

Low homology between 2019-nCoV Orf8 protein and its SARS-CoV counterparts questions their identical function

Viktoria Iatsenko*, Konstantin Boyarshin, Olga Bespalova, Violetta Klyueva, Yuliya Kurkina, and Irina Batlutskaya
Belgorod State National Research University, Belgorod, Russia

Abstract. SARS-CoV accessory protein Orf8b is involved in suppressing interferon-mediated immune response of the infected cell and this might lead to supposition that the corresponding protein 2019-nCoV Orf8 shares the same role. But the tertiary structures of these proteins are still unknown, and the primary structures demonstrate very low homology and different calculating parameters. This time they both are affected by stabilizing selection and in natural viral populations do not tend to be deleted. The question whether in this case very different proteins could share the same function rises from the present data.

1 Introduction

Single-stranded positive-sense RNA viruses of the family *Coronaviridae* are capable to infect amphibians, birds and mammals [1]. In particular, they cause such infectious human diseases as Severe Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS) [2-4] and CoVID-19 [5]. Viruses close to 2019-nCoV also infect pangolin *Manis javanica* [6-8] and bats, particularly *Rhinolophus affinis* [9,10]. Several other human infecting coronaviruses, such as HCoV-NL63, HCoV229E, HCoV-OC43 and HCoVHKU1 are well known [11]. These viruses are quite different, so 2019-nCoV and SARS-CoV share only near 70% genomic identity [12]. The closest animal counterpart of 2019-nCoV seems to be the bat virus RaTG13 with 96% genomic identity [10].

So-called accessory proteins of coronaviruses in many cases are involved in blocking of the interferon-mediated immune response of the infected cell [13]. Among them polymorphic protein Orf8 is present in 2019-nCoV and SARS-CoV. In the CoViD-19 pathogen and early strains of SARS coronavirus it exists in the form of a single 121-122 aa protein. In many other SARS strains it is splitted for two different proteins Orf8a and Orf8b. They still are coded by one mRNA, and their coding sequences partially overlap (Fig. 1).

On some different models it was shown that the SARS proteins Orf8b and joint Orf8ab are involved in blocking of interferon induction [15]. They interact directly with interferon regulatory factor 3 (IRF3) [16] leading to its ubiquitination and degradation, that prevents the induction of interferon synthesis by its activated form. Orf8a this time prevents the fast degradation of Orf8b. Besides in one of models it some way prevents expression of Orf8b [15], that itself is difficult to interpret.

*Corresponding author: viktoria-y-99@mail.ru

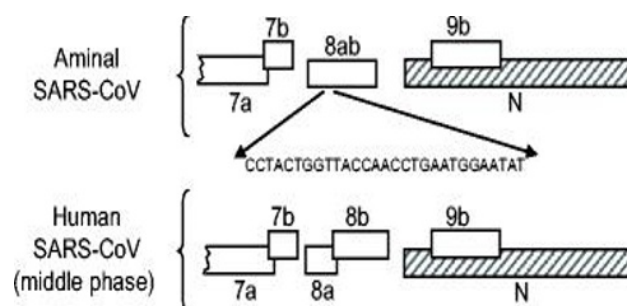


Fig. 1. Open reading frames for the joint protein Orf8ab (or simply Orf8) in animal and early human SARS-CoV and splitted Orf 8a and Orf8b in the late human-adapted strains. The nucleotide sequence absent in the strains with splitted Orf8 is marked by arrows. Figure is from the work of Keng and Tan [14]. 2019-nCoV resembles upper variant, but lacks Orf9b and possess Orf10 right of gene N.

It seems that comparing the primary sequences of the coronaviral proteins similar to Orf8b we should achieve high homology and the presence of some invariant structural features, ensuring its function that is really important and supported by stabilizing selection. Otherwise, we should see an example of very fast evolution, changing whether function or the structural base of its performing.

2 Materials and methods

Sequences of the proteins Orf8 and Orf8a are from the public database of National Institute of Health of USA. Information about genomic polymorphisms of 2019-nCoV is from the database described in the work of von Dorp and coauthors [17].

