



www.bioinformatics.net
Volume 18(4)

Research Article

Received March 1, 2022; Revised April 30, 2022; Accepted April 30, 2022, Published April 30, 2022

DOI: 10.6026/97320630018432

Declaration on Publication Ethics:

The author's state that they adhere with COPE guidelines on publishing ethics as described elsewhere at <https://publicationethics.org/>. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Declaration on official E-mail:

The corresponding author declares that lifetime official e-mail from their institution is not available for all authors

License statement:

This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Comments from readers:

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

Edited by P Kanguane

Citation: Yashvardhini *et al.* Bioinformatics 18(4): 432-437 (2022)

Genome sequence analysis of nsp15 from SARS-CoV-2

Niti Yashvardhini^{1*}, Deepak Kumar Jha², Amit Kumar³, Manjush Gaurav³ & Kumar Sayrav⁴

¹Department of Microbiology, Patna Women's College, Patna, 800 001, Bihar, India; ²Department of Zoology, S.M.P. Girls Degree College, Ballia, 277401, Uttar Pradesh, India; ³Department of Botany, Patna University, Patna-800 005, Bihar, India; ⁴Department of Chemistry, V.K.S. University, Ara-802301, Bihar India; *Corresponding author - E-mail: nitiyashvardhini@gmail.com

Abstract:

SARS-CoV-2 (Severe Acute Respiratory Syndrome), a causative agent of COVID-19 disease created a pandemic situation worldwide. Nsp15 is a uridine specific endoribonuclease encoded by the genome of SARS-CoV-2. It plays important role in processing viral RNA and, thus evades the host immune system. Therefore, it is of interest to identify mutants of nsp15 amongst Asian SARS-CoV-2 isolates, where a total of 1795 mutations, from 7793 sequences of Asia submitted till 31st January 2022, amongst which A231V, H234Y, K109N, K259R and S261A mutations were found frequent. Hence, we report data on the predicted secondary structure of wild type form followed by hydrophathy plot, physiochemical properties, Ramachandran plot, B-cell epitopes prediction and protein modeling of wild type and mutant of nsp15 protein. Data shows that nsp15 of SARS-CoV-2 is a potential candidate for the development of vaccine to control the infections of SARS-CoV-2.

Keywords: SARS-CoV-2, Pandemic, COVID-19, Nsp15, Mutation.

Background:

SARS-CoV-2, a novel betacoronavirus, belonging to the family Coronaviridae and order nidovirales is an enveloped, positive-sense, largest known RNA virus (genome size~30kb) and the causative agent of coronavirus disease 2019 (COVID-19) [1,2]. Coronavirus outbreak initiated from a small sea food market in Wuhan, China [3]. Although, several vaccines candidates are approved and are used on a mass scale, however the contagious nature of this virus are still a great threat to human beings whose ripple effect poses a huge human health crisis worldwide. As of 31st January 2022, confirmed cases of COVID-19 have been reported to be 450, 229 and 635, including 6,019,085 casualties, globally by WHO. As a consequence of multiple mutations this virus showed high rate of variability which enable them to evade host immunity [4]. Due to its ability to show high rate of mutations and poor fidelity of its RNA polymerase, these viruses shows greater antigenic variability [5,6] and these high rates are greatly linked to virulence and evolution of this virus, traits prerequisite for viral adaptation [7].

