Vaccine Research
ISSN: 2423-4923
eISSN: 2383-2819

In Silico Study to Predict and Characterize SARS-CoV-2 Surface Glycoprotein

Kumar Avnish*[1], Sharma Bhuvnesh[2]

[1]Department of Biotechnology, School of Life Sciences, Dr. Bhimrao Ambedkar University, Agra, Uttar Pradesh, India.
[2]Independent Researcher, 64/218, Madhu Nagar, Agra Uttar Pradesh, India

ARTICLE INFO

Original Article

VacRes, 2020
Vol. 7, No. 1, 10-16
Received: March 17, 2020
Accepted: June 11, 2020
Pasteur Institute of Iran

***Corresponding Author**: Avnish Kumar, Department of Biotechnology, School of Life Sciences, Dr. Bhimrao Ambedkar University, Agra, Uttar Pradesh, India
Email: avnishkumar81@gmail.com
Tel/Fax: (+91)9410410378

KEYWORDS: SARS- CoV-2, COVID-19, Spike protein, Vaccine, Bioinformatics, Antigenic sites

ABSTRACT

Introduction: Coronavirus family member SARS- CoV-2 is a current worldwide threat. It enters into the epithelium membrane of respiratory tract with the help of its antigenic spike proteins and cause Coronavirus disease 2019 (COVID -19). **Methods:** Considering SARS- CoV-2 a potent vaccine or diagnostic candidate, a bioinformatical study was done to determine its structure homology modeling, physiological properties and structure validation with presence of antigenic sites. **Results:** The surface glycoprotein of SARS-CoV-2 was found to be a stable protein with stereochemically good structure. It also contains 65 antigenic sites. **Conclusion:** The present study suggests further wet-lab research to develop a vaccine or diagnostic kit using this promising surface glycoprotein.

Citation:
Kumar A, Sharma B P. In silico study to predict and characterize of SARS CoV 2 Surface glycoprotein. vacres. 2020; 7 (1) :10-16. DOI: 10.29252/vacres.7.1.10

INTRODUCTION

Coronavirus disease 2019 (COVID -19) is recent pandemic threat of the world. Worldwide, millions of COVID-19 patients have been diagnosed after first case in December, 2019 at Wuhan city of Hubei Province in China. A WHO and China joint-venture had identified approximately 80,000 new cases from Hubei and surrounding provinces till first week of March, 2020. It is said about this novel coronavirus that disease symptoms appears on an average four days (practically within two to seven days). However, according to a modeling study, 2.2 days incubation period could be only in 2.2 percent people while 11.5 days in remaining percent of individuals[1-4].

The growth and economic productivity of world has been constrained by this virus to slow down. The AIDS associated stigma was more concerned with family members[5]; however, the financial impact did not affect the society.

COVID-19 is a latest mutant member of coronavirus group of family Coronaviridae. They have appeared with a crown or halo-like structure under electron microscopy. Their outer envelope is decorated with glycoprotein-studs. Coronaviruses cause seasonal infection in human beings and cattle[6]. Their transmission occurs via air droplets but there are no evidences of spread of human coronavirus via cattle or other animals. The human coronaviruses can be divided into two groups, 229E-like and OC43-like. Both groups diverge from each other by antigenic determinants and culturing requirements and also shown very less cross reactivity and thus associated with independent epidemics of indistinguishable disease[6].

The well-studied coronavirus species, known as Severe Acute Respiratory Syndrome-related coronavirus may be a poor representative for COVID-19. It is thus our current knowledge and understanding is limited about of typical aspects associated with this species and as a result we are unable to manage zoonotic spillovers to humans. COVID-19 can kill patients by severe pneumonia[7-9] and due to severity and rapid spreading speed of virus, drugs repurposing has been in trial worldwide. In in-vitro experiments, chloroquine derivative (hydroxy) chloroquine have shown optimistic prophylactic and/or remedial effects. Previously, Chloroquine is in use against rheumatoid arthritis and lupus erythematosus and malaria as an anti inflammatory agent. A study has indicated its possible antiviral activities against SARS-CoV. According to Yan et al., chloroquine could increase endosomal pH and obstruct the glycosylation of cellular receptors of SARS-CoV[10, 11].

