Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Chapter 691

Arterivirus Serine Endopeptidase

Databanks

MEROPS name: equine arteritis virus serine peptidase
MEROPS classification: clan PA, subclan PA(S), family S32, peptidase S32.001
Tertiary structure: Available
Species distribution: known only from equine arteritis virus
Reference sequence from: equine arteritis virus (UniProt: P19811)

MEROPS name: porcine reproductive and respiratory syndrome virus nsp4 peptidase
MEROPS classification: clan PA, subclan PA(S), family S32, peptidase S32.002
Tertiary structure: Available
Species distribution: known only from porcine reproductive and respiratory syndrome virus
Reference sequence from: porcine reproductive and respiratory syndrome virus (UniProt: Q04561)

Name and History

The family *Arteriviridae* currently includes equine arteritis virus (EAV; the family prototype), porcine reproductive and respiratory syndrome virus genotypes I and II (PRRSV-I and PRRSV-II), lactate dehydrogenase-elevating virus (LDV), and simian hemorrhagic fever virus (SHFV) [1]. EAV is the best-characterized arterivirus, although recent studies have increasingly been focused on PRRSV due to its economic importance. Arteriviruses are enveloped viruses with a polycistronic plus-strand RNA genome (12−15 kb; [2−6]). Their replicase proteins are expressed from open reading frames (ORFs) 1a and 1b that encode two large polyproteins: pp1a (187−260 kDa) and pp1ab (345−422 kDa), the latter resulting from a C-terminal extension of pp1a via ribosomal frameshifting. Both polyproteins are processed extensively by three or four ORF1a-encoded endopeptidases to generate nonstructural proteins (nsps) [7−15]. The arterivirus proteases and proteolytic pathways should be compared with those of the distantly related coronaviruses (see Chapters 494 and 546) and roniviruses, all of which are united in the order *Nidovirales* [16,17].

The ***arterivirus serine proteinase*** (***SP***) domain is located in non-structural protein 4 (nsp4), a 21 kDa cleavage product from the central region of the ORF1a-encoded polypeptide. The SP was first identified and linked structurally to cellular chymotrypsin-family proteinases by comparative sequence analysis of the ORF1a protein of the arterivirus prototype, equine arteritis virus EAV [2], a prediction that was experimentally verified in subsequent expression experiments in cultured cells [10]. Due to a number of striking similarities with picornavirus cysteine 3C proteinases, arterivirus nsp4 is also called a 3C-like proteinase (see below).

Activity and Specificity

The activity of the EAV SP domain has been detected *in vivo*, both in infected cells and in transient expression systems upon expression of appropriate parts of the ORF1a polyprotein [8,10]. Moreover, *in vitro* activity assays were developed for both EAV and PRRSV using recombinant nsp4 purified from *E. coli* and synthetic peptides or recombinant proteins as substrates [18−20]. In EAV-infected cells, in addition to the mature nsp4, an nsp3-4 processing intermediate with a long half-life is synthesized. Nsp4 itself and all other proteins located downstream in the replicative polyproteins are derived through nsp4-mediated cleavages. The EAV SP is now known to mediate (at least) six cleavages at Glu↓Ser/Gly/Ala sites in the C-terminal half of the ORF1a protein [10,21] and three additional sites in the ORF1b-encoded polyprotein [12,13]. These nine sites are conserved in arteriviruses and can be described by the formula (Glu/Gln)↓(Gly/Ser/Ala/Asn/Lys). A similar substrate specificity was originally described for picornavirus cysteine 3C proteinases (picornains; see Chapter 537). The EAV nsp4 SP domain has been shown to be able to cleave a separately expressed ORF1b polyprotein in *trans* [13].

