Processing of the SARS-CoV pp1a/ab nsp7-10 region

Boris Krichel[1], Sven Falke[2], Rolf Hilgenfeld[3,4], Lars Redecke[3,5]; Charlotte Uetrecht[1,6*];

Affiliation: [1]Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Martinistraße 52, 20251 Hamburg, Germany; [2]University of Hamburg, Institut für Biochemie und Molekularbiologie, Martin-Luther-King-Platz 6, 20146 Hamburg, Germany; [3]University of Lübeck, Institute of Biochemistry, Center for Structural and Cell Biology in Medicine, Ratzeburger Allee 160, 23562 Lübeck, Germany; [4]German Center for Infection Research (DZIF), Hamburg-Lübeck-Borstel-Riems Site, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany; [5]Deutsches Elektronen Synchrotron (DESY), Notkestraße 85, 22607 Hamburg, Germany; [6]European XFEL GmbH, Holzkoppel 4, 22869 Schenefeld, Germany.

* correspondence to: charlotte.uetrecht@xfel.eu

1.1. Abstract

Severe acute respiratory syndrome coronavirus (SARS-CoV) is the causative agent of a respiratory disease with a high case fatality rate. During the formation of the coronaviral replication/transcription complex (RTC), essential steps include processing of the conserved polyprotein nsp7-10 region by the main protease M^{pro} and subsequent complex formation of the released nsp's. Here, we analyzed processing of the coronavirus nsp7-10 region using native mass spectrometry showing consumption of substrate, rise and fall of intermediate products and complexation. Importantly, there is a clear order of cleavage efficiencies, which is influenced by the polyprotein tertiary structure. Furthermore, the predominant product is an nsp7+8(2:2) hetero-tetramer with nsp8 scaffold. In conclusion, native MS, opposed to other methods, can expose the processing dynamics of viral polyproteins and the landscape of protein interactions in one set of experiments. Thereby, new insights into protein interactions, essential for generation of viral progeny, were provided, with relevance for development of antivirals.

1.2. Introduction

The human pathogenic potential of zoonotic infections by coronaviruses was exposed by the discovery of severe acute respiratory syndrome coronavirus (SARS-CoV) as causative agent of the SARS epidemic in 2003 [1, 2]. The replicase gene ORF 1ab occupies two thirds of the CoV single-strand (+)-sense RNA genome. Initially, ORF1ab is directly translated into either replicase polyprotein pp1a (nsp1-11) or pp1ab (nsp1-16), depending on a ribosomal (-1)-frameshift [3]. Subsequently, the polyproteins undergo proteolytic processing into 11 or 16 individual nsp's. Eventually, the nsp's take part in forming the RTC, a membrane-anchored, highly dynamic protein-RNA complex facilitating the replicative processes [4-7].

The main chymotrypsin-like protease (M^{pro}; nsp5) facilitates processing of the C-terminal part of the polyprotein [8]. Following maturation by auto-cleavage, two protomers of M^{pro} associate into one active dimeric unit that specifically targets the cleavage sites at the nsp inter-domain junctions of nsp5-11 or nsp5-16. These sites primarily contain LQ↓S or LQ↓A residues at the positions P2, P1 and P1′ , but strictly conserved is only Q at P1 [9]. Being a possible drug target, structure and function of this protease were extensively investigated [10, 11].

The timely coordinated processing of pp1a is crucial for replication [12]. This was specifically shown for the processing order of the nsp7-10 region [13]. In general, the order of processing is directly dependent on the substrate specificity of M^{pro}, which has been tested in peptide-based biochemical assays [14-16]. However, peptide assays interrogate merely the substrate specificity for the isolated amino-acid sequence but disregard the polyprotein's structural layout, i.e. overall conformation or accessibility.

Upon processing of the nsp7-10 region by M^{pro} , the four small proteins nsp7, nsp8, nsp9 and nsp10 are released [17]. They can form functional complexes with CoV core enzymes and thereby stimulate replication. Most importantly, a complex of nsp7+8 acts as processivity factor for nsp12, the RNA-

dependent RNA-polymerase (RdRp) [18]. Particularly, various oligomeric states of nsp7+8 have been reported [19-22].

Detailed studies of nsp processing and complex formation require a method that follows these dynamic processes at a molecular level. Native mass spectrometry (MS) probes and detects protein complexes in mixtures, while keeping their non-covalent interactions intact through nano-electrospray ionization (nanoESI) [23, 24]. Furthermore, collision-induced dissociation (CID) enables to decipher the stoichiometry and topology of natively sprayed protein complexes, by dissociating non-covalent interactions *in vacuo*. Thereby, it allows to reveal the landscape of protein species from a heterogeneous sample solution and to observe dynamic processes over time.

