Advances in Bioresearch
Adv. Biores., Vol 12 (1) January 2021: 216-220

Print ISSN 0976-4585; Online ISSN 2277-1573
Journal's URL:http://www.soeagra.com/abr.html
CODEN: ABRDC3
DOI: 10.15515/abr.0976-4585.12.1.216220

Advances in Bioresearch

ORIGINAL ARTICLE

In-silico analysis on Non-Structural Proteins of Severe Acute Respiratory Syndrome Coronavirus 2(SARS-CoV2)

Debolina Chatterjee[1], Anooj E.S[2], B.V.Vibala[*]
[1]Department of Industrial Biotechnology, Manipal Institute of Technology, MAHE
[2,*]Xcellogen Biotech India Pvt Ltd, Thiruvananthapuram, Kerla-695035

ABSTRACT
COVID-19 has taken a huge toll on everyday life of every human being around the globe and the world scientists are converging towards a single goal of drug and vaccine discovery against the SARS-Cov2 virus which is a single stranded RNA virus with 4 structural proteins and 16 non-structural proteins. In this study we have carried out an In silico analysis on the several the several NSPs and found that each of the NSPs plays a very significant role in the viral replication inside the host as they belong to the "proteinase/transcriptase" category. Amongst them NSP 13 which exhibits helicase activity along with NTPase and ATPase properties and NSP 12 which has RNA-dependent RNA polymerase activity could be an important drug target without causing much side effect to the host.
Keywords: *COVID-19, SARS-COV-2, NSP,NSP 12, NSP 13, drug development*

Received 29.11.2020 Revised 21.12.2020 Accepted 04.01.2021

How to cite this article:
D Chatterjee, Anooj E.S, B.V.Vibala. *In-silico* analysis on Non-Structural Proteins of Severe Acute Respiratory Syndrome Coronavirus 2(SARS-CoV2). Adv. Biores., Vol 12 (1) January 2021: 216-220

INTRODUCTION

COVID 19 has affected the world like wildfire since 2019 and it's already been a year where every scientistis looking for noble treatments including vaccine discovery and drug developments [1]. The causative agent is already known to many and it is SARS CoV2 which shows close Phylogenetic relationship with Bat corona virus RaTG13 with more than 91.3% identity in their sequences and comparatively it is less similar to SARS-CoV and SARSr-CoV with less than 80% sequence identity[2].Coronaviruses belong to the order of Nidovirales which possess a single stranded RNA genome of length 26-31 Kb [3]. The +ssRNA contains 5'-cap and 3'-poly-A tail and contains six open reading frames (ORFS). Out of the two ORFs, $ORF1_{a/b}$ codes for 16 non structural proteins 1-16. (nsp 1-16) whereas the four main structural proteins including spike protein (S), membrane (M), envelope (E) and the nucleocapsid (N) is coded by the ORFs near 3'-end of the genome. [4] In this study we primarily focus on the non-structural proteins (NSPS) or 500 KDa polyproteins translated from the 800 KDa ORF 1ab which has replicase or transcriptase activity. [5]. As the cell binds to the human ACE2 (Angiotensin Converting Enzyme 2) and enters the host cell by cathepsin dependent mechanism, ORF1 of viral genome gets translated and processed which is essential for the viral replication.[6] Amongst the NSPs, NSP 5 also known as the 3CL-proteinase (3CL pro) that processes the C-terminus of nsp-4 through nsp 16, whereas nsp 1-3 cleaves to form one or two papain like proteases within NSP 3.[7] Activities of few NSPs have been studied, where NSP 12 exhibits RNA-dependent RNA polymerase activity, NSP-13 shows helicase, NSP 14 has 3'-5' exoribonuclease and NSP 16 manifests 2'- o-methyltransferase activity.[8] The following research work will provide us a better understanding about the sequence and structural aspects of the NSPs under consideration[9,10].

Combining our high-resolution crystal structure with existing data on the C-terminus of Nsp1 from SARS-CoV-2, we propose a model of the full-length protein. Our results provide insight into the molecular structure of a major pathogenic determinant of SARS-CoV-2. but also some unique structural features that likely contribute to increased stability of the β-barrel fold in SARS-CoV-2 Nsp1. Comparative analysis reveals additional structural homologs in Nsp1 proteins from *Alphacoronaviruses*, despite low levels of

shared sequence identity. These results highlight the critical role this unique protein fold plays in facilitating viral infection and suppression of host gene expression.

