ISSN: 2834-880X

Review Article

Current Research in Vaccines Vaccination

In-Silico Study of Virus-Host Protein-Protein Interactions (PPIs) can Anticipate and Cure Viruses by Reducing Time-Duration of Vaccine Trials: A Review

Avnish Kumar[1*], Bhuvnesh Prasad Sharma[2], Kavita Singh[3] , Vinita Gupta[3], Monika Asthana[1]

[1]*Department of Biotechnology, School of Life Sciences, Dr. Bhimrao Ambedkar University, Agra, Uttar Pradesh, India.*

[2]*Department of Biotechnology, Bhagwant University, Ajmer, Rajasthan, India,*

[3]*Department of Chemistry, Agra College Agra, Affiliated to Dr. Bhimrao Ambedkar University Agra, Uttar Pradesh, India*

[*]Corresponding author
Avnish Kumar, Department of Biotechnology, School of Life Sciences, Dr. Bhimrao Ambedkar University, Agra, Uttar Pradesh, India.

Submitted: 28 Feb 2023; **Accepted**: 23 Mar 2023: **Published**: 28 Mar 2023

Citation: Kumar, A., Sharma, B. P., Singh, K., Gupta, V., Asthana, M. (2023). *In-Silico* Study of Virus-Host Protein-Protein Interactions (PPIs) can Anticipate and Cure Viruses by Reducing Time-Duration of Vaccine Trials: *A Review. Curr Res Vaccines Vaccination, 2*(1), 31-35.

Abstract

COVID-19, a pandemic, has slowed global growth. This virus' fury and fast spread necessitated worldwide drug repurposing experiments. SARS-CoV-2 surface glycoprotein triggers host innate immunological and inflammatory responses. In-silico research characterised 6VSB B as more virulent than 6VSB A and important in pathophysiology. The worldwide assessment of SARS-CoV-2 virus sequences includes analysis from hundreds of high-coverage, high-quality genomes. Protein structure prediction predicts protein three-dimensional structures using amino acid sequences. SARS-CoV-2 was examined for isoelectric point, stability, GRAVY, amino acid, and atomic composition. Protein extinction coefficients quantify light absorption at a wavelength. Sequence data can compute protein molar extinction coefficients. The present review supports in-silico investigations using an immunological and molecular biology knowledge database to create novel vaccines and determine therapeutic targets when repurposing an existing drug.

Keywords: SARS-CoV-2, Coronavirus, Vaccine, Homology Modeling, EMBOSS

Introduction

Coronavirus disease 2019 (COVID-19) is the most recent mutation of the coronavirus family, Coronaviridae, that has become a global pandemic threat. The family shows a characteristic crown or halo like structure under electron microscopy as its envelope surface is adorned with glycoprotein studs. Since the first incidence was identified in December 2019 in Wuhan, Hubei Province, China, millions of COVID-19 patients have been diagnosed globally. On average, symptoms of this new coronavirus are expected to appear four days after infection (practically within two to seven days). According to modelling research, the incubation time may be as little as 2.2 days in 2.2 percent of people and 11.5 days in the rest [1-4].This virus has slowed global growth and economic productivity. On comparing its onset with the first onset of AIDS-related stigma in 1981, AIDS was more of a worry for family members but the financial burden had little effect on society [5].

There are two families of human coronaviruses: 229E-like and OC43-like. Both groups differ in their antigenic determinants and their culturing needs, as well as their cross-reactivity, and are thus linked with distinct outbreaks of the same disease [6]. The well-studied coronavirus species known as severe acute respiratory syndrome-related coronavirus may not adequately reflect COVID-19. As a result of our inadequate information and comprehension of the usual characteristics of this species, we are unable to regulate zoonotic disease transmission to humans. It is considered that its transmission is by airborne droplets; however, there is a lack of evidence of human coronavirus infection by cattle or other animals. COVID-19 can kill patients through severe pneumonia [7, 8], and due to the virus's virulence and quick rate of dissemination, medication repurposing has been tested globally.

The chloroquine derivative (hydroxy) chloroquine has demonstrated promising preventive and/or therapeutic actions in *vitro*. As an anti-inflammatory, chloroquine was formerly used for rheumatoid arthritis, lupus erythematosus, and malaria. Research suggests that

it may have antiviral properties against SARS-CoV. According to Yan & Savarino et al [9–10], chloroquine might elevate endosomal pH and inhibit the glycosylation of SARS-CoV cellular receptors. *In-silico* research is seen as a crucial step in resolving the existing top pandemic risks. Pathologically significant SARS-CoV-2 proteins are glycoproteins designated as 6VSB A, 6VSB B, and 6VSB C. These spike proteins target and damage the epithelium of the respiratory system resulting in lethal pneumonia. The involvement of these three glycoproteins in pathophysiology is being investigated further. A study revealed that 6VSB B is more virulent than 6VSB A and has a larger role in triggering inflammatory and innate immune responses in the host. Bioinformatics methods could predict the structure of these spike glycoproteins in order to understand their other characteristics. An antigenic peptide of a protein was thought to be beneficial for diagnostic and vaccination purposes [11]. The present mini-review is based on our research paper by Kumar and Sharma to support the in-silico examinations of various virulent proteins to combat novel infections and drug resistance; here we have taken the study of the surface glycoprotein of SARS-CoV-2 as an example [12].

