COVID

MDPI

Review

The Potential Functions of Protein Domains during COVID Infection: An Analysis and a Review

Haitham Sobhy [1,2]

1   Laboratory of Dynamics Host-Pathogen Interactions (DHPI), IUT Louis Pasteur, University of Strasbourg, 67300 Schiltigheim, France; haitham.sobhy@colorado.edu or haithamsobhy@gmail.com
2   Department of Biochemistry, JSC Biotechnology Building, University of Colorado Boulder, Campus Box 596, Boulder, CO 80303-0596, USA

Abstract: Coronaviruses (CoVs) are a large viral family that can evolve rapidly emerging new strains that cause outbreaks and life-loss, including SARS-CoV, MERS-CoV, and SARS-CoV-2 (COVID-19). CoVs encode a diverse number of proteins, ranging from 5 proteins in bat CoV, to 14 in SARS CoV, which could have implication on viral tropism and pathogenicity. Here, we highlight the functional protein motifs (domains) that could contribute in the coronavirus infection and severity, including SARS-CoV-2. For this role, we used the experimentally validated domain (motif) datasets that are known to be crucial for viral infection. Then, we highlight the potential molecular pathways and interactions of SARS-CoV-2 proteins within human cells. Interestingly, the C-terminal of SARS-CoV-2 nsp1 protein encodes MREL motif, which a signature motif of the tubulin superfamily, and regulate tubulin expression. The C-terminal region of nsp1 protein can bind to ribosome and regulation viral RNA translation.

Keywords: protein functional domains; short linear motifs; coronaviruses; COVID-19; severe acute respiratory syndrome-related coronavirus; 2019-nCoV; virus outbreak; tubulin autoregulation motif

1. Introduction

The coronavirus outbreak (coronavirus disease-19, COVID-19) thought to be initiated by a zoonotic virus that transmitted to human. Genome sequencing reveals that the causative virus is named severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2, or SARS-2 for short), and belongs to genus *Betacoronavirus*, family *Coronaviridae* [1–4]. SARS-2 shares sequence homology with coronaviruses infecting bats and murine, such as bat SARS (bat-SL-CoVZC45), and classified in so-called subgenus *Sarbecovirus*. Furthermore, the first SARS-CoV (SARS-1) outbreak occurred in 2003, from zoonotic origin as well. The phylogenetic analysis shows a close relationship with bat coronavirus (e.g., BtCoV_279/2005 and BtRs-BetaCoV/YN2018). In 2012, another zoonotic virus was transmitted to human, so-called Middle East respiratory syndrome coronavirus (MERS-CoV). The virus belongs to subgenus *Merbecovirus*, and shares sequence homology with bat CoV HKU4, HKU5, and *Erinaceus* CoV. The third subgenus *Embecovirus*, which includes human CoV HKU1 and OC43 and share homology with murine CoV MHV-1 and betacoronavirus HKU24.

The phylogenetic analysis shows coronaviruses are conserved within the same subgroup [1]. However, coronaviruses encode diverse number of proteins, ranging from 5 protein in bat CoV, to 14 in SARS CoV, Supplementary Table S1. They evolve rapidly emerging new strains that cause outbreaks and life-loss. Although the high rate of homology of CoV genomes, mutation of one or more nucleotides may lead to significant changes on the short protein motifs or domains. Particularly, the newly isolated viruses (despite of evolutionary relationships with other CoVs) may encode new proteins of unknown functions and utilize different molecular interactions and pathways within the

host cell [5–9]. These domains could contribute in viral virulence and ability to infect wide-range of host cells.

2. Materials and Methods

Here, we highlight the role of functional protein domains and the potential pathways that could be triggered during CoVs infection. To perform this analysis, the full protein sequences of 32 coronaviruses, including SARS-CoV and MERS-CoV, in addition to 10 protein sequences of the newly isolated SARS-CoV-2 are obtained from NCBI database, during 2020, listed in Supplementary Table S1. It worth to note that for precise genome coordination, only sequences that are reviewed and annotated in UniProt database were selected. Additionally, some proteins are small to detect functional or structural motifs; therefore, these proteins are excluded.

The functional motifs in the coronavirus proteomes identified using exact text-finding (mining) implemented in Shetti and Shetti-Motif tools, for detailed method [6]. For short, we used the datasets of functional motifs that experimentally validated to be involved in viral infection and virulence, see reviews [7–9]. Additional protein domains are downloaded from PROSITE database (https://prosite.expasy.org/ (accessed on 15 April 2020)). Together these construct a dataset of over 1500 experimentally validated motifs and domains, Supplementary Table S2. It worth to note that, the pattern of the proteins motifs are automatically listed and loaded into Shetti-Motif tool. The whole proteomes were loaded to the tool as well. The tool searches for the motifs within the whole proteome of CoVs [6]. The output of the tool is collected, which is then used to construct a table contains columns with CoVs proteomes and the rows represent the pattern (motif), in addition to the number of occurrence of each pattern. If a protein encode (harbor) multiple instance of a pattern, it considered as one instance, Table S3. If the protein does not encode this motif, the cell filled with zero. Then, the number of motif-containing proteins normalized to number of the total proteins in the proteome (in percent to the total number of proteins). The results are shown as a matrix of proteomes and number of motifs-containing proteins (in percent) in each proteome, Tables S3–S5. Details of the motifs-containing proteins, protein names and locations are shown in Tables S6–S8.

