Proteolytic processing of mesonivirus replicase polyproteins by the viral 3C-like protease

Sandra Blanck and John Ziebuhr

Institute of Medical Virology, Justus Liebig University, Giessen, Germany

Mesoniviridae are a family of insect RNA viruses that diverged profoundly from other families of the Nidovirales. Mesonivirus replicative proteins are produced from large polyprotein (pp) precursors (pp1a and pp1ab) through proteolytic cleavage by the viral 3C-like protease (3CLpro) and, possibly, other proteases. Using recombinant forms of the Cavally virus 3CLpro and pp1a/pp1ab-derived substrates, we characterized 3CLpro cleavage sites in mesonivirus polyproteins. Our data lead us to suggest that 3CLpro cleaves the central and C-proximal regions of mesonivirus pp1a/pp1ab at 12 conserved sites. Compared to other nidovirus homologues, the mesonivirus 3CLpro features a distinct substrate specificity, with asparagine at P2 being a major specificity determinant. Furthermore, we provide evidence that expression of the ORF1b-encoded part of pp1ab involves a -1 ribosomal frameshift at a conserved GGAUUUU heptanucleotide sequence in the ORF1a/1b overlap region. Taken together, the study identifies critical steps in the expression and maturation of mesonivirus replicative proteins.

Nidovirales (families Coronaviridae, Arteriviridae, Roniviridae and Mesoniviridae) are a phylogenetically diverse order of plus-strand RNA viruses that share similar genome organization and expression strategies. They include viruses with the largest RNA genomes known to date (>30 kb) and have been identified in vertebrate and invertebrate hosts (de Groot et al., 2012a, b; Gorbalenya et al., 2006). To date, the biology of invertebrate nidoviruses (Mesoniviridae and Roniviridae) has not been studied very well and our current knowledge is largely based on predictions derived from sequence analyses or studies of distantly related protein homologues from other (vertebrate) nidovirus families (Cowley et al., 2000; de Groot et al., 2012b; Lauber et al., 2012, 2013; Nga et al., 2011; Zirkel et al., 2011, 2013).

Previous studies identified a 3C-like (main) protease (3CLpro) encoded by the replicase ORF 1a of the mesonivirus genome (Nga et al., 2011; Zirkel et al., 2011, 2013). By analogy with other nidoviruses (Ziebuhr et al., 2000), 3CLpro is thought to cleave the central and C-terminal regions of the replicase polyproteins (pp) 1a and 1ab at multiple sites. A recent study suggested that the Cavally virus (CavV) 3CLpro undergoes autoproteolytic cleavage at two N-terminal and one C-terminal cleavage site(s) to generate a proteolytically active form of 3CLpro composed of 314 residues (Blanck et al., 2014). Also, there is evidence that mesonivirus 3CLpro substrate specificities differ from those of other nidovirus 3CLpro’s (Ziebuhr et al., 2000) in that they do not use Gln/Glu and small residues (Ala/Ser/Gly), respectively, as main P1 and P1’ specificity determinants (P numbering according to Schechter & Berger, 1967). The present study aimed to obtain additional insight into the enzyme’s substrate specificity and the role of 3CLpro in the proteolytic processing of the mesonivirus replicase polyproteins.

