RESEARCH ARTICLE Indian Journal of Animal Research, Volume 60 Issue 4: 644-651 (April 2026)

Evolutionary Conservation, Functional Motifs and Molecular Dynamics of SARS-CoV-2 NSP6

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], H.H. Osman[4], Mohamed El-Sayed Hasan[6]

10.18805/IJAR.BF-2096

ABSTRACT

Background: SARS-CoV-2 Non-Structural Protein 6 (NSP6) is pivotal for viral replication, but a comprehensive understanding of its evolutionary stability, functional sites and dynamic behavior has been limited. Leveraging a newly established high-confidence 3D model, this study provides an integrative analysis of NSP6's biology and biophysics.

Methods: The validated NSP6 structure was analyzed for conserved regions and evolutionary history using BioEdit and MEGA11 for phylogenetic tree construction. Functional motifs and post-translational modification (PTM) sites were predicted using PROSITE, SMART, MotifFinder and MotifScan. Structural classification was performed using CATH, SCOP and SUPERFAMILY. A 100 ns molecular dynamics (MD) simulation was conducted using GROMACS with the CHARMM27 force field to evaluate structural stability through RMSD, RMSF, Rg and SASA.

Result: Phylogenetic analysis revealed NSP6's close relationship to bat coronaviruses and identified a single, fully conserved domain across its entire 290-amino-acid length. Motif analysis identified the definitive Coronavirus replicase NSP6 domain and predicted critical PTM sites, including Casein Kinase II phosphorylation and N-myristoylation sites. Structural classification revealed an unexpected homology to cobalamin adenosyltransferase-like folds. The 100 ns MD simulation demonstrated outstanding model stability, with low RMSD (0.2-0.35 nm after 20 ns), a consistent radius of gyration (2.04±0.01 nm) and stable solvent-accessible surface area.

Key words: Functional motifs, Molecular dynamics simulation, NSP6, Phylogenetic tree, Post-translational modifications, RMSD, RMSF, SARS-CoV-2.

INTRODUCTION

The relentless spread and evolutionary trajectory of SARS-CoV-2 have highlighted the importance of understanding the molecular nuances of its constituent proteins (Huang *et al*., 2020; Santerre *et al*., 2021). Non-Structural Protein 6 (NSP6) is a key viral agent involved in remodeling host cell membranes to form replication organelles, a process critical for the virus to evade immune surveillance and efficiently replicate its genome (Abdelkader *et al*., 2022; Altman and Dugan, 2003). While the genomic sequence of NSP6 is known, a deeper comprehension of its evolutionary history, functional regions and dynamic behavior is essential to fully appreciate its role in the viral lifecycle and identify potential vulnerabilities.

The high-confidence tertiary structure of NSP6, which we recently determined through an integrative computational pipeline [Companion Paper, currently under review], provides a unique opportunity to probe its biology beyond the primary sequence. With this robust structural model in hand, we can now investigate fundamental questions about the protein's characteristics. Phylogenetic analysis allows us to trace the evolutionary relationships of NSP6 across coronaviruses, revealing conserved domains and residues under purifying selection that are likely critical for function (Zhang *et al*., 2005; Creighton 1990). Furthermore, elucidating functional motifs and post-translational modification (PTM) sites may provide insights into potential regulatory mechanisms; for instance, predicted phosphorylation and myristoylation sites are known to influence protein-protein interactions, membrane association and subcellular localization in viral proteins (Floudas *et al*., 2006; Cheng and Baldi, 2007).

However, a static protein structure represents only a structural snapshot. To gain insight into its dynamic

[1]Helwan University, Faculty of Science, Physics Department, Medical Biophysics Division, 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, Saudi Arabia.
[4]Department of Human Anatomy, College of Medicine, Taif University, 21944, Saudi Arabia.
[5]Department of Pathology, College of Medicine, Taif University, Saudi Arabia.
[6]University of Sadat City, Genetic Engineering and Biotechnology Research Institute, Bioinformatics Department, Sadat City 32897, Egypt.

Corresponding Author: Ashraf Albrakati, Department of Human Anatomy, College of Medicine, Taif University, 21944, Saudi Arabia. Email: a.albrakati@tu.edu.sa

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., Osman, H.H. and Hasan, M.E. (2026). Evolutionary Conservation, Functional Motifs and Molecular Dynamics of SARS-CoV-2 NSP6. *Indian Journal of Animal Research.* **60(4):** 644-651. **doi:** 10.18805/IJAR.BF-2096.

Submitted: 11-12-2025 **Accepted:** 18-03-2026 **Online:** 17-04-2026

behavior under near-physiological conditions, molecular dynamics (MD) simulations can be employed (Cheng and Baldi, 2007; Montgomery and Rogers, 2003). MD simulations can assess the structural stability and flexibility of the predicted NSP6 model over time, revealing rigid core regions and flexible loops that may be important for its function. Parameters such as root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), radius of gyration (Rg) and solvent-accessible surface area (SASA) provide quantitative measures of the protein's conformational stability and compactness (Stecher *et al*., 2020; Xue *et al*., 2013).

