WORLD JOURNAL OF PHARMACEUTICAL RESEARCH
SJIF Impact Factor 8.084
Volume 9, Issue 5, 1306-1325. Research Article ISSN 2277– 7105

IN SILICO STUDY AND VALIDATION OF SARS CORONAVIRUS-2 (COVID19) PROTEINS FOR BETTER UNDERSTANDING THE CONTROL MEASURES

Ashokan Kannarath*, Koshti V. Vijay[1], Mundaganur D. S. [2] and Mundaganur Y. D.[3]

*Department of Biological Science (Zoology), P.V.P. College, K.Mahankal, Sangli, Maharashtra, India-416405.

[1]Department of Statistics, P.V.P. College, K.Mahankal, Sangli.

[2]Mundaganur D.S., Department of Zoology, Willingdon College, Sangli-4146416.

[3]Mundaganur Y.D., Department of Zoology, Miraj College, Miraj, Sangli 416410.

Article Received on
01 March 2020,
Revised on 22 March 2020,
Accepted on 12 April 2020
DOI: 10.20959/wjpr20205-17338

***Corresponding Author**
Dr. Ashokan Kannarath
Department of Biological Science (Zoology), P.V.P. College, K.Mahankal, Sangli, Maharashtra, India-416405.

ABSTRACT

SARS Corona Virus-2 (COVID19) protein sequences of nucleocapsid Phosphoprotein (N), membrane glycoprotein (M), envelope protein (E) and surface (spike) glycoprotein (S) was evaluated with bioinformatics servers and tools. The study revealed that membrane protein shows highest percentage of amino acid (>100) and the least (97.9) in S-protein. The bit score also shows the same trend but envelope protein shows the least bit score and maximum E-value, a positive correlation. The amino acid composition shows that all the proteins studied are rich in Leucine except N- protein which contains more glycine. Pyrrolysine, selenocystein and cysteine are absent in all the proteins studied, an indication of low thermo stability. The physico-chemical studies shows that except spike protein all other proteins studied are positively charged due to aspartic acid and glutamic acid, but S-protein is negatively charged. EC is maximum in S protein and minimum in E protein. Instability index (II) shows that N and S proteins are more stable. Further all proteins are hydrophobic except E- protein. We also generate 3D model and the model was evaluated and observed that the 3D structure falls in the accepted limits. The study also focused on phosphorylation site, transmembrane sequence and hydrophobic residues. The N-glycosylation site prediction showed that it is very poor in N, E protein and totally absent in M-protein and more in S-protein. The practical implication of the study is

that the result will assists the scientist and technicians related to control the SARS Corona Virus -2 (COVID19) in better way.

KEYWORDS: SARS Coronavirus-2, Physico-chemical parameters, phosphorylation site, N-Glycosylation, hydrophobic residues and transmembrane protein.

INTRODUCTION

A novel virus causing a cluster of pneumonia was detected in Wuhan City, Hubei province of China in January, 2019. This virus was named as Sever Acute Respiratory Syndrome coronavirus-2 (SARS Coronavirus-2) (WHO). Coronaviruses are a family of viruses named for the crown-like spikes found on their surface. They carry their genetic material in single positive strand of RNA, rather than DNA. They infect a variety of human and animal hosts, causing mostly upper-respiratory symptoms like those of the common cold. Until recently, two coronaviruses have been known to have caused severe disease in humans: Middle East Respiratory Syndrome, or MERS, identified in 2012, and Severe Acute Respiratory Syndrome, or SARS, which was identified in humans in 2002. The SARS-CoV-2 that causes the currently circulating COVID-19 disease is the third example. Phylogenetic data suggesting zoonotic origin[1], and rapid spreading suggest person to person transmission.[2,3,4,5,6,7] COVID-19 was initially identified in Wuhan, China in December 2019 as pneumonia case and now became global pandemic affected more than 170 countries.[8,9,10,5,11] The SARS Coronavirus-2 has genetic material as single strand positive RNA. The genome share 80% similarity with SARS-CoV and about 96 % similarity with bat coronavirus Bat CoV RaTG13.[12] The structure of the SARS Coronavirus-2 is very identical in molecular organization of SARS coronavirus. The main proteins are membrane protein, Nucleo capsid protein, envelope protein, surface (spike) protein etc. The spike glycoproteins (S- proteins) on the surface of the virus (SARS CoV) help to recognize the receptor and fusion of the virus with the host cell.[13,14] During the viral infection the S-protein splits into S_1 and S_2. S_1 contain Receptor binding protein (RBD) which bind to the peptidase domain of angiotensin converting enzyme-2[15], but S_2 responsible for membrane fusion. When S_1 bind to the receptor another cleavage site of S_2 is exposed and cleaved by host protease, this is essential for infection.[16,17,18] SARS CoV2 also exploits the same mechanism as SARS Coronavirus to establish infection.[12,19,20] The other three structural proteins like M, C and E together forms the envelope of the virus.[11] In the present study we analyzed the selected four

proteins by *in silico* methods with the aim of assisting the concerned to develop better control measures.

METHODS

Extraction of protein sequences

The word search "Covid19 protein" in NCBI retrieve 28 reference sequences of proteins. Out of the twenty eight reference proteins the sequence in fasta format of four Covid19 (SARS Corona Virus- 2) proteins was extracted form NCBI (National center for Biotechnology). NCBI is the part of US national Library of Medicine (NLM), a branch of the national Institute of Health. It has a series of data base collection necessary to biotechnology and bioinformatics. The sequences extracted from NCBI and its various parameters are given in (Table-1).

Table 1: Retrieval of protein sequences and its specificity for SARSCoronovirus-2(Covid19).

Identifier	Accession number(Blast-P-NCBI)	Protein	Score in bits	Expect	Identities	E-value	Sequence length in Number of amino acids
YP_009724397	YP_009724397.2	Nucleocapsid Phosphoprotein (N)	854	854	100	0.0	419
YP_009724393	YP_009724393.1	Membrane glycoprotein (M)	452	452	100	1e-160	222 aa
YP_009724392	YP_009724392.1	Envelope protein (E)	144	144	100	3e-43	75 aa
YP_009724390	YP_009724390.1	Surface (Spike)glycoprotein(S)	2637	2637	100	0.0	1273 aa

Identification of amino acid composition

These sequences was used for the *in silico* studies. The SARS CoV 2 has four structural proteins like S (Spike), E (Envelope), M (Membrane), and N (nucleocapsid) proteins. The N protein holds the RNA genome and the other three proteins form the viral envelope. We analyzed the amino acid composition of the selected proteins (Table-2), its homology then selected the nearest homologous protein considering the E value and identity score. The proteins sequences were then subjected to various bioinformatics analysis and created its 3D structure. The structure was validated for its authentication and constructed the Ramchandran plot for further validation.

