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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Structural basis for catalysis and substrate specificity of a 3C-like cysteine protease from a mosquito mesonivirus

Manuel Kanitza,b, Sandra Blanckc, Andreas Heineb, Anastasia A. Gulyaevad, Alexander E. Gorbalenya d,e,f, John Ziebuhr c,**, Wibke E. Diederich a,b,*

* Center for Tumor Biology and Immunology, Philipps University, Marburg, Germany
** Institute of Medical Virology, Justus Liebig University, Giessen, Germany
*a,b, Center for Tumor Biology and Immunology, Philipps University, Marburg, Germany
b, Institute of Pharmaceutical Chemistry, Philipps University, Marburg, Germany
c, Institute of Medical Virology, Justus Liebig University, Giessen, Germany
d, Belloserky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia
e, Institute of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia
f, Department of Medical Microbiology, Leiden University Medical Center, Leiden, the Netherlands

Abstract

Cavally virus (CavV) is a mosquito-borne plus-strand RNA virus in the family Mesoniviridae (order Nidovirales). We present X-ray structures for the CavV 3C-like protease (3CLpro), as a free enzyme and in complex with a peptide aldehyde inhibitor mimicking the P4-to-P1 residues of a natural substrate. The 3CLpro structure (refined to 1.94 Å) shows that the protein forms dimers. The monomers are comprised of N-terminal domains I and II, which adopt a chymotrypsin-like fold, and a C-terminal α-helical domain III. The catalytic Cys-His dyad is assisted by a complex network of interactions involving a water molecule that mediates polar contacts between the catalytic His and a conserved Asp located in the domain II-III junction and is suitably positioned to stabilize the developing positive charge of the catalytic His in the transition state during catalysis. The study also reveals the structural basis for the distinct P2 Asn-specific substrate-binding pocket of mesonivirus 3CLpros.

1. Introduction

Proteases with a two-β-barrel fold prototyped by chymotrypsin form one of the largest clans of proteolytic enzymes, called the PA clan, and are encoded by pro- and eukaryotes as well as many viruses (Rawlings et al., 2012). The cellular chymotrypsin-like enzymes invariably employ a canonical catalytic Ser-His-Asp triad (Hedstrom, 2002; Polgar, 2005) and have evolved, by duplication and diversification, a large spectrum of substrate specificities facilitating their diverse roles in many biological processes. Likewise, RNA viruses with single-stranded RNA genomes of positive polarity (ssRNA + viruses) often employ chymotrypsin-like enzymes to control genome expression and remodel host cell functions by targeting different proteins.

One of the major monophyletic groups of RNA virus chymotrypsin-like enzymes is known as 3C and 3C-like proteases (3C/3CLpro; see below), which evolved by extensive diversification during virus speciation. In contrast to cellular and other viral homologs, these enzymes uniquely diversified the principal catalytic nucleophile residue to employ either the canonical Ser or an unconventional Cys (depending on the virus lineage) (Bazan and Fletterick, 1988; Gorbalenya et al., 1986, 1989a; Malcolm, 1995), while (somewhat counterintuitively) sharing a conserved narrow substrate specificity. In those proteases that employ Cys as the principal nucleophile, the catalytic Asp may be replaced with another residue, either another acidic residue (Glu) (Matthews et al., 1989a; Malcom, 1995), while (somewhat counterintuitively) sharing a conserved narrow substrate specificity. In those proteases that employ Cys as the principal nucleophile, the catalytic Asp may be replaced with another residue, either another acidic residue (Glu) (Matthews et al., 1989a; Malcom, 1995), while (somewhat counterintuitively) sharing a conserved narrow substrate specificity. In those proteases that employ Cys as the principal nucleophile, the catalytic Asp may be replaced with another residue, either another acidic residue (Glu) (Matthews et al., 1989a; Malcom, 1995), while (somewhat counterintuitively) sharing a conserved narrow substrate specificity. In those proteases that employ Cys as the principal nucleophile, the catalytic Asp may be replaced with another residue, either another acidic residue (Glu) (Matthews et al., 1989a; Malcom, 1995), while (somewhat counterintuitively) sharing a conserved narrow substrate specificity. In those proteases that employ Cys as the principal nucleophile, the catalytic Asp may be replaced with another residue, either another acidic residue (Glu) (Matthews et al., 1989a; Malcom, 1995), while (somewhat counterintuitively) sharing a conserved narrow substrate specificity. In those proteases that employ Cys as the principal nucleophile, the catalytic Asp may be replaced with another residue, either another acidic residue (Glu) (Matthews et al., 1989a; Malcom, 1995), while (somewhat counterintuitively) sharing a conserved narrow substrate specificity.
residues are accommodated in a protease substrate pocket that generally includes a highly conserved His residue (which differs from the catalytic His) and another residue, often Ser or Thr, which together define the S1 subsite homologous to that of cellular chymotrypsin-like proteases (Bazan and Fletterick, 1988; Gorbalenya et al., 1989a). These residues are generally considered valuable targets for the development of broadly acting antiviral drugs (Kuo et al., 2009; Yang et al., 2005; Banerjee et al., 2018; Pillaiyar et al., 2016).

Viral 3C/3CLpro’s are produced as part of large polyproteins that contain the key replicative domains and several other proteins (for reviews see (Kräusslich and Wimmer, 1988; Ziebuhr et al., 2000; Dougherty and Semler, 1993; Ryan and Flint, 1997)). The individual proteins are released by the 3C/3CLpro in both cis- and trans-cleavage reactions, thereby regulating the formation of viral replication complexes in a timely coordinated manner. In addition, some 3C/3CLpro’s may mediate the processing of capsid polyproteins to generate the proteins required for virus particle formation as well as cleaving cellular proteins to facilitate virus reproduction. Proteases encoding 3C/3CLpros belong to two large monophyletic orders, the Picornavirales (which infect uni- and multicellular eukaryotes) and the Nidovirales (which infect vertebrates and invertebrates), but also include several other virus families or groups, such as the plant Potyviridae and Sobemovirus, and the vertebrate Caliciviridae and Astroviridae, displaying more distant affinities to one of these two virus orders. Over the past years, our appreciation of the natural 3C/3CLpro diversity and conservation of specific residues has been steadily improved by comparative genomics, while our understanding of the roles and properties of these proteases in viral replication was largely based on just a few viruses of the above taxa, with a strong focus on viruses that (may) infect humans. Specifically, the structural characterization of 3C/3CLpros remained limited to those encoded by several mammalian viruses and one avian virus of the Picornaviridae, Coronaviridae, Arteriviridae, Caliciviridae, and Astroviridae, and plant Potyviridae and Sobemovirus (Matthews et al., 1994; Banerjee et al., 2018; Phan et al., 2002; Nunn et al., 2005; Speroni et al., 2009; Gayathri et al., 2006; Damalanka et al., 2018; Weerawarna et al., 2016; Galasiti Kankanamalage et al., 2017; Takahashi et al., 2013; Kim et al., 2012; Muhaxhirli et al., 2013; Allaire et al., 1994; Anand et al., 2002, 2003; Bergmann et al., 1997; Xue et al., 2008; Yang et al., 2003; Barrette-Ng et al., 2002). All these proteases have substrate specificities that are critically determined by the P1 position.

