

Original Article



In-silico comparison of post-translational modifications of SARS-CoV and SARS-CoV-2 structural proteins

Naghmeh Soltani¹, Fatemeh Pakniya¹, Niloofar Parchami¹, Mandana Behbahani¹, Hassan Mohabatkar^{1*}

¹Faculty of Biological Science and Technology, University of Isfahan, Iran

***Corresponding Author:** Hassan Mohabatkar, Faculty of Biological Science and Technology, University of Isfahan, Iran. Tel: 0098 (313) 793 4391, Email: h.mohabatkar@ast.ui.ac.ir

Abstract

Background and aims: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a newly discovered coronavirus which causes an infectious disease. The severe acute respiratory syndrome (SARS-CoV) and the Middle East respiratory syndrome (MERS) broke out in 2003 and 2012, respectively. These viruses have some structural proteins, including spike (S), membrane (M), envelope (E), and nucleocapsid (N) proteins. These proteins assist the virus in infecting cells through interaction with cell receptors, penetration into the cell, and proliferation. These coronavirus proteins are modified by post-translational modifications (PTMs) which activate various functional and interactional activities of proteins. This study aimed to investigate the PTMs in SARS-CoV/CoV-2, as well as to examine the effect of these PTMs on the pathogenicity of these two viruses.

Methods: In this study, PTMs sites were detected by different bioinformatics tools. Evaluation and comparison of PTMs were performed and their roles in structural proteins activities of SARS-CoV/CoV-2 coronaviruses were examined in order to gain a richer understanding of these modifications' relationships with the protein activities.

Results: The PTMs sum and percentages of four structural proteins of SARS-CoV/CoV-2 were evaluated, with a focus on their effects on viral replication and pathogenesis in order to develop a method for treating these diseases. According to our study results, some of the PTMs in SARS-CoV/CoV-2 were different from each other.

Conclusion: It was concluded that SARS-CoV-2 had more pathogenicity than SARS-CoV.

Keywords: PTMs, Structural proteins, Coronaviruses, Bioinformatics tools

Received: 25 April 2021, Accepted: 17 July 2021, ePublished: December 6 2021

Introduction

Since the outbreak of acute respiratory syndrome, several other coronaviruses have been observed in bats and other natural hosts. Scientists studying these coronaviruses has concluded that human may become infected with these coronaviruses in the future. In the last two decades, coronaviruses have caused two widespread pandemics, namely severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) (1). In late December 2019, new human cases of pneumonia were investigated by Chinese health officials, most of whom having caught it from those visiting the seafood market or coming into contact with live animals in Wuhan, Hubei Province, China. In the earlier phase of the disease, patients exhibited symptoms of fever and cough, but then they started complaining of chest discomfort and shortness of breath. The chest radiographs (CT) scans from the patients showed pulmonary sclerosis (2).

The new 2019 coronavirus (CoV) caused an infection in the lower respiratory tract. It was first called the new coronavirus pneumonia (NCP) disease by the Chinese government, but then was renamed COVID-19 by the World Health Organization (WHO). Finally, the

2019-nCoV virus was renamed SARS-CoV-2 by The International Committee for Taxonomy of Viruses (3).

After the outbreak of SARS-CoV in 2002 and 2003, a number of Chinese medical practitioners examined the genome of SARS-CoV to explain how the virus bound to its victims and how it worked; however, they failed in their initial endeavors. Then bioinformatics researches were carried out in various regions. Researches of bioinformatics nature were instrumental in studying and diagnosing the disease in the given particular circumstances. They were conducted around the world to slow down the spread of the disease and to gain global support against it, resulting in publication of more than one hundred scientific articles based on bioinformatics data. In the meantime, many researchers became interested in some active research areas including cellular and molecular biology of viruses and their mutations. Pathology was involved in the prognosis of epitopes and immune mechanisms, drug design, vaccine design and production expansion, early diagnosis, prevention, and treatment (4). At present the world is facing with a widespread epidemic that first afflicted more than one thousand Chinese couple of years ago but then was rapidly spread around the world. WHO

declared the disease as a global crisis in late January 2020. The prevalence of COVID-19 disease is extremely high and unprecedented, and the scientific community must develop a sensible approach to respond to this global crisis. Recent findings can assist in understanding and comprehending the content and cause(s) of the disease's dynamic transfer.