In-silico studies are considered as a highly useful step for resolution of the present top pandemic threats. The pathologically important proteins of SARS-CoV-2 are glycoproteins known as 6VSB_A, 6VSB_B and 6VSB_C. These spike proteins target and disrupt the respiratory tract epithelium, leading to deadly pneumonia. Further role of these three glycoproteins in the pathogenesis are under study. A study showed that 6VSB_B is more virulent than 6VSB_A and more important for inducing the host inflammatory and innate immune responses.The bioinformatics tools could predict a structure to understand the other properties of these spike

[DOI: 10.29252/vacres.7.1.10]

glycoproteins. It was considered that antigenic peptide of a protein could be useful in diagnostic and vaccine purpose[12, 13]. In the present work, we evaluated and characterized the surface glycoprotein of SARS-CoV-2.

MATERIALS AND METHODS

Data Set

The sequence of surface glycoprotein (accession number QHD43416.1) was obtained from NCBI Protein database. Thereafter, a surface glycoprotein 6 VSB template sequence was applied for pBlast run to identify template protein with high percentage of similarity[14]. At the same time, the PDB file of 6 VSB [15] were downloaded from the PDB database[16] and to homology modeling crystal structure of prefusion 2019-nCoV spike glycoprotein was utilized [15].

Sequence Analysis

The Protparam tool[17] was applied to evaluate isoelectric point,instability, grand average of hydropathicity(GRAVY), and amino acid and atomic composition for SARS-CoV-2[18].

Homology Modeling

Protein structure homology was done by automatic homology modeling server SWISS-MODEL accessible on a web-server[19]. Three protein models were prepared as per given instruction on SWISS MODEL web server. These three models were checked for its energy level and a model with low energy was considered best for further studies. The PDB format files of an unknown protein (Surface glycoprotein SARS- CoV-2) and the template protein (Prefusion 2019-nCoV spike glycoprotein; 6VSB) was visualized in PyMoL, a molecular visualization system (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC). PyMoL based alignment of homologous sequences between 6VSB and Surface glycoprotein SARS-CoV-2 was recorded and thereafter, surface glycoprotein SARS-CoV-2 was used to construct spatial structure file based on the template protein[20].

Validation of Structure

The protein properties were validated by constructing Ram Chandran Plot (ProSA-web), which is a plot of torsional angles. It is an approach to visualize energetically-allowed regions for backbone dihedral angles[21, 22]. Stereochemical quality of protein model was evaluated using Procheck[23, 24]. Z-score plot (ProSA-web), which is also called standard score, provided an idea of how far from the mean a data point is and indicated the overall model quality[21, 22].

Antigenic Sites In SARS-Cov-2 Surface Glycoprotein

Antigenic sites for this predicted protein sequence was identified by using single parameter based antigenic Emboss tool[25]. The tool was applied as per instruction of developer.

RESULTS

The PROTPARAM Observations

Physical and chemical parameters of homology model of protein demonstrated that spike glycoprotein of SARS-CoV-2 has 1273 amino acids (Fig. 1) and molecular weight of 141178.47 Da with theoretically computed value of isoelectric point (pI) 6.24 that plays important role to define the pH dependent characteristics (stability, catalysis) of protein. Percentage of various amino acid is shown in Fig. 1. Themodel protein had a high percentage of Leucine (an alpha amino acid and non polar aliphatic amino acid) for having a pharmacological activity. The protein has 110 negatively charged residues (Asp + Glu) and 103 positively charged residues (Arg + Lys). Protparam also provided extinction coefficient that narrate how much light could be absorbed by model protein at a specific wavelength (considering Cystine for calculation due to cysteine's property to not absorb appreciably at wavelengths >260 nm) for a given wavelength. The resulted absorption was 1.055 (1g/l).

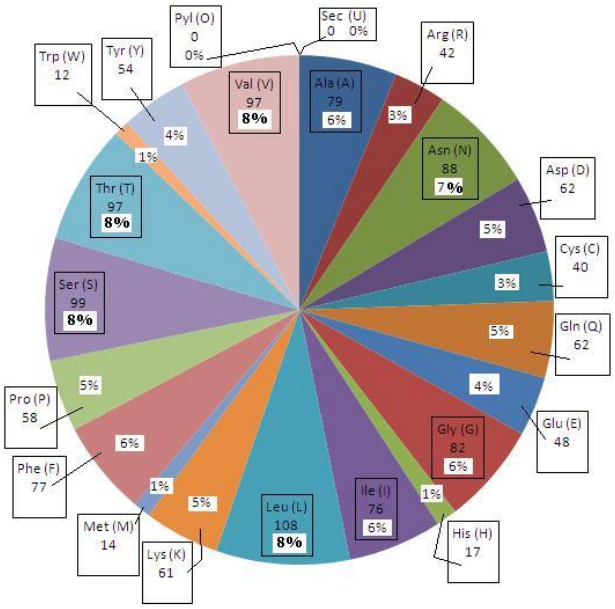

Fig. 1. The *in-silico* based analysis gives amino acid's percent wise composition

