

## **Comparative Evolutionary and Structural Analysis of Conserved Surface-Protein Regions in SARS-CoV, MERS-CoV, Influenza A and SARS-CoV-2 for Broad-Spectrum Antiviral Target Identification**

Roma Sharma<sup>1</sup>, Dr. Surabhi Singh<sup>2</sup>

*Research scholar, School of Basic and Applied Science, Career Point University, Kota, India*  
*Research Supervisor, School of Basic and Applied Science, Career Point University, Kota, India*

---

### **Abstract:**

*Respiratory viruses like SARS-CoV, MERS-CoV, SARS-CoV-2, and influenza A are constantly evolving, which complicates the identification of conserved antiviral targets. In the present study, we combined sequence-based evolutionary and structural analyses of viral surface proteins to identify potentially conserved and functionally relevant regions. We carried out multiple sequence alignment and phylogenetic analysis of 13 representative viral sequences and comparative structural analysis of representative surface proteins. The analysis showed that the coronavirus surface proteins were more closely related to each other evolutionarily, while influenza A formed a separate group. Combining evolutionary and structural information, candidate conserved regions that could be potential targets for broad-spectrum antiviral development were identified. These results establish a computational framework for selecting conserved regions of viral surface proteins for further experimental and structure-based antiviral research.*

**Keywords:** *Respiratory viruses; SARS-CoV; MERS-CoV; SARS-CoV-2; Influenza A; Surface proteins; Evolutionary analysis; Phylogenetic analysis; Structural analysis; Antiviral broad spectrum.*

---

Date of Submission: 18-09-2026

Date of Acceptance: 01-10-2026

---

### **I. Introduction**

Respiratory viral infections are a major health problem worldwide, with large associated morbidity and mortality and a significant socioeconomic burden. Respiratory viruses include influenza viruses, coronaviruses, respiratory syncytial virus, and adenoviruses, among others, and can cause clinical manifestations ranging from mild respiratory illness to severe and potentially fatal disease.<sup>1</sup> Their efficient transmission through respiratory pathways has resulted in repeated outbreaks and, in some cases, worldwide pandemics. Influenza viruses alone cause a substantial disease burden annually. The emergence of SARS-CoV-2 in late 2019 has demonstrated the significant effect an emerging respiratory virus can have on global health systems and economies.<sup>2,3</sup> Despite advances in vaccines and antiviral therapeutics, treating respiratory viral infections is still difficult because of rapid viral evolution, the emergence of novel variants, and the development of antiviral resistance. These challenges underscore the need for standard treatments that can maintain activity against genetically diverse and evolving viral populations. One of the important steps of respiratory viral infection is the entry of the virus into host cells, which is mediated by the interactions between viral surface proteins and specific receptors on host respiratory epithelial cells. Viral tropism, host susceptibility and disease severity are in part determined by the distribution and availability of cellular receptors. Coronaviruses and influenza viruses are among the major respiratory viruses that use spike (S) glycoproteins and hemagglutinin (HA), respectively, for attachment to host-cell receptors such as sialic-acid-containing receptors. For instance, the coronavirus spike proteins attach to host receptors, including angiotensin-converting enzyme 2 (ACE2), while influenza HA interacts with sialic acid receptors to trigger viral attachment and entry. These surface proteins are therefore important determinants for viral infectivity and host-cell entry, and have thus become important targets for antiviral intervention.<sup>4,5</sup> However, viral surface proteins are under constant evolutionary pressure, resulting in the accumulation of sequence variations that can alter receptor recognition, antigenicity and susceptibility to therapeutic intervention. Consequently, targeting highly variable regions may have limited durability, especially when viral variants emerge that have mutations that reduce inhibitor binding or alter functional properties. Conversely, essential viral functions may limit the number of tolerated mutations, thus leading to more evolutionary conservation. Identification of such conserved regions is an attractive strategy for developing broad-spectrum antiviral agents. In this respect, comparative evolutionary analysis can give insight into the relationships among viral proteins and