Proteins can undergo various post-translational modifications (PTMs) (5). PTMs occur in almost all proteins and play important roles in activation and structure of the proteins. Many biological processes are controlled by PTMs. They perform important functions such as placing proteins in specific locations, stabilizing proteins, and regulating enzymatic activity (6). The coronaviruses contain four structural proteins: the spike (S) protein helping the virus bind to the host cell receptor and infect the host cell (7); the membrane (M) protein facilitating virus budding (8); the envelope (E) protein playing a major role in the assembly of viral particles (9); and the nucleocapsid (N) protein binding to the genomic RNA to form the nucleocapsid. The N protein may be involved in regulating the RNA production of the virus (10). Due to the role of the PTMs in the function of proteins, this study aimed to investigate the PTMs in SARS-CoV/CoV-2, and to examine the effect of PTMs on the pathogenicity of the given two viruses.

Methods

In this study, some in-silico platforms were applied to predict different PTMs sites of structural proteins of SARS-CoV and SARS-CoV-2 (Figure 1). The sequences of structural proteins (N, M, S, and E) of coronaviruses (SARS-CoV and SARS-CoV-2) were obtained from NCBI database (<https://www.ncbi.nlm.nih.gov/>). The accession numbers for SARS-CoV N, S, M, and E proteins were QIH55228.1, QHR63290.2, QIA98586.1, and QIA98598.1, respectively; and the accession numbers for SARS-CoV-2 N, S, M, and E proteins were NP_828858.1, NP_828851.1, AAT76152.1, and NP_828854.1, respectively. The FASTA

formats of these proteins were downloaded to survey their PTMs. These structural proteins went through some PTMs to form the mature protein products. These PTMs were predicted by several bioinformatics servers (11) (Table 1).

N-Glycosylation

The N-Glycosylation occurs in S, M, and E structural proteins of SARS-CoV and SARS-CoV-2. NetNGlyc server (<http://www.cbs.dtu.dk/services/NetNGlyc>) was used in order to identify the potential of N-Glycosylation sites in these structural proteins. This server predicted N-Glycosylation by examining the sequence background of Asn-Xaa-Ser/Thr and using artificial neural networks. With an average accuracy of 76 percent, the networks were able to classify 86% of glycosylated, and 61 percent of non-glycosylated sequence. The protein sequences were examined using FASTA format and considering the thresholds equal to or higher than 0.5 (12).

Phosphorylation

Another modification is phosphorylation that occurs in N protein. To identify these sites, the NetPhos server (<http://www.cbs.dtu.dk/services/NetPhos>) was adopted. The server was a neural network that predicted phosphorylation sites in independent sequences with a sensitivity of 69% to 96%. The input was FASTA format, and the server could examine the Ser, Thr, and Tyr amino acids that may have been modified at a threshold equal to or higher than 0.5. The predictions were made for the following 17 kinases: ATM, CKI, CKII, CaMII, DNAPK, EGFR, GSK3, INSR, PKA, PKB, PKC, PKG, RSK, SRC, cdc2, cdk5 and p38MAPK (13).

