


Bioinformatics Analysis Unveils Certain Mutations Implicated in Spike Structure Damage and Ligand-Binding Site of Severe Acute Respiratory Syndrome Coronavirus 2

Bioinformatics and Biology Insights
Volume 15: 1–10
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11779322211018200


Emre Aktas 

Faculty of Art and Science, Department of Molecular Biology and Genetics, Bioinformatic Section, Afyon Kocatepe University, Afyonkarahisar, Turkey.

ABSTRACT: There are certain mutations related to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In addition to these known mutations, other new mutations have been found across regions in this study. Based on the results, in which 4,326 SARS-CoV-2 whole sequences were used, some mutations are found to be peculiar with certain regions, while some other mutations are found in all regions. In Asia, mutations (3 different mutations in QLA46612 isolated from South Korea) were found in the same sequence. Although huge number of mutations are detected (more than 70 in Asia) by regions, according to bioinformatics tools, some of them which are G75V (isolated from North America), T95I (isolated from South Korea), G143V (isolated from North America), M177I (isolated from Asia), L293M (isolated from Asia), P295H (isolated from Asia), T393P (isolated from Europe), P507S (isolated from Asia), and D614G (isolated from all regions) (These color used only make correct) predicted a damage to spike protein structure. Furthermore, this study also aimed to reveal how binding sites of ligands change if the spike protein structure is damaged, and whether more than one mutation affects ligand binding. Mutations that were predicted to damage the structure did not affect the ligand-binding sites, whereas ligands' binding sites were affected in those with multiple mutations. It is thought that this study will give a different perspective to both the vaccine SARS-CoV studies and the change in the structure of the spike protein belonging to this virus against mutations.

KEYWORDS: SARS-CoV-2, mutations by regions, spike structure, ligand

RECEIVED: September 9, 2020. **ACCEPTED:** April 27, 2021.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Emre Aktas, Faculty of Art and Science, Department of Molecular Biology and Genetics, Bioinformatic Section, Afyon Kocatepe University, ANS Campus, Afyonkarahisar 03200, Turkey. Email: eaktas@aku.edu.tr

Introduction

In 2 decades, mankind has accosted with at least one lethal outbreak from the betacoronaviruses.¹ The first was severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002, which infected more than 8,000 people, with nearly 800 deaths.² In 2012, Middle East Respiratory Syndrome (MERS)-CoV resulted in 2,294 cases.³ The last one is that severe acute respiratory syndrome–coronavirus 2 causes the contagious disease COVID-19 (coronavirus disease 2019), which was first reported in Wuhan, in December 2019. Despite wide efforts to control the disease, COVID-19 has now spread to more than 100 countries and result in a worldwide pandemic.⁴ Until now, recorded cases are more than 119,267,000, and the number of deaths has exceeded 2,647,000 (<https://covid19.who.int/>, March 15, 2020). Information about viral mutations for COVID-19 will give important insights into assessing viral drug resistance, immune escape, and pathogenesis-related mechanisms.⁵ Moreover, this information may play a vital role in the design of new vaccines, antiviral drugs, and diagnostic assays. However, mutagenic process is complex, and many factors are implicated in this process such as replication of nucleic acids influenced by few or no proofreading capability and/or postreplicative nucleic acid repair, host enzymes, spontaneous nucleic acid damages due to physical and chemical mutagens, recombination events, and other particular genetic elements.⁵

Some mutations which belong to different proteins of SARS-CoV were already found.⁶ Along these mutations,⁶ some combined factors are thought to make COVID-19 dangerous. One of these factors may be that humanity has no direct immunological experience with SARS-CoV-2, making humans prone to the infection.⁷ COVID-19, which has a rapid global spread, may provide the virus with a higher chance for natural selection of mutations. As with the case of influenza (where mutations slowly accumulate in the hemagglutinin protein), there is a complex interplay between mutations that can confer immune resistance to the virus and the fitness landscape of the particular variant in which they arise. Severe acute respiratory syndrome–coronavirus 2, which has a remarkably high mutation rate and many characterized variations, has been shown to have undergone certain mutations in its structural and nonstructural proteins, within several months of its global spread.^{8–11} Virus-related mutations are a concern because mutations can both affect the transmission rate of the virus and affect possible vaccine studies, and mutations can belong to regions such as Europe and North America.^{5,12} For example, SARS-CoV-2 variants with G614 in the S protein have replaced the original D614 variants and have become the dominant form circulating globally.⁹ Like some aforementioned important studies,^{8–12} this study focused on mutations and some of their characteristics to affect on spike structure damage. This study is focused on



determining mutations that occurred based on regions and evaluating whether these new mutations affect the structure of spike proteins. In addition, this study also predicts how mutations affect the ligand-binding site of SARS-CoV-2. Characterization of these detected variants may give new insight for designing new candidate vaccine studies, treatments, and diagnostic approaches SARS-CoV-2.

Materials and Methods

Data set construction

NCBI Virus website (www.ncbi.nlm.nih.gov/labs/virus) was used for obtain 4,526 whole sequences surface glycoprotein of COVID-19 (taxid: 2697049) isolated from humans, and NCBI Virus website has been adjusted for the respective based on their geographic region.