Therefore, the objective of this study was to leverage the validated NSP6 structure to conduct an in-depth analysis of its evolutionary conservation, functional annotation and dynamic properties. We performed extensive phylogenetic reconstruction, identified conserved domains and potential PTM sites and conducted a 100-ns MD simulation. This integrative approach provides a comprehensive biophysical and evolutionary profile of NSP6, offering valuable insights that bridge the gap between its atomic structure and its biological function in SARS-CoV-2 pathogenesis.

MATERIALS AND METHODS

The methodology used accurately analyzes the SARS-CoV-2 NSP6 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 and conserved region analysis

Multiple sequence alignments were carried out for the SARS-CoV-2 NSP6 protein against various sequences and the conserved region was predicted using BioEdit software (Alzohairy, 2011) and MEGA11 software (Tamura *et al*., 2021). The conserved regions in NSP6 were scanned using BIOEDIT software, which is an advanced editor of biological sequence alignment that tends to implement the fundamental editing, alignment, manipulation and analysis capabilities for protein sequences.

Molecular evolutionary and phylogenetic analysis

To study the evolutionary history of the NSP6 protein, neighbor-joining alignment was applied and to determine the phylogenetic trees, maximum likelihood and unweighted pair group mean (UPGMA) alignments were used, which enabled us to determine the topology of the trees and the length of the branches according to the primary sequence of the protein. The three algorithms of multiple sequence alignment in MEGA11 software were employed to generate evolutionary phylogenetic trees.

Domain separation and analysis

Domain separation and differentiation are the milestones in the process of tertiary structure prediction in SARS-CoV-2 NSP6 using ThreaDom (Xue *et al*., 2013), GalaxyDom (Ko *et al*., 2012), Interpro (Hunter *et al*., 2009), ProDom and the NCBI (Conserved Domain Database). ThreaDom is ranked as one of the most accurate servers for CASP8, CASP9 and CASP10. The template-based algorithm is used to predict the boundaries of domains of proteins whose locations are determined by multiple threading alignments (Xue *et al*., 2013). The NCBI Conserved Domain Database (CDD) provides the ability to annotate queries with the positions of conserved protein domain footprints, as well as the functional sites and motifs inferred from these footprints.

Functional motif and post-translational modification (PTM) prediction

To detect the motifs in the SARS-CoV-2 NSP6 protein, we used the PROSITE web server (Sigrist *et al*., 2013), SMART (Letunic *et al*., 2021), MotifFinder (Kanehisa, 1997), HITS (Liu *et al*., 2020), MEME (Bailey *et al*., 2015) and MotifScan (Pagni *et al*., 2007). SMART applies statistical models such as hidden Markov models (HMMs), profiles, or position-specific scoring matrices (PSSMs), in which multiple sequence alignments are used to preserve and express the sequence data as probabilistic models (Letunic *et al*., 2021). The MotifFinder web server is a de novo motif identification program that finds statistically significant patterns in a collection of protein sequences while considering their evolutionary connections (Kanehisa, 1997). MotifScan scans the sequences and searches for recognized motifs when given a set of input genomic regions (Pagni *et al*., 2007).

Structural classification

To find the structural classification and functional annotations of the SARS-CoV-2 NSP6 protein, we used the CATH (Sillitoe *et al*., 2021), InterPro, SCOP (Chandonia *et al*., 2019) and super family (Gough *et al*., 2001; Pandurangan *et al*., 2019) databases. CATH uses the PDB database to identify the domains in proteins and correctly clarify their evolutionary superfamilies. The database includes 151 million (59%) predicted domains, 500238 structural domains and 65351 (15%) fully classified domain structures, which are assigned to 5481 superfamilies (Wallner and Elofsson, 2003).

Molecular dynamics simulation

In the simulation process, every protein-ligand complex was centered in a cubic box with 10 Å distance to the surrounding edges (of boundary condition was defined in the x, y and z directions) and surrounded by TIP3P water molecules model. The salt concentration of the complex was maintained by adding 0.15M of Na^+ and Cl^- ions to the simulation box (Huang *et al*., 2017; Kutzner *et al*., 2019; Vanommeslaeghe *et al*., 2012) and particle-Ewald summation was used to evaluate the electrostatic interactions (Darden *et al*., 1993) and for the estimation of VdW interactions a 10 Å cut-off was used. Then the resulting systems were energy-minimized for 50000 steps through the steepest descent and conjugate gradient algorithms.

