BioLuster Research Center Ltd

Medicinal Chemistry and Therapeutics

Research Article

Open Access

In Silico Structural and Functional Characterization of the HMPV M2-1 Protein Reveals Its Potential as a Therapeutic Target

Saifullah Sarkar[1], Sanjay Das[1], Md. Hasan Ali[1]

[1]Department of Agriculture, Gopalganj Science and Technology University, Gopalganj-8100, Bangladesh

Correspondence
Md. Hasan Ali
Email: hasan.19agr019@gstu.edu.bd

Academic Editor
Muhammad Torequl Islam, PhD
Email: dmt.islam@blrcl.org

Received: 11 August 2025
Revised: 17 November 2025
Published: Advanced online

Abstract: Human metapneumovirus (HMPV) is a leading cause of acute respiratory infections in children, the elderly, and immunocompromised individuals, yet no licensed antivirals or vaccines exist. This study employed in-silico approaches to characterize the HMPV M2-1 protein, a key regulator of viral transcription, and to identify potential natural inhibitors. Physicochemical analysis revealed that M2-1 is a hydrophilic, thermally stable, and basic protein with properties favorable for RNA binding and transcriptional regulation. Conserved domain and motif analyses identified a transcription processivity factor and a zinc finger C3H1 motif, confirming its role in RNA stabilization. Gene Ontology and protein–protein interaction analyses positioned M2-1 as a multifunctional protein associated with viral replication, assembly, and host immune modulation. Homology modeling using SWISS-MODEL and I-TASSER generated high-quality 3D structures, with the SWISS-MODEL showing superior stereochemical quality and enabling accurate active site prediction. CASTpFold analysis revealed 22 potential binding pockets, with the largest active site measuring 279.006 $Å^3$. Molecular docking using AutoDock Vina in PyRx with 58 garlic (*Allium sativum*) compounds identified IMPHY010911 as the most potent binder (-7.7 kcal/mol). These findings highlight M2-1 as a promising antiviral drug target and suggest garlic bulb-derived phytochemicals may serve as potential natural inhibitors against HMPV infection.

Keywords: Human metapneumovirus (HMPV), Pneumovirus matrix protein 2 (M2), In silico, Homology modeling, Molecular docking.

1. Introduction

Human metapneumovirus (HMPV) is an enveloped respiratory pathogen of the Paramyxoviridae family, first identified in 2001 in the Netherlands. Since then, it has been reported worldwide and is recognized as a major cause of acute respiratory infections (ARIs), particularly among infants, young children, the elderly, and immunocompromised individuals (Pandey et al., 2025; Uddin & Thomas, 2025).

Morphologically, the virus is pleomorphic, appearing in spherical or filamentous forms with an average diameter of ~209 nm (Das & Dunbar, 2024; Hamelin et al., 2004). Its genome is a negative-sense, single-stranded RNA of approximately 13.3 kb that encodes nine proteins, including the fusion (F) protein that mediates host cell entry (Leyrat et al., 2014; Mozaffari et al., 2025; Soto et al., 2018). HMPV replicates in the cytoplasm of airway epithelial cells, thriving under physiological temperature (37 °C) and pH conditions of the respiratory tract (Bao et al., 2007; Tollefson et al., 2010).

The viral life cycle involves attachment to host cell receptors, membrane fusion, RNA replication, and budding from the cell membrane (Dong et al., 2025; Krüger et al., 2025). It exhibits clear seasonal patterns, peaking in late winter and early spring in temperate climates (Adedokun et al., 2025; Bhattacharya et al., 2025). Genetically, HMPV is classified into two major lineages (A and B), each with sub-lineages (A1, A2, B1, B2), reflecting significant antigenic and genetic diversity (Alotaibi et al., 2025; Krüger et al., 2025; Ullah et al., 2025).

Transmission occurs primarily through respiratory secretions via droplets, direct contact, or contaminated surfaces, and the virus can survive long enough on fomites to cause hospital outbreaks (Bhattacharya et al., 2025; Mohammadi et al., 2025). HMPV predominantly targets the respiratory epithelium, causing clinical features similar to respiratory syncytial virus (RSV), including fever, cough, wheezing, rhinorrhea, hypoxia, and radiographic abnormalities (Papenburg & Boivin, 2010; Satapathy et al., 2025; Williams et al., 2009). Severe cases may progress to bronchiolitis, pneumonia, asthma exacerbations, or complications in patients with chronic lung disease, particularly those with weakened immunity (Adedokun et al., 2025; Hejran et al., 2020).

https://doi.org/10.71193/jmct.20250008
Journal Home Page: Journal of Medicinal Chemistry and Therapeutics

Despite its substantial global health burden, no licensed vaccines or specific antiviral treatments are currently available for HMPV (Acharya & Byrareddy, 2025). Management remains supportive, while prevention relies on standard infection control practices such as hand hygiene, patient isolation, and cohorting during outbreaks (Demmler-Harrison, 2018; Foley et al., 2025). This therapeutic gap underscores the need for deeper structural and functional studies of viral proteins. Therefore, the present study aims to perform an in-silico structural and functional annotation of the HMPV M2-1 protein, a key regulator of viral replication and transcription, to evaluate its potential as a therapeutic target and provide a foundation for future drug or vaccine development.

2. Methodology

2.1. Sequence Retrieval of Hypothetical Protein

The HMPV protein sequence, consisting of 187 amino acids, was retrieved from the UniProt database in FASTA format using the accession number Q6WB97 ("UniProt: The Universal Protein Knowledgebase in 2023.," 2023). This hypothetical protein's primary sequence is FASTA format, and additional analysis was conducted using this protein's FASTA sequence.

2.2. Physicochemical Properties Determination

The physicochemical characteristics of the chosen hypothetical protein were analyzed using the ProtParam tool available on the ExPASy server (Wilkins et al., 1999). The instability index evaluates protein stability (Guruprasad et al., 1990), the aliphatic index measures the volume percentage occupied by aliphatic side-chain amino acids (Ikai, 1980), and the GRAVY value determines the overall balance of hydrophobicity and hydrophilicity (Kyte & Doolittle, 1982). Amino acid composition (Downard & Cody, 2024), atomic composition (Baudouin-Cornu et al., 2001), theoretical PI (Sillero & Ribeiro, 1989), molecular weight (Fischer et al., 2004), total number of negative and positive residues (Monné et al., 1998), Grand average hydropathicity (GRAVY), extinction coefficient (Gill & Von Hippel, 1989), aliphatic index (AI), estimated half-life (Rahman & Sadygov, 2017), and instability index (II) were calculated by this web based tool.

2.3. Functional Annotation

Conserved domains in the chosen protein were found using the CD search tool on the NCBI website. Protein domain and protein family models obtained from numerous sequence alignments are included in the CDD (Wang et al., 2023). Another resource that makes it easier to find and annotate conserved areas in protein sequences is ScanProsite (Cuche et al., 2013). These conserved sections fall into two primary categories: patterns that emphasize short sequence motifs usually associated with residues that are functionally or structurally relevant and generalized profiles that delineate modular protein domains and protein families (Bork & Koonin, 1996; Sun et al., 2004; Van Roey et al., 2014).

2.4. Gene Ontology Analysis

The Gene ontology (GO) is a standardized vocabulary for describing gene and protein functions in bioinformatics (Consortium, 2004; Du Plessis et al., 2011). It classifies them into molecular function (the activities of a gene product), biological process (the broader goals it supports), and cellular component (its location in the cell), enabling consistent annotation and cross-species genomic analysis (Dolan et al., 2005; Primmer et al., 2013). The GO of the selected protein was identified by gene ontology section of UniProt database.