On considering N terminal as methionine, the half-life (time taken for half of the amount of synthesized protein in a cell) of model protein was also predicted i.e. 30 h for mammalian reticulocytes in *vitro*>20 h (yeast, *in vivo*). >10 h (*Escherichia coli*, *in vivo*).The instability index value of model protein was found to be 33.01, indicating a stable protein. Being a cut off, value of instability index should be below 40 for a stable protein in test tube.Aliphatic index of protein is narrated about the amino acids which have an aliphatic side chain (Alanine, Valine,Isoleucine and Leucine) in their structure. It may also regard as positive factor for thermostability of globular protein. Modeled protein had Aliphatic Index of 84.67. GRAVY (Grand Average of Hydropathicity) index shows the solubility of the protein. Positive GRAVY shows hydrophobic and negative GRAVY shows hydrophilic. The modeled protein had negative value means protein is hydrophilic.

Structure of the Protein

The PYmol tool is an open source molecular visualization system developed by Warren Lyford DeLano. It produces a high-quality 3D image of protein (Fig. 2) and protein with template alignment (Fig. 3).

[DOI: 10.29252/vacres.7.1.10]

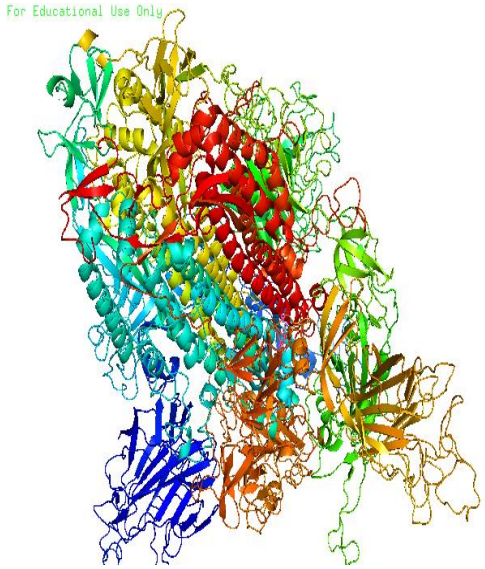

Fig. 2. SARS-CoV-2 surface glycoprotein homology model

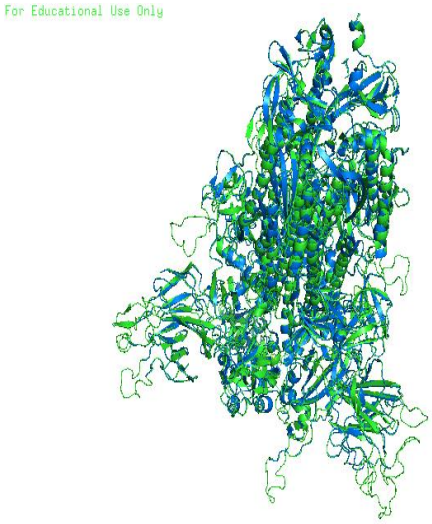

Fig.3. Structure of Surface glycoprotein with template alignment

Validation of Protein Structure

(I) Ramchandran Plot

Analysis and understanding of the reliability of a protein model is extremely important in terms of overall structure as well as its details. Ramachandran plot could authenticate the co-ordinates and quality of the protein structure[26]. Statistically Ramchandran plot of 6VSB homology model has 2533 out of 2976 non-Proline and non-Glycine residues i.e. 85.1 % residues in the core regions which indicated a stereochemically good structure quality of the protein (Fig. 4). Other qualities indicated in Ramchandran plot has been mentioned in Table 1.

