

Research Article

Sequence Analysis and Structure Prediction of SARS-CoV-2 Accessory Proteins 9b and ORF14: Evolutionary Analysis Indicates Close Relatedness to Bat Coronavirus

Chittaranjan Baruah ¹, Papari Devi,² and Dhirendra K. Sharma³

¹Bioinformatics Laboratory (DBT-Star College), P.G. Department of Zoology, Darrang College, Tezpur, 784 001 Assam, India

²TCRP Foundation, 781006, Guwahati, India

³School of Biological Sciences, University of Science and Technology, Meghalaya, Baridua-793101, India

Correspondence should be addressed to Chittaranjan Baruah; chittaranjan_21@yahoo.co.in

Received 8 June 2020; Revised 12 September 2020; Accepted 30 September 2020; Published 21 October 2020

Academic Editor: Gyrgy Schneider

Copyright © 2020 Chittaranjan Baruah et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has a single-stranded RNA genome that encodes 14 open reading frames (ORFs), eight of which encode accessory proteins that allow the virus to infect the host and promote virulence. The genome expresses around 29 structural and nonstructural protein products. The accessory proteins of SARS-CoV-2 are not essential for virus replication but do affect viral release, stability, and pathogenesis and finally contribute to virulence. This paper has attempted the structure prediction and functional analysis of two such accessory proteins, 9b and ORF14, in the absence of experimental structures. Sequence analysis, structure prediction, functional characterization, and evolutionary analysis based on the UniProtKB reviewed the amino acid sequences of SARS-CoV-2 9b (P0DTD2) and ORF14 (P0DTD3) proteins. Modeling has been presented with the introduction of hybrid comparative and *ab initio* modeling. QMEANDisCo 4.0.0 and ProQ3 for global and local (per residue) quality estimates verified the structures as high quality, which may be attributed to structure-based drug design targets. Tunnel analysis revealed the presence of 1-2 highly active tunneling sites, perhaps which will be able to provide certain inputs for advanced structure-based drug design or to formulate potential vaccines in the absence of a complete experimental structure. The evolutionary analysis of both proteins of human SARS-CoV-2 indicates close relatedness to the bat coronavirus. The whole-genome phylogeny indicates that only the new bat coronavirus followed by pangolin coronaviruses has a close evolutionary relationship with the novel SARS-CoV-2.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a positive-sense, single-stranded RNA virus with a genome size of 29,903 nucleotides in length. The 5' terminus of the SARS-CoV-2 genome encodes a polyprotein (pp1ab), which is further cleaved into 15 nonstructural proteins (nsp-1 to nsp-10 and nsp-12 to nsp-16), whereas the 3' terminus encodes four structural proteins (spike, envelope, membrane, and nucleocapsid) and eight accessory proteins (3a, 3b, p6, 7a, 7b, 8b, 9b, and ORF14) [1, 2]. The virus is the causative agent of coronavirus disease 2019 (COVID-

19) and is contagious through human-to-human transmission. Previously identified human CoVs that cause human disease include alphaCoVs hCoV-NL63 and hCoV-229E and the betaCoVs HCoV-OC43, HKU1, severe acute respiratory syndrome CoV (SARS-CoV), and Middle East respiratory syndrome CoV (MERS-CoV) [3]. Among the seven strains coronaviruses (CoVs) discovered so far, three strains proved to be highly pathogenic (SARS-CoV, MERS-CoV, and 2019-nCoV), which caused endemic to severe CoV disease [4, 5]. The viruses can be classified into four genera: alpha, beta, gamma, and deltaCoVs [6]. The SARS-CoV and MERS-CoV infections can result in life-threatening

diseases and have pandemic potential. SARS-CoV-2 is responsible for infection with special reference to the involvement of both the lower and upper respiratory tract [5, 7]. Furthermore, the potential for close contact between bats, civets, and humans in the wildlife trade in southern China, coupled with a possible propensity of these bats to foster CoV host-shifts, could explain SARS-like CoVs as the source of SARS-CoV [8].

To accommodate the wide spectrum of clinical presentations and outcomes of infections caused by SARS-CoV-2 [9], the WHO recently introduced the name COVID-19 (World Health Organization, 2020) to denote this disease. The acronym COVID-19 stands for “CO - corona,” “VI - viruses,” “D - disease,” and “19 - the year 2019” [10]. Despite the fact that COVID-19 has a death rate of 3.27% as of September, 27,236,916 confirmed cases with 891,031 confirmed deaths in a few months (December 8, 2019, to September 08, 2020) across 216 countries or territories are terrifying. Indeed, this virus is highly contagious, and the number of infected people can be doubled in less than seven days with a basic reproductive number (R_0) of 2.2–2.7 [11]. In humans, SARS-CoV and SARS-CoV-2 are rapidly spread by respiratory droplets, airborne routes, or direct contact [12].

The viral genome encodes 29 proteins (*Nature* doi:10.1038/s41433-020-0790-7). The functions of a large number of SARS-CoV-2 ORFs are poorly understood or unknown. The accessory proteins are unique to SARS-CoV, as they have little homology in amino acid sequences with accessory proteins of other coronaviruses [13]. An accessory ORF14 was first described in SARS-CoV by Marra et al. [14, 15]. Understanding the complete proteome of SARS-CoV-2, including the accessory proteins, is the need of the hour for the final destination of drug/medicine. Although the complete genome of SARS-CoV-2 has been made available in the public domain databases, it has been observed in our previous study on SARS-CoV-2 proteome analysis [2] that the two “accessory” ORFs ORF13 (9b) and ORF14 are poorly studied in SARS-CoV-2, as both are not annotated in most of the completed genome sequences [2]. Given the similarity of SARS-CoV-2 to bat SARS-CoV-like coronaviruses, it is likely that bats serve as reservoir hosts for its progenitor. The SARS-CoV-2 spike protein optimized for binding to human-like ACE2 is the result of natural selection [16]. Therefore, the present study reports the *in silico* sequence analysis, structure prediction, and evolutionary analysis of two such accessory proteins, 9b and ORF14, of the newly emerged SARS-CoV-2.

2. Materials and Methods

2.1. Acquisition and Analysis of Sequences. UniProtKB reviewed the amino acid sequences of SARS-CoV-2 9b (accession no. P0DTD2), and ORF14 protein (accession no. P0DTD3) was used in the present study. A conceptual framework of the workflow in the current study is represented in Figure 1. The amino acid sequences for different taxa were downloaded from UniProtKB for phylogenetic analysis based on BLASTp [17] and FASTA hits [18]. Data mining and sequence analyses were carried out using ExPASy

proteomic tools (<https://www.expasy.org/tools>). The physicochemical parameters were computed using ProtParam [19] and BioEdit [20].

Genome sequences (16 nos.) of different coronavirus genomes with NCBI-IDs AY572034, AY572035, KF569996, MH734115, MG92481, MG923467, MT040335, MT040333, MT072864, MN996532, MT791905, MT451886, MT973427, DQ412042, DQ648856, and AY321118 were retrieved from NCBI genome database for construction of whole-genome phylogeny.

2.2. Comparative and Ab Initio Modeling. BlastP and FASTA searches were performed independently with PDB to know the existing structure from the PDB, for a suitable template for comparative modeling and to decide *ab initio* modeling requirements (Table 1). The significance of the BLAST results was assessed the e-value generated by the BLAST family of search algorithm and query coverage. The comparative modeling was carried out in the Modeller9.24 program [21], and *ab initio* modeling was done in Baker Rosetta Server (<https://rosetta.bakerlab.org/>). The loop regions were modeled using the ModLoop server [22]. The final 3D structures with complete coordinates were obtained by optimization of the molecular probability density function of the Modeller 9.24 [23]. The computational protein structures were verified by using global and local (per residue) quality estimates of ProQ3 and QMEANDisCo 4.0.0 [24]. All the graphic presentations of the 3D structures were prepared using Chimera version 1.8.1 [25] and pyMOL 0.97rc [26].

2.3. Proteomics Analysis and Functional Annotation. Sequence-based functional annotation was carried out using Pfam ([pfam.sanger.ac.uk/-](http://pfam.sanger.ac.uk/)), Hmmer version 3.3 [27], PFam, PROSITE, and InterProScan. ProFunc server [28] was used to identify the likely biochemical function of proteins from the predicted 3D structures. MOLE 2.0 [29] and Caver Web 1.0 [30] were used for the advanced analysis of biomacromolecular channels. The tunnel bottleneck radius and lengths were calculated in Ångström (Å) and throughput (estimated tunnel importance) calculated as e^{-cost} , where e is Euler's number. Active site prediction of protein server [31] has been used for the computation of cavities in the target proteins.

