In Silico Identification of Nsp12 and Nsp13 as Potential Targets for Development of Broad-spectrum Antiviral Agents Against SARS-CoV, MERS-CoV, and 2019-nCoV (SARS-CoV-2)

Shengjie Dong

Guangdong Baiyun University

Jiachen Sun

Tianjin University

Zhuo Mao

Tianjin University

Yilin Lu (✉ yilinlu@tju.edu.cn)

College of New Energy, Bohai University, Jinzhou 121007, China

Jiesen Li

Foshan University

Research

Abstract

Background

2019 novel coronavirus (2019-nCoV) is officially named severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2), and is a positive-sense, single-stranded RNA coronavirus. The virus is the pathogen of coronavirus disease 2019 (COVID-19) and is infectious through human-to-human transmission. The fact that 2019-nCoV is very close to SARS-CoV has been proved by several evidences, but there are significant differences between MERS-CoV and them. Therefore, in this work, we used MERS-CoV as a probe to find the homology proteins with conserved sequences among these three known human highly pathogenic coronaviruses.

Methods

The primary protein sequences of three viruses translated from the complete genome were downloaded from National Center for Biotechnology Information (NCBI). The sequence alignments of ORF1ab proteins of three viruses were done by using Clustal Omega. The assessments of the feasibility of homology modeling were performed by using SWISS-MODEL.

Results

Here, by using computational biology, we propose that four nonstructural proteins nsp12, nsp13, nsp14, and nsp16 exhibit considerable homology among SARS-CoV, MERS-CoV, and 2019-nCoV. Among them, nsp12 and nsp13 amino acid sequences are more conserved. Considering the crucial role of these two proteins in the process of virus invasion and pathological response, we first proposed these two proteins as priority targets to design new or screen existing broad-spectrum antiviral drugs. The high consistency of primary sequence indicates the great similarity of three-dimensional structure and similar targets are likely to be inhibited by the same inhibitor. The inhibitors designed for these targets are likely to have broad-spectrum antiviral effect.

Conclusion

Very recently, some clinical trial reports preliminarily proved that Favipiravir and Remdesivir are effective for COVID-19. These clinical data provide some proof and basis for our conjecture in some degree. It is believed that the effective broad-spectrum antiviral drugs are not only helpful for the current epidemic situation, but also more beneficial for the future unpredictable epidemic situation.

1. Introduction And Background

Coronavirus is a general term for a large class of viruses. Under the electron microscope, it can be observed that their surfaces have similar corona-like protuberances, which look like the crown, so they are named as coronavirus. The diameter of coronavirus is about 80-120 nm, which is a large class of viruses widely existing in nature. So far, scientists have found about 15 different coronaviruses, seven of which

can infect people, but not necessarily cause pneumonia, but cause colds, upper respiratory tract infections and other diseases. Seven of the coronaviruses found so far can sicken people. In addition to SARS-CoV, MERS-CoV and 2019-nCoV, four other human coronaviruses (HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1) usually cause mild or moderate upper respiratory diseases, such as cold.

In November 2002, a viral severe acute respiratory syndrome (SARS) first appeared in Guangzhou. It spread rapidly to other countries, causing more than 8000 people to be infected, with a mortality rate of 9.6%. This pathogen is named severe acute respiratory syndrome coronavirus (SARS-CoV), which is a β-coronavirus from the *Chiroptera*, then infects the *Paguma larvata* (the intermediate host), and finally infects the humans. In 2017, Shi *et al.* reported the abundant bat related coronavirus gene pool found in Yunnan, China.[1] In June 2012, SARS-like respiratory diseases caused by Middle East respiratory syndrome coronavirus (MERS-CoV) occurred in Saudi Arabia. Although the interpersonal transmission of MERS-CoV is limited, it has caused outbreaks in Saudi Arabia (2012) and South Korea (2015), with more than 2000 confirmed cases worldwide and a mortality rate of about 35%. Similar to SARS-CoV, MERS-CoV also comes from the *Chiroptera*, but their intermediate host is *Camelus dromedarius*. In January 2020, new coronavirus 2019-nCoV was found and confirmed. 2019-nCoV is a kind of coronavirus which is different from the one that caused severe acute respiratory syndrome (SARS) in 2003 and the one that caused Middle East respiratory syndrome (MERS) in 2012-2015. SARS-CoV, MERS-CoV, and 2019-nCoV are more serious than four human coronaviruses which only cause mild to moderate respiratory diseases, and their genomes are similar to each other.[2] At present, there is no vaccine or specific antiviral drug for SARS-CoV, MERS-CoV, or 2019-nCoV.

