Journal of Entomology and Zoology Studies

Available online at www.entomoljournal.com

E-ISSN: 2320-7078
P-ISSN: 2349-6800
www.entomoljournal.com
JEZS 2020; 8(2): 06-10

Received: 03-01-2020
Accepted: 07-02-2020

Swagatika Priyadarsini
Division of Biochemistry, Indian Veterinary Research Institute, Bareilly, Uttar Pradesh, India

Snehasmita Panda
Division of Animal Genetics and Breeding, Indian Veterinary Research Institute, Bareilly, Uttar Pradesh, India

Rohit Singh
Division of Pathology, Indian Veterinary Research Institute, Bareilly, Uttar Pradesh, India

Amrita Behera
Division of Biochemistry, Indian Veterinary Research Institute, Bareilly, Uttar Pradesh, India

Prachurya Biswal
Livestock Production and Management Section, Indian Veterinary Research Institute, Bareilly, Uttar Pradesh, India

Puja Mech
Division of Microbiology, Indian Veterinary Research Institute, Bareilly, Uttar Pradesh, India

Karthikeyan Ramaiyan
Division of Biochemistry, Indian Veterinary Research Institute, Bareilly, Uttar Pradesh, India

Corresponding Author:
Snehasmita Panda
Division of Animal Genetics and Breeding, Indian Veterinary Research Institute, Bareilly, Uttar Pradesh, India

In silico structural delineation of nucleocapsid protein of SARS-CoV-2

Swagatika Priyadarsini, Snehasmita Panda, Rohit Singh, Amrita Behera, Prachurya Biswal, Puja Mech and Karthikeyan Ramaiyan

Abstract
Among the four structural proteins of SARS-CoV-2, nucleocapsid (N) is a highly conserved protein based on its structure and sequence. In the recent past, X-ray crystallographic and NMR studies performed by various researchers on the N protein of SARS-CoV revealed that it is a multifunctional protein with the primary role of binding to the genomic RNA of the virus and forming ribonucleoprotein complex. The protein has been observed to be involved in the regulation of host cellular homeostasis in addition to its function in regulating the viral genome replication and transcription. An attempt has been made to reveal the biochemical and structural characteristics of the nucleocapsid protein of SARS-CoV-2 using several online tools. The predicted secondary and tertiary structures of the protein can be used to map various epitopes and also to hunt for several drug binding sites in future.

Keywords: Nucleocapsid, Domain, Tertiary Structure, NBD

Introduction
Like the etiological agents of previous Coronavirus (CoV) outbreaks in human in the past two decades- SARS-CoV and MERS-CoV, the SARS-CoV-2 that causes COVID-19 is also a *Betacoronavirus* with approximately 30kb of single-stranded (ss)-RNA as its genomes [1].
Four major structural proteins of the SARS-CoV-2 include spike (S) glycoprotein, small envelope (E) protein, matrix (M) protein and nucleocapsid (N) protein. Apart from the structural proteins, CoV also has 15 non-structural proteins (nsp) such as nsp1-10 and nsp12-16 and 8 accessory proteins such as 3a, 3b, p6, 7a, 7b, 8b, 9b, and orf14 [2]. In brief, the S-protein is a transmembrane protein trimer that helps viral entry into host cell by attaching to the host cell angiotensin-converting enzyme (ACE)-2 receptor of the type-II pneumocyte of the lungs. E-protein is a tri-domain integral membrane protein which forms viro-porin. M-protein binds to nucleocapsid while viral packaging and hence gives shape to the viral envelope and lastly the N-protein is one of the most conserved proteins across several families of CoV that has multiple functions [3]. N-protein of SARS-CoV was found to have three specific domains- RNA-binding site present in the N-terminal domain (NTD), protein dimerization site in the carboxy terminal domain (CTD) and a linker region that connects these former two domains [4].
The NTD of the protein binds in a sequential manner to the viral RNA to form a ribonucleoprotein (RNP) complex before viral packaging inside host cell. Other secondary functions of this protein are to differentially regulate viral genome replication and transcription and, in the host, to inhibit translation machinery and cellular proliferation, alter cell cycle and apoptosis etc [5].
A single antiviral drug has not been discovered yet for the treatment of COVID-19 specifically, which forces the healthcare system to utilize the existing antivirals in a hit and trial method. Most of the studies are performed to invent a ligand that can target and inhibit the spike (S) protein of the virus. However, spike protein is highly evolutionary as any occurrence of mutation its receptor binding domain (RBD) determines the host for the virus [6]. Therefore, the most conserved protein of the virus that is nucleocapsid protein can be explored for finding the drug-binding sites [4]. The objective of this study was to analyze the physical and biochemical properties of the nucleocapsid protein of SARS-CoV-2 and to decipher its secondary and tertiary structure by the use of various online bioinformatic tools.