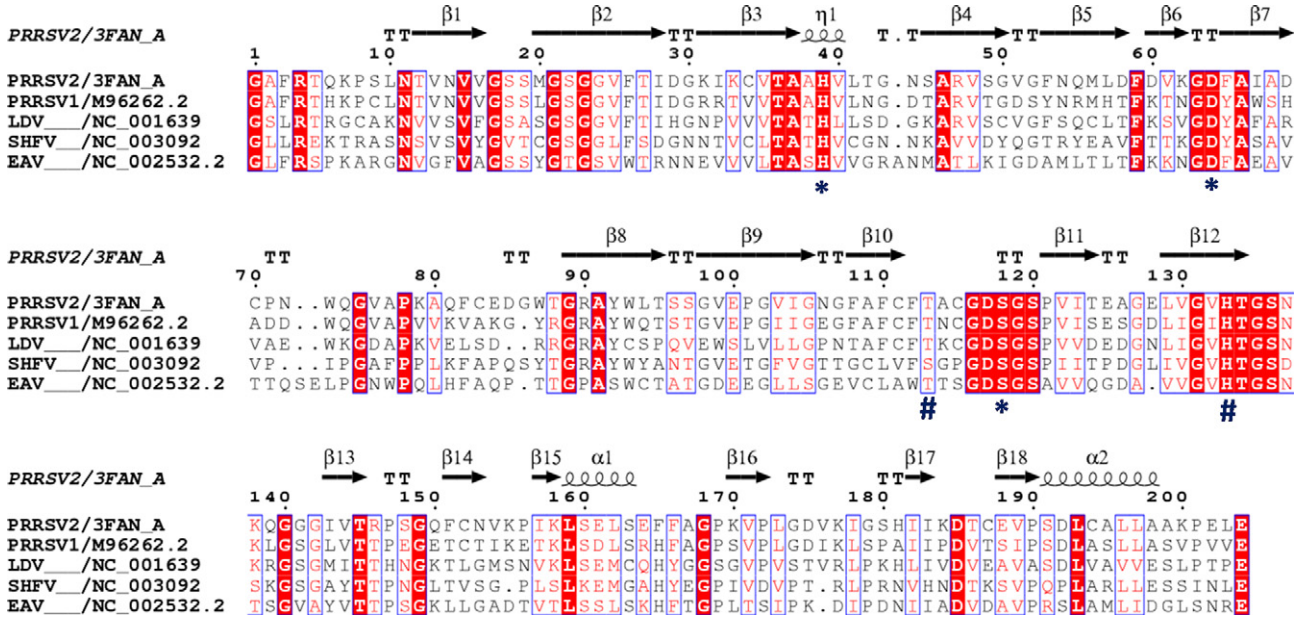

FIGURE 691.1 Multiple sequence alignment of the nsp4 SP domains of arteriviruses. Shown is an alignment of nsp4 of equine arteritis virus (EAV), porcine reproductive and respiratory syndrome virus genotypes I and II (PRRSV1 and PRRSV2, respectively), lactate-dehydrogenase-elevating virus (LDV) and simian hemorrhagic fever virus (SHFV). The alignment was produced with Muscle [24] and Clustal [25] using the Viralis platform [26] and was prepared for publication with ESPript 2.2 [27]. The positions of the catalytic Ser, His and Asp residues, and substrate-binding Thr/Ser and His residues of the protease domain are marked with * and #, respectively. Conserved and identical residues in all viruses are colored. Secondary structure elements for PRRSV-II nsp4 [19] are schematically shown on top of the alignment. GenBank and/or RefSeq accession numbers of respective virus genome sequences as well as the PDB accession number for PRRSV-II nsp4 are presented next to the virus acronyms.

Structural Chemistry

The arterivirus nsp4 SP is a (predicted) 196−204 residue proteolytic enzyme whose conserved sequence elements in the N-terminal ∼160 residue domain (Figure 691.1) match functionally important residues of clan PA enzymes belonging to two different groups. The SP utilizes the canonical His-Asp-Ser catalytic triad found in classical chymotrypsin-family proteinases. On the other hand, its putative substrate-binding region contains Thr/Ser and His residues which are conserved in the so-called viral 3C(-like) cysteine peptidases (Chapter 537) [10] and which determine their specificity for Gln (sometimes Glu) as the P1 residue of their cleavage sites and mainly Gly, Ala or Ser at the P1′ position.

The replacement of the members of the predicted catalytic triad of EAV nsp4 (His1103, Asp1129 and Ser1184; here and hereafter polyprotein numbering is used) confirmed their essential nature [10]. The Ser1184 to Cys mutant did not show any proteolytic activity. However, the replacement of Asp1129 by Glu, a substitution which can also be found in certain 3C-like cysteine peptidases, was partially tolerated. The substrate specificity of nsp4 and the putative role of Thr1179 and His1199 in substrate recognition were supported by the results of characterization of the predicted cleavage sites and substrate pocket mutants [10].