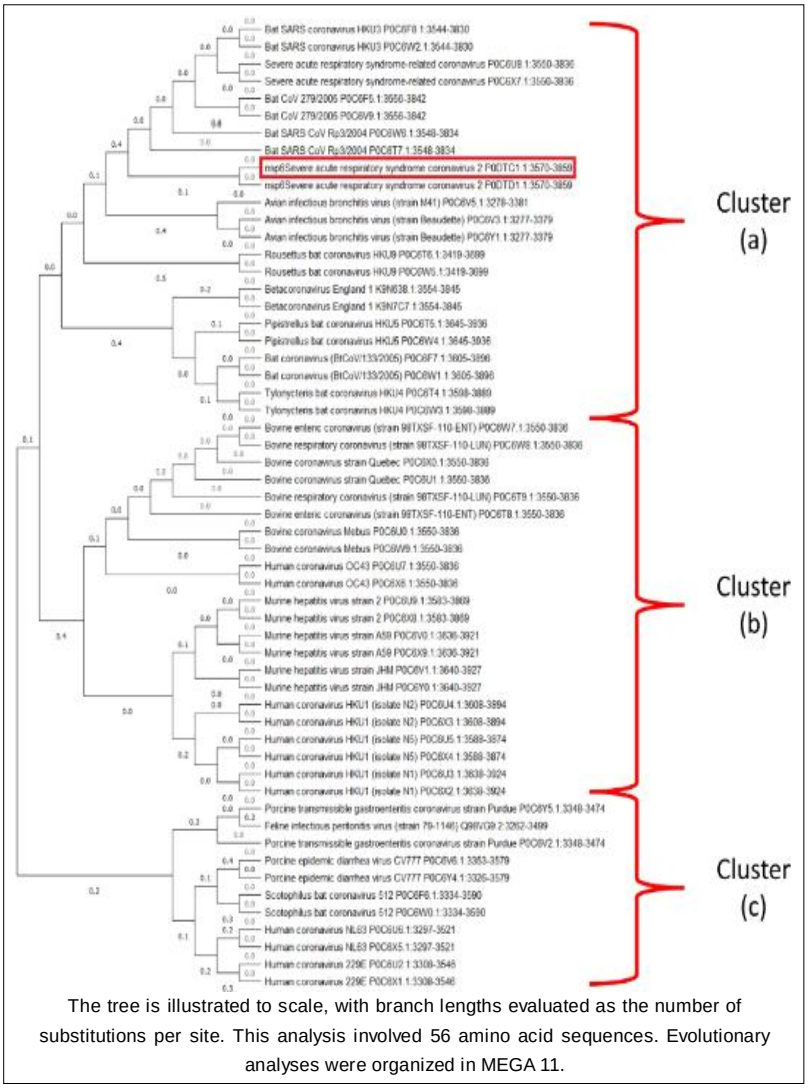
After that, the systems were subjected to an NPT and NVT equilibrium phases for 100 ps for each process with integration step of 2fs (Bussi *et al*., 2007; Singh *et al*., 2020). Finally, the MD simulations production run for 100 ns with snapping every 10 ps for the analysis (Singh *et al*., 2020; Prakash *et al*., 2018a). All production runs were performed on The Bibliotheca Alexandrina Supercomputing unit. GROMACS-2021.3 was used for the molecular dynamic simulation of the protein and protein-ligand complexes, with applying CHARMM27 forcefield in the estimation of the interactions (Huang *et al*., 2017; Kutzner *et al*., 2019; Vanommeslaeghe *et al*., 2012). GROMACS-2021.3 package was used for analyzing the results of the MD simulations including the protein root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (RG) and solvent accessible surface area (SASA) (Laberge and Yonetani, 2008; Prakash *et al*., 2018b).

RESULTS AND DISCUSSION

Conserved regions and evolutionary analysis

We used BioEdit software to detect the conserved region of the SARS-CoV-2 NSP6 protein. The results showed that there was only one conserved region that encompassed the entire full length of the protein, with a segment length of 290 and an average entropy of 0.2.

The maximum likelihood, neighbor-joining and unweight pair group method with arithmetic mean (UPGMA) methods were used to establish the evolutionary history of the NSP6 protein of SARS-CoV-2 based on the Jones-Taylor-Thornton (JTT) matrix-based model (Fig 1). The phylogenetics study shows that NSP6 is related to the Bat SARS CoV, severe acute respiratory syndrome-related coronavirus, Rousettus bat coronavirus HKU9, Betacoronavirus England 1, Pipistrellus bat coronavirus HKU5, Bat coronavirus (BtCoV/133/2005), Tylonycteris bat coronavirus



The tree is illustrated to scale, with branch lengths evaluated as the number of substitutions per site. This analysis involved 56 amino acid sequences. Evolutionary analyses were organized in MEGA 11.

Fig 1: Molecular phylogenetic analysis of the NSP6 using the neighbor-joining method.

