Exploring Structural Variations in Covid-19 Variants: A Comprehensive 3D Analysis

Aryan Kothari
AIE22005
Amrita Vishwa Vidyapeetham
Bangalore, India

V.R.N.S Nikhil
AIE22062
Amrita Vishwa Vidyapeetham
Bangalore, India

V Harish
AIE22063
Amrita Vishwa Vidyapeetham
Bangalore, India

Namana Rohit
AIE22071
Amrita Vishwa Vidyapeetham
Bangalore

Dr. Amrita Thakur
Assistant Professor
Amrita Vishwa Vidyapeetham
Bangalore

***Abstract*—The dynamic evolution of SARS-CoV-2 introduces critical alterations in viral function, wherein residue replacements exert influence over physiochemical properties and folding conformations. Despite the abundance of available data on SARS-CoV-2, the effective analysis of how mutations reshape viral functions has been hindered by the absence of robust structural methodologies. In this study, we leverage the advancements in protein structure fingerprint technology to scrutinize the folding conformational changes induced by mutations. Through the integration of protein sequences and folding conformations, we align structures across SARS-CoV to SARS-CoV-2, encompassing the original Alpha variant (lineage B.1.1.7) to the latest Variant JN.1(lineage BA.2.75.2.). Our findings underscore the evolution of the virus, demonstrating shifts in mutational positions and physicochemical properties that enhance the affinity between the spike protein and ACE2—an integral factor in coronavirus entry into human cells. Our analysis unveils the progressive evolution of both structure and function within the coronavirus, elucidating the substantial contributions of SARS-CoV-2 variants to the global prevalence of severe acute disease.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

The three-dimensional structure of biological macromolecules, notably proteins, plays a pivotal role in governing their functions and interactions within cellular environments. Understanding the intricacies of these structures is essential for unravelling the molecular mechanisms that underlie diverse biological processes. In this study, we focus on the 3D structure of proteins, recognizing its significance in providing insights into the physicochemical properties of these macromolecules and facilitating the exploration of structural changes induced by genetic variations, such as mutations. To achieve this, we employ sophisticated computational tools, including Chimera and AlphaFold, harnessing their capabilities to model and analyze protein structures with high precision and efficiency.
There are many experimental ways of finding the structure of a protein, namely Xray crystallography or nuclear magnetic resonance (NMR) spectroscopy. One challenge associated with these methods lies in their time-intensive and cost-intensive nature, restricting their applicability to highly specific proteins.We now Harness the power of extremely powerful GPUs present in the cloud/locally to improve the rate of development in this field .
The selection of the coronavirus, specifically SARS-CoV-2, for our structural analysis holds scientific merit rooted in several key factors. First and foremost, SARS-CoV-2 is responsible for a global pandemic, making it paramount to comprehensively understand its molecular characteristics. Additionally, the virus exhibits unique structural features, particularly within its spike protein, which plays a central role in host cell entry. By focusing on coronavirus proteins, we aim to contribute valuable insights into the structural aspects that govern viral infectivity and pathogenicity. Moreover, studying the 3D structure of coronavirus proteins allows us to explore the impact of genetic variations on the virus's ability to interact with host cells, providing crucial information for the development of targeted therapeutic interventions. The global significance of SARS-CoV-2, combined with its distinctive structural attributes, makes it a compelling subject for our research, offering insights that extend beyond fundamental science to address urgent public health concerns.

II. Problem Statement

SARS-CoV-2 undergoes dynamic evolution, leading to crucial changes in its function through residue replacements that impact physiochemical properties and folding conformations. However, the analysis of how these mutations reshape viral functions faces obstacles due to the lack of robust structural methodologies. This study addresses these challenges by utilizing advanced protein structure fingerprint technology and computational tools like Chimera and AlphaFold. The aim is to comprehensively analyze the 3D structural variations induced by mutations in SARS-CoV-2, overcoming limitations in traditional approaches.