Nsp4 structures have now been obtained by X-ray crystallography for both EAV and PRRS-II [19,22] (Figure 691.2). The protein consists of three domains, with domains I and II forming the typical chymotrypsin-like two-β-barrel fold of the SP. In addition, arterivirus nsp4 SP also contains a C-terminal domain III of ∼40 residues with an uncharacterized function [19,22]. This domain III is connected to domain II through a flexible hinge region, and in EAV it was shown to be dispensable for proteolytic activity [23].

Preparation

To allow the biochemical and structural characterization of EAV and PRRSV nsp4 protocols for the large-scale production of recombinant nsp4 in *E. coli* were developed [19,20,22,23]. These protocols involve expression of nsp4 either carrying an N- or C-terminal hexahistidine tag, or fused to maltose binding protein or glutathione *S*-transferase, which was subsequently cleaved off. These recombinant proteinases were shown to be proteolytically active in different *trans*-cleavage assays.

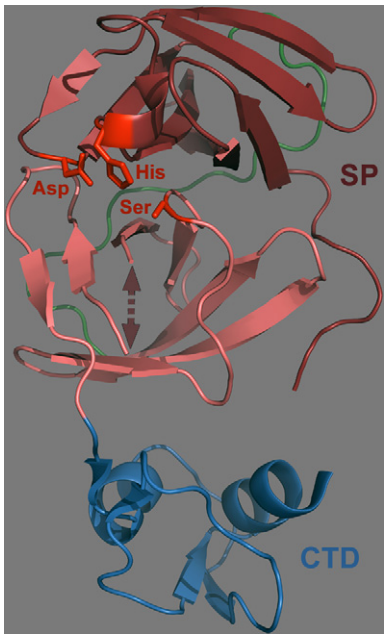

FIGURE 691.2 Ribbon diagram of the crystal structure of PRRSV-II nsp4. Representations were made using the Pymol molecular graphics system [28] and PDB entry 3FAN [19]. The protein consists of three domains of which the first two (in dark red and pink) form the typical chymotrypsin-like two-β-barrel fold. A long loop connecting domains I and II is highlighted in green, whereas the arrow represents a loop consisting of residues 136−140, which was not resolved in the structure. The residues of the catalytic triad are highlighted in red. The C-terminal domain (CTD or domain III) was shown to be dispensable for proteolytic activity in EAV nsp4 and is depicted in blue [22,23].

Biological Aspects

The extensive proteolytic processing of the replicase polyproteins is a major mechanism of regulation of genome expression of arteriviruses. The nsp4 SP is the main proteinase responsible for the production of the majority of nsps from the pp1a and pp1ab polyproteins (reviewed in [15−17]. Among the proteins produced are key viral enzymes like the nsp9 RNA-dependent RNA polymerase and the nsp10 RNA-helicase. The suppression of the production of the latter proteins, or the downstream-located nsp11 and nsp12, by mutagenesis of the respective nsp4-cleavage sites blocked EAV replication [13]. Likewise, mutagenesis of the SP cleavage sites in the C-terminal half of the ORF1a-encoded polyprotein invariably blocked or severely inhibited EAV RNA synthesis [21]. In EAV, the poorly understood fine-tuning of replicase polyprotein proteolysis by nsp4 may involve its domain III [23]. The proteolytic activity of nsp4 is strongly influenced by other processing steps mediated by two or three cysteine protease activities residing in the nsp1-2 region of the replicase polyproteins (see Chapters 495−497). In particular, cleaved nsp2 was reported to act as a co-factor in the cleavage by SP of the nsp4↓5 site at the C-terminus of the SP domain [14]. The presence of nsp2 probably promotes the correct spatial organization that is required for cleavage of this site in the polyprotein. In the absence of nsp2, the nsp5↓6 and nsp6↓7 sites rather than the nsp4↓5 site are cleaved by the SP [14]. As yet, the functional significance of the use of alternative processing pathways by the SP to process the C-terminal part of the ORF1a-encoded polyprotein remains unclear although it may limit the variety of proteins that could otherwise be derived from the polyprotein [16].

