Biochemical Journal (2020) **477** 1009–1019
https://doi.org/10.1042/BCJ20200029

PORTLAND PRESS

Research Article

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

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

[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: Charlotte Uetrecht (charlotte.uetrecht@xfel.eu)

Severe acute respiratory syndrome coronavirus is the causative agent of a respiratory disease with a high case fatality rate. During the formation of the coronaviral replication/transcription complex, 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.

Introduction

The discovery of severe acute respiratory syndrome coronavirus (SARS-CoV) as the causative agent of the SARS epidemic in 2003 brought the world's attention to the human pathogenic potential of zoonotic infections by coronaviruses [1,2]. Two-thirds of their single-strand (+)-sense RNA genome exhibit the overlapping replicase genes ORF 1a and ORF 1ab, and the other one third encode for a set of subgenomic mRNAs which are required for accessory proteins as well as for the structural proteins. Initially, ORF 1a and ORF 1ab are directly translated into either replicase polyprotein pp1a (nsp1–11) or pp1ab (nsp1–16), respectively, 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 replication/transcription complex (RTC), a membrane-anchored, highly dynamic protein–RNA complex facilitating the replicative processes [4–7].

Two CoV proteases facilitate the processing of the polyprotein, the papain-like protease (PL^{pro}; nsp3) between nsp1–4 and main chymotrypsin-like protease (M^{pro}; $3CL^{pro}$, nsp5) between nsp4–11/16. Due to their essential roles, these proteases were extensively investigated regarding their interplay of structure and function, and their suitability as drug target [8–11].

The activity of M^{pro} is regulated over mechanisms comprising maturation from the polyprotein by autoprocessing of nsp4–5 and nsp5–6, as well as concentration and substrate induced self-assembly into an active dimeric unit [12–14]. After freeing itself, M^{pro} liberates nsp's 6–16 from the polyprotein by specifically targeting their nsp inter-domain junctions. These primarily contain LQ↓S or LQ↓A residues at the positions P2, P1 and P1′, but the strictly conserved requirement is only Q at P1 [15].

Received: 13 January 2020
Revised: 19 February 2020
Accepted: 21 February 2020

Accepted Manuscript online: 21 February 2020
Version of Record published: 13 March 2020

PORTLAND PRESS

Biochemical Journal (2020) 477 1009–1019
https://doi.org/10.1042/BCJ20200029

Temporal and spatial coordination of polyprotein processing is a way to regulate specific functions in the life-cycle of +ssRNA viruses [16,17]. The coordinated processing of pp1a is crucial for replication [18]. This was shown in detail for the processing order of the nsp7–10 region, where domain deletions or switching as well as cleavage site mutations were lethal to the virus replication [19]. 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 [20–22]. 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 [23]. They can form functional complexes with CoV core enzymes and thereby stimulate replication. Most importantly, a complex of nsp7+8 acts as a processivity factor for nsp12, the RNA-dependent RNA-polymerase (RdRp) [24]. Furthermore, point mutations in SARS-CoV nsp7 and nsp8 delayed virus growth or were identified as critical for virus replication [24]. Particularly, various oligomeric states of nsp7+8 have been reported for SARS-CoV and Feline infectious peritonitis virus (FIPV) [25–29]. Therefore, the interaction of nsp7 and nsp8 appears to be conserved across the coronavirus family making these proteins interesting drug targets, which highlights the need of understanding their structural flexibility.

Detailed studies of polyprotein 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 protein mixtures, while keeping their non-covalent interactions intact through nano-electrospray ionization (nanoESI) [30,31]. Furthermore, collision-induced dissociation (CID) enables the stoichiometry and topology of natively sprayed protein complexes to be deciphered, by dissociating non-covalent interactions *in vacuo*. Thereby, it allows the landscape of protein species from a heterogeneous sample solution to be revealed 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-labeled 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.

Materials and methods

Cloning and gene constructs

To generate inserts for expression plasmids for the CoV ORF1ab nsp7–10 region, DNA was amplified by PCR from commercially obtained cDNA (Eurofins scientific SE) and from DNA plasmids existing from earlier work [23,27]. To create ends for directed cloning, the PCR products were digested by Eco31I (BsaI) (Thermo Fisher Scientific) and ligated into IBA pASK33+ and pASK35+, encoding for C- and N-terminal His-tag, respectively. The expression plasmid for SARS-CoV M^{pro} with authentic ends was created as described by Xue et al. [32].

Expression and purification

Main protease SARS-CoV M^{pro} was produced with authentic ends [32]. Briefly, the protein constructs had a GST-tag connected via an auto-cleavage site at the N-terminus and a His-tag connected via a PreScission protease cleavage site at the C-terminus. The protease was purified by affinity chromatography using Ni^{2+}-NTA beads (Thermo Fisher Scientific) and dialyzed into storage buffer (50 mM HEPES, 10% glycerol pH 7.5). The His-tag was cleaved by dialysis into PreScission cleavage buffer (20 mM Tris–HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT at pH 7.0) while incubating with PreScission protease (protease to protein ratio 1 : 2000) (GE Healthcare) overnight. PreScission protease was removed by binding to GST-Sepharose (GE Healthcare) for 2 h. Finally, M^{pro} was buffer exchanged with centrifugal filter devices (10 000 MWCO, Amicon, Merck Millipore) into storage buffer. Purified M^{pro} was frozen in liquid nitrogen and stored at −80°C.