help discriminate between conserved and highly variable regions. Multiple sequence alignment (MSA) and phylogenetic analysis are useful characterization tools for sequence diversity, evolutionary relationships and conserved motifs across viral strains and species. Here, we compared the surface-protein sequences of SARS-CoV, MERS-CoV, SARS-CoV-2 and influenza A to examine their evolutionary relationships and to identify regions under potential functional constraint. Sequence conservation reveals important aspects of evolutionary stability, but sequence similarity alone may not fully reflect functional or therapeutic similarity. Proteins with low sequence conservation may have structurally related elements. Structural features related to essential biological functions can be preserved despite sequence divergence. Thus, structural analysis can complement evolutionary analysis by identifying conserved three-dimensional features, functional motifs and potentially accessible binding sites. Thus, a comparative structural analysis of viral surface proteins can offer an additional layer of information to prioritise antiviral targets. In this study, representative structures of the influenza A hemagglutinin and the spike glycoproteins of MERS-CoV, SARS-CoV and SARS-CoV-2 were evaluated to investigate structural characteristics and conserved features among these respiratory viral surface proteins. Structural assessment was carried out after a sequence-based evolutionary analysis, thus allowing candidate regions to be considered within the framework of both evolutionary and structural information.<sup>6</sup> The paper methodology involved multiple sequence alignment and maximum-likelihood phylogenetic analysis, followed by structural superimposition of representative viral surface proteins to identify potentially conserved structural motifs. One important factor in such a comparative analysis is the evolutionary distinction between influenza A and coronaviruses. The phylogenetic analysis done in this study showed that influenza surface-protein sequences grouped as a distinct clade, whereas SARS-CoV, MERS-CoV and SARS-CoV-2 fell into the coronavirus group, indicating a closer evolutionary relatedness amongst the beta-coronaviruses. Hence, identification of broad-spectrum antiviral targets need not be based solely on direct sequence homology among distantly related viruses, but could also benefit from the evolutionary conservation within related viral groups combined with structurally or functionally constrained features among different viral surface proteins. Therefore, in the present study, we sought to integrate comparative evolutionary and structural analyzes of surface proteins from SARS-CoV, MERS-CoV, SARS-CoV-2 and influenza A, so as to identify conserved and structurally relevant regions that could be candidate targets for broad-spectrum antiviral development. Multiple sequence alignment and phylogenetic analysis of viral surface-protein sequences were used to characterize evolutionary relationships and sequence conservation. The representative experimentally determined protein structures were analyzed and compared in an effort to find conserved structural features and functionally relevant regions. This integrated approach offers a platform for the transition from viral evolutionary diversity to conserved molecular features and finally to structurally defined antiviral targets, and thus provides an avenue for the identification of candidate regions for broad-spectrum antiviral intervention.

## **II. Materials and Methods**

### **2.1 Sequence Retrieval**

The surface protein sequences involved in substrate binding in respiratory viruses, including SARS-CoV, MERS-CoV, Influenza and SARS-CoV-2, were systematically retrieved in this study. We performed a systematic retrieval process using the National Center for Biotechnology Information (NCBI) virus repository (<https://www.ncbi.nlm.nih.gov/labs/virus/>).<sup>7</sup> This repository provides a valuable resource for researchers when searching for and accessing sequences of interest and contains a large collection of viral genomic sequences. Relevant sequences of surface proteins associated with substrate binding were selected by using specific search terms. The terms included the virus name and the surface protein name if available in the NCBI virus repository search interface. This allowed us to specifically pull out the sequences of surface proteins that are involved in substrate binding.<sup>8</sup> After retrieval, a comprehensive preprocessing step was used to clean the dataset and improve its quality. Redundant and incomplete sequences were removed and ambiguous or discrepant sequences were filtered out. These measures were taken to obtain a high-quality dataset that accurately represents the surface protein sequences required for substrate binding in the targeted respiratory viruses. In addition to this dataset, other data sources were used to improve its reliability and completeness, such as curated virus databases and relevant scientific literature. These sources were useful for cross validation of the retrieved sequences, and confirmed their accuracy and integrity. Finally, the curated and validated surface protein sequences represented a robust dataset that served as a basis for downstream analyzes. These sequences were used in later studies to elucidate the substrate-binding mechanisms and investigate their relevance in the setting of respiratory viral infections. The dataset provided key inputs for the study of viral surface proteins interaction with host receptors, which would aid in better understanding the molecular mechanisms of viral pathogenesis and in designing possible therapeutics targeting substrate binding.