MATERIAL AND METHODS

Protein property identification

Through research studies, RefSeq accession ids of the NSPs were collected following which the FASTA sequence was retrieved from the NCBI. Chou and Fasman Secondary Structure Prediction tool was used for the identification of the percentage of helix, turns and sheets, GLOBPLOT was used for predicting the globular domains and disordered domains according to Russell/Linding definition and TMpred predicted models for transmembrane helices. The results regarding all the NSPs are reported in a tabular form.

Protein Modifications identification methodology

For the prediction of the protein modification, ChloroP was used to identify the chloroplast transit sites, PredGPI was utilized to predict the GPI-anchor signals, identify phosphorylation sites by NETPhos, cleavage sites by NetCorona and Sulfation sites by Sulfinator tool.

Protein structure modeling

Most of the structures of NSPs have not been deposited in the PDB and hence we carried out homology modelling by SWISSMODELLER and reported the template, Molprobity score, clash score, Ramachandran plot favored percentage, Ramachandran plot outliers percentage and bad angles percentage in a table format for the better understanding.

Phylogenetic relation between the NSPS:

Clustal Wserver was used to find out the Phylogenetic relationship between the NSPS. Clustal omega has been widely used to estimate the multiple sequence alignment utilizing seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences.

RESULTS

NSPs Secondary Structure through sequence analysis

Table 1. Nonstructural proteins of coronaviruses and their secondary structure sequence property analysis

NAME OF THE PROTEIN	TOTAL CHAIN LENGTH	HYDROGEN RESIDUES PERCENT	SEETS PERCENT	TURNS PERCENT	POTENTIAL GLOBULAR DOMAINS
NSP 4	500	63.8	65.6	8.8	1-500
NSP 2	500	74.5	50	12.1	1-250; 265-635
NSP 1	638	73.9	61.1	12.2	1-125
NSP 16	298	63.4	54.4	10.7	2-295
NSP 15	346	86.1	81.8	10.4	1-346
NSP 14	527	59.6	76.5	8.3	1-120;149-406; 422-527
NSP 13	601	58.4	54.7	10	85-276
NSP 12	932	67.6	79.4	10.5	1-210;235-670; 687-825; 840-932
NSP 11	13	84.6	46.2	7.7	NONE
NSP 10	139	40.3	64	10.8	3-7;
NSP 9	113	61.9	70.8	8.8	1-113
NSP 8	198	81.8	42.2	9.6	1-198
NSP 7	83	91.6	49.4	8.4	1-83;
NSP 6	290	77.9	80.7	7.2	1-290
NSP 5	306	57.5	60.8	10.5	2-114; 184-306

Structural analysis of NSPs

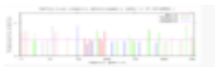
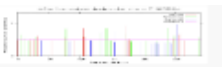
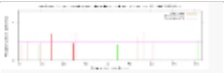
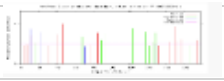
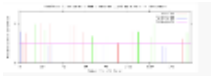
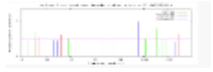
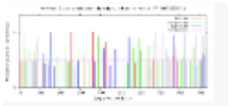
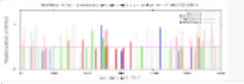
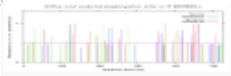
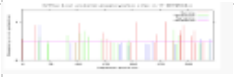
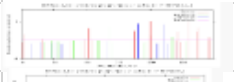
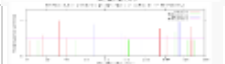
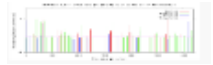
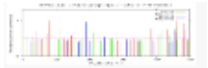
Table 1. Nonstructural proteins of coronaviruses and their structural features

PROTEIN NAMES	GMQE	QMEAN	MOLPROBITY SCORE	PERCENTAGE FAVOURED BY RAMACHANDRAN PLOT	CLASH SCORE	RAMACHANDRAN DRANOUTLIERS	BAD ANGLES
NSP 1	0.05	-8.65	2.27	75.00%	8.99	8.93%	23/1204
NSP 2	0.98	-3.93	2.35	89.39%	3.87	0.67%	56/6478
NSP 7	0.98	-0.58	1.02	100%	2.39	0	4/828
NSP 9	0.99	-0.51	1.43	94.04	0.58	0	15/2359
NSP 8	0.67	-0.31	1.02	97.39%	0.55	0	4/1254
NSP 10	0.98	-2.54	1.8	96.90%	1.61	0	17/1329
NSP 11	0.99	-2.38	2.35	94.05%	0.94	0.54%	45/10430

