# Computational Modeling and High-confidence Tertiary Structure Prediction of the SARS-CoV-2 NSP6 Protein: Implications for Viral Pathogenesis and Host Interaction

Mohammed Mostafa Salama[1], Medhat Wahba Shafaa[1], Mohamed El-Sayed El-Nagdy[1], Manal F. El-Khadragy[2], Ahmed E. Abdel Moneim[3], Ashraf Albrakati[4], Khalid Ebraheem Hassan[5], Elham H. Alrubai[6], H.H. Osman[4], Mohamed El-Sayed Hasan[7]

10.18805/IJAR.BF-2093

## ABSTRACT

**Background:** The non-structural protein 6 (NSP6) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a critical transmembrane protein essential for the formation of viral replication organelles. Despite its importance as a potential drug target, the absence of an experimentally solved crystal structure has hindered structure-based antiviral discovery. This study aimed to predict, refine and validate the tertiary structure of NSP6 (YP_009742613) using a comprehensive computational pipeline.

**Methods:** The amino acid sequence of NSP6 was obtained from UniProtKB. Its secondary structure was predicted using a consensus from eleven servers. Tertiary structure models were generated using eight distinct prediction servers (SWISS-MODEL, Phyre2, AlphaFold, C-Quark, Galaxyweb, I-Tasser, LOMETS and Robetta). The resulting models were subsequently refined using six different servers (3D-refine, ModRefiner, ReFOLD3, DeepRefiner, GalaxyRefine, GalaxyRefine2), producing 48 refined models. All models were rigorously evaluated using multiple quality assessment tools (SWISS-MODEL Structure Assessment, PROSA, PROQ, SAVES, TM-align) analyzing parameters including ERRAT, Ramachandran plot, Z-score and TM-score.

**Result:** Secondary structure analysis confirmed NSP6 as a highly alpha-helical (~68-78%) transmembrane protein. The model refinement process significantly enhanced model quality, with RMSD decreasing to 0.25-0.3 Å and TM-score increasing to 0.9952 for the top models. The evaluation demonstrated that the model generated by the AlphaFold server and refined by DeepRefiner was of the highest quality, with an overall ERRAT score of 99.64%, 94.4% of residues in the core Ramachandran regions and a PROSA Z-score of -1.33, confirming its placement within the range of native protein structures.

**Key words:** Ab-initio, Homology modeling, Model refinement, Model validation, NSP6, Protein structure prediction, SARS-CoV-2, Tertiary structure.

## INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has underscored the urgent need for a profound understanding of viral proteins to facilitate the development of effective therapeutics (Huang *et al*., 2020; Li *et al*., 2020). The SARS-CoV-2 genome encodes a series of non-structural proteins (NSPs) that are fundamental to viral replication and pathogenesis within the host cell (Cucinotta and Vanelli, 2020; Zou *et al*., 2020). Among these, Non-Structural Protein 6 (NSP6) has emerged as a critical yet structurally enigmatic component. NSP6, a 290-amino-acid transmembrane protein, plays a pivotal role in the formation of double-membrane vesicles (DMVs) by anchoring the viral replication-transcription complexes (RTCs) to the endoplasmic reticulum (ER) in conjunction with NSP3 and NSP4 (Abdelkader *et al*., 2022; Lu *et al*., 2020). Furthermore, NSP6 is implicated in subverting host cell autophagy, a key innate immune defense mechanism, thereby creating a protected niche for viral replication (Lubin *et al*., 2022). This central role in the viral life cycle makes NSP6 an attractive target for antiviral drug design.

A significant barrier to rational drug discovery against NSP6 is the absence of an experimentally determined high-resolution three-dimensional structure. Unlike several other

[1]Medical Biophysics Division, Physics Department, Faculty of Science, Helwan University, Cairo, Egypt.
[2]Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia.
[3]Unit of Scientific Research, Applied College, Qassim University, Buraydah, Saudi Arabia.
[4]Department of Human Anatomy, College of Medicine, Taif University, Taif 21944, Saudi Arabia.
[5]Department of Pathology, College of Medicine, Taif University, Taif, Saudi Arabia.
[6]Security Forces Hospital Program, General Directorate of Medical Services, Ministry of Interior, Riyadh, Saudi Arabia.
[7]Department of Bioinformatics, Genetic Engineering and Biotechnology Research Institute, University of Sadat City, Sadat City 32897, Egypt.

**Corresponding Author:** Ashraf Albrakati, Department of Human Anatomy, College of Medicine, Taif University, Taif 21944, Saudi Arabia. Email: a.albrakati@tu.edu.sa
ORCIDs: 0000-0003-0866-3688, 0000-0003-1807-6516, 0000-0003-4132-782X, 0000-0001-5457-6791, 0000-0002-3148-6782, 0000-0002-4116-7865, 0000-0003-3085-0981, 0009-0009-5571-1364, 0000-0003-3080-4774, 0000-0002-6325-3393

**How to cite this article:** Salama, M.M., Shafaa, M.W., El-Nagdy, M.E., El-Khadragy, M.F., Moneim, A.E.A., Albrakati, A., Hassan, K.E., Alrubai, E.H., Osman, H.H. and Hasan, M.E. (2026). Computational Modeling and High-confidence Tertiary Structure Prediction of the SARS-CoV-2 NSP6 Protein: Implications for Viral Pathogenesis and Host Interaction. *Indian Journal of Animal Research.* **60(Special issue):** 52-29. doi: 10.18805/IJAR.BF-2093.