In this study, we sought to address these knowledge gaps by determining the structure of a 3Cpro encoded by Cavally virus, an invertebrate nidovirus. The protease of this virus represents family C107, the member of the genus Alphamesonivirus in the family Mesoniviridae which, together with the families Arteriviridae, Coronaviridae, and Roniviridae as well as five others just approved by ICTV (Siddell et al., 2019), forms the order Nidovirales (Fig. 1). A distinct feature of previously characterized proteases from this virus order is the presence of an extra C-terminal domain with regulator function. This domain was shown to adopt different folds in arteri- and coronavirus enzymes (representing protease families S32 and C30, respectively) (Anand et al., 2002; Barrette-Ng et al., 2002) while there is currently no structural information on other nidovirus 3CLpro’s, such as those of roniv-, toro- and bafiniviruses, all of which representing distinct protease families in the MEROPS database (C62, S65, and S75, respectively); toro- and bafiniviruses belong to distinct subfamilies of the newly established family Tobaniviridae.

Here, we present the first crystal structures of a mesonivirus 3CLpro, both for the free enzyme and in complex with a covalently bound inhibitor. The protein structure comprises a two-β-barrel fold that is linked to a large C-terminal helical domain of > 100 residues. The structural analysis identifies critical residues involved in substrate binding and suggests that mesonivirus 3CLpros employ a catalytic system that depends on (i) Cys153 (principal nucleophile), (ii) His48 (general base), and (iii) a water molecule mediating a polar contact between (iv) Asp216 and the catalytic His48. A strikingly similar arrangement was previously reported for coronavirus 3CLpro structures, which share very low sequence identity (< 10%) with 3CLpro orthologs from other nidovirus families, further supporting the universal key roles of these residues in the respective enzymes (Fig. 2).

2. Results

2.1. Overall structure of CavV 3CLpro

The CavV 3CLpro crystal structure determined at 1.94 Å resolution shows two 3CLpro molecules in the asymmetric unit (Fig. 3A). The two molecules, named A (residues 1–314) and B (residues 1–305), form a dimer with an overall contact surface of 1728 Å², involving 4 salt bridges and 24 hydrogen bonds as calculated by PDBePISA (Krisinel and Henrick, 2007). Molecules A and B are quite similar, with a root-mean-square deviation (RMSD) of 0.59 Å (as determined for all equivalent Cα atoms).

Each molecule is comprised of three domains, called I, II, and III, respectively. Domains I and II have a chymotrypsin-like, two-β-barrel fold that is formed by seven and six β-strands, respectively (ß2-ß8 and ß9-ß13 plus ß15). The principal catalytic residues, Cys153 and His48 (Blanck et al., 2014), are part of strand ß9 and ß10 helix ß6, respectively, and are positioned in the center of the active-site cleft located between domains I and II. Domain II is connected to the C-terminal domain III, the latter being formed by seven helices (ß7 and α2 to α7) (Fig. 3A). Due to partially poor electron density, residues 51–55, 191–194, and 214–220 of chain A as well as residues 27–30 and 306–314 of chain B were omitted from the model. The C-terminal part of chain A wraps around chain B and thereby comes in close contact to the active site of chain B.

The N-terminal domains I and II of CavV 3CLpro resemble those of related enzymes from the Coronaviridae and Arteriviridae and the 3C/ 3CLpro enzymes from other RNA + viruses that lack an extra C-terminal domain (Fig. 3B) (Matthews et al., 1994; Mosimann et al., 1997; Phan et al., 2002; Speroni et al., 2009; Muhaxhirli et al., 2013; Allaire et al., 1994; Anand et al., 2002, 2003; Bergmann et al., 1997; Xue et al., 2008; Yang et al., 2003; Barrette-Ng et al., 2002; Zhao et al., 2006).

2.2. CavV 3CLpro active site

The S1 atom of the catalytic Cys153 in chain B is in H-bond distance to the carbonyl oxygen of Phe150 (3.5 Å), the carbonyl oxygen of
Fig. 1. Characteristics of nidovirus 3CLpros in a virus phylogeny and host context. Left, RdRp-based phylogeny of representatives of established nidovirus species. Right, presumed primary host of the respective virus species, 3CLpro principal nucleophile residue, and 3CLpro PDB structure availability. Information on CavV is highlighted using a gray background. Information on nidovirus genomes used for phylogeny reconstruction is provided in the Supplemental Information. Box, SH-aLRT branch support is shown using three ranges of values.
Fig. 2. Structure-based MSA of mesonivirus and coronavirus 3CL\textsuperscript{pro}s. Strictly conserved residues are indicated using red background color, partially conserved residues are indicated in red. Secondary structure elements are shown for CavV (PDB ID: 5LAC, chain B) and TGEV (PDB ID: 1LVO, chain A) in blue above the alignment. Residues that were omitted from the respective structure models are indicated in green. Residue numbers given above the alignment refer to the CavV 3CL\textsuperscript{pro} sequence corresponding to the CavV 3CL\textsuperscript{pro} structure (PDB ID: 5LAC, chain B). Crucial CavV 3CL\textsuperscript{pro} residues are indicated below the alignment: catalytic residues – by asterisks; key residues in the S1 subsite of the substrate-binding pocket – by filled circles; the Asp residue that interacts with a conserved water molecule in the active site – by a black triangle. PDB IDs used to produce the structure-based MSA are given in the Supplemental Information.
The Nδ of catalytic His48 makes also contact to the carboxylate side chain of Asp216 through water molecule 558 with distances of 2.7 Å and 2.9 Å, respectively, forming an angle of 118.9° (chain B). This water molecule establishes another H-bond to water molecule 548 (2.7 Å), which is also hydrogen-bonded to Oγ of Thr88 (3.0 Å) and the side chain oxygen of Tyr215 (2.7 Å) (Fig. 4). Furthermore, Asp216 forms a salt bridge to Arg47 with H-bond distances of 2.8 and 3.0 Å, respectively. In monomer A, the residues of the active site exhibit a higher flexibility compared to the residues of chain B, which is indicated by (i) a slightly more disordered electron density for His48, (ii) an alternative side chain conformation for Cys153 (56% conformer A, 44% conformer B), and (iii) the lack of visible Asp216.

