International Journal of Pharmaceutical and Bio-Medical Science

ISSN(print): 2767-827X, ISSN(online): 2767-830X

Volume 03 Issue 04 April 2023

Page No: 173-177

DOI: https://doi.org/10.47191/ijpbms/v3-i4-04, Impact Factor: 6.858

In Silico Approach to Construct the 3D Structures of Spike Glycoproteins of Novel Variants of Severe Acute Respiratory Syndrome Coronavirus 2

Shalini Swami

Department of Microbiology, Ram Lal Anand College, University of Delhi, Benito Jaurez Marg, South Campus, New Delhi

ABSTRACT

Objective: This study aims in constructing a three-dimensional modeled Spike glycoprotein structure of novel variants of SARS CoV-2.

Methods: The protein models were constructed using SWISS-Model online tool. The constructed protein models were submitted in online database called Protein Model Database (PMDB) for public access to the structures.

Results: A total of 70 protein sequences of Spike glycoprotein of novel variants of SARS CoV-2 were retrieved from NCBI virus database and were subjected for sequence similarity search and homology model construction. The constructed models were subjected for Ramachandran plot analysis to validate the quality of the structures. A total of 40 structures were considered to be of significant quality and were submitted to the online database PMDB.

Conclusion: These predicted structures would help greatly in identification and drug design. This would greatly help in drug development and personalized drug treatment against different variants of the pathogen. This database would significantly support the structure-based computational drug design applications toward personalized medicine against the variants of concern of SARS CoV-2.

ARTICLE DETAILS

Published On:
19 April 2023

KEYWORDS: SARS CoV-2, SWISS homology modeling, PMDB, Ramachandran Plot, Variants of concern, Spike glycoproteins, 3D structure.

Available on:
https://ijpbms.com/

INTRODUCTIONS

The drug discovery process in pharmaceutical industries relies on structure-based computer-aided drug design (SBCADD). Developing novel interventions with potential interaction with therapeutic targets is of paramount significance. The availability of the 3D structures of target proteins has led the foundation to design target-specific drugs based on structure-based drug design methods (1). Usually, analytical techniques like X-ray crystallography and nuclear paramagnetic resonance (NMR) are employed to construct the 3D dimensional structures of the target proteins. Conventionally, these methods are too expensive and time-consuming. To confound this problem, homology modeling aims to build the 3D structures of the protein based on the protein sequence similarity for which crystallographic structures are already available in the repository for the different organisms. The notion of an online tool SWISS Model, available to build the 3D structures of drug target proteins, is employed in this study (2-3, 4). The inbuilt computational algorithm in the SWISS Model is used to compare, match and analytically predict the 3D coordinates of amino acid residues with the pre-existing protein structures based on sequence similarity. (2,5-6).

The outbreak of novel severe acute respiratory syndrome coronavirus-2 (SARS CoV-2) infection has undergone significant mutation since it originated in Wuhan, Hubei province, China, in December 2019 (7). The explosion of SARS CoV-2 has severe morbidity and mortality levels reported by World Health Organization (WHO) in 2021. The mutation and adaptation to the existing environment have ignited concern about the spread of SARS-CoV-2 (8). Significantly numerous SARS-CoV-2 variants were produced ascribable to various mutations emerging within the RBD of the spike (9-10). Among the SARS CoV-2 variants produced, variants of concern (VOC) were identified to have increased transmissibility and virulence (11). It has also increases the flexibility of the spike proteins to interact with the host receptors (12-13).

Additionally, the VOC has reduced responses to known therapeutics (14). The mutational changes in the RBD of spike glycoproteins have caused structural modifications, eventually the product of amino acid alteration, which may affect viral pathogenesis (15). A better understanding of variant of concern possible structural alteration and similarity is required. To date, no structural data is available about these novel variants of SARS-CoV-2. Thus the computational approach has been used to model the structure of these proteins that pave the way to novel therapeutic strategies.

