

## Commentary

# Expanding our understanding of the role polyprotein conformation plays in the coronavirus life cycle

 Samantha Gildenhuis

Department of Life and Consumer Sciences, College of Agriculture and Environmental Sciences, University of South Africa, Private Bag X6, Florida, Roodepoort 1710, South Africa

**Correspondence:** Samantha Gildenhuis (gildes@unisa.ac.za)



Coronaviruses are the causative agents in many globally concerning respiratory disease outbreaks such as severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS) and coronavirus disease-2019 (COVID-19). It is therefore important that we improve our understanding of how the molecular components of the virus facilitate the viral life cycle. These details will allow for the design of effective interventions. Krichel and coauthors in their article in the *Biochemical Journal* provide molecular details of how the viral polyprotein (nsp7–10) produced from the positive single stranded RNA genome, is cleaved to form proteins that are part of the replication/transcription complex. The authors highlight the impact the polyprotein conformation has on the cleavage efficiency of the main protease ( $M^{pro}$ ) and hence the order of release of non-structural proteins 7–10 (nsp7–10) of the SARS-CoV. Cleavage order is important in controlling viral processes and seems to have relevance in terms of the protein–protein complexes formed. The authors made use of mass spectrometry to advance our understanding of the mechanism by which coronaviruses control nsp 7, 8, 9 and 10 production in the virus life cycle.

## Introduction

Coronaviruses have gained much attention as the causative agents for the outbreaks of human respiratory syndromes such as severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS) and most recently coronavirus disease-2019 (COVID-19). These respiratory diseases arose from zoonotic transfer from animals to humans that have produced strains of the virus that have not previously circulated in the human population. The resulting respiratory diseases have produced high fatality rates as is the case for MERS with a case fatality rate reported by the World Health Organization of  $\approx 35\%$  in January 2020. In addition, COVID-19 has shown rapid global spread resulting in many infected individuals as cases increase above 1,200,000 and counting (7 April 2020) [1]. The need for specific treatment options targeting these coronaviruses is high and in order to provide these options detailed understanding is required regarding the structure of the molecular components that are essential for the viral life cycle. Analysis of the SARS-CoV2 (COVID-19) sequence indicated that it shows 77.2% amino acid identity with SARS-CoV (SARS) with the non-structural proteins (nsp's) 7–10 showing between 97.1% to 98.8% identity [2]. Therefore, molecular details pertaining to SARS-CoV proteins and regulation should apply broadly to the family of viruses.

