Mapping COVID-19 Mutations:

Evolution of SARS-CoV-2 non structural protein 14 during the first 6 months of the global pandemic

Aaliyah Khan[1], Jitendra Singh[2], Kevin Catalfano[3], Elliot Dolan[4], Changpeng Lu[4], Vidur Sarma[4], Zhoufan Shen[4], Maria Szegedy[4], Lingjun Xie[4], Sagar Khare[4], Christina Zardecki[4 5], and Stephen K. Burley[4 5]

[1]University of Maryland Baltimore County, [2]New York City College of Technology, [3]Hope College, [4]Institute for Quantitative Biomedicine, Rutgers University, [5]RCSB Protein Data Bank

OVERVIEW

The World Health Organization reported that the number of coronavirus disease 2019 (COVID-19) cases worldwide exceeded 10 million with over 500,000 deaths as of June 30, 2020. Severe acute respiratory syndrome 2 (SARS CoV-2) is the virus that causes COVID-19. Genome sequencing of the SARS-CoV-2 showed the virus to be a single-stranded, positive sense, 5'-capped RNA genome that encodes ~30 proteins. Additional genomic sequencing of SARS-CoV-2 isolates documented evolution of the virus as amino acid changes accumulated in each one of the viral proteins. We studied mutations of key viral proteins to provide insight into the virus life cycle; allowing the discovery and development of new antiviral drugs and vaccines.

METHODOLOGY

We studied how SARS-CoV-2 proteins evolved during the first six months of the COVID-19 with bioinformatics tools:

- Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo/) for sequence alignments and phylogenetic trees
- Mol* (molstar.org) for 3D molecular visualization; and
- Foldit (fold.it) and PyRosetta (pyrosetta.org) for structural/energetic effects of sequence mutations.

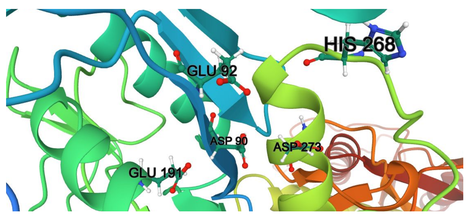


Fig. 1: Active site residues of exoribonuclease proofreading domain of Nsp14 shown in ball and stick representation with Mol*

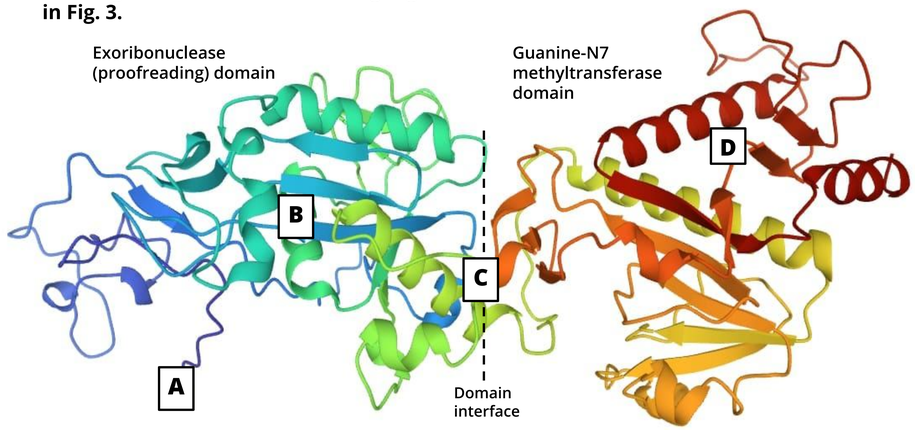
ACKNOWLEDGEMENTS

RISE at Rutgers, New Jersey Space Consortium

RCSB PDB is funded by the NSF (DBI-1832184), the US DOE (DE-SC0019749), and the National Cancer Institute, National Institute of Allergy and Infectious Diseases, and National Institute of General Medical Sciences of the National Institutes of Health under grant R01GM133198.

FOCUS: NON-STRUCTURAL PROTEIN 14 (NSP14)

Figure 2: 3D structure of Nsp14 with blue N terminal to red C terminal rainbow coloring in Mol*. A, B, C, and D highlight the locations of mutations illustrated in Fig. 3.

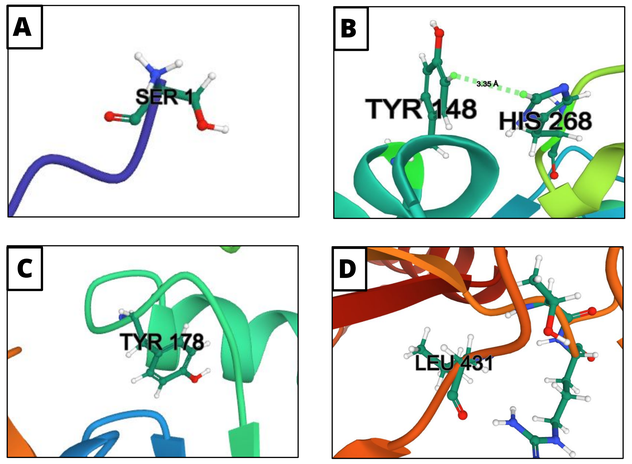


Non-structural protein 14 (Nsp14, Fig 2.) plays a dual role in the viral replication process. The methyltransferase domain caps the viral RNA at the 5' end with a 5' 7-methyl-guanine cap to prevent RNA degradation. The exoribonuclease domain performs proofreading on replicated viral RNA. 307 unique variants of Nsp14 were modeled using structural bioinformatics tools and analyzed using PyRosetta. Figures 3 shows selected mutations.

CONCLUSION

- Focus mutations for Nsp14 were chosen on the bases of proximity to active sites and domain interfaces, energetic impact, and number of times observed in independent genome sequences.
- The active site of proofreading exoribonuclease domain showed no evidence of mutations to catalytic residues..
- High level of conservation of the Nsp14 active site makes it a potential target for discovery and development of antiviral agents.

Figure 3 . Selected mutations of Nsp14 shown in ball and stick representation in Mol*. See Fig. 2 for their locations within the protein 3D structure.



(A) Mutation A1S alanine 1 to serine was located at the N-terminal of Nsp14. It had a negative energy change of -2.3 REUs, which suggests it is a stabilizing mutation.

(B) Mutation S178Y serine (polar) 178 to tyrosine (non-polar) had a high energy change of +1493 REUs which reflects its location in the domain interface core layer

(C) Mutation H148Y histidine 148 to tyrosine was close to proofreading active site (Fig 1) . The green dotted line shows the distance of 3.35 Å from the active site catalytic residue histidine 268.

(D) Mutation F431L phenylalanine 431 to leucine was located in the methyltransferase domain (Fig 2). It was observed in 19 independently isolated viral genomes.