



ISSN: 0067-2904

Investigation of Mutations in Cleavage and Fusion regions within Spike gene of SARS-CoV-2

Noor Saber^{1,2*}, Nuha J Kandala¹

¹Department of Biotechnology, College of Science, University of Baghdad, Baghdad, Iraq

² National Center of Hematology, Mustansiriyah University, Iraq

Received: 16/4/2024

Accepted: 24/8/2024

Published: 30/8/2025

Abstract

The Furin cleavage site and fusion peptide are critical regions in the Spike protein of SARS-CoV-2. The activation of the spike protein relies on furin cleavage, whereas the fusion peptide is critical for merging the viral envelope with host cell membranes. This study aims to map the mutations in the S1/S2 Furin, S2' cleavages sites, and fusion peptide regions within Iraqi isolates, examining the variability and conservation of these critical sites. Our study includes 14 SARS-CoV-2 Sanger sequenced fragments coupled with data mining of the 1043 genomic dataset of Furin/ S2' cleavage sites and fusion peptide regions extracted from whole genome sequences, representing the entire entry of Iraq to GISAID during the pandemic. The current analysis identifies a mutational hotspot located within the Furin cleavage region at codon P681. The proline at this position changed to arginine in the Delta variant and subsequent sublineages. However, this change was displaced by P681H in Omicron variants and related newly emerging strains. Furthermore, the combination of P681H and N679K mutations enhances the basic nature of the site, making it a more favourable target for Furin, thereby boosting the Spike protein activation process. The investigation emphasizes the prevalence of D614G, H655Y, and D796Y mutations as substantial factors influencing viral transmissibility and pathogenicity. These mutations persist in global strains, indicating their role in the virus's evolutionary trajectory. The conservative nature of the Furin and S2' cleavage sites underscores its critical role in viral transmissibility and pathogenicity and adaptability in various proteases environment in the host cell. Mutations such as A688S, S698L, A701V, and S704L, which can impact vaccine-induced antibody sensitivity, were also detected, highlighting the ongoing challenge of maintaining vaccine efficacy against evolving strains.

Keywords: SARS-CoV-2, Furin Cleavage Region, Fusion Peptide Region, Genomic Surveillance, Iraqi Isolates.

التحري عن الطفرات في مناطق الانقسام والالتحام ضمن الجين الشوكي لفايروس SARS-CoV-2

نور صابر^{1,2*}, نهى جوزيف قندلا¹

¹التقنيات الاحيائية, كلية العلوم, جامعة بغداد, بغداد, العراق

²المركز الوطني لبحوث وعلاج امراض الدم, الجامعة المستنصرية, بغداد, العراق

الخلاصة

*Email: noorsabeer1993@gmail.com

يعد موقع انقسام الفيورين وبيتيد الاندماج من المناطق المهمة في البروتين الشوكي لفايروس SARS-CoV-2. حيث يعتمد تنشيط البروتين الشوكي على انقسام الفيورين بينما يلعب بيتيد الاندماج دورا مهما في دمج الغلاف الفيروسي واغشية المضيف. تهدف هذه الدراسة إلى رسم خريطة للطفرات في مناطق انقسام الفيورين S1 / S2 ومنطقة انقسام الـ S2 وبيتيد الاندماج داخل العزلات العراقية ، وفحص مدى تغاير هذه المواقع الحرجة. تتضمن دراستنا 14 قطعة متسلسلة من طريقة Sanger إلى جانب تنقيب البيانات الجينومية لـ 1043 منطقة انقسام فيورين ومنطقة انقسام الـ S2 وبيتيد الاندماج المستخرجة من تسلسل الجينوم الكامل الذي يمثل مدخول العراق بالكامل إلى مستوعب GISAID أثناء الوباء. يحدد التحليل الحالي موضع منطقة الطفرات الساخنة داخل منطقة انقسام الفيورين في الكودون P681. تغيّر البرولين في هذا الموضع إلى أرجينين في متغير دلتا والسلالات الفرعية اللاحقة. ومع ذلك ، فإن هذا التغيير استبدل بـ P681H في متغيرات Omicron والسلالات الناشئة حديثا ذات الصلة. علاوة على ذلك، فإن الجمع بين طفرات P681H وN679K يعزز الطبيعة القاعدية للموقع، مما يجعله هدفاً أكثر ملاءمة للفيورين، وبالتالي تعزيز عملية تنشيط البروتين الشوكي. يوضح هذا التحري الطفرات D614G و H655Y و D796Y كمساهمين مهمين في انتقال الفيروس والإمراضية. تستمر هذه الطفرات في السلالات العالمية ، مما يشير إلى دورها في المسار التطوري للفايروس. تؤكد الطبيعة المحافظة لمواقع انقسام الفيورين و S2 على دورها الحاسم في قابلية انتقال الفيروس والقدرة على إحداث المرض والقدرة على التكيف في بيئة البروتيازات المختلفة في الخلية المضيفة. تسلط الطبيعة المحافظة لموقع انشقاق الفيورين الضوء على دوره الحاسم في انتقال الفيروس والإمراضية. كما تم الكشف عن طفرات مثل A688S و S698L و A701V و S704L و N764K ، والتي يمكن أن تؤثر على استجابة الأجسام المضادة التي يولدها اللقاح ، مما يؤكد التحدي المستمر المتمثل في الحفاظ على فعالية اللقاح ضد السلالات المتطورة.

1. Introduction

SARS-CoV-2, an enveloped ssRNA virus responsible for COVID-19, has caused the ongoing global health crisis [1]. Since first reported case in Wuhan 2019, the WHO has declared 774,395,593 infected cases and 7,023,271 deaths globally [2]. SARS-CoV-2 virus possesses a unique spike gene that codes for a transmembrane spike glycoprotein, spike protein exhibits a binding affinity for the human angiotensin-converting enzyme 2 (ACE2) receptor. ACE-2 receptor is prevalent in lung cells, especially type II alveolar cells, and is also expressed in various other tissues of the human body. This wide expression attributes of ACE-2 receptor suggests diverse routes for viral entry and highlights the complexity of SARS-CoV-2 transmission and pathogenesis [3, 4]. The S protein structure is characterized as a trimer, consisting of S1 and S2 subunits, S1 binds to the ACE-2 receptor, whereas S2 domain includes elements necessary for viral fusion process [5, 6]. The S1 subunit comprises the N-terminal and receptor-binding domains, is linked to the S2 subunit. A specific amino acids segment with a polybasic in nature, cleavable by host cell Furin protease, is located between the S1 and S2 subunits [7-10]. The S1/S2 cleavage site comprises six amino acid sequences including proline, arginine, arginine, alanine, arginine, and serine that are abbreviated as PRRARS. This amino acid fingerprint does not present in closely related coronaviruses; RRAR↓S represents the specific sequence targeted by Furin to promote primary cleavage, with cleavage occurring between arginine and serine. This cleavage induces a significant conformational change that shifts the spike protein from prefusion (closed) to a post fusion (open) state. This structural transition enables SARS-CoV-2 to fuse with targeted cell membrane, permitting the entry of RNA and initiation of the infection. The presence of this Furin cleavage site has been linked to the enhanced transmissibility and pathogenicity of SARS-CoV-2, due to its ability to facilitate broader tropism and efficient viral spread [11-13]. Upon interaction of viral receptor binding domain to the host ACE-2 receptor and subsequent protease cleavage, the S protein undergoes a conformational change ended with releasing of viral content into the cytosol of targeted cells [14, 15]. The mutations