Materials and Methods

Sequence retrieval

Amino acid sequences of nucleocapsid proteins of SARS-CoV-2, SARS-CoV and MERS-CoV were downloaded in FASTA format from the UniProt database (https://www.uniprot.org/) with accession numbers as NCAP_SARS2, Q6S8E1_CVHSA and NCAP_CVEMC respectively. Genomic sequence of SARS-CoV2 was retrieved in FASTA format from NCBI genome database (https://www.ncbi.nlm.nih.gov/genome/?term=SARS+COV2) with accession number NC_045512.2.

Sequence alignments

Protein sequence with accession number NCAP_SARS2 was used as query to run a protein BLAST against all available proteins in the database as subject using UniProt blast (https://www.uniprot.org/blast/). Further all the three protein sequences NCAP_SARS2, Q6S8E1_CVHSA and NCAP_CVEMC were subjected to BLASTp in NCBI server (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins).

Physicochemical features of SARS-CoV-2 N protein

The FASTA sequence of SARS-CoV-2 N-protein was subjected to ProtParam server (https://web.expasy.org/protparam/) to obtain several characteristics of the protein. Hydropathy of the protein was identified by Kyte and Doolittle algorithm and linear weight variation model in ProtScale server (https://web.expasy.org/protscale/) keeping window size at 9 (1). Intrinsically disordered motifs were identified by using DEPICTER online tool (http://biomine.cs.vcu.edu/servers/DEPICTER/) (3).

Domain identification

Protein domains were predicted using InterProScan database from EMBL (https://www.ebi.ac.uk/interpro/search/sequence/). Further, several motifs of the protein were revealed online by ScanProsite tool from ExPASy (https://prosite.expasy.org/scanprosite/).