## 2.2 Phylogenetic Tree and Multiple Sequence Alignment

To analyze the genetic relationships and evolutionary patterns among different strains and species of SARS, MERS, Influenza and SARS-CoV-2, a multiple sequence alignment (MSA) was performed. The Molecular Evolutionary Genetics Analysis software version (MEGA) was used to perform multiple sequence alignment (MSA).<sup>11,9</sup> MEGA is a bioinformatics tool for annotation and analysis of nucleotide or proteomic sequence information. It has a user-friendly user interface and a broad array of statistical methods and models. The MSA process included alignment of surface protein sequences of the respiratory viruses by ClustalW algorithm implemented into the MEGA software. This algorithm is recognized for its accuracy and efficiency in the alignment of multiple sequences. The output alignment was able to show conserved areas and differences between the sequences and helped to find evolutionary relationships and similarities between the sequences. Finally, the phylogenetic tree was constructed using the aligned protein sequences.<sup>10</sup> We employed the maximum likelihood (ML) method, a well-established and robust approach for inferring phylogenetic relationships. The parameters of the ML were set to default values. The robustness and reliability of the tree topology was tested by the bootstrap method (1,000 replicates), which gives statistical support for the order of branching in the phylogenetic tree. Gaps found in the sequence alignment were treated as missing or unknown characters in the alignment. These gaps correspond to the areas where there are no conserved amino acid residues among the aligned sequences. This could indicate variability in the regions of the surface proteins that bind to substrates. The resulting phylogenetic tree was visualized using FigTree v1.4.4, a tool specifically designed for the visualization and exploration of phylogenetic trees. The tree visualization clearly showed the evolutionary relationships between the respiratory viruses and allowed for the identification of distinct clades or clusters within each virus group.<sup>11</sup> The multiple sequence alignment and the following phylogenetic analysis gave useful information about the genetic diversity, relatedness and evolutionary patterns of the surface protein sequences involved in the binding of the substrate across different strains and species of SARS, MERS, Influenza and SARS-CoV-2. This information is important for understanding the evolutionary dynamics of these respiratory viruses. It can be used to find conserved regions or motifs involved in substrate binding and gives insights for targeted therapeutic interventions.

## 2.4 Target Finalisation and Structure Modelling

The study was a comparative structural analysis of the surface proteins of four different viruses, including Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), Middle East Respiratory Syndrome Coronavirus (MERS-CoV), Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), and Influenza A virus. For this purpose, the Protein Data Bank (PDB) identifiers of the surface protein of each virus were carefully chosen (Table 3.1). The individual protein superimposition analyzes were performed by Swiss-PDB Viewer (SPDBV) software suite. The first aim of this approach was to find conserved structural motifs across the viral surface proteins. These conserved regions are potentially functionally important elements important for viral activity.<sup>12</sup>

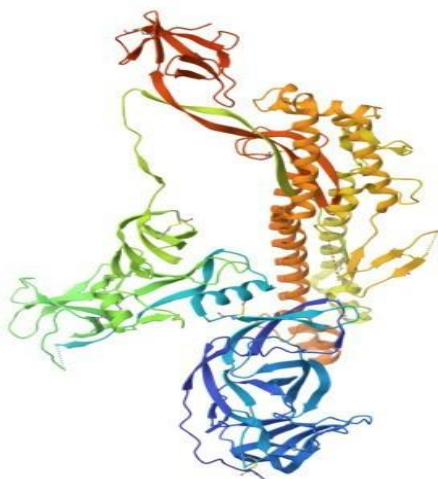
**Table - 2.1**  
**Details of Crystal structures**

| PDB ID | Description                                                                   | Method               | Resolution<br>Å | Ramachandran<br>outlier |
|--------|-------------------------------------------------------------------------------|----------------------|-----------------|-------------------------|
| 1RU7   | 1934 human<br>H1hemagglutinin                                                 | X-Ray<br>Diffraction | 2.30            | 1.0%                    |
| 5X59   | Prefusion structure of MERS-CoV spike<br>glycoprotein, three-fold<br>symmetry | Electron Microscopy  | 3.70            | 1.3%                    |
| 6CRV   | SARS Spike<br>Glycoprotein, Stabilized variant, C3<br>symmetry                | Electron Microscopy  | 3.20            | 0%                      |
| 7DWY   | S protein of SARS-CoV-<br>2 in the locked conformation                        | Electron Microscopy  | 2.70            | 0.4%                    |