Extinction coefficients

The extinction coefficient indicates how much light a protein absorbs at a certain wavelength. It is useful to have an estimation of this coefficient for analyzing a protein with a spectrophotometer when purifying it. It has been shown[21] that it is possible to estimate the molar extinction coefficient of a protein from knowledge of its amino acid composition. From the molar extinction coefficient of tyrosine, tryptophan and cysteine (cysteine does not absorb appreciably at wavelengths >260 nm, while cysteine does) at a given wavelength, the extinction coefficient of the native protein in water can be computed using the following equation:

E (Prot) = Numb (Tyr)*Ext (Tyr) + Numb (Trp)*Ext (Trp) +Numb (Cysteine)*Ext (Cysteine)

Where (for proteins in water measured at 280 nm):

Ext (Tyr) = 1490, Ext (Trp) = 5500, Ext (Cysteine) = 125;

The absorbance (optical density) can be calculated using the following formula:

Absorb (Prot) = E (Prot) / Molecular weight

Instability Index

The instability index provides an estimate of the stability of a protein in a test tube. Statistical analysis of 12 unstable and 32 stable proteins has revealed[22] that there are certain dipeptides, the occurrence of which is significantly different in the unstable proteins compared with those in the stable ones. The authors of this method have assigned a weight value of instability to each of the 400 different dipeptides (DIWV). Using these weight values it is possible to compute an instability index (II) which is defined as:

$$II = \frac{10}{L} * \sum_{i=1}^{i=L-1} DIWV(x[i]x[i+1])$$

Where: L is the length of sequence

DIWV(x[i] x[i+1]) is the instability weight value for the dipeptide starting in position i.

A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable.

Aliphatic Index

The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, Valine, isoleucine, and Leucine). It may be regarded as a positive factor for

the increase of thermo stability of globular proteins. The aliphatic index of a protein is calculated according to the following formula[23]:

Aliphatic index = X (Ala) + a * X (Val) + b * (X (Ile) + X(Leu))

Where X (Ala), X (Val), X (Ile), and X (Leu) are mole percent (100 X mole fraction) of alanine, Valine, isoleucine, and Leucine. The coefficients a and b are the relative volume of Valine side chain (a = 2.9) and of Leu/Ile side chains (b = 3.9) to the side chain of alanine.

3D Model generation

The 3D Model of the protein was generated by using SWISSMODEL package. SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server, or from the program Deep View (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modeling accessible to all biochemists and molecular biologists worldwide.[24,25,26] The three Models were created with the help of YASARA software by using the pdb template generated from SWISSMODEL.

Validation of 3D model

The 3D model was predicted by various parameters given by SWISS model server as sequence identity percentage, E-value and QMEAN Z Square.

Gravy (Grand Average Hydropathy)

The GRAVY value for a peptide or protein is calculated as the sum of hydropathy values.[27] of all the amino acids, divided by the number of residues in the sequence.

Prediction of phosphorylation site

Phosphorylation sites were predicted by using NetPhos. The NetPhos 2.0 server produces neural network Predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins.[28]

Transmembrane sequence analysis

Transmembrane domains were predicted by using SOSUI server.[29,30,31] SOSUI which distinguishes between membrane and soluble proteins from amino acid sequences, and predicts the transmembrane helices for the former.

Prediction of hydrophobic residues

The hydrophobic residues were predicted by using Pepwheel program. Pepwheel draws a helical wheel diagram for a protein sequence. This displays the sequence in a helical

representation as if looking down the axis of the helix. It is useful for highlighting amphipathicity and other properties of residues around a helix. By default, aliphatic residues are marked with squares; hydrophilic residues are marked with diamonds, and positively charged residues with octagons, although this can be changed.[32]

Glycosylation site prediction

The glycosylation sites were predicted by NetNglyc bioinformatics tool. N-glycosylation is known to occur on Asparagines along with other factors. NetNglyc attempt to distinguish glycosylated seqouns from non-glycosylated ones.

RESULTS

The sequence (Table1) form NCBI revealed that the spike protein contain maximum percentage of amino acids (1273aa), the least in envelope protein (73aa). The score in bits also shows the same trend with 100% identities in all. When come to the E- value maximum is observed in envelope and membrane protein and least in spike and nucleocapsid phosphor protein. This shows that spike protein is intracellular in location.

Table 2: Amino acid composition of selected four proteins.

No	Amino acids	Number				Percentage			
		Nucleocapsid Phosphoprotein	Membrane glycoprotein	Envelope protein	Surface glycoprotein	Nucleocapsid Phosphoprotein	Membrane glycoprotein	Envelope protein	Surface glycoprotein
1	Ala	37	19	4.0	22	8.8	8.6	5.3	6.3
2	Arg	29	14	3.0	9.0	6.9	6.3	4.0	2.6
3	Asn	22	11	5.0	21	5.3	5.0	6.7	6.0
4	Asp	24	6.0	1.0	20	5.7	1.8	1.3	5.7
5	Cys	00	4.0	3.0	14	0.0	2.7	4.0	4.0
6	Gln	35	4.0	0.0	20	8.4	1.8	0.0	5.7
7	Glu	12	7.0	2.0	17	2.9	3.2	2.7	4.9
8	Gly	43	14	1.0	21	10.3	6.3	1.3	6.0
9	His	4.0	5.0	0.0	8.0	1.0	2.3	0.0	2.3
10	Ile	14	20	3.0	24	3.3	9.0	4.0	6.9
11	Leu	27	35	14	32	6.4	15.8	18.7	9.2
12	Lys	31	7.0	2.0	20	7.4	3.2	2.7	5.7
13	Met	7.0	4.0	1.0	5.0	1.7	1.8	1.3	1.4
14	Phe	13	11	5.0	15	3.1	5.0	6.7	4.3
15	Pro	28	5.0	2.0	11	6.7	2.3	2.7	3.2
16	Ser	37	15	8.0	28	8.8	6.8	10.7	8.0
17	Thr	32	13	4.0	19	7.6	5.9	5.3	5.4
18	Trp	5.0	7.0	0.0	4.0	1.2	3.2	0.0	1.1
19	Tyr	11	9.0	4.0	10	2.6	4.1	5.3	2.9

20	Val	8.0	12	13	29	1.9	5.4	17.3	6.3
21	Pyl	0.0	00	0.0	0.0	0.0	0.0	0.0	0.0
22	Sec	0.0	00	0.0	0.0	0.0	0.0	0.0	0.0

The amino acid composition and various chemical and physical parameters (Table2) reveals that membrane (15.8%), envelope (18.7) and spike (9.2%) protein is rich in Leucine and nucleocapsid protein is rich in glycine (10.3%).In all the protein selected shows pyrolysin and selenocystein is totally absent. The absence of pyrolysin and selenocystein indicate the virus is more thermo liable. The amino acids pyrolysin and selenocystein are also less in MERS coronavirus but it is very abundant in Archae bacteria *Methanosacrcina barhkeri.*[33,34] The lack of cysteine in nucleocapsid and poor percentage in in other protein indicate the less stability of the virus due to lack of disulphide bonds. Leucine is more abundant in all protein except nucleocapsid protein. It is reported that in HIV Nef protein contain Leucine motif down-regulated CD4 from the cell surface and enhanced viral replication, the same mode action may be involved in SRAS Cornavirus-2 (COVID19), hence it is more extended in membrane proteins. The abundance of serine (8.0%) in spike protein may be due to the fact that the virus multiplication by lytic cycle is predominant as in Epstein bar virus.[35]

The physico-chemical parameters (Table3) showed that alpha helix is more, but 3_{10}helix, pi Helix, beta helix is nil, an indication instability of the proteins-N, M, C and S. It is supported by the lack of end regions. Extended strands and beta turns is also comparatively less. The only predominant one is random coils to give medium stability for the concerned proteins.