2.4. Molecular Phylogenetic Analysis. The amino sequences used for phylogenetic analysis were aligned using ClustalW 1.6 [32] integrated in the MEGA X software [33]. The evolutionary history was inferred using maximum likelihood methods [34]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) [35]. The initial tree(s) for the heuristic search were obtained automatically by applying the Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model and then selecting the topology with a superior log likelihood value. To verify the reliability of protein phylogeny (9b and ORF proteins), a whole-genome ML phylogenetic tree was constructed using the General Time Reversible model with Gamma distribution (G).

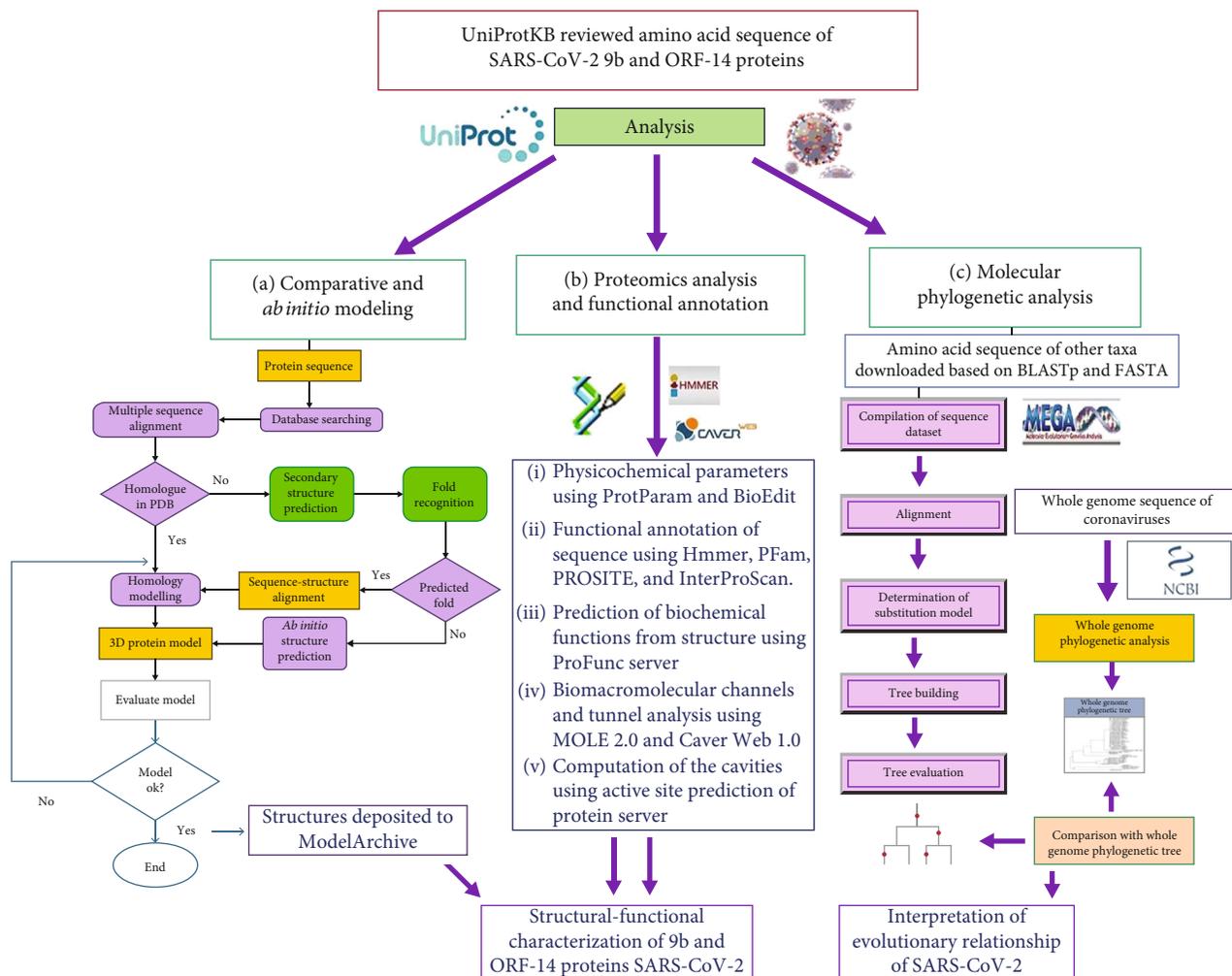


FIGURE 1: A conceptual framework of the present study for analysis of SARS-CoV-2 accessory proteins ORF9b and ORF14.

TABLE 1: BLAST results against available PDB structures for selection of the modeling method, template selection for the structures of 9b and ORF14 proteins.

Sl no.	Protein name and UniProtKB accession number	Length (aa residue)	PDB Template (S)	Identity with template (%)	E-value	Query coverage	The final structure/modeling method selected
1	ORF9b protein (P0DTD2)	97	2CME_B (79 aa)	70.93%	2e-34	89%	Comparative modelling
2	ORF14 protein (P0DTD3)	73	3A32_A	39.13%	9.0	31%	<i>Ab initio</i> modelling

3. Results and Discussion

3.1. Tertiary Structures of 9b and ORF14. The 9b protein (P0DTD2) with molecular weight = 10796.10 Daltons is rich in leucine (12.37%) and valine (10.31%) (Figure 2). The ORF14 (P0DTD3) molecular weight = 8049.27 Daltons is rich in leucine (20.55%) and alanine (12.33%) (Figure 3). ProMotif analysis of the final predicted structure of 9b, modeled using comparative modeling, calculated 2 sheets, 2 beta hairpins, 7 strands, 4 helices, 9 beta turns, and 1 gamma turn (Table 1; Figures 4 and 5; Figure S1). ProMotif analysis of the predicted *ab initio* structure of ORF14 calculated 5 helices, 9 helix-helix interacts,

and 2 beta turns (Table 1; Figures 6 and 7; Figure S2). The verified structures of the ORF9b and ORF14 proteins had qmean4 z scores of -1.64 (Global Score : 0.67 ± 0.09) and -1.18 (Global Score : 0.52 ± 0.11), respectively (Annexures 1 & 2). Precheck verification showed 94.0% and 95.5% residues in the most favored regions (A, B, L) of Ramachandran plot in 9b and ORF14 proteins, respectively. Structural verification in ERRAT revealed good quality of the models with quality factors of 97.56 and 100 for ORF9b and ORF14 proteins, respectively. The verification reports indicate the high reliability of the theoretical structures (Figures S3-S6).

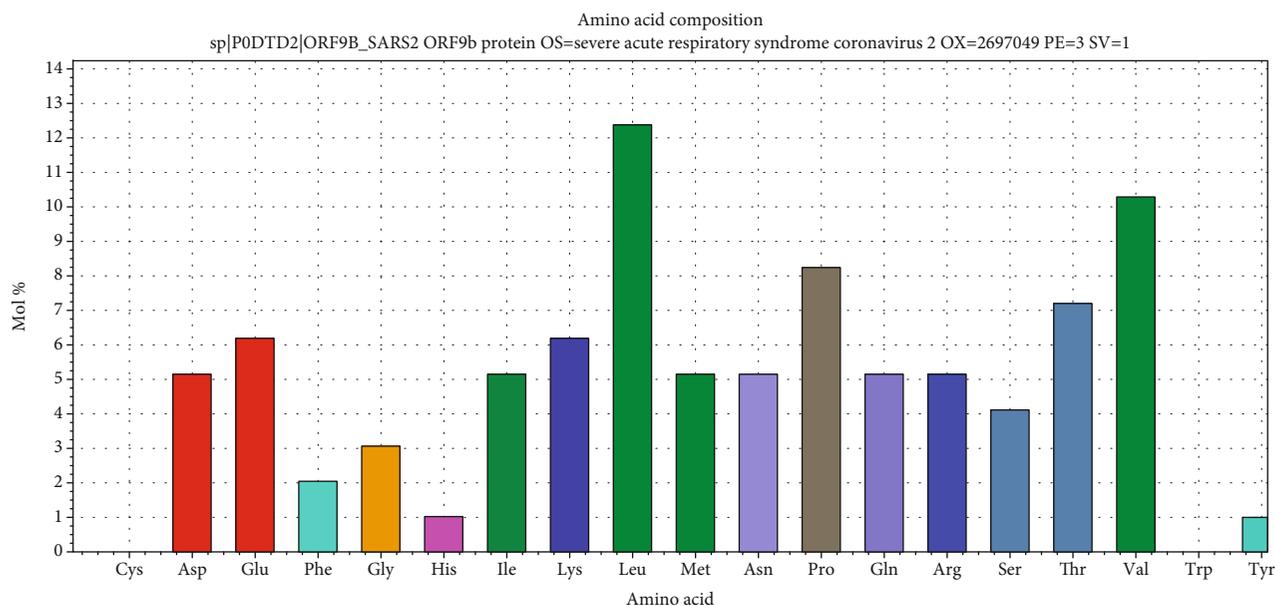


FIGURE 2: Amino acid composition of SARS-CoV-2 9b protein (leucine and valine-rich).

