Research article

Koprululu Kucuk, G. and N. I. Giritlioglu, Homology Modeling of L18F Mutation on SARS-CoV-2 Spike Protein Receptor-Binding Domain. International Journal of Life Sciences and Biotechnology, 2022. 5(3): p. 591- 601. DOI: 10.38001/ijlsb.1159060

Homology Modeling of L18F Mutation on SARS-CoV-2 Spike Protein Receptor-Binding Domain

Gizem Koprululu Kucuk [1]*, Nazlı Irmak Giritlioglu[2]

ABSTRACT

Proteins have unique properties to participate in many structural and physiological processes. Knowledge of the three-dimensional structure of proteins is important to understand their roles in the physiological processes and the functions of these processes. Any structural defect in proteins due to mutations can cause diseases, treatment unresponsiveness, and drug resistance development. The recent emergence of the new SARS-CoV-2 variants containing mutations that accelerate the spread of the virus by affecting infectiousness has been of concern. In the study, visualization of the homology model and investigation of the chemical properties of L18F mutation responsible for the formation of mutant type SARS-CoV-2 spike protein via in silico approach was intended. In this study, amino acid number, molecular weight, theoretical pI value, the percentage composition of amino acids, total negatively charged residue number, total positively charged residue number, atomic composition, formula, total atomic number, molar extinction coefficient, aliphatic index, and the average hydropathy were calculated via ProtParam. The FASTA amino acid sequence was used for visualization of the homology models via UCSF Chimera in wild-type and mutant-type spike proteins. Basic chemical calculations also were displayed on BIOVIA Discovery Studio Visualizer. ΔΔG value and the changes in the stability in L18F mutation were predicted via I-Mutant Suite software. We detected that location of the mutant residue is near a highly conserved position and the L18F mutation may not cause the damage.

ARTICLE HISTORY
Received
8 August 2022
Accepted
2 November 2022

KEYWORDS
in silico, spike protein, homology modeling, SARS-CoV-2, mutation

Introduction

Proteins are molecules that participate in many physiological processes and are found in all organisms. Proteins participate in many physiological processes such as cell structure, cell viability, cell signals, ligand binding, and enzyme catalysis [1]. All proteins consist of amino acid sequences called polypeptide chains. In proteins in biological organisms, there are 20 types of amino acids with different characteristics, all of which have a common central carbon atom, an amino group, and a carboxyl group [2]. Each amino acid has its

[1] Department of Radiotherapy, Istanbul Sisli Vocational School Istanbul, Turkey
[2] Department of Molecular Biology and Genetics, Graduate School of Science and Engineering, Yildiz Technical University, Istanbul, Turkey
*Corresponding Author: Gizem Koprululu Kucuk, e-mail: gzm0507@gmail.com

properties [3]. The three-dimensional structure (3D) is important for understanding protein structure [4]. The specific order of amino acids determines the 3D structure of a protein [5]. The folding of a protein into a certain conformation depends on physicochemical properties such as hydrogen bonds, ionic interactions, Van der Waals forces and hydrophobic interactions, and covalent interactions between amino acids [6]. The amino acid sequence determines how these interactions occur. The way a protein folds also depends on environmental influences such as the presence of water or lipids and the pH of the environment [7]. Knowing this structure is vital in understanding the function of the protein. Changing any amino acid can disrupt the entire structure or function of the protein by disrupting the protein and its environment, as well as the forces of interaction within the protein [8].

Knowing the three-dimensional structure of proteins is important in understanding their role in the physiological process and the functioning of this process [9]. Three-dimensional structures of proteins can be revealed by various experimental studies. However, besides these experimental studies, there are also important developments in studies on determining the three-dimensional structural models of proteins using bioinformatics methods. Any structural defect in proteins due to mutations can cause diseases, treatment unresponsiveness, and drug resistance development [10].

The emergence of new variants of SARS-CoV-2, which have emerged recently and contain mutations that affect transmissibility and accelerate the spread of the virus, has been a matter of concern [11]. Especially with the identification of two independent strains that emerged in the United Kingdom and South Africa, studies on the detection of new mutations gained momentum. SARS-CoV-2 strains B.1.1.7 emerged in the UK (HV 69-70 deletion, Y144 deletion, N501Y, A570D, P681H, T716I, S982A, D1118H) and B.1.351 in South Africa (L18F, D80A, D215G, R246I, K417N, E484K, N501Y, and A701V) were detected in the spike protein receptor binding site (RBD) [12].

