Next Generation Sequencing Analysis of NSP5 NSP6 and NSP7 SARS CoV-2 Nonstructural Protein with Bio-Python

Vineeta Johri [1], Ekta Tyagi [2], Uma Kumari [3*]

[1] *Project Trainee at Bioinformatics Project and Research Institute, Noida, India*
[2] *Project Trainee at Bioinformatics Project and Research Institute, Noida, India*
[3] *Senior Bioinformatics Scientist, Bioinformatics Project and Research Institute, Noida-201301, India. Email: uma27910@gmail.com*
**Corresponding Author: Uma Kumari*

KEYWORDS
Next generation sequencing, Molecular Docking, SARS CoV-2, Bio-python, Deep fold, Bipepired, ORF finder

ABSTRACT
Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) an extremely pathogenic and transmissible virus that emerged in 2019, and caused global pandemic. The rapid mutation or evolution of various strains necessitates the analysis of its non-structural proteins using bioinformatics tools for therapeutic and drug design approaches. The virus consists of structural proteins like spike (S), nucleocapsid (N), membrane (M), and envelope (E). It binds to the ACE2 receptor via the S protein and uses the host's machinery to replicate and causes an immune response. This response can lead to cytokine storms and organ damage. Bioinformatics tools are used to analyse these proteins, and developed therapeutic development. Next-generation sequencing (NGS) and molecular docking tools accelerate drug discovery, while computational approaches help in predicting immune responses and developing vaccines. Initially protein samples of NSP 5 and NSP 6 and 7 were analysed with tool Prosite, to understand protein function, significant biological sites, and motifs. The ORF finder tool identified open reading frames in the RNA sequence, indicating potential viral replication sites. Potential docking sites for protein-ligand interactions were identified using the CB-Dock server, which evaluates binding affinity. The Deep Fold tool with GNN and RNN was used to predict protein structures, compatibility and compactness. Bio-python provided detailed structural information, including atom and residue distances and angles. For epitope prediction, the Bepipered linear epitope prediction tool identified B cell epitopes and developed reverse vaccinology and diagnostics. The analysis of protein samples 8OKB and 7MB6 of (NSP5, NSP6 and NSP7) for predicting protein structures and developing targeted drug delivery. This comprehensive approach offers significant insights into the structural and functional characteristics of SARS-CoV-2 non-structural proteins.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS CoV-2) is one of the deadliest pathogenic and highly transmissible virus that emerged in 2019 leading to a global pandemic creating a threat to human life. The continuous emergence of various coronavirus strains since the beginning of its very first case has created a scenario where we have to analyse its non-structural proteins using bioinformatics tools which can be useful in therapeutic and drug designing approach. [1] Coronavirus is a positive sense RNA virus belonging to family Coronaviridae and subfamily Coronavirinae whose members are capable of infecting variety of mammals, which can cause diverse clinical symptoms. Based on their sequence homology data they are grouped in 4 sub group: alpha, beta, gamma and delta. [2] Seven coronaviruses have been shown to infect humans since the 1960s. Two of these seven human coronaviruses (HCoVs) are classified as alpha-HCoVs, while the other five are classified as beta-HCoVs (OC43, SARS, HKU1, MERS, and SARS-2) .Three of them the newly identified SARS-CoV-2, the Middle East respiratory syndrome coronavirus (MERS-CoV), and the severe acute respiratory syndrome coronavirus (SARS-CoV)—cause serious or fatal sickness, while the other four (two α-CoVs and two β-CoVs) produce mild cold-like symptom. [3] The primary mode of transmission of SARS-CoV2 occurs through infected person's droplets through direct, indirect or via close contact with person infected with the virus. The main target of the virus is nasal epithelia, conjunctival or oral mucosa where its receptors can be found allowing their entry and replication to various organs. Host becomes contagious after 5 days and is able to transmit the virus to other person. The main symptoms associated with the infection include transient damage to various organs like lungs, kidneys, intestine and blood vessels.[4] Mechanism of the pathogenesis of the virus has been incompletely understood but the clinical phase of the virus has been proposed to consist of 3 phases-Viremia Phase, Acute Phase

and Recovery Phase**.**[5] The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virion consists of the following structural proteins: spike protein (S), nucleocapsid protein (N), membrane protein (M) and envelope protein (E). In stage 1 S protein, using its S1 domain gets attached to ACE2 (present on host lungs, kidney, GI) allowing TMPRSS2 to cleave Spike protein which is stage 2 leading to the fusion of S2 domain and its activation during stage 3 then this activated S2 domain fuse host lipid bilayer with viral protein which allows the deposition of positive sense ssRNA genome of the virus in the genomic content of the host cell during stage 4. Now using the host machinery virus replication begins creating a double stranded RNA intermediate which ultimately activates the host innate immune response through MDA5 and RIG1 activation which initiate a cascade of protein activation and reaction leading to the release of some paracrine factors like type 1 and 3 interferon via receptor genes present in the plasma membrane of host cell and stimulation of JAK STAT signalling pathway which is responsible for the activation of interferon specific genes which have direct or indirect effect on the host antiviral mechanism response.[6] SARS CoV 2 triggers the host's immune system to generate overproduction of proinflammatory cytokines which eventually leads to cytokine storm if they not remedied in time it can also severely destroy various organs. Even in asymptomatic individuals there might be a large amount of viral load contributing to the spread of infection. [7] Severe COVID-19 is characterized by organ dysfunction like hypercytokinemia and lymphopenia. Immune dysfunction in patients of COVID-19, including lymphopenia, decreased numbers of $CD4^+$ T cells and abnormal cytokine levels, is a common feature and may be a crucial factor associated with disease severity and worse outcomes [8]**.**Genomic analysis of SARS CoV2 has been sequenced several times to identify and detect its variants. It has been shown to contain 29,811 nucleotides which codes for 29 different viral proteins. [9] These viral proteins has been classified into 3 groups-structural proteins, non-structural proteins and accessory proteins. The viral RNA contains 70 bases at 5'leader sequence in which 7-10 are those bases which have transcription regulated sequence (TRS-L) encoding 13-15 reading frame in which ORF1a is the longest region of genome consists of ~13,200 of the 29,674 bases which encodes genes such as papain like protease and 3CL protease. ORF1b which comprises 14,442–21,563 bases encodes enzymes which are responsible for viral replication such as RNA dependent RNA polymerase (RdRP), a helicase and an endonuclease, overlaps with ORF1a.[10] Regions spanning from nucleotide 21,562-29,674 encode structural proteins like spike (S), membrane(M), envelope(E), Nucleocapsid(N) protein. Structure and content of the virus promote its survival within the host which may contribute to the disease progression. Genome of SARS CoV 2 contains CpG islands which allows virus cellular detection through its association with TLR9 in endosome which can trigger IFN response inside the host. Several studies have elucidated that during replication to evade the host's immune response there is a selective pressure to lower the CpG content. Another survival mechanism includes in the large scale internal RNA base pairing in the virus structure which can potentially shield viral RNA from recognition by host immune response and might contribute towards its progression. [11] There are 2 ORF in the genome of SARS CoV 2 ORF1a and ORF 1b their translated polypeptides are non-structural proteins which encompasses poly proteins, nucleoproteins and membrane proteins with different functions. For instance NSP 1 counteract antiviral response and supress apoptosis during early stage of infection in the host. NSP 2 disrupts intracellular host signalling. NSP3 facilitates mRNA synthesis and supress host's protein synthesis. NSP 3 and 4 induce membrane rearrangement mechanism and their role is crucial in viral replication. In NSP 5 several enzymatic and structural studies has been shown that SARS-CoV papain like protease is able to cleave peptide bond thereby acting as a peptidase as well as can cleave polyubiquitin chain by acting as a deubiquinating (DUB) enzyme. This viral protease and DUB activity enzyme can modulate or block innate immune response pathways. NSP 6 , located in ER region and is a component of alpha and beta coronavirus , generates autophagosomes .It was speculated that mutation in NSP6 might be able to change the expression of SARS CoV particularly concerning viral defence system of the host , for instance autophagolysosomal machinery.[12]. In the outer membrane of NSP 6 there is a stretch of phenylalanine residue which is favourable for binding affinity between ER membrane and

