Using data mining to identify variations within the proteases of Coronaviridae

Cassandra Olivas[1], Christine Zardecki[2], Sagar Khare[3], Stephen Burley[4]

[1]California State University, Stanislaus [2]Rutgers Proteomics, RCSB Protein Data Bank, [3]Institute for Quantitative Biomedicine, [4]RCSB Protein Data Bank, Rutgers University

colivas1@csustan.edu

Coronaviruses (Coronaviridae) have been the cause of several recent outbreaks, including SARS-CoV-1, MERS-CoV, and SARS-CoV-2, which is the causative agent for COVID-19. Although efforts to vaccinate against the current pandemic are in effect, there is a question of the efficacy of available vaccines against additional strains . Therefore, focusing efforts on a conserved area within the viral genome may provide a way to offer broad-spectrum coverage against future variants. In this study, we focused on the two proteases that are conserved across all species: the main protease and the papain-like protease. Using sequences from the NCBI Virus database, we created 3D models using the Rosetta software package and experimentally determined structures from the Protein Data Bank (PDB, https://www.rcsb.org/). By focusing on residues adjacent to or directly bonded with the catalytic sites within each protease, we will attempt to identify possible targets for small-molecule binding. In this poster, I will discuss our methodology including how we are refining our search criteria and addressing discrepancies.