Palmitoylation

The S and E proteins undergo the Palmytoilation. The CSS-Palm server (<http://csspalm.biocuckoo.org/online.php>) was designed to identify this kind of modification. It was a novel computational method/software the results

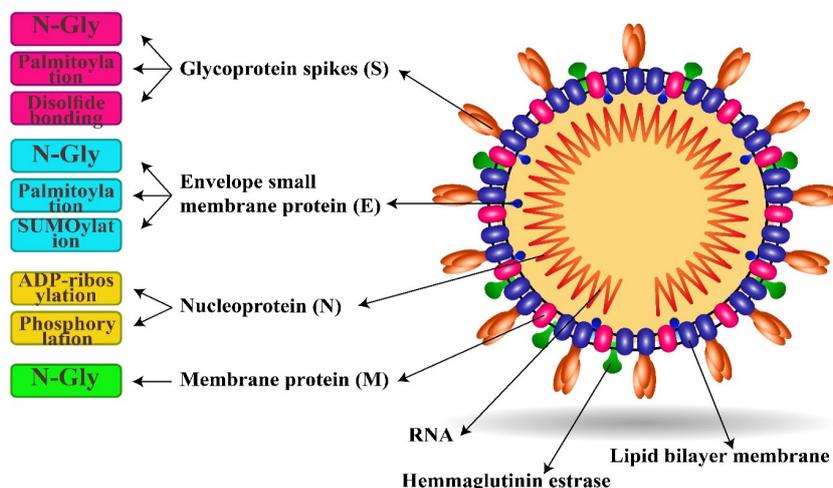


Figure 1. Schematic diagram showing the PTMs of coronavirus S, E, N, and M proteins.

Table 1. Servers used for predicting the PTMs of SARS-CoV and SARS-CoV-2

PTM	Protein	Server
N-Glycosylation	S	http://www.cbs.dtu.dk/services/NetNGlyc
	M	
	E	
Palmitoylation	S	http://csspalm.biocuckoo.org/online.php
	E	
Disulfide Bonding	S	http://disulfind.dsi.unifi.it
ADP-ribosylation	N	https://www.adpredict.net/index.php
Phosphorylation	N	http://www.cbs.dtu.dk/services/NetPhos
Sumoylating	N	http://predictor.nchu.edu.tw/SUMOgo
Ubiquitination	E	http://www.ubpred.org/

of which were based on a clustering and scoring strategy. The medium threshold was selected for examination (14).

Disulfide bonding

Disulfide bond is another modification that occurs in S protein. The DISULFIND server (<http://disulfind.dsi.unifi.it>) was an indicator of cysteine disulfide bonding state and connectivity. The server made prediction in two computational stages: 1) A BRNN-SVM binary classifier predicting each cysteine's disulfide bonding state, and 2) A recursive neural network pairs; Cysteines that were recognized in stage one to produce a connectivity pattern (15).

ADP-ribosylation

In this study, some PTMs were specific for proteins such as ADP-ribosylation for N protein. ADPredict (<https://www.adpredict.net/index.php>) was a web application and predicted ADP-ribosylation of glutamic and aspartic residues. The threshold score in this server was equal to or higher than 0.4. In the present study, however, a score equal to or higher than 0.5 was selected, which was the same score selected in previous studies (16).

Sumoylating

The N protein undergoes the sumoylating modification. This PTM was predicted by GPS-SUMO server (<http://predictor.nchu.edu.tw/SUMOgo>). The server was able to investigate the relationship between sumoylating and SUMO interaction processes. The sumoylating results with medium threshold were selected (17).

Ubiquitination

Ubiquitination occurs in E protein. UbPred (<http://www.ubpred.org/>) was the server employed to predict ubiquitination sites in the protein. UbPred achieved 72%

class-balanced precision, with 80 percent field under the ROC curve. Predictions were made about all lysine residues of a query sequence. The server considered only ubiquitination for lysins with a score of equal to or higher than 0.62 (18).

The percentage of PTMs in each terminal domain of proteins was detected to make more comparisons among the proteins of viruses. The protein sequences were divided into three areas: the first one was the N-terminal; the second one was the Middle domain; and the last one was the C-terminal.