To produce proteins (Protein ID: P0C6X7) of different-length of the nsp7–10 region, BL21 Rosetta2 (Merck Millipore) were transformed, grown in culture flasks to 0.4–0.6 OD_{600} in 1 l, then induced with 50 μM anhydrotetracycline and continued to grow at 20°C overnight. To pellet, cultures were centrifuged (6000×***g*** for 20 min) and cells were frozen at −20°C. To separate soluble nsp's, pelleted cells were lysed in 1 : 5 (v/v) buffer B1 (40 mM phosphate buffer, 300 mM NaCl) with one freeze–thaw cycle, sonicated (micro tip, 70% power, 6 times on 10 s, off 60 s; Branson digital sonifier SFX 150) and then centrifuged (20 000×***g*** for 45 min). The proteins were isolated with Ni^{2+}-NTA beads (Thermo Fisher Scientific) in gravity flow columns (Bio-Rad). First, the beads were equilibrated with 20 column volumes (CV) B1 + 20 mM imidazole, then bound to nsp's

by incubation with crude extract for 60 min and finally to remove unspecifically bound proteins washed with 20 CV B1 + 20 mM imidazole followed by 10 CV of B1 + 50 mM imidazole. To elute the nsp's, eight fractions of 0.5 CV B1 + 300 mM imidazole were collected. Immediately after elution, the fractions were supplemented with 4 mM DTT. For quality analysis, SDS–PAGE was performed.

FRET peptide assays

For SARS-CoV M^{pro} activity assays, Förster resonance energy transfer (FRET) peptide substrates (FPS) were commercially purchased (Eurogentec), designed as SARS-CoV cleavage site analogs of twelve amino-acid residues (P6–P6′) with a FRET pair labeling, namely the fluorophore (F) HiLyte488 at the N-terminus and the Quencher (Q) QXL520 at the C-terminus (Supplementary Tables S1 and S2).

Initially, the freeze-dried FPS were dissolved in DMSO to 2 mM and stored at −20°C, protected from light. For sample preparation, the peptides were serially diluted in 50 mM HEPES, 10% (v/v) glycerol, 1 mM DTT, pH 7.5 and final concentrations were adjusted by measuring A_{280} from five independent droplets, using ε = 73 000 mol^{-1} cm^{-1}, the absorption coefficient of the HiLyte488 fluorophore as given by the manufacturer.

Assays were carried out using a 96-well plate reader (Infinite200, Tecan) with the following parameters: excitation wavelength 485 nm (bandwidth 9 nm), emission wavelength 535 nm (bandwidth 20 nm), gain 80, 10 flashes and 40 μs integration time. Furthermore, flat bottom black polystyrol 96-well plates (Greiner) were used for fluorescence applications.

Fluorescence coefficients k_{sx} and inner filter effect of the peptides were determined for each substrate according to previous protocols [33] (Supplementary Figure S1). The kinetic parameters were determined as suggested by Grum-Tokars et al. [34]. Briefly, 20 μl SARS-CoV M^{pro} (final conc. c_{Mpro} of 0.5 μM) was mixed with 80 μl FPS of six different concentrations (final conc. $c_{X,S}$ of 2 μM, 1 μM, 0.5 μM, 0.25 μM, 0.125 μM and 0 μM) and fluorescence (AFU) was monitored every 30 s for 5 min. To calculate kinetic parameters, triplicate measurements were performed.

The fluorescence coefficients k_{sx} were used to convert the initial increase in AFU into the initial reaction rate, which was then normalized by c_{Mpro} to rate/[Enzyme]. For comparing substrate specificities, the rates/[Enzyme] were plotted against the substrate concentrations $c_{X,S}$ resulting in process curves. The slopes of the process curves correspond to the apparent k_{cat}/K_M.

The influence of salt conditions on enzyme activity was determined by incubating 20 μl M^{pro} (final concentration 0.1 μM) with 80 μM FRET peptide substrate FPS4–5 (final concentration 1 μM) in varying NH_4OAc (10% Glycerol, pH 7.5 and 50 mM, 165 mM, 280 mM, 395 mM, 510 mM, 625 mM, 740 mM, 855 mM, 970 mM, 1085 mM and 1200 mM) and NaCl (50 mM HEPES, 10% Glycerol, pH 7.5 and 0 mM, 120 mM, 240 mM, 360 mM, 480 mM, 600 mM, 720 mM, 840 mM, 960 mM, 1080 mM and 1200 mM). The fluorescence was followed every 60 s by monitoring AFU and initial slopes (AFU/s) were converted by the fluorescence coefficient k_{sx} into the initial rate (Supplementary Figure S2).

SDS–PAGE

SDS–PAGE was performed with a 4%–12% gradient acrylamide Bis-tris gel with XT MES running buffer. For *in vitro* processing, 3.2 μM SARS-CoV M^{pro} was incubated with 13 μM SARS-CoV nsp7–9-His (ratio ∼1 : 4) at 4°C in 20 mM phosphate buffer, 150 mM NaCl, 1 mM DTT at pH 8.0.

