Identification of protease and ADP-ribose 1″-monophosphatase activities associated with transmissible gastroenteritis virus non-structural protein 3

Ákos Putics, Alexander E. Gorbalenya and John Ziebuhr

1Institute of Virology and Immunology, University of Würzburg, Versbacher Straße 7, 97078 Würzburg, Germany
2Molecular Virology Laboratory, Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands

The replicase polyproteins, pp1a and pp1ab, of porcine Transmissible gastroenteritis virus (TGEV) have been predicted to be cleaved by viral proteases into 16 non-structural proteins (nsp).

Correspondence
John Ziebuhr
j ziebuhr@uni-wuerzburg.de
or j ziebuhr@qub.ac.uk

Received 9 October 2005
Accepted 12 November 2005

Porcine Transmissible gastroenteritis virus (TGEV) is a member of the Coronaviridae, a family of enveloped, positive-strand RNA viruses comprising the genera Coronavirus and Torovirus (González et al., 2003; Siddell et al., 2005). Based on genetic and antigenic criteria, TGEV has been classified as belonging to coronavirus group I, which also includes Human coronavirus 229E (HCoV-229E), human coronavirus NL63, Porcine epidemic diarrhea virus, Canine coronavirus (CCoV) and Feline coronavirus (FCoV) (González et al., 2003; van der Hoek et al., 2004).

Coronaviruses feature exceptionally large RNA genomes of about 30 kb and employ a unique transcription mechanism, called discontinuous extension of minus strands, to synthesize an extensive set of 5′ leader-containing, subgenomelength RNAs encoding the viral structural proteins and several species-specific (‘accessory’) proteins (Sawicki & Sawicki, 1995, 1998; Zúñiga et al., 2004; Siddell et al., 2005). Except for the nucleocapsid protein, which has recently been shown to be involved in coronavirus replication (Almazán et al., 2004; Schelle et al., 2005), all of the functions required for viral RNA synthesis are thought to be encoded by the replicase gene (Ziebuhr, 2005). This gene occupies about two-thirds of the genome (more than 20 000 nt) and comprises two large ORFs called 1a and 1b. ORF1a encodes the replicase polyprotein pp1a and ORFs 1a and 1b together encode a C-terminally extended version of pp1a, which is called pp1ab and requires ribosomal frameshifting for expression of its ORF1b-encoded portion (Brierley et al., 1987). Proteolytic processing by two or three proteases yields up to 16 processing end products called non-structural proteins (nsp) 1–16 that, together with several cellular proteins and the nucleocapsid protein, form the active replicase–transcriptase (Ziebuhr et al., 2000; Shi & Lai, 2005; Ziebuhr, 2005). The replicase–transcriptase has been demonstrated (or predicted) to include protease, RNA polymerase, helicase, endo- and exoribonuclease, ribose 2′-O-methyltransferase and ADP-ribose 1″-phosphatase (ADRP) activities (Gorbalenya et al., 1989; Lu et al., 1995; Ziebuhr et al., 1995, 2000; Seybert et al., 2000; Snijder et al., 2003; Bhardwaj et al., 2004; Ivanov & Ziebuhr, 2004; Ivanov et al., 2004a, b; Putics et al., 2005). The biologically relevant reactions catalysed by these enzymes are only slowly beginning to emerge and our understanding of the complex enzymology involved in coronavirus RNA synthesis is limited. For TGEV, none of the replicase gene-encoded activities, except for the 3C-like (main) protease (3CLpro, Mpro) (Anand et al., 2002, 2003; Hegyi & Ziebuhr, 2002), have so far been characterized.
In this study, we characterized two enzymic activities residing in TGEV nsp3, the largest pp1a/pp1ab processing product. Previous sequence analysis revealed that, despite significant sequence diversity in this part of the genome (Ziebuhr et al., 2001; Snijder et al., 2003), there are several domains in nsp3 that, with few exceptions, are conserved in coronaviruses. The typical arrangement of these domains (from N to C terminus) is as follows: acidic domain, papain-like protease 1 (PL1\textsuperscript{pro}), X domain, papain-like protease 2 (PL2\textsuperscript{pro}) and Y domain (Fig. 1) (Ziebuhr et al., 2001). The papain-like protease activities have previously been characterized for Murine hepatitis virus (MHV) (Dong & Baker, 1994; Bonilla et al., 1997; Kanjanahaluethai & Baker, 2000), HCoV-229E (Herold et al., 1998, 1999; Ziebuhr et al., 2001) and Severe acute respiratory syndrome coronavirus (SARS-CoV) (Thiel et al., 2003; Harcourt et al., 2004). Whilst most coronaviruses encode two papain-like protease domains, SARS-CoV nsp3 contains only one of these domains. Because its position in nsp3 corresponds to that of the PL2\textsuperscript{pro} domains of other coronaviruses, the SARS-CoV papain-like protease is called PL2\textsuperscript{pro} (Snijder et al., 2003; Thiel et al., 2003). Also, avian Infectious bronchitis virus (IBV) employs only PL2\textsuperscript{pro} to process the N-terminal pp1a/pp1ab region (Lim & Liu, 1998; Lim et al., 2000). In this case, however, remnants of an (inactive) PL1\textsuperscript{pro} domain have been identified, indicating that IBV has lost its PL1\textsuperscript{pro} activity in the course of evolution (Fig. 1) (Ziebuhr et al., 2001).