Results

PTM in S protein

N-linked glycosylation

N-linked glycosylation may assist in protein folding and plays an essential role in the functionality of S protein (19). The amino acid sequence of S protein from the SARS-CoV and SARS-CoV-2 were analyzed by the NetNGlyc server. The 17 N-linked glycosylation sites were predicted in the S protein of both viruses.

Palmitoylation and disulfide bond formation

The results from the bioinformatics studies of the above-mentioned PTM are presented in Table 2. These results showed that the content of amino acid units with palmitoylation and disulfide bonds of SARS-CoV-2 was higher than that of SARS-CoV. This analysis predicted 7 and 27 palmitoylation and disulfide bonds sites in the S protein of SARS-CoV. The numbers for amino acid units undergone these two PTMs in SARS-CoV-2 were 10 and 37, respectively.

PTM of E, M and N proteins

The PTM in E, M, and N proteins are shown in Table 2. The results indicated that the numbers for Palmitoylation and N-Glyc sites of E protein in SARS-CoV, and SARS-CoV-2 were the same. Estimating the ways the PTMs were distributed in the N-terminal and C-terminal domain of S, M, N, and E proteins of SARS-CoV and SARS-CoV-2 (Figure 2) demonstrated that S protein glycosylation sites had been predicted in the N-terminal, middle and C-terminal domain of both viruses. Therefore, the rates of glycosylation in the N-terminal, middle and C-terminal domain of SARS-CoV, and SARS-CoV-2 were the same. Palmitoylation was not seen in the middle domain of SARS-CoV and SARS-CoV-2. Additionally, this PTM in the C-terminal domain of SARS-CoV-2 was higher than that in the SARS-CoV.

The M protein Glycosylation in both viruses was

Table 2. PTMs of S, E, M, and N proteins of SARS-CoV and SARS-CoV-2

Protein	S			M		E		N		
	N-Glyc	Disulfide Bonding	Palm	N- Glyc	N- Glyc	Palm	Ubiq	Phos	ADP-ribosylation	SUMO
SARS-CoV	17	27	7	1	2	3	0	109	3	2
SARS-CoV-2	17	37	10	1	2	3	0	65	3	2

predicted only in the N-terminal domain. The prediction showed that the sumoylation of the N protein in SARS-CoV and SARS-CoV-2 was found only in the Middle domain. The N protein ADP-ribosylation of two viruses in the N-terminal domain and Middle domain was anticipated to be identical. The rate of N protein phosphorylation in the C-terminal and N-terminal domains of SARS-CoV-2 was slightly higher than that of the SARS-CoV.

The E protein palmitoylation in SARS-CoV and SARS-CoV-2 was predicted only in the middle domain. Glycosylation of E protein in SARS-CoV and SARS-CoV-2 was distributed in the Middle and C-terminals domains.

Discussion

This study revealed that the infectivity of SARS-CoV-2 was higher than that of SARS-CoV. Palmitoylation of viral proteins play an essential role in virus assembly and infection (20). Due to the increasing level of S protein palmitoylation in SARS-CoV-2, the infectivity of SARS-CoV-2 might be higher than SARS-CoV. N-linked glycosylation of S protein facilitates the formation of coronavirus S protein, followed by the binding of S protein to the host cell receptor (11). The number for N-linked glycosylation sites of S protein in SARS-CoV and SARS-

CoV-2 was the same. Therefore, the ability of both viruses to bind to the host cell receptor was expected to be the same.

The E protein Palmitoylation helps regulate protein function within the membrane, and leads to increased hydrophobicity and easier passage of the virus through the cell membrane (21). The E protein Palmitoylation in SARS-CoV and SARS-CoV-2 was found to be the same. It was predictable that both viruses would be similar in performing this function. Previous studies have shown that glycosylation of the E protein plays role in increasing the molecular weight and oligomerization of E protein which leads to increased pathogenicity (21). Due to the similar level of glycosylation in E protein of both viruses, the E protein-induced pathogenicity in both viruses was predicted to be similar.