For multiple sequence alignment the program Clustal Omega was used [18]. The program Protein BLAST [19] was used for search and pairwise alignment of similar sequences by the chosen query. The standard settings were applied.

To assess amino acid composition and determine theoretical values of computational protein parameters, for instance isoelectric point and grand average of hydrophobicity, the program ProtParam [20] was applied.

3 Results and discussion

Among 11 Orf8b protein sequences belonging to SARS-CoV isolated from humans only 2 polymorphic positions were detected. They both are close to C-terminus. That are Lys81Asn and Thr83Ile substitutions. They tend to coincide and are the minority. In position 81 both residues are positively charged, although to varying degrees, in position 83 properties of residue change. Other positions seem to be monomorphic in such a sampling.

14 sequences of the joint SARS-CoV Orf8 protein all were found to be identical. This sample obviously does not exhibit all the variation abilities, but it lets achieve the sequence to be relatively conserved in human samples.

Joint Orf8 sequences of SARS-like bat coronaviruses are divided to three completely different types. The most numerous cluster is including 10 closely homologous sequences possessed 17 polymorphic positions, 11 of them in the N-terminal region, if to call so 84 N-terminal residues basing on the length of presumably homologous Orf8b.

Aligning 13 2019-nCoV Orf8 sequences, 2 close bat and 2 pangolin ones, 15 polymorphic positions were found in 84 aa arbitrary N-terminal region, besides 16 aa deletion in one of the pangolin sequences.

These four groups of protein sequences, aligned and roughly evaluated in terms of their variability, were aligned to visualize common features of their regions presumably corresponding to SARS-CoV Orf8b. The most characteristic sequences were selected to show them, giving an idea of the overall picture (fig. 2).

Thus, sequences of joint SARS-CoV and SARS-CoV like Orf8 from samples taken from human (hS), civet (cS) and bat (bS), of human SARS Orf8b (8b), of 2019-nCoV and nCoV-like viruses Orf8 from human (hC), bat (bC) and pangolin (pC) were included in the multiple alignment.

SARS-CoV Orf8b has strong homology with some of SARS-like viruses from bats, weak homology with Orf8 from 2019-nCoV and related coronaviruses, and very little similarity to the joint Orf8 (Orf8ab) from some SARS-CoV forms from human and their close relatives from bat and civet. With SARS-CoV with joint Orf8 it shares essentially different positions then with 2019-nCoV. The proteins in our sampling share universally only two amino acid positions: Cys40 (Cys83 by numeration of 2019-nCoV) and Leu56 (Leu98).

Such a way, Orf8 protein from SARS pathogens are absolutely divergent not only by splitting of Orf8 coding sequence but also by the history of this sequence. It may be closer to 2019-nCoV in the splitted variant and very

far from it in the joint one. However, in both cases the similarities do not seem completely irregular.

Percentage of identity between Orf8b and parts of SARS-CoV and 2019-nCoV Orf8 aligned with it is only 17% for SARS fused form and 25% for 2019-nCoV protein (tab. 1). The most impressing is completely different electric charge of both types SARS-CoV and 2019-nCoV molecules, that mirrors isoelectric point (pI) value. The indexes related to hydrophobicity also differ strongly. These data do not testify in favor of functional homology between SARS-CoV and 2019-nCoV Orf8 proteins, even to a greater extent than low levels of identity.

It is known that the capabilities of the joint Orf8 and the splitted proteins to bind other viral proteins are different [21]. Keng and Tan deduce from these data some significant conformational rearrangement that occurred because of transition from single to splitted protein [14]. But then only the data about interactions between SARS-CoV proteins were used. Adopting the version of functional homology between 2019-nCoV Orf8 and its SARS-CoV counterparts, we should turn attention to the conclusions of Wong and coauthors, that postulate direct interaction between SARS-CoV Orf8b and Orf8ab with human IRF3 [15]. So, we might expect that their counterpart in 2019-nCoV Orf8 shares this function, but too different protein parameters question this point.