III. Literature Review

The paper titled "Homology Modeling and Molecular Dynamics-Driven Search for Natural Inhibitors That

Universally Target Receptor-Binding Domain of Spike Glycoprotein in SARS-CoV-2 Variants" proposes a system for identifying potential drug candidates that can efficiently inhibit the Receptor Binding Domain (RBD) of spike glycoproteins from different variants of SARS-CoV-2.The paper discusses the binding affinities of RBDs to ACE2 for different SARS-CoV-2 variants. The Alpha, Beta, and Gamma variants demonstrate enhanced affinities to ACE2 compared to Omicron and Delta variants. Homology modelling can be used to build new models for emerging variants, such as new clades of Omicron. The structures of the complexes showed fluctuations during simulations but were stabilized, with deviations not exceeding 2 A, except for the Delta RBD-ACE2 complex. The shortcoming of the paper is that it does not talk about the accuracy of the models and computational resources required.[1]

The paper titled "Highly accurate protein structure prediction with AlphaFold" delves into the CASP14 assessment, which is a community-wide competition in the field of computational biology and bioinformatics, with the software called AlphaFold.The architecture involves training with labelled and unlabeled data, including self-distillation and a Bidirectional Encoder Representations from Transformers (BERT)-style objective for interpreting phylogenetic relationships. The neural network is iteratively refined using recycling, contributing to its accuracy.On the CASP14 benchmark for protein structure prediction, AlphaFold achieved a median backbone accuracy of 0.96 Å, far surpassing all other methods.Despite its success, AlphaFold has limitations related to MSA depth and challenges in predicting proteins with few intra-chain or homotypic contacts compared to heterotypic contacts.[2]
The paper titled "Exploring the structural distribution of genetic variation in SARS-CoV-2 with the COVID-3D online resource" investigates the challenges posed by the COVID-19 pandemic and the ensuing global efforts to comprehend and combat the SARS-CoV-2 virus, declared a pandemic on March 11, 2020. A critical challenge lies in the accumulation of genetic variations in the virus, with approximately two variants emerging per month. This literature introduces the COVID-3D resource, designed to analyze and interpret over 11,000 detected variants in SARS-CoV-2 genomic sequences. The resource provides a three-dimensional viewer, facilitating the spatial visualization of variants and enabling the identification of co-evolutionary relationships and potential compensatory mutations. The authors delve into specific genetic variations, such as the prevalent QHD43416 p.Asp614Gly variant, emphasizing its location and potential impact on protein dynamics. The survey also explores therapeutic targets, particularly the spike protein and main proteinase, highlighting selective pressures and identifying genes under purifying selection for potential drug targets. Overall, this comprehensive literature survey underscores the urgency of understanding SARS-CoV-2 genetic variations and the significance of tools like COVID-3D in guiding therapeutic discovery efforts amid the ongoing global health crisis.[3]

The research paper "Biologically inspired ChaosNet architecture for Hypothetical Protein Classification" proposes using a biologically inspired neural network architecture called ChaosNet to classify hypothetical proteins (HPs). HPs are proteins with unknown functions that are important to study. The ChaosNet architecture uses chaotic neurons for feature extraction and classification. On a dataset of HPs classified by function, a single-layer ChaosNet model achieved 96% accuracy, comparable to other machine learning methods, while using less training data. Feature selection was also conducted and showed ChaosNet can reach 96% accuracy on a subset of features. Overall, ChaosNet is promising for HP classification with limited data.[5]

The paper titled "Uncertainty Visualization for Secondary Structures of Proteins" presents a technique to visualize the uncertainty in the secondary structure of proteins. The secondary structure is an abstraction of the 3D molecular structure based on the atomic coordinates. It represents local structures like helices and sheets. Since there is no consistent definition of secondary structure, assignments can vary across methods. The paper conveys the confidence per structure element and amino acid from multiple assignment methods. An uncertainty model is proposed to aggregate discrepancies across methods and simulation time steps. The model can flexibly map various sources of uncertainty to a single value per amino acid. The technique extends commonly used representations like sequence diagrams and ribbon diagrams using visual variables like distortion, transparency, frequency etc. to facilitate qualitative analysis of uncertainty. The effectiveness of the technique was evaluated through expert reviews on two applications - analyzing deviations between assignments and visualizing time-dependent changes from dynamics simulations.[6]

