International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

In silico comparative analysis "In silico comparative analysis of spike proteins of seven variants of Coronavirus"

[1a] Sajalendu Ghosh*, [2a] Medha Kumari, [3a] Sandip Mandal, [4b] Koel Mukherjee
[1a] *Asst. Professor, Department of Zoology, Dr. Shyama Prasad Mukherjee University, Ranchi, Jharkhand, India*
[2a] *Research scholar, Department of Zoology, DSPMU, Ranchi,*
[3a] *Research scholar, Department of Zoology, DSPMU, Ranchi, Jharkhand*
[4b] *Assistant Professor, Department of Biotechnology, Birla Institute of Technology, Mesra, Ranchi*

Submitted: 09-01-2023 Accepted: 19-01-2023

ABSTRACT

INTRODUCTION- In the last two decades, Corona Virus outbursts have threatened the world thrice: SARS in 2002, MERS in 2012, and again SARS in 2019. The present study focuses on investigating the differences and similarities in the spike protein of variants of coronaviruses using the In-Silico method. Seven variants have been selected, which include SARS-CoV-2002, MERS, SARS-CoV-2019, OMICRON, BETA, DELTA, and GAMMA. The amino acid sequence of the variants was retrieved in FASTA format from the Protein database of NCBI and analyzed for primary, secondary, and tertiary structures using different In-Silico tools.

RESULT- Analysis of the physiochemical parameters such as instability index (30.42-36.60) and aliphatic index (80.42 to 83.32) had shown the stable and hydrophobic behavior of the spike protein in the variants. The higher percentage of secondary structures alpha helix (ranges from 29.81% -26.15%) aids in its stability. Tertiary structure prediction represents the differences in the sequence number of domains in the spike protein.

CONCLUSION- The variation in the retrieved data of variants helps to understand the virulence and transmission of the coronavirus. Such as MERS's highest stability can be a possible reason for its lethality. Omicron having medium stability can spread fast and is less lethal than MERS. Delta and Gamma stability is somewhat similar to MERS and that's why it possibly shows high mortality rate than omicron.

Keywords: Corona Virus, spike protein, N-terminal domain,C-terminal domain, Heptad Repeats

I. INTRODUCTION:

In the last two decades Corona Virus outbreak had threatened the world thrice; SARS in 2002, MERS in 2012, and again SARS in 2019 which has its impact till the present date.

SARS CoV-19 has been detected in Wuhan, China for the first time in December 2019 (Taherizadeh M., et.al., 2020), after diagnosing the causative agent of a sudden upsurge of Pneumonia-like symptoms in a wide range of people (Shu Y., et.al., 2021).

In the first half of 2020, the novel Coronavirus was named 'Severe Acute Respiratory Syndrome Corona Virus-2' (SARS-CoV-2), and its severity was declared a global pandemic by the World Health Organization (Aleem A., et.al., 2022).

COVID-19 and influenza viruses have similar disease presentations as they cause respiratory disease with fever. Although the influenza virus can spread faster than COVID19 mortality for COVID-19 appears higher than for influenza, especially seasonal influenza. (https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf as accessed on10/9/2022).

Coronavirus had attacked human beings previously but not so severely as this time in 2019. Most coronaviruses that are pathogenic to humans were associated with mild clinical symptoms except for two namely SARS-CoV (severe acute respiratory syndrome coronavirus), and MERS Middle East respiratory syndrome coronavirus (MERS-CoV). SARS-CoV emerged in Southern China, in 2002, and resulted in 774 deaths while MERS-CoV was first detected in Saudi Arabia in 2012 and was responsible for 858 fatalities in the

International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

Middle East and 38 deaths in South Korea (Lu R. 2020).

SARS-CoV-2 is a zoonotic disease that first spread in animals such as bats and then transferred to human beings (Zhang J., et.al., 2021). Genomic analysis revealed that SARS-CoV-2 of 2019 was phylogenetically related to severe acute respiratory syndrome-like (SARS-like) bat viruses. Thus bats have been considered the possible primary site of origin. Bats are unlikely to be the animal that is directly responsible for the transmission of the virus to humans for several reasons (Lu R.2020) but the culprit intermediate source of origin and transfer to humans is not clear, though the rapid human-to-human transfer with strain-specific infectivity has been established extensively (Shereen M.A. 2020).

SARS-CoV-2 is an enveloped and positive-sense single-stranded RNA (+ssRNA) virus that belongs to the family Coronaviridae, subfamily Coronavirinae, genera Betacoronavirus together with two highly pathogenic viruses, SARS-CoV and MERS-CoV. SARS-CoV-2 is considered a novel human-infecting coronavirus genetically distinct from SARS-CoV (with about 79% similarity) and MERS-CoV (Harapen H 2020). To date there is seven types of coronaviruses were detected that cause respiratory syndrome in humans: these are HCoV-OC43 and HCoV-HKU1 which belong to lineage A of genera beta Corona Virus, HCoV-229E and HCoV-NL63 of genera alpha Corona Virus (Lim Y., et.al., 2016), SARS-CoV 2002 and MERS-CoV and SARS-CoV-19 from the lineage B and C of beta Corona Virus.

SARS-CoV-2 is a highly spreading coronavirus that has outburst into many countries only in a few months and is responsible for the death of more than 5 million people (Negi S., et.al., 2022). The common symptoms of the disease are severe pneumonia, cold, cough, diarrhea, fever, sore throat, shortness of breath, and loss of smell and taste (Alberca G., et.al., 2021). Even asymptomatic patient has also been recognized in SARS-CoV and MERS-CoV. These patients are more dangerous as they spread it to many unknowingly.

As per an Epidemical update by the WHO as of December 11, 2021, there are 5 types of SARS –CoV-2 variants identified since the start of the pandemic (Aleem A., et.al., 2022) namely: ALPHA (B.1.1.7), BETA (B.1.351 lineage), GAMMA (P.1 lineage), DELTA (B.1.617.2 lineage), and OMICRON (B.1.1.529 lineage).

Worldwide Severely pathetic loss of human life, and devastating economic and social disruption raised before humanity an unprecedented challenge to public health. The far-reaching impacts of COVID-19 on the entire humanity have initiated ordering diverse fields to work in coordination to prevent similar pandemics in near future. Promoting an integrated understanding of the complex interconnections and mutual dependencies between wildlife and people, knowledge of the molecular biology and bioinformatics on the mode of viral infection through common human cell receptors to reduce the risk of future virus-mediated pandemics are prime needs of this particular post-pandemic time. (https://www.fao.org/documents/card/en/c/cb1163en accessed on 10/9/2022).

It is a well-known fact that virus continuously mutates and especially in the spike protein gene sequences (Tian D., et.al., 2022). This part of the viral protein is involved in the recognition of host cell receptors and mediates a high infection rate and is the major target of the humoral immune response. This research, on the comparative analysis of all seven variants of Coronavirus, with retrieved spike protein sequences, was to get a molecular insight of mutation and related changes in the strains with their consequent change in infectivity and virulence with the help of bioinformatics.

II. MATERIALS AND METHODS:

1. Retrieval of Amino Acid Sequence:

Amino acid sequences of spike proteins of all seven selected variants of COVID viruses (SARS - CoV 2002, MERS, SARS - CoV 2019, DELTA, GAMMA, BETA, and OMICRON) had been collected in FASTA format from the NCBI-BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Each amino acid sequence was analyzed for primary, secondary and tertiary structure with the help of various servers and tools. A table was prepared containing name of the proteins, accession numbers and amino acid sequence lengths.