phenylalanine region may be favourable for causing coronavirus infection by disrupting the ability of autophagosome to deliver viral component to the lysosome for its degradation, thereby limiting autophagosome expansion, by starving or chemical inhibition by mTOR signalling. Various bioinformatics analysis have shown the highly conserved interface residue between NSP7 and NSP8 shows that both NSP7 and NSP 8 are essential cofactor which bind to NSP12 and this complex helps in stabilizing the polymerase domain. NSP 7 can significantly supress the IFN-α signalling which can lower the immune response against the virus.[13] Bioinformatics can provide valuable insights in SARS CoV 2 related research by providing various computational strategies to analyse the complex experimental data. Although Non-structural proteins are not a part of the main virus structure but they can help us perceive virus evolution and its pathogenesis mechanism. By interpreting these proteins at molecular level we can develop various strategies to combat the virus. There are several methods and applications relating to COVID-19 based on currently available reports of bioinformatics technologies, focusing on future research for overcoming the virus pandemic. Next generation Sequencing (NGS) provides a detailed and comprehensive view of virus detection and high throughput screening data for genome sequence of SARS CoV 2 can be obtained fairly quickly.[14,15,16] Several biological characteristics of the virus, type of mutation it undergoes and their analysis helps us in determining non-structural, structural and various accessory protein it uses for its replication. Immune response elucidation in the host by viral entry can be understood by single cell RNA sequencing. Angiotensin converting enzyme (ACE2) which could be the potential target for drug delivery. Linear sequence of the protein has been determined and after that secondary and tertiary details were determined using computational algorithm and the process is largely based on homology modelling. Various Molecular docking tools like CB dock2 can accelerate the development of drug by identifying potential binding affinity between protein and ligand in short amount of time.[17] Moreover, various computational approach such as epitope prediction tool has been used to prevent Covid 19 infection by using reverse vaccinology, immunoinformatics by predicting the value of peptide score which can be used in identifying B and T cell immunogenicity.[18] Bio-python is a dedicated library in the Python for bioinformatics, offering a versatile range of tools for sequence analysis to structural bioinformatics.[19] Sequence analysis in Bio-python methods plays important role to recapture of genetic sequences associated with proteins from public and experimental database. In structural bioinformatics, Bio-python integrates with tools for the predicting of the three-dimensional structure of protein, providing insights into its functional domains and potential binding sites. Furthermore, Bio.PDB module provide the visualization of protein structures, helping in the interpretation of protein's structural characteristics relevant to its function in SARS CoV-2 Disease. [20] The rapid evolution of computational methods has modify drug discovery, enabling the identification of new therapeutic compounds with extraordinary speed and accuracy. Molecular docking and virtual screening are important tools for predicting how small molecules like ligands interact with their target protein. With advanced computational resources, researcher can easily explore extensive chemical spaces, finding the promising drug candidates to effectively fight COVID 19. [21]

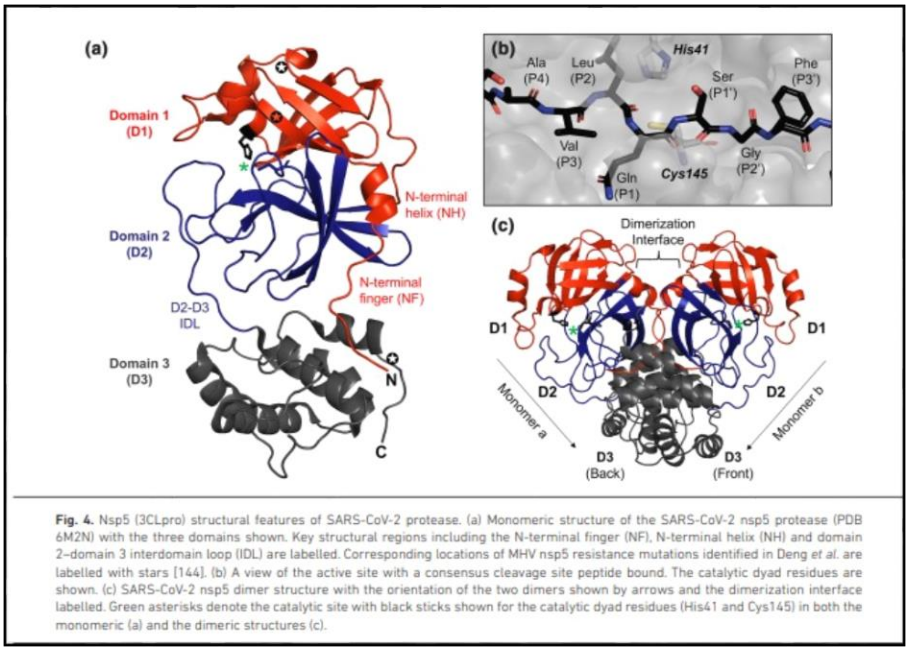


Fig. 4. Nsp5 (3CLpro) structural features of SARS-CoV-2 protease. (a) Monomeric structure of the SARS-CoV-2 nsp5 protease (PDB 6M2N) with the three domains shown. Key structural regions including the N-terminal finger (NF), N-terminal helix (NH) and domain 2–domain 3 interdomain loop (IDL) are labelled. Corresponding locations of MHV nsp5 resistance mutations identified in Deng *et al.* are labelled with stars [144]. (b) A view of the active site with a consensus cleavage site peptide bound. The catalytic dyad residues are shown. (c) SARS-CoV-2 nsp5 dimer structure with the orientation of the two dimers shown by arrows and the dimerization interface labelled. Green asterisks denote the catalytic site with black sticks shown for the catalytic dyad residues (His41 and Cys145) in both the monomeric (a) and the dimeric structures (c).