Secondly, we aimed to visualize the differences between SARS-CoV-2 and its evolutionary closely relative viruses, such as bat CoV, human CoV, SARS-CoV, and MER-CoV. Together, it constructs a matrix of 17 proteomes. We collected the number of occurrence of the motifs in each of these 17 proteomes, Table S4. The matrix consists of motif-proteome enrichment (i.e., motifs as rows and their number of occurrence in each proteome represented as columns). The data clustered using hierarchical clustering and the heat-map is constructed using the default settings (MeV tool, http://mev.tm4.org/ (accessed on 15 April 2020)). The heat-map clusters the viruses that harbor similar motif patterns, as well as it visualizes the number of occurrence in each proteome. In principle, the viruses encode similar motifs (domains), could share the same viral tropism and molecular pathways. Finally, as a proof of concept, we visualize structure of some of the viral protein-containing domain bind to host proteins.

3. Results

3.1. Coronaviruses Encode Wide-Range of Functional Motifs

In the first part, we aimed to highlight the different type of protein motif (domain) each of coronavirus, which could help to identify the domain(s) that contribute in viral virulence, as shown in method and Figure 1A. The results suggests that CoVs encode wide-range of functional motifs that differ from one virus to another; the closely related viruses encode different motifs, Tables S3–S5. Clustering based on the number of occurrence of the motifs in the proteome could reveal the potential virus-host protein interactions. In addition, viruses encode similar motifs could share similar viral tropism. We found that SARS-CoV-2 and bat-SL-CoVZC45 clustered in one clade close to SARS-CoV-1, Figure 1B,C. This finding is consistent with the fact that SARS-CoV 1 and 2 are evolutionary closely

related, infect the same host, and could transmitted from the same zoonotic animal. Details on these motifs and proteins found in Tables S3–S8.
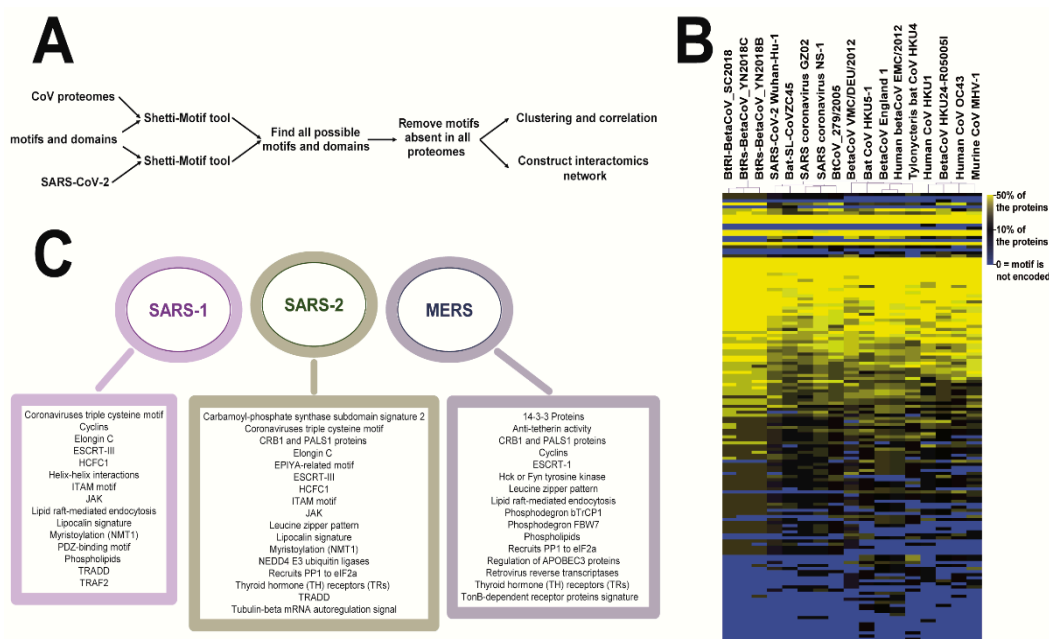
Figure 1. (**A**) The pipeline used to identify the motifs and counting number of motifs per proteome. (**B**) The heatmap based on hierarchical clustering represents the occurrence of motifs (rows) in CoV species (column). The color scale is shown, the blue color is 0%, i.e., absent of the motif in this species; whereas, the yellow color means one or more proteins harbor at least one instance of this motif. The numbers represent the percent of proteins normalized to total number of proteins in the proteome. The heatmap constructed from Supplementary Table S4. (**C**) The list of potential pathways that could be triggered SARS-CoV, SARS-CoV-2, and MERS-CoV.

Interestingly, although SARS-CoV-1 and 2 are evolutionary-related, they encode some different motifs, which suggest that they have different viral tropism within the host cells. For example, two motifs are required for cell signaling transduction, which can recognize TRAF2 protein (PxQxT motif) or PDZ domain-containing proteins (KTxxx[W/I]), where x means any residue, and [W/I] or [WI] means W or I residue. These two motifs are deleted in SARS-CoV-2, but encoded in the closely related bat-SL-CoVZC45 virus. However, SARS-CoV-2 encodes multiple domains that can recognize ubiquitination proteins, such as E3 ubiquitin ligases (E3 Ub), Elongin C (ELOC), TRAF6, and SIAH1 (SLxxxLxxxI, PxExxE or PxAxV motifs). Noting that the closely related viruses does not harbor these motifs. Additionally, the canonical PPxY motif is largely encoded and utilized by viruses to hijack the cellular machinery, reviewed [7]. PPxY motif is needed to recruit NEDD4 E3 ubiquitin ligases for protein degradation, and endosomal sorting complexes required for the transport (ESCRT) pathway. ESCRT pathway is crucial for budding of HIV-1 and paramyxoviruses and exit from the cell. Additionally, adenoviruses utilize PPxY motif during cell entry and cellular trafficking. In coronaviruses, surface glycoprotein (S) of SARS-CoV-2 harbor PPxY motif. Two proteins of MERS-CoVs and three proteins of *Erinaceus* CoV harbor PPxY motif, whereas other human CoVs and SARS-CoV-1 do not encode these motifs, Table S3.

