JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 06 Issue 02
2024

Research Article
https://doi.org/10.56979/602/2024

Insilico Structural and Functional Analysis of Covid-19 Genomic Sequences

Moeil Waseem[1*], Ahmad Naeem[1], Naeem Aslam[1], Kamran Abid[1], Muhammad Tariq Pervez[2], Abdul Mannan[3,4], and Abdul Majid Soomro[5]

[1]Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Punjab, Pakistan.
[2]Department of Bioinformatics and Computational Biology, Virtual University of Pakistan.
[3]Department of Biomedical Engineering, NFC Institute of Engineering and Technology Multan, Punjab, Pakistan.
[4]Department of Electrical Engineering, NFC Institute of Engineering and Technology Multan, Punjab, Pakistan.
[5]Department of Computer Science, National College of Business Administration and Economics, Multan, Pakistan.
[*]Corresponding Author: Moeil Waseem. Email: moeilwasim456@gmail.com

Received: November 01, 2023 **Accepted:** February 04, 2024 **Published:** March 01, 2024

Abstract: The Covid-19 outbreak classified as a pandemic by World Health Organization, is caused by the severe acute respiratory (SARS-CoV-2) syndrome coronavirus 2. The surface glycoprotein or spike (S) protein is one of the four primary structural proteins that SARS-CoV-2 encodes, and it has generally been the focus of research for potential novel vaccines. A number of bioinformatics methods were used in the suggested in silico study to examine the mutational, physicochemical, phylogenetic, structural, and immunological characteristics of spike protein. 81 SARS-CoV-2 sequences from different regions of Pakistan were retrieved from the NCBI database. From all retrieved sequences, the sequences of spike protein were screened out for further analysis. blastp suit available on NCBI web server was employed to align the sequences with reference sequence to analyze mutations in Spike region. SIFT, MUpro, SNAP2, PolyPhen2, PhD-SNP, and I-Mutant were utilized to predict deleterious mutations. Through ProtParam, physiochemical characteristics were analyzed. MEGA11 was used for the phylogenetic analysis, SOPMA and I-TASSER server employed for the investigation of protein secondary and tertiary 3D structures. The position of B-Cell epitopes was predicted using IEDB ElliPro tool. The result of mutation analysis showed 67 mutations in spike region out of which 51 mutations were predicted to be deleterious and affects proteins' functional stability. Physiochemical analysis showed that this protein was stable in nature. Structural analysis showed that this protein contains random coil as major structure with stable distribution of alpha helix and beta strands. Model 3 was selected for 3D structure build by I-TASSER and verified by 3 tools. ElliPro predicted 21 candidates for epitopes out of which 6 were predicted as potential epitopes by VaxiJen2.0 and AllerTop v2.0. The proposed study makes it easier to understand spike protein in depth, which will be helpful in future research on the infections of SARS-CoV-2 and the construction of diagnostics, antiviral drugs and treatments.

Keywords: SARS-CoV-2; Spike, Mutation; B-Cell Epitopes; Insilico Study; Bioinformatics.

1. Introduction

Humans and birds can get sick from a broad group of animal viruses called coronaviruses, which belong to the Coronaviridae. These viruses have a long history and are still connected to previous viral outbreaks. The IBV known as infectious bronchitis virus, which causes infections in hens, was the first coronavirus identified in 1930 [1]. Two further coronaviruses in animals, MHV known as mouse hepatitis virus or transmittable gastroenteritis virus (TGEV), found as the following years. 229E or the coronavirus in human OC43 were first covid in human to be isolated from humans in 1960 [2]. The broad family of animal viruses known as the Coronaviridae was further divided into beta, alpha, gamma and delta [3]. According to scientific research and the literature that is currently available, all coronaviruses that appeared as a result of a viral outbreak and caused sickness in people belong to the beta group.

Coronaviruses that typically infect birds fall under the gamma and delta classification categories. Although it has been suggested that coronaviruses are naturally found in wild animals, numerous coronavirus species are also found in animals. When beta covid infect people, they mainly cause the higher respiratory tract infections and the illnesses in cases of chronic illness. When coronaviruses infect a human, the term "severe acute respiratory syndrome" (SARS) is frequently used [4]. These coronaviruses have a good attraction to the hACE2 receptor, which allows them to enter the human respiratory tract. The severity of illnesses is also determined by the alteration of hACE2 in grown-ups, kids and elders as well. Other tissues, the gastrointestinal tract, kidney, liver and heart, as well as respiratory airways, show ACE2 expression, exposing them to SARS-CoV infection. The type of virus, such as wild type versus mutant, also affects the rate and severity of infection [1].

In late December 2019, a pneumonia outbreak with no known reasons occurred in Wuhan, Hubei state, China. The outbreak had considerably spread until the end of January 2020, infecting 106 people in 19 different nations and killed 213 individuals in China, where it had infected 9720 people [4]. A few days later, several autonomous laboratories resolute current covid is the major reason of pneumonia [5]. The World Health Organization provide the pandemic names extreme SARS-CoV-2 known as severe acute respiratory syndrome coronavirus type 2 or Covid-19 respectively. SARS-CoV-2 is a highly contagious and virulent coronavirus and is the seventh coronavirus discovered to infect humans [6]. It is an encapsulated, single-stranded, positive-sense RNA virus that is a member of the Beta coronavirus genus [7].

The etiology of COVID-19 is complicated, although evidence suggests that its primary method is identical to those of MERS-CoV and SARS-CoV. [8]. Coronaviruses mediate viral entrance into target cells through the spike protein. Angiotensin converting enzyme 2 is a biological receptor that the S1 unit of S protein binds to in order to cause this entry (ACE2). The amino acid homology between SARS and SARS-2 is considerable (> 70%). [9]. The interaction of viral proteins with the receptors of cell membrane is among the most crucial stages in the pathological process of viruses [10]. The upper respiratory tract's main passageways, particularly the mucosa of the nose and larynx, are likely where the virus entered the body [11]. Although other organs that produce type 2 transmembrane serine protease (TMPRSS2) and ACE2 receptor protein are also attacked and invaded by viruses, the lungs are their primary target for entry through the respiratory system. Infected host cells emit an excessive amount of proinflammatory cytokines, which results in a cytokine storm [12].

Patients infected with Covid-19 present a variety of symptoms that make it difficult to differentiate them from those of other respiratory disorders. SARS-CoV-2 infection is divided from mild to moderate and severe to critical categories depending on how severe the symptoms are [13]. Most diseases, which may appear after one week of virus exposure, include fever, dry cough, shortness of breath, chills, headaches, muscle aches, and a loss of taste or smell [14]. The most common Covid-19 adverse effects recorded were pneumonia, heart damage, liver and kidney failure, severe fever, and respiratory distress [15]. According to recent figures, underlying disorders accounted for around 50% of COVID-19 deaths, with hypertensive disorder (46%) having higher occurrence. The next most prevalent underlying disorders were (26%) diabetes mellitus, (21%) cardiovascular disorder, (11%) cancer, (8%) obstructive pulmonary disorder, (7%) renal disorder, and (3%) liver disorder [16].

The study analyzes spike protein sequences in Pakistan and other countries, identifying 67 and 92 mutations. Using SIFT, DUET, SNAP2, DynaMut, PhD-SNP, and I-Mutant tools, the study predicts 55 and 62 deleterious mutations. Spike protein is characterized as acidic, thermally stable, and hydrophilic. The study predicts 6 potential B-cell epitopes using I-TASSER, providing better results with 3D structure predictions.

The contribution of this study is this investigation have allowed us to fully understand spike protein. There are 51 predicted mutations that result in significant phenotypic damage and might alter the structure and functional behavior of the protein. The spike protein is a very potent epitope on the surface of SARS-CoV-2 with a number of qualities that make it a strong contender for the creation of novel diagnostic tools, antiviral drugs, and therapeutic interventions.

2. Related Work

Over the past few decades, bioinformatics has emerged as a popular tool for investigating bacterial or viral genomes, predicting the function and structure of proteins, and developing new vaccines [18] [19].

The insilico study conducted in [17] was aim to examine few immunological or the structural features of the spike protein and sequences of covid-19 recovered from different countries listed in NCBI GenBank. Sequences were translated and aligned using the CLC Sequence Viewer, and numerous tools were used to predict physiochemical properties, phylogenetic analysis and B-Cell epitopes. Phosphorylation, glycosylation, and covalent linkages were designated as modification sites. All sequences' primary and secondary structures were also calculated.