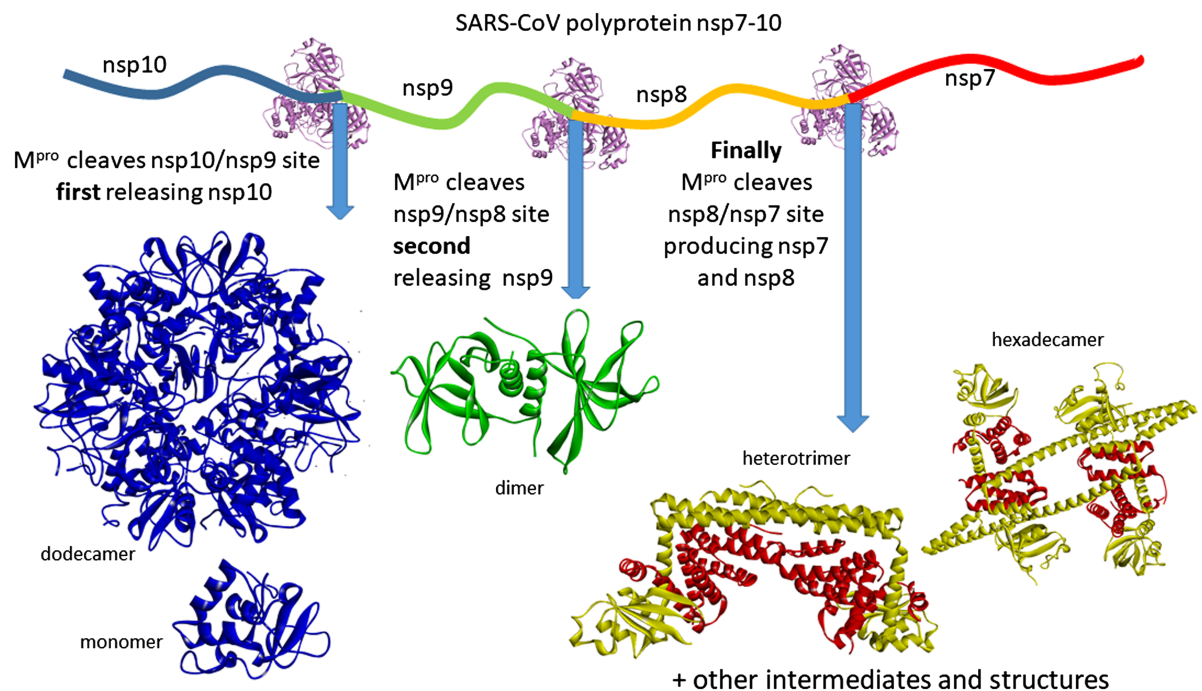
## Polyprotein nsp7–10 processing

SARS is caused by the coronavirus known as SARS-CoV which belongs to the order *Nidovirales*, the family *Coronaviridae*, the subfamily *Coronavirinae* and the genera *betacoronavirus*. These viruses are enveloped and have some of the largest single stranded RNA genomes. The SARS-CoV positive

Received: 17 March 2020  
Revised: 8 April 2020  
Accepted: 9 April 2020

Version of Record published:  
29 April 2020

stranded RNA genome encodes two open reading frames (ORF), ORF1a and ORF1b. When ORF1b is translated it produces two polyproteins (pp), pp1a and pp1ab by ribosomal frameshifting [3]. The latter has a higher degree of sequence conservation than ORF1a, most likely due to the functional significance of the proteins produced by this region [4]. The polyproteins produced from ORF1b are auto catalytically processed to produce 15 to 16 nsp's essential for viral RNA synthesis which include two proteases (nsp3, PL<sup>Pro</sup>; nsp5, M<sup>Pro</sup>) and the viral RNA dependent polymerase (nsp12, RdRp). The processing of the polyprotein region is a point of posttranslational control that is essential for virus replication. SARS-CoV uses both proteases to process polyproteins, the papain-like protease (PL<sup>Pro</sup>) and the main chymotrypsin-like protease (M<sup>Pro</sup>). Non-structural proteins 7, 8, 9 and 10 are cleaved by M<sup>Pro</sup> which recognizes specific cleavage sites. Sequence conservation is high in the nsp's within the *Coronavirinae*, which is expected, since these proteins play a critical role in the viral life cycle. They have been proposed as interacting partners for nsp12, nsp14 (exoribonuclease, ExonN) and nsp16 (ribose 2'-O-MTase) as well as RNA [5], but their exact functions are not well characterized. Krichel and coauthors in their article [6] investigate the processing of the nsp7–10 polyprotein. Their results show for the first time the critical role protein conformational structure plays in the order in which the nsp's are released from the polyprotein. Their data shows that using peptides representing joining segments between nsp variants (nsp10 with 9, nsp9 with 8 and nsp8 with 7) that the order in which nsp's are released from the polyprotein would be nsp10, then nsp7, then nsp9 from nsp8, based on the rates of the enzymatic peptide cleavage reaction. When they use a full-length polyprotein (nsp7–10) in its native conformation, M<sup>Pro</sup> cleaves the polyprotein in a different order, nsp10 is cleaved first then nsp9 and then nsp8 from nsp7 (Figure 1). Sequence analysis [7] and crystal structures [8,9] of coronavirus protease M<sup>Pro</sup> have indicated that the substrate binding pocket of this enzyme has many key amino acid residues that make many contacts with its substrate. It is likely that the need for so many key interactions is for distinguishing protein conformation of the polyprotein.