Amino acid substitution analysis

The data set that is downloaded from the NCBI Virus website was aligned using MEGAX (align by MUSCLE) program. Geographical regions were evaluated separately, and amino acid substitutions were found manually.¹³

Predicted protein structure

Phyre2 (www.sbg.bio.ic.ac.uk/phyre2/html), a suite of tools available on the web, was used to predict and analyze the protein structure, function, and mutations. All the predicted structures were obtained using this tool one by one.¹⁴

Predicted structure models

MISSENSE3D online tool was used to predict the structure of missense variants in relation to normal structure. All the results (from Phyre2) obtained were analyzed one by one.¹⁵

Predicted phylogenetic clusters and genotypes

Genome Detective Coronavirus Typing Tool was used for the prediction of phylogenetic clusters of virus. This application identifies the phylogenetic clusters and genotypes from assembled genomes in amino acid FASTA format.¹⁶

Prediction of ligand site

3DLigandSite online method was used for an automated prediction of the ligand-binding sites.¹⁷

Results and Discussion

Finding mutations based on regions and discussion

Eighty four whole spike protein sequences isolated from Africa are used, and the most common mutations are found to be Q667H (5 mutations), D614G (3 mutations), R408I (2 mutations), and others (1 mutations), which was found containing 8

Table 1. Eight different types of mutation results based on whole sequences of surface glycoprotein based on Africa region (data taken from NCBI Virus website and results obtained using MEGAX manually).

ACCESS NUMBER	THE MUTATION
QKR84285	S12F
QJX45356	T29I
QJX45344	A288T
QJX45344	Q314R
QKT21014	R408I
QKR84321	A570S
QJX45321	D614G
QKW95051	S640A

different mutations (Table 1). One of these sequences, QJX45344, which was isolated from Tunisia, has 2 mutations which are A288T and Q314R (Table 1). This result may increase the likelihood of being infected with 2 different mutations at the same time. The predicted structure damage for both did not influence the structure damage to the spike proteins, based on the result of the MISSENSE3D online tool.¹⁵ In this area, only D614G mutation predicted a damaged structure. For others, there was no prediction of structure damage (Table 1). But it is known that compared with the D614 variant, higher viral loads were found in patients infected with the G614 variant, but clinical data suggested no significant link between the D614G alteration and disease severity, and also suggesting the alteration may have increased the infectivity of SARS-CoV-2.¹⁸ According to this interpretation, it may not be possible to draw a clear conclusion about how other found mutations will affect this epidemic process in terms of severity. Because, even D614G mutation predicted a damaged structure (Table 1), no significant link between the D614G alteration and disease severity was found.¹⁸

347 whole sequences of the spike protein isolated from Europe were used to predict possible mutations on SARS-CoV-2 surface proteins. The most common mutations were found to be D614G (39 mutations), H49Y (3 mutations), Y453F (8 mutations), G261D (6 mutations), A845S (4 mutations), T676I (2 mutations), S254F (2 mutations), and I197V (2 mutations), respectively, while the others that have only one mutation are shown on Table 2. According to results of Table 2, the same mutation can occur at different positions. For instance, Alanine can change to Serine at 2 different positions such as A845S and A892S (Table 2). In addition, Threonine (T) can change to Isoleucine at 3 different positions T22I, T240I, and T676I. Only 2 (T393P and D614G) of these mutant sequence predicted a structure damage (Table 2). Korber et al¹⁸ suggested that the alteration (D614G) may have increased the infectivity of SARS-CoV-2, and higher viral loads were found

Table 2. Thirty five different types of mutations (obtained using MEGAX program) results only for Europe region.

ACCESS NUMBER	THE MUTATION	ACCESS NUMBER	THE MUTATION
QKM76366	T22I	QJS39507	N501T
QJT72134	L5F	QJT73034	T553N
QHU79173	H49Y	QJC19455	K558R
QJD23141	Q115R	QJT72470	T572I
QJT72086	M153I	QJT72278	L611F
QJT72350	L176I	QKM76846	D614G
QJS53410	N188D	QJT72614	T676I
QJS53494	I197V	QJZ28203	M740I
QKJ68364	V213L	QJS54286	G769V
QJT73010	T240I	QJS53386	Y789D
QKM76906	S254F	QIC53204	F797C
QJS39543	G261D	QJT72710	A845S
QJS39627	V367F	QJS53578	A892S
QJT72806	V382E	QJT72242	A1020V
QJT72386	C379F	QJS53506	H1101Y
QJS54106	T393P	QJS53398	V1122L
QJS39603	Y453F	QJZ28203	D1260N
QJS39567	F486L		

in patients infected with the G614 variant, and Toyoshima et al¹⁹ said that this variant has also higher fatality rate., Like this suggestion, when T393P occurs at the spike protein, it may affect on both infectivity of SARS-CoV-2 and higher viral loads. A study showed that 4 mutations (at the nucleotide level) are common in the SARS-CoV-2 European isolates genomes, where the severity of the infection is mostly more intense than in the other geographical regions.⁶ T393P mutation, which the other predicted a structure damage, is found only in Europe (Table 2). It is conceivable that this mutation is more likely to be found in Europe. Although F486L and N501T is not predicted that does not damage the structure of the spike protein (Table 2), it has been stated that the N501T and F486L mutations affect the stability of the spike protein.²⁰ It is known that stability is a fundamental property affecting function, activity, and regulation of biomolecules, and stability also is very important for vaccine study.^{12,21}

