IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Analysis of Structural proteins of Novel Corona Virus 2 and the identification of mutational sites responsible for Viral Drug Resistance

[1]Dr.D.Gayathri, [2]K.Suchitha Reddy, [3]K.Sumila Reddy, [4]B. Jhanavi

[1]Associate Professor, [2] Associate Professor, [3]Associate Professor, [4]Associate Professor
[1]Department of Biotechnology,
[1]Ideal Degree College for Women, Hyderabad, India

Abstract:

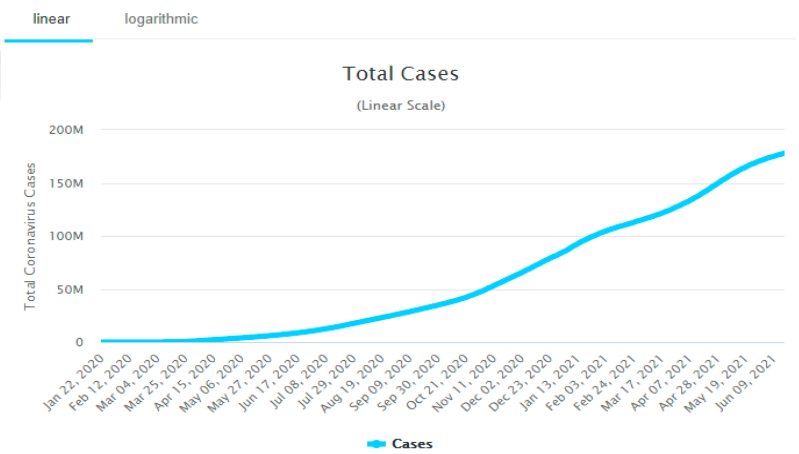
Novel Corona Virus 2019 Wuhan isolate has marked the year 2019 with its dreadful, pathogenic and contagious disease COVID. It has threatened the lives of millions of people on a global scale. This caused pooling of all the researchers and scientific world into the COVID research and succeeded in identifying several drugs towards its treatment. However the worst part of these positive stranded RNA viruses is its efficacy in developing drug resistance and self-mutation resulting in the current pandemic. Though there are he archives of research accumulated related to this virus in the 2 years the concept of this drug resistance by the virus is not much investigated. The present research aimed to identify the cause of this drug resistance development based on proteomic analysis of the virus structural genes.

Keywords:
COVID, Pandemic, Proteomic Analysis, Drug resistance

I.Introduction:

Novel corona virus 2019 was first identified in the Wuhan city of China infecting the local people of the city [1]. It was thought to be produced from the bats. This virus came into limelight because of the highly contagious and life threatening infection COVID. Rapid spread of this virus was irrespective to the climatic conditions, Environmental factors, temperature humidity etc. It was universally fast spreading without limitations [2]. This indicates that versatility of the organism. This is a Positive stranded RNA virus encapsulated in a protein rich capsid appearing to be a crown shaped head under the electron microscope [3]. This virus belongs to the beta corona virus group of coronaviridae family. The virus is also termed as SARS corona virus 2 in view of its high degree (89%) of similarity to bat SARS-like-CoVZXC21 [4].
According to the records of COVID cases as on June 17th 2021 a total of 177,885,880 positive cases have been recorded globally. Number of deaths reported due to COVID are 3,850,529 on a global scale [5].

Fig 1: Graph showing the statistics of COVD Infection globally as on 17th June 2021.



The above graphical representation of COVID infection shows a gradual increase in the infected cases from 50Million in Nov 2020 to greater than 150 million by June 2021. This statistics emphasizes the need for a solution to this disease.

