 doi:10.20944/preprints202106.0238.v1

Role of N501Y mutation in SARS-CoV-2 spike protein structure

Urmi Roy[*]

[*]Department of Chemistry & Biomolecular Science

Clarkson University

8 Clarkson Avenue, Potsdam, NY 13699-5820, United States

[*]To whom correspondence should be addressed. E-mail: urmi@clarkson.edu; Phone: 315/212-7346

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2021 doi:10.20944/preprints202106.0238.v1

Abstract

It has been more than a year since the first case of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was found. This coronavirus has infected more than 110 million people worldwide by the end of February, 2021, and several virulent as well as more spreadable mutant forms of SARS-CoV-2 have emerged subsequently. In the latter group, three variants B.1.1.7, B.1.351, and P1 lineages, have been reported. Using computer simulation, the present paper investigates the structural differences between the wild type SARS-CoV-2 spike protein and its Asn501Tyr (N501Y) mutant variant. Time-based structural changes between the receptor binding domains of these two species are also examined. The N501Y mutation is common to all the three aforesaid mutant variants.

Keywords: Immunoinformatics • Molecular dynamics simulation • N501Y mutation • SARS-CoV-2 • Spike protein • Virus structure

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2021 doi:10.20944/preprints202106.0238.v1

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has already infected more than 110 million people and 2.4 million deaths globally [1]. This virus, first detected in December 2019, is particularly deadly due to its strong adaptability of different clades and its fast mutations. In December of 2020, a more contagious mutant version of B.1.1.7 has emerged. Another variant strain, B.1.351 lineage appeared around the same time. In January 2021, a further variant of aggressive nature, P1 was reported [2]. As of now, it is still not fully confirmed if these new strains represent antibody resistant viruses or antibody neutralization escape mutants.

The, SARS-CoV-2 pathogen contains several structural proteins including Spike (S), Envelope (E), Membrane (M) and Nucleaocapsid (N) with negative sense RNA. From structural perspectives, the transmembrane spike protein consists of two subunits, S1 and S2; the N-terminal S1 binds to the angiotensin converting enzyme 2 (ACE2) receptors in the epithelial airway cells of human lungs, while the C-terminal S2 operates in membrane fusion. S1 is considered the receptor binding domain (RBD). A recently found mutation of the S protein, namely, the B.1.1.7 lineage or 20B/501Y.V./VOC 202012/01 mutations have shown 70% greater spreadability compared to the original virus. Along with several other mutations the Asn501Tyr (N501Y) mutation of S protein is observed in all three lineages mentioned above [2]. This 501 residue is located within the S1 subunit of the S protein.

In our earlier studies we have analyzed several immunologically relevant protein structures, specifically focusing on the correlation between their structural changes and their functions [3-7]. In the present paper, we examine the structures of the wild type (wt) S1 RBD

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2021 doi:10.20944/preprints202106.0238.v1

and its N501Y variant, which is a mutation common to all the three virulent strains mentioned above.

Materials and Methods

There are several structures of the spike RBDs in the protein databank. For our study, we have selected the E chain of 6M0J, the SARS-CoV-2 S protein's RBD as described by Lan et al [8]. 6M0J is the X ray crystal structure of SARS-CoV-2 RBD bound ACE2 receptor at 2.45 Å resolutions. We have chosen the E subunit of 6M0J as wt RBD, and used the subsequent mutant variant, N501Y for simulations. We performed two sets of simulations, one for the wt and another one for its mutant N501Y variant of RBD S1 protein.

Nanoscale Molecular Dynamics (NAMD), quickMD and Visual Molecular Dynamics (VMD) [9-11] programs were employed for time-based simulation studies of the wt and mutant spike S1 RBDs. Mutant variants were generated based on the wt proteins using the mutator plugin of VMD software. The "structure manipulation" feature was utilized to prepare the protein structures and to generate the psf structure files. The implicit solvation method was used in combination with the Generalized Born implicit solvent (GBIS) models [12], and the CHARMM36 force field. Energy minimization was performed in 2000 steps, and annealing was performed for 0.03 ns. The system temperature was gradually increased from 60 to 300K for 0.24 ns, while equilibration was carried out for 0.04 ns at 300 K. The final production run was performed for 30 ns at 300 K using the Langevin dynamics. No atoms were constrained for the production run. For all simulations the time step was set to 2 femtoseconds (fs) using the NVT ensemble. In total two sets of simulation data were obtained, wt strain and one single point