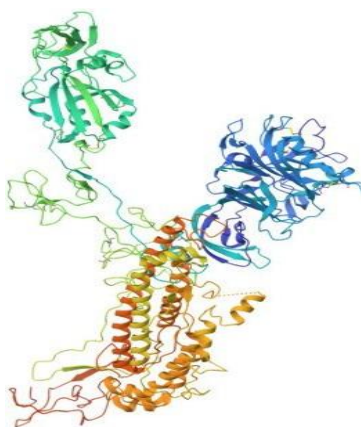


**Fig. 2.1: 1RU7\_1: Represented by Chain A**

Furthermore, a comprehensive quality assessment was conducted for all four crystal structures: 1RU7 (Influenza A virus), 5X59 (MERS-CoV), 6CRV (SARS-CoV), and 7DWY (SARS-CoV-2). This evaluation, detailed in Table 3.1, considered factors such as crystallisation method, resolution, and Ramachandran plot statistics to ensure structure integrity. Based on these metrics, 1RU7 and 7DWY were ultimately chosen for further analysis due to their superior quality parameters. It is important to note that while a single target could have been selected for this investigation, the deliberate inclusion of diverse viral surface proteins enriches the analysis by providing a broader perspective. This approach

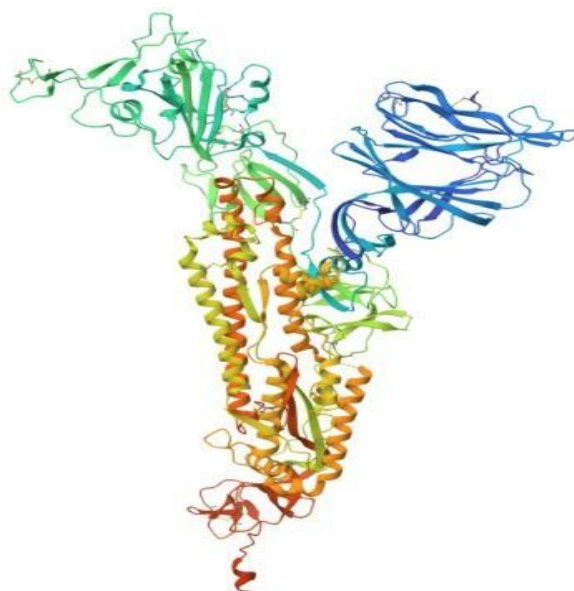


**Fig. 2.2: 6CRV\_1: Represented by Chain A**



**Fig. 2.3: 5X59\_1: Represented by Chain A**

Further, the quality of all four crystal structures was comprehensively evaluated, namely 1RU7 (Influenza A virus), 5X59 (MERS-CoV), 6CRV (SARS-CoV) and 7DWY (SARS-CoV-2). This assessment included crystallization method, resolution and Ramachandran plot statistics to ensure structure integrity (Table 3.1). Based on these parameters, 1RU7 and 7DWY were finally selected for the further analysis, as they presented superior quality parameters. It should be mentioned that, although a single target could have been chosen for the present study, the intentional inclusion of different viral surface proteins expands the view of the analysis.<sup>13</sup>