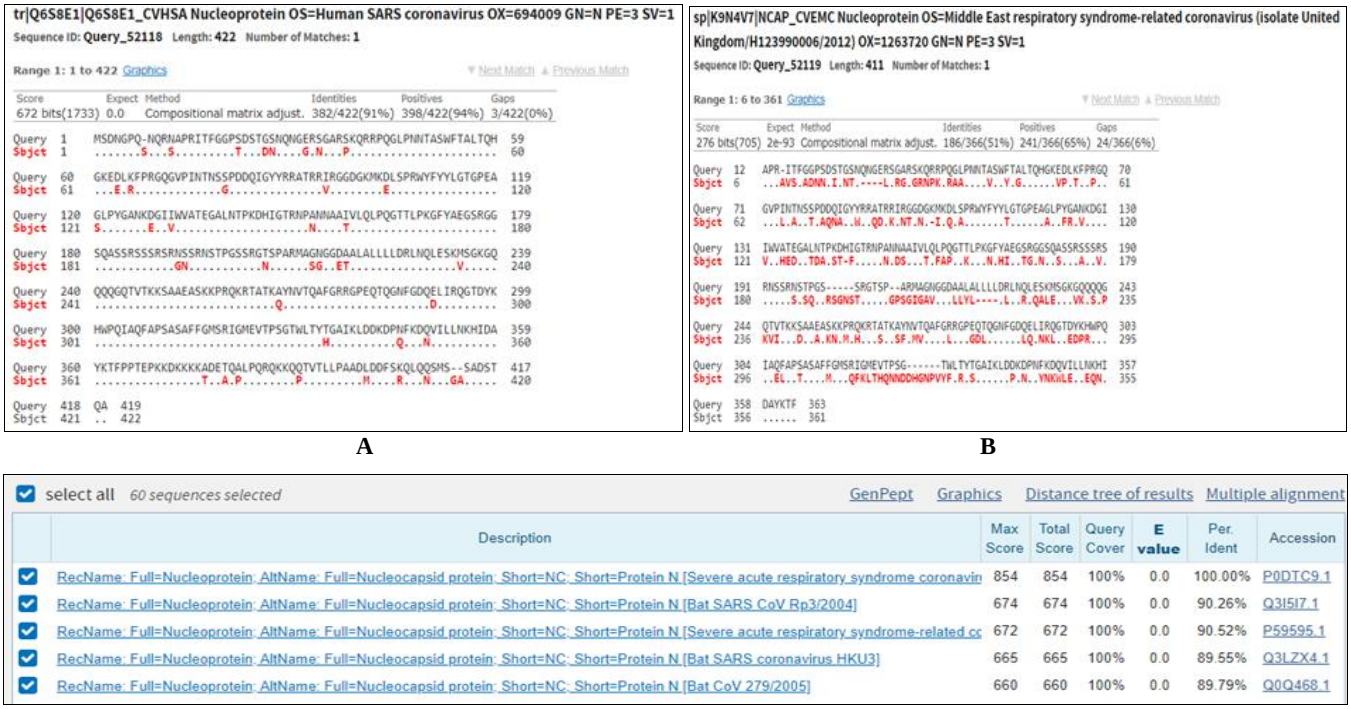
Structure prediction

Secondary structure of the nucleocapsid protein was predicted by subjecting the sequence to SOPMA (https://npsa-prabi.ibcp.fr/cgi-in/npsa_automat.pl?page=/NPSA/npsa_sopma.html), keeping the similarity threshold at 8. Further the structure was validated in PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) [7]. For prediction of protein conformations and its allowed regions, the structure was analyzed in SAVES server ((http://servicesn.mbi.ucla.edu/SAVES/) developed by Wodak *et al*. [8]. Tertiary structure for the sequence was predicted by Phyre2 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) with intensive modelling mode [9].

Results

Sequence alignment

Local alignment of amino acid sequences of the three nucleocapsid proteins of SARS-CoV-2, SARS-CoV and MERS-CoV performed using BLASTp from the NCBI server revealed that there is 90.52% identity (E value = 0) between the first two proteins and 50.82% identity between the first and third protein. Further protein BLAST of the first sequence was done subjecting to all the proteins in the UniProt database and it was found that the N-protein of SARS-CoV2 has a maximum of 90.52% identity with human SARS-CoV (UniProt accession number- P59595.1) followed by 90.26% identity with bat CoV Rp3/2004 (UniProt accession number- Q3I5I7.1). The BLAST results are shown in fig.1.



A B

Fig 1: Amino acid sequence alignment for nucleocapsid protein of SARS-CoV-2. (A) Pairwise alignment with Human-SARS-CoV, (B) Pairwise alignment with MERS-CoV, (C) Pairwise alignment with proteins in the uniport database.

Physicochemical properties

The nucleocapsid protein of SARS-CoV-2 has 419 amino acids and its sequence was analyzed in the ProtParam server which revealed that the protein has sixty numbers of positively charged amino acids and thirty-six negatively charged amino acids. Presence of 14% positively charged amino acids in the complete sequence makes the protein basic and hence explains its ability to bind to RNA and form RNP complex. Other characteristics are explained in table 1.

Table 1: Physicochemical properties of nucleocapsid protein of SARS-CoV-2

Physicochemical properties	Values
Number of amino acids:	419
Positively charged amino acids:	60
Negatively charged amino acids:	36
Molecular weight:	45.6kDa
Isoelectric pH:	10.07
Ext. coefficient:	43890
Theoretical half-life:	30hrs (mammalian reticulocytes, *in vitro*)
Instability index (Degree of instability):	55.09
Average of hydropathicity:	-0.971

Protein motifs and domains

Amino acid sequence of the protein was uploaded to the InterPro server to predict the domains and it was found that sequence ranging from 27 to 180 contain RNA-binding site and sequence 251 to 364 contain protein dimerization domain. Detailed report is given in fig. 2. ScanProsite analysis revealed that the protein has a serine-rich region (176-206) and also sites for N-myristoylation, N-glycosylation, casein kinase II phosphorylation, cAMP- and cGMP-dependent protein kinase phosphorylation, protein kinase C phosphorylation and amidation.