Table 1. Properties based on Ramchandran plot analysis

***| All Ramachandrans:**	192 labelled residues (out of3354)					
+| Chi1-chi2 plots:	35 labelled residues (out of1954)					
| Side-chain params:	5 better	0 inside	0 worse			
***| Residue properties: Max.deviation:**	6.9		Bad contacts:	0		
*|**Bond len/angle:**	8.1		Morris et al class:	1	1	2 |
+|1 cis-peptides|						
| **G-factors**	Dihedrals:	-0.40	Covalent:	-0.02	Overall:	-0.23 |
*| Planar groups:	88.2%		within limits	11.8%	highlighted	41 off graph

[DOI: 10.29252/vacres.7.1.10]

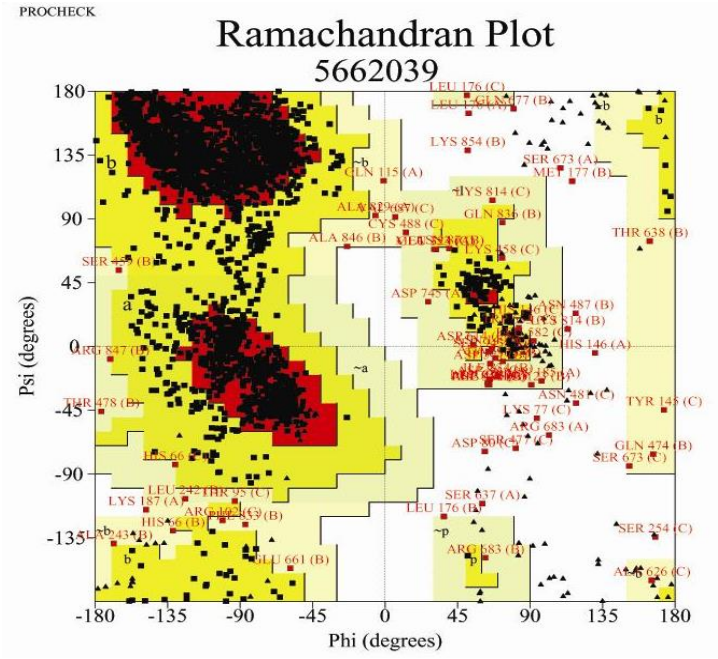

Fig.4. Ramchandranplot of 6VSB protein from SARS-CoV-2 surface spikes. The white coded areas verify the conformations of closer atoms in the polypeptide than the sum of their van der Waals radii. White regions allowed exclusively for glycine due to absence of a side chain. The yellow coded regions are specific for shorter van der Waals radii atoms that can come to little closer together. This brings out an additional region which corresponds to the left-handed alpha-helix. The red region contains alpha-helical and beta-sheet of proteins.

(II) Z-score

The overall model quality Z-Score from ProSA-web is shown in Fig. 5.The positive Z-score shows that the structure is not stabilized while zero and negative score is an indication of one of the best ideal structures[27, 28]. The Z-score of the template sequence (6VSB) was -6.57 as shown in Fig. 5. This negative score of the target demonstrated that the model is structurally reliable.

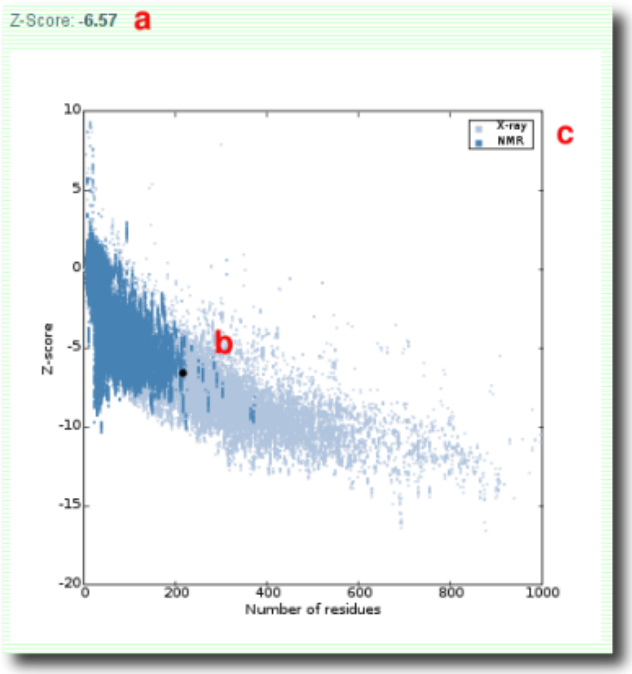

Fig. 5. Z-score of the template sequence (6VSB) from SARS-CoV-2 surface spikes. The group of X Ray sources (light blue) and NMR sources (dark blue) are shown.

