Helix (2022) 12 (3): 9-14
https://doi.org/10.29042/2022-12-3-9-14

Proteomic annotation and identification of potential peptide vaccine of SARS omicron variant

*[1]Madan Mohan, [2]Jyothsna Gundlapally
[1] Department of Industrial Biotechnology, Bharath Institute of Higher Education and Research, Selaiyur, Chennai-600073,India, [2] BioAxis DNA Research Centre Pvt.Ltd, Hyderabad
mmadhanmohanreddy@gmail.com

Received: 12th April 2022, Accepted: 25th May 2022, Published: 30th June 2022

Abstract
With the incidence of COVID 19 rapid spread and huge mortality rate vaccines were developed against the primary strain of the virus. At most care and multiple measures were taken to make every individual vaccinated ad immune towards the virus. However this could not be the final solution as he virus evolved further to provide additional infections by strain variation and vaccine resistant. Several variants of Corona virus evolved which includes alpha, beta, gamma, delta and omicron all of which has a diverse pathological effects but a very high spread rate. Omicron is a new variant of corona virus that was first reported in November 2021 and is highly immunogenic. It is rapid in spreading when compared to previous variants. Its incubation period is 6 days and exhibits early symptoms. omicron affects the immune system of an infected person and makes the body a hub for several other heart ailments. Further the current day vaccines against COVID 19 failed to prevent the infection by omicron variant, however, it is expected to minimize the pathogenic influence on the vaccinated host. The present research aimed to screen and compare the proteome of the Omicron variant and the wild strain of COVID, also identify the variable proteins which may be the cause of vaccine resistance. Further, a deep proteomic analysis is to be performed to identify and propose a suitable peptide vaccine against the omicron variant of SARS Corona virus.

Keywords: SARS Coronavirus, COVID, Proteome, Omicron, Pathogenicity

Introduction
COVID 19 has ceased a havoc or the years 2019 and 2020. It was a multi variant and multiple times spread of the infections that people faced in various times. The highest concern was the spread of infection even to the pre vaccinated individuals owing to its strain and variants that have resistance to the vaccine and can cause reinfection [5]. There were several variants of COVID 19 which may be Alpha, Beta, Gamma, Delta and Omicron. All the variants were a little variable n the type of infection they cause and the symptoms in spite of their parent organism being same. Alpha variant was reported in May 2020 that has a rapid spread of the infection [2]. Beta variant was identified in August 2020 from South Africa, with a similar spreading ability and also shown to possess resistant against some vaccinations. The third variant Gamma was reported first in Brazil by November 2020 with comparatively high spread rate and low prevention with Vaccines, India in October 2020 experienced the infections by Delta variant with very high rate of spread and the symptoms were found to be highly variable with very low protection rate from Vaccination. The final variant observed was the Omicron type which started in November 2021 in several countries and

The Omicron variant spreads more easily than earlier variants of the virus that cause COVID-19, including the Delta variant [7]. CDC expects that anyone with Omicron infection, regardless of vaccination status or whether or not they have symptoms, can spread the virus to others. Symptoms of Persons infected with the Omicron variant can present with symptoms similar to previous variants [13]. The presence and severity of symptoms can be affected by COVID-19 vaccination status, the presence of other health conditions, age, and history of prior infection. Studies have also found that the Omicron variant has a shorter incubation period, compared to other variants. For the Omicron variant, the incubation period is 2 to 4 days. The most common symptoms of the new COVID Variant "Omicron" are fever, cough, tiredness, and loss of taste or smell. Less common symptoms for the new COVID Variant "Omicron" are sore throat, headache, aches, pains, diarrhea, a rash on the skin, discoloration of fingers or toes, or irritated eyes [10]. With this high variability in symptoms the diagnosis and proper medication is a challenging task. The current work aims to annotate the complete proteome of the new

variant with the focus on identification of pathogenic protein which shows no similarity with the previous strains and thus vaccine resistant.