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2021 doi:10.20944/preprints202106.0238.v1

mutant variant. The proteins' structural illustrations were developed with Biovia's Discovery Studio Visualizer [13]

Results

Fig. 1A displays the structural representation of the E chain of 6M0J. PDB, the SARS-CoV-2 RBD, where the wt Asn501 residue is displayed in the stick mode. The RBD in Figure 1A is 333-526 amino acids (AAs) long. Fig. 1B displays the two-dimensional (2-D) diagram of wt Asn501 residue interactions with its neighbor E chain residues. Fig. 1C denotes the mutant variant with Tyr501 and Fig. 1D represents a 2-D diagram of structural interactions between the mutant residue Tyr501 and its neighbor residues within the SARS-Cov-2 S1 RBD. The initial structural analyses results from Fig. 1 reveal that the mutant residue within the SARS-CoV-2 RBD is slightly more active than its wt species, since Tyr501 initially has some supplementary interactions with its nearest neighbors. Nonetheless the amide pi stacked interaction and the steric clash in Fig.1D are not manifested in the MD simulation timeline as demonstrated in Fig. 2. The mutant residue has two intra-chain H bonds, seven vdW interactions, two covalent bonds (the peptide bonds connecting to residues 500 and 502), and one amide pi stacked interaction. The wt 501 residue has four intra-chain H bonds, and four vdW interactions. Similar to the mutant residue, two covalent bonds are present in the wt protein.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2021 doi:10.20944/preprints202106.0238.v1

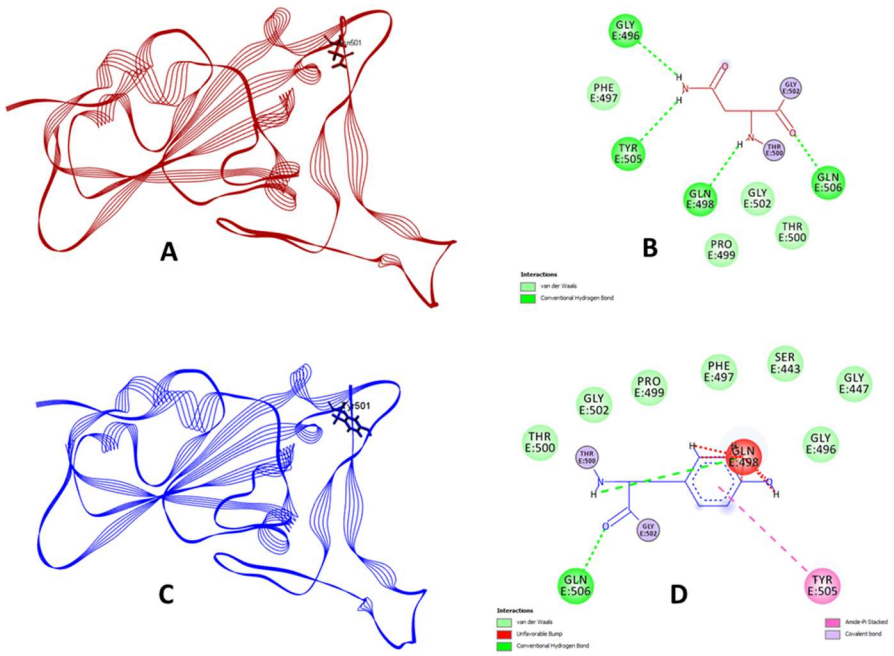

Fig. 1A. The structural presentation of 6M0J.PDB, E chain. The wt Asn501 residue is displayed in the stick mode **B.** The 2D diagram of the interactions between the wt Asn501 residue and its neighbor E chain residues. **C.** The single mutant variant of 6M0J.PDB E chain. The mutant Tyr501 residue is highlighted. **D.** A 2-D diagram of the initial interactions between the mutant Tyr501 residue and its neighbor E chain residues.

Although a small steric hindrance with Gln498 is observed for the mutant version, following the initial simulation protocols (minimization/ annealing and equilibration) the temporal profile has changed, and the steric hindrance is no longer detected. We have presented the preliminary interaction data of Tyr501 (Fig. 1D) recorded before the minimization process; notably, this is not a commonly recommended procedure, since a minimization step would frequently remove steric overlapping.