Native MS

To prepare samples for native MS measurements, they were buffer-exchanged into nanoESI-compatible solution. Protease M^{pro} was buffer-exchanged into 250–500 mM NH_4OAc, 1 mM pH 8.0 by two cycles of centrifugal gel filtration (Biospin mini columns, 6000 MWCO, Bio-Rad). The nsp's were buffer-exchanged into 250–500 mM NH_4OAc, 1 mM DTT, pH 8.0 by five rounds of dilution and concentration in centrifugal filter units (Amicon, 10 000 MWCO, Merck Millipore).

NanoESI capillaries were pulled in-house from borosilicate capillaries (1.2 mm outer diameter, 0.68 mm inner diameter, with filament, World Precision Instruments) with a micropipette puller (P-1000, Sutter Instruments) using a squared box filament (2.5 × 2.5 mm, Sutter Instruments) in a two-step program. Subsequently, capillaries were gold-coated using a sputter coater (Q150R, Quorum Technologies) with 40 mA, 200 s, tooling factor 2.3 and end bleed vacuum of 8×10^{-2} mbar argon.

Native MS was performed at a nanoESI quadrupole time-of-flight (Q-TOF) instrument (Q-TOF2, Micromass/Waters, MS Vision) modified for higher masses [35]. Samples were ionized in positive ion mode

Biochemical Journal (2020) 477 1009–1019
https://doi.org/10.1042/BCJ20200029

with voltages applied at the capillary of 1300–1500 V and at the cone of 130–135 V. The pressure in the source region was kept at 10 mbar throughout all native MS experiments. For the purpose of desolvation and dissociation, the pressure in the collision cell was adjusted to $1.3–1.5\times10^{-2}$ mbar argon. Native mass spectra were obtained at an accelerating voltage of 10–30 V while for CID-MS/MS these voltages were increased to 30–200 V. In ESI-MS overview spectra for nsp's, the quadrupole profile was 1–10 000 *m/z*. In tandem MS, for precursor selection, LMres and HMres were adjusted at 10–30 V collisional voltage until a single peak was recorded, and then dissociation was induced.

To calibrate raw data, CsI (25 mg/ml) spectra were acquired and calibration was carried out with MassLynx (Waters) software. Data were analyzed using MassLynx (Waters) and Massign (by Morgner [36]). All determined masses are provided (Supplementary Table S3).

To start the processing reactions, nsp's and M^{pro} were mixed. Three independent reactions were started in parallel and incubated at 4°C. To acquire mass spectra at defined time-points, sample aliquots of 1–3 μl were taken by means of a microliter syringe (5 μl, Hamilton) with flexible fused silica tubing (Optronics) and loaded into in-house fabricated nanoESI capillaries, which were mounted on the nanoESI source, all within 2 min. Then, raw spectra were acquired in the first 300 scans (1 scan/s).

To analyze the data, the raw spectra were smoothed (2 × 5) in Masslynx 4.1 (Waters) and then nsp's were assigned to peak series. For each assigned mass species, signal responses (SRs), using the combined relative intensities of peaks, were summarized and normalized. This was done independently for each spectrum. Finally, the average and standard deviation for SR and time-points were calculated from three independent spectra. Poorly resolved spectra were not included in the data analysis.

To determine mass fractions (MFs), the SRs of assigned peaks for monomers and monomers in complexes were combined (Supplementary Table S4).

Results

Specific substrate efficiencies from FRET peptide assay

Initially, the substrate specificity of SARS-CoV M^{pro} for cleavage of the nsp7–10 inter-domain junctions was studied by means of 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 labeled with a fluorophore and a quencher at their N- and C-terminus, respectively (Supplementary Table S1). Substrate specificities were determined by monitoring the increasing fluorescence upon M^{pro} cleavage of the FPS's (Figure 1, Supplementary Figure S1 and Table S2). 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 μM^{-1} min^{-1}. Compared with 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 the results of similar experiments [20,34]. Considering that short peptides represent merely the primary structure of the cleavage site, more native-like substrates were prepared and tested to verify results obtained with these FPS's.

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 *Escherichia coli* and purified. This protein construct had a non-cleavable N-terminal His-tag to reduce the number of cleavage product species and thereby simplified interpretation of the mass spectra.

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 nanoESI compatible ammonium acetate (NH_4OAc) solution at pH 8.0, which proved equally suitable to maintain protease activity compared with buffers with NaCl (Supplementary Figure S2). Processing products were observed as peaks in the mass spectra and assigned to protein mass species (Figure 2, Supplementary Table S3). 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} (~1 h); subsequently, intermediate products (~6 h), representing partially

Biochemical Journal (2020) **477** 1009–1019
https://doi.org/10.1042/BCJ20200029

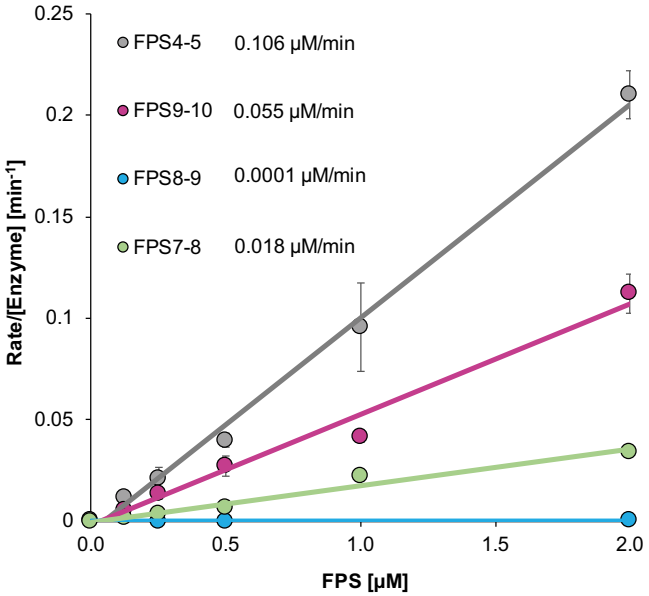

Figure 1. M^{pro} assay with FRET peptide substrates.