2. Comparison of Primary Sequences:

Primary sequence data of each amino acid of all the selected variants were retrieved through the EXPASY server using Protparam tool (https://web.expasy.org/protparam/). The physiochemical parameters such as composition of different amino acids, theoretical pI (isoelectric focusing point), extinction coefficient, half-life, instability index, aliphatic index and GRAVY

International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

(grand average of hydropathy) had been recorded and compared.

3. Secondary Structure Prediction:

Using SOPMA (Self-Optimized Prediction Method with Alignment), the secondary structure data of each spike protein was collected (https://npsa-prabi.ibcp.fr/cgibin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) and examined for different conformational properties like alpha helix, beta turn, extended strand and random coil.

4. Domain and active site comparison:

After predicting the secondary structure of spike proteins, domains and active sites were examined with the help of Scanprosite tool available online (https://prosite.expasy.org/scanprosite/). Active sites are the part of protein that help it to bind to other molecules, ions etc. The active domains and motifs as well as the disulfide bonds present were analyzed for the spike proteins that are depicted in the table 5 and 6.

5. Modelling of Phylogenetic Tree:

Mega-11 tool was used to do phylogenetic analysis of the SARS-CoV-2002, MERS,SARS-COV-2019, DELTA, BETA, GAMMA, and OMICRON strains (https://www.megasoftware.net/dload_mac_beta).

III. RESULTS:

Retrieval Of Amino Acid Sequence

Amino acid sequence of spike proteins (S proteins) of all the selected seven variants of corona virus have been collected from the protein database of the NCBI site(https://blast.ncbi.nlm.nih.gov/Blast.cgi) and shown in table 1. The selected corona virus variants with their accession number mentioned in the parenthesis were SARS-CoV 2002 **(**AAU81608.1), MERS **(**YP_009047204.1**)**, SARS-CoV 2019 **(**QOU86714.1**)**, DELTA**(**PDB: 7TOU_A), GAMMA **(**PDB: 7M8K_A), BETA **(**PDB: 7LYK_A), and OMICRON **(**7QO9_B). The number of amino acids of all the variants was measured between 1238 – 1353 amino acid residues.

Table 1: Representing the accession number and no of amino acid of the retrieved spike protein sequence.

NAME	AccessionNumber	Amino acid number	Reference
SARS-CoV-2002	5XLR_1	1203	Gui M, et.al., 2016
MERS	5W9P_1	1329	Pallesen J, et.al., 2017
SARS-CoV-2019	6VXX_1	1281	Walls AC, et.al., 2020
OMICRON	7WVO_1	1258	Hong Q, et.al., 2022
BETA	PDB: 7LYK_A	1277	Gobeil SM, et.al., 2021
DELTA	PDB: 7TOU_A	1275	Gobeil SM, et.al., 2022
GAMMA	PDB: 7M8K_A	1238	Wang P, et.al., 2021

Table 1 showing Name of different strains of Corona Virus their accession number, total number of amino acid present, with the reference of published articles are given in this table.

Primary Sequence Comparison

The primary structures data retrieved from ProtParam gives the physicochemical parameters (which include Instability index, aliphatic index, theoretical pI, molecular weight, extinction coefficient, half-life, and GRAVY value) of all the seven variants is shown in table 2 and graph 1-5. Physicochemical analysis of Corona virus variants can help to investigate the similarities anddissimilarities of the physical and chemical properties of the spike proteins. The pI value or the Isoelectric point is the point where the positive charges in the protein become equal to the negative charges showing no net charge on the protein. Theoretical pI value of 7 different variants was found slightly acidic in the range of 5.43 to 6.52 and was different for all the variants. Average molecular weight of all the selected variants was found to be 137974.4, with highest molecular weight of MERS corona virus (146217.98) and lowest of SARS-CoV-2002 variant (133568.34). Half-life is the time taken for a protein to degrade by 50% of its concentration in a cell, whether in vitro or in vivo. Half-life of nearly all the strains

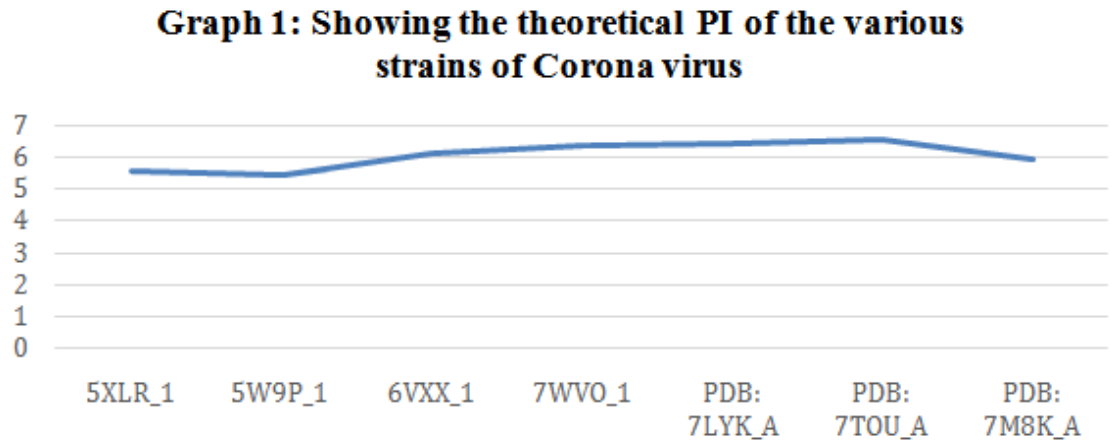
International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

wasobserved to be 30 hours in human reticulocytes.The instability index indicates proteins in vitro stability if the value comes under 40. Instability index of all the variants have been detected between the 30.21- 36.15, means all the spike proteins are stable enough. The volume that the protein's aliphatic side chains occupy is known as the aliphatic index, and it indicates thermo stability of the protein. The aliphatic index of the seven variants ranged from 80.42 to 83.32, indicating that all the proteinsare thermostable. The hydropathy scale known as GRAVY value (grand average hydropathy) is theoverall average of protein's hydrophobicity and hydrophilicity. All our experimental strains have negative GRAVY values, indicating the non-polar and hydrophobic behavior of the S proteins in general.

Table2:Physiochemical properties of retrieved amino acid sequences using ProtParam

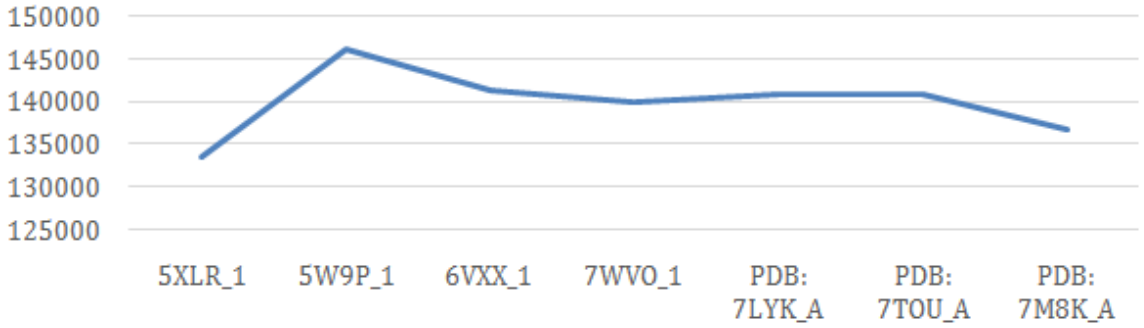
Common Name	Theoretical pI	Molecular Weight	Half life	Instability Index	Aliphatic Index	GRAVY
SARS-CoV-2002	5.55	133568.34	30 hours	30.21	81.36	-0.098
MERS	5.43	146217.98	30 hours	36.15	82.01	-0.095
SARS-CoV-2019	6.09	141410.94	30 hours	31.26	83.32	-0.139
OMICRON	6.31	140082.63	30 hours	32.52	82.90	-0.170
BETA	6.38	140928.01	30 hours	31.33	80.83	-0.169
DELTA	6.52	140842.83	30 hours	30.42	80.42	-0.199
GAMMA	5.94	136837.63	30 hours	31.32	82.75	-0.135

Table 2 showingdifferent physicochemical properties like theoretical pI (TpI), molecular weight of the spike protein, half-life, instability index, aliphatic index and grand average hydropathy values of different strains of Corona Virus. MERS showing lowest values of TpI while maximum value of GRAVY.