MATERIALS AND METHODS

NCBI bibliographic database

The research status of protein modeling concerning novel variants of concern of SARS CoV-2 and Omicron was analyzed using the NCBI and GISAID database. The NCBI database and GISAID (https://www.ncbi.nlm.nih.gov; https://gisaid.org) provide a robust knowledge of literature and scientific experiments. This allows us to establish a suitable strategy for this study. A search query of the virus generated all the information needed to know. (2)

Sequence retrieval

The NCBI virus is an integrative, value-added resource devised to endow the retrieval, display, and analysis of a curated collection of virus sequences and large sequence datasets. Protein sequences of the spike proteins of the novel variants of concern (VOC) were
retrieved from this database. It contains the protein sequences that are available as a. fasta file and stored for further study. The sequence analysis was conducted using the highlight sequence feature. The database search query had the gene accession number and pango lineage information with other parameters [3-4].

Sequence alignment

The pBlast (Protein Basic Local Alignment Search Tool) is an online search tool (https://blast.ncbi.nlm.nih.gov.). That finds the local similarity between nucleotide and protein sequences. (5, 16). The retrieved sequences were statistically analyzed and compared to the pre-existing repository of protein sequences of the Protein Database Bank (PDB). The aligned sequences sharing the highest percentage of similarity were shortlisted.

Structure Prediction

SWISS-MODEL is an entirely automated web-based homology-modeling server. It is accessible via the ExPASy web server, or the program DeepView (Swiss Pdb-Viewer) predicts the 3D structures of proteins. (21-22) This server aims to make protein modeling attainable globally for all life science researchers. Constructing a homology model comprises four significant steps: template structure identification (s), target sequence and template structure alignment(s), model-building, and model quality evaluation. (23). The retrieved sequences in fasta file format were used to construct the 3D protein structure using SWISS-MODEL. The template quality used to build the protein structure was analyzed using an inbuilt global and local quality estimation tool. (27-28) (https://swissmodel.expasy.org/). Multiple models were generated for each submitted protein and stored in the PDB format.

Model Analysis

The qualitative analysis of 3D protein structures was done via the MolProbity tool offered by a SWISS-MODEL. (2) The reliability and stereochemical quality of the modeled protein were inspected using the Ramachandran plot for qualitative estimation. (24-25) The favorability score, called Ramachandran's favored region, is designated as the confidence score for each modeled protein. (30) The Ramachandran region was also analyzed, for the conformation of *phi* and *psi* angles of the peptide bond, placing them in the favored region. (30-31)
The modeled protein having residues in the permitted region (Ramachandran outlier) and Ramachandran favored score was considered for screening the modeled protein. Only proteins that preconize 90% of the residues in the favorable, permitted region and having the highest confidence score were selected.

Model Submission

The predicted models were submitted to PMDB (http:// srv00.recas.ba.infn.it/PMDB/main.php),the public repository database. The manually built 3D structures of the protein are stored in this database. The models published in the scientific literature and validating experimental data can be accessed using this public repository database [2-3].

RESULTS

Building Homology Model

A comprehensive literature search and analysis was done using CoVsurver of GISAID and NCBI virus. The retrieved sequences of the novel variants of SARS CoV-2 from the SARS CoV-2 data hub of the NCBI virus were analyzed and subjected to homology model construction using the SWISS model web tool. The *in silico* tool yielded 1-5 structural protein models for each entry. The best Model for each protein was selected using the Ramachandran plot.

Ramachandran Plot Validation

The stereochemical quality of the 3D structure of proteins was validated using Ramachandran plot analysis. The reliability of the modeled protein structure was investigated using Ramachandran's favored score obtained by the Molprobity
inbuilt within the SWISS-Model online tool. The predicted models having 90% of its residues in the Ramachandran favored or permitted region were considered significant. Consequently, among the multiple models generated for each protein, the proteins exhibiting the highest percentage of residues in the Ramachandran favorable region. Fig. 1 shows

the Ramachandran plot analysis of a preferred model with >90% favored region and a least preferred model with <90% Ramachandran favored region. A reckoning of 40 modeled proteins depicted a substantial score in Ramachandran plot analysis and was selected for additional processing.