In a first set of experiments, we employed a ‘self-cleavage’ approach in which the 3CLpro domain and a range of pp1a-derived sequences were co-expressed in Escherichia coli as N-terminal maltose-binding protein (MBP) and C-terminal glutathione S-transferase (GST) fusion protein constructs (Fig. 1a). Induction of expression in E. coli Rosetta(DE3) cells at 18°C for 4 h resulted in autocatalytic cleavage of the 3CLpro, as previously shown (Blanck et al., 2014), and additional proteolytic cleavage(s) in the C-terminal part of the expressed protein. The C-terminal cleavage product was purified by GST-affinity chromatography and its N-terminal sequence was determined by five cycles of Edman degradation as described previously (Blanck et al., 2014). The identity of the N-terminal sequence, N-H-V-E-N, confirmed that cleavage had occurred between pp1a/pp1ab residues 2386\(\text{E}\rangle\text{N}2387\) (Fig. 1a, b). Next, we successively shortened the C-terminal pp1a/pp1ab region expressed from these constructs and determined, one after the other, 3CLpro cleavage sites.
**Fig. 1.** Determination of CavV 3CL\textsuperscript{pro} cleavage sites in pp1a and pp1ab. (a) Schematic representation of the 588 kDa pp1ab (with major functional domains indicated) and fusion protein constructs used in this study for the identification of 3CL\textsuperscript{pro} cleavage sites in CavV pp1a/pp1ab by cis-cleavage (top) and trans-cleavage assays (bottom). Numbers indicate amino acid positions in the CavV pp1ab (NCBI accession no. YP_004598981.2). Cleavage sites identified in pp1a and pp1ab are indicated by arrows and the five N-terminal amino acids as determined by Edman degradation analysis of the respective C-terminal cleavage products are given. Also shown are representative trans-cleavage assays using purified fusion protein substrates (S1 to 3) and recombinant 3CL\textsuperscript{pro} (P) (bottom). Reaction products were analyzed in 14% SDS-polyacrylamide gels stained with Coomassie brilliant blue (bottom). Protein substrates, N-terminal (MBP-containing) cleavage products and the recombinant, purified 3CL\textsuperscript{pro} are indicated by arrowheads. The C-terminal (GST-containing) cleavage products used for N-terminal sequence analysis either directly (S3-derived cleavage product) or after GST-affinity purification (S1 and S2-derived cleavage products) are indicated by asterisks. (b) Functional domains and 3CL\textsuperscript{pro}-mediated proteolytic processing of the CavV replicase polyprotein pp1ab. 3CL\textsuperscript{pro} autoprocessing sites (grey arrowheads) and other 3CL\textsuperscript{pro} cleavage sites (black arrowheads) determined in this and a previous study (Blanck et al., 2014) are indicated. ExoN, Exoribonuclease; Hel, helicase; N, nidovirus RdRp-associated nucleotidyltransferase domain; NMT, guanosine-N\textsuperscript{7} methyltransferase; OMT, ribose-2'-O methyltransferase; RdRp, RNA-dependent RNA polymerase; T, transmembrane domains; Z, helicase-associated zinc-binding domain.
present in the expressed pp1a/pp1ab sequences (Fig. 1a, b). In all cases, GST-containing cleavage products generated by 3CL\textsuperscript{pro} could be readily purified by GST-affinity chromatography and N-terminally sequenced using previously established protocols (Blanck et al., 2014). The N-termini determined for the processing products obtained from the respective protein constructs are summarized in Fig. 1(a) (top). The data revealed a total of five 3CL\textsuperscript{pro} processing sites in the pp1a region downstream of the previously determined 3CL\textsuperscript{pro} autoprocessing sites (Blanck et al., 2014) (see Fig. 1a, b). Surprisingly, our data suggested that the most C-terminal 3CL\textsuperscript{pro} cleavage site in CavV pp1a (\(\text{ggattttc}\)) was located more than 110 residues upstream of the CavV pp1a C-terminus. In most other nidovirus polyproteins (except for toroviruses, which have a phosphodiesterase domain in this polyprotein region; Smits et al., 2006; Snijder et al., 2003; Zhao et al., 2012), conserved 3CL\textsuperscript{pro} cleavage sites are located a few residues upstream of the pp1a C-terminus (Ziebuhr et al., 2000) and the ribosomal frameshift (RFS) signal involved in pp1ab expression (Brierley, 1995; Brierley et al., 1987, 1989; Herold & Siddell, 1993). Proteolytic cleavage at this most C-terminal pp1a cleavage site has two major consequences: it removes a small peptide from the pp1a C-terminal processing product and generates the N-terminus of the nidovirus RdRp-containing processing product, the latter containing a few ORF1a-encoded N-terminal residues (reviewed by de Groot et al., 2012b; Gorbalenya et al., 2006; Snijder et al., 2013; Ziebuhr et al., 2000).

To identify potential 3CL\textsuperscript{pro} cleavage sites near the ORF1a/1b junction, we went on to express a slightly larger protein construct containing pp1ab residues 1799–2530 (Fig. 1a). To do this, we added one nucleotide (T) to the predicted slippery sequence (\(\text{ggattttc} \rightarrow \text{ggatttttgc}\)) to join the ORF1a/1b coding sequences and produce an authentic pp1ab sequence at the 1a/1b junction site, assuming that the programmed –1 frameshift takes place at the predicted site (Nga et al., 2011; Zirkel et al., 2011). Following expression of this fusion protein construct, we identified and purified the C-terminal cleavage product and determined its N-terminal sequence as N-V-A-R-Y (Fig. 1a, top). The sequence data led us to conclude that 3CL\textsuperscript{pro} cleaves the CavV pp1ab between 2386\textsuperscript{ni}2387 in an ORF1b-encoded sequence downstream of the CavV RFS site, suggesting that the mesonivirus N+RdRp is encoded (exclusively) by ORF1b.