Materials and Methods

Collection of all proteins of SARS corona virus variant Omicron from NCBI [3]:

NCBI is a premier website and a master data base hat has redundant collection if biological data which includes genes, proteins, SNP, Journals, Article etc along with their required annotations. Being the major and Primary data bases all the researchers would have a primary priority fr submission of Ge and protein data to this data base for their recognition and identity. Each data in the data base is provided a unique code for access and its universal usage. Thus Omicron of Corona Virus being a new organism a huge attention was given for its gene and protein sequencing followed by data submission to NCBI. Thus to obtain the gene and protein related data of this strain NCBI is a best option and thus used for collection of Genetic information about the strain. NCBI is a free and public library for this kind of data.

All the proteins of the Omicron variant were collected from NCBI protein data base with collection restricted to only non redundant data. All the proteins of Omicron variant were collected with their basic annotations.

Foreign Protein Identification

Based on all the proteins collected, each was separately submitted for BLAST P analysis for its comparison to Human proteome to derive their degree of convergence to human proteins. All the protein which share identity at least by 35% are said to be similar and would not be capable of causing an infection or disease n the human host. However the proteins that share no or less than 35% identity would ac as a foreign pathogen and lead to an excited immune response and disease. All the proteins of Omicron were subjected for BLAST P [8] analysis to human proteome and pathogenic protein cluster was gathered.

Identification of Antigenic propensity of proteins:

Once the antigenic proteins were identified and selected they are subjected for further screening to identify the exact antigenic region with its degree of antigenicity called propensity using PVS [11]. All the proteins were screened to obtain a list of peptides (regions within the protein) with high antigenic propensity based on the peaks provided by the graph. This forms the basis of Peptide Vaccine prediction. This server performs the accurate calculation of sequence variability with multiple sequence alignment with the use of some variability matrices.

EMBOSS [4] antigenic analysis:

It is an open source software, an analysis package capable of predicting potentially antigenic regions of user entered protein sequences along with their scoring patterns and comparison. In addition to PVS this site would locate the exact antigenic site capable of eliciting an immune response and binding to an antibody. The tool accepts the input either as a FASTA sequence or the accession number.

VAXIJEN [1] analysis:

It is another approach for the prediction of antigen independent of alignment unlike any other previous tools. However it work on the principle of Autocross co variance ACC transformation of protein sequences. It classifies the antigens purely based on the physicochemical properties. Thus an add on to the previous methodologies of antigenic region prediction.

Prediction of Transmembrane regions of Protein [6]:

Prior to the selection of a protein region as antigenic it has to be considered that the region falls in the exterior of the structure and should not be a trans membrane regions as they are comparatively in accessible. The tool TMHMM provides a pictographic representation of outliner regions, internal as well as ransmembrane regions in a given protein sequence.

NET CTL analysis:

Once the antigenic sites have been identified it becomes one of the top priorities to screen them for the presence of T cell epitopic regions that are involved in sensitizing the T cells and production of an immune response. NetCTL 1.2 predicts the CTL epitopes in protein sequence [9]. It predicts the MHC Class 1 binding to 12 MHC super types which includes A26 and B39.