affect the S2' cleavage site, located at codons 814 KR↓S 816, which is positioned just before the internal fusion peptide. The KR↓S S2' cleavage site is the favourable target of host proteases such as Transmembrane Serine Protease 2 (TMPRSS2) or the cathepsins that derive the second cleavage events. Cleavage at S2' site is a crucial step for exposing the fusion peptide, which is a hydrophobic region that inserted to the host cell membrane promoting viral entry [9, 16,17]. The COVID-19 pandemic is global; there is a paucity of comprehensive surveillance and detailed virological data in regions with limited resources. Although of limited number of studies that pointed to diversity of circulating variants in Iraq and their impact [18, 19]. More efforts are needed to address such gap. This study aims to investigating the unique and common genetic variations within the S1/S2 cleavage and fusion peptide regions in circulating viral variants in Iraq. Molecular tracking was conducted via conventional Sanger sequencing, targeting a 767-nucleotide region encompassing the S1/S2 cleavage site and internal fusion peptide region. In addition, a total of 1043 SARS-CoV-2 genomes retrieved from the GISAID repository were analyzed to monitor mutations within the S1/S2 and fusion peptide regions, assessing their persistence and occurrence in circulating variants.

2. Methodology

2.1 Specimen collection

A total of 18 nasopharyngeal specimens were collected by the Iraqi Central Public Health Laboratory/National Influenza Center from February 17 to March 5, 2022, using viral transport media (VTM). These specimens were collected from patients across various provinces, including Baghdad, Wassit, Basra, Diyala, Muthana, Maysan, and Diwaniyah.

2.2 RNA extraction and RT-PCR Assay

Viral RNA was extracted using the automated Bioneer ExiPrep™ 96 Lite extraction machine (A-5250, BIONEER) with the Exiprep 96 Viral DNA/RNA kit (K-4614, BIONEER). The RNA served as a template for the TaqPath COVID-19 assay, a commercial viral diagnostic kit by Thermo Fisher. Assays were conducted on the Applied Biosystems 7500 Fast RT-PCR system, in accordance with the recommended manufacturer procedure.

2.3 Complementary DNA synthesis

The complementary DNA (cDNA) was synthesized using Promega GoScript™ Reverse Transcription Mix, Random Primers kit (A2800) following the manufacturer's guidelines. The cDNA quality was evaluated using the Promega Quantus Fluorometer.

2.4 Primer Design

One set of forward and reverse primers is used to amplify the cleavage regions and fusion peptide regions of the SARS-CoV-2 Spike gene, the used primers are part of ARTICV3 primers pool <https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye> for the next generation sequencing of SARS-CoV-2. In the current study, a combination of primers was used, consisting of the forward primer SubA_23290F from one set and the reverse primer SubA_24057R from another set, to produce a 767-base pair amplicon. The forward primer (5' CCGTGATCCACAGACACTTGAGAT 3') extended from 23290 to 23313 and the reverse primer (5' CCAGCATCTGCAAGTGTCAC TTT 3') extended from 24035 to 24057. These primers cover the Spike S1-S2 cleavage region (from codon 672-709), Furin cleavage site (685-686), fusion peptide (788-806), 814 KR↓S 816 S2' site and partially covers internal fusion peptide (816-833) till codon 831 (NCBI Reference Sequence: NC_045512.2).

2.5 PCR Assay and Sanger Sequencing

Amplification was performed using the SimpliAmp™ Thermal Cycler (Applied Biosystem) with the following program settings: an initial denaturation at 95°C for 5 minutes; 40 amplification cycles consisting of 30 seconds of denaturation at 95°C, 30 seconds of primer annealing at 60°C, and 30 seconds of extension at 72°C; and a final extension at 72°C for seven minutes. Gel electrophoresis and a gel documentation system were utilized to inspect and visualize the amplicons. The amplicons were submitted to Macrogen Company in Korea for sequencing using the Applied Biosystems ABI3730XL machine.

2.6 Data Mining and Analysis

The resulting FASTA files were trimmed, and all low-quality base calls were removed. The processed FASTA files were then aligned to the Wuhan parental strain using the Map to Reference Assembly tool powered by Geneious Prime software. Annotations for the parental strain were retrieved from the NCBI RefSeq (NCBI Reference Sequence: NC_045512.2). Phylogenetic analysis was performed using Galaxy Pangolin (<https://usegalaxy.org/>) and Nextstrain Auspice (<https://auspice.us/>). Protein prediction was conducted using the Phyre2 platform [20] and CoVsurver, visual representations were created with PyMOL (2.5.2). Iraqi genomic sequences were retrieved from the Global Initiative on Sharing All Influenza Data (GISAID) [21-23], including a total of 1043 whole genome sequences. Spike gene sequences were extracted and annotated to determine the related region using Geneious Prime. Further genomic investigation was performed using GISAID/CoVsurver (v1.22.06). All analyzed sequences have been deposited in GenBank/NCBI and EpiCoV™/GISAID. The metadata for all specimens, as well as the accession numbers for the FASTA files in NCBI GenBank and GISAID, are provided in Supplementary File 1.

