 doi:10.20944/preprints202004.0171.v1

Title: The potential roles of protein functional motifs in coronavirus infection

Author: Haitham Sobhy

Address: Laboratory of dynamics host-pathogen interactions (DHPI), IUT Louis Pasteur, University of Strasbourg, Schiltigheim, France.

E-mail: sobhy@unistra.fr, and haithamsobhy@gmail.com

Abstract

Although phylogenetic analysis shows coronaviruses (CoV) share similar genome sequences, CoVs encode different number of proteins (5 to 14). The newly isolated viruses harbour more proteins than the old ones. Therefore, identifying the functional protein unites will benefits to understand the molecular interactions of the virus, and then identify molecular targets for antiviral drug. Here, the comparative *in-silico* analysis of 33 coronavirus proteomes show that coronaviruses harbour diverse number of protein functional motifs. Coronaviruses harbour wide-range of motifs including those involved in integrin-binding and ESCRT pathway before virus budding. For example, SARS-CoV-2, but not SARS-CoV-1, encodes PPxY motif, which is required for virus entry and budding of HIV, influenza and adenoviruses. The quinolone inulding the antiviral FGI-104 is able to block ESCRT pathway and viral budding and has been used against HIV, HCV and Ebola virus.

Main text

The coronavirus outbreak (coronavirus disease-19, COVID-19) were initiated by a zoonotic virus that has been transmitted to human. Genome sequencing reveals that the causative virus is named severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2), and belongs to genus *Betacoronavirus*, family *Coronaviridae* [1]. The virus shares sequence homology with coronaviruses infecting bats and murine, such as bat SARS-like CoV bat-SL-CoVZC45, so-called subgenus *Sarbecovirus*. The first SARS-CoV outbreak occurred in 2003. The zoonotic virus was found to share sequence homology 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 belong to subgenus *Merbecovirus*, and share sequence homology with bat CoV

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 April 2020 doi:10.20944/preprints202004.0171.v1

HKU4, HKU5, and betacoronavirus Erinaceus. The third subgenus *Embevovirus* include human CoV HKU1 and OC43 share homology with murine CoV MHV-1 and betacoronavirus HKU24.

The phylogenetic analysis shows coronaviruses of the same subgroups are highly conserved and share high sequence homology within the same clade [1]. Coronaviruses encode diverse number of proteins, ranging from 5 protein in bat CoV, to more than 12 in human CoVs, table 1. These reveal that coronaviruses characterize by high mutation rates. Therefore, the molecular proteomics interactions and virus-host cell interactions are different from one virus to another. Generally, mutations introduce short protein motifs or domains, which lead to emergence of new strains; these short motifs could contribute in viral virulence and ability to infect wide-range of host cells [2,3].

This analysis highlights the role of functional protein domains during the viral infection. Particularly, the proteome of the newly isolated coronaviruses harbour more number of proteins, which have unknown functions. In this analysis, the full protein sequences of 33 coronaviruses, including SARS-CoV-2 are obtained from NCBI database. Then, the functional motifs in the coronavirus proteomes identified using exact search text-mining implemented in Shetti-Motif tool (https://sites.google.com/site/haithamsobhy/software), for detailed discussion and method see [2-4]. Additionally, the comparative genomics approach between the human CoV and zoonotic CoV reveals particular proteins motifs that were deleted or acquired by the new emerging viruses. It worth to note that the multiple and pairwise alignment may not help to find these short functional and linear motifs. Therefore, the motifs should be validated experimentally and find them by exact search text-mining.

With that in mind, PPxY motif (x means any residue) are largely encoded by coronaviruses the surface glycoprotein (S) encoded by SARS-CoV-2; two proteins of MERS-CoV EMC/2012 (1A and 1AB polyprotein), three proteins of *Erinaceus* CoV (protein S, ORF 3b and orf4a), table 1 and supplementary table 1 and 2. The motif is not encoded by human CoV nor by SARS-CoV-1. The motif is crucial for HIV-1 and paramyxoviruses budding and exit from the cell. The viruses utilize the motif to recruit Nedd4 E3 ubiquitin ligases and then endosomal sorting complexes required for the transport (ESCRT) pathway, which lead to virus egress and budding, reviewed in [3]. Additionally, adenoviruses utilize PPxY motif during cell entry and cellular trafficking. Coronaviruses encode other canonical ESCRT-interacting motifs, such as P[T/S]AP, [F/I/L/V]PxV, YxxL, and LYPxL. For example, the orf1ab polyprotein of SARS-CoV-2 harbours LYPTL, and 2 LPGV and 2 VPFV motifs, whereas the virus does not code P[T/S]AP motif. The P[T/S]AP domain can recruit TSG101 protein, a component of ESCRT-I, whereas LYPxnL can recruit Alix, a component of ESCRT-II complex, reviewed in [3,5-8]. The interactions between viral proteins and ESCRT, were considered as potential