Coronaviruses encode other canonical ESCRT-interacting motifs, such as P[T/S]AP, [F/I/L/V]PxV, YxxL, and LYPxL. Although ORF1ab polyprotein of SARS-2 harbors LYPTL, LPGV and VPFV motifs, the virus does not encode P[T/S]AP motif, which is encoded only by MERS-CoV EMC/2012, Table S3. P[T/S]AP motif recruits TSG101 protein, a component of ESCRT-I, whereas LYPxnL recruits ALIX, a component of ESCRT-III complex, reviewed in [7,10–14]. In fact, quinolones antiviral therapeutics can target ESCRT pathway and viral budding, such as FGI-104, FGI-103, FGI-106, and chloroquine.

SARS-1 and 2, but not MERS encode motif that recognizes host cell factor 1 (HCFC1), which is crucial to regulate the cell cycle. Additionally, two Cys-rich motifs are predicted to link between spike and envelope proteins [15]. SARS-1 and 2, but not MERS subgroup encode these two motifs. Other C-rich motifs, which are needed for baculovirus virions production and nucleocapsid assembly, are encoded by multiple coronaviruses, Table S3. On the other hand, coronaviruses harbor multiple integrin-binding (RGD) motif. In absence of RGD, viruses may utilize other motifs to attach to cellular receptors and enter into host cells, such as KGE, LDV, LDI, and SDI, Table S3.

3.2. The Potential CoV-Human Protein Interactions Pathways Based on Functional Motif

We used the functional motif encoded by each of SARS-CoV-2 proteins to predict the potential molecular interactions and pathways within the host cell. For this part of analysis, we searched for the motif pattern within CoV proteomes using Shetti-Motif tool, as described before. Then, we constructed a matrix of SARS-2 proteins (columns) versus the motifs and their number of occurrence (rows), Table S9. The table used to construct a schematic diagram (manually curated), which represents SARS-CoV-2 proteins, the motifs encoded, and the potential interactions that validated to be utilized during viral infections, Figure 2. We observed that SARS-2 encodes multiple Ub-, and SUMO-binding motif, including the PPxY motif, which is localized on surface (S) protein. Recruiting ubiquitin proteins are essential to degrade antiviral proteins and hijack the immune response imposed by the cells. SARS-2 encodes motifs to recognize heparan sulfate (HS), which is required for post-internalization events of viral entry, discussed in [7]. Lung epithelium and endothelium are covered with a layer of heparin sulfate (closely related in structure to heparan), glycoproteins and glycolipids, so-called endothelial glycocalyx. Our results support the hypothesis that SARS-CoV-2 could adhere to heparin sulfate of glycocalyx layer, which could potential drug target [16,17]. As mentioned above ORF1ab harbors MREI-like tubulin motif, which will be discussed later. Besides, it harbors the canonical motif for protein trafficking and nuclear localization signal (NLS), which is essential for trafficking through the nuclear membrane.

On the other hand, furin endoprotease belongs to group of proprotein convertases that cleave the precursor proteins to the activated form. It binds to the canonical motif RxRK/R||x, where || denotes the cleavage site. Coronaviruses encode R||S motif, such as RRRR||S motif, which interact with furin leading to proteolytic activation of the spike protein and viral entry into host cells [18]. In addition, the R||S motif is essential for syncytium formation [18]. ORF1ab is the largest polyprotein encoded by SARS-2, which is auto-proteolytically processed into 16 non-structural proteins (nsp1 to nsp16), reviewed in [4], Figure 3A. ORF1ab harbors at least seven $Rx_{0-3}RS$ motifs, which could correspond to the cleavage of the polyprotein to the non-structural proteins.