It is necessary to thoroughly analyze the 1st wave of the SARS-CoV-2 or find the different mutations that have a major impact on viral fitness. For viruses, mutation is to be advantageous and help in viral adaptation to pathogenicity, signal transduction, and treatment resistance [20]. Numerous mutations impact the virulence of different pathogenic viruses and lead to medication resistance. As a result, they have a significant effect on human health. It will be crucial to determine how these mutations affect human pathogenesis and transmissibility because their functional and structural effects are unknown [21]. Crucial aspects of Covid-19, with developing mutations worldwide, are revealed through analysis of the sequence information accessible at NCBI and GISAID [23]. As a result, we continued to concentrate on SARS-CoV-2 mutations that affected the RdRp, 3C-like protease (3CLpro), and spike proteins in an effort to gauge the spread of new viral variants across countries as well as the actual functional and structural effects of these mutations on the pathogenicity of SARS-CoV-2. The development of vaccines and antiviral medications is thought to primarily focus on these viral proteins. By analysing sequence data gnomically, it is possible to gain a more thorough understanding of pathogenic mutations and their evolution, which can help direct various experiments. Researchers apply various bioinformatics tools to uncover buried clinical and molecular information because of the availability of such detailed data [24]. The easily available data must be used to identify harmful mutations and their pathogenic variations, and their effects on the molecular level must be further studied. With the help of insilico methods, it is possible to select various variants economically and research some effects about particular changes [25]. In order to gain understanding into the pathogenic landscape of different mutations on specified viral proteins, several in-silico methodologies have been applied in this study to collect all of the genomic data of the 1st wave of Pandemic that is currently available [22]. Understanding and predicting different pathogenic mutations in the first wave of SARS-CoV-2 RdRp, 3C-like protease (3CLpro), and spike proteins was the study's main objective.

SARS-CoV [26], MERS-CoV [27], and SARS-CoV-2 have all infected humans and caused fatal pneumonia by crossing the species barrier. The replication enzyme coding region (ORF 1a and ORF 1b), the coronavirus-specific Envelop protein gene, the Membrane protein gene, the Spike protein gene, and the Nucleocapsid protein gene are all included in the SARS-CoV-2 genome's six basic open reading frames (ORFs) [28]. The production of the structurally complete virion requires the structural proteins membrane protein, spike protein, nucleocapsid protein, and envelope protein [29] [30] [31] [32]. Spike glycoprotein directs coronavirus entry into host cells. A replication enzyme made up of 16 highly conserved non-structural proteins (nsp1-6) is encoded by the coronavirus ORF 1a and 1b. Main protease (Mpro or 3CLpro), one of the important nsps produced by ORF 1a and 1b, is essential for breaking down polyproteins and controlling coronavirus replication [33] [34]. RNA-dependent RNA polymerase (RdRp), also known as nsp12, is another essential replicase that uses the viral genomic RNA template to catalyse RNA replication [35]. RNA viruses have a mutation rate that is up up to 1 million times greater than that of their host, which increases their pathogenicity and capacity for evolution (creation of new taxa) [36]. The replication of coronavirus is more error-prone than other RNA viruses, and its predicted genetic variation is 4x10-4 nucleotide substitutions/site/year [37] From one country to the next, the mortality and SARS-CoV-2 induced illness spread rates vary. One of the primary factors influencing the pace of virus transmission and lethality, among other contributing factors, is assumed to be mutations within the SARS-CoV-2 strains. The goal of this study was to get more information about SARS-CoV-2 genomes from distinct COVID-19-infected geographical areas. Genome analysis of the SARS-CoV-2 strains from 13 different countries showed a substantial range of mutations in the critical structural proteins. The impact of these changes on the pathogenicity, replication, and entry of viral particles has never before been thoroughly discussed. This research provides information on the origins of these mutations in the important structural and nsp genes represented by the SARS-CoV-2 genome in different countries. The impact of mutations on the behavior of Mpro was investigated in this case using molecular simulation and other in silico methods. The results

of this study offer a hint for the creation of prospective COVID19 vaccine candidates or treatment designs in the future [38].

The covid-19 pandemic has made researchers from across the world to do research on various aspects of the virus. Many literatures reporting Covid-19 have already been published. To the best of our knowledge only limited literatures were available that have reported Covid-19 genome analysis from Pakistan. The present study suggests the analysis of locally transmitted Covid-19 genome sequences from various regions within Pakistan. The required data was available on NCBI data base. The analysis of Covid-19 genome has revealed the fact that surface spike protein is one of the 4 main structural proteins that facilitates the entry of virus into the host cells. So, an in-depth analysis of spike protein is required to understand the Covid-19 pathophysiology, modifications and for the development of new medications and treatments.

3. Materials and Methods

The proposed methodology for our study is given in Figure 1.

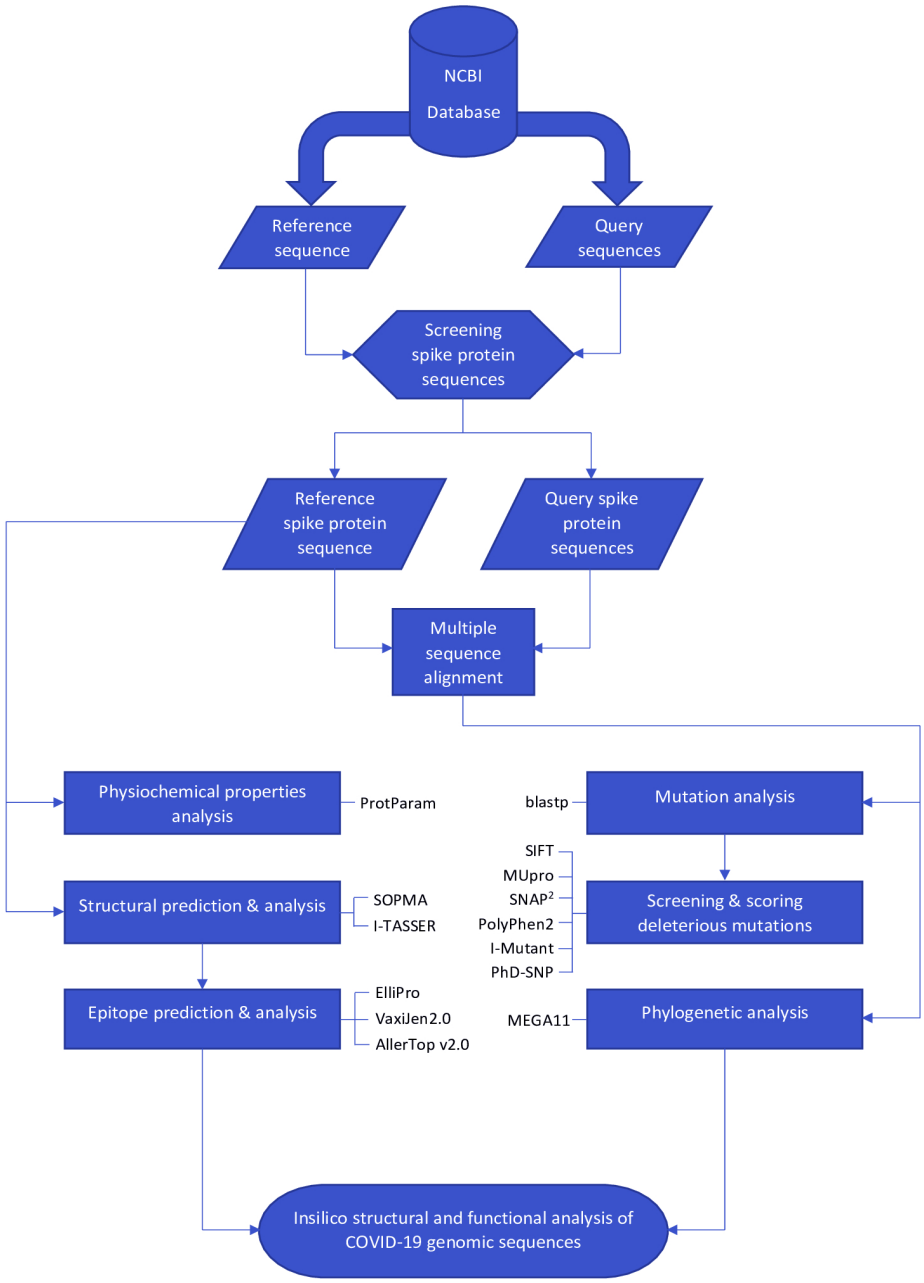

Figure 1. Flow Diagram of Proposed Methodology.

3.1. Data Set

We downloaded genome sequences of locally transmitted SARS-CoV-2 from different regions of Pakistan registered in NCBI (National Center for Biotechnology Information) virus database (from 04-03-2020 to 29-09-2022). The genome sequence with accession number NC_045512 was used as refence sequence. From all download genome sequences, the sequences of spike protein were screened out. The accession numbers along with protein_ids that were utilized in this Insilco study were listed in table 1.

Table 1. All 81 Genome Sequences with Accession Numbers Along with Associated Surface Glycoprotein or Spike (S) Protein_id and NC_045512 as Reference Sequence.