The obtained data do not claim to be the final solution to this question, but they draw attention to significant differences among coronaviruses designated as SARS-CoV, as well as to the deep divergence of Orf8 proteins. In this regard, attention should be paid to the independent adaptation of SARS-CoV and 2019-nCoV to reproduction in human cells. In the case of commonality of their functions, we have an example of convergent evolution, which went significantly different ways. If the functions have changed, experimental studies can help clarify this complex picture. The expectation of structural information about the considered proteins becomes more intriguing in the presented context.

Table 1. Comparison of identity percentage between SARS-CoV ORF8B (8B), another SARS-CoV joint ORF8 (HS, designation as on alignment on fig. 2) and 2019-nCoV ORF8 (HC), and some calculated parameters of these sequences.

	Length: full (identity) , homolo- gous part (identity)	pI	Aliphatic index	GRAVY
8b	84	9.45	88.21	-0.029
hS	122 (11%)	8.24	96.72	0.218
	84 (17%)	8.55	88.33	0.038
hC	121 (17%)	5.42	97.36	0.219
	81 (25%)	5.18	99.75	0.143

```

hS MKIILFLTLIVFTSCEL--YHYQECVRGTTVLLKEPCPSGTYEGNS-----PFHPLADNKF 54
cS MKIILFLTLIVFTSCEL--YHYQECVRGTTVLLKEPCPSGTYEGNS-----PFHPLADNKF 54
bS MKIILFLTLIALASCEL--YHYQECVRGTTVLLKEPCPSGTYEGNS-----PFHPLADNKF 54
8b -----MCLKLLVRYNTRGNTYSTAWLC--ALGKVLPPH--RWHTM 36
bS MKLLIVLTCISLCS CIR--TVVQRCA SNKPHVLEDPCPTGYQPEWNI RYNTRGNTYSTAWLC--ALGKVLPPH--RWHTM 74
bS MKFLIVMTCISLCC CIR--TVVQRCA SNTPYVLENPCPTGYQPEWNI RYNTRGNTYSTAWLC--ALGKVLPPH--RWHTM 74
bS MKLLIVFGLLTSVYCIHKESIQECCENQPYQIEDPCPIHYSDWFIKIGSRKSARL-VQLCEGDYGKRIPHYEMFGNY 79
hC MKFLVFLGIITTVAAFHQECSLQSQCTQHQP YVVDDPCPIHFYSKWI RVGARKSAPL-IELCVDEAGSKSPIQYIDIGNY 79
bC MKFLVFLGILTTVAAFHQECSLQSCAQHQPYVVDDPCPIHFYSRWYIRVGARKSAPL-IELCVDEVGSKSPIQYIDIGNY 79
bC MKLLVFLGILTTVTAFAHQECSLQSCAQHQPYVVDDPCPIHFYSKWI RVGARKSAPL-IELCVDEVGSKSPIQYIDIGNY 79
pC MKFLVLLGILTTVHTFHQECSLQSCQFNSPYVVDDPCPIHFYSKWI RVGARKSAPL-IELCVDEVGSKTPIKYIDIGNY 79
pC MKFLVFLGILTTVTAFAHQECSLQSCAQHQPYVVDDPCPIHFYSRWYIRVGARKSAPL-I-----DIGNY 63

hS ALTCTSTHFACADGTRHTYQLRARSVSPKLFIRQEEVQOELY SPLFLIVAALVFLILCFTIKR KTE 122
cS ALTCTSTHFACADGTRHTYQLRARSVSPKLFIRQEEVQOELY SPLFLIVAALVFLILCFTIKR KTE 122
bS ALTCTSTHFACADGTRHTYQLRARSVSPKLFIRQEEVHVELY SPLFLIVAALVFLITL CFTIKR KTE 122
8b VQTCTP---NVTINCQDPAGGALIARC-----WYLHEGHQTAAFRDVLV-----VLNKR TN- 84
bS VQTCTP---NVTINCQDPVGGALIARC-----WYLHEGHQTAAFRDILV-----VLNKR TN- 122
bS VQACTP---NVTINCQDPVGGALVARC-----WYFYQGPQTATFRDIHV-----DLFFKRT- 122
bS TISCEP---LEINCQAPPVGS LVVRC-----SYDYDFVE---HHDV RVV--LDFI----- 121
hC TVSCLP---FTINCQEPKLGSLVVRC-----SFYEDFLE---YHDV RVV--LDFI----- 121
bC TVSCSP---FTINCQEPKLGSLVVRC-----SFYEDFLE---YHDV RVV--LDFI----- 121
bC TVSCSP---FTINCQEPKLGSLVVRC-----SFYEDFLE---YHDV RVV--LDFI----- 121
pC TVSCSP---FTINCQEPKLGSLVVRC-----SFYEDFV D---YHDV RVV--LDFI----- 121
pC TVSCSP---FTINCQEPKLGSLVVRC-----SFYEDFLE---YHDV RVV--LDFI----- 105
    
```