**Figure 1. Polyprotein nsp7–10 cleavage order and protein product structures.**

Krichel and coauthors [6] indicated that conformation of the nsp7–10 polyprotein resulted in nsp's being cleaved from the polyprotein by M<sup>Pro</sup> (PDB code 1UK3 [8]—purple) in order; nsp10 then nsp9 then nsp8 from nsp7. Structures available for the nsp's are depicted as ribbon structures visualized using Discovery Studio software version 19.1.0.18287 (Accelrys, San Diego, U.S.A.) in corresponding colors. Blue represents nsp10, the dodecamer structure (PDB 2G9T) [11] as well as monomer structure (PDB 2FYG) [12]. Green represents nsp9, the dimer structure (PDB 1QZ8) [14]. Yellow represents nsp8 and red represents nsp7 with hexadecamer (PDB 2AHM) [15] and heterotrimer (PDB 3UB0) [10] structures. Krichel and coauthors also reported a potential hetero-tetramer (nsp7:nsp8 2:2) which indicates more protein complexes could be present.

## The order in which nsp's are released links to different functional quaternary structures formed

The polyprotein chain has been shown to contain sufficient structure to catalyze RNA synthesis in the uncleaved form [10]. It is therefore not surprising that the polyprotein adopts specific tertiary structural characteristics that modulate the enzymatic cleavage by  $M^{pro}$ . The conformational structure formed by the polyprotein nsp9–7 by design causes the nsp's to be cleaved in what seems to be a logical order. One would expect that these nsp's are cleaved in an order that reflects their associations with one another and other key proteins in line with the viral needs during the production of new viruses. Nsp10 and nsp9 have no reported structures involving either nsp7 or nsp8 (Figure 1) this was also confirmed by Krichel and coauthors [6] therefore it is not surprising that it is released from the polyprotein first as it requires no other partners to carry out its function. A spherical dodecameric structure has been reported that is composed of trimeric units of nsp10 alone [11] and a monomeric structure of nsp10 has also been reported [12]. Both structures show the presence of zinc finger motifs. Proteins containing zinc finger motifs are often associated with DNA or RNA binding functions indicating the potential role nsp10 plays in RNA synthesis and due to it being cleaved first this role must be required prior to the other nsp's. Similarly, nsp9 has been shown to form dimeric structures that bind RNA [13,14] (Figure 1). Krichel and coauthors also confirmed the presence of nsp9 monomers and dimers in solution but no complexes of nsp9 with the other nsp's [6]. Therefore, it seems logical that nsp9 will be cleaved next rather than nsp7. Studies have shown that nsp8 and nsp7 interact and form different structures (Figure 1), Zhai and coauthors [14] reported the formation of a hexadecamer that possesses two structural conformations of nsp8 termed the 'golf club' and 'golf club with a bent shaft'. The hexadecamer forms a doughnut shape structure while other reports indicated the formation of a heterotrimer consisting of two copies of nsp7 and one of nsp8 [10]. The nsp7/nsp8 heterotrimer structure has been shown to produce RNA pointing to a role as an RNA polymerase [10]. It therefore makes sense that nsp7–8 would be cleaved last producing the proteins that can then form various quaternary structures. Krichel and coauthors discovered the presence of a novel arrangement composed of two nsp7 and two nsp8 monomers in a hetero-tetramer [6]. They also found that the complex could dissociate forming complexes with one nsp7 and two nsp8 monomers as well as two nsp7 and one nsp8 monomer. It would therefore appear based on the variety of structures reported for nsp7 interacting with nsp8 that these proteins form multiple functional quaternary structures and possibly adopt many stable intermediate structures.

## Conclusion

Coronavirus continue to pose a very real danger globally. Our detailed understanding of the proteins encoded by these viral genomes and the posttranslational control mechanisms will be critical in the development of interventions. Although literature provides detailed structures for the nsp7, 8, 9 and 10 it is evident that many possible forms exist. These different structural forms could be part of a detailed posttranslational control mechanism and it would be important to determine the function of these different structures, be that enzymatic or as key intermediates driving virus formation. In the paper by Krichel and coauthors, they also highlight the importance of understanding polyprotein structure and its role in the viral life cycle. The virus seems to employ multiple levels of control and each protein could play multiple roles. Intermediates may also be off pathway structures used to modulate rates of reactions. Knowledge of the existence of structures would then provide the information necessary to understand the molecular details of the role these proteins play in the viral life cycle and would thus provide information that would allow manipulation of the viral proteins for control and treatment options.