SARS-2 harbors clathrin-binding motifs and clathrin adaptor protein (AP)-binding motifs, which are required for endocytosis. Coronaviruses enter the cells by fusion or endocytosis, which may require clathrin for some strains, reviewed in [19]. The integrins (ITGs) and heparan sulfates could have role in the entry into host cell. Noteworthy, HIV-1 utilizes AP-binding motifs to direct anti-tetherin (BST2) to the lysozyme and antagonizes the antiviral immune response, reviewed in [7]. Regarding the cellular signaling, ORF1b polyprotein harbors motifs involved in multiple cellular signaling, including JAK, MAPK, TRADD, TRAF6, and caspases-binding motifs. Caspases and TRAFs are linked with inflammation and apoptosis [20,21], therefore multiple viruses (e.g., herpesviruses and influenza) hijack the caspase pathways to regulate the programmed cell death.
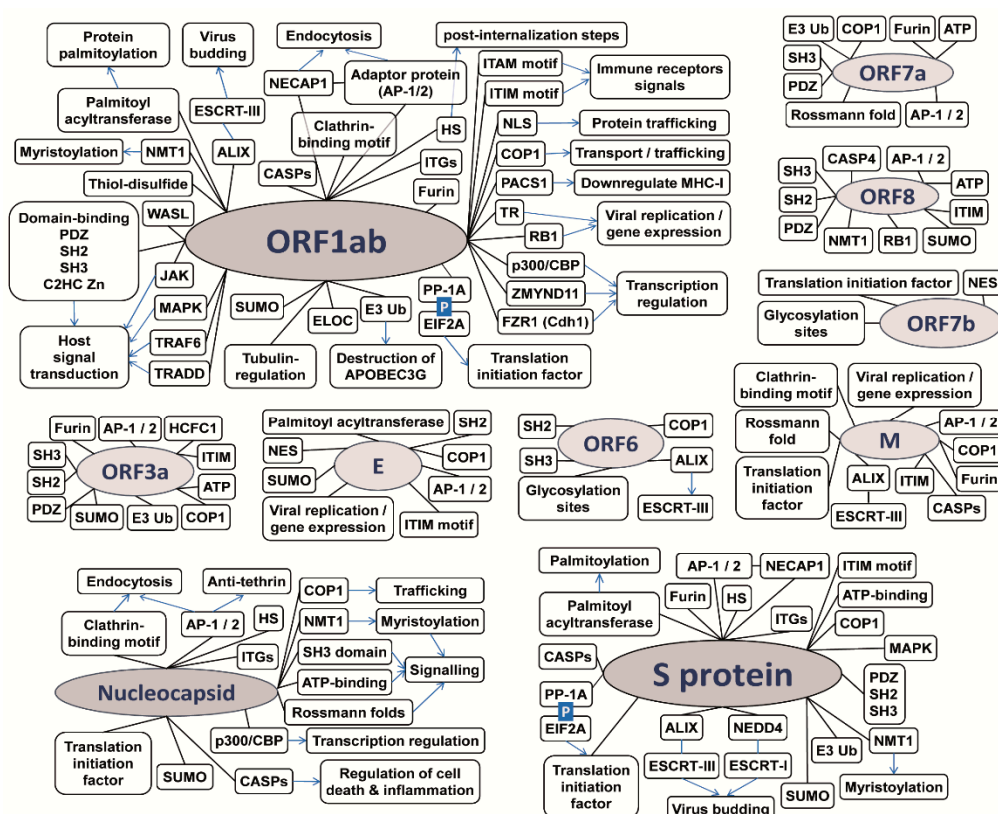
Figure 2. Schematic diagram shows SARS-CoV-2 proteins. The black arrows refer to encoded motif (such as PDZ or SH2/3-binding motifs), while blue lines refer to the potential function/pathways of the motif. Abbreviations: ALIX = programmed cell death 6-interacting protein; Anti-tetherin = motif antagonizes the bone marrow stromal antigen 2 (BST2); AP-1/2 = adaptor protein complex AP-1 or AP-2; APOBEC3G = apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G; ATP-binding = Walker's motifs (ATP-binding motif); C2HC Zn = motif recognizes the CysCysHisCys (C2HC) type zinc finger domain; CASPs = caspases; COP1 = E3 ubiquitin-protein ligase COP1; E3 Ub = E3 ubiquitin ligases; EIF2A = eukaryotic translation initiation factor 2A; ELOC = Elongin C; ESCRT = endosomal sorting complexes required for the transport; FZR1 (Cdh1) = Fizzy-related protein homolog; HCFC1 = host cell factor 1; HS = heparan sulfate; ITGs = integrins; ITAM = immunoreceptor tyrosine-based activation motif; ITIM = immunoreceptor tyrosine-based inhibition motif; JAK = tyrosine-protein kinase; MAPK = mitogen-activated protein kinase; NECAP1 = adaptin ear-binding coat-associated protein 1; NEDD4 = E3 ubiquitin-protein ligase NEDD4; NES = nuclear export signal; NLS = nuclear localization signal; NMT1 = Glycylpeptide N-tetradecanoyltransferase 1; P in black box = phosphorylation; p300/CBP = histone acetyltransferase/CREB-binding protein; PACS1 = phosphofurin acidic cluster sorting protein 1; PDZ = motif recognizes the PDZ-domain, post-synaptic density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1), and zonula occludens-I protein (zo-1); PP-1A = serine/threonine-protein phosphatase PP1-alpha catalytic subunit; RB1 = retinoblastoma-associated protein; SH2 and SH3 = motif recognizes the SRC Homology 2 and 3 domain; SUMO = small ubiquitin-related modifier binding motif; TR = thyroid hormone (TH) receptors (TRs); TRADD = tumor necrosis factor receptor type 1-associated DEATH domain protein; TRAF6 = TNF receptor-associated factor 6; WASL = WAS/WASL-interacting protein family member 1; ZMYND11 = zinc finger MYND domain-containing protein 11.

Among the interesting motifs, ORF1ab harbors the canonical motif for binding with palmitoyl acyltransferase, in addition to multiple thiol disulphide and Cys-rich motifs, which are needed for protein palmitoylation [22]. Noteworthy, the envelope (E) protein of some strains of coronaviruses are shown to be palmitoylated [23–25]. Myristoylation is another lipid post-translational modification event. ORF1ab harbors the canonical MGxxxS motif for binding with N-myristoyltransferase (NMT1), which adds a myristoyl group to the N-terminal glycine residue of the proteins [26]. Noteworthy, myristoylation is crucial for egress of some viruses, including coronaviruses [27], as well as activation or inhibition of the immune response by phosphorylation of tyrosine residues in ITAM and