In general, coronaviruses may originate from animals, and in the long evolution, they form pathogens that infect animals and humans respectively. Nevertheless, at present, the origin of the coronavirus is not completely clear. The origin of coronavirus can be traced back to about 8000 BC. Early recognition is that non cold-blooded flying animals, such as bats and birds, are the best hosts (source hosts) of coronaviruses, that is, the relationship between viruses and them is parasitic and symbiotic. Based on the early serological and subsequent genomic evidence, the coronavirus subfamily can be divided into four major genera: α, β, γ, and δ. The mammalian coronaviruses are mainly α- and β-coronaviruses, which can infect a variety of animals including pigs, dogs, cats, mice, cattle, horses, etc. The avian coronaviruses mainly comes from the coronaviruses of γ and δ genera, which can cause the diseases of many kinds of birds such as chickens, turkeys, sparrows, ducks, geese, pigeons, etc. Bats carry alpha and beta coronaviruses; birds carry gamma and delta coronaviruses.[3] In particular, β-coronavirus can be divided into four lineages: A, B, C, and D. Among the six previously known human coronaviruses (HCoVs), HCoV-229E and HCoV-NL63 belong to α-coronavirus, while HCoV-OC43 and HCoV-HKU1 belong to pedigree A of β-coronavirus, SARS-CoV belongs to pedigree B of β-coronavirus, and MERS-CoV belongs to pedigree C of β-coronavirus. Research shows that 2019-nCoV and SARS-CoV have about 80% similarity in genome sequence. At amino acid level, most of the 2019-nCoV genome coding proteins are consistent with SARS-CoV. In contrast, 2019-nCoV differs greatly from MERS-CoV, and is not closely related to MERS-CoV.[4,5] For

example, the receptors of SARS-CoV and 2019-nCoV are both angiotensin converting enzyme 2 (ACE2), while the receptor of MERS-CoV is dipeptidyl peptidase 4 (DPP4).

It is believed that the research achievements from bioinformatics and computational biology would be important in the process of resistance to new coronavirus.[6-11] In this work, using bioinformatics online tools and databases, we evaluate the sequence consistency between SARS-CoV, MERS-CoV, and 2019-nCoV. And then, we look for proteins with similar structures encoded and expressed by SARS-CoV, MERS-CoV, and 2019-nCoV. In other words, we try to find a common target of these three viruses, which can help to develop broad-spectrum antiviral drugs against these three viruses. As is known to all, when an unprecedented virus appears in human society, it is often too late to develop new antiviral drugs to fight against the epidemic and save lives. Therefore, the development of broad-spectrum antiviral drugs is not only helpful for the current epidemic situation, but also more beneficial for the future unpredictable epidemic situation.

2. Materials, Methods, And Procedures

2.1 Experimental Materials

The primary protein sequences of three viruses translated from the complete genome were downloaded from National Center for Biotechnology Information (NCBI). The GenBank serial numbers of SARS-CoV, MERS-CoV, and 2019-nCoV are AY390556.1, JX869059.2, and MN908947.3, respectively. The ORF1ab primary sequences with details of SARS-CoV and MERS-CoV were downloaded from Protein Information Resource (PIR). The UniProtKB serial numbers of SARS-CoV and MERS-CoV are P0C6X7 and K9N7C7, respectively.

2.2 Computational procedures

In this work, we did sequence alignments of ORF1ab proteins of three viruses using Clustal Omega.[12-14] First of all, the ORF1ab proteins of SARS and MERS viruses were sequenced. Then, the ORF1ab proteins of SARS-CoV and 2019-nCoV were sequenced. Finally, we have done the corresponding sequence alignment of MERS-CoV and 2019-nCoV. Then, we discussed the sequence consistency of S protein of these three viruses. First, the S proteins of SARS and MERS viruses were sequenced. Then, the S proteins of SARS-CoV and 2019-nCoV were sequenced. Finally, we have done the corresponding sequence alignment of MERS-CoV and 2019-nCoV. Next, we found that the length of ORF3 protein of MERS virus was significantly shorter than that of SARS virus and 2019-nCoV. Thus, we only performed the sequence alignment between SARS-CoV and 2019-nCoV. Then we focused on the sequence consistency of E protein of these three viruses and then we examined the homology of M protein of three viruses. At last, we examine the similarity of N protein of three viruses. Here, using MERS-CoV as a probe, we identified proteins with structural similarity between SARS-CoV, MERS-CoV, and 2019-nCoV. The identifications were performed using SWISS-MODEL.[15-17]

3. Results, Main Findings, And Discussion

3.1 Sequence alignments among SARS-CoV, MERS-CoV, and 2019-nCoV

First of all, we found that for the first 2000 amino acids, the sequence consistency of SARS-CoV ORF1ab and MERS-CoV ORF1ab was relatively low. For amino acids 2000-4000, the sequence consistency was greatly improved. For amino acids after 4000, the sequence consistency is very high (see Figure S1). Then, we found that the sequence of SARS-CoV ORF1ab and 2019-nCoV ORF1ab is surprisingly consistent. The only significant mismatch occurs around amino acid 1000. There are about 25 amino acids in 2019-nCoV ORF1ab having no corresponding sequences in SARS-CoV ORF1ab (see Figure S2). Finally, for MERS-CoV and 2019-nCoV, we found that for the first 1500 amino acids, the mismatch was obvious. For amino acids 1500-4000, the sequence consistency was significantly improved. For amino acids after 4000, the sequence consistency has been considerable (see Figure S3). It can be seen that SARS-CoV ORF1ab and 2019-nCoV ORF1ab are closely related. The differences between SARS-CoV ORF1ab and MERS-CoV ORF1ab are similar to those between MERS-CoV ORF1ab and 2019-nCoV ORF1ab.