Based on 760 whole sequences from Oceania and South America, the most common mutations are found to be G1124V (25 mutations) and D614G (20 mutations), while other different mutations tend to increase, such as S50L (10 mutations), A262T (11 mutations), L5F (5 mutations), D138H (3 mutations), S221L (3 mutations), G485R (3 mutations) (Table 3). As in Europe and Africa, there are similar mutations that occurred at

different positions, such as T29I, T76I, and T791I. Besides, QKV37632 sample has 2 mutations which are T29I and S704 (Table 3). As seen in sequences from all regions, the D614G mutation predicted structure damage for this region. Even the D936Y mutation did not predict the damage to the spike protein structure, however, this mutation is predicted to reduce the stability of spike proteins.⁶ Stability is already mentioned that it is quite important for function and vaccine study.^{12,21}

The maximum D614G mutation rates are found in North America in 2,700 complete sequences of only spike proteins. Based on results (some are shown on Table 4), more than 255 mutations for D614G was determined. In the sample sequences isolated for this study, some other mutations were found, such as L5F (19 mutations), D138H (18 mutations), E554D (13 mutations), and P631L (10 mutations). Like Tables 2 and 3, two different mutations were found at the same position. For instance, QKG89654 (A845D) and QKV35819 (A845V) have different mutations at the same position. Other examples are QKG91034 (Q836P) and QKG81751 (Q836L) (Table 4). These 2 examples may be a proof that some positions are more vulnerable to mutations. For both, there was no predicted structure damage according to MISSENSE3D online tool.¹⁵ As in Europe and Africa, Threonine (T) changed to Isoleucine (I) at 3 different positions; however, there was no predicted

Table 3. Thirty one different types of mutations. 760 whole sequences from Oceania and South America (20 of them belong to South America) of spike protein were used.

ACCESS NUMBER	THE MUTATION	ACCESS NUMBER	THE MUTATION
QJR90681	L5F	QHR84449	D614G
QKV37632	T29I	QJR87501	P621S
QKV38004	H49Y	QJR93417	A626V
QJR87081	S50L	QKR84925	Q675H
QJR88113	T76I	QJR87477	Q701H
QJR92637	I128F	QKV37632	S704L
QJR93237	D138H	QJR85593	M731I
QJR93801	L176F	QJR88113	T791I
QJR89217	S221L	QKV38208	P812S
QHR84449	S247R	QJR87261	A846V
QJR87129	W258L	QJR93861	D936Y
QJR87465	A262T	QJR88221	P1079S
QJR86937	I468T	QKV37548	G1124V
QKR86245	G485R	QJR85701	D1163G
QKR85081	H519Q	QJR85833	D1260N
QJR85965	P561L		

structure damage (Table 4). The interaction of the mutations in the spike protein with the antibody was examined, and as a result, it was determined that possible mutations affect the functions of the antibodies.²² Not all mutations might have a negative effect on the spike protein, some are known to affect them negatively. Among the results in this table (Table 4), it is possible that the spike protein will be adversely affected.^{12,21,22} The presence of some mutations in a particular region can be mentioned as a regional effect in the formation of mutations (Tables 1–5).

Asia is the region where most mutation types were seen (Table 5). As seen in all regions, D614G was the most variant for all regions (Tables 1 to 5); 240 isolated samples had this variant in Asia (Table 5). In addition, mutations more than 3 were found, such as L54F (40 mutations), R78M (15 mutations), V367F (5 mutations), A829T (10 mutations), H1083Q (4 mutations), T791I (12 mutations), Q677H (4 mutations), E583D (15 mutations), T572I (10 mutations), and L8V (4 mutations). Moreover, some other regions have 1 or 2 mutations. QLA46612 isolated from South Korea has 4 different mutations L54F, F86S, T95I, and QKY60177, whereas India has 4 mutations Q506H, P507S, Y508N, and K786N, respectively (Table 5). None of these mutations predicted structure damage according to MISSENSE3D online tool.¹⁵ Also, some mutations were found in more than one. As in Table 5, Threonine (T) changes to Isoleucine (I) at different positions

such as T22I, T76I, T95I, T572I, T791I, and T827I; whereas Glutamine (Q) changed to Histidine (H) (QLA10116 and QKW92184). In this region, some mutations found include C1243F, Q1201K, K1191N, D1153Y, P507S, among others. Another example where 3 mutations occurred at the same isolated sequence is QKY60177, which has Q506H, Y508N, and P507S mutations. Like QLA46612 isolated from South Korea, QJD23249 isolated from Wilayah Persekutuan Malaysia has 4 mutations which includes L293M, D294I, P295H, and H519Q. Interestingly, QJD23249 isolated sample' mutations are predicted no structure damages (Table 5). There are situations that are anticipated to increase the periodicity of encounters between SARS-CoV-2 and antibodies that could effect the dawn of antibody Millions of individuals have already been infected with SARS-CoV-2 and among them, neutralizing antibody titers are extremely changeable.^{23,24} In addition, it may be predicted that the effects of mutations may worsen this situation. It will be important to identify mutations and monitor their prevalence in a way that is analogous to antiviral and antibiotic resistance monitoring.²³ It has been stated that environmental factors also affect the spread of SARS-CoV. In the same study, it was determined that both the temperature and the environment affected the spread of the virus.²⁵ Viral factors might contribute to transmissibility too. For example, a distinct rise in the prevalence of SARS-CoV-2 bearing a D614G mutation has been noted over time.¹⁸ Whether this mutation

Table 4. Fifty two different mutations based on whole sequences (2,500 sequences) of spike proteins in North America.