In view of the complete dependence of COVID Pandemic control on Pharmacological inventions, it becomes a challenging task for drug scientists. Most important challenges for drug development are the availability of limited information about the disease pathogen in addition to the limited time available for product release. Thus an alternative to drug discovery is the Repurposing technology [6] of drugs. Base on the genetic similarity of the pathogen to the known pathogens an already available drug can be repurposed and used to face the condition. Plethora of drugs has been studied and suggested to target either viral replication cycle or symptoms observed in COVID infection. Chloroquine, Hydroxychloroquine, Lopinavir, Ritonavir, Nafamostat, Camostat, Famotidie [7] etc have been proposed and tried till date in the treatment for this dreadful infection. However the nothing could be proved efficient. One of the major reasons for this failure can be the development of drug resistance by these RNA viruses.

II. Methodology

The detailed study of the Structural Proteins coded by Novel Corona Virus 2019 was performed from NCBI data base. Total number of proteins coded by SARS CoV 2019 is 26 as per the records of NCBI: (NC_045512). These proteins are coded by a total of 10 genes. This is because of the fact that some of the genes are capable of coding for more than 1 protein [8]. Among the viral genes ORF1ab is the largest gene coding for a total of 16 proteins by its internal cleavage by proteases. These proteases are also produced by the same gene. Major proteins of SARS CoV2 include both structural and non-structural proteins. The current study included all the structural proteins due to their vital role in the viral pathogenicity and its efficacy in host cell binding. Structural genes of SARS CoV include Spike Surface Glycoprotein [S] (functions in binding to the cognate receptor on a human or animal cell, Matrix Protein [M], Envelope Protein [E] and Nucleocapsid Protein [N] (internal packaging of the genome).

Sequences of all the 4 structural proteins were collected from one of the premier primary databases, NCBI [9]. This is a major data base that contains information about all the known proteins, genes, functions etc. It has access to a huge archive of medical and research journals and books. All the sequences along with the corresponding annotations were collected from this database. The annotations include Length of the sequence, Organism source, internal motifs and domains, classification of the organism etc.

Once the sequences were collected they were compared with the Human proteome to check for their similarities to the Humans. Only the proteins with zero identity to the host can act as foreign particles and provoke immune system thereby causing pathogenicity. BLASTP [10] is used for the sequence similarity search and all the structural proteins are screened for their foreignness. All the five proteins were identified to be foreign to humans thus may be pathogenic in nature.

Being identified to be pathogenic these sequences were subjected for detailed annotation for the identification of localized antigenic regions. An online tool kit protein Variability Server (PVS) [11] was used which could identify the antigenic regions on the protein sequences along with their antigenic propensity represented as graph. These regions were further cross conformed to another antigenic site identification tool EMBOSS ANTIGENIC [12]. Based on the annotation from both the tools the antigenic sites have been identified. After the identification of all the pathogenic sites the study was extended to check the influence of SNP at these regions by replacing the existing wild amino acid at the target site to all the other possible 19 amino acids. I MUTANT [13] was used for this purpose and the results are tabulated. This tool predicts the changes in the stability of a protein at both sequence and structure level upon change in the single amino acid. In this context of work a decreased stability indicates that the pathogen loses its efficacy, whereas an increased stability upon mutation at the pathogenic region indicates a drug resistance and more stability of the pathogen. Increase in the stability of protein due to mutation at a specific site is considered fatal in this case.

Thus the study can pave a way for the identifying the cause of Drug resistance thereby developing a new drug that can treat COVID.

III. Results and Discussion

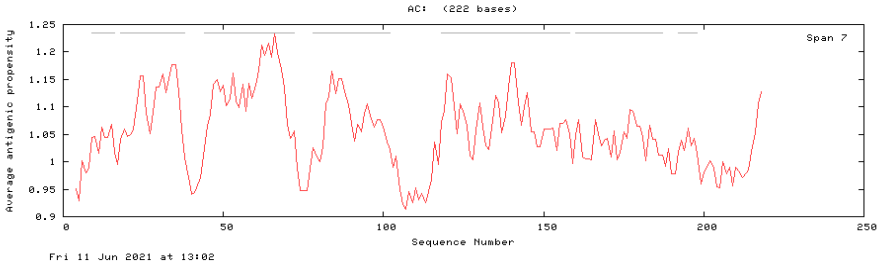
Antigenic Propensity results based on PVS for all the 4 proteins are depicted in the Figure 3.1 below.