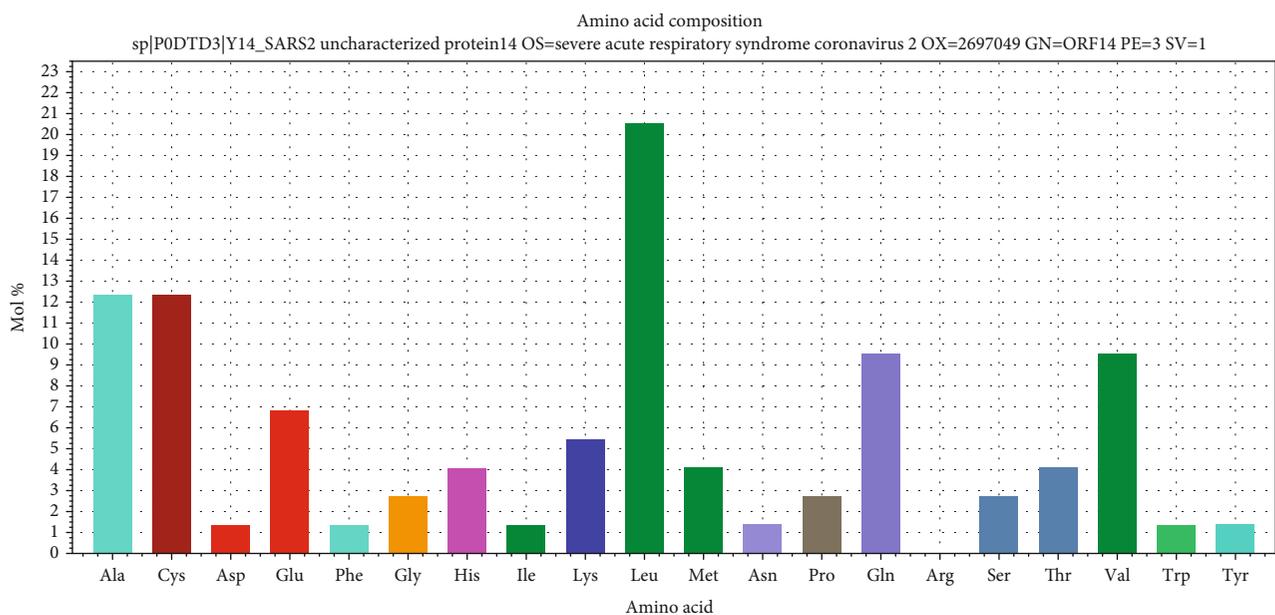


FIGURE 3: Amino acid composition of SARS-CoV-2 ORF14 protein (leucine and valine-rich).

The estimated high throughput tunnel-1 (blue) in the 9b protein is a bottleneck radius of 1.9 Å, length of 1.5 Å, distance to the surface of 1.5 Å, curvature of 1.0, throughput of 0.92, and number of residues 11; tunnel-2 (green) is a bottleneck radius of 1.5 Å, length of 5.4 Å, distance to surface of 4.9 Å, curvature of 1.1, throughput of 0.78, and number of residues 14 (Figure 5, Figures S7-S8). The estimated high throughput tunnel-1 (blue) in ORF14 protein is a bottleneck radius of 1.2 Å, length of 5.8 Å, distance to the surface of 4.9 Å, curvature of 1.2, throughput of 0.69, and number of residues 14 (Figure 7; Figure S9).

Structural comparison of SARS-CoV-2 9b protein (97 aa residues) with the crystal structure of SARS-CoV ORF9b

protein (79 aa residues; PDB ID 2 CME), which shared 70.93% sequence identity (89% query coverage) showed minor differences in the number of strands, helices, and beta turns (SARS-CoV: 6 strands, 1 helix, 7 beta turns, 3 gamma turns; SARS-CoV-2: 7 strands, 4 helices, 5 beta turns). This difference in the increase in the number of helices and strands may be due to increased sequence length in the SARS-CoV-2 9b protein. Changes in the DNA sequence will therefore affect both the conventional and alternative ORF, limiting the rate and extent to which the corresponding proteins can evolve [14]. The structure of SARS-CoV ORF9b is a 2-fold symmetric dimer constructed from two adjacent twisted β sheets [36]. Each of these sheets is formed from β

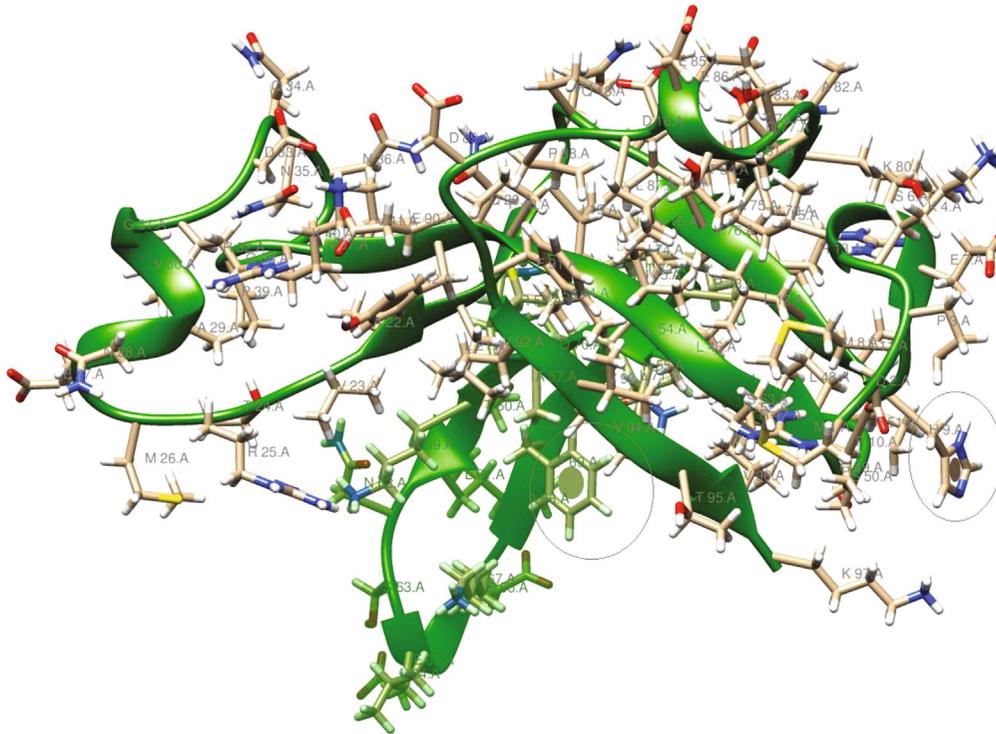


FIGURE 4: Structure of SARS-CoV-2 9b protein along with major active sites.

strands contributed by both monomers, which form a highly interlocked architecture reminiscent of a handshake. The interdigitated nature of the ORF9b dimer rests on a highly unusual topology of largely antiparallel β sheets in which monomers wrap around each other [36].

The accessory ORF14 described only by Marra et al. [14] is still an uncharacterized protein, and very little is known about its structure and interactions. ORF14 has no significant sequence homology to proteins in other coronaviruses. It belongs to the group of proteins, named as predicted unknown proteins (PUPs), and is unique to SARS-CoV [15, 37]. Interactions among SARS-CoV accessory proteins were studied using a bimolecular fluorescence complementation assay [15, 37]. Self-interactions were observed with 9b and ORF14, indicating the formation of dimeric or multimeric complexes in the nucleus, similar to the findings of von Brunn et al. [15]. ORF9b and ORF14 interacted with themselves, indicating the formation of dimeric or multimeric complexes [15]. ORF9b and ORF14 self-interactions were also found in the co-Immunoprecipitation (CoIP) assay. α -Galactosidase and β -galactosidase assays of protein interactions of 9b-9b, 8a-9b, and ORF14-ORF14 demonstrated self-interactions [15, 37].

3.2. Proteomics Profiles of 9b and ORF14. The InterProScan Search Result of ORF9b protein has revealed that it belongs to protein family—protein 9b and SARS-like (IPR018542) (Figure 8). This is a family of proteins found in SARS and SARS-like coronaviruses. It includes protein 9b from SARS coronavirus 2 (SARS-CoV-2), human SARS coronavirus (SARS-CoV), and bat coronaviruses. Protein 9b is one of 8

accessory proteins in SARS-CoV [38]. The gene (ORF9b, also known as ORF13) that encodes this protein is included within the nucleocapsid (N) gene (alternative ORF) [39]. The ORF9b accessory protein is associated with the spike and nucleocapsid proteins and has unusual membrane-binding properties [14, 36]. SARS-CoV ORF9b has been shown to localize to the outer mitochondrial membrane and target mitochondrial antiviral signaling proteins (MAVS), suppressing innate immunity [40, 41]. Antibodies against SARS-CoV ORF9b have been found in patients, demonstrating that it is produced during infection [36, 42]. Protein 9b from SARS-CoV comprises 98 amino acids, the structure of which has a novel fold that forms a dimeric tent-like beta structure with an amphipathic surface, and a central hydrophobic cavity that binds lipid molecules [36]. This cavity is likely involved in membrane attachment [36]. The sequence of ORF9b is well conserved in different SARS isolates; however, there is little homology between protein 9b from SARS-CoV and the I-protein (protein 9b homologue) present in other coronaviruses [39, 43].