Within the scope of the study, homology modelling studies of the L18F mutation responsible for the formation of mutant type coronavirus occurring in the receptor-binding domain (RBD) of spike protein by Swiss Model and amino acid number, molecular weight,

theoretical pI value, the percent composition of amino acids, the total amino acid content of wild and mutant type were calculated via ProtParam.

Materials and Methods

The L18F mutation was detected in the B chain in the receptor-binding domain (RBD) of the spike protein of the Sars CoV2 virus that causes Severe Acute Respiratory Syndrome (SARS). The homology model of L18F mutation was created by bioinformatics tools. In the study, the Uniprot website and GenBank database of the National Center for Biotechnology Information (USA, NCBI) website were used detection of the amino acid sequence of the SARS CoV2 spike RBD protein [13]. Sequence information was arranged according to wild type and mutant type. Homology models of Spike protein types (wild and mutant types) were created with the Swiss Model Program which is a web-based bioinformatics tool, and the obtained three-dimensional models were examined with the UCSF Chimera program which is a visualization tool [14, 15]. Physico-chemical properties of models were obtained by ExPASy-ProtParam which is a bioinformatic tool and results were evaluated. sequence with accession numbers P0DTC2 was used for wild type in this study [16].

Swiss model analysis

The homology modelling study was carried out using Swiss-Model [17] database and the Chimera program. Wild and mutant type sequence sets were loaded into the system separately and their three-dimensional structures were obtained. The selection was made by looking at the Qualitative Model Energy Analysis (QMEAN) values of the three-dimensional structures obtained.

Homology Modelling by Chimera Program

Using UCSF Chimera program tools, all proteins were superimposed with each other, and visualization of three-dimensional structures was provided. The structural differences of the mutant protein were observed by visualizing the wild-type and mutant protein structures with a ribbon display.

Detection of Physico-chemical Properties

In the study, the Physico-chemical properties of Spike protein were calculated to understand the functional diversity. For this purpose, the ProtParam program of the ExPASy database, which is one of the bioinformatics tools, was used. FASTA formats of wild and mutant types were obtained on the GeneBank database. Physico-chemical properties of models were calculated. In both models (wild and mutant type); amino acid number, molecular weight, theoretical pI value, the percent composition of amino acids, total number of negatively charged residues, the total number of positively charged residues, atomic composition, formula, total atomic number, molar extinction coefficient, aliphatic index, and average hydropathy were calculated.

Result and Discussion

In homology modelling, protein sequence and three-dimensional structure information previously obtained by methods such as X-ray are used. Although a nature-like structure is not fully met in modelling, these studies are of great importance for clinical drug development studies. Mutations occurring in the RBD region are an important issue affecting sustained ligand binding and thus virus infectivity [18]. The changes caused by these mutations on the binding surface give us information about the degree of danger of the variation.

The amino acid sequence format that will be used as a basis in bioinformatics studies has been created based on the NCBI-P30518 accession number sequence. These arranged sequence sets were used to create the three-dimensional structure and extract the pysico-chemical properties of the models. The sequence of wild-type and mutant-type proteins is shown in Figures 1 and 2.