ACCESS NUMBER	THE MUTATION	ACCESS NUMBER	THE MUTATION
KKG81847	L5F	KKG90866	A570V
KKG81475	S12C	KKE61636	D614G
KKG90662	Q14H	KKG81571	P631L
KKV07471	T29I	KKG89666	A647V
KKG90530	F32L	QLC93320	Q677R
KKV38905	S50L	KKV39263	T732A
KKG89918	H69Y	KKV35279	N751D
QLA47679	G75V	KKG90590	A783S
KKG27877	T95I	KKG90614	P812S
KKW89191	E132D	KKG91034	Q836P
KKG86505	D138H	KKG81751	Q836L
KKV38905	G143V	QLB39201	G838D
QLB39236	R158S	KKG89654	A845D
QLC91400	R214L	KKV35819	A845V
QLC47920	F220L	KKV38964	L922F
KKY77964	L229F	KKS65656	S922F
KKS65788	H245R	QLC92852	A1078V
KKX46227	D253G	QLC47920	R1091L
QLC48052	A262S	KKG90434	T1120I
KKG90986	V267L	KKG86529	V1129A
KKV35267	R273S	KKV35279	L1141F
KKV37031	P330S	QLC91196	P1162S
KKV39455	T345S	KKG91082	E1195Q
QLC48016	N354K	KKS65584	G1219V
KKV08239	P384L	QLC93524	V1228L
KKI30376	E554D	QLC92372	P1263L

provides a selective odds to the virus has been debated,²⁶ it has now been known that this variant infects human ACE2 cell lines more efficiently than wild-type virus, that offspring virus has increased expression of S protein, that the S protein has a higher rate of binding to ACE2.^{27,28} As in these studies, the obtained mutations may affect the interaction of the spike protein with ACE-2 and may affect the transmission rate of the virus under certain environmental conditions. In a study of household transmission in China, opening windows to allow better air movement led to lower secondary household transmission.²⁹ Poor ventilation has been implicated in numerous transmission clusters, including those in bars, churches, and other location.^{30,31} Even such specific areas can affect the

distribution of the virus, while geographically, it may affect the distribution of the virus where there are special climates and conditions.

Predicted reasons for structure damages. All missense mutations were used to predict structure damage and the results are shown in Figure 1. Predicted structure damage for D614G mutation (found in all regions) is due to substitution, which replaces glycine originally located in a bend curvature in this area (Figure 1A). T393P isolated from Europe substitutes and introduces a buried proline which triggers disallowed phi/psi alert. The phi/psi angles are found in the favored region of the wild-type residue but not in outlier region of the mutant residue (Figure 1B).

Table 5. Seventy six mutation results based on whole sequences (635 sequences) of spike proteins in Asia region.

ACCESS. NUMBER	THE MUTATION	ACCESS. NUMBER	THE MUTATION
QJX44586	F2L	QJD23249	H519Q
QIT07011	L8V	QIU81885	A570V
QJX44430	S13I	QJT43608	T572I
QKO25614	Q14H	QKJ68545	D574Y
QJQ84843	T22I	QJR84537	E583D
QIA20044	Y28N	QKJ68497	Q613H
QKO25770	H49Y	QIT06999	D614G
QIU80913	S50L	QIU81873	A653V
QKO25770	T76I	QKT20894	H655Y
QLA46612	L54F	QKW92184	Q675H
QJY40517	R78M	QLA10116	Q677H
QLA46612	F86S	QKV49386	R682Q
QLA46612	T95I	QKN61217	R682W
QKO25758	D138H	QKU37093	A684V
QKE61684	N148Y	QJX44634	A706S
QKV27551	W152	QJD47800	R765L
QKQ30162	M153I	QIZ16509	V772I
QJT43452	E156D	QJD20632	T791I
QJY40469	S162I	QKY60177	K786N
QKJ68737	Q173H	QKY65277	K795Q
QJW00291	M177I	QKO00486	P809S
QLA09870	K188N	QJQ84831	A829T
QKO25794	N211Y	QJT43584	T827I
QHZ00379	S221W	QJX44466	A879S
QLA10140	W258L	QJD47718	S884F
QKY60121	A262S	QJT43572	A892V
QKX47933	G261R	QIA98583	A930V
QJC19491	Q271R	QKK12815	S939Y
QJD23249	L293M	QKF95522	Q1002E
QJD23249	D294I	QJY40517	H1083Q
QJD23249	P295H	QKI31226	F1109L
QKV49386	V367F	QKO25782	V1104L
QJX44562	E471Q	QJR84369	K1181R
QKY60177	Q506H	QKJ68545	D1153Y
QKY60177	Y508N	QKO25674	K1191N
QKY60177	P507S	QKJ68605	Q1201K
QKY60189	P507H	QJR84429	C1243F

Although a large number of the same result is obtained in number, several are represented as a representation.