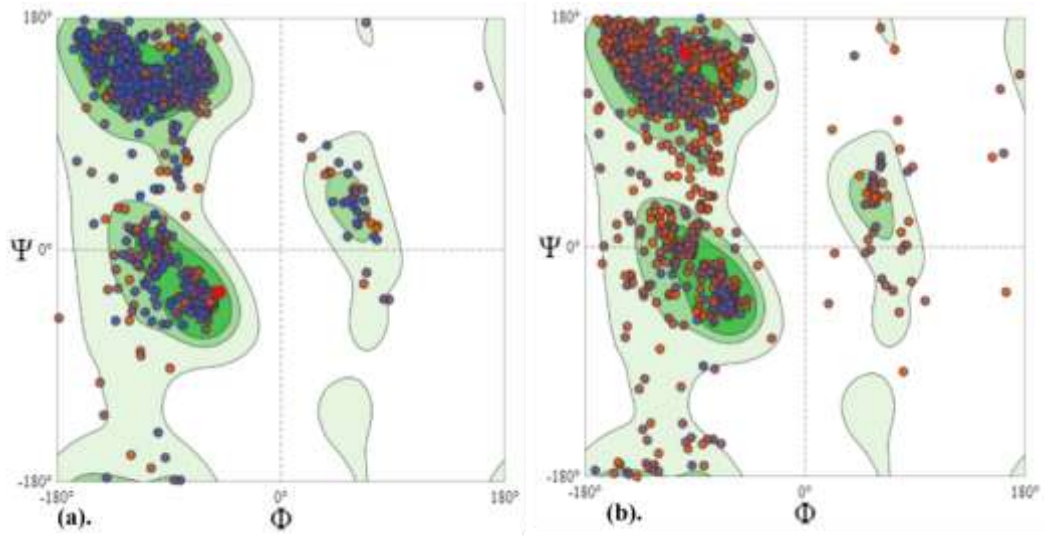

Figure 1: Ramachandran Plot Analysis (a). Showing the 3D model with >90% confidence score. (b). Rejected 3D model with <90% confidence score.

Submission to PMDB

The 40 proteins were selected based on their Ramachandran plot analysis and submitted to the online database of PMDB (https:// bioinformatics.cineca.it/PMDB/), made available to the public access for research purposes. Table 1. summarizes the details of the constructed Model, PMDB entry ID, and confidence score

Table 1: List of modeled Spike glycoproteins of novel variants of SARS CoV-2 with their PMDB ID

S.NO.	GENE ACCESSION NO..	VARIANT	CHAIN	PMDB ID	CONFIDENCE SCORE
1.	OQ338920.1	BQ.1	C	PM0084446	95%
2.	OQ338922.1	BQ.1.8	A	PM0084448	95%
3.	OQ338922.1	BQ.1.8	C	PM0084449	95%
4.	OQ338938.1	XBB.1	A	PM0084450	94%
5.	OQ338938.1	XBB.1	C	PM0084451	95%
6.	OQ327946.1	XBB	A	PM0084452	94%
7.	OQ327946.1	XBB	C	PM0084453	95%
8.	OQ332013.1	B.1.526	A	PM0084454	95%
9.	OQ332013.1	B.1.526	C	PM0084455	95%
10.	OQ344199.1	B.1.529	A	PM0084456	95%
11.	OQ314212.1	BA.1	A	PM0084457	95%
12.	OQ314212.1	BA.1	C	PM0084458	95%
13.	OQ333266.1	BA.2	A	PM0084459	95%
14.	OQ333266.1	BA.2	C	PM0084460	95%
15.	OQ343866.1	BA.2	A	PM0084461	95%
16.	OQ341918.1	B. 1.351	A	PM0084462	94%
17.	OQ326841.1	B.2.75.1	A	PM0084463	91%
18.	OQ316323.1	P.1	A	PM0084464	94%
19.	OQ344786.1	BF.11	A	PM0084466	95%
20.	OQ318434.1	BA.1.1	A	PM0084468	95%
21.	OQ345048.1	BA.5.2.1	A	PM0084470	95%
22.	ON286816.1	BA.2.9	A	PM0084473	92%
23.	ON286816.1	BA.2.9	C	PM0084474	90%
24.	OQ341628.1	BA.5.2	A	PM0084475	95%