The 2-D interaction data of the 501 mutant residues from the initial, intermediate and final trajectories of the MD simulation timescale are displayed in Fig. 2. The corresponding PDB coordinates from the dcd file were generated using the stride command of the VMD window. wt Asn is hydrophilic in nature whereas mutant Tyr is hydrophobic. An aromatic residue usually

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2021 doi:10.20944/preprints202106.0238.v1

interacts with its neighbor residues to make the structure stronger and stable. Therefore, based on initial data of Fig. 1, it is reasonable to assume that, within the spike protein the mutant residue may be more interactive than its wild form.

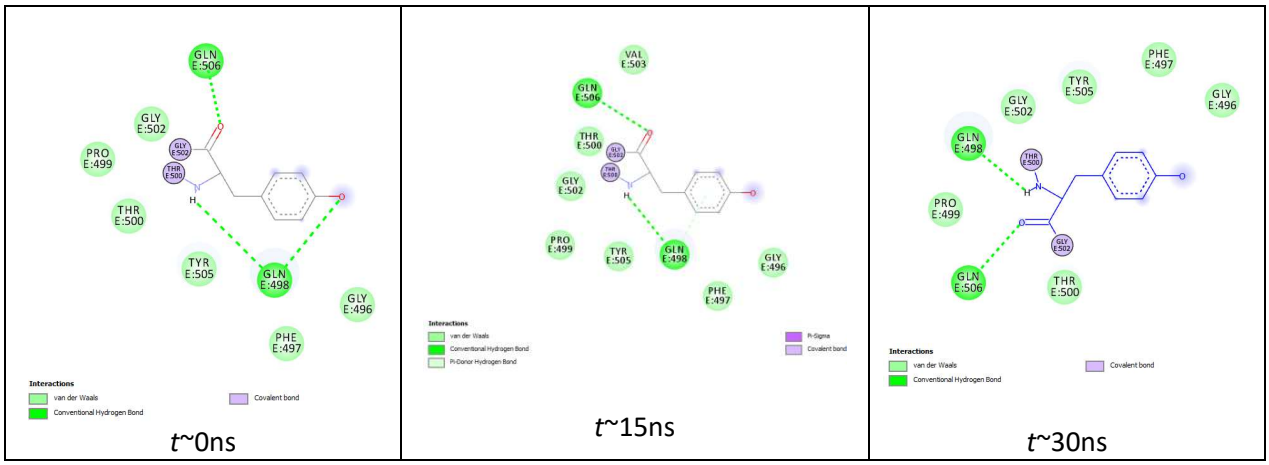

Fig. 2. The 2-D interactions of the mutant 501 residue from different timeframes of MD simulation.

Fig. 3A represents root mean square deviation (RMSD) plots for the wt and the N501Y mutant variant proteins as functions of time. The wt species displays slightly higher (all-atom) RMSD than the N501Y mutant species. Fig. 3B displays the root mean square fluctuation (RMSF) plots for the wt and mutant proteins as functions of their residue numbers. From Fig. 3B it is evident that the mutant species is more stable than its wt version. The RMSF values of the mutant (501Y) and wt (501N) residues are 1.837Å and 3.27 Å respectively. While wt Asn501 residue has high RMSF values, some of the nearby residues (around residue 484 and terminal residues around 520) also show higher fluctuation values than those of their mutant version. Overall, and throughout the residues, the mutant species maintains a relatively lower RMSF.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2021 doi:10.20944/preprints202106.0238.v1