Fig. 2. Aligned sequences of joint SARS-CoV and SARS-CoV like Orf8 from samples taken from human (hS), civet (cS) and bat (bS), of human SARS Orf8b (8b), of 2019-nCoV and nCoV-like viruses Orf8 from human (hC), bat (bC) and pangolin (pC). Amino acid residues matching with those of SARS-CoV Orf8b (8b) are marked gray.

References

- L.P. Shaw, A.D. Wang, D. Dylus [et al.], *Mol. Ecol.* **29**, 1 (2020)
- Z. Song, Y. Xu, L. Bao, L. Zhang [et al.], “Thrusting Coronaviruses into the Spotlight”, From SARS to MERS Viruses, **11** (2019).
- Y. Yin, R.G. Wunderink, *Respirology* **23**, 130 (2018)
- E. de Wit, N van Doremalen, D Falzarano, V.J. Munster, *Nat Rev Microbiol.* **14**, 523 (2016)
- C. Huang, Y. Wang, X. Li [et al.], *Lancet* **395**, 497 (2020.)
- K. Xiao, J. Zhai, Y. Feng, [et al.], *Nature online-publication* (2020)
- L. Wahba, N. Jain, A. Fire [et al.], *bioRxiv online-publication* (2020)
- T. Lam, M. Shum, H.-H. Zhu, [et al.], *Nature* **583**, 282 (2020)
- W. Ji, W. Wang, X. Zhao, J. Zai, X. Li, *J. Med. Virol.* **92**, 433 (2020)
- P. Zhou, X.-L. Yang, X.-G. Wang [et al.], *Nature*, **579**, 270 (2020)
- Z. Song, Y. Xu, L. Bao [et al.], *Viruses* **11**, (2019)
- Y. Z. Zhang, E. C. Holmes, *Cell* **181**, 223 (2020)
- J. Cui, F. Li, Z.L. Shi, *Nat. Rev. Microbiol.* **17**, 181 (2019)
- CT. Keng, YJ. Tan, Berlin, Heidelberg, 177-191, (2009)
- H.H. Wong, T.S. Fung, Sh. Fang [et al.], *Virology* **515**, 165 (2018)
- K.A. Fitzgerald, S.M. McWhirter, K.L. Faia, [et al.], *Nat. Immunol.* **4**, 491 (2003)
- L. van Dorp, M. Acman, D. Richard [et al.], *Infect. Genet. Evol.* **83**, (2020)
- F. Sievers, A. Wilm, D. Dineen [et al.], *Mol. Syst. Biol.* **7**, (2011)
- S. Altschul, W. Gish, W. Miller, E. Myers, D. Lipman, *J. Mol. Biol.* **215**, 403 (1990)
- E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, *Protein Identification and Analysis Tools on the ExPASy Server*, (John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press, 2005)
- C.T. Keng, Y.W. Choi, M.R. Welkers, [et al.], *Virology* **354**, 132 (2006)