InterProScan Search Result of ORF14 protein revealed its family of membership as protein 14, SARS-like (IPR035113) (*Protein 14_SARS-like*) (Figure 9). This is a family of unknown functions found in SARS and SARS-like coronaviruses. It includes uncharacterized protein 14 from SARS coronavirus 2 (SARS-CoV-2), human SARS coronavirus (SARS-CoV), and bat coronavirus Rp3/2004 (SARS-like coronavirus Rp3) [14]. In SARS-CoV, ORF14 is completely contained within the ORF encoding the nucleocapsid protein (N) [38]. In SARS-CoV-2, uncharacterized protein 14 was predicted to contain one transmembrane helix. The ORF14

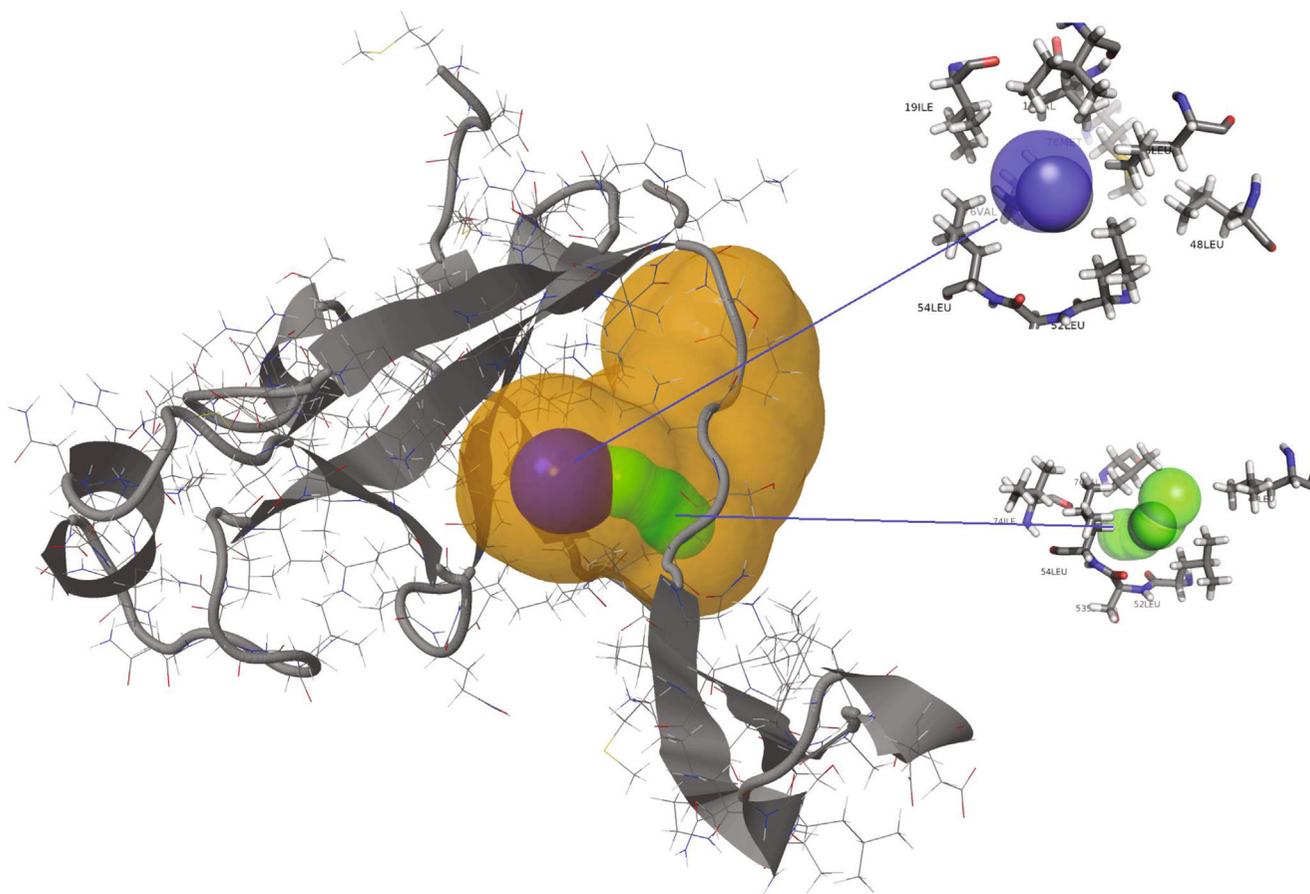


FIGURE 5: Structure of SARS-CoV-2 9b protein and its two high-throughput tunnels. Tunnels are colored on the basis of preferences in throughput values, i.e., tunnel-1 (blue) and tunnel-2 (green). The high relevance pocket is shown (yellow).

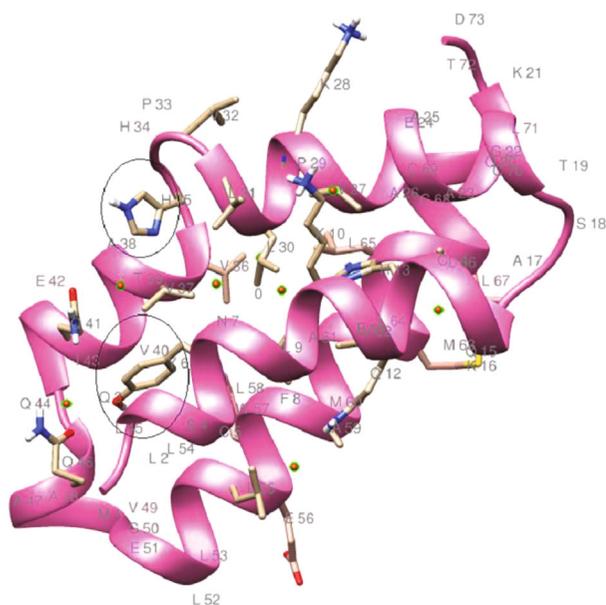


FIGURE 6: Structure of SARS-CoV-2 ORF14 protein along with major active sites.

protein is with three domains: (i) noncytoplasmic domain (1-51), (ii) transmembrane region (52-72), and (iii) cytoplasmic domain (73-73).

Protein 9b shows its subcellular location as a host cytoplasmic vesicle membrane, peripheral membrane protein, and host cytoplasm that binds noncovalently to intracellular lipid bilayers. Gene ontology revealed the cellular components of the host cell cytoplasmic vesicle membrane, and the subunit structure is homodimer with binary interactions. ORF14 protein may play a role in host-virus interaction-subcellular location: membrane sequence analysis and single-pass membrane protein sequence analysis. The topology of gene ontology exhibits cellular components, integral components of the membrane, transmembrane, and transmembrane helices.

3.3. Functional Annotation of 9b and ORF14. PROSITE analysis of the 9b protein revealed three sites: (i) PS00006 CK2_PHOSPHO_SITE Casein kinase II phosphorylation site (24-27; 63-66; 83-86), (ii) PS00008 MYRISTYL N-myristoylation site (49-54), and (iii) PS00005 PKC_PHOSPHO_SITE Protein kinase C phosphorylation site (95-97).

The domain profile of ORF9b resembles the Sarbecovirus 9b domain profile (PROSITE entry PS51920). Coronaviruses are divided into four genera: α -coronavirus, β -coronavirus,

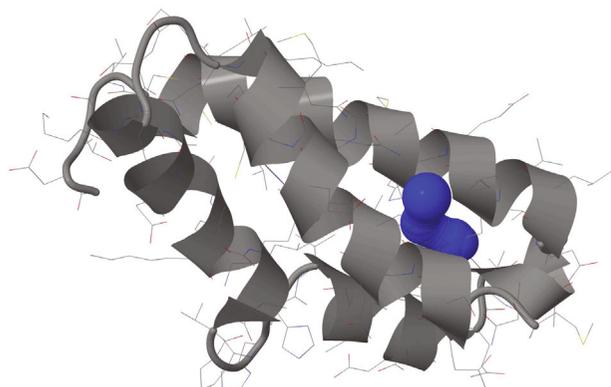


FIGURE 7: Structure of SARS-CoV-2 ORF14 protein along with its throughput tunnels (blue).