Figure 3.1: Results of PVS showing antigenic regions Surface glycoprotein

n	Start Position	Sequence	End Position
1	4	FLVLLPLVSSQCVNL	18
2	22	TQLPPAY	28
3	33	TRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHV	70
4	80	DNPVLPFNDGVYFA	93
5	114	TQSLLIV	120
6	122	NATNVVIKVCEFQFCNDPFLGVYY	145
7	156	EFRVYSS	162
8	167	TFEYVSQPFLM	177
9	199	GYFKIYSK	206
10	208	TPINLVRDLPQGFSALEPLVDLPIG	232
11	237	RFQTLLALHRSYLT	250
12	261	GAAAYYVGYLQPRTFLL	277
13	285	ITDAVDCALDP	295
14	297	SETKCTLKSFTVEK	310
15	316	SNFRVQPTESIVRF	329
16	331	NITNLCPFGE	340
17	346	RFASVYA	352
18	357	RISNCVADYSVLYNS	371
19	373	SFSTFKCYGVSP	384
20	386	KLNDLCFTNVYADSFVIR	403

Best antigenic sites present on the above selected protein are 22 to 28 and 33 to 70 with the peptides: TQLPPAY and TRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHV which can be further annotated.

Figure 3.2: PVS results for Matrix protein M



There are 7 antigenic determinants in your sequence:

n	Start Position	Sequence	End Position
1	9	TVEELKKL	16
2	18	EQWNLVIGFLFLTWICLLQFA	38
3	44	RFLYIIKLIFLWLLWPVTLACFVLAAVYR	72
4	78	GGIAIAMACLVGLMWLSYFIASFRL	102
5	118	ILLNVPLHGTILTRPLLESELVIGAVILRGHLRIAGHHLGR	158
6	160	DIKDLPKEITVATSRTLSYYKLGASQRV	187
7	192	GFAAYSR	198

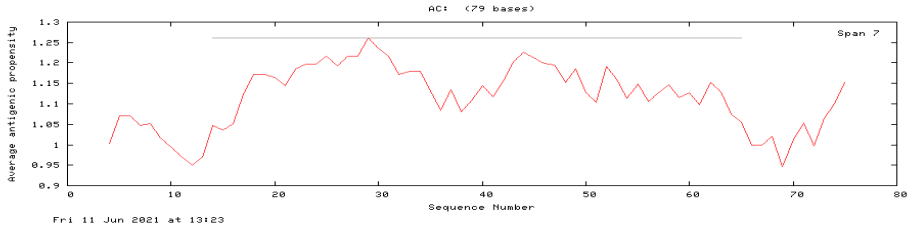
From the above table and graph the region with highest pathogenicity was identified to be 44 to 72 with the peptide sequence RFLYIIKLIFLWLLWPVTLACFVLAAVYR

Figure 3.3 : Antigenic peptide identification for Envelope protein E

Your sequence is 79 residues long

Average antigenic propensity for this protein is 1.1135

Antigenic plot for sequence



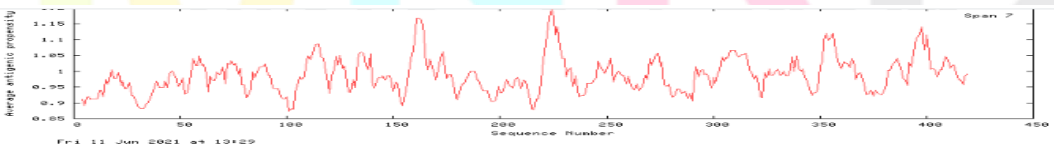
There are 1 antigenic determinants in your sequence:

n	Start Position	Sequence	End Position
1	14	GTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSR	65