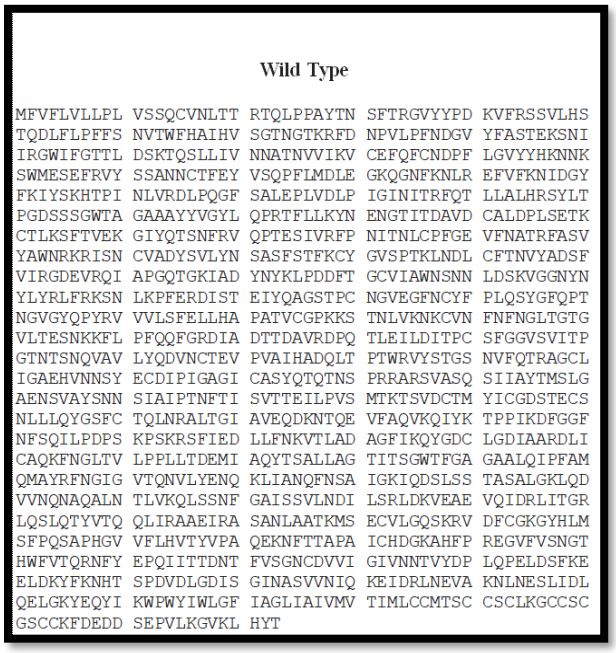

Fig 1 Wild-type sequence for the formation of the three-dimensional structure

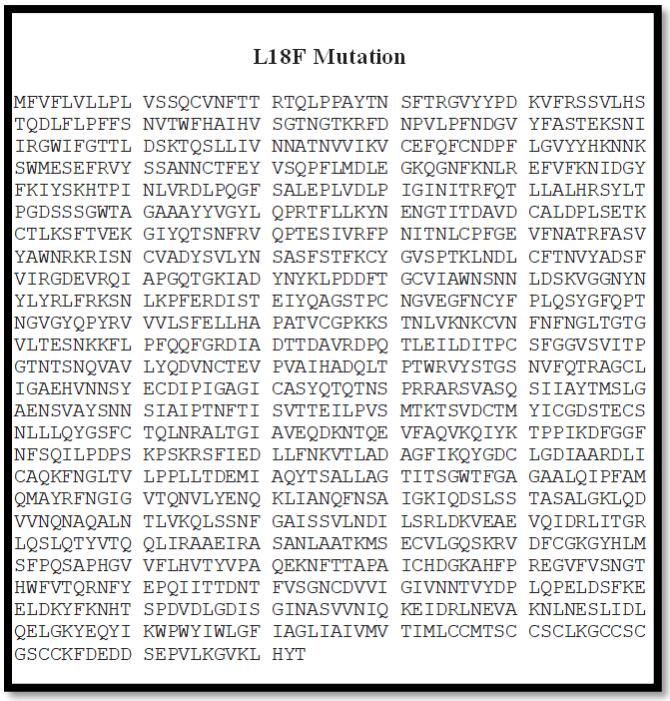

Fig 2 Mutant-type sequence for the formation of the three-dimensional structure

3D structures of the spike protein were obtained by using wild and mutant-type sequence sets. 3D structure of Spike protein homology modelling was performed using the Swiss-Model (www.swissmodel.expasy.org) database and the Chimera program,

By examining the QMEAN values of the three-dimensional structures obtained, high models were selected and used as models (Table 1). We used the P0DTC2 accession number sequence for the wild type as the basis. The selection was made by evaluating the QMEAN values of the 3D structures obtained as a result of Swiss Homology modelling. The QMEAN for the wild type was detected at -1.59 and the mutant type was detected at -1.58.

Table 1 QMEAN values of the models that homology modeling was built by Swiss-Model

	QMEAN
Wild-type	-1,59
L18F	-1,58

Basic chemical calculations were displayed on BIOVIA Discovery Studio Visualizer (Table 2). Leucine is a more hydrophobic amino acid than phenylalanine (Hydrophobicity values are 3.8 and 2.8, respectively). Protein-ligand binding needs hydrophobic interactions. L18F mutation at the RBD site of the spike protein gains more hydrophilic character to the protein. Changing the stability and ΔΔG value in L18F mutation was predicted via I-Mutant software. The conditions were selected as 25°C, pH: 7.0 (default settings), and the software showed that the stability of the mutant spike protein decreased. Also, the graph of temperature–ΔΔG was created in Microsoft Excel. According to the data, ΔΔG is increased when the temperature is decreased. The increased temperature causes a more unstable mutation at the B: 18F site.

Table 2 Calculation of hydrophobicity by BIOVIA Discovery Studio Visualizer

	Full name	**Hydrophobicity**	**Secondary structure**
Wild-type	B:Leu18	3.8	Coil
Mutant-type	B:Phe18	2.8	Coil

Also HOPE web service was utilized for determining the structural effects of the mutant protein in this study. The location of the mutant residue is near a highly conserved position and L18F mutation may not cause damage. Receptor-ligand interactions could be changed negatively in the structural base because the mutant residue is bigger than the wild-type residue. The amino acid change was detected for the mutant type (Figure 3).

