The analysis of conservative and non-conservative amino acid mutations in SARS-CoV-2 variants provides insights into the early evolution of the virus in humans

Since the end of 2019, the world is experiencing a threatening and global health emergency associated with a Severe Acute Respiratory Syndrome caused by a hitherto unknown coronavirus (SARS-CoV-2). Despite the enormous efforts made globally, the development of effective therapeutic or preventive approaches for this disease is still an ongoing process. The ability of SARS-CoV-2 to mutate rapidly represents, among others, a remarkable complicacy. In this scenario, quantitative evaluations of the effects that these mutations have on the virus structure/function could have a profound impact in in this field. Moreover, the availability of a large number of SARS-CoV-2 sequences since the early phases of the pandemic represents a unique opportunity to follow the adaptation of the virus to humans (evolution in action). Here, we evaluated the SARS-CoV-2 amino acid mutations and their progression by analyzing genomes deposited at different stages of the pandemic (March 15th and October 7th). These mutations were classified in conservative and non-conservative ones based on the probability to be accepted during the evolution according to the Point Accepted Mutation substitution matrices and on the similarity of the encoding codons. The comparative analysis of mutations detected at these two stages of the pandemic unravels significant analogies despite the large difference in their overall content. These include the most frequent substitutions (T>I, L>F, and A>V) and the accumulation of hydrophobic residues associated, on average, with the mutation events. On the other hand, the analysis of the most recent dataset indicates that non-conservative mutations present lower frequencies than conservative ones. This observation may be ascribed to a progressive adaptation of the virus to the host. In conclusion, the present study provides some interesting indications of the early evolution of the virus and useful tools for the global and genome-specific evaluation of the impact that mutations could have on the structure/function of SARS-CoV-2 variants that emerged or will emerge in the pandemic [1].

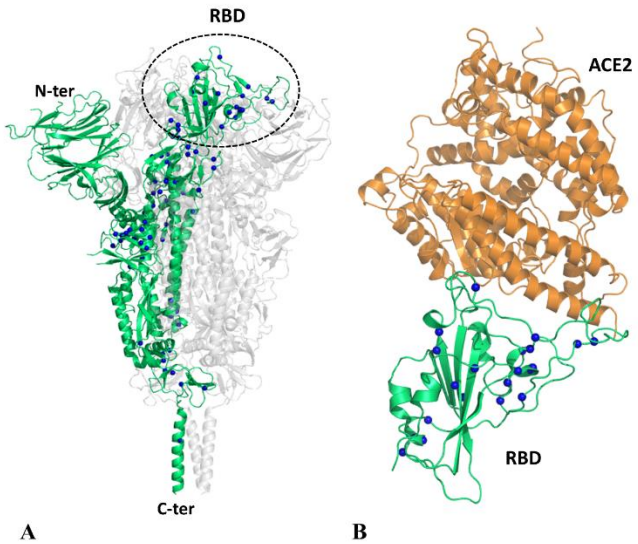

Figure. Cartoon representation of the three-dimensional structure of the the SARS-CoV-2 Spike protein. The protein trimer (PDB ID 6xr8) and the complex of the Spike Receptor Binding Domain (RBD) with the cell receptor ACE2 (PDB ID 6m0j) are shown in the panels A and B, respectively. The location of the residues that have never been found to be changed up to October 7th 2020 is shown as blue balls.

References:
[1] N. Balasco, G. Damaggio, L. Esposito, F. Villani, R. Berisio, V.Colonna, L. Vitagliano *Submitted*.

Keywords: SARS-CoV-2, Substitution matrices, Amino acid Mutations.
Contacts: Luigi Vitagliano (luigi.vitagliano@unina.it), Nicole Balasco, Luciana Esposito, Rita Berisio, Gianluca Damaggio (IGB-CNR), Flavia Villani (IGB-CNR), Vincenza Colonna (IGB-CNR).
Website(s): http://www.ibb.cnr.it/?command=viewu&id=86
Other: The present activities have been performed in the framework of the project “RicErCa e sviluppO VERsus COVID19 in Campania RECOVER-COVID19” funded by the Campania Region.