2.5. Protein-Protein Interaction

A protein-protein interaction (PPI) is the physical contact between proteins, essential for cellular functions such as signaling, metabolism, and cell cycle control. PPIs may be permanent or transient, and studying their networks offers key insights into cellular physiology, biochemistry, and disease mechanisms (Athanasios et al., 2017; Lu et al., 2020; Safari-Alighiarloo et al.,

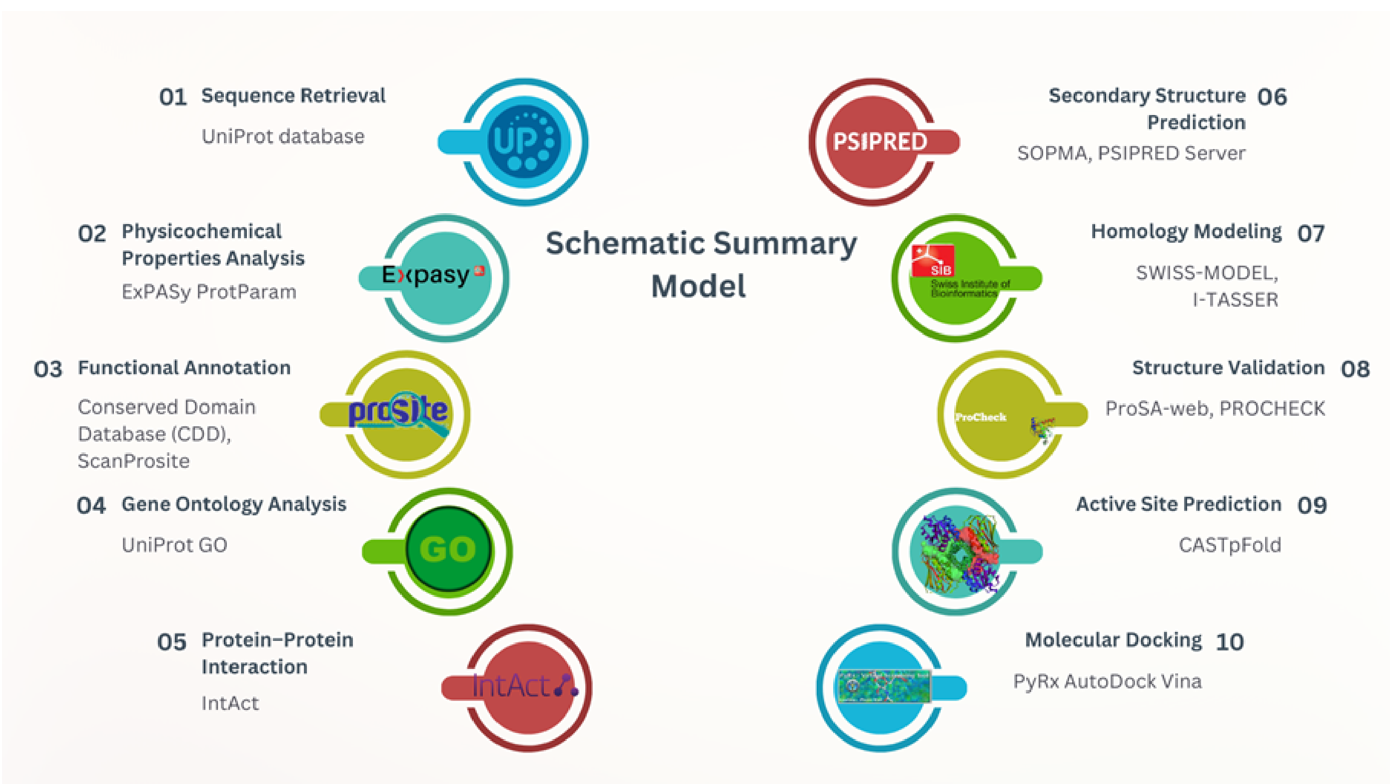


Fig. 1. Workflow diagram of HMPV M2-1 protein *in silico* study.

2014). PPI network analysis was determined by the IntAct Molecular Interaction Database (Kerrien et al., 2012).

2.6. Secondary Structure Protein

The protein's secondary structure was meticulously determined. Utilizing the SOMPA server, the structural elements were resolved (Geourjon & Deleage, 1995). Subsequently, a robust prediction of the targeted protein's secondary structure was achieved via the PSIPRED server (Buchan & Jones, 2019). The SOMPA server is one of the methods available through the NPSA server for the prediction of secondary structural components of the protein. The PSIPRED server is a widely used and relatively accurate method for predicting protein secondary structure (Buchan & Jones, 2019; Geourjon & Deleage, 1995).

2.7. Homology Modeling and Quality Assessment

By aligning a protein's amino acid sequence with a known, homologous protein structure (template) and then creating a model based on that alignment, homology modeling, sometimes referred to as comparative modeling, is a computational technique that predicts a protein's three-dimensional structure (Morgnanesi et al., 2015). By entering the protein's amino acid sequence in FASTA format, the I-TASSER (Roy et al., 2010) and SWISS-MODEL servers (Arnold et al., 2006) were able to determine the protein's possible three-dimensional structure. Using a sequence-to-structure-to-function technique, I-TASSER is an automated tool for predicting protein function and structure. It builds three-dimensional atomic models through iterative structural assembly simulations and various threading alignments (Kumar & Kim, 2024; Roy et al., 2010). On the other hand, SWISS-MODEL is an online bioinformatics tool that is mostly used for 3D protein structure homology modeling using template-based data (Arnold et al., 2006; Bordoli et al., 2009a). Moreover, the structures were then evaluated using the PROCHECK tool from the UCLA SAVES program (v.6.1) (Laskowski et al., 1996) and the ProSA-web server (Wiederstein & Sippl, 2007).

2.8. Active Site Prediction

Active site prediction involves computationally pinpointing the exact three-dimensional regions on a protein or enzyme where ligands like substrates or drugs bind to enable biological function (Sankararaman et al., 2010). The active site of the target protein was predicted using CASTpFold 3.0, based on the high-quality model produced by SWISS-MODEL. CASTpFold is an online tool that detects, analyzes, and measures concave features on protein 3D structures, such as surface pockets and internal cavities (Barnsley & Ondrechen, 2022).

2.9. Molecular Docking

Molecular docking is a computational method used to predict the structure of a receptor-ligand complex. Molecular docking seeks to predict the binding modes and affinities of small molecules within the binding sites of specific receptor targets. It is widely used as a standard computational technique in drug design for lead compound optimization and in virtual screening to discover new biologically active molecules (Guedes et al., 2014; Pagadala et al., 2017). Molecular docking was carried out using the AutoDock Vina wizard integrated in PyRx to evaluate the binding interactions of 58 bioactive compounds derived from garlic (*Allium sativum*) bulb within the active site of the target protein.

3. Results and Discussion

3.1. Sequence Retrieval of Hypothetical Protein

The 187 amino acid residues of the chosen protein, which has the accession number Q6WB97, was recovered from the viral species Human metapneumovirus. The amino acid sequences were retrieved and verified through UniProt, a publicly available database containing protein sequences and functional annotations. Information gathered from the UniProt database has been compiled and presented in **Table 1.**

Table 1: Properties of the chosen protein.

Protein identity	Properties of the targeted protein
Locus	M21_HMPVC
Definition	Protein M2-1
Accession	Q6WB97
Version	Q6WB97 .1
Amino Acid	187
Gene	M2-1
Organism	Human metapneumovirus (strain CAN97-83)
FASTA format	>sp|Q6WB97|M21_HMPVC Protein M2-1 OS=Human metapneumovirus (strain CAN97-83) OX=694067 GN=M2-1 PE=1 SV=1 MSRKAPCYVRGKCNRGSECKFNHNYWSWP-DRYLLIRSNYLLNQLLRNTDRADGLSIISGAG REDRTQDFVLGSTNVVQGYIDDNQSITKAAA CYSLHNIIKQLQEVEVRQARDSKLSDSKHVAL HNLILSYMEMSKTPASLINNLKRLPREKLKKL

3.2. Physicochemical Properties Determination

From ProtParam tool results, the protein's molecular weight was 21,234.09 Da, with a theoretical isoelectric point (pI) of 9.14. Its molecular formula was determined to be $C_{922}H_{1494}N_{274}O_{287}S_7$. The predicted half-life was 30 hours in mammalian reticulocytes (*in vitro*), over 20 hours (>20) in yeast (*in vivo*), and more than 10 hours (>10) in Escherichia coli (*in vivo*). Additionally, 27 positively charged residues (Arg and Lys) were counted, compared to 21 negatively charged residues (Asp and Glu). This indicates that the protein is positively charged and classified as basic (Khan et al., 2017; Requião et al., 2017). The aliphatic index (AI) was calculated to be 90.27, indicating a high proportion of aliphatic side chains, contributing to thermal stability in globular proteins (Dill et al., 1989; Jaenicke, 2000; Privalov, 1979). The selected protein instability index (II) of 46.27. The GRAVY score was computed as -0.582, suggesting that the protein is hydrophilic, implying a greater potential for water interaction (Chou & Morr, 1979). The extinction coefficient at 280 nm was calculated as 23,170 M^{-1} cm^{-1} assuming all cysteine residues form cystine bonds, and 22,920 M^{-1} cm^{-1} assuming all cysteines are in their reduced form, offering insight into the protein's absorbance characteristics. The physicochemical properties of the selected protein are shown in **Fig. 2.** and **Table 2.**

3.3. Functional Annotation

A domain is an extended, conserved sequence segment that serves as a distinct functional and structural unit within a protein (Taylor, 1999; Whisstock & Lesk, 2003), whereas a motif is a shorter, conserved sequence pattern typically associated with a specific function (Bork & Koonin, 1996; Nevill-Manning et al., 1998). Conserved domain analysis using the NCBI CD-search tool identified the Pneumovirus matrix protein 2 (M2) superfamily (accession ID: cl07866), spanning residues 29-179 with an E-value of 6.59e-74. This domain functions as a transcription processivity factor, which is essential for maintaining the efficiency of viral RNA synthesis during replication.