Distinguishing Features

The SP combines properties of two different groups of chymotrypsin-like enzymes: it utilizes the His-Asp-Ser catalytic triad found in classical chymotrypsin-like proteinases (clan PA, subclan S), and its putative substrate-binding region contains Thr/Ser and His residues which are typical of peptidases of clan PA subclan C, including diverse picornains (also known as cysteine 3C and 2A proteases of picornaviruses and 3C-like proteases of positive-stranded RNA viruses distantly related to picornaviruses) (Chapters 537−546). As such, it can be considered as the prototype of a novel subgroup, the 3C-like serine proteinases [10]. Similar proteinases were predicted and partially characterized in plant sobemoviruses (Chapter 692), luteoviruses and pea enation mottle virus, and animal astroviruses and toroviruses, a subfamily of the *Coronaviridae*.

Polyclonal peptide antisera against EAV nsp4 have been raised in rabbits and are available from the authors for research purposes on request [8].

Further Reading

For further information, the following papers are recommended: Snijder *et al.* [10], Ziebuhr *et al.* [16], van Aken *et al.* [18,21,23], Tian *et al.* [19], Xu *et al.* [20], and Barrette-Ng *et al.* [22].

References

[1] Faaberg, K.S., Balasuriya, U.B., Brinton, M.A., Gorbalenya, A.E., Leung, F.C., Nauwynck, H.J., Snijder, E.J., Stadejek, T., Yang, H., Yoo, D. (2012). Family Arteriviridae, in: *Virus Taxonomy,* the 9th Report of the International Committee on Taxonomy of Viruses, King, A., Adams, M., Carstens, E., Lefkowitz, E.J., eds., Academic Press, pp. 796−805.

[2] den Boon, J.A., Snijder, E.J., Chirnside, E.D., de Vries, A.A.F., Horzinek, M.C., Spaan, W.J.M. (1991). Equine arteritis virus is not a togavirus but belongs to the coronavirus like superfamily. *J. Virol.* 65, 2910−2920.

[3] Godeny, E.K., Chen, L., Kumar, S.N., Methven, S.L., Koonin, E.V., Brinton, M.A. (1993). Complete genomic sequence

and phylogenetic analysis of the lactate dehydrogenase-elevating virus (LDV). *Virology* 194, 585−596.

[4] Meulenberg, J.J.M., Hulst, M.M., de Meijer, E.J., Moonen, P.L.J.M., den Besten, A., de Kluyver, E.P., Wensvoort, G., Moormann, R.J.M. (1993). Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS), is related to LDV and EAV. *Virology* 192, 62−72.

[5] Snijder, E.J., Meulenberg, J.J.M. (1998). The molecular biology of arteriviruses. *J. Gen. Virol.* 79, 961−979.

[6] Snijder, E.J., Spaan, W.J.M. (2006). Arteriviruses, in: *Fields Virology*, 5th ed., Knipe, D.M., Howley, P.M., eds., Philadelphia, Pa.: Lippincott, Williams & Wilkins, pp. 1337−1355.

[7] Snijder, E.J., Wassenaar, A.L.M., Spaan, W.J.M. (1992). The 5′ end of the equine arteritis virus replicase gene encodes a papainlike cysteine protease. *J. Virol.* 66, 7040−7048.

[8] Snijder, E.J., Wassenaar, A.L.M., Spaan, W.J.M. (1994). Proteolytic processing of the replicase ORF1a protein of equine arteritis virus. *J. Virol.* 68, 5755−5764.

[9] Snijder, E.J., Wassenaar, A.L.M., Spaan, W.J.M., Gorbalenya, A.E. (1995). The arterivirus nsp2 protease. an unusual cysteine protease with primary structure similarities to both papain-like and chymotrypsin-like proteases. *J. Biol. Chem.* 270, 16671−16676.

[10] Snijder, E.J., Wassenaar, A.L.M., van Dinten, L.C., Spaan, W.J.M., Gorbalenya, A.E. (1996). The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases. *J. Biol. Chem.* 271, 4864−4871.

[11] den Boon, J.A., Faaberg, K.S., Meulenberg, J.J.M., Wassenaar, A.L.M., Plagemann, P.G.W., Gorbalenya, A.E., Snijder, E.J. (1995). Processing and evolution of the N-terminal region of the arterivirus replicase ORF1a protein: identification of two papainlike cysteine proteases. *J. Virol.* 69, 4500−4505.

[12] van Dinten, L.C., Wassenaar, A.L.M., Gorbalenya, A.E., Spaan, W.J.M., Snijder, E.J. (1996). Processing of the equine arteritis virus replicase ORF1b protein: identification of cleavage products containing the putative viral polymerase and helicase domains. *J. Virol.* 70, 6625−6633.