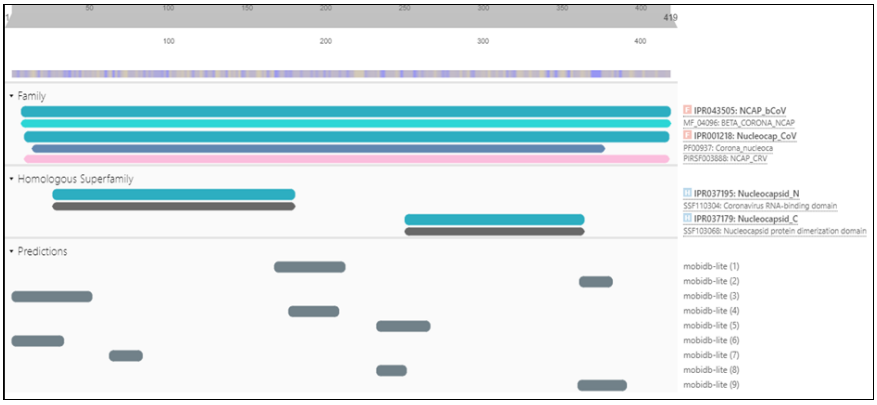


Fig 2: Domains predicted for nucleocapsid protein of SARS-CoV-2 using InterPro server showing NTD, Linker region and CTD from left to right direction of figure in the homologous superfamily section.

Predicted structure for nucleocapsid protein:

Kang *et al*. [10] has already developed the crystal structure for the NTD of N-protein of SARS-CoV-2 and submitted to the PDB database. But here we have subjected the whole sequence to SOPMA site which revealed that the protein contains 20.05% alpha helix, 17.18% beta sheet, 10.74% turns and 52.03% coil in its secondary structure, further details are shown in fig.3A. The result obtained from PSIPRED is depicted in fig. 3B. To estimate the accuracy of the secondary structural components of the protein, Ramachandran plot was developed for the same using SAVES online tool. It was observed from the plot that 24.4% of the conformational region lies in the allowed region and 5.5% of the conformation lies in the disallowed region and the detailed report is shown as in table 2 (Fig.4). The tertiary structure of the protein was predicted by heuristics using *ab-initio* model and assessed by subjecting the FASTA sequence to Phyre2 (Fig.5). The tertiary model can be explored for finding of drug and ligand binding sites using a number of softwares.

Table 2: Report of Ramachandran plot for nucleocapsid protein of SARS-CoV-2 via PROCHECK in SAVES server.

```
+----------<<< P R O C H E C K   S U M M A R Y >>>----------+
|                                                            |
| /var/www/SAVES/Jobs/3047870/3047870.pdb   1.5   419 residues |
|                                                            |
*| Ramachandran plot:   61.6% core   24.9% allow   8.1% gener   5.5% disall |
|                                                            |
*| All Ramachandrans:   94 labelled residues (out of 417)     |
*| Chi1-chi2 plots:      8 labelled residues (out of 234)     |
 | Side-chain params:    5 better     0 inside     0 worse    |
|                                                            |
*| Residue properties: Max.deviation:    25.0      Bad contacts:  144 |
*|                     Bond len/angle:   48.6    Morris et al class:  3  2  3 |
+|     1 cis-peptides                                         |
*| G-factors           Dihedrals:  -0.57  Covalent:  -6.53   Overall:  -2.68 |
|                                                            |
+| Planar groups:    98.1% within limits   1.9% highlighted   |
|                                                            |
+------------------------------------------------------------+
  + May be worth investigating further.  * Worth investigating further.
```