**Fig. 2.4: 7DWY\_1: Represented by Chain A**

### III. Results and Discussion

#### 3.1 Sequence Retrieval

In this study, a thorough and careful selection process was conducted to determine the most suitable surface protein sequences for further analysis. Thirteen sequences were selected after careful consideration of factors such as the prevalence of infection in the Asian population, highly pathogenic strains that can infect humans and pose a serious threat to public health, and the recency of sequence collection to ensure the most up-to-date information. Table 4.1 lists the top thirteen selected sequences, the organisms, genotypes/variants, and National Center for Biotechnology Information (NCBI) sequence IDs from which they are derived. The selected sequences correspond to a variety of respiratory viruses such as SARS-CoV, MERS-CoV, Influenza (H1N1 subtype) and SARS-CoV-2 (Delta variant, B.1.617.2 lineage).<sup>14,15</sup>

**Table - 3.1**  
**List of the top thirteen chosen sequences**

| Organism   | NCBI sequence ID                                           |
|------------|------------------------------------------------------------|
| SARS-CoV   | ADC35483.1, ADC35497.1                                     |
| MERS-CoV   | QDP16206.1, AZK15900.1                                     |
| Influenza  | AND61370.1, AIT93147.1, AHB59395.1, AWH65845.1             |
| SARS-CoV-2 | UNA09639.1, UOS65895.1, UZC40670.1, UZU73061.1, WAZ69178.1 |

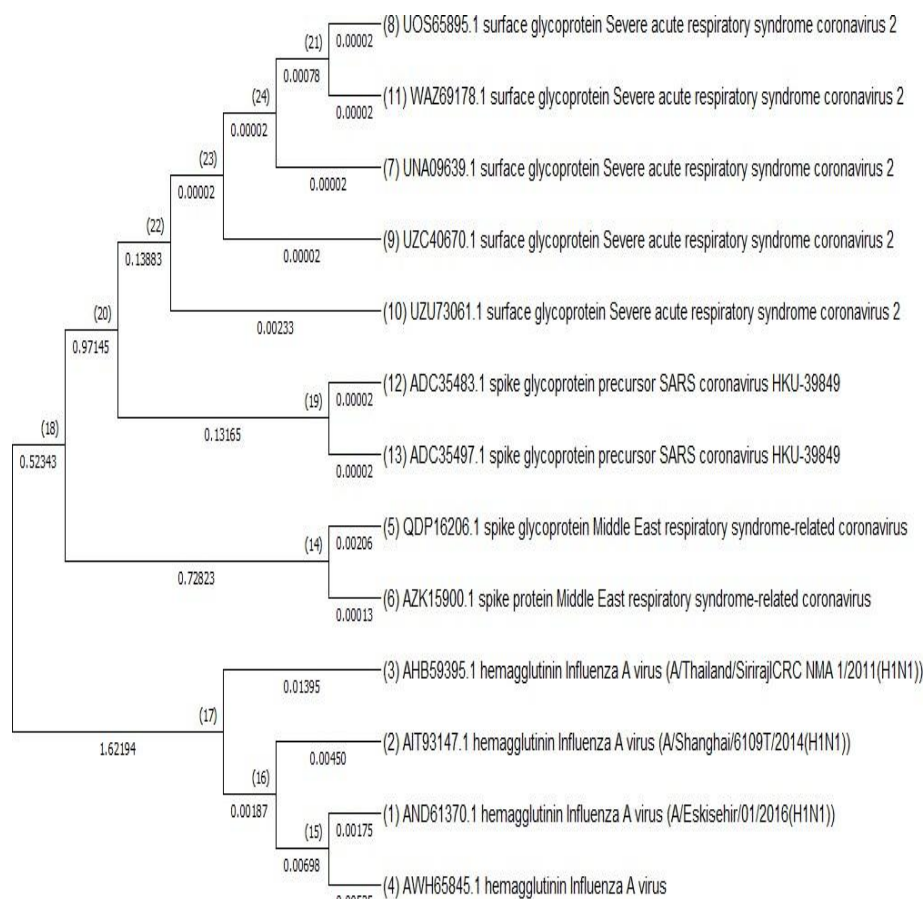
#### 3.2 Multiple Sequence Alignment & Phylogenetic Tree

After conducting a detailed analysis of the phylogenetic tree, notable observations were made regarding the relationships among the surface proteins of Influenza (hemagglutinin), SARS, MERS, and SARS-CoV-2. The phylogenetic tree revealed distinct clades, indicating evolutionary connections between these viruses.<sup>16</sup>