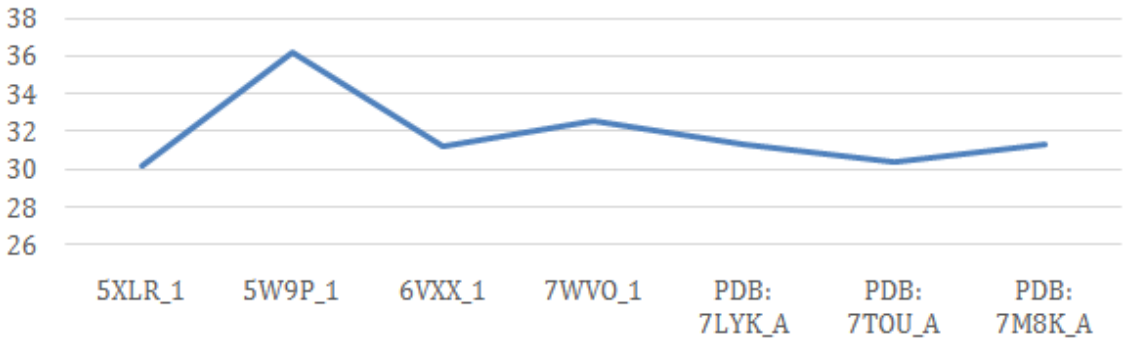


International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

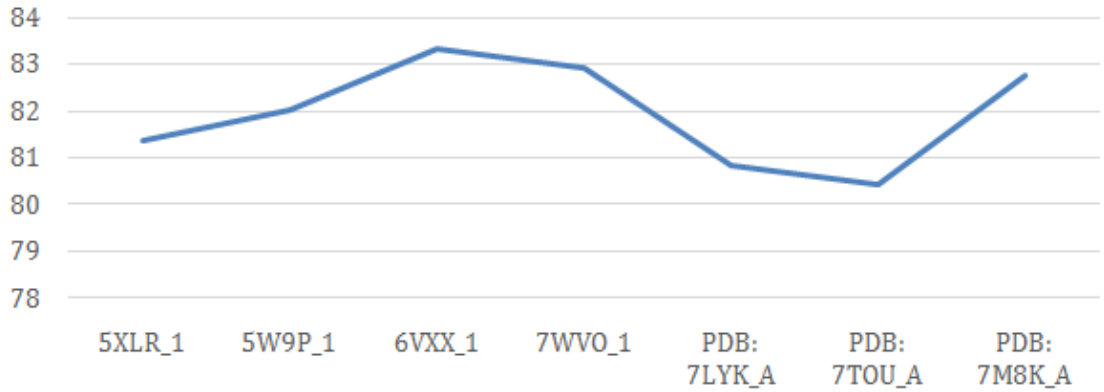
Graph 2: Showing the molecular weight of the various strains of Corona virus



Graph 3: Showing the instability index of the various strains of Corona virus

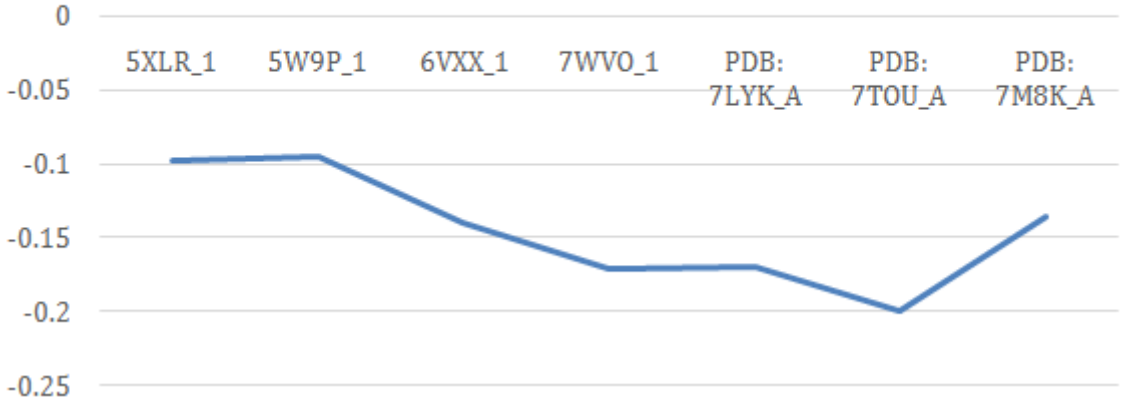


Graph 4: Showing the aliphatic index of the various strains of Corona virus



International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

Graph 5: Showing the GRAVY of the various strains of Corona virus



Secondary Structure Prediction

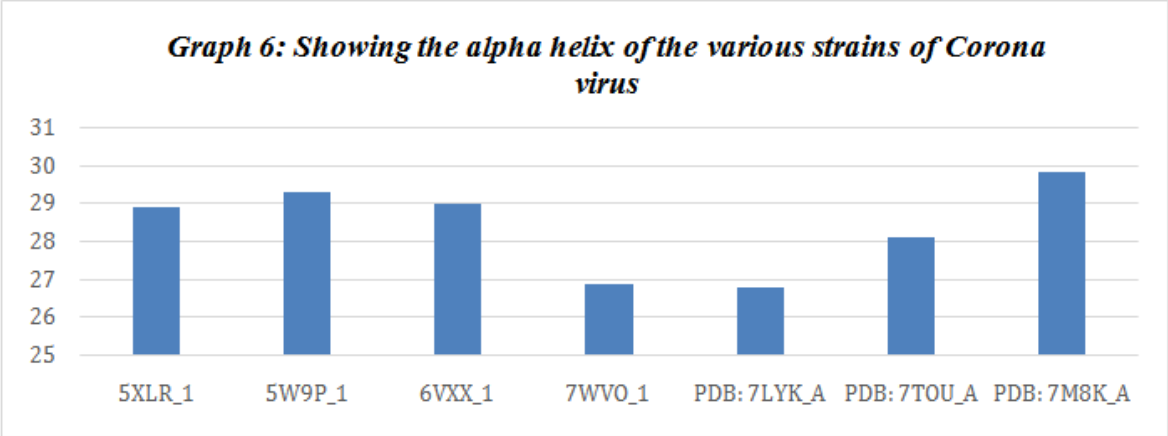
The secondary structure investigation helped to predict the major conformation of the spike protein of seven variants. Stability of the protein can be recognized via percentage alpha helix, beta- turns, and extended strands. All the variants showed different percentages of alpha helix (26.87%-29.81%-), extended strains(20.55%-23.53%) and beta-turns (3.39 %-5.42%). Highest percentage of alpha helix was detected in Gamma variants followed by MERS, SARS-CoV-2019, SARS-CoV-2002, DELTA, OMICRON AND BETA. High percentage of random coils (ranges between 44.78%-46.82%) in our data depicted that the major portion of the S protein was present in an unfolded state and so could be considered as the unstable part(Leikina, E, et.al, 2002). Variation in the amino acid sequences involved in the formation of secondary structure in the seven variants have been represented in the following table 3 and graph 6-9.The composition and percentage of 20 amino acids in the spike protein of seven variants have been shown in table 4. Graph 10 shows differences in the amino acid percentage of several amino acids.