3. Results

The FASTA files were processed using Geneious Prime, which involved trimming and filtering to ensure high-quality data, resulting in the exclusion of four files that failed to meet the quality threshold. The trimmed files were assembled to the Wuhan-Hu-1 genome (NC_045512.2), representing the parental strain, via map-to-reference assembly. The average length of the trimmed files was 565 bp, and the pairwise identity for the 14 processed sequences was 99.6%. The maximum length observed was 693 bases, with the shortest being 174 bases. A total of 13 reads covered the S1/S2 cleavage region (from codon 672 to 709), eight covered the fusion peptide (from codon 788 to 806), and six partially covered the internal fusion peptide (from codon 816 to 831). All variations observed in the sequences revealed a total of 27 polymorphisms, comprising 22 Single nucleotide polymorphisms (SNPs) transversions, four SNP transitions, and one substitution. Supplementary file 2 lists all identified mutations, including their lengths, frequencies, and locations. Despite the variable length of the analysed reads, two mutations occurred with 100% frequency, resulted from substitution of aspartic acid to glycine at locus 614 (D614G) due to a transition from alanine to guanine (GAT>GGT), and the changes of histidine with tyrosine at locus 655 (H655T) due to a transition from cytosine to thymine (CAT>TAT), as shown in Figure 1. Additional significant mutations observed in the extended sequences included N679K, P681H, N684K, and D796Y. Additionally, various low-frequency mutations were noted across different loci within the studied fragments which indicated in supplementary file 2.



Figure 1: Illustrative view of main dominant changes within the studied reads, 1: D614G, 2: H655Y, 3: N679K/P681H, 4: K764N and 5: D796Y. Some of the low frequency changes are also represented including the Q613R and L763F.

The lineage identification of SARS-CoV-2 was performed using Nextclade Pango/Galaxy, which revealing that all 14 fragments belonged to the Omicron variant. These findings were further validated by phylogenetic analysis conducted with Auspice, confirming that all mutations were identical with those harbored by BA.1.1 in clade 21K and BQ.1.19 in clade 22E, as presented in Figure 2.

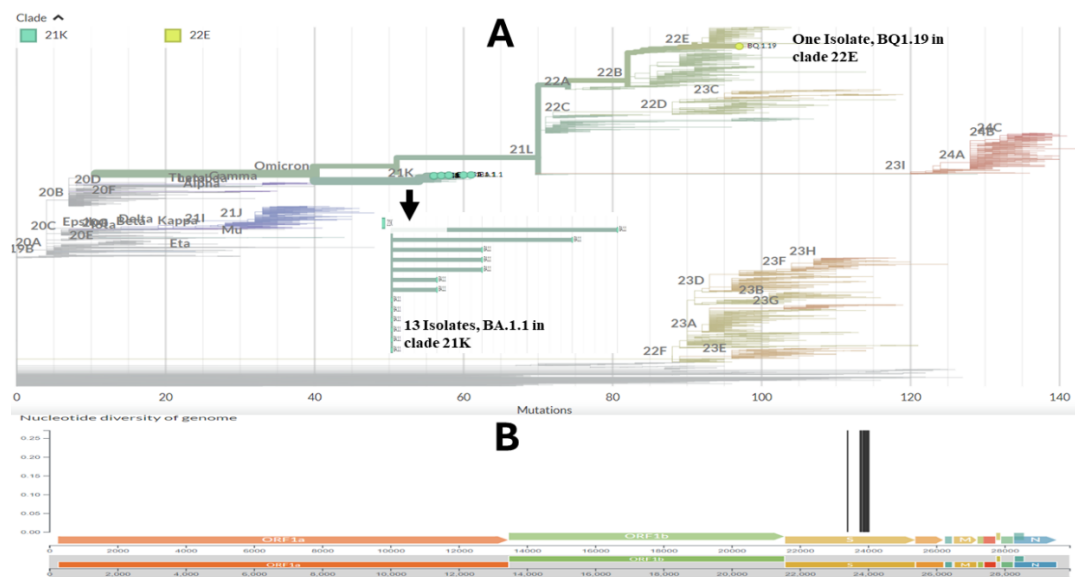


Figure2: Phylogenetic analysis depicts the evolutionary relationships among various SARS-CoV-2 variants. A: illustrates the phylogenetic placement of SARS-CoV-2 samples within the Omicron variant, particularly focusing on the 21K and 22E clades, B: locations of analysed mutations in the SARS-CoV-2 genome.

SARS-CoV-2 Spike partial amino acid sequence containing investigated mutations has been modelled using Phyre2 platform with 100% confidence percentage and visualized using the PyMol as indicated in Figure 3.

