SARS-CoV-2 Encodes a PPxY Late Domain Motif Known to Enhance Budding and Spread in Enveloped RNA Viruses

Halim Maaroufi
Institut de biologie intégrative et des systèmes (IBIS), Université Laval, Quebec City, Canada
Halim.maaroufi@ibis.ulaval.ca

ABSTRACT

The current COVID-19 (Coronavirus Disease-2019) pandemic is affecting the health and/or socioeconomic welfare of almost everyone in the world. Finding vaccines and therapeutics is therefore urgent, but elucidation of the molecular mechanisms that allow some viruses to cross host species boundaries, becoming a threat to human health, must also be given close attention. Here, analysis of all proteins of SARS-CoV-2 revealed a unique PPxY Late (L) domain motif, ^{25}PPAY^{28}, in a spike (S) protein inside a predicted hot disordered loop subject to phosphorylation and binding. PPxY motifs in enveloped RNA viruses are known to recruit Nedd4 E3 ubiquitin ligases and ultimately the ESCRT complex to enhance virus budding and release, resulting in higher viral loads, hence facilitating new infections. Interestingly, proteins of SARS-CoV-1 do not feature PPxY motifs, which could explain why SARS-CoV-2 is more contagious than SARS-CoV-1. Should an experimental assessment of this hypothesis show that the PPxY motif plays the same role in SARS-CoV-2 as it does in other enveloped RNA viruses, this motif will become a promising target for the development of novel host-oriented antiviral therapeutics for preventing S proteins from recruiting Nedd4 E3 ubiquitin ligase partners.

KEYWORDS SARS-CoV-2; Spike protein; PPxY Late Domain Motif; WW-domain; Nedd4 E3 ubiquitin ligases; Budding; Spread

INTRODUCTION

The current COVID-19 (Coronavirus Disease-2019) pandemic caused by SARS-CoV-2 from Wuhan is affecting the health (both physical and psychological) and/or socioeconomic welfare of almost everyone in the world. Finding vaccines and therapeutics to fight SARS-CoV-2 is therefore urgent. In addition, it is imperative, now and in the future, to maintain research efforts aimed at elucidating the molecular mechanisms (tropism, cell entry, multiplication and spread, etc.) used by emerging viruses such as SARS-CoV-2 to infect humans. An improved understanding of these mechanisms should be viewed as an investment in our future to prevent and/or control the emergence of dangerous viruses that could have as or more severe consequences for humanity than SARS-CoV-2. The family *Coronaviridae* (CoVs) is large and comprises enveloped, positive-sense, single-stranded RNA viruses (Graham and Baric, 2010; Li, 2013). Among all RNA viruses, CoVs possess the largest genomes, ranging in size between 27 and 32 kb. The subfamily *Coronavirinae* comprises four genera, *Alpha-*, *Beta-*, *Delta-* and *Gammacoronavirus*, which can infect both mammalian and avian species. Phylogenetically, SARS-CoV-2 is more closely related to SARS-CoV-1 than to MERS-CoV, but all three are betacoronaviruses (betaCoVs) (de Wit et al., 2016).

The envelope-anchored SARS-CoV-2 spike (S) is a multifunctional glycoprotein with 1273 amino acid residues. Its structure was solved as a homotrimeric molecule with several domains. The S protein is subdivided into S1 and S2 subunits. S1 is formed by two domains, an N-terminal domain (NTD) and a C-terminal receptor-binding domain (RBD) that binds specifically to the host receptor angiotensin-converting enzyme 2 (ACE2) on the host cell surface (Li, 2012). The S2 subunit, which is also multidomain, induces viral-host membrane fusion.

The viral proline-rich Pro–Pro-x-Tyr (PPxY) Late or L-domain motif (Chen and Sudol, 1995) interacts with host proteins containing the WW-domain. The term “Late” reflects a late function in the virus life cycle (Garnier et al., 1996; Freed, 2002). L-domain motifs are identified in enveloped RNA viruses such as in the Gag protein of a number of retroviruses

and in the matrix of arenaviruses, rhabdoviruses and filoviruses (Baillet et al., 2019; Bieniasz, 2006; Wirblich et al., 2008). The WW-domain is a small domain of 35–40 amino acids, named after two conserved tryptophan (W) residues in the sequence (Bork and Sudol, 1994). It is a fold of three-stranded antiparallel β-sheets that contains two ligand-binding grooves (Huang et al., 2000; Kanelis et al., 2001). RNA viruses use PPxY L-domain motifs to recruit specific WW-domain of host proteins of either Nedd4 (Neuronal precursor cell-Expressed Developmentally Down-regulated 4), or Nedd4 family members to facilitate their egress and spread (Bieniasz, 2006). Nedd4 family proteins have three domains, N-terminal lipid binding (C2) domain; WW modules, present in multiple copies; and a C-terminal Hect (for homologous to E6-associated protein C terminus) domain that contains the ubiquitin ligase activity.