In addition, ScanProsite analysis revealed the presence of a C3H1-type zinc finger motif (ZnF_C3H1, residues 1-28; accession: PS50103; score 13.216). This motif plays a critical role in RNA binding and stabilization, allowing M2-1 to interact with viral RNA and regulatory elements. Its structural configuration supports the protein's involvement in viral transcriptional regulation and nucleocapsid interaction, processes that are indispensable for viral assembly and budding.

Together, the matrix domain and the zinc finger functional motif highlight M2-1 as a multifunctional regulator of HMPV pathogenesis, linking transcriptional activity with structural assembly. These conserved elements also underscore its potential as a therapeutic target, since disrupting either the transcription processivity factor or the RNA-binding zinc finger motif could significantly impair viral replication.

3.4. Gene Ontology Analysis

The Gene Ontology (GO) analysis reveals that the protein is associated with diverse roles across biological process, molecular

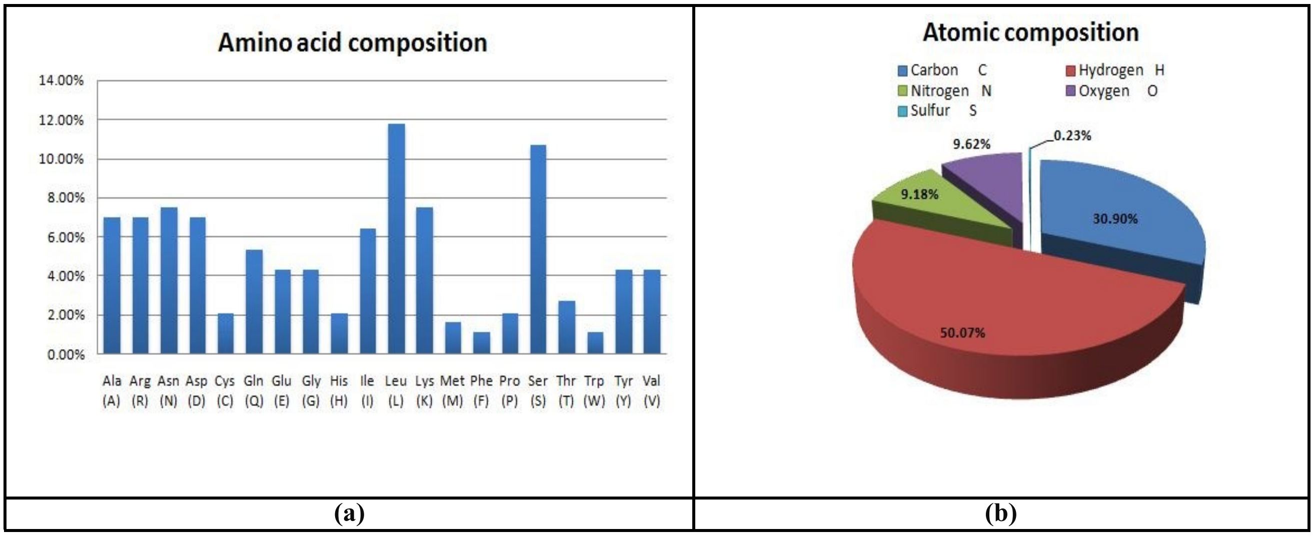


Fig. 2. (a), Analysis of the amino acid composition of the selected protein revealed that Leu (22, 11.8%) was the predominant amino acid, followed by Ser (20, 10.7%), Asn (14, 7.5%), Lys (14, 7.5%), Ala (13, 7.0%), Arg (13, 7.0%), Asp (13, 7.0%), Ile (12, 6.4%), Gln (10, 5.3%), Glu (8, 4.3%), Gly (8, 4.3%), Tyr (8, 4.3%), Val (8, 4.3%), Thr (5, 2.7%), Cys (4, 2.1%), His (4, 2.1%), Pro (4, 2.1%), Met (3, 1.6%), Phe (2, 1.1%), and Trp (2, 1.1%). **(b),** The atomic composition showed that Hydrogen was the most abundant element, with 1494 atoms (50.07%), followed by Carbon with 922 atoms (30.90%), Oxygen with 287 atoms (9.62%), Nitrogen with 274 atoms (9.18%), and Sulfur with 7 atoms (0.23%).

Table 2. Physicochemical characterization of the targeted protein.

Characteristics	**Value**
Molecular weight	21,234.09 Da
Theoretical isoelectric point (pI)	9.14
Total number of negatively charged residues (Asp + Glu)	21
Total number of positively charged residues (Arg + Lys)	27
Formula	$C_{922}H_{1494}N_{274}O_{287}S_7$
Total number of atoms	2,984
The estimated half-life	(a) 30 hours (mammalian reticulocytes, *in vitro*). (b) >20 hours (yeast, *in vivo*). (c) >10 hours (*Escherichia coli*, *in vivo*).
Instability index	46.27
Aliphatic index	90.27
GRAVY	-0.582

function, and cellular component categories (**Table 3**). The GO annotations highlight the multifunctional nature of the M2-1 protein. Its localization in the host cell cytoplasm, nucleus, and virion components indicates its dynamic role in different stages of the viral life cycle. The RNA-binding and metal ion-binding activities reflect its contribution to stabilizing RNA interactions and regulating transcriptional processes, consistent with the zinc finger motif identified earlier. Moreover, its classification under structural molecule activity suggests that M2-1 not only functions as a regulatory factor but also contributes to maintaining the architecture of viral complexes. Importantly, its role in the regulation of viral transcription positions M2-1 as a central factor in ensuring efficient genome replication and viral propagation.

Collectively, these GO terms link M2-1 directly to viral pathogenesis, emphasizing its dual role in structural stability and functional regulation. Such multi-functionality reinforces its value as a potential therapeutic target, as disrupting these biological functions could simultaneously impair viral transcription, replication, and assembly.

3.5. Protein-Protein Interaction

Protein-protein interaction (PPI) analysis revealed several experimentally supported associations involving the HMPV M2-1 protein. Notably, M2-1 interacts with the matrix (M) protein (UniProt IDs: P04545 and P0DOE7) through physical association, colocalization, and proximity assays, with a moderate confidence score (MI Score = 0.58). These interactions were observed in HeLa and HEK293T cells using confocal microscopy and co-immunoprecipitation (Co-IP) assays.

Additionally, M2-1 was found to interact with the human transcription factor RELA (NF-κB p65 subunit; UniProt Q04206), suggesting a potential role in modulating host immune signaling pathways. This interaction was supported by Co-IP and fluorescence imaging assays in A549 lung carcinoma cells with an MI Score of 0.46, indicating a biologically relevant host-virus interaction.

Furthermore, M2-1 also associated with the viral fusion (F) protein (UniProt P03420) as demonstrated by colocalization and proximity assays with a moderate MI Score of 0.38. Together, these findings highlight the central role of M2-1 in viral assembly, replication, and host interaction, reflecting its multifunctional nature in the HMPV life cycle **Fig. 3**.

