Understanding SARS-CoV-2 Protein Evolution in 3D During the COVID-19 Pandemic: ORF7a

Luz Helena Alfaro Alvarado[4], Thejasvi Venkatachalam[2], Helen Zheng[5], Elliott Dolan[3], Changpeng Lu[3], Vidur Sarma[3], Zhuofan Shen[3], Maria Szegedy[3], Lingjun Xie[3], Christine Zardecki[1], Sagar Khare[3], Stephen K. Burley[1]

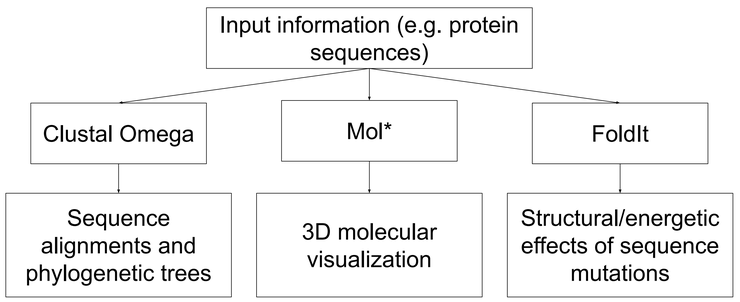
[1]RCSB PDB. [2]Rutgers University. [3]Rutgers IQB. [4]Grinnell College. [5]Watchung Hills Regional.

OVERVIEW

The year 2019 took an unprecedented turn as the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), an RNA virus that causes the Coronavirus Disease 2019 (COVID-19), took a toll on lives, economies, bodies, and minds. By examining specific proteins and their respective mutations in SARS-CoV-2, scientists can achieve a better understanding of its structural biology. New insights gained can aid scientists in drug and vaccine discovery and in obtaining a more nuanced understanding of coronaviruses in general.

METHODOLOGY

- Bioinformatics Tools
  - Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo/)
  - Mol* molecular visualization (molstar.org)
  - Foldit (fold.it)



OPEN READING FRAME 7a (ORF7a)

SARS-CoV ORF7a is a viral transmembrane protein that interferes with N-linked glycosylation of the cellular protein BST-2 (tetherin), an inhibitor of coronavirus release (Taylor et al. 2015). In SARS-CoV ORF7a transfected cells, cells have been found to be arrested in the G0/G1 checkpoint, preventing cell cycle progression. The cyclin D3/pRb pathway is involved in arresting the cell since SARS-CoV ORF7a transfected cells have reduced levels of cyclin D3 (perhaps due to the inhibition of translation of mRNA) as well as lower levels of phosphorylated Rb (retinoblastoma) (Yuan et al. 2005).

TWO MUTATIONS OF INTEREST (S81L, P99S) AND MUTANT SUMMARY STATISTICS

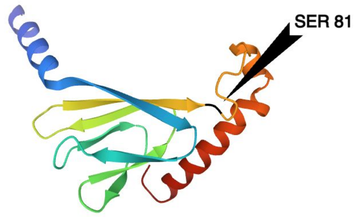


Fig. 1. Serine (residue 81)

A mutation of interest is S81L (Serine to Leucine at residue 81). The residue change is polar to non-polar, has a count of 128, and is a nonconservative mutation occuring on the protein boundary with a large positive energy change of 56.6163245.

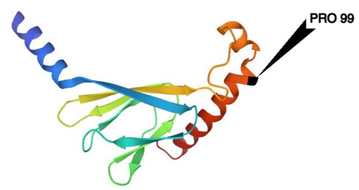


Fig. 2. Proline (residue 99)

A mutation of interest is P99S (Proline to Serine). The residue change is nonconservative, non-polar to polar, and occurs on the surface. P99S has a favorable energy change (-13.11844644) and a count of 6.

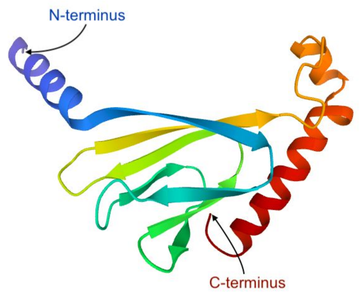
single mutations	185
double mutations	10
multi-point mutations	1
average number of mutations per residue	208/121 = 1.72

Table 1. Summary statistics for variant Orf7a sequences.

More than 40,000 ORF7a sequences were evaluated for mutations following transmission of the virus from Central China to other parts of the world in December 2020. Most ORF7a sequences were unchanged. A total of 196 variant sequences were identified and analyzed for spatial location, conservation, impact on function, and impact on protein stability.

ORF7a is a monomeric protein that contains 121 residues, is triangle-shaped and has measurements 58Å x 48Å x 17Å (height x width x depth). The secondary structure makeup of ORF7a is as follows: alpha helices (34%), beta sheets (34%), and random coils(32%). There are 2 domains. Domain I (residues 1-8) is 89 residues long, and is composed of 7 beta strands and 1 alpha helix, and has the shape of a utility funnel. Domain II (90-12) is 31 residues long, and consists of two alpha helices, and has the shape of the letter L.

Fig. 3. ORF7a 3D Model



CONCLUSION

By analyzing SARS-CoV-2 ORF7a mutations using bioinformatics tools, ORF7a has been shown to undergo mutation that can possibly alter the pathogenicity of SARS-CoV-2. Studying the evolution of SARS-CoV-2 proteins over time using bioinformatics tools can lead to insights into the effects of certain mutations on virus pathogenicity, virulence, and other factors that contribute to its potential to do harm.

Funding is gratefully acknowledged from RSCB PDB, RISE at Rutgers, and the NJ Space Grant Consortium (NJSGC). RCSB PDB is funded by the National Science Foundation (DBI-1832184), the US Department of Energy (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. The David Baker laboratory is to thank for the ORF7a structural model.