Previous and current investigations are focusing on understanding the first step that allows SARS-CoV-1 and SARS-CoV-2 to enter host cells. This research examines primarily the interaction between the C-terminal receptor-binding domain (RBD) of S proteins and the host receptor angiotensin-converting enzyme 2 (ACE2). Here, the molecular mechanism of the last step (late function virus life cycle) of budding and release of virus is presented. The N-terminal of SARS-CoV-2 spike protein contains a PPxY L-domain motif that is known to hijack host WW-domain of Nedd4 E3 ubiquitin ligases and ultimately the ESCRT (Endosomal Sorting Complex Required for Transport) complex to enhance virus budding and spread. Importantly, this motif is absent in SARS-CoV-1, which could explain why SARS-CoV-2 is more contagious than SARS-CoV-1. Development of a novel host-oriented inhibitor targeting molecular interactions between PPxY and the host WW-domain holds potential in the fight against SARS-CoV-2 infection.

RESULTS AND DISCUSSION

The N-terminal of S protein contains a PPxY L-domain motif

Sequence analysis of SARS-CoV-2 spike protein using the Eukaryotic Linear Motif (ELM) resource (http://elm.eu.org/) revealed a short linear motif (SLiMs) known as PPxY Late (L-) domain motif, ^{25}PPAY^{28} in the N-terminus of the S1 subunit. The PPxY L-domain motif is known to interact with proteins containing WW-domain(s). L-domain motifs have been identified in single-stranded enveloped RNA viruses (Table 1), including in the Gag proteins of several retroviruses and in the matrix of arenaviruses, rhabdoviruses and filoviruses (Baillet et al., 2019; Bieniasz, 2006; Wirblich et al., 2008). RNA viruses use PPxY L-domain motifs to recruit specific WW-domains of host Nedd4 family members and ultimately the ESCRT complex to facilitate their budding and egress (Bieniasz, 2006). Interestingly, the SARS-CoV-2 S protein and Nedd4 can be S-palmitoylated, suggesting that they can localize to similar membrane subdomains (Petit et al., 2007; Gordon et al., 2020; Chesarino et al., 2014) and may therefore be able to interact. The ScanProsite tool (https://prosite.expasy.org/scanprosite/) showed that the PPxY motif is present in SARS-CoV-2 but absent from SARS-CoV-1 proteins. It has been reported in enveloped RNA viruses that PPxY motif recruits Nedd4 E3 ubiquitin ligases and ultimately the ESCRT complex to enhance virus budding and release that means a high viral load, thus facilitating new infections. This suggests in SARS-CoV-2 that the PPxY motif, through its role in enhancing the viral load, could play a role in making SARS-CoV-2 more contagious than SARS-CoV-1 (To et al., 2020).

The PPxY motif is also present in the S protein of bat RaTG13 and pangolin GX-P5L, two SARS-related coronaviruses (Fig. 1A). Moreover, S proteins with PPxY motif are phylogenetically close (Fig. 1C). Interestingly, alignment of S protein sequences of two pangolin coronaviruses showed only two differences in the S1 subunit (amino acid at position 25 (P/Q) and deletion of ^{448}GY^{449} in pangolin GX-P1E) (Lam et al., 2020; Zhang et al., 2020). In the future, it would be interesting to compare virulence of pangolin GX-P1E and pangolin GX-P5L coronaviruses to know if the loss of PPxY motif is important for virulence. Figure 1B shows differences among coronaviruses in nucleotide composition

for the codon corresponding to Pro25 in SARS-CoV-2. Transition mutations (purine (A, G) to purine or pyrimidine (C, T) to pyrimidine mutations) are more likely than transversion (purine to pyrimidine or vice versa) mutations. Additionally, most substitutions occur in the third nucleotide of codons (silent mutations), less frequently in the first nucleotide, and the least frequently in the second nucleotide (Wang et al., 2007). Therefore, a mutation from Ala (GCT) to Pro (CCT) is more likely to occur than one from Gln (CAG) to Pro (CCG). Moreover, a double mutation Asn (AAT) to Proline (CCT) is even less probable (Fig. 1B).

PPxY L-domain motif is in hot region

Scanning the SARS-CoV-2 S protein sequence with the DisEMBL software (Linding et al., 2003) revealed, in the N-terminal region, a unique hot disordered loop ^{19}TTRTQL**PP**A**Y**TNSFT^{33} that includes the PPxY motif (Table 2), and this loop contains additional motifs. Indeed, the PPxY motif shares its tyrosine with an overlapping ^{28}YTNSF^{32} motif (Fig. 1A) that enables binding to the SH2 domain of STAP1 when its Tyr28 is phosphorylated. STAP1 (Signal Adaptor Protein 1) is an adaptor protein with an SH2 and a PH class of lipid-binding domain (Wang et al., 1994). It is expressed in lymphoid cells and is phosphorylated by the Tec tyrosine kinase (Tec TK), which participates in B cell antigen receptor signaling (Ohya et al., 1999; Yang et al., 2001). The middle T-antigen of the polyoma virus, with its tyrosine residue in a phosphorylated state, has been shown to act as a binding site for the SH2 domain of the phosphatidylinositol-3'OH kinase 85K subunit (Dilworth et al. 1994). Another site, ^{28}YTNS^{31} in the S protein, is also Tyr-phosphorylation-dependent for binding SH2 domain (Lin et al., 2006).

