

Mutations in the SARS CoV-2 spike protein may cause functional changes in the protein quaternary structure

[SARS CoV-2 spike proteinindeki mutasyonlar, protein kuaterner yapısında fonksiyonel değişikliklere neden olabilir]

Abstract

Objectives: This study aimed to model the changes resulting from mutations in surface (spike/S) glycoproteins, which play a key role in the entry of the severe acute respiratory syndrome coronavirus-2 (SARS CoV-2) into host cells, in a protein quaternary structure and to evaluate their possible effects on the functional structure.

Methods: Genome sequence information of SARS CoV-2-infected patients located in Turkey was obtained from the GISAID EpiCoV database. Structural analysis of spike proteins was done using bioinformatics tools (MAFFT, PSIPRED, ProMod3, PyMoL and DynOmics).

Results: We identified 76 Thr>Ile mutations in the N-terminal domain; 468 Ile>Val mutations in the receptor binding site and 614 Asp>Gly, 679 Asn>Lys, 771 Ala>Val and 772 Val>Ile mutations in the S1 subunit. It has been observed that the mutations, except those of residues 771 and 772, may cause significant conformational, topological and electrostatic changes in a protein quaternary structure. It has been determined that the mutations in the receptor binding site transform the protein structure into a formation that can mask the binding site and affect receptor affinity.

Conclusions: It has been considered that SARS CoV-2 S glycoprotein mutations may cause changes in a protein functional structure that can affect the severity of disease.

Keywords: COVID19; mutation; quaternary structure; S glycoprotein; SARS CoV-2.

Öz

Amaç: Bu çalışmada, şiddetli akut solunum sendromu koronavirüs-2’nin (SARS CoV-2) konakçı hücrelere girişinde anahtar rol oynayan yüzey (spike/S) glikoproteini, mutasyonların proteinin kuaterner ve fonksiyonel yapısı üzerindeki olası etkilerini değerlendirmek amacı ile modellendi.

Gereç ve yöntem: Türkiye’de yerleşik SARS CoV-2 ile enfekte hastaların genom sekans bilgileri GISAID EpiCoV veri tabanından elde edildi. Spike proteinlerin yapışal analizi biyoinformatik araçlar (MAFFT, PSIPRED, ProMod3, PyMoL ve DynOmics) kullanılarak yapıldı.

Tartışma: N-terminal bölgesinde 76 Thr>Ile, reseptör bağlanma bölgesinde 468 Ile>Val, S1 alt biriminde 614 Asp>Gly, 679 Asn>Lys, 771 Ala>Val ve 772 Val>Ile mutasyonlarını tanımladık. 771 ve 772 rezidüleri dışındaki mutasyonların, protein kuaterner yapısında önemli konformeänder, topolojik ve elektrostatik değişikliklere neden olabileceği gözlenmiştir. Reseptör bağlanma bölgesindeki mutasyonların, protein yapısı, bağlanma sahasını maskeleyebilecek ve reseptör afinitesini etkileyebilecek bir formasyona dönüştürüldüğü belirlenmiştir.

Sonuç: SARS CoV-2 S glikoprotein mutasyonlarının, protein fonksiyonel yapısında hastalık şiddetini etkileyebilecek değişikliklere neden olabileceği düşünülmektedir.
In our study, the changes in the protein structure of mutations in the sequence encoding S glycoprotein from the Turkish SARS CoV-2 isolates, which is a significant drug target, were modeled. The results that may be ascertained by mutations in the functional structure of the protein have been evaluated in terms of changes in virulence effect and risk of drug resistance.

Materials and methods

Sequence data and mutation analysis

Nucleotide sequence information of 61 isolates from Turkey were obtained from GISAID EpiCoV database [15]. S glycoprotein nucleotide sequences were transformed into protein sequences with MegaX software [16]. Reference spike glycoprotein accession code is YP_009724390 [17]. Protein sequence information of 61 isolates were aligned with the MAFFT (v7.633) multiple sequence alignment program FFT-NS-2 algorithm [18]. The scoring matrix BLOSUM 80 was chosen for the amino acid sequences [19]. Gap opening penalty was used 1.53. The mutated residues were analyzed with MegaX bioinformatic workbench [16].

Homology modeling and structural analysis

Three-dimensional model of wild and mutant spike proteins was generated by the method of homology modeling using Swiss-Model [20], 6vxx (protein data bank code) was selected as template. Models are built based on the target-template alignment using ProMod3 [21]. MolProbity and QMEAN (Qualitative Model Energy Analysis) were used for structural validation and model quality (covalent geometry, torsion angle, optimized hydrogen placement and whole atom contact analysis) of wild and mutant spike proteins [22, 23]. Physiochemical properties of wild and mutant spike proteins were estimated by ProtParam tool from ExPASy portal [24]. Secondary structure component (random coils, beta strands alpha helices) of spike protein were identified by using PSIPRED web server [25]. Protein structure imaging and comparison was performed with PyMol (Ver2.3.4 Schrödinger).