HKU4, murine hepatitis virus strain JHM, bovine enteric coronavirus, human coronavirus HKU1 (isolate N2), murine hepatitis virus strain JHM, human coronavirus OC43, bovine coronavirus Mebus, porcine epidemic diarrhea virus CV777, porcine transmissible gastroenteritis coronavirus strain Purdue, Scottophilus bat coronavirus 512, avian infectious bronchitis virus and Feline infectious peritonitis virus. The total of amino acid substitutions per site among sequences were organized using the Poisson correction model (Huang *et al*., 2020). This analysis implied 57 amino acid sequences. All confusing positions were removed for each sequence pair (pairwise deletion option). Here was a total of 305 locates in the final dataset. The incidence of 'n/c' in the results indicates cases where it was not probable to estimate evolutionary distances.

Domain architecture

The NCBI conserved domain server revealed that the SARS-CoV-2 NSP6 is composed of only one domain that covers all sequences of the protein, with an accession number (cd21560), interval (1-290) and E value (3.29e+129).

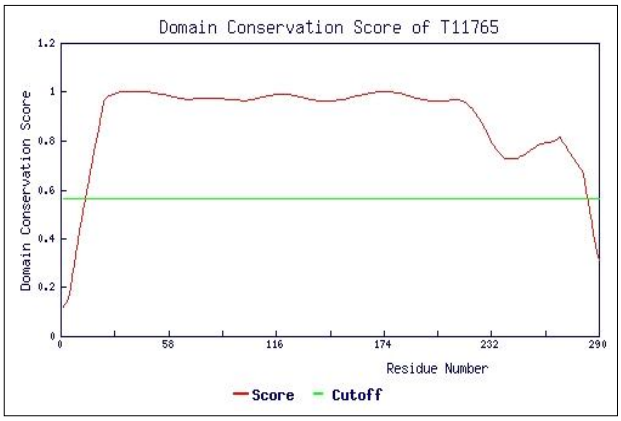


Fig 2: Domain separation of SARS-CoV-2 NSP6 using the ThreaDom server.

Moreover, the ThreaDom, InterPro, GalaxyWEB Dom and ProDom servers support the results of having only one domain with a cutoff of 0.56, a score of 1290 (572.8 bits), an E value of 1e+165 and an identity of 287/290 (99%), (Fig 2).

Functional motifs and post-translational modifications

The PROSITE web server and the SMART, 3D-blast, Dali, TM, MotifFinder, HITS, HMMER, MEME, MotifScan and Smotif servers were used to analyze the motifs of the NSP6 of SARS-CoV-2 (Table 1, 2). To study the motifs of the NSP6 model, The MotifFinder server revealed two motifs. The first motif was Coronavirus replicase NSP6 (PF19213), which spans amino acids 29 to 290, a score of 206.2 and an E-value of 1.10E-60. The second most common sequence was TM1506 (PF08973), which had an alignment range of 91:133 amino acid residues, a score of 11.6 and an E value of 0.18 (Table 3). The HMMER server is compatible with the MotifFinder server, as it contains only one motif, the coronavirus replicase NSP6 (PF19213).

Prediction of post-translational modification site (PTMs)

To figure out the nature of the interactions of the protein, the MotifScan server identified five PTM predictions: an N-glycosylation site at 174-177, a casein kinase II phosphorylation site at 96-99, 130-133 and 247-250, an N-myristoylation site at 177-182 and 279-284, a protein kinase C phosphorylation site at 6-8 and 127-129 and a tyrosine kinase phosphorylation site at 109-116, (Table 3).

Myristyl groups permit weak interactions between proteins and lipids as well as between proteins (Farazi *et al*., 2001). It is crucial for membrane targeting and protein protein interactions and has a wide range of applications in signal transduction pathways. Thus far, N-myristoylation, a posttranslational modification of proteins, is believed to be essential for the anchoring of proteins to membranes (Hayashi and Titani, 2010). Remarkably, the anticipated myristoylated proteins included a significant number of myristoylated proteins assumed to be involved in signal

Table 1: Motif analysis of the SARS-CoV-2 NSP6 using the motif finder server.

#Query	DBNAME	ID	POS	Score	EVAL
MMO	Pfam	CoV_NSP6	29..290	206.2	1.10E-60
MMO	Pfam	TM1506	91..133	11.6	0.18

MMO: Motif model output; Pfam: Protein families database; POS: Position of the motif in the sequence; Score: Alignment score generated by the server; E-value: Expectation value indicating statistical significance.

Table 2: Motif analysis of the SARS-CoV-2 NSP6 protein using the SMART server.