Process curves of M^{pro} (0.5 μM) with different concentrations of FRET peptide substrates (FPS) in 50 mM HEPES, 10% Glycerol at pH 7.5. The slopes correspond to the apparent k_{cat}/K_M as given in the legend. Standard deviations given in Supplementary Table S2.

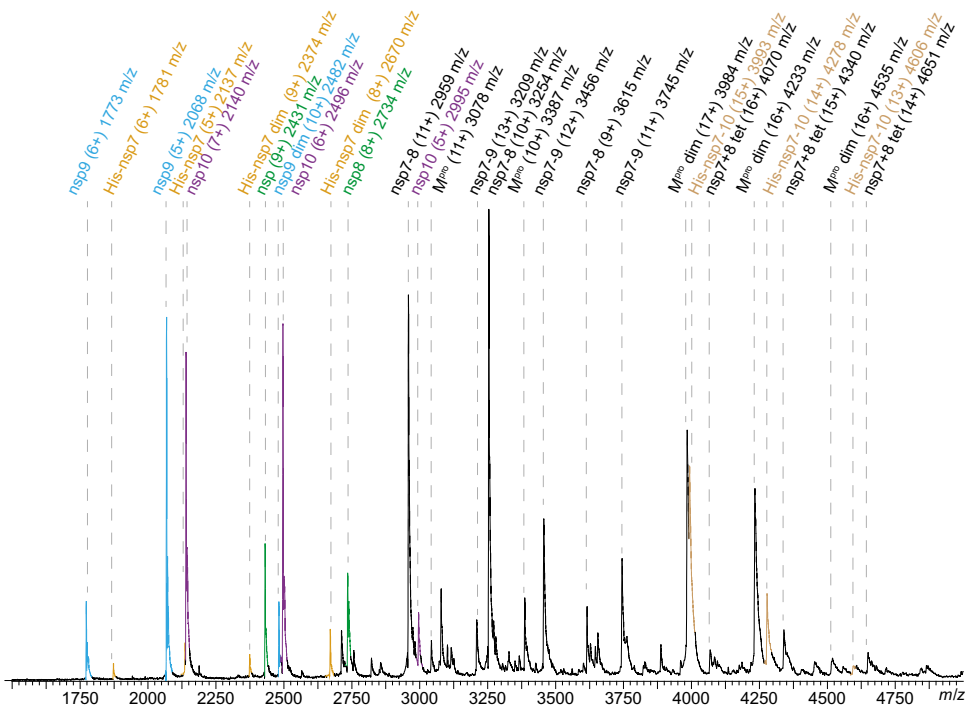

Figure 2. SARS-CoV nsp7–10 processing in native MS.

Peak assignment in native MS after 6 h. Mass species, charge states and their *m/z* are labeled. 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.

Biochemical Journal (2020) 477 1009–1019
https://doi.org/10.1042/BCJ20200029

cleaved substrate; and in the end mature products encompassing nsp monomers and protein complexes (~20 h) (Supplementary Figure S3). 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 3A), indicating that M^{pro} depleted nsp7–10. After 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 (Supplementary Figure S4). To correct for this bias, signal intensities were converted into MFs, which include the relative signal of each released subunit over time (Supplementary Table S4). Thereby, MFs demonstrated a processing order of the nsp9–10 site first, the nsp8–9 site second, and the nsp7–8 site third (Figure 3B). MFs were similar at the end-point of processing (~20 h) showing that the individual nsp's were all cleaved off nsp7–10 (Supplementary Figure S4).

To determine the relative cleavage efficiencies in the full-length substrate, we used the relative MFs between nsp10, nsp9, nsp8 and nsp7 after 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 3C). 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 3C, Supplementary Figure S5). Additionally, shorter substrates were also analyzed in native MS, fully in line with the results obtained for SARS-CoV nsp7–10 (Supplementary Figure S6).

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 [22]. In fact, two observations made by other researchers provide an 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 [37,38]. Moreover, Chuck et al. [39] reported that β-strands in substrates promote binding to M^{pro}. The results shown in our work indicate that the structural layout of the cleavage site is different when comparing peptide to the 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.

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 (Supplementary Figure S7). Most importantly, we identified an nsp7+8 (2 : 2) complex, undescribed before, which is 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 (Supplementary Figure S8). 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 [40]. So far, nsp7's ability of dimerization in the absence of other proteins has already been noted, but no function has been suggested [27] 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.