TM-score and RMSD (root mean square deviation) values were calculated with i-Tasser web service to detect topological and structural differences between wild and mutant S proteins [26]. The changes caused by mutations in the dynamic structure of proteins were simulated in elastic network models (ENM). After mutation, the movement and dynamic changes of molecules in the S glycoprotein structure were analyzed by DynOmics. By comparing the experimentally crystallographic data with the wild and mutant model data obtained in this study, both the collective motion of the proteins and the mean square fluctuation of each residue were calculated by Gaussian network model (GNM) and anisotropic network model (ANM) [27].

Results

A total of six mutations have been determined in the S glycoproteins from the Turkish SARS CoV-2 isolates. We
identified 76 Thr>Ile mutations in the N-terminal domain; 468 Ile>Val mutations in the receptor binding site and 614 Asp>Gly, 679 Asn>Lys, 771 Ala>Val and 772 Val>Ile mutations in the S1 subunit. Mutant S glycoprotein sequence information was established in the MegaX software using the identified mutation data. Secondary structure prediction for wild type and mutant S glycoprotein was performed using the PSIPRED workbench. Furthermore, 65 strands, 20 helices and 84 coils were identified in the wild type glycoproteins, whereas 62 strands, 25 helices and 87 coils were identified in the mutant glycoproteins. The wild type and mutant protein models for SARS CoV-2 S glycoprotein were created using the ProMod3 (ver3.0.0). The coverage value and sequence identity ratio in the wild type structure, which is used for the homotrimeric SARS CoV-2 S glycoprotein model, were 0.94 and 99.50%, respectively. The coverage value and sequence identity ratio in the mutant structure were 0.94 and 99.08%, respectively. The quality of the model was evaluated using the QMEAN and MolProbity software. The MolProbity scores in the mutant and wild type structures were 1.39 and 1.42, respectively. The MolProbity score of the model used as a pattern was 2.8. The fact that the MolProbity score in the model is lower than that in the model used as a pattern shows that it is better than the average structures at this resolution [28]. For the wild type protein sequences, the GMQE score was 0.75, the QMEAN score was −2.07, the clash score was 1.57, the Ramachandran favored score was 91.5% and the Ramachandran outliers score was 1.82%. For the mutant protein sequences, the clash score was 1.34, the Ramachandran Favored score was 91.69%, the Ramachandran Outliers was 1.52%, the GMQE score was 0.74 and the QMEAN score was −2.13. Ramachandran plots indicate that the generated protein models have acceptable polypeptide backbone phi (φ) and psi (ψ) torsion angles for alpha helix and beta strand regions [29]. The quality scores show that the models are obtained within appropriate limits in terms of model quality.

It was found that Thr>Ile mutation in S glycoprotein residue 76 may result in changes in the protein conformation and tertiary structure (Figure 1). It was observed that 76 Thr>Ile mutations caused changes in the region between 68 Ile and 166 Cys. Furthermore, residues 154–156 (coils in the wild type) transformed into a shorter strand structure in the mutant type (Figure 2). We observed that Ile>Val mutation in the area where SARS CoV-2 S glycoprotein interacts with hACE2 receptor can result in changes in the tertiary structure between residues 468–490, which exhibit coil structural properties (Figure 3). In the wild type, the distance between 456 Phe and 482 Gly was 23.9 angstroms (Å), whereas in the mutant type, this distance was found to be 17.2 Å, which is closer to the major protein backbone by 6.7 Å. In the same region, the distance

Figure 1: Cartoon representation of the conformational changes in the tertiary structure caused by the 76Thr>Ile mutation in the SARS CoV-2 spike protein. Blue color = wild spike protein, yellow color = mutant spike protein, red color = mutant residue, magenta color = residue considered in distance calculations, a) Representation of the area with conformational change in wild protein, b) Representation of the region between Ile68-Cys166 residues with conformational changes in the mutant protein, c) Aligned wild-mutant protein structures.

