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.
Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein

Jian Leia, Yuri Kusova, Rolf Hilgenfelda,b,

a Institute of Biochemistry, Center for Structural and Cell Biology in Medicine, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany
b German Center for Infection Research (DZIF), Hamburg – Lübeck – Borstel – Riemser Säue, University of Lübeck, Germany

ARTICLE INFO

Keywords:
Ubiquitin-like domain
Papain-like protease
Macrodomain
Nucleic-acid binding domain
Innate immunity
Structural biology

ABSTRACT

The multi-domain non-structural protein 3 (Nsp3) is the largest protein encoded by the coronavirus (CoV) genome, with an average molecular mass of about 200 kD. Nsp3 is an essential component of the replication/transcription complex. It comprises various domains, the organization of which differs between CoV genera, due to duplication or absence of some domains. However, eight domains of Nsp3 exist in all known CoVs: the ubiquitin-like domain 1 (Ubl1), the Glu-rich acidic domain (also called “hypervariable region”), a macrodomain (also named “X domain”), the ubiquitin-like domain 2 (Ubl2), the papain-like protease 2 (PL2pro), the Nsp3 ectodomain (3Ecto, also called “zinc-finger domain”), as well as the domains Y1 and CoV-Y of unknown functions. In addition, the two transmembrane regions, TM1 and TM2, exist in all CoVs. The three-dimensional structures of domains in the N-terminal two thirds of Nsp3 have been investigated by X-ray crystallography and/or nuclear magnetic resonance (NMR) spectroscopy since the outbreaks of Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) in 2003 as well as Middle-East Respiratory Syndrome coronavirus (MERS-CoV) in 2012. In this review, the structures and functions of these domains of Nsp3 are discussed in depth.

1. Introduction

This review of published research on the coronavirus non-structural protein 3 (Nsp3) forms part of a series in Antiviral Research on “From SARS to MERS: research on highly pathogenic human coronaviruses” (Hilgenfeld and Peiris, 2013). Two excellent earlier papers dealt with aspects of Nsp3. The first described the state of knowledge of the papain-like protease (PLpro) (Báez-Santos et al., 2015), while the second adopted a bioinformatics viewpoint when describing Nsp3 and other non-structural proteins involved in anchoring the coronavirus replication/transcription complex (RTC) to modified membranous structures originating from the endoplasmic reticulum (ER) (Neuman, 2016). We build on these fine reviews, focusing on recent results and discussing the structures and functions of the individual Nsp3 domains in sequential order.

Coronavirus (CoV) is a member of the subfamily Coronavirinae within the family Coronaviridae of the order Nidovirales. It is the enveloped positive-sense single-stranded RNA (+ssRNA) virus with the largest genome of all known RNA viruses thus far (Brian and Baric, 2005; Gorbalenya et al., 2006). The genomes of different CoVs comprise between 26 and 32 kilobases; however, the overall organization of the genomes is similar. The 5'-terminal two thirds of the genome include two open reading frames (ORFs), 1a and 1b, that together encode all non-structural proteins for the formation of the RTC, whereas the 3'-proximal third encodes the structural and accessory proteins (Fig. 1A; Brian and Baric, 2005). ORF1a encodes polyprotein (pp) 1a containing...
Nsp1-11, while ORF1a and ORF1b together produce pp1ab containing Nsp1-16 through a (−1) ribosomal frameshift overreading the stop codon of ORF1a (Fig. 1A; Brierley et al., 1989). Coronaviruses are divided into four genera: Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus (Adams and Cartensen, 2012). CoVs can infect many species (Fehr and Perlman, 2015); currently, the coronaviruses infecting humans are all from the genera alphacoronavirus or betacoronavirus. HCoV 229E and HCoV NL63 belong to the former (Tyrrell and Oldstone, 1965; van der Hoek et al., 2004), whereas HCoV OC43, HKU1, SARS-CoV, and MERS-CoV belong to the latter genus (Hamre and Bystøl, 2016). The domain organization of Nsp3 from HCoV NL63 as a representative of alphacoronaviruses, and of SARS-CoV in clade B of the genus betacoronavirus, is zoomed out. The question mark within HCoV-NL63 Nsp3 indicates a region of unknown function and structure. (B) Summary of the functions and domain organization of SARS-CoV Nsp3. Nsp3 is bound to double-membrane vesicles recruited from the endoplasmic reticulum (ER) membrane. The protein passes through this membrane twice, via the two transmembrane regions TM1 and TM2. AH1 is possibly an amphipathic helix attached to the ER membrane, next to TM2. Except for the 3Ecto domain, all other Nsp3 domains are located in the cytosol. All domains with known three-dimensional structures are indicated in light green (X-ray structures) or orange (NMR structures), whereas parts with unknown structure are in red. The best characterized functions of each domain of Nsp3 are shown.:*; glycosylation sites in the 3Ecto domain (Asn1431 and Asn1434; Harcourt et al., 2004).
Table 1