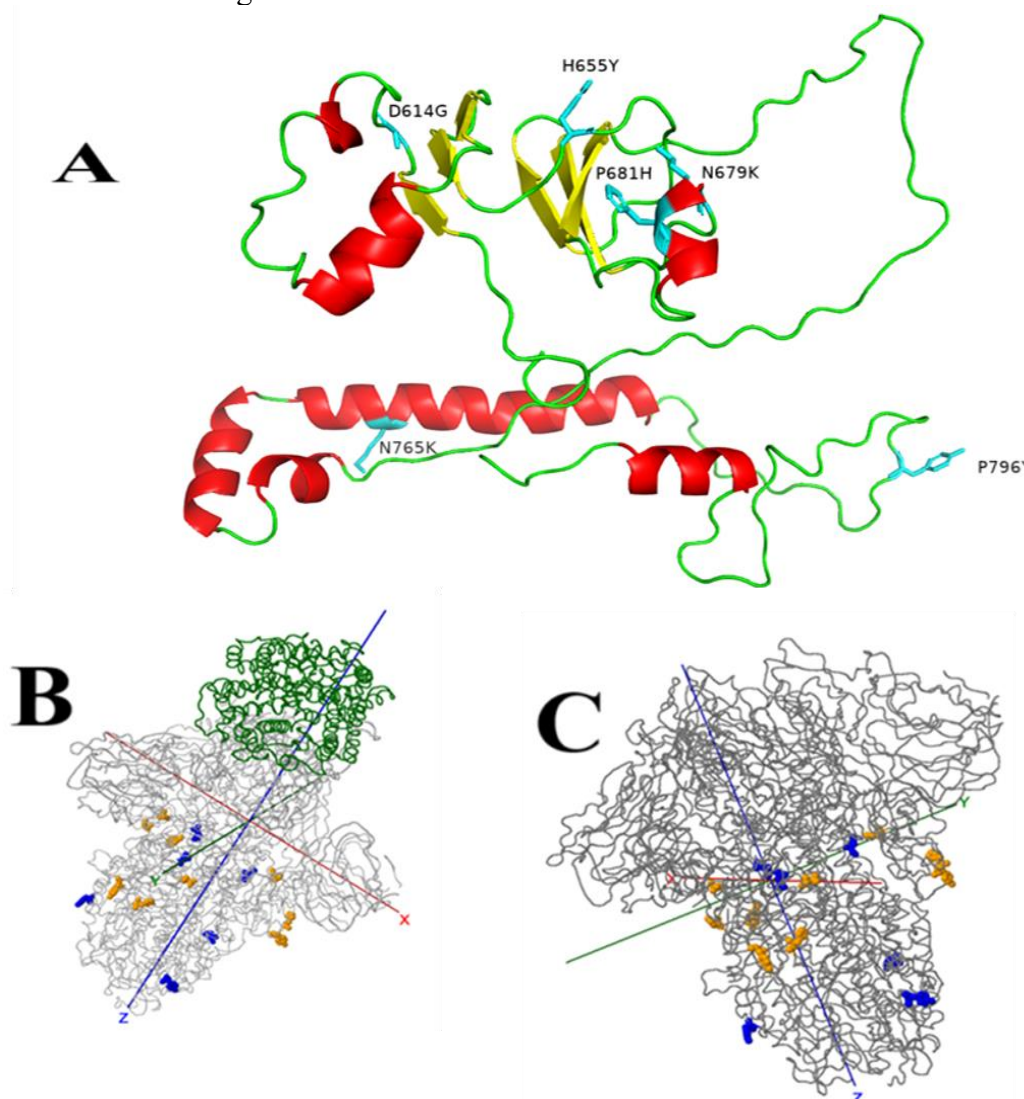


Figure 3: SARS-CoV-2 Spike partial sequence (based on c7tprc template), A: BA.1.1 green colour indicates loops, red represent alpha helix and yellow colour indicates beta sheets, cyan colour represents changed amino acids. B: Mutational representation complexed with ACE-2 receptor using CoVsurver, green colour indicates ACE-2 receptor, blue, red and green arrows indicate axes, orange and blue macromolecules indicate changed amino acids. C: Mutational visualization without ACE-2

To further investigate the durability of these mutations, an analysis was conducted using the CoVsurver tool from GISAID. This analysis concentrated on evaluating the prevalence of mutations in the fragments under study. The persistence analysis demonstrated that all identified mutations continue to be present in the currently circulating SARS-CoV-2 strains worldwide, as outlined in Table 1.

Table 1: Dominancy of investigated key mutations in the global SARS-CoV-2 lineages and sublineages.

Mutation ID	Description
D614G	D614G already occurred 15823548 times within 99.28% of all global reported data among 220 countries. The first viral strain with this change which collected in January 2020 in Argentina. Notably this change persists among current circulating strains.
H655Y	H655Y already occurred 8777174 times within 55.07% of all global reported data among 216 countries. The first viral strain with this change which collected in January 2020 in Thailand. Notably this change persists among current circulating strains.
N679K	N679K already occurred 8616147 times within 54.06% of all global reported data among 218 countries. The first viral strain with this change which collected in January 2020 in the USA. Notably this change persists among current circulating strains.
P681H	P681H already occurred 9862477 times within 61.88% of all global reported data among 215 countries. The first viral strain with this change which collected in January 2020 in the Niger. Notably this change persists among current circulating strains.
N764K	N764K already occurred 8192981 times within 51.40% of all global reported data among 214 countries. The first viral strain with this change which collected in January 2020 in the USA. Notably this change persists among current circulating strains.
D796Y	N764K already occurred 8504046 times within 53.35% of all global reported data among 214 countries. The first viral strain with this change which collected in January 2020 in the USA. Notably this change persists among current circulating strains.

Further investigation was conducted through data mining on February 11, 2024, involving, 1043 complete SARS-CoV-2 genomes from the EpiCoV™/GISAID repository. These genomes represent the entirety of SARS-CoV-2 sequences uploaded from Iraq during the COVID-19 pandemic, covering the period from June 6, 2020, to November 19, 2023. The study reveals a wide spectrum of SARS-CoV-2 lineages, with lineage AY.33 emerging as the most prevalent, accounting for 227 of the sequenced genomes. Close behind were the B.1.617.2 lineage, with 217 instances, and the BA.1.1 lineage, with 214 occurrences. The Alpha variant, B.1.1.7, identified in 132 cases, was previously labelled as a Variant of Concern due to its heightened transmissibility. Other variants, such as BA.1 and AY.122, were significantly present, with 36 and 34 sequences, respectively. Notably, B.1.428.1 was detected in 29 genomes, while BA.2 appeared in 12. The dataset also includes lineages of lower frequency, such as B.1.36.1, which was identified in 5 genomes, as well as several unique lineages like B.1.1.374, B.1.351, and B.1.533. Furthermore, recently emerging lineages like the XBB, EG.4, and FL.10, were also observed among the analysed genomic data. This diversity highlights the genetic variability of SARS-CoV-2 in Iraq and demonstrates the dynamic nature of the virus as it continues to evolve. The viral evolution is represented by a plethora of a presented in the curated genomes, including the D614G, which were dominant with 100% frequency. In the S1/S2 cleavage region, various mutations were observed across different viral lineages and sublineages, including Q675H, T676S, Q677H, N679K, S680T, P681H/R/L, A688S, S698L, A701V, and S704L. The S1/S2 cleavage site remained conserved across all lineages, with three amino acid changes observed in the fusion peptide and one in the internal fusion peptide. Hotspot regions were identified upstream and downstream of the S1/S2 Furin cleavage region, including two changes downstream of the S1/S2 cleavage region (T716I and N764K) and one change upstream (H655Y). Table 2 lists all changes within the investigated Iraqi genomes.

Table 2: Distribution of SARS-CoV-2 S1/S2 cleavage region and Fusion peptide amino acids changes across Iraqi reported lineages and sublineages.