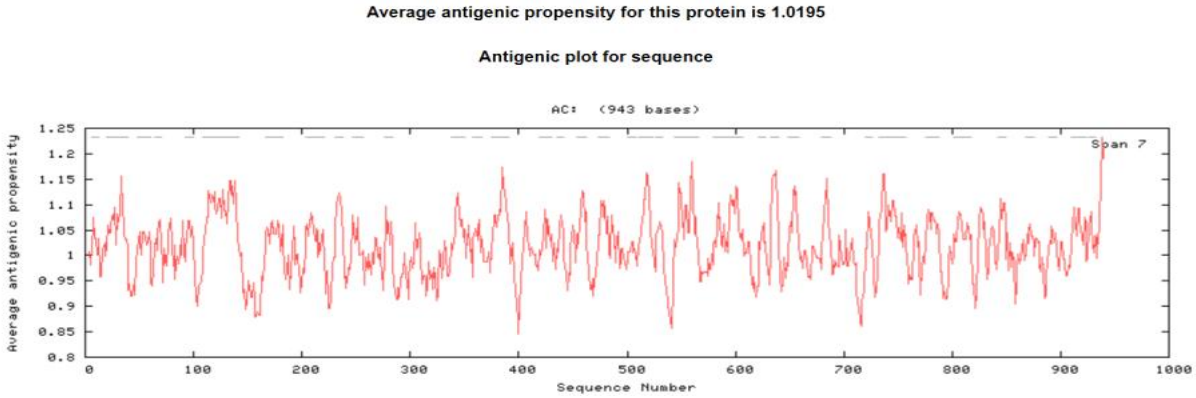
3D structure development of Vaccine peptide:
Once the final list of peptides is selected to be the vaccine candidates, their 3D structure development is the final protocol. For this purpose several software applications are available. The current research utilized Argus lab [12] for structure development.

Results and Discussion

Sequence retrieval of omicron protein using NCBI database:
A total of 21 proteins were collected from Omicron variant which were screened for Foreign protein identification in comparison to human proteome. All the 28 proteins collected were subjected for BLASTP comparison to humans. Among all the screened ones only one protein intimin omicron was antigenic to human proteome. This was further subjected for screening.

Antigenic peptide prediction based on Protein Variability Server (PVS):

Fig 1: The below Graph shows the antigenic sites present in the Intimin protein



Inference: The above result of PVS shows that the peptide falling in the regions above 900 bp is showing a high degree of antigenecity. This can further be tested. Antgenic propensiy of highest antigenic peak (peptide) is approxmetely : 1.25

Fig 2: The result of EMBOSS antigenic for INTIMIN:

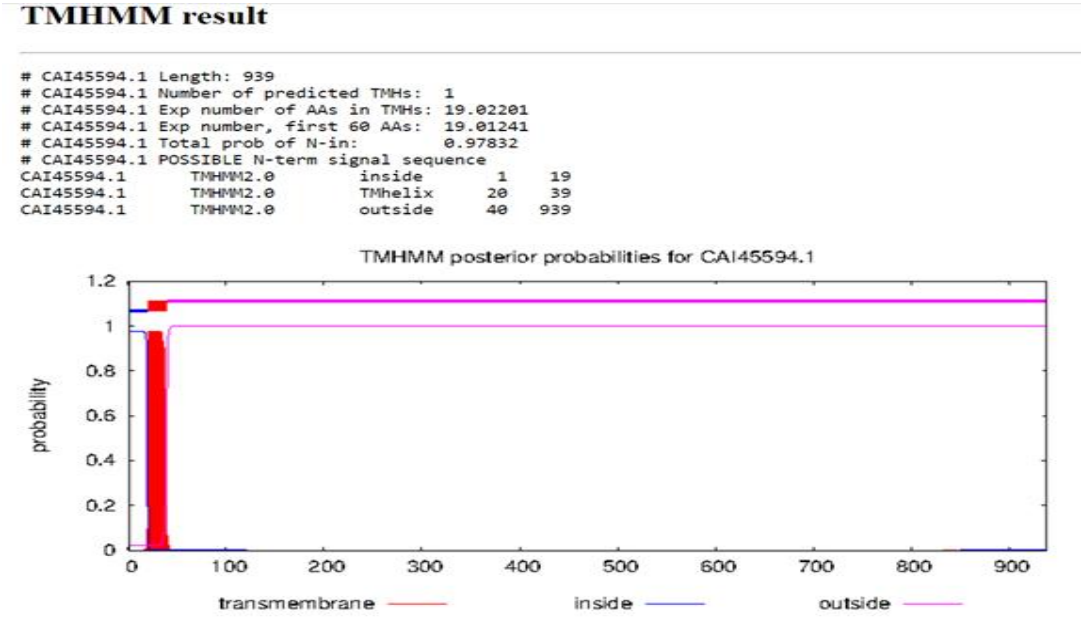
```
# Program: antigenic
# Rundate: Mon  4 Apr 2022 09:03:31
# Commandline: antigenic
#    -auto
#    -sequence /var/lib/emboss-explorer/output/567335/.sequence
#    -minlen 6
#    -outfile outfile
#    -rformat2 motif
# Report_format: motif
# Report_file: outfile
#################################################

#=======================================
#
# Sequence:     from: 1   to: 939
# HitCount: 43
#=======================================

Max_score_pos at "*"

(1) Score 1.233 length 6 at residues 931->936
          *
 Sequence: NYAYAV
           |  |
       931    936
 Max_score_pos: 935

(2) Score 1.184 length 23 at residues 541->563
                *
 Sequence: NVQLTITVLSNGQVVDQVGVTDF
```