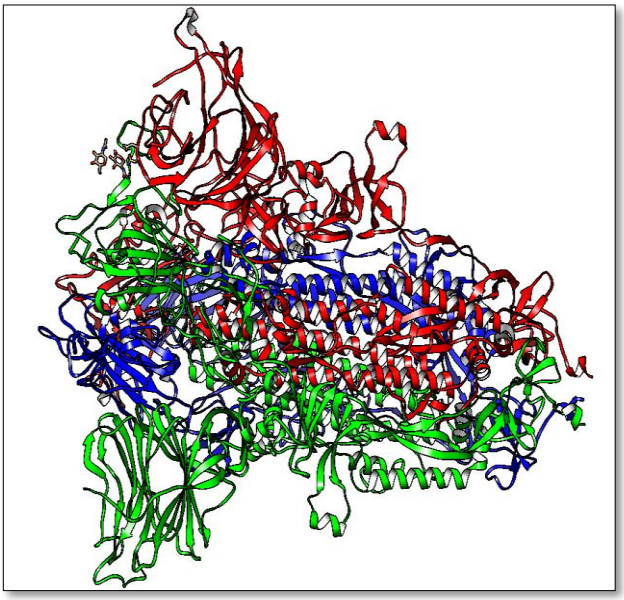

Fig 3 Wild-type ribbon structure

The change in the receptor-binding domain of the mutant type Spike protein is shown in Figure 4. The amino acid Leucine at position 18 in the wild-type receptor-binding domain of the spike protein was converted to mutant-type as phenylalanine. When the ribbon structure of both models was examined, conformational change was not observed.