Location	Accession	Protein_id
Ref_Seq	NC_045512	YP_009724390
Federal Capital	MW421988, MW421989	QQH15738, QQH15750
	MW421990, MW421991	QQH15762, QQH15774
	MW421992, MW421994	QQH15786, QQH15810
	MW421995, MW421996	QQH15822, QQH15834
	MW421997, MW421998	QQH15846, QQH15858
	MW421999, MW422000	QQH15870, QQH15882
	MW422001, MW422002	QQH15894, QQH15906
	MW422003	QQH15918
Punjab	MW421982, MW421984	QQH15666, QQH15690
	MW421985, MW421986	QQH15702, QQH15714
	MW421987, MW421993	QQH15726, QQH15798
	MW422005, MW422007	QQH15942, QQH15966
	MW422013, MW422020	QQH16038, QQH16122
	MW422025, MW422030	QQH16182, QQH16242
	MW422032	QQH16266
Karachi	OP550240, OP550241 OP550242,	UXN86996, UXN87008
	OP550243 OP550244, OP550245	UXN87020, UXN87032
	OP550246, OP550247	UXN87044, UXN87056
	OP024248, OP024249	UXN87068, UXN87080
	OP024250, OP024261	UTS98465, UTS98477
	OP024262, OP024263	UTS98489, UTS98621
	OP024264, OP024265	UTS98633, UTS98645
		UTS98657, UTS98669
Islamabad	ON115820, ON115821	UNY81286, UNY81289
	ON115822, MW960273	UNY81292, QTY34199
	MW960274, MW960275	QTY34211, QTY34223
	MW960276, MW960277	QTY34235, QTY34247
	MW960278, MW960279	QTY34259, QTY34271
	MW960280, MW960281	QTY34283, QTY34295
	MW960282, MW960283	QTY34307, QTY34319
	MW960284, MW960285	QTY34331, QTY34343
	MW960286, MW960287	QTY34354, QTY34365
Lahore	MW947211, MW947212	QTX93634, QTX93635
KPK	MW421983, MW422034	QQH15678, QQH16290

	MW422035, MW422070	QQH16302, QQH16722
	MW422073, MW422077	QQH16758, QQH16806
	MW422081, MW422088	QQH16854, QQH16938
	MW422094, MW422116	QQH17010, QQH17274
	MT262993	QIS60276
Gilgit	MW422066, MT240479	QQH16674, QIQ22760
AJK	MW422071, MW422086	QQH16734, QQH16914
Rawalpindi	MW422080	QQH16842

3.2. Multiple Sequence Analysis

BLAST (Basic Local Alignment Search Tool) is a tool form NCBI resources was used to perform multiple sequence analysis by generating multiple sequence alignment. BLAST is a program or tool used to identify similarity between query sequence and similar sequences within NCBI database or input subject sequences. It is a very popular tool which utilizes vigorous statistical framework to identify the regions with local similarity between two or more aligned sequences. It's framework has the ability to statistically calculate the correctness of the aligned sequences. The tool's interface has many options including blastn, blastp, blastx, tblastn and tblastx. In this study we used blastp with 'Align two or more sequences' option checked for multiple sequence alignment. The reference sequence was inserted in the query sequence box and remaining 80 sequences were inserted in the subject sequence box and then the tool was run using default parameters.

3.3. Mutation Analysis

The multiple sequence alignment generated by blastp is viewed on the BLAST interface with using query anchored with dots for identities option selected for the visualization of mutations. The position of each amino acid changes (mutation) is assessed and recorded.

The recorded mutations were analyzed further by using 6 (SIFT, MUpro, SNAP2, PolyPhen2, PhD-SNP, I-Mutatnt 3.0) different tools of bioinformatics to implicit the impact identified mutations on the structural and functional stability of spike protein.

3.4. Screening Of Deleterious Mutations

For the screening of recorded mutations, the predictions of all 6 tools mentioned above were recorded. A score criteria was set from 0 to 6. The mutation with score 0 was predicted to be neutral (does not affect protein stability) by all 6 tools. Though it would be predicted to effect protein stability if it would get any score by all 6 tools. The mutations with score equal or greater than 3 were assumed to be deleterious and effects structural and functional stability of spike protein.

3.5. Physiochemical Properties Analysis

The physiochemical properties of the spike protein of were predicted using ProtParam tool. ProtParam is an online tool available on the Expasy which is Swiss Bioinformatics Resource portal. ProtParam is used to determine the various physiochemical properties that can be inferred using a protein sequence. ProtParam only require the protein as a Swiss-Prot/TrEMBL id, or in the raw sequence format. The physiochemical properties of spike protein including Molecular weight, aliphatic index, theoretical isoelectric point (pI), extinction coefficient, instability index, and grand average hydropathy (GRAVY) were predicted by the ProtParam.The physiochemical properties analysis play very important role in predicting the function and structure of protein sequences.

3.6. Phylogenetic Analysis

MEGA (Molecular Evolutionary Genetics Analysis) is an offline tool used to carried out phylogenetic analysis to examine similarities and differences between the 80 SARS-CoV-2 genome sequences with reference genome sequence downloaded from NCBI. The optimal phylogenetic tree was built by using Neighbour-Joining method and bootstrap (1000 replicate) tests. The associated taxa that are clustered together are show in the replicate trees adjacent to the branches. Length of the branch of tree show the evolutionary distances in the phylogenetic tree. The evolutionary distances in units of the number of amino acid substitutions per site were calculated using the Poisson correction method. For each sequence pair, all unclear places were eliminated (pairwise deletion option). The final dataset contained 1273 locations

altogether. MEGA11, an offline software application, was used to conduct evolutionary analyses. First multiple sequence alignment profile was generated by MEGA using Clustalw or MUSCLE method and exported as .meg file. The phylogenetic tree was then build using multiple sequence alignment .meg file.

3.7. Structural Analysis

The local folded structures that appear as a result of interactions between backbone atoms are represented by the secondary structure of proteins. The α helix and the β sheet are the two most well-known classes of secondary structures. These two structures differ from one another due to hydrogen bonds that form between the carbonyl O of one amino acid and the amino H of another. The secondary structure of spike protein was build using SOPMA (Self-Optimized Prediction Method with Alignment) software.

A protein three-dimensional structure can also be referred as its tertiary structure. In order to reach optimal stability or the minimal energy state, the protein molecule flexes and twists in this shape. A protein can preserve its three-dimensional shape with the help of various bonds and forces. They are hydrophobic interactions, which play a major role in preserving a protein's structure. Protein structure is stabilized by hydrogen bonds that form inside the chain of polypeptides and between amino acid "R" groups. Ionic bonding occurs in between the positively and negatively charged "R" groups of cysteine amino acids, whereas covalent bonds between the "R" groups of cysteine amino acids results in folding. The three dimensional or tertiary model of SARS-CoV-2 spike protein was inferred by I-TASSER (Iterative Threading Assembly Refinement) provided by Zhang Lab.

3.8. Epitope Prediction And Analysis

An epitope (or antigenic determinant) is the specific part of an antigen (foreign substance or protein, in our case SARS-CoV-2) identified by our immune system. An epitope is a particular area of an antigen to which an antibody bind. Predicting potential epitope candidates is one of the most important steps in vaccine design. For this purpose, ElliPro tool available on IEDB (Immune Epitope Database) Analysis Resource was used. ElliPro takes protein 3d structure or PDB id as input for the prediction of epitopes.

VaxiJen 2.0 software was used for the identification of antigenic epitopes from above predicted results. For further analyzing the allergenicity of identified antigenic epitopes, AllerTop v2.0 was used.

4. Results

4.1. Sequence Alignment And Mutation Analysis

Multiple sequence alignment of 80 protein sequences with reference sequence NC_045512 showed that Spike protein contain mostly conserved regions with 67 mutations that occurred in various samples from different regions. Table 2 contains the list of the mutations that were recorded from the samples used in this study. These mutations may be resulted to increase the viral fitness or may be the as a result of the effects of various environmental factors.

Table 2. Single Site Amino Acid Substitutions that were Found in the Multiple Sequence Alignment.