The genome of SARS-CoV-2 comprises of 14 open reading frame (ORF) sequences which encode 29 proteins including 4 types of structural proteins; S (spike), E (envelope), M (membrane), and N (nucleocapsid) that are essential for the assembly of a complete virion particle. Among 16 nonstructural proteins, nsp15 contains uridylylate specific endoribonuclease (EndoU) in the catalytic C-terminal domain [8-10], as it has only been detected in Nidovirales, the virus EndoU was given the name nidovirus EndoU (NendoU) [11-13]. The high-resolution crystal structure of Nsp15 from SARS-CoV-2 was determined. The N-terminal oligomerization domain, the middle domain, and the C-terminal catalytic domain constitute Nsp15. In addition, crystal structural analyses revealed a hexameric configuration made up of two trimers with a peculiar spatial configuration [8, 14-17]. NendoU hydrolyzes the phosphodiester bonds at uridine sites of single- and double-stranded RNA, with 2', 3'-cyclic phosphodiester and 5'-hydroxyl termini produced [18, 19]. The pattern recognition receptor MDA5 is used by the host innate immune system to recognize the polyU sequence in negative-sense viral RNA, which is copied from the polyA sequence in viral RNA [20-21]. To limit the build up of the polyU-containing sequence in cells and so dodge the innate immune response, nsp15 cleaves its own negative-sense RNA [20, 22-24]. nsp15, for example, decreases viral RNA in host cells and hinders RNA-activated antiviral responses by preventing the production of antiviral cytoplasmic stress granules [25]. Nsp15 also has an essential role in the replication of corona-viruses in processing of the viral genome [14, 26, 27]. Nsp15 deletion suppresses viral replication substantially [23, 25, 28]. Thus, inhibiting Nsp15 activity at the non-symptomatic stage may be a potential strategy for preventing and controlling viral infection by interrupting viral genome replication and activating the host innate immune response [29]. The SARS-CoV-2 Endo RNase (nsp15) contains 345 amino acids which are located in the region of ORF1ab polyprotein, ranging from 6453 to 6798 amino acids. Therefore, it is of interest to document the the potential effects of mutations on the activity of nsp15 protein from Asian SARS-CoV-2 and compare it with those of Wuhan isolates.

Considerable alterations in its structural conformations have been observed. Nsp15 plays pivotal role in the life cycle of SARS-CoV-2 hence; it can be used as a major antiviral therapeutic target.

Methods:

Sequence retrieval: nsp15 protein sequences of SARS-CoV-2 were retrieved from NCBI virus database, submitted from Asia from the onset of this disease till 31st January, 2022. A reference sequence of Wuhan type virus with accession number YP_009724389 was used for identification of nsp15 mutants.

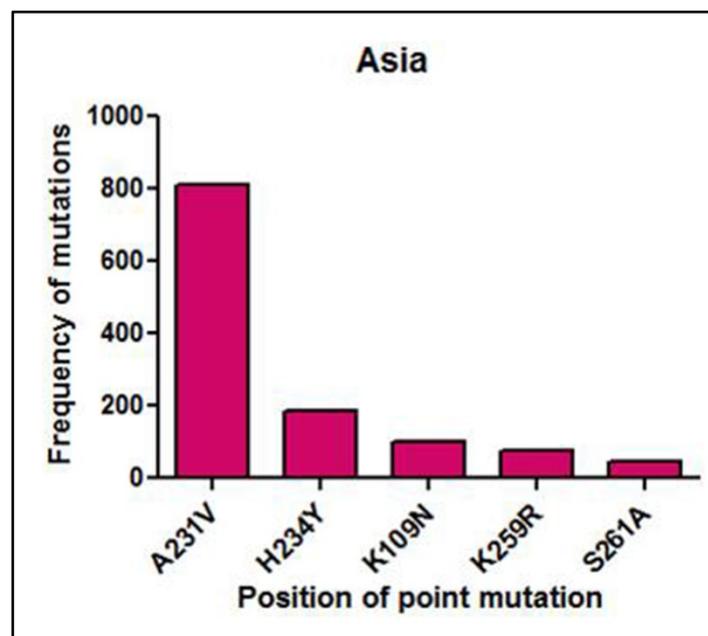


Figure 1: Showing the total number of mutations occurring in the nsp15 protein sequences of Asian isolates

Identification of nsp15 mutants:

The nsp15 protein sequences (345 amino acids residues) were downloaded and aligned using Clustal Omega online server and the alignment was viewed using Jalview. The differences occurring in the nsp15 protein sequence with respect to Wuhan type virus sequence was noted. The frequency of different mutations was calculated.

Calculation of physicochemical properties and hydropathy index of nsp15 protein:

The physicochemical properties including molecular weight, amino acid composition, overall pH, instability index, aliphatic index, extinction coefficient, half life and average of hydrophobicity (GRAVY) was estimated using ProtParam tool of ExPasy online program. Hydropathy plot of nsp15 protein was prepared using ProtScale tool of expasy [30].

Prediction of 3D structure of nsp15 protein:

The 3D structures of both the wild type and mutated nsp15 proteins were built using Chimera software [31]. Chimera is a tool for visualization of and analysis of molecular structures.