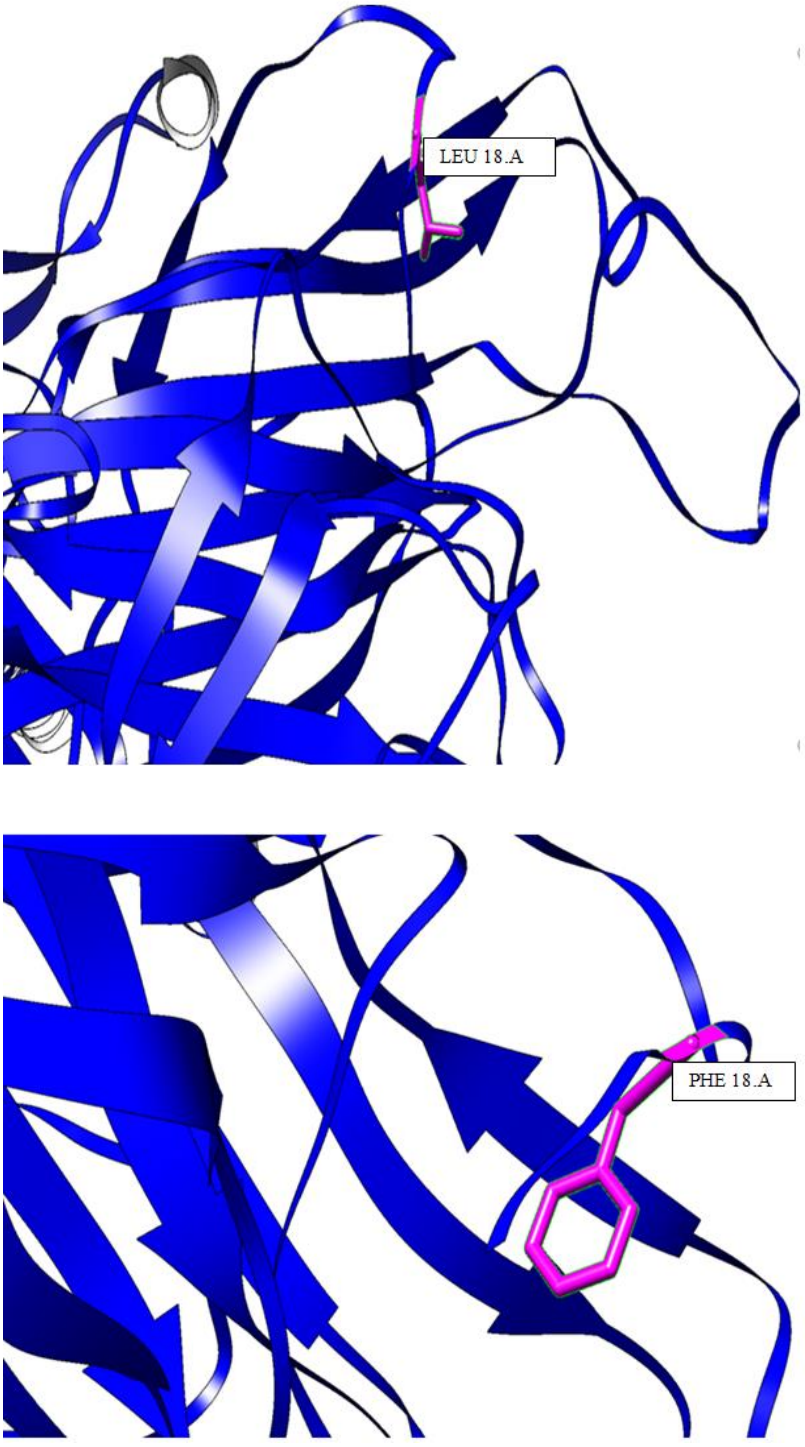

Fig 4 Amino acid changes

Leucine is a more hydrophobic amino acid than phenylalanine (Hydrophobicity values are 3.8 and 2.8, respectively). Protein-ligand binding needs hydrophobic interactions. L18F

mutation at the RBD of the spike protein gains more hydrophilic character to the protein. Also HOPE web service was utilized for determining the structural effects of the mutant protein in this study. The location of the mutant residue is near a highly conserved position and L18F mutation may not cause damage. Receptor-ligand interactions could change negatively in the structural base because the mutant residue is bigger than the wild-type residue.

Travel restrictions have been imposed in most countries since the pandemic was declared. However, the SARS-CoV-2 virus continued to mutate. New SARS-CoV-2 variants occurred and continued to spread across continents.

Substitution of phenylalanine with aromatic ring instead of leucine with aliphatic chain impairs steric elasticity in the protein chain. It may cause the ligand not to cleave. It breaks conformational elasticity sterically. Aromatic rings increase stability [19]. As the increased stability also strengthens the ligand-protein bonds, continuous binding takes place. This increases the damage of the mutation. It is a negative situation in methods such as xray in the study of protein ligand interactions. Because the relaxed structure of the protein cannot be visualized. There is no such disadvantage in homology modelling [20].

In a study, Kuzmina A. et al. detected that the L18F mutation (B. 1. 351 variant) is more resistant to greater infectivity and antibody neutralization [21]. Moreover, they showed that when K417T/E484K mutants were compared with L18F, it was detected that these mutants resisted the post-vaccination serum neutralization similarly to L18F.

On the contrary, Thomson EC. et al. found that N439K facilitated resistance to antibody neutralization [22]. Boon S. et al. speculated that N439S, T478S, and N501K mutations gave the virus a chance to infect host cells more efficiently and with low antigenicity. They detected that these mutations allow SARS-Cov2 to enter the host cell easily and make it easier to infect the cell [23].

Shahhosseini N. et al. used the SWISS-MODEL database to create homology models for SARS-CoV-2 Spike protein mutants [24]. We also used the SWISS-MODEL database for homology modelling. However, we cannot obtain hydrophobic interactions, hydrophobic structure, or atom-bond visualizing for protein by SWISS-MODEL. Because of this situation, we used the Chimera program and we backed up our models with Chimera.