NSP 12	0.98	-3.93	1.44	89.39%	3.87	0.67%	56/6478
NSP 13	0.98	-3.18	0.63	89.44%	0.97	1.34%	48/5841
NSP 14	0.99	0.58	1.42	98.40%	0	0	82/22746
NSP 15	0.99	-1.14	1.83	96.28%	0	1.86%	18/3286
NSP 16	0.13	0.36	0.91	100%	0	0	3/368
NSP 6	0.05	-2.73	1.83	92.11%	6.31	5.26%	9/421
NSP 5	0.99	0.45	1.05	97.67%	0.21	1.17%	35/6493
NSP 4	0.1	-1.65	1.99	95.40%	3.91	2.30%	14/1968

Analysis of NSPs Protein modification sites

Table 3.Nonstructural proteins of coronaviruses and their structural modification sites

NAME OF THE PROTEIN	GPI ANCHOR PREDICTION	OMEGA SITE POSITION	PHOSPHORYLATION SITES	SULFATION SITES	CHLORO P SCORE	NETCORONA CLEAVAGE SITES
NSP 5	NOT GPI ANCHORED	277		NA	0.447	NONE
NSP6	NOT GPI ANCHORED	265		NA	0.466	NONE
NSP 7	NOT GPI ANCHORED	63		NA	0.454	NONE
NSP 8	NOT GPI ANCHORED	173		NA	0.442	NONE
NSP 9	NOT GPI ANCHORED	91		NA	0.445	NONE
NSP 10	NOT GPI ANCHORED	119		NA	0.467	NONE
NSP 11	NOT GPI ANCHORED			NA	0.432	NONE
NSP 12	NOT GPI ANCHORED	909		453(e value: 53;455 (E VALUE: 54); 456 (E VALUE: 54); 458 (E VALUE 54); 521(E VALUE: 48); 826 (E VALUE:30); 828 (E VALUE: 54); 831 (54)	0.468	NONE
NSP 13	NOT GPI ANCHORED	577		205 (E VALUE: 36); 541 (E VALUE= 53); 543 (E VALUE=43)	0.451	NONE
NSP 14	NOT GPI ANCHORED	502		NA	0.443	NONE
NSP 15	NOT GPI ANCHORED	324		NA	0.435	NONE
NSP 16	NOT GPI ANCHORED	277		NA	0.449	NONE
NSP 1	NOT GPI ANCHORED	160		154 (E VALUE= 46)	0.437	NONE
NSP 2	NOT GPI ANCHORED	615		NA	0.429	NONE
NSP 4	NOT GPI ANCHORED	480		220(E VALUE= 51)	0.437	NONE