Herein, we analyzed the processing of SARS-CoV nsp7-10 and complex formation of its sequentially released nsp's. First, the substrate specificity of SARS-CoV M^{pro} for the cleavage sites within nsp7-10 was determined in an assay with FRET-labelled peptides. Then, we explored native MS to reveal time-resolved nsp release and complex formation upon processing of a full-length SARS-CoV nsp7-10 substrate. Specifically, the analysis via CID-MS gave insights into a preferred complex of SARS-CoV nsp7+8.

1.3. Results

1.3.1. Specific substrate efficiencies from FRET peptide assay

Initially, the substrate specificity of SARS-CoV M^{pro} for nsp7-10 inter-domain junction cleavage was studied by means of FRET peptide substrates (FPS's). The substrates FPS7-8, FPS8-9 and FPS9-10 represented the cleavage sites nsp7-8, nsp8-9 and nsp9-10, respectively. Additionally, FPS4-5 served as a positive control analogous to the highly efficient N-terminal auto-cleavage site nsp4-5 of SARS-CoV M^{pro} (nsp5). The FPS's encompassed 12 amino-acid residues spanning positions P6 to P6' around the cleavage site and were labelled with a fluorophore and a quencher at their N- and C-terminus, respectively. Substrate specificities were determined by monitoring the increasing fluorescence upon M^{pro} cleavage of the FPS (Figure 1 and Table S 4). As expected, the assay showed FPS4-5 to be most efficiently cleaved amongst the FPS's tested here, with an apparent efficiency (K_{cat}/K_M) of 0.105±0.005 $\mu M^{-1}min^{-1}$. Compared to FPS4-5 (100 %), substrates FPS7-8 (17.1 %) and FPS9-10 (51.8 %) were less efficiently cleaved while FPS8-9 (0.1 %) was almost non-cleavable. The absolute and relative substrate efficiencies determined here agree with results of similar experiments [14, 25]. Considering that short peptides represent merely the primary structure of the cleavage site, more native-like substrates were prepared and tested to verify these results obtained with the FPS's.

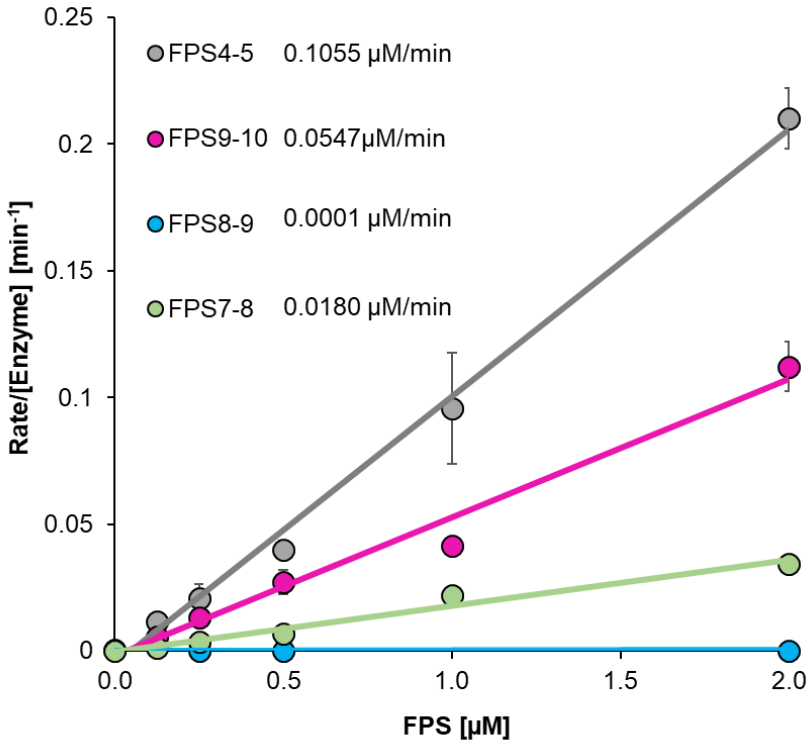

Figure 1: Process curves of M^{pro} (0.5 µM) with different concentrations of FRET peptide substrates (FPS) in 50 mM HEPES, 10 % Glycerol pH 7.5. The slopes correspond to the apparent k_{cat}/K_M as given in the legend.

Processing of nsp7-10

Next, we studied the substrate specificity of SARS-M^{pro} for folded proteins, which exhibit the structural environment of the cleavage sites beyond the amino-acid sequence. Therefore, full-length SARS-CoV His-nsp7-10 was recombinantly produced in *E.coli* and purified. This protein construct had a non-cleavable N-terminal His tag to reduce number of cleavage product species and thereby simplified interpretation of the mass spectra.