While the graph in Fig. 6, below for the target is mostly below from the zero on x-axis, based on this graph, the modeled 3D structure for the target can be constructed [27].

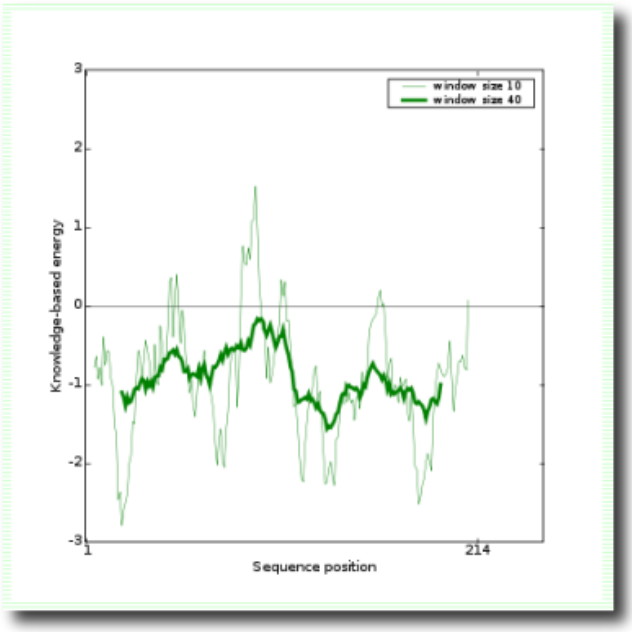

Fig.6. Plot of residue scores of a native protein structure. Plot of residue scores of protein structure showed the energy graph.

Analysis of physicochemical parameters discloses that SARS-CoV-2 glycoprotein has 1273 amino acids and molecular weight of 141178.47 Da with GRAVY score −0.079. A negative GRAVY score belongs to stable, hydrophilic protein molecule which is capable of establishing hydrogen bonds[28].

[DOI: 10.29252/vacres.7.1.10]

Antigenic Sites in SARS-CoV-2 Surface Glycoprotein

On experimental basis, hydrophobic residues Cys, Leu and Val are a part of antigenic sites on surface of proteins [25]. Antigenic site prediction tool Emboss [25] predicted a total of 65 antigenic sites on SARS-CoV-2 surface glycoprotein (Table 2).

Table 2. Antigenic sites are predicted in SARS-CoV-2 surface

Start	End	Score	Max_score_pos
4	19	1.261	8
1215	1256	1.221	1250
1039	1071	1.215	1063
504	528	1.214	510
123	146	1.213	129
358	372	1.183	364
607	629	1.182	610
1123	1136	1.175	1129
857	867	1.167	861
535	541	1.152	536
262	278	1.151	267
209	233	1.147	228
485	497	1.144	491
34	71	1.143	45
1262	1270	1.143	1267
387	404	1.140	393
168	178	1.139	173
115	121	1.139	117
749	765	1.138	752
332	341	1.132	338
286	296	1.131	294
846	855	1.131	852
429	436	1.126	432
998	1016	1.126	1010
780	792	1.126	784
912	921	1.125	914
1029	1037	1.125	1035
374	385	1.123	380
959	969	1.122	965
238	251	1.120	242
581	600	1.115	587
1138	1146	1.115	1140
684	698	1.113	690
1172	1180	1.112	1177
1078	1086	1.110	1080
973	996	1.107	974
735	746	1.107	741
661	675	1.106	671
872	881	1.101	876
632	654	1.101	652
932	955	1.100	949
891	899	1.100	895
768	773	1.099	769
834	844	1.097	843
347	353	1.095	350
450	456	1.094	454
1195	1203	1.091	1200
719	731	1.089	727
558	564	1.088	562
157	163	1.085	160
818	832	1.084	824
298	311	1.076	303
23	29	1.075	25
1091	1098	1.074	1093
81	94	1.071	85
317	330	1.070	329
802	810	1.064	805
548	554	1.064	549
200	207	1.062	203
407	412	1.061	410
1161	1167	1.058	1167
470	478	1.050	477
702	708	1.044	704
1206	1213	1.036	1212
1186	1192	1.022	1186