The hot region of the S protein also contains a Ser/Thr phosphorylated motif. Indeed, the PPxY motif shares its Pro25 with an overlapping ^{19}TTRTQL**P**^{25} motif that could be phosphorylated at TQ by a phosphatidylinositol 3- and/or 4-kinase (PIKKs). The Gln (Q) beside to the target Ser/Thr is critical for substrate recognition. PIKKs are atypical Ser/Thr kinases found only in eukaryotes (Imseng et al., 2108; Angira et al., 2020). Interestingly, Yang et al. (2012) showed that phosphatidylinositol 4-kinase IIIβ (PI4KB) is required for cellular entry by pseudoviruses bearing the SARS-CoV-1 S protein. This cell entry is

highly inhibited by knockdown of PI4KB. Moreover, the same authors showed that PI4KB does not affect virus entry at the SARS-CoV-1 S-ACE2 interface. Finally, the motif ^{26}PAY**T**NSF**T**^{33} is recognised by the Ser/Thr kinase GSK3 (Glycogen synthase kinase 3). Interestingly, the lipid kinase phosphatidylinositol 4-kinase II α (PI4KIIα) was reported to be a substrate of GSK3 (Robinson et al. 2014). This finding establishes a link with ^{19}TTRTQL**P**^{25}motif, described above, that is predicted to be phosphorylated by GSK3. In summary, proteins that interact with this hot disordered loop are connected to lipids (membrane), particularly with phosphatidylinositol, and bindings is regulated by phosphorylation. Recently, Besson et al. (2019) showed that the mitogen-activated protein kinase pathway and phosphatidylinositol metabolism are involved in the later stage of RABV infection.

Interactions of PPxY L-domain motif with Nedd4-WW3-domain Ub ligase

To compare the interaction between the PPAY S protein and the Nedd4WW-domain with those of other PPxY-Nedd4WW-domain complexes, molecular docking was performed using the software AutoDock vina (Trott and Olson, 2010). Fig. 2 shows that the ^{23}QL**PP**A**Y**TNS^{31} peptide of SARS-CoV-2 S protein is localized in the same region as ^{112}TA**PP**E**Y**MEA^{120} motif of the Zaire Ebola virus Matrix protein VP40 (PDBid: 2KQ0), and important amino acid residues of the SARS-CoV-2 peptide are superposed with corresponding residues of the Ebola VP40 peptide, with the exception of different orientations for the aromatic ring of tyrosine (Fig. 2B and C). Iglesias-Bexiga et al. (2019) reported that the Nedd4-WW3 binding site is highly plastic and can accommodate PPxY-containing ligands with varying orientations. This structural superposition with PPxY of VP40 confirms the reliability of the AutoDock vina peptide docking module. Detailed knowledge of the molecular interaction between the ^{23}QL**PP**A**Y**TNS^{31} motif and the Nedd4WW-domain will help design a peptide able to mimic the surface of the Nedd4 WW-domain and strongly bind the ^{23}QL**PP**A**Y**TNS^{31} peptide, thus preventing the recruitment of Nedd4 E3 ubiquitin ligase by the SARS-CoV-2 S protein (Zaidman and Wolfson, 2016). Indeed, Han et al. (2014) have developed two lead compounds that specifically block the VP40 PPxY-host Nedd4 interaction, displaying anti-budding activity in the nanomolar range.

MATERIALS AND METHODS

Sequence analysis

To search probable short linear motifs (SLiMs), SARS-CoV-2 spike protein sequence was scanned using the eukaryotic linear motif (ELM) resource (http://elm.eu.org/). The PPxY L-domain motif was searched in proteins of SARS-CoV-2 and SARS-CoV-1 using https://prosite.expasy.org/scanprosite/

3D modeling and molecular docking

Unfortunately, the region containing the ^{23}QL**PP**A**Y**TNS^{31}peptide has not been resolved in any known structure of spike S protein. Therefore, the Pep-Fold (Thevenet et al., 2012) software was used to generate a *de novo* model of this peptide. Model quality of the peptide was assessed by analysis of a Ramachandran plot through PROCHECK (Vaguine et al., 1999).

Docking of the peptide into the human Nedd4 3rd WW-domain (PDBid: 2KQ0_A) was performed using the software AutoDock vina (Trott and Olson, 2010). The 3D complex containing the human Nedd4 3rd WW-domain and the peptide was refined using FlexPepDock (London et al., 2011), which allows full flexibility to the peptide and side-chain flexibility to the receptor.

Phylogeny

Amino acid sequences of different spike (S) proteins, namely those of SARS-CoV-2 (accession: YP_009724390.1), SARS-CoV-1 (P59594.1), bat_coronavirus_RaTG13 (QHR63300.2), pangolin_PCoV_GX-P5L (QIA48632.1), pangolin_PCoV_GX-P1E (QIA48623.1), bat_CoVZC45_CoV (AVP78031.1) and bat_Rs4874_CoV (ATO98205.1), were aligned using the T-coffee software (Rius et al., 2011). This multiple alignment was then used to estimate the phylogenetic relationships between SARS-CoV-2 and others coronaviruses. The phylogenetic tree was constructed using MrBayes (Huelsenbeck and Ronquist, 2001) with the following parameters: Likelihood model (Number of substitution

types: 6(GTR); Substitution model: Poisson; Rates variation across sites: Invariable + gamma); Markov Chain Monte Carlo parameters (number of generations: 100 000; Sample a tree every: 1000 generations) and Discard first 500 trees sampled (burnin).