Figure 2: Cartoon representation of tertiary changes between 154 and 156 residues caused by the 76Thr>Ile mutation in the SARS CoV-2 spike glycoprotein. Blue color = wild spike protein, yellow color = mutant spike protein, red color = mutant residue, magenta color = 154, 155 and 156 residues.
between 476 Gly and 490 Phe in the wild and mutant types were found to be 4.4 Å and 13.8 Å, respectively (Figure 4). It has been observed that 614 Asp > Gly mutations in S glycoprotein may cause conformational changes between residues 620 and 641 (Figure 5). It was observed that the structure between residues 623 and 638 with a hairpin appearance in the wild type (Figure 5A) varied in the mutant protein (Figure 5B). The distance between residues 623 and 638 in the wild type and mutant proteins was found to be 8 and 10.9 Å, respectively. It has been determined that S glycoprotein 771 Ala>Val and 772 Val>Ile mutations did not exhibit conformational deterioration in these positions,
which may cause a change in the functional structure (Figure 6). It was observed that the mutations caused changes in the electrostatic structure of S glycoprotein (Figure 7). Between the wild and mutant S proteins, the RMSD and TM-score value were found to be 0.62 Å and 0.975, respectively. Gaussian network model (GNM) and anisotropic network model (ANM) were calculated with DynOomics to analyze the changes in protein dynamic structure (Figure 9). The B-factor correlation between the experimental crystallographic structure and the wild protein was 0.67, and 0.62 with the mutant protein.

Discussion

Higher infection and mortality rates associated with SARS CoV-2 compared with other pathogen coronaviruses prioritise the clarification of its molecular organisation and pathogenesis.

The present study revealed that the mutation in S glycoproteins in the Turkish SARS CoV-2 isolates may cause changes in the protein quaternary structure, topology and conformation (RMSD 0.62 and TM-score 0.975). S glycoprotein is known to play an important role in the entry of the virus into target cells, cellular transmission and infection severity. The virus, which binds to the host cell receptors with the S1 subunit of S glycoproteins, initiates the membrane fusion process required for the passage of the virion into the host cell via S2 subunit activity [30–32]. Two conformational conditions are observed in the S glycoprotein receptor binding site. One of them is down-formation, in which the receptor binding site is hidden/masked, whereas the other is up-formation, in which the receptor binding site is accessible and exhibits a less stable structure [33–35]. In this study, it was observed that the region that exhibits an up-formation-like structure turned into a down-formation-like structure because of the changes caused by 468 Ile>Val mutation between residues 468 and 490 in the region in which the S glycoprotein is bound to hACE2 (Figure 8). The region with formational change (between residues 468 and 490) covers 8 of 14 residues in which the S glycoprotein interacts with the hACE2 receptor [36]. It is known that the mutations in the receptor binding site have an impact on the S glycoprotein-receptor affinity [37–39]. The data obtained in this study point to the structure evolving into a down-formation-like structure. This can result in a decrease in the S glycoprotein receptor affinity, viral infection severity and transmission. Moreover, the data obtained from the study may be associated with the fact that the basic reproduction
number (R0) in Turkey is decreased from double digit numbers to less than -1- [40].

In the present study, it was observed that 614 Asp>Gly and 76 Thr>Ile mutations may cause conformational changes in the region between residues 620 and 641 and between 68 Ile and 166 Cys residues in the S1 subunit of S glycoprotein, respectively. It was observed that residues 154–156 (coils in the wild type) transformed into a shorter strand structure in the mutant type. Following the binding of the S1 subunit to the host cell receptor, large-scale conformational changes occur in its secondary, tertiary and quaternary structures. These changes are intended to ensure the transition of the S2 subunit to a stable conformation following the membrane fusion [34, 35]. The conformational changes following 614 Asp>Gly and 76 Thr>Ile mutations are considered to affect the mechanism of such multilayer exchange and the membrane fusion process in the S1 subunit.

The S glycoprotein dynamic network models indicated a decrease in molecular fluctuation in the mutant protein compared to the wild protein (Figure 9). It is known that the molecular fluctuation frequency is closely related to the functional properties of biomolecules [41, 42].

Theoretical fluctuation data of static molecular structures contribute to high accuracy results in the analysis of collective movements and dynamic changes of large biomolecules [43]. The reduction in molecular fluctuation
in the receptor binding region of the mutant S glycoprotein can be evaluated from two aspects. Firstly, decreased molecular motion can result in a decrease in ACE2 affinity. Secondly, reduced molecular motion can enable the S-ACE2 complex to transform into a more stable and rigid structure after binding. After mutation, considering the transformation of the receptor binding site into an upformation-like structure, molecular stability and rigid structure may adversely affect the effective coupling of S glycoprotein and ACE2 and the fusion process.

In conclusion, the mutations in the Turkish SARS CoV-2 isolates may result in changes in the conformational and topological structure of S glycoprotein. It is considered that conformational changes in the receptor binding site may result in a decrease in the receptor affinity. The data obtained in our study provide a structural framework to understand the possible effects of changes caused by S glycoprotein mutations in a protein quaternary structure on the functional properties.