First, the surface protein of Influenza (hemagglutinin) constituted a well-formed clade, which demonstrates a close genetic relationship among different strains of Influenza. This clustering suggested a common origin and shared evolutionary history. Secondly, another clade was identified in which SARS, MERS and SARS-CoV-2 were grouped together.<sup>17</sup> This clustering suggests a strong evolutionary link between these beta coronaviruses. This group was supported by a relatively high bootstrap value, indicating a strong support for this clade, further emphasizing the close genetic relationships of these respiratory viruses. In addition, the full-length

sequence comparison showed that the surface proteins of the beta coronavirus family had a high sequence similarity. This result is indicative of a common evolutionary history and gives us an insight into the evolutionary relationships of these viruses. The high sequence similarity indicates potential common functional features and conserved structural elements of their surface proteins.

Figure 4.1 shows the maximum likelihood phylogenetic tree with bootstrap validation scores and branch lengths. The bold numbers above each branch are the bootstrap validation scores indicating support for the respective branching patterns. The numbers under the branches are branch lengths, and thus genetic distances between the sequences.



**Fig. 3.1: Maximum likelihood phylogenetic tree with bootstrap validation scores and branch lengths**

This study aimed to improve the understanding of the evolutionary relationships of Influenza, SARS, MERS and SARS-CoV-2 through phylogenetic analysis and sequence comparison. Such findings have implications for the understanding of possible origins of these viruses and their modes of transmission. The phylogenetic tree shows common evolutionary history and genetic links that offer useful insights into the interplay of respiratory viruses and aid in the creation of specific interventions and preventive measures.

### 3.3 Finalisation of Target, Structure, and Active Site Identification

In our investigation, we used the Protein Data Bank (PDB) to search related structures. 1RU7 was chosen for further investigation after a quality assessment based on metrics like X-ray diffraction resolution (with a priority for structures at or below 2.30 Å). However, upon detailed analysis of the 1RU7 structure (Influenza A virus), some discrepancies were found, which required a more detailed literature review. Our extensive literature search indicated 1RVZ as a better alternative structure for further analyzes, possibly containing a better binding pocket for our study. mainly on its putative binding pocket formed by the amino acid residues Lys222, Asp225, Ala227, Gln226, Ala137, Thr136, Glu190, Asn193 and Val135.<sup>18</sup>

## IV. Conclusion

Herein, we integrated evolutionary and structural analyses of SARS-CoV, MERS-CoV, SARS-CoV-2 and influenza A surface proteins to identify potential conserved and structurally important antiviral target regions. Phylogenetic analysis showed that the coronavirus surface proteins and influenza A were separated into a distinct evolutionary group. Structural analysis also reinforced the prioritisation of functional relevance. To summarise,

the study demonstrates that the combination of viral evolution, sequence conservation and structural analysis can provide a useful framework for identifying candidate targets for future broad-spectrum antiviral development.