Table 3: Physical parameters-1 of the selected proteins (SOPAM).

No	Protein	Parameters Percentage*							
		Hh %	Gh%	Ii%	Bb%	Ee%	Tt%	Ss%	Cc%
1	Nucleocapsid Phosphoprotein	21.24	0.00	0.00	0.00	16.71	6.92	0.00	55.13
2	Membrane glycoprotein	34.68	0.00	0.00	0.00	21.17	6.76	0.00	37.39
3	Envelope protein	44.00	0.00	0.00	0.00	26.67	9.33	0.00	20.00
4	Surface glycoprotein	55.59	0.00	0.00	0.00	15.47	2.87	0.00	26.07

** Hh (Alpha helix), Gh (3_{10} Helix), Ii (pi Helix), Bb (Beta Helix), Ee(Extended Strand), Tt(Beta Turn), Ss(bend Region), Cc(random Coil)*

The other physico-chemical parameters of proteins-membrane, nucleocapsid and envelope proteins are positively charged and spike protein is negatively charged, this is response to abundance of respective amino acids (Table4). EC is more in M, N and E protein but less in S protein. Instability index is less than 40 in membrane and envelope protein comparing to N

and S protein, it shows the M and E proteins are less stable[22], it may be interpret that the virus is less stable in environment. The aliphatic index of a protein is an index of relative volume occupied by aliphatic side chains of the amino acids- alanine, Valine, Leucine and isoleucine. An increase in the aliphatic index increases the thermo stability of globular protein. In the present study the COVID 19 shows low AI in spike protein and nucleocapsid protein comparing to E and M protein is very less but comparing to other thermophile bacteria it is too less. The present study showed that M and E protein have more AI and hence more heat stable and N and S protein have low aliphatic index and hence less stable (Table4). Grand Average Hydrophobicity (GRAVY) shows all proteins are hydrophobic with GRAVY below zero except envelope protein which has more than one GRAVY (Table4). The hydrophobic residues were predicted by helical wheel prediction by Pepwheel Programme (Fig.8a-c).

Table 4: Physical parameters-2 of the selected proteins.

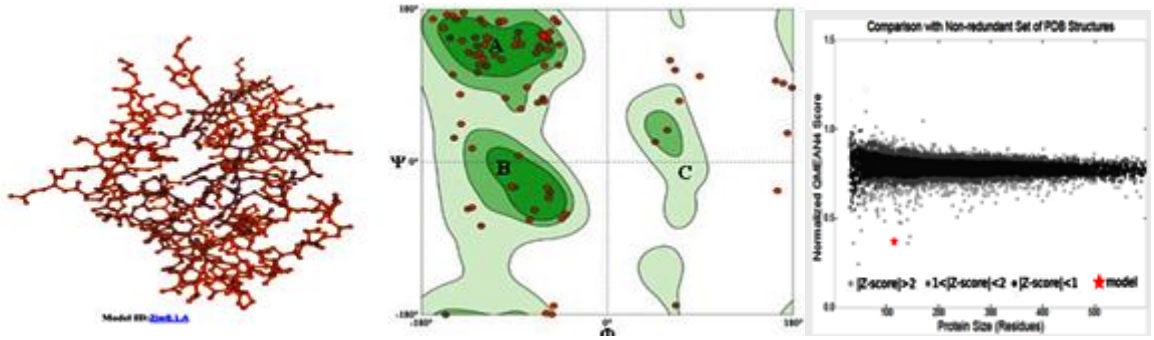
No	Protein	Parameter**							
		Mol.wt	pI	No.of (-)ve charged residues	No.of (+)ve charged residues	EC	II	AI	GRAVY
1	Nucleocapsid Phosphoprotein	45625.70	10.07	(Asp+Glu): 36	(Arg+Lys): 60	43890	55.09	52.53	-0.971
2	Membrane glycoprotein	25146.62	9.51	(Asp + Glu): 13	(Arg + Lys): 21	52160	39.14	120.86	0.446
3	Envelope protein	8365.04	8.57	(Asp + Glu): 3	(Arg + Lys): 5	5960	38.68	144.00	1.128
4	Surface glycoprotein	38625.08	5.51	(Asp + Glu): 37	(Arg + Lys): 29	36900	41.75	92.98	-0.051

Table 5: Global Quality of 3D model of the proteins generated by using Swiss-Model.

No	Protein	ID	QMEAN	Cβ	All atoms	Solvation	Torsion	Sequence Identity
	nucleocapsid Phosphoprotein	2jw8.1.A	-7.62	-2.25	-2.76	-2.40	-6.70	91.23%
	membrane glycoprotein	6qpk.1.A	-6.09	-2.43	-3.43	-9.39	-2.45	91.24%
	surface glycoprotein	6vyb.1.A	0.15	1.03	3.09	-0.72	0.10	91.26%

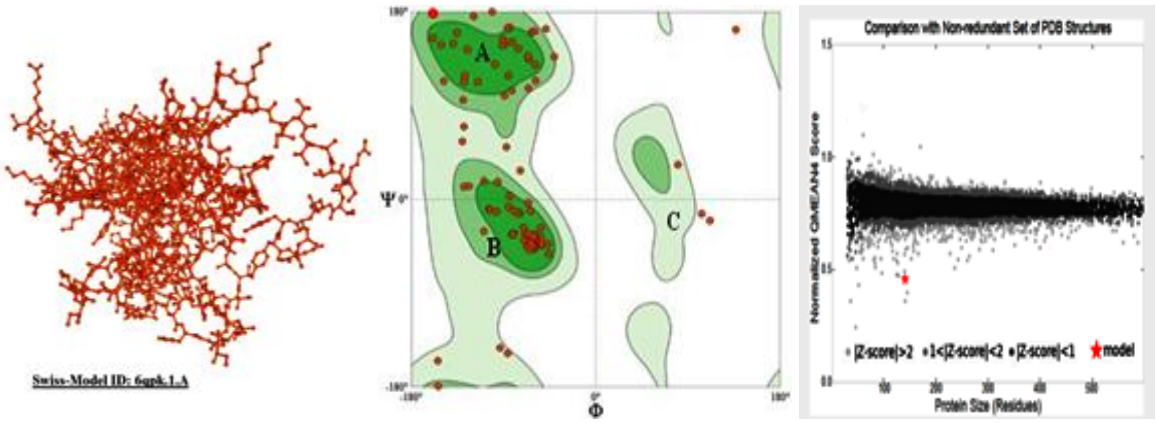
For more clarity about the proteins of COVID19 we generated 3D model (Fig 1-3-a,b&c) by using SWISSMODELLER and YASAR software (Fig1 & 2), it revealed that the protein N and M shows model score lower than experimental structure s on average. The QMEAN (Qualitative Model Energy Analysis) score of S protein shows that the model scores higher

than the experimental structure on average. The negative Cβ score for N and M protein and Positive close to '0' indicate that the atomic distance is in acceptable distance as also in all atoms score. Solvation score and torsion score indicate the accuracy of QMEAN and Cβ scores in acceptable limits.