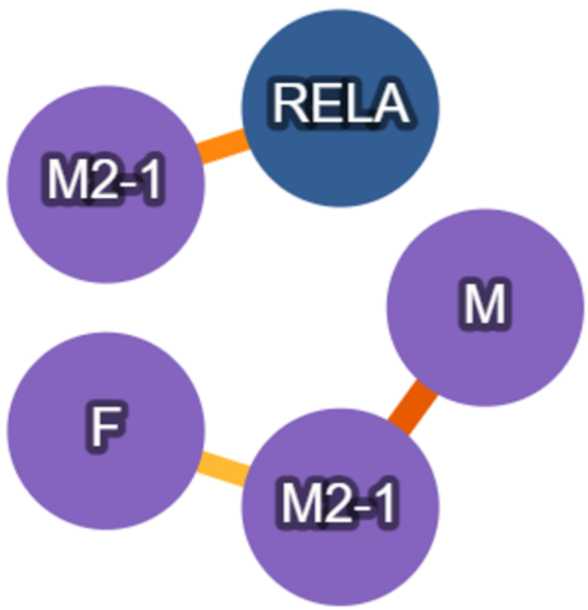


Fig. 3. The PPI network analysis revealed that the target protein interacts with F, M, and RELA proteins.

Table 3. Gene ontology analysis.

Accession	Name	Ontology
GO:0030430	Host cell cytoplasm	Cellular component
GO:0042025	Host cell nucleus	Cellular component
GO:0044423	Virion component	Cellular component
GO:0046872	Metal ion binding	Molecular function
GO:0003723	RNA binding	Molecular function
GO:0005198	Structural molecule activity	Molecular function
GO:0046782	Regulation of viral transcription	Biological process

3.6. Secondary Structure Determination

Understanding protein secondary structure comprising alpha helices, beta sheets, and beta turns is key to studying protein function. Predicting these structures aims to accurately identify important features using statistical methods and to clarify protein folding mechanisms. This process improves analysis by considering factors like residue patterns, hydrophobicity, sequence variations, and conservation, enhancing prediction accuracy through weighted evaluation and incremental information (Brandt, 2015; King & Sternberg, 1996). The SOPMA server analysis identified various secondary structural elements within the protein, including alpha helices, extended strands, and random coils. Specifically, alpha helices were the most prevalent, comprising 92 residues (49.20%) of the protein. Extended strands accounted for 25 residues (13.37%), while random coils made up 70 residues (37.43%) (**Fig. 4.**). These structural components provide important insights into the protein's overall conformation and potential functional regions.

3.7. Homology Modelling and Quality Assessment

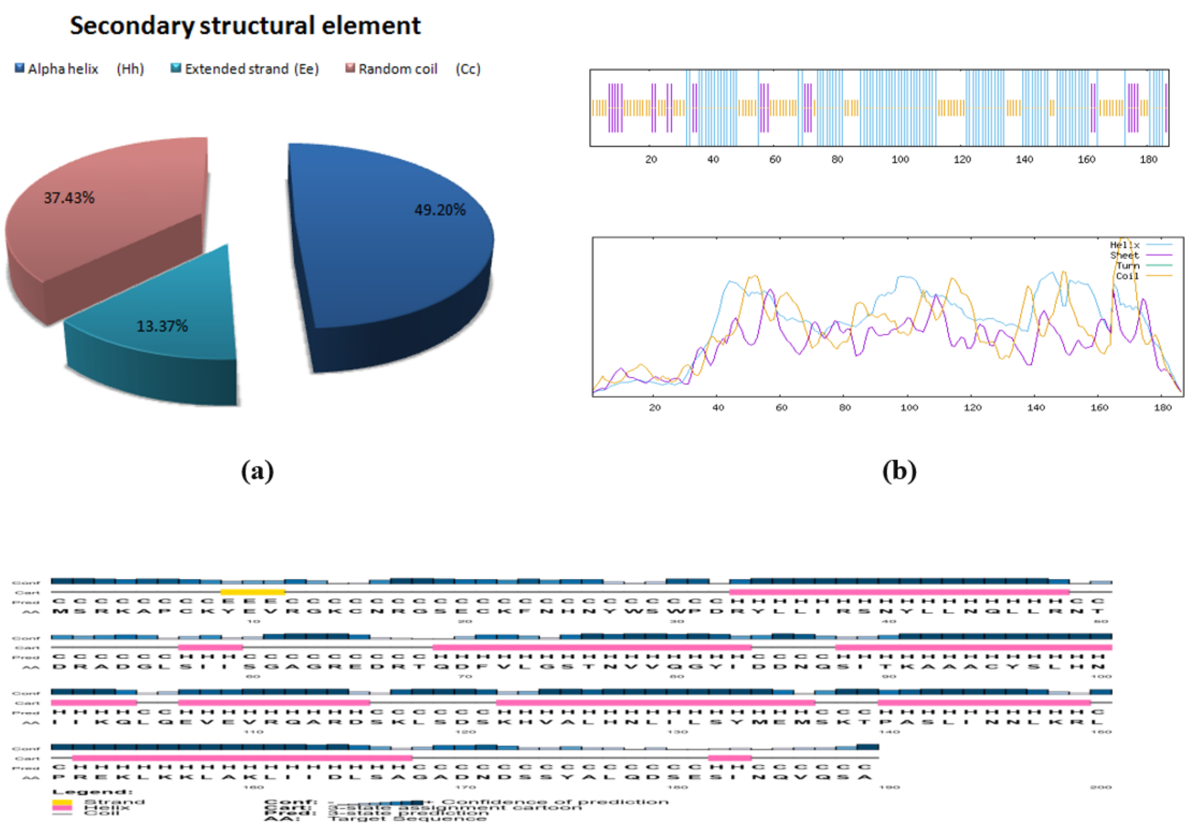
Homology modeling estimates the 3D structure of a target protein by aligning its sequence with that of a known template protein (Bordoli et al., 2009b). This process generally involves four key steps: identifying the target, aligning sequences, building the model, and refining it. Protein structure is crucial for regulating various biological functions, as its 3D conformation provides valuable insights into functionality and molecular interactions. Understanding protein structures aids in designing targeted experiments, such as studying disease-related mutations, developing specific inhibitors, or conducting site-directed mutagenesis (Vijayanirmala et al., 2024; Waterhouse et al., 2018).

In the SWISS-MODEL analysis, a total of 48 template structures were identified, among which the best-fitting template showed a sequence identity of 97.86% and a Global Model Quality Estimation (GMQE) score of 0.89 with the target HMPV M2-1 protein. The predicted model demonstrated a Z-score of -5.4, indicating high structural reliability and excellent model quality. Similarly, the I-TASSER server identified four potential template models, with the first model selected as the best based on its confidence score, RMSD value, and resolution. The resulting I-TASSER model also exhibited a Z-score of -5.4, consistent with the SWISS-MODEL prediction, thereby confirming the robustness and consistency of the homology modeling results across both platforms (**Fig. 5.**).

According to the ramachandran plot study, the favored areas (A, B, and L) of the plot contained 96.1% of the residues in the SWISS-MODEL projected structure. On the other hand, only 75.1% of residues were discovered in the I-TASSER model's favored regions. Apart from this, in SWISS-MODEL, the percentages of residues in additional allowed regions, generously allowed regions, and disallowed regions were 3.3%, 0%, and 0.7%, respectively. In the I-TASSER model, these percentages were 16.2%, 4.0%, and 4.6%, respectively (**Table 4**). This analysis demonstrates that the SWISS-MODEL structure has significantly better stereo-chemical quality than the I-TASSER model, making it more suitable for accurate 3D structure prediction of the target protein.