**Submitted:** 04-12-2025 **Accepted:** 20-06-2026 **Online:** 14-07-2026

SARS-CoV-2 proteins, NSP6 has proven recalcitrant to crystallization, leaving a critical gap in our structural knowledge (Morais *et al*., 2020). In such scenarios, computational protein structure prediction becomes an indispensable tool in structural biology (Pandey *et al*., 2020; Zhou *et al*., 2020). Modern prediction methods have advanced dramatically, ranging from homology modeling and threading to sophisticated *ab initio* and deep learning approaches like AlphaFold, which have revolutionized the field by achieving unprecedented accuracy (Guo *et al*., 2020). However, the reliability of any predicted model is contingent upon rigorous validation and refinement to ensure its physical realism and structural integrity (Altman and Dugan, 2003; Zhang *et al*., 2005).

To address this gap, we undertook a comprehensive computational study to generate, refine and validate the tertiary structure of SARS-CoV-2 NSP6. We employed a multi-faceted strategy, leveraging eight distinct structure prediction servers-SWISS-MODEL, Phyre2, AlphaFold, C-Quark, Galaxyweb, I-Tasser, LOMETS and Robetta-to generate a diverse set of initial models. Recognizing that initial predictions often require optimization, we subjected these models to a systematic refinement process using six different servers (3D-refine, ModRefiner, ReFOLD3, DeepRefiner, GalaxyRefine and GalaxyRefine2). The quality of the resulting 48 refined models was then meticulously assessed using a battery of validation tools.

The primary objective of this study was to establish a robust pipeline for determining the most accurate and reliable tertiary structure model of NSP6. This work provides the scientific community with a high-confidence, validated structural framework for NSP6, which serves as an essential foundation for subsequent functional studies, mechanistic insights and structure-based antiviral design, as explored in our companion paper.

## MATERIALS AND METHODS

The methodology used accurately analyzes the SARS-CoV-2 NSP6 protein through the prediction of domains, conserved regions, secondary structures, tertiary structures, posttranslational modification sites, signatures and motifs. In addition, the structural classification and functional annotations of the target proteins were identified.

### Sequence retrieval and preparation

The amino acid sequence of SARS-CoV-2 NSP6 (UniProtKB/NCBI accession number YP_009742613) was obtained from the UniProtKB and NCBI databases. The length of the protein (290 amino acids).

### Prediction of secondary structure and solvent accessibility

To ensure high-confidence prediction of the secondary structure and solvent accessibility of the SARS-CoV-2 NSP6 protein, we employed multiple state-of-the-art servers that rely on neural-network–based algorithms. These platforms represent standardized and widely validated approaches for sequence-based structural prediction and their collective use strengthens the reliability of the generated models (Ismi *et al*., 2022; Meng *et al*., 2025).

### Tertiary structure prediction

To date, no experimentally resolved crystal structure is available for SARS-CoV-2 NSP6. Therefore, its tertiary structure was predicted using a comprehensive computational strategy that integrates both homology-based and ab initio modeling approaches. Multiple state-of-the-art prediction platforms based on threading, fragment assembly and deep learning frameworks were employed to ensure structural reliability. The integration of multiple computational platforms has become a widely adopted strategy in bioinformatics because combining complementary algorithms improves the robustness and reproducibility of in silico analyses across diverse biological applications (Rajith *et al*., 2023). These tools represent well-established and widely validated methodologies in modern protein structure prediction and their combined use has been shown to significantly improve model accuracy (Ismi *et al*., 2022).

### Tertiary structure model refinement

To enhance the structural quality of the tertiary models, refinement was performed using energy minimization and molecular-dynamics-based optimization. The refinement workflow followed widely adopted strategies that adjust hydrogen-bonding networks, optimize local stereochemistry and apply knowledge-based force fields to improve atomic-level accuracy. These approaches are well established in modern protein structure refinement and have consistently been shown to improve RMSD, stereochemical quality and overall model reliability (Adiyaman and McGuffin, 2021).

### Evaluation of tertiary structure models

Evaluation of tertiary structure models is a critical step in protein modeling, requiring the use of multiple quantitative quality assessment (QA) metrics such as TM-score, GDT-TS, Z-score, RMSD, stereochemical validation and global/local model quality indices. In this study, the refined NSP6 models were assessed using widely established QA principles that combine statistical potentials, structural consistency checks and template-based similarity measures. These validation strategies have been demonstrated to reliably distinguish near-native folds from incorrectly predicted models and are considered the foundation of modern model-evaluation pipelines (Benkert *et al*., 2011).

## RESULTS AND DISCUSSION

The NSP6 protein of SARS-CoV-2 is composed of 290 amino acid sequences; accession number YP_009742613 was obtained from the UniProtKB and NCBI databases.