The research paper "Protein structure similarity based on multi-view images generated from 3D molecular visualization" proposes a new approach for protein structure comparison without the need for alignment. It uses canonical angles between subspaces generated from multiple views of different 3D visualizations of the protein structure like backbone, ribbons, rockets. By considering multiple visualizations, the protein structure is represented more elaborately. For each visualization, images are rendered from different viewpoints by rotating the protein model. Features are extracted from the images and Principal Component Analysis (PCA) is applied to create a subspace. The similarity between two proteins is measured by the canonical angles between their subspaces using the Mutual Subspace Method. Experiments on classifying four protein types from the SCOP database show that the proposed method performs better than common methods like Combinatorial Extension alignment and Gauss Integral Tuning descriptor.[7]

The research paper ”Comparative Analysis of Protein Complex Identification Methods Using Topological and Biological Information” discusses the importance of protein complexes and explores various methods for detecting them in Protein-Protein Interaction Networks (PPINs). It emphasizes the need for integrating both topological and biological context information for accurate complex detection. The study compares popular methods like MCL, COACH, MCODE, iPAC, and DPClus, highlighting the improvement in performance when incorporating gene expression data along with topological features. Evaluation measures and functional analysis demonstrate the significance of combining multiple types of information for predicting biologically relevant protein complexes.[8]

In this research article titled ”Visualization of COVID Bimodal Scan Using DNN,” a Bimodal Deep Neural Network (DNN) classifier is proposed for the rapid and accurate identification of COVID-19 through the fusion of chest X-ray and CT scan images. The model utilizes three datasets: COVID X-ray chest images, CT-scan images of SARS-CoV-2, and X-ray images of pneumonia cases. By leveraging deep learning and Grad-CAM for color visualization, the model achieves a total accuracy of 92.33 percent, with precision and recall rates of 94 percent and 93 percent, respectively. The significance of this work lies in its potential to detect positive COVID-19 cases more swiftly than traditional RT-PCR tests, aiding in early diagnosis and potentially saving lives. The study also establishes a relationship between COVID-19 and pneumonia cases. The integration of artificial intelligence, machine learning, and deep training is crucial for achieving efficient results in medical image classification. The paper concludes by presenting the proposed framework’s methods, performance measures, experimental results, and analysis.[9]

In this groundbreaking study titled ”Centralized CNN–GRU Model by Federated Learning for COVID-19 Prediction in India,” the researchers introduce a novel approach for the robust estimation of COVID-19 cases across 36 different provinces in India. The proposed federated-convolutional neural network–gated recurrent unit (Fed-CNN–GRU) model leverages the principles of transfer and federated learning to enhance its adaptability and accuracy in capturing the diverse transmission dynamics observed across the provinces. Unlike prior research that focused predominantly on deep learning models without addressing robustness concerns, this study takes a significant step forward by incorporating a federated learning approach, which allows the model to adapt to the evolving trends of the pandemic.[10]

This paper ”Protein Loop Modeling Using AlphaFold2” explores the performance of various AlphaFold2 variants on well-established loop modeling benchmark datasets. The authors introduce an efficient protocol named IAFLoop, specifically designed for loop modeling using AlphaFold2. The IAFLoop protocol involves providing a moderately extended segment of the target loop region as input to AlphaFold2, utilizing a fast version of AlphaFold2 with a reduced database, and employing RMSD-based consensus scores to select final output models.[11]

The paper ”Comparative Visualization of Protein Structure-Sequence Alignments” addresses challenges in protein fold recognition, specifically in visualizing and analyzing the three-dimensional structures of proteins based on similarity to known protein structures. The authors introduce a low-resolution molecular graphics tool that represents amino acid residues using abstract shapes or glyphs, reminiscent of Lego blocks. By focusing on major structural aspects and eliminating detailed side-chain orientations, the tool allows for the display of approximately double the information compared to high-resolution depictions. [12]

The paper ”A Visual Environment for Data Driven Protein Modeling and Validation” and is authored by a researcher who holds a chair in scientific visualization at Linköping University. The paper discusses the development of a visual environment for protein modeling and validation, focusing on the evaluation of protein structures, particularly in the context of cryo-electron microscopy (cryo-EM). It covers topics such as the development of metrics for evaluating protein structures, the use of visualization tools for structure validation, and the integration of various validation metrics into a visual framework. The framework consists of three main components: molecular visualization, a linear overview of the model provided by the residue quality visualization with metrics, and supporting plots like Ramachandran plots, scatter plots, and parallel coordinate plots. The paper also discusses the use of different data sources and processing methods, as well as the visual attributes and color maps used in the framework. The authors provide use cases to demonstrate the functionality and concepts of the framework, including the comparison between different protein structure models. The document is intended for publication in IEEE Transactions on Visualization and Computer Graphics[13]

The paper ”COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings” examines using AI and speech analysis of forced-cough recordings to screen for COVID-19, including asymptomatic cases. This is highly relevant given the need for large-scale, low-cost screening. The authors created a large balanced dataset of 5320 cough recordings to train machine learning models, using crowdsourcing. Preprocessing steps like chunking and MFCC feature extraction are explained. A convolutional neural network architecture using transfer learning from acoustic biomarker models is proposed. The use of complementary biomarkers related to muscular degradation, vocal cords, sentiments and lungs/respiratory tract is motivated well. Results demonstrate strong discrimination ability, with the sensitivity of 98.5 per cent and specificity of 94 per cent on

officially confirmed positive subjects. 100 per cent sensitivity on asymptomatics highlights utility as a screening tool. Saliency maps are generated to enable model interpretability and longitudinal tracking of individuals. A comparison with prior work on Alzheimer's diagnosis using the same biomarkers is discussed.[14]

The paper "Artificial Intelligence-based System in Protein Folding using Alphafold" discusses protein folding is a complex and unsolved problem in molecular biology. Understanding protein structures is crucial for drug discovery and biotechnology. The paper provides a good background on the protein folding process, associated challenges, and the relevance of computational structure prediction methods. AlphaFold has recently achieved significant breakthroughs in protein structure prediction using AI/deep learning. The paper summarizes AlphaFold's methodology and performance in CASP competitions. Comparison with conventional experimental techniques is provided. The paper explains AlphaFold's usage of attention-based neural networks and iterative refinement of predictions. Integration of evolutionary information through multiple sequence alignment is also highlighted as a key technique. Recent availability of the AlphaFold prediction database with 200 million protein structures is discussed. Applicability in various biological domains like drug discovery, antivirals etc is described. Comparative analysis of AlphaFold 1 and 2 is presented. Enhancements like the addition of a pre-processing pipeline and Evoformer neural architecture in AlphaFold 2 are covered.[15]

The paper titled "TemPred: A Novel Protein Template Search Engine to Improve Protein Structure Prediction" by Asmita Tripathi, Rajkrishna Mondal, Tapobrata Lahiri, Deepak Chaurasiya, and Manoj Kumar Pal. The paper introduces a new protein template search engine, TemPred, which utilizes a novel similarity criteria called ProtPCV2 to identify template proteins with known structures. The authors compare the performance of TemPred with the conventional search engine BLAST and demonstrate that the accuracy of predicted structural models can be significantly improved by combining the performances of both search engines. The paper provides a detailed methodology for the collection and conversion of protein sequence data into ProtPCV2, as well as the construction of the TemPred search engine. The authors also discuss the benefits of using fixed dimensional numerical vectors of protein sequences and the process of computing the Periodicity Count Values (PCV) for protein sequences.[16]

IV. Methodology

A. Software

This study utilizes advanced computational tools, primarily Chimera and AlphaFold, for comprehensive structural analysis of proteins. Chimera serves as a versatile platform for molecular visualization and analysis. At the same time, AlphaFold, a state-of-the-art deep learning model, employs deep learning techniques to analyze genetic sequences and predict the 3D structures of proteins, including spike proteins, by interpreting the complex relationship between amino acid sequences and their folded configurations with remarkable accuracy.

