Conservation of substrate specificities among coronavirus main proteases

Annette Hegyi and John Ziebuhr

Institute of Virology and Immunology, University of Würzburg, Versbacher Straße 7, 97078 Würzburg, Germany

The key enzyme in coronavirus replicase polyprotein processing is the coronavirus main protease, 3CL\textsuperscript{pro}. The substrate specificities of five coronavirus main proteases, including the prototypic enzymes from the coronavirus groups I, II and III, were characterized. Recombinant main proteases of human coronavirus (HCoV), transmissible gastroenteritis virus (TGEV), feline infectious peritonitis virus, avian infectious bronchitis virus and mouse hepatitis virus (MHV) were tested in peptide-based trans-cleavage assays. The determination of relative rate constants for a set of corresponding HCoV, TGEV and MHV 3CL\textsuperscript{pro} cleavage sites revealed a conserved ranking of these sites. Furthermore, a synthetic peptide representing the N-terminal HCoV 3CL\textsuperscript{pro} cleavage site was shown to be effectively hydrolysed by non-cognate main proteases. The data show that the differential cleavage kinetics of sites within pp1a/pp1ab are a conserved feature of coronavirus main proteases and lead us to predict similar processing kinetics for the replicase polyproteins of all coronaviruses.

Coronaviruses are positive-strand RNA viruses with exceptionally large genome sizes. Based on a similar polycistronic genome organization, common gene expression strategies and a conserved array of homologous domains in the viral polyprotein, the Coronaviridae have been united with the Arteriviridae in the order Nidovirales (den Boon et al., 1991; Cavanagh, 1997). Both coronaviruses and arteriviruses produce a nested set of subgenomic mRNAs using a unique discontinuous transcription mechanism (Spaan et al., 1983). Most probably, the noncontiguous 5\textsuperscript{‘} and 3\textsuperscript{‘} sequences are fused during negative-strand RNA synthesis (Sawicki & Sawicki, 1998; Sawicki et al., 2001) in a process that, based on arterivirus reverse-genetics data, has been suggested to resemble similarity assisted, copy-choice RNA recombination (van Marle et al., 1999).

It has been shown recently that all the protein functions required for coronavirus replication and transcription are encoded by the replicase gene (Thiel et al., 2001b). This gene occupies the 5\textsuperscript{‘}-proximal two-thirds of the genome and comprises two open reading frames (ORFs), ORFs 1a and 1b, which are connected by a ribosomal frameshift site (Brierley et al., 1987). Thus, two overlapping polyproteins are translated from the genome RNA: ORF1a encodes a ~450 kDa protein, called pp1a, and ORFs 1a and 1b together encode the C-terminally extended frameshift protein, pp1ab, with a molecular mass of ~750 kDa. The two replicate polyproteins are processed extensively by viral proteases. The N-proximal region of pp1a/pp1ab is cleaved by virus-encoded papain-like proteases (Baker et al., 1989; Bonilla et al., 1997; Herold et al., 1998; Kanjanahaluethai & Baker, 2000; Lim et al., 2000; Ziebuhr et al., 2001), while the C-proximal region is processed by the coronavirus main protease, also called 3C-like protease (3CL\textsuperscript{pro}). 3CL\textsuperscript{pro} cleaves the replicase polyproteins at 11 conserved interdomain junctions (reviewed by Ziebuhr et al., 2000) and shares a remote similarity with the picornavirus 3C proteases (Gorbalenya et al., 1989b; Liu et al., 1994, 1997, 1998; Liu & Brown, 1995; Lu et al., 1995, 1998; Ziebuhr et al., 1995, 1997; Grötzinger et al., 1996; Tibbles et al., 1996; Heusipp et al., 1997a, b; Denison et al., 1999; Ziebuhr & Siddell, 1999). The substrate specificity of coronavirus main proteases is determined mainly by the P1, P2 and P1\textsuperscript{’} positions (the amino acids flanking the protease cleavage sites are numbered from the N to the C terminus as follows: –P3–P2–P1 \downarrow P1\textsuperscript{’}–P2\textsuperscript{’}–P3\textsuperscript{’}–; Schechter & Berger, 1967), which are occupied preferentially by LQ[S or LQ[A. Sequence comparisons and mutagenesis data revealed that coronavirus main proteases most probably employ a catalytic dyad of conserved His and Cys residues, rather than the catalytic His–Asp(Glu)–Cys triad present in other RNA virus 3C-(like) proteases (Bazan & Fletterick, 1988; Gorbalenya et al., 1989b; Liu & Brown, 1995; Lu & Denison, 1997; Ziebuhr et al., 1997, 2000). Furthermore, coronavirus main proteases have a unique C-terminal domain (Gorbalenya et al., 1989b) that appears to be involved in proteolytic activity. Thus, truncations of this domain reduced significantly or abolished completely the proteolytic activities of the avian infectious bronchitis virus
and the 3C-like main proteases are indicated by arrows. TB1(pMALc2-FIPV-3CL) cells (Hegyi et al., 1995; Seybert et al., 1997) were cloned into the bacterial expression plasmid pMAL-c2 (New England Biolabs) and expressed as fusions with the Escherichia coli maltose-binding protein (MBP). This strategy has been proven previously to be suitable for the expression of the HCoV and MHV main proteases (Ziebuhr et al., 1995; Seybert et al., 1997). The fusion proteins were partially purified on amylose–agarose columns as described previously (Ziebuhr et al., 1995; Herold et al., 1996) and the 3CLpro domains with their authentic N- and C-termini were released by factor Xa cleavage (Fig. 1).