### References

- [1]. Global, regional, and national burden of influenza-associated lower respiratory infections, 1990–2021: a systematic analysis from the Global Burden of Disease Study 2021 | BMC Infectious Diseases | Springer Nature Link. <https://link.springer.com/article/10.1186/s12879-025-12282-7> (accessed 2026-08-24).
- [2]. Zheng, X.; Li, Y.; Ruan, D.; Lin, L.; Meng, R.; Zhao, D.; Yu, J.; Li, X.; Huang, H.; Zhou, M. Burden of Non-COVID-19 Lower Respiratory Infections and Etiologies in China and Globally: An Analysis for the Global Burden of Disease Study 2021. *Chin. Med. J. Pulm. Crit. Care Med.* **2025**, 3 (3), 193–208. <https://doi.org/10.1016/j.pccm.2025.08.006>.
- [3]. Sharma, R. Exploring the Binding Affinity Landscape of SARS-CoV-2 Variants: A Computational Approach. *J. Med. Pharm. Allied Sci.* **2024**, 13 (3), 6566–6569. <https://doi.org/10.55522/jmpas.V13I3.6555>.
- [4]. Oh, H.; Thi Thuy Tien, V.; Ahmed, S.; Choi, J.; Ryu, K.-J.; Yang, J. Host Glycan–Lectin Interplay in SARS-CoV-2 Infection. *Int. J. Mol. Sci.* **2026**, 27 (3), 1608. <https://doi.org/10.3390/ijms27031608>.
- [5]. Frontiers | Viral glycoprotein-mediated entry and antibody-mediated immunity in HIV-1 and SARS-CoV-2 infection. <https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2025.1733684/full> (accessed 2026-08-27).
- [6]. Shuaibu, B. I.; Momodu, A.; Muhibi, M. A.; Dutsinma, U. A.; Yelwa, I. S. A Comparative Review on the Epidemiology of Influenza, COVID-19 and Respiratory Syncytial Viruses. *Discov. Med.* **2025**, 2 (1), 274. <https://doi.org/10.1007/s44337-025-00518-z>.
- [7]. Brister, J. R.; Ako-adjei, D.; Bao, Y.; Blinkova, O. NCBI Viral Genomes Resource. *Nucleic Acids Res.* **2015**, 43 (D1), D571–D577. <https://doi.org/10.1093/nar/gku1207>.
- [8]. Gurung, R.; Kumari, U. NGS and Structure Based Drug Designing of Natural Compound against Drug Resistant EGFR in Non-Small Cell Lung Cancer.
- [9]. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, 38 (7), 3022–3027. <https://doi.org/10.1093/molbev/msab120>.
- [10]. Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res.* **1994**, 22 (22), 4673–4680. <https://doi.org/10.1093/nar/22.22.4673>.
- [11]. Bianchini, G.; Sánchez-Baracaldo, P. TreeViewer: Flexible, Modular Software to Visualise and Manipulate Phylogenetic Trees. *Ecol. Evol.* **2024**, 14 (2), e10873. <https://doi.org/10.1002/ece3.10873>.
- [12]. Burley, S. K.; Berman, H. M.; Duarte, J. M.; Feng, Z.; Flatt, J. W.; Hudson, B. P.; Lowe, R.; Peisach, E.; Piehl, D. W.; Rose, Y.; Sali, A.; Sekharan, M.; Shao, C.; Vallat, B.; Voigt, M.; Westbrook, J. D.; Young, J. Y.; Zardecki, C. Protein Data Bank: A Comprehensive Review of 3D Structure Holdings and Worldwide Utilization by Researchers, Educators, and Students. *Biomolecules* **2022**, 12 (10), 1425. <https://doi.org/10.3390/biom12101425>.
- [13]. Bank, R. P. D. Exploring Structure Quality. <https://www.rcsb.org/news/5764422199cccf72e74ca391> (accessed 2026-08-27).
- [14]. NCBI Virus. <https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/> (accessed 2026-08-27).
- [15]. Yuan, Y.; Cao, D.; Zhang, Y.; Ma, J.; Qi, J.; Wang, Q.; Lu, G.; Wu, Y.; Yan, J.; Shi, Y.; Zhang, X.; Gao, G. F. Cryo-EM Structures of MERS-CoV and SARS-CoV Spike Glycoproteins Reveal the Dynamic Receptor Binding Domains. *Nat. Commun.* **2017**, 8 (1), 15092. <https://doi.org/10.1038/ncomms15092>.
- [16]. Sharma, R. Exploring the Binding Affinity Landscape of SARS-CoV-2 Variants: A Computational Approach. *J. Med. Pharm. Allied Sci.* **2024**, 13 (3), 6566–6569. <https://doi.org/10.55522/jmpas.V13I3.6555>.
- [17]. Sharma, R.; Kumari, U.; Hasan, K. Next-Generation Sequencing Workflow Study Of SARS-Cov-2 Beta RBD in Complex with Human ACE2 With Biopython. In Review November 5, 2024. <https://doi.org/10.21203/rs.3.rs-5308384/v1>.
- [18]. Bank, R. P. D. RCSB PDB - 1RU7: 1934 Human H1 Hemagglutinin. <https://www.rcsb.org/structure/1RU7> (accessed 2026-08-27).