Prediction of secondary structure and Ramachandran plot:

To know the number of alpha helices, beta sheets and turns, the secondary structure prediction of the wild type nsp15 protein was done using CFSSP (Chou and Fasman secondary structure prediction) online software [32]. Ramachandran plot of nsp15 protein was predicted using Swiss model software.

Identification of linear B-cell epitopes:

The linear B-cell epitopes of the nsp15 protein were predicted using IEDB webserver [33] which identifies B-cell epitopes based on properties like flexibility, accessibility, hydrophilicity, turns, polarity and antigenic propensity of the protein using amino acid scales and HMMs.

Table 1: Physicochemical properties of nsp15 protein (wild type)

Physicochemical properties	nsp15	Amino acid composition	No.	Percent composition (%)
Molecular weight	38813.4	Ala (A)	16	4.6
No. of amino acids	346	Arg (R)	9	2.6
Theoretical pI	5.06	Asn (N)	17	4.9
Instability index	36.28	Asp (D)	23	6.6
No. of negatively charged (Asp+ Glu)	46	Cys (C)	5	4.4
No. of positively charged (Arg+ Lys)	34	Gln (Q)	15	4.3
aliphatic index	95.09	Glu (E)	23	6.6
Grand average of hydropathicity	-0.076	Gly (G)	22	6.4
Estimated half-life (mammalian reticulocytes, in vitro)	1.9 hours	His (H)	6	1.7
Atomic composition		Ile (I)	23	6.6
C	1759	Leu (L)	29	8.4
H	2743	Lys (K)	25	7.2
N	445	Met (M)	5	1.4
O	523	Phe (F)	21	6.1
S	10	Pro (P)	14	4
Formula	C ₁₇₅₉ H ₂₇₄₃ N ₄₄₅ O ₅₂₃ S ₁₀	Ser (S)	20	5.8
Total number of atoms	5480	Thr (T)	21	6.1
		Trp (Y)	3	0.9
		Val (V)	38	11
		Phy (O)	0	0
		Sec (U)	0	0

Table 2: B-cell epitope prediction of nsp15 protein sequence of SARS-CoV-2 and its allergenicity

No.	Start	End	Peptide	Length	Allergenicity
1	12	20	KGHFDGQQG	9	Non-allergen
2	44	48	ENKTT	5	Allergen
3	60	72	KRNIKPVPEVKIL	13	Allergen
4	89	97	KRDAPAHIS	9	Allergen
5	106	115	DIAKKPTETI	10	Non-allergen
6	126	136	RVDGQVDLFRN	11	Allergen
7	147	161	SVKGLQPSVGPQKAS	15	Allergen
8	183	187	DGVVQ	5	Allergen
9	196	230	QSRNLQEFKPRSQMEIDFLELAMDEFIERYKLEGY	35	Non-allergen
10	257	271	RFKESPFLEDFIPM	15	Allergen
11	304	312	EIISQDLS	9	Allergen

Results and Discussion:

A total of 7793 ORF1ab full length protein sequences were submitted from Asia till 31st January, 2022 from onset of this pandemic. These sequences were downloaded along with a reference sequence of Wuhan type virus from NCBI virus database.

The alignments were observed using Jalview to mark the differences occurring in the nsp15 region. A total of 1795 point mutations were detected in the nsp15 protein region of ORF1ab polyprotein amongst the sequences. Amongst these point mutations, A231V, H234Y, K109N, K259R and S261A were the most

frequently occurring mutations and hence used for further characterization in this study (Figure 1). The estimation of physicochemical properties of nsp15 protein revealed that nsp15 protein is 345 amino acids long with a molecular weight 38813.40, aliphatic index 95.09, instability index 36.28, and GRAVY score of -0.076 (Table 1). The hydropathy plot showed C-terminal amino acid to be more hydrophobic as compared to the N-terminal end of nsp15 protein (Figure 2). The nsp15 protein sequence was modeled using Chimera software. The models of both wild type and mutated protein sequences were built as shown in Figure 3. Ramachandran plot analysis showed that most of the amino acid sequences were in the favoured region of the plot (Figure 4). Secondary structure analysis was done using CFSSP online program to detect the number of alpha helix, beta sheet and turns in nsp15 protein which is shown in Figure 5.