Figure 1: NSP5 structural features of SARS CoV2 protease

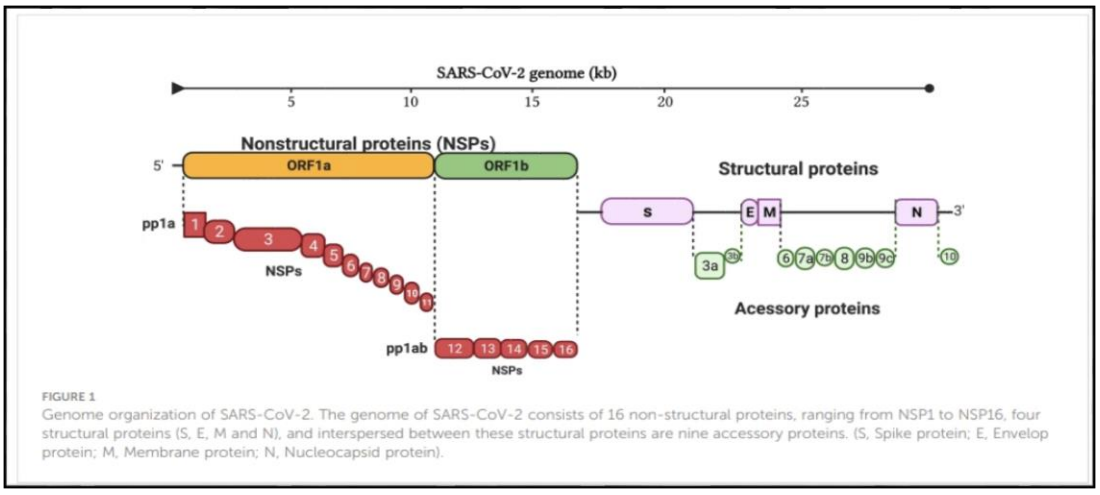


FIGURE 1
Genome organization of SARS-CoV-2. The genome of SARS-CoV-2 consists of 16 non-structural proteins, ranging from NSP1 to NSP16, four structural proteins (S, E, M and N), and interspersed between these structural proteins are nine accessory proteins. (S, Spike protein; E, Envelop protein; M, Membrane protein; N, Nucleocapsid protein).

Figure:-2 Non-structural protein of SARS CoV 2 genome organization

2. Material and Method

The study entreat the use of various computational tools to analyse the functioning and structural details of Non-structural protein 5,6 and 7 of SARS CoV2 using NGS and Bio-python. The study begins by finding the protein sample of NSP 5, NSP6, NSP7 with PDB ID – 8OKB and of NSP5, NSP6, NSP7 with PDB ID- 7MB6 which is analysed with the help of Scan Prosite which enables us to have better understanding of protein function, biologically significant site and pattern to identify from which protein family or motif. ORF finder tool was used to search open reading frame in the RNA sequence using the genetic codes to identify the possible sites from where virus can start its replication which can be useful in developing drugs against that site. We used CB Dock for identifying potential docking sites where our protein sample and ligand can interact and detect the binding affinity between them which can provide valuable insights into feasibility and efficacy of the potential drug candidate. To predict the structure of protein we used Deep Fold tool which is an advanced deep learning tool to analyse GNN (graphical neural network) for protein folding, RNN (Recurrent Neural Network) used in sequence based model capture the sequence-based amino acid which is integrated with folding kinetics and experimental data as well as various spatial based constrains like contact map

or distance map to check the folding compactness and structural compatibility of our protein sample (8OKB). Moreover, bio Python was implied to determine structural details like various atoms, residue, their distance and angle. To predict epitope region of our protein sample (7MB6) and B cell response against the invasion of SARS CoV2 we used an epitope prediction and analysis tool which is a companion site of curated databases of immune complexes curated by experimental data. B cell epitope prediction tool Bepipered linear epitope prediction 2.0 used to predict linear B cell epitopes based on sequence characteristics of the antigen using amino acid scales and HMMs which can be quite useful in reverse vaccinology , diagnostic application and antibody epitope predict. This comprehensive approach establish a thorough investigation of SARS CoV2 NSPs to offer insights in Coronavirus research.

3. Result

Prosite

PDB-8OKB-A (306 aa)

SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNF

LVQAGNVQLRVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRP

NFTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDT

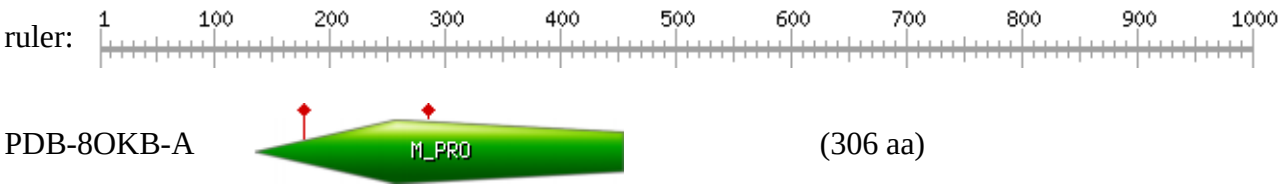
TITVNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLDM

CASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQ

Please note that the graphical representations of domains displayed hereafter are for illustrative purposes only, and that their colors and shapes are not intended to indicate homology or shared function.
For more information about how these graphical representations are constructed, go to https://prosite.expasy.org/mydomains/.

hits by profiles: [1 hit (by 1 profile) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.