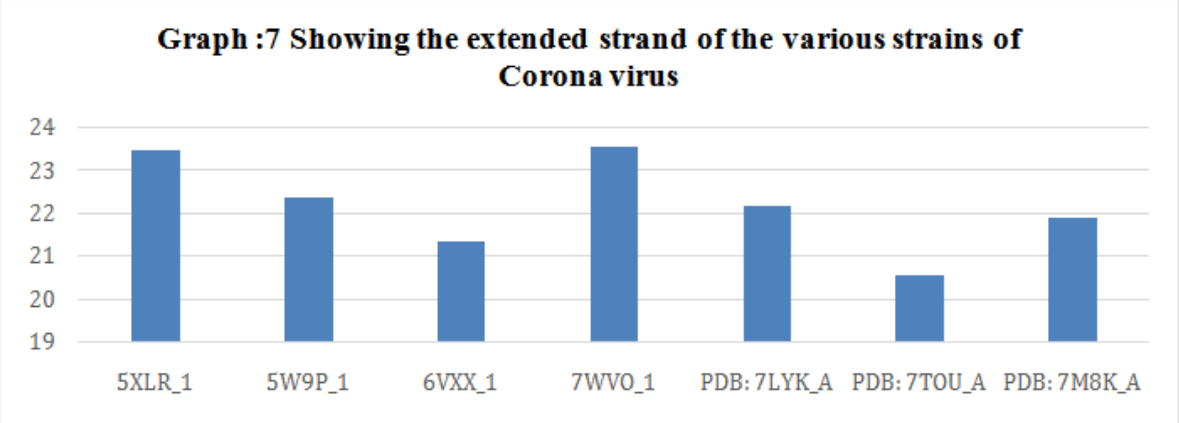
Table 3: Percentage of predicted secondary structure in seven variants of corona virus

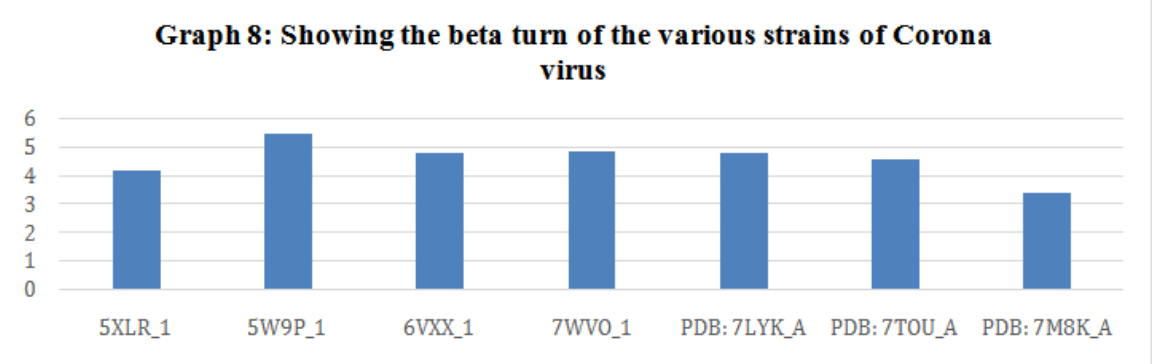
Common Name	Percentage of Alpha helix	Percentage of Extended strand	Percentage of Beta turn	Percentage of Random coil
SARS-CoV-2002	28.9	23.44	4.16	43.47
MERS	29.27	22.35	5.42	42.96
SARS-CoV-2019	28.96	21.31	4.76	44.96
OMICRON	26.87	23.53	4.85	44.75
BETA	26.78	22.16	4.78	46.28
DELTA	28.08	20.55	4.55	46.82
GAMMA	29.81	21.89	3.39	44.91

Table 3 showingcomparison of different forms of secondary structure among different strains of Corona Virus. Gamma showing highest percentage of alpha helix whereas delta showing highest percentage of Random coils.

International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781







International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

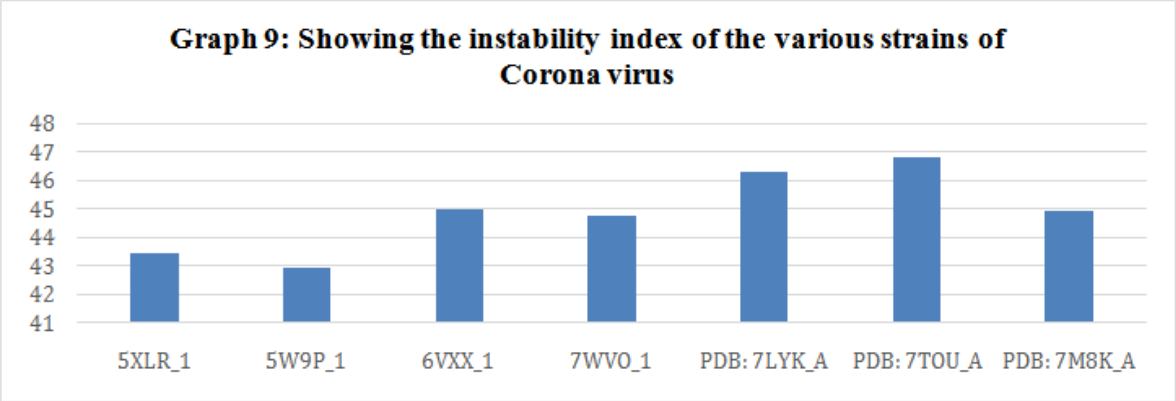


Table 4:Showing composition and percentage of all 20 amino acid in the seven strains of corona virus