Complex formation of nsp7+8

Finally, the complex of nsp7+8 was investigated in detail. 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

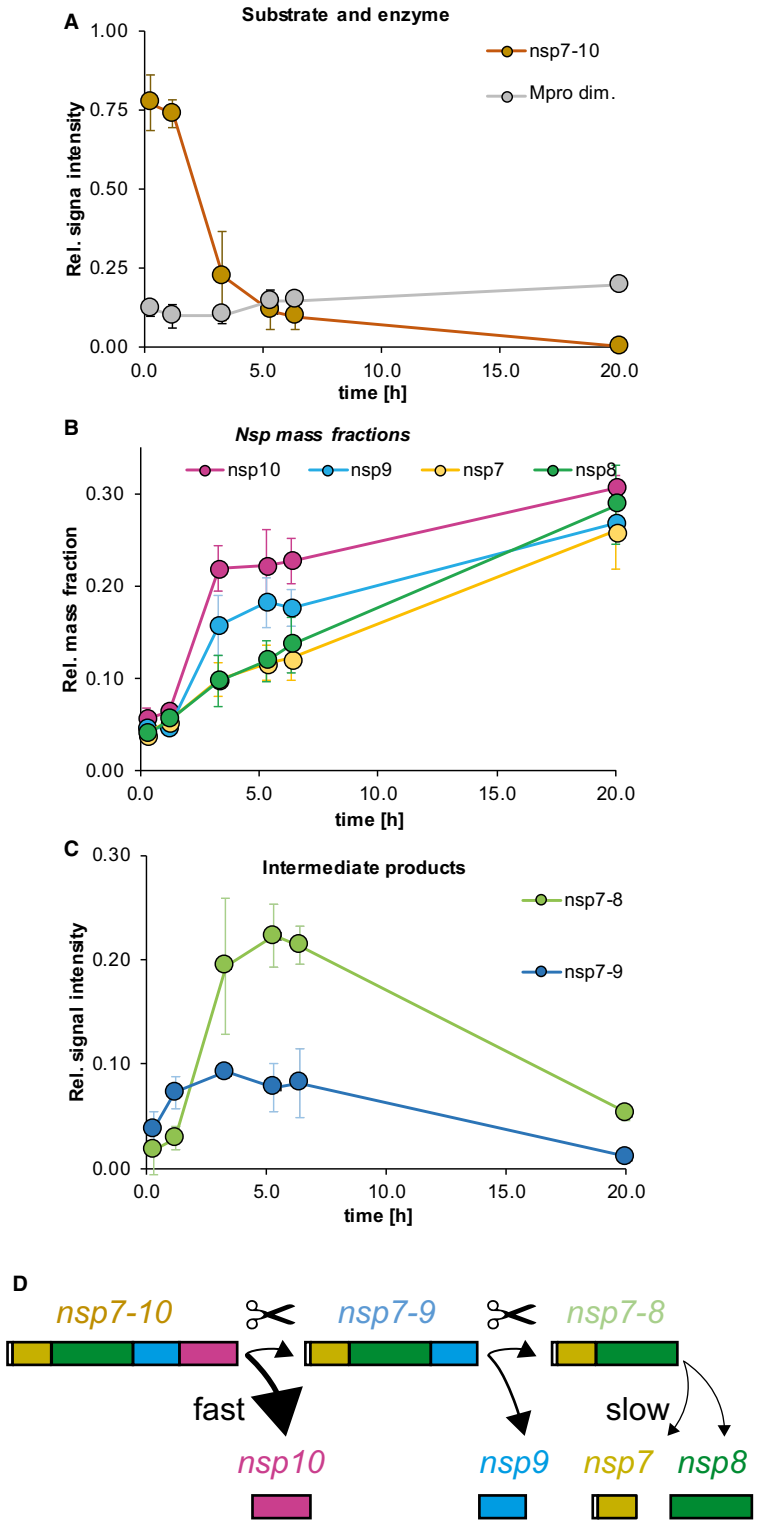

Figure 3. SARS-CoV nsp7–10 processing monitored by native MS: signal over time of protease, substrate and products.

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 and 20 ± 0.5 h. (**D**) Schematic illustration of cleavage order and efficiency.