PS51442 **M_PRO** *Coronavirus main protease (M-pro) domain profile* :

1 - 306: score = 154.193

SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTsEDMLNPNYEDLLIR

KSNHNFLVQAGNVQLRVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQTFSVLACYNG

SPSGVYQCAMRPNFTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGN

FYGPFVDRQTAQAAGTDTTITVNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYE

PLTQDHVdiLGPLSAQTGIAVLDMCASLKELLqNGMNGRTILGSALLEDEFTPFDVVRQC

SGVTFQ

Predicted features:

DOMAIN	1	306	/note="Peptidase C30 "	[condition: none]	
ACT_SITE	41		/note="For 3CL-PRO activity"	[condition: H]	[group: 1]
ACT_SITE	145		/note="For 3CL-PRO activity"	[condition: C]	[group: 1]

In computational analysis prosite database identifying protein domain family, functional site by utilising patterns and profile to better understanding of protein function and adding in the classification of protein based on sequence data.

ORF FINDINGS

ORF Finder searches for open reading frames in a RNA sequence using the standard or alternative genetic codes to identify all the possible open reading frames in a sequence

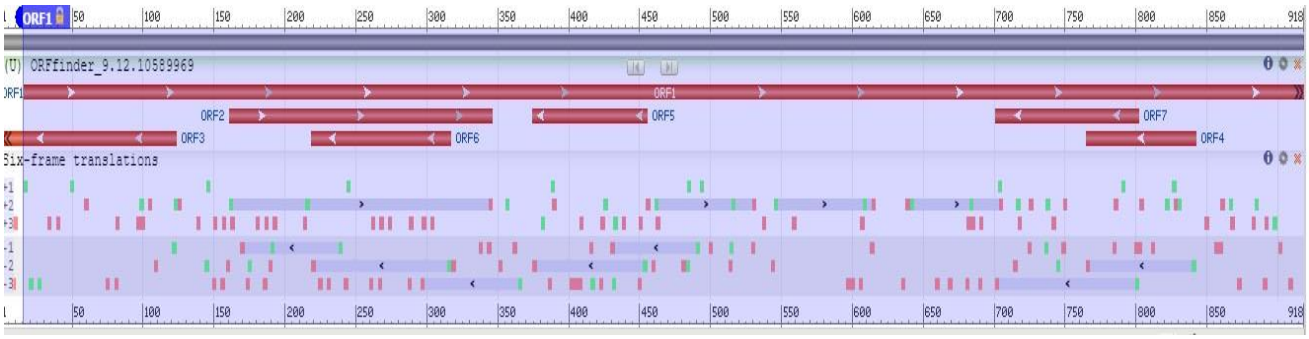


Figure 3: - Analysis of frame find all open reading frame (orf) and predicate protein sequences in forward and reverse direction in six translation frame.

Six horizontal bars corresponding to one of the possible reading frame. In each direction of the DNA there would be 3 possible reading frames(22). So total 6 possible reading frame (6 horizontal bars) would be there for every DNA sequence. The 6 possible reading frames are +1, +2, +3 and -1, -2 and -3 in the reverse strand. The resultant amino acids can be saved and search against various protein databases using blast for finding similar sequences or amino acids. The result displays the possible protein.

NSP5, NSP6 and NSP7 ligand With Target protein (7KMB6) CB docking

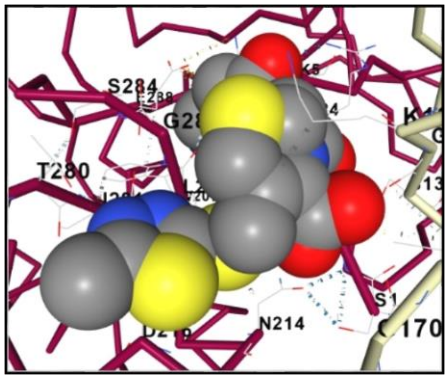


Figure 4

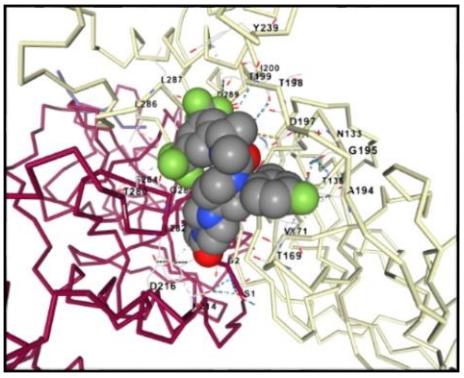


Figure 5

Figure 4: - protein ligand interaction between ligand (7MB6) and drug Cefazoline of NSP5 and
Figure 5: - protein ligand interaction between ligand (7MB6) and drug Casopitant of NSP5

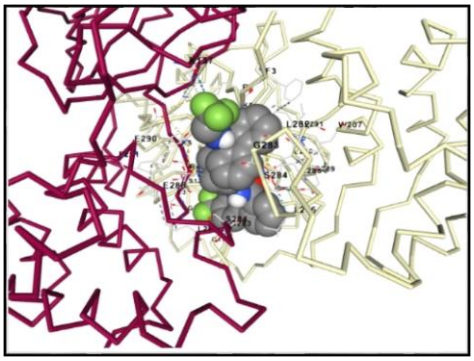


Figure 6

Figure 7

Figure 6: - protein ligand interaction between ligand (7MB6) and drug lomitapide of NSP6 and
Figure 7: - protein ligand interaction between ligand (7MB6) and drug Methaqualone of NSP7

Table 1: - Auto dock vina score of ligand NSP5, NSP6, NSP7 and target protein 7MB6 with their effective Drug

Sr no	PDB File Name (7MB6)	Drug Name	Autodock Vina Score	Center Axis	Docking size	Binding Affinity (protein- ligand interaction
1	NSP5	Cefazoline	-8.0	22,1,30	24,31,24	Good binding affinity
2	NSP5	Casopitant	-9.2	22,1,30	27,27,27	Stronger binding Affinity
3	NSP5	Progesterone	-7.3	22,1,30	27,31,20	Moderately strong binding
4	NSP6	Lomitapide	-9.5	22,1,30	30,30,30	Stronger binding affinity
5	NSP7	Methaqualone	-8.5	31,12,33	19,19,30	Good binding affinity
6	NSP7	Setiptiline	-8.3	31,12,33	19,19,30	Good binding affinity

Docking score analysis -7.3 in CB dock indicate moderately strong binding affinity between the ligand (NSP5, NSP6 and NSP7) and the target protein (7MB6). Docking score analysis -8,-8.3,-8.5 in CB dock (for any molecular docking software) indicate good binding affinity between the ligand (NSP5, NSP6, and NSP7) and the target protein (7MB6).Docking score analysis -9,-9.2,-9.5 in CB dock indicate an even stronger binding affinity between the ligand (NSP5, NSP6, and NSP7) and the target protein (7MB6). The target protein score -8 to-9.5 reflects high affinity implying that ligand fit well into the binding site. While -7.3 it's not as strong as -8 or -9.5 score but -7.3 still suggest a favorable interaction between target protein and the ligand. Expected validation and comparison can help to determine its true potential.