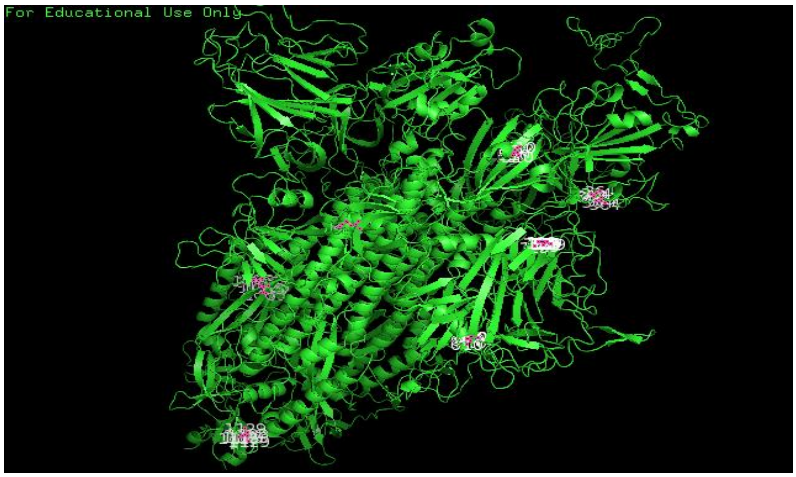

Fig.6. SARS-CoV-2 with different antigenic sites.

[DOI: 10.29252/vacres.7.1.10]

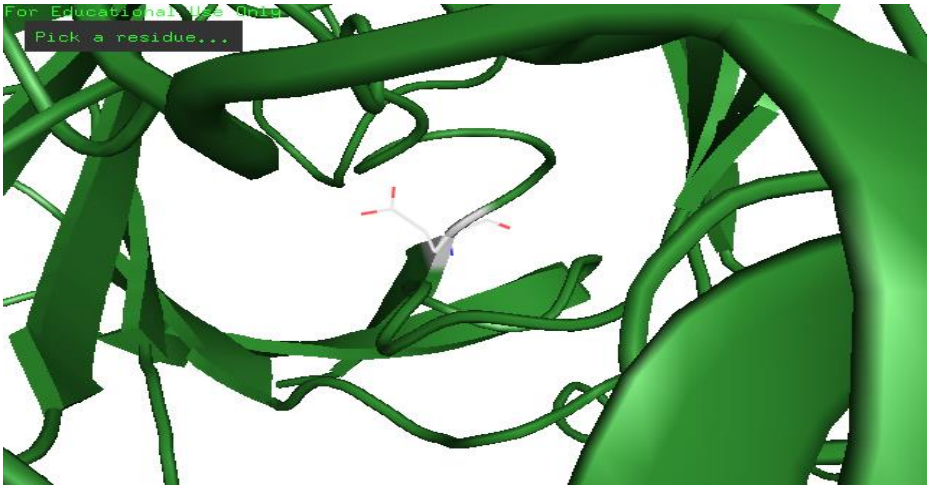

Fig. 7. SARS-CoV-2 surface glycoprotein with antigenic sites. Fig. 6 shows different antigenic sites while position view of antigenic site has been shown in Fig. 7.

DISCUSSION

Due to the epidemic emergency, bioinformatics studies are highly significant to analyze the protein structure of SARS-CoV-2 surface glycoprotein. While the conventional vaccines are able to stimulate an immune reaction using the whole pathogen, virulent or antigenic part into the host, bioinformatics could help the vaccine researchers to identify antigenic regions or epitopes present in a protein which are potential candidate to provoke different arms of the immune system [29]. Moreover, surface-exposed peptides study can reduce the investigation time and cost while adding higher specificity to a vaccine research [30].

Here we analyzed the structure of SARS-CoV-2 surface glycoprotein by a homology modeling method along with different other tools. We identified 65 antigenic sites present in SARS-CoV-2 spike glycoprotein which makes it a potential candidate for the vaccine design. This antigenic spike protein had a molecular weight of 141178.47 Da with a pI value of 6.24. Investigating the pI is considered important for the pH-dependent characteristics like a protein's stability and catalysis properties. The investigated spike protein had also a high percentage of Leucine and an estimated half-life of 30 h in mammalian cells with an absorption of light was (>260 nm) 1.055 (1g/l). Moreover, the negative value of GRAVY of this spike glycoprotein suggested its hydrophilic nature. It can also be envisaged that this protein can form hydrogen bond easily. Altogether, the present study suggests that SARS-CoV-2 appears to be highly stable protein with different antigenic sites. Based on the present findings and upon further analyses, the antigenic sites of these surface glycoproteins could be a potential candidate for peptide or DNA vaccines as well as pertaining diagnostics kits.