2.3. Structure of a CavV 3CLpro/inhibitor complex

Overall structure. The structure of CavV 3CLpro in complex with Bz-YYNQ-H, a peptide aldehyde inhibitor representing the P1–P4 residues of the C-terminal CavV 3CLpro autoprocessing site, contains four rather than two molecules in the asymmetric unit. The molecules form a tetramer (a dimer of dimers) with RMSDs of 0.53 Å (for dimer A-B) and 0.65 Å (for dimer C-D), respectively. RMSD values of all four monomers are low and range between 0.33 and 0.56 Å, suggesting that the four monomers adopt the same fold. The total contact areas between monomers A and B and between C and D were calculated to be 1770 Å² and 1640 Å², respectively. Dimer A/B is stabilized by four salt bridges and 29 hydrogen bonds whereas, in the contact surface of C/D, three salt bridges and 24 hydrogen bonds are present as determined by PDBePISA (Krissinel and Henrick, 2007). With only 494 Å², the contact area between the two dimers is significantly smaller and involves three salt bridges and eight hydrogen bonds. In both monomers A and C, Arg28 was found to reach into the substrate-binding pocket of another molecule (C and A, respectively), resulting in different interactions between enzyme and ligand in those cases (see below). (In monomer C, the side chain of Arg28 was omitted due to missing electron density, but it would interact with molecule A if present.) As expected, the ligand Bz-YYNQ-H has reacted with the sulfhydryl moiety of the catalytic Cys153 and the resulting thiohemiacetal is clearly detectable in the active site with well-defined density in the 2mFo-DFc map for monomers A, B, and C (Fig. 5A). Also for monomer D, the Fo-Fc map indicates that the ligand is bound but, in this case, the electron density is relatively poor. We therefore omitted the ligand from the final model. The binding mode of the ligand to the three subsites S1, S2, and S3 is nearly identical for monomers A, B, and C, while slight variations exist among the three monomers with respect to the S4 subsite (Fig. 5B).

Binding mode. In monomers A, B, and C, the ligand is covalently...
bound to Cys153 with C-S bond lengths of 1.9 Å (chain A), 1.8 Å (chain B), and 1.8 Å (chain C), respectively. The oxyanion hole that stabilizes the noncovalent Michaelis complex (Otto and Schirmeister, 1997) and the transition state includes the amide nitrogens of Gly151 (2.9 Å to the oxygen of the thiohemiacetal) and Cys153 (3.1 Å to the oxygen of the thiohemiacetal) in molecule A. Similar distances are observed in molecule C for the oxygen atom of the thiohemiacetal (2.8 Å to Gly151 and 2.9 Å to Cys153), while the distances in molecule B are much longer (4.4 Å to Gly 151 and 4.5 Å to Cys153) due to R-configuration. The ligand is suitably positioned to facilitate the nucleophilic attack of the sulfhydryl moiety at the prochiral carbonyl carbon, resulting in the formation of the (S)-configured thiohemiacetal in molecules A and C. Overall, the binding pockets of the monomers have a similar shape, except for two flexible regions, Tyr215-Val222 and a small region around the oxyanion hole (Phe150-Cys153).

**S1 subpocket.** In monomer A, the S1 subpocket is formed by His168 and Thr148, plus Arg28 of chain C, with the main chain carboxyl oxygen of Arg forming a hydrogen bond to the Nε of the P1 Gln with a distance of 2.8 Å. As the side chain of Arg28(C) was not clearly defined in the electron density, it was not considered further.

In contrast, the side chain of Arg28 of chain A is well defined in the density of the S1 site of monomer C (Fig. 6A), which is organized in a manner similar to that of monomer A. The side chain closes the S1 site and the main chain carboxyl oxygen of Arg28 forms a hydrogen bond to the Nε of the P1 Gln (2.8 Å). The Oε of the P1 Gln establishes H-bonds to the Nε of His168 (C) at a distance of 2.7 Å and to the Oε of Thr148 (C) at a distance of 2.6 Å.

In monomer B, the ligand is bound slightly differently (Fig. 6B). The S1 subpocket is formed by His168 (B), Thr148 (B), and Val149 (B). The polar interactions between ligand and S1 subsite are very similar to those observed for monomers A and C. The binding is stabilized by H-bonds between the Oε of the P1 Gln to both the Oε of Thr148 (B; 2.4 Å) and the Nε of His168 (B; 2.8 Å). We consider it likely that the observed interactions of Arg28 (of chain C and A, respectively) with the S1 subpockets of another monomer (A and C, respectively) result from the crystal packing and thus represent crystallographic artifacts that are unlikely to occur in solution.

**S2 subpocket.** The P2 Asn residue in monomer C of the Bz-YYNQ-H ligand points toward the S2 subpocket (Fig. 6C) formed by Asp216 and Ser52. Polar contacts are established between the Nε of the ligand P2 Asn and the carbonyl oxygen of Asp216 (chain A: 2.8 Å; chain B: 2.9 Å; chain C: 2.9 Å). In monomers A and B, a second hydrogen bond is formed with the Oε of Ser52 (chain A: 2.9 Å; chain B: 3.4 Å). Because of insufficient electron density, the latter interaction could not be confirmed for monomer C.

**S3 subpocket.** The S3 pocket (Fig. 6D) is formed by Asp173, with a hydrogen bond between Oε of Asp173 and the oxygen of the P3 tyrosine (monomer A: 2.7 Å; monomer C: 2.5 Å) stabilizing the binding to the ligand. In monomer C, an additional hydrogen bond (3.1 Å) to Nε of Arg28 (A) is formed. A similar binding mode was observed for the P3 tyrosine in molecule B, albeit the side chain OH was not included in the model.