Acknowledgement: We gratefully acknowledge the authors, originating and submitting laboratories of the sequences from GISAID’s EpiFlu™ Database on which this research is based. The list is detailed below. Shaikh Terkis Islam Pavel, Hazel Yetiskin, Günsu Aydin, Can Holayvkin, Muhammet Ali Uygut, Zehra Dursun, Ilhami Celik, Alper Iseri, Aykut Ozlarendeli from Vaccine Research, Development and Application Center, Erciyes University (submitting laboratory Gen Era Diagnostics Inc.), Ilker Karacan, Tugba Kizilboga Akgün, Bugra Agaoglu, Gizem Alkurt, Jale Yildiz, Betsi Köse, Elifnaz Celik, Mehtap Aydin, Levent Doganay, Gizem Dinler Doganay from Genomic Laboratory (GLAB) (Conjoint lab of Health Directorate of Istanbul and Istanbul Technical University), Fatma Bayrakdar, Ayşe Başak Alttaş, Yasemin Coşgun, Gülay Korukluoğlu, Selçuk Kılıç, Tulin Demir, Süleyman Yaşın from Ministry of Health Turkey.

Research funding: None declared.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of Interest: The author has no conflict of interest.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China 2019. N Engl J Med 2020;382:727–33.
2. Worldometer. Coronavirus cases. https://www.worldometers.info/coronavirus/ [accessed June 2020].
23. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 2011;27:343–50.

24. Wilkins MR, Gasteiger E, Bairoch A, Hoogland C, Gattiker A, Appel RD. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 1999;112:531–52.

25. Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. Scalable web services for the PSIPRED protein analysis workbench. Nucleic Acids Res 2013;41:349–57.

26. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins 2004;57:702–10.

27. Li H, Chang YY, Lee JY, Bahar I, Yang LW. DynOmicS: dynamics of structural proteome and beyond. Nucleic Acids Res 2017;45:374–80.

28. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007;35:375–83.

29. Lovell SC, Davis IW, Arendall WB, De Bakker PI, Word JM, Prisant MG, et al. Structure validation by Cα geometry: φ, ψ and Cβ deviation. Proteins Struct Funct Genet 2003;50:437–50.

30. Bosch BJ, van der Zee R, de Haan CAM, Rottier PJ. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J Virol 2003;77:8801–11.

31. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 2020;181:271–80.

32. Li F. Structure, function, and evolution of coronavirus spike proteins. Annu Rev Virol 2016;3:237–61.

33. Gui M, Song W, Zhou H, Xu J, Chen S, Xiang Y, et al. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. Cell Res 2017;27:119–29.

34. Walls AC, Tortorici MA, Snijder J, Xiong X, Bosch BJ, Rey FA, et al. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. Natl Acad Sci 2017;114:11157–62.

35. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 2020;367:1260–3.

36. Li F, Li W, Farzan M, Harrison SC. Structural biology: structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 2005;309:1864–8.

37. Jia Y, Shen G, Zhang Y, Huang KS, Ho HY, Hor WS, et al. Analysis of the mutation dynamics of SARS-CoV-2 reveals the spread history and emergence of RBD mutant with lower ACE2 binding affinity. BioRxiv 2020;1:1–12.

38. Kim YS, Algerim A, Park U, Kim Y, Rhee JY, Choi JP, et al. Sequential emergence and wide spread of neutralization escape middle east respiratory syndrome coronavirus mutants South Korea 2015. Emerg Infect Dis 2019;25:1161–8.

39. Ou J, Zhou Z, Dai R, Zhang J, Lan W, Zhao S, et al. Emergence of RBD mutations from circulating SARS-CoV-2 strains with enhanced structural stability and higher human ACE2 receptor affinity of the spike protein. BioRxiv 2020;1:1–30.

40. Göksel T, Çınar Y. COVID-19 salgınında Türkiye için normalleşme döneminde geçiş izlenmesi: Türkiye COVID-19 gelişim barometresi. https://www.ankara.edu.tr/wp-content/uploads/sites/6/2020/04/Ankara-Universitesi-COVID-19-Barometre-30-04-2020.pdf [accessed May 2020].

41. Bahar I, Lezon TR, Yang LW, Eyal E. Global dynamics of proteins: bridging between structure and function. Annu Rev Biophys 2010;39:23–42.

42. Haliloglu T, Bahar I. Adaptability of protein structures to enable functional interactions and evolutionary implications. Curr Opin Struct Biol 2015;35:17–23.

43. Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. Protein Eng Des Sel 2001;14:1–6.