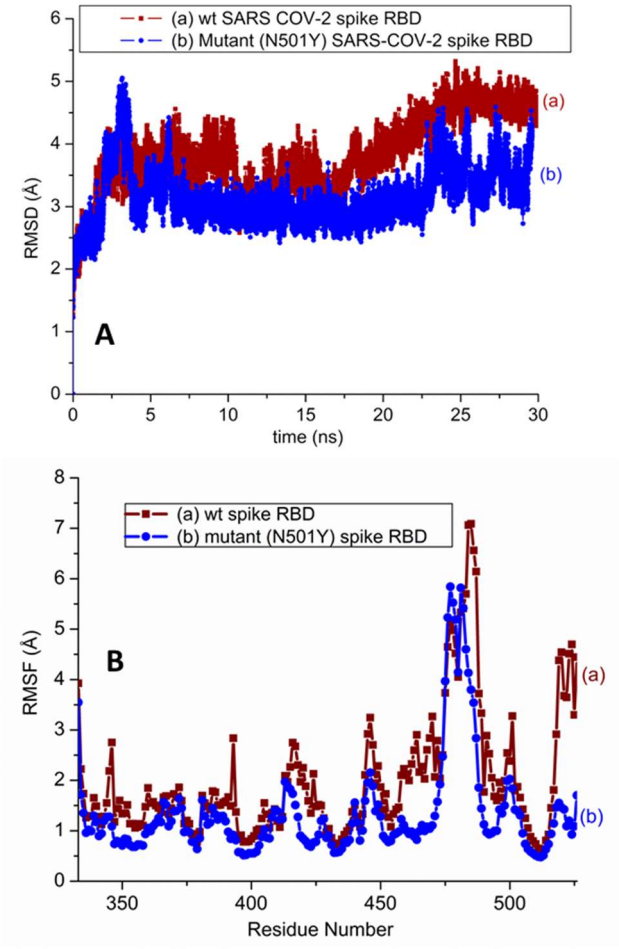

Fig. 3. **A** RMSD plots of wt SARS-CoV-2 spike RBD and its N501Y mutant variant with functions of time. **B** RMSF in wt and single mutant (N501Y) SARS-CoV-2 RBD with functions of residue number.

Fig. 4A-A' shows time dependent secondary structural changes of the wt SARS-CoV-2 RBD, together with those of its single mutant N501Y variant. Time-based structural changes of the wt Asn501 and mutant Tyr501 residues are plotted in Fig. 4B-B'. Interchangeable 3-10 helix to coils are observed in both wt and mutant residues in Fig. 4A-A'. The individual residue 501 shown in Fig. 4B-B' demonstrates that the mutant is marginally more stable, as accompanied by fewer structural changes during the middle and later temporal segments of the simulation. Major structural changes, observed around residue 480 in the wt species, led to the conclusion that N501Y mutation might play an active role in stabilizing the overall mutant RBD structure.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2021 doi:10.20944/preprints202106.0238.v1

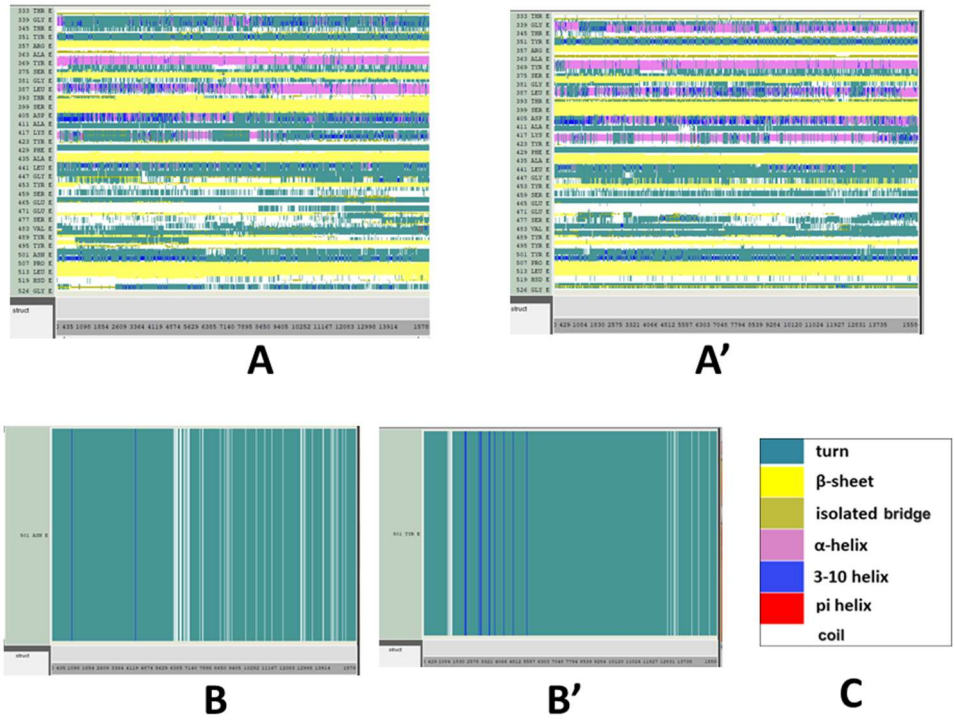

Fig. 4. A-A' Secondary structure changes of the wt SARS-CoV-2 RBD and single mutant N501Y RBD proteins with time. **B-B'** The time-based structure changes of the wt Asn501 and mutant Tyr501 residues within the wt and mutant variant of SARS-CoV-2 RBD are displayed. **C** The color code of this figure.