To further corroborate our conclusions on 3CL\textsuperscript{pro} cleavage sites in the ORF1a-encoded polyprotein regions, we performed a \textit{trans}-cleavage reaction (see below) using an affinity-purified MBP-pp1a-1799-2134-GST fusion protein as substrate and recombinant 3CL\textsuperscript{pro} (Blanck et al., 2014). Following incubation for 16 h at 22 °C, the reaction mixtures were separated by SDS-PAGE, transferred onto PVDF membrane, and the relevant processing products were isolated and subjected to N-terminal sequence analysis. Consistent with the fusion protein autoprocessing data summarized in Fig. 1(a) (top), the N-termini of the cleavage products obtained in this additional \textit{trans}-cleavage assay were found to be N-L-N-I-A and A-Q-K-I-M, respectively (not shown), suggesting that both the recombinant form of CavV 3CL\textsuperscript{pro} used in the \textit{trans}-cleavage experiments and the 3CL\textsuperscript{pro}-containing fusion protein constructs exhibited robust protease activities that produced consistent CavV 3CL\textsuperscript{pro} specificity data.

Mesonivirus ORF1a-encoded polyprotein regions diverged profoundly from other nidoviruses while there was a higher degree of conservation for ORF1b, which allowed major replicative enzymes to be identified, including the nidovirus RdRp-associated nucleotidyl transferase (N), RdRp, helicase, exoribonuclease and methyltransferase domains (Lauber et al., 2012, 2013; Lehmann et al., 2015; Nga et al., 2011; Zirkel et al., 2011, 2013) (Fig. 1). Based on these earlier studies, we delimited putative interdomain regions and then determined experimentally the scissile peptide bonds cleaved by 3CL\textsuperscript{pro} in the C-terminal part of pp1ab. To do this, we performed \textit{trans}-cleavage reactions in vitro using recombinant CavV 3CL\textsuperscript{pro} (Blanck et al., 2014) and a range of fusion protein substrates expressed in \textit{E. coli}. Based on comparative sequence analyses of mesonivirus pp1ab sequences (Zirkel et al., 2013), three fusion protein substrates (S1, S2 and S3) were designed. The substrates comprised putative CavV pp1ab interdomain sequences that were fused to MBP (at the N-terminus) and GST or His tag at the C-terminus (Fig. 1a, bottom). The substrate proteins S1 and S2 encoded by plasmids pMAL-c2-[pp1ab-3534-3617-GST] and pMAL-c2-[pp1ab-4282-4356-GST], respectively, were expressed in \textit{E. coli} TB1 cells at 18 °C for 6 h, while substrate protein S3 was expressed from pMAL-c2-[pp1ab-4820-5088-His\textsuperscript{pro}]. Following incubation for 16 h (Fig. 1a, bottom). SDS-PAGE analysis of the reaction mixtures confirmed 3CL\textsuperscript{pro}-cleavage reactions, the substrates (S1, S2, S3) were subjected to Edman degradation analysis, which identified G-R-L-I-T and A-E-K-M-P as the N-terminal sequences of the processing products derived from S1 and S2, respectively (Fig. 1a, bottom). 3CL\textsuperscript{pro}-mediated cleavage of substrate S3 yielded a C-terminal His tag-containing a processing product of approximately 28 kDa (Fig. 1a and data not shown) that, following electrophoretic transfer onto PVDF membrane, was subjected directly to N-terminal sequence analysis. The N-terminus of the 28 kDa processing product was determined to be A-A-L-A-K (Fig. 1a).
Taken together, our *in vitro* study identified a total of nine (new) 3CL<sup>pro</sup> cleavage sites in the CavV pp1a/pp1ab sequence, the precise positions of which could be confirmed by N-terminal sequence analysis of relevant cleavage products (Fig. 1b, black arrowheads). Together with information on 3CL<sup>pro</sup> autoprocessing sites (Fig. 1b, grey arrowheads) (Blanck et al., 2014), our data suggest that CavV pp1a/pp1ab is processed by the virus-encoded 3CL<sup>pro</sup> at 12 sites (Fig. 1 and Fig. 2a). The data support and extend previous conclusions on the unique specificity of the CavV 3CL<sup>pro</sup> employing P2-Asn (rather than P1-Gln/Glu) as the major specificity determinant (Blanck et al., 2014; Ziebuhr et al., 2000). Comparative sequence analysis of the 12 CavV 3CL<sup>pro</sup> cleavage sites in pp1a/