25.	OQ333097.1	BA 2.12.1	A	PM0084476	95%
26.	OQ359575.1	CK.2.1.1	A	PM0084477	95%
27.	ON196873.1	B.1.351.3	A	PM0084478	95%
28.	OQ351080.1	BM.2	A	PM0084480	95%
29.	OQ344115.1	BM.4.1.1	A	PM0084481	95%
30.	OM304632.1	B.55	A	PM0084482	95%
31.	OQ380862.1	BA.2.21	A	PM0084483	92%
32.	OQ380862.1	BA.2.21	C	PM0084484	86%
33.	OQ380863.1	DF.1	A	PM0084485	93%
34.	OQ380878.1	BW.1	A	PM0084486	93%
35.	OQ380877.1	BA.5.2.34	A	PM0084487	92%
36.	OQ380879.1	BN.1.7	A	PM0084489	94%
37.	OQ380879.1	BN.1.7	C	PM0084490	91%
38.	OQ380879.1	BN.1.7.	A	PM0084491	92%
39.	OQ381306.1	XBF	A	PM0084492	92%
40.	BS007008.1	BA 5.1.1	A	PM0084495	93%

CONCLUSION

The study aimed at constructing the 3D protein structures of spike glycoprotein of the novel variants of SARS CoV-2 using a computational approach. The homology modeling was employed to build the modeled structures with a significant confidence score based on Ramachandran plot analysis. The modeled protein structures were deposited to the PMDB database computational protein model made available to the public. Hence, in this study, 3D structures of novel SARS CoV-2 variants of concern have been predicted and deposited to PMDB for public access. This could be significant for drug discovery and the development of targeted drugs that could inhibit the binding of spike glycoproteins with host receptor proteins. The ignited concern in the treatment of the SARS CoV-2 infection is the significant mutations found in the RBD of the spike glycoprotein, which eventually increases the virus transmissibility and evasiveness. Consequently, it is of utmost priority to understand and model the 3D structures of the novel variants of SARS-CoV-2. Thus, the homology modeling approach could help overcome the issue of rapid mutation development and sensitivity. This will also provide an advantage to the structure-based computational drug design studies on coronavirus organism, obliging in developing an effective drug variant to overcome the current challenges healthcare faces to impede the infection.

ACKNOWLEDGMENTS

The author would like to express special gratitude to Prof. Rakesh Kumar Gupta, Principal, and the University of Delhi, for supporting the research conducted. The author declares no competing interests.

DECLARATION OF CONFLICTING INTEREST

The author declares no conflict of interest.

FINANCIAL DISCLOSURE

No financial support was received for this study.

REFERENCES

I. Anderson AC. The process of structure-based drug design. Chem. Biol. 2003; 10:787–797
II. Sreenivas KA, Imran S, Ravi L. Elucidation of computational 3D Models of protein drug targets for Colletotrichum falcatum a fungal plant pathogen causing red rod of sugarcane. Biomed Pharmacol J 2020; 13:627-33.
III. Jindam D, Ravi L, Krishnan K. Construction of computational protein structure data base by Homology Modeling for the aquatic pathogen Perkinsus marinus for targeted drug design and development. Research J. Pharm. and Tech 2018; 11: 2203-2208.
IV. Feolo M, Helmberg W, Sherry ST, Maglott DR. NCBI genetic resources supporting immunogenetic research. Rev Immunogenet 2000; 2: 461-7.
V. Patel N, Prajapati N, Patel K, Patel R, Kalasariya H. Sequence Homology, Primer Designing and Homology Modeling Prediction by In Silico Pursuit. New Delhi: Microbiology in Services of Mankind; 2019.
VI. Kopp J. The SWISS-MODEL repository of annotated three-dimensional protein structure homology models. Nucleic Acids Res 2004; 32:230-4.
VII. Callaway E. Delta coronavirus variant: scientists brace for impact. Nature 2021; 595:17-
VIII. Salvatore M, Bhattacharyya R, Soumik Purkayastha S, et al. Resurgence of SARS-CoV-2 in India: Potential role of the B.1.617.2 (Delta) variant and delayed interventions. medRxiv 2021;25: 9405.