Discussions

Compared to other similar viruses like SARS-CoV found in 2003 and MERS-CoV in 2012, the SARS-CoV-2 spreads much more rapidly, and its newly emerged strains propagate even faster. Although numerous studies have been performed since the initial outbreaks of Covid 19, it is too early to draw any conclusion about its structural variants and their impacts, because this new virus can rapidly mutate and/or adapt.

Based on Fig. 1 we can assume that mutant N501Y is more interactive. However, during the time of MD simulations, changes may occur in the spike protein's interactions and/or its capability to bind with neighboring residues. Such changes could potentially alter, with the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2021 doi:10.20944/preprints202106.0238.v1

possibility of actually improving in extended runs. Since these "variants of concern" and their mutations have only been identified recently, and since experimental data on this subject is seldom, drawing a definite conclusion on the interactive nature of N501Y mutant is not straight forward. Moreover, N501Y within the S protein RBD shows detectable interactions. As the variants of SARS-CoV-2 transmit faster, stronger interactivity between N501Y and hACE2 receptor cannot be ruled out.

A recently published paper suggests the presence of efficient binding affinity between S1 RBD and ACE2, as determined by structural and functional analyses [14]. Another preprint, that also became available rather recently, indicates certain differences found between the receptor binding interfaces in N501Y mutant and wt protein ecto-domain by cryoEM [15]. An additional preprint uses computational methods to describe the structural dynamics of the mutant bound receptor [16]. While these aforesaid results are now available, the present work strictly focuses on the detailed structure based analyses of N501Y mutant residue and the S1 protein RBD. The exploratory computational results presented here also suggest that, the N501Y mutant variant exhibiting a detectable level of higher stability than its wt variant (Fig. 3). According to these results, one may infer that the mutant N501Y residue definitely plays a critical role in stabilizing the overall structure of spike protein's RBD. However considering the few experimental data on this subject, it is difficult to conclude with certainty whether or not N501Y itself can qualify as a stabilizing mutation. More direct experimental evidence will likely be necessary to further adequately address this question

Previous authors have suggested that spread-rates would be relatively higher in the D614G mutated SARS-CoV-2 variant as a result of epidemiological distribution [17]. It has also been suggested that, for mice, "favored" interactions would be observed between the mutant

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2021 doi:10.20944/preprints202106.0238.v1

RBD and the ACE2 receptor [18]. However, it is still not clear whether such strong bindings are related to their epidemiological distribution. Recently, a new variant, known as B.1.526 lineage has emerged with approximately 25% more spread-rate. This lineage has some shared mutations with the other previously detected strains [19]. Another new strain, B.1.427/B.1.429 lineages have been found, though, little information is available about these new lineages, published reports indicate that these lineages contain multiple S protein mutations, among which L452R is notable [20]. The B.1.427/B.1.429 strain is on the list of "variant of concern" by CDC as of March 16, 2021.

Conclusions

The results presented here predict greater stability for the mutant species. This inference is based on certain differences found between the two groups in their RMSD/RMSF calculations and secondary structure analyses. In a follow-up work of the present study, we will discuss the structural impact of these emerging variants on human receptor bindings and assessment of their time-based stability. It is still a subject of future investigation whether there are any other factors that promote the SARS-COV-2 variants' access to a host cell. It is also not clear at this time whether these new strains may emerge as antibody resistance viruses, or may escape from the current vaccinations, or weaken the vaccination efficacy; the answers may lie in the near future.

Conflict of interest statement

The author declares no financial conflict of interest.

Acknowledgments

The author acknowledges utilization of the following simulation and visualization software packages: 1) NAMD and 2) VMD: NAMD and VMD, developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois, Urbana-Champaign. 3) Discovery Studio Visualizer: Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, San Diego, CA: Dassault Systèmes (2015).