Several studies have determined that sumoylation of the N protein may play an important role in nucleocapsid formation and in causing disorganization in host cell division (22). Bioinformatics predictions in this study suggested that the numbers for N protein sumoylation sites in SARS-CoV and SARS-CoV-2 were equal. In SARS-CoV-2, the number for N protein phosphorylation sites decreased. The SARS-CoV-2 N protein is targeted

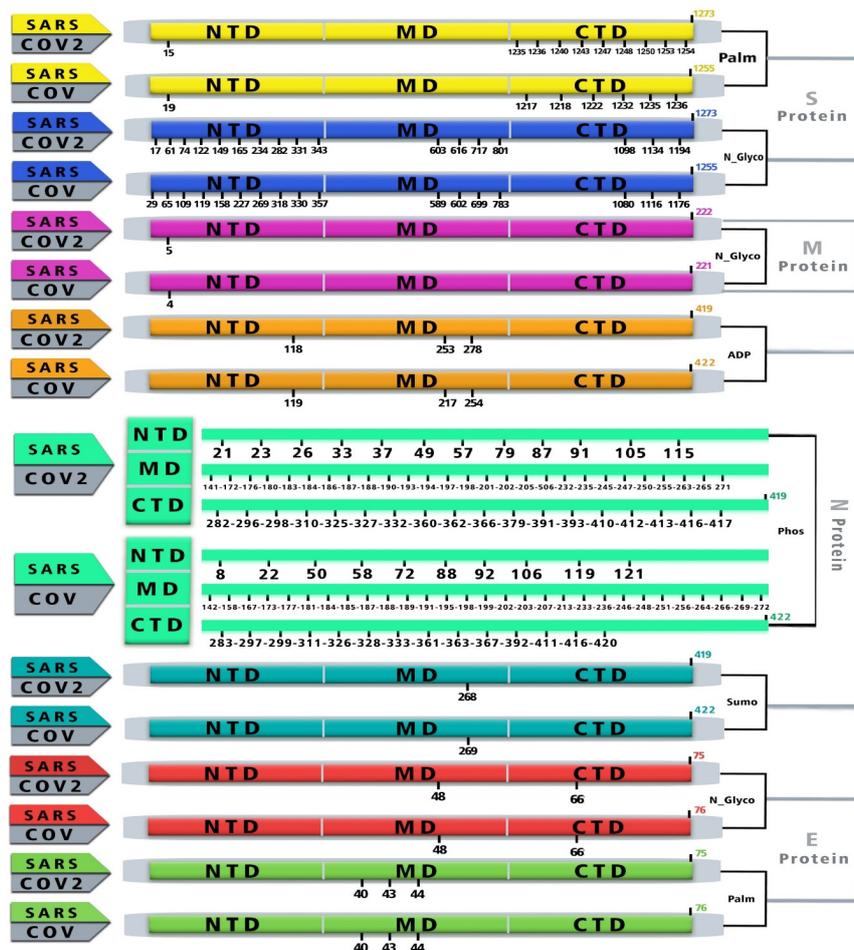


Figure 2. Schematic diagram showing functional domain and PTMs on SARS-CoV and SARS-CoV-2 S, M, N, and E proteins (not to scale). NTD: N-terminal domain, MD: Middle domain, CTD: C-terminals domain, S: Spike, M: Membrane, E: Envelope, N: Nucleocapsid, Palm: Palmitoylation, N- Glyco: N-Glycosylation, ADP: ADP-ribosylation, Phos: Phosphorylation, Sumo: Sumoylating.

by the host cell protein kinase. The functional role of N protein phosphorylation is not precisely understood (23).