Inference: The above result of EMBOSS ANTIGENIC shows that the region 931-936 is highly anigenic with the pattern NYAYAV has the score 1.233. This can be the preferred peptide vaccine

Overall Prediction by TMHMM for the Protective Antigen was found to be 0.6851 (Probable ANTIGEN). Thus TMHMM states the protein to be antigenic.

Inference: From the results of TMHMM it can be observed that the selected protein exhibits high antigenic propensity, thus has potential in eliciting an immune response in the host.

Fig 3: Prediction of trans membrane helices in a protein using TMHMM



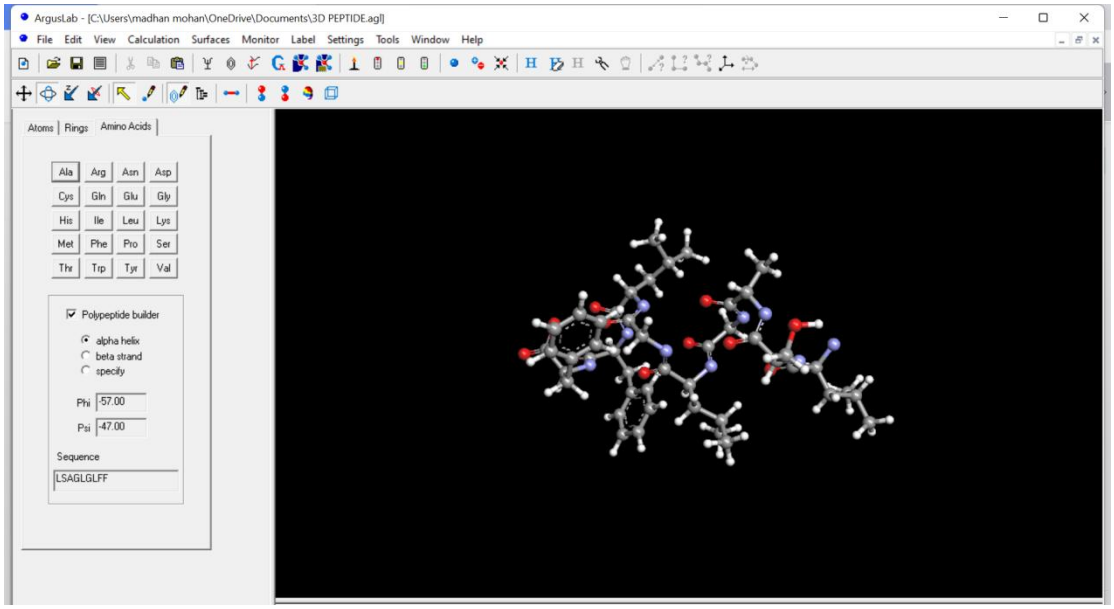
Inference: From the above TMHMM result it can be observed that protien has a trans membrane region from 20 to 39. And the region falling inside the membrane include 1to 19. The region that falls out side includes 40 to 939, thus indicating the selected peptide to be extracellular in region.

Fig 4: Prediction of CTL Epitops in protein sequences using net CTL server

868	ID	Sequence	pep	PSSIKELKD	aff	0.0533	aff_rescale	0.2262	cle 0.0443	tap	-2.2110 COMB	0.1223	
869	ID	Sequence	pep	SSIKELKDL	aff	0.0597	aff_rescale	0.2534	cle 0.6069	tap	0.9720 COMB	0.3930	
870	ID	Sequence	pep	SIKELKDLY	aff	0.1277	aff_rescale	0.5422	cle 0.5824	tap	3.1580 COMB	0.7875	<-E
871	ID	Sequence	pep	IKELKDLYD	aff	0.0528	aff_rescale	0.2244	cle 0.0234	tap	-1.8330 COMB	0.1362	
872	ID	Sequence	pep	KELKDLYDD	aff	0.0515	aff_rescale	0.2188	cle 0.0305	tap	-1.8170 COMB	0.1325	
873	ID	Sequence	pep	ELKDLYDDW	aff	0.0490	aff_rescale	0.2080	cle 0.6224	tap	0.7990 COMB	0.3413	
874	ID	Sequence	pep	LKDLYDDWG	aff	0.0600	aff_rescale	0.2547	cle 0.0236	tap	-1.6850 COMB	0.1740	
875	ID	Sequence	pep	KDLYDDWGA	aff	0.0493	aff_rescale	0.2095	cle 0.1036	tap	-0.6700 COMB	0.1916	
876	ID	Sequence	pep	DLYDDWGAA	aff	0.0584	aff_rescale	0.2479	cle 0.3674	tap	-0.5440 COMB	0.2758	
877	ID	Sequence	pep	LYDDWGAAN	aff	0.0577	aff_rescale	0.2450	cle 0.1307	tap	-1.3620 COMB	0.1965	
878	ID	Sequence	pep	YDDWGAANK	aff	0.0768	aff_rescale	0.3261	cle 0.7638	tap	0.0090 COMB	0.4412	
879	ID	Sequence	pep	DDWGAANKY	aff	0.0517	aff_rescale	0.2194	cle 0.9777	tap	2.4210 COMB	0.4871	
880	ID	Sequence	pep	DWGAANKYQ	aff	0.0471	aff_rescale	0.1998	cle 0.0578	tap	-0.4770 COMB	0.1847	
881	ID	Sequence	pep	WGAANKYQH	aff	0.0506	aff_rescale	0.2149	cle 0.0326	tap	-0.9510 COMB	0.1723	
882	ID	Sequence	pep	GAANKYQHY	aff	0.2909	aff_rescale	1.2353	cle 0.9196	tap	2.7590 COMB	1.5112	<-E
883	ID	Sequence	pep	AANKYQHYS	aff	0.0701	aff_rescale	0.2977	cle 0.0619	tap	-2.1730 COMB	0.1983	
884	ID	Sequence	pep	ANKYQHYSQ	aff	0.0418	aff_rescale	0.1775	cle 0.0315	tap	0.0380 COMB	0.1841	
885	ID	Sequence	pep	NKYQHYSQE	aff	0.0480	aff_rescale	0.2039	cle 0.0394	tap	-1.1380 COMB	0.1529	
886	ID	Sequence	pep	KYQHYSQES	aff	0.0487	aff_rescale	0.2070	cle 0.0530	tap	-1.7650 COMB	0.1267	
887	ID	Sequence	pep	YQHYSQESI	aff	0.0657	aff_rescale	0.2791	cle 0.1885	tap	0.7210 COMB	0.3435	
888	ID	Sequence	pep	QHYSQESIT	aff	0.0510	aff_rescale	0.2164	cle 0.0380	tap	-0.5500 COMB	0.1946	