Conclusion

Detection of the effects of changes in the protein sequences of the SARS-CoV-2 virus is an important element in determining the spread and severity of COVID-19. In particular, since spike protein is closely related to receptor proteins such as ACE2, FURIN, and TMPRSS2 in humans, it is extremely necessary to determine the rate and direction of change in this protein. The change from leucine to phenylalanine at amino acid 18 is found in the VOC strain that has a replicative advantage, and it is clear that such mutations result in subspecies with novel properties. When we consider the drug development process from a technological point of view, vaccine studies against COVID-19 are at the beginning and it is a matter of debate whether they are sufficient against variants. We think that it is important to model all newly formed variants just like in our study and to examine them from all chemical aspects. In addition to our in silico study, performing in vitro studies will be important in terms of understanding the beta coronavirus species and subspecies that are likely to occur in the future.

References

1. Lee, M. J., and M.B. Yaffe, Protein regulation in signal transduction. Cold Spring Harbor Perspectives in Biology, 2016. 8(6).
2. Alberts, B., A. Johnson, and J. Lewis, The Shape and Structure of Proteins. 2002, United States: Molecular Biology of the Cell.
3. Lopez, M J. and S.S. Mohiuddin, Biochemistry; Essential Amino Acids. StatPearls. Retrieved, 2020.
4. Schmidt, T., A. Bergner, and T. Schwede, Modelling three-dimensional protein structures for applications in drug design. Drug Discovery Today Elsevier Ltd., 2014.
5. Suryanto, C. H., H. Saigo, and K. Fukui, Structural Class Classification of 3D Protein Structure Based on Multi-View 2D Images. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2018. 15(1): p. 286–299.
6. Englander, S. W., and L. Mayne, The nature of protein folding pathways. Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 2014.
7. Sun, P. D., C.E. Foster, and J.C. Boyington, Overview of protein structural and functional folds. Current Protocols in Protein Science, 2004.
8. Fuentes-Prior, P., and G.S. Salvesen, The protein structures that shape caspase activity, specificity, activation, and inhibition. Biochemical Journal, 2004.
9. Torrisi, M., G. Pollastri, and Q. Le, Deep learning methods in protein structure prediction. Computational and Structural Biotechnology Journal, 2020.

10. Longley, D. B., and P. G. Johnston, Molecular mechanisms of drug resistance. Journal of Pathology, 2005.
11. Khandia, R., et al., Emergence of SARS-CoV-2 Omicron (B.1.1.529) variant, salient features, high global health concerns and strategies to counter it amid ongoing COVID-19 pandemic. Environmental Research, 2022.
12. Naveca, F., et al., Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the Spike protein. Virological, 2021. p. 1–12.
13. https://www.ncbi.nlm.nih.gov/ Access date: 15.06.2022.
14. Waterhouse, A., et al., SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research, 2018. p.296-303
15. Pettersen, EF, et al., UCSF Chimera—A Visualization System for Exploratory Research and Analysis. Journal of Computational Chemistry, 2004. 25(13): p.1605-12.
16. Gasteiger, E., et al., Protein Identification and Analysis Tools on the ExPASy Server, J.M. Walker, Editor. 2005, Humana Press. U.S.: U.S. p. 571-607.
17. www.swissmodel.expasy.org. Access date: 15.06.2022.
18. https://bmcmolcellbiol.biomedcentral.com/articles/10.1186/s12860-021-00403-4, Accessed: 18.10.2022 .
19. Biancalana, M., et al., Aromatic cluster mutations produce focal modulations of β-sheet structure. Protein Science, 2015. 24(5): p. 841–849.
20. Celej, MS., G.G. Montich, and G.D. Fidelio, Protein stability induced by ligand binding correlates with changes in protein flexibility. Protein Science, 2003. 12(7): p. 1496–1506.
21. Kuzmina, A., et al., SARS-CoV-2 spike variants exhibit differential infectivity and neutralization resistance to convalescent or post-vaccination sera. Cell Host Microbe, 2021. 29:p. 522–528.
22. Thomson, EC., et al., COVID-19 Genomics UK (COG-UK) Consortium, et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. Cell, 2021. 184: p. 1171–1187.
23. Boon, S. S., et al., Temporal-Geographical Dispersion of SARS-CoV-2 Spike Glycoprotein Variant Lineages and Their Functional Prediction Using in Silico Approach. American Society for Microbiology, 2021. 12(5).
24. Shahhosseini, N., et al., Mutation signatures and in silico docking of novel sars-cov-2 variants of concern. Microorganisms, 2021. 9(5).