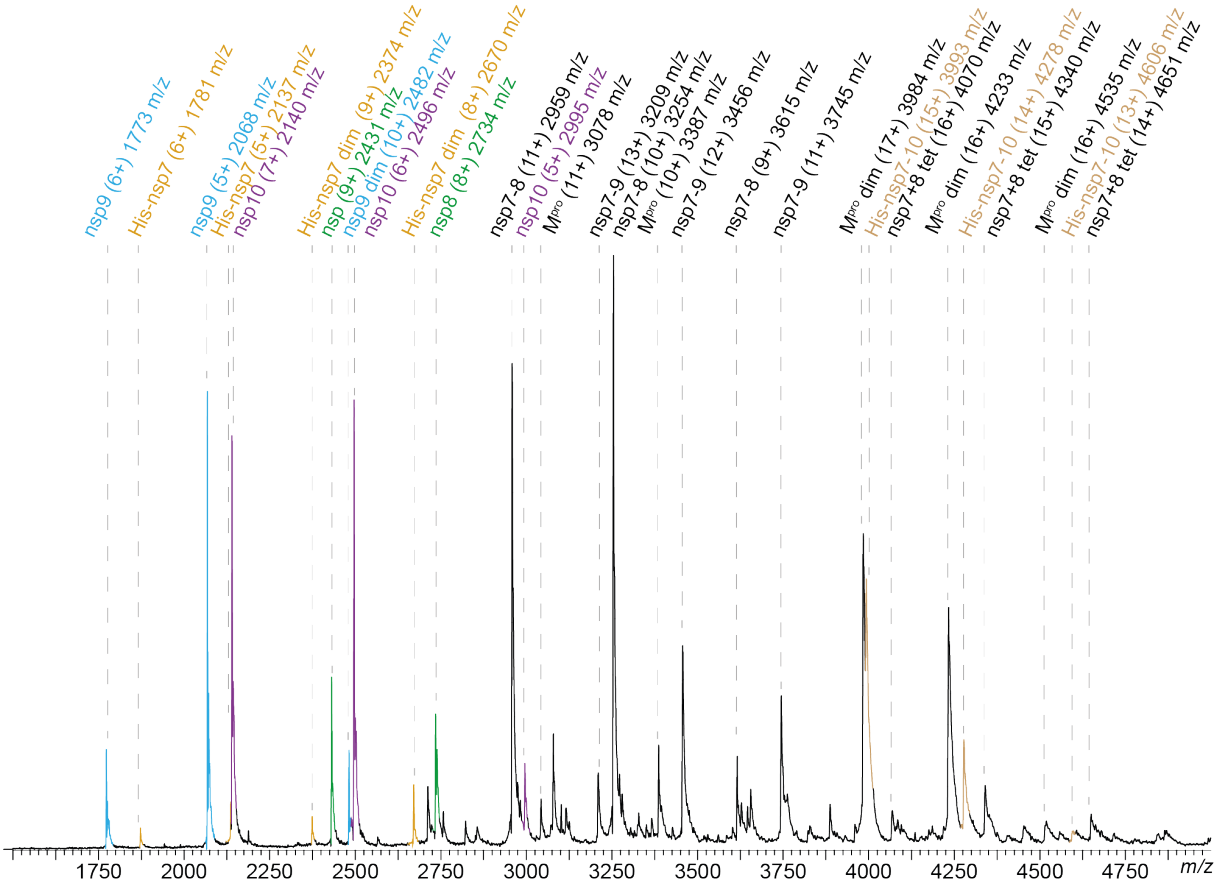

Figure 2: SARS nsp7-10 processing in native MS. Peak assignment in native MS after 6h. Mass species, charge states and their *m/z* are labelled. For processing, 1.25 µM SARS-CoV M^{pro} and 12.5 µM SARS-CoV His-nsp7-10 (~1:10 ratio) were incubated at 4°C in 250 mM NH_4OAc, 1 mM DTT at pH 8.0. The components were mixed and after a defined incubation time, samples were injected into an electrospray capillary and mass spectra were acquired.

To monitor the processing dynamics, SARS-CoV His-nsp7-10 was cleaved by SARS-CoV M^{pro} in a test tube and samples were subjected to native MS. Cleavage was performed in a nano ESI compatible ammonium acetate (NH_4OAc) solution at pH 8.0, which proved equally suitable to maintain protease activity compared to buffers with NaCl (Figure S 1). Processing products were observed as peaks in the mass spectra and assigned to protein mass species (Figure 2, Table S 3). When acquiring mass spectra at different times during processing, signals of three protein categories dominated successively, in the beginning full-length nsp7-10 substrate and M^{pro} (~1h); subsequently, intermediate products (~6 h), representing partially cleaved substrate; and in the end mature products encompassing nsp monomers and protein complexes (~20 h) (Figure S 6). Covalent and non-covalent products were distinguished via CID-MS/MS.

The order of reactions became evident from monitoring peak intensities over time. The relative signal intensity of M^{pro} remained stable while nsp7-10 decreased (Figure 3 A), indicating that M^{pro} depleted nsp7-10. At 20 h, no signal of nsp7-10 was detected anymore and therefore, cleavage was considered complete.