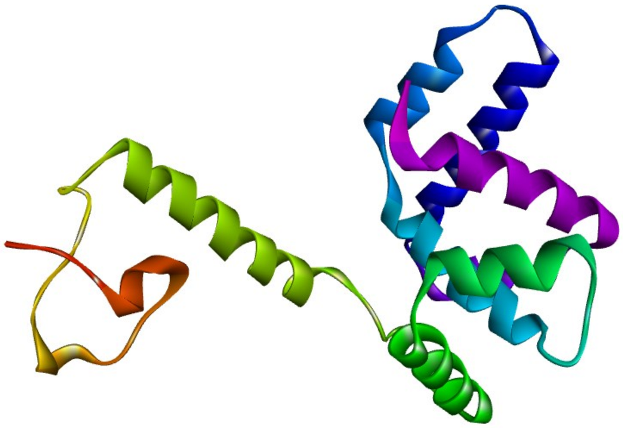
3.8. Active Site Prediction

Active site prediction involves identifying pockets on a protein's surface that are accessible to small molecules but not to larger ones, suggesting potential ligand-binding regions. These pockets include both functionally active residues and others that may not directly participate in ligand binding (Cammisa et al., 2013; Volkamer et al.,

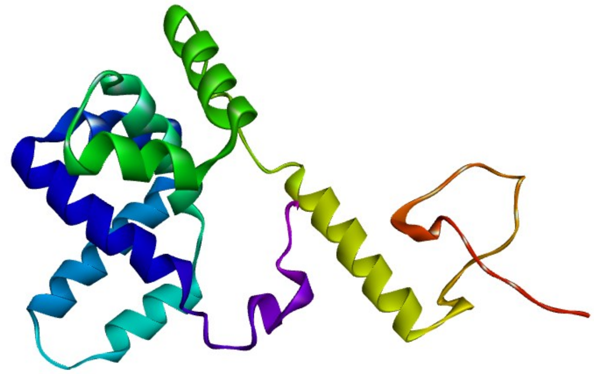


(c)

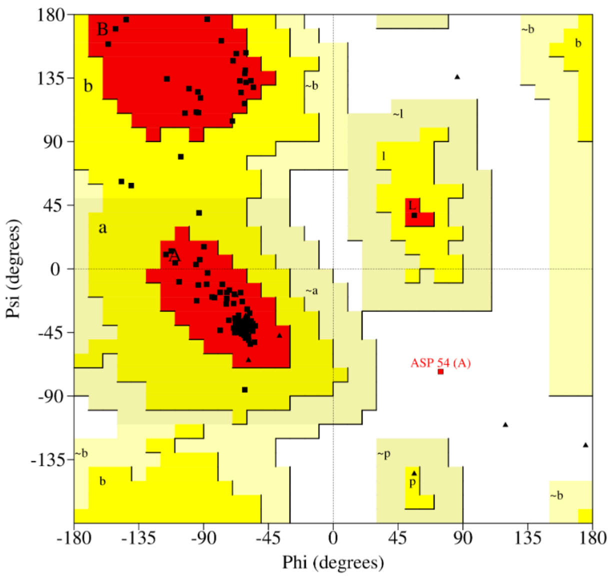
Fig. 4. (a), The SOPMA server was used to predict the 2D structural elements of the selected protein. **(b),** The schematic illustration of secondary structures, including helices, sheets, turns, and coils in the SOPMA model, is based on the analyzed protein sequence. **(c),** Additionally, the PSIPRED server was employed to predict the secondary structure of the specified protein.



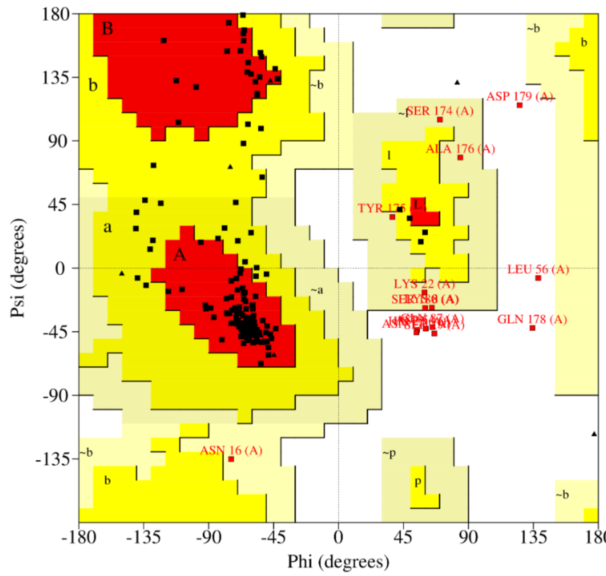
(a)



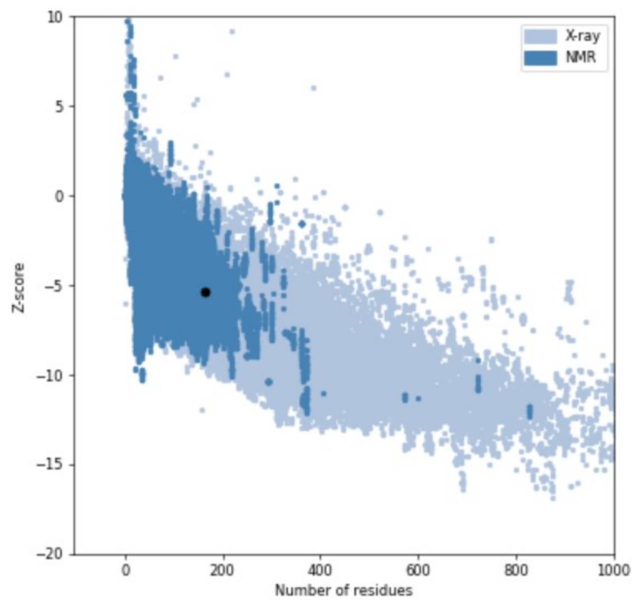
(b)


Psi (degrees)
Phi (degrees)
ASP 54 (A)

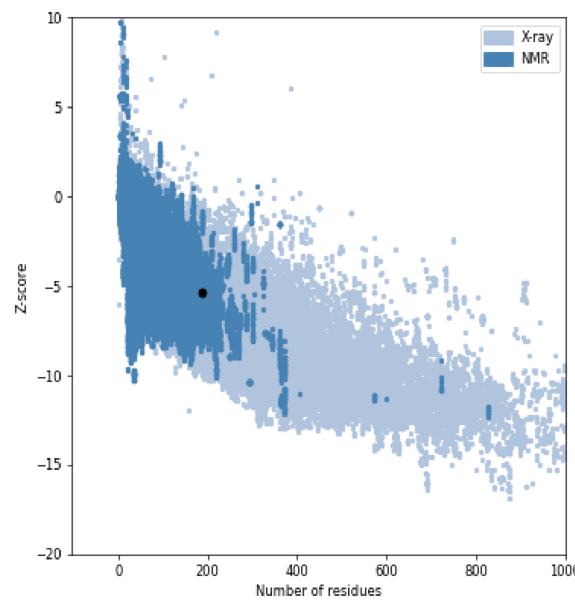
(c)


Psi (degrees)
Phi (degrees)
ASP 179 (A)
SER 174 (A)
ALA 176 (A)
LEU 56 (A)
GLN 178 (A)
ASN 16 (A)

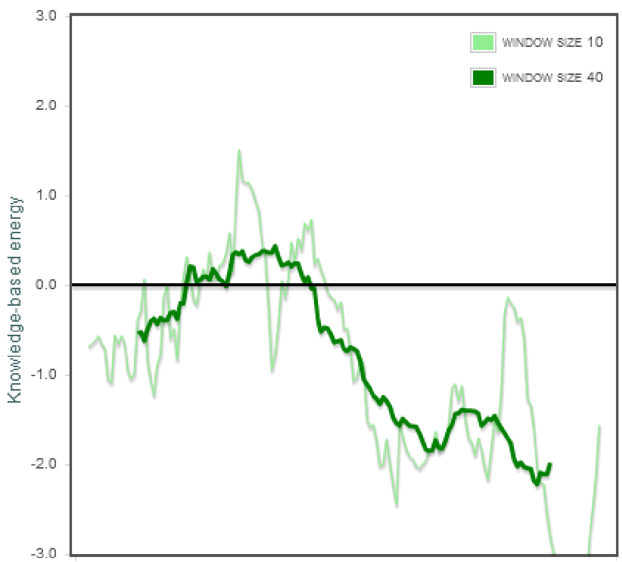
(d)