Data Set

Computer systems and their ability for natural language processing and speech recognition have been radically altered by the advent of deep learning [13]. Recent applications in bioinformatics range from clinical image categorization to protein-DNA interaction prediction [12, 14]. Because of the availability of common data sets, traditional machine learning tasks are now more approachable for people without specific domain expertise. Data is information that has been prepared for transmission or processing. A previous study referred to NCBI's protein database [15]. The sequence of a surface glycoprotein with the accession number QHD43416.1 was obtained from the NCBI protein database. As a result of this database, authors were able to make early discoveries. The Global Evaluation of SARS-CoV-2 and hCoV-19 Sequences (GESS) (https://wan-bioinfo.shinyapps.io/GESS/) is a database that contains analysis findings from hundreds of high-coverage and high-quality SARS-CoV-2 whole genomes. In addition to displaying time-dependent patterns for single nucleotide variant (SNV) occurrences that represent the evolution of SARS-CoV-2 genomes, GESS provides regional distributions of SNVs around the world.

Table 1: Different type of Databases

S. No.	Data Type	Data base	Category	Example References used this Database
1.	RNA Sequencing Data	Open Science Framework: accession number doi:10.17605/OSF.IO/8F6N9	SARS-CoV-2 Transcriptomic Map	[16]
2.	Single cell transcriptomics data	https://www.covid19cellatlas.org/	COVID-19 Cell Atlas	[17]
3.	Genomic Epidemiology	https://nextstrain.org/sars-cov-2/	SARS-CoV-2 Strains	[18]
4.	Host Genetics Data (GWAS, WES, WGS)	https://www.covid19hg.org/	The COVID-19 Host Genetics Initiative	[19]
5.	Clinical Trial Related Information	https://clinicaltrials.gov/ct2/home	List of Clinical Trials	[20]
6.	DNA Sequencing Data	https://www.ncbi.nlm.nih.gov/ genbank/sars-cov-2-seqs/	SARS-CoV-2 Genome Sequencing Data	[21]
7.	Mass Spectrometry Raw Data	http://proteomecentral. proteomexchange.org/cgi/ GetDataset?ID=PXD018117	SARS-CoV-2 and Human Protein Interactions	[22]

Sequence Analysis

Sequence analysis is an important part of bioinformatics. It can be used to learn about the structure, function, or evolution of a biomolecule, like DNA, RNA, or a peptide. It is standard practice to deduce the structure, function, and evolutionary history of new or uncharacterized proteins by comparing them to experimentally documented proteins with which they share significant sequence similarities. It is well accepted that determining homologous relationships can begin with a comparison of sequences. Many high-quality sequence search algorithms have been created over the years, including BLAST HMMER HHblits and HHpred [23-27].

In earlier works the authors used the Protparam programme to analyse SARS-CoV-2 for its isoelectric point, stability, grand average of hydropathicity (GRAVY), amino acids, and atomic composition [12, 28]. If you have a protein sequence or know where it can be found in Swiss-Prot or TrEMBL, the software called ProtParam (http://www.expasy.org/tools/protparam.html) is useful to determine its physical and chemical properties. In ProtParam, protein molecular weight is determined by summing the average isotopic weights of the amino acids in the given protein and the average isotopic mass of a single water molecule. How much light a protein takes in at a particular wavelength is measured by its extinction coefficient. Protein molar extinction coefficients have been shown to be calculable from sequence data alone.

Protein pI is computed using amino acid pKa values. An amino acid's side chain is responsible for its pKa value, which is important in determining how a protein responds to changes in pH. The half-life of a protein is how long it is expected to take for half of the protein in a cell to vanish after it has been synthesised. Adding up the hydropathy values of each amino acid residue and dividing by the number of residues in the sequence or the length of the sequence yields the GRAVY value of a protein or peptide. Increasing hydrophobicity is represented by a positive score.