Location	Mutation ID	Nucleotide Change	Lineage	Clade	Percentage
S1/S2 cleavage region	Q675H	CAG→CAT	B.1	GH	0.38%
S1/S2 cleavage region	T676S	ACT→TCT	AY.122	GK	0.10%
S1/S2 cleavage region	Q677H	CAG→CAT	B.1.617.2	GK	0.19%
S1/S2 cleavage region	N679K	AAT→AAG	XBB.2, XBB.1.9.1, XBB.1.5.4, XBB.1.5, XBB.1.42.1, XBB.1.37, XBB.1.22.1, XBB.1.16.2, XBB.1.16.11, XBB.1.16.1, XBB.1.16, XBB.1, GW.5, FY.5, FL.4.8, FL.4, FL.2, FL.10, EG.4.3, CV.1, BN.1, BA.5.2.16, BA.5.2.1, BA.5.2, BA.5.1.30, BA.2, BA.1.20, BA.1.17.2, BA.1.17, BA.1.14, BA.1.13, BA.1.1, BA.1, B.1.617.2, B.1.1.529	GRA, GR	31.45%
S1/S2 cleavage region	S680T	TCT→ACT	AY.122	GK	0.48%
S1/S2 cleavage region (PPAR locus)	P681R	CCT→CGT	B.1.617.2, AY.103, AY.33, AY.126, AY.106, AY.127, AY.122, AY.43, AY.121, AY.98, AY.5, AY.65, AY.46, B.1.533, B.1.1, AY.122	GK	49.57%
S1/S2 cleavage region (PPAR locus)	P681H	CCT→CAT	XBB.1.16.2, XBB.1.16.11, XBB.1.16, FL.4.8, XBB.1.22.1, FY.5, XBB.1.16.1, EG.4.3, XBB.1.42.1, XBB.1.9.1, FL.2, GW.5, FL.10, FL.4, XBB.1.5.4, XBB.1.5, XBB.2, XBB.1, BN.1, XBB.1.37, CV.1, BA.5.2, BA.2, BA.5.2.16, BA.5.1.30, BA.5.2.1, BA.1, BA.1.1, BA.1.17.2, BA.1.17, BA.1.20, BA.1.13, BA.1.14, B.1.1.7, B.1.1.529, B.1.36.1, FL.5, AY.5	GRA, GR, GRY, O, GH, GK	44.01%
S1/S2 cleavage region (PPAR locus)	P681L	CCT→CTT	B.1.617.2	GK	0.10%
S1/S2 cleavage region	A688S	GCT→TCT	B.1.1.7	GRY	0.10%
S1/S2 cleavage region	S698L	TCA→TTA	B.1.617.2	GK	0.10%
S1/S2 cleavage region	A701V	GAC→FTA	BA.1.17.2, B.1.351, B.1.1.7	GRA, GH	0.96%
S1/S2 cleavage region	S704L	TCA→TTA	XBB.1.16, B.1.1.7,	GRA,	0.58%

			B.1.617.2	GRY, GK	
Fusion Peptide	P792T	CCA→ACA	B.1.617.2	GK	0.10%
Fusion Peptide	D796Y	GAT→TAT	XBB.1.16.2, XBB.1.16.11, XBB.1.16, FL.4.8, XBB.1.22.1, FY.5, XBB.1.16.1, EG.4.3, XBB.1.42.1, XBB.1.9.1, FL.2, GW.5, FL.10, FL.4, XBB.1.5.4, XBB.1.5, XBB.2, XBB.1, BN.1, XBB.1.37, CV.1, BA.5.2, BA.2, BA.5.2.16, BA.5.1.30, BA.5.2.1, BA.1, BA.1.1, BA.1.17.2, BA.1.17, BA.1.20, BA.1.13, BA.1.14, B.1.1.7, B.1.1.529, B.1.36.1, FL.5, AY.5, AY.98	GRA, GK, G	30.97%
Fusion Peptide	F800Y	TTT→TAT	AY.122	GK	0.10%
Internal Fusion Peptide	N824D	AAC→GAC	AY.122	GK	0.10%
Downstream of S1/S2 Cleavage region	T716I	ACA→ATA	XBB.1.16, B.1.1.7, B.1.1, B.1.1.89, B.1.617.2, AY.122,	GRA, GRY, GK, GR, O,	12.37%
Downstream of S1/S2 Cleavage region	N764K	AAC→AAA	XBB.1.16.2, XBB.1.16.11, XBB.1.16, FL.4.8, XBB.1.22.1, FY.5, XBB.1.16.1, EG.4.3, XBB.1.42.1, XBB.1.9.1, FL.2, GW.5, FL.10, FL.4, XBB.1.5.4, XBB.1.5, XBB.2, BN.1, XBB.1.37, XBB.1, BA.5.2, BA.2, BA.5.2.16, BA.5.1.30, BA.5.2.1, BA.1, BA.1.1, BA.1.17.2, BA.1.17, BA.1.13, BA.1.20	GRA, GR	30.58%
Upstream of S1/S2 Cleavage region	H655Y	CAT→TAT	XBB.1.16.2, XBB.1.16.11, XBB.1.16, FL.4.8, XBB.1.22.1, FY.5, XBB.1.16.1 EG.4.3, XBB.1.42.1, XBB.1.9.1, FL.2, GW.5, FL.10, FL.4, XBB.1.5.4, XBB.1.5, XBB.2, XBB.1, BN.1, XBB.1.37, CV.1, BA.5.2, BA.2, BA.5.2.16, BA.5.1.30, BA.5.2.1, BA.1, BA.1.1, BA.1.17.2,	GRA, GR, G	31.38%

			BA.1.17. BA.1.13. BA.1.20, BA.1.14, B.1.1.529		
--	--	--	---	--	--

This combined analysis of public data and processed fragments reveals a tendency toward conserving the Furin cleavage site across different viral lineages. The high degree of conservation at this site highlights its essential function in facilitating efficient viral infection of host cells. However, a notable exception was observed in six genomes of the B.1.617.2 (Delta) variant, which exhibited a silent mutation at codon 682. Specifically, the nucleotide sequence CGG (coding for arginine) was changed to CGT, which still encodes for arginine. This silent mutation highlights the subtle genetic variability within the virus, even in regions that are generally conserved. In addition to these findings, the study focused on the S2' cleavage site, represented by codons 814-816 (KR↓S). The S2' cleavage site did not show any variability in the processed data. However, a rare mutation was noticed in the one Sanger-generated read, represented by R815T. This mutation changes the charged arginine to threonine. The R815T has been reported only 19 times on GISAID and was primarily harbored by Omicron sublineages, including BA.1.1, BA.1, BA.2.38, BA.4, BA.5.6, and BA.5.2, as well as Delta (B.1.617.2) and its sublineages AY.4, AY.9, and AY.122.