**Fig. 2.** 3CL<sup>pro</sup> cleavage sites in mesonivirus replicase polyproteins. (a) P8 to P8′ positions of 3CL<sup>pro</sup> cleavage sites in the CavV pp1a and pp1ab determined in this and a previous study (Blanck et al., 2014). Numbers indicate amino acid positions in CavV pp1ab (NCBI accession no. YP_004598981.2). (b) Sequence logo presentations (Schneider & Stephens, 1990) of 12 experimentally confirmed CavV 3CL<sup>pro</sup> processing sites (left panel) and 132 putative mesonivirus 3CL<sup>pro</sup> cleavage sites (right panel), the latter identified by comparative sequence analysis of 11 mesonivirus replicase polyprotein sequences [see legend for (c)]. Residues that flank the scissile bond (between P1 and P1′) are given using the nomenclature introduced by Schechter & Berger (1967). The height of each letter (amino acid residue) is proportional to the frequency of a specific residue at a given position. (c) Sequence alignment of CavV pp1ab residues 4265 to 4365 and corresponding pp1ab sequences of 11 representative mesoniviruses. The sequence alignment was produced using Clustal Omega (Sievers et al., 2011) and converted using ESPript (http://espript.ibcp.fr) (Robert & Gouet, 2014). A black background colour indicates invariant residues. The putative 3CL<sup>pro</sup> cleavage site is indicated, together with the partially conserved P2 residue (indicated by a filled diamond). Also shown is the conserved superfamily I helicase (Hel) motif VI (Gorbalenya & Koonin, 1993) and motif I of the putative exoribonuclease (ExoN) domain (Nga et al., 2011; Snijder et al., 2003; Zirkel et al., 2013). CavV, Cavally virus (isolate C79); NDiV, Nam Dinh virus (isolate 02VN178); Houston virus (strain V3982); HanaV, Hana virus (strain A4/Ci/2004); NseV, Nse virus (strain F4/Ci/2004); NgewotanV, Ngewotan virus (strain JKT9982); BontangV, Bontang virus (strain JKT7774); KsaV, Karang Sari virus (strain JKT10701); KPhV, Kamhang Phet virus (strain KP84-0344); CASV, Casuarina virus (isolate 0071).
pp1ab revealed that, except for one site, Asn is strictly conserved at the P2 position (Fig. 2a, b, left). Further studies showed that these sites are conserved in other mesonivirus polyproteins (Fig. 2b, c), suggesting a mesonivirus-wide conservation of 3CL\text{pro} substrate specificity and supporting the biological significance of the processing sites identified in this and a previous study (Blanck et al., 2014). The logo representations in Fig. 2(b) reveal Asn at the P2 position as a key specificity determinant for mesonivirus 3CL\text{pro}'s. Only at a single 3CL\text{pro} cleavage site, Asn was less conserved (Fig. 2c) and replaced with Thr (or Ile) in about half of the mesonivirus polyproteins included in this analysis. Possible implications of the Asn-to-Thr/Ile substitution at this particular site remain to be investigated. We note that the putative helicase/exoribonuclease cleavage site is flanked by highly or even strictly conserved residues including the P6, P5, P4 and P1' positions, possibly indicating that other residues contribute to substrate binding in this case. Interestingly, there are other examples of nidovirus helicase/exoribonuclease cleavage sites with 'noncanonical' substitutions. For example, the conserved P1-Gln is replaced by His in the equivalent nsp13/14 cleavage sites of the human coronaviruses HCoV-HKU1 and HCoV-NL63 (van der Hoek et al., 2004; Woo et al., 2005a, b), whereas mouse hepatitis virus and closely related betacoronaviruses have Cys at the P1' position, possibly indicating specific constraints at this particular site that remain to be identified in further studies. Together, our analysis of putative mesonivirus 3CL\text{pro} cleavage sites confirmed a strong conservation of Asn at P2 and partial conservation at other positions (Fig. 2b, right), suggesting that, in addition to the P2-Asn, residues at the P4, P1', P1 and (possibly) P6 positions may interact with the 3CL\text{pro} substrate-binding pocket.