To obtain the sequence of nsp release, the ratio of the mature nsp monomers was followed over time. However, some non-covalent interactions of the nsp's introduced a bias to the relative peak intensities among monomers (Figure S 5). To correct for this bias, signal intensities were converted into mass fractions (MFs), which include the relative signal of each released subunit over time (Table S 2). Thereby, MFs demonstrated a processing order of the nsp9-10 site first, the nsp8-9 site second, and the nsp7-8 site third (Figure 3 B). MFs were similar at the end-point of processing (~20 h) showing that the individual nsp's were all cleaved off nsp7-10 (Figure S 5).

To determine the relative cleavage efficiencies in the full-length substrate, we used the relative MFs between nsp10, nsp9, nsp8 and nsp7 at 3.3 h (100%, 66%, 36 % and 36 %, respectively). These values suggest decreasing cleavage efficiency from the C- to the N-terminus of SARS nsp7-10. Consistently, the cleavage pattern of the intermediate products revealed that nsp7-10 is predominantly cleaved to nsp7-9 and only then to nsp7-8 (Figure 3 C). The time-course of processing of the SARS-CoV full-length substrate as determined here via native MS corresponded well with results from SDS-PAGE using the shorter SARS-CoV His-nsp7-9 substrate (Figure 3 C, Figure S 7). Additionally, shorter substrates were also analyzed in native MS, fully in line with the results obtained SARS-CoV nsp7-10 (Figure S 8).

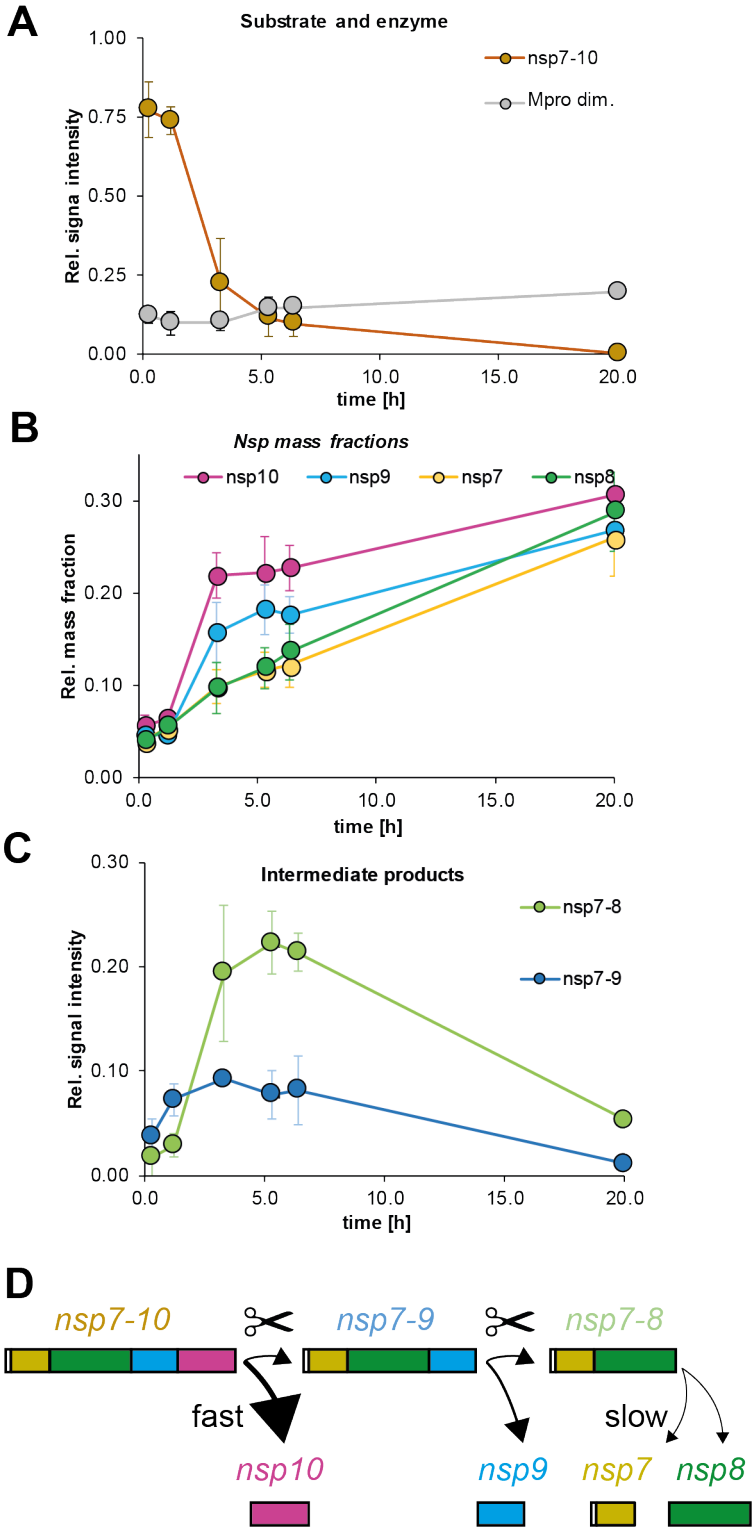

Figure 3: SARS-CoV nsp7-10 processing monitored by native MS: Signal over time of protease, substrate and products. **(A)** Substrate and enzyme. **(B)** Mass fractions over time show the order of nsp release. **(C)** Intermediate products. Error bars depict standard deviation (*N*=3). Time points (AVG±SD, *N*=3): 0.25±0.1 h, 1.2±0.1 h, 3.3±0.2 h, 5.3±0.3 h, 6.4±0.3 h±0.5 h. **(D)** Schematic illustration of cleavage order and efficiency.