ACKNOWLEDGMENTS

I would like to thank the IBIS bioinformatics group for their assistance. I am grateful to Dr. Michel Cusson, Université Laval, for the revision of the manuscript.

CONFLICT OF INTERESTED

The author declares that he has no conflicts of interest.

REFERENCES

Angira, D., Shaik, A., Thiruvenkatam, V., 2020. Structural and strategic landscape of PIKK protein family and their inhibitors: an overview. Front. Biosci. 25, 1538–1567.

Baillet, N., Krieger, S., Carnec, X., Mateo, M., Journeaux, A., Merabet, O., Caro, V., Tangy, F., Vidalain, P.-O., Baize, S., 2019. E3 Ligase ITCH Interacts with the Z Matrix Protein of Lassa and Mopeia Viruses and Is Required for the Release of Infectious Particles. Viruses 12. https://doi.org/10.3390/v12010049

Besson, B., Kim, S., Kim, T., Ko, Y., Lee, S., Larrous, F., Song, J., Shum, D., Grailhe, R., Bourhy, H., 2019. Kinome-Wide RNA Interference Screening Identifies Mitogen-Activated Protein Kinases and Phosphatidylinositol Metabolism as Key Factors for Rabies Virus Infection. mSphere 4. https://doi.org/10.1128/mSphere.00047-19

Bieniasz, P.D., 2006. Late budding domains and host proteins in enveloped virus release. Virology 344, 55–63. https://doi.org/10.1016/j.virol.2005.09.044

Bork, P., Sudol, M., 1994. The WW domain: a signalling site in dystrophin? Trends Biochem. Sci. 19, 531–533. https://doi.org/10.1016/0968-0004(94)90053-1

Bouamr, F., Melillo, J.A., Wang, M.Q., Nagashima, K., de Los Santos, M., Rein, A., Goff, S.P., 2003. PPPYVEPTAP motif is the late domain of human T-cell leukemia virus

type 1 Gag and mediates its functional interaction with cellular proteins Nedd4 and Tsg101 [corrected]. J. Virol. 77, 11882–11895. https://doi.org/10.1128/jvi.77.22.11882-11895.2003

Chen, H.I., Sudol, M., 1995. The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. Proc. Natl. Acad. Sci. U. S. A. 92, 7819–7823. https://doi.org/10.1073/pnas.92.17.7819

Chesarino, N.M., Hach, J.C., Chen, J.L., Zaro, B.W., Rajaram, M.V., Turner, J., Schlesinger, L.S., Pratt, M.R., Hang, H.C., Yount, J.S., 2014. Chemoproteomics reveals Toll-like receptor fatty acylation. BMC Biol. 12, 91. https://doi.org/10.1186/s12915-014-0091-3

de Wit, E., van Doremalen, N., Falzarano, D., Munster, V.J., 2016. SARS and MERS: recent insights into emerging coronaviruses. Nat. Rev. Microbiol. 14, 523–534. https://doi.org/10.1038/nrmicro.2016.81

Dilworth, S.M., Brewster, C.E., Jones, M.D., Lanfrancone, L., Pelicci, G., Pelicci, P.G., 1994. Transformation by polyoma virus middle T-antigen involves the binding and tyrosine phosphorylation of Shc. Nature 367, 87–90. https://doi.org/10.1038/367087a0

Dolnik, O., Kolesnikova, L., Stevermann, L., Becker, S., 2010. Tsg101 is recruited by a late domain of the nucleocapsid protein to support budding of Marburg virus-like particles. J. Virol. 84, 7847–7856. https://doi.org/10.1128/JVI.00476-10

Freed, E.O., 2002. Viral late domains. J. Virol. 76, 4679–4687. https://doi.org/10.1128/jvi.76.10.4679-4687.2002

Garnier, L., J. W. Wills, M. F. Verderame and M. Sudol (1996). "WW domains and retrovirus budding." Nature 381(6585): 744-745. https://doi.org/10.1038/381744a0

Graham, R.L., Baric, R.S., 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J. Virol. 84, 3134–3146. https://doi.org/10.1128/JVI.01394-09

Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., O'Meara, M.J., Guo,

J.Z., Swaney, D.L., Tummino, T.A., Huettenhain, R., Kaake, R.M., Richards, A.L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., Kim, M., Haas, P., Polacco, B.J., Braberg, H., Fabius, J.M., Eckhardt, M., Soucheray, M., Bennett, M.J., Cakir, M., McGregor, M.J., Li, Q., Naing, Z.Z.C., Zhou, Y., Peng, S., Kirby, I.T., Melnyk, J.E., Chorba, J.S., Lou, K., Dai, S.A., Shen, W., Shi, Y., Zhang, Z., Barrio-Hernandez, I., Memon, D., Hernandez-Armenta, C., Mathy, C.J.P., Perica, T., Pilla, K.B., Ganesan, S.J., Saltzberg, D.J., Ramachandran, R., Liu, X., Rosenthal, S.B., Calviello, L., Venkataramanan, S., Liboy-Lugo, J., Lin, Y., Wankowicz, S.A., Bohn, M., Sharp, P.P., Trenker, R., Young, J.M., Cavero, D.A., Hiatt, J., Roth, T.L., Rathore, U., Subramanian, A., Noack, J., Hubert, M., Roesch, F., Vallet, T., Meyer, B., White, K.M., Miorin, L., Rosenberg, O.S., Verba, K.A., Agard, D., Ott, M., Emerman, M., Ruggero, D., García-Sastre, A., Jura, N., von Zastrow, M., Taunton, J., Ashworth, A., Schwartz, O., Vignuzzi, M., d'Enfert, C., Mukherjee, S., Jacobson, M., Malik, H.S., Fujimori, D.G., Ideker, T., Craik, C.S., Floor, S., Fraser, J.S., Gross, J., Sali, A., Kortemme, T., Beltrao, P., Shokat, K., Shoichet, B.K., Krogan, N.J., 2020. A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. bioRxiv. https://doi.org/10.1101/2020.03.22.002386

Han, Z., Lu, J., Liu, Y., Davis, B., Lee, M.S., Olson, M.A., Ruthel, G., Freedman, B.D., Schnell, M.J., Wrobel, J.E., Reitz, A.B., Harty, R.N., 2014. Small-molecule probes targeting the viral PPxY-host Nedd4 interface block egress of a broad range of RNA viruses. J. Virol. 88, 7294–7306. https://doi.org/10.1128/JVI.00591-14

Harty, R.N., Brown, M.E., McGettigan, J.P., Wang, G., Jayakar, H.R., Huibregtse, J.M., Whitt, M.A., Schnell, M.J., 2001. Rhabdoviruses and the cellular ubiquitin-proteasome system: a budding interaction. J. Virol. 75, 10623–10629. https://doi.org/10.1128/JVI.75.22.10623-10629.2001

Harty, R.N., Brown, M.E., Wang, G., Huibregtse, J., Hayes, F.P., 2000. A PPxY motif within the VP40 protein of Ebola virus interacts physically and functionally with a ubiquitin ligase: implications for filovirus budding. Proc. Natl. Acad. Sci. U. S. A. 97, 13871–13876. https://doi.org/10.1073/pnas.250277297

Huang, X., Poy, F., Zhang, R., Joachimiak, A., Sudol, M., Eck, M.J., 2000. Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. Nat. Struct. Biol. 7, 634–638. https://doi.org/10.1038/77923

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Iglesias-Bexiga, M., Palencia, A., Corbi-Verge, C., Martin-Malpartida, P., Blanco, F.J., Macias, M.J., Cobos, E.S., Luque, I., 2019. Binding site plasticity in viral PPxY Late domain recognition by the third WW domain of human NEDD4. Sci. Rep. 9, 15076. https://doi.org/10.1038/s41598-019-50701-3

Imseng, S., Aylett, C.H., Maier, T., 2018. Architecture and activation of phosphatidylinositol 3-kinase related kinases. Curr. Opin. Struct. Biol. 49, 177–189. https://doi.org/10.1016/j.sbi.2018.03.010

Kanelis, V., Rotin, D., Forman-Kay, J.D., 2001. Solution structure of a Nedd4 WW domain-ENaC peptide complex. Nat. Struct. Biol. 8, 407–412. https://doi.org/10.1038/87562

Lam, T.T.-Y., Shum, M.H.-H., Zhu, H.-C., Tong, Y.-G., Ni, X.-B., Liao, Y.-S., Wei, W., Cheung, W.Y.-M., Li, W.-J., Li, L.-F., Leung, G.M., Holmes, E.C., Hu, Y.-L., Guan, Y., 2020. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. Nature. https://doi.org/10.1038/s41586-020-2169-0

Li, F., 2013. Receptor recognition and cross-species infections of SARS coronavirus. Antiviral Res. 100, 246–254. https://doi.org/10.1016/j.antiviral.2013.08.014

Li, F., 2012. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. J. Virol. 86, 2856–2858. https://doi.org/10.1128/JVI.06882-11

Lin, W., Kim, S.S., Yeung, E., Kamegaya, Y., Blackard, J.T., Kim, K.A., Holtzman, M.J., Chung, R.T., 2006. Hepatitis C virus core protein blocks interferon signaling by interaction with the STAT1 SH2 domain. J. Virol. 80, 9226–9235. https://doi.org/10.1128/JVI.00459-06

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., Russell, R.B., 2003. Protein disorder prediction: implications for structural proteomics. Structure 11, 1453–1459. https://doi.org/10.1016/j.str.2003.10.002

London, N., Raveh, B., Cohen, E., Fathi, G., Schueler-Furman, O., 2011. Rosetta FlexPepDock web server--high resolution modeling of peptide-protein interactions. Nucleic Acids Res. 39, W249–53. https://doi.org/10.1093/nar/gkr431