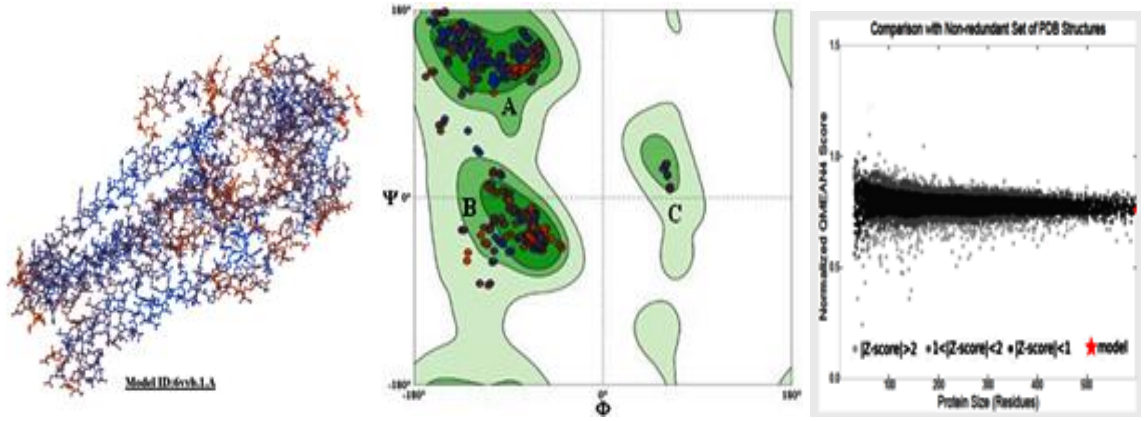
a- 3D structure **b:Ramchandran Plot** **c: 3D –validation**

Fig: 1-2jw8.1.A (N-protein).



a- 3D structure **b:Ramchandran Plot** **c: 3D –validation**

Fig: 2-6qpk.1.A (M-protein)



a- 3D structure **b: Ramchandran Plot** **c: 3D –validation**

Fig: 3-6vyb.1.A (S-protein).

A further support for this finding is the Ramachandran plot (Fig: 1-3) also shows the three areas which is in alignment with findings of QMEAN, Cβ scores, all atoms score, torsion score and solvation data. The (Tables-5) shows the global qualities of 3D model, it is in acceptable range.

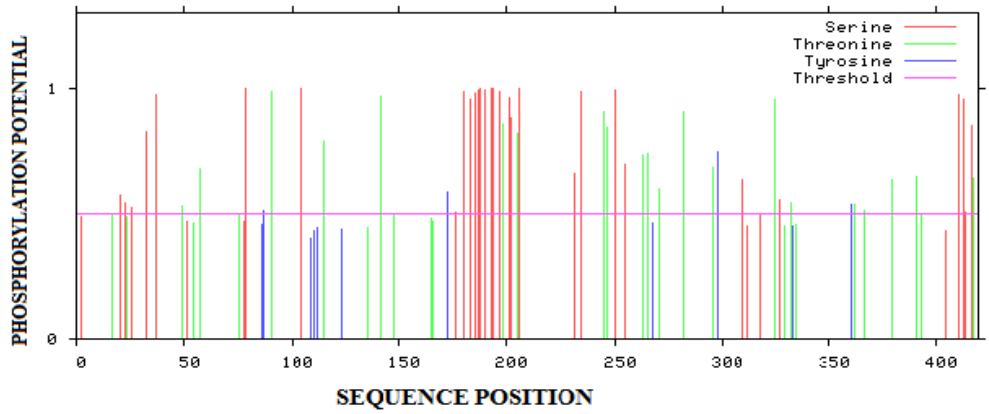


Fig4: Phosphorylation site in N-protein.

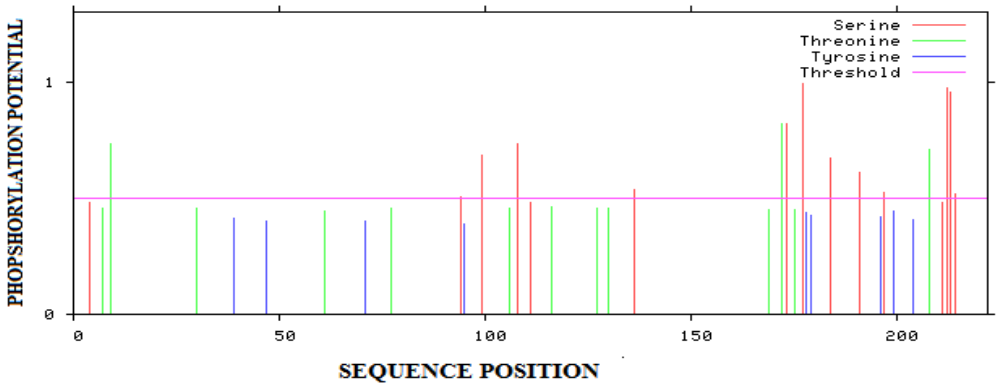


Fig5: Phosphorylation site in M-protein.

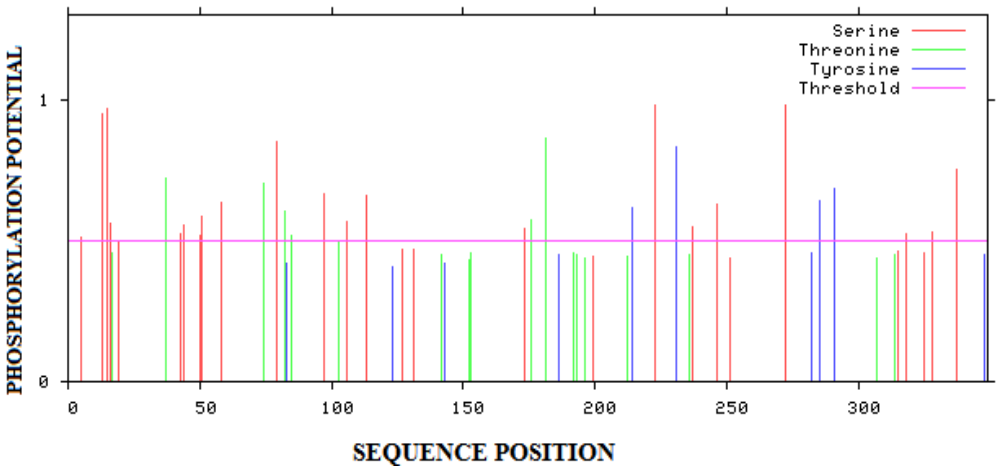


Fig: 6 Phosphorylation site in S protein.