NSP5, NSP6 AND NSP7 ligand with Target Protein (8OKB)

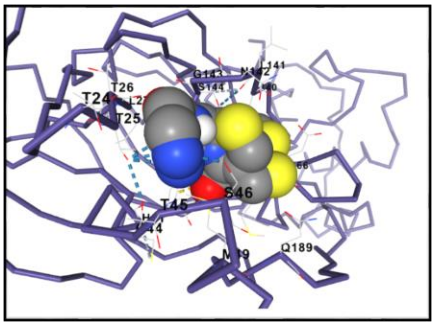


Figure 8

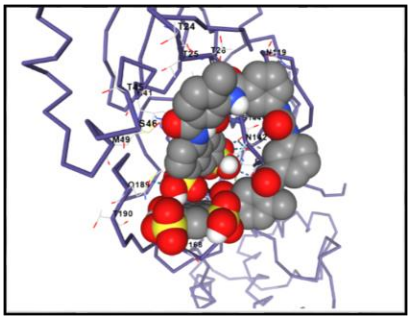


Figure 9

Figure 8:- Protein ligand interaction between ligand 8OKB and drug Cephazolin (NSP5)

Figure 9:- 8OKB with drug Suramin (NSP5)

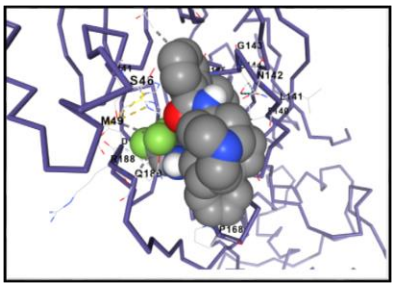


Figure 10

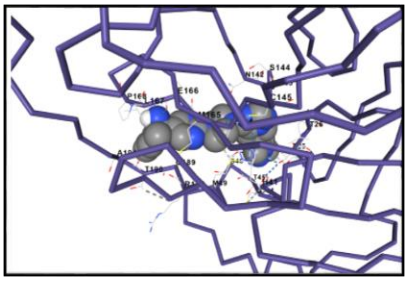


Figure 11

Figure 10: - protein ligand interaction between ligand (80KB) & lomitapide of NSP6 and

Figure 11:- protein ligand interaction between 80KB & Temoporfin of NSP6

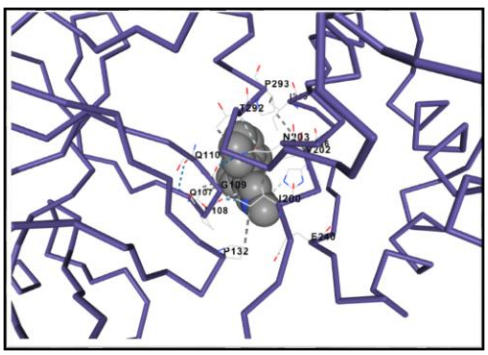


Figure 12

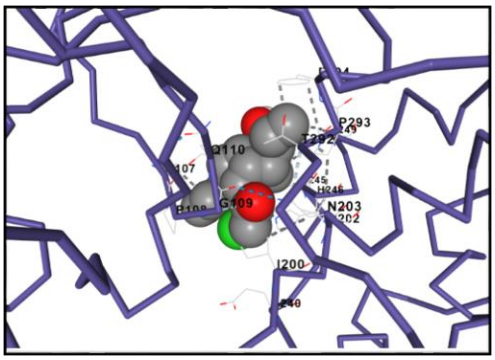


Figure 13

Figure 12: - protein ligand interaction between ligand (80KB) & Setiptiline of NSP7 and Figure 13 : - protein ligand interaction between ligand (80KB) & Lorajmine of NSP7

Interpreting the docking score using CB dock 2 shows the biological significance of target site and validation of the data obtained through experiments by utilizing computational algorithm. Our sample 8OKB with drug casopitant shows binding affinity of -9.5 and with Dactinomycin shows -10.5 docking score which indicates a stronger binding affinity. Docking of sample (8OKB) with Lomitapide and Temoporfin which are the drug molecules used as NSP6 shows binding affinity of -9.1 and -9.9 respectively which indicates very strong binding affinity. Docking of 8OKB with setiptiline and methaquinone shows binding affinity of -7.1 and -7.7 respectively which indicates moderately between protein (8OKB) and the ligand (NSP7).

Table 2: - Autodock vina score of ligand NSP5, NSP6, NSP7 and target protein 8OKB with their effective Drug

S.No	PDB ID	Drug Name	Docking Score	Remarks
1.	8OKB (NSP5)	Cephazolin	-7.5	Moderately strong binding
2.		Suramin	-9.5	Stronger Binding affinity
3.		Casopitant	-9.5	Stronger binding affinity
4.		Dactinomycin	-10.1	Very strong binding affinity
5.		Progesteron	-7.4	Moderately strong binding
6.	8OKB(NSP6)	Lomitapide	-9.1	Stronger binding affinity
7.		Temoporfin	-9.9	Very strong binding affinity
8.		Avapritinib	-8.9	Stronger binding affinity
9.	8OKB(NSP7)	Setiptiline	-7.1	Moderately strong binding
10.		Metapramine	-7.1	Moderately strong binding
11.		Lorajmine	-7.1	Moderately strong binding
12.		Methaqualone	-7.7	Moderately strong binding

Deep fold

Deep fold enhancing protein structure prediction alpha full protein structure database provide protein structure prediction for human proteome for deep learning on computational biology Deep fold analysis refers to the using advanced computational technique Graph Neural network: GNN represent the graph (with amino acid node) (protein folding) RNN (Recurrent Neural Network) used in sequence based model capture the sequence-based nature amino acid Future Direction Integration with experimental data Folding kinetics (pathway and time scale of folding)Distance map or contact map is a matrix represent distance between two amino acid .here analysing this to check the folding compactness and structural flexibility. By interpreting the main distance low mean distance indicate most of the residue are closely packed and in high mean distance it might be elongated or disorder protein represent multi domain protein.