Next, we found that, for the S proteins of SARS and MERS viruses, the overall similarity was acceptable, among which the consistency of the first 600 amino acids was relatively low, and that of the last 600 amino acids was relatively high (see Figure S4). Then, we found that, for the S proteins of SARS-CoV and 2019-nCoV, the overall similarity was quite high, especially the amino acids after 600 showed a high degree of consistency (see Figure S5). Finally, we found that, for MERS-CoV and 2019-nCoV, the overall consistency was relatively low, and the sequence similarity after 800 was higher than that before 800 (see Figure S6).

We found that the ORF3 protein of SARS-CoV is highly homologous with ORF3a protein of 2019-nCoV (see Figure S7). Then we found that the E protein of SARS-CoV is highly homologous with that of 2019-nCoV, while the E protein of MERS-CoV is not very homologous with both (see Figures S8-S10). We also found that the M protein of SARS-CoV and 2019-nCoV was highly homologous, while the M protein of MERS-CoV was slightly less consistent with the above two cases (see Figures S11-S13). We still get similar results, that is, the N protein of SARS-CoV is highly homologous with that of 2019-nCoV, while the N protein of MERS-CoV is slightly less homologous with that of the two viruses (see Figures S14-S16).

3.2 Searching for proteins from MERS-CoV which are homologous with those from SARS-CoV and 2019-nCoV

The 2019 novel coronavirus encodes at least 27 proteins, including 15 nonstructural proteins (nsp1-nsp10, nsp12-nsp16), 4 structural proteins (S, E, M, and N) and 8 accessory proteins (3a, 3b, p6, 7a, 7b, 8b, 9b, and ORF14).[4,18,19] Although the protein composition of 2019-nCoV is mostly the same as that of SARS-CoV, there are also some differences. For example, 2019-nCoV lost the 8a protein encoded by SARS-CoV. The length of 8b protein (121 amino acids) encoded by 2019-nCoV is longer than that of SARS-CoV (84 amino acids). There was also a significant difference in 3b protein between them.

Differences in the composition of these accessory proteins means that 2019-nCoV and the SARS-CoV that erupted before may have some differential pathogenesis. And it also illustrates the fact that these auxiliary proteins are not preferred targets for the development of broad-spectrum antiviral drugs. In the previous section, we have explained that although the four structural proteins of 2019-nCoV are very close to those of SARS-CoV, they are quite different from those of MERS-CoV. Consequently, four structural proteins are also not suitable for the development of broad-spectrum antiviral drugs, although S protein can be used as a key target for the development of specific vaccines. So next, we focus on ORF1ab protein. As we have discussed before, the homology of ORF1ab of SARS-CoV and SARS-CoV-2 (2019-nCoV) is very high, so we use ORF1ab of MERS-CoV as a probe to estimate the conservation of ORF1ab among the three viruses considered in this work.

For nonstructural protein 1 (nsp1),[20] the searching results indicate that the best template for modeling nsp1 of MERS-CoV is 2HSX_A, which is the NMR structure of the nsp1 from the SARS-CoV. However, the identity is only 24.32% (see Figure S17), demonstrating that nsp1 is not a highly conservative protein among different coronaviruses.

For nonstructural protein 2 (nsp2), no suitable template was found.

For papain-like proteinase (nsp3),[21] it is found that three parts of nsp3 were achieved experimentally. Using the sequences of these three parts, we estimated the possibility of nsp3 as a universal target. We found that the identities between the nsp3 proteins from MERS-CoV and the corresponding proteins from SARS-CoV are 45% approximately (see Figures S18 and S19). It demonstrates that nsp3 is not a highly conservative protein among different coronaviruses.

For nonstructural protein 4 (nsp4), two templates from feline coronavirus and mouse hepatitis virus are suggested, and the identities are 43.01% and 51.14%, respectively (see Figure S20). It demonstrates that nsp4 is not a highly conservative protein among different coronaviruses.

For proteinase 3CL-Pro (nsp5), we found that the identity between nsp5 from MERS-CoV and that from SARS-CoV is 52.98% (see Figure S21). It should be noted that nsp5 is highly conserved between MERS-CoV and some other coronaviruses. Nevertheless, these coronaviruses are not human highly pathogenic coronaviruses. Therefore, we will not focus on this target here.

For nonstructural protein 6 (nsp6), no suitable template was found.

For nonstructural protein 7 (nsp7), two templates from SARS coronavirus and feline coronavirus are suggested, and the identities are 55.42% and 40.96%, respectively (see Figure S22). It demonstrates that nsp7 is not a highly conservative protein among different coronaviruses.

For nonstructural protein 8 (nsp8), two templates from SARS coronavirus and feline coronavirus are suggested, and the identities are 53.30% and 44.85%, respectively

(see Figure S23). It demonstrates that nsp8 is not a highly conservative protein among different coronaviruses.

For nonstructural protein 9 (nsp9), two representative templates from SARS coronavirus and human coronavirus 229E are suggested, and the identities are 53.64% and 45.87%, respectively (see Figure S24). It demonstrates that nsp9 is not a highly conservative protein among different coronaviruses.