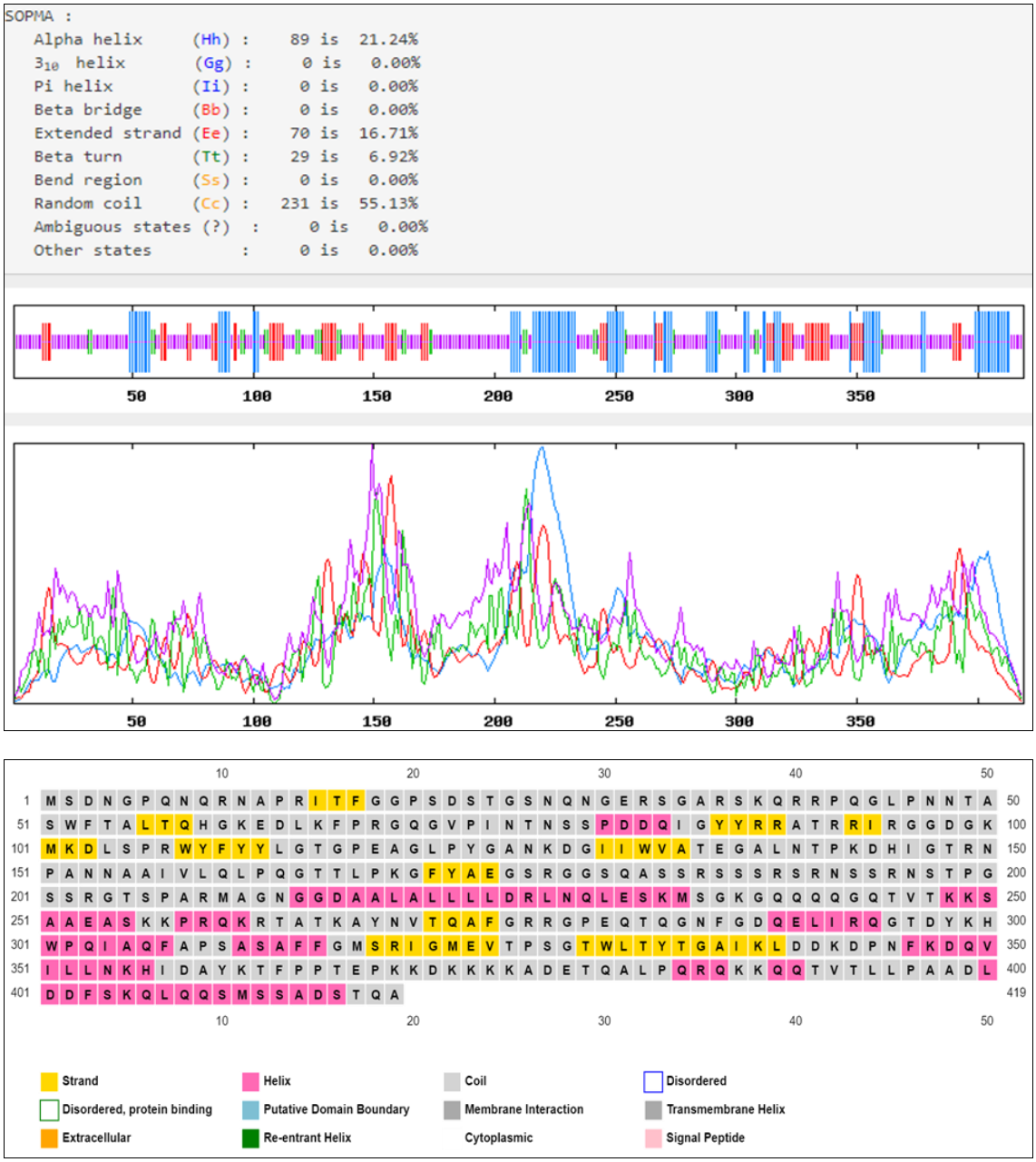


Fig 3: Predicted secondary structure for the nucleocapsid protein of SARS-CoV-2. (A) Structural composition predicted in SOPMA that indicates helices as blue, strands as red, turns as green and coil as orange, (B) Structural composition predicted in PSIPRED.

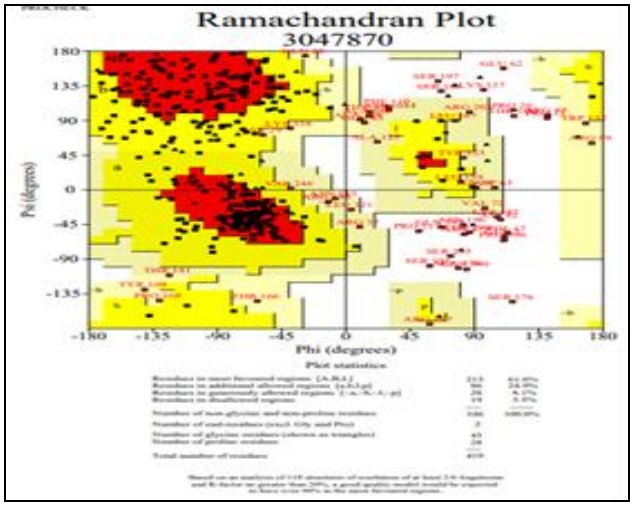


Fig 4: Ramachandran plot developed for the nucleocapsid protein of SARS-CoV-2 using SAVES server