4. Discussion

Throughout the evolution of SARS-CoV-2, numerous mutations have been identified, significantly contributing to the emergence of various variants. The D614G mutation stands out for its role in enhancing viral transmissibility and is linked to distinct clinical manifestations such as impairments in taste and olfactory functions [24, 25]. In the upstream of S1/S2 cleavage region, the H655Y mutation, identified within all analyzed fragments, and within 31.38% of the investigated Iraqi genomes. This mutation is one of distinguishing features of the Omicron, which persists in subsequent sublineages, such results indicating its impact on enhanced entry mechanism and fusogenicity, increased transmissibility and enhanced cleavage of the S protein [26, 27]. Notably, other addressed changes include the P681, this site has emerged as a hotspot for mutations across different lineages, with P681R significantly enhancing Furin site cleavability in Delta variants. However, these changes were displaced by P681H. The evolution of the virus is characterized by a strong preference for mutations that may enhance viral fitness. In the Omicron variant and its related sublineages, the co-occurrence of N679K and P681H near the Furin scissile bond has been observed in analyzed sequences. Mutational pairing of P681H and N679K enhances the site's basic nature that represents a superior site targeted by Furin [28, 29]. Tendency toward shifting the proline arises from the fact that Furin's predilection for flexible substrates, with its catalytic site that favors such configurations for more effective cleavage. The inherent rigidity of proline, due to its unique cyclic structure, limits its presence at Furin cleavage sites, underscoring a selection for mutations that confer greater structural flexibility, thereby enhancing Furin's access to crucial cleavage sites [30, 31]. Mutations at the SARS-CoV-2 spike protein's P681 site have been reported on GISAID database for nearly every amino acid, albeit at significantly lower frequencies compared to P681H and P681R mutations. Notably, mutations to methionine (P681M) and glutamic acid (P681E) at this site have not been reported. Comprehensive genomic analysis of Iraqi viral isolates revealed the conservative status of the polybasic motif 682RRAR↓S686, this indicates continuing dependency of SARS-CoV-2 on Furin and aligning with other reported results [28, 32]. Furthermore, the S2' cleavage site (KR↓S) show a similar conservative attitude. The conserved nature of Furin and S2' cleavage sites coupled with their vital roles in viral entry can shed the lights on their importance in maintaining SARS-CoV-2 infectivity and enhanced adaptability toward different proteases environments in host cells [33]. While mutations such as A688S, S698L,

A701V, and S704L were detected at lower frequencies in the current study, their impact on vaccine efficacy cannot be overlooked. Particularly, the dual mutations of S698L and A701V have been implicated in diminishing vaccine-induced antibody sensitivity [34, 35]. Meanwhile, S704L possesses a deleterious effect which reduces Omicron BA.2 Spike-mediated infection [36], however, S704L still reported on GISAID among newer circulating SARS-CoV-2 variants at low frequency. In the context of fusion peptide, current results showed occurrence of 796Y among 30.97% of Iraqi genomes, specifically among Omicron and subsequent sublineages and were prevalent in current investigated fragments, D796Y, is situated near the beginning of the S2 subunit, just before the fusion peptide. This mutation could potentially influence the spike proteins interaction with the TMPRSS2 enzyme. This mutation could alter the presentation of nearby glycan epitopes, impacting the neutralization capacity of antibodies targeting the spike protein's S1 region [37]. Furthermore, the emergence of the T716I mutation in recent variants like XBB.1.16, associated with enhanced transmissibility and potential immune evasion, [38, 39], alongside the high-frequency circulation of N764K in Omicron dissented variants, introduces new cleavage sites recognized by SKI-1/S1P. This alteration may hinder the exposure of the internal fusion peptide, critical for viral entry and syncytia formation. The specific expression pattern of SKI-1/S1P, absent in lung tissue but present in the bronchus and nasopharynx, suggests a nuanced shift in the virus's tissue targeting, driven by these mutations [40].

Conclusion

This investigation into SARS-CoV-2 spike protein mutations, particularly within Iraqi isolates, reveals crucial adaptations and variability tendency toward boosting viral fitness. Key mutations across different regions of the S1/S2 Furin cleavage site, specifically the hot spot locus of P681 illuminate the selective pressures that shape viral transmissibility and immune evasion in Iraq and around the global. Despite that, the tendency towards maintaining cleavage sites can offer strategic point of intervention.

Data Availability

All GISAID accession numbers of the 14 Fasta files are indicated in supplementary file 1, and the analysis results of all 14 files are depicted supplementary file 2.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Funding

The authors didn't receive any specific fund for this work.

Ethical Statement

This work was approved by the Ethics committee at College of Science, University of Baghdad under the reference number CSEC/0222/0155 on January 28, 2022.

Acknowledgement

Authors would like to thank all researchers and contributors that make the genomic data available on GISAID. Furthermore, we would like to acknowledge the Central Public Health Laboratory/National Influenza Center in Baghdad for their support.