For nonstructural protein 10 (nsp10), we found that the identity between nsp10 from MERS-CoV and that from SARS-CoV is 59.42% (see Figure S25). Therefore, we do not think that nsp10 is a good choice as a universal target for the design of broad-spectrum antiviral drugs.

For RNA-directed RNA polymerase (nsp12),[22] we found that the identity between nsp12 from MERS-CoV and that from SARS-CoV is 72.14% (see Figure S26). It indicates that nsp12 would be a wonderful target for the development of broad-spectrum antiviral drugs against human highly pathogenic coronavirus. We also found that the identity between MERS-CoV and foot and mouth disease virus RNA-dependent RNA polymerase is 14.55%. This shows that the conservation of RdRp among coronaviruses is much higher than that among different types of viruses. We examined the current structural biology achievements of RdRp from these three viruses. We found that for SARS-CoV and SARS-CoV-2, a hetero-oligomeric complex with nsp7 and/or nsp8 is available. Experimental structures of hetero-oligomeric complexes exist. For MERS-CoV, the protein structure has not been achieved experimentally and thus should be predicted via homology modeling. The results are summarized in Table 1 and Figure 1.

For helicase (nsp13),[23-26] we found that the identity between nsp13 from MERS-CoV and that from SARS-CoV is 72.37% (see Figure S27). It indicates that nsp13 would be also a wonderful target for the development of broad-spectrum antiviral drugs against human highly pathogenic coronavirus. We examined the current structural biology achievements of helicase from these three viruses. For all these three viruses, experimental structures were realized experimentally. The results are summarized in Table 2 and Figure 2.

For guanine-N7 methyltransferase (nsp14),[27-29] we found that the identity between MERS-CoV nsp14 and SARS-CoV nsp14 is 63.22% (see Figure S28). This indicates that nsp14 is also a potential target for the development of broad-spectrum anti-coronavirus drugs. We examined the current structural biology achievements of proofreading exoribonuclease from these three viruses. We found that for SARS-CoV, the protein is structurally achieved while for MERS-CoV and SARS-CoV-2, the protein could be only realized by using homology modeling. The results are summarized in Table 3 and Figure 3.

For uridylate-specific endoribonuclease (nsp15), we found that the identity between MERS-CoV nsp15 and SARS-CoV-2 (2019-nCoV) nsp15 is 51.62%; the identity between MERS-CoV nsp15 and human coronavirus 229E nsp15 is 47.04% (see

Figure S29). Therefore, we do not think that nsp15 is a good choice as a universal target for the design of broad-spectrum antiviral drugs.

For 2'-O-methyltransferase (nsp16),[28,30] we found that the identity between MERS-CoV nsp16 and SARS-CoV nsp16 is 65.32% (see Figure S30). This indicates that nsp16 is also a potential target for the development of broad-spectrum anti-coronavirus drugs. We examined the current structural biology achievements of 2'-O-methyltransferase from these three viruses. We found that for all these three viruses, the protein structure has been resolved experimentally. The results are summarized in Table 4 and Figure 4.

3.3 Is it possible to use structural proteins as general targets?

Some HCoVs have used cell surface enzymes as receptors, such as SARS-CoV receptor ACE2, MERS-CoV receptor DPP4. Generally, S protein of coronavirus is further cleaved into S1 and S2 subunits by host protease, and S1 / S2 cleavage is mediated by one or more host protease. For example, the activation of S protein of SARS-CoV requires sequential cleavage by endocystine cathepsin L and another trypsin-like serine protease. Unlike SARS-CoV, the S protein of MERS-CoV contains two furin cleavage sites. Host factors may also limit the attachment and entry of HCoV.

It can be seen that the S protein is very important for virus attachment and entry. Nevertheless, we found that the identity between MERS-CoV S protein and SARS-CoV-2 (2019-nCoV) S protein is 30.26% (see Figure S31). In addition, we also found that the identity between SARS-CoV S protein and SARS-CoV-2 (2019-nCoV) S protein is 76.47% (see Figure S32). This also shows that the molecular mechanism of SARS-CoV and SARS-CoV-2 using S protein to invade cells is the same, but MERS-CoV using S protein to invade cells uses a different molecular mechanism. Therefore, we think that S protein, as the most important structural protein, should not be used as a general target of coronavirus, because it is so different in different kinds of coronavirus.

Next, we discussed the other three structural proteins. For E protein, the identity between SARS-CoV-2 and SARS-CoV was only 32.76% (see Figure S33). The research of M protein structure biology is always blank, so we have not got valuable data. For N protein, the identity between SARS-CoV-2 and SARS-CoV was less than 60% (see Figure S34). Therefore, these three structural proteins are not ideal targets for the development of broad-spectrum anti coronavirus drugs.