Biochemical Journal (2020) **477** 1009–1019
https://doi.org/10.1042/BCJ20200029

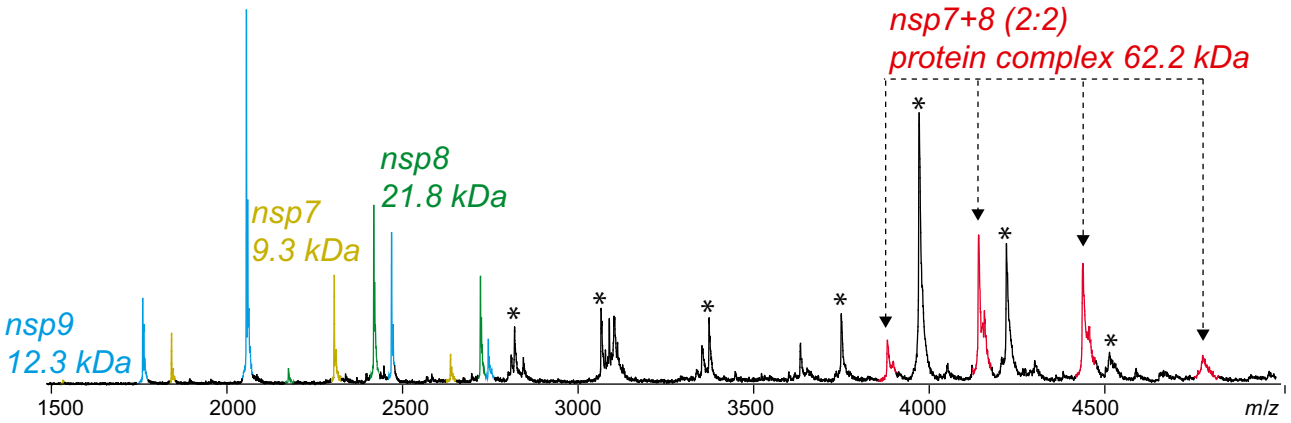

Figure 4. SARS-CoV nsp7+8 forms a hetero-tetramer.

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.

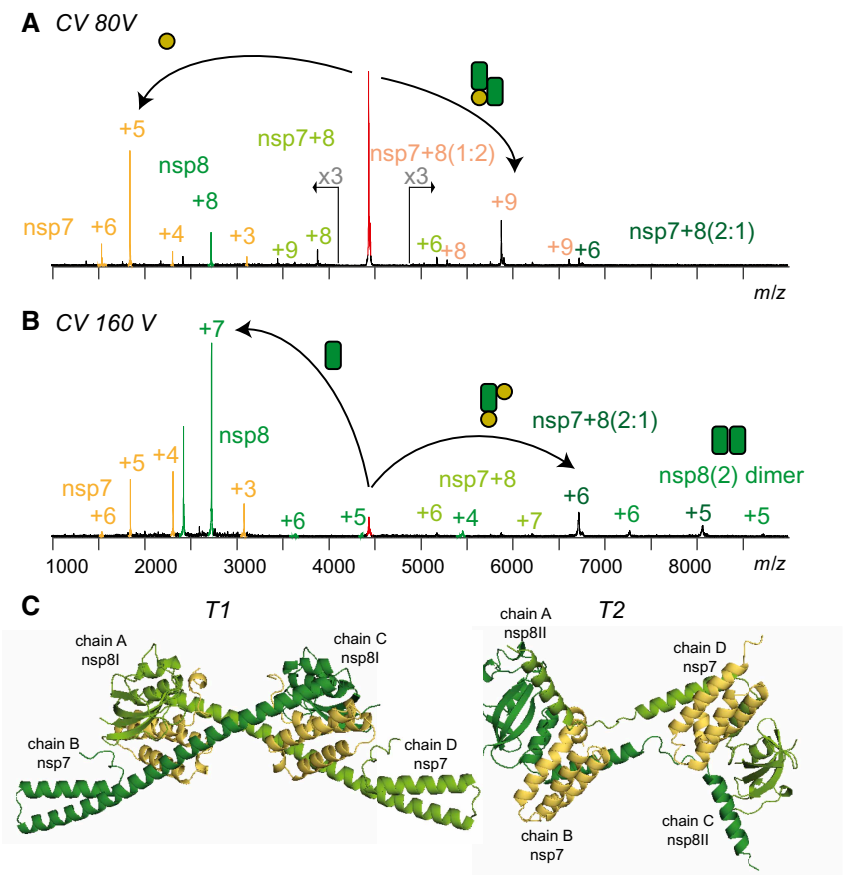

Figure 5. CID-MS/MS reveals the architecture of the nsp7+8 complex.

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 2AHM) [29] and both in line with the native MS results.

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 5A) 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 5A and Supplementary Figure S9), which confirmed the hetero-tetrameric (2 : 2) complex stoichiometry.

Two alternative dissociation pathways revealed the arrangement of subunits within the complex (Figure 5B). 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.

Discussion

Here, we analyzed the processing of the SARS-CoV polyprotein region nsp7–10 *in vitro*. The mechanistic system of processing resembles a limited proteolysis of a folded protein, and we used native MS to monitor the covalent and non-covalent products. Thereby, we gained insights into the coordinated release of nsp monomers, intermediate products and the formation of complexes. A relatively long-lived nsp7–8 product and a short-lived nsp8–9 intermediate product were observed, which were not expected based on protease activities measured on isolated peptide substrates. Therefore, we conclude that the processing of nsp7–10 is not only coordinated by the amino-acid sequence itself but also the structural environment of the nsp inter-domain junctions. This, in turn, may lead to different processing intermediates than expected based on results from protease assays at isolated peptides. To collect evidence of such intermediates *in vivo* appears challenging by more traditional techniques. To reveal their relevance in the viral life-cycle, novel MS techniques may be advantageous, e.g. single-cell proteomics and *in vivo* cross-linking MS [41,42].

The major identified protein complex was a hetero-tetramer of nsp7+8 (2 : 2). Although, molecular structures of SARS-CoV nsp7+8 have been reported as dimer (1 : 1) containing N-terminally chopped nsp8 and as a hexa-decamer (8 : 8) [26,29] 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 hexa-decamer resemble our results [29]. Specifically, in two conformers, T1 and T2 (Figure 5C), 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 the 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 (Supplementary Table S5). 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.