Table 6: Transmembrane region in N- protein.

No	N-Terminal	Transmembrane region	C-Terminal	Type	Length
1	16	LLEQWNLVIGFLFLTWICLLQF	37	SECONDARY	22
2	47	YIIKLIFLWLLWPVTLACFVLAA	69	PRIMARY	23
3	79	GIAIAMACLVGLMWLSYFIASFR	101	PRIMARY	23

Table 7: Transmembrane region in M-protein.

No	N-Terminal	Transmembrane region	C-Terminal	Type	Length
1	11	TLIVNSVLLFLAFVVFLLVTLAI	33	PRIMARY	23
2	37	LRLCAYCCNIVNVSLVKPSFYVY	59	PRIMARY	23

Table 8: Transmembrane region in S-protein.

No	N-Terminal	Transmembrane region	C-Terminal	Type	Length
1	299	GLIAIVMVTIMLCCMTSCCSCLK	321	PRIMARY	23

The phosphorylation site in N, M and S proteins are also shown in the (Fig-4-6). The transmembrane regions of protein N, M and S are shown in the (Table: 7-9), the M and E proteins showed two transmembrane region and S protein showed one transmembrane region. The transmembrane regions correlate with the function of the proteins.

The glycosylation site predicted by NetNglyc (Fig7a-c) shows that N-protein is showed only one potential N-glycosylation site, M-proteins has no N-glycosylation site at all, S-protein showed four potential N-glycosylation sites.

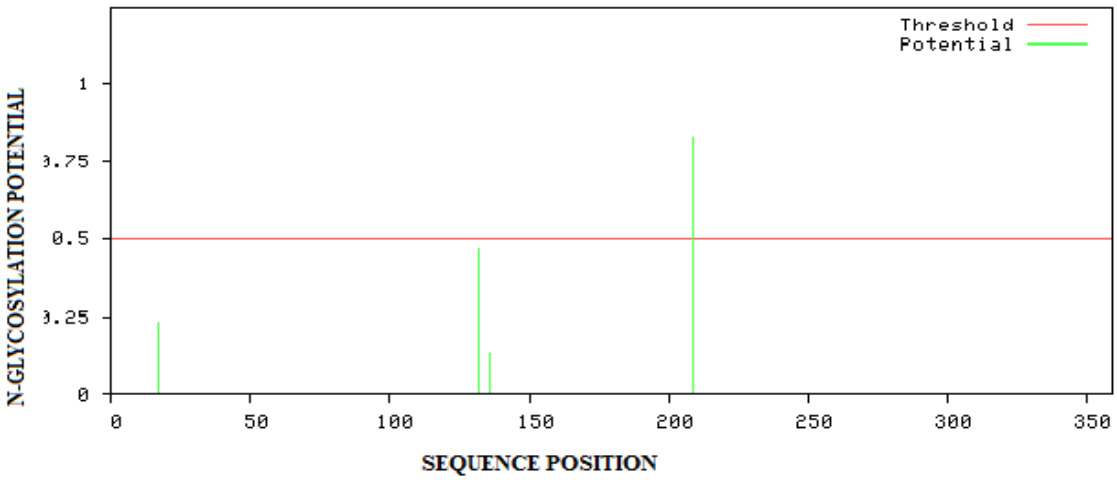


Fig: 7-(a) N-glycosylation site predictedd for Nucleocapsid protein (N)

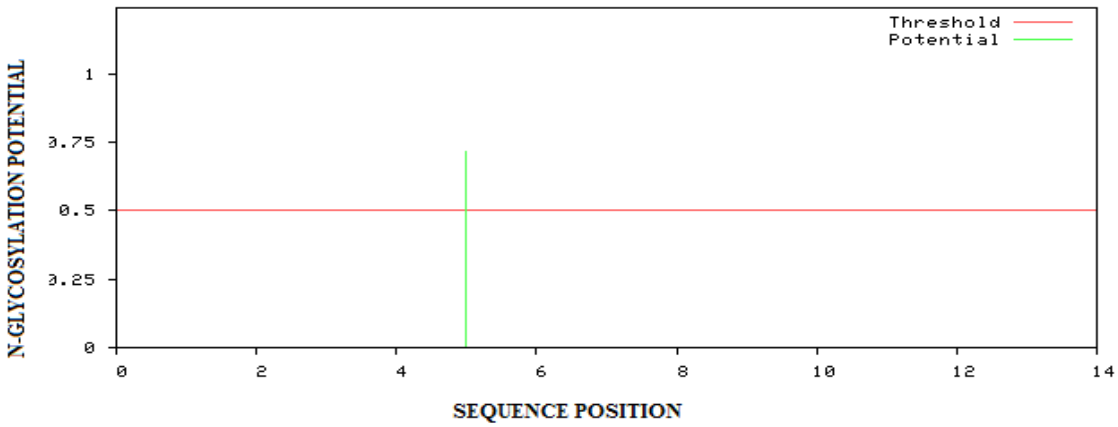
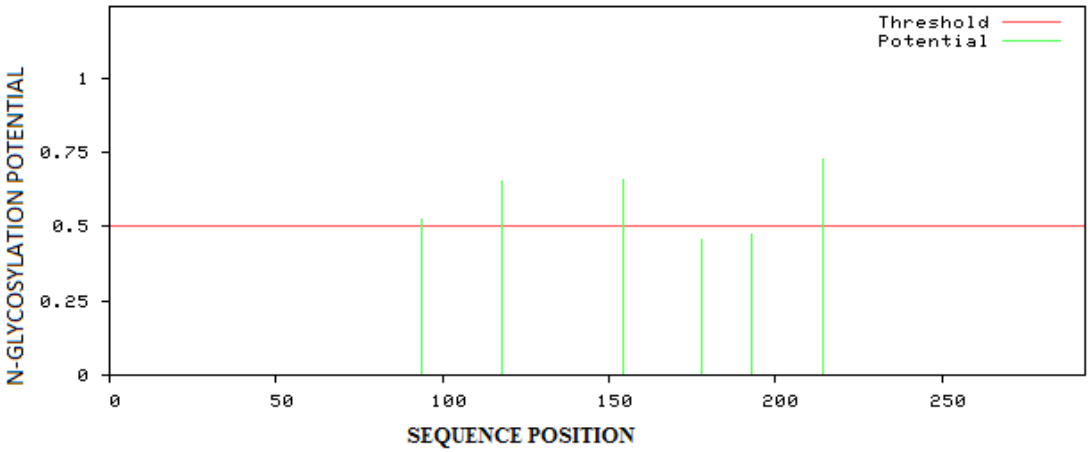


Fig (7b): N-glycosylation site predicted for Envelope protein (E).