Three sets of synthetic 15-mer peptides (Table 1), which represented corresponding 3CLpro cleavage sites in the replicase polyproteins of HCoV, TGEV and MHV, were synthesized by solid-phase chemistry (Merrifield, 1965) and used in competition experiments to determine relative cleavage efficiencies, expressed as $K_{\text{m-rel}}$(V_max/K_mrel). The identity and purity of the peptides were confirmed by mass spectroscopy and HPLC (Jerini Bio-Tools). Two of the peptides, SP1 and SP4, represented the sites flanking the 3CLpro domain (P1P2 and P2P3; the mature proteins, which are released from pp1a/pp1ab by 3CLpro, are numbered continuously from P1 to P13, with P1 being the most N-terminal product; Ziebuhr et al., 2000) and two other peptides, SP5 and SP6, represented the sites flanking the processing product immediately upstream of the putative growth factor-like domain (P5P6; Gorbunova et al., 1989b; Lu et al., 1998; Ziebuhr & Siddell, 1999). It should be noted that the P5P6 junction is a so-called noncanonical cleavage site because its P1 position is occupied by Asn, rather than Ala, Ser or Gly, which are usually found at this position. The competitive cleavage assays were done as described previously (Ziebuhr & Siddell, 1999). Briefly, two peptides (SP1 and another peptide, each at 300 µM) were incubated with recombinant 3CLpro (0.3 µM HCoV 3CLpro, 0.3 µM TGEV 3CLpro and 1.8 µM MHV 3CLpro) in 10 mM Bis-Tris–HCl buffer (pH 7.0) at 25 °C. Aliquots were removed, added to an equal volume of 2% trifluoroacetic acid and stored at −80 °C prior to analysis. The reaction products were separated by reverse-phase HPLC on a Delta Pak C18 column (3.9 × 150 mm; Waters) using a 5–90% linear gradient of acetonitrile in 0.1% trifluoroacetic acid, as previously described (Ziebuhr et al., 1997), and the elution was monitored at an absorbance wavelength of 215 nm. The 3CLpro cleavage efficiencies of specific sites in relation to a standard cleavage site (P1P2, represented by SP1) were calculated using the methods described by Pallai et al. (1989). The data of these experiments are summarized in Table 1. We obtained clear evidence for differential kinetics in the cleavage of the P1P2, P2P3, P5P6 and P6P7 sites. The sites flanking the 3CLpro-based cleavage assays. To this end, the 3CLpro-coding sequences of HCoV (strain 229E; Herold et al., 1993), porcine transmissible gastroenteritis virus (TGEV, strain Purdue-115; Eleouet et al., 1995), feline infectious peritonitis virus (FIPV, strain 79–1146; GenBank accession no. AF3265755), IBV (strain Beaudette; Boursnell et al., 1987) and MHV (strain JHM; Lee et al., 1991) were cloned into the bacterial expression plasmid pMAL-c2 and expressed as fusions with the Escherichia coli maltose-binding protein (MBP). In a previous study, we have shown that peptides representing different cleavage sites in the HCoV pp1a/pp1ab are not equally susceptible to proteolysis by recombinant 3CLpro (Ziebuhr & Siddell, 1999). To gain additional insights into the substrate preferences of coronavirus main proteases, we have now extended these studies and characterized recombinant main proteases of five coronaviruses in peptide-based cleavage assays. To this end, the 3CLpro-coding sequences of HCoV (strain 229E; Herold et al., 1993), porcine transmissible gastroenteritis virus (TGEV, strain Purdue-115; Eleouet et al., 1995), feline infectious peritonitis virus (FIPV, strain 79–1146; GenBank accession no. AF3265755), IBV (strain Beaudette; Boursnell et al., 1987) and MHV (strain JHM; Lee et al., 1991) were cloned into the bacterial expression plasmid pMAL-c2 (New England Biolabs) and expressed as fusions with the Escherichia coli maltose-binding protein (MBP). These experiments are summarized in Table 1. We obtained clear evidence for differential kinetics in the cleavage of the P1P2, P2P3, P5P6 and P6P7 sites. The sites flanking the 3CLpro-
domain were found to be cleaved most effectively in all three viruses analysed. In contrast, the noncanonical P5|P6 site was hydrolysed far less efficiently. Based on these results, a ranking of pp1a/pp1ab cleavage sites can be inferred for each of the viruses tested. The conservation of this ranking among prototypic viruses from the coronavirus groups I and II supports the biological significance of the data and leads us to predict that the order in which pp1a/pp1ab cleavage events occur may be very similar in all coronaviruses. It is also reasonable to suggest that the structural properties residing in the 15-mer peptides are critically involved in determining the half-life of specific processing intermediates (Pallai et al., 1989). Furthermore, secondary structure predictions (Gorbalenya et al., 1989b) have indicated that most (if not all) of the coronavirus 3CLpro cleavage sites are located at (solvent-exposed) interdomain junctions, which should make these sites easily accessible to the trans-acting protease. However, we do not wish to exclude that, at least in some cases, higher order structures and folding of the replicase polyproteins may modulate the cleavage kinetics of specific sites.