ITIM (immunoreceptor tyrosine-based activation and inhibition motifs, respectively) [26]. SARS-2 proteins encode both ITAM and ITIM motifs.
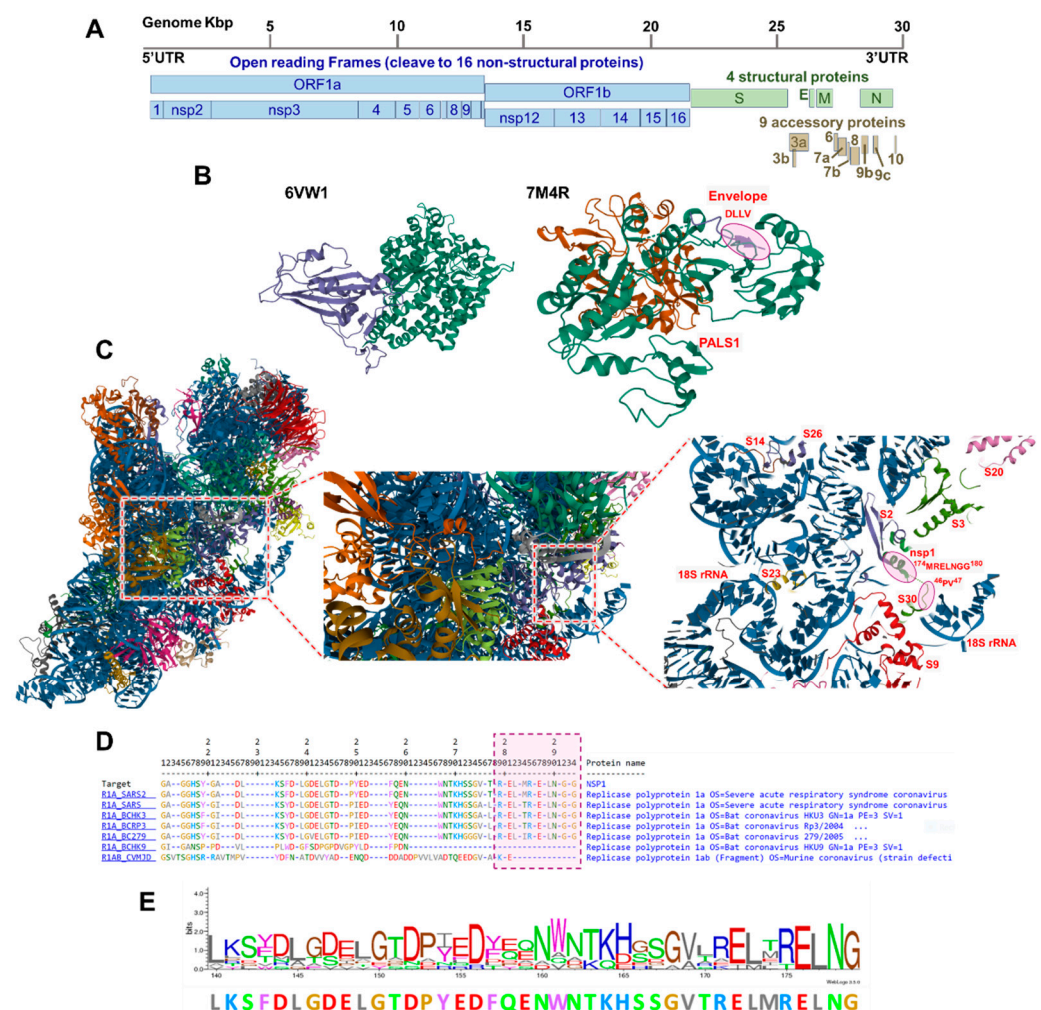
Figure 3. (**A**) Schematic diagram of the COVID genome and protein encoded. (**B**) Two example of viral—host cell protein interactions, through SH2, SH3 and PDZ domains. The surface spike protein binds to human ACE2 protein (PDB ID: 6VW1). Spike protein harbor multiple motifs the confer SH2- and PDZ domain binding, e.g., TGV, TSV, FLGV, GIGV, SNVYA, PSVYA, and QPYRVVVL. Additionally, envelope protein (**E**) harbors DLLV motif that binds to host PALS1 protein (PDB: 7M4R). (**C**) SARS-CoV-2 nsp1 protein binds to S40 ribosome complex (PDB: 6ZLW). (**D**) Sequence alignment shows that MRELNGG motif are conserved within SARS-CoV-2, but not other coronaviruses. (**E**) The sequence logo and consensus sequences of nsp1 C-terminus of coronaviruses. The figures (**A**–**C**) adopted from [28–32]; (**D**–**E**) from EBI portal (https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/VarSite/GetPage.pl?home=TRUE, last accessed on 30 August 2021, UniProt ID: P0DTD1_1).

SARS-2 proteins harbor multiple domain-interacting motifs, such as motifs recognizing PDZ, SH2, SH3 domains [29–31,33], Figure 3B. These motifs are essential for cellular signaling and protein trafficking within host cells. For example, the PDZ-binding and DLLV motifs in SARS-1 E protein can influence the subcellular localization of PALS1 (MPP5), which may disrupt the tight junction and apicobasal polarity of the cell [30,31,34], Figure 3B. The L-rich and PDZ-binding motifs are used by multiple viruses to recruits mTORC1, kinase signaling to initiate translation [7]. In SARS-2, S, N, M, ORF1ab, and ORF7b proteins harbor additional L-rich motifs, Table S9. ORF7b harbors LxxLL motif, which is crucial for HIV retro-transposition, and papillomavirus-induced oncogenesis and

cell transformation. Additional motifs such as [RKY]xxPxxP or RxxK can interact with host SH3 leading to endosome sorting. For example, HIV and HCV interact with tyrosine kinase through the SH2 and SH3 domains-interacting motifs, e.g., PxxPxR [8,9]. Although some human CoVs and MERS-CoV encode PxxPxR or [RKY]xxPxxP motifs, SARA-CoV-1 and 2 do not encode the same motifs.