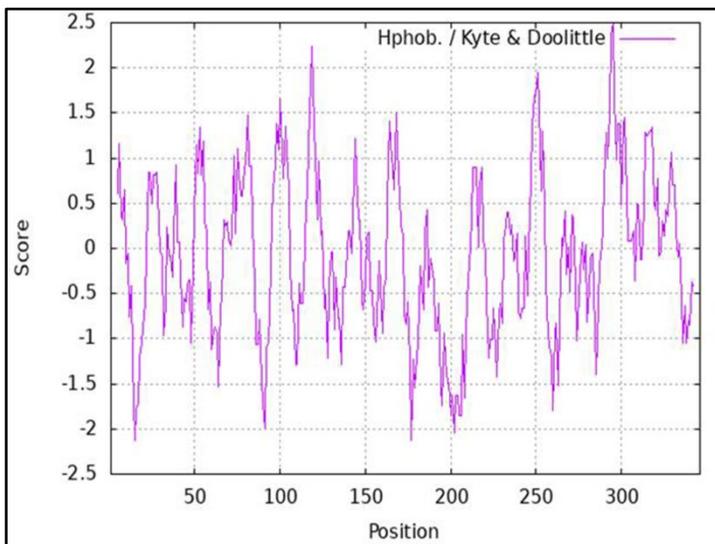


Figure 2: Hydropathy plot of wild type nsp15 protein showing hydrophobic amino acid residues.

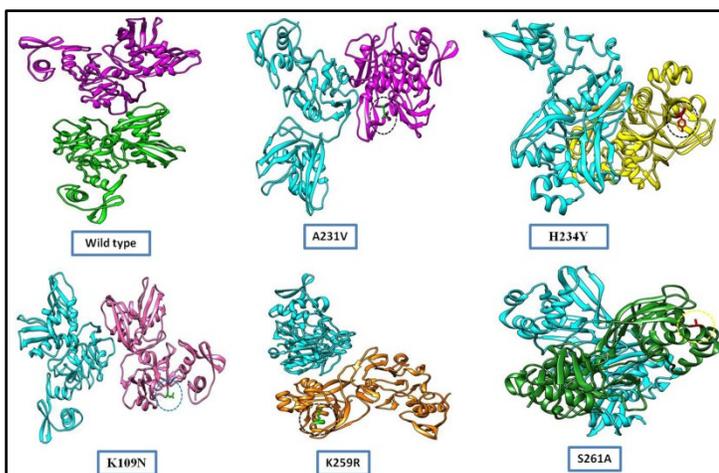


Figure 3: Protein modeling of wild type and mutant nsp15 protein. Models were prepared by Chimera software.

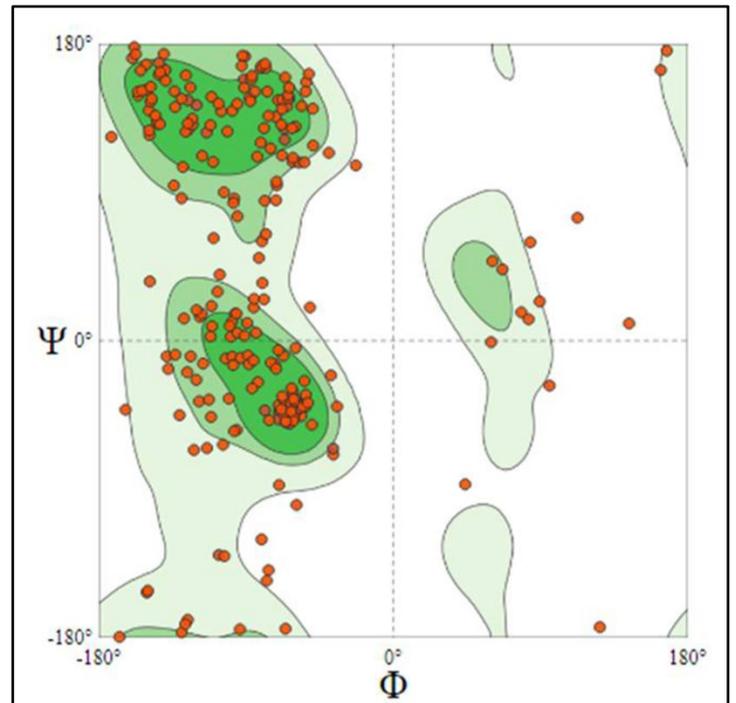


Figure 4: Ramachandran plot analysis of template protein, showing 89% of amino acid residues in favored region, rest in outlier region.