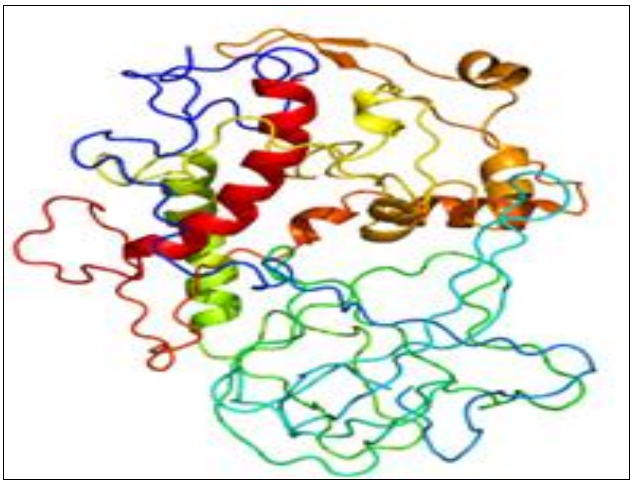


Fig 5: Assessment of tertiary structure for the nucleocapsid protein of SARS-CoV-2 by Phyre2 using *ab-initio* model. The structure is coloured in RAINBOW from N to C terminal.

Discussion

The primary function of the nucleocapsid protein of Coronavirus is to bind to its genomic RNA and regulate its multiplication inside the eukaryotic host [3]. Besides the protein have several other functions like manipulation of homeostasis, proliferation and translation machinery of the host cell etc. [11]. In present investigation, results obtained from InterPro and SOPMA analysis reveled that the protein has a rich portion of intrinsically disordered structures. Hence, multifunctional biological ability of this protein can be hypothesized [12, 13]. Similar results were also obtained in this study from the Ramachandran plot predicted for the protein structure. In our investigation, based on the sequence alignment using BLASTp, it can be suggested that bat is somehow involved in the origin of SARS-CoV-2 as the identity reaches 90.26% for a bat SARS-COV-like coronavirus. Our results were in agreement with Andersen *et al*. [14] who reported a natural selection of this virus in an animal host like bat before zoonotic transfer. However, it still needs much extensive surveillance to trace the real source of virus. Presence of more basic amino acids (positively charged) as compared to the contribution of acidic amino acids, defines the ability of the N-protein to bind to the negatively charged ribonucleic acid.

Further, the data obtained from InterPro analysis agree with the fact that NTD is responsible for RNA binding and CTD is for protein dimerization as seen in the case of N-protein of SRAS-CoV [15]. As the protein contains a serine region and several other sites for biochemical modifications, it is suggested that the protein is rich in active sites and hence can be explored for several drug binding sites.

Conclusion

Nucleocapsid protein of SARS-CoV-2 is rich in basic amino acids which justifies its role in binding to ribonucleic acid and forming the RNA-protein complex. Presence of highly intrinsic disordered regions signifies the protein's ability to fold differentially and bind to various ligands. This piece of work may help in delineating the three-dimensional tertiary structure of the protein that can be more helpful in finding its active sites, mapping for epitopes and also hunting for the ligand-docking regions.

References

1. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. Journal of virology, 2020, 94(7).
2. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P *et al*. Genome composition and divergence of the novel Coronavirus (2019-nCoV) originating in China. Cell Host Microbe, 2020.
3. Dhama K, Sharun K, Tiwari R, Sircar S, Bhat S, Malik YS *et al*. Coronavirus disease 2019–COVID-19, 2020.
4. Tilocca B, Soggiu A, Sanguinetti M, Musella V, Britti D, Bonizzi L *et al*. Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. Microbes and infection, 2020.
5. Surjit M, Lal SK. The SARS-CoV nucleocapsid protein: a protein with multifarious activities. Infection, genetics and evolution. 2008; 8(4):397-405.
6. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S *et al*. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020; 581(7807):215-220.
7. Buchan DW, Jones DT. The PSIPRED protein analysis workbench: 20 years on. Nucleic acids research. 2019; 47(1):402-407.
8. Wodak SJ, Vagin AA, Richelle J, Das U, Pontius J, Berman HM. Deviations from standard atomic volumes as a quality measure for protein crystal structures, 2012.
9. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nature protocols. 2015; 10(6):845.
10. Kang S, Yang M, Hong Z, Zhang L, Huang Z, Chen X *et al*. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. Acta Pharmaceutica Sinica B, 2020.
11. Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH. The SARS coronavirus nucleocapsid protein–forms and functions. Antiviral research. 2014; 103:39-50.
12. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nature reviews Molecular cell biology. 2015; 16(1):18-29.
13. Oldfield CJ, Dunker AK. Intrinsically disordered proteins and intrinsically disordered protein regions. Annual review of biochemistry. 2014; 83:553-584.
14. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nature medicine. 2020; 26(4):450-452.
15. Chen CY, Chang CK, Chang YW, Sue SC, Bai HI, Riang L *et al*. Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. Journal of molecular biology. 2007; 368(4):1075-1086