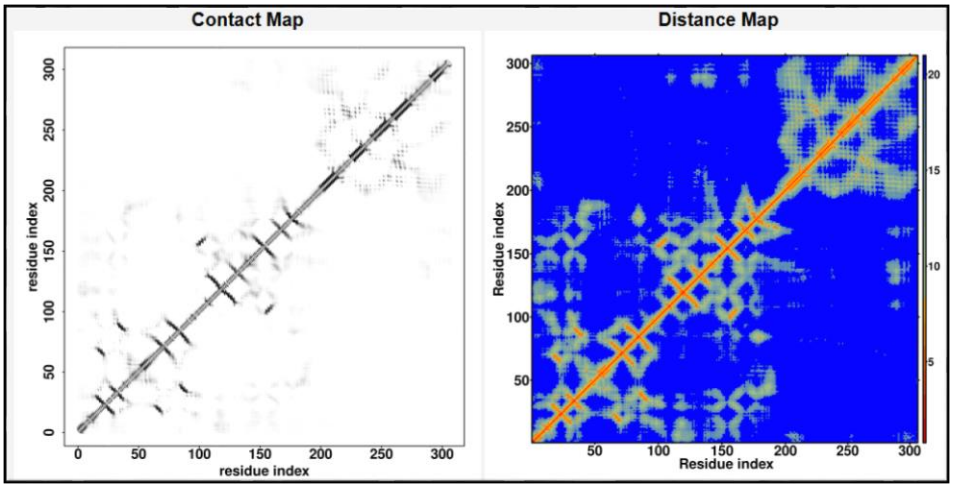


Figure 14:- The Deep Fold simulations are guided by the consensus contact map (left figure), distance map (middle figure). In the contact and distance maps, the axes mark the residue index along the sequence. For the contact map, each dot represents a residue pair predicted to form a contact, while for the distance, a colour scale represents a distance of 1-20+ angstroms or 0-360 degrees, respectively

Bepipred Linear Epitope Prediction 2.0 Results (B-Cell epitope prediction)

Input Sequences

1 SGFRKMAFPS GKVEGCMVQV TCGTTTLNGL WLDDVVYCPR HVICTSEDML NPNYEDLLIR

61 KSNHNFLVQA GNVQLRVI GHSMQNCVLK LKVDTANPKT PKYKFVRIQP GQTFSVLACY

121 NGSPSGVYQC AMRPNFTIKG SF LNGSCG SVGFNIDYDC VSFCYMHHME LPTGVHAGTD

181 LEGNFYGPFV DRQTAQAAGT DTTITVNVLA WLYA AVIN GDRWFLNRFT TTLNDFNLVA

241 MKYNYEPLTQ DHVDILGPLS AQTGIAVLDM CASLKELLQN GMNGRT IL GSALLEDEFT

301 PFDVVRQCSG VTFQ

Predicted peptides:

No.	Start	End	Peptide	Length
1	5	13	KMAFPSGKV	9
2	47	57	EDMLNPNYEDL	11
3	93	109	TANPKTPKYKFVRIQPG	17
4	170	196	GVHAGTDLEGNFYGPFVDRQTAQAAGT	27
5	225	228	TTLN	4
6	236	247	KYNYEPLTQDHV	12
7	273	278	QNGMNG	6
8	290	298	EFTPFDVVR	9
9	301	302	SG	2

Predicted residue scores:

Position	Residue	Score	Assignment
0	**S**	0.269	.
1	**G**	0.323	.
2	**F**	0.385	.
3	**R**	0.453	.
4	**K**	0.524	E
5	**M**	0.541	E
6	**A**	0.552	E
7	**F**	0.568	E
8	**P**	0.564	E
9	**S**	0.564	E
10	**G**	0.559	E
11	**K**	0.532	E
12	**V**	0.507	E
13	**E**	0.458	.

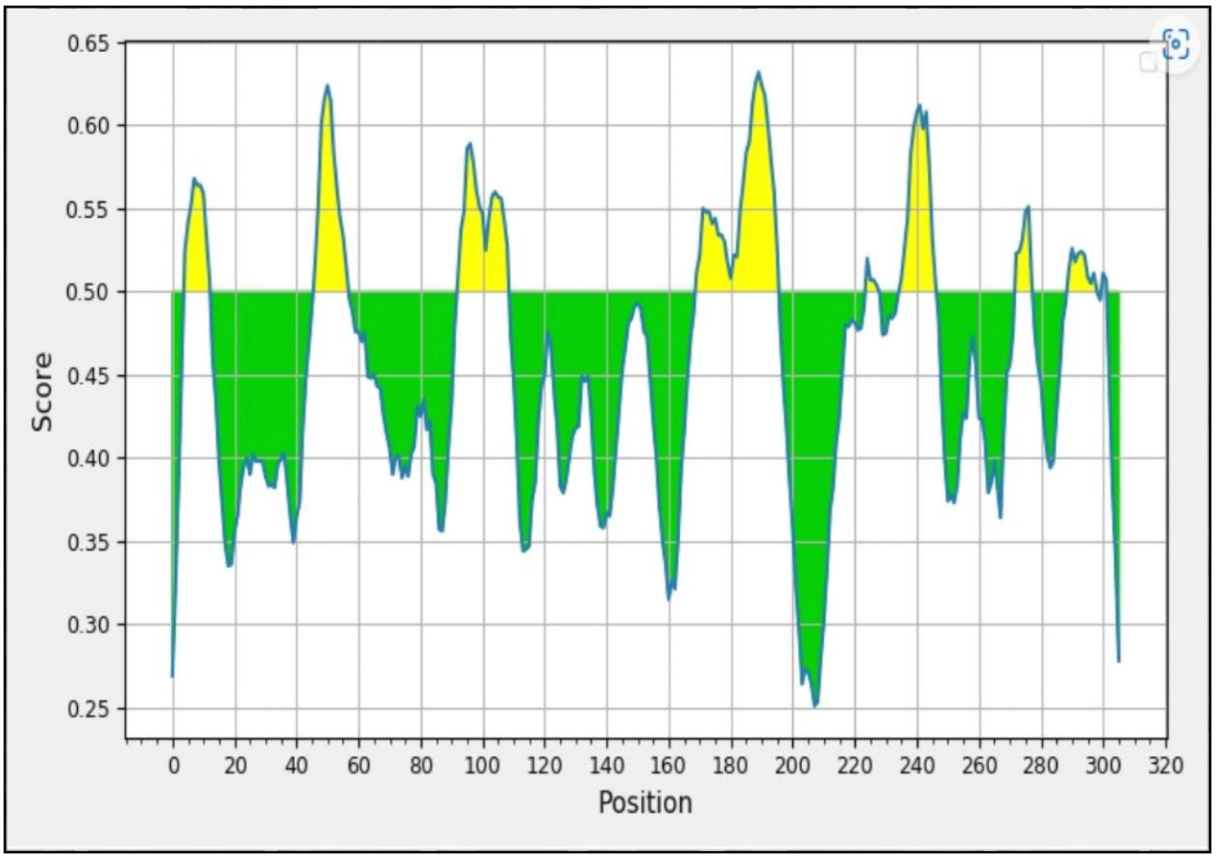


Figure 15:-Graph shows Bepiperd linear epitope prediction graphical analysis (yellow color shows B cell epitope)

B-cell epitope prediction is critical to understanding to immune response and we conclude that adding in the design of vaccine, diagnostic and antibody therapy focusing on the most immunogenic region of the antigen.