γ -coronavirus, and delta-coronavirus. SARS, SARS-CoV-2, BatCoV RaTG13, and Bat-SARS-like coronavirus (BAT-SL-CoVZXC21 and BAT-SL-CoVZC45) belong to the Sarbecovirus subgenus of β -coronavirus.

Coronaviruses code for the characteristic proteins replicase polyprotein (pp1ab), spike (S), membrane (M), envelope (E), and nucleocapsid (N) proteins. In addition, Sarbecoviruses code for subgroup-specific accessory proteins that are thought to be dispensable for viral replication in cell culture but may be important for virus-host interactions and thus contribute to virus fitness.

To achieve the optimum output from their limited genomes, viruses frequently make use of alternative open reading frames, in which translation is initiated from a start codon within an existing gene and, being out of frame, gives rise to a distinct protein product. ORF9b codes for a small accessory protein of 98 amino acid residues, which are found in Sarbecovirus-infected cells. The ORF9b protein (p9b) has been shown to self-interact and interact with nsp5, nsp14, and the accessory protein p6. The function of p9b is unknown, although it has been suggested that it specifically recognizes and binds to intracellular vesicular. The 9b protein could have a role in membrane interactions during the assembly of the virus membranes [36, 44, 45].

The 9b domain has a fold with seven β -strands (PDB ID: 2CME). The β -strands from two molecules form two adjacent twisted β -sheets, resulting in a highly interlocked handshake structure that contains a hydrophobic central cavity, which binds to lipids and stabilizes the molecule (Meier et al., 2006) [36]. Protein 9b is a homodimer that plays a role in membrane interactions during the assembly of the virus.

PROSITE analysis of ORF14 protein also estimated three sites: (i) PS00005 PKC_PHOSPHO_SITE Protein kinase C phosphorylation site (aa 19-21), (ii) PS00008 MYRISTYL N-myristoylation site (aa 22-27), and (iii) PS00006 CK2_PHOSPHO_SITE Casein kinase II phosphorylation site (aa 39-42).

Casein kinase II (CK-2) is a protein serine/threonine kinase whose activity is independent of cyclic nucleotides and calcium. CK-2 phosphorylates many different proteins [46]. N-myristoylation site, an appreciable number of eukaryotic proteins are acylated by the covalent addition of

myristate (a C14-saturated fatty acid) to their N-terminal residue via an amide linkage [47, 48]. The sequence specificity of the enzyme responsible for this modification, myristoyl CoA:protein N-myristoyl transferase (NMT), has been derived from the sequence of known N-myristoylated proteins and from studies using synthetic peptides [48]. In vivo, protein kinase C exhibits a preference for the phosphorylation of serine or threonine residues found close to a C-terminal basic residue (Meier et al., 2006; Liu et al., 2014). The presence of additional basic residues at the N- or C-terminal of the target amino acid enhances the V_{max} and K_m of the phosphorylation reaction [49].

The instability index values of the ORF9b and ORF14 proteins of SARS-CoV-2 were 33.11 and 32.56, respectively, which classifies both proteins were stable (Table 2). The aliphatic indices of SARS-CoV2 9b and ORF14 were 105.46 and 125.62, respectively, indicating high thermal stability in both proteins (Table 2). The grand average of hydropathicity (GRAVY) values of the 9b and ORF14 proteins are computed as -0.085 and 0.603, respectively, which indicates that protein 9b is hydrophilic and ORF14 is hydrophobic in nature (Table 2; Figures S10 and S11).

A comparison made in this study on the physicochemical parameters of ORF9b and ORF14 proteins among the different coronaviruses showed that ORF9 protein of SARS-CoV-2 has 76.53% sequence identity with *Rhinolophus affinis* coronavirus, 74.23% with human CoV, and 73.20% with bat SARS-CoV. The ORF14 protein of SARS-CoV-2 has 92.86% identity with the ORF14 protein of bat coronavirus, 78.57% with human SARS-CoV, 77.14% with civet and *Rhinolophus affinis* coronavirus. The ORF9 protein showed a wide range of isoelectric points from 4.9 (human and civet SARS-CoV) to 6.56 (human SARS-CoV-2). The instability index values of ORF9b of coronaviruses ranged from 33.11 (human SARS-CoV-2) to 41.80 (bat CoV) (Table 2). The instability index values of ORF14 of coronaviruses ranged from 25.59 (bat CoV) to 32.81 (*R. affinis* CoV) (Table 2). This indicates higher stability of the ORF14 protein than the ORF9 protein. The grand average of hydropathicity (GRAVY) values were computed in the range of -0.176 to -0.012 in ORF9b protein and 0.196 to 0.603 in ORF14, indicating that ORF9b protein is hydrophilic and ORF14 is hydrophobic in nature (Table 2). The physicochemical parameters, including amino acid composition, pI, instability index and hydropathicity of SARS-CoV-2, showed higher identity with bat SARS-CoV (Table 2).

The functional analysis results of Profunc have been presented in Table 3. Of the nine (09) estimated cavity points in the structure of the 9b protein, the cavity-1 produced by amino acids "NPQVDKGEYTAMIFRLS" is with xyz coordinates of 10.606, -3.416, and -5.832 and a volume of 1261 Ångström cube; the cavity-2 produced by the amino acids "DKQPRVELNTFYIAM" is with cavity point 10.014, -0.066, and 5.372 and a volume of 971 Ångström cube (Figure 5; Table S1). Out of the 10 potential cavities for computed active sites for the function of ORF14 protein, cavity-1 is represented by amino acids "HEPIATVLKWCMDMY," with xyz coordinates of -7.341, 17.615, and -5.427 and a volume of 662 Ångström cube; the cavity-2 with amino acids

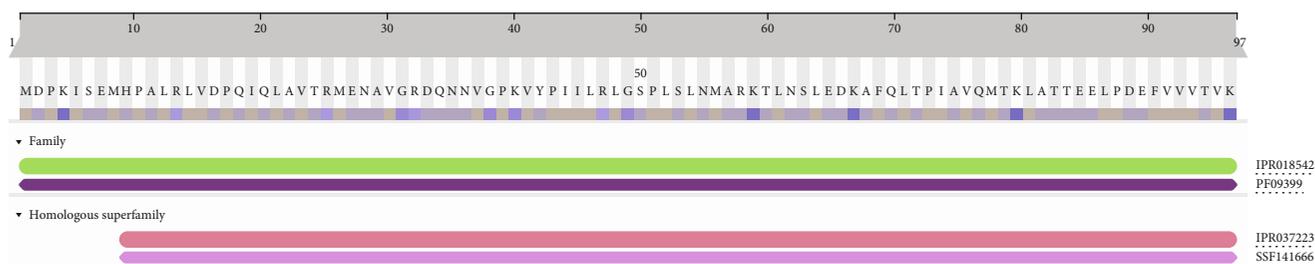


FIGURE 8: Protein family membership of SARS-CoV-2 protein 9b resembles the SARS-like protein (IPR018542) and the 9b SARS InterPro homologous superfamily (9-97; IPR037223).

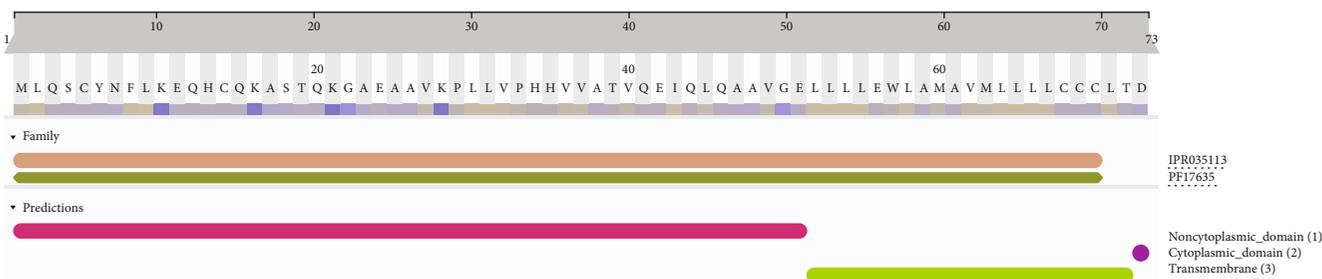


FIGURE 9: Protein family membership of SARS-CoV-2 protein 14 resembles SARS-like protein (IPR035113). The protein is with three domains: (i) noncytoplasmic domain (1-51), (ii) transmembrane region (52-72), and (iii) cytoplasmic domain (73-73).

TABLE 2: Physicochemical parameters of SARS-CoV-2 9b and ORF14 proteins and comparison with other coronaviruses.