### Prediction of secondary structure and solvent accessibility

The prediction of the secondary structure of the SARS-CoV-2 NSP6 protein was investigated using a consensus result

that was generated from the eleven servers Hhpred, pep2d, Jpred4, Porter, PredictProtein, Psipred, SOPMA, Raptorx, SSPRO, Yaspin and APSSP2 (Table 1). The predicted secondary structure of the SARS-CoV-2 NSP6 protein from the Psipred server may be composed of 77.93% α-helices, 2.76% β-strands and 19.31% coils. In addition, the predict protein server predicted that the secondary structure of the NSP6 protein of SARS-CoV-2 consists of 68% alpha helices, 19% coils and 13% beta strands. The Helices were analyzed to be 8-35, 41-85, 88-100, 107-130, 136-149, 158-172, 176-196, 216-236, 248-256 and 265-272; Coils, 1-7, 36-40, 86, 87, 101-103, 131-135, 155-157, 173-175, 197, 198, 204-206, 237-241, 247, 257-264, 273, 274 and 278-281; and strands, 104-106, 150-154, 199-203, 207-215, 242-246, 275-277 and 282-289. Additionally, the PredictProtein server could predict solvent accessibility; 76.2% and 23.79% of the proteins were mostly buried intermediately and exposed, respectively. The NSP6 protein of SARS-CoV-2 is found to be a transmembrane protein with a transmembrane helix of 56.5%, an extracellular helix of 16.5% and a cytoplasmic helix of 27%.

**Prediction and refinement of tertiary structure**

To construct a tertiary structure model, we built many models, refined them and evaluated the full models to select the model with the highest quality.

**Construction of the initial model using target-template alignment**

Several servers, such as the GalaxyWEB, SWISS-MODEL and LOMETS, are based on the structuring of aligned regions of proteins, while the I-TASSER, Robetta and Phyre2 servers are designed to predict the structures of proteins that have low similarity with known structured proteins. We used AlphaFold which is a server that uses deep learning techniques and neural networks to predict the distance and torsion of proteins. Its data depends on training on known schemes of PDB structures with known amino acid sequences and MSA features. Additionally, the AlphaFold model is based on a template-free modeling algorithm that can predict new unstructured proteins. In addition, the Robetta, C-Quark, LOMETS and I-TASSER servers can predict the five full models for the query according to the confidence score (C-score), which was -1.90 for the best model and a TM-score of 0.45±0.15. Additionally, GalaxyWEB predicts five full models for the query arranged according to the HHsearch results rescoring system. In addition, the Phyre2 server was able to construct eight models of small regions of the query, as was the SWISS-MODEL server.

**Reduced-level structure assembly and refinement simulations**

To improve the quality of the tertiary structure prediction process, we needed to refine the predicted models of the NSP6 protein of SARS-CoV-2. As a result, we used the 3D-REFINE, ModRefiner, Refold3, DeepRefiner, GalaxyWEB Refine and GalaxyWEB Refine2 servers to refine the models by refining the backbone structure, side chain positioning and hydrogen bonds. The model refinement results revealed that the RMSD decreased to 0.25 and 0.3 A and the TM-score increased to 0.9952 for the AlphaFold model of the NSP6 protein of SARS-CoV-2, which indicates an increase in the quality of the predicted structure compared to that of the initially predicted structure.

**Model evaluation and selection**

Fig 1 (a, b) shows the best predicted three-dimensional structure of the NSP6 protein of SARS-CoV-2 using the AlphaFold server, which was generated by Autodock Vina software different tools of view ribbon view, solvent-accessible surface and atomic representation of the protein. Fig 1 (c) shows the binding sites of SARS-CoV-2 NSP6 using Autodock Vina software. Finding the closest superimposed model is an important process in which we can construct a structural framework for obtaining additional information in protein design, interpreting functional data, determining the molecular function of the protein and designing and targeting drugs. I-TASSER server was used to predict the closest superimposed experimentally structured models to the NSP6 protein of SARS-CoV-2. I-TASSER server revealed that the experimental structure 5aexJ template

**Table 1:** Predicted secondary structure of the SARS-CoV-2 NSP6 protein using different servers.

| | Helix | | Strand | | Coil | | Buried | | Exposed | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Count | % | Count | % | Count | % | Count | % | Count | % |
| ASSP2 | 140 | 48.61 | 75 | 26.04 | 73 | 25.35 | - | - | - | - |
| JPRED4 | 217 | 74.83 | 12 | 4.14 | 61 | 21.03 | 222 | 76.55 | 68 | 23.45 |
| PEP2D | 155 | 53.45 | 46 | 15.86 | 89 | 30.69 | - | - | - | - |
| Porter | 217 | 74.83 | 4 | 1.38 | 69 | 23.79 | - | - | - | - |
| Predict protien | 198 | 68 | 38 | 13 | 54 | 19 | 221 | 76.2 | 69 | 23.79 |
| Psipred | 226 | 77.93 | 8 | 2.76 | 56 | 19.31 | - | - | - | - |
| RAPTORX | 209 | 72.07 | 13 | 4.48 | 68 | 23.45 | - | - | - | - |
| SOPMA | 149 | 51.38 | 73 | 25.17 | 68 | 23.45 | - | - | - | - |
| SSPRO | 216 | 74.48 | 7 | 2.41 | 67 | 23.1 | 161 | 55.52 | 129 | 44.48 |
| YASPIN | 59 | 20.34 | 170 | 58.62 | 61 | 21.34 | - | - | - | - |

was the closest predicted structural model of the SARS-CoV-2 NSP6 protein.