References

- [1] M. Pal, G. Berhanu, C. Desalegn, and V. Kandi, "Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2): An Update," *Cureus*, vol. 12, no. 3, p. e7423, 2020, doi: 10.7759/cureus.7423.
- [2] World Health Organization, "COVID-19 Dashboard," 2023. [Online]. Available: <https://data.who.int/dashboards/covid19/cases>.
- [3] A. Gheware, A. Ray, D. Rana, P. Bajpai, A. Nambirajan, S. Arulselvi, P. Mathur, A. Trikha, S. Arava, P. Das, A. R. Mridha, G. Singh, M. Soneja, N. Nischal, S. Lalwani, N. Wig, C. Sarkar, and D. Jain, "ACE2 protein expression in lung tissues of severe COVID-19 infection," *Scientific Reports*, vol. 12, no. 1, p. 4058, 2022, doi: 10.1038/s41598-022-07918-6.
- [4] H. Xu, L. Zhong, J. Deng, J. Peng, H. Dan, X. Zeng, T. Li, and Q. Chen, "High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa," *International Journal of Oral Science*, vol. 12, no. 1, p. 8, 2020, doi: 10.1038/s41368-020-0074-x.
- [5] P. Zhou, X. L. Yang, X. G. Wang, B. Hu, L. Zhang, W. Zhang, H. R. Si, Y. Zhu, B. Li, C. L. Huang, H. D. Chen, J. Chen, Y. Luo, H. Guo, R. D. Jiang, M. Q. Liu, Y. Chen, X. R. Shen, X. Wang, X. S. Zheng, K. Zhao, Q. J. Chen, F. Deng, L. L. Liu, B. Yan, F. X. Zhan, Y. Y. Wang, G. F. Xiao, and Z. L. Shi, "A pneumonia outbreak associated with a new coronavirus of probable bat origin," *Nature*, vol. 579, no. 7798, pp. 270-273, 2020, doi: 10.1038/s41586-020-2012-7.
- [6] F. Li, "Structure, Function, and Evolution of Coronavirus Spike Proteins," *Annual Review of Virology*, vol. 3, no. 1, pp. 237-261, 2016, doi: 10.1146/annurev-virology-110615-042301.
- [7] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N. H. Wu, A. Nitsche, M. A. Müller, C. Drosten, and S. Pöhlmann, "SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor," *Cell*, vol. 181, no. 2, pp. 271-280.e8, 2020, doi: 10.1016/j.cell.2020.02.052.
- [8] R. Essalmani, J. Jain, D. Susan-Resiga, U. Andréo, A. Evagelidis, R. M. Derbali, D. N. Huynh, F. Dallaire, M. Laporte, A. Delpal, P. Sutto-Ortiz, B. Coutard, C. Mapa, K. Wilcoxon, E. Decroly, T. Nq Pham, É. A. Cohen, and N. G. Seidah, "Distinctive Roles of Furin and TMPRSS2 in SARS-CoV-2 Infectivity," *Journal of Virology*, vol. 96, no. 8, p. e0012822, 2022, doi: 10.1128/jvi.00128-22.
- [9] M. Hoffmann, H. Kleine-Weber, and S. Pöhlmann, "A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells," *Molecular Cell*, vol. 78, no. 4, pp. 779-784.e5, 2020, doi: 10.1016/j.molcel.2020.04.022.
- [10] A. G. Wrobel, D. J. Benton, P. Xu, C. Roustan, S. R. Martin, P. B. Rosenthal, J. J. Skehel, and S. J. Gamblin, "SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects," *Nature Structural and Molecular Biology*, vol. 27, no. 8, pp. 763-767, 2020, doi: 10.1038/s41594-020-0468-7.
- [11] Y. A. Chan and S. H. Zhan, "The Emergence of the Spike Furin Cleavage Site in SARS-CoV-2," *Molecular Biology and Evolution*, vol. 39, no. 1, p. msab327, 2022, doi: 10.1093/molbev/msab327.
- [12] T. P. Peacock, D. H. Goldhill, J. Zhou, L. Baillon, R. Frise, O. C. Swann, R. Kugathasan, R. Penn, J. C. Brown, R. Y. Sanchez-David, L. Braga, M. K. Williamson, J. A. Hassard, E. Staller, B. Hanley, M. Osborn, M. Giacca, A. D. Davidson, D. A. Matthews, and W. S. Barclay, "The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets," *Nature Microbiology*, vol. 6, no. 7, pp. 899-909, 2021, doi: 10.1038/s41564-021-00908-w.
- [13] M. G. Hossain, Y. D. Tang, S. Akter, and C. Zheng, "Roles of the polybasic furin cleavage site of spike protein in SARS-CoV-2 replication, pathogenesis, and host immune responses and vaccination," *Journal of Medical Virology*, vol. 94, no. 5, pp. 1815-1820, 2022, doi: 10.1002/jmv.27539.
- [14] M. Sadarangani, A. Marchant, and T. R. Kollmann, "Immunological mechanisms of vaccine-induced protection against COVID-19 in humans," *Nature Reviews Immunology*, vol. 21, no. 8, pp. 475-484, 2021, doi: 10.1038/s41577-021-00578-z.
- [15] P. A. Koenig and F. I. Schmidt, "Spike D614G - A Candidate Vaccine Antigen Against Covid-19," *The New England Journal of Medicine*, vol. 384, no. 24, pp. 2349-2351, 2021, doi: 10.1056/NEJMcibr2106054.

- [16] A. C. Walls, Y. J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veessler, "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein," *Cell*, vol. 181, no. 2, pp. 281-292.e6, 2020, doi: 10.1016/j.cell.2020.02.058.
- [17] J. Shang, Y. Wan, C. Luo, G. Ye, Q. Geng, A. Auerbach, and F. Li, "Cell entry mechanisms of SARS-CoV-2," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 21, pp. 11727-11734, 2020, doi: 10.1073/pnas.2003138117.
- [18] T. A. A. Alhussien and H. Y. Fadhil, "Analysis of Mutations in Conserved and Susceptible Regions Across the Whole Genome Sequencing Analysis for SARS-CoV-2 in Iraqi Patients," *Iraqi Journal of Science*, vol. 64, no. 1, pp. 56–64, 2023, doi: 10.24996/ij.s.2023.64.1.6.
- [19] J. J. Ghazzi, H. Y. Fadhil, and I. M. Aufi, "Impact of SARS-COV-2 Variants on the Infection Severity among Iraqi Patients," *Iraqi Journal of Science*, vol. 64, no. 7, pp. 3263–3272, 2023, doi: 10.24996/ij.s.2023.64.7.7.
- [20] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis," *Nature Protocols*, vol. 10, no. 6, pp. 845-858, 2015, doi: 10.1038/nprot.2015.053.
- [21] S. Elbe and G. Buckland-Merrett, "Data, disease and diplomacy: GISAID's innovative contribution to global health," *Global Challenges*, vol. 1, no. 1, pp. 33-46, 2017, doi: 10.1002/gch2.1018.
- [22] C. Khare, Gurry, L. Freitas, M. B. Schultz, G. Bach, A. Diallo, N. Akite, J. Ho, R. T. Lee, W. Yeo, G. C. Curation Team, and S. Maurer-Stroh, "GISAID's Role in Pandemic Response," *China CDC Weekly*, vol. 3, no. 49, pp. 1049–1051, 2021, doi: 10.46234/ccdcw2021.255.
- [23] Y. Shu and J. McCauley, "GISAID: Global initiative on sharing all influenza data - from vision to reality," *Eurosurveillance*, vol. 22, no. 13, p. 30494, 2017, doi: 10.2807/1560-7917.ES.2017.22.13.30494.
- [24] R. R. Butowt, K. Bilinska, and C. S. Von Bartheld, "Chemosensory Dysfunction in COVID-19: Integration of Genetic and Epidemiological Data Points to D614G Spike Protein Variant as a Contributing Factor," *ACS Chemical Neuroscience*, vol. 11, no. 20, pp. 3180–3184, 2020, doi: 10.1021/acschemneuro.0c00596.
- [25] F. Hajizadeh, S. Khanizadeh, H. Khodadadi, Y. Mokhayeri, M. Ajourloo, A. Malekshahi, and E. Heydari, "SARS-COV-2 RBD (Receptor binding domain) mutations and variants (A sectional-analytical study)," *Microbial Pathogenesis*, vol. 168, p. 105595, 2022, doi: 10.1016/j.micpath.2022.105595.
- [26] A. Escalera, A. S. Gonzalez-Reiche, S. Aslam, I. Mena, M. Laporte, R. L. Pearl, A. Fossati, R. Rathnasinghe, H. Alshammary, A. van de Guchte, K. Farrugia, Y. Qin, M. Bouhaddou, T. Kehrer, L. Zuliani-Alvarez, D. A. Meekins, V. Balaraman, C. McDowell, J. A. Richt, G. Bajic, E. M. Sordillo, M. Dejoze, T. P. Zwaka, N. J. Krogan, V. Simon, R. A. Albrecht, H. van Bakel, A. García-Sastre, and T. Aydiello, "Mutations in SARS-CoV-2 variants of concern link to increased spike cleavage and virus transmission," *Cell Host & Microbe*, vol. 30, no. 3, pp. 373-387.e7, 2022, doi: 10.1016/j.chom.2022.01.006.
- [27] M. Yamamoto, K. Tomita, Y. Hirayama, J. Inoue, Y. Kawaguchi, and J. Gohda, "SARS-CoV-2 Omicron spike H655Y mutation is responsible for enhancement of the endosomal entry pathway and reduction of cell surface entry pathways," *bioRxiv*, 2022, doi: 10.1101/2022.03.21.485084.
- [28] L. Cassari, A. Pavan, G. Zoia, M. Chinellato, E. Zeni, A. Grinzato, S. Rothenberger, L. Cendron, M. Dettin, and A. Pasquato, "SARS-CoV-2 S Mutations: A Lesson from the Viral World to Understand How Human Furin Works," *International Journal of Molecular Sciences*, vol. 24, no. 5, p. 4791, 2023, doi: 10.3390/ijms24054791.
- [29] B. Lubinski, J. A. Jaimes, and G. R. Whittaker, "Intrinsic furin-mediated cleavability of the spike S1/S2 site from SARS-CoV-2 variant B.1.1.529 (Omicron)," *bioRxiv*, 2022, doi: 10.1101/2022.04.20.488969.
- [30] S. O. Dahms, J. W. Creemers, Y. Schaub, G. P. Bourenkov, T. Zögg, H. Brandstetter, and M. E. Than, "The structure of a furin-antibody complex explains non-competitive inhibition by steric exclusion of substrate conformers," *Scientific Reports*, vol. 6, p. 34303, 2016, doi: 10.1038/srep34303.
- [31] J. R. Casey, S. Grinstein, and J. Orłowski, "Sensors and regulators of intracellular pH," *Nature Reviews Molecular Cell Biology*, vol. 11, no. 1, pp. 50-61, 2010, doi: 10.1038/nrm2820.