B. Input

Alpha variant (B.1.1.7): Alpha variant is characterized by increased transmissibility and a higher risk of hospitalization.Was one of the earliest variants.

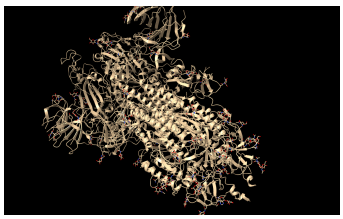


Fig. 1. Alpha Variant

Beta variant (B.1.351): Beta variant is known for mutations that impact vaccine efficacy and are associated with higher resistance to neutralization by antibodies.

Delta variant (B.1.617.2): Delta variant gained prominence for its heightened transmissibility and potential increased severity of illness, contributing to a surge in COVID-19 cases worldwide.

Omicron variant (BA.2.12.1): The Omicron variant carries a significant number of mutations, particularly in the spike protein, resulting in increased transmissibility and concerns about vaccine effectiveness.

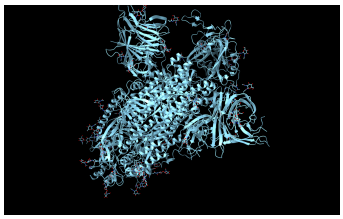


Fig. 2. Omicron Variant

JN.1 variant (BA.2.75.2): This Omicron subvariant exhibits specific genetic differences, prompting ongoing research to understand its implications for transmissibility, severity, and potential impact on public health measures.

For the structural analysis of all the variants, we make use of the genome structural code and spike protein of a variant to analyse the differences and similarities among the proteins. The variants we use are present in Table 1.

C. Genetic Sequence

The genetic sequences for the selected COVID-19 variants were obtained from the National Center for Biotechnology

TABLE I
VARIANTS OF SARS-COV-19

Variant	Lineage	Date of Origin
Wuhan-Hu-1	-	December 2019
Alpha	B.1.1.7	September 2020
Beta	B.1.351	May 2020
Delta	B.1.617.2	October 2020
Omicron	BA.2.12.1	November 2021
JN.1	BA.2.75.2	June 2023

Information (NCBI) database, a reputable repository for biological information. The genetic information was extracted in the form of FASTA files from the NCBI website. These FASTA files contain the primary nucleotide sequences of the viral genomes. To analyze the three-dimensional structures of the viral proteins, we focused on the spike (S) protein, specifically its subunits S1 and S2. These two structures are combined to form a comprehensive representation of the entire spike protein structure. These subunits were identified as critical regions for host cell entry and are commonly targeted in structural studies.

V. RESULTS AND DISCUSSION

Following the acquisition of the structural data for individual COVID variants through diverse sources including the AlphaFold, NCBI, RCSB, and GSID databases, we employed the visualization tool known as ChimeraX. The resultant structural representation is presented below in Fig 1.