The results from FRET peptide assay and full-length protein cleavage are in conflict regarding the suggested cleavage order. This is particularly interesting for nsp8-9, because of its *Coronaviridae*-wide conserved NNE motif at P1'-P3'. Previously, it was suggested that this motif could confer poor substrate properties to the nsp8-9 cleavage site, which would result in a considerable influence on processing order and RTC assembly [16]. In fact, two observations made by other researchers provide indication for a peptide-specific effect: Fan et al. analyzed the secondary structure of nsp peptides and observed that nsp8-9, comprising the NNE motif and closely resembling the FPS8-9 tested here, had a higher propensity for α-helices than other nsp peptides [26, 27]. Moreover, Chuck et al. reported that β-strands in substrates promote binding to M^{pro} [28]. The results shown in our work indicate that the structural layout of the cleavage site is different when comparing peptide to full-length protein, which plays a role in endowing the nsp8-9 cleavage site with decent cleavability. Since the full-length protein possesses the structural environment of the cleavage sites, we consider results with this substrate more relevant for studying the processing pattern of the CoV polyprotein.

1.3.3. Protein complex formation upon processing of nsp7-10

In parallel, we applied CID-MS/MS to identify non-covalent interactions. Thereby, covalent processing intermediates were distinguished from newly formed complexes, for instance nsp7-8 intermediate and nsp7+8(1:1) dimer overlap in mass spectra (Figure S 3). Most importantly, we identified an nsp7+8 (2:2) complex, undescribed before, and subject to detailed analysis later on. Here, this complex' intensity in native MS remained low, most likely due to interference by the non-cleavable His-tag N-terminal of nsp7. Furthermore, homo-dimers of nsp7 and nsp9 were found (Figure S 4). The MFs of nsp9 monomers and dimers (AVG±SD, N=3; 75±11% and 25±2%, respectively) were determined, based on the intensity ratio of the assigned peaks. Although, the X-ray structure of nsp9 was reported as a dimer, its low abundance found here is consistent with a low dimer affinity in solution [29]. So far, nsp7's ability of dimerization in absence of other proteins has already been noted, but no function has been suggested [21] and therefore, more detailed information remains elusive. Many interactions between components of nsp7-10 have been reported in the past, however, nsp

interactions other than those reported above were completely absent, despite the high sensitivity of native MS.

1.3.4. Complex formation of nsp7+8

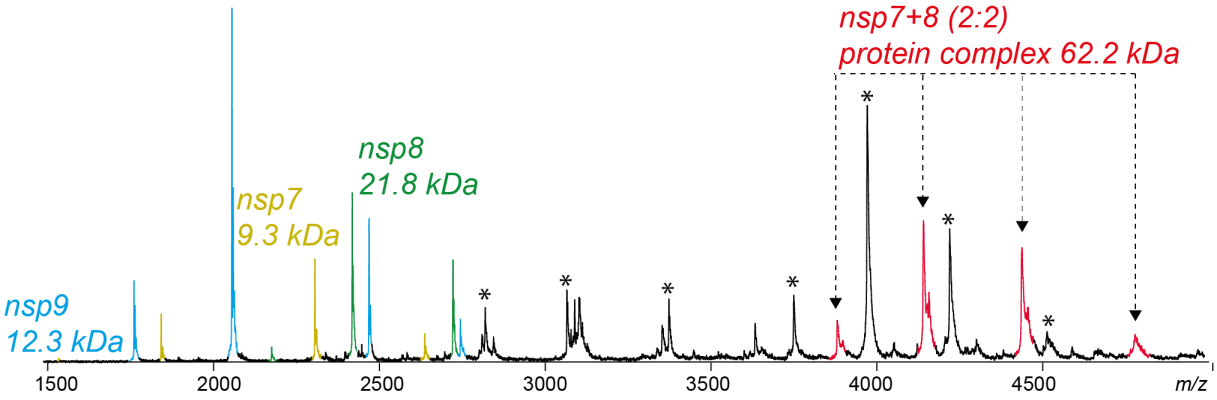

Figure 4: Upon cleavage of nsp7-9-His, a putative nsp7+8 (2:2) tetramer (red) emerged. Other assigned peaks correspond to M^{pro} (asterisk), nsp7 (yellow), nsp8 (green) and nsp9 (blue). For cleaving the precursor, 2 µM SARS-CoV M^{pro} was incubated with 14 µM SARS-CoV nsp7-9-His (ratio ~1:7) at 4°C in 300 mM NH_4OAc, 1 mM DTT at pH 8.0 overnight.