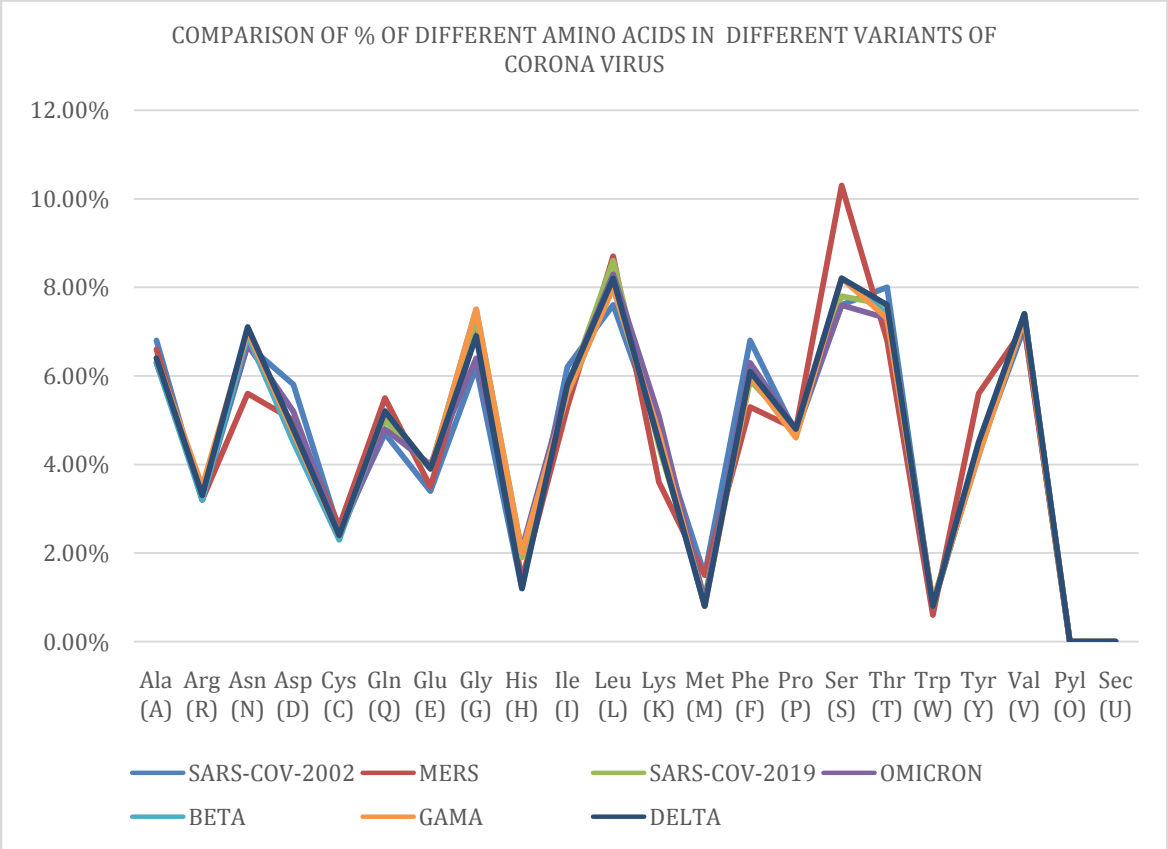
AMINO ACID	SARS-COV-2002		MERS		SARS-COV-2019		OMICRON		BETA		GAMMA		DELTA	
	Composition	Percentage	Composition	Percentage	Composition	Percentage	Composition	Percentage	Composition	Percentage	Composition	Percentage	Composition	Percentage
Ala (A)	82	6.8%	88	6.6%	81	6.3%	79	6.3%	81	6.30%	81	6.40%	79	6.40%
Arg (R)	38	3.2%	43	3.2%	42	3.3%	43	3.4%	41	3.20%	44	3.50%	41	3.30%
Asn (N)	81	6.7%	74	5.6%	89	6.9%	84	6.7%	88	6.90%	89	7.00%	88	7.10%
Asp (D)	70	5.8%	66	5.0%	61	4.8%	66	5.2%	58	4.50%	60	4.70%	59	4.80%
Cys (C)	30	2.5%	34	2.6%	31	2.4%	30	2.4%	30	2.30%	30	2.40%	30	2.40%
Gln (Q)	56	4.7%	73	5.5%	64	5.0%	61	4.8%	66	5.20%	66	5.20%	64	5.20%
Glu (E)	41	3.4%	46	3.5%	50	3.9%	50	4.0%	50	3.90%	50	3.90%	48	3.90%
Gly (G)	74	6.2%	93	7.0%	93	7.3%	80	6.4%	96	7.50%	95	7.50%	86	6.90%
His (H)	15	1.2%	18	1.4%	24	1.9%	26	2.1%	26	2.00%	26	2.00%	15	1.20%
Ile (I)	74	6.2%	71	5.3%	73	5.7%	73	5.8%	72	5.60%	71	5.60%	72	5.80%

International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

Leu (L)	92	7.6%	116	8.7%	110	8.6%	105	8.3%	102	8.00%	102	8.00%	102	8.20%
Lys (K)	57	4.7%	48	3.6%	57	4.4%	64	5.1%	59	4.60%	60	4.70%	56	4.50%
Met (M)	18	1.5%	20	1.5%	13	1.0%	11	0.9%	10	0.80%	10	0.80%	10	0.80%
Phe (F)	82	6.8%	71	5.3%	75	5.9%	79	6.3%	78	6.10%	76	6.00%	76	6.10%
Pro (P)	57	4.7%	64	4.8%	62	4.8%	61	4.8%	60	4.70%	59	4.60%	59	4.80%
Ser (S)	92	7.6%	137	10.3 %	100	7.8%	96	7.6%	105	8.20%	105	8.20%	105	8.20%
Thr (T)	96	8.0%	91	6.8%	97	7.6%	92	7.3%	95	7.40%	93	7.30%	94	7.60%
Trp (W)	10	0.8%	8	0.6%	10	0.8%	10	0.8%	12	0.90%	12	0.90%	10	0.80%
Tyr (Y)	52	4.3%	74	5.6%	55	4.3%	55	4.4%	54	4.20%	53	4.20%	56	4.50%
Val (V)	86	7.1%	94	7.1%	94	7.3%	93	7.4%	94	7.40%	93	7.30%	92	7.40%
Pyl (O)	0	0%	0	0%	0	0%	0	0%	0	0.00%	0	0.00%	0	0.00%
Sec (U)	0	0%	0	0%	0	0%	0	0%	0	0.00%	0	0.00%	0	0.00%

Table 4 showingdifferent composition and amino acid percentage present in the different strains of the corona virus. Changes in the percentage of asparagine and threonine has been observed in MERS that may have some role in its increased lethality.

Graph10: Showing the amino acid percentage of the seven strains of Corona virus



International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

Domain And Active Site Comparison

Polypeptide chain becomes functional after undergoing specialized folding or arrangingitself in three-dimensional structure and can interact with other specific molecules (Lewis T., et.al., 2021).These particular areas are known as active sites, which often correspond to the polypeptide chain's domains and motifs (https://bio.libretexts.org/@go/page/16427). The active sites of uncharacterized polypeptides can be identified by analyzing the tertiary structure data retrieved from the ScanProsite. It gives details on the various patterns and domains that a polypeptide chain might take on after folding (Gattiker A., et.al., 2002).

The domain and active site prediction has shown that all the seven variants of corona virus contain four active sites namely - N-terminal domain andC-terminal domain of S1 subunit, HR-1 or Heptad Repeat-1, and HR-2or Heptad Repeat-2 of S2 subunit. But differences had been found in the sequence length of all the major domains present in the spike protein of the seven variants (table5).

Most of the variation in amino acid number of different domains of corona viruses was centered on S1-NTD. MERS-CoV showed maximum number with 335 residues while minimum number was in SARS-CoV-2002 with 279 residues. SARS-CoV-2019 contains 303 amino acids starting from 20-322, while Omicron had 11 amino acids less than SARS-CoV-2019 ranging from 9 to 300from N terminal. Both Beta and Gamma strains contain 292 amino acids of same range from 1-292 and Delta contains 290 amino acids. In the C terminal domain (BCOV_S1_CTD) variation of amino acid number was observed in SARS-CoV-2002 (192) and MERS-CoV (207) than others with same count (194). In the S2-HR1 domain total 106 amino acids with different ranges have been observed in all the variants. Amino acid number in S2-HR2 varies a little i.e., 83 amino acids in SARS-CoV-2019, 79 in SARS-CoV-2002 and 84 in rest of the strains but all with different ranges except in Beta and Gamma strains (1132-1215). Different domains of corona virus variants: Omicron (A), Gamma (B), Delta (C) and Beta (D) is represented in fig 1

Table5: Difference of number and sequence range of amino acids in four different domains of all seven strains of Corona viruses under investigation.