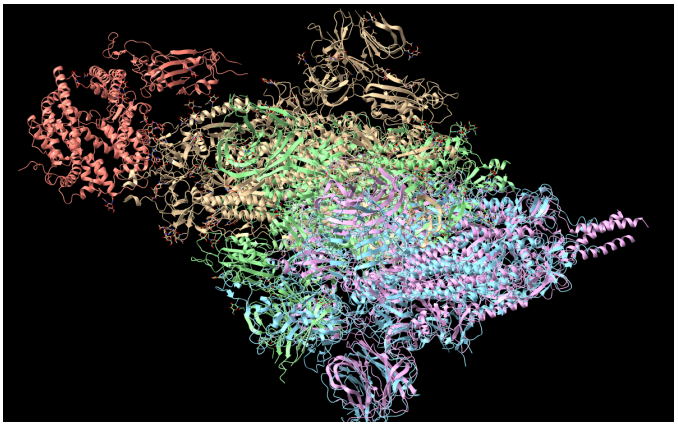


Fig. 3. Overall Comparision

A. Sequence Alignment

A sequence alignment score is a number that measures how similar two biological sequences, like DNA, RNA, or proteins, are to each other.
Fig 2.We can see that the sequence alignment score of 4613.1 suggests a significant degree of similarity among all the variant protein structures, indicating a shared origin and common elements throughout. This supports the verification of mutation among newer variants of SARS-Cov-2.

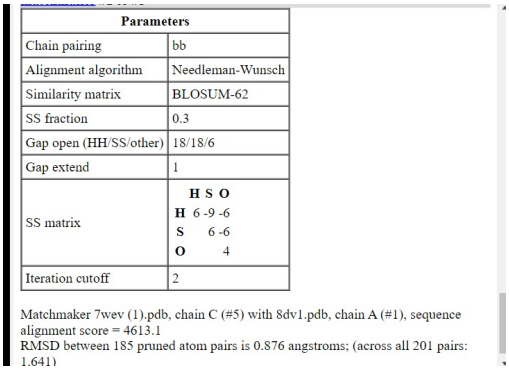


Fig. 4. ChimeraX Matchmaker

B. Similarity Matrix

A Similarity matrix can be seen in chimraX Matchmaker.The SS matrix visualizes how similar each part of a molecule is to every other part, like a color-coded map revealing patterns and relationships.We notice the following information from our SS matrix.

1) High Overall Similarity: The predominantly yellow and orange colors suggest a high degree of sequence conservation between the compared virus strains. This indicates that most of the amino acids in their genomes are identical, likely reflecting core viral functions essential for survival and replication.

2) Local Variations: The scattered regions of green and light yellow represent areas with lower similarity, hinting at the presence of mutations. These mutations could be

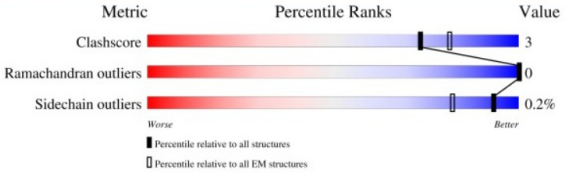
3) Neutral: Having minimal impact on viral fitness or transmission.

4) Adaptive: Conferring some advantage to the virus, such as enhanced immune evasion or transmissibility.

5) Deleterious: Clumsy mistakes that weaken the virus, sometimes leading to dead-end lineages.

C. Ramachandran plot

Ramachandran plot is a graphical representation of the dihedral angles in a protein's backbone. Dihedral angles describe the rotation about a bond between two atoms, and in the context of proteins, they are crucial for understanding the conformation of the polypeptide chain.
The Omicron variant boasts stellar structural stability, evidenced by its zero Ramachandran outliers, a miniscule 0.2 percent sidechain outlier rate, and a likely top-tier Clashscore. Compared to the vast archive of protein structures, Omicron shines: its Ramachandran and sidechain outlier counts rank among the top 0.1 percent, and its Clashscore likely lands it within the high percentiles. This exceptional Ramachandran plot paints a picture of a well-defined backbone and sidechains, suggesting a high-quality, structurally sound Omicron variant protein. While percentile ranks for Clashscore and Sidechain outliers relative to all and EM structures would give even more context, the available data strongly hints at a structurally robust Omicron variant.



Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

Fig. 5. Ramchandran outline on Omicron Variant

D. RMSD Number

The Root Mean Square Deviation (RMSD) value of 0.876 Å in the comparison of protein structures related to different variants of COVID-19 indicates a favorable alignment, suggesting structural similarity among these variants. The use of Pruned RMSD, which excludes uncertain regions, enhances the precision of the comparison. With 185 pruned atom pairs involved in the RMSD calculation, the analysis focuses on confidently aligned residues, providing insights into conserved regions across COVID-19 variants. However, the absence of complete alignment details, such as specific protein names, alignment algorithm, and additional scores, limits a comprehensive understanding of the structural differences and similarities in the context of these variants..

VI. Research Gaps

A. Computational Resources:

AlphaFold requires substantial computational resources. Implementing it on standard hardware may be resource-intensive, limiting access for smaller research groups or institutions with limited computational capabilities.

B. Function Prediction:

While AlphaFold predicts the 3D structure of proteins, it does not directly provide information about protein function. Experimental validation is often necessary to understand the functional implications of a predicted structure.

C. Biological Context:

Protein function is not solely determined by structure; it also involves interactions with other molecules and cellular environments. Both Chimera and AlphaFold primarily focus on structure and may not provide a comprehensive understanding of a protein's biological context.

D. Dependence on Input Quality:

The accuracy of predictions is heavily dependent on the quality of the input data. If the initial protein structure data is of low quality, the predictions generated by these tools may not be reliable.

VII. Future Enhancements

A. Genomic Surveillance and Sequencing

Increasing investment in genomic surveillance is vital to continually monitor the virus's evolution. Advanced sequencing technologies must be developed and implemented to improve the speed and accuracy of identifying variants. These enhancements are critical for timely detection, analysis, and response, enabling a more effective public health strategy to address the evolving nature of COVID-19 and its variants.

B. Data Integration and Analysis

Integration of genomic data with clinical, epidemiological, and demographic information is essential for a holistic understanding of variant behavior. Developing advanced data analysis techniques, such as machine learning and artificial intelligence, is crucial for identifying patterns and predicting potential future variants. This comprehensive approach enhances our ability to preemptively address emerging threats, enabling a more informed and adaptive public health response to the evolving landscape of COVID-19 and its variants.

C. Virus Evolution Studies

Conducting thorough virus evolution studies is imperative to anticipate future changes. In-depth research should explore evolutionary patterns, including factors influencing variant emergence, like host immune responses and viral interactions. Understanding these dynamics aids in predicting potential developments, informing proactive measures. By unraveling the intricacies of viral evolution, the scientific community can better prepare for and mitigate the impact of emerging variants, contributing to more effective strategies for managing the ongoing challenges posed by the COVID-19 virus.

VIII. Conclusion

In conclusion, this study has successfully demonstrated the utility of advanced computational tools like Chimera and AlphaFold for analyzing the three-dimensional protein structures of SARS-CoV-2 variants, obtaining molecular models of the spike proteins from major variants including Alpha, Beta, Delta and Omicron. Structural analysis revealed a high degree of sequence and structure similarity pointing to shared origins, as well as local variations likely representing adaptive mutations. Ramachandran plot analysis supported Omicron's structural stability, while RMSD values indicated favorable alignments between variants, highlighting conserved regions amidst mutational changes. Overall, by applying computational methodologies, this research has enriched understanding of how mutations reshape the physicochemical properties and folding patterns of SARS-CoV-2 variants, underscoring the ongoing evolution of structure-function relationships and providing molecular perspectives into epidemiological trends. Moving forward, this integrated framework can be expanded through collaborations and further research into evolution patterns and emergence factors. This study exemplifies the indispensable role of computational molecular modeling in

deciphering viral mechanisms and informing public health strategies against the dynamically mutating COVID-19 virus.

REFERENCES

[1] Ovchynnykova O, Kapusta K, Sizochenko N, Sukhyy KM, Kolodziejczyk W, Hill GA, Saloni J. "Homology Modeling and Molecular Dynamics-Driven Search for Natural Inhibitors That Universally Target Receptor-Binding Domain of Spike Glycoprotein in SARS-CoV-2 Variants". Molecules. 2022 Oct 28;27(21):7336. doi: 10.3390/molecules27217336. PMID: 36364158; PMCID: PMC9657887.