Finally, the complex of nsp7+8 was investigated in detail. In order to promote interactions of nsp's with authentic amino acid sequence, we engineered a SARS-CoV nsp7-9-His containing a cleavable His-tag at the C-terminus. Upon processing of this substrate by M^{pro}, we tested the landscape of mass species via native MS. Peak envelopes were detected for the monomeric products nsp7, nsp8 and nsp9 as well as for M^{pro} monomer and dimer. While nsp9 did not interact with any of the other subunits, even after cleavage of the tag, nsp7+8 formed a hetero-tetramer of nsp7+8(2:2) (62.2 kDa) (Figure 4). This stoichiometry was predicted based on the molecular weights of the subunits (2×9.3 kDa + 2×21.8 kDa = 62.2 kDa). However, for further evidence, molecular ions of three different charge states of the putative nsp7+8 hetero-tetramer (4449 *m/z*, +14; 4153 *m/z*, +15; 2893 *m/z*, +16) (Figure 5 A) were selected as precursor and subjected to CID-MS/MS. To successively induce subunit dissociation, we increased the collision voltage (CV) stepwise. Thereby, the overall dissociation pattern was exposed (Figure 5 A and Figure S 9), which confirmed the hetero-tetrameric (2:2) complex stoichiometry.

Two alternative dissociation pathways revealed the arrangement of subunits within the complex (Figure 5 B). The hetero-tetramer dissociated at lower activation energy into nsp7 and nsp7+8(1:2),

and alternatively, at elevated collisional energy into nsp8 and nsp7+8(2:1). Additionally, the product ions nsp7+8(1:1) and nsp8(2) provided evidence that their subunits interacted in the complex. The alternative dissociation routes may have resulted from analogous energy absorption paths, either due to similar binding contribution of nsp7 and nsp8 in the complex or due to different gas phase conformers.

