

## Mapping COVID-19 Mutations in 3D: How SARS-CoV-2 has evolved during the first 6 months of the global pandemic

A Khan<sup>1</sup>, C Zardecki<sup>2</sup>, S Burley<sup>3</sup>, S Khare<sup>4</sup>

<sup>1</sup>*RCSB Protein Data Bank, Takoma Park, MD*, <sup>2</sup>*Rutgers Proteomics, RCSB Protein Data Bank, Piscataway, NJ*, <sup>3</sup>*RCSB Protein Data Bank, Rutgers University, Piscataway, NJ*, <sup>4</sup>*Institute for Quantitative Biomedicine*  
*aaliyah.khan@rcsb.org*

The World Health Organization reported that the number of COVID-19 cases worldwide was over 10 million with over 500,000 deaths as of June 30, 2020. The causative agent of COVID-19, SARS-CoV-2, has a very long single-strand, positive sense, 5'-capped RNA genome that encodes ~30 proteins. As the virus spread worldwide following its exit from central China and viral genomes isolated from infected individuals were sequenced, many mutations of viral proteins were observed. Studying mutations of key viral proteins can give insight into the virus life cycle and may impact discovery and development of new antiviral drugs and vaccines. In order to do this, our team, as part of a virtual summer research experience with the RCSB PDB, studied how SARS-CoV-2 proteins evolved during the first six months of the COVID-19 pandemic. This was done by exploring amino acid sequence and 3D atomic-level structure using various structural bioinformatics tools, including Clustal Omega ([www.ebi.ac.uk/Tools/msa/clustalo/](http://www.ebi.ac.uk/Tools/msa/clustalo/)) for sequence alignments and phylogenetic trees; Mol\* ([molstar.org](http://molstar.org)) for 3D molecular visualization; and Foldit ([fold.it](http://fold.it)) for structural/energetic effects of sequence mutations. One key protein for SARS-CoV-2 is non-structural protein 14 (Nsp14). This protein supports both viral genome capping with a 5' 7-methyl-guanine cap and proofreading. It shares 95% sequence identity with Nsp14 of SARS-CoV, demonstrating a high level of conservation. More than 300 unique variants of Nsp14 were modeled using the structural bioinformatics tools and analyzed using PyRosetta. We then identified certain mutations for further study on the basis of the affected amino acid's role in protein function and its location within the 3D structure.

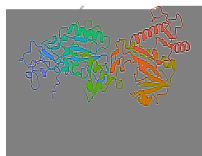


Figure 1