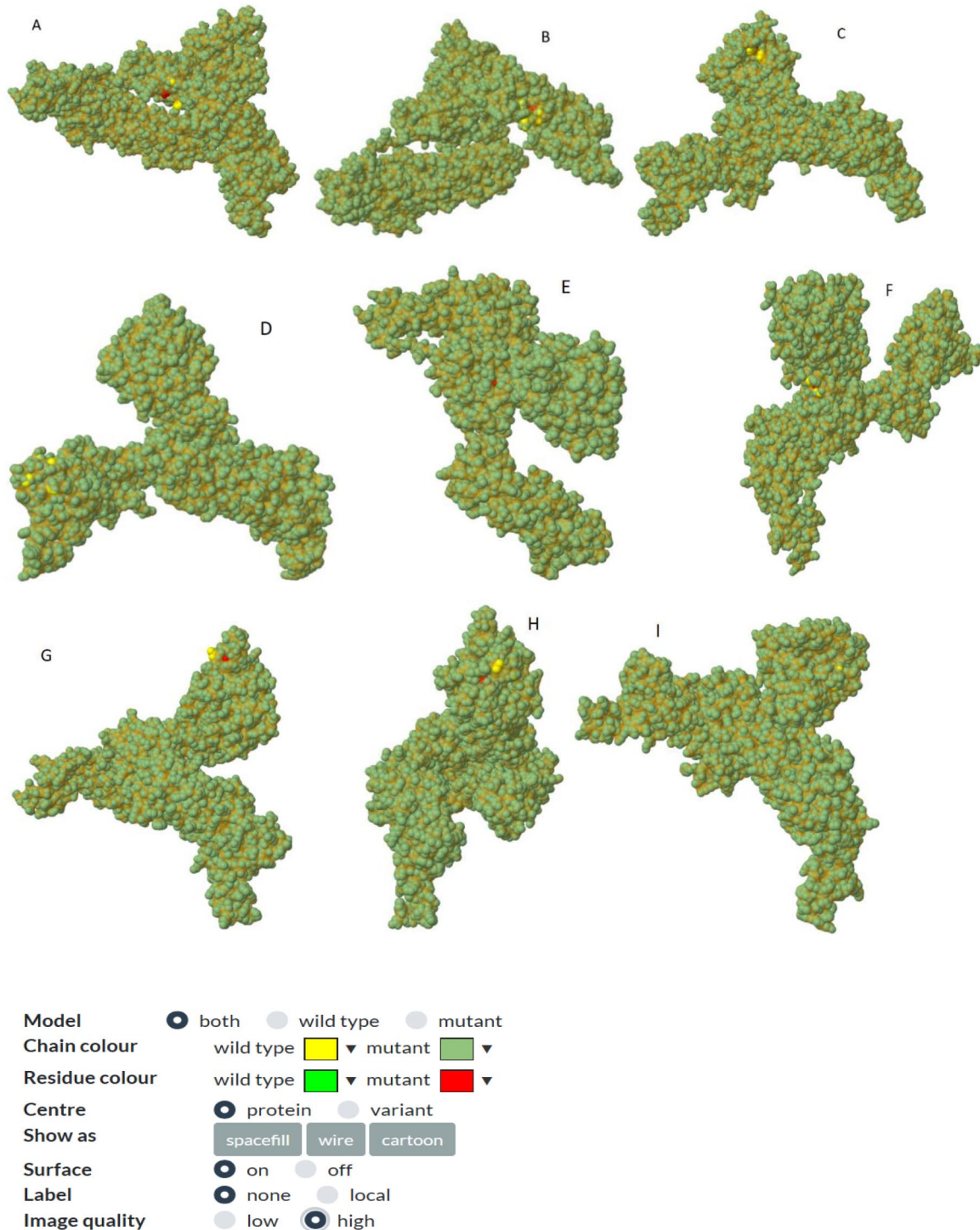


Figure 1. All mutations found in Tables 1-5 were analyzed one by one based on their region by using bioinformatic tools. And mutations that predicted to may affect the structure of the spike protein are shown. These mutations are D614G(A), T393P(B), M177I(C), P507S(D), P295H(E), L293M(F), G75V(G), G143V(H), T95I(I) respectively. According to bioinformatics analysis, these mutations might affect the structure of the spike protein. Both structures are given and illustrates with colors. While yellow color shows wild-type chains, dark green color shows mutant chains. Light green color shows wild-type residue, and red color shows mutant residues. The reason why the light green color does not appear is that it remained inside the shape.

The predicted reason for M177I isolated from Asia is that substitution results in a change between the buried and exposed state of the target variant residue. Metiyonin is buried (relative solvent accessibility (RSA) = 1.0%) and Arginine is exposed (RSA = 16.9%). RSA for buried has to be <9% and difference between Relative Solvent Accessibility has to be $\geq 5\%$ (Figure 1C). The substitution in the P507S mutant sequence isolated

from Asia replaces a buried uncharged residue (Proline, RSA 0.0%) with a charged residue Histidine (Figure 1D). The substitution in the P295H mutant sequence isolated from Asia replaces a buried uncharged residue (Proline, RSA 0.7%) with a charged residue Histidine and leads to the expansion of cavity volume by 142.128 \AA^3 (Figure 1E). The substitution in the L293M mutant sequence resulted in a change between buried