Ohya, K., Kajigaya, S., Kitanaka, A., Yoshida, K., Miyazato, A., Yamashita, Y., Yamanaka, T., Ikeda, U., Shimada, K., Ozawa, K., Mano, H., 1999. Molecular cloning of a docking protein, BRDG1, that acts downstream of the Tec tyrosine kinase. Proc. Natl. Acad. Sci. U. S. A. 96, 11976–11981. https://doi.org/10.1073/pnas.96.21.11976

Petit, C.M., Chouljenko, V.N., Iyer, A., Colgrove, R., Farzan, M., Knipe, D.M., Kousoulas, K.G., 2007. Palmitoylation of the cysteine-rich endodomain of the SARS-coronavirus spike glycoprotein is important for spike-mediated cell fusion. Virology 360, 264–274. https://doi.org/10.1016/j.virol.2006.10.034

Rius, J., Cores, F., Solsona, F., van Hemert, J.I., Koetsier, J., Notredame, C., 2011. A user-friendly web portal for T-Coffee on supercomputers. BMC Bioinformatics 12, 150. https://doi.org/10.1186/1471-2105-12-150

Robinson, J.W., Leshchyns'ka, I., Farghaian, H., Hughes, W.E., Sytnyk, V., Neely, G.G., Cole, A.R., 2014. PI4KIIα phosphorylation by GSK3 directs vesicular trafficking to lysosomes. Biochem. J 464, 145–156. https://doi.org/10.1042/BJ20140497

Strack, B., Calistri, A., Accola, M.A., Palu, G., Gottlinger, H.G., 2000. A role for ubiquitin ligase recruitment in retrovirus release. Proc. Natl. Acad. Sci. U. S. A. 97, 13063–13068. https://doi.org/10.1073/pnas.97.24.13063

Thevenet, P., Shen, Y., Maupetit, J., Guyon, F., Derreumaux, P., Tufféry, P., 2012. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. Nucleic Acids Res. 40, W288–93. https://doi.org/10.1093/nar/gks419

To, K.K.-W., Tsang, O.T.-Y., Leung, W.-S., Tam, A.R., Wu, T.-C., Lung, D.C., Yip, C.C.-Y., Cai, J.-P., Chan, J.M.-C., Chik, T.S.-H., Lau, D.P.-L., Choi, C.Y.-C., Chen, L.-L., Chan, W.-M., Chan, K.-H., Ip, J.D., Ng, A.C.-K., Poon, R.W.-S., Luo, C.-T., Cheng, V.C.-C., Chan, J.F.-W., Hung, I.F.-N., Chen, Z., Chen, H., Yuen, K.-Y., 2020. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. Lancet Infect. Dis. https://doi.org/10.1016/S1473-3099(20)30196-1

Trott, O., Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461. https://doi.org/10.1002/jcc.21334

Vaguine, A.A., Richelle, J., Wodak, S.J., 1999. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallogr. D Biol. Crystallogr. 55, 191–205. https://doi.org/10.1107/S0907444998006684

Wang, G.-Z., Chen, L.-L., Zhang, H.-Y., 2007. Phase-dependent nucleotide substitution in protein-coding sequences. Biochem. Biophys. Res. Commun. 355, 599–602. https://doi.org/10.1016/j.bbrc.2007.01.006

Wang, D.S., Shaw, R., Winkelmann, J.C., Shaw, G., 1994. Binding of PH domains of beta-adrenergic receptor kinase and beta-spectrin to WD40/beta-transducin repeat containing regions of the beta-subunit of trimeric G-proteins. Biochem. Biophys. Res. Commun. 203, 29–35. https://doi.org/10.1006/bbrc.1994.2144

Wirblich, C., Tan, G.S., Papaneri, A., Godlewski, P.J., Orenstein, J.M., Harty, R.N., Schnell, M.J., 2008. PPEY motif within the rabies virus (RV) matrix protein is essential for efficient virion release and RV pathogenicity. J. Virol. 82, 9730–9738. https://doi.org/10.1128/JVI.00889-08

Yang, N., Ma, P., Lang, J., Zhang, Y., Deng, J., Ju, X., Zhang, G., Jiang, C., 2012. Phosphatidylinositol 4-kinase IIIβ is required for severe acute respiratory syndrome coronavirus spike-mediated cell entry. J. Biol. Chem. 287, 8457–8467. https://doi.org/10.1074/jbc.M111.312561

Yang, W.C., Ching, K.A., Tsoukas, C.D., Berg, L.J., 2001. Tec kinase signaling in T cells is regulated by phosphatidylinositol 3-kinase and the Tec pleckstrin homology domain. J. Immunol. 166, 387–395. https://doi.org/10.4049/jimmunol.166.1.387

Zaidman, D., Wolfson, H.J., 2016. PinaColada: peptide–inhibitor ant colony ad-hoc design algorithm. Bioinformatics 32, 2289–2296. https://doi.org/10.1093/bioinformatics/btw133

Zhang, T., Wu, Q., Zhang, Z., 2020. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. Curr. Biol. 30, 1346–1351.e2. https://doi.org/10.1016/j.cub.2020.03.022