Fig, (7c): N-glycosylation site predicted for Spike protein (S).

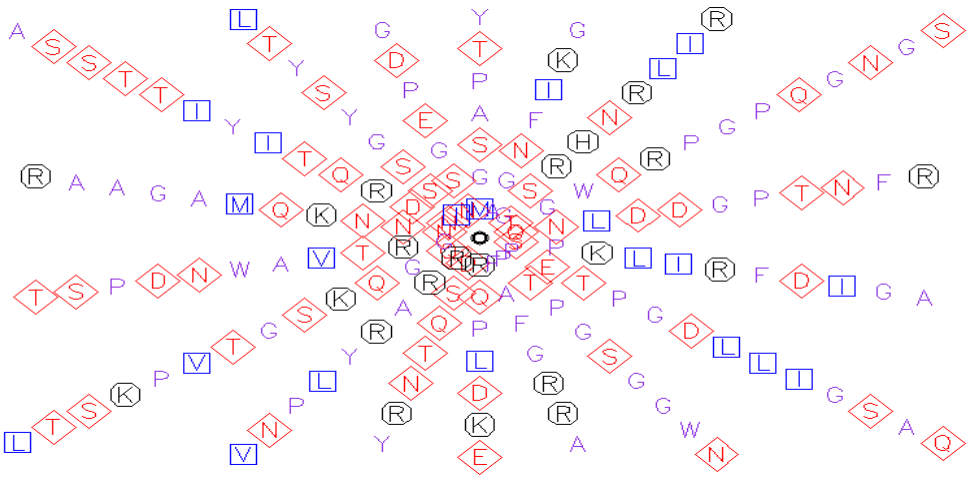


Fig. 8(a): Helical wheel predicted for Nucleocapsid protein.

Octagon-(+) vely charged, Square-Aliphatic residues, Diamond-(-) vely charged residues, Free letters-hydrophobic residues, Overrides- Amphipathic residues

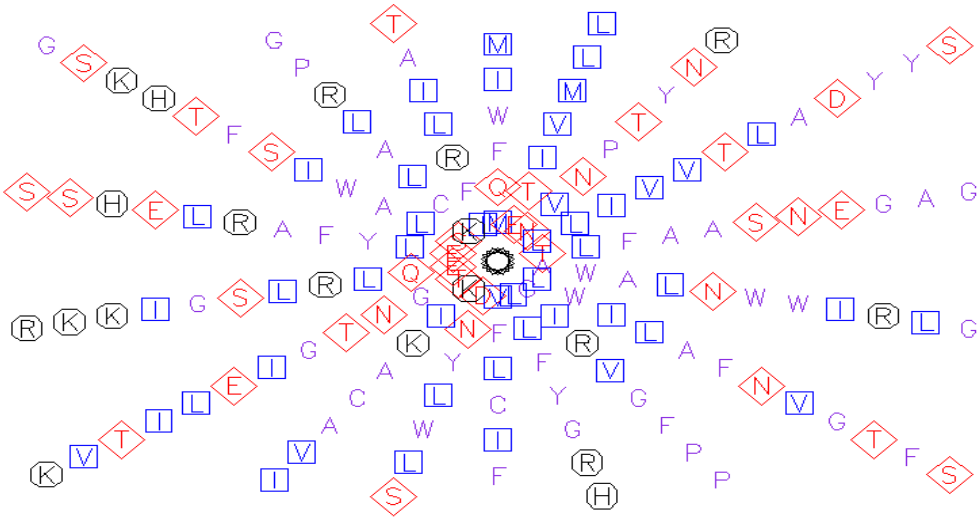
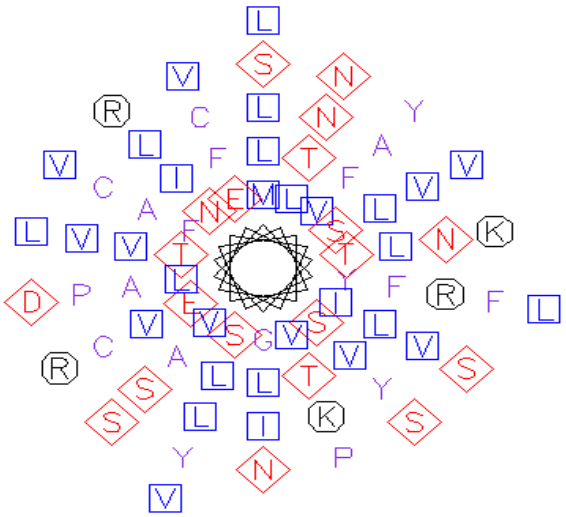


Fig. 8(b): Helical wheel predicted for membrane protein.



a-N-Protein **b- Membrane protein**

Fig 10: Helical Predicted by Pepwheel*

**Octagon-(+) vely charged, Square-Aliphatic residues, Diamond-(-) vely charged residues, Free letters-hydrophobic residues, Overrides- Amphipathic residues.*

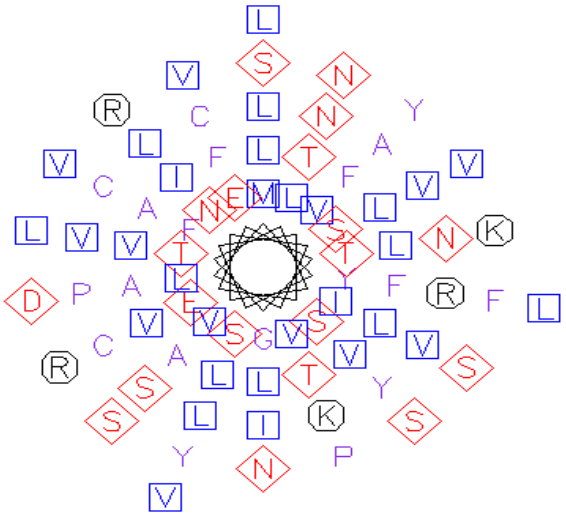


Fig-8(c) Helical wheel predicted for Envelope protein

**Octagon-(+) vely charged, Square-Aliphatic residues, Diamond-(-) vely charged residues, Free letters-hydrophobic residues, Overrides- Amphipathic residues*

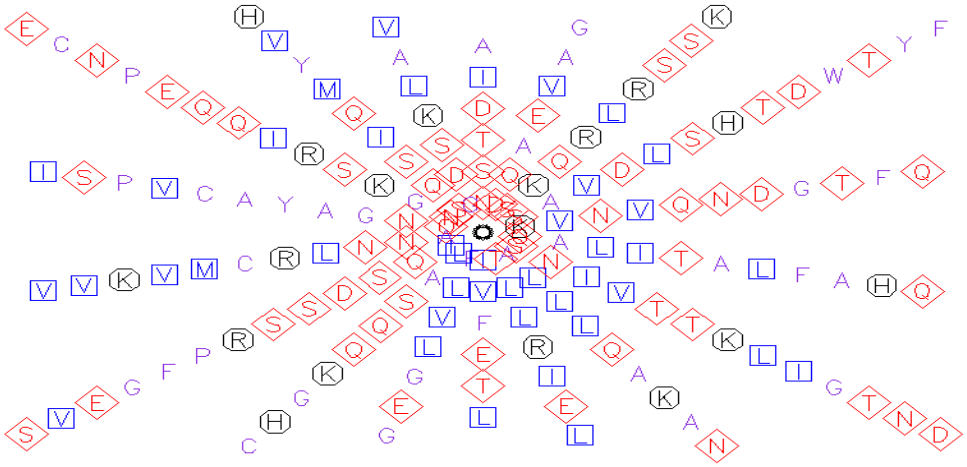


Fig. 10(d): Helical wheel predicted for Spike protein.