Contact Pedro Reche
Last Update: 11 June 2021

From the above analysis the only pathogenic peptide present in the protein is 14-65 with its sequence GTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSR

Figure 3.4: Identification of antigenic regions in Nucleocapsid Protein N



There are 16 antigenic determinants in your sequence:

n	Start Position	Sequence	End Position
1	55	SWFTALTQ	62
2	72	RGQGVPI	78
3	86	DQIGYYR	92
4	109	SPRWYFYYLG	118
5	133	GIIWVAT	139
6	157	NNAAIVLQLPQGT	169
7	220	DAALALLLLDR	230
8	246	QGQTVTK	252
9	270	KAYNVTQ	276
10	302	YKHWPQIAQFAPSASAF	318
11	326	MEVTPSGT	333
12	336	TYTGAIK	342
13	350	FKDQVILLNKHIDAYKT	366
14	382	ETQALPQ	388
15	392	KQQTVTLLPAADL	404
16	406	DFSKQLQQS	414

Contact Pedro Reche
Last Update: 11 June 2021

The above analysis shows that the region of the peptide 220 to 230 has highest antigenic propensity and pathogenicity with its sequence: DAALALLLLDR

Table 3.1: Summary of the complete annotation work:

S.No	Protein (length)	Antigenic Sites in PVS	Antigenic sites in EMBOSS Antigenic
1	surface glycoprotein S (1273 aa)	TQLPPAY 22-28	FLVL**L**PLVSSQCVNLT 4-19
2	surface glycoprotein S (1273 aa)	TRGVYYPDKVFRSSVLHST QDLFLPFFSNVTWFHAIHV 33-70	RGVYYPDKVFR**S**SVLHSTQDLFLPFFSNV TWFHAIHVS 34-71
3	Matrix protein M (222 aa)	RFLYIIKLIFLWLLWPVTLA CFVLAAVYR 44-72	FLYIIKLIFLWLLWPVTLACFV**L**AAVYRI 45-73
4	Envelope protein E (75 aa)	GTLIVNSVLLFLAFVVFLLVTLAILT ALRLCAYCCNIVNVSLVKPSFYVYS R 14-65	TLIVNSVLLFLAFVV**F**LLVTLAILTALRLC AYCCNIVNVSLVKPSFYVYSRVKNL 11-65
5	Nucleocapsid Protein N (419 aa)	DAALALLLLDR 220-230	AALA**L**LLLDRL 217-227

According to the above table a total of 5 peptides were identified to be pathogenic based on the analysis. However the first peptide in the table shows a discrepancy in EMBOSS and PVS analysis. Thus it can be omitted from further study.

Among the remaining 4 peptides with corresponding antigenic sites identified, mutation prediction was performed. I mutant 2.0 was used to predict the effect of SNP in the selected 4 sites and their effect on the stability of the protein. Being a pathogenic peptide an SNP causing decreased stability of peptides is a positive indication and the one causing increased stability is hazardous and may be the cause of increased virulence in the strain. This study can be a potential data for analysing the Drug resistance of Novel Corona Virus 2.

Fatal mutations identified based on IMUTANT results are tabulated below:

Table 3.2: Summary of I Mutant results for all the selected peptides:

S.No	Protein	Peptide	Selected Fatal SNP's based on Increased Stability of mutant
1	Surface glycoprotein S	TRGVYYPDKVFR**S**SVLHSTQDL FLPFFSNVTWFHAIHV (45 position)	S45V, S45I, S45F, S45W, S45P, S45E and S45D
2	Matrix protein M	FLYIIKLIFLWLLWPVTLACFV**L** AAVYRI (67 position)	There are no fatal mutations identified in the above peptide
3	Envelope protein E	TLIVNSVLLFLAFVV**F**LLVTLAI LTALRLCAYCCNIVNVSLVKPS FYVYSRVKNL (26 position)	F26P and F26E
4	Nucleocapsid Protein N	AALA**L**LLLDRL (221 position)	No Fatal mutations are identified in the above peptide.