An interesting phenomenon in viruses is the ability of a virus to disturb the host transcription for the sake of the viral gene expression. An example is the adenovirus E1A oncoprotein that regulates host transcription by binding to transcription regulators, histone acetyltransferase/CREB-binding protein (p300/CBP), through the Fx[DE]xxxL motif [35,36]. Similarly, the adenoviral E1A and the Epstein–Barr virus EBNA2 oncoproteins interact with the C-terminal Mynd domain of ZMYND11 (BS69) transcription regulator, which is facilitated by PxLxP motif in the viral proteins [37]. The transcription regulation Fx[DE]xxxL and PxLxP motifs are observed on ORF1ab and N protein sequences (but not ORF1a), suggesting their potential roles in virus replication. Furthermore, oncoproteins with conserved LxCxE motif can phosphorylate and inactivate retinoblastoma protein (RB1), leading to initiation of the gene expression and virus replication, reviewed in [7]. Although the motif is a five residues long, ORF1ab is the only protein that harbors this motif. An additional contributor on viral replication is ORF3a protein, which harbors the HCFC1-binding motif. HCFC1 regulates the cell cycle by recruiting the regulator p300 and histone deacetylase (HDACs).

3.3. The SARS-CoV-2 Nsp1 Protein Encodes for a Tubulin MREL Motif

ORF1ab polyprotein is a long protein that cleaved in host cells into 16 non-structural proteins (nsp1 to nsp16). Among all coronaviruses, only nsp1 of SARS-2 harbors the tubulin-beta mRNA autoregulation signal motif (PROSITE accession: PS00228) in the C-terminal. Tubulins are able to auto-regulate their expression through the binding between the polymerized tubulin protein to the MREI motif of the nascent tubulin peptides, which, by an unknown mechanism, can terminate the translation of tubulin [38]. The recent studies suggest that the polymerized tubulin may bind to an unknown mediator or ribonucleases, which cause ribosome stalling and terminate translation, causing constant level of the tubulin in the cells [38]. The exact biochemical and structural mechanism by which the tubulins recognize the motif and regulate the translation is remain to be discovered.

It worth to note that all tubulins harbor MR[E/D][I/L] motif in the N-terminal end of the tubulin, which thought to be a signature of tubulin superfamily. The motif tends to be found in few proteins, for example it is detected in 219 proteins belong to tubulin family, among them 11 are human tubulins. The motif can be found in 269 other proteins (do not belong to tubulins), among them seven proteins are encoded by human, including the RNA-binding TNRC6A, the tyrosine-protein kinase MUSK, and the phospholipid phosphatase PLPP4 proteins.

To test the hypothesis that this motif is ubiquitous in viruses, we searched for the tubulin motif in herpesvirus, iridovirus, and poxvirus proteins. The search shows that some of these large viruses (the genome sizes are 10–20 times larger than CoV) do not encode the tubulin motif, while some of these large proteomes harbor only one version. Together, this shows that the motif is not widely encoded by human nor viruses. Moreover, the motif could have additional regulatory functions other than regulation of tubulin expression, for example the motif could have a roles in regulation of translation.

On the other hand, nsp1 protein is thought to have a role in viral RNA translation. It can bind to 40S ribosome complex to inhibit host RNA translation via endonucleolytic cleavage at the 5′UTR [32,39]. Structure analysis shows the binding between MREL domain with the P residue of S30 protein, Figure 3C. Interestingly, the MRELNGG is conserved among some coronaviruses, including SARS-CoV-2. However, the closely-related coronaviruses, such as SARS-CoV and MERS-CoV encode T, L, or I residues, instead of M, Figure 3D,E. The function of C-terminal domain and MREL tubulin-motif, and their roles in viral virulence deserve to be studied by future studies.

4. Discussion

This analysis highlights the fact that coronaviruses encode wide-range of motif that could help virus to trigger new molecular functions and interactions. Therefore, the protein domain/motifs datasets could help future research to discover new aspects about MERS, SARS-1, and SARS-2 infection. Furthermore, we used functional motifs to predict the potential interactions SARS-CoV-2 proteins, Figure 2.

In consistent with our analysis, a recent interatomics study shows that SARS-CoV-2 proteins interacts with ubiquitin ligases, kinases, lipid modifications, as well as proteins contain zinc finger, SH2, SH3, and PDZ domains [28], Supplementary Table S10. Additionally, PDZ-binding motifs (PBM) is crucial during SARS-CoV infection. To study the implication of PBM on infection, PBM located in SARS-1 E protein was mutated and deleted [40]. Astonishingly, the virus restored the motif after several passages in vitro or in mice, however mutated PBM in nsp1 protein leads to virus attenuation. We noticed in our results that a virus could harbor multiple copies of the same motif, such as PDZ-binding motifs and SH2 and SH3 domains, which may function in the severity of the virus. Noteworthy, ORF1ab is the largest protein and it contains almost a copy of all binding motifs. One could suggest that multiple copies of the same motifs are kept on ORF1ab protein for restoration of these motifs in case of their drastic loss; however, this possibility remains to be validated.