Corona Variants	BCOV_S1_NTD	BCOV_S1_CTD	COV_S2_HR 1	COV_S2_HR 2
SARS-CoV-2002	12-290 (279)	321 – 513 (193)	878 – 983 (106)	1125 – 1203(79)
MERS-CoV	17 – 351 (335)	381 – 587 (207)	970 – 1075 (106)	1226 – 1309(84)
SARS-CoV-2019	20 – 322 (303)	353 – 546(194)	915-1020(106)	1162-1244(83)
OMICRON	9 – 300 (292)	331 – 524 (194)	893 – 998 (106)	1140 – 1223 (84)
BETA	1 – 292 (292)	323 – 516 (194)	885 – 990 (106)	1132 – 1215 (84)
DELTA	1 – 290 (290)	321 – 514 (194)	883 – 988 (106)	1130 – 1213 (84)
GAMMA	1 – 292 (292)	323 – 516 (194)	885 – 990 (106)	1132 – 1215 (84)

Table 5showing four domains (S1 – NTD, S1 – CTD, HR1, HR2) of the spike protein and their corresponding number of amino acid sequence in different variants of corona virus. Amino acid numbers are shown within parenthesis.

International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

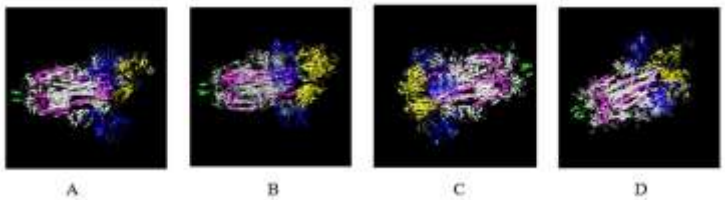


Fig1: comparison of 3d structure highlighting four domains of four strains of corona virus

Fig1 Showing ribbon model of 3D structure constructed using RasMol software of different domains of corona virus variants:Omicron (A), Gamma (B), Delta (C) and Beta (D). BCOV_S1_NTD is represented in Blue, BCOV_S1_CTD is represented in Yellow, COV_S2_HR 1 is represented in Magenta and COV_S2_HR 2 is represented in Green.

Phylogenetic analysis

Phylogenetic tree of spike proteins of selected seven variants of Corona virus was build using MEGA11 software and result has shown in fig 2. The Tree helped in analyzing the similarity between spike proteins of seven selected strains of the corona viruses. Delta variant and Omicron had shown 92% similarity whereas Gamma and Beta variant showed 90% of similarity. Delta, Omicron, Gamma and Beta are closely related with SARS-CoV-2019 and had common ancestors. SARS-CoV-2019 and its all the variants (Delta, Omicron, Gamma and Beta) had 90% resemblance with SARS-CoV-2002. MERS-CoV had shown list resemblance with rest of the six strains.

Fig 2:Phylogenetic tree build using MEGA 11 software

International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

Fig 2 showing distance between different strains of corona virus according to their amino acid composition of spike protein in the constructed phylogenetic tree by MEGA 11 software. MERS shows distant relation to other strains of corona virus. While DELTA and OMICRON has close phylogenetic relationship.

Analysis of Ramachandran's Plot:

Stability of an unknown protein usually derived by the analysis of Ramachandran's plot. It defines by means of polypeptide backbone rotation around the comparatively rigid bond between Cα-C (psi bond) and N- Cα (phi bond). Over 90% residues should belong to core region and less than 2% in allowed region would usually made the protein stereo chemically stable.

For all the strains we retrieved percentage of amino acid present in most favored region of Ramachandran Plot separately using Procheck PDBsum online tool (table 6 and fig 3). Our analysis showed that in SARS-CoV-2022, 74.9% of amino acid residues was found under most favored region comprises of core alpha (A), core left alpha (L) and core beta (B). MERS CoV strain showed 88.7%, Delta strain shows 87.1%, beta shows 85.5% and gamma shows 85.5% of amino acid residues in the most favorable region (A, B, L) of the Ramachandran's plot. 90.0% and 91.8% of amino acid residues were found under most favored region of the plot in SARS-CoV-2019 and Omicron.

Table 6: Amino acid percentage found in most favored region of Ramachandran Plot.

Corona Variants	Percentage of amino acid
SARS-CoV-2002	74.9%
MERS-CoV	88.7%
SARS-CoV-2019	90.0%
OMICRON	91.8%
BETA	86.9%
DELTA	87.1%
GAMMA	85.5%

Table 6 showingdifferent amino acid percentage present in the most favored region of the Ramachandran Plot of different strains of the corona virus. SARS-CoV-2019 and OMICRON is found in the core region with 90% amino acid.

Fig 3: Ramachandran plot of all seven variants build using PDBsum

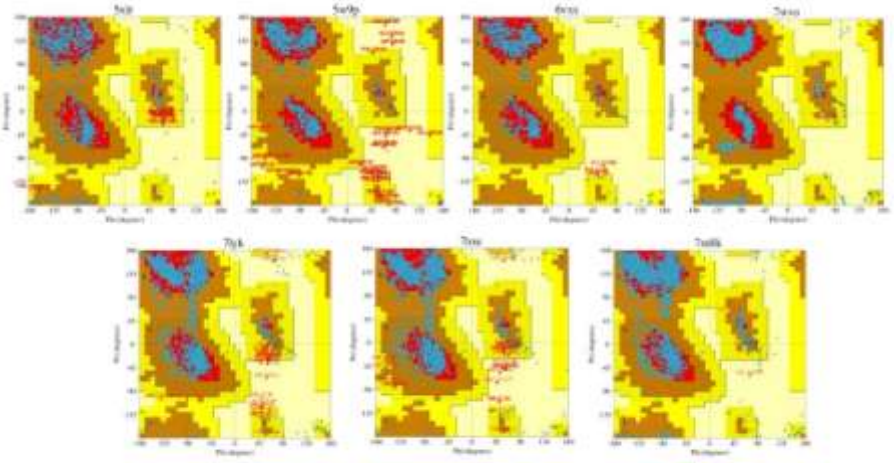


Fig3Showing Ramachandran Plot build using PDBsum of all seven variants of Corona virus- (5xlr) SARS-CoV-2002, (5w9p) MERS, (6vxx) SARS-CoV-2019, (7wvo) OMICRON, (7lyk)BETA, (7tou)DELTA, (7m8k) GAMMA.

International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

IV. DISCUSSION:

C fragment (259 amino acids from 441 to 700) of the spike protein of SARS CoV 2002 were reported to act as main immune- dominant antigen and could be utilized as the serological detection of SARS CoV 2002 infection (1). Likewise same stretch of spike proteins in different strains could be tested to find out their serological detection procedure.

Compared to the original Wuhan virus delta strain showed changes in spike proteins and other proteins by amino acid substitution and certain deletions in N terminal domain of spike protein that made them to escape from antibodies in vaccinated individuals (2).