**Model evaluation and selection of the best model**

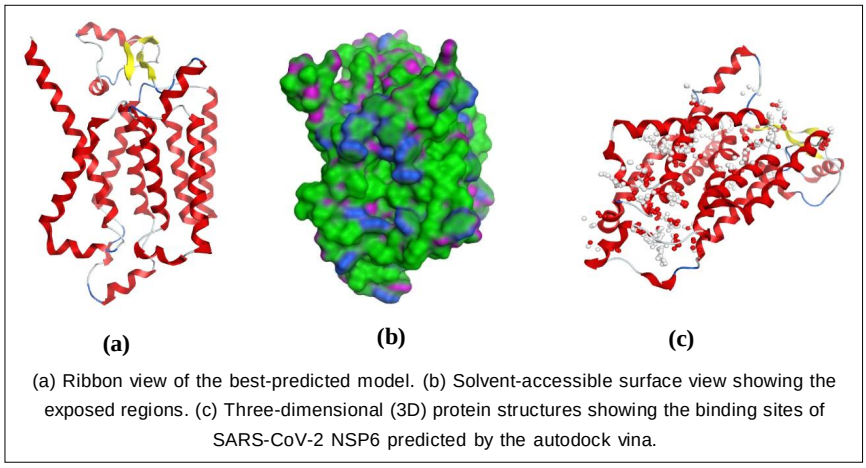
To determine the best model of the NSP6 protein with good topology, all the predicted and refined models were evaluated, Table 2. The model predicted by AlphaFold and then refined by DeepRefiner presented the best results, with an overall quality of 99.64%, according to the ERRAT server compared to all the other prediction models. The Ramachandran plot generated with the PROCHECK server revealed 94.4% of the residues in the core regions in addition to 5.6% in the allowed regions and no residues in the disallowed regions. The SWISSMODEL structure Assessment server supported the results of the Ramachandran plot generated by the PROCHECK server, which included 98.35% of the residues in the favored regions and 0.41% of the residues in the outgroup regions. Additionally, according to the PROSA server, the model had a molprobity value of 2.61 and a clash score of 139.01, the mean Z score was -1.33 and the Z score was -4.12. These Z scores clarified that the modeled protein was within the range of X-ray-solved protein structures, which indicated that it was a good model. The PROQ server showed that the LG score was 8.807, which indicates an extremely good model according to the evaluation rules and the maximum substitution value was -0.544.

The absence of an experimental structure for SARS-CoV-2 NSP6 has been a significant impediment to targeted antiviral development. This study was designed to bridge this gap by implementing a rigorous, multi-step computational pipeline to generate a high-confidence tertiary structure model. Our approach was predicated on the rationale that a consensus from multiple, independent prediction and refinement methods, followed by stringent validation, would yield a more reliable model than any single method alone. The integration of multiple computational platforms has also been successfully adopted in comparative genomic investigations, where combining complementary bioinformatics approaches improves the robustness and biological interpretation of computational analyses (Liu *et al*., 2025).

Our initial analysis confirmed that NSP6 is a highly helical transmembrane protein, a finding consistent with its known biological role in anchoring replication complexes to endoplasmic reticulum membranes (Abdelkader *et al*., 2022; Lu *et al*., 2020). This secondary structure profile provided a crucial sanity check for the subsequent tertiary models, as a plausible model would need to reflect this fundamental characteristic.

The cornerstone of our strategy was the generation of a diverse model set using eight distinct prediction servers, encompassing both template-based and *ab initio* approaches. The initial models displayed considerable variation, underscoring the challenge of predicting transmembrane protein structures and the necessity for a comparative approach. The subsequent refinement step proved critical, significantly enhancing model quality by resolving steric clashes and optimizing hydrogen-bonding networks. The marked improvement in metrics such as RMSD and TM-score post-refinement highlights that this is not a superfluous step but an essential one for achieving atomic-level realism (Gordon *et al*., 2020; Zhang *et al*., 2020).

The comprehensive validation phase, utilizing a battery of assessment tools, provided clear and convergent evidence for the superiority of the model generated by AlphaFold and refined by DeepRefiner. This model consistently ranked highest across multiple independent quality metrics. Its exceptional ERRAT score (99.64%) signifies a high-quality non-bonded atomic interaction profile, while its Ramachandran plot (94.4% core, 0% disallowed) confirms the stereochemical excellence of its backbone dihedral angles (Zhou *et al*., 2011). Furthermore,




(a) Ribbon view of the best-predicted model. (b) Solvent-accessible surface view showing the exposed regions. (c) Three-dimensional (3D) protein structures showing the binding sites of SARS-CoV-2 NSP6 predicted by the autodock vina.

**Fig 1:** The best predicted three-dimensional structure of the NSP6 protein of SARS-CoV-2 using the AlphaFold server, which was generated by autodock vina.