4. Outlook And Conclusions

In this work, we identified four proteins with high conservation in three known human highly pathogenic coronaviruses. The four proteins are RNA-directed RNA polymerase, helicase, guanine-N7 methyltransferase, and 2'-O-methyltransferase, respectively. We found that these four proteins are very important nonstructural proteins of coronavirus, which play important roles in the process of viral transcription and replication as well as escape from the immune system. We speculate that the inhibitors designed for these four proteins are likely to have a broad-spectrum anti-coronavirus effect. In particular,

we think that RNA-directed RNA polymerase and helicase should be paid more attention first because their sequence conservation and structural similarity between different coronaviruses are more prominent. Moreover, some recent reports on Favipiravir and Remdesivir seem to be clinically effective encouraged our point of view. The genetic material of RNA virus is RNA (RNA ribonucleic acid), and thus for the replication of RNA virus genome, most of them need intermediate synthesis, which requires virus specific polymerase (RNA dependent RNA polymerase, RdRp, or RNA dependent DNA polymerase, RdDp). Consequently, these polymerases are ideal targets for the design of antiviral drugs. Favipiravir and Remdesivir work by interfering with RdRp, which are precursors of adenosine analogues. After triphosphate *invivo*, they would be incorporated into the newly synthesized RNA chain of the virus as a substrate, and then the synthesis of the virus genome is interrupted quickly. In addition to influenza virus, Favipiravir also showed good antiviral effect on a variety of RNA viruses, such as Ebola virus, sand virus, Bunia virus, rabies virus, etc. Remdesivir has a good effect on Ebola virus in animal experiments, and it is effective on many viruses of filoviridae, adenoviridae and coronaviridae *invitro*. Cell test showed that Remdesivir could inhibit the replication of SARS coronavirus and MERS coronavirus in human airway epithelial cells. Animal experiments also show that it is effective for SARS virus and MERS virus infection.[31-36]

Abbreviations

2019-nCoV: 2019 novel coronavirus

COVID-19: Coronavirus disease 2019

SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2

SARS-CoV: Severe acute respiratory syndrome coronavirus

MERS-CoV: Middle East respiratory syndrome coronavirus

HCoVs: Human coronaviruses

ACE2: Angiotensin converting enzyme 2

DPP4: Dipeptidyl peptidase 4

NCBI: National Center for Biotechnology Information

PIR: Protein Information Resource

PDB: Protein Data Bank

Declarations

Availability of data and materials

Data Availability Statement

The data used to support the findings of this study are included within the article.

Acknowledgements

Not applicable.

Funding

Author information

Author Contributions

SD designed the project. SD and YYL analyzed the data. SD and JL prepared the manuscript. SD and JS performed the calculations. SD and ZM discussed the results.

Ethics declarations

Ethics approval and consent to participate

Compliance with Ethics Requirements

This article does not contain any studies with human or animal subjects.

Consent for publication

Not applicable.

Competing interests

Declaration of Competing Interest

The authors declare that they have no conflict of interests.

References

1. Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, Li B, Xie JZ, Shen XR, Zhang YZ, Wang N, Luo DS, Zheng XS, Wang MN, Daszak P, Wang LF, Cui J, Shi ZL. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog*. 2017;13(11):e1006698. https://doi.org/10.1371/journal.ppat.1006698

2. Lim YX, Ng YL, Tam JP, Liu DX. Human coronaviruses: A review of virus-host interactions. *Diseases*. 2016;4(3):26. https://doi.org/10.3390/diseases4030026

3. Fung TS, Liu DX. Human coronavirus: Host-pathogen interaction. *Annu Rev Microbiol*. 2019;73:529-557. https://doi.org/10.1146/annurev-micro-020518-115759

4. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J, Sheng J, Quan L, Xia Z, Tan W, Cheng G, Jiang T. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe*. 2020;27(3):325-328. https://doi.org/10.1016/j.chom.2020.02.001

5. Xu XT, Chen P, Wang JF, Feng JN, Zhou H, Li X, Zhong W, Hao P. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci*. 2020;63(3):457–460. https://doi.org/10.1007/s11427-020-1637-5

6. Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S, Ciccozzi M. COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *J Med Virol*. 2020;1–5. https://doi.org/10.1002/jmv.25719

7. Liu Z, Xiao X, Wei X, Li J, Yang J, Tan H, Zhu J, Zhang Q, Wu J, Liu L. Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. *J Med Virol*. 2020;1–7. https://doi.org/10.1002/jmv.25726

8. Conway MJ. Identification of coronavirus sequences in carp cDNA from Wuhan, China. *J Med Virol*. 2020. https://doi.org/10.1002/jmv.25751

9. Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, Zhang Z. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J Med Virol*. 2020. https://doi.org/10.1002/jmv.25762

10. Lung J, Lin YS, Yang YH, Chou YL, Shu LH, Cheng YC, Liu HT, Wu CY. The potential chemical structure of anti-SARS-CoV-2 RNA-dependent RNA polymerase. *J Med Virol*. 2020. https://doi.org/10.1002/jmv.25761

11. Cárdenas-Conejo Y, Liñan-Rico A, García-Rodríguez DA, Centeno-Leija S, Serrano-Posada H. An exclusive 42 amino acid signature in pp1ab protein provides insights into the evolutive history of the 2019 novel human-pathogenic coronavirus (SARS-CoV2). *J Med Virol*. 2020. https://doi.org/10.1002/jmv.25758

12. Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence

alignments using Clustal Omega. *Mol Systems Biology*. 2011;7:539. https://doi.org/10.1038/msb.2011.75

13. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci*. 2018;27(1):135-145. https://doi.org/10.1002/pro.3290

14. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 2019;47(W1):W636-W641. https://doi.org/10.1093/nar/gkz268

15. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*. 2011;27:343-350. https://doi.org/10.1093/bioinformatics/btq662

16. Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, Schwede T. The SWISS-MODEL Repository—New features and functionality. *Nucleic Acids Res*. 2017;45(D1):D313-D319. https://doi.org/10.1093/nar/gkw1132

17. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296-W303. https://doi.org/10.1093/nar/gky427

18. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269. https://doi.org/10.1038/s41586-020-2008-3

19. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-273. https://doi.org/10.1038/s41586-020-2012-7

20. Lokugamage KG, Narayanan K, Nakagawa K, Terasaki K, Ramirez SI, Tseng CT, Makino S. Middle East respiratory syndrome coronavirus nsp1 inhibits host gene expression by selectively targeting mRNAs transcribed in the nucleus while sparing mRNAs of cytoplasmic origin. *J Virol*. 2015;89(21):10970-10981. https://doi.org/10.1128/JVI.01352-15

21. Báez-Santos YM, Mielech AM, Deng X, Baker S, Mesecar AD. Catalytic function and substrate specificity of the papain-like protease domain of nsp3 from the Middle East respiratory syndrome coronavirus. *J Virol*. 2014;88(21):12511-12527. https://doi.org/10.1128/JVI.01294-14

22. Ahn DG, Choi JK, Taylor DR, Oh JW. Biochemical characterization of a recombinant SARS coronavirus nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. *Arch Virol*. 2012;157:2095-2104. https://doi.org/10.1007/s00705-012-1404-x

23. Tanner JA, Watt RM, Chai YB, Lu LY, Lin MC, Peiris JS, Poon LL, Kung HF, Huang JD. The severe acute respiratory syndrome (SARS) coronavirus NTPase/helicase belongs to a distinct class of 5' to 3' viral helicases. *J Biol Chem*. 2003;278:39578-39582. https://doi.org/10.1074/jbc.C300328200

24. Adedeji AO, Marchand B, Te Velthuis AJ, Snijder EJ, Weiss S, Eoff RL, Singh K, Sarafianos SG. Mechanism of nucleic acid unwinding by SARS-CoV helicase. *PLoS One*. 2012;7(5):e36521. https://doi.org/10.1371/journal.pone.0036521

25. Adedeji AO, Singh K, Calcaterra NE, DeDiego ML, Enjuanes L, Weiss S, Sarafianos SG. Severe acute respiratory syndrome coronavirus replication inhibitor that interferes with the nucleic acid unwinding of the viral helicase. *Antimicrob Agents Chemother*. 2012;56(9):4718-4728. https://doi.org/10.1128/AAC.00957-12

26. Adedeji AO, Singh K, Sarafianos SG. Structural and biochemical basis for the difference in the helicase activity of two different constructs of SARS-CoV helicase. *Cell Mol Biol (Noisy-le-grand)*. 2012;58(1):114-121.

27. Minskaia E, Hertzig T, Gorbalenya AE, Campanacci V, Cambillau C, Canard B, Ziebuhr J. Discovery of an RNA virus 3'→5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc Natl Acad Sci U S A*. 2006;103(13):5108-5113. https://doi.org/10.1073/pnas.0508200103

28. Bouvet M, Debarnot C, Imbert I, Selisko B, Snijder EJ, Canard B, Decroly E. *In vitro* reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog*. 2010;6(4):e1000863. https://doi.org/10.1371/journal.ppat.1000863

29. Bouvet M, Imbert I, Subissi L, Gluais L, Canard B, Decroly E. RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc Natl Acad Sci U S A*. 2012;109(24):9372-9377. https://doi.org/10.1073/pnas.1201130109

30. Decroly E, Imbert I, Coutard B, Bouvet M, Selisko B, Alvarez K, Gorbalenya AE, Snijder EJ, Canard B. Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'*O*)-methyltransferase activity. *J Virol*. 2008;82:8071-8084. https://doi.org/10.1128/JVI.00407-08

31. Ferron F, Subissi L, Silveira De Morais AT, Le NTT, Sevajol M, Gluais L, Decroly E, Vonrhein C, Bricogne G, Canard B, Imbert I. Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc Natl Acad Sci U S A*. 2018;115(2):E162-E171. https://doi.org/10.1073/pnas.1718806115

32. Sakabe S, Sullivan BM, Hartnett JN, Robles-Sikisaka R, Gangavarapu K, Cubitt B, Ware BC, Kotliar D, Branco LM, Goba A, Momoh M, Sandi JD, Kanneh L, Grant DS, Garry RF, Andersen KG, de la Torre JC, Sabeti PC, Schieffelin JS, Oldstone MBA. Analysis of CD8[+]T cell response during the 2013−2016 Ebola epidemic in West Africa. *Proc Natl Acad Sci U S A*. 2018;115(32):E7578-E7586.https://doi.org/10.1073/pnas.1806200115