X-ray
NMR
Z-score
Number of residues

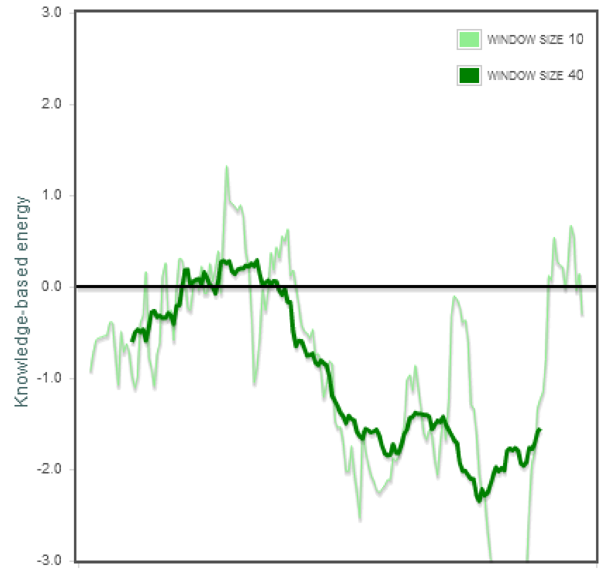
(e)


X-ray
NMR
Z-score
Number of residues

(f)



(g)



(h)

Fig. 5. The predicted three-dimensional structure of the identified protein was generated using **(a)** SWISS-MODEL and **(b)** I-TASSER. The Ramachandran plots for the selected protein were also produced by **(c)** SWISS-MODEL and **(d)** I-TASSER to assess the stereochemical quality of the models. Additionally, the Z-scores for the 3D models, obtained through the ProSA-web server, were reported for both **(e),** SWISS-MODEL (-5.4) and **(f),** I-TASSER (-5.4), providing an overall evaluation of model accuracy. Furthermore, local quality assessments of the predicted structures were performed for both **(g),** SWISS-MODEL and **(h),** I-TASSER to identify regions of varying confidence within the models.

Table 4. Plot statistics of SWISS-MODEL and I-TASSER.

Characteristics	SWISS-MODEL	I-TASSER
Residues in most favored regions [A,B,L]	146 (96.1%)	130 (75.1%)
Residues in additional allowed regions [a,b,l,p]	5 (3.3%)	28 (16.2%)
Residues in generously allowed regions [-a,-b,-l,-p]	0 (0%)	7 (4.0%)
Residues in disallowed regions	1 (0.7%)	8 (4.6%)
Number of non-glycine and non-proline residues	152	173
Number of end-residues (excl. Gly and Pro)	2	2
Number of glycine residues	8	8
Number of proline residues	4	4

2010). The active site prediction was carried out using the SWISS-MODEL-derived structure, as it exhibited superior stereo-chemical quality and accuracy compared to the I-TASSER model. This high-quality 3D model enabled reliable identification of potential binding pockets and catalytically important residues within the target protein.

Putative active sites of the protein were identified using the CASTpFold server, which utilizes a probe radius of 1.4 Å to calculate solvent-accessible surface areas. The analysis evaluated the mouth openings, surface areas, and volumes of the pockets, detecting a total of 22 active pockets. The highest-ranked active site was chosen based on its surface area and volume, measuring 282.422 Å^2 and 279.006 Å^3, respectively, indicating the most likely functional site (**Table 5**, **Fig. 6**, and **Supplementary file 1**).

3.9. Molecular Docking

Molecular docking was carried out using the AutoDock Vina wizard integrated in PyRx to evaluate the interaction of 58 bioactive compounds derived from garlic (*Allium sativum*) bulb with the active site of the target protein. The docking grid was precisely defined to encompass the active pocket, with the following coordinates: center_x = 11.7228, center_y = 41.9787, center_z = 71.9775, and grid box dimensions of size_x = 23.8131, size_y = 36.4284, and size_z = 26.7374. The exhaustiveness value was set to 8 to ensure adequate sampling of ligand conformations within the binding site.

Among all the docked compounds, IMPHY010911 exhibited the highest binding affinity with a docking score of -7.7 kcal/mol,

Table 5. Top 3 pockets of targeted protein and its area and volume.

Pocket ID	Area (SA) (Å^2)	Volume (SA) (Å^3)
1	282.422	279.006
2	50.895	72.844
3	80.915	48.319

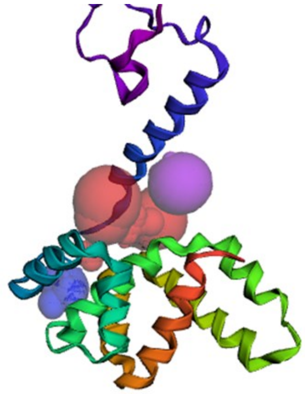


Fig. 6. The active site predicted by CASTpFold. The negative volume of pocket-1, pocket-2, pocket-3 indicated respectively by red, paste, and blue color. The shape of negative volume is cartoon.

indicating a strong and stable interaction with the target protein's active site residues and also highlights IMPHY010911 as a potential lead molecule for further optimisation and validation. The resulting binding affinities for all compounds are illustrated in **Fig. 7**. and **Supplementary file 2**. This figure was generated by using Python programming.

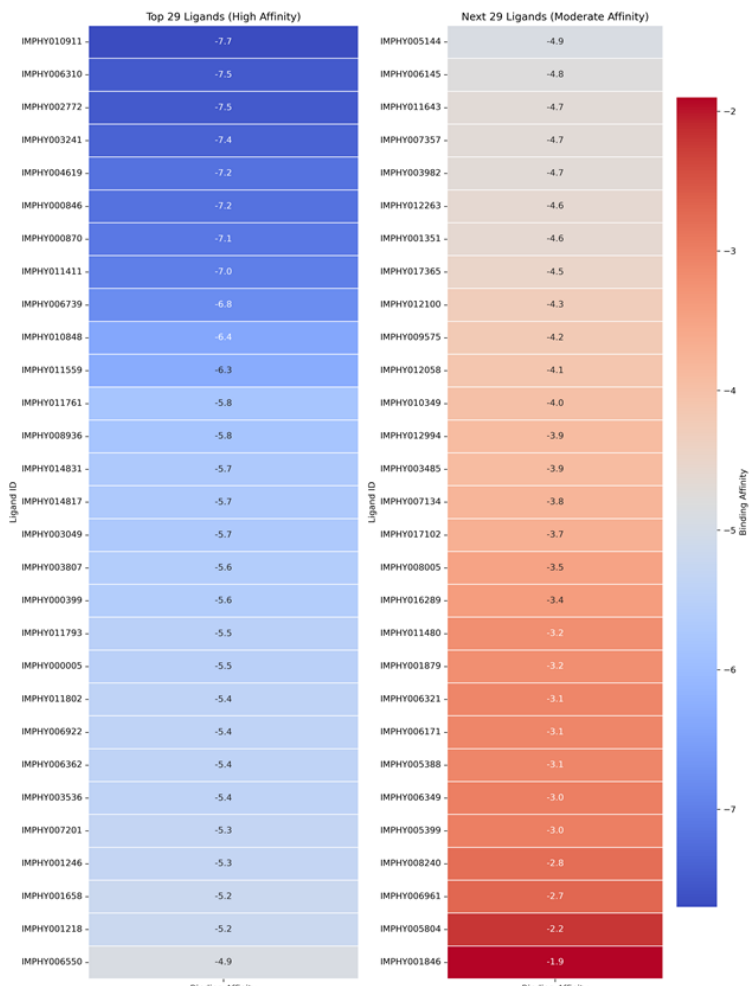


Fig. 7. Binding affinity of 58 bioactive compounds derived from garlic (Allium sativum) bulb with the active site of the targeted protein as determined by molecular docking using PyRx AutoDock Vina.

4. Conclusions with future perspective

This study reveals the structural and functional significance of the HMPV M2-1 protein as a key regulator of viral replication, transcription, and assembly. Computational analyses confirmed its stability, conserved RNA-binding motifs, and central role in the viral replication complex. Structural modeling and molecular docking identified potential active sites and promising natural inhibitors, emphasizing its therapeutic potential. Further in-silico optimization and in-vitro validation are required to confirm the efficacy of these potential inhibitors and to explore M2-1 targeted antiviral strategies for effective HMPV control.

List of Abbreviations

Abbreviation	Full Form
CDD	Conserved Domain Database.
2D	Two Dimension.
3D	Three Dimension.
NPSA	Network Protein Sequence Analysis.