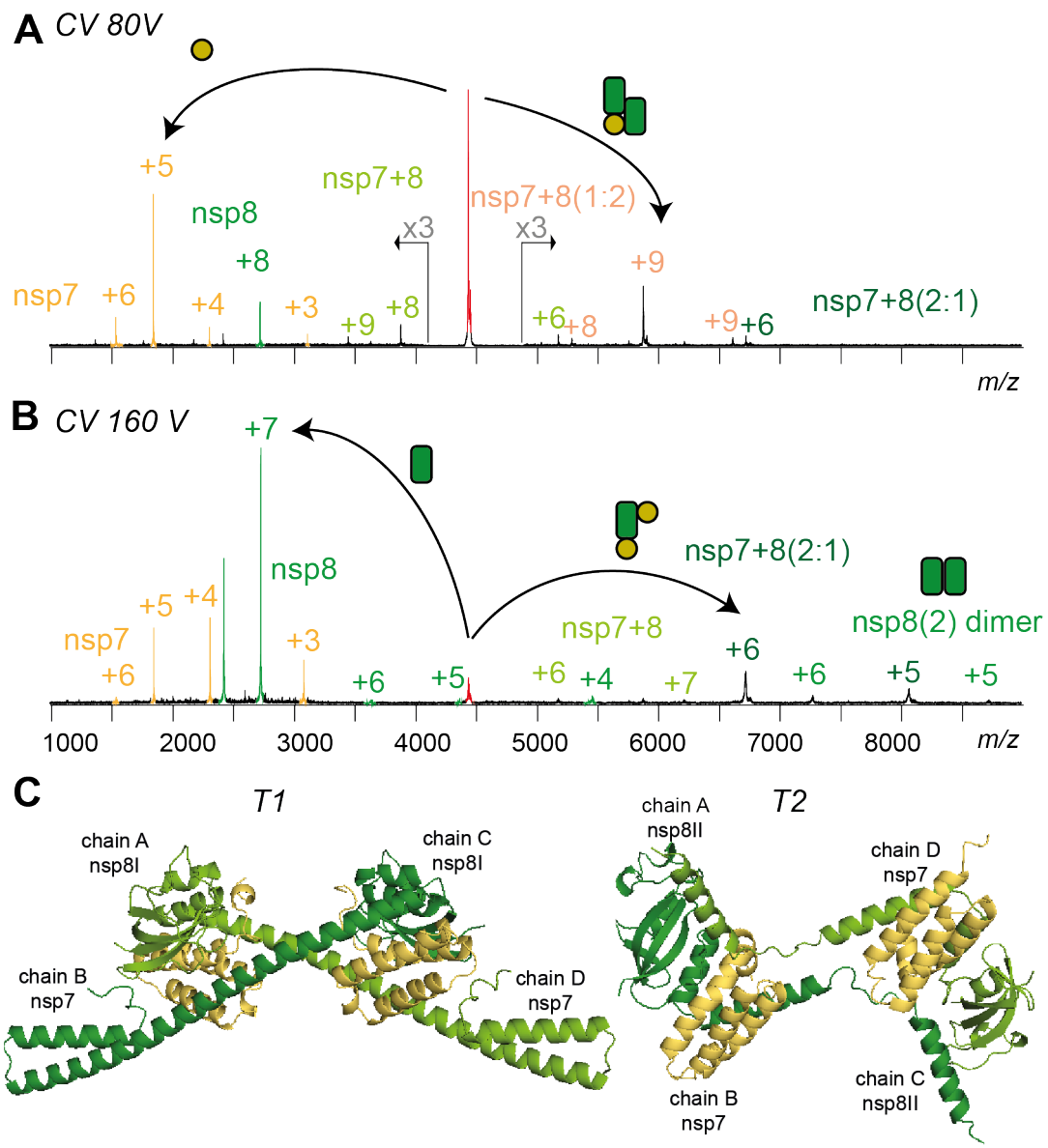

Figure 5: Product ion spectra showing two CID dissociation pathways of SARS-CoV nsp7+8 hetero-tetramer (+14, 4449 *m/z*) at (A) 50 V and (B) 80 V collisional energy. (C) Two conformers denoted as T1 and T2 extracted from a hexa-decameric SARS-CoV nsp7+8 (PDB 2G9T) [30] and both in line with the native MS results.

Recently, the structure of the SARS-CoV RdRp complex was reported as nsp(7+8)+8+12 ((1:1):1:1)[22]. Nevertheless, the quaternary structure in the absence of other proteins could be different and an important intermediate for either stability or subsequent interactions in solution. Although, molecular structures of isolated SARS-CoV nsp7+8 have been reported as dimer (1:1) containing N-terminally chopped nsp8 and as a hexa-decamer (8:8) [20, 30]. Up to date, a hetero-tetrameric interaction has not been shown.

In terms of stoichiometry and subunit connectivity, only nsp7+8(2:2) subcomplexes of the hexadecamer resemble our results [30]. Specifically, in two conformers, T1 and T2 (Figure 5 C), nsp8 builds a scaffold with head-to-tail interaction and nsp7 subunits sandwich the scaffold without self-interaction. All interfaces are mainly hydrophobic. However, three polar bonds within or in proximity of the nsp8 shaft domain facilitate binding to nsp7. Polar bonds more readily persist in the gas phase than hydrophobic interactions, which explains higher gas phase stability for nsp7+8(1:1) over nsp8(2). The total buried surface areas of T1 and T2 are virtually equal (Table S 5). Nevertheless, the nsp8(2) dimer scaffold in T2 is compact because a loop within the N-terminal α-helix allows bending and interacting with the other nsp8, whereas in T1 the α-helix extends to interact with another tetramer in the crystal lattice. Hence, T2 represents a more likely structure of nsp7+8(2:2) in solution.

In summary, we performed a comprehensive analysis of specific processing and complex formation of a folded viral polyprotein with native MS. While it remains a challenge to perform quantitative MS from heterogeneous protein mixtures, the usage of mass fractions disentangled the mass species and allowed following the processing dynamics. Intriguingly, our results suggest that the processing pattern of SARS-CoV nsp7-10 could have evolved to tailor a remaining nsp7-8 intermediate product. This intermediate product may have independent enzymatic function [21] but finally serves as an efficient precursor for nsp7+8(2:2) hetero-tetramer formation, as stimulant of the CoV RdRp [18]. Extending our approach to processing of different CoV species and longer polyproteins will increase understanding of the formation of the functional CoV-RTC.

1.4. Acknowledgement:

The Heinrich Pette Institute, Leibniz Institute for Experimental Virology is supported by the Free and Hanseatic City of Hamburg and the Federal Ministry of Health. C.U. and B.K. are supported by EU Horizon 2020 ERC StG-2017 759661. C.U. and B.K.. also received funding through Leibniz Association through SAW-2014-HPI-4 grant. SF and LR thank the German Federal Ministry for Education and Research (BMBF) for funding (grants 01KX0806, 01KX0807, and 05K18FLA). RH acknowledges funding from the SILVER project of the European Commission (contract HEALTH-F3-2010-260644).

1.5. Author contributions

Conceptualization and methodology B.K and C.U.; Cloning constructs S.F.; providing research materials R.H.; investigation B.K., discussion of results B.K., C.U,, S.F. and L.R.; formal analysis, visualization and original draft B.K and C.U; writing - review and editing B.K., C.U., S.F., L.R., and R.H..

1.6. Supplement

- Material and Methods
- Supplementary Figures and Tables
- Supplementary References

1.7. References:

1. Drosten, C., et al., *Identification of a novel coronavirus in patients with severe acute respiratory syndrome.* New. Engl. J. Med., 2003. **348**(20): p. 1967-1976.