Sr#	Mutations	Frequency No. (%)
1	L18F	1(1.3%)
2	T19R	3(6%)
3	T19I	16(20%)
4	P25T	1(1.3%)
5	P26L	1(1.3%)
6	A27S	16(20%)
7	D80Y	1(1.3%)
8	T95I	1(1.3%)
9	E96V	1(1.3%)
10	R102S	4(5%)
11	G142D	19(23.8%)

12	E156G	3(3.8%)
13	F157L	1(1.3%)
14	M177V	1(1.3%)
15	V213G	16(20%)
16	A222V	4(5%)
17	S254F	1(1.3%)
18	G339D	9(11.3%)
19	S371F	4(5%)
20	S373P	14(17.5%)
21	S375F	14(17.5%)
22	T376A	14(17.5%)
23	D405N	14(17.5%)
24	R408S	12(15%)
25	K417N	11(13.6%)
26	D420N	1(1.3%)
27	N440K	12(15%)
28	K444N	3(3.8%)
29	K444M	2(2.5%)
30	K444T	2(2.5%)
31	L452R	14(17.5%)
32	N460K	3(3.8%)
33	D467N	1(1.3%)
34	S477N	16(20%)
35	T478K	19(23.8%)
36	E484A	16(20%)
37	E484K	1(1.3%)
38	F486V	16(20%)
39	Q498R	16(20%)
40	N501Y	17(21.3%)
41	Y505H	16(20%)
42	N556Y	1(1.3%)
43	D614G	71(88.6%)
44	H655Y	16(20%)
45	N658S	1(1.3%)
46	S673T	2(2.5%)
47	Q675H	1(1.3%)
48	Q677H	2(2.5%)
49	N679K	16(20%)
50	P681H	19(23.8%)
51	P681R	3(3.8%)
52	T716I	2(2.5%)
53	N764K	5(6.3%)
54	D796Y	16(20%)

55	S813N	1(1.3%)
56	D950H	1(1.3%)
57	D950N	2(2.5%)
58	Q954H	16(20%)
59	N969K	16(20%)
60	T1117I	3(3.8%)
61	D1153Y	1(1.3%)
62	D1163Y	1(1.3%)
63	I1169V	1(1.3%)
64	N1192H	1(1.3%)
65	Q1208H	1(1.3%)
66	G1219V	1(1.3%)
67	V1228L	1(1.3%)

4.2. Screening And Scoring Deleterious Mutations

For predicting mutations to be deleterious or neutral for the functional stability of protein, following 6 tools including SIFT, MUpro, SNAP2, PolyPhen2, PhD-SND and I-Mutant were utilized.

SIFT tool processes given query sequence by using multiple sequence alignment data and sorts tolerable substitutions from intolerable substitutions on the basis of probability score. Substitutions with probability score less than 0.5 are predicted intolerable and substitutions with probability score greater than 0.5 are predicted to be tolerable. SIFT tool predicts 9 mutations to be intolerable out of 67 recorded mutations. Remaining 58 mutations were predicted to be tolerable.

MUpro uses machine learning based algorithms to with 20-fold cross validations to predict the single site amino acid substitutions with 84% accuracy. Predictions with negative free energy change value or values less than 0 are supposed to decrease the stability of protein (i.e., deleterious) while predictions with positive free energy values or values greater than 0 are supposed to increase protein stability (i.e., neutral). MUpro analysis predicted only 1 out of 67 mutations to be neutral. While remaining 66 mutations were predicted to be deleterious by MUpro.

SNAP2 is based on machine learning algorithm called neural network which utilizes different sequence and variant features to predict whether single site amino acid substitution will affect protein functional stability or not. The mutation will affect the functional stability of the protein if the associated predicted score is greater than 50 and if the score is less than -50 the mutation will have no effect (i.e., neutral) on the protein functional stability. The analysis of SNAP2 revealed that 16 mutations out of 67 will affect the protein functional stability while remaining 51 mutations predicted neutral for the functional stability of protein.

PolyPhen2 predicts the effect of single site amino acid substitution on the structure and function of protein by using Simple physical and evolutionary comparative assessment with score 0 to 1. The mutations with score less than 0.5 are considered to be benign (i.e., neutral) and the mutations predicted with score equal or greater than 0.5 are predicted to be damaged (i.e., effect protein stability). Out of 67, 34 mutations were predicted damaged by the polyphen2 and remaining 33 mutations were predicted benign.

PhD-SNP is a support vector machine based classifier which predicts whether mutation is diseased or neutral with score 0 to 1. Score less than 0.5 are predicted diseased while score equal or greater than 0.5 are predicted neutral. Total 36 mutations out of 67 were predicted to be diseased by the PhD-SNP and remaining 31 mutations were predicted neutral for protein functional stability.

I-Mutant 3.0 predicts whether the single site amino acid substitution will increase or decreases the functional stability of the protein based on free energy change value (DDG value). The mutation will decrease the stability of protein if its DDG value is less than 0 and increases the stability of the protein if its DDG value is greater than 0. I-Mutant identified 55 mutations out of 67 mutations that decreases the stability of the protein. Remaining 12 mutations were identified to increase the protein stability.

For identifying truly deleterious mutations out of the 67 recorded mutations, the combined scores of all 6 tools that were used for the prediction of mutation were calculated and analyzed. The mutations with greater scores can be the potential candidates. The mutations that have been predicted deleterious by three or more than 3 tools were identified as high-risk mutations. Table 3 Figure 2, and Figure 3 summarizes the results of all 6 prediction tools. Based on the criteria (defined in materials and methods i.e., score ≥ 3), 51 mutations out of 67 mutations were identified to be truly deleterious and affects the functional and structural stability of the spkike protein.

Table 3. Summary of the Predictions made by All 6 Tools with Associated Scores.

Sr#	Mutations	SIFT	MUpro	SNAP[2]	PolyPhen2	PhD-SNP	I-Mutant	Score
1	L18F	-	✓	-	-	-	✓	2
2	T19R	-	✓	✓	-	✓	✓	4
3	T19I	-	✓	✓	-	-	✓	3
4	P25T	-	✓	-	-	-	✓	2
5	P26L	-	-	-	-	-	✓	1
6	A27S	-	✓	-	-	-	✓	2
7	D80Y	-	✓	✓	✓	✓	✓	5
8	T95I	-	✓	-	✓	✓	✓	4
9	E96V	-	✓	-	✓	✓	-	3
10	R102S	-	✓	-	✓	-	✓	3
11	G142D	-	✓	✓	-	✓	✓	4
12	E156G	-	✓	-	✓	-	✓	3
13	F157L	-	✓	-	✓	✓	✓	4
14	M177V	-	✓	-	-	-	✓	2
15	V213G	-	✓	✓	✓	✓	✓	5
16	A222V	-	✓	-	-	-	-	1
17	S254F	-	✓	✓	-	✓	-	3
18	G339D	-	✓	-	-	✓	✓	3
19	S371F	-	✓	-	✓	✓	-	3
20	S373P	-	✓	-	✓	✓	-	3
21	S375F	-	✓	-	✓	✓	-	3
22	T376A	-	✓	-	✓	-	✓	3
23	D405N	-	✓	-	✓	✓	✓	4
24	R408S	-	✓	-	✓	✓	✓	4
25	K417N	-	✓	-	✓	-	✓	3
26	D420N	-	✓	-	✓	✓	✓	4
27	N440K	-	✓	✓	-	✓	✓	4
28	K444N	-	✓	-	-	✓	✓	3
29	K444M	-	✓	-	✓	-	-	2
30	K444T	-	✓	-	-	✓	✓	3
31	L452R	-	✓	-	-	✓	✓	3
32	N460K	-	✓	-	-	✓	✓	3
33	D467N	✓	✓	-	✓	-	✓	4
34	S477N	-	✓	-	-	-	-	1
35	T478K	-	✓	-	-	✓	✓	3

36	E484A	-	✓	-	✓	✓	✓	4
37	E484K	-	✓	-	-	✓	✓	3
38	F486V	-	✓	-	-	✓	✓	3
39	Q498R	-	✓	-	-	✓	✓	3
40	N501Y	-	✓	✓	-	-	-	2
41	Y505H	✓	✓	✓	✓	✓	✓	6
42	N556Y	✓	✓	✓	✓	✓	-	5
43	D614G	-	✓	-	-	✓	✓	3
44	H655Y	-	✓	-	-	✓	-	2
45	N658S	-	✓	-	-	-	✓	2
46	S673T	-	✓	-	-	-	✓	2
47	Q675H	-	✓	-	-	-	✓	2
48	Q677H	-	✓	-	-	-	✓	2
614	N679K	-	✓	✓	-	-	✓	3
50	P681H	-	✓	-	✓	-	✓	3
51	P681R	-	✓	-	✓	-	✓	3
52	T716I	✓	✓	✓	✓	-	✓	5
53	N764K	✓	✓	✓	✓	✓	✓	6
54	D796Y	-	✓	-	-	✓	✓	3
55	S813N	-	✓	-	-	-	✓	2
56	D950H	✓	✓	✓	✓	-	✓	5
57	D950N	✓	✓	-	-	✓	✓	4
58	Q954H	-	✓	-	✓	-	✓	3
59	N969K	-	✓	✓	✓	✓	✓	5
60	T1117I	-	✓	-	✓	-	✓	3
61	D1153Y	✓	✓	✓	✓	-	-	4
62	D1163Y	-	✓	-	✓	✓	✓	4
63	I1169V	-	✓	-	-	-	✓	2
64	N1192H	-	✓	-	✓	-	✓	3
65	Q1208H	-	✓	-	✓	✓	✓	4
66	G1219V	✓	✓	-	✓	✓	✓	5
67	V1228L	-	✓	-	✓	-	✓	3

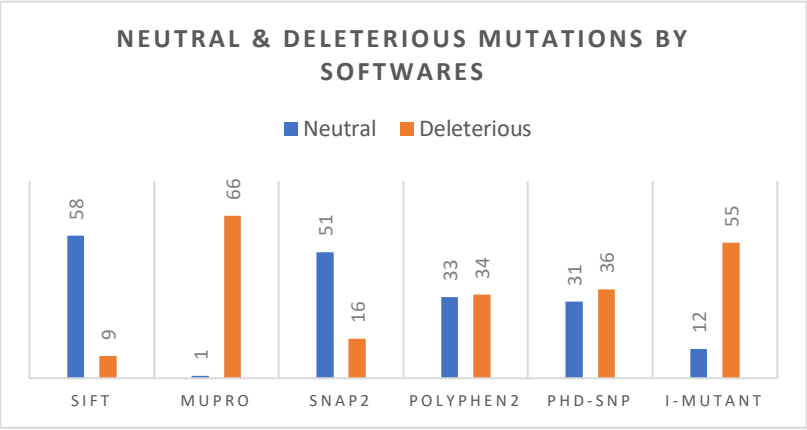

Figure 2. Neutral and Deleterious Mutations Predicted by each Software Separately.

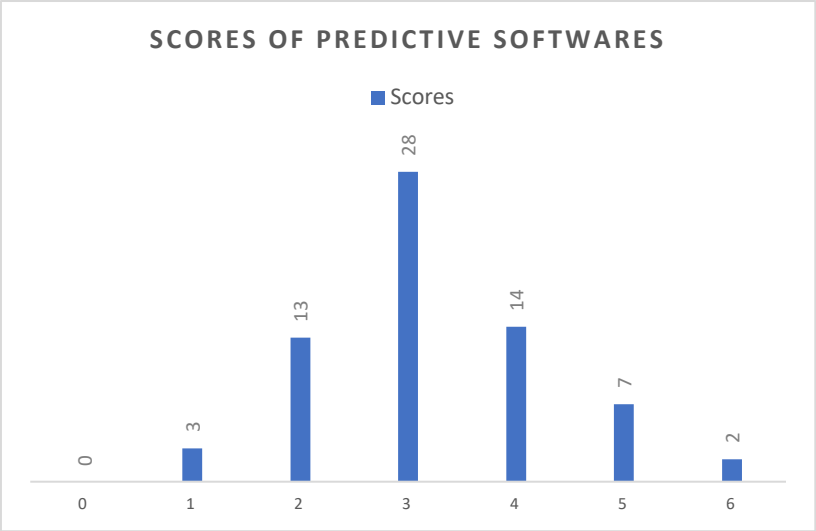

Figure 3. The Combined Scores of 6 Software for Recorded Mutations.

4.3. Physiochemical Properties Analysis Results

The physiochemical analysis of spike protein using ProtParam tool revealed that it has molecular weight of 141178.47kDa. The protein was thermally stable acidic peptide due to high proportion of amino acids that are acidic in nature with aliphatic index of 84.67 and theoretical pI of 6.24 repetitively. Extinction coefficient, proteins' ability to strongly absorbs or reflects the particular wavelength of light or radiation was 148960. The physiochemical analysis also sheds light on the fact that spike protein was hydrophilic one (water loving) with grand average of hydropathicity (GRAVY) value of -0.079 and remains stable in the test tube with instability index of 33.01.

4.4. Phylogenetic Analysis Results

The phylogenetic tree for all downloaded 81 sequences (Figure 4) was built by using neighbor joining method and 1000 bootstrap replication tests on MEGA11 software tool. The analysis of the phylogenetic tree revealed that all sequences from Karachi are located in the upper cluster with some sequences from the Islamabad. The Sequences from Punjab shows more similarity with the sequences from Islamabad and some Federal Capital. The sequences from AJK and Islamabad show similarity with the reference sequence. Sequences from KPK shows more similarity with the sequences from Federal Capital and Gilgit than Islamabad. Lahore sequences shows similarity only with Islamabad sequences. The Rawalpindi sequence is located along with the sequences from KPK and Federal Capital. Gilgit sequences shows more similarity to KPK and Federal Capital sequences than Islamabad. The sequences from Federal Capital were mostly located in the lower cluster with sequence from Rawalpindi and shows more similarity with sequences from KPK and Islamabad than Punjab.

4.5. Structural Analysis Results

The secondary structure of a protein is the representation of the structures that emerge from interactions between the atoms of the backbone. The two most well-known categories of secondary structures are α helix and the β sheet. The hydrogen bonds that develop between the carbonyl O of one amino acid and the amino H of another are what distinguish these two structures from one another. SOPMA is an online web server used to predict the secondary structure of the spike protein. The result of the analysis (Figure 5 & Figure 6) showed that spike protein contains random coil accounted for 44.78% of the major structure, followed by the alpha helix 28.59%, extended strand 23.25%, and beta turn 3.38% respectively. The greater percentage of alpha helix than extended strands highlighted that this protein is stable in nature.

The term "tertiary structure" can also be used to describe a protein's three-dimensional structure. The protein molecule flexes and twists in this shape to reach the lowest energy state or best stability. With the help of several bonds and forces, a protein may maintain its three-dimensional form. They are hydrophobic interactions, which are crucial for maintaining the structural integrity of proteins. Hydrogen bonds that develop between polypeptide chains and between amino acid "R" groups help to stabilise protein structure. Between the positively and negatively charged "R" groups of cysteine amino acids, ionic bonding takes

place, whereas folding is caused by covalent bonds between the "R" groups of cysteine amino acids. I-TASSER is an online web server famous for predicticting protien 3D structures. The 3D structure of spike protein was predicted using I-TASSER which predicts 5 models for protein 3D structure. These 5 models were further evaluated using QMEAN, ERRRAT and Ramachandran Plot to identify the best quality model out of 5 models. Table 4 showes the result of all 3 tools for every model along with I-TASSER C-Score. The model 3 which is bolded in the was selected. Figure 7 showes final model for the Spike protein predicted by the I-TASSER.

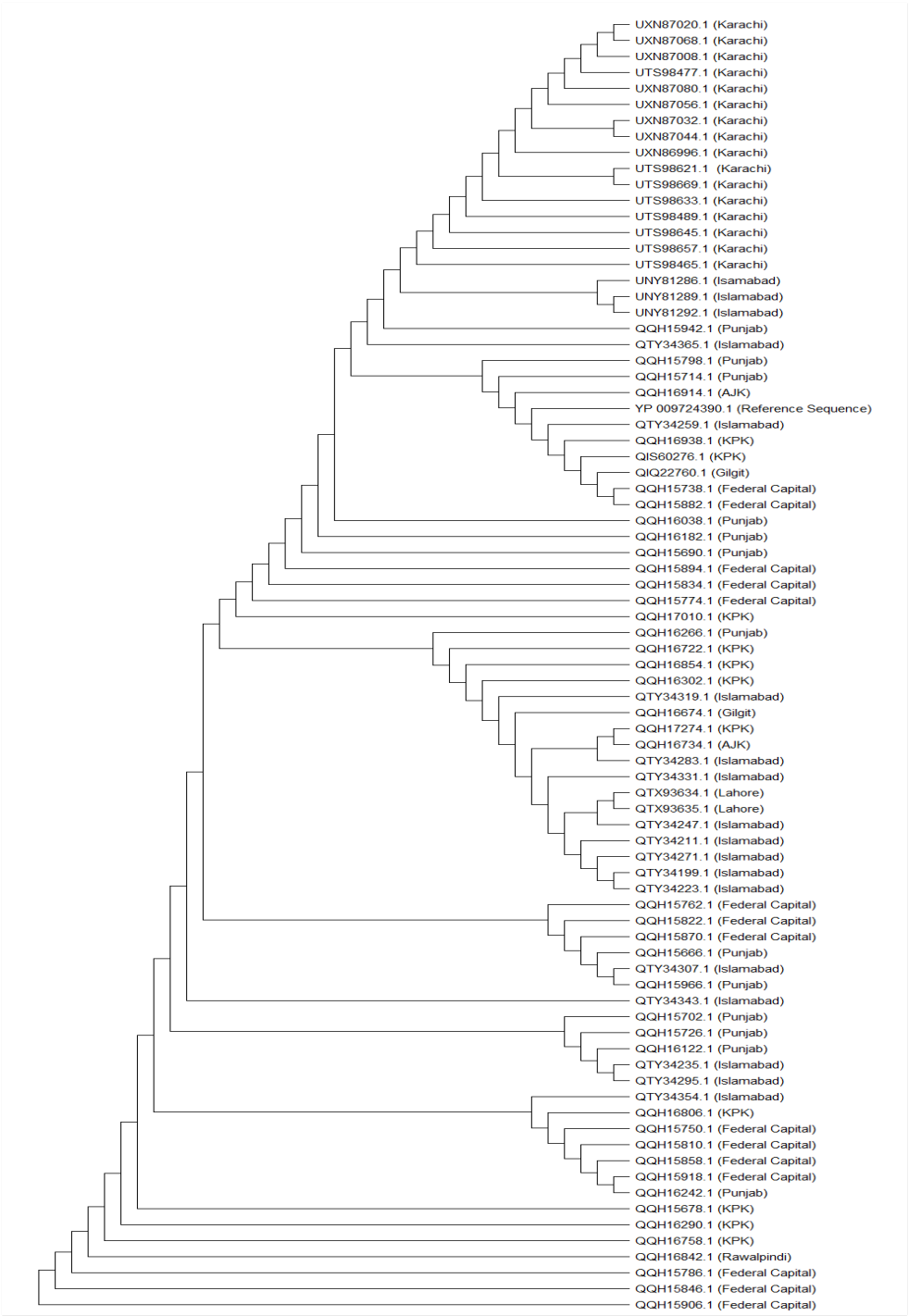

Figure 4. The Phylogenetic Tree for All Downloaded Sequences Built Using Neighbor Joining Method with 1000 Bootstrap Tests on MEGA11.

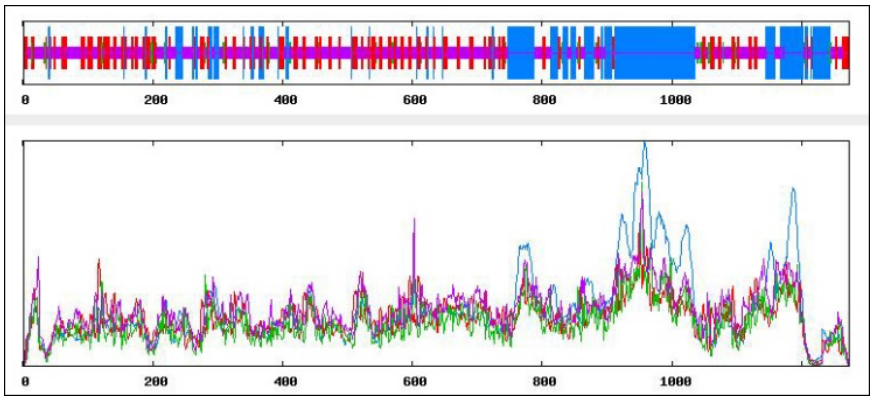

Figure 5. SOPMA Results. Orange Bars Indicate Random Coil. Blue Bars Indicate Alpha Helix. Red Bars Indicate Extended Strand. Green Bars Indicate Beta Turn.

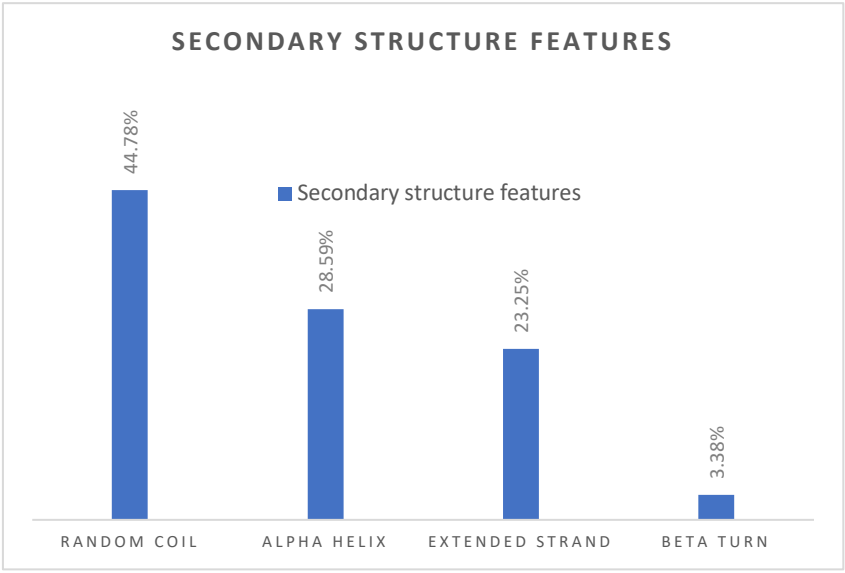

Figure 6. Secondary Structure Features Predicted by SOPMA.

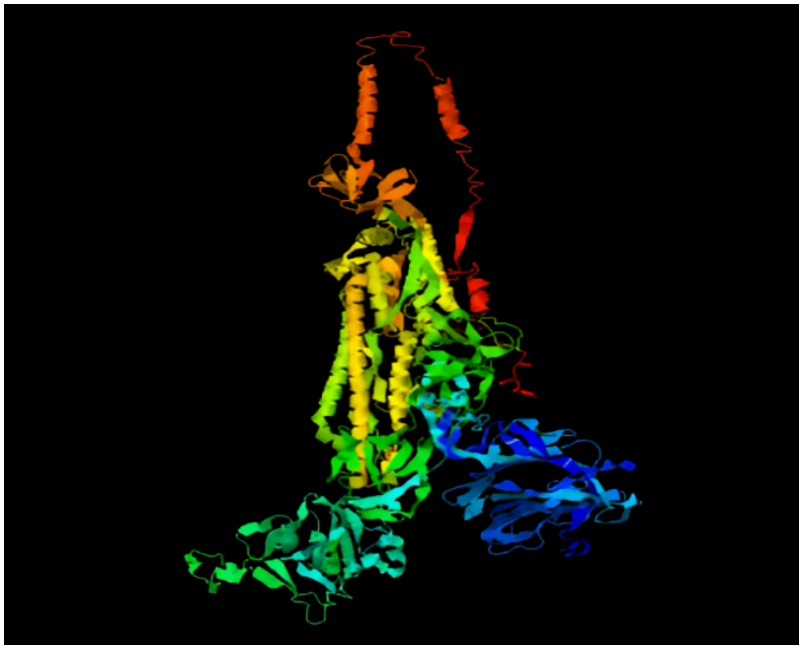

Figure 7. Final Predicted 3D Model of Spike Protein from I-TASSER.

Table 4. Results of all 3 Tools for Predicting the 3D Structure of Spike Protein.

Server	Models	C-Score	QMEAN	ERRAT	Ramachandran Plot
I-TASSER	1	-2.00	-6.00	84.06%	1005 (88.86%)
	2	-1.98	-7.65	75.48%	969 (85.68%)
	3	-1.90	-6.09	83.08%	1009 (89.21%)
	4	-2.74	-7.39	81.02%	985 (87.09%)
	5	-2.98	-6.78	81.77%	999 (88.32%)

4.6. Epitope Prediction Results

The specific part of an antigen (a foreign substance or protein, in our case SARS-CoV-2) that is recognized by human immune system is known as an epitope (also known as an antigenic determinant). An epitope is a region of an antigen where an antibody can attach. One of the most crucial tasks in vaccine design is predicting potential epitope candidates. IEDB (Immune Epitope Database) Analysis Resource's ElliPro tool was utilized for predicting the SARS-CoV-2 spike protein. The final predicted 3D model of Spike protein by I-TASSER was used as input for ElliPro. ElliPro predicts 21 epitope candidates (Figure 8).

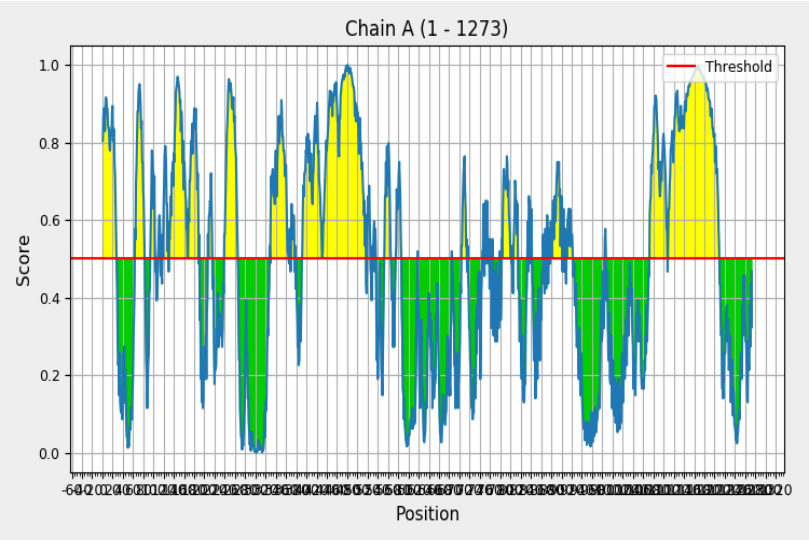

Figure 8. The Regions Above the Threshold Line Indicate Epitopes Predicted by The Ellipro.

For screening potential epitopes from these predicted epitopes, Vaxijen2.0 and AllerTop v2 tools were used. Vaxijen2.0 predicts the antigenicity of the predicted epitope and AllerTop v2 predicts the allergenicity of the predicted epitope. The epitope predicted as the antigen and non-allergen by these two tools was selected as the potential epitope candidate. Table 5 shows all the epitopes predicted by ElliPro and the combined results of Vaxijen2.0 and AllerTop v2. The screened epitopes were bolded in the table. ElliPro also shows the mapping of predicted epitopes on the protein 3D structure (Figure 9).

Table 5. Epitopes Predicted by Ellipro Including Vaxijen2.0 And Allertop V2 Results for the Predicted Epitopes.

Sr#	Start	End	Residues	Antigenicity	Allergenicity
1	1	27	27	Antigen	Non-Allergen
2	64	84	21	Antigen	Allergen
3	94	105	12	Non-Antigen	Non-Allergen
4	108	114	7	Antigen	Non-Allergen
5	118	190	73	Non-Antigen	Non-Allergen
6	207	216	10	Non-Antigen	Non-Allergen

7	239	265	27	Antigen	Non-Allergen
8	327	381	55	Non-Antigen	Non-Allergen
9	392	526	135	Antigen	Allergen
10	528	537	10	Antigen	Non-Allergen
11	553	566	14	Antigen	Non-Allergen
12	574	586	13	Antigen	Allergen
13	703	716	14	Antigen	Allergen
14	744	758	15	Non-Antigen	Allergen
15	783	803	21	Non-Antigen	Non-Allergen
16	806	815	10	Non-Antigen	Allergen
17	832	342	11	Non-Antigen	Allergen
18	863	869	7	Non-Antigen	Allergen
19	879	904	26	Antigen	Non-Allergen
20	909	922	14	Antigen	Allergen
21	1073	1210	138	Non-Antigen	Non-Allergen

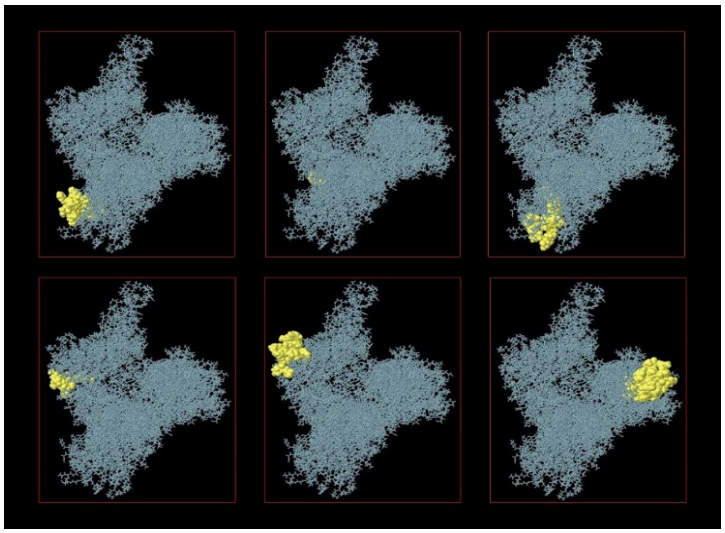

Figure 9. Ellipro 3D Structure Mapping of a Predicted Potential Epitopes.

5. Discussion

The proposed study is based on insilico analysis of spike protein. The mutation analysis of all retrieved sequences from different regions of Pakistan identified 67 mutations in spike protein region. In the research conducted by [23], identified 92 mutations in spike protein region from the first wave sequences from 20 countries. Researchers in [23] utilizes SIFT, DUET, SNAP2, DynaMut, PhD-SNP, and I-Mutant tools to predict the deleterious mutations from identified mutations. We used SIFT, MUpro, SNAP2, PolyPhen2, PhD-SNP, and I-Mutant for the prediction of deleterious mutations. In [23] SIFT identified 34, SNAP2 29, PhD-SNP 20 I-Mutant identified 71 mutations as deleterious mutations. In our SIFT identified 9, SNAP2 16, PhD-SNP 36 I-Mutant identified 55 mutations as deleterious mutations. Only 5 mutations were identified similar to the mutations found in [23] remaining 62 mutations were found new in the sequences and were not identified previous in [23]. In previous research, D80Y mutation was found with score 2 while in our studies D80Y found with score 5. While D614G, H655Y, Q675H and Q677H were found with scores 4, 3, 4 and 3 in previous research and with scores 3, 2, 2 and 2 in our studies which means some mutations predicted deleterious in [23] were predicted neutral in our studies. Spike protein was characterized as acidic, thermally stable, and hydrophilic by present study. However, it seems that yeast and mammalian cells can express this protein more efficiently since spike protein requires some post-modification procedures. Researchers in [39] Expressed spike protein using several cell lines, which

supported our ProtParam prediction by demonstrating the protein's stability in mammalian cells. The projected secondary structure of the spike protein revealed that random coil is the major structure, followed by alpha helix and beta strand fair distributions, which support the protein's stability. 3D structure predicted by I-TASSER was refined by the scores obtained by 3 tools and C-Score for each model individually, as a result model 3 predicted by I-TASSER was selected as final tertiary structure for spike protein. By using linear sequence, researchers in [17] specified 4 B-cell epitopes (249-259,674-687,807-816, and 1254-1265) while our study predicted the 6 potential B-Cell epitopes (1-27, 108-114, 239-265, 528-537, 553-566, and 879-904) by utilizing the final 3D structure predicted by I-TASSER. Thus, our study provides better results with 3D structure by identifying more exposed regions predicted for B-Cell epitopes.

The fact that knowledge about the Covid-19 problem is constantly changing and that online databases are updated with new sequences on a daily basis may be the limitation of this work. Our analysis might not provide a complete overview of spike protein as a result. However, as a pilot study, our results provide recommendations for future investigation.

6. Conclusions

Using a several of bioinformatics tools, six different types of insilico experiments were conducted in this study. Using the blastp suit, which is available on the NCBI web server for mutation analysis, multiple sequence alignment was generated which identified 67 mutations in Spike protein region. The spike protein was screened for harmful mutations using SIFT, MUpro, SNAP2, PolyPhen2, PhD-SNP, and I-Mutant. The analysis predicted 51 mutations out of 67 as deleterious mutations by having effect on protein structural and functional stability. Through ProtParam, physiochemical characteristics were analyzed, which showed that protein is acidic (pH 6.24), thermostable (aliphatic index 84.67), hydrophilic (GRAVY -0.079), stable in test tube (instability index 33.01) and molecular weight 141178.47kDa. Phylogenetic tree build on MEGA11 by using neighbor joining method and bootstrap 1000 tests shows that sequences from AJK and Islamabad show similarity with the reference sequence. The result of SOPMA for protein secondary structure analysis showed that spike protein contains random coil accounted for 44.78% of the major structure, followed by the alpha helix 28.59%, extended strand 23.59%, and beta turn 3.38% respectively. The greater percentage of alpha helix than extended strands highlighted that this protein is stable in nature. Out of all 5 models predicted by I-TASSER, Model 3 was selected for final 3D structure after verified by 3 tools QMEAN, ERRAT and Ramachandran Plot including C-Score from I-TASSER. Finally, the IEDB (Immune Epitope Database) tool ElliPro predicted 21 candidates for epitopes out of which 6 were predicted as potential epitopes by VaxiJen2.0 and AllerTop v2.0.

The findings of this study gave us a thorough grasp of spike protein. In total, 51 mutations that cause significant phenotypic harm and may change the protein's structure and functional behaviour have been predicted. On the surface of the SARS-CoV-2, the spike protein is a very effective epitope that has various characteristics that make it a good candidate for the development of new diagnostics, antiviral medications, and treatments.

As most of these pathogenic mutations may influence the viral protein's affinity for different medications, it may be important to assess the effects of these found pathogenic mutations using various in vitro and molecular methodologies in the future.

Funding: Please add: This research received no external funding.

Data Availability Statement: The authors declare that all data supporting the findings of this study are available within the article.

Conflicts of Interest: According to the authors, there is no conflict of interest in this study.