Protein name	UniProt KB accession no. organism	Length	MW (Da)	pI	Chemical formula	Instability index	Aliphatic index	Gravy*
ORF9b protein	P0DTD2 human SARS-CoV-2	97	10796.66	6.56	C ₄₇₈ H ₇₉₆ N ₁₃₀ O ₁₄₂ S ₅	33.11	105.46	-0.085
	Q3LZX3 bat CoV	97	10722.54	6.05	C ₄₇₅ H ₇₈₆ N ₁₂₈ O ₁₄₂ S ₅	41.80	104.43	-0.012
	Q6RD12 human SARS-CoV	98	10802.45	4.90	C ₄₇₂ H ₇₇₈ N ₁₃₀ O ₁₄₈ S ₅	38.95	98.47	-0.122
	A0A023PUR2 R. <i>affinis</i> CoV	98	10781.50	5.69	C ₄₇₅ H ₇₈₇ N ₁₃₁ O ₁₄₅ S ₄	39.97	105.41	-0.050
	Tr Q3ZTD0 SARS-CoV civet010	98	10790.40	4.90	C ₄₇₀ H ₇₇₄ N ₁₃₀ O ₁₄₉ S ₅	38.95	94.49	-0.176
ORF14 protein	P0DTD3 human SARS-CoV-2	73	8049.65	5.79	C ₃₅₉ H ₅₈₈ N ₉₂ O ₁₀₀ S ₈	32.56	125.62	0.603
	AVP78040 bat CoV	70	7690.13	6.38	C ₃₄₁ H ₅₅₅ N ₉₁ O ₉₆ S ₇	25.59	117.14	0.466
	ARO76392 human SARS-CoV	70	7842.29	6.25	C ₃₅₄ H ₅₇₁ N ₉₃ O ₉₇ S ₅	32.40	119.86	0.321
	AAU04674 SARS-CoV civet	70	7868.37	6.25	C ₃₅₇ H ₅₇₇ N ₉₃ O ₉₆ S ₅	23.92	125.43	0.387
	AHX37568 R. <i>affinis</i> CoV	70	7810.21	6.39	C ₃₅₂ H ₅₆₃ N ₉₃ O ₉₇ S ₅	32.81	111.57	0.196

*GRAVY: Grand average of hydropathicity.

“PATIHQVLWKYENMCSF” is with a cavity point of -1.634, 12.400, and 3.280 and a volume of 494 Ångström cube (Figure 7; Table S2). ORF9b is an unusual membrane-binding protein with a long hydrophobic lipid-binding tunnel.

3.4. Molecular Phylogeny of the 9b and ORF14 Proteins. Evolutionary analysis of the 9b and ORF14 proteins of SARS-CoV-2 was based on the Maximum Likelihood (ML) method and the JTT matrix-based model. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The ML phylogenetic tree, based on the amino acid sequence of the 9b protein, revealed that it has close evolutionary relatedness with human SARS coronavirus (UniProtKB accession number APO40587) followed by the bat SARS-CoV (UniProtKB accession numbers AAZ67037, AAZ41338, and Q3LZX3 (Figure 10)).

However, the ML phylogenetic tree based on the amino acid sequence of the human SARS-CoV2 ORF14 protein showed the closest evolutionary relationship with bat SARS-like coronaviruses (accession number AVP78040) with 100% boot strap support (Figure 11).

The whole-genome phylogenetic tree strongly supports the protein phylogeny based on ORF9b and ORF14 proteins, indicating that the close evolutionary SARS-CoV-2 has very closely evolutionarily related to newly sequenced bat coronavirus RaTG13 genome/March 2020 from China (MN996532) followed by the pangolin coronavirus genome (MT040333, MT040335, MT072864) (Figure 12). Bat SARS-CoV Rf1/2004 (DQ412042) and bat CoV 273/2005 (DQ648856) along with human SARS-CoV and horseshoe bat (*Rhinolophus affinis*) formed a different clade in the whole-genome phylogeny, indicating rapid evolution of coronavirus.

TABLE 3: Predicted functions of SARS-CoV-2 9b and ORF14 proteins with respective ProFunc score (shown within parenthesis).

Protein name	Protein name terms	Summary of predicted function		
		Cellular component	Gene ontology (GO) terms Biological process	Biochemical function
9b protein	SARS coronavirus ORF9b (0.90) aquifex aeolicus trbp111 structure-specific (0.50) aeolicus trbp111 structure-specific trna (0.50) trbp111 structure-specific trna binding (0.50) ustilago maydis lipase um03410 (0.50) maydis lipase um03410 short (0.50) lipase um03410 short form (0.50) um03410 short form without (0.50)	Extracellular region (1.33) cytoplasm (0.85)	Cellular process (1.66) cellular metabolic process (1.66)	tRNA binding (0.50) RNA binding (0.50) aminoacyl-tRNA ligase activity (0.50) binding (0.50)
ORF14 protein	Human (1.72) domain (1.56) atcc (1.00) nmr (1.00) aminoimidazole riboside kinase (0.70) phycocyanin (0.57) ccm3 (0.50) c-terminal regulatory domain stk25 (0.50)	Cytoplasm (1.86) cell (1.86) cell part (1.86) intracellular (1.86)	Metabolic process (3.06) cellular process (2.27) cellular metabolic process (2.27) primary metabolic process (1.53)	Catalytic activity (2.91) binding (2.48) metal ion binding (1.61) ion binding (1.61)

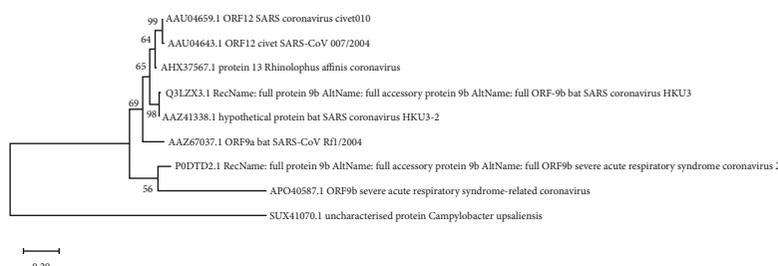


FIGURE 10: Evolutionary analysis of 9b protein of SARS-CoV-2 by Maximum Likelihood method and JTT matrix-based model (Jones and Taylor, 1992). The tree with the highest log likelihood (-1113.75) is shown. This analysis involved 9 amino acid sequences. There were a total of 141 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

Moreover, all the pangolin coronavirus genomes sequenced in April 2020 (MT04033, MT072864, MT040335) were found to be sister taxa in the whole-genome phylogeny. The findings indicate that only the new bat coronavirus followed by pangolin coronaviruses have close evolutionary related with the novel SARS-CoV-2. The present study strongly supports that like the human host the coronavirus had undergone rapid evolution in bats and pangolin as an amplifying host (Figure 12).

Earlier research claimed that snakes or pangolins may be intermediate hosts for creating the coronavirus by recombination events [50]. Cross-species transmission of zoonotic coronaviruses (CoVs) can result in disease outbreaks [51]. Molecular analysis supported bats as natural hosts for SARS-CoV, but palm civets (*Paguma larvata*) had a critical role in the transmission to humans [52, 53]. Bats are implicated in SARS-CoV-2 origin. A very similar SARS-CoV-2 strain (RaTG13 CoV) was detected in *Rhinolophus affinis* bat with 96% genome similarity compared with SARS-CoV-2 genome sequence. Considering that bats were in hibernation when the outbreak occurred, the virus is more likely to have been transmitted via other species [54]. Both protein (9b and ORF14) genome phylogeny results of the present study are supported by the hypothesis for the zoonotic transmission route was constructed based on contact

with Malayan pangolins (*Manis javanica*) by visitors of Huanan seafood market in Wuhan, China [55]. The close phylogenetic relationship to RaTG13 provides evidence that 2019-nCoV may have originated in bats [10]. Differently from bats, which are able to suppress viral replication, pangolin is an amplifying host which allows the increase of viral load and accelerated SARS-CoV-2 jump to human host and human-to-human transmission subsequently [56].

Another study, which supports the results of present finding, showed that the bat and pangolin coronaviruses were the most related to SARS-CoV-2 with 96% and 86% of identity all along the genome [57]. The comparison study from bat and pangolin by Li and his friends explains that BetaCoV/bat/Yunnan/RaTG13/2013 virus was more similar to the SARS-CoV-2 virus than the coronavirus obtained from the two pangolin samples (SRR10168377 and SRR10168378). The human SARS-CoV-2 virus, which is responsible for the recent outbreak of COVID-19, did not come directly from pangolins [13].