Name	Start	End	E-value	Confidence
Low complexity	16	32	N/A	Confident
Solute_trans_a	19	243	308000	Threshold
VKc	42	133	7810	Threshold
LAB_N	50	97	120000	Threshold
B561	66	159	443	Threshold
WR1	198	227	1070	Threshold

transduction between the membrane and cytoplasmic fractions. N-glycosylation is a vital protein modification that impacts numerous cellular processes and is required for the appropriate folding of membrane and secretory proteins (Yasuda *et al*., 2015). Protein kinase C phosphorylates serine or threonine residues adjacent to a basic residue in the C-terminal region via the enzyme. The Vmax and km of phosphorylation increase when the target amino acid has extra basic residues at its N- or C-terminus (Lim *et al*., 2015). As a constitutively active Ser/Thr protein kinase, Casein kinase II is involved in numerous human disorders and phosphorylates hundreds of substrates in addition to controlling other signaling pathways. Its function in cancer is better understood because it controls almost all the telltale signs of this disease. Human infections are among the other well-known CK2 infections; in particular, a variety of viruses use CK2 in host cells for reproduction (Borgo *et al*., 2021).

Structural classification

To study the family and the full lineage of the NSP6, the superfamily server displayed only one matching protein, 2g2d A, at 69-173, with a family e-value of 0.025. This protein was found to belong to the superfamily of cobalamin adenosyltransferase-like proteins, with a superfamily e-value of 9.66e-02. For more advanced details about the 2g2d A: protein, we used the SCOPe server to track the full lineage of the protein (Table 4, 5).

Molecular dynamics simulations demonstrate model stability

The simulation was applied to the unbounded NSP6 at a 100 ns time function. The trajectories and parameters of NSP6 were analyzed according to different parameters, which are Root Mean Square Deviations (RMSDs), gyration radius, Root Mean Square Fluctuations (RMSFs) and the solvent-accessible surface area (SASA). The molecular dynamics simulation was applied to the NSP6 at a 100 ns time function.

RMSD

The structural and conformational stability of the SARS-CoV-2 protein was estimated by evaluating the RMSD of the protein backbone atoms against the time of simulation. The NSP6 exhibited a sharp increase in RMSD up to 20 ns, stabilizing between 0.2 and 0.35 nm which reveals a

Table 3: Prediction of PTMs of the NSP6 in SARS-CoV-2 using the motif scan server.

Category	Signature	Matching positions
	ASN_GLYCOSYLATION	174-177
	N-glycosylation site	
RNA	CK2_PHOSPHO_SITE	96-99
Associated	*Casein kinase II phosphorylation site*	130-133
Protein	MYRISTYL	247-250
Domain	*N-myristoylation site*	177-182
Posttranslational	PKC_PHOSPHO_SITE	279-284
Modifications	*Protein kinase C phosphorylation site*	6-8
	TYR_PHOSPHO_SITE	127-129
	Tyrosine kinase phosphorylation site	109-116

Table 4: Structural Classification of Protein (SCOP) of the NSP6 of SARS-CoV-2 using the superfamily server.

Sequence	Start residue	End residue	Super family	Super family E-value	Family	Family E-value
SARS-CoV-2 NSP6	69	173	Cobalamin adenosyltransferase-like	9.66e-02	SCOPe (2g2d A:)	0.025

Table 5: The full lineage of the 2g2d A: protein.

Root	SCOPe 2.08
Class	a: All alpha proteins [46456] (290 folds)
Fold	a. 25: Ferritin-like_[47239] (6 superfamilies)
	Core: 4 helices; bundle, closed, left-handed twist; 1 crossover connection
Superfamily	a. 25.2: Cobalamin adenosyltransferase-like [89028] (3 families)
	Crossover loop goes across a different side of the 4-helical bundle; no internal metal-binding site
Family	a. 25.2.0: automated matches [191442] (1 protein) not a true family
Protein	automated matches [190652] (6 species) not a true protein
Species	Mycobacterium tuberculosis [TaxId:1773] [187731] (5 PDB entries)
Domain	d2g2da_: 2g2d A: [164561] automated match to d1rtyb_

significant stability (Fig 3). The analysis revealed low fluctuations in the core helical regions, with higher flexibility observed in the loop regions, particularly between residues X and Y.

RMSF

RMSF is the average estimation of the displacement of a specific group of atoms or a structure to a reference structure. It is important in the prediction of the flexibility of the residues and the protein backbone (Fig 4). The low standard deviation indicates a stable, compact conformation throughout the simulation.

Radius of gyration

Radius of gyration (Rg) is an important parameter in the investigation of the compactness and integrity of the complex's structure which is used in the evaluation of the stability of the system. It is the mass-weighted RMSD of atoms from their center of mass. The average Rg values of NSP6 is 2.04±0.01 nm (Fig 5). The minimal variation in SASA confirms the stability of the protein's folded state and hydrophobic core.

SASA analysis

To have precise knowledge about the complex stability, folding and compactness of the hydrophobic core, the solvent-accessible surface area (SASA) was used to estimate the exposed area of the complex that interacts with the surrounding solvent molecules. The average SASA value for the NSP6 is 163.4±2.49 nm^2 (Fig 6).

With a high-confidence tertiary structure of SARS-CoV-2 NSP6 now established in our companion paper [Manuscript in preparation], we leveraged this model to perform a deep dive into its evolutionary history, functional landscape and dynamic behavior. Our integrative analysis reveals NSP6 to be a highly conserved, stable protein with features that elucidate its crucial role in the viral lifecycle.