**S4 subpocket.** The S4 pocket (Fig. 6E) is formed by Leu209, Tyr215, and Val222 and flanked by a loop consisting of residues Asp216 to Asn221. The hydroxyl functionality of the ligand establishes hydrogen bonds to the carbonyl oxygen of Tyr215 (monomer A: 2.7 Å; monomer C: 2.6 Å) and to the carbonyl oxygen of Glu218 (chain A: 3.3 Å; chain C: 2.3 Å). The observed binding mode of the ligand differs in monomer B, where the residues of the loop region adopt a different conformation, resulting in a smaller S4 pocket, which, in turn, leads to a relocation of the P4 tyrosine. For the ligand in monomer B, an intramolecular hydrogen bond to the P2 Asn residue is observed (2.5 Å).

For the N-terminal benzoyl group of the ligand, different conformations were observed for each monomer, resulting in different interactions. While the ligand does not establish any H-bonds in monomer A, H-bonds are formed to the amide nitrogen of Val222 (3.6 Å) in
monomer B and to the carbonyl oxygen atom of Thr220 (2.8 Å) in monomer C.

3. Discussion

In this study, we present the first crystal structure of a 3CLpro from a mosquito-borne nidovirus representing the family Mesoniviridae in the Nidovirales, a profoundly divergent order of RNA + viruses. The structure supports and extends previous studies suggesting that nidovirus 3C-like main proteases share a three-domain domain organization involving a chymotrypsin-like fold (domains I and II) and a C-terminal domain III (Anand et al., 2002; Barrette-Ng et al., 2002; Blanck et al., 2014; Blanck and Ziebuhr, 2016; Lauber et al., 2012; Nga et al., 2011; Zirkel et al., 2011, 2013). Domain III is absent (or much smaller) in the related 3C/3CLpros of other RNA + viruses, while the N-terminal chymotrypsin-like fold (Birktoft and Blow, 1972; Matthews et al., 1967) is conserved in all these enzymes (Matthews et al., 1994; Mosimann et al., 1997; Muhaxhiri et al., 2013; Allaire et al., 1994). Despite their conserved structural organization, nidovirus 3CLpros are remarkably diverse with respect to their catalytic residues, again illustrating the profoundly divergent evolution of the various subfamilies of the order Nidovirales. Thus for example, arterivirus, torovirus and bafinivirus 3CLpros employ a canonical Ser-His-Asp triad (Barrette-Ng et al., 2002; Smits et al., 2006; Ulferts et al., 2011), while coronavirus, mesonivirus, and ronivirus 3CLpros use a Cys-His catalytic dyad that, as shown previously for coronavirus (Anand et al., 2002) and, in this study, for mesonivirus 3CLpro, is assisted by a water molecule interacting with an Asp residue located in a noncanonical position in the primary structure, C-proximal of the catalytic Cys (see below and Fig. 2).

Database searches for related structures using the DALI server (Holm and Rosenström, 2010) revealed the 3C-like cysteine proteases of the subfamily Coronavirinae as the closest structural homologs of the CavV 3CLpro (PDB ID: 5LAC, chain A). Using the HCoV-HKU1 3CLpro structure (PDB ID: 3D23, chain A) (Zhao et al., 2008), an RMSD (Cα) of 3.8 Å (sequence identity 8%; Z-score: 15.0) was calculated (239 residues of 299 from 3D23, chain A, were aligned).

**Dimerization and interfaces.** Similar to what was shown previously for coronavirus 3CLpros, the CavV 3CLpro forms a tight dimer in the crystal structure. The arrangement of the two molecules in dimer A/B and C/D, respectively, is similar to the orientation reported for the two protomers that form the coronavirus 3CLpro dimer (Anand et al., 2002, 2003; Yang et al., 2003). Each monomer of the two dimers of the
CavV 3CL\textsuperscript{pro} structure reaches into the other, with its N-terminus (residues 1–16) interacting with residues of the dimer mate, forming 6 hydrogen bonds (A2-B147, A5-B136, A6-B136, A7-B147, A10-B134 (2x)) and 1 salt bridge (A1-B296) in the case of monomer A. For the entire A/B dimer, 20 hydrogen bonds and 5 salt bridges are formed. The large contact interfaces observed for dimers A/B (1764 Å\textsuperscript{2}) and C/D (1640 Å\textsuperscript{2}) suggest that CavV 3CL\textsuperscript{pro} dimer formation is of functional relevance which remains to be confirmed in further studies, such as those performed for coronavirus 3CL\textsuperscript{pro}s, in which the role of dimerization for trans-cleavage activity was confirmed and characterized in significant detail (for a review, see (Xia and Kang, 2011)). In contrast to the A/B and C/D dimers described above, the formation of the tetramer A/B/C/D with an interface area of 494 Å\textsuperscript{2} (involving three salt bridges and eight hydrogen bonds between monomers A and C) likely represents a crystallographic artifact.