IX. Shu Y and McCauley J. Gisaid: global initiative on sharing all influenza data—from vision to reality. Euro Surveill. 2017; 22:3.

X. Khateeb J, Li Y and Zhang H. Emerging SARS-CoV-2 variants of concern and potential intervention approaches. Crit. Care. 2021; 25: 244.

XI. Koyama T, Platt D, and Parida L. (2020). Variant Analysis of SARS-CoV-2 Genomes. Bull. World Health Organ. 2020; 98: 495–504.

XII. Teruel N, Mailhot O, Najmanovich RJ. Modelling conformational state dynamics and its role on infection for SARS-CoV-2 Spike protein variants. PLoS Comput. Biol. 2021; 17:8.

XIII. Noh JY, Jeong HW, Shin EC. SARS-CoV-2 mutations, vaccines, and immunity: implication of variants of concern. Sig Transduct Target Ther. 2021; 6: 203.

XIV. Chen SC and Bahar I. Mining Frequent Patterns in Protein Structures: a Study of Protease Families. Bioinformatics. 2004; 20: 77–85.

XV. Gurung AB. *In silico* structure modelling of SARS-CoV-2 Nsp13 helicase and Nsp14 and repurposing of FDA approved antiviral drugs as dual inhibitors. Gene Rep. 2020; 21:100860.

XVI. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. The SWISSMODEL repository and associated resources. Nucleic Acids Res 2009;37: 387-92

XVII. Wu F, Zhao S, Yu B, et al. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* 2020; 579: 265-269.

XVIII. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403-10.

XIX. Ye J, McGinnis S, Madden TL. BLAST: improvements for better sequence analysis. Nucleic Acids Res. 2006; 34: W6-W9.

XX. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008; 36: W5-W9

XXI. Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. Electrophoresis. 2009; 30: 162-173.

XXII. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018; 46: W296-W303

XXIII. Gasteiger, E., et al., Protein Identification and Analysis Tools on the ExPASy Server, J.M. Walker, Editor. 2005, Humana Press. U.S.: U.S. p.571-607

XXIV. Arg E. MolProbity Ramachandran analysis. Proteins 2003; 437:180.

XXV. Hooft RW, Sander C, Vriend G. Objectively judging the quality of a protein structure from a Ramachandran plot. Bioinformatics 1997; 13:425-30.

XXVI. Hollingsworth SA, Karplus PA. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. Biomol Concepts 2010; 1:271-83.

XXVII. Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo - distance constraints applied on model quality estimation. Bioinformatics 2020; 36: 1765-1771.

XXVIII. Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Scientific Reports 7 2017.

XXIX. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 2011; 27: 343-350.

XXX. Laskowski R A, MacArthur MW, Moss DS. PROCHECK - a Program to Check the Stereochemical Quality of Protein Structures. *J. App. Cryst 1993;* 26:283-291.

XXXI. Wiederstein M & Sippl MJ. ProSA-web: Interactive Web Service for the Recognition of Errors in Three-dimensional Structures of Proteins. *Nucleic Acids Research* 2007; 35: 407-410.

XXXII. Sippl MJ. Recognition of errors in three-dimensional structures of proteins**.** Proteins 1993; 4:355-62.

XXXIII. Mirza MU, Froeyen M. Structural elucidation of SARS-CoV-2 vital proteins: Computational methods reveal potential drug candidates against main protease, Nsp12 polymerase and Nsp13 helicase. J Pharm Anal. 2020; 4:320-328.