The evolutionary analysis paints a picture of remarkable conservation. The presence of a single, continuous conserved domain spanning all 290 residues is unusual and indicates that NSP6 functions as an integrated, indivisible unit. This high degree of evolutionary constraint suggests that disruptive mutations anywhere in the protein are likely to be deleterious, making it an attractive and potentially resilient drug target. The phylogenetic clustering with bat coronaviruses reinforces the zoonotic origin of SARS-CoV-2 and indicates that the core function of NSP6 was already well-established in ancestral strains (Zhang *et al*., 2005).

The functional motif analysis provides mechanistic insights into how NSP6 operates. The identification of the definitive "Coronavirus replicase NSP6" domain (PF19213) confirms its specific role in the viral replication complex. More intriguing are the predicted post-translational modification sites. The presence of multiple Casein Kinase II (CK2) phosphorylation sites is highly significant, as CK2 is a host kinase frequently hijacked by viruses to regulate viral protein function and to subvert host cell processes (Borgo *et al*., 2021). The dual N-myristoylation sites strongly suggest a stable membrane anchoring mechanism, which is entirely consistent with NSP6's role in remodeling the ER into double-membrane vesicles (Abdelkader *et al*., 2022; Alshahrani, 2025). These PTMs point to a sophisticated

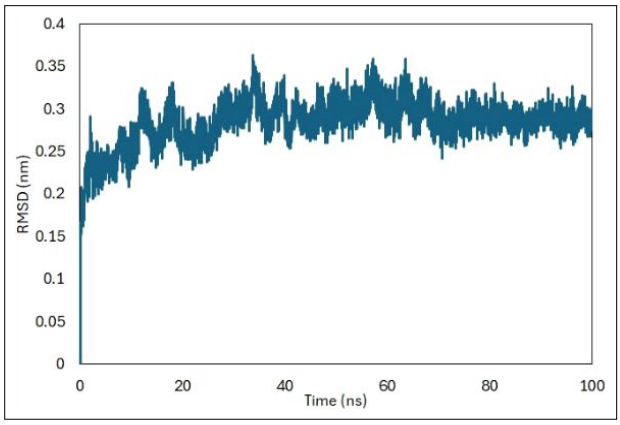


Fig 3: RMSD plot of docked complexes generated through MDS at 100 ns.

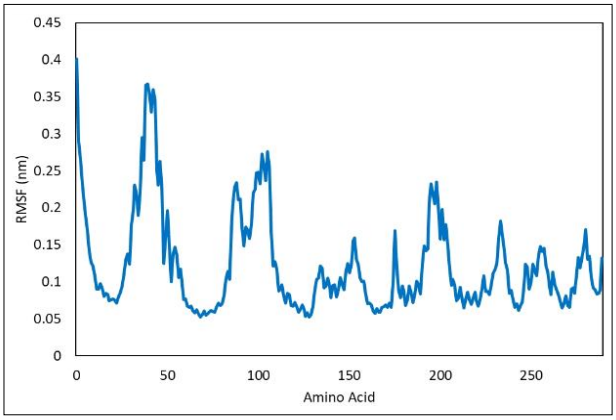


Fig 4: RMSF plot of docked complexes generated through MDS at 100 ns.

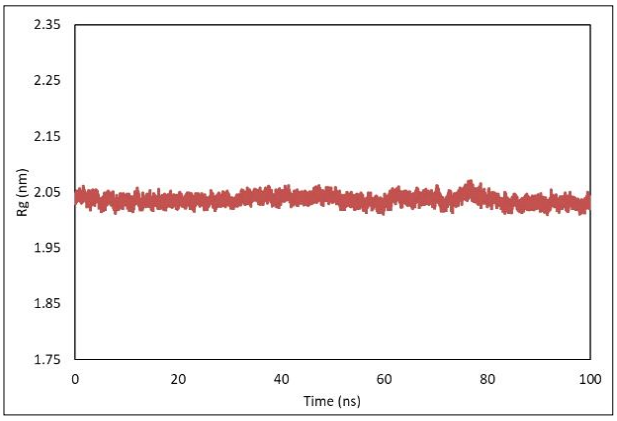


Fig 5: Rg plot of docked complexes generated through MDS at 100 ns.

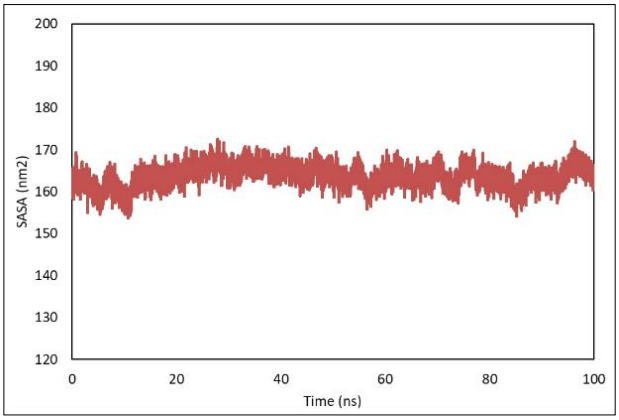


Fig 6: SASA plot of docked complexes generated through MDS at 100 ns.

level of host-pathogen interaction, where NSP6's function is potentially regulated by host cell signals.