The poor activities of the recombinant enzymes towards the SP5 peptides, which represent the noncanonical P5|P6 cleavage site, Gln|Asn, strongly suggests structural constraints for the coronavirus 3CLpro S1′-binding sites, as discussed previously (Ziebuhr & Siddell, 1999; Ziebuhr et al., 2000), and provides a plausible explanation for the conservation of small aliphatic residues (Ala or Ser) at the P1′ position in coronavirus 3CLpro cleavage sites (Ziebuhr et al., 2000). The exceptional degree of sequence conservation among coronavirus P5|P6 sites leads us to speculate that (i) either the slow cleavage of the P5|P6 site has been preserved during evolution to extend the half-life of a precursor protein containing both P5 and P6 or (ii) the N terminus of P6, which is the conserved sequence NNE(L/I)MP, is required for the biological activity of the coronavirus P6 protein, which remains to be determined. In other words, functional constraints of the mature proteins may have dictated the conservation of this sequence and even accepted unfavourable cleavage kinetics in this case. Genetically engineered coronavirus mutants (Almazán et al., 2000; Thiel et al., 2001a) carrying amino acid substitutions at the P5|P6 site should be extremely informative in identifying the selective forces that determined the conservation of this sequence.

The rapid proteolysis of the SP1 and SP4 substrates suggests that the autocatalytic release of 3CLpro from the viral polyproteins is an early processing event. If this conclusion is correct, most 3CLpro cleavages within pp1a/pp1ab should occur in trans. A similar conclusion has been reached in another study using alternative approaches (Lu et al., 1996). The conserved ranking of pp1a/pp1ab cleavage sites, with the N-terminal 3CLpro autoprocessing sites being, in most cases, the most efficiently cleaved substrates, prompted us to address the
question of whether this site would be a suitable substrate for main proteases from all coronavirus groups. To answer this question, we incubated the partially purified and factor Xa-activated HCoV, IBV, FIPV, MHV and TGEV main proteases from all coronavirus groups. To answer this question, we incubated the partially purified and factor Xa-activated HCoV, IBV, FIPV, MHV and TGEV main proteases with the HCoV SP1 peptide, which represents the N-terminal HCoV 3CLpro cleavage site, and analysed the reaction products by reverse-phase chromatography. As shown in Fig. 2, all recombinant coronavirus main proteases tested in this experiment cleaved this peptide, albeit with slightly different kinetics. Surprisingly, FIPV 3CLpro proved to be even more active towards the HCoV-derived substrate than the cognate enzyme (Fig. 2, B and D). We conclude from this experiment that, despite the considerable sequence diversity among coronavirus main proteases (Ziebuhr et al., 2000), the substrate specificities are highly conserved. The results lead us to believe that both the development of universally applicable 3CLpro assays and the design of broad-spectrum inhibitors blocking all coronavirus main proteases should be feasible.

We thank Paul Britton, Rosa Casais and David Brown for the cDNA clone pFRAG-2 (Casais et al., 2001) containing the IBV 3CLpro sequence. The work was supported by grants from the Deutsche Forschungsgemeinschaft (GR 222/2-2, ZI 618/2-1) and the Bayerische Forschstiftung (Neue Antinfektiva).

References


Received 21 September 2001; Accepted 20 November 2001