| Domain/region | Res. no. /MW\(^{\#}\) | Method | Coronavirus | Reference |
|---------------|-------------------------|--------|-------------|-----------|
| Ubl1          | 1-112/12.6              | NMR    | SARS-CoV    | Serrano et al. (2007) |
| Acidic domain (HVR) | 113-183/8.3     | n. d.  | SARS-CoV    | Serrano et al. (2007) |
| PL1\(^{\#2}\) | n. a./23.6              | X-ray  | SARS-CoV    | Wojdyla et al. (2010) |
| Mac1 (X domain) | 184-365/19.5         | X-ray  | SARS-CoV    | Saitkandu et al. (2005) |
|                |                       | X-ray  | SARS-CoV    | Egloff et al. (2006) |
|                |                       | X-ray  | HCoV-299E, IBV | Xu et al. (2009) |
|                |                       | X-ray  | HCoV-299E, IBV | Piotrowski et al. (2009) |
|                |                       | X-ray  | FG\(\beta\) | Wojdyla et al. (2009) |
| Mac2 (SUD-N)   | 389-524/15.2           | X-ray  | SARS-CoV    | Tan et al. (2009) |
| Mac3 (SUD-M)   | 525-652/14.0           | NMR    | SARS-CoV    | Saikatendu et al. (2005) |
|                |                       | X-ray  | SARS-CoV    | Egloff et al. (2006) |
|                |                       | X-ray  | HCoV-229E, IBV | Xu et al. (2009) |
|                |                       | X-ray  | HCoV-229E, IBV | Piotrowski et al. (2009) |
|                |                       | X-ray  | F\(\alpha\) | Wojdyla et al. (2009) |
| DPUP (SUD-C)   | 653-720/7.8            | X-ray  | SARS-CoV    | Saikatendu et al. (2005) |
|                |                       | X-ray  | HCoV-229E, IBV | Xu et al. (2009) |
|                |                       | X-ray  | HCoV-229E, IBV | Piotrowski et al. (2009) |
|                |                       | X-ray  | FCoV | Wojdyla et al. (2009) |
|                |                       | X-ray  | MERS-CoV    | Cho et al. (2016) |
| Ubl2−PL2\(^{\#5}\) | 723-1036/35.2         | X-ray  | SARS-CoV    | Ratia et al. (2006) |
|                |                       | X-ray  | SARS-CoV + human Ub | Chou et al. (2014) |
|                |                       | X-ray  | SARS-CoV + dUb | Béhéès et al. (2016) |
|                |                       | X-ray  | SARS-CoV + hISG15\(^{\#5}\) | Dazzkowski et al. (2017a) |
|                |                       | X-ray  | SARS-CoV + mISG15\(^{\#6}\) | Dazzkowski et al. (2017a) |
|                |                       | X-ray  | MERS-CoV    | Lei et al. (2014) |
|                |                       | X-ray  | MERS-CoV + human Ub | Lee et al. (2015) |
| Mac2 (SUD-N)   | 389-524/15.2           | X-ray  | SARS-CoV    | Tan et al. (2009) |
| Mac3 (SUD-M)   | 525-652/14.0           | NMR    | SARS-CoV    | Saikatendu et al. (2005) |
|                |                       | X-ray  | SARS-CoV    | Egloff et al. (2006) |
|                |                       | X-ray  | HCoV-229E, IBV | Xu et al. (2009) |
|                |                       | X-ray  | HCoV-229E, IBV | Piotrowski et al. (2009) |
|                |                       | X-ray  | F\(\alpha\) | Wojdyla et al. (2009) |
| NAB            | 1066-1180/13.0         | NMR    | SARS-CoV    | Ratia et al. (2006) |
| βSM (G2M)      | 1203-1318/12.5         | n. d.  | SARS-CoV    | Chou et al. (2014) |
| TM1            | 1391-1413/2.4          | n. d.  | SARS-CoV + human Ub | Ratia et al. (2014) |
| TM2            | 1496-1518/2.7          | n. d.  | SARS-CoV + human Ub | Béhéès et al. (2016) |
| AH1            | 1523-1545/2.7          | n. d.  | SARS-CoV + human Ub | Béhéès et al. (2016) |
| Y1 + CoV-Y     | 1546-1922/41.9         | n. d.  | SARS-CoV    | Serrano et al. (2007) |

