Mutation Prone Point Within Spike Protein Molecule of SARS-CoV2: A Nanoinformatics Study

Beuy Joob 1,*1, Viroj Wiwanitkit 2

1 Priavate Academic Consultant, Bangkok Thailand; beuyjoob@hotmail.com (B.J.);
2 Department of Community Medicine, Dr DY Patil University, Pune, India; wviroj@yahoo.com (V.W.);
* Correspondence: beuyjoob@hotmail.com;

Abstract: Coronavirus disease 2019 (COVID-19) is a new emerging infection caused by a novel coronavirus, SARS CoV2. The pandemic of COVID-19 causes several problems worldwide. The interesting data from epidemiological surveillance show that the mutated virus already occurs and might result in a more serious outbreak. In this study, a standard combined bioinformatics analysis for SARS CoV2 nanostructure analysis was performed. The present study aims to determine the risk for mutation within the amino acid sequence of SARS CoV2. The standard tool, namely GlobPlot, was used for sequence analysis. According to the present study, the prediction of the positions resistant to a mutation within the spike protein of SARS CoV2 is made. There are many identified prone positions. This finding can imply that there will be a further new mutated type of SARS CoV2 pathogen. Therefore, the system for continuous molecular epidemiology surveillance is required.

Keywords: mutation; COVID-19; SARS CoV2; bioinformatics; spike; nanostructure.

1. Introduction

Nanotechnology is applicable to the management of medical problems. In nanomedicine, nanoinformatics approach can help clarify the problem of the pathogenesis of the disease. A possible advantage of the nanoinformatics analysis is analysis for mutation resistance within a molecule. Generally, the mutation is an important genetic phenomenon that might lead to increased severity of the disease. The mutation of a pathogen is usually mentioned for the association with change in pathogenesis.

In the present study, the authors focus interest on coronavirus disease 2019 (COVID-19), a newly emerging infectious disease. COVID-19 already causes pandemic worldwide, affecting more than 22 million people worldwide (data on 18 August 2020) [1–3]. Many factors lead to difficulty in disease control. An important consideration is on the molecular genetic change, the mutation, of the pathogen. Theoretically, predicting the mutation resistance/susceptibility of the pathogen is useful for risk monitoring and planning a surveillance system. To achieve those data, the applied nanoinformatics technology for identification of the peptide motifs within an amino acid sequence of the pathogen is possible. This approach can give the data on the vulnerable points or weak linkages in a protein of the pathogen [4]. In the present study, the authors use a nanoinformatics analysis for assessing the amino acid sequence of the spike protein of SARS CoV2. The objective is a prediction for vulnerable points of mutation within the studied molecule.
2. Materials and Methods

This work is a nanoinformatics study. First, data mining using database searching was performed. An international reference database, PubMed (www.pubmed.com), was used to find the amino acid sequence of spike protein S1 of SARS CoV2. The derived amino acid sequences were further used in weak linkage analysis. Regarding the weak linkage analysis, after getting the referencing sequence, another bioinformatics tool, namely GlobPlot, a web-based tool for analysis of the tendency for order/globularity and disorder, was used [5]. The standard protocol for analysis of mutation prone position within the amino acid sequence was used. This protocol is the same as fully described in the previous studies on other molecules by the team of authors [6-8].

3. Results and Discussion

The derived referencing amino acid sequences for further nanoinformatics analysis are 7BZ5_A, 6W41_C, 6XE1_E, 6XC3_C, 6XC7_A, 6XC4_Z, 6XC4_A, 6XC2_Z and 6XC2_A. After the analysis, according to the protocol, the prediction shows that many prone positions for mutation within the studied molecules (Table 1). From overall positions, the prone positions are identified in different percentages for each studied molecules. Given the fact that there are 46 kinds of amino acid, a permutation possibility calculation can show that there are numerous possible mutated variants.

| molecules  | Overall positions (N) | Prone positions (N) | Percentage of prone positions (%) |
|------------|-----------------------|---------------------|----------------------------------|
| 7BZ5_A     | 229                   | 175                 | 76.42                            |
| 6W41_C     | 231                   | 174                 | 75.32                            |
| 6XE1_E     | 273                   | 205                 | 75.09                            |
| 6XC3_C     | 231                   | 174                 | 75.32                            |
| 6XC7_A     | 231                   | 175                 | 75.76                            |
| 6XC4_Z     | 231                   | 175                 | 75.76                            |
| 6XC4_A     | 231                   | 175                 | 75.76                            |
| 6XC2_Z     | 231                   | 175                 | 75.76                            |
| 6XC2_A     | 231                   | 175                 | 75.76                            |

3.1. Discussion.

COVID-19 is a newly emerging viral infection that is still not successfully manage (situation in August 2020, 8 months after the first appearance of the disease). Lack of knowledge is a basic problem. To achieve the knowledge on the molecular aspect of the pathogen can help improve recognition of the new virus's pathogenesis. Some new reports already show that there are new emerging mutated variants of the new coronavirus. The clinical problem due to the variant is already confirmed [9–13].