Finding a link between the plasticity of nsp7+8 and its defined functions, particularly in the context of the CoV-RTC, remains to be addressed. This understanding was furthered by the identification of the SARS-CoV RdRp complex as nsp(7+8)+8+12 ((1: 1) : 1 : 1) [28]. However, in this complex, the stoichiometry of nsp7+8 is not consistent with the quaternary structures that have been previously reported or ours presented here [27,29]. Nevertheless, the quaternary structure of the nsp7+8 complex in the absence of other RTC subunits could be different and may be an important intermediate for either stability, functionality or subsequent interactions 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 MFs 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 an independent enzymatic function [27] but finally serves as an efficient precursor for nsp7+8(2 : 2) hetero-tetramer formation, as stimulant of the CoV RdRp [24]. Extending our approach to processing of different CoV species and longer polyproteins will increase understanding of the formation of the functional CoV-RTC.

Biochemical Journal (2020) **477** 1009–1019
https://doi.org/10.1042/BCJ20200029

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

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 the Leibniz Association SAW-2014-HPI-4 grant. S.F. and L.R. thank the German Federal Ministry for Education and Research (BMBF) for funding [grant numbers 01KX0806, 01KX0807 and 05K18FLA]. R.H. acknowledges funding from the SILVER project of the European Commission (contract HEALTH-F3-2010-260644).

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.

Abbreviations

CID, collision-induced dissociation; CV, collision voltage; FPS, FRET peptide substrates; FRET, Förster resonance energy transfer; MFs, mass fractions; MS, mass spectrometry; nanoESI, nano-electrospray ionization; RdRp, RNA-dependent RNA-polymerase; RTC, replication/transcription complex; SARS-CoV, severe acute respiratory syndrome coronavirus; SRs, signal responses.

References

1 Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.R., Becker, S. et al. (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967–1976 https://doi.org/10.1056/NEJMoa030747
2 Hilgenfeld, R. and Peiris, M. (2013) From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. *Antiviral Res.* **100**, 286–295 https://doi.org/10.1016/j.antiviral.2013.08.015
3 Brierley, I., Digard, P. and Inglis, S.C. (1989) Characterization 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
4 Marra, M.A., Jones, S.J., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S. et al. (2003) The genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399–1404 https://doi.org/10.1126/science.1085953
5 Thiel, V., Ivanov, K.A., Putics, A., Hertzig, T., Schelle, B., Bayer, S. et al. (2003) Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84**, 2305–2315 https://doi.org/10.1099/vir.0.19424-0
6 Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L. et al. (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* **331**, 991–1004 https://doi.org/10.1016/S0022-2836(03)00865-9
7 Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J. and Snijder, E.J. (2006) Nidovirales: evolving the largest RNA virus genome. *Virus Res.* **117**, 17–37 https://doi.org/10.1016/j.virusres.2006.01.017
8 Ratia, K., Pegan, S., Takayama, J., Sleeman, K., Coughlin, M., Baliji, S. et al. (2008) A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16119–16124 https://doi.org/10.1073/pnas.0805240105
9 Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R. and Hilgenfeld, R. (2003) Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **300**, 1763–1767 https://doi.org/10.1126/science.1085658
10 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
11 Hilgenfeld, R. (2014) From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS J.* **281**, 4085–4096 https://doi.org/10.1111/febs.12936
12 Hsu, M.-F., Kuo, C.-J., Chang, K.-T., Chang, H.-C., Chou, C.-C., Ko, T.-P. et al. (2005) Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.* **280**, 31257–31266 https://doi.org/10.1074/jbc.M502577200
13 Cheng, S.-C., Chang, G.-G. and Chou, C.-Y. (2010) Mutation of Glu-166 blocks the substrate-induced dimerization of SARS coronavirus main protease. *Biophys. J.* **98**, 1327–1336 https://doi.org/10.1016/j.bpj.2009.12.4272
14 Gonzalez, L.S., Anson, B. and Mesecar, A. (2017) Effect of allosteric changes in MERS 3CL protease enzymatic activity and dimerization. *FASEB J.* **31**, 601.7–601.7 https://doi.org/10.1096/fj.201600713R
15 Ziebuhr, J. and Siddell, S.G. (1999) 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.* **73**, 177–185 https://doi.org/10.1128/JVI.73.1.177-185.1999
16 Emmott, E., de Rougemont, A., Hosmillo, M., Lu, J., Fitzmaurice, T., Haas, J. et al. (2019) Polyprotein processing and intermolecular interactions within the viral replication complex spatially and temporally control norovirus protease activity. *J. Biol. Chem.* **294**, 4259–4271 https://doi.org/10.1074/jbc.RA118.006780
17 Kräusslich, H.-G., Nicklin, M.J., Lee, C.K. and Wimmer, E. (1988) Polyprotein processing in picornavirus replication. *Biochimie* **70**, 119–130 https://doi.org/10.1016/0300-9084(88)90166-6