\#: Nsp3 of the SARS-CoV strain TOR2 (Genbank: AY274119.3); %: molecular mass (kD); n. d.: structure is not determined; \(^{\#1}\): absent in SARS-CoV; n. a.: does not apply (residue numbers are only given for SARS-CoV); \(^{\#2}\): Mac2−Mac3 structure; \(^{\#3}\): Mac3−DPUP structure; \(^{\#4}\): DPUP−Ubl2−PL2\(^{\#5}\)−structure; \(^{\#5}\): Ubl2−PL2\(^{\#5}\)−C-terminal Ubl domain of human ISG15 structure; \(^{\#6}\): Ubl2−PL2\(^{\#5}\)−C-terminal Ubl domain of mouse ISG15 structure; \(^{\#7}\): regions predicted by TMHMM server v. 2.0 (Krogh et al., 2001). TM1 and TM2 are transmembrane regions while AH1 is not (Oostra et al., 2008).

---

Fig. 2. Structures (in cartoon view) of the ubiquitin-like domain 1 (Ubl1) and Ubl2 in SARS-CoV, Ubl1 in MHV, as well as their structural homologues. (A) Ubl1 (residues 20–108) of SARS-CoV (PDB entry: 2IDY; Serrano et al., 2007). (B) Ubl1 (19–114) of MHV (PDB entry: 2MOA; Keane and Giedroc, 2013). (C) Ubl2 (residues 1–60) of SARS-CoV (PDB entry: 2FA8; Ratia et al., 2006). (D) Human ubiquitin (PDB entry: 1UBQ; Vijay-Kumar et al., 1987). (E) Ras-interacting domain of RalGDS (PDB entry: 1LFD; Huang et al., 1998). (F) The N-terminal ubiquitin-like domain of the Ras-interacting domain of RalGDS (PDB entry: 1LFD; Huang et al., 1998). The N and C termini of all structures are marked. All α and 3\(\eta\) (\(\eta\)) helices are labeled and shown in cyan. β strands are in purple and loops are in brown. This figure and Figs. 3 and 5, as well as 6 and 8 were generated by using Chimera (Pettersen et al., 2004).
CoV Ubl1 are replaced by one long continuous helix (α3) in MHV Ubl1 (Fig. 2A and B).

The known functional roles of Ubl1 in CoVs are related to single-stranded (ssRNA) binding and interacting with the nucleocapsid (N) protein (Fig. 1B; Serrano et al., 2007; Hurst et al., 2010, 2013). The Ubl1 of SARS-CoV binds ssRNA containing AUA patterns. Surprisingly, many negatively charged regions (such as the 310 helix, α1) show obvious conformational changes in the NMR spectra when RNA is added to the protein solution (Serrano et al., 2007), indicating that RNA binding has long-range effects on the protein conformation. In view of the presence of several AUA repeats in the 5′-untranslated region (UTR) of the SARS-CoV genome, the Ubl1 likely binds to this region.