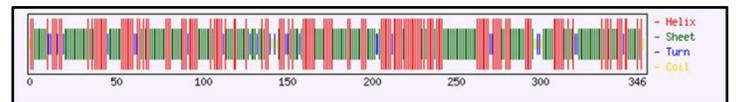


Figure 5: Secondary structure prediction of nsp15 protein.

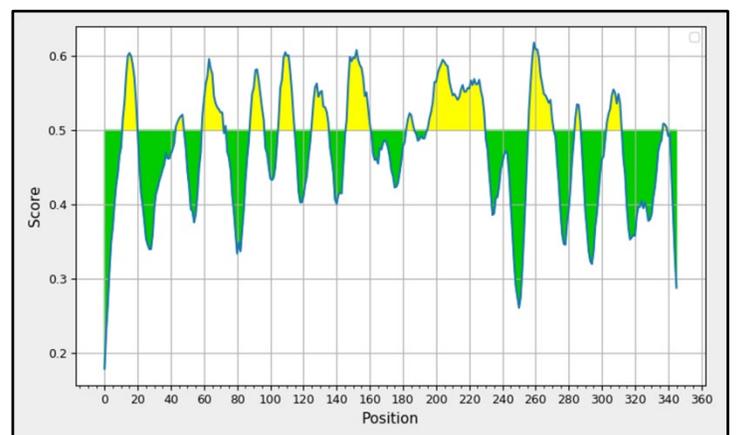


Figure 6: B-cell epitope prediction of nsp15 protein. The threshold cutoff is 0.5 above which the residues are epitopes.

A total of nine linear B-cell epitopes were predicted for 345 amino acid long nsp15 protein as shown in Figure 6 (Table 2). These epitopes can induce antibody production and hence play a crucial role in humoral immunity. Coronavirus poses humanitarian health crisis globally. Considering its infectivity, World Health Organization on 11th March 2020 has declared public health

emergency internationally (WHO 2020). SARS-CoV-2 belongs to RNA viruses family and has remarkable capacity to mutate their genome in a very short span of time [34-35]. Notably, majority of viral mutation shows harmful effects, moreover, a mutation is essential for viral evolution and adaptability, these traits are found as the key determinants for these viruses to better survive in the dynamic host environment. The findings of the present study, emphasizes mainly on the occurrence of recurrent mutations in the nsp15 protein of Asian SARS-CoV-2 isolates; by comparing it with that of Wuhan SARS-CoV-2 isolate. Nsp15 is indispensable for viral RNA cleavage and also involved in the process of pathogenicity. The results of multiple sequence alignment revealed the presence of 1795 mutations in the SARS-CoV-2 nsp15 protein, out of which A231V, H234Y, K109N, K259R and S261A occurred most frequently as shown in **Figure 1**. The Ramachandran plot analysis showed the presence of major amino acids in the favored region of the quadrangle. Prediction of secondary structure showed the presence of helices and sheets in the wild type protein. Protein modeling of wild type as well as mutated protein showed significant changes in the wild type nsp15 protein upon mutation. B-cell epitope prediction confirmed the efficacy of nsp15 protein to induce humoral immune response. Therefore, these mutations might induce conformational changes and considered as major challenges in the designing of vaccine to curb SARS-CoV-2 infections. Data suggests that the nsp15 of SARS-CoV-2 is an important protein for coronavirus replication and can be used as potential antiviral solution. Earlier studies have shown that the viral nsp15 appears as challenging druggable targets: Interestingly, nsp15 protein induces humoral immunity. Therefore, SARS-CoV-2 nsp15 is a suitable target for anti-viral drug discovery, as there is paucity of available information regarding designing of SARS-CoV-2 nsp15 inhibitors [36, 37].

Conclusion:

Data provides insights into impact of mutation in developing antiviral drugs to curb this pandemic because these mutations might have functional consequences which need to be incorporated in ongoing future research work.

Acknowledgement:

The corresponding author would like to acknowledge Patna Women's College (Autonomous), Patna University, Patna for providing research grant in the form of seed money to complete this research work.