Tunnels are access paths connecting the interior of molecular systems with the surrounding environment. The presence of tunnels in proteins influences their reactivity, as they determine the nature and intensity of their interactions. Tunnel analysis of the newly predicted structures of the present study has estimated the presence of multiple tunnels in

It is reasonable to propose that ORF9b in SARS-CoV-2 may contribute to viral pathogenesis as I-protein in mouse hepatitis virus (MHV) does. The interactions between S, N, and ORF9b may help to localize ORF9b inside the particles but closer to the envelope. The structure of ORF9b, an intertwined dimer with an amphipathic outer surface and a long hydrophobic lipid binding tunnel, suggests how this protein may interact, via an unusual anchoring mechanism, with compartments of the ER-Golgi network to act as an accessory protein during the assembly of the SARS virion (Meier et al., 2006). However, further analyses of the properties and functions of ORF9b and ORF14 proteins are still necessary to understand its contribution to virus pathogenesis. All current studies on accessory proteins of coronaviruses, including SARS-CoV-2, suggest that they are not essential for virus replication [63] but do affect viral release, stability, and pathogenesis and finally contribute to virulence [64].

4. Conclusion

The RNA genome of SARS-CoV-2 has 29.9 kb nucleotides, encoding 14 open reading frames (ORFs) for 29 proteins, although one may not be expressed. Studying these different components of the virus as well as how they interact with human cells has already yielded some clues but much remains to be explored. The present study reported theoretical modeling, sequence-based, and structure-based functional characterization of two accessory protein-9b and ORF14 of SARS-CoV-2 p. Phylogenetic analysis of both proteins revealed a close evolutionary relationship between the newly emerged human SARS-CoV-2 and bat SARS-like corona virus. The whole-genome phylogeny indicates that 2019-nCoV may have originated in bat, undergone rapid evolution in bats, and pangolin may more likely to be an amplifying host. The presence of a large number of tunnels in the 9b protein indicates its high reactivity. The theoretical structures and statistical verification reports were successfully deposited in the Model Archive. The theoretical structures would perhaps be useful for advanced computational analysis of interactions of each protein for detailed functional analysis, understanding of viral pathogenesis and virulence for structure-based drug design, or to study potential vaccines, if at all, towards to prevent epidemics and pandemics in the absence of a complete experimental structure.

Data Availability

(1) The resultant protein structures are deposited in ModelArchive (<https://www.modelarchive.org/>). The same data has been provided in a supplementary file (folder name: Data availability). (2) The supplementary file for the data generated in the project has been deposited to ChemRxiv. Preprint. doi:10.26434/chemrxiv.12424958.v1 (supplementary file). (3) All the above data are also included along with this manuscript.

Additional Points

Commercial Declarations. It is made available under a CC-BY-NC-ND 4.0 International license.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

P Devi and DKS were involved in planning, designing, and realizing the present work. Data acquisition and analysis were done by C. Baruah. After in-depth discussion of the results with DKS, both CB and PD prepared the manuscript.

Acknowledgments

The authors are grateful to DBT-Govt. of India for supporting Bioinformatics Laboratory (under the DBT-Star College scheme) at the Post Graduate Department of Zoology, Darrang College, Tezpur, Assam. The authors are thankful to the Principal, Darrang College (Gauhati University), Tezpur (Assam), India, and Head of the Post Graduate Department of Zoology, Darrang College, and the University of Science and Technology, Meghalaya, India, for supporting the research laboratory facility.

Supplementary Materials

Table S1: computed cavities in the 3D structure of ORF9b protein for active sites. Table S2: computed cavities in the 3D structure of ORF14 protein for active sites. Figure S1: secondary structure profile of 9b protein. Figure S2: secondary structure profile of ORF14 protein. Figure S3: QMEANDisCo local quality estimate for 9b protein. Figure S4: QMEANDisCo local quality estimate for ORF14 protein. Figure S5: protein 9b structure verification in ERRAT. Figure S6: protein ORF14 structure verification in ERRAT. Figure S7: profile of tunnel 1 in 9b protein. Figure S8: profile of tunnel 2 in 9b protein. Figure S9: tunnel-profile of ORF14 protein. Figure S10: hydropathicity plot for 9b protein. Figure S11: hydrophobicity plot for ORF14 protein. Annexure 1: protein 9b structure verification. Annexure 2: ORF14 protein structure verification. (*Supplementary Materials*)

References

- [1] H. Yang, M. Bartlam, and Z. Rao, "Drug design targeting the main protease, the Achilles' heel of coronaviruses," *Current Pharmaceutical Design*, vol. 12, no. 35, pp. 4573–4590, 2006.
- [2] C. Baruah, P. Devi, and D. K. Sharma, "In silico proteome analysis of severe acute respiratory syndrome coronavirus 2(SARS-CoV-2)," in *BioRxiv*, 2020.
- [3] G. Lu, Q. Wang, and G. F. Gao, "Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond," *Trends in microbiology*, vol. 23, no. 8, pp. 468–478, 2015.

- [4] G. Lu and D. Liu, "SARS-like virus in the Middle East: a truly bat-related coronavirus causing human diseases," *Protein & Cell*, vol. 3, no. 11, pp. 803–805, 2012.
- [5] C. I. Paules, H. D. Marston, and A. S. Fauci, "Coronavirus Infections—More than just the common cold," *JAMA*, vol. 323, no. 8, pp. 707–708, 2020.
- [6] P. C. Woo, S. K. Lau, C. S. Lam et al., "Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus," *Journal of Virology*, vol. 86, no. 7, pp. 3995–4008, 2012.
- [7] C. K. Chang, S. Jeyachandran, N. J. Hu et al., "Structure-based virtual screening and experimental validation of the discovery of inhibitors targeted towards the human coronavirus nucleocapsid protein," *Molecular BioSystems*, vol. 12, no. 1, pp. 59–66, 2016.
- [8] J. Cui, N. Han, D. Streicker et al., "Evolutionary relationships between bat coronaviruses and their hosts," *Emerging infectious diseases*, vol. 13, no. 10, pp. 1526–1532, 2007.
- [9] C. Huang, Y. Wang, X. Li et al., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.
- [10] P. Zhou, X.-L. Yang, X.-G. Wang et al., "A pneumonia outbreak associated with a new coronavirus of probable bat origin," *Nature*, vol. 579, no. 7798, pp. 270–273, 2020.
- [11] S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, and R. Ke, "High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2," *Emerging Infectious Diseases*, vol. 26, no. 7, pp. 1470–1477, 2020.
- [12] H. A. Rothan and S. N. Byrareddy, "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak," *Journal of Autoimmunity*, vol. 109, p. 102433, 2020.
- [13] X. Li, J. Zai, Q. Zhao et al., "Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2," *Journal of medical virology*, vol. 92, no. 6, pp. 602–611, 2020.
- [14] M. A. Marra, S. J. Jones, C. R. Astell et al., "The genome sequence of the SARS-associated coronavirus," *Science*, vol. 300, no. 5624, pp. 1399–1404, 2003.
- [15] A. von Brunn, C. Teepe, J. C. Simpson et al., "Analysis of intraviral protein-protein interactions of the SARS coronavirus ORF3," *PLoS One*, vol. 2, no. 5, p. e459, 2007.
- [16] K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, and R. F. Garry, "The proximal origin of SARS-CoV-2," *Nature Medicine*, vol. 26, no. 4, pp. 450–452, 2020.
- [17] S. F. Altschul, T. L. Madden, A. A. Schaffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic acids research*, vol. 25, no. 17, pp. 3389–3402, 1997.
- [18] W. R. Pearson, "Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms," *Genomics*, vol. 11, no. 3, pp. 635–650, 1991.
- [19] E. Gasteiger, C. Hoogland, A. Gattiker et al., "Protein identification and analysis tools on the ExPASy server," in *The Proteomics Protocols Handbook*, pp. 571–607, Humana Press, 2005.
- [20] T. A. Hall, "Bio edit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT," *Nucleic Acids Symposium Series*, vol. 41, pp. 95–98, 1999.
- [21] B. Webb and A. Sali, "Comparative protein structure modeling using Modeller," *Current Protocols in Bioinformatics*, vol. 15, pp. 5–6, 2016.
- [22] A. Fiser, R. K. Do, and A. Sali, "Modeling of loops in protein structures," *Protein Science*, vol. 9, pp. 1753–1773, 2008.
- [23] N. Eswar, B. Webb, M. A. Marti-Renom et al., "Comparative protein structure modeling using MODELLER," *Current Protocols in Protein Science*, J. E. Coligan, B. M. Dunn, D. W. Speicher, and P. T. Wingfield, Eds., 2006.
- [24] J. Haas, R. Gumienny, A. Barbato et al., "Introducing "best single template" models as reference baseline for the continuous automated model evaluation (CAMEO)," *Proteins*, vol. 87, no. 12, pp. 1378–1387, 2019.
- [25] E. Pettersen, J. Lamotte-Brasseur, J. P. Chessa et al., "UCSF Chimera?A visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.
- [26] W. DeLano and L. Pymol, "An open-source molecular graphics tool," *CCP4 Newsletter On Protein Crystallography*, vol. 40, pp. 82–92, 2002.
- [27] R. D. Finn, J. Clements, and S. R. Eddy, "HMMER web server: interactive sequence similarity searching," *Nucleic Acids Research*, vol. 39, suppl, pp. W29–W37, 2011.
- [28] R. A. Laskowski, J. D. Watson, and J. M. Thornton, "ProFunc: a server for predicting protein function from 3D structure," *Nucleic Acids Research*, vol. 33, no. Web Server, pp. W89–W93, 2005.
- [29] D. Sehnal, R. Svobodová Vařeková, K. Berka et al., "MOLE 2.0: advanced approach for analysis of biomacromolecular channels," *Journal of cheminformatics*, vol. 5, no. 1, p. 39, 2013.
- [30] J. Stourac, O. Vavra, P. Kokkonen et al., "Caver web 1.0: identification of tunnels and channels in proteins and analysis of ligand transport," *Nucleic Acids Research*, vol. 47, no. W1, pp. W414–W422, 2019.
- [31] T. Singh, D. Biswas, and B. Jayaram, "AADS- An automated active site identification, Docking, and scoring protocol for protein targets based on Physicochemical Descriptors," *Journal of chemical information and modeling*, vol. 51, no. 10, pp. 2515–2527, 2011.
- [32] J. D. Thompson, D. G. Higgins, and T. J. Gibson, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice," *Nucleic acids research*, vol. 22, no. 22, pp. 4673–4680, 1994.
- [33] S. Kumar, G. Stecher, M. Li, C. Nuyez, and K. Tamura, "MEGA X: molecular evolutionary genetics analysis across computing platforms," *Molecular Biology and Evolution*, vol. 35, no. 6, pp. 1547–1549, 2018.
- [34] D. T. Jones, W. R. Taylor, and J. M. Thornton, "The rapid generation of mutation data matrices from protein sequences," *Computer Applications in the Biosciences*, vol. 8, no. 3, pp. 275–282, 1992.
- [35] J. Felsenstein, *Inferring phylogenies*, Sinauer Associates, Sunderland, Massachusetts, 2003.
- [36] C. Meier, A. R. Aricescu, R. Assenberg et al., "The crystal structure of ORF-9b, a lipid binding protein from the SARS coronavirus," *Structure*, vol. 14, no. 7, pp. 1157–1165, 2006.
- [37] J. Kong, Y. Shi, Z. Wang, and Y. Pan, "Interactions among SARS-CoV accessory proteins revealed by bimolecular fluorescence complementation assay," *Acta Pharmaceutica Sinica B*, vol. 5, no. 5, pp. 487–492, 2015.