 doi:10.20944/preprints202004.0171.v1

antiviral therapeutics (drug), [9,10]. Quinolones were shown to be able to block ESCRT pathway. FGI-104 is a quinolone that is used as an antiviral drug against HIV, hepatitis B and C, and Ebola viruses [11,12]. Noting that, quinolones include antiviral drugs (FGI-103 and FGI-106), and chloroquine.

Additionally, RGD motif is the crucial integrin-binding motif, reviewed in [3]. All the coronaviruses, except the proteome of human CoV OC43, harbour RGD motif. In absence of RGD, viruses may utilize other motifs to attach to cellular receptors and enter into host cells, such as LDV, LDI and SDI, which are encoded by coronaviruses. Noteworthy, SARS-CoV-2 harbours KGE motif. Although the closely related bat CoV-CoVZC45 does not harbour KGE motif, the distant proteome of bat CoV, and *Erinaceus* and murine CoVs harbour the same motif. KGE is required for stabilization of the virus on cell surface before entry into cytoplasm [13]. Recently, it was shown that SARS-CoV-2 harbour RGD motif, which means that the virus is able to interact with integrins to enter into host cells [14].

That said, two proteins could have crucial roles in entry of SARS-CoV-2 virus. i) The orf1ab polyprotein, 7096 residues, which harbours integrin-binding motifs, such as DGE, DLxxL, RGD, and SDI; and clathrin-binding motifs, such as PxxP, L[FILV]x[FILV][DE], YxYxx[FILV] and WxxF. The protein homologue, orf1a polyprotein, 4405 aa, harbours similar protein functional motifs, supplementary table 2. ii) The protein S, which harbours RGD motif and clathrin-binding motifs. Both S and orf1ab proteins harbour KR-rich motifs, which needed to localize proteins and genome into the nucleus.

To conclude, COVID-19 outbreak highlights the importance of studying the protein motifs to predict the functional units that contribute in viral virulence. Although RGD-like motifs and PPxY motif are short motifs (3-4 residues), they are deleted in some coronaviruses. Surprisingly, coronaviruses encode motifs that functionally resemble those encoded by viruses that infect respiratory tract, e.g. paramyxoviruses and adenoviruses. SARS-CoV-2 harbours functional motifs differ from the closely related bat-SL-CoVZC45, but in some motifs are similar to the distant bat CoVs and mammalian CoVs, which open question regarding the evolution of SARS-CoV-2. Finally, high-throughput analyses benefit developing vaccine or antiviral drug by offering appropriate molecular targets.

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

3. Sobhy, H. A review of functional motifs utilized by viruses. *Proteomes* **2016**, *4*.
4. Sobhy, H. Virophages and their interactions with giant viruses and host cells. *Proteomes* **2018**, *6*.
5. Williams, R.L.; Urbe, S. The emerging shape of the escrt machinery. *Nat Rev Mol Cell Biol* **2007**, *8*, 355-368.
6. 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.
7. 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 Suppl 2*, S138-145.
8. Wolff, S.; Ebihara, H.; Groseth, A. Arenavirus budding: A common pathway with mechanistic differences. *Viruses* **2013**, *5*, 528-549.
9. 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.*, et al.* Small-molecule probes targeting the viral ppxy-host nedd4 interface block egress of a broad range of rna viruses. *J Virol* **2014**, *88*, 7294-7306.
10. Tavassoli, A.; Lu, Q.; Gam, J.; Pan, H.; Benkovic, S.J.; Cohen, S.N. Inhibition of hiv budding by a genetically selected cyclic peptide targeting the gag-tsg101 interaction. *ACS Chem Biol* **2008**, *3*, 757-764.
11. Zhu, J.D.; Meng, W.; Wang, X.J.; Wang, H.C. Broad-spectrum antiviral agents. *Front Microbiol* **2015**, *6*, 517.
12. Kinch, M.S.; Yunus, A.S.; Lear, C.; Mao, H.; Chen, H.; Fesseha, Z.; Luo, G.; Nelson, E.A.; Li, L.; Huang, Z.*, et al.* Fgi-104: A broad-spectrum small molecule inhibitor of viral infection. *Am J Transl Res* **2009**, *1*, 87-98.
13. Berryman, S.; Clark, S.; Kakker, N.K.; Silk, R.; Seago, J.; Wadsworth, J.; Chamberlain, K.; Knowles, N.J.; Jackson, T. Positively charged residues at the five-fold symmetry axis of cell culture-adapted foot-and-mouth disease virus permit novel receptor interactions. *J Virol* **2013**, *87*, 8735-8744.
14. Sigrist, C.J.; Bridge, A.; Le Mercier, P. A potential role for integrins in host cell entry by sars-cov-2. *Antiviral Res* **2020**, *177*, 104759.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 April 2020 doi:10.20944/preprints202004.0171.v1