**Table 2:** Evaluation of the predicted models of the NSP6 protein in SARS-CoV-2 infection using different bioinformatics tools.

| Evaluation server | | | Saves | | | | | TM-score | | | TM-align | | | | PROSA | Swissmodel structure assessment | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tertiary structure prediction server | Model refinement server | Model no. | ERRAT Overall quality | PROCHECK RAM CORE | PROCHECK RAM Allowed | Number of residues in common | RMSD of the common residues | TM-score | GDT-TS | GDT-HA | Aligned length | RMSD | TM-score Normalized by length of Chain_1 | TM-score Normalized by length of Chain_2 | Z-score | Molprobity | Clash score | Ramachandran favoured (%) | Qmean disco | Qmean Z-score |
| AlphaFold | 3D-refine | 5 | 99.4539 | 94.7 | 4.9 | 276 | 22.07 | 0.142 | 0.062 | 0.036 | 209 | 5.49 | 0.471 | 0.349 | -3.93 | 1.66 | 11.26 | 97.57 | 0.56± 0.05 | -0.81 |
| | ModRefiner | 1 | 98.556 | 95.9 | 4.1 | 276 | 22.074 | 0.141 | 0.063 | 0.037 | 208 | 5.48 | 0.468 | 0.346 | - | 2.02 | 34.43 | 98.61 | 0.54± 0.05 | -2.32 |
| | GalaxyRefine | 2 | 99.6324 | 97 | 3 | 276 | 22.11 | 0.141 | 0.063 | 0.037 | 211 | 5.71 | 0.466 | 0.346 | -4.29 | 1.15 | 3.61 | 98.96 | 0.58± 0.05 | -2.03 |
| | DeepRefiner | 1 | 99.6403 | 94.4 | 5.6 | 276 | 22.2 | 0.141 | 0.063 | 0.036 | 208 | 5.48 | 0.466 | 0.346 | -4.12 | 2.61 | 139.01 | 98.35 | 0.58± 0.05 | -1.33 |
| | refold3 | 4 | 94.5652 | 94.4 | 5.3 | 276 | 22.26 | 0.142 | 0.064 | 0.038 | 206 | 5.39 | 0.469 | 0.347 | -3.8 | 0.99 | 2.13 | 98.61 | 0.61± 0.05 | -1.67 |
| | GalaxyRefine 2 | 1 | 99.6337 | 96.2 | 3.4 | 276 | 22.05 | 0.14 | 0.062 | 0.037 | 211 | 5.8 | 0.462 | 0.343 | - | 0.5 | 0 | 99.23 | 0.57± 0.05 | -1.68 |
| C-quark | 3D-refine | 1 | 63.8298 | 71.1 | 18.4 | 276 | 20.51 | 0.153 | 0.062 | 0.042 | 158 | 5.99 | 0.333 | 0.249 | -4.4 | 4.08 | 63.39 | 75 | 0.31± 0.05 | -10.73 |
| | ModRefiner | 1 | 62.766 | 77.4 | 15 | 276 | 20.59 | 0.152 | 0.062 | 0.04 | 145 | 5.75 | 0.321 | 0.238 | - | 3.75 | 150.3 | 82.64 | 0.34± 0.05 | -9.48 |
| | GalaxyRefine | 1 | 71.8412 | 78.2 | 16.2 | 276 | 20.57 | 0.151 | 0.061 | 0.039 | 144 | 5.65 | 0.322 | 0.238 | -4.56 | 3.25 | 36.56 | 82.64 | 0.35± 0.05 | -8.23 |
| | DeepRefiner | 1 | 84.6715 | 78.6 | 15.4 | 276 | 20.88 | 0.148 | 0.061 | 0.039 | 155 | 6.36 | 0.31 | 0.235 | -3.71 | 3.3 | 168.73 | 87.19 | 0.33± 0.05 | -6.7 |
| | refold3 | 1 | 68.3799 | 78.6 | 12.8 | 276 | 20.87 | 0.147 | 0.06 | 0.039 | 159 | 6.63 | 0.313 | 0.236 | -3.77 | 3.19 | 15.94 | 81.6 | 0.36± 0.05 | -7.95 |
| | GalaxyRefine 2 | 1 | 70.3571 | 77.1 | 16.2 | 276 | 20.37 | 0.152 | 0.062 | 0.041 | 142 | 5.55 | 0.32 | 0.237 | - | 2.48 | 14.45 | 78.08 | 0.33± 0.05 | -9.16 |
| Robetta | 3D-refine | 5 | 99.2908 | 91 | 7.9 | 276 | 22.2 | 0.149 | 0.068 | 0.039 | 191 | 5.65 | 0.419 | 0.312 | -2.72 | 2.14 | 19.35 | 94.79 | 0.51± 0.05 | -2.41 |
| | ModRefiner | 1 | 95.3901 | 92.1 | 7.1 | 276 | 22.27 | 0.148 | 0.069 | 0.04 | 185 | 5.62 | 0.408 | 0.304 | - | 2.32 | 41.66 | 98.61 | 0.52± 0.05 | -3.23 |
| | GalaxyRefine | 3 | 95.6835 | 94.4 | 4.9 | 276 | 22.26 | 0.148 | 0.069 | 0.039 | 187 | 5.72 | 0.413 | 0.306 | -2.7 | 1.45 | 4.89 | 97.22 | 0.53± 0.05 | -3.21 |
| | DeepRefiner | 1 | 98.5816 | 91.7 | 7.5 | 276 | 22.32 | 0.146 | 0.07 | 0.04 | 193 | 5.97 | 0.416 | 0.31 | -2.84 | 2.72 | 145.89 | 97.52 | 0.52± 0.05 | -2.68 |
| | refold3 | 1 | 96.6418 | 89.5 | 9.4 | 276 | 22.37 | 0.146 | 0.067 | 0.038 | 203 | 6.27 | 0.414 | 0.312 | -2.23 | 1.37 | 0.21 | 93.75 | 0.53± 0.05 | -4.23 |
| | GalaxyRefine 2 | 1 | 97.4359 | 92.9 | 6 | 276 | 22.33 | 0.147 | 0.067 | 0.04 | 188 | 5.85 | 0.402 | 0.3 | - | 1.15 | 1.49 | 96.15 | 0.53± 0.05 | -3.42 |