Substrate-binding pocket. The S1 pockets of CavV 3CL\textsuperscript{pro} and coronavirus 3CL\textsuperscript{pro}s are similar and can accommodate a Gln residue as shown here for a peptide aldehyde corresponding to the P1-to-P4 residues of the C-terminal CavV 3CL\textsuperscript{pro} autoprocessing site. The structure shows that, in addition to His168, a conserved Thr residue (Thr148) is located in the S1 subsite and suitably positioned to establish interactions with Gln (or Glu), both of which are common in the P1 position of a substrate (including CavV 3CL\textsuperscript{pro} substrates (Blanck and Ziebuhr, 2016). Interactions between the equivalent Thr/His residues conserved in other viral 3C/3CL\textsuperscript{pro}s, including arterivirus ns4p (Barrette-Ng et al., 2002; Snijder et al., 1996), and the side chains of P1 Glu/Gln residues have been shown previously to be critically involved in the binding of Gln and Glu residues, respectively. Interestingly, mesonivirus 3CL\textsuperscript{pro}s do not appear to have a strong preference for Gln in the P1 position of substrates (which is a typical feature in coronavirus 3CL\textsuperscript{pro}s) nor do they have a preference for Gln over Gln as shown for arterivirus ns4p enzymes (Ziebuhr et al., 2000). A possible explanation for this less pronounced specificity for Gln or Glu in mesonivirus 3CL\textsuperscript{pro}s may be that His and Thr do not establish the same type of interactions with nearby located residues that appear to be required for fine-tuning the P1 specificity for Gln and Glu, respectively (Mosimann et al., 1997; Anand et al., 2002; Bergmann et al., 1997). Thus, for example, the equivalent His residue in the S1 subsite of coronavirus 3CL\textsuperscript{pro}s was reported to establish interactions with two other residues (Phe and Tyr) (Anand et al., 2002). These latter interactions are thought to keep the His residue over a wide pH range in the neutral state required for interacting with the P1 Gln side chain. Compared to other viral 3C/3CL\textsuperscript{pro}s, mesonivirus 3CL\textsuperscript{pro}s have a less pronounced specificity toward the P1 residue and tolerate a range of different residues in this position. In addition to Gln and Glu, Lys and several other residues are found at the P1 position of predicted cleavage sites in the replicase polyprotein of CavV and several other mesonivirus 3CL\textsuperscript{pro}s (Blanck and Ziebuhr, 2016). Interestingly, a replacement of His168 with Ala in a bacterial fusion protein construct containing the CavV 3CL\textsuperscript{pro} was previously shown to have differential effects on the cleavage of the N- and C-terminal 3CL\textsuperscript{pro} autoprocessing sites (Blanck et al., 2014). While cleavage at the N-terminal autoprocessing site was retained, cleavage of the C-terminal processing site was abolished in a CavV 3CL\textsuperscript{pro} H168A mutant. As the N-terminal autoprocessing site contains a Lys residue in the P1 position, while the C-terminal autoprocessing site contains Gln in this position, it is tempting to speculate that His168 is required for the cleavage of substrates containing Gln (and, likely, Gln) in the P1 position, while this His residue may be (largely) dispensable for the cleavage of substrates containing Lys in the P1 position. To test this hypothesis and, more generally, establish the structural details of substrate binding for 3CL\textsuperscript{pro}s with other residues at the P1 (and other) position(s), additional studies of enzyme/inhibitor complexes with suitable peptidic inhibitors should be performed. Also, it may be worth testing if (and to what extent) specific P1 and other residues flanking the scissile bond affect the cleavage efficiency at specific polypeptide cleavage sites, thereby possibly contributing to the timely coordinated release of specific processing products from the viral replicase polyproteins. As indicated above, the P1 Gln of molecules A and C was found to (also) interact with the backbone carbonyl oxygen of Arg28 from a molecule of the other dimer. We however consider this a crystallographic artifact and, therefore, will not discuss this further.

Biochemical studies and comparative sequence analyses of mesonivirus replicase polyproteins (Blanck et al., 2014; Blanck and Ziebuhr, 2016) identified the P2 position as a key specificity determinant of mesonivirus 3CL\textsuperscript{pro}s. In CavV and most other members of the genus Alphamesonivirus, the P2 position of putative 3CL\textsuperscript{pro} substrates is predominantly occupied by Asn. The crystal structure of the 3CL\textsuperscript{pro}/inhibitor complex shows that the Asn side chain fits perfectly into the S2 subpocket, with its carboxamide functionality acting as hydrogen bond donor and acceptor in interactions with the main chain carbonyl oxygen of Asp216 and the O\text{y} atom of Ser252, respectively. The position of the strictly conserved Asp216 is stabilized by a salt bridge with the conserved Arg47 residue, suggesting that Asp216 has a dual functional role: besides its involvement in the coordination of the water molecule that interacts with the catalytic His48 residue, it is part of the S2 pocket, where it interacts with the carboxamide moiety of the Asn side chain. In addition to its specificity for Asn, the CavV 3CL\textsuperscript{pro} S2 pocket would be suitable (and large enough) to accommodate other residues, such as Thr, which is found at the P2 position of one (out of 12) predicted CavV 3CL\textsuperscript{pro} cleavage sites. In contrast, the P2 position of 3CL\textsuperscript{pro} cleavage sites is either predominantly occupied by a Leu residue (in coronaviruses) or varies considerably (in arteriviruses).

The direct interaction partner of the peptidic ligand in the S3 sub-pocket is Asp173, which forms a hydrogen bond to the phenolic hydroxyl of the tyrosine P3 residue. In this subpocket, also basic residues, such as Lys or Arg, as well as residues carrying a donor/acceptor function could establish a similar interaction pattern. This view is also supported by the occurrence of a range of other residues in the P3 position of CavV 3CL\textsuperscript{pro} substrates (Blanck and Ziebuhr, 2016). Compared to the S3 and S1 subsites of coronavirus 3CL\textsuperscript{pro}s, the respective subsites are more spacious in the CavV 3CL\textsuperscript{pro} (see Fig. 7).

The rather hydrophobic S4 subpocket, consisting of Ile99, Ile207, Leu209, and Val222, is able to accommodate small hydrophobic residues including Leu, which is the most common residue at the P4 position of CavV 3CL\textsuperscript{pro} substrates. However, as the loop formed by residues 216–221 covering the S4 pocket results in a semi-open pocket, larger residues with hydrogen bond donor functionalities, such as Tyr, can also be accommodated. In contrast to coronavirus 3CL\textsuperscript{pro}s, which offer only space for smaller residues, such as Ser, Val or Thr, in a fairly closed S4 pocket (Ziebuhr et al., 2000; Yang et al., 2003, 2005; Anand et al., 2002, 2003), the S4 subsite is larger in the CavV 3CL\textsuperscript{pro}.

The main difference between the different chains present in the CavV 3CL\textsuperscript{pro}/inhibitor crystal structure is that, in chain B (but not in chains A and C), a flipped Glu218 blocks the access to the S4 pocket, resulting in a deviating position of the benzoyl group which in turn enables an intramolecular hydrogen bond to the P2 substituent (Fig. 5). Also, no major differences are observed between the structures of the free enzyme and the enzyme/inhibitor complex. The superposition of 293 Ca atoms of SLAC (chain B) and SLAK (chain A) revealed an RMSD of 0.76 Å. The only major difference observed upon complex formation is a peptide flip of Phe150 enabling a hydrogen bond between the carboxyl oxygen of the inhibitor (Gln5) and the amide backbone of residue Gly151 (2.9 Å).