Authors' contributions

Conceptualization, SS and SD; **methodology**, SS, SD and MHA; **investigation**, SS and SD; **software**, MHA; **validation**, SS and SD; **formal analysis**, SS, SD and MHA; **data curation**, SS and SD; **resources**, MHA; **writing—original draft preparation**, SS, SD and MHA; **writing—review and editing**, MHA; **visualization**, SS and SD; **supervision**, MHA. All authors have read and agreed to the published version of the manuscript.

Data availability statement

Data is contained within the article.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical statement

This study did not involve the use of human participants, animals, or clinical trials. All analyses were performed using publicly available computational tools and databases; therefore, no ethical approval was required.

Clinical trial number: Not applicable.

References

Acharya, A., & Byrareddy, S. N. (2025). Immunological insights into the re-emergence of human metapneumovirus. Current Opinion in Immunology, 94, 102562. https://doi.org/https://doi.org/10.1016/j.coi.2025.102562

Adedokun, K. A., Adekola, S. A., Tajudeen, A., Bello-Ibiyemi, A. A., Babandina, M. M., Magwe, E. A., & Bello, A. (2025). Rising global threat of human metapneumovirus (hMPV in 2024/2025): pathogenesis, immune dynamics, vulnerabilities in immunocompromised individuals, and lessons from past pandemics. Journal of Rare Diseases, 4(1), 16. https://doi.org/10.1007/s44162-025-00079-w

Alotaibi, B. S., Habib, M., Habib, A., Syed, F., Shah, N. N., & Tantry, B. A. (2025). Human Metapneumovirus: A Review of Epidemiology, Pathogenesis, Diagnosis, and Future Prospects. Journal of Pure and Applied Microbiology, 19(2), 755–765. https://doi.org/10.22207/JPAM.19.2.07

Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics (Oxford, England), 22(2), 195–201. https://doi.org/10.1093/bioinformatics/bti770

Athanasios, A., Charalampos, V., Vasileios, T., & Md. Ashraf, G. (2017). Protein-protein interaction (PPI) network: recent advances in drug discovery. Current Drug Metabolism, 18(1), 5–10.

Bao, X., Liu, T., Spetch, L., Kolli, D., Garofalo, R. P., & Casola, A. (2007). Airway epithelial cell response to human metapneumovirus infection. Virology, 368(1), 91–101. https://doi.org/10.1016/j.virol.2007.06.023

Barnsley, K. K., & Ondrechen, M. J. (2022). Enzyme active sites: Identification and prediction of function using computational chemistry. Current Opinion in Structural Biology, 74, 102384. https://doi.org/10.1016/j.sbi.2022.102384

Baudouin-Cornu, P., Surdin-Kerjan, Y., Marliere, P., & Thomas, D. (2001). Molecular evolution of protein atomic composition. Science, 293 (5528), 297–300.

Bhattacharya, S., Bhattacharjee, S., & Singh, A. (2025). Human Metapneumovirus: A Comprehensive Epidemiological Analysis of a Global Respiratory Threat. Infect Chemother, 57(2), 194–202.

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., & Schwede, T. (2009a). Protein structure homology modeling using SWISS-MODEL workspace. Nature Protocols, 4(1), 1–13.

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., & Schwede, T. (2009b). Protein structure homology modeling using SWISS-MODEL workspace. Nature Protocols, 4(1), 1–13. https://doi.org/10.1038/nprot.2008.197

Bork, P., & Koonin, E. V. (1996). Protein sequence motifs. Current Opinion in Structural Biology, 6(3), 366–376.

Brandt, G. S. (2015). Secondary Structure. Molecular Life Sciences, 1–9. https://doi.org/10.1007/978-1-4614-6436-5_8-1

Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench : 20 years on. 47(April), 402–407. https://doi.org/10.1093/nar/gkz297

Cammisa, M., Correra, A., Andreotti, G., & Cubellis, M. V. (2013). Identification and analysis of conserved pockets on protein surfaces. BMC Bioinformatics, 14 Suppl 7(Suppl 7), S9. https://doi.org/10.1186/1471-2105-14-S7-S9

Chou, D. H., & Morr, C. V. (1979). Protein-water interactions and functional properties. Journal of the American Oil Chemists' Society, 56(1), A53–A62.

Consortium, G. O. (2004). The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research, 32(suppl_1), D258–D261.

Cuche, A., Sigrist, C. J. A., Castro, E. De, Cerutti, L., Hulo, N., Bridge, A., Bougueleret, L., & Xenarios, I. (2013). New and continuing developments at PROSITE. 41(November 2012), 344–347. https://doi.org/10.1093/nar/gks1067

Das, S., & Dunbar, S. A. (2024). Chapter 119 - Respiratory syncytial virus and metapneumovirus (Y.-W. Tang, M. Y. Hindiyeh, D. Liu, A. Sails, P. Spearman, & J.-R. B. T.-M. M. M. (Third E. Zhang (Eds.); pp. 2429–2449). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-818619-0.00138-6

Demmler-Harrison, G. J. (2018). Healthcare-associated viral infections: considerations for nosocomial transmission and infection control. In Healthcare-Associated Infections in Children: A Guide to Prevention and Management (pp. 229–257). Springer.

Dill, K. A., Alonso, D. O. V, & Hutchinson, K. (1989). Thermal stabilities of globular proteins. Biochemistry, 28(13), 5439–5449.

Dolan, M. E., Ni, L., Camon, E., & Blake, J. A. (2005). A procedure for assessing GO annotation consistency. Bioinformatics, 21(suppl_1), i136–i143.

Dong, Y., Xie, Z., & Xu, L. (2025). Receptors and host factors: key players in human metapneumovirus infection. Frontiers in Cellular and Infection Microbiology, 15, 1557880.

Downard, K. M., & Cody, R. B. (2024). Amino Acid Composition Determination From the Fractional Mass of Peptides. Journal of Mass Spectrometry : JMS, 59(10), e5089. https://doi.org/10.1002/jms.5089

Du Plessis, L., Škunca, N., & Dessimoz, C. (2011). The what, where, how and why of gene ontology—a primer for bioinformaticians. Briefings in Bioinformatics, 12(6), 723–735.

Fischer, H., Polikarpov, I., & Craievich, A. F. (2004). Average protein density is a molecular-weight-dependent function. Protein Science, 13 (10), 2825–2828.

Foley, D. A., Smith, D. W., Barr, I., Moore, H. C., Nicol, M., & Blyth, C. C. (2025). Determinants of respiratory syncytial virus and human metapneumovirus transmission. Clinical Microbiology Reviews, e00203-24.

Geourjon, C., & Deleage, G. (1995). SOPMA : significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. 11(6), 681–684.

Gill, S. C., & Von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Analytical Biochemistry, 182(2), 319–326.

Guedes, I. A., de Magalhães, C. S., & Dardenne, L. E. (2014). Receptor–ligand molecular docking. Biophysical Reviews, 6(1), 75–87. https://doi.org/10.1007/s12551-013-0130-2

Guruprasad, K., Reddy, B. V. B., & Pandit, M. W. (1990). Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Engineering, Design and Selection, 4 (2), 155–161. https://doi.org/10.1093/protein/4.2.155

Hamelin, M.-E., Abed, Y., & Boivin, G. (2004). Human metapneumovirus: a new player among respiratory viruses. Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America, 38(7), 983–990. https://doi.org/10.1086/382536

Hejran, A. B., Sangin, M. Y., Niazay, F., Saken, K., Sawicka, B., Atti, N. Z., & Niazi, P. (n.d.). Human metapneumovirus: An emerging threat among respiratory pathogens with rising cases in infants and young children.

Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. Journal of Biochemistry, 88(6), 1895–1898.

Jaenicke, R. (2000). Stability and stabilization of globular proteins in solution. Journal of Biotechnology, 79(3), 193–203.

Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., & Hinz, U. (2012). The IntAct molecular interaction database in 2012. Nucleic Acids Research, 40(D1), D841–D846.

Khan, R. H., Siddiqi, M. K., & Salahuddin, P. (2017). Protein structure and function. Basic Biochemistry, 5, 1–39.