Homology Modeling

Computational structure prediction, the process of inferring the 3D structure of proteins from their amino acid sequences, often uses homology modelling, which is widely considered the most exact approach for predicting computational structures. The sequence alignment of template proteins allows homology modelling to provide an educated guess as to the 3D structure of a query protein. The four steps of homology modelling are target identification, sequence alignment, model development, and model refinement. Homology modelling relies on the fact that proteins with shared evolutionary heritage often have similar three-dimensional structures [5]. In molecular biology, homology modelling may be used to anticipate drug-design hypotheses, find ligand binding sites, establish substrate specificity, and annotate biological activities. The online homology modelling tool SWIS-MODEL was used by the authors to predict the effectiveness of the SARS-CoV-2 vaccine. [29]. we used the SWISS MODEL web service to create three different models of proteins. The energy expenditures of these three models were evaluated, and the one with the lowest total was chosen for further investigation. To compare the template protein (Surface glycoprotein SARS-CoV-2) and the unknown protein (Surface glycoprotein SARS-CoV-2) (Perfusion 2019-nCoV spike glycoprotein; 6VSB), the molecular visualisation software PyMoL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC) was used. PyMoL was used to keep track of the 6VSB and surface glycoprotein SARS-CoV-2 sequence homologs, and the latter was used as a template to generate a spatial structure file.

Validation of Structure

As part of the process of figuring out the structures of macromolecules, their structures are checked to make sure they are right. Some of the local validation criteria used in structure analysis before structure deposition are bond length and bonding angle aberrations, outliers in the Ramachandran plot, and clashing contacts. Local, national, and international rules about packing may not have been fully put into place yet [6]. In our previous article [12], we analysed torsional angles to validate protein properties, and for this, we developed the RamChandran Plot (ProSA-web). It's a way to illustrate the safe dihedral angles for the backbone in terms of energy [30]. Using Procheck, the stereochemical quality of the protein model was analysed [31–32]. Z-score plots (ProSA-web), also known as standard scores, demonstrated the extent to which individual data points deviated from the mean and provided insight into the overall quality of the model [30].

Antigenic Sites Prediction

Several methods are now being used to identify antigenic determinants and other protein features. Most of these calculations require using scales for each of the 20 amino acids. A different number on the scale represents each amino acid's likelihood of appearing on the protein surface or participating in the formation of secondary structures (such as helices or bends). All of the most reliable techniques employ a grading system based on the solubility of individual amino acids in water. Success rates are highest for hydrophilicity scales that prioritise charged and polar amino acids without being too discriminating for either positive or negative charges. Using this method, critical antigenic sites on proteins of well-studied pathogens may be identified with high accuracy. Many distinct types of protein interaction sites and membrane-spanning regions are included in the hydrophilicity profiles. A crucial part of developing a vaccine is identifying potential antigenic epitopes on the protein surface. The great majority of the epitope prediction algorithms that are now available make use of protein sequences in order to generate predictions regarding continuous epitopes that have a linear structure.

The authors predicted the SARS CoV-2 antigenic sites using the antigenic Emboss programme (https://www.bioinformatics.nl/cgi-bin/emboss/antigenic) [33]. Antigenic sites in a protein were predicted by feeding its sequence into the EMBOSS input module. Statistical analysis of experimentally identified antigenic sites on proteins suggest that the hydrophobic residues Cys, Leu, and Val are more likely to be part of antigenic sites if they are located on the protein surface [34]. EMBOSS protein analysis makes use of three different prediction tools: antigenic for predicting antigenic regions, sigcleave for predicting signal cleavage sites, and Garnier for predicting secondary structures. The EMBOSS Nucleotide Analysis plugin includes both the "tfscan" programme, which does a search for transcription factors, and the ‘tcode’ programme, which provides a prediction graph for protein coding. The "tfscan" tool looks for transcription factors, and the "tcode" programme predicts how proteins will be coded.

Protein-Protein Interaction Studies

Protein-protein interaction networks, also known as PPI networks, are platforms that provide us the opportunity to systematically find disease-related genes by looking at the connections that exist between genes that perform comparable activities [35]. It is necessary for SARS-CoV-2 to interact with the proteins of the host in the same way that other viruses do in order to enter host cells and copy its genome. Therefore, the discovery of virus-host protein-protein interactions (also known as PPIs) might be beneficial in predicting how the virus will behave and in the development of antiviral medications [36]. The experimental methods required for the identification of virus-host PPIs are highly time consuming and costly. For many different types of early research, computational methods could be a viable option.

Conclusion

The present review paper concludes that the identification of vi-

rus-host protein-protein interactions (PPIs) is useful for predicting the behaviour of viruses and developing antiviral drugs. In-silico researchers agree that computational approaches may be a feasible alternative to expensive and time-consuming experimental procedures.

