using an in Silico Approach. *KnE Medicine*. 2022. <https://doi.org/10.18502/kme.v2i3.11871>