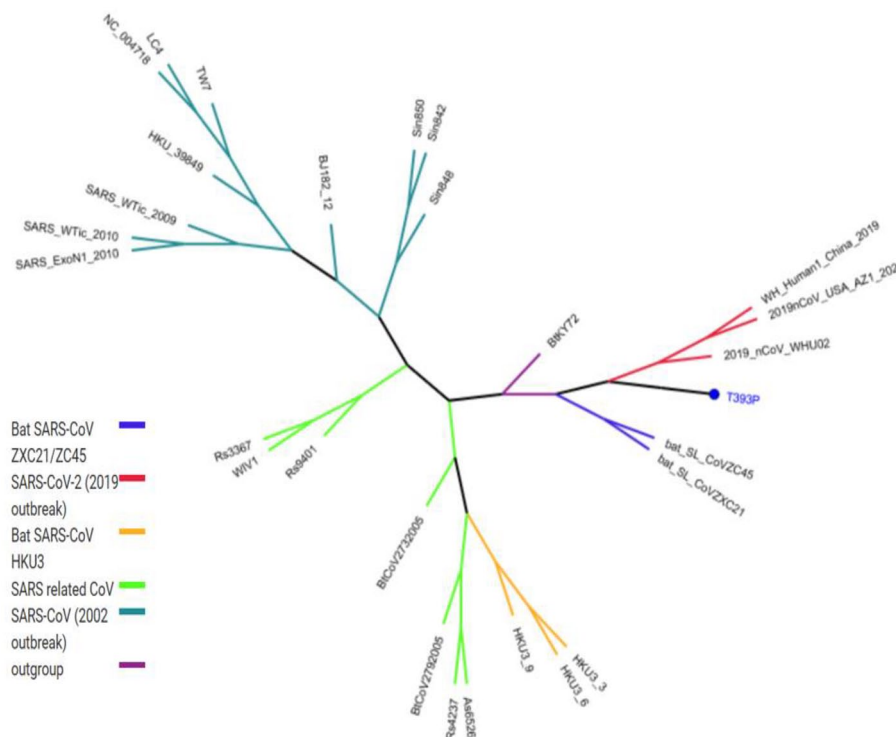


Figure 2. The Phylogenetic tree of one mutation (T393P, in blue color) predicted to may play a role in structure damage according to Genome Detective Coronavirus Typing Tool is shown. All mutations have same location.

and exposed state of the target variant residue. Leucine was buried (RSA 2.4%) and Metiyonin was exposed (RSA 13.2%) (Figure 1F). The substitution in the G75V mutant sequence isolated from North America replaces a buried GLY residue (RSA 3.5%) with a buried Valine residue (RSA 0.0%) (Figure 1G). This (G143V) substitution triggers a disallowed phi/psi alert. The phi/psi angles are in the allowed region of wild-type residue, but not the outlier region of the mutant residue, and it replaces glycine originally located in a bending curvature (Figure 1H). The substitution in T95I mutant sequence isolated from Asia and North America disrupts all side-chain/side-chain H-bond(s) and/or side-chain/main-chain H-bond(s) formed by a buried Threonine residue (RSA 0.0%) (Figure 1I). The phylogenetic tree of mutations according to bioinformatics tools in shown in Figure 2. The phi (ϕ) values of amino acid residues and the psi (ψ) values and H-bond(s) are important to create homolog models and 3D structures of the envelope protein.³² Therefore, finding these results may be used by bioinformaticians to possible vaccine studies and obtaining predicted protein structure. They tend to closely relate to both bats SARS-CoV and outgroup, according to Genome Detective Coronavirus Typing Tool which is assembled genomes in FASTA format.¹⁶ Amino acid forms of spike protein were used to obtain pyhlogenetic some samples. This allows proper identification of other coronavirus types and the chasing of new viral mutations as the outbreak expands globally.¹⁶ Interestingly, mutations that damage the structure did not affect the ligand-binding sites (Figure 3); however, ligands' binding sites were affected in those with

multiple mutations (Figure 4). The results for all mutations detected to affect the structure were the same and are shown in Figure 3. For example, the same source structure (2dd8_S,2ajf_E.pdb) was taken for the structure predicted for all ligand-binding sites. Moreover, all amino acids were the same (Figure 3). QJX45344 that is isolated from Africa has 2 mutations at the same sequence, and the first (Figure 4) represents the sequence result. As seen in Figure 3, the source used to predict the structure was 2dd8_S,2ajf_E; however, the predicted binding sites were different from those of Figure 3. These binding sites includes 338 Phenylalanine (contact: 1, Av distance: 0.00), 339 Glycine (contact: 1, Av distance: 0.00), 342 Phenylalanine (contact: 2, Av distance: 0.18), 343 Asparagine (contact: 2, Av distance: 0.00). The second (Figure 4) is QLA46612 (which has 3 mutations) isolated from South Korea. The source used to predict the structure was 1ww6_A,1ulf_A,1ulc_B; while that used for predicting the binding site was 118 Leucine (contact: 3, Av distance: 0.27), 120 Valine (contact: 3, Av distance: 0.16) 127 Valine (contact: 3, Av distance: 0.05), 129 Lysine (contact: 3, Av distance: 0.00), 157 Phenylalanine (contact: 3, Av distance: 0.169), 159 VAL (contact: 3, Av distance: 0.00), 160 Tyrosine (contact:2, Av distance: 0.00), 169 Glutamic Acid (contact: 2, Av distance: 0.54). However, these predicted binding sites are different from those in Figure 3. It can be said that more than one mutation affects the ligand-binding site based on 3DLigandSite analysis. The phi (ϕ) values of amino acid residues and the psi (ψ) values and H-bond(s) are important to create homolog models and 3D structure of envelope protein.³² Therefore,