As discussed above, our data suggest an unusual N-terminal N+RdRp processing site encoded by ORF1b, downstream of the putative RFS. Both the position of this cleavage site and the sequence of the RFS are not conserved in other nidoviruses (Brierley, 1995; Brierley et al., 1987; Cowley et al., 2000; Ziebuhr et al., 2000). The heptanucleotide sequence is conserved among mesoniviruses (Fig. 3a) and located upstream of a putative RNA structural element (Nga et al., 2011; Vasilakis et al., 2014; Warrilow et al., 2014; Zirkel et al., 2011, 2013). To provide direct experimental evidence for the presumed functional role of this sequence, we expressed fusion protein constructs containing the CavV ORF1a/1b overlap region. Competent E. coli TBI cells were transformed with pMAL-c2-[pp1a/1ab-2401-2694-GST] (WT) plasmid DNA or mutant derivatives containing specific nucleotide substitutions (FS1 to FS4) in the ORF1a/1b overlap region, including the putative slippery sequence (Fig. 3b). Protein expression was induced with IPTG and bacteria were grown at 18°C for 5 h. As a control, we inserted an additional thymidine in the putative slippery sequence, resulting in a continuous/fused ORF1ab that encodes a C-terminally extended translation product whose expression does not depend on ribosomal frameshifting (Fig. 3b, Fus). The predicted size of the full-length fusion protein (MBP-pp1ab-2401-2694-GST) was 108 kDa and that of the MBP-pp1a-2401-2499 fusion protein (terminated at the ORF1a stop codon) was 54 kDa. Protein expression was analyzed by SDS-PAGE and Western blotting using total cell lysates and antibodies specific for the C-terminal GST-tag and the N-terminal MBP domain (Fig. 3c). Expression of the fusion protein containing the wild-type sequence resulted in a ~100 kDa protein that was detectable by using GST- (Fig. 3c, top) and MBP-specific (Fig. 3c, bottom) antibodies. Also, induction of expression gave rise to a ~55 kDa protein that could be detected by MBP-specific (but not GST tag-specific) antibodies (Fig. 3c). The sizes observed for the major translation products corresponded to those calculated for the proteins whose expression was terminated at the ORF1a stop codon or continued by RFS into the ORF1b sequence. Consistent with this interpretation, the 100 kDa 'frameshift protein' in lane WT+ was found to comigrate with our control protein expressed from a fused ORF1ab reading frame (Fig. 3c, lane Fus+), further supporting the identity of the ~100 kDa translation product observed in lane WT+. Nucleotide substitutions in the ORF1a/1b overlap region immediately upstream of the slippery sequence (FS1, FS2) or within the spacer sequence (FS4) that links the putative slippery sequence with the downstream RNA structural element (Nga et al., 2011; Vasilakis et al., 2014) resulted in translation products that comigrated with those expressed from the wild-type construct, suggesting that frameshifting was not affected by these mutations. In contrast, in the FS3 construct, expression of the 100 kDa protein was not detectable, while expression of the 55 kDa ORF1a termination product remained detectable (Fig. 3c), suggesting that the nucleotide substitution in the predicted slippery sequence (at the ribosomal A site) abolished RFS. Together, these data provide strong support for previous predictions on GGAUUUU being the site at which programmed −1 RFS into ORF1b occurs to ensure translation of ORF1b.

Taken together, the study confirms and extends previous conclusions on the unique specificity of mesonivirus main proteases and suggests a 3CL\text{pro} processing map for the replicase polyproteins of CavV and other mesoniviruses. Furthermore, the study provides insight into the expression of the second ORF in the mesonivirus genome, which is shown to occur at a unique slippery sequence not conserved in other nidoviruses. We also show that, in contrast to other nidoviruses, the mesonivirus N+RdRp is encoded by ORF1b alone. Together, the data obtained in this and previous studies provide an excellent basis for the heterologous expression and characterization of individual subunits of the mesonivirus replication/transcription complex.
Fig. 3. Analysis of the putative CavV RFS site. (a) Multiple sequence alignment of mesonivirus ORF1a/1b overlap regions. Nucleotides 7828 to 7863 of the CavV genome were aligned with the corresponding genome regions of 10 representative mesoniviruses (for the abbreviations of the virus names, see Fig. 2). The ORF1a stop codon, the 5'-terminal part of the predicted RNA secondary structure and the putative slippery sequence are indicated (Nga et al., 2011; Zirkel et al., 2011). (b) Mutations in the ORF1a/1b overlap region expressed as part of fusion protein constructs in E. coli transformed with pMAL-c2-[pp1a/1ab-2401-2694-GST] plasmid DNA or mutant derivatives containing the indicated mutations. Nucleotide substitutions and insertions are indicated by a grey background colour. The putative slippery sequence is indicated in bold. (c) Western blot analysis of total cell lysates obtained from E. coli cells expressing recombinant fusion proteins containing the indicated WT or mutant ORF1a/1b overlap regions. Fusion proteins were detected by Western blotting using GST-specific (top) and MBP-specific (bottom) antibodies. FS1 to 4, Frameshift variants 1 to 4 as indicated in (b); Fus, construct containing a single-nucleotide insertion to 'fuse' ORFs 1a and 1b; +, induced; −, not induced. The positions of protein size markers are indicated to the right.
Acknowledgements

The work was supported by the Deutsche Forschungsgemeinschaft (DFG11021, A01). We would like to thank Karin Schulteß for excellent technical help.

References