- [32] B. A. Johnson, X. Xie, A. L. Bailey, B. Kalveram, K. G. Lokugamage, A. Muruato, J. Zou, X. Zhang, T. Juelich, J. K. Smith, L. Zhang, N. Bopp, C. Schindewolf, M. Vu, A. Vanderheiden, E. S. Winkler, D. Swetnam, J. A. Plante, P. Aguilar, K. S. Plante, V. Popov, B. Lee, S. C. Weaver, M. S. Suthar, A. L. Routh, P. Ren, Z. Ku, A. An, K. Debbink, M. S. Diamond, P. Y. Shi, A. N. Freiberg, and V. D. Menachery, "Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis," *Nature*, vol. 591, no. 7849, pp. 293-299, 2021, doi: 10.1038/s41586-021-03237-4.
- [33] M. Lavie, J. Dubuisson, and S. Belouzard, "SARS-CoV-2 Spike Furin Cleavage Site and S2' Basic Residues Modulate the Entry Process in a Host Cell-Dependent Manner", *Journal of Virology*, vol. 96, no. 13, pp. e0047422, 2022. doi: 10.1128/jvi.00474-22.
- [34] K. Subramoney, N. Mtileni, J. Giandhari, Y. Naidoo, Y. Ramphal, S. Pillay, U. Ramphal, A. Maharaj, D. Tshiabula, H. Tegally, E. Wilkinson, T. de Oliveira, B. C. Fielding, and F. K. Treurnicht, "Molecular Epidemiology of SARS-CoV-2 during Five COVID-19 Waves and the Significance of Low-Frequency Lineages," *Viruses*, vol. 15, no. 5, p. 1194, 2023, doi: 10.3390/v15051194..
- [35] Q. Wang, S. B. Ye, Z. J. Zhou, A. L. Song, X. Zhu, J. M. Peng, R. M. Liang, C. H. Yang, X. W. Yu, X. Huang, J. Yu, Y. Qiu, and X. Y. Ge, "Key mutations in the spike protein of SARS-CoV-2 affecting neutralization resistance and viral internalization," *Journal of Medical Virology*, vol. 95, no. 1, p. e28407, 2023, doi: 10.1002/jmv.28407.
- [36] C. Pastorio, S. Noettger, R. Nchioua, F. Zech, K. M. J. Sparrer, and F. Kirchhoff, "Impact of mutations defining SARS-CoV-2 Omicron subvariants BA.2.12.1 and BA.4/5 on Spike function and neutralization," *iScience*, vol. 26, no. 11, p. 108299, 2023, doi: 10.1016/j.isci.2023.108299.
- [37] E. A. Elko, H. L. Mead, G. A. Nelson, J. A. Zaia, J. T. Ladner, and J. A. Altin, "Recurrent SARS-CoV-2 mutations at Spike D796 evade antibodies from pre-Omicron convalescent and vaccinated subjects," *Microbiology Spectrum*, vol. 12, no. 2, p. e0329123, 2024, doi: 10.1128/spectrum.03291-23.
- [38] I. Lazarevic, V. Pravica, D. Miljanovic, and M. Cupic, "Immune Evasion of SARS-CoV-2 Emerging Variants: What Have We Learnt So Far?" *Viruses*, vol. 13, no. 7, p. 1192, 2021, doi: 10.3390/v13071192.
- [39] S. Eslami, M. C. Glassy, and S. Ghafouri-Fard, "A comprehensive overview of identified mutations in SARS CoV-2 spike glycoprotein among Iranian patients," *Gene*, vol. 813, p. 146113, 2022, doi: 10.1016/j.gene.2021.146113.
- [40] H. Maaroufi, "The N764K and N856K mutations in SARS-CoV-2 Omicron S protein generate potential cleavage sites for SKI-1/S1P protease," *bioRxiv*, 2022, doi: 10.1101/2022.01.21.477298.