**Table 2: Continue..**

**Table 2: Continue..**

| Evaluation server | | | Saves | | | | | TM-score | | | TM-align | | | | PROSA | Swissmodel structure assessment | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tertiary structure prediction server | Model refinement server | Model no. | ERRAT Overall quality | PROCHECK RAM CORE | PROCHECK RAM Allowed | Number of residues in common | RMSD of the common residues | TM-score | GDT-TS | GDT-HA | Aligned length | RMSD | TM-score Normalized by length of Chain_1 | TM-score Normalized by length of Chain_2 | Z-score | Molprobity | Clash score | Ramachandran favoured (%) | Qmean disco | Qmean Z-score |
| I-Tasser | 3D-refine | 1 | 93.2624 | 85 | 10.5 | 276 | 21.68 | 0.126 | 0.057 | 0.036 | 161 | 5.77 | 0.36 | 0.266 | -3.74 | 3.83 | 54.68 | 87.15 | 0.43± 0.05 | -8.3 |
| | ModRefiner | 1 | 68.3274 | 91.4 | 6 | 276 | 21.8 | 0.125 | 0.054 | 0.033 | 160 | 5.71 | 0.36 | 0.266 | - | 3.47 | 119.02 | 93.75 | 0.37± 0.05 | -7.18 |
| | GalaxyRefine | 4 | 87.8136 | 90.6 | 7.1 | 276 | 21.74 | 0.125 | 0.057 | 0.036 | 162 | 5.93 | 0.36 | 0.265 | -3.78 | 2.73 | 31.24 | 93.75 | 0.41± 0.05 | -6.37 |
| | DeepRefiner | 1 | 89.0071 | 83.8 | 14.3 | 276 | 22.01 | 0.121 | 0.054 | 0.033 | 168 | 6.23 | 0.352 | 0.263 | -3.28 | 3.18 | 157.1 | 90.91 | 0.45± 0.05 | -5.7 |
| | refold3 | 4 | 93.9716 | 87.6 | 10.5 | 276 | 21.97 | 0.124 | 0.057 | 0.035 | 148 | 4.67 | 0.377 | 0.273 | -3.38 | 2.77 | 10.2 | 90.62 | 0.45± 0.05 | -5.39 |
| | GalaxyRefine 2 | 1 | 91.6058 | 88 | 9.8 | 276 | 21.79 | 0.124 | 0.057 | 0.036 | 159 | 5.75 | 0.353 | 0.262 | - | 1.96 | 7.44 | 90 | 0.42± 0.05 | -6.36 |
| Lomets | ModRefiner | 1 | 93.2624 | 93.6 | 4.5 | 276 | 21.92 | 0.131 | 0.06 | 0.038 | 156 | 6.21 | 0.33 | 0.245 | -3.85 | 2.57 | 53.99 | 94.79 | 0.79± 0.05 | -4.47 |
| | 3D-refine | 1 | 94.3262 | 90.2 | 6 | 276 | 21.95 | 0.131 | 0.062 | 0.038 | 152 | 6.05 | 0.326 | 0.242 | - | 2.28 | 14.66 | 92.71 | 0.51± 0.05 | -3.89 |
| | GalaxyRefine | 2 | 96.0993 | 90.6 | 6.8 | 276 | 21.99 | 0.131 | 0.061 | 0.038 | 154 | 6.18 | 0.327 | 0.243 | -3.93 | 2.04 | 13.17 | 93.75 | 0.49± 0.05 | -3.7 |
| | DeepRefiner | 1 | 95.0178 | 86.4 | 10.6 | 276 | 22.07 | 0.131 | 0.06 | 0.038 | 145 | 6.01 | 0.313 | 0.232 | -3.35 | 3.09 | 146.37 | 92.53 | 0.49± 0.05 | -3.56 |
| | refold3 | 1 | 91.3043 | 90.2 | 6 | 276 | 22.15 | 0.129 | 0.059 | 0.036 | 150 | 6.04 | 0.322 | 0.239 | -3.06 | 1.24 | 0.64 | 92.36 | 0.45± 0.05 | -5.45 |
| | GalaxyRefine 2 | 1 | 95.7447 | 89.8 | 7.9 | 276 | 22.83 | 0.131 | 0.062 | 0.039 | 156 | 6.24 | 0.329 | 0.245 | - | 1.5 | 2.55 | 92.69 | 0.52± 0.05 | -4.46 |
| Swiss | ModRefiner | 1 | 70.5882 | 95 | 5 | 25 | 5.47 | 0.04 | 0.027 | 0.016 | 20 | 2.61 | 0.379 | 0.043 | - | 1.75 | 17.63 | 100 | 0.38± 0.05 | -0.18 |
| Model | GalaxyRefine | 3 | 66.6667 | 95 | 5 | 25 | 5.44 | 0.04 | 0.028 | 0.017 | 20 | 2.66 | 0.43 | 0.043 | 0.12 | 0.5 | 0 | 100 | 0.43± 0.05 | -0.03 |
| | DeepRefiner | 1 | 100 | 95 | 5 | 25 | 5.43 | 0.041 | 0.028 | 0.016 | 20 | 2.73 | 0.41 | 0.042 | -0.18 | 2.55 | 122.11 | 100 | 0.55± 0.05 | -0.84 |
| | refold3 | 1 | 90.9091 | 95 | 5 | 25 | 5.51 | 0.04 | 0.028 | 0.016 | 20 | 2.53 | 0.404 | 0.043 | -0.02 | 0.8 | 0 | 95.65 | 0.34± 0.05 | -0.09 |