**Octagon-(+) vely charged, Square-Aliphatic residues, Diamond-(-) vely charged residues, Free letters-hydrophobic residues, Overrides- Amphipathic residues.*

DISCUSSION

A new coronavirus, commonly known as novel coronavirus or Severe acute respiratory syndrome Coronaviurs-2 (SARS Coronavirus-2), has been identified in Wuhan, China which causes a pneumonia outbreak in this area in late 2019.[36,8] This pneumonia like disease was later named as COVID19 (Coronavirus diesease19). This new virus belongs to the beta coronavirus family in which SARS coronavirus and SARS related bat coronavirus.[1,2] It is a single stranded RNA virus (+ssRNA) that can be isolated in different animal species.[37] It is speculated that the SARS coronavirus-2 originated form bat and transferred to human through

an intermediate host probably civet cat or pangolin. Various study demonstrated that the virus spread through droplet and people remain asymptomatic could transmit the virus. This fact suggests that isolation is the best method to curtain this epidemic. The incubation time is about 3 -7 days and up to 7 days.[1] The coronavirus are spherical to pleomorphic enveloped entities. The envelope have projecting glycoproteins, and surrounded by matrix protein enclosing a positive -sense RNA (Mr 6 x 10^6) associated with nucleoprotein. The envelope glycoproteins are responsible for infection to host cell. Thus the virus contain membrane protein (M), nucleocapsid protein (N), Envelope protein (E), spike protein and a positive strand sense RNA. In the study we extracted the Fasta sequence of all these protein and subjected to *in silico* study. The study shows the various amino acid compositions of these proteins, its various physico-chemical parameters, transmembrane strands, phosphorylation regions, and finally we generate the 3D model and validated the model. In our study we observed that structural protein is positively charged due to more arginine and lysine. The M-protein plays an important role in host interactions. It glycosylated at several region though it is not essential for viral infectivity.[38] Mutation in M protein at glycosylation site decrease the infectivity of the virus is proposed[39] but it is also demonstrated that the glycosylation state does not alter the replication in vitro. The present study shows that phosphorylation is more predominant in membrane protein and nucleocapsid protein and it is very less in spike protein, it may correlate that the viral infection is more depends on m and N protein rather than S protein. The coronavirus N protein is localized in nucleolus as well as to the cytoplasm.[40] These reports suggest that N protein induces a cell cycle delay or arrest, most probably in G_2-/ M phase by inhibition of cytokinin. Corona virus E protein is an integral membrane protein.[41] Our study showed that the M protein have three transmembrane strands, a proof for the integral nature of M protein. The corona spike virus protein is a type I glycoprotein that form peplomeres on the corona virus particle. The amino terminal S_1 subunit forms the globular head. The head contain a receptor binding domain (RBD) within first 330 amino acids.[42] The spike protein plays an important role in viral entry, cell- to cell spread and determining tissue tropism. The coronavirus entry is not pH dependent hence entry occurs directly at plasma membrane and not through endosomal route. But there are reports that SARS coronavirus enter through endosomal route because its entry is prevented by lysosomotropic agents. The plasma membrane route is enhanced by the cleavage in the spike due to[43] angiotensin converting enzyme (ACE) from the inflammatory cells present in the lungs. In the present study we proposed the hydrophobicity and instability index, extinction coefficient isoelectric points and composition of four different proteins[44]

mentioned earlier. The N-glycosylation site prediction shows that the N-glycosylation is on only few position and potential of N-glycosylation also very poor, the most potential is shown by spike protein. Glycosylation plays a key role in protein stabilization.[45,46] In addition glycosylation also plays a key role in stabilizing glycoprotein in microenvironment through alteration in half-life.[47] A further study and confirmation required to establish a final conclusion and designing of better treatment and control measures to fight with SARS Coronavirus 2 spreading and devastating the virus from the environment.

CONCLUSION

In silico analysis of the four proteins including membrane protein (M), Nucleocapsid protein (N), Envelope protein (E) and Surface (Spike) protein (S) of SRAS Coronavirus-2 (COVD19) reveals that the virus is positively charged due to aspartic acid and glutamic acid, but S-protein is negatively charged due to arginine and lysine residues. It may help to precipitate the virus particles by free radical like O^- and OH^- a method proposed to control the spreading of Coronavirus. The study shows that it is thermo liable due to the lack of some amino acids like pyrolysin and selenocystein, since these are abundant in Archae bacteria.[33,34]

Transmembrane strands are varied in different proteins studied. M protein has 3 transmembrane strands, E proteins have two and spike proteins have one. The spike protein shows comparatively more N-glycosylation though other proteins are very poor in N-glycosylation. So we here speculate that, though the virus is heat liable the spike protein is more stable in the microenvironment. A further study to substantiate this finding is required. One of the other predictions is that N-glycosylation is totally absent in M-protein. Nucleocapsid protein not has shown any transmembrane strands. Phosphorylation region is more in capsid protein and least in spike protein may correlate with functional status.

REFERENCES

1. Lu R, Zhao X, Li Juan, Niu P et al. Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*, 2020; 395: 565-574.
2. Chan JF-W, Yuan S, Kok K-H et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*, 2020; 395: 514-523.

3. Phan L, Nguyen T, Luong Q et al. Importation and human-to-human transmission of a novel coronavirus in Vietnam. *N Engl J Med.*, 2020; 382: 872-874
4. Li Q, Guan X, Wu P et al.Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia *Engl J Med.* 2020; (published online Jan 29.)
5. Chen N, Zhou M, Dong X et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*, 2020 395: 507-513.
6. WHO: Statement on the meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV).World Health Organization, Jan 23, 2020: https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)
7. Park SW, Bolker BM, Champredon D et al.(2020). Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus (2019-nCoV) outbreak. *medRxiv,* (published online Feb 7.) (Preprint).
8. Zhu N, Zhang D, Wang W, et al A Novel Coronavirus from Patients with Pneumonia in China, 2019a. *N Engl J Med.*, 2020; **294**(5): 1351-1362
9. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modeling study. *Lancet,* 2020a; 1-3.
10. He F, Deng Y, Li W. Coronavirus Disease 2019 (COVID-19): What we know? *J Med Virol.*, 2020; 2019: 0-2.
11. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *Jama.*, 2020b; 2019: 3-6.
12. Zhou P, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, pmid: 32015507.
13. Gallagher T.M and Buchmeier. M.j. Coronavirus spike proteins in viral entry and pathogenesis. *Virology,* 2001; 279: 371–374.