The angiotensin-converting enzyme 2 (ACE2) helps functional receptor to insert SARS-CoV and SARA-CoV-2 into the host cell. The receptor-binding domain of the S proteins and ACE2 are available for several cysteines, which leads to the formation of disulfide bonds. This connection is associated with the increase in the ability to reproduce the virus. If disulfide groups are changed into thiol and sulfhydryl groups, the binding of viral S protein into ACE2 is impaired (24). Taking into account the previous research results and the increase in disulfide bonds in SARS-CoV-2 found in our study, it was suggested that this increase could have been considered as the strength and weakness of SARS-CoV-2 compared to SARS-CoV. As the disulfide bonds increase, the probability of the binding to ACE2 also increases, which leads to an improvement in the pathogenicity of the virus. If the sulfhydryl groups replace the disulfide groups, however, the binding of the S protein to ACE2 is disrupted.

Conclusion

According to the results from collected data, it was concluded that coronavirus proteins underwent a variety of PTMs. The PTMs play an important role in coronavirus proteins' functions. The extent and location of these PTMs in different species of coronavirus could vary. In sum, the numbers for palmitoylation and disulfide bonds sites in the S protein and N protein phosphorylation sites of SARS-CoV and SARS-CoV-2 were different. Due to this difference in modification, a change was observed in the rate and severity of coronavirus pathogenesis. Overall, it was found that the scale of PTMs in SARS-CoV and SARS-CoV-2 was different, and SARS-CoV-2 had more pathogenicity than SARS-CoV. The practical implications of some of these PTMs on the structural and non-structural proteins of the coronavirus are still unclear. Studying these PTMs with different methods and tools can facilitates resolving the existing ambiguities. Examining in vivo and in vitro models as well as investigating these PTMs at the genetic and epigenetic levels could increase our knowledge about the coronavirus in the future.

Conflict of Interests

The authors declare that they have no conflict of interests.

Ethical Approval

This study did not include human participants and/or animal subjects. All procedures were approved by the Ethics committee of University of Isfahan, Iran (Ethic code: IR.UI.REC.1400.NN6).

Authors' Contribution

NS and FP and NP contributed equally to this study. HM and MB conceived of the presented idea and supervised the project. NS and FP and NP developed and performed the theory and wrote the manuscript. All authors discussed the results and contributed to the final manuscript. All figures are designed by the authors.

Funding/Support

This study was supported by the University of Isfahan.

Acknowledgments

This project was derived from the bioinformatics research. The project did not receive a grant and did not need an approval from the University of Isfahan.