References

1. WHO, Weekly epidemiological update. https://www.who.int/publications/m/item/weekly-epidemiological-update---23-february-2021 (2021). Accessed 25 February 2021.
2. CDC, SARS-CoV-2 Variants (2021). https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html Accessed 1st March 2021.
3. Roy U (2020) Structural and Molecular Analyses of Functional Epitopes and Escape Mutants in Japanese Encephalitis Virus Envelope Protein Domain III. Immunol Res 68: 81-9. doi: 10.1007/s12026-020-09130-y
4. Roy U (2020) Insight into the Structures of Interleukin-18 Systems. Comput Biol Chem 88: 107353.
5. Roy U (2019) 3D Modeling of Tumor Necrosis Factor Receptor and Tumor Necrosis Factor-bound Receptor Systems. Mol Inform 38: e1800011. doi: 10.1002/minf.201800011
6. Roy U (2017) Structural modeling of tumor necrosis factor: A protein of immunological importance. Biotechnol Appl Biochem 64: 454-63. doi: 10.1002/bab.1523
7. Roy U (2016) Structural Characterizations of the Fas Receptor and the Fas-Associated Protein with Death Domain Interactions. Protein J 35: 51-60. doi: 10.1007/s10930-015-9646-6
8. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. (2020) Structure of the SARS-CoV-2 Spike Receptor-binding Domain Bound to the ACE2 Receptor. Nature 581: 215-20. doi: 10.1038/s41586-020-2180-5
9. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26: 1781-802. doi: 10.1002/jcc.20289
10. Ribeiro JV, Bernardi RC, Rudack T, Stone JE, Phillips JC, Freddolino PL, et al. (2016) QwikMD - Integrative Molecular Dynamics Toolkit for Novices and Experts. Sci Rep 6: 26536. doi: 10.1038/srep26536
11. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graph 14: 33-8. doi: DOI: 10.1016/0263-7855(96)00018-5
12. Tanner DE, Phillips JC, Schulten K (2012) GPU/CPU Algorithm for Generalized Born/Solvent-Accessible Surface Area Implicit Solvent Calculations. J Chem Theory Comput 8: 2521-30. doi: 10.1021/ct3003089
13. Dassault Systèmes BIOVIA. Discovery Studio Modeling Environment, San Diego, CA: Dassault Systèmes (2015).

14. Supasa P, Zhou D, Dejnirattisai W, Liu C, Mentzer AJ, Ginn HM, et al. (2021) Reduced Neutralization of SARS-CoV-2 B.1.1.7 Variant by Convalescent and Vaccine Sera. Cell 184: 2201-11.e7. doi: 10.1016/j.cell.2021.02.033
15. Zhu X, Mannar D, Srivastava SS, Berezuk AM, Demers J-P, Saville JW, et al. (2021) Cryo-EM Structure of the N501Y SARS-CoV-2 Spike Protein in Complex with a Potent Neutralizing Antibody. bioRxiv: 2021.01.11.426269. doi: 10.1101/2021.01.11.426269
16. Ahmed W, Philip AM, Biswas KH (2021) Stable Interaction Of The UK B.1.1.7 lineage SARS-CoV-2 S1 Spike N501Y Mutant With ACE2 Revealed By Molecular Dynamics Simulation. bioRxiv: doi: 10.1101/2021.01.07.425307
17. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. (2020) Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell 182: 812-27.e19. doi: 10.1016/j.cell.2020.06.043
18. Gu H, Chen Q, Yang G, He L, Fan H, Deng Y-Q, et al. (2020) Adaptation of SARS-CoV-2 in BALB/c Mice for Testing Vaccine Efficacy. Science 369: 1603. doi: 10.1126/science.abc4730
19. West AP, Barnes CO, Yang Z, Bjorkman PJ (2021) SARS-CoV-2 lineage B.1.526 emerging in the New York Rregion Detected by Software Utility Created to Query the Spike Mutational Landscape. bioRxiv: doi: 10.1101/2021.02.14.431043
20. Zhang W, Davis BD, Chen SS, Sincuir Martinez JM, Plummer JT, Vail E (2021) Emergence of a Novel SARS-CoV-2 Variant in Southern California. Jama: doi: 10.1001/jama.2021.1612