Biopyhton

```
from prody import *
Chain_8okb = parsePDB('8okb')
showProtein(Chain_8okb)
@> Connecting wwPDB FTP server RCSB PDB (USA).
DEBUG:.prody:Connecting wwPDB FTP server RCSB PDB (USA).
@> 8okb downloaded (8okb.pdb.gz)
DEBUG:.prody:8okb downloaded (8okb.pdb.gz)
@> PDB download via FTP completed (1 downloaded, 0 failed).
DEBUG:.prody:PDB download via FTP completed (1 downloaded, 0 failed).
@> 2457 atoms and 1 coordinate set(s) were parsed in 0.05s.
DEBUG:.prody:2457 atoms and 1 coordinate set(s) were parsed in 0.05s.
<Axes3D: xlabel='x', ylabel='y', zlabel='z'>
```

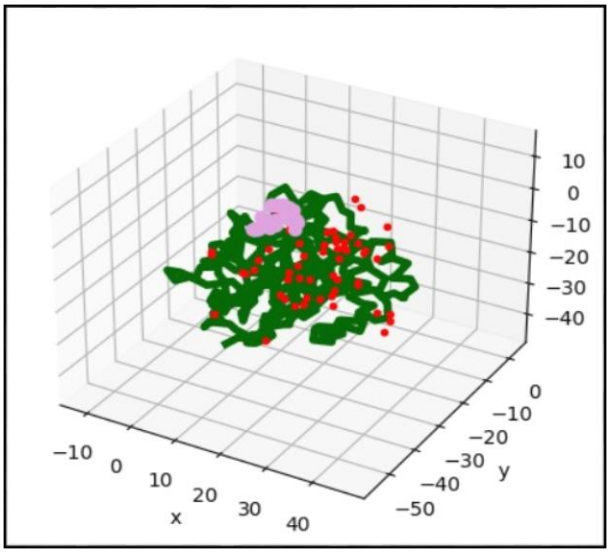


Figure 16:- Graphical representation of 3D structure with bio python code by using PDBParser` for parsing these files to extract atomic coordinates, residue position, and chain details.

4. Discussion

The result indicate various structural and functional features of NSP5, 6 and 7 of SARS CoV2. Understanding the genome sequence and analysing protein sample 8OKB of NSP5 and 7MB6 protein sample of NSP6 and 7 is crucial for predicting protein structure and in the development of targeted drug delivery. The objective of our study was to computationally analyse how protein sample with PDB ID 8OKB and 7MB6 isolated from human blood sample. For better classification of our protein sample, identifying its domain and active site and for the comprehensive understanding of the particular function of the protein sample we used Scan prosite which gave us valuable information about motif and profile. ORF finder tool helped us in identification of various possible open reading frame in the RNA sequence of the sample by using alternative genetic code to analyse all the possible

open reading frames in our sample sequence which is quite useful in healthcare industry for personalized medicine and improvement of the patient outcome. Then CB dock 2 tool is utilized which uses several computational algorithm to dock protein sample and the ligand from which we were able to analyse the binding affinity between the two. By interpreting the docking score we showed the biological relevance of 8OKB and 7MB6 against the drugs identified through the pubchem and its validation through the computational data for further analysis. Deep fold analysis tool which is a deep learning tool to predict the protein structure we were able to analyse protein folding compactness and structural flexibility by interpreting the contact or distance map. Then IEDB analysis tool which is an Epitope Prediction tool was used to predict the epitope in our protein sample against which B and T cell could possibly induce an immune response which can be further find its application in reverse vaccinology and development of targeted drug delivery against SARS CoV2. Lastly Bio python script was used to import various libraries and sequence data used to analyse various atoms, residues, their distance and angle in our protein sample.

5. Conclusion

The study explores the structural and functional features of NSP5, NSP6, and NSP7 of SARS-CoV-2 by the analysis of genome sequences and protein samples 8OKB (NSP5, NSP6,NSP7) and 7MB6 (NSP5, NSP6 and NSP7), this research focused to predict protein structures and develop targeted drug with the help of bioinformatics tools. We used different tools like Expasy Scan Prosite for motif and profile identification. ORF Finder for identifying open reading frames in DNA sequences. CB-Dock 2 for docking protein samples and ligands to analyze binding affinities. Deep Fold for predicting protein structures, compactness, and flexibility. IEDB for predicting B and T cell epitopes. Bio-python scripts were used to import libraries and sequence data to analyze atoms, residues, their distances, and angles. This comprehensive approach provides significant insights into the structural and functional characteristics of SARS-CoV-2 non-structural proteins, aiding in the development of therapeutic strategies.

Acknowledgment:

We thank the whole BPRI team for their knowledge transfer to us. We thank to Dr Uma Kumari for structural based drug designing, prediction of epitope, next generation sequencing and molecular docking technique research and Atul Sir to their insight on bio python.