- [38] K. Xu, B. J. Zheng, R. Zeng et al., "Severe acute respiratory syndrome coronavirus accessory protein 9b is a virion-associated protein," *Virology*, vol. 388, no. 2, pp. 279–285, 2009.
- [39] I. Moshynskyy, S. Viswanathan, N. Vasilenko et al., "Intracellular localization of the SARS coronavirus protein 9b: evidence of active export from the nucleus," *Virus Research*, vol. 127, no. 1, pp. 116–121, 2007.
- [40] C. S. Shi, H. Y. Qi, C. Boullaran et al., "SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome," *Journal of Immunology*, vol. 193, no. 6, pp. 3080–3089, 2014.
- [41] L. Zhang, Y. Qin, and M. Chen, "Viral strategies for triggering and manipulating mitophagy," *Autophagy*, vol. 14, no. 10, pp. 1665–1673, 2018.
- [42] M. Qiu, Y. Shi, Z. Guo et al., "Antibody responses to individual proteins of SARS coronavirus and their neutralization activities," *Microbes and Infection*, vol. 7, no. 5–6, pp. 882–889, 2005.
- [43] K. Sharma, S. Akerstrom, A. K. Sharma et al., "SARS-CoV 9b protein diffuses into nucleus, undergoes active Crm 1 mediated nucleocytoplasmic export and triggers apoptosis when retained in the nucleus," *PLoS One*, vol. 6, article e19436, 2011.
- [44] D. X. Liu, T. S. Fung, K. K. Chong, A. Shukla, and R. Hilgenfeld, "Accessory proteins of SARS-CoV and other coronaviruses," *Antiviral Research*, vol. 109, pp. 97–109, 2014.
- [45] A. Shukla and R. Hilgenfeld, "Acquisition of new protein domains by coronaviruses: analysis of overlapping genes coding for proteins N and 9b in SARS coronavirus," *Virus Genes*, vol. 50, no. 1, pp. 29–38, 2015.
- [46] L. A. Pinna, "Casein kinase 2: an 'eminence grise' in cellular regulation?," *Biochimica et Biophysica Acta*, vol. 1054, no. 3, pp. 267–284, 1990.
- [47] D. A. Towler, J. I. Gordon, S. P. Adams, and L. Glaser, "The biology and enzymology of eukaryotic protein acylation," *Annual Review of Biochemistry*, vol. 57, no. 1, pp. 69–97, 1988.
- [48] R. J. Grand, "Acylation of viral and eukaryotic proteins," *The Biochemical Journal*, vol. 258, no. 3, pp. 625–638, 1989.
- [49] A. Kishimoto, K. Nishiyama, H. Nakanishi et al., "Studies on the phosphorylation of myelin basic protein by protein kinase C and adenosine 3': 5'-monophosphate-dependent protein kinase," *The Journal of Biological Chemistry*, vol. 260, no. 23, pp. 12492–12499, 1985.
- [50] H. Brüßow, "The Novel Coronavirus - A Snapshot of Current Knowledge," *Microbial Biotechnology*, vol. 13, no. 3, pp. 607–612, 2020.
- [51] S. Agnihothram, B. L. Yount Jr., E. F. Donaldson et al., "A mouse model for Betacoronavirus subgroup 2c using a bat coronavirus strain HKU5 variant," *mBio*, vol. 5, no. 2, 2014.
- [52] W. Li, S. K. Wong, F. Li et al., "Animal origins of the severe acute respiratory syndrome coronavirus: insight from ACE2-S-protein interactions," *Journal of Virology*, vol. 80, no. 9, pp. 4211–4219, 2006.
- [53] L. F. Wang and B. T. Eaton, "Bats, civets and the emergence of SARS," *Current Topics in Microbiology and Immunology*, vol. 315, pp. 325–344, 2007.
- [54] J. Sun, W. T. He, L. Wang et al., "COVID-19: epidemiology, evolution, and cross-disciplinary perspectives," *Trends in Molecular Medicine*, vol. 26, no. 5, pp. 483–495, 2020.
- [55] T. T. Lam, N. Jia, Y. W. Zhang et al., "Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins," *Nature*, vol. 583, no. 7815, pp. 282–285, 2020.
- [56] L. R. Lopes, C. G. de Mattos, and P. B. Paiva, "Molecular evolution and phylogenetic analysis of SARS-CoV-2 and hosts ACE2 protein suggest Malayan pangolin as intermediary host," *Brazilian Journal of Microbiology*, pp. 1–7, 2020.
- [57] R. Touati, S. Haddad-Boubaker, I. Ferchichi et al., "Comparative genomic signature representations of the emerging COVID-19 coronavirus and other coronaviruses: High identity and possible recombination between Bat and Pangolin coronaviruses," *Genomics*, vol. 112, no. 6, pp. 4189–4202, 2020.
- [58] J. Brezovsky, P. Babkova, O. Degtjarik et al., "Engineering a de novo transport tunnel," *ACS Catalysis*, vol. 6, no. 11, pp. 7597–7610, 2016.
- [59] K. Halpin, P. L. Young, H. E. Field, and J. S. Mackenzie, "Isolation of Hendra virus from pteropid bats: a natural reservoir of Hendra virus," *The Journal of General Virology*, vol. 81, no. 8, pp. 1927–1932, 2000.
- [60] J. M. Yob, H. Field, A. M. Rashdi et al., "Nipah virus infection in bats (order Chiroptera) in peninsular Malaysia," *Emerging Infectious Diseases*, vol. 7, no. 3, pp. 439–441, 2001.
- [61] B. Hu, X. Ge, L. F. Wang, and Z. Shi, "Bat origin of human coronaviruses," *Virology journal*, vol. 12, no. 1, p. 221, 2015.
- [62] D. Vijaykrishna, G. J. Smith, J. X. Zhang, J. S. Peiris, H. Chen, and Y. Guan, "Evolutionary insights into the ecology of coronaviruses," *Journal of Virology*, vol. 81, no. 8, pp. 4012–4020, 2007.
- [63] C. A. de Haan, P. S. Masters, X. Shen, S. Weiss, and P. J. Rotter, "The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host," *Virology*, vol. 296, no. 1, pp. 177–189, 2002.
- [64] S. R. Weiss and S. Navas-Martin, "Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus," *Microbiology and molecular biology reviews*, vol. 69, no. 4, pp. 635–664, 2005.