Funding: Patna Women's College (Autonomous), Patna University

Reference:

- [1] Lu R *et al.* *Lancet*. 2020 395:10224 [PMID:32007145]
- [2] Yashvardhini N *et al.* *Arch Microbiol*. 2021 203:9 [PMID: 34410443]
- [3] Zhu N *et al.* *N Engl J Med*. 2020 382:8 [PMID: 31978945]
- [4] Sackman AM *et al.* *Mol Biol Evol*. 2017 34:12 [PMID: 29029274]
- [5] Domingo E & Holland JJ. *Annu Rev Microbiol*. 1997:51 [PMID: 9343347]
- [6] Jha DK *et al.* *Indian J Pathol Microbiol*. 2021 64:4 [PMID: 34673601]
- [7] Duffy S *PLoS Biol*. 2018 16(8):e3000003 [PMID: 30102691]
- [8] Kim Y *et al.* *Protein Sci*. 2020 29:7 [PMID: 32304108]
- [9] Sharma A *et al.* *J Biomol Struct Dyn*. 2022 40:1 [PMID: 32896226]
- [10] Kim Y *et al.* *Commun Biol*. 2021 4:1 [PMID: 33564093]
- [11] Ivanov KA *et al.* *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101:34 [PMID: 15304651]
- [12] Motwalli O & Alazmi M. *J Mol Model*. 2021 27:6 [PMID: 33963942]
- [13] Zheng A *et al.* *J Biol Chem*. 2018 293:31 [PMID: 29887523]
- [14] Pillon MC *et al.* *Nat. Commun*. 2021 12:1 [PMID: 33504779]
- [15] Zhang L *et al.* *J Virol*. 2018 92:22 [PMID: 30135128]
- [16] Cao J & Zhang X. *Virus Res*. 2012 167:2 [PMID: 22617024]
- [17] Cao J *et al.* *Inter virology*. 2008 51:5 [PMID: 19023218]
- [18] Ulferts R & Ziebuhr J. *RNA Biol*. 2011 8:2 [PMID: 21422822]
- [19] Frazier MN *et al.* *Nucleic Acids Res*. 2021 49:17 [PMID: 34403466]
- [20] Hackbart M *et al.* *Proc. Natl. Acad. Sci. U.S.A.* 2020 117: 14 [PMID: 32198201]
- [21] Yuen CK *et al.* *Emerg Microbes Infect*. 2020 9:1 [PMID: 32529952]
- [22] Deng X *et al.* *Proc. Natl. Acad. Sci. U.S.A.* 2017 114:21 [PMID: 28484023]
- [23] Kindler E *et al.* *PLoS Pathog*. 2017, 13(2):e1006195 [PMID: 28158275]
- [24] Bhardwaj K *et al.* *J Biol Chem*. 2008 283:6 [PMID: 18045871]
- [25] Gao B *et al.* *PLoS Pathog*. 2021 17(2):e1008690 [PMID: 33635931]
- [26] Batool A *et al.* *Eur J Pharmacol*. 2021 892:173779 [PMID: 33275961]
- [27] Krishnan DA *et al.* *Inform Med Unlocked*. 2020 20:100392 [PMID: 32835078]
- [28] Kang H *et al.* *J Virol*. 2007 81:24 [PMID: 17898055]
- [29] Liu X *et al.* *Mol Immunol*. 2019 114 [PMID: 31351410]
- [30] Larsen JE *et al.* *Immunome Res*. 2006 2:2 [PMID: 16635264]
- [31] Goddard TD *et al.* *Structure*. 2005 13:3 [PMID: 15766548]
- [32] Ashok KT *Wide. Spec*. 2013 1:9
- [33] Gasteiger E *et al.* *Prot. Proto. Hand*. 2005 112 [PMID: 10027275]
- [34] Benvenuto DJ *et al.* *Med Virol*. 2020 92:4 [PMID: 31994738]
- [35] Islam MR *et al.* *Sci. Rep*. 2020 10:1 [PMID: 32814791]
- [36] Khan RJ *et al.* *J Biomol Struct Dyn*. 2022 40:1 [PMID: 32885740]
- [37] Sixto-López Y & Martínez-Archundia M. *J Comput Chem*. 2021 42:13 [PMID: 33713492]