the PROSA Z-score of -1.33 definitively places this model within the range of scores typically observed for experimentally determined X-ray structures of similar size (Wiederstein and Sippl, 2007). While other servers like Robetta also produced good models, the AlphaFold-DeepRefiner combination demonstrated superior and consistent performance across all validation parameters.

It is noteworthy that AlphaFold, a deep learning system, outperformed traditional homology modeling and threading-based servers. This aligns with the broader revolution in protein structure prediction, where neural networks trained on known structures can now accurately predict distances and angles, even for proteins with few homologous templates (Guo *et al*., 2020; van der Hoeven *et al*., 2016). Our study provides a concrete example of this superiority for a challenging viral transmembrane protein.

A potential limitation of any computational study is the lack of experimental validation. However, our multi-pronged validation strategy, demonstrating convergence across multiple independent quality metrics, provides a high degree of confidence in the selected model. Collectively, these findings support the reliability and structural plausibility of the predicted NSP6 model. Importantly, the availability of a high-confidence NSP6 structure may facilitate future structure-based drug discovery approaches, including molecular docking, virtual screening and molecular dynamics simulations aimed at identifying potential inhibitors targeting NSP6-mediated host-virus interactions. Similar structure-guided in silico approaches have been successfully applied for identifying and evaluating potential therapeutic compounds against multiple biological targets, highlighting the value of reliable structural models in rational drug discovery (Veni *et al*., 2022). Future molecular dynamics simulations in membrane-mimetic environments may further validate the structural stability of the predicted NSP6 structure.

## CONCLUSION

In conclusion, through a systematic pipeline of prediction, refinement and validation, we have identified the AlphaFold-DeepRefiner model as the best-predicted tertiary structure of SARS-CoV-2 NSP6. This model is not merely a prediction but a rigorously validated structural hypothesis. It serves as an indispensable resource for the scientific community, providing a solid foundation for future functional investigations, molecular dynamics analyses and structure-based drug discovery against this critical viral target.

## ACKNOWLEDGEMENT

The authors would like to acknowledge the Deanship of Graduate Studies and Scientific Research at Taif University for funding this work.

### Author contributions

Conceived and designed the experiments: Mohammed Mostafa Salama, Ashraf Albrakati, Medhat Wahba Shafaa, Mohamed El-Sayed El-Nagdy and Mohamed El-Sayed Hasan; Performed the experiments: Mohammed Mostafa Salama, Medhat Wahba Shafaa, Elham H Alrubai, Mohamed El-Sayed El-Nagdy and Mohamed El-Sayed Hasan; Analyzed and interpreted the data: Mohammed Mostafa Salama, Medhat Wahba Shafaa, Mohamed El-Sayed El-Nagdy and Mohamed El-Sayed Hasan; Contributed reagents, materials, analysis tools or data: Ahmed E. Abdel Moneim and Manal F. El-Khadragy; Wrote the paper: Mohammed Mostafa Salama, Medhat Wahba Shafaa, Mohamed El-Sayed El-Nagdy and Mohamed El-Sayed Hasan; All authors have read and agreed to the published version of the manuscript.

### Data availability statement

We confirm that all original raw data is available at the time of submission. As per the Data Policy, this data will be stored for a minimum of 10 years and will be made available to the Editorial Office, Editors and readers upon request.

### Funding

The authors would like to acknowledge the Deanship of Graduate Studies and Scientific Research at Taif University for funding this work.

### Disclaimers

Not applicable.

### Informed consent statement

Not applicable.

### Conflict of interest

The authors declare no conflicts of interest.

## REFERENCES

Abdelkader, A., Elzemrany, A.A., El-Nadi, M., Elsabbagh, S.A., Shehata, M.A., Eldehna, W.M., El-Hadidi, M. and Ibrahim, T.M. (2022). *In silico* targeting of SARS-CoV-2 NSP6 for drug and natural products repurposing. *Virology.* **573:** 96-110.

