A Study on the Relationship between the 3-D Structure of Spike Proteins and Infectiousness of SARS-CoV-2 Delta Variant

Ziyang Tang*

The High School Affiliated to Renmin University of China, Beijing, 100080, China
*Corresponding author: ziyangtang2022@163.com

Abstract. Since the outbreak of novel coronavirus pneumonia in Wuhan in 2019, the SARS-CoV-2 epidemic has become a hot topic. Over time, SARS-CoV-2 has evolved many variants. The diversity of the 3-D structure of the variant’s proteins resulted in the difference in the binding ability and infectious differences between different virus variants and human angiotensin-converting enzyme 2 (ACE2) receptors. In 2020, an evolutionary analysis of the Delta and Delta Plus variants of SARS-CoV-2 provided a three-dimensional model of the protein of the delta variant. However, it only focused on the delta variant and Delta plus variant themselves and did not compare the delta variant or delta plus variant with the original strain. It is hard to give a direct or apparent reason why the delta variant is more infectious and difficult to cure than the original strain. Therefore, this paper further compared the 3-D structures of homologous trimeric spike glycoproteins (S-proteins) and the receptor-binding domain between the SARS-CoV-2 original strain and the SARS-CoV-2 delta variant. By observing and analyzing the models of the above proteins in the PyMOL Molecular Graphics System, the reasons for the increase of infectivity of the delta variant can be interpreted in a direct way. This article also focuses on the data of the Indian cases from the JHU database to deeply analyze the relationship between the structure and transmission ability of the SARS-CoV-2 delta variant. Last but not least, the reproductive ability of SARS-CoV-2 can be reflected by the number of NAG (2-acetamido-2-deoxy-beta-D-glucopyranose). Through data analysis and protein structure research, we can better understand the characteristics of the binding of SARS-CoV-2 to the human receptor, thus providing a theoretical basis for accurately predicting virus variation. Through the comparative study of virus structure and infectiousness, this paper will provide a scientific basis for the relevant departments to improve epidemic prevention and improve the public’s vigilance against virus variants.

Keywords: SARS-CoV-2; Delta Variant; 3-D Structure of Spike Protein; Transmission.

1. Introduction

Since the discovery of pneumonia in 2019, a pandemic caused by a novel virus has severely affected the lives of people all around the world. According to the official website of the World Health Organization (WHO) data source, most people infected with the virus will experience mild to moderate respiratory illness and recover without requiring special treatment. However, some will become seriously ill and require medical attention. Older people and those with underlying medical conditions like cardiovascular disease, diabetes, chronic respiratory disease, or cancer are more likely to develop serious illnesses. Anyone can get sick with the virus and become seriously ill or die at any age. Therefore, the novel virus pneumonia has caused a worldwide pandemic and has become a hot topic of attention among all the academic circles. In 2020, through separation, second-generation sequencing, and testing, the Chinese researchers had identified the pathogen, a newly discovered coronavirus, and named it SARS-CoV-2[1].

The SARS-CoV-2 is a typical example of the single-stranded RNA virus. It is identified as Coronaviridae, Nidovirales due to its simplistic RNA single-chain structure.

All viruses, including novel coronavirus pneumonia, will mutate over time, and a simplistic RNA single strain allows the virus to mutate rapidly. Phylogenetic analysis of full-length genomes of 2019-nCoV and representative viruses of the genus Betacoronavirus [2] is shown in Figure 1.
Although most changes have little impact on the characteristics of the virus, some changes will affect the features of the virus to a certain extent, such as the difficulty of virus transmission. Because of this, the relevant authoritative organizations, such as WHO, expert groups, research institutes, and government researchers, have been working together to monitor and evaluate the evolution of SARS-CoV-2 since January 2020. By 2020, the latest variants of SARS-CoV-2 are classified as Variant of Interest (VOI) for specific needs and Variant of concern (VOC) that need to be noted. The details of VOI and VOC are shown in Table 1 and Table 2 [3].
We all know that alterations of the 3-D structure of the S-proteins, which are coded by the mutated RNA strain, help understand the biological characteristics of the virus and give direct indicators related to infectivity. Prediction by computational modeling method is an effective way to study the biological factors of a specific virus.

Therefore, this research aims to explore the relationship between the 3-D structure of the spike proteins and the transmission rate of the SARS-CoV-2 delta variant.

2. Analysis and results

2.1. Comparison and analysis of the geometries of S-proteins of original strain and delta variant

It is now discovered that Coronavirus uses S-proteins to bind to ACE2 receptors to infect the cell. This binding triggers a series of events, leading to the fusion between the cell and the virus membrane to enter the cell.