33. de Wit E, Feldmann F, Cronin J, Jordan R, Okumura A, Thomas T, Scott D, Cihlar T, Feldmann H. Prophylactic and therapeutic remdesivir (GS-5734) treatment in the rhesus macaque model of MERS-CoV infection. *Proc Natl Acad Sci U S A*. 2020:117(12);6771-6776. https://doi.org/10.1073/pnas.1922083117

34. Mulangu S, Dodd LE, Davey RT, Tshiani Mbaya O, Proschan M, Mukadi D, Lusakibanza Manzo M, Nzolo D, Tshomba Oloma A, Ibanda A, Ali R, Coulibaly S, Levine AC, Grais R, Diaz J, Lane HC,

Muyembe-Tamfum JJ. A randomized, controlled trial of Ebola virus disease therapeutics. *N Engl J Med*. 2019;381:2293-2303. https://doi.org/10.1056/NEJMoa1910993

35. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, Spitters C, Ericson K, Wilkerson S, Tural A, Diaz G, Cohn A, Fox L, Patel A, Gerber SI, Kim L, Tong S, Lu X, Lindstrom S, Pallansch MA, Weldon WC, Biggs HM, Uyeki TM, Pillai SK. First case of 2019 novel coronavirus in the United States. *N Engl J Med*. 2020;382:929-936. https://doi.org/10.1056/NEJMoa2001191

36. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, Shi Z, Hu Z, Zhong W, Xiao G. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) *in vitro*. *Cell Res*. 2020;30:269–271. https://doi.org/10.1038/s41422-020-0282-0

Tables

Table 1 The current structural biology achievements of nsp12 from three considered coronaviruses.

Protein description	Present situation	PDB ID
Nsp12 of SARS-CoV	Available experimentally	6NUR, 6NUS
Nsp12 of MERS-CoV	Not available experimentally	6NUR (identity: 72.01%), 7C2K (identity: 71.40%)
Nsp12 of SARS-CoV-2	Available experimentally	6M71, 6XEZ, 6XQB, 6YYT, 7BTF, 7BV1, 7BV2, 7BW4, 7BZF, 7C2K

Table 2 The current structural biology achievements of nsp13 from three considered coronaviruses.

Protein description	Present situation	PDB ID
Nsp13 of SARS-CoV	Available experimentally	6JYT
Nsp13 of MERS-CoV	Available experimentally	5WWP
Nsp13 of SARS-CoV-2	Available experimentally	6ZSL, 6XEZ

Table 3 The current structural biology achievements of nsp14 from three considered coronaviruses.

Protein description	Present situation	PDB ID
Nsp14 of SARS-CoV	Available experimentally	5C8S, 5C8T, 5C8U, 5NFY
Nsp14 of MERS-CoV	Not available experimentally	5NFY (identity: 63.22%), 5C8S (identity: 63.22%)
Nsp14 of SARS-CoV-2	Not available experimentally	5NFY (identity: 94.88%), 5C8S (identity: 95.07%)

Table 4 The current structural biology achievements of nsp16 from three considered coronaviruses.

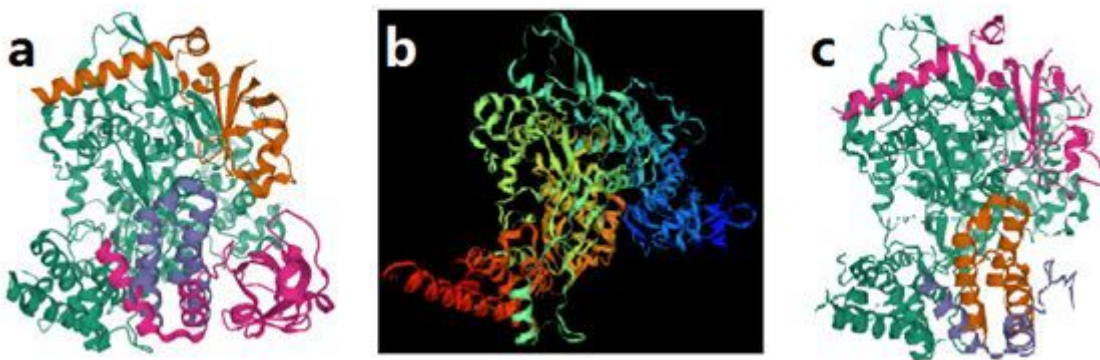| Protein description	Present situation	PDB ID
Nsp16 of SARS-CoV	Available experimentally	3R24, 2XYQ, 2XYR
Nsp16 of MERS-CoV	Available experimentally	5YN5
Nsp16 of SARS-CoV-2	Available experimentally	6W4H, 6W61, 6W75, 6WJT, 6WKQ, 6WKS, 6WQ3, 6WRZ, 6WVN, 6XKM, 6YZ1, 7BQ7, 7C2I, 7C2J