FIGURES LEGEND

Figure 1. Multiple alignment of spike (S) protein of betacoronaviruses using the T-coffee software. **(A)** Localization of PPxY Late domain in N-terminal. PPxY Late domain motifs are colored in blue. The ^{28}YTNSF^{32} motif is indicated by green arrows. When its Tyr is phosphorylated, it is able to bind Src Homology 2 (SH2) domain. GenBank and UniProt accession numbers are indicated at the start of each sequence. The figure was prepared using ESPript (http://espript.ibcp.fr). (B) Alignment of codons of PPxY motifs. If nucleotides in red in codons could mutate into cytosine, they would produce proline residue. (C) Unrooted phylogenetic tree of S protein sequences of representative betacoronaviruses. The tree was constructed using Mr Bayes method based on the multiple alignment of complete S protein sequences by T-coffee. Betacoronaviruses in blue possess a PPxY motif. GenBank and UniProt accession numbers are showed at the start of each sequence.

Figure 2. Cartoon and surface representation of human Nedd4 WW3-domain (PDBid: 2KQ0_A) with docked peptides. (A) ^{23}QL**PP**A**Y**TNS^{31} peptide of SARS-CoV-2 S protein (cyan) and (B) ^{23}QL**PP**A**Y**TNS^{31} superposed to Zaire Ebola virus Matrix protein VP40 (^{112}TA**PP**E**Y**MEA^{120}) peptide (green, PDBid: 2KQ0_B). (C) Surface representation of Nedd4 WW3-domain with peptides depicted as sticks. Images were generated using PyMol (www.pymol.org).

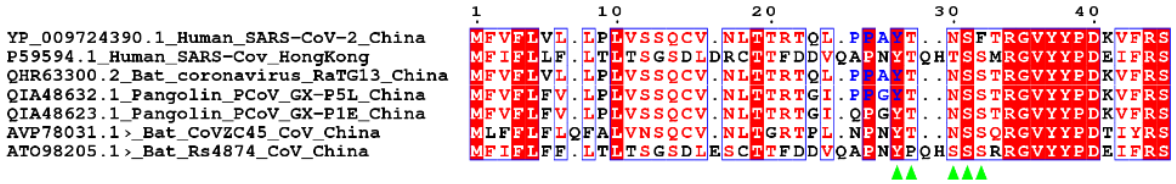
Table 1. PPxY late-domain motif in single stranded enveloped RNA viruses.

Viruses	Family	disease in human	Geography	PPxY protein	Reference
HIV-1	Retroviruses	AIDS	World	Gag	Strack et al., 2000
HTLV-1	Retroviruses	Aggressive lymphoma	World	Gag	Bouamr et al., 2003
Rhabies	Rhabdoviruses	Deadly brain inflammation	World	M	Harty et al., 2001
Marburg	Filoviruses	Deadly hemorrhagic fever	West Africa	VP40	Dolnik et al., 2010
Ebola	Filoviruses	Deadly hemorrhagic fever	West Africa	VP40	Harty et al., 2000
Lassa virus	Arenaviruses	Deadly hemorrhagic fever	West Africa	Z matrix	Baillet et al., 2019

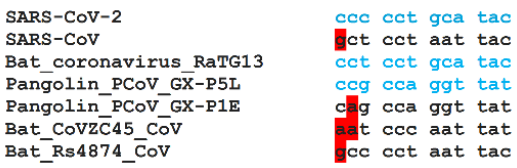
Table. 2. ELM motifs of hot disordered loop in SARS-CoV-2 S protein.

Elm Name	Instances	Positions	Elm Description	Cell compartment	Pattern
LIG_WW_1	PPAY	25-28	PPXY is the motif recognized by WW domains of Group I.	cytosol, membrane	PP.Y
LIG_SH2_STAP1	YTNSF	28-32	STAP1 Src Homology 2 (SH2) domain Class 2 binding motif.	plasma membrane	(Y)[DESTA][^GP][^GP][ILVFMWYA]
LIG_SH2_STAT5	YTNS	28-31	STAT5 Src Homology 2 (SH2) domain binding motif.	cytosol	(Y)[VLTFIC]..
MOD_GSK3_1	PAYTNSFT	26-33	GSK3 phosphorylation recognition site.	cytosol, nucleus	...([ST])...[ST]
MOD_PIKK_1	TTRTQLP	19-25	(ST)Q motif which is phosphorylated by PIKK family members.	nucleus, cytosol	...([ST])Q..