Therefore, the binding between S-proteins and the ACE2 receptors is crucial for the virus to successfully infect the cell and greatly impacts the transmission of the virus. This paper searched and collected 3-D structures of S-proteins of SARS-CoV-2 original strain and delta variant from the RCSB Protein Data Bank [5]. After that, PyMOL [6] was used to open corresponding files and show the protein interface, and then these 3-D structures on PyMOL were rotated, observed, and studied [6]. Moreover, this paper worked on measuring distances between atoms in the S-proteins and atoms in the corresponding ACE2 receptors. Benton et al. [7] did a research on the structural mechanism of the spike proteins of SARS-CoV-2 binding to human ACE2 receptors, which shows that the S-proteins have to be more open if the Coronavirus wants to use the S-proteins to bind to more ACE2.

| WHO label | Pango lineage* | GISAID clade | Nextstrain clade | Additional amino acid changes monitored* | Earliest documented samples | Date of designation |
|-----------|----------------|--------------|-----------------|-----------------------------------------|-----------------------------|---------------------|
| Alpha     | B.1.1.7        | GRY          | 20l (V1)        | +S:484K +S:452R                         | United Kingdom, Sep-2020    | 18-Dec-2020         |
| Beta      | B.1.351        | GH/501Y.V2   | 20H (V2)        | +S:L18F                                | South Africa, May-2020     | 18-Dec-2020         |
| Gamma     | P.1            | GR/501Y.V3   | 20J (V3)        | +S:681H                                | Brazil, Nov-2020           | 11-Jan-2021         |
| Delta     | B.1.617.2      | G/478K.V1    | 21A, 21l, 21J   | +S:417N +S:484K                        | India, Oct-2020            | VOF: 4-Apr-202 VOC: 11-May-2 |
| Omicron*  | B.1.1.529      | GR/484A      | 21K             | -                                       | Multiple countries, Nov-2021| VUM: 24-Nov-2 VOC: 26-Nov-2 |

| WHO label | Pango lineage* | GISAID clade | Nextstrain clade | Earliest documented samples | Date of designation |
|-----------|----------------|--------------|-----------------|-----------------------------|---------------------|
| Lambda    | C.37           | GR/452Q.V1   | 21G             | Peru, Dec-2020              | 14-Jun-2021         |
| Mu        | B.1.621        | GH           | 21H             | Colombia, Jan-2021          | 30-Aug-2021         |

Table 1. The label for VOC [4].

Table 2. The label for VOI [5].
receptors. Hence, it is reasonable and possible to deduce that the more open structure of the S-protein represents higher infectiousness. Therefore, this paper located the partial protein structures from similar positions of the SARS-CoV-2 original strain and SARS-CoV-2 delta variant to study the open situation of these structures. The 3-D structures of this S-protein of SARS-CoV-2 original strain and ACE2 receptor [8, 9] are shown in Figure 2. Then the 3-D structures of the S-protein of the SARS-CoV-2 delta variant and ACE2 receptor [8, 10] are shown in Figure 3. In addition, the partial structures of similar positions are shown in Figure 4 and Figure 5.

Figure 2. Overall structures of S-protein (shown in purple) of SARS-CoV-2 original strain and ACE2 receptor.

Figure 3. Overall structures of S-protein (shown in green) of SARS-CoV-2 delta variant and ACE2 receptor (shown in orange).
First of all, Figures 2 and 3 show that the structure of the S-protein of the SARS-CoV-2 delta variant is more open than the structure of the S-protein of the SARS-CoV-2 original strain. Moreover, the more open structure of the S-protein of the SARS-CoV-2 delta variant can also be reflected in Figures 4 and 5. One simple example can be given that the distance between two specific atoms in the partial structure of the SARS-CoV-2 original strain in Figure 4 is 8.5 Å and the distance between two corresponding atoms in the partial structure of the SARS-CoV-2 delta variant in Figure 5 is 9.2 Å. Above two points indicate that SARS-CoV-2 delta variant can use its S-proteins to bind to more ACE2 receptors. In other words, the SARS-CoV-2 delta variant has higher infectiousness than SARS-CoV-2 original strain does.

In addition, the number of NAG (2-acetamido-2-deoxy-beta-D-glucopyranose, shown in blue cube) in Figure 2 and Figure 3 may indicate the infectiousness of SARS-CoV-2 since it may interact with the receptor-binding domain (RBD) of SARS-CoV-2, and then it could affect the binding of RBD of SARS-CoV-2 by ACE2 [11]. Therefore, it needs or requires more NAG to be added to the RBD that is easy to bind to more ACE2 receptors. Based on this, Figure 2 and Figure 3 can also reflect that the SARS-CoV-2 delta variant binds with more NAGs. Consequently, it indicates that the SARS-CoV-2 delta variant can easily use its S-proteins to bind to more ACE2 receptors than SARS-CoV-2 original strain can do from another aspect.