In the present study we had investigated the similarities and differences of the spike protein in the seven corona virus variants which include - SARS-CoV-2002, MERS, SARS-CoV-2019, OMICRON, BETA, DELTA, and GAMMA. Using different In-Silico tools the primary structure, secondary structure and tertiary structure were analyzed. Stability of spike proteins was investigated by analysing different parameters such as physiochemical characteristics, secondary structures and phylogenetic tree analysis etc. Physiochemical investigation of all the variants had shown that all the proteins were stable in nature as their instability index detected well below 40 and aliphatic index ranges from 80.42 to 84.82.High percentage of secondary structure mainly alpha helix is also responsible for the stability of the spike proteins.

All the variant of Spike proteins were hydrophobic in nature as confirmed by thenegative GRAVY values. Furthermore, the study has shown that spike protein structure (as predicted by Swiss Model) of all the seven variants looks similar apparently but there were variations in primary, secondary and tertiary structures as depicted by the detailed data analysis. These variations may further influence the rate of transmission and severity of the corona virus disease. By examining the MERS corona virus detailed structure, it was revealed that high percentage alpha helix, large aliphatic index and low instability index of the spike protein in this strain confirms its high stability and plausible reason for its severe infectivity.

Domain and active site investigation of all the seven variants contains four domains (NTD and CTD of S1 subunit and HR-1 and HR-2 of S2 subunit) in their Spike protein. Major differences have been observed in the amino acid sequence of NTD whereas CTD of S1 subunit shows minor variation. MERS-CoV shows highest differences in both the domains of S1 subunit. In S2 subunit very little variation has been seen in the HR-2 subunit and no variation in the HR-1.

SARS-CoV-2002 spread fast but was not as lethal as MERS-CoV, whereas SARS-CoV-2019 and its variants were deadly and life threatening as well as highly transmittable. May be this was due to changes in the spike protein stability, to asses this we have analysed Ramachandran plot of all seven variants to find out their relative stability. Ramachandran plot of all the variants revealed that spike protein of SARS-CoV-2019 and Omicron was more stable than other variants as 90.0% and 91.8% of its amino acid falls under the most favorable region of the Ramchandran plot, whereas spike protein of SARS-CoV-2002 shows least stability with 74.9% of its amino acid was found under the favorable region that may speak of their relative instability. Spike protein of rest of the variants showed moderately stable amino acid residues in the most favorable region of the Ramachandran plot (85.5%-88.7%).

Difference in the physiochemical parameters revealed that changes has been occurred in amino acid sequence of the spike protein of all seven variants. By analysing the secondary structure and Ramachandran plot we have observed that MERS-CoV was more stable than other variants of corona virus and this can be possible reason for its lethality. Domain and active site analysis had shown that major variation had been occurred in the NTD of S1 subunit of the spike protein which helps binding with the host receptor. Changes in the S1 subunit might had played a major role in both differential transmission and lethality of various strains of the same virus.

V. CONCLUSION:

Accordingly, considering all the parameters that we checked for differentiating seven strains of corona virus for infectivity and severity, future strains that might come out in near future could be analyzed and handled beforehand to prevent setting on an outburst of future pandemic. MERS's highest stability can be a possible reason for its lethality. Omicron having medium stability can spread fast and is less lethal than MERS. Delta and Gamma stability is somewhat similar to MERS and that's why it possibly shows high mortality rate than omicron.Further extensive study and analysis in this field might be helpful to build therapeutics to combat diseases from Corona or related viruses.

International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

Ethical Statement:

Acknowledgements: Authors are indebted to the DSPM University for providing tools and internet support to perform this work in the department of Zoology.

Funding: (if any): The authors did not receive support from any organization for the submitted work.

Financial or non-financial interests: The authors have no competing interests to declare that are relevant to the content of this article.

Ethical approval: This is an observational study. The Institutional Research Ethics Committee of Dr Shyama Prasad University has confirmed that no ethical approval is required.

Informed consent: Bioinformatics of the Amino acid Sequences available in the NCBI were the raw data for this work and Authors are responsible for correctness of the statements provided in the manuscript.

Author contribution: All authors contributed to the study conception and design, Data collection and analysis were performed by Medha Kumari, and Sandip Mandal under the guidance of Dr Koel Mukherjee. The first draft of the manuscript was written by Sajalendu Ghosh and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

REFERENCES

[1]. Taherizadeh, M., Tabibzadeh, A., Panahi, M., Tameshkel, F. S., Golahdooz, M., & Niya, M. H. K. (2020). An introduction to SARS coronavirus 2; comparative analysis with MERS and SARS coronaviruses: a brief review. Iranian Journal of Public Health, 49(Suppl 1), 30.

[2]. Shu, Y., He, H., Shi, X., Lei, Y., & Li, J. (2021). Coronavirus disease-2019. World Academy of Sciences Journal, 3(2), 1-1.

[3]. Aleem, A., Samad, A. B. A., & Slenker, A. K. (2022). Emerging variants of SARS-CoV2 and novel therapeutics against coronavirus (COVID-19). In StatPearls [Internet]. StatPearls Publishing.

[4]. Shi, Z., & Hu, Z. (2008). A review of studies on animal reservoirs of the SARS coronavirus. Virus research, 133(1), 74-87.

[5]. Feikin, D. R., Alraddadi, B., Qutub, M., Shabouni, O., Curns, A., Oboho, I. K., ... & Madani, T. A. (2015). Association of higher MERS-CoV virus load with severe disease and death, Saudi Arabia, 2014. Emerging infectious diseases, 21(11), 2029.

[6]. Taherizadeh, M., Tabibzadeh, A., Panahi, M., Tameshkel, F. S., Golahdooz, M., & Niya, M. H. K. (2020). An introduction to SARS coronavirus 2; comparative analysis with MERS and SARS coronaviruses: a brief review. Iranian Journal of Public Health, 49(Suppl 1), 30.

[7]. Lim, Y. X., Ng, Y. L., Tam, J. P., & Liu, D. X. (2016). Human coronaviruses: a review of virus–host interactions. Diseases, 4(3), 26.

[8]. Zumla, A., Hui, D. S., & Perlman, S. (2015). Middle East respiratory syndrome. The Lancet, 386(9997), 995-1007.

[9]. Negi, S. S., Schein, C. H., & Braun, W. (2022). Regional and temporal coordinated mutation patterns in SARS-CoV-2 spike protein revealed by a clustering and network analysis. Scientific reports, 12(1), 1-10.

[10]. Alberca, G. G. F., Solis-Castro, R. L., Solis-Castro, M. E., & Alberca, R. W. (2021). Coronavirus disease–2019 and the intestinal tract: An overview. World Journal of Gastroenterology, 27(13), 1255.

[11]. Aleem, A., Samad, A. B. A., & Slenker, A. K. (2022). Emerging variants of SARS-CoV2 and novel therapeutics against coronavirus (COVID-19). In StatPearls [Internet]. StatPearls Publishing.

[12]. Lancet, K. S. K. Q. (2021). 398: 2126–2128. doi: 10.1016. S0140-6736 (21), 02758-6.

[13]. Mlcochova, P., Kemp, S. A., Dhar, M. S., Papa, G., Meng, B., Ferreira, I. A. T. M., ... & Indian, S. A. R. S. (2021). CoV-2 Genomics Consortium (INSACOG). Genotype to Phenotype Japan (G2P-Japan) Consortium, 114-119.