2. Hilgenfeld, R. and M. Peiris, *From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses.* Antiviral research, 2013. **100**(1): p. 286-295.

3. Brierley, I., P. Digard, and S.C. Inglis, *Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot.* Cell, 1989. **57**(4): p. 537-47.

4. Marra, M.A., et al., *The Genome sequence of the SARS-associated coronavirus.* Science, 2003. **300**(5624): p. 1399-404.

5. Thiel, V., et al., *Mechanisms and enzymes involved in SARS coronavirus genome expression.* J. Gen. Virol., 2003. **84**(Pt 9): p. 2305-15.

6. Snijder, E.J., et al., *Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage.* J Mol Biol, 2003. **331**(5): p. 991-1004.

7. Gorbalenya, A.E., et al., *Nidovirales: evolving the largest RNA virus genome.* Virus Res., 2006. **117**(1): p. 17-37.

8. Hsu, M.F., et al., *Mechanism of the maturation process of SARS-CoV 3CL protease.* J Biol Chem, 2005. **280**(35): p. 31257-66.

9. Ziebuhr, J. and S.G. Siddell, *Processing of the human coronavirus 229E replicase polyproteins by the virus-encoded 3C-like proteinase: identification of proteolytic products and cleavage sites common to pp1a and pp1ab.* J Virol, 1999. **73**(1): p. 177-85.

10. Anand, K., et al., *Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs.* Science, 2003. **300**(5626): p. 1763-7.

11. Yang, H., et al., *The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor.* Proc. Natl. Acad. Sci. U.S.A., 2003. **100**(23): p. 13190-13195.

12. Sawicki, S.G., et al., *Functional and genetic analysis of coronavirus replicase-transcriptase proteins.* PLoS Pathog, 2005. **1**(4): p. e39.

13. Deming, D.J., et al., *Processing of open reading frame 1a replicase proteins nsp7 to nsp10 in murine hepatitis virus strain A59 replication.* J Virol, 2007. **81**(19): p. 10280-91.

14. Fan, K., et al., *Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase.* J Biol Chem, 2004. **279**(3): p. 1637-42.

15. Wu, A., et al., *Prediction and biochemical analysis of putative cleavage sites of the 3C-like protease of Middle East respiratory syndrome coronavirus.* Virus Res., 2015. **208**: p. 56-65.

16. Hegyi, A. and J. Ziebuhr, *Conservation of substrate specificities among coronavirus main proteases.* J. Gen. Virol., 2002. **83**(Pt 3): p. 595-9.

17. Falke, S., *(Doctoral Thesis) Coronaviral Polyprotein Nsp7-10 : Proteolytic Processing and Dynamic Interactions within the Transcriptase/Replicase Complex*. 2014, Staats- und Universitätsbibliothek Hamburg.

18. Subissi, L., et al., *One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities.* Proc. Natl. Acad. Sci. U.S.A., 2014. **111**(37): p. E3900-E3909.

19. te Velthuis, A.J., S.H. van den Worm, and E.J. Snijder, *The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension.* Nucleic Acids Res., 2012. **40**(4): p. 1737-47.

20. Li, S., et al., *New nsp8 isoform suggests mechanism for tuning viral RNA synthesis.* Protein Cell, 2010. **1**(2): p. 198-204.

21. Xiao, Y., et al., *Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1 heterotrimer that exhibits primer-independent RNA polymerase activity.* J Virol, 2012. **86**(8): p. 4444-54.

22. Kirchdoerfer, R.N. and A.B. Ward, *Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors.* Nat Commun, 2019. **10**(1): p. 2342.

23. Chandler, S.A. and J.L. Benesch, *Mass spectrometry beyond the native state.* Curr Opin Chem Biol, 2018. **42**: p. 130-137.

24. Dülfer, J., et al., *Structural mass spectrometry goes viral.* Adv. Virus Res., 2019. **105**: p. 189-238.

25. Grum-Tokars, V., et al., *Evaluating the 3C-like protease activity of SARS-Coronavirus: recommendations for standardized assays for drug discovery.* Virus Res., 2008. **133**(1): p. 63-73.

26. Huang, C., et al., *3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism.* Biochemistry, 2004. **43**(15): p. 4568-74.

27. Fan, K., et al., *The substrate specificity of SARS coronavirus 3C-like proteinase.* Biochem. Biophys. Res. Commun., 2005. **329**(3): p. 934-40.

28. Chuck, C.-P., et al., *Profiling of substrate specificity of SARS-CoV 3CLpro.* PloS one, 2010. **5**(10): p. e13197.

29. Ponnusamy, R., et al., *Variable oligomerization modes in coronavirus non-structural protein 9.* J Mol Biol, 2008. **383**(5): p. 1081-96.

30. Zhai, Y., et al., *Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer.* Nat. Struct. Mol. Biol., 2005. **12**(11): p. 980-6.