As reported previously, the primary structures of 3C-like cysteine proteases encoded by Coronavirinae, Roniviridae, and Mesoniviridae are poorly conserved except for two strictly conserved sequence signatures, RH and GxCG, that include the catalytic His and Cys residues (underlined) (Blanck et al., 2014; Nga et al., 2011; Zirkel et al., 2011, 2013; Ziebuhr et al., 2003). Our study suggests that this list of highly conserved residues may be extended by an Asp residue, Asp187 in the SARS-CoV-3CL\textsuperscript{pro} and Asp216 in the CavV 3CL\textsuperscript{pro}, that is shown to be conserved among the 3CL\textsuperscript{pro}s of all known coronavirus and mesonivirus
3CL\textsuperscript{pro} s in a similar position of the sequence (Fig. 2). Furthermore, previous comparative sequence analyses revealed that an equivalent Asp and Arg residues in coronavirus 3CL\textsuperscript{pro} was suggested to indicate served to form a salt bridge with Arg40. The strict conservation of the catalytic triad (typically Asp) (Anand et al., 2002). Asp186 was observed to form a salt bridge with a conserved Arg that immediately precedes the catalytic His in the primary structure (Fig. 2). Furthermore, as revealed by the CavV 3CL\textsuperscript{pro}/inhibitor structure (Fig. 6), the main chain carbonyl oxygen of this Asp residue contributes also to the S2 subsite of the substrate-binding pocket.

Taken together, the study provides interesting new insight into the molecular basis of substrate binding and catalysis of nidovirus 3CL\textsuperscript{pro}s. With less than 10% amino acid sequence identity, nidovirus 3CL\textsuperscript{pro}s are extremely diverse and include enzymes that either employ the classical catalytic triad of chymotrypsin-like serine proteases (Ser-His-Asp) as shown for members of the Arteriviridae and (former) Torovirinae (Barrette-Ng et al., 2002; Smits et al., 2006; Ulferts et al., 2011) or a noncanonical Cys-His catalytic dyad assisted by a water molecule that is oriented by a complex network of interactions involving a conserved Asp residue located in a region that connects domains II and III. Also, this work reveals the structural basis for the distinct specificities reported for coronavirus and mesonivirus 3CL\textsuperscript{pro}s, respectively.

4. Material and methods

Phylogeny reconstruction. Nidovirus phylogeny was re-constructed based on an MSA of the conserved core of the RdRp which was generated using the Viralis platform (Gorbatenya et al., 2010). Representatives of 67 species, 67 out of 88 delineated by DEmARC (Lauber and Gorbalenya, 2012) and recognized now by ICTV (Siddell et al., 2019; Adams et al., 2016; de Groot et al., 2013), were included in this analysis (Supplementary Table 1). IQ Tree 1.5.5 (Nguyen et al., 2015) with automatically selected LG + F + R7 evolutionary model was used. To estimate branch support, the Shimodaira-Hasegawa-like approximate likelihood ratio test (SH-aLRT) with 1000 replicates was conducted (Guindon et al., 2010).

Structure-based multiple sequence alignment. The structure-based MSA of coronavirus and CavV 3CL\textsuperscript{pro}s was built using the PyMOL Molecular Graphics System, version 2.2.3 (Schrödinger, LLC), command “extra_fit”, alignment method “cealign”, with only alpha carbon atoms considered. The TGEV 3CL\textsuperscript{pro} structure (PDB entry 1LVO, chain A) served as a reference to which other structures (Supplementary
Table 2) were aligned independently. For non-CavV mesoniviruses (Supplementary Table 1), the 3CL*pro MSA was downloaded from the Viralis platform (Gorbalenya et al., 2010). The two MSAs were aligned manually adjusted to improve regional sequence and structure similarity, mostly in the C-terminal domain. This refinement took into account another structure-based alignment of the same enzymes which was generated using the protein structure comparison service ProDomFold at the European Bioinformatics Institute (Krissinel and Henrick, 2004) which superimposes multiple structures simultaneously (Krissinel et al., 2005). Secondary structures were retrieved from the DSSP database (Hekkelman and Vriend, 2005; Touw et al., 2015). The resulting alignment was visualized by ESPript 2.1 (Gouet et al., 2003; Robert and Gouet, 2014).

Production and purification of wild-type CavV 3CL*pro. Wild type CavV 3CL*pro was produced as described before (Blanck and Ziebuhr, 2016) using a slightly modified purification protocol. Briefly, a fusion protein, in which the maltose-binding protein (MBP) sequence was fused to the complete 3CL*pro coding sequence (corresponding to the CavV pp1a/pp1ab residues 1387 to 1700, GenBank NC_015668), was produced in E. coli and purified by amylose affinity chromatography. Following cleavage with factor Xa, MBP and 3CL*pro (the latter containing no extra residues) were separated by anion-exchange chromatography. Pooled peak fractions containing CavV 3CL*pro were further purified by size-exclusion chromatography in buffer containing 20 mM Tris (pH 8.0), 150 mM NaCl, 0.1 mM EDTA, and 1 mM DTT and using an ÄktaPURIFIER 10 chromatography system equipped with a HiLoad 16/60 Superdex 75 column (GE Healthcare). CavV 3CL*pro eluted as a single peak after 60 mL. Following adjustment of the protein concentration to 7 mg/mL using Amicon Ultra Filters (10 kDa NMWL), the protein was immediately used for crystallization experiments.

Production and purification of SeMet-CavV 3CL*pro. The MBP-pp1a-1387-1700 coding sequence (Blanck et al., 2014) was subclonated into pET11d (Novagen) and selenomethionine-labeled MBP-3CL*pro fusion protein was produced using the methionine auxotroph E. coli strain B834(DE3). Freshly transformed cells were grown in SM medium (2x M9 salts, 0.4% glucose, amino acids except methionine [each at 40 μg/mL], vitamins [1 μg/mL], adenine, thymine, uracil, guanine [each at 200 μg/mL] and trace elements) containing carbenicillin (75 μg/mL) and methionine (40 μg/mL). At an optical density at 600 nm of 0.6, the cells were harvested by low-speed centrifugation, resuspended in pre-warmed SM medium without methionine and incubated for another 4 h at 37°C. Then, seleno-L-methionine (40 μg/mL) was added to the medium and the culture was incubated for 30 min at 37°C. Next, protein expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the culture was incubated overnight at 16°C under vigorous shaking (225 rpm). SeMet-CavV 3CL*pro was purified as described above, concentrated to 7 mg/mL and used in crystallization trials.