References

- Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-3. doi: 10.1038/s41586-020-2012-7.
- Binnicker MJ. Emergence of a novel coronavirus disease (COVID-19) and the importance of diagnostic testing: why partnership between clinical laboratories, public health agencies, and industry is essential to control the outbreak. *Clin Chem*. 2020;66(5):664-6. doi: 10.1093/clinchem/hvaa071.
- Yuen KS, Ye ZW, Fung SY, Chan CP, Jin DY. SARS-CoV-2 and COVID-19: the most important research questions. *Cell Biosci*. 2020;10:40. doi: 10.1186/s13578-020-00404-4.
- Hao P, Chen M, Zhang G, He W, Li Y. Bioinformatics research on the SARS coronavirus (SARS_CoV) in China. *Curr Pharm Des*. 2006;12(35):4565-72. doi: 10.2174/138161206779010404.
- Cleemput S, Dumon W, Fonseca V, Abdool Karim W, Giovanetti M, Alcantara LC, et al. Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes. *Bioinformatics*. 2020;36(11):3552-5. doi: 10.1093/bioinformatics/btaa145.
- Audagnotto M, Dal Peraro M. Protein post-translational modifications: in silico prediction tools and molecular modeling. *Comput Struct Biotechnol J*. 2017;15:307-19. doi: 10.1016/j.csbj.2017.03.004.
- Xia X. Domains and functions of spike protein in SARS-Cov-2 in the context of vaccine design. *Viruses*. 2021;13(1). doi: 10.3390/v13010109.
- Masters PS. Coronavirus genomic RNA packaging. *Virology*. 2019;537:198-207. doi: 10.1016/j.virol.2019.08.031.
- Mukherjee S, Bhattacharyya D, Bhunia A. Host-membrane interacting interface of the SARS coronavirus envelope protein: Immense functional potential of C-terminal domain. *Biophys Chem*. 2020;266:106452. doi: 10.1016/j.bpc.2020.106452.
- Lin Z, Gao Q, Qian F, Jinlian M, Lishi Z, Yu Q, et al. The nucleocapsid protein of SARS-CoV-2 abolished pluripotency in human induced pluripotent stem cells. *SSRN Electronic Journal*. 2020. doi: 10.2139/ssrn.3561932.
- Fung TS, Liu DX. Post-translational modifications of coronavirus proteins: roles and function. *Future Virol*. 2018;13(6):405-30. doi: 10.2217/fvl-2018-0008.
- Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput*. 2002;310-22.
- Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol*. 1999;294(5):1351-62. doi: 10.1006/jmbi.1999.3310.
- Zhou F, Xue Y, Yao X, Xu Y. CSS-Palm: palmitoylation site prediction with a clustering and scoring strategy (CSS). *Bioinformatics*. 2006;22(7):894-6. doi: 10.1093/bioinformatics/btl013.
- Ceroni A, Passerini A, Vullo A, Frascioni P. DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. *Nucleic Acids Res*. 2006;34(Web Server issue):W177-81. doi: 10.1093/nar/gkl266.
- Lo Monte M, Manelfi C, Gemei M, Corda D, Beccari AR. ADPredict: ADP-ribosylation site prediction based on physicochemical and structural descriptors. *Bioinformatics*. 2018;34(15):2566-74. doi: 10.1093/bioinformatics/bty159.
- Zhao Q, Xie Y, Zheng Y, Jiang S, Liu W, Mu W, et al. GPS-SUMO: a tool for the prediction of sumoylation sites and

- SUMO-interaction motifs. *Nucleic Acids Res.* 2014;42(Web Server issue):W325-30. doi: 10.1093/nar/gku383.
18. Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, et al. Identification, analysis, and prediction of protein ubiquitination sites. *Proteins.* 2010;78(2):365-80. doi: 10.1002/prot.22555.
 19. Zheng J, Yamada Y, Fung TS, Huang M, Chia R, Liu DX. Identification of N-linked glycosylation sites in the spike protein and their functional impact on the replication and infectivity of coronavirus infectious bronchitis virus in cell culture. *Virology.* 2018;513:65-74. doi: 10.1016/j.virol.2017.10.003.
 20. McBride CE, Machamer CE. Palmitoylation of SARS-CoV S protein is necessary for partitioning into detergent-resistant membranes and cell-cell fusion but not interaction with M protein. *Virology.* 2010;405(1):139-48. doi: 10.1016/j.virol.2010.05.031.
 21. Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. *Virology.* 2019;16(1):69. doi: 10.1186/s12985-019-1182-0.
 22. Li FQ, Xiao H, Tam JP, Liu DX. Sumoylation of the nucleocapsid protein of severe acute respiratory syndrome coronavirus. *FEBS Lett.* 2005;579(11):2387-96. doi: 10.1016/j.febslet.2005.03.039.
 23. Tugaeva KV, Hawkins D, Smith JLR, Bayfield OW, Ker DS, Sysoev AA, et al. The mechanism of SARS-CoV-2 nucleocapsid protein recognition by the human 14-3-3 proteins. *J Mol Biol.* 2021;433(8):166875. doi: 10.1016/j.jmb.2021.166875.
 24. Hati S, Bhattacharyya S. Impact of Thiol-Disulfide Balance on the Binding of Covid-19 Spike Protein with Angiotensin-Converting Enzyme 2 Receptor. *ACS Omega.* 2020;5(26):16292-8. doi: 10.1021/acsomega.0c02125.