Table 1. Show the functional motif canonical sequence, the function of the motifs, and number of proteins harbour these motif. Number of proteins encoded by viruses are also indicated.

		Bat CoV BtCoV79005	CoV BtRs-BetaCoV_YN2018B	CoV BtRs-BetaCoV_YN2018C	SARS CoV GZ02	SARS CoV NS-1		bat-SL-CoVZC45	SARS-CoV-2 isolate Wuhan-Hu-1		Human betaCoV 2c EMC/2012	Bat CoV HKU5-1	BetaCoV Erinaceus	Tylonycteris bat CoV HKU4	Human CoV HKU1	Human CoV OC43	BetaCoV HKU24	Murine CoV MHV-1
Total number of proteins		**9**	**5**	**5**	**11**	**11**		**11**	**12**		**11**	**9**	**12**	**9**	**8**	**9**	**10**	**11**
Function	**Motif**																	
Viral attachment to cellular receptors	KGE	1	1	1	1	2		0	2		0	0	2	0	0	0	0	1
Binding to integrins, and viral attachment to cellular receptors	RGD	1	1	1	1	2		2	3		2	1	1	1	1	0	1	2
	DGE	1	1	1	1	2		1	2		2	3	2	1	0	0	0	0
	IDA	2	3	2	3	4		2	3		3	2	3	2	1	2	2	2
	LDI	2	1	1	1	1		2	3		0	0	0	1	1	1	3	1
	LDV	1	1	1	3	4		0	0		3	1	3	1	2	3	2	2
	SDI	1	1	1	1	1		1	2		2	3	0	3	3	1	4	1
	YGL	0	0	0	0	0		0	0		2	2	3	2	2	3	2	2
Binding to phospholipids, lipid raft-mediated endocytosis	RxLR	0	1	0	1	1		0	0		1	2	0	0	0	1	1	0
Enhance virion-release, anti-tetherin activity	DSGxxS	0	0	0	0	0		0	0		2	1	1	0	1	1	0	0
Helix-helix interactions	Ax3Ax3Ax3W	0	0	0	1	1		0	0		0	0	0	0	0	0	0	0
	Vx3IxLx3L	0	0	0	1	1		1	0		0	0	1	0	0	1	0	2
Heparan sulfate-binding motif, post-internalization of adenovirus infection	bbbxxb	0	0	0	1	1		0	0		0	1	1	1	1	1	0	1
Interact with host SH3 domain	[RKY]xxPxxP	0	0	0	0	0		0	0		1	2	2	2	1	1	2	1
ITAM motif, positive signal of immune receptors	Yxx[LI]x(6,8)Yxx[LI]	2	1	1	1	2		1	2		0	1	2	0	1	1	1	2
ITIM motif, negative signal of immune receptors	[SIVL]xYxx[IVL]	4	3	3	5	6		6	6		5	2	7	5	5	5	4	6
Recruit ESCRT pathway	PPxY	0	0	0	0	0		0	1		2	1	3	2	0	0	0	0
Recruit ESCRT pathway, mediates viral budding and release	P[TS]AP	0	0	0	0	0		0	0		2	0	0	0	0	0	0	0