Finally, the advantage of our in silico analysis is the usage of dataset of the motifs (domains) that have been experimentally validated by multiple methods for other viruses, which increases the robustness, discussed in [5–7]. The resulting datasets can make the future functional studies easy and can enrich the attempts to understand coronaviruses (e.g., COVID-19) infection and the attempts to find antiviral drug. Most of the domains used in the analysis are confer to a structure that increases their importance during binding with protein domains of host cells. Interestingly, viruses could acquire the motifs from an evolutionary distant virus or organism. It is of interest to study the molecular mechanisms govern the transfer of the protein motifs and domains, which helps to predict the future and emerging pathogens.

References

1.   Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N.; et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet* **2020**, *395*, 565–574. [CrossRef]

2.   Narayanan, K.; Huang, C.; Makino, S. Sars coronavirus accessory proteins. *Virus Res.* **2008**, *133*, 113–121. [CrossRef] [PubMed]

3.   Schoeman, D.; Fielding, B.C. Coronavirus envelope protein: Current knowledge. *Virol. J.* **2019**, *16*, 69. [CrossRef] [PubMed]

4.   Song, Z.; Xu, Y.; Bao, L.; Zhang, L.; Yu, P.; Qu, Y.; Zhu, H.; Zhao, W.; Han, Y.; Qin, C. From sars to mers, thrusting coronaviruses into the spotlight. *Viruses* **2019**, *11*, 59. [CrossRef]

5.   Sobhy, H. Virophages and their interactions with giant viruses and host cells. *Proteomes* **2018**, *6*, 23. [CrossRef] [PubMed]

6.   Sobhy, H. A bioinformatics pipeline to search functional motifs within whole-proteome data: A case study of poxviruses. *Virus Genes* **2017**, *53*, 173–178. [CrossRef] [PubMed]

7.   Sobhy, H. A review of functional motifs utilized by viruses. *Proteomes* **2016**, *4*, 3. [CrossRef]

8.   Davey, N.E.; Trave, G.; Gibson, T.J. How viruses hijack cell regulation. *Trends Biochem. Sci.* **2011**, *36*, 159–169. [CrossRef] [PubMed]

9.   Hraber, P.; O'Maille, P.E.; Silberfarb, A.; Davis-Anderson, K.; Generous, N.; McMahon, B.H.; Fair, J.M. Resources to discover and use short linear motifs in viral proteins. *Trends Biotechnol.* **2020**, *38*, 113–127. [CrossRef] [PubMed]

10.   Williams, R.L.; Urbe, S. The emerging shape of the escrt machinery. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 355–368. [CrossRef]

11.   Sette, P.; Jadwin, J.A.; Dussupt, V.; Bello, N.F.; Bouamr, F. The escrt-associated protein alix recruits the ubiquitin ligase nedd4-1 to facilitate hiv-1 release through the lypxnl l domain motif. *J. Virol.* **2010**, *84*, 8181–8192. [CrossRef]

12.   Han, Z.; Madara, J.J.; Liu, Y.; Liu, W.; Ruthel, G.; Freedman, B.D.; Harty, R.N. Alix rescues budding of a double ptap/ppey l-domain deletion mutant of ebola vp40: A role for alix in ebola virus egress. *J. Infect. Dis.* **2015**, *212*, S138–S145. [CrossRef]

13.   Wolff, S.; Ebihara, H.; Groseth, A. Arenavirus budding: A common pathway with mechanistic differences. *Viruses* **2013**, *5*, 528–549. [CrossRef]

14.   Conrad, K.P. Might proton pump or sodium-hydrogen exchanger inhibitors be of value to ameliorate SARS-CoV-2 pathophysiology? *Physiol. Rep.* **2021**, *8*, e14649. [CrossRef]

15.   Wu, Q.; Zhang, Y.; Lu, H.; Wang, J.; He, X.; Liu, Y.; Ye, C.; Lin, W.; Hu, J.; Ji, J.; et al. The e protein is a multifunctional membrane protein of SARS-CoV. *Genom. Proteom. Bioinform.* **2003**, *1*, 131–144. [CrossRef]

16.   Okada, H.; Yoshida, S.; Hara, A.; Ogura, S.; Tomita, H. Vascular endothelial injury exacerbates coronavirus disease 2019: The role of endothelial glycocalyx protection. *Microcirculation* **2020**, *28*, e12654. [CrossRef]

17.   Clausen, T.M.; Sandoval, D.R.; Spliid, C.B.; Pihl, J.; Perrett, H.R.; Painter, C.D.; Narayanan, A.; Majowicz, S.A.; Kwong, E.M.; McVicar, R.N.; et al. SARS-CoV-2 infection depends on cellular heparan sulfate and ace2. *Cell* **2020**, *183*, 1043–1057.e1015. [CrossRef] [PubMed]

18.   Yamada, Y.; Liu, D.X. Proteolytic activation of the spike protein at a novel rrrr/s motif is implicated in furin-dependent entry, syncytium formation, and infectivity of coronavirus infectious bronchitis virus in cultured cells. *J. Virol.* **2009**, *83*, 8744–8758. [CrossRef] [PubMed]

19.   Belouzard, S.; Millet, J.K.; Licitra, B.N.; Whittaker, G.R. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* **2012**, *4*, 1011–1033. [CrossRef] [PubMed]

20.   Man, S.M.; Karki, R.; Kanneganti, T.D. Molecular mechanisms and functions of pyroptosis, inflammatory caspases and inflammasomes in infectious diseases. *Immunol. Rev.* **2017**, *277*, 61–75. [CrossRef]

21.   Nainu, F.; Shiratsuchi, A.; Nakanishi, Y. Induction of apoptosis and subsequent phagocytosis of virus-infected cells as an antiviral mechanism. *Front. Immunol.* **2017**, *8*, 1220. [CrossRef]