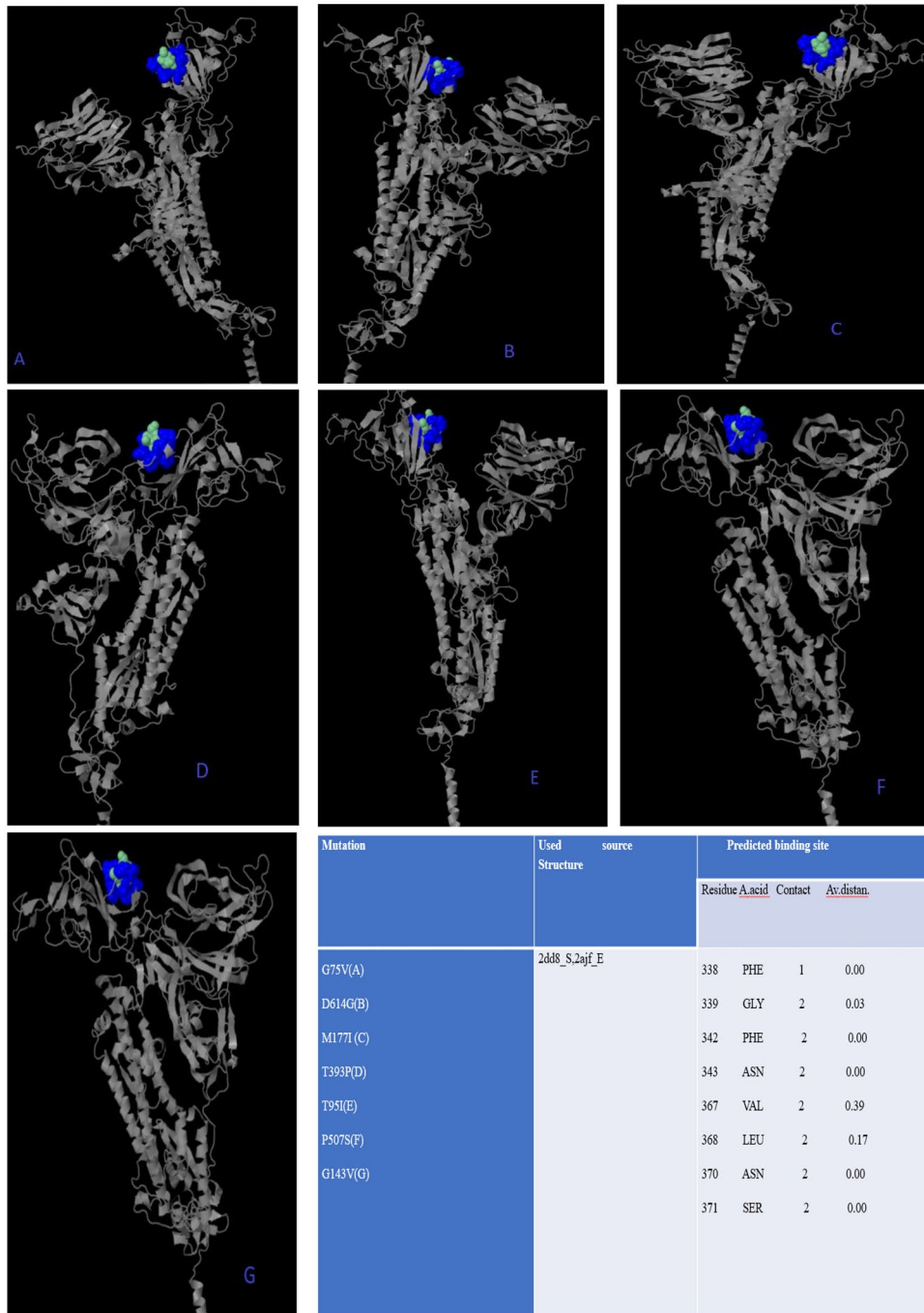


Figure 3. The mutations predicted to might affect ligand-binding site results are obtained by 3DLigandSite.17 They are QJX45344(A, B), QKY60177(C, D), QLA46612(E), QJD23249(F), QKV37632(G). Besides it is predicted that mutations (are shown in Figure 1) damage the structure but did not affect ligand-binding sites.

All results (predicted ligand-binding sites) were the same, but the structures are different. While blue color represents predicted residues, cyan represents heterogens based on 3DLigandSite analysis.

finding this results may be used by bioinformatician to possible vaccine studies and obtaining predicted protein structure

Conclusion

In this study, it was determined that some of the mutations obtained affect the structure of the spike both the protein structure and the binding sites of the ligand, and some did not. In addition, some of these mutations were found in all regions,

while others were found to be only in a certain region. According to this result, mutations may be region specific and can be thought to be affected by environmental factors belonging to that region. Another important result of the study is that more than one mutation is seen in a sample. It can be concluded that more than one mutants maybe found in individuals at the same time. Therefore, attention should be paid to travel as there may be a risk of different mutations according to the regions. It

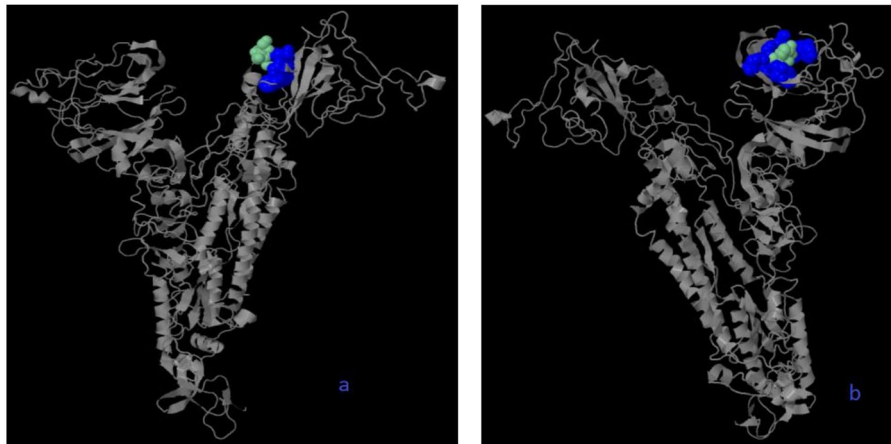


Figure 4. The ligand-binding site and some varying features when 2 or more mutations occur. In these two samples, it was determined that when two mutations occur at the same time, both the structure of the spike surface protein and the ligand-binding sites may be affected (QLA46612(a), QJX45344(b)). While blue color represents predicted residues, cyan represents heterogens based on 3DLigandSite analysis.

should also be known that there is a possibility that one mutation may have different forms in a single person, and accordingly, this possibility should be considered in possible vaccine studies.