A surprising finding was the structural homology to cobalamin adenosyltransferase-like proteins. While the sequence identity is low, the conservation of a four-helix bundle fold suggests that NSP6 may have evolved from an ancient cellular enzyme scaffold. This could be an example of viral "exaptation," where a structural motif is repurposed for a novel function-in this case, potentially mediating protein-protein interactions essential for the assembly of the viral replication organelle (Orengo *et al*., 1997).

Finally, the 100-ns molecular dynamics simulation provided a critical test of the model's biophysical realism and offered a view of the protein's behavior in a near-physiological state. The rapid stabilization and maintenance of low RMSD and consistent Rg values are hallmarks of a stable, well-folded protein. This structural rigidity is likely essential for its role as a scaffold within the viral replication complex. The RMSF analysis further illuminates this picture, showing that while the core helical bundles are rigid, specific loop regions display flexibility. These flexible regions often correspond to sites of protein-protein interaction or, notably, the locations of several predicted PTM sites, suggesting these areas are functionally dynamic.

CONCLUSION

Our study moves beyond a static structure to provide a dynamic and functional profile of NSP6. We show it to be an evolutionarily constrained, mono-domain protein that is strategically modified by the host cell and exhibits the structural stability required for its essential scaffolding role in viral replication. The convergence of evolutionary, functional and biophysical data strongly supports the biological relevance of our model and highlights specific features-such as the CK2 phosphorylation and myristoylation sites-as promising avenues for future experimental investigation and therapeutic intervention.

ACKNOWLEDGMENT

"The authors would like to extend their sincere appreciation to the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R23), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia."

Funding

We kindly request that the same statement be included under the Funding section, as it is essential for institutional documentation and reimbursement purposes.

Author contributions

Conceived and designed the experiments: Mohammed Mostafa Salama, Khalid Ebraheem Hassan, Ashraf Albrakati, Medhat Wahba Shafaa, Mohamed El-Sayed El-Nagdy and Mohamed El-Sayed Hasan; Performed the experiments: Mohammed Mostafa Salama, Medhat Wahba Shafaa, 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, H.H. Osman 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

All raw data are available upon request.

Conflict of interest

The authors declare no conflict 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.

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

Alshahrani, M.M. (2025). Inhibition of human N myristoyltransferase 1 as a strategy to suppress cancer progression driven by myristoylation. *Sci. Rep.* **15:** 28927.

Alzohairy, A.M. (2011). BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences.* **2(1):** 60-61.

Bailey, T.L., Johnson, J., Grant, C.E. and Noble, W.S. (2015). The MEME suite. *Nucleic Acids Res.* **43:** W39-49.

Borgo, C., D'Amore, C., Sarno, S., Salvi, M. and Ruzzene, M. (2021). Protein kinase CK2: A potential therapeutic target for diverse human diseases. *Signal Transduct Target Ther.* **6:** 183.

Bussi, G., Donadio, D. and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126:** 014101.

Chandonia, J.M., Fox, N.K. and Brenner, S.E. (2019). SCOPe: Classification of large macromolecular structures in the structural classification of proteins-extended database. *Nucleic Acids Res*. **47:** D475-d481.

Cheng, J. and Baldi, P. (2007). Improved residue contact prediction using support vector machines and a large feature set. *BMC Bioinformatics.* **8:** 113.

Creighton, T.E. (1990). Protein folding. *Biochem J.* **270:** 1-16.

Darden, T., York, D. and Pedersen, L. (1993). Particle mesh ewald: An nlog (N) method for ewald sums in large systems. *The Journal of Chemical Physics.* **98:** 10089-10092.

Farazi, T.A., Waksman, G. and Gordon, J.I. (2001). The biology and enzymology of protein N-myristoylation. *J. Biol. Chem.* **276:** 39501-39504.

Floudas, C.A., Ho Ki, F., McAllister, S.R., Mönnigmann, M. and Rajgaria, R. (2006). Advances in protein structure prediction and de novo protein design: A review. *Chemical Engineering Science*. **61:** 966-988.

Gough, J., Karplus, K., Hughey, R. and Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J. Mol. Biol.* **313:** 903-919.

Hayashi, N. and Titani, K. (2010). N-myristoylated proteins, key components in intracellular signal transduction systems enabling rapid and flexible cell responses. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* **86:** 494-508.