Adiyaman, R. and McGuffin, L.J. (2021). ReFOLD3: Refinement of 3D protein models with gradual restraints based on predicted local quality and residue contacts. *Nucleic Acids Research.* **49(W1):** W589-W596.

Altman, R.B. and Dugan, J.M. (2003). Defining bioinformatics and structural bioinformatics. *Methods Biochem. Anal.* **44:** 3-14.

Benkert, P., Biasini, M. and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.* **27(3):** 343-350.

Cucinotta, D. and Vanelli, M. (2020). WHO declares COVID-19 a pandemic. *Acta Biomed*. **91(1):** 157-160.

Gordon, D.E, Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., Tummino, T.A. and Hüttenhain, R. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. **583(7816):** 459-468.

Guo, Y.R., Cao, Q.D., Hong, Z.S., Tan, Y.Y., Chen, S.D., Jin, H.J., Tan, K.S., Wang, D.Y. and Yan, Y. (2020). The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak-an update on the status. *Mil. Med. Res.* **7(1):** 11. doi: 10.1186/s40779-020-00240-0.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y. *et al*., (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* **395(10223):** 497-506.

Ismi, D.P., Pulungan, R. and Afiahayati, (2022). Deep learning for protein secondary structure prediction: Pre and post-AlphaFold. *Comput. Struct. Biotechnol. J.* **20:** 6271-6286.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., Xing, X., Xiang, N. *et al*., (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382(13):** 1199-1207.

Liu, Z., Zhi, Y., Liu, Y., Mei, C. and Wang, H. (2025). Comprehensive genomic analysis of two *Avibacterium paragallinarum* strains with significantly different virulence. *Indian Journal of Animal Research.* **59(12):** 2006-2017. doi: 10.18805/IJAR.BF-2033.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., *et al*. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet.* **395(10224):** 565-574.

Lubin, J.H., Zardecki, C., Dolan, E.M., Lu, C., Shen, Z., Dutta, S., Westbrook, J.D., Hudson, B.P., Goodsell, D.S., Williams, J.K., *et al.* (2022). Evolution of the SARS-CoV-2 proteome in three dimensions (3D) during the first 6 months of the COVID-19 pandemic. *Proteins.* **90(5):** 1054-1080.

Meng, Y., Zhang, Z., Zhou, C., Tang, X., Hu, X., Tian, G., Yang, J. and Yao, Y. (2025). Protein structure prediction *via* deep learning: An in-depth review. *Front Pharmacol.* **16:** 1498662.

Morais, I.J., Polveiro, R.C., Souza, G.M., Bortolin, D.I., Sassaki, F.T. and Lima, A.T.M. (2020). The global population of SARS-CoV-2 is composed of six major subtypes. *Sci. Rep*. **10(1):** 18289.

Pandey, P., Prasad, K., Prakash, A. and Kumar, V. (2020). Insights into the biased activity of dextromethorphan and haloperidol towards SARS-CoV-2 NSP6: *in silico* binding mechanistic analysis. *J. Mol. Med. (Berl).* **98(12):** 1659-1673.

Rajith, R.B., Ekambaram, B., Laxmi, P., Harikrishna, C.H., Bhattacharya, T. and Sushma, G. (2023). CRISPR/Cas genome editing single guide RNA (sgRNA) design using three different web tool platforms. *Indian Journal of Animal Research.* **58(1):** 13-20. doi: 10.18805/IJAR.B-5200.

van der Hoeven, B., Oudshoorn, D., Koster, A.J., Snijder, E.J., Kikkert, M. and Bárcena, M. (2016). Biogenesis and architecture of arterivirus replication organelles. *Virus. Res.* **220:** 70-90.

Veni, P., Thangapandiyan, M., Paramasivam, R. and Rao, G.V. (2022). Exploring inhibitory potential of curcumin against multiple targets involved in the cancer progression, metastasis and apoptosis pathways by *in silico* molecular docking. *Indian Journal of Animal Research.* **59(11):** 1841-1845. doi: 10.18805/IJAR.B-4917.

Wiederstein, M. and Sippl, M.J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35(Web server issue):** W407-410.

Zhang, J., Lan, Y. and Sanyal, S. (2020). Membrane heist: Coronavirus host membrane remodeling during replication. *Biochimie.* **179:** 229-236.

Zhang, Y., Arakaki, A.K. and Skolnick, J. (2005). TASSER: An automated method for the prediction of protein tertiary structures in CASP6. *Proteins.* **61 Suppl 7:** 91-98.

Zhou, A.Q., O'Hern, C.S. and Regan, L. (2011). Revisiting the Ramachandran plot from a new angle. *Protein Sci.* **20(7):** 1166-1171.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., Chen, H.D., Chen, J., Luo, Y., *et al.* (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. **579(7798):** 270-273.

Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., Yu, J., Kang, M., Song, Y., Xia, J., Guo, Q., Song, T., He, J., Yen, H.L., Peiris, M. and Wu, J. (2020). SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N. Engl. J. Med.* **382(12):** 1177-1179.