ORCID iD

Emre Aktas  <https://orcid.org/0000-0002-9422-3402>

REFERENCES

- Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270-273. doi:10.1038/s41586-020-2012-7.
- Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol*. 2010;84:3134-3146. doi:10.1128/JVI.01394-09.
- Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol*. 2019;17:181-192. doi:10.1038/s41579-018-0118-9.
- Shi J, Wen Z, Zhong G, et al. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Nat Rev Microbiol*. 2020;368:1016-1020. doi:10.1126/science.abb7015.
- Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med*. 2020;18:179.
- Bakhshandeh B, Jahanafrooz Z, Abbasi A, et al. Mutations in SARS-CoV-2; Consequences in structure, function, and pathogenicity of the virus. *Microb Pathog*. 2021;154:104831. doi:10.1016/j.micpath.2021.104831.
- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*. 2020;5:536-544.
- Wang C, Liu Z, Chen Z, et al. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J Med Virol*. 2020;92:667-674.
- Hu B, Guo H, Zhou P, et al. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*. 2021;19:141-154. doi:10.1038/s41579-020-00459-7.
- Lv M, Luo X, Estill J, et al. Coronavirus disease (COVID-19): a scoping review. *Eurosurveillance*. 2020;25:2000125.
- Islam MR, Hoque MN, Rahman MS, et al. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Sci Rep*. 2020;10:14004. doi:10.1038/s41598-020-70812-6.
- Can H, Köseoğlu AE, Erkunt Alak S, et al. In silico discovery of antigenic proteins and epitopes of SARS-CoV-2 for the development of a vaccine or a diagnostic approach for COVID-19. *Sci Rep*. 2020;10:22387. doi:10.1038/s41598-020-79645-9.
- Kumar S, Stecher G, Li M, Kynaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35:1547-1549.
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10:845-858.
- S Islam SA, Khanna T, Alhuzimi E, David A, Sternberg MJE. Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated. *J Mol Biol*. 2019;431:2197-2212.
- Cleemput S, Dumon W, Fonseca V, et al. Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes. *Bioinformatics*. 2020;36:3552-3555.
- Wass MN, Kelley LA, Sternberg MJ. 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res*. 2010;38:W469-W473.
- Korber B, Fischer WM, Gnanakaran S, et al; Sheffield COVID-19 Genomics Group. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*. 2020;182:812-827e19. doi:10.1016/j.cell.2020.06.043.
- Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J Hum Genet*. 2020;65:1075-1082. doi:10.1038/s10038-020-0808-9.
- Ahamad S, Kanipakam H, Gupta D. Insights into the structural and dynamical changes of spike glycoprotein mutations associated with SARS-CoV-2 host receptor binding [published online ahead of print August 27, 2020]. *J Biomol Struct Dyn*. doi:10.1080/07391102.2020.1811774.
- Khan S, Vihinen M. Performance of protein stability predictors. *Hum Mutat*. 2010;31:675-684. doi:10.1002/humu.21242.
- Weisblum Y, Schmidt F, Zhang F, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife*. 2020;9:e961312. doi:10.7554/eLife.61312.
- Robbiani DF, Gaebler CM, et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature*. 2020;584:437-442. doi:10.1038/s41586-020-2456-9.
- Luchsinger LL, Ransgnola BP, Jin DK, et al. Serological assays estimate highly variable SARS-CoV-2 neutralizing antibody activity in recovered COVID-19 patients. *J Clin Microbiol*. 2020;58:e02005-e02020. doi:10.1128/JCM.02005-20.
- Meyerowitz EA, Richterman A, Gandhi RT, et al. Transmission of SARS-CoV-2: a review of viral, host, and environmental factors. *Ann Intern Med*. 2020. doi:10.7326/M20-5008.
- Volz EM, Hill V, McCrone JT, et al. Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity. medRxiv. Preprint posted online. doi:10.1101/2020.07.31.20166082.
- Daniloski Z, Guo X, Sanjana NE. The D614G mutation in SARS-CoV-2 Spike increases transduction of multiple human cell types. bioRxiv. Preprint posted online. doi:10.1101/2020.06.14.151357.
- Zhang L, Jackson CB, Mou H, et al. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv. Preprint posted online. doi:10.1101/2020.06.12.148726.
- Wang Y, Tian H, Zhang L, et al. Reduction of secondary transmission of SARS-CoV-2 in households by face mask use, disinfection and social distancing: a cohort study in Beijing, China. *BMJ Glob Health*. 2020;5. doi:10.1136/bmjgh-2020-002794.
- James A, Eagle L, Phillips C, et al. High COVID-19 attack rate among attendees at events at a church—Arkansas, March 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69:632-635. doi:10.15585/mmwr.mm6920e2.
- Furuse Y, Sando E, Tsuchiya N, et al. Clusters of coronavirus disease in communities, Japan, January-April 2020. *Emerg Infect Dis*. 2020;26:2176-2179. doi:10.3201/eid2609.202272.
- Azeez SA, Alhashim ZG, Al Otaibi WM, et al. State-of-the-art tools to identify druggable protein ligand of SARS-CoV-2. *Arch Med Sci*. 2020;16:497-507. doi:10.5114/aoms.2020.94046.