A


1 10 20 30 40
YP_009724390.1_Human_SARS-CoV-2_China
P59594.1_Human_SARS-Cov_HongKong
QHR63300.2_Bat_coronavirus_RaTG13_China
QIA48632.1_Pangolin_PCoV_GX-P5L_China
QIA48623.1_Pangolin_PCoV_GX-P1E_China
AVP78031.1>_Bat_CoVZC45_CoV_China
ATO98205.1>_Bat_Rs4874_CoV_China

B


SARS-CoV-2 ccc cct gca tac
SARS-CoV gct cct aat tac
Bat_coronavirus_RaTG13 cct cct gca tac
Pangolin_PCoV_GX-P5L ccg cca ggt tat
Pangolin_PCoV_GX-P1E cag cca ggt tat
Bat_CoVZC45_CoV aat ccc aat tat
Bat_Rs4874_CoV gcc cct aat tac

C

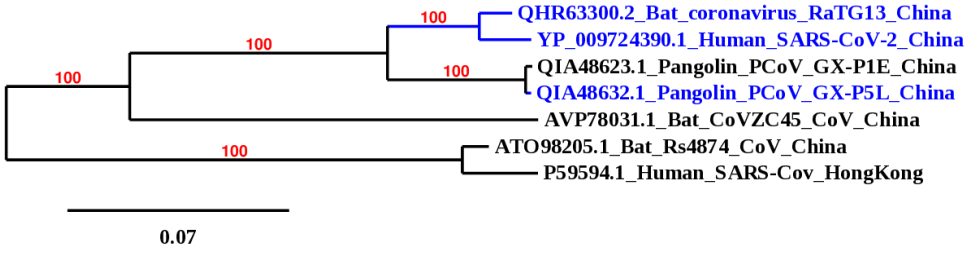

QHR63300.2_Bat_coronavirus_RaTG13_China
YP_009724390.1_Human_SARS-CoV-2_China
QIA48623.1_Pangolin_PCoV_GX-P1E_China
QIA48632.1_Pangolin_PCoV_GX-P5L_China
AVP78031.1_Bat_CoVZC45_CoV_China
ATO98205.1_Bat_Rs4874_CoV_China
P59594.1_Human_SARS-Cov_HongKong
100
100
100
100
100
0.07

Fig. 1

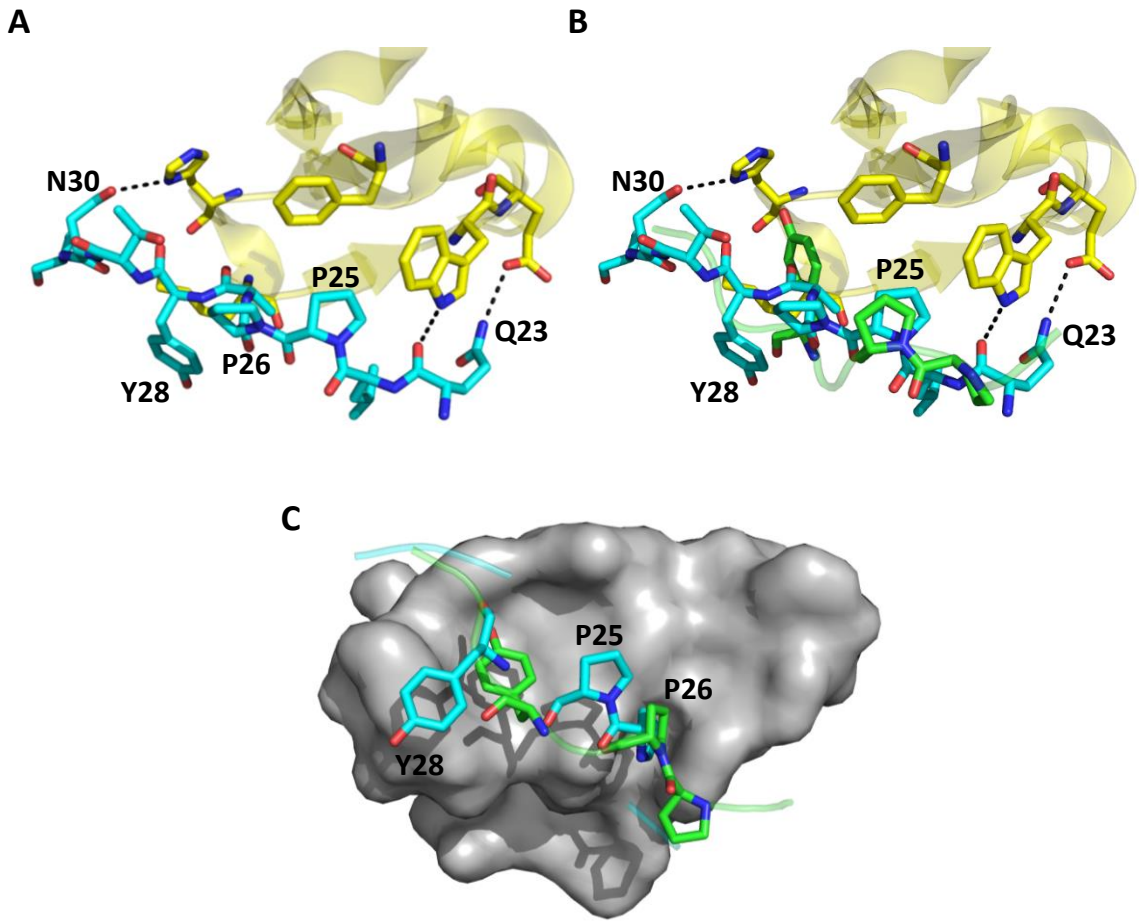


Fig. 2