Crystallization. Crystallization screens (1152 different conditions) were performed at the MarXtal laboratory (Philippines University Marburg) using CavV 3CL*pro (7 mg/mL) or a mixture of CavV 3CL*pro and N-benzoyl peptide (Bz-YNYQ-H; ThinkPeptides, Oxford, UK). The sequence of the peptide was derived from the CavV 3CL*pro C-terminus and represents the P4–P3–P2–P1 residues of the C-terminal 3CL*pro autoprocessing site. For crystallization of the enzyme/inhibitor complex, the protein solution (7 mg/mL) was pre-incubated with the peptide (1.5 mM) for 2 h to facilitate the formation of the thiohemiacetal with the sulfhydryl group of the catalytic Cys153. For the free enzyme, crystals suitable for diffraction experiments were obtained using 0.2 M lithium acetate, 24% polyethylene glycol (PEG3350). Crystals grew overnight using a mixture of 0.5 μL reservoir solution and 1 μL protein solution in a 15-well hanging-drop plate (Qiagen). Before flash-cooling in liquid nitrogen, crystals were incubated for 10–15 s in reservoir solution containing 20% glycerol. SeMet-CavV 3CL*pro was crystallized using identical conditions. First crystals of the 3CL*pro/inhibitor complex were observed after 6 days at 18°C. The best crystals that were used for subsequent diffraction experiments were obtained by using 0.1 M bicine (pH 8.5) with 20% PEG6000 as the reservoir. Before flash-cooling, the crystals were incubated for 10–15 s in reservoir solution containing 20% PEG400.

X-ray diffraction. Preliminary diffraction experiments were carried out using a CuKa-X-ray source from Incoatec 4μ with a Mar345db image plate detector, which rendered diffraction up to 2.8 Å for the
CavV 3CL\textsuperscript{pro}, the CavV 3CL\textsuperscript{pro}/inhibitor complex, and the SeMet-CavV 3CL\textsuperscript{pro} crystals. High resolution data sets for the CavV 3CL\textsuperscript{pro} and SeMet-labeled CavV 3CL\textsuperscript{pro} crystals, respectively, were collected at the MX beam line 14.2 at BESSY II (Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung, Helmholtz-Zentrum Berlin für Energie und Materialien, Berlin, Germany) (Mueller et al., 2015). Measurements were carried out at 100 K and a wavelength of 0.91841 Å. Data were processed using XDS\textsuperscript{APP} (Krug et al., 2012) resulting in a CavV 3CL\textsuperscript{pro} SeMet dataset diffracting up to a resolution of 1.89 Å (see Table 1).

In order to obtain phase information, a MAD experiment using crystals of SeMet-CavV 3CL\textsuperscript{pro}, belonging to the space group P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1}, (unit cell dimensions: a = 94.2 Å, b = 111.4 Å, c = 58.2 Å) with two molecules in the asymmetric unit, was carried out. Data were collected at the peak wavelength (0.97971 Å) with a significant anomalous signal up to 2.27 Å (\(d''/\sigma(d'') = 1.083-4.647\)), at the inflection point (\(\lambda = 0.97990\) Å; 2.77 Å; \(d''/\sigma(d'') = 1.015-2.574\)), and at a high energy remote wavelength (\(\lambda = 0.91841\) Å) with a resolution up to 1.89 Å. The anomalous signal allowed us to unambiguously determine 12 out of 14 possible SeMet sites using SHELD (CC\textsubscript{max} = 58.2) (Schneider and Sheldrick, 2002), followed by phasing performed with SHELEX (Sheldrick, 2002), all implemented in HKL2MAP (Pape and Schneider, 2004). An initial electron density map was calculated and a model (581 amino acids; 10 chains; 208 amino acids in the longest chain; score 0.960) was built by ARP/wARP (Langer et al., 2008). This model was further modified using COOT (Emsley et al., 2010) and refined using PHENIX (Adams et al., 2010).

A dataset for the co-crystallized CavV 3CL\textsuperscript{pro}/inhibitor complex was collected at beam line BM30-A at the ESRF (European Synchrotron Radiation Facility, Grenoble, France). Measurements were carried out at 100 K at a wavelength of 0.979742 Å. The dataset was processed to a resolution of 2.3 Å using XDS (Kabsch, 2010). The crystals belong to space group P2\textsubscript{1} with cell dimensions a = 72.1 Å, b = 110.0 Å, c = 98.4 Å and \(\beta = 105.9^\circ\), featuring four molecules in the asymmetric unit.

Structure determination, model building, and refinement. Due to its higher resolution and superior data quality, the SeMet-CavV 3CL\textsuperscript{pro} dataset was used for further model building and refinement. The refinement of the SeMet-CavV 3CL\textsuperscript{pro} structure was carried out in PHENIX using xyz refinement, TLS refinement (from the TLSMD-Server; 12 groups), individual atomic displacement parameters (ADPs; isotropic), occupancy refinement, and finally NCS refinement (torsion angles). Water molecules were added in the automatic model building of ARP/wARP (Schiebel et al., 2016). For molecular replacement and first refinement cycles of the 3CL\textsuperscript{pro}/inhibitor complex structure, an automated refinement script obtained by using the Grade Web Server (http://grade.globalphasing.org/cgi-bin/grade/server.cgi). Average B-values were calculated using

Table 1

| Parameter | SLAC | SLAK |
|-----------|------|------|
| Wavelength (Å) | 0.91841 | 0.97990 |
| Space group | P2(1)2(1)2 | P2(1) |
| Unit-cell parameters (Å) | a = 94.1, b = 111.3, c = 58.2 | a = 72.1, b = 110.0, c = 98.4, \(\beta = 105.9^\circ\) |
| Resolution range (Å) | 50.0-0.200 (2.12-2.00) | 50.0-0.207 (2.19-2.07) |
| No. of unique reflections | 78454 (12573) | 71073 (11493) |
| Multiplicity\textsuperscript{a} | 2.4 (2.3) | 2.4 (2.3) |
| Rsym (%)\textsuperscript{a} | 5.3 (27.5) | 5.7 (30.0) |
| CC (1/2) (%) | 99.7 (86.2) | 99.7 (83.1) |
| Completeness (%)\textsuperscript{a} | 98.1 (97.5) | 98.5 (98.8) |
| (I/\sigma(I)) | 14.0 (3.7) | 13.6 (3.5) |

\textsuperscript{a} Values in brackets are statistics for the highest resolution shell.