[13] van Dinten, L.C., Rensen, S., Spaan, W.J.M., Gorbalenya, A.E., Snijder, E.J. (1999). Proteolytic processing of the open reading frame 1b-encoded part of arterivirus replicase is mediated by nsp4 serine protease and is essential for virus replication. *J. Virol.* 73, 2027−2037.

[14] Wassenaar, A.L.M., Spaan, W.J.M., Gorbalenya, A.E., Snijder, E.J. (1997). Alternative proteolytic processing of the arterivirus replicase ORF1a polyprotein: evidence that NSP2 acts as a cofactor for the NSP4 serine protease. *J. Virol.* 71, 9313−9322.

[15] Fang, Y., Snijder, E.J. (2010). The PRRSV replicase: Exploring the multifunctionality of an intriguing set of nonstructural proteins. *Virus Res.* 154, 61−76.

[16] Ziebuhr, J., Snijder, E.J., Gorbalenya, A.E. (2000). Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* 81, 853−879.

[17] Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., Snijder, E.J. (2006). Nidovirales: Evolving the largest RNA virus genome. *Virus Res.* 117, 17−37.

[18] van Aken, D., Benckhuijsen, W.E., Drijfhout, J.W., Wassenaar, A.L.M., Gorbalenya, A.E., Snijder, E.J. (2006). Expression, purification, and *in vitro* activity of an arterivirus main proteinase. *Virus Res.* 120, 97−106.

[19] Tian, X., Lu, G., Gao, F., Peng, H., Feng, Y., Ma, G., Bartlam, M., Tian, K., Yan, J., Hilgenfeld, R., Gao, G.F. (2009). Structure and cleavage specificity of the chymotrypsin-like serine protease (3CLSP/nsp4) of porcine reproductive and respiratory syndrome virus (PRRSV). *J. Mol. Biol.* 392, 977−993.

[20] Xu, A.T., Zhou, Y.J., Li, G.X., Yu, H., Yan, L.P., Tong, G.Z. (2010). Characterization of the biochemical properties and identification of amino acids forming the catalytic center of 3C-like proteinase of porcine reproductive and respiratory syndrome virus. *Biotechnol. Lett.* 32, 1905−1910.

[21] van Aken, D., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Snijder, E.J. (2006). Proteolytic maturation of replicase polyprotein pp1a by the nsp4 main proteinase is essential for equine arteritis virus replication and includes internal cleavage of nsp7. *J. Gen. Virol.* 87, 3473−3482.

[22] Barrette-Ng, I.H., Ng, K.K.S., Mark, B.L., van Aken, D., Cherney, M.M., Garen, C., Kolodenko, Y., Gorbalenya, A.E., Snijder, E.J., James, M.N.G. (2002). Structure of arterivirus nsp4 − The smallest chymotrypsin-like proteinase with an alpha/beta C-terminal extension and alternate conformations of the oxyanion hole. *J. Biol. Chem.* 277, 39960−39966.

[23] van Aken, D., Snijder, E.J., Gorbalenya, A.E. (2006). Mutagenesis analysis of the nsp4 main proteinase reveals determinants of arterivirus replicase polyprotein autoprocessing. *J. Virol.* 80, 3428−3437.

[24] Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.

[25] Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. *BMC Bioinformatics* 23, 2947−2948.

[26] Gorbalenya, A.E., Lieutaud, P., Harris, M.R., Coutard, B., Canard, B., Kleywegt, G.J., Kravchenko, A.A., Samborskiy, D.V., Sidorov, I.A., Leontovich, A.M., Jones, T.A. (2010). Practical application of bioinformatics by the multidisciplinary VIZIER consortium. *Antiviral Res.* 87, 95−110.

[27] Gouet, P., Courcelle, E., Stuart, D.I., Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15, 305−308.

[28] The PyMOL Molecular Graphics System., DeLano Scientific: 2002.

Eric J. Snijder

Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, LUMC E4-P, PO Box 9600, 2300 RC Leiden, The Netherlands. Email: e.j.snijder@lumc.nl

Alexander E. Gorbalenya

Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, LUMC E4-P, PO Box 9600, 2300 RC Leiden, The Netherlands. Email: a.e.gorbalenya@lumc.nl

Handbook of Proteolytic Enzymes, 3rd Edn
ISBN: 978-0-12-382219-2

DOI: http://dx.doi.org/10.1016/B978-0-12-382219-2.00691-8