18 Sawicki, S.G., Sawicki, D.L., Younker, D., Meyer, Y., Thiel, V., Stokes, H. et al. (2005) Functional and genetic analysis of coronavirus replicase-transcriptase proteins. *PLoS Pathog.* **1**, e39 https://doi.org/10.1371/journal.ppat.0010039
19 Deming, D.J., Graham, R.L., Denison, M.R. and Baric, R.S. (2007) Processing of open reading frame 1a replicase proteins nsp7 to nsp10 in murine hepatitis virus strain A59 replication. *J. Virol.* **81**, 10280–10291 https://doi.org/10.1128/JVI.00017-07
20 Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L. et al. (2004) Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J. Biol. Chem.* **279**, 1637–1642 https://doi.org/10.1074/jbc.M310875200
21 Wu, A., Wang, Y., Zeng, C., Huang, X., Xu, S., Su, C. et al. (2015) Prediction and biochemical analysis of putative cleavage sites of the 3C-like protease of Middle East respiratory syndrome coronavirus. *Virus Res.* **208**, 56–65 https://doi.org/10.1016/j.virusres.2015.05.018
22 Hegyi, A. and Ziebuhr, J. (2002) Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.* **83**(Pt, 595–599 https://doi.org/10.1099/0022-1317-83-3-595
23 Falke, S. 2014. (Doctoral Thesis) Coronaviral Polyprotein Nsp7-10: Proteolytic Processing and Dynamic Interactions within the Transcriptase/Replicase Complex, Staats- und Universitätsbibliothek Hamburg
24 Subissi, L., Posthuma, C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E. et al. (2014) One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E3900–E3909 https://doi.org/10.1073/pnas.1323705111
25 te Velthuis, A.J., van den Worm, S.H. and Snijder, E.J. (2012) The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic Acids Res.* **40**, 1737–1747 https://doi.org/10.1093/nar/gkr893
26 Li, S., Zhao, Q., Zhang, Y., Zhang, Y., Bartlam, M., Li, X. et al. (2010) New nsp8 isoform suggests mechanism for tuning viral RNA synthesis. *Protein Cell* **1**, 198–204 https://doi.org/10.1007/s13238-010-0028-8
27 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
28 Kirchdoerfer, R.N. and Ward, A.B. (2019) Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* **10**, 2342 https://doi.org/10.1038/s41467-019-10280-3
29 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
30 Chandler, S.A. and Benesch, J.L. (2018) Mass spectrometry beyond the native state. *Curr. Opin Chem. Biol.* **42**, 130–137 https://doi.org/10.1016/j.cbpa.2017.11.019
31 Dülfer, J., Kadek, A., Kopicki, J.-D., Krichel, B. and Uetrecht, C. (2019) Structural mass spectrometry goes viral. *Adv. Virus Res.* **105**, 189–238 https://doi.org/10.1016/bs.aivir.2019.07.003
32 Xue, X., Yang, H., Shen, W., Zhao, Q., Li, J., Yang, K. et al. (2007) Production of authentic SARS-CoV M pro with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J. Mol. Biol.* **366**, 965–975 https://doi.org/10.1016/j.jmb.2006.11.073
33 Liu, Y., Kati, W., Chen, C.-M., Tripathi, R., Molla, A. and Kohlbrenner, W. (1999) Use of a fluorescence plate reader for measuring kinetic parameters with inner filter effect correction. *Anal. Biochem.* **267**, 331–335 https://doi.org/10.1006/abio.1998.3014
34 Grum-Tokars, V., Ratia, K., Begaye, A., Baker, S.C. and Mesecar, A.D. (2008) Evaluating the 3C-like protease activity of SARS-coronavirus: recommendations for standardized assays for drug discovery. *Virus Res.* **133**, 63–73 https://doi.org/10.1016/j.virusres.2007.02.015
35 van den Heuvel, R.H., van Duijn, E., Mazon, H., Synowsky, S.A., Lorenzen, K., Versluis, C. et al. (2006) Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* **78**, 7473–7483 https://doi.org/10.1021/ac061039a
36 Morgner, N. and Robinson, C.V. (2012) Massign: an assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies. *Anal. Chem.* **84**, 2939–2948 https://doi.org/10.1021/ac300056a
37 Huang, C., Wei, P., Fan, K., Liu, Y. and Lai, L. (2004) 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry* **43**, 4568–4574 https://doi.org/10.1021/bi036022q
38 Fan, K., Ma, L., Han, X., Liang, H., Wei, P., Liu, Y. et al. (2005) The substrate specificity of SARS coronavirus 3C-like proteinase. *Biochem. Biophys. Res. Commun.* **329**, 934–940 https://doi.org/10.1016/j.bbrc.2005.02.061
39 Chuck, C.-P., Chong, L.-T., Chen, C., Chow, H.-F., Wan, D.C.-C. and Wong, K.-B. (2010) Profiling of substrate specificity of SARS-CoV 3CLpro. *PLoS One* **5**, e13197 https://doi.org/10.1371/journal.pone.0013197
40 Ponnusamy, R., Moll, R., Weimar, T., Mesters, J.R. and Hilgenfeld, R. (2008) Variable oligomerization modes in coronavirus non-structural protein 9. *J. Mol. Biol.* **383**, 1081–1096 https://doi.org/10.1016/j.jmb.2008.07.071
41 Liu, F., Rijkers, D.T.S., Post, H. and Heck, A.J.R. (2015) Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **12**, 1179–1184 https://doi.org/10.1038/nmeth.3603
42 Specht, H., Emmott, E., Petelski, A., Huffman, R.G., Perlman, D.H., Serra, M. et al. (2019) Single-cell mass-spectrometry quantifies the emergence of macrophage heterogeneity. *bioRxiv*, 665307 https://doi.org/10.1101/665307