Since the binding between S-proteins and ACE2 receptors is crucial for the virus to successfully infect the cell, we can figure out why the SARS-CoV-2 delta variant has higher infectiousness than SARS-CoV-2 original strain does from the above points.

All in all, through comparing, studying, and analyzing the 3-D structures of S-proteins of SARS-CoV-2 original strain and delta variant, we can conclude that the SARS-CoV-2 delta variant can easily use its S-proteins to bind to more ACE2 receptors than SARS-CoV-2 original strain can do, and it has higher infectiousness than SARS-CoV-2 original strain does. In other words, the people who carry the SARS-CoV-2 delta variant are more likely to infect other people than those carrying the SARS-CoV-2 original strain. Based on this, we can realize that it is necessary to take immediate
actions and strict policies to prevent the spread of the virus, especially facing the one with higher infectiousness after mutations. At the same time, this work can also help people learn the importance of studying on 3-D structures of proteins of the SARS-CoV-2 original strain and relevant variants.

2.2. Impact of mutations of SARS-COV-2 on the transmission rate of SARS-CoV-2 in India

In this work, the data about SARS-CoV-2 cases in India before and after the appearance of the SARS-CoV-2 delta variant was collected, and the 3-D structures of S-proteins of SARS-CoV-2 original strain and delta variant were studied. This is because the 3-D structure of the spike protein of SARS-CoV-2 original strain and delta variant both reflect the micro aspect. In contrast, the newly confirmed SARS-CoV-2 cases or other data reflect the severity in a macro aspect, intuitively indicating that the SARS-COV-2 delta variant is highly infectious and must be paid more attention to. Therefore, combining these two aspects with studying the SARS-COV-2 delta variant will be able to provide a novel, reasonable, and practical idea for analyzing the problems caused by SARS-CoV-2.

Since there are some rest days on weekends and holidays in some hospitals and institutions that can do the SARS-CoV-2 tests in India, the data related to the newly confirmed SARS-CoV-2 cases for each week are better to be tackled and explain some problems than the data of the newly confirmed SARS-CoV-2 cases for each day. Based on this situation, this paper combined the consideration of the maximum incubation period of SARS-CoV-2, 14 days. Hence, when handling the data related to SARS-CoV-2 in India before and after the appearance of the SARS-CoV-2 delta variant, two weeks was used as one unit for time. It means that the data related to the newly confirmed SARS-CoV-2 cases in India before and after the appearance of the SARS-CoV-2 delta variant every two weeks were used for the y-axis when creating the scatter plot with the time as x-axis (a time unit: two weeks).

Based on the above considerations and after making the scatter plot for the number of cases in India for every two weeks versus (vs.) the time, this research tested different regression or fitting methods by Matlab to let the computer learn from the past data in India, solve different optimization problems for different methods, find the best model, and output the best fitting results by using the best model that the computer trained. In other words, these data were handled to obtain the predicted number of cases, which may predict the data related to SARS-CoV-2 in the future to help figure out the proper and scientific policies or actions before the problems become serious because of mutations of the virus.

After testing various regression methods, the Gaussian Process Regression model based on the rational quadratic covariance function can give the best performance for the data in India before and after the appearance of the SARS-CoV-2 delta variant collected. The Gaussian Process Regression model is based on the Gaussian distribution (or called normal distribution) and is a kind of nonparametric regression model, which has a better adaptability and can provide good results for solving nonlinear problems. To some extent, it explains why the computer automatically tested different models and finally chose the Gaussian Process Regression model for the collected data instead of using the methods that work well for linear problems. The situation of the prediction by the Gaussian Process Regression (GPR) model for the number of newly confirmed COVID-19 cases every two weeks in India before and after the appearance of the SARS-CoV-2 delta variant is shown in Figure 6, which also presents the original data in blue and the predicted data in orange.
From analyzing Figure 6, we can deduce that the GPR model works well for predicting the number of newly confirmed SARS-CoV-2 cases every two weeks in India from May 5th, 2020, to October 24th, 2021. The first point means the first two weeks beginning on May 5th, 2020, corresponding to the period May 5th, 2020 – May 18th, 2020, and the second point corresponds to the second two weeks, and all points are obtained and shown in this way. During the second peak, there are two points with big errors, one is the 24th point, and the other one is the 27th point. It is found that these two points show the same phenomenon that the value of original data is larger than the value of predicted data, which means that the severity of SARS-CoV-2 could be much worse than what we could predict sometimes.

Besides, even though the delta variant was detected in India in October 2020, when it was the descending phase of the first peak, there was still the second peak after two months with so many cases in India. Also, it spread very fast for several months and caused a severe increase in SARS-CoV-2 cases worldwide.