Identifying the points of mutation within the sequence of the pathogen can give basic data for a further clinical, epidemiological study. The new advanced nanoinformatics approach can help solve the problem of time-consuming of the classic approach. In the present study, the author successfully performed an in silico analysis. The data on the mutation prone positions of the pathogenic virus is revealed. In the present study, many vulnerable positions of the pathogen structure are proven prone to mutation. It can say that the spike protein S1 of SARS CoV2 is highly labile, and mutation can easily occur. This might lead to new atypical clinical features as well as difficulty in vaccine development.
Further additional ontology analysis to find a possible significant effect on molecular expression is planned in the next study. According to our best knowledge, this is the first world report that estimates the magnitude of a possible mutation of the new coronavirus.

4. Conclusions

The applied nanoinformatics was used to analyze mutation prone positions within the spike protein of the SARS CoV2 molecule. From amino acid sequence analysis, there are many identified prone positions. Hence, the new mutated type of SARS CoV2 pathogen is likely to occur continuously and might cause a more COVID-19 severe outbreak. It is necessary to set a good system for continuous molecular epidemiology surveillance.

Funding

This research received no external funding.

Acknowledgments

This research has no acknowledgment.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Pascarella, G.; Strumia, A.; Pliego, C.; Bruno, F.; Del Buono, R.; Costa, F.; Scarlata, S.; Agrò, F.E. COVID-19 diagnosis and management: a comprehensive review. J Intern Med 2020, 288, 192-206, https://doi.org/10.1111/joim.13091.
2. Garrido, I.; Liberal, R.; Macedo, G. Review article: COVID-19 and liver disease-what we know on 1st May 2020. Aliment Pharmacol Ther 2020, 52, 267-275, https://doi.org/10.1111/apt.15813.
3. Jan, H.; Faisal, S.; Khan, A.; Khan, S.; Usman, H.; Liaqat, R.; Shah, S.A. COVID-19: Review of Epidemiology and Potential Treatments Against 2019 Novel Coronavirus. Discoveries (Craiova) 2020, 26, https://doi.org/10.15190/d.2020.5.
4. Lee, C; Wang, Q. Bioinformatics analysis of alternative splicing. Brief Bioinform. 2005, 6, 23–33, https://doi.org/10.1093/bib/bj6.1.23.
5. Linding, R.; Russell, R.B.; Nedova, V.; Gibson, T.J. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res. 2003, 31, 3701–8, https://doi.org/10.1093/nar/gkg519.
6. Wiwanitkit, V. Mutation in Rh48: Assessment for possible mutation prone point. Indian J Hematol Blood Transfus 2010, 26, 6-7, https://doi.org/10.1007/s12288-010-0003-9.
7. Wiwanitkit, V. (pro)renin receptor: A stable molecule. J Nat Sci Biol Med 2011, 2, 209-10, https://doi.org/10.4103/0976-9668.92321.
8. Wiwanitkit, S.; Wiwanitkit, V. Disrupted-in-schizophrenia 1 Mutation Prone Positions: A Pathogenesis of Schizophrenia. Indian J Psychol Med 2013, 35, https://doi.org/10.4103/0253-7176.112221.
9. Koyama, T.; Platt, D.; Parida, L. Variant analysis of SARS-CoV-2 genomes. Bull World Health Organ 2020, 98, 495-504, https://doi.org/10.2471/BLT.20.253591.
10. Ogawa, J.; Zhu, W.; Tonnu, N.; Singer, O.; Hunter, T.; Ryan, A.L.; Pao, G.M. The D614G mutation in the SARS-CoV2 Spike protein increases infectivity in an ACE2 receptor dependent manner. bioRxiv 2020, https://doi.org/10.1101/2020.07.21.214932.
11. Kaushal, N.; Gupta, Y.; Goyal, M.; Khaiboullina, S.F.; Baranwal, M.; Verma, S.C. Mutational Frequencies of SARS-CoV-2 Genome during the Beginning Months of the Outbreak in USA. Pathogens 2020, 9, https://doi.org/10.3390/pathogens9070565.
12. Matyášek R, Kovařík A. Mutation Patterns of Human SARS-CoV-2 and Bat RaTG13 Coronavirus Genomes Are Strongly Biased Towards C>U Transitions, Indicating Rapid Evolution in Their Hosts. Genes (Basel) 2020, 11, https://doi.org/10.3390/genes11070761.
13. Hassan, S.S.; Choudhury, P.P.; Basu, P.; Jana, S.S. Molecular conservation and differential mutation on ORF3a gene in Indian SARS-CoV2 genomes. Genomics 2020, 112, 3226-3237, https://doi.org/10.1016/j.ygeno.2020.06.016.