[2] Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

[3] Portelli, S., Olshansky, M., Rodrigues, C.H.M. et al. Exploring the structural distribution of genetic variation in SARS-CoV-2 with the COVID-3D online resource. Nat Genet 52, 999–1001 (2020). https://doi.org/10.1038/s41588-020-0693-3

[4] Jin-Hong Kim, Geon-Tae Ahn, Myung-Joon Lee and Su-Hyun Lee, "Protein structure comparison basis on structural classification information," 7th Korea-Russia International Symposium on Science and Technology, Proceedings KORUS 2003. (IEEE Cat. No.03EX737), Ulsan, South Korea, 2003, pp. 173-177 vol.2.

[5] S. K. H, A. Sudeesh, P. P. Nair and P. Suravajhala, "Biologically inspired ChaosNet architecture for Hypothetical Protein Classification," 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, 2023, pp. 1-6, doi: 10.1109/ICECCT56650.2023.10179833.

[6] C. Schulz, K. Schatz, M. Krone, M. Braun, T. Ertl and D. Weiskopf, "Uncertainty Visualization for Secondary Structures of Proteins," 2018 IEEE Pacific Visualization Symposium (PacificVis), Kobe, Japan, 2018, pp. 96-105, doi: 10.1109/PacificVis.2018.00020.

[7] C. H. Suryanto, S. Jiang and K. Fukui, "Protein structure similarity based on multi-view images generated from 3D molecular visualization," Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 2012, pp. 3447-3451.

[8] K. R. Saranya and E. R. Vimina, "Comparative Analysis of Protein Complex Identification Methods Using Topological and Biological Information," 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 2022, pp. 1-6, doi: 10.1109/INDICON56171.2022.10040192.

[9] A. Tripathi, A. Basavapattana, R. R. Nair and T. Singh, "Visualization of COVID Bimodal scan using DNN," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 01-07, doi: 10.1109/ICCCNT51525.2021.9579878.

[10] M. S. Pandianchery, V. Sowmya, E. A. Gopalakrishnan, V. Ravi and K. P. Soman, "Centralized CNN–GRU Model by Federated Learning for COVID-19 Prediction in India," in IEEE Transactions on Computational Social Systems, doi: 10.1109/TCSS.2023.3250656.

[11] J. Wang, W. Wang and Y. Shang, "Protein Loop Modeling Using AlphaFold2," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 20, no. 5, pp. 3306-3313, 1 Sept.-Oct. 2023, doi: 10.1109/TCBB.2023.3264899.

[12] M. Hansen, D. Meads and A. Pang, "Comparative visualization of protein structure-sequence alignments," Proceedings IEEE Symposium on Information Visualization (Cat. No.98TB100258), Research Triangle, CA, USA, 1998, pp. 106-110, doi: 10.1109/INFVIS.1998.729566.

[13] M. Falk, V. Tobiasson, A. Bock, C. Hansen and A. Ynnerman, "A Visual Environment for Data Driven Protein Modeling and Validation," in IEEE Transactions on Visualization and Computer Graphics, doi: 10.1109/TVCG.2023.3286582.

[14] Laguarta, Jordi, Ferran Hueto, and Brian Subirana. "COVID-19 artificial intelligence diagnosis using only cough recordings." IEEE Open Journal of Engineering in Medicine and Biology 1 (2020): 275-281.

[15] Srivastava, Pragya, Shreyansh Suyash, and N. Jayapandian. "Artificial Intelligence based System in Protein Folding using Alphafold." 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS). IEEE, 2022.

[16] Tripathi, Asmita, et al. "TemPred: A novel protein template search engine to improve protein structure prediction." IEEE/ACM Transactions on Computational Biology and Bioinformatics (2023).