Evolutionary analysis of NSPs

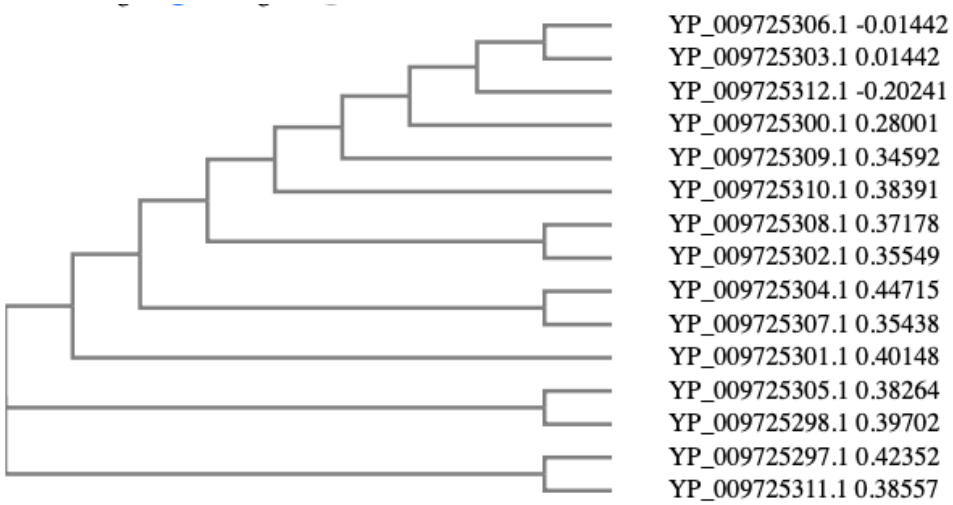


Figure 1: Graphical representation of Phylogenetic tree constructed from NSPs

DISCUSSION

The target protein NSP 13 (RNA Helicase) of SARS-CoV2 was considered in our research as a target after surveying several research papers which revealed its potentiality of being a major drug target causing much side effect to the host. Firstly, we have downloaded the FASTA sequence of the protein (PDB ID: 6ZSL) and using NCBI BLAST we identified it shares a 100% identity with ORF1Ab polyprotein (Accession Id: QLL35519.1). Using SWISSMODEL, we have successfully developed a model with template 6XEZ.1.F with which it shared an identity of 100% and obtained a QMEAN value of -3.93, GMQE of 0.98, Ramachandran plot favored percentage to be 89.39%, clash score of 3.87 and bad angles of 0.86% (56/6478). With the application of Net Corona 1.0 Server, we predicted that none of the amino acids were scored beyond the threshold value of 0.5 and hence we can say there were no cleavage sites around the amino acids being scored by the tool in the target protein. Dynomics is a software used the Protein (6ZSL) X-ray crystallographic structure properties are analysed. The protein molecular motion, Mean square fluctuation residues, Theoretical and experimental factor, cross correlation between residue fluctuation, Inter residue contact map, Potential functional site, Perturbation Response map all properties are analyzed.The Graphical representation of Phylogenetic tree constructed from 22 sequences (NSP 13 (RNA Helicase) of SARS-CoV2).With the hypothetical root of straight line denoted 100% sequence similarity and branch length denoted non-sequence similarity. Dendogram is generated using Neighbor-joining method. With the help of cladogram as we obtained from CLUSTAL OMEGA server, we found out that 6ZSL is closely linked with the proteins of accession id: QHR63289.1 and QHR63299 and distantly related to NP_828849.7.

CONCLUSION

Due to the ongoing pandemic situation, every scientistisconverged to contributing for the same cause. In this work we have used in-silico methods to understand the RNA-helicase of SARS-CoV2 where we have developed homology model, Ramachandran plot, used NetCorona to analyze the cleavage sites in the protein followed by analysis of the dynamics and Phylogenetic relationships. This in-silico work could be further used by the scientists to carry out in-vitro and in-vivo research in the following days for successfully developing a drug against the novel corona virus, causative agent of COVID 19.

ABBREVIATIONS

NSP: Non-Structural Proteins, SARS-CoV2: Severe Acute Respiratory Syndrome-Coronavirus 2.

ACKNOWLEDGEMENT

We would like to thank the xcellogen biotech India Pvt Ltd team for supporting towards the completion of the full work.

REFERENCES

1. Kyoung-Jin Jang, SeonghwanJeong, Dong Young Kang, NipinSp, Young Mok Yang& Dong-Eun Kim,(2020). A high Atp concentration enhances the cooperative translocation of the SARS coronavirus helicase nsP13 in the unwinding of duplex RnA,, SCIENTIFIC REPORTS, 10:4481
2. Betz, U. A. K., Fischer, R., Kleymann, G., Hendrix, M. & Rubsamen-Waigmann, H. (2002). Potent in vivo antiviral activity of the herpes simplex virus primase-helicase inhibitor BAY 57-1293.Antimicrob. Agents Chemother. 46, 1766–1772.
3. Kwong, A. D., Rao, B. G. &Jeang, K. T. (2005). Viral and cellular RNA helicases as antiviral targets. Nat. Rev. Drug. Discovery 4, 845–853.
4. Subissi L, Imbert I, Ferron F, Collet A, Coutard B, Decroly E, Canard B. (2014). SARS-CoV ORF1b-encoded nonstructural proteins 12-16: replicative enzymes as antiviral targets. Antiviral Res.; 101(1):122-30.
5. Ivanov KA, Thiel V, Dobbe JC, van der Meer Y, Snijder EJ, Ziebuhr J. (2004). Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. J Virol. 78:5619–5632. doi: 10.1128/JVI.78.11.5619-5632.2004.
6. Consortium., T. C. S. M. E. (2002). Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Sci.* 303,1666–1669, https://doi.org/10.1126/science.109-220.
7. Marra, M. A. *et al*. (2003). The genome sequence of the SARS-associated coronavirus. *Sci.* 300, 1399–1404, https://doi.org/10.1126/science.1085953.
8. Rota, P. A. *et al*. (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Sci.* 300, 1394–1399,https://doi.org/10.1126/science.108952.
9. Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. (2000). Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virology* 81, 853–879, https://doi.org/10.1099/0022-1317-81-4-853.
10. Brierley, I., Digard, P. & Inglis, S. C. (1989). Characterization of an efficient coronavirus ribosomal frame shifting signal: requirement for an RNA pseudoknot. *Cell* 57, 537–547, https://doi.org /10.1016/0092-8674 (89)90124-4.