889	ID	Sequence	pep	HYSQESITA	aff	0.0610	aff_rescale	0.2590	cle 0.8144	tap	-0.4150 COMB	0.3604	
890	ID	Sequence	pep	YSQESITAW	aff	0.1248	aff_rescale	0.5297	cle 0.9085	tap	0.8500 COMB	0.7085	
891	ID	Sequence	pep	SQESITAWT	aff	0.0670	aff_rescale	0.2845	cle 0.0392	tap	-0.7730 COMB	0.2517	
892	ID	Sequence	pep	QESITAWTL	aff	0.0685	aff_rescale	0.2908	cle 0.9716	tap	0.8820 COMB	0.4806	
893	ID	Sequence	pep	ESITAWTLQ	aff	0.1032	aff_rescale	0.4381	cle 0.0304	tap	-0.2490 COMB	0.4302	
894	ID	Sequence	pep	SITAWTLQT	aff	0.0821	aff_rescale	0.3486	cle 0.4514	tap	-0.7270 COMB	0.3799	
895	ID	Sequence	pep	ITAWTLQTS	aff	0.0814	aff_rescale	0.3455	cle 0.0838	tap	-2.1730 COMB	0.2494	
896	ID	Sequence	pep	TAWTLQTSE	aff	0.0573	aff_rescale	0.2433	cle 0.0983	tap	-1.3530 COMB	0.1904	
897	ID	Sequence	pep	AWTLQTSEN	aff	0.0529	aff_rescale	0.2246	cle 0.0375	tap	-0.9510 COMB	0.1827	
898	ID	Sequence	pep	WTLQTSENK	aff	0.0759	aff_rescale	0.3224	cle 0.1509	tap	0.5880 COMB	0.3745	
899	ID	Sequence	pep	TLQTSENKV	aff	0.0751	aff_rescale	0.3190	cle 0.7915	tap	0.2250 COMB	0.4490	
900	ID	Sequence	pep	LQTSENKVQ	aff	0.0490	aff_rescale	0.2078	cle 0.0301	tap	-0.0520 COMB	0.2098	
901	ID	Sequence	pep	QTSENKVQG	aff	0.0646	aff_rescale	0.2745	cle 0.0522	tap	-1.3600 COMB	0.2143	
902	ID	Sequence	pep	TSENKVQGV	aff	0.1363	aff_rescale	0.5789	cle 0.7110	tap	0.1410 COMB	0.6926	
903	ID	Sequence	pep	SENKVQGVA	aff	0.0595	aff_rescale	0.2526	cle 0.1976	tap	-0.7020 COMB	0.2471	
904	ID	Sequence	pep	ENKVQGVAS	aff	0.0449	aff_rescale	0.1908	cle 0.0245	tap	-2.5500 COMB	0.0669	
905	ID	Sequence	pep	NKVQGVAST	aff	0.0523	aff_rescale	0.2219	cle 0.0460	tap	-0.7080 COMB	0.1934	
906	ID	Sequence	pep	KVQGVASTY	aff	0.1573	aff_rescale	0.6678	cle 0.9766	tap	3.2790 COMB	0.9783	<-E
907	ID	Sequence	pep	VQGVASTYD	aff	0.0550	aff_rescale	0.2337	cle 0.0399	tap	-1.7400 COMB	0.1527	
908	ID	Sequence	pep	QGVASTYDL	aff	0.0571	aff_rescale	0.2423	cle 0.6822	tap	0.8260 COMB	0.3860	
909	ID	Sequence	pep	GVASTYDLV	aff	0.0889	aff_rescale	0.3773	cle 0.3575	tap	0.2420 COMB	0.4430	