Table 2

| Data collection, refinement, and model statistics. |
| Parameter | SLAC | SLAK |
|-----------|------|------|
| Wavelength (Å) | 0.91841 | 0.97974 |
| Space group | P2(1)2(1)2 | P2(1) |
| Unit-cell parameters (Å) | a = 94.1, b = 111.3, c = 58.2 | a = 72.1, b = 110.0, c = 98.4, \(\beta = 105.9^\circ\) |
| Resolution range (Å) | 50.0-0.200 (2.12-2.00) | 50.0-0.207 (2.19-2.07) |
| No. of unique reflections | 78454 (12573) | 71073 (11493) |
| Multiplicity\textsuperscript{a} | 2.4 (2.3) | 2.4 (2.3) |
| Rsym (%)\textsuperscript{a} | 5.3 (27.5) | 5.7 (30.0) |
| CC (1/2) (%) | 99.7 (86.2) | 99.7 (83.1) |
| Completeness (%)\textsuperscript{a} | 98.1 (97.5) | 98.5 (98.8) |
| (I/\sigma(I)) | 14.0 (3.7) | 13.6 (3.5) |

\(\text{a Values in brackets are statistics for the highest resolution shell.}

\text{b} R_{sym} = \frac{\sum|I| - \sum|F_o|}{\sum|F_o|}

\text{c} CC = 100 \times \frac{|I|}{|\sigma(I)|}

\text{d} R_{sym} = \frac{\sum|I| - \sum|F_o|}{\sum|F_o|}

\text{e} CC_{max}\text{ is calculated in the same manner as } R_{sym}\text{, but on 5% of the data that were not used for refinement.}

5% data fraction was used. COOT was used to add water molecules and for the fitting of amino acid side chains using \(\sigma\)-weighted 2Fo-Fc and Fo-Fc difference electron density maps. Restraints for ligands were obtained by using the Grade Web Server (http://grade.globalphasing.org/cgi-bin/grade/server.cgi). Average B-values were calculated using

\text{Armed water molecules and}

\text{for the fitting of amino acid side chains using } \sigma\text{-weighted 2Fo-Fc and Fo-Fc difference electron density maps. Restraints for ligands were obtained by using the Grade Web Server (http://grade.globalphasing.org/cgi-bin/grade/server.cgi). Average B-values were calculated using.
MOLEMAN (Kleywegt et al., 2006) and Ramachandran plot statistics were analyzed with PROCHECK (Laskowski et al., 1993) (see Table 2).

Coordinate accession numbers

Coordinates and structure factors for the CaV 3CLpro and the CaV 3CLpro inhibitor complex have been deposited in the Protein Data Bank under the PDB IDs 5LAC and 5LAK, respectively.

Declarations of interest

None.

Acknowledgments

The work was supported by the Deutsche Forschungsgemeinschaft (SFB1021 A01, to JZ). The contribution of AAG and AEG was funded in part by the EU Horizon2020 EVAG 653316 project and the LUMC MoBiLe program (to AEG). We are very grateful to Prof. Dr. Wolfgang Buckel, Head of the Laboratory of Microbial Biochemistry and Max Planck Fellow of the MPI for Terrestrial Microbiology, Marburg, for helpful discussions regarding mechanistic aspects. We thank the beamline staff of BESSY II (Helmholtz-Zentrum Berlin) and ESRF for their outstanding help and support during data collection and acknowledge the generous support by travel grants from the Helmholtz-Zentrum für Materialien und Energie, Berlin. We also thank Ralf Pöscke (MaxPlanck, Philips University Marburg) and Karin Schultheiß (Medical Virology, Justus Liebig University Giessen) for excellent technical assistance, and Igor Sidorov and Dmitry Samborskiy for assistance with the Viralis platform.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2019.05.001.

References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, W.L., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213–221.

Adams, M.J., Lefkowitz, E.J., King, A.M., Harrach, B., Harrison, R.L., Knowles, N.J., Kropinski, A.M., Kuhn, J.H., Mushegian, A.R., Nibert, M., Kuo, C.J., Liu, H.G., Lo, Y.K., Seong, C.M., Lee, K.I., Jung, Y.S., Liang, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213–221.

Anand, K., Palm, G.J., Mesters, J.R., Siddell, S.G., Ziebuhr, J., Hilgenfeld, R., 2002. Characterization of an alphanemovirus 3C-like protease defines a special group of nidoviruses main proteases. J. Virol. 76, 11747–11758.

Balogh, E.J., Weiss, M.S., Heinemann, U., Mueller, U., 2012. Structure-based design, synthesis and evaluation of oxazolidinone-based inhibitors of norovirus 3CL protease. Eur. J. Med. Chem. 143, 881–890.

Bocker, S., An, S., Zilke, P., Hilgenfeld, R., Marker, B., 2014. Characterization of an alphanemovirus 3C-like protease defines a special group of nidoviruses main proteases. J. Virol. 76, 11747–11758.

Brockman, E.J., Ziebuhr, J., 2013. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. J. Virol. 87, 7790–7792.

Elmer, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of COOT. Acta Crystallogr D Biol Crystallogr 66, 486–501.

Falk, M., Galiano, M., Gorbalenya, A.E., Memish, Z.A., Samborskiy, A.M., Jones, A.E., 2018. Toward development of generic inhibitors of coronavirus and picornavirus main proteases. J. Med. Virol. 90, 1749–1758.

Gayathri, P., Satheshkumar, P.S., Prasad, K., Nair, S., Savithri, H.S., Murthy, M.R., 2006. Crystal structure of the nsp3 proteinase domain of Semeia virus: rationalization of the PA200 pocket. J. Mol. Biol. 357, 1353–1362.

Kim, Y., Lovell, S., Tiew, K.C., Mandadapu, S.R., Alliston, K.R., Battaile, K.P., Groutas, E.M., 2004. MUSCLE: a fast and accurate multiple sequence alignment method. J. Mol. Biol. 325, 1383–1390.

Krissinel, E., Henrick, K., 2005. Multiple alignment of protein structures in three dimensions. Acta Crystallogr D Biol Crystallogr 61, 2256–2268.

Kuo, C.J., Liu, H.G., Lo, Y.K., Seong, C.M., Lee, K.I., Jung, Y.S., Liang, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213–221.

Langer, G., Cohen, S.X., Langan, J.S., Perras, A.K., 2018. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. 3, 1171–1179.

M. Kanitz et al.  
Virology 533 (2019) 21–33  
32