References:

[1] Nancy ,Uma Kumari BIOINFORMATICS PIPELINE TO STUDY THE VIRAL SEQUENCE OF HUMAN CORONAVIRUS NL63 PROTEASE 2022 JETIR May 2022, Volume 9, Issue 5 www.jetir.org (ISSN-2349-5162)

[2] Cui J., Li F., Shi Z.-L. Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. 2019;17(3):181–192. doi: 10.1038/s41579-018-0118-9. [DOI] [PMC free article] [PubMed] [Google Scholar]

[3] Jauregui A.R. Identification of residues of SARS-CoV nsp1 that differentially affect inhibition of gene expression and antiviral signaling. PloS One. 2013;8(4) doi: 10.1371/journal.pone.0062416. [DOI] [PMC free article] [PubMed] [Google Scholar][Ref list] jahangir Emrani [a,*], Maryam Ahmed [b], Liesl Jeffers-Francis [c], John C Teleha [d], Nathan Mowa [b], Robert H Newman [c], Misty D Thomas

[4] Gu J., Korteweg C. Pathology and pathogenesis of severe acute respiratory syndrome. Am. J. Pathol. 2007;170(4):1136–1147. doi: 10.2353/ajpath.2007.061088. [DOI] [PMC free article] [PubMed] [Google Scholar]

[5] Lin L, Lu L, Cao W, Li T. Hypothesis for potential pathogenesis of SARS-CoV-2 infection-a review of immune changes in patients with viral pneumonia. Emerg Microbes Infect. 2020;9:727–732. doi: 10.1080/22221751.2020.1746199. [DOI] [PMC free article] [PubMed] [Google Scholar]

[6] Mart M. Lamers & Bart L. Haagmans *Nature Reviews Microbiology* **volume 20**, pages270–284 (2022)

[7] Liu Y., Zhang C., Huang F., Yang Y., Wang F., Yuan J., Zhang Z., Qin Y., Li X., Zhao D., Li S., Tan S., Wang Z., Li J., Shen C., Li J., Peng L., Wu W., Cao M., Xing L., Xu Z., Chen L., Zhou C., Liu W.J., Liu L., Jiang C. Elevated plasma level of selective cytokines in COVID-19 patients reflect viral load and lung injury. Natl. Sci. Rev. 2020;7(6):1003–1011. doi: 10.1093/nsr/nwaa037. [DOI] [PMC free article] [PubMed] [Google Scholar]

[8] Chen G, Wu D, Guo W, Cao Y, Huang D, Wang H, Wang T, Zhang X, Chen H, Yu H, et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. J Clin Invest. 2020;130:2620–2629. doi: 10.1172/JCI137244. [DOI] [PMC free article] [PubMed] [Google Scholar]

[9] Adam S.B., Sukhramani G.S., Bishnoi P., Pable A.A., Barvkar V.T. SARS-CoV-2, the pandemic coronavirus: molecular and structural insights. J. Basic Microbiol. 2021;61(3):180–202. doi: 10.1002/jobm.202000537. [DOI] [PMC free article] [PubMed] [Google Scholar][Ref list]

[10] Homs M., Buschauer R., Ameismeier M., Koepke L., Denk T., Hirschenberger M., Kratzat H., Hayn M., Mackens-Kiani T., Cheng J., Straub J.H., Stürzel C.M., Fröhlich T., Berninghausen O., Becker T., Kirchhoff F., Sparrer K.M.J., Beckmann R. Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. Science. 2020;369(6508):1249–1255. doi: 10.1126/science.abc8665. [DOI] [PMC free article] [PubMed] [Google Scholar][Ref list]

[11] Di Gioacchino A., 2021) Šulc P., Komarova A.V., Greenbaum B.D., Monasson R., Cocco S. The heterogeneous landscape and early evolution of pathogen-associated CpG dinucleotides in SARS-CoV-2. Mol. Biol. Evol. 2021;38(6):2428–2445. doi: 10.1093/molbev/msab036. [DOI] [PMC free article] [PubMed] [Google Scholar][Ref list])

[12] Benvenuto D. Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. J. Infect. 2020;81(1):e24–e27. doi: 10.1016/j.jinf.2020.03.058. [DOI] [PMC free article] [PubMed] [Google Scholar][Ref list

[13] Raj, R. (2020). Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing[1]. *Biochem Biophys Rep*, 25, 100847. doi: 10.1016/j.bbrep.2020.100847[2]. PMCID: PMC7750489[3]. PMID: 33364445

[14] Kumari Uma, and Shruti Gupta. "NGS and Sequence Analysis with Biopython for Prospective Brain Cancer Therapeutic Studies." 2023

[15] Uma Kumari, Abhishek Ebenezar,etal;2024, Next Genration sequencing and Proteomics analysis of BRD2-BD1 complex with MDP5 in Medulloblastoma: COMPUTATIONAL ANALYSIS AND NGS PIPELINE FOR HUMAN GRK2 IN COMPLEX WITH G-BETA-GAMMA IN CARDIOVASCULAR DISEASE WITH BIOPYTHON,2024; Volume 6 , Issue - 14 : Page: 1993-2012 African Journal of Biological Sciences doi: 10.48047/AFJBS.6.14.2024.1992-2012

[16] Uma Kumari, Unnayni Joshi, 2024 "Targeted NGS Proteomics analysis on M342R Mutation and 17+1mer Oligonucleotide with Triplet GGT", International Journal of Emerging Technologies and Innovative Research,Vol.11, Issue 6, page no. ppj785-j793, June-2024, http://doi.one/10.1729/Journal.40303

[17] Roma Sharma, Uma Kumari, 2024. Exploring the binding affinity landscape of SARS-CoV-2 variants: a computational approach, V 13 - I 3, Pages- 6566 – 6569. Doi: https://doi.org/10.55522/jmpas.V13I3.6555.

[18] Uma Kumari, Vineeta Johri, Swarali Dhopate, Tijil Jha Structure Based Drug Designing for the prediction of epitope for targeting Malignant brain Tumor, 2024 JETIR July 2024, Volume 11, Issue 7 www.jetir.org (ISSN-2349-5162)

[19] Vipasha Rathi,Uma Kumari,2024 "Identification of active compound and NGS analysis BRD4-BD2 in complex with SF2523 of human malignant brain tumor", International Journal of Emerging Technologies and Innovative Research Vol.11, Issue 8,ppb597-b606, https://doi.one/10.1729/Journal.40957

[20] Uma Kumari,Gurpreet Kaur *et al*,2024,"Biopython/Network Of Protein Identification And NGS Analysis Of Glioma Cancer ATP Competitive Type III C-MET Inhibitor : 7.367 (Calculated by Google Scholar) : Volume 11, Issue 2 : 27-Jun-2024 :pp 41-51 : http://doi.one/10.1729/Journal.40229

[21] Uma Kumari, Shipra Chaudhary, next generation sequencing, molecular docking and network pharmacology reveal potent inhibitor for the treatment of lung cancer, 2024 JETIR September 2024, Volume 11, Issue 9, www.jetir.org (ISSN-2349-5162) ,JETIR2409544 , Journal of Emerging Technologies and Innovative Research (JETIR) www.jetir.org

[22] Uma Kumari,Stuti,;2021 Genome Sequence Analysis of Beta Coronavirus by Applying Bioinformatics Tools ***BJBS 1(1): 49-54***,**DOI**: 10.32996/bjbs.2021.1.1.4