

<https://doi.org/10.7124/bc.000B5B>

O.V. Ketsa

Yuriy Fedkovych Chernivtsi National University
2, Kotsiubynskoho Str., Chernivtsi, Ukraine, 58002
o.ketsa@chnu.edu.ua

BIOINFORMATIC ANALYSIS OF VARIABILITY AND CONSERVATION OF THE SARS-COV-2 SPIKE PROTEIN USING MULTIPLE SEQUENCE ALIGNMENT

Background/Aim. The Spike (S) protein of SARS-CoV-2 is the main surface protein and key antigen responsible for viral infectivity and immune response. Due to its high mutation rate, there is a need to identify conserved regions as potential biomarkers and targets for vaccines and targeted therapy. The aim of this study was to analyze the variability of Spike protein based on the protein sequence alignment and to identify conserved regions. **Methods.** Amino acid sequences of the SARS-CoV-2 Spike protein from different viral variants, including a reference sequence and clinical isolates, were obtained from the NCBI GenBank database. Multiple sequence alignment (MSA) was performed using the MAFFT algorithm, while conservation of amino acid positions was assessed in the Jalview software environment using entropy-based analysis. **Results.** Multiple sequence alignment of the Spike protein amino acid sequences revealed a clearly heterogeneous distribution of variability along the molecule, reflecting its functional and structural differentiation. The highest level of variability was observed in the receptor-binding domain (RBD), where a substantial number of amino acid substitutions were identified among major viral variants (Alpha, Delta, and Omicron). This confirms the high evolutionary plasticity of this region, which is under strong selective pressure from the host immune system and plays a key role in binding to the ACE2 receptor. In contrast, MSA and entropy-based analysis identified three well-defined conserved regions with low entropy values (<0.2), predominantly localized within the S2 subunit of the Spike protein. These regions exhibit a high degree of evolutionary conservation across variants, indicating their critical structural and functional importance. The most stable amino acid positions are concentrated in the C-terminal region of S2, which is directly involved in the membrane fusion process between the virus and the host cell, a key step in viral infection. Furthermore, conserved regions were maintained across evolutionarily distant SARS-CoV-2 variants, suggesting limited mutational tolerance and functional indispensability. Entropy analysis further confirmed that the S2 subunit exhibits significantly higher structural stability compared to the S1 and RBD domains, making it a more suitable source of stable molecular targets for diagnostic, vaccine, and therapeutic applications. **Conclusions.** The obtained results confirm that conserved regions of the S2 subunit represent the most promising molecular targets for further development of diagnostic, vaccine, and therapeutic approaches.

Keywords: SARS-CoV-2, Spike protein, multiple sequence alignment, bioinformatics, conserved regions, entropy analysis.

Received: 07.07.2026; Accepted: 08.09.2026; Published: 29.09.2026

Citation: Ketsa O.V. (2026) Bioinformatic analysis of variability and conservation of the SARS-CoV-2 spike protein using multiple sequence alignment. *Biopolymers & Cell*, Special issue(42), 34. <https://doi.org/10.7124/bc.000B5B>
© Publisher PH "Akadempriodyka" of the NAS of Ukraine, 2026. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited