

Original Paper

In Silico Comparative Analysis of the Functional, Structural, and Evolutionary Properties of SARS-CoV-2 Variant Spike Proteins

Renukaradhya K Math, PhD; Nayana Mudennavar, MSc; Palaksha Kanive Javaregowda, PhD; Ambuja Savanur, MSc
SDM Research Institute for Biomedical Sciences, Shri Dharmasthala Manjunatheshwara University, Dharwad, India

Corresponding Author:

Renukaradhya K Math, PhD
SDM Research Institute for Biomedical Sciences
Shri Dharmasthala Manjunatheshwara University
5th floor, Manjushree building
SDM College of Medical Sciences & Hospital Campus
Dharwad, 580009
India
Phone: 91 7019982929
Email: aradhya.swamy@gmail.com

Abstract

Background: A recent global outbreak of COVID-19 caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) created a pandemic and emerged as a potential threat to humanity. The analysis of virus genetic composition has revealed that the spike protein, one of the major structural proteins, facilitates the entry of the virus to host cells.

Objective: The spike protein has become the main target for prophylactics and therapeutics studies. Here, we compared the spike proteins of SARS-CoV-2 variants using bioinformatics tools.

Methods: The spike protein sequences of wild-type SARS-CoV-2 and its 6 variants—D614G, alpha (B.1.1.7), beta (B.1.351), delta (B.1.617.2), gamma (P.1), and omicron (B.1.1.529)—were retrieved from the NCBI database. The ClustalX program was used to sequence multiple alignment and perform mutational analysis. Several online bioinformatics tools were used to predict the physiological, immunological, and structural features of the spike proteins of SARS-CoV-2 variants. A phylogenetic tree was constructed using CLC software. Statistical analysis of the data was done using jamovi 2 software.

Results: Multiple sequence analysis revealed that the P681R mutation in the delta variant, which changed an amino acid from histidine (H) to arginine (R), made the protein more alkaline due to arginine's high pKa value (12.5) compared to histidine's (6.0). Physicochemical properties revealed the relatively higher isoelectric point (7.34) and aliphatic index (84.65) of the delta variant compared to other variants. Statistical analysis of the isoelectric point, antigenicity, and immunogenicity of all the variants revealed significant correlation, with *P* values ranging from <.007 to .04. The generation of a 2D gel map showed the separation of the delta spike protein from a grouping of the other variants. The phylogenetic tree of the spike proteins showed that the delta variant was close to and a mix of the *Rousettus* bat coronavirus and MERS-CoV.

Conclusions: The comparative analysis of SARS-CoV-2 variants revealed that the delta variant is more aliphatic in nature, which provides more stability to it and subsequently influences virus behavior.

(*JMIR Bioinform Biotech* 2022;3(1):e37391) doi: [10.2196/37391](https://doi.org/10.2196/37391)

KEYWORDS

spike protein variants; NCBI; bioinformatics tools; pI; isoelectric point; 2D map; phylogenetic tree; COVID-19; COVID therapy; SARS-CoV-2 treatment; therapeutic; spike protein; protein; prophylactic; sequence analysis; genomic; bioinformatics; viral protein

Introduction

The constant evolution of new variants of SARS-CoV-2 is a substantial challenge to people from every sector, particularly those in health care and research and development in the areas

of diagnostics, prophylactics, and therapeutics development, as well as policy makers and administrators. The virus was first observed in December 2019 in China and later spread throughout the globe causing a pandemic. However, continuous mutations in the genome of the virus have created several new variants among the circulating viruses in different geographical regions

worldwide. These mutations have led to increased transmissibility, antibody evasion, and severity in patients. Several research studies have cited that the spike protein of the coronavirus is responsible for interacting with human cells to penetrate inside [1,2]; however, adaptive mutations accumulated in the gene responsible for the spike protein have allowed the virus to acclimatize and escape the host immune system. A number of variants have been isolated and identified worldwide; the World Health Organization classified them as variants of concern (alpha, beta, gamma, delta, and omicron) and variants of interest (VOIs; eta, iota, kappa, and lambda) [3].

A single mutation in a gene can lead to a change of an amino acid in a protein, which can drastically affect an organism's ability in many ways, such as transmissibility, quick adaptability, stability evading the host immune system, and pathogenicity [4]. The availability of high throughput second generation sequencing technologies like next-generation sequencing in the last decade has helped in the timely sequencing of the genomes of SARS-CoV-2 variants to identify mutations occurring among new genetic variants. Rapid sequencing technology has not only facilitated sequencing of the viral genome but also helped accelerate the development of diagnostic tools, vaccines, and therapeutics for COVID-19. Furthermore, the availability of genome sequence data and databases helped the scientific community understand the virus through the molecular, biochemical, genome, and proteome analyses of the virus. However, the continuous evolution of SARS-CoV-2 compels the scientific community to continuously strive toward new and thorough understanding of the properties of the virus variants and develop counter strategies against the virus to safeguard humankind.

The isoelectric point (pI) of a protein is crucial in determining the physicochemical properties of the protein [5]. The exposure of charged amino acids on the protein surface to solvents,

hydration, and dehydration also influences the pI of a protein [6,7]. Thus, evaluation of the pI of the spike proteins of wild-type and mutated variants is quintessential in understanding the influence of the spike protein on virus behavior and the transmission rate of viruses, as well as developing prophylactic and therapeutic agents. Additionally, the roles of posttranslational modification, phosphorylation, methylation, and alkylation also influence the pI of a protein, which cannot be ignored. Similarly, the genome of SARS-CoV-2 is prone to genetic evolution or genetic shift while adapting to a human host's microenvironment. Such mutations result in the emergence of new variants that might have different characteristics compared to its ancestral strains. Therefore, in this study, we aim to adopt an *in silico* method to analyze the spike protein sequences of wild-type and other variants of SARS-CoV-2 to know their physicochemical, functional, and evolutionary properties, which might help in the surveillance of SARS-CoV-2 through a systematic understanding of the continuously evolving properties as well as to develop the diagnostic tools and standardization of prophylactics and therapeutics strategies.

Methods

Retrieval of Spike Protein Sequences and Multiple Sequence Alignment

The protein sequences of wild-type SARS-CoV-2 and its 6 variants—D614G, alpha (B.1.1.7), beta (B.1.351), delta (B.1.617.2), gamma (P.1), and omicron (B.1.1.529)—were obtained from the NCBI database. The list of all the variants with the NCBI accession numbers are mentioned in Table 1. Multiple sequence alignment of all the protein was done using ClustalX2 software [8] to identify and confirm common and specific mutations among all the sequences listed in Table 1.

Table 1. List of SARS-CoV-2 spike proteins and the comparison of mutational analysis data of the wild type and its variants.

SARS-CoV-2 variants	First identified	NCBI accession	Linear SeqVrl	Pango lineage	Mutation in spike protein variants sequence	Specific mutation in spike protein variants
Wild type	Wuhan, China	YP_009724390	__a	—	—	—
D614G	United States	QTA38988	21-Mar-21	D614G	D614G	D614G
Alpha (B.1.1.7)	United Kingdom	P0DTC2	02-Jun-21	B.1.1.7	E484K, A570D, D614G, P681H	N501Y
Beta (B.1.351)	South Africa	7LYQ_C	09-Jul-21	B.1.351	K417N, E484K, N501Y, D614G	R246I
Delta (B.1.617.2)	India	QWP92316	12-Jun-21	B.1.617.2	E156, K417N, D614G, N501Y	T19R, L452R, T478K, P581R, P681R, D950N
Gamma (P.1)	Brazil	7M8K_C	28-May-21	P.1	K417N, E484K, N501Y, D614G	T20N, P26S, D138Y, R190S, T1027I
Omicron (B.1.1.529)	South Africa	7QO7_C	19-Jan-2022	B.1.1.529	N501Y, Y505H, T547K, D614G, H655Y, etc	A67V, del69-70, T95I, del142-144, Y145D, etc

^aNot available.

Physicochemical Properties and Posttranslational Modification Prediction

Physicochemical properties such as total amino acids, molecular weight, pI, grand average of hydropathicity (GRAVY), aliphatic index, etc, were predicted using the ProtParam tool [9]. Posttranslational modifications for all spike protein variants were predicted as follows: phosphorylation sites using the NetPhos 2.0 server [10]; glycosylation sites using the NetNGlyc 1.0 server [11]; and disulfide bonds using the Scratch Protein Predictor server [12].

Prediction of Immunoproperties

B-cell epitopes of all the variants were predicted using the ABCpred server [13] and exposed B-cell epitopes were predicted using the BepiPred 2.0 server [14]. T-cell epitopes and their immunogenicity were predicted using the IEBD Analysis Resource server [15], strong binding T-cells were identified using the NetMHCpan 4.1 server [16], and predications of cytotoxic T-lymphocytes were identified by the NetCTL 4.0 server [17].

Secondary and Tertiary Structure Prediction

Predications of secondary structures were identified by the PHYRE2 server [18] and the percentage of the following parameters were assessed: alpha helix, beta stand, transmembrane helix, and disorder. Subsequently, the tertiary structure was predicted using the Swiss model server [19] and the global model quality estimation, confidence, and coverage scores were collected. The structure was also analyzed for Ramachandran plot-allowed regions.

Generation of Phylogenetic Tree and 2D Gel Reference Map

The spike protein sequences of the wild type and variants were collected from NCBI, and the protein sequences of MERS-CoV and *Rousettus* bat coronavirus GCCDC1 were also obtained from NCBI with accession numbers AHY22525 and QKF94914, respectively. The phylogenetic tree was constructed using CLC sequence viewer and DNAMAN software [20] by neighbor joining method, and a virtual 2D protein map of all the spike protein variants was generated using the JVirGel V2.0 software [21].

Ethical Considerations

The study is registered with the institutional ethics committee of the Shri Dharmasthala Manjunatheshwara University (registration no. ECR/950/Inst/KA/2017/RR-21), Sattur, India. According to the institutional ethics committee guidelines, in compliance with the National guidelines/regulation on ethics in Biomedical Research, ethical clearance is not required for studies not involving human subjects/animals/patient medical records/tissues/biologicals fluids/pathogenic microorganism.

Results

In Silico Analysis

The results of an in silico analysis of the wild-type and variant protein sequences of spike protein revealed several common and specific mutations as listed in Table 1. Interestingly, the mutation P681H in the alpha variant [22] changed an amino acid from histidine (H) to arginine (R). Further, the analysis of physicochemical properties revealed that the pI and antigenicity of the delta variant were relatively high compared to other variants, and immunoproperties like B- and T-cell epitope sequences were different in the beta and delta variant compared to others. The phylogenetic tree and 2D gel maps clearly separated the delta variant from others.

Physicochemical Properties and Posttranslational Modification Prediction

The analysis of the physicochemical properties of all the spike proteins was interesting, especially the pI and aliphatic index, which ranged from 6.28 to 7.34 and from 82.04 to 84.65, respectively (Table 2). Among all the proteins, the delta variant had high pI (7.34) and high aliphatic index (84.65) when compared to other proteins (Table 2). However, all the proteins were stable. Meanwhile, the predicted total number of phosphorylation sites in the wild type was 133 whereas this prediction relatively decreased in the delta variant protein (Table 3). In addition, the number of serine, threonine, and tyrosine in phosphorylation sites varied among the variants compared to the wild type.

Predictions of the total number of N-glycosylation and disulfide bonds among the spike protein variants revealed numbers ranging from 16 to 20 and from 14 to 15 sites, respectively (Table 3).

Table 2. Comparison of predicted physicochemical property values of the spike protein of SARS-CoV-2 and its variants.

SARS-CoV-2 variants	Total amino acids	MW ^a	pI ^b	Extinction coefficient (M ⁻¹ cm ⁻¹)	EC/A ^c	Half-life (h)	Instability index	Classification of protein	Aliphatic index	GRAVY ^d
Wild type	1273	141178	6.24	148960	1.055	30	33.01	stable	84.67	-0.079
D614G	1252	138712	6.49	147345	1.062	30	32.06	stable	85.01	-0.067
Alpha (B.1.1.7)	1273	141178	6.24	148960	1.055	30	33.01	stable	84.67	-0.079
Beta (B.1.351)	1288	142201	6.38	148335	1.043	30	31.29	stable	82.03	-0.142
Delta (B.1.617.2)	1271	140895	7.34	148960	1.057	30	32.58	stable	84.65	-0.08
Gamma (P.1)	1257	139207	6.18	140315	1.008	30	31.15	stable	83.43	-0.127
Omicron (B.1.1.529)	1285	142424	6.63	146845	1.018	30	33.10	stable	81.84	-0.164

^aMW: molecular weight.^bpI: isoelectric point.^cEC/A: extinction coefficient/absorbance for 1% solutions.^dGRAVY: grand average of hydropathicity.**Table 3.** Comparison of phosphorylation, N-glycosylation, and disulfide bonds values of the spike protein of SARS CoV-2 and its variants.

SARS-CoV-2 variants	Phosphorylation		N-glycosylation		Disulfide bonds		
	Phosphorylation sites	Predicted Ser ^a , Thr ^b , and Tyr ^c sites	Predicted number	N-glycosylation sites	Predicted number	Total number of cysteine	Predicted number
Wild type	Ser, Thr, Tyr	Ser-64, Thr-44, Tyr-22	133	Asn ^d -Ser/Thr	17	40	15
D614G	Ser, Thr, Tyr	Ser-64, Thr-45, Tyr-22	131	Asn-Ser/Thr	17	38	15
Alpha (B.1.1.7)	Ser, Thr, Tyr	Ser-67, Thr-44, Tyr-22,	133	Asn-Ser/Thr	17	40	15
Beta (B.1.351)	Ser, Thr, Tyr	Ser-71, Thr-43, Tyr-24	136	Asn-Ser/Thr	17	30	14
Delta (B.1.617.2)	Ser, Thr, Tyr	Ser-65, Thr-43, Tyr-22	130	Asn-Ser/Thr	16	40	15
Gamma (P.1)	Ser, Thr, Tyr	Ser-66, Thr-43, Tyr-24	133	Asn-Ser/Thr	20	30	14
Omicron (B.1.1.529)	Ser, Thr, Tyr	Ser-68, Thr-43, Tyr-21	132	Asn-Ser/Thr	18	30	14

^aSer: serine.^bThr: threonine.^cTyr: tyrosine.^dAsn: asparagine.

Prediction of Immunoproperties

Predictions of the number of exposed B-cell epitopes for spike protein varied from 38 to 41 (Table 4). The B-cell epitope sequence for the wild-type, alpha, and gamma variants was QTQTNSPRRARSV, whereas this sequence changed for the D614G, beta, and delta variants. Among all the variants, the delta variant showed the highest score for protective antigen (0.4709) and antigenicity (0.7440; Table 4). Predictions for C-cell epitopes revealed that the number of epitopes ranged from 35 to 38 and the number of strong binders in T-cell were

from 21 to 23. The predicted T-cell epitopes for the wild-type and gamma spike variants were identical to IGINITRFQTLALH but changed in other spike protein variants. However, except in the alpha variant, the sequence QTLALH was supposed to be conserved (Table 4). The immunogenicity prediction scores for the spike protein variants varied. Meanwhile, statistical analysis of pI, antigenicity, and immunogenicity scores revealed the significance of the data, especially between alpha and wild type ($P=.007$), delta and wild type ($P=.02$), and delta and alpha ($P=.02$). The correlation matrix was positive with P values ranging from $<.007$ to $.04$ (Table 5).

Table 4. Comparison of the immunological properties of the spike protein of SARS-CoV-2 and its variants.

SARS-CoV-2 variants	Exposed B-cell epitopes	B-cell epitopes	Protective antigen prediction score	Predicted probability of antigenicity score	Number of epitopes identified in CTL ^a	Number of strong binders in T-cell	Predicted epitopes in T-cell	Immuno-genicity predication score
Wild type	40	QTQTNSPRRARSV	0.4646	0.717053	37	21	IGINITRFQTLALH	0.3751
D614G	40	YHKNNKS	0.4583	0.741478	35	22	ITRFQTLA	0.96257
Alpha (B.1.1.7)	40	QTQTNSPRRARSV	0.4646	0.717053	37	21	NGTHWFVTQRN-FYEP	0.3019
Beta (B.1.351)	40	HPQFEKGGSGGGGSG	0.4542	0.643558	38	23	QPYRVVLSFELLHA	1.23216
Delta (B.1.617.2)	38	SLGAENSVAYSN	0.4709	0.744007	35	22	IRAAEIRASANLAAT	0.0304
Gamma (P.1)	41	QTQTNSPRRARSV	0.4583	0.596261	36	22	IGINITRFQTLALH	1.07515
Omicron (B.1.1.529)	33	QTQTKSHGSASSVA	0.4646	0.717053	31	20	IGINITRFQTLALH	0.49637

^aCTL: cytotoxic T-lymphocyte.

Table 5. The *P* values of isoelectric point, antigenicity, and immunogenicity of the spike protein of SARS-CoV-2 and its variants.

SARS-CoV-2 variants	Wild type	D614G	Alpha	Beta	Delta	Gamma	Omicron
Wild type							
<i>r</i>	— ^a	0.996	1.000	0.989	0.999	0.992	1.000
<i>P</i> value	—	.06	.007	.09	.02	.08	.01
D614G							
<i>r</i>	0.996	—	0.995	0.998	0.992	0.999	0.998
<i>P</i> value	.06	—	.06	.04	.08	.03	.04
Alpha							
<i>r</i>	1.000	0.995	—	0.988	1.000	0.990	1.000
<i>P</i> value	.007	.06	—	.10	.02	.09	.02
Beta							
<i>r</i>	0.989	0.998	0.988	—	0.983	1.000	0.992
<i>P</i> value	.09	.04	.10	—	.12	.01	.08
Delta							
<i>r</i>	0.999	0.992	1.000	0.983	—	0.986	0.998
<i>P</i> value	.02	.08	.02	.12	—	.11	.04
Gamma							
<i>r</i>	0.992	0.999	0.990	1.000	0.986	—	0.994
<i>P</i> value	.08	.03	.09	.01	.11	—	.07
Omicron							
<i>r</i>	1.000	0.998	1.000	0.992	0.998	0.994	—
<i>P</i> value	.01	.04	.02	.08	.04	.07	—

^aNot applicable.

Secondary and Tertiary Structure Prediction

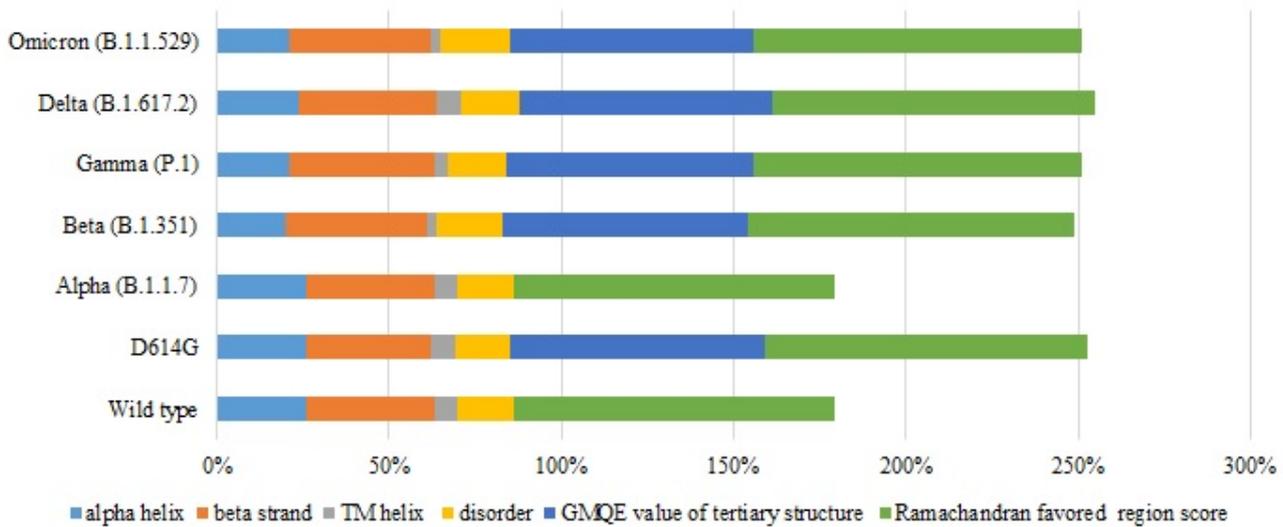
Secondary and tertiary prediction of all the spike protein variants showed that all protein structures were stable (Figure 1). The predicted proportion of alpha helices, beta strands,

transmembrane helices, and disorders of all the variants are shown in Figure 1. The alpha helix percentages varied from 26% in the wild type to 21% in the gamma variant, whereas the beta strand percentages varied from 37% in the wild type to 42% in the gamma variant; however, analysis of alpha and beta

percentages revealed that the proteins were stable. The global model quality estimate scores and Ramachandran favored regions—of which all the spike protein variants’ percentages

were similar to the wild type’s—indicated that there was no significant change in the stability of the variants compared to the wild type (Figure 1).

Figure 1. Graphical illustration of the predicted percentages of the secondary and tertiary structure and disorder of the spike proteins of SARS-CoV-2 wild type and its variants. GMQE: global model quality estimate; TM: transmembrane.

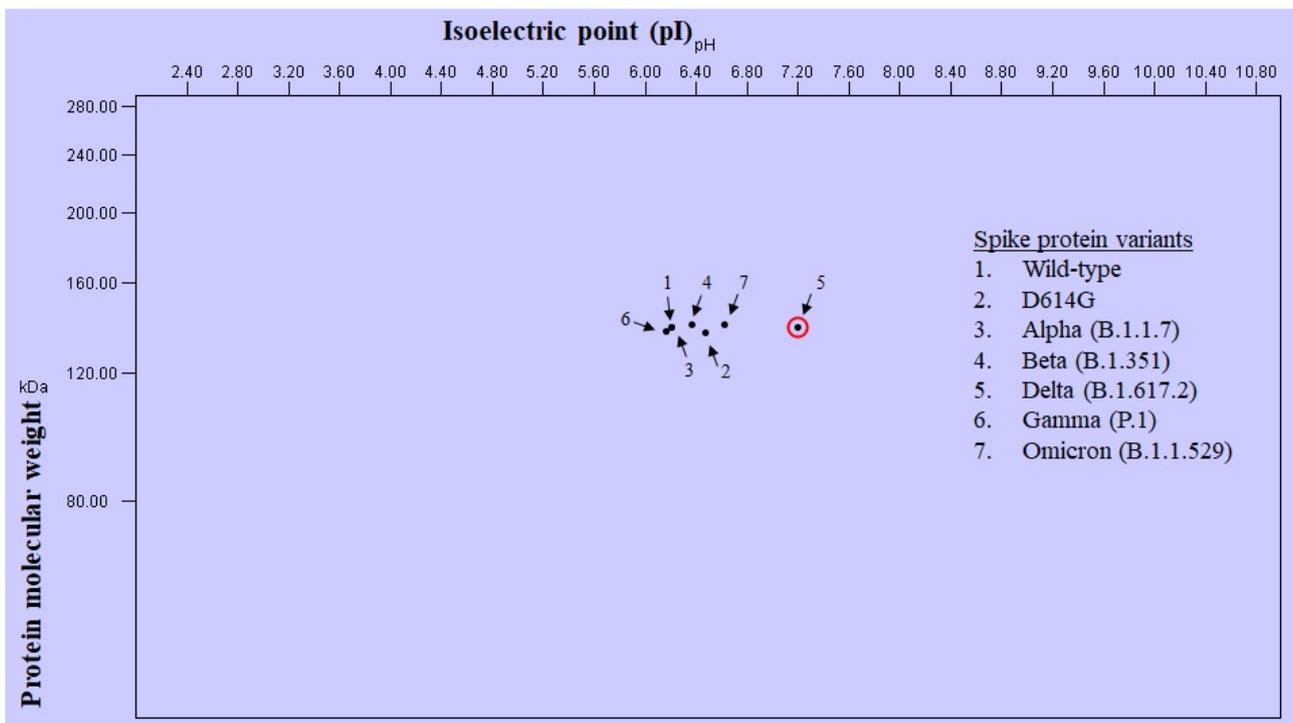


Generation of 2D Gel Reference Map and Phylogenetic Tree

The 2D reference map of the spike protein and its variants revealed a grouping of the wild-type, D614G, alpha, beta, gamma, and omicron variants whereas the delta variant was clearly separated (Figure 2). The phylogenetic tree of the spike protein of the wild-type and variant proteins along with MERS-CoV and bat coronavirus was constructed as seen in Figure 3. Construction of the phylogenetic tree grouped the

wild-type and alpha variants as one cluster and the beta, gamma, and omicron variants as another cluster. The D614G variant stood in between these 2 groups. Interestingly, the delta variant was closely grouped with MERS-CoV and *Rousettus* bat coronavirus. Meanwhile, the omicron variant showed 99% bootstrap values with the D614G variant, showing the closest relationship among all others. However, the tree has 2 branches: one with the wild type and variants (except delta) and another with the delta variant along with MERS-CoV and *Rousettus* bat coronavirus (Figure 3).

Figure 2. The 2D gel map of SARS-CoV-2 and its variants.



was observed in the D614G variant spike protein [27,28], where an acidic amino acid (D) was replaced with neutral one (G).

To see the effect of the pI on the biochemical properties, the wild-type and variant spike proteins were evaluated by predicting the 2D gel reference map, which interestingly demonstrated grouping except the separation of the delta protein on the map (Figure 2). The results of the physicochemical properties and 2D gel analysis revealed that the pI of a protein plays an important role in the behavior of proteins, especially where high pI turns protein more positive or aliphatic. Additionally, it is interesting to know that this property of protein is used in antibody preparation—to make a protein or its subunit more immunogenic [29,30]. Hence, we suppose that this property of proteins should be taken into consideration when designing and preparing prophylactic and therapeutic agents. Similarly, a recent study on the VOIs of SARS-CoV-2 virus particles that measured pI using chemical force microscope revealed VOIs have lower surface charge and hydrophobicity than the wild type, which might have played a role in VOIs having increased transmission ability [29]. Therefore, the pI of protein and the surface charge and hydrophobicity of viral particles are important factors to consider when designing prophylactics and therapeutics for any viral diseases [4,28,30,31].

Substantial changes in the physicochemical properties of the spike protein of wild-type SARS-CoV-2 and its variants has not been observed; however, a high number of phosphorylation sites (136) was observed in the beta variant and a low number (130) was observed in the delta variant. Meanwhile, a high number of N-glycosylation sites (20) was observed in the gamma variant and a low number (16) was observed in the delta variant. The total number of disulfide bonds among the wild type and variants did not vary much; however, it ranged from 14 to 15 and the number bonds were enough to give structural stability. Overall, the physicochemical properties did not differ markedly; however, the differences found were enough to influence the change in protein behavior and, subsequently, the behavior of the virus.

The antigenicity and immunogenicity scores of wild-type SARS-CoV-2 and its variants were interesting, especially the delta variant which showed high and low scores, respectively. Comparison of pI, antigenicity, and immunogenicity was made to evaluate the significance of the data; the *P* values were .007

between the alpha variant and wild type, .02 between the delta variant and wild type, and .02 between the delta and alpha variants. Overall, these *P* values suggest that the predicted values and the antigenic and immunogenic sequences are reliable. The number of exposed B-cell and cytotoxic T-lymphocyte epitopes were low in the delta variant compared to the wild-type and other variants of SARS-CoV-2. The exposed B-cell epitope sequence for the delta variant was SLGAENSVAYSN, and the change in epitope sequence showed that mutations did have an effect on the morphology of the virus.

The evolutionary relationship among the wild type and its variants was evaluated by constructing a phylogenetic tree using protein sequences. Interestingly, we found a clear grouping of the D614G, beta, and gamma variants, with bootstrap values of 100% for beta and gamma and 93% between D614G and beta/gamma. The delta variant was close to *Rousettus* bat coronavirus and MERS-CoV (Figure 3) with a bootstrap value of 68%. The alpha variant was close to the wild type with an 83% bootstrap value. Interestingly, we could observe a hybrid position of the delta variant between the wild type and the MERS-CoV and *Rousettus* bat coronavirus. Therefore, we suppose that the delta variant might be carrying characteristics from the wild type and other wild bat coronaviruses (Figure 3). In another interesting observation, the omicron variant seems to have evolved from the D614G variant with a bootstrap value 99%, as well as having imported some characteristics from the beta and gamma variants since omicron was also seen branching from these 2 variants with a 60% bootstrap value (Figure 3).

In conclusion, our study highlights that the accumulation of adaptive mutations in SARS-CoV-2 influenced the change in pI of the spike protein and, subsequently, the behavior of the virus. The prediction and comparison of the physicochemical properties of spike proteins of the wild type and its variants revealed that the delta variant displayed unique changes compared to the wild type. Evolutionary features showed a clear separation of the wild-type, alpha, and delta spike proteins and the grouping of the D614G, beta, and gamma spike proteins. Nevertheless, the continuous evolution of new SARS-CoV-2 strains demands further systematic understanding of its variants, which would not only help in developing the improvised rapid “antigen test” market but also in developing vaccines and therapeutics. Thus, more similar studies would unravel the important biochemical properties, evolutionary history, immunological behavior, and physiological properties of viruses.

Conflicts of Interest

None declared.

References

1. Brielle ES, Schneidman-Duhovny D, Linial M. The SARS-CoV-2 exerts a distinctive strategy for interacting with the ACE2 human receptor. *Viruses* 2020 Apr 30;12(5):497 [FREE Full text] [doi: [10.3390/v12050497](https://doi.org/10.3390/v12050497)] [Medline: [32365751](https://pubmed.ncbi.nlm.nih.gov/32365751/)]
2. Walls AC, Park Y, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020 Apr 16;181(2):281-292.e6 [FREE Full text] [doi: [10.1016/j.cell.2020.02.058](https://doi.org/10.1016/j.cell.2020.02.058)] [Medline: [32155444](https://pubmed.ncbi.nlm.nih.gov/32155444/)]
3. Tracking SARS-CoV-2 variants. World Health Organization. 2022 May 18. URL: <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/> [accessed 2022-05-19]

4. Mastriani E, Rakov AV, Liu SL. Isolating SARS-CoV-2 strains from countries in the same meridian: genome evolutionary analysis. *JMIR Bioinform Biotech* 2021 Jan 22;2(1):e25995 [FREE Full text] [doi: [10.2196/25995](https://doi.org/10.2196/25995)] [Medline: [33497425](https://pubmed.ncbi.nlm.nih.gov/33497425/)]
5. Ebrahim-Saraie HS, Dehghani B, Mojtahedi A, Shenagari M, Hasannejad-Bibalan M. Functional and structural characterization of SARS-Cov-2 spike protein: an in silico study. *Ethiop J Health Sci* 2021 Mar 01;31(2):213-222 [FREE Full text] [doi: [10.4314/ejhs.v31i2.2](https://doi.org/10.4314/ejhs.v31i2.2)] [Medline: [34158771](https://pubmed.ncbi.nlm.nih.gov/34158771/)]
6. Youden WJ, Denny FE. Factors influencing the pH equilibrium known as the isoelectric point of plant tissue. *Am J Bot* 1926 Dec 01;13(10):743-753. [doi: [10.1002/j.1537-2197.1926.tb05907.x](https://doi.org/10.1002/j.1537-2197.1926.tb05907.x)]
7. Shaw KL, Grimsley GR, Yakovlev GI, Makarov AA, Pace CN. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci* 2001 Jun;10(6):1206-1215 [FREE Full text] [doi: [10.1110/ps.440101](https://doi.org/10.1110/ps.440101)] [Medline: [11369859](https://pubmed.ncbi.nlm.nih.gov/11369859/)]
8. Higgins DG. CLUSTAL V: multiple alignment of DNA and protein sequences. *Methods Mol Biol* 1994;25:307-318. [doi: [10.1385/0-89603-276-0:307](https://doi.org/10.1385/0-89603-276-0:307)] [Medline: [8004173](https://pubmed.ncbi.nlm.nih.gov/8004173/)]
9. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. *The Proteomics Protocols Handbook*. Totowa, NJ: Springer Humana Press; 2005:571-607.
10. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 1999 Dec 17;294(5):1351-1362. [doi: [10.1006/jmbi.1999.3310](https://doi.org/10.1006/jmbi.1999.3310)] [Medline: [10600390](https://pubmed.ncbi.nlm.nih.gov/10600390/)]
11. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Singapore: World Scientific; 2001 Dec Presented at: Pacific Symposium on Biocomputing 2002; January 3-7, 2002; Kauai, HI p. 310-322. [doi: [10.1142/9789812799623_0029](https://doi.org/10.1142/9789812799623_0029)]
12. Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res* 2005 Jul 01;33(Web Server issue):W72-W76 [FREE Full text] [doi: [10.1093/nar/gki396](https://doi.org/10.1093/nar/gki396)] [Medline: [15980571](https://pubmed.ncbi.nlm.nih.gov/15980571/)]
13. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 2006 Oct 01;65(1):40-48. [doi: [10.1002/prot.21078](https://doi.org/10.1002/prot.21078)] [Medline: [16894596](https://pubmed.ncbi.nlm.nih.gov/16894596/)]
14. Jespersen MC, Peters B, Nielsen M, Marcantili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 2017 Jul 03;45(W1):W24-W29 [FREE Full text] [doi: [10.1093/nar/gkx346](https://doi.org/10.1093/nar/gkx346)] [Medline: [28472356](https://pubmed.ncbi.nlm.nih.gov/28472356/)]
15. Nielsen M, Lundegaard C, Lund O, Keşmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 2005 Apr;57(1-2):33-41. [doi: [10.1007/s00251-005-0781-7](https://doi.org/10.1007/s00251-005-0781-7)] [Medline: [15744535](https://pubmed.ncbi.nlm.nih.gov/15744535/)]
16. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res* 2020 Jul 02;48(W1):W449-W454 [FREE Full text] [doi: [10.1093/nar/gkaa379](https://doi.org/10.1093/nar/gkaa379)] [Medline: [32406916](https://pubmed.ncbi.nlm.nih.gov/32406916/)]
17. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* 2007 Oct 31;8(1):424 [FREE Full text] [doi: [10.1186/1471-2105-8-424](https://doi.org/10.1186/1471-2105-8-424)] [Medline: [17973982](https://pubmed.ncbi.nlm.nih.gov/17973982/)]
18. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 2015 Jun;10(6):845-858 [FREE Full text] [doi: [10.1038/nprot.2015.053](https://doi.org/10.1038/nprot.2015.053)] [Medline: [25950237](https://pubmed.ncbi.nlm.nih.gov/25950237/)]
19. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018 Jul 02;46(W1):W296-W303 [FREE Full text] [doi: [10.1093/nar/gky427](https://doi.org/10.1093/nar/gky427)] [Medline: [29788355](https://pubmed.ncbi.nlm.nih.gov/29788355/)]
20. Wang W. Erratum to: the molecular detection of *Corynespora Cassiicola* on cucumber by PCR assay using DNAMAN software and NCBI. In: Li D, Li Z, editors. *Computer and Computing Technologies in Agriculture IX*. CCTA 2015. IFIP Advances in Information and Communication Technology, vol 479. Cham, Switzerland: Springer; 2016.
21. Hiller K, Schobert M, Hundertmark C, Jahn D, Münch R. JVirGel: calculation of virtual two-dimensional protein gels. *Nucleic Acids Res* 2003 Jul 01;31(13):3862-3865 [FREE Full text] [doi: [10.1093/nar/gkg536](https://doi.org/10.1093/nar/gkg536)] [Medline: [12824438](https://pubmed.ncbi.nlm.nih.gov/12824438/)]
22. Liu Y, Liu J, Johnson BA, Xia H, Ku Z, Schindewolf C, et al. Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. *bioRxiv Preprint* posted online on September 05, 2021. [FREE Full text] [doi: [10.1101/2021.08.12.456173](https://doi.org/10.1101/2021.08.12.456173)] [Medline: [34462752](https://pubmed.ncbi.nlm.nih.gov/34462752/)]
23. Math RK, Kambiranda D, Yun HD, Ghebreyessus Y. Binding of cloned Cel enzymes on clay minerals related to the pI of the enzymes and database survey of cellulases of soil bacteria for pI. *Biosci Biotechnol Biochem* 2020 Feb;84(2):238-246. [doi: [10.1080/09168451.2019.1679613](https://doi.org/10.1080/09168451.2019.1679613)] [Medline: [31625450](https://pubmed.ncbi.nlm.nih.gov/31625450/)]
24. Grimsley GR, Scholtz JM, Pace CN. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci* 2009 Jan;18(1):247-251 [FREE Full text] [doi: [10.1002/pro.19](https://doi.org/10.1002/pro.19)] [Medline: [19177368](https://pubmed.ncbi.nlm.nih.gov/19177368/)]
25. Pace CN, Grimsley GR, Scholtz JM. Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J Biol Chem* 2009 May 15;284(20):13285-13289 [FREE Full text] [doi: [10.1074/jbc.R800080200](https://doi.org/10.1074/jbc.R800080200)] [Medline: [19164280](https://pubmed.ncbi.nlm.nih.gov/19164280/)]

26. Masunov A, Lazaridis T. Potentials of mean force between ionizable amino acid side chains in water. *J Am Chem Soc* 2003 Feb 19;125(7):1722-1730. [doi: [10.1021/ja025521w](https://doi.org/10.1021/ja025521w)] [Medline: [12580597](https://pubmed.ncbi.nlm.nih.gov/12580597/)]
27. Isabel S, Graña-Miraglia L, Gutierrez JM, Bundalovic-Torma C, Groves HE, Isabel MR, et al. Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. *Sci Rep* 2020 Aug 20;10(1):14031 [FREE Full text] [doi: [10.1038/s41598-020-70827-z](https://doi.org/10.1038/s41598-020-70827-z)] [Medline: [32820179](https://pubmed.ncbi.nlm.nih.gov/32820179/)]
28. Math RK, Goni M, Javaregowda PK, Khandagale AS, Oli A. SARS-CoV19-2 spike protein mutation patterns: a global scenario. Preprints Preprint posted on July 7, 2020. [doi: [10.20944/preprints202007.0132.v1](https://doi.org/10.20944/preprints202007.0132.v1)]
29. Areo O, Joshi PU, Obrenovich M, Tayahi M, Heldt CL. Single-particle characterization of SARS-CoV-2 isoelectric point and comparison to variants of interest. *Microorganisms* 2021 Jul 28;9(8):1606 [FREE Full text] [doi: [10.3390/microorganisms9081606](https://doi.org/10.3390/microorganisms9081606)] [Medline: [34442686](https://pubmed.ncbi.nlm.nih.gov/34442686/)]
30. Apple RJ, Domen PL, Muckerheide A, Michael JG. Cationization of protein antigens. IV. Increased antigen uptake by antigen-presenting cells. *J Immunol* 1988 May 15;140(10):3290-3295. [Medline: [3258879](https://pubmed.ncbi.nlm.nih.gov/3258879/)]
31. Domen PL, Muckerheide A, Michael JG. Cationization of protein antigens. III. abrogation of oral tolerance. *J Immunol* 1987 Nov 15;139(10):3195-3198. [Medline: [3680943](https://pubmed.ncbi.nlm.nih.gov/3680943/)]

Abbreviations

GRAVY: grand average of hydropathicity

pI: isoelectric point

VOI: variant of interest

Edited by A Mavragani; submitted 18.02.22; peer-reviewed by BS Chrisman, M Giri; comments to author 10.04.22; revised version received 25.04.22; accepted 16.05.22; published 30.05.22

Please cite as:

Math RK, Mudennavar N, Javaregowda PK, Savanur A

In Silico Comparative Analysis of the Functional, Structural, and Evolutionary Properties of SARS-CoV-2 Variant Spike Proteins
JMIR Bioinform Biotech 2022;3(1):e37391

URL: <https://bioinform.jmir.org/2022/1/e37391>

doi: [10.2196/37391](https://doi.org/10.2196/37391)

PMID: [35669291](https://pubmed.ncbi.nlm.nih.gov/35669291/)

©Renukaradhya K Math, Nayana Mudennavar, Palaksha Kanive Javaregowda, Ambuja Savanur. Originally published in JMIR Bioinformatics and Biotechnology (<https://bioinform.jmir.org>), 30.05.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Bioinformatics and Biotechnology, is properly cited. The complete bibliographic information, a link to the original publication on <https://bioinform.jmir.org/>, as well as this copyright and license information must be included.