Figures

Figure 1

Diagrammatic sketch of nsp12 from three considered coronaviruses. a. SARS-CoV; b. MERS-CoV; c. SARS-CoV-2
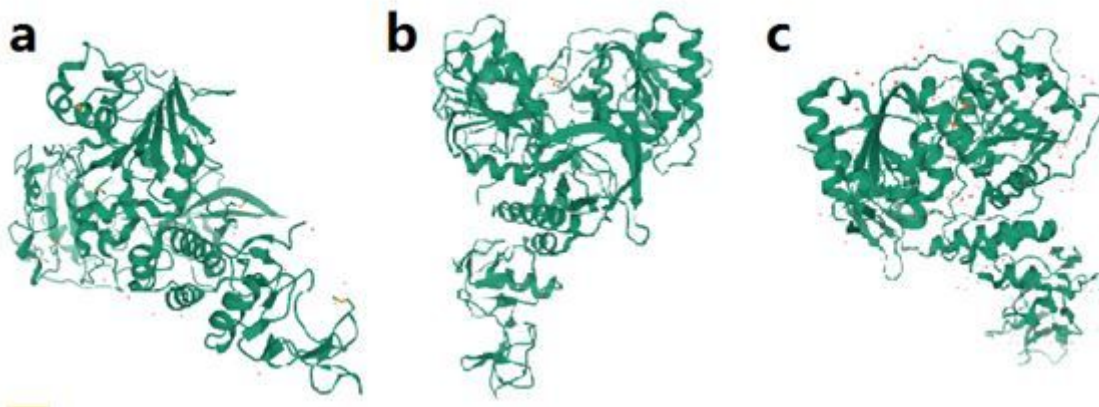
Figure 2

Diagrammatic sketch of nsp13 from three considered coronaviruses. a. SARS-CoV; b. MERS-CoV; c. SARS-CoV-2
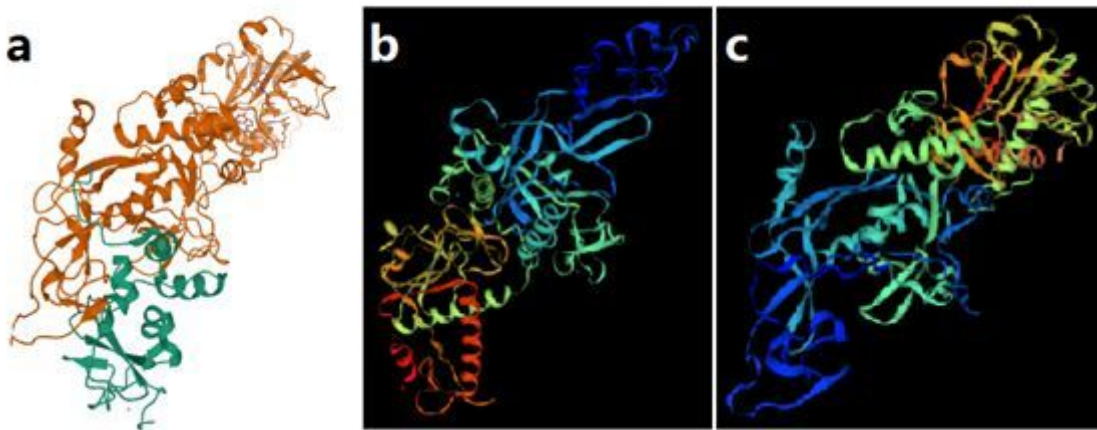

Figure 3

Diagrammatic sketch of nsp14 from three considered coronaviruses. a. SARS-CoV; b. MERS-CoV; c. SARS-CoV-2
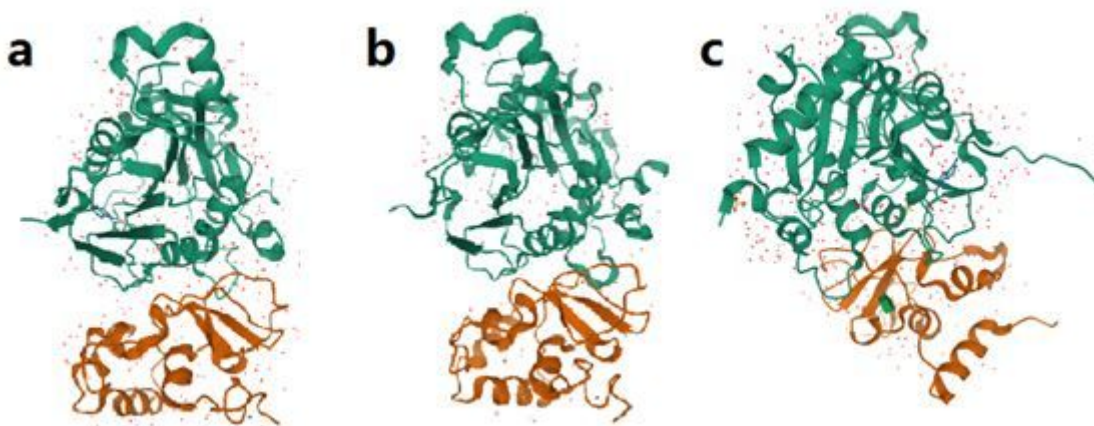
Figure 4

Diagrammatic sketch of nsp16 from three considered coronaviruses. a. SARS-CoV; b. MERS-CoV; c. SARS-CoV-2

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supportinginformationforMilitaryMedicalResearch.pdf