References

1. Al-Rohaimi, A. H., & Al Otaibi, F. (2020). Novel SARS-CoV-2 outbreak and COVID19 disease; a systemic review on the global pandemic. Genes & Diseases, 7(4), 491-501.
2. Department of Computer Science, National College of Business Administration and Economics, Multan, Pakistan, Multan,
3. Pakistan.
4. Junejo, Y., Ozaslan, M., Safdar, M., Khailany, R. A., Rehman, S., Yousaf, W., & Khan, M. A. (2020). Novel SARS-CoV-2/COVID-19: origin, pathogenesis, genes and genetic variations, immune responses and phylogenetic analysis. Gene reports, 20, 100752.
5. Malik, H., Anees, T., Din, M., & Naeem, A. (2022). CDC_Net: multi-classification convolutional neural network model for detection of COVID-19, pneumothorax, pneumonia, lung Cancer, and tuberculosis using chest X-rays. Multimedia Tools and Applications, 1-26.
6. Naeem, A. B., Senapati, B., Chauhan, A. S., Kumar, S., Orosco Gavilan, J. C., & F. Abdel-Rehim, W. M. (2023, February 17). Deep Learning Models for Cotton Leaf Disease Detection with VGG-16 | International Journal of Intelligent Systems and Applications in Engineering. Deep Learning Models for Cotton Leaf Disease Detection With VGG-16 | International Journal of Intelligent Systems and Applications in Engineering. https://www.ijisae.org/index.php/IJISAE/article/view/2710.
7. Naeem, A. B. ., Senapati, B. ., Chauhan, A. S. ., Makhija, M. ., Singh, A. ., Gupta, M. ., Tiwari, P. K. ., & Abdel-Rehim, W. M. F. (2023). Hypothyroidism Disease Diagnosis by Using Machine Learning Algorithms. International Journal of Intelligent Systems and Applications in Engineering, 11(3), 368–373. Retrieved from https://www.ijisae.org/index.php/IJISAE/article/view/3178
8. Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., ... & Ippodrino, R. (2020). Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. Journal of translational medicine, 18(1), 1-9.
9. Zheng, J. (2020). SARS-CoV-2: an emerging coronavirus that causes a global threat. International journal of biological sciences, 16(10), 1678.
10. Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., ... & Zhong, N. S. (2020). Clinical characteristics of coronavirus disease 2019 in China. New England journal of medicine, 382(18), 1708-1720.
11. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., ... & Pöhlmann, S. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. cell, 181(2), 271-280.
12. Maginnis, M. S. (2018). Virus–receptor interactions: the key to cellular invasion. Journal of molecular biology, 430(17), 2590-2611.
13. Di Gennaro, F., Pizzol, D., Marotta, C., Antunes, M., Racalbuto, V., Veronese, N., & Smith, L. (2020). Coronavirus diseases (COVID-19) current status and future perspectives: a narrative review. International journal of environmental research and public health, 17(8), 2690.
14. Naeem, A.B., Soomro, A.M., Saim, H.M. et al. Smart road management system for prioritized autonomous vehicles under vehicle-to-everything (V2X) communication. Multimed Tools Appl (2023). https://doi.org/10.1007/s11042-023-16950-1
15. Abbasi-Oshaghi, E., Mirzaei, F., Farahani, F., Khodadadi, I., & Tayebinia, H. (2020). Diagnosis and treatment of coronavirus disease 2019 (COVID-19): Laboratory, PCR, and chest CT imaging findings. International Journal of Surgery, 79, 143-153.
16. Lee, Y., Min, P., Lee, S., & Kim, S. W. (2020). Prevalence and duration of acute loss of smell or taste in COVID-19 patients. Journal of Korean medical science, 35(18).
17. Renu, K., Prasanna, P. L., & Gopalakrishnan, A. V. (2020). Coronaviruses pathogenesis, comorbidities and multi-organ damage–A review. Life sciences, 255, 117839.
18. Javanmardi, F., Keshavarzi, A., Akbari, A., Emami, A., & Pirbonyeh, N. (2020). Prevalence of underlying diseases in died cases of COVID-19: A systematic review and meta-analysis. PloS one, 15(10), e0241265.
19. Ebrahim-Saraie, H. S., Dehghani, B., Mojtahedi, A., Shenagari, M., & Hasannejad-Bibalan, M. (2021). Functional and structural characterization of SARS-Cov-2 spike protein: an in silico study. Ethiopian Journal of Health Sciences, 31(2).
20. Hashempour, T., Dehghani, B., Mousavi, Z., Akbari, T., Hasanshahi, Z., Moayedi, J., ... & Davarpanah, M. A. (2021). Association of mutations in the NS5A-PKRBD region and IFNL4 genotypes with hepatitis C interferon responsiveness and its functional and structural analysis. Current Proteomics, 18(1), 38-49.
21. Dehghani, B., Hasanshahi, Z., Hashempour, T., & Motamedifar, M. (2020). The possible regions to design Human Papilloma Viruses vaccine in Iranian L1 protein. Biologia, 75(5), 749-759.
22. Sanjuán, R., & Domingo-Calap, P. (2016). Mechanisms of viral mutation. Cellular and molecular life sciences, 73(23), 4433-4448.
23. Khailany, R. A., Safdar, M., & Ozaslan, M. (2020). Genomic characterization of a novel SARS-CoV-2. Gene reports, 19, 100682.
24. Naeem, A. B., Khalid, F., Soomro, A. M., Del Mundo, A. D., Zaidi, A., Senapati, B., & Doshi, O. P. (2023, March 29). Early Gender Identification of Date Palm Using Machine Learning | Journal of Computing & Biomedical Informatics. Early Gender Identification of Date Palm Using Machine Learning | Journal of Computing & Biomedical Informatics. https://www.jcbi.org/index.php/Main/article/view/147
25. Ikram, A., Naz, A., Awan, F. M., Rauff, B., Obaid, A., Hakim, M. S., & Malik, A. (2021). The impact of mutations on the structural and functional properties of SARS-CoV-2 proteins: A comprehensive bioinformatics analysis. bioRxiv.

26. Naeem, Awad & Senapati, Biswaranjan & Zaidi, Abdelhamid & Maaliw III, Renato & Sudman, Md & Das, Debabrata & Almeida, Friban & Sakr, Hesham. (2024). Detecting Three Different Diseases of Plants by Using CNN Model and Image Processing. Journal of Electrical Systems. 20. 1519-1525. 10.52783/jes.1455.
27. Soomro, A. M., Naeem, A. B., Shahzad, K., Madni, A. M., Del Mundo, A. D., Sajid, M., & Baloch, M. A. (2022, December 29). Forecasting Cotton Whitefly Population Using Deep Learning | Journal of Computing & Biomedical Informatics. Forecasting Cotton Whitefly Population Using Deep Learning | Journal of Computing & Biomedical Informatics. https://doi.org/10.56979/401/2022/67
28. Naeem, A.B. et al. (2023) 'Intelligent Four-way Crossroad Safety Management for autonomous, non-autonomous and VIP vehicles', 2023 IEEE International Conference on Emerging Trends in Engineering, Sciences and Technology (ICES&T). doi:10.1109/icest56843.2023.10138829.
29. Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. nature, 579(7798), 270-273.
30. Schoeman, D., & Fielding, B. C. (2019). Coronavirus envelope protein: current knowledge. Virology journal, 16(1), 1-22.
31. Soomro, A.M.; Naeem, A.B.; Senapati, B.; Bashir, K.; Pradhan, S.; Maaliw, R.R.; Sakr, H.A. Constructor Development: Predicting Object Communication Errors. In Proceedings of the 2023 IEEE International Conference on Emerging Trends in Engineering, Sciences and Technology (ICES&T), Bahawalpur, Pakistan, 9–11 January 2023; pp. 1–7.
32. Wang, C., Zheng, X., Gai, W., Zhao, Y., Wang, H., Wang, H., ... & Xia, X. (2017). MERS-CoV virus-like particles produced in insect cells induce specific humoural and cellular imminity in rhesus macaques. Oncotarget, 8(8), 12686.
33. Mortola, E., & Roy, P. (2004). Efficient assembly and release of SARS coronavirus-like particles by a heterologous expression system. FEBS letters, 576(1-2), 174-178.
34. Lindner, H. A., Fotouhi-Ardakani, N., Lytvyn, V., Lachance, P., Sulea, T., & Ménard, R. (2005). The papain-like protease from the severe acute respiratory syndrome coronavirus is a deubiquitinating enzyme. Journal of virology, 79(24), 15199-15208.
35. Shimamoto, Y., Hattori, Y., Kobayashi, K., Teruya, K., Sanjoh, A., Nakagawa, A., ... & Akaji, K. (2015). Fused-ring structure of decahydroisoquinolin as a novel scaffold for SARS 3CL protease inhibitors. Bioorganic & medicinal chemistry, 23(4), 876-890.
36. Naeem, A. B. ., Senapati, B. ., Mahadin, G. A. ., Ghulaxe, V. ., Almeida, F. ., Sudman, S. I. ., & Ghafoor, M. I. . (2024). Determine the Prevalence of Hepatitis B and C During Pregnancy by Using Machine Learning Algorithm . International Journal of Intelligent Systems and Applications in Engineering, 12(13s), 744–751. Retrieved from https://www.ijisae.org/index.php/IJISAE/article/view/4704
37. A. B. Naeem et al., "Heart Disease Detection Using Feature Extraction and Artificial Neural Networks: A Sensor-Based Approach," in IEEE Access, vol. 12, pp. 37349-37362, 2024, doi: 10.1109/ACCESS.2024.3373646.
38. Salemi, M., Fitch, W. M., Ciccozzi, M., Ruiz-Alvarez, M. J., Rezza, G., & Lewis, M. J. (2004). Severe acute respiratory syndrome coronavirus sequence characteristics and evolutionary rate estimate from maximum likelihood analysis. Journal of virology, 78(3), 1602-1603.
39. Khan, M. I., Khan, Z. A., Baig, M. H., Ahmad, I., Farouk, A. E., Song, Y. G., & Dong, J. J. (2020). Comparative genome analysis of novel coronavirus (SARS-CoV-2) from different geographical locations and the effect of mutations on major target proteins: An in silico insight. PLoS One, 15(9), e0238344.
40. Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell, 181(2), 281-292.