

## Amino Acid Occurrence Rates in the SARS-CoV-2 Proteome

Yekbun Adiguzel      yekbun.adiguzel@atilim.edu.tr Turkey Atilim University

This study aims to investigate SARS-CoV-2 amino acid (aa) occurrence rates (i.e., frequencies) to compare those with that of humans, assuming/expecting higher number of host-pathogen common sequences in the respective proteomes with closer aa frequencies. It can be related to adaptation, and can put susceptible individuals into the molecular mimicry-based autoimmunity risk upon infection.

With the given aim, frequencies of aa in the viral proteome were obtained from 12 proteins in the reference genome of SARS-CoV-2 in NCBI (NC\_045512.2). Overall aa frequencies were then compared to the aa frequencies in the reference genome of human (GRCh38.p13). Redundancy in the human proteome, e.g., in the form of isoforms, were not taken into account. Respective data of the other viruses (SARS-CoV, MERS, Influenza A, HIV 1, Hepatitis C, Zika, Dengue) were also calculated through the information at NCBI and compared. Statistical tests were finally performed and p-value less than 0.05 was deemed significant.

Results of SARS-CoV-2 with single letter abbreviations of the aa, corrected for the overlapping parts of Orf 1ab and Orf 1a proteins are as follows: 0.07 (A), 0.03 (C), 0.05 (D), 0.05 (E), 0.05 (F), 0.06 (G), 0.02 (H), 0.05 (I), 0.06 (K), 0.10 (L), 0.02 (M), 0.05 (N), 0.04 (P), 0.04 (Q), 0.04 (R), 0.07 (S), 0.07 (T), 0.08 (V), 0.01 (W), and 0.05 (Y).

Amino acid frequencies of SARS-CoV-2 are not significantly different than of humans or deviating more frequently than that of the other viruses. This indifference can be related to the humans being the host. However, assuming higher possible pathogenicity due to the lack of a significant difference between the aa frequencies in the viral proteome and the human proteome would be a premature deduction at this stage, regarding the aim of this study. Comparisons should involve viruses with distinct hosts.