Inference: The above results show that a total of 32 CTL epitopes were predicted and the region selected for the vaccine development also falls in the region of CTL epitope as shown in the fig 4 adding to its immunogenecity.

Fig 6: 3D structure of the selected peptide in Argus lab:



Inference: The above structure shows the 3D design of the selected peptide. The above 3D structure of the peptide can further be used for the production of vaccine.

Conclusion

The current work was aimed to design a suitable peptide vaccine to the omicron variant of SARS Corona virus. For this the pathogenic protein of omicron was selected. Out of the 28 proteins screened protein that was foreign and pathogenic was intimin of omicron. Initially 28 protein sequences of Omicron were collected from NCBI which were screened for foreigness identification using BLASTP against homo sapiens proteome. The results indicated only 1 protein to be foreign to humans which was identified as intimin. This protein was further analyzed in detail to develop a potential peptide vaccine of the protein. PVS was used to calculate the antigenic propensity of the protein whose results reveled it to be 1.25. For further analysis emboss and vaxijen tools were used to predict antigenic site and antigenic region. Using the Net CTL server the final peptide selected for the peptide vaccine development is LKKTFVMLSAGLGLFFYVNQ (20 to 39). The 3D structure of the peptide was constructed in Argus lab. This can be further synthesized in laboratory and tested for its efficacy in animals.

Reference:

1) Adeyemi NK, Adepoju EG, Adeyemi MA. Distribution Patterns of Primary Health Care Centers in Osun State, Southwestern Nigeria: Implications for Sustainable Development Goals and Containment of COVID-19 Pandemic. Int J Health Serv. 2022 Oct;52(4):512-522. doi: 10.1177/00207314221114533. Epub 2022 Jul 25. PMID: 35876353; PMCID: PMC9325677.
2) Aleem A, Akbar Samad AB, Slenker AK. Emerging Variants of SARS-CoV-2 And Novel Therapeutics Against Coronavirus (COVID-19). 2022 Oct 10. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan–. PMID: 34033342.

3) Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402. PubMed
4) Bitencourt-Ferreira G, de Azevedo WF Jr. Molecular Docking Simulations with ArgusLab. Methods Mol Biol. 2019;2053:203-220. doi: 10.1007/978-1-4939-9752-7_13. PMID: 31452107.

5) Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007 Jan 5;8:4. doi: 10.1186/1471-2105-8-4. PMID: 17207271; PMCID: PMC1780059.
6) Duong BV, Larpruenrudee P, Fang T, Hossain SI, Saha SC, Gu Y, Islam MS. Is the SARS CoV-2 Omicron Variant Deadlier and More Transmissible Than Delta Variant? Int J Environ Res Public

Health. 2022 Apr 11;19(8):4586. doi: 10.3390/ijerph19084586. PMID: 35457468; PMCID: PMC9032753.

7) Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. Nucleic Acids Res. 2008 Jul 1;36(Web Server issue):W35-41. doi: 10.1093/nar/gkn211. Epub 2008 Apr 27. PMID: 18442995; PMCID: PMC2447719.

8) Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001 Jan 19;305(3):567-80. doi: 10.1006/jmbi.2000.4315. PMID: 11152613.

9) Menni C, Valdes AM, Polidori L, Antonelli M, Penamakuri S, Nogal A, Louca P, May A, Figueiredo JC, Hu C, Molteni E, Canas L, Österdahl MF, Modat M, Sudre CH, Fox B, Hammers A, Wolf J, Capdevila J, Chan AT, David SP, Steves CJ, Ourselin S, Spector TD. Symptom prevalence, duration, and risk of hospital admission in individuals infected with SARS-CoV-2 during periods of omicron and delta variant dominance: a prospective observational study from the ZOE COVID Study. Lancet. 2022 Apr 23;399(10335):1618-1624. doi: 10.1016/S0140-6736(22)00327-0. Epub 2022 Apr 7. PMID: 35397851; PMCID: PMC8989396.

10) Rahman, S., Rahman, M.M., Miah, M. *et al.* COVID-19 reinfections among naturally infected and vaccinated individuals. *Sci Rep* **12**, 1438 (2022). https://doi.org/10.1038/s41598-022-05325-5

11) Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022 Jan 7;50(D1):D20-D26. doi: 10.1093/nar/gkab1112. PMID: 34850941; PMCID: PMC8728269.

12) Aleem A, Akbar Samad AB, Slenker AK. Emerging Variants of SARS-CoV-2 And Novel Therapeutics Against Coronavirus (COVID-19). 2022 Oct 10. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan–. PMID: 34033342.