Upstream of PL2\textsuperscript{pro}, all coronaviruses encode a so-called X domain (Gorbalenya et al., 1991), which has recently been shown for HCoV-229E and SARS-CoV to mediate ADRP activity (Putics et al., 2005).

The activity and substrate specificity of the TGEV PL1\textsuperscript{pro} were characterized by expressing the domain together with flanking regions in rabbit reticulocyte lysates. Nucleotides 315–4319 of the TGEV (strain Purdue 46) genomic sequence (GenBank accession no. AJ271965) were amplified by RT-PCR using poly(A) RNA isolated from TGEV-infected cells as a template. The amplicon was inserted into pBlueScript II KS\textsuperscript{+} DNA (Stratagene) and the resultant plasmid, p31/32, was used as template to amplify the coding sequence of the pp1a/pp1ab residues Thr\textsuperscript{785}–Gln\textsuperscript{1335}. The PCR forward primer contained a bacteriophage T7 RNA polymerase promoter sequence followed by a translation initiation codon, and the reverse primer contained a translation stop codon. From this PCR product, m\textsuperscript{7G(5′)ppp(5′)G}-capped RNA was transcribed by using T7 RNA polymerase and subsequently translated in rabbit reticulocyte lysates in the presence of \textsuperscript{[\textsuperscript{35}S]}methionine by using previously described protocols (Ziebuhr et al., 2001). The expressed protein, 785–1335, essentially contained the nsp3 acidic and PL1\textsuperscript{pro} domains preceded by a predicted cleavage site, G\textsuperscript{879}GGG\textsuperscript{880}.

---

**Fig. 1.** Domain organizations and proteolytic-processing pathways of coronavirus N-terminal replicase polyprotein regions. Active protease domains are indicated in black and other domains are indicated in dark grey. The P1 and P1\textsuperscript{′} residues of characterized papain-like protease cleavage sites are given. Predicted sites (Ziebuhr et al., 2001) are indicated by question marks. Abbreviations: Ac, acidic domain; PL1, papain-like protease 1; X, X domain (with predicted or confirmed ADRP activity); PL2, papain-like protease 2; Y, Y domain (Ziebuhr et al., 2001); SUD, SARS-CoV unique domain (Snijder et al., 2003). The PL1 of IBV is crossed out to indicate that the domain has no proteolytic activity.
(Fig. 2a). Following incubation of the translation products for up to 8 h, the proteins were analysed by SDS-PAGE and autoradiography. As a negative control in this experiment, a mutant version of 785–1335, 785–1335_C1093A, was used, in which the codon of the presumed active-site nucleophile, Cys_{1093}, was replaced with an alanine codon by using PCR-based mutagenesis methods. The data shown in Fig. 2(b) suggest that 785–1335 (apparent molecular mass, 78 kDa)
undergoes autoproteolysis to produce cleavage products of approximately 68 and 10 kDa. By contrast, the control protein, 785–1335_C1093A, remained stable. The data are consistent with previous predictions (Eleouet et al., 1995; Ziebuhr et al., 2001) and lead us to conclude that the TGEV PL1Pro domain has proteolytic activity and Cys1093 has an essential (probably catalytic) function. Surprisingly, the apparent mass of the larger processing product of 68 kDa differed significantly from its calculated mass (51 kDa). To obtain additional evidence for cleavage at the predicted sequence, we expressed the predicted processing products, 785–879 and 880–1335, in separate reactions. The two proteins were analysed together with the products obtained from the 785–1335 autoprocessing reaction in the same gel (data not shown). We observed co-migration of the autoprocessing products with the two separately expressed (predicted) cleavage products, further supporting the predictions on the site of cleavage and confirming an aberrant migration of the C-terminal cleavage product in SDS-PAGE. Independent support for the site of cleavage also came from site-directed mutagenesis. Substitution with Ala of each of the three consecutive Gly residues at positions 878–880 abolished cleavage, suggesting a requirement for Gly residues at the cleavage site (Fig. 2c). To determine unambiguously the scissile bond, we performed an N-terminal radiosequence analysis of the C-terminal cleavage product (Ziebuhr et al., 2001). In this case, we used a mutant form of the protein, 785–1335_V884M/V890M, in which two Val residues downstream of the presumed cleavage site were replaced with Met. The mutated RNA was translated in the presence of [35S]methionine to produce two cleavage products that co-migrated with those derived from the parental protein, 785–1335 (data not shown), confirming that the two substitutions did not affect the autoprocessing activity of the protein. The processing products obtained from protein 785–1335_V884M/V890M were separated by SDS-PAGE and transferred electrophoretically onto a PVDF membrane (Bio-Rad). The C-terminal processing product was identified by autoradiography, excised from the membrane and subjected to radiosequence analysis using previously described protocols (Ziebuhr et al., 2001). The sequence data (Fig. 2d) provided conclusive evidence that TGEV PL1Pro cleaves the triple-glycine sequence between Gly879 and Gly880.