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

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmüller, H. and MacKerell, A.D.J. (2017). CHARMM 36 m: An improved force field for folded and intrinsically disordered proteins. *Nat Methods.* **14:** 71-73.

Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., *et al*. (2009). InterPro: The integrative protein signature database. *Nucleic Acids Res.* **37:** D211-215.

Kanehisa, M. (1997). Linking databases and organisms: GenomeNet resources in Japan. *Trends Biochem. Sci.* **22:** 442-444.

Ko, J., Park, H., Heo, L. and Seok, C. (2012). GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.* **40:** W294-297.

Kutzner, C., Páll, S., Fechner, M., Esztermann, A., de Groot, B.L. and Grubmüller, H. (2019). More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *J. Comput. Chem.* **40:** 2418-2431.

Laberge, M. and Yonetani, T. (2008). Molecular dynamics simulations of hemoglobin A in different states and bound to DPG: Effector-linked perturbation of tertiary conformations and HbA concerted dynamics. *Biophys J.* **94:** 2737-2751.

Letunic, I., Khedkar, S. and Bork, P. (2021). SMART: Recent updates, new developments and status in 2020. *Nucleic. Acids Res.* **49:** D458-d460.

Lim, P.S., Sutton, C.R. and Rao, S. (2015). Protein kinase C in the immune system: From signalling to chromatin regulation. *Immunology.* **146:** 508-522.

Liu, B., Jiang, S. and Zou, Q. (2020). HITS-PR-HHblits: Protein remote homology detection by combining pagerank and hyperlink-induced topic search. *Brief Bioinform.* **21:** 298-308.

Montgomery, R. and Rogers, T. (2003). North carolina association challenges the state medicaid agency and wins! *Caring: National Association for Home Care magazine.* **22:** 72-74.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B. and Thornton, J.M. (1997). CATH-A hierarchic classification of protein domain structures. *Structure.* **5:** 1093-1108.

Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C.V., Hau, J., Martin, O., Kuznetsov, D. and Falquet, L. (2007). My hits: Improvements to an interactive resource for analyzing protein sequences. *Nucleic Acids Res.* **35:** W433-437.

Pandurangan, A.P., Stahlhacke, J., Oates, M.E., Smithers, B. and Gough, J. (2019). The super family 2.0 database: A significant proteome update and a new webserver. *Nucleic Acids Res.* **47:** D490-d494.

Prakash, A., Dixit, G., Meena, N.K., Singh, R., Vishwakarma, P., Mishra, S. and Lynn, A.M. (2018a). Elucidation of stable intermediates in urea-induced unfolding pathway of human carbonic anhydrase IX. *J. Biomol Struct Dyn.* **36:** 2391-2406.

Prakash, A., Kumar, V., Meena, N.K. and Lynn, A.M. (2018b). Elucidation of the structural stability and dynamics of heterogeneous intermediate ensembles in unfolding pathway of the N-terminal domain of TDP-43. *RSC Adv.* **8:** 19835-19845.

Santerre, M., Arjona, S.P., Allen, C.N., Shcherbik, N. and Sawaya, B.E. (2021). Why do SARS-CoV-2 NSPs rush to the ER? *J. Neurol.* **268:** 2013-2022.

Sigrist, C.J., de Castro, E., Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L. and Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Res.* **41:** D344-347.

Sillitoe, I., Bordin, N., Dawson, N., Waman, V.P., Ashford, P., Scholes, H.M., Pang, C.S.M. *et al.* (2021). CATH: Increased structural coverage of functional space. *Nucleic Acids Res.* **49:** D266-d273.

Singh, R., Meena, N.K., Das, T., Sharma, R.D., Prakash, A. and Lynn, A.M. (2020). Delineating the conformational dynamics of intermediate structures on the unfolding pathway of β-lactoglobulin in aqueous urea and dimethyl sulfoxide. *J. Biomol Struct Dyn.* **38:** 5027-5036.

Stecher, G., Tamura, K. and Kumar, S. (2020). Molecular evolutionary genetics analysis (MEGA) for macOS. *Mol. Biol. Evol.* **37:** 1237-1239.

Tamura, K., Stecher, G. and Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38:** 3022-3027.

Vanommeslaeghe, K., Raman, E.P. and MacKerell, A.D.J. (2012). Automation of the CHARMM general force field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **52:** 3155-3168.

Wallner, B. and Elofsson, A. (2003). Can correct protein models be identified? *Protein Sci.* **12:** 1073-1086.

Xue, Z., Xu, D., Wang, Y. and Zhang, Y. (2013). ThreaDom: Extracting protein domain boundary information from multiple threading alignments. *Bioinformatics.* **29:** 247-256.

Yasuda, D., Imura, Y., Ishii, S., Shimizu, T. and Nakamura, M. (2015). The atypical N-glycosylation motif, Asn-Cys-Cys, in human GPR109A is required for normal cell surface expression and intracellular signaling. *Faseb J*. **29:** 2412-2422.

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.