Hence, the analysis of the past data can tell that we need to be careful and pay more attention for an extended period. We have to experiment to estimate the results caused by new variants of SARS-CoV-2 or other viruses to be as accurate as possible once we detect them, like studying and analyzing the 3-D structures of proteins of viruses.

The reason why combining above aspects to study on SARS-COV-2 delta variant is a good proposal for analyzing the problems caused by SARS-CoV-2 can be explained: (1) when we detect a new variant, we can study and quantify some properties of the 3-D structure of the variant; (2) then we can compare every 3-D structures of every variants before and original strain; (3) after that, we may use different regression methods or even various neural networks to learn from the past data about SARS-CoV-2 cases caused by different viruses and train the best models for corresponding viruses; (4) in the next step, we can find some clues from comparing the 3-D structure of each variant and its corresponding fitting model obtained from step (3), which will be able to allow us to deduce the appropriate model once we master the 3-D structure of a new variant in the future; (5) finally,
using the proper model to make some predictions for this new variant can approximately point out when it will be possible for the peak to shown up, when the speed of the increase of new cases is highest, and so on. Then we could adjust our relevant policies, provide feasible suggestions for people, and make some preparations to prevent the situation from getting worse early.

Therefore, combining the micro aspect of analyzing 3-D structures of proteins of viruses and the macro aspect of using proper prediction methods for relevant data can give feasible and reasonable applications for controlling SARS-COV-2.

3. Conclusions

Through comparing, studying, and analyzing the 3-D structures of S-proteins of the SARS-CoV-2 original strain and delta variant, we can conclude that the SARS-CoV-2 delta variant can easily make the usage of its S-proteins to bind to more ACE2 receptors than the original strain can do, and it has higher infectiousness than the original strain does.

Based on the rational quadratic covariance function, the GPR model can make the best fit and prediction for the data about SARS-CoV-2 cases from May 5th, 2020, to October 24th, 2021, in India.

Combining the micro aspect of studying and analyzing 3-D structures of proteins of viruses with the macro aspect of building appropriate machine learning models for the data related to the viruses is a novel, reasonable, and feasible direction for making contributions to scientific epidemic prevention in the future.

From the perspective of survival and evolution, reaching the maximum matching conditions with the receptor is the ultimate goal of the virus and, if possible, an exact match like a lock and its key. Therefore, by comparing and analyzing the geometries of S-proteins of the original strain and the delta variant, we should be able to connect the transmission rate of the viruses with the ability of the S-proteins to bind with the ACE2 receptors. This will be very scientific data that will contribute to scientific epidemic prevention and even predict whether the virus can infect humans in the future.

However, under the background of rapid virus variation, the function of 3-D structure prediction and protein analysis is limited because changes have brought some difficulties to human monitoring and analysis.

Therefore, only by keeping vigilant at all times and continuous dynamic monitoring of viruses that infect humans and livestock can cut off the occurrence of public health events similar to the novel coronavirus pandemic at the origin.

References

[1] Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new CORONAVIRUS associated with human respiratory disease in China. Nature, 579(7798), 265–269.

[2] Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., Chen, J., Meng, Y., Wang, J., Lin, Y., Yuan, J., Xie, Z., Ma, J., Liu, W. J., Wang, D., Xu, W., Holmes, E. C., Gao, G. F., Wu, G., Chen, W., Shi, W., Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet (London, England), 395(10224), 565-574. https://doi.org/10.1016/S0140-6736(20)30251-8

[3] Information from: https://www.who.int/zh/activities/tracking-SARS-CoV-2-variants

[4] RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences (2021) Nucleic Acids Research 49: D437–D451 doi: 10.1093/nar/gkaa1038
[5] D. Sehnal, S. Bittrich, M. Deshpande, R. Svobodová, K. Berka, V. Bazgier, S. Velankar, S.K. Burley, J. Koča, A.S. Rose (2021) Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. Nucleic Acids Research. doi: 10.1093/nar/gkab314

[6] DeLano W.L., The PyMOL Molecular Graphics System, 2002. DeLano Scientific, Palo Alto, CA, USA. PyMOL | www.pymol.org.

[7] Benton, D.J., Wrobel, A.G., Xu, P., Roustan, C., Martin, S. R., Rosenthal, P. B., Skehel, J. J., Gamblin, S. J. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. Nature 588, 327–330 (2020). https://doi.org/10.1038/s41586-020-2772-0

[8] Information from: https://www.rcsb.org/3d-view/6VW1

[9] Information from: https://www.rcsb.org/3d-view/7V8B/1

[10] Lan, J., Ge, J., Yu, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220 (2020). https://doi.org/10.1038/s41586-020-2180-5

[11] Information from https://gisanddata.maps.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6