Evolution of SARS CoV-2 Proteins in 3D During the COVID-19 Pandemic: Nsp3e NAB domain

Erika McCarthy[1], Lindsey Whitmore[2], Sophia Staggers[3], Elliott M. Dolan[4,5], Changpeng Lu[4], Vidur Sarma[4], Zhuofan Shen[4,5], Lingjun Xie[4,5], Joseph L. Lubin[4,5], Christine Zardecki4[4,6], Sagar D. Khare[4,5], Stephen K. Burley[4,5,6]

[1]Department of Chemistry and Chemical Biology, Stevens Institute of Technology, Hoboken, NJ 07030. [2]Department of Chemistry, Howard University, Washington DC 20059. [3]Department of Chemistry, Frostburg State University, Frostburg, MD 21532. [4] Institute for Quantitative Biomedicine, Rutgers University, Piscataway, NJ 08854. [5]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854. [6]RCSB Protein Data Bank, Rutgers University, Piscataway, NJ 08854

OVERVIEW

Covid-19, is the result of severe acute respiratory syndrome coronavirus 2 (SARS CoV-2), and it has caused a global health and economic crisis. SARS CoV-2 is a single stranded RNA (ssRNA) virus, which prone to mutation. The GISAID database has accumulated variant amino acid sequences obtained from viral genome sequences obtained around the world. In this study we sought to model in three-dimensions and analyze the impact of prevalent mutations in the NAB domain of the Nsp3 protein of the virus. These structural changes may inform the efforts to develop anti-viral countermeasures..

METHODOLOGY

We studied how SARS-CoV-2 proteins evolved during the first six months of the COVID-19 pandemic using structural bioinformatics tools

- Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo/): sequence alignments and phylogenetic trees
- Mol* (molstar.org): 3D molecular visualization
- Foldit (fold.it): structural/energetic effects of sequence mutations
- Reference PDB structure ID: 2K87 (Serrano, P. et al. (2009) J Virol 83: 12998-13008) from SARS-CoV

Nsp3e NAB domain: Nucleic Acid Binding

Nonstructural protein 3 (NSP3) is a 15-domain protein of SARS CoV-2. The nucleic acid binding (NAB) domain is highly conserved from the SARS CoV virus, with expectedly similar ssRNA binding and dsDNA unwinding capabilities. A SARS CoV-2 homology model based on the SARS CoV NAB domain was constructed to analyze the effects of reported mutations on the structure and function of the protein. Positively charged surface residues are essential for NAB function, which is supported by the properties of frequently occurring mutations.

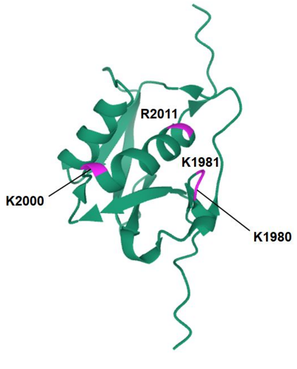

Figure 1. Nsp3 NAB region with RNA-binding residues in pink

Frequently occurring mutations in the NSP3 NAB domain

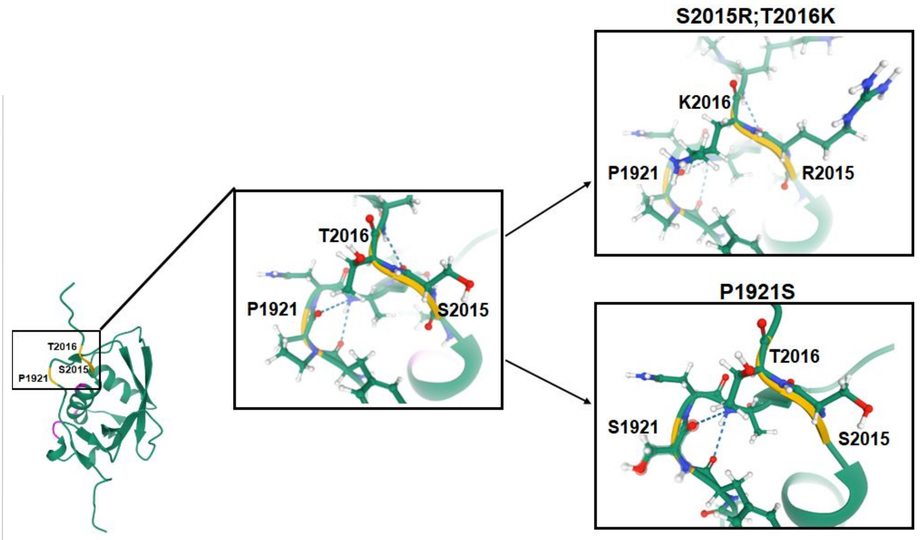

Figure 2. NAB domain of SARS CoV-2: notable mutation sites in yellow

The bound structure of the NAB domain is unknown; however, positively charged surface residues are essential for ssRNA binding. The S2015R, T2016K, and 1921S mutations have occurred 10, 251, and 64 times respectively. The S2105R:T2016K double mutant has occurred 188 times with a calculated ΔG value of 0.5794 kcal/mol from the wild-type to mutant structure.

The frequent occurrence of the S2015R;T2016K mutation with minimal energetic penalty suggests that positively charged residues at this location could aid ssRNA binding.

The P1921S mutation is a nonconservative surface mutation, with a ΔG value of 2.2961 kcal/mol. This is a small energetic penalty, and could potentially stabilize this loop if it were involved in RNA binding.

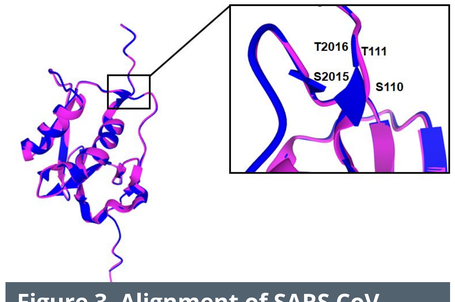

Figure 3. Alignment of SARS CoV (pink) and SARS CoV-2 (blue) NAB regions

The NAB domain of SARS CoV and SARS CoV-2 is highly conserved; however, aligning the structures shows discrepancy in the loop where the above mutations are located. This region was not implicated in RNA binding in SARS CoV, but the differences in structures and frequent mutations to positive residues may indicate a difference in RNA binding patterns.

CONCLUSION

The most frequently occuring mutations in the Nsp3 NAB domain have replaced uncharged surface residues with positively charged ones, This is consistent with the RNA-binding function of this domain. Having a thorough understanding of the structure of the viral proteome and how it is changing is essential for vaccine and drug development.

RCSB PDB is funded by the National Science Foundation (NSF; 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 (NIH) under grant R01GM133198. Khare lab is funded by NSF (CBET 1929237) and NIH (R01GM132565). This work was supported by the Research Intensive Summer Experience (RISE) at Rutgers program and the New Jersey Space grant.