[14]. Tatham, L., Sharp, J., Kijak, E., Herriott, J., Neary, M., Box, H., ... & Owen, A. (2022). Lack of Ronapreve (REGN-CoV; casirivimab and imdevimab) virological efficacy against the SARS-CoV-2 Omicron variant (B. 1.1. 529) in K18-hACE2 mice. bioRxiv.

[15]. Gu, H., Krishnan, P., Ng, D. Y., Chang, L. D., Liu, G. Y., Cheng, S. S., ... & Poon, L. L. (2022). Probable transmission of SARS-CoV-2 omicron variant in quarantine hotel, Hong Kong, China,

International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 1 Jan-Feb 2023, pp: 747-762 www.ijprajournal.com ISSN: 2249-7781

November 2021. Emerging infectious diseases, 28(2), 460.

[16]. Li, F. (2016). Structure, function, and evolution of coronavirus spike proteins. Annual review of virology, 3(1), 237.

[17]. Leikina, E., Mertts, M. V., Kuznetsova, N., & Leikin, S. (2002). Type I collagen is thermally unstable at body temperature. Proceedings of the National Academy of Sciences, 99(3), 1314-1318

[18]. Lee, C., & Choi, W. J. (2021). Overview of COVID-19 inflammatory pathogenesis from the therapeutic perspective. Archives of pharmacal research, 44(1), 99-116.

[19]. Sanches, P. R., Charlie-Silva, I., Braz, H. L., Bittar, C., Calmon, M. F., Rahal, P., & Cilli, E. M. (2021). Recent advances in SARS-CoV-2 Spike protein and RBD mutations comparison between new variants Alpha (B. 1.1. 7, United Kingdom), Beta (B. 1.351, South Africa), Gamma (P. 1, Brazil) and Delta (B. 1.617. 2, India). Journal of virus eradication, 7(3), 100054.

[20]. Wink, P. L., Ramalho, R., Monteiro, F. L., Volpato, F. C. Z., Willig, J. B., Lovison, O. V. A., ... & Martins, A. F. (2022). Genomic Surveillance of SARS-CoV-2 Lineages Indicates Early Circulation of P. 1 (Gamma) Variant of Concern in Southern Brazil. Microbiology spectrum, 10(1), e01511-21.

[21]. Tian, D., Sun, Y., Xu, H., & Ye, Q. (2022). The emergence and epidemic characteristics of the highly mutated SARS-CoV-2 Omicron variant. Journal of Medical Virology, 94(6), 2376-2383.

[22]. Huang, Y., Yang, C., Xu, X. F., Xu, W., & Liu, S. W. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacologica Sinica, 41(9), 1141-1149.

[23]. Tian, D., Sun, Y., Xu, H., & Ye, Q. (2022). The emergence and epidemic characteristics of the highly mutated SARS-CoV-2 Omicron variant. Journal of Medical Virology, 94(6), 2376-2383.

[24]. Gu, H., Krishnan, P., Ng, D. Y., Chang, L. D., Liu, G. Y., Cheng, S. S., ... & Poon, L. L. (2022). Probable transmission of SARS-CoV-2 omicron variant in quarantine hotel, Hong Kong, China, November 2021. Emerging infectious diseases, 28(2), 460.

[25]. Zhang, J., Xiao, T., Cai, Y., & Chen, B. (2021). Structure of SARS-CoV-2 spike protein. Current Opinion in Virology, 50, 173-182. 32

[26]. Huang, Y., Yang, C., Xu, X. F., Xu, W., & Liu, S. W. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacologica Sinica, 41(9), 1141-1149.

[27]. Tortorici, M. A., Walls, A. C., Lang, Y., Wang, C., Li, Z., Koerhuis, D., ... & Veesler, D. (2019). Structural basis for human coronavirus attachment to sialic acid receptors. Nature structural & molecular biology, 26(6), 481-489.

[28]. LuG, W. (2015). GaoGF. BatGtoGhuman: spikefeaturesdeterG mining'hostjump'ofcoronavirusesSARSG CoV, MERSGCoV, andbeyond, 23(8), 468G478.32.

[29]. https://blast.ncbi.nlm.nih.gov/Blast.cgi

[30]. https://web.expasy.org/protparam/

[31]. https://www.megasoftware.net/dload_mac_beta

[32]. https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html

[33]. https://swissmodel.expasy.org/interactive

[34]. https://www.ncbi.nlm.nih.gov

[35]. Pramanik, K., Ghosh, P. K., Ray, S., Sarkar, A., Mitra, S., & Maiti, T. K. (2017). An in silico structural, functional and phylogenetic analysis with three dimensional protein modeling of alkaline phosphatase enzyme of Pseudomonas aeruginosa. Journal of Genetic Engineering and Biotechnology, 15(2), 527-537.

[36]. Lewis, T., & Stone, W. L. (2021). Biochemistry, Proteins Enzymes. In StatPearls [Internet]. StatPearls Publishing.

[37]. https://bio.libretexts.org/@go/page/16427

[38]. https://prosite.expasy.org/prosuser.html#general

[39]. Sigrist, C. J., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Bulliard, V., Bairoch, A., & Hulo, N. (2010). PROSITE, a protein domain database for functional characterization and

annotation. Nucleic acids research, 38(suppl_1), D161-D166.

[40]. Gattiker, A., Gasteiger, E., & Bairoch, A. M. (2002). ScanProsite: a reference implementation of a PROSITE scanning tool. Applied bioinformatics, 1(2), 107-8.

[41]. Lewis T, Stone WL. Biochemistry, Proteins Enzymes. [Updated 2021 May 4]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan- . Available from:https://www.ncbi.nlm.nih.gov/books/NBK554481/

[42]. Lu, L., Manopo, I., Leung, B. P., Chng, H. H., Ling, A. E., Chee, L. L., ... & Kwang, J. (2004). Immunological characterization of the spike protein of the severe acute respiratory syndrome coronavirus. Journal of Clinical Microbiology, 42(4), 1570-1576.

[43]. Moghaddar, M., Radman, R., & Macreadie, I. (2021). Severity, pathogenicity and transmissibility of delta and lambda variants of SARS-CoV-2, toxicity of spike protein and possibilities for future prevention of COVID-19. Microorganisms, 9(10), 2167.

[44]. Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., & Wang, X. (2017). Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. Cell research, 27(1), 119-129.

[45]. Pallesen, J., Wang, N., Corbett, K. S., Wrapp, D., Kirchdoerfer, R. N., Turner, H. L., ... & McLellan, J. S. (2017). Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. PNAS 114. E7348–E7357.

[46]. Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell, 181(2), 281-292.

[47]. Hong, Q., Han, W., Li, J., Xu, S., Wang, Y., Xu, C., ... & Cong, Y. (2022). Molecular basis of receptor binding and antibody neutralization of Omicron. Nature, 604(7906), 546-552.

[48]. Gobeil, S. M. C., Janowska, K., McDowell, S., Mansouri, K., Parks, R., Stalls, V., ... & Edwards, R. J. (2021). Effect of natural mutations of SARS-CoV-2 on spike structure, conformation and antigenicity. bioRxiv Prepr. Serv. Biol, 6226, 1-24.

[49]. Gobeil, S. M. C., Henderson, R., Stalls, V., Janowska, K., Huang, X., May, A., ... & Acharya, P. (2022). Structural diversity of the SARS-CoV-2 Omicron spike. Molecular cell.

[50]. Wang P, Casner RG, Nair MS, et al. Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. Cell Host Microbe. 2021;29(5):747-751.e4. doi:10.1016/j.chom.2021.04.007