22.   Sobocinska, J.; Roszczenko-Jasinska, P.; Ciesielska, A.; Kwiatkowska, K. Protein palmitoylation and its role in bacterial and viral infections. *Front. Immunol.* **2017**, *8*, 2003. [CrossRef]

23.   Corse, E.; Machamer, C.E. The cytoplasmic tail of infectious bronchitis virus e protein directs golgi targeting. *J. Virol.* **2002**, *76*, 1273–1284. [CrossRef] [PubMed]

24.   Liao, Y.; Yuan, Q.; Torres, J.; Tam, J.P.; Liu, D.X. Biochemical and functional characterization of the membrane association and membrane permeabilizing activity of the severe acute respiratory syndrome coronavirus envelope protein. *Virology* **2006**, *349*, 264–275. [CrossRef]

25.   Lopez, L.A.; Riffle, A.J.; Pike, S.L.; Gardner, D.; Hogue, B.G. Importance of conserved cysteine residues in the coronavirus envelope protein. *J. Virol.* **2008**, *82*, 3000–3010. [CrossRef]

26.   Udenwobele, D.I.; Su, R.C.; Good, S.V.; Ball, T.B.; Varma Shrivastav, S.; Shrivastav, A. Myristoylation: An important protein modification in the immune response. *Front. Immunol.* **2017**, *8*, 751. [CrossRef] [PubMed]

27.   Du, Y.; Zuckermann, F.A.; Yoo, D. Myristoylation of the small envelope protein of porcine reproductive and respiratory syndrome virus is non-essential for virus infectivity but promotes its growth. *Virus Res.* **2010**, *147*, 294–299. [CrossRef] [PubMed]

28.   Lechuga, G.C.; Souza-Silva, F.; Sacramento, C.Q.; Trugilho, M.R.O.; Valente, R.H.; Napoleão-Pêgo, P.; Dias, S.S.G.; Fintelman-Rodrigues, N.; Temerozo, J.R.; Carels, N.; et al. SARS-CoV-2 proteins bind to hemoglobin and its metabolites. *Int. J. Mol. Sci.* **2021**, *22*, 9035. [CrossRef]

29.   Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **2020**, *581*, 221–224. [CrossRef]

30.   Javorsky, A.; Humbert, P.O.; Kvansakul, M. Structural basis of coronavirus e protein interactions with human pals1 pdz domain. *Commun. Biol.* **2021**, *4*, 724. [CrossRef]

31.   Chai, J.; Cai, Y.; Pang, C.; Wang, L.; McSweeney, S.; Shanklin, J.; Liu, Q. Structural basis for SARS-CoV-2 envelope protein recognition of human cell junction protein pals1. *Nat. Commun.* **2021**, *12*, 3433. [CrossRef]

32. Teoh, K.T.; Siu, Y.L.; Chan, W.L.; Schluter, M.A.; Liu, C.J.; Peiris, J.S.; Bruzzone, R.; Margolis, B.; Nal, B. The sars coronavirus e protein interacts with pals1 and alters tight junction formation and epithelial morphogenesis. *Mol. Biol. Cell* **2010**, *21*, 3838–3852. [CrossRef]

33. Pelka, P.; Ablack, J.N.; Fonseca, G.J.; Yousef, A.F.; Mymryk, J.S. Intrinsic structural disorder in adenovirus e1a: A viral molecular hub linking multiple diverse processes. *J. Virol.* **2008**, *82*, 7252–7263. [CrossRef] [PubMed]

34. Ferreon, J.C.; Martinez-Yamout, M.A.; Dyson, H.J.; Wright, P.E. Structural basis for subversion of cellular control mechanisms by the adenoviral e1a oncoprotein. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13260–13265. [CrossRef] [PubMed]

35. Ansieau, S.; Leutz, A. The conserved mynd domain of bs69 binds cellular and oncoviral proteins through a common pxlxp motif. *J. Biol. Chem.* **2002**, *277*, 4906–4910. [CrossRef] [PubMed]

36. Gordon, D.E.; Jang, G.M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K.M.; O'Meara, M.J.; Rezelj, V.V.; Guo, J.Z.; Swaney, D.L.; et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020**, *583*, 459–468. [CrossRef]

37. Thoms, M.; Buschauer, R.; Ameismeier, M.; Koepke, L.; Denk, T.; Hirschenberger, M.; Kratzat, H.; Hayn, M.; Mackens-Kiani, T.; Cheng, J.; et al. Structural basis for translational shutdown and immune evasion by the nsp1 protein of SARS-CoV-2. *Science* **2020**, *369*, 1249–1255. [CrossRef]

38. Gasic, I.; Mitchison, T.J. Autoregulation and repair in microtubule homeostasis. *Curr. Opin. Cell Biol.* **2019**, *56*, 80–87. [CrossRef]

39. Min, Y.-Q.; Mo, Q.; Wang, J.; Deng, F.; Wang, H.; Ning, Y.-J. SARS-CoV-2 nsp1: Bioinformatics, potential structural and functional features, and implications for drug/vaccine designs. *Front. Microbiol.* **2020**, *11*, 7317. [CrossRef]

40. Jimenez-Guardeno, J.M.; Regla-Nava, J.A.; Nieto-Torres, J.L.; DeDiego, M.L.; Castano-Rodriguez, C.; Fernandez-Delgado, R.; Perlman, S.; Enjuanes, L. Identification of the mechanisms causing reversion to virulence in an attenuated SARS-CoV for the design of a genetically stable vaccine. *PLoS Pathog.* **2015**, *11*, e1005215. [CrossRef]