ACKNOWLEDGEMENT

Authors are highly grateful to the software tools developers and their teams for SWISS MODEL, ProCheck, ProSA, Antigenic Emboss, and PyMOL.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New England Journal of Medicine. 2020.
2. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Annals of internal medicine. 2020;172(9):577-82.
3. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J et al. China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease. 2019:1708-20.
4. Chan JF-W, Yuan S, Kok K-H, To KK-W, Chu H, Yang J et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The Lancet. 2020;395(10223):514-23.
5. Herek GM, Capitanio JP. Public reactions to AIDS in the United States: a second decade of stigma. American journal of public health. 1993;83(4):574-7.
6. Tyrrell DA, Myint SH. Coronaviruses. Medical Microbiology. 4th edition. University of Texas Medical Branch at Galveston; 1996.
7. Gao J, Tian Z, Yang X. Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. Bioscience trends. 2020.
8. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell research. 2020;30(3):269-71.
9. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J et al. China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382(8):727-33.
10. Yan Y, Zou Z, Sun Y, Li X, Xu K-F, Wei Y et al. Anti-malaria drug chloroquine is highly effective in treating avian influenza A H5N1 virus infection in an animal model. Cell research. 2013;23(2):300-2.
11. Savarino A, Boelaert JR, Cassone A, Majori G, Cauda R. Effects of chloroquine on viral infections: an old drug against today's diseases. The Lancet infectious diseases. 2003;3(11):722-7.
12. Kumar A, Girdhar BK, Parkash O. Immunoproteomic analysis of Mycobacterium leprae derived cell membrane antigen. Int J Biol Med Res. 2010;1(4):242-7.
13. Kumar A, Parkash O, Girdhar BK. Analysis of antigens of Mycobacterium leprae by interaction to sera IgG, IgM, and IgA response to improve diagnosis of leprosy. BioMed research international. 2014;2014.
14. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W et

[DOI: 10.29252/vacres.7.1.10]

al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997;25(17):3389-402.
15. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020;367(6483):1260-3.
16. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H et al. The protein data bank. Nucleic acids research. 2000;28(1):235-42.
17. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook. Springer; 2005. p. 571-607.
18. ul Qamar MT, Alqahtani SM, Alamri MA, Chen L-L. Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. Journal of pharmaceutical analysis. 2020.
19. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic acids research. 2018;46(W1):W296-W303.
20. Liu W, Li H. COVID-19: attacks the 1-beta chain of hemoglobin and captures the porphyrin to inhibit human heme metabolism. Preprint revised on. 2020;10(04).
21. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research. 2007;35(suppl_2):W407-W10.
22. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins: Structure, Function, and Bioinformatics. 1993;17(4):355-62.
23. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography. 1993;26(2):283-91.
24. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of biomolecular NMR. 1996;8(4):477-86.
25. Kolaskar A, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS letters. 1990;276(1-2):172-4.
26. Tyagi C, Singh A, Singh IK. Mechanistic insights into mode of action of rice allene oxide synthase on hydroxyperoxides: An intermediate step in herbivory-induced jasmonate pathway. Computational biology and chemistry. 2016;64:227-36.
27. Darabi M, Seddigh S. Structural, functional, and phylogenetic characterization of phosphoenolpyruvate carboxylase (PEPC) in C4 and CAM plants. Caryologia. 2018;71(3):272-88.
28. Darabi M, Seddigh S. Bioinformatic characterization of aspartic protease (AP) enzyme in seed plants. Plant systematics and evolution. 2015;301(10):2399-417.
29. Dhanda SK, Usmani SS, Agrawal P, Nagpal G, Gautam A, Raghava GP. Novel in silico tools for designing peptide-based subunit vaccines and immunotherapeutics. Briefings in Bioinformatics. 2017;18(3):467-78.
30. Ul Qamar MT, Saleem S, Ashfaq UA, Bari A, Anwar F, Alqahtani S. Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. Journal of translational medicine. 2019;17(1):362.

[DOI: 10.29252/vacres.7.1.10]