References

1. Qun, L., Xuhua, G., Peng, W., Xiaoye, W., Lei, Z., Yeqing, T., ... & Zijian, F. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New England journal of medicine, 382(13), 1199-1207.
2. Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., ... & Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Annals of internal medicine, 172(9), 577-582.
3. 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.
4. Chan, J. F. W., Yuan, S., Kok, K. H., To, K. K. W., Chu, H., Yang, J., ... & Yuen, K. Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The lancet, 395(10223), 514-523.
5. Herek, G. M., & Capitanio, J. P. (1993). Public reactions to AIDS in the United States: a second decade of stigma. American journal of public health, 83(4), 574-577.
6. Baron, S., Fons, M., & Albrecht, T. (1996). Viral pathogenesis. Medical Microbiology. 4th edition.
7. Gao, J., Tian, Z., & Yang, X. (2020). Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. Bioscience trends, 14(1), 72-73.
8. Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., ... & Xiao, G. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell research, 30(3), 269-271.
9. Yan, Y., Zou, Z., Sun, Y., Li, X., Xu, K. F., Wei, Y., ... & Jiang, C. (2013). Anti-malaria drug chloroquine is highly effective in treating avian influenza A H5N1 virus infection in an animal model. Cell research, 23(2), 300-302.
10. Savarino, A., Boelaert, J. R., Cassone, A., Majori, G., & Cauda, R. (2003). Effects of chloroquine on viral infections: an old drug against today's diseases. The Lancet infectious diseases, 3(11), 722-727.
11. Kumar, A., Parkash, O., & Girdhar, B. K. (2014). Analysis of antigens of Mycobacterium leprae by interaction to sera IgG, IgM, and IgA response to improve diagnosis of leprosy. BioMed research international, 2014.
12. Kumar, A., & Sharma, B. P. (2020). In silico study to predict and characterize of SARS CoV 2 Surface glycoprotein. Vaccine Research, 7(1), 10-16.
13. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.
14. Ting, D. S., Liu, Y., Burlina, P., Xu, X., Bressler, N. M., & Wong, T. Y. (2018). AI for medical imaging goes deep. Nature medicine, 24(5), 539-540.
15. Alipanahi, B., Delong, A., Weirauch, M. T., & Frey, B. J. (2015). Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. Nature biotechnology, 33(8), 831-838.
16. Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. Cell, 181(4), 914-921.
17. Wilk, A. J., Rustagi, A., Zhao, N. Q., Roque, J., Martínez-Colón, G. J., McKechnie, J. L., ... & Blish, C. A. (2020). A single-cell atlas of the peripheral immune response in patients with severe COVID-19. Nature medicine, 26(7), 1070-1076.
18. Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., ... & Neher, R. A. (2018). Nextstrain: real-time tracking of pathogen evolution. Bioinformatics, 34(23), 4121-4123.
19. Kousathanas, A., Pairo-Castineira, E., Rawlik, K., Stuckey, A., Odhams, C. A., Walker, S., ... & Peterborough City Hospital team Butcher Deborah 153 154 O'Sullivan Susie 153 154 Butterworth-Cowin Nicola 153 154. (2022). Whole-genome sequencing reveals host factors underlying critical COVID-19. Nature, 607(7917), 97-103.
20. Gao, Q., Bao, L., Mao, H., Wang, L., Xu, K., Yang, M., ... & Qin, C. (2020). Development of an inactivated vaccine candidate for SARS-CoV-2. Science, 369(6499), 77-81.
21. Chiara, M., D'Erchia, A. M., Gissi, C., Manzari, C., Parisi, A., Resta, N., ... & Pesole, G. (2021). Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. Briefings in Bioinformatics, 22(2), 616-630.
22. Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., ... & Krogan, N. J. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature, 583(7816), 459-468.
23. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research, 25(17), 3389-3402.
24. Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server: 2018 update. Nucleic acids research, 46(W1), W200-W204.
25. Prakash, A., Jeffryes, M., Bateman, A., & Finn, R. D. (2017). The HMMER web server for protein sequence similarity search. Current protocols in bioinformatics, 60(1), 3-15.
26. Remmert, M., Biegert, A., Hauser, A., & Söding, J. (2012). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nature methods, 9(2), 173-175.
27. Nakamura, H., Schultz, E. E., & Balskus, E. P. (2017). A new strategy for aromatic ring alkylation in cylindrocyclophane biosynthesis. Nature Chemical Biology, 13(8), 916-921.
28. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server (pp. 571-607). Humana press.

29. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., ... & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic acids research, 46(W1), W296-W303.
30. Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research, 35(suppl_2), W407-W410.
31. Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography, 26(2), 283-291.
32. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of biomolecular NMR, 8, 477-486.
33. Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., ... & Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019. New England journal of medicine. 382:727–733.
34. Kolaskar, A. S., & Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS letters, 276(1-2), 172-174.
35. Pazos, F., Juan, D., Izarzugaza, J. M., Leon, E., & Valencia, A. (2008). Prediction of protein interaction based on similarity of phylogenetic trees. Functional Proteomics: Methods and Protocols, 523-535.
36. Llabrés, M., & Valiente, G. (2020). Alignment of virus-host protein-protein interaction networks by integer linear programming: SARS-CoV-2. Plos one, 15(12), e0236304.