14. Simmons G, Zmora P, Gierer S, Heurich A and Pöhlmann S, *Proteolytic activation of the SARS-coronavirus spike protein: Cutting enzymes at the cutting edge of antiviral research. Antiviral Res., 2013; 100: 605–614.*
15. Li F, Li W, Farzan M and Harrison S.C, Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science, 2005; 309: 1864–1868.*
16. Belouzard S, Chu V.C and, Whittaker G.R. Activation of the SARS coronavirus spike protein via sequential Proteolytic cleavage at two distinct sites. *Proc. Natl. Acad. Sci. U.S.A.*, 2009; *106: 5871–5876.*
17. Millet J.K and Whittaker G.R. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Res.*, 2015; 202: 120–134.
18. Simmons G, Gosalia D.N, Rennekamp A.J, Reeves J.D, Diamond S.L and Bates P, Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc. Natl. Acad. Sci. U.S.A., 2005; 102: 11876–11881*.
19. Hoffmann M, Kleine-Weber H, Krüger N,Müller M, Drosten C and Pöhlmann S. (2020). The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells. BioRxiv .01.31.929042 [Preprint]. 31 January.
20. Guruprasad, K., Reddy, B.V.B. and Pandit, M.W. 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 Eng.*, 1990; 4: 155-161.
21. Ikai, A.J. hermo stability and aliphatic index of globular proteins. *J. Biochem*, 1980; 88: 1895-1898.
22. Kuba K, Imai Y Rao S, Gao H et al. A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat. Med.*, 2005; 11: 875–879.
23. Gill SC and von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem*, 1989; 182: 319-326.
24. Arnold K, Bordoli L., Kopp J, and Schwede T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. *Bioinformatics,* 2006; **22**: 195-201.
25. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*, 2009; **37:** D387-D392.
26. Arnold K, Bordoli L., Kopp J, and Schwede T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. *Bioinformatics,* 2006; **22**: 195-201.

27. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*, 2009; **37:** D387-D392.
28. Peitsch C.M. Protein Modeling by E-mail, Nature Biotechnology, 1995; 13(7): 658-660.
29. Kyte, J and Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, 1982; **157:** 105-132.
30. Blom, N, Gammeltoft, S, and Brunak, S. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology*, 1999.
31. Hirokawa, T., Seah, B.-C. and Mitaku, S. SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics, 1998; 14: 378–379.
32. Mitaku, S., Ono,M., Hirokawa,T., Seah, B.-C. and Sonoyama, M. Proportion of membrane proteins in proteomes of15 single-cell organisms analyzed by a prediction system SOSUI.*Biophys. Chem*, 1999; 82: 165–171.
33. Mitaku1S, Hirokawa Tand Tsuji T. Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane–water interfaces. *Bioinformatics*, 2002; 18(4): 608–616
34. Ramachandran, G.N.; Sasiskharan, V. "Conformation of polypeptides and proteins". *Advances in Protein Chemistry*. Advances in Protein Chemistry, 1968; **23**: 283–437.
35. Srinivasan G, James CM, Krzycki JA. Pyrrolysine encoded by UAG in Archaea: charging of a UAG- decoding specialized tRNA 296 (5572). *Science*, 2002; 1459–1462.
36. Hao.B, Weimin G, Tsuneo K. F, Carey M. J, Joseph A. K, Michael K. C. A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase, 2002; 296 (5572). *Science*, 1462–1466.
37. Amy F, Tobias R, Lyn G, Lee, Ayman E.G, Yoshimi E and George M. Amino Acid Substitutions Reveal Distinct Functions of Serine 186 of the ZEBRA Protein in Activation of Early Lytic Cycle Genes and Synergy with the Epstein - Barr virus R Transactivator, *J. Virol*, 1999; **73**(6): 4543-4551.
38. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. Washington State 2019-nCoV Case Investigation Team. Washington State 2019-nCoV Case Investigation Team. First case of 2019 novel coronavirus in the United States. *N Engl J Med*., 2020; 382: 929–36.
39. Perlman S, Netland J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat. Rev. Microbiol.* Jun, 2009; 7(6): 439-50.
40. Haan de, C. A., M. de Wit, L. Kuo, C. Montalto-Morrison, B. L. Haagmans, S. R. Weiss, P. S. Masters, and P. J. Rottier. The glycosylation status of the murine hepatitis

coronavirus M protein affects the interferogenic capacity of the virus in vitro and its ability to replicate in the liver but not the brain. *Virology*, 2003; 312**:** 395-406.

41. Laude, H., J. Gelfi, L. Lavenant, and B. Charley. Single amino acid changes in the viral glycoprotein M affect induction of alpha interferon by the coronavirus transmissible gastroenteritis virus. *J. Virol*., 1992; 66**:** 743-749.
42. Wurm, T., H. Chen, T. Hodgson, P. Britton, G. Brooks, and J. A. Hiscox**.** Localization to the nucleolus is a common feature of coronavirus nucleoproteins, and the protein may disrupt host cell division. *J. Virol*., 2001; 75**:** 9345-9356.
43. Yu, X., W. Bi, S. R. Weiss, and J. L. Leibowitz. Mouse hepatitis virus gene 5b protein is a new virion envelope protein. *Virology*, 1994; 202**:** 1018-1023.
44. Kubo, H., Y. K. Yamada, and F. Taguchi. Localization of neutralizing epitopes and the receptor-binding site within the amino-terminal 330 amino acids of the murine coronavirus spike protein. *J. Virol*, 1994; 68**:** 5403-5410.
45. Simmons, G., J. D. Reeves, A. J. Rennekamp, S. M. Amberg, A. J. Piefer, and P. Bates**.** Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. *Proc. Natl. Acad. Sci*. USA., 2004; 101**:** 4240-4245.
46. Ashokan .K.V, Mundaganore.D.S and Mundaganore. Y.D. *In silico* validation of Middle East respiratory syndrome (MERS) virus proteins for better drug development. International Journal of Applied Science and Biotechnology, 2013; 1(4): 272-278.
47. Bechor S. D.Levy. Y. (2008). Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *PNAS 10582568261*.
48. Bechor S.D. Levy.Y. Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. *Curr Opin Struct Biol*., 2009; 1952453.
49. Arey B.J. (2012). The Role of Glycosylation in Receptor Signaling, Glycosylation, Stefana Petrescu, Intech Open, DOI: 10.5772/50262. Available from: https://www.intechopen.com/books/glycosylation/the-role-of-glycosylation-in-receptor-signaling