King, R. D., & Sternberg, M. J. E. (1996). Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. Protein Science, 5(11), 2298–2310. https://doi.org/10.1002/pro.5560051116

Krüger, N., Laufer, S. A., & Pillaiyar, T. (2025). An overview of progress in human metapneumovirus (hMPV) research: Structure, function, and therapeutic opportunities. Drug Discovery Today, 30(5), 104364. https://doi.org/https://doi.org/10.1016/j.drudis.2025.104364

Kumar, H., & Kim, P. (2024). Artificial intelligence in fusion protein three-dimensional structure prediction: Review and perspective. Clinical and Translational Medicine, 14(8), e1789.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology, 157(1), 105–132. https://doi.org/10.1016/0022-2836(82)90515-0

Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. Journal of Biomolecular NMR, 8(4), 477–486. https://doi.org/10.1007/BF00228148

Leyrat, C., Paesen, G. C., Charleston, J., Renner, M., & Grimes, J. M. (2014). Structural insights into the human metapneumovirus glycoprotein ectodomain. Journal of Virology, 88(19), 11611–11616. https://doi.org/10.1128/JVI.01726-14

Lu, H., Zhou, Q., He, J., Jiang, Z., Peng, C., Tong, R., & Shi, J. (2020). Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. Signal Transduction and Targeted Therapy, 5(1), 213.

Mohammadi, K., Faramarzi, S., Yaribash, S., Valizadeh, Z., Rajabi, E., Ghavam, M., Samiee, R., Karim, B., Salehi, M., Seifi, A., & Shafaati, M. (2025). Human metapneumovirus (hMPV) in 2025: emerging trends and insights from community and hospital-based respiratory panel analyses-a comprehensive review. Virology Journal, 22(1), 150. https://doi.org/10.1186/s12985-025-02782-y

Monné, M., Nilsson, I., Johansson, M., Elmhed, N., & von Heijne, G. (1998). Positively and negatively charged residues have different effects on the position in the membrane of a model transmembrane helix. Journal of Molecular Biology, 284(4), 1177–1183.

Morgnanesi, D., Heinrichs, E. J., Mele, A. R., Wilkinson, S., Zhou, S., & Kulp, J. L. (2015). A computational chemistry perspective on the current status and future direction of hepatitis B antiviral drug discovery. Antiviral Research, 123, 204–215. https://doi.org/https://doi.org/10.1016/j.antiviral.2015.10.014

Mozaffari, F., Fakhari, S., Elikaei, A., Ahmadi, Z., Farahmand, B., & Maleki, A. (2025). Divergent origins, divergent pandemic potential: A comparative analysis of HMPV and SARS-CoV-2. Journal of Clinical Virology Plus, 5(3), 100223. https://doi.org/https://doi.org/10.1016/j.jcvp.2025.100223

Nevill-Manning, C. G., Wu, T. D., & Brutlag, D. L. (1998). Highly specific protein sequence motifs for genome analysis. Proceedings of the National Academy of Sciences, 95(11), 5865–5871.

Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking: a review. Biophysical Reviews, 9(2), 91–102. https://doi.org/10.1007/s12551-016-0247-1

Pandey, V., Shahi, P., Kolios, G., Uddin, M. I., Spathakis, M., Collins, A. R., Paspaliaris, V., & Rai, A. K. (2025). Human metapneumovirus (hMPV): the virus who came with the common cold. Infection. https://doi.org/10.1007/s15010-025-02626-5

Papenburg, J., & Boivin, G. (2010). The distinguishing features of human metapneumovirus and respiratory syncytial virus. Reviews in Medical Virology, 20(4), 245–260.

Primmer, C. R., Papakostas, S., Leder, E. H., Davis, M. J., & Ragan, M. A. (2013). Annotated genes and nonannotated genomes: cross-species use of Gene Ontology in ecology and evolution research. Molecular Ecology, 22(12), 3216–3241.

Privalov, P. L. (1979). Stability of proteins small globular proteins. Advances in Protein Chemistry, 33, 167–241.

Rahman, M., & Sadygov, R. G. (2017). Predicting the protein half-life in tissue from its cellular properties. PloS One, 12(7), e0180428. https://doi.org/10.1371/journal.pone.0180428

Requião, R. D., Fernandes, L., de Souza, H. J. A., Rossetto, S., Domitrovic, T., & Palhano, F. L. (2017). Protein charge distribution in proteomes and its impact on translation. PLoS Computational Biology, 13(5), e1005549.

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols, 5(4), 725–738.

Safari-Alighiarloo, N., Taghizadeh, M., Rezaei-Tavirani, M., Goliaei, B., & Peyvandi, A. A. (2014). Protein-protein interaction networks (PPI) and complex diseases. Gastroenterology and Hepatology from Bed to Bench, 7(1), 17.

Sankararaman, S., Sha, F., Kirsch, J. F., Jordan, M. I., & Sjölander, K. (2010). Active site prediction using evolutionary and structural information. Bioinformatics (Oxford, England), 26(5), 617–624. https://doi.org/10.1093/bioinformatics/btq008

Satapathy, T., Satapathy, A., Satapathy, A., Yadav, N., Chandrakar, M., & Chandrakar, K. (2025). Recent outbreaks of Human Metapneumovirus (HMPV): Prevention, Diagnosis and Therapeutic insights. Journal of Drug Delivery & Therapeutics, 15 (2).

Sillero, A., & Ribeiro, J. M. (1989). Isoelectric points of proteins: theoretical determination. Analytical Biochemistry, 179(2), 319–325.

Soto, J. A., Gálvez, N. M. S., Benavente, F. M., Pizarro-Ortega, M. S., Lay, M. K., Riedel, C., Bueno, S. M., Gonzalez, P. A., & Kalergis, A. M. (2018). Human Metapneumovirus: Mechanisms and Molecular Targets Used by the Virus to Avoid the Immune System. Frontiers in Immunology, 9, 2466. https://doi.org/10.3389/fimmu.2018.02466

Sun, P. D., Foster, C. E., & Boyington, J. C. (2004). Overview of protein structural and functional folds. Current Protocols in Protein Science, 35(1), 11–17.

Taylor, W. R. (1999). Protein structural domain identification. Protein Engineering, 12(3), 203–216.

Tollefson, S. J., Cox, R. G., & Williams, J. V. (2010). Studies of culture conditions and environmental stability of human metapneumovirus. Virus Research, 151(1), 54–59. https://doi.org/10.1016/j.virusres.2010.03.018

Uddin, S., & Thomas, M. (2025). Human Metapneumovirus.

Ullah, F., Ullah, S., Amin, M., Ullah, W., Ullah, S., & Wang, S. (2025). An in-silico design of a multi-epitope vaccine candidate against Human metapneumovirus (HMPV) through prediction of B- and T-cell epitopes and molecular dynamics simlation. Human Immunology, 86(5), 111557. https://doi.org/https://doi.org/10.1016/j.humimm.2025.111557

UniProt: the Universal Protein Knowledgebase in 2023. (2023). Nucleic Acids Research, 51(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052

Van Roey, K., Uyar, B., Weatheritt, R. J., Dinkel, H., Seiler, M., Budd, A., Gibson, T. J., & Davey, N. E. (2014). Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. Chemical Reviews, 114(13), 6733–6778.

Vijayanirmala, M. B., Padukone, S. S., Sinchana, B. L., Swetha, V., & Swathi, V. (2024). 3D PROTEIN STRUCTURE PREDICTION SYSTEM. 11(5), 264–271.

Volkamer, A., Griewel, A., Grombacher, T., & Rarey, M. (2010). Analyzing the topology of active sites: on the prediction of pockets and subpockets. Journal of Chemical Information and Modeling, 50(11), 2041–2052. https://doi.org/10.1021/ci100241y

Wang, J., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, S., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-bauer, A. (2023). The conserved domain database in 2023. 51(December 2022), 384 –388.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., Beer, T. A. P. De, Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL : homology modelling of protein structures and complexes. 46(May), 296–303. https://doi.org/10.1093/nar/gky427

Whisstock, J. C., & Lesk, A. M. (2003). Prediction of protein function from protein sequence and structure. Quarterly Reviews of Biophysics, 36(3), 307–340.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research, 35(Web Server issue), W407-10. https://doi.org/10.1093/nar/gkm290

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. Methods in Molecular Biology (Clifton, N.J.), 112, 531–552. https://doi.org/10.1385/1-59259-584-7:531

Williams, J. V, Piedra, P. A., & Englund, J. A. (2009). Respiratory syncytial virus, human metapneumovirus, and parainfluenza viruses. Clinical Virology, 817–847.