The data obtained in this study, together with our previous predictions on an upstream site, 110GlyAla111 (Fig. 1) (Ziebuhr et al., 2001), led us to predict that TGEV nsp2 comprises the pp1a/pp1ab residues Ala111–Gly879. To further corroborate this idea, we performed an additional set of experiments. First, we sought to produce an nsp2-specific antiserum. To this end, residues Lys526–Asp713 were expressed in Escherichia coli as a fusion with the maltose-binding protein (MBP). The protein was purified by amylose-affinity chromatography and used to immunize rabbits. The resultant antiserum, T3, was used in immuno-precipitation experiments as described previously (Ziebuhr et al., 1995). Briefly, 7 x 10⁶ ST cells (ATCC no. CRL-1746) were mock-infected or infected with TGEV at an m.o.i. of 100 TCID₅₀ per cell. Newly synthesized proteins were labelled metabolically with 30 μCi (1110 kBq) [35S]methionine.
(ml medium)\(^{-1}\) between 8 and 11 h post-infection. Subsequently, cytoplasmic lysates were prepared and incubated with either T3 serum or the corresponding preimmune serum for 45 min at 4 °C. Immune complexes were isolated and analysed by SDS-PAGE using 17-3 % gels. As shown in Fig. 2(e), antiserum T3 specifically precipitated a protein of \(~85\) kDa from virus-infected cells. The protein was not detectable in lysates obtained from mock-infected cells or if the corresponding preimmune serum was used. The size of the protein corresponded to the calculated size of fully processed nsp2, indicating that, besides the \(87^\text{G}\/G^{880}\) site, the predicted \(\text{GI}^{116}\text{A}^{111}\) site was also cleaved in infected cells. The corresponding N-terminal processing product, nsp1, remains to be identified.

In a final set of experiments, we characterized the so-called X domain, which is located between the PL1\(^{\text{pro}}\) and PL2\(^{\text{pro}}\) domains in nsp3 (Fig. 3). Recently, we have shown that the X-domain homologues of HCoV-229E and SARS-CoV are highly specific phosphatases that act to convert ADP-ribose 1′′-phosphate (Appr-1′′-p) to ADP-ribose (Putics \textit{et al.}, 2005). Intriguingly, the enzymic activity (although being conserved among coronaviruses) proved to be dispensable for HCoV-229E RNA synthesis and production of virus progeny in cultured cells, suggesting that the activity is involved in a non-essential regulatory function or provides a selective advantage only in the infected host (Putics \textit{et al.}, 2005). To further evaluate the biological significance of this newly identified activity, we investigated whether the TGEV X domain is functional. The coding sequence of the TGEV X domain (pp1a/pp1ab residues 1312–1486, Fig. 3) was expressed as an MBP fusion as described previously for the homologues from SARS-CoV and HCoV-229E. Examination of the ADRP activity of the amylose affinity-purified and factor Xa-cleaved protein provided convincing evidence that the protein had the predicted activity, although it was less active than the bacterially expressed homologue from HCoV-229E. To test whether this low activity is an intrinsic property of the TGEV protein, we expressed three additional versions of this protein with slightly extended or shortened N- and C-terminal protein segments. As shown in Fig. 3, variations of the domain borders strongly affected the ADRP activities of the expressed proteins. Under the conditions used in our assay, the protein TGEV-X-2 (and, as expected, our positive control, HCoV-229E X) dephosphorylated the substrate, Appr-1′′-p, completely. TGEV-X-1 proved to be less active, whilst TGEV-X-3 and TGEV-X-4 were completely inactive [also after longer incubation times (data not shown)]. The data confirm unambiguously that the TGEV X domain has the predicted ADRP activity and delimit the minimal catalytically active domain to pp1a/pp1ab residues Asp\(^{1320}\)--Ser\(^{1486}\). The data further suggest that the enzymic activity of the TGEV X domain critically depends on the integrity of N-terminal structure elements and thus could be modulated (or even inhibited) by the adjacent domains in nsp3.

In conclusion, X domain-associated ADRP activities have now been identified in members of three different coronavirus (sub)groups (Putics \textit{et al.}, 2005). Furthermore, X-domain homologues are conserved in the genomes of toroviruses, hepatitis E viruses, rubellaviruses and alphaviruses (Gorbalenya \textit{et al.}, 1991; Draker \textit{et al.}, 2006) and exist in many cellular organisms (so-called ‘macro-domain proteins’) (Karras \textit{et al.}, 2005). The function of these ubiquitous domains, which is probably promoted by ADP–ribose binding and the associated ADRP activity, remains to be identified.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (GK587/2, SFB479).

References