## Competing Interests

The author declares that there are no competing interests associated with this manuscript.

## Abbreviation

COVID-19, coronavirus disease-2019; MERS, Middle East respiratory syndrome; ORF, open reading frames; SARS, severe acute respiratory syndrome

## References

- 1 WHO. Novel Coronavirus (2019-nCoV) situation report-77. [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200406-sitrep-77-covid-19.pdf?sfvrsn=21d1e632\\_2](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200406-sitrep-77-covid-19.pdf?sfvrsn=21d1e632_2) (Accessed on 7 April 2020)

- 2 Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G. et al. (2020) A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 <https://doi.org/10.1038/s41586-020-2008-3>
- 3 Brierley, I., Digard, P. and Inglis, S.C. (1989) Characterisation of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**, 537–547 [https://doi.org/10.1016/0092-8674\(89\)90124-4](https://doi.org/10.1016/0092-8674(89)90124-4)
- 4 Hu, L.D., Zheng, G.Y., Jiang, H.S., Xia, Y., Zhang, Y. and Kong, X.Y. (2003) Mutation analysis of 20 SARS virus genome sequences: evidence for negative selection in replicase ORF1b and spike gene. *Acta Pharmacol. Sin.* **24**, 741–745 PMID:12904271
- 5 Snijder, E.J., Decroly, E. and Ziebuhr, J. (2016) The nonstructural proteins directing coronavirus RNA synthesis and processing. *Adv. Virus Res.* **96**, 59–126 <https://doi.org/10.1016/bs.aivir.2016.08.008>
- 6 Krichel, B., Falke, S., Hilgenfeld, R., Redecke, L. and Utrecht, C. (2020) Processing of the SARS-CoV pp1a/ab nsp7-10 region. *Biochem. J.* **477**, 1009–1019 <https://doi.org/10.1042/BCJ20200029>
- 7 Ziebuhr, J., Snijder, E.J. and Gorbalenya, A.E. (2000) Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* **81**, 853–879 <https://doi.org/10.1099/0022-1317-81-4-853>
- 8 Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z. et al. (2003) The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl Acad. Sci. U.S.A.* **100**, 13190–13195 <https://doi.org/10.1073/pnas.1835675100>
- 9 Zhu, L., George, S., Schmidt, M.F., Al-Gharabli, S.I., Rademann, J. and Hilgenfeld, R. (2011) Peptide aldehyde inhibitors challenge the substrate specificity of the SARS-coronavirus main protease. *Antiviral. Res.* **92**, 204–212 <https://doi.org/10.1016/j.antiviral.2011.08.001>
- 10 Xiao, Y., Ma, Q., Restle, T., Shang, W., Svergun, D.I., Ponnusamy, R. et al. (2012) Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1 heterotrimer that exhibits primer-independent RNA polymerase activity. *J. Virol.* **86**, 4444–4454 <https://doi.org/10.1128/JVI.06635-11>
- 11 Su, D., Lou, Z., Sun, F., Zhai, Y., Yang, H., Zhang, R. et al. (2006) Dodecamer structure of severe acute respiratory syndrome coronavirus nonstructural protein nsp10. *J. Virol.* **80**, 7902–7908 <https://doi.org/10.1128/JVI.00483-06>
- 12 Joseph, J.S., Saikatendu, K.S., Subramanian, V., Neuman, B.W., Brooun, A., Griffith, M. et al. (2006) Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs. *J. Virol.* **80**, 7894–7901 <https://doi.org/10.1128/JVI.00467-06>
- 13 Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J. et al. (2004) The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure* **12**, 341–353 <https://doi.org/10.1016/j.str.2004.01.016>
- 14 Egloff, M.P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H. et al. (2004) The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proc. Natl Acad. Sci. U.S.A.* **101**, 3792–3796 <https://doi.org/10.1073/pnas.0307877101>
- 15 Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M. et al. (2005) Insights into SARS-CoV transcription and replication from the structure of the nsp7–nsp8 hexadecamer. *Nat. Struct. Mol. Biol.* **12**, 980–986 <https://doi.org/10.1038/nsmb999>