The above table highlights the possible SNP's among the structural proteins that can cause an increased stability of the mutant leading to an enhanced pathogenicity for the virus. These are the regions though to be the cause of Drug resistance by strain improvement.

IV. Conclusion:

One of the challenges faced by Researches working for COVID drug development is the inherent drug resistance of the virus. Irrespective of the many medications proposed for COVID therapy, none of them proved to be efficient to resolve the problem. The current work aimed to identify the cause of drug resistance among the SARS CoV 2019 virus. Study was limited to the structural proteins of the virus, targeting which the actual anchoring of the virus can be disturbed thereby eliminating the disease onset. All the 4 structural proteins were screened for their foreignness to the human proteome after which they are selected for further annotation. Using insilico tools like PVS and Antigenic EMBOSS the various antigenic sites present on the proteins were identified. Among the identified sites only those sites with high antigenicity were selected based on EMBOSS score and PVS antigenic propensity values. All the selected antigenic peptides were subjected for further analysis for the identification of virulent SNP's at the selected antigenic site. Based on the antigenic sites and Stability of SNP's among the 4 structural proteins Surface glycoprotein S and Envelope protein E are identified to be more virulent based on stability of the mutant strains. The final SNP's identified to be one of the causes of drug resistance are S45V, S45I, S45F, S45W, S45P, S45E and S45D of Surface glycoprotein S and F26P and F26E of Envelope protein E. These studies can further be confirmed based on laboratory analysis of mutants and testing their efficacy with the selected drugs.

ACKNOWLEGMENT

I am immensely thankful to our Principal Mr.Madhusudhan for unwavering and enthusiastic encouragement in bringing out in work in the School of Lifesciences.Throughout its preparation, many well wishers have freely given their help and advice and colleages their knowledge and insight.

REFERENCES

1) Maolin You, Zijing Wu, Yong Yang, Jun Liu, Dehua Liu, Spread of Coronavirus 2019 From Wuhan to Rural Villages in the Hubei Province, Open Forum Infectious Diseases, Volume 7, Issue 7, July 2020, ofaa228, https://doi.org/10.1093/ofid/ofaa228

2) V'kovski, P., Kratzel, A., Steiner, S. et al. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol 19, 155–170 (2021). https://doi.org/10.1038/s41579-020-00468-6

3) Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. Annu Rev Virol. 2016;3(1):237-261. doi:10.1146/annurev-virology-110615-042301

4) Cascella M, Rajnik M, Aleem A, et al. Features, Evaluation, and Treatment of Coronavirus (COVID-19) [Updated 2021 Apr 20]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-Available from: https://www.ncbi.nlm.nih.gov/books/NBK554776/

5) https://www.worldometers.info/coronavirus/coronavirus-cases/

6) Yadi Zhou, Fei Wang, Jian Tang, Ruth Nussinov, Feixiong Cheng,Artificial intelligence in COVID-19 drug repurposing, The Lancet Digital Health, Volume 2, Issue 12,2020,Pages e667-e676

7) Leah Shaffer, "15 drugs being tested to treat COVID-19 and how they would work", Nature Medicine, 15 May 2020

8) Bar-On YM, Flamholz A, Phillips R, Milo R. SARS-CoV-2 (COVID-19) by the numbers. Elife. 2020; 9:e57309. Published 2020 Apr 2. doi:10.7554/eLife.57309

9) National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2017 Apr 06]. Available from: https://www.ncbi.nlm.nih.gov/

10) Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402. PubMed

11) Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res*. 2008;36(Web Server issue):W35-W41. doi:10.1093/nar/gkn211

12) Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 276(1-2): 172–174. [PubMed] [Google Scholar]

13) Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 2005;33(Web Server issue):W306-W310. doi:10.1093/nar/gki375