In MHV, the Ubl1 domain efficiently binds the cognate nucleocapsid (N) protein; thus it seems to be important for virus replication as well as initiation of viral infection. There is a critical relationship between Nsp3 interaction with the N protein and infectivity, as this interaction serves to tether the viral genome to the newly translated RTC at an early stage of coronavirus infection (Hurst et al., 2010, 2013). Deletion of the Ubl1 core (residues 19–111) of MHV abrogates viral replication (Hurst et al., 2013). The major interface regions of the complex Ubl1 − N involve acidic residues of Ubl1 helix α2 and the serine- and arginine-rich region (SR-rich region) of the N protein, as shown by NMR titration experiments (Keane and Giedroc, 2013). However, the acidic residues in helix α2 are not absolutely conserved among different CoVs, implying that the details of the interactions between Ubl1 and N protein will not be the same. In addition, the binding affinity between the bovine coronavirus (BCoV) N (residues 57–216) and MHV Ubl1 is about 260-fold lower compared to MHV N (residues 60–219) and its cognate Ubl1 (Keane and Giedroc, 2013). A structure of the Ubl1 − N complex would help understand why non-cognate Ubl1 and N protein bind weakly to each other. Thus far, only a computer docking model of the MHV Ubl1 − N complex was reported (Tatar and Tok, 2016). This model proposes that residues of β1, α1, the loop between β1 and α1, β3, and β4 of MHV Ubl1 interact with the N-terminal domain (NTD) as well as the SR-rich region of the N protein. Differently from what was suggested above, most acidic residues of Ubl1 helix α2 do not interact with the SR-rich region of N in the docking model (Tatar and Tok, 2016).

The interaction between the N protein and nucleic acid is essential for CoV genome transcription (Chang et al., 2014). The NTD plus the SR-rich region (residues 60–219) of MHV N play an important role in interacting with transcriptional regulatory sequence (TRS) RNA (Grossehme et al., 2009). The N − TRS RNA complex prevents the formation of the Ubl1 − N complex (Keane and Giedroc, 2013). The competition between N protein binding to either the TRS or the Ubl1 might be connected to the switch between viral transcription and replication. It has been shown that the SR region of N protein can be phosphorylated (Peng et al., 2008). Each of two phosphomimetic substitutions of serine residues predicted to be phosphorylated (S207D and S218D) in the SR region of MHV N decreases the binding affinity to Ubl1 by about 3-fold, compared to wild-type N (Keane and Giedroc, 2013).

The overall structure of the SARS-CoV Ubl1 domain is similar to human ubiquitin (Ub) and that of each of the two ubiquitin-like domains of human or mouse interferon-stimulated gene 15 (hISG15 or mISG15) (Fig. 2D and E; Vijay-Kumar et al., 1987; Narasimhan et al., 2005; Daczkowski et al., 2017a). In human Ub as well as in the ISG15s, only a short 310 helix is found at the position of α1-α3 or α3 in Ubl1 of SARS-CoV or MHV (Fig. 2D and E). Ub and ISG15 are important for innate antiviral immunity (Heaton et al., 2016; Morales and Lenschow, 2013); therefore, viruses tend to not only inhibit the conjugation of Ub or ISG15 to targets but also remove Ub or ISG15 from ubiquitinated or ISGylated proteins, respectively (Yuan and Krug, 2001; Bakshi et al., 2013; Yang et al., 2014). Thus, in CoVs, one or two papain-like protease (PLpro) domain(s) within Nsp3 possess deubiquitinating (DUB) and delISGylating activities (see below; for a recent review on the role of viral proteases in counteracting the host-cell’s innate immune system, see Lei and Hilgenfeld (2017)). Interestingly, two ubiquitin-like domains (Ubl1 and Ubl2) exist in all CoVs (see below; Neuman, 2016). Considering that ubiquitin-like modules are often involved in protein–protein interactions to regulate various biological processes (Hochstrasser, 2009), such as the MHV Ubl1 − N interaction mentioned above, novel possible function of Ub-like domains in CoVs might be the interaction with target proteins of Ub (or ISG15) by mimicking the shape of these two molecules. The purpose of such mimicry could be to somehow interfere with pathways involving ubiquitinated or ISGylated host targets, thereby leading to disruption of host anti-viral signal transduction or protein degradation.

The Ubl1 of SARS-CoV is also similar to the Ras-interacting domain (RID) of RalGDS (Ral guanine nucleotide dissociation stimulator; Fig. 2F; Serrano et al., 2007). Ras regulates cell-cycle progression via binding to the RID of Ras-interacting proteins (Hofer et al., 1994; Huang et al., 1998; Coleman et al., 2004). By mimicking the RID, the Ubl1 might interrupt the interactions between Ras and its effectors, thus affecting the cell cycle to support virus replication. In agreement with this, it is known that both MHV and SARS-CoV induce cell-cycle arrest in the G0/G1 phase (Chen and Makino, 2004; Yuan et al., 2005).

Following the Ubl1, the second subdomain of Nsp3a in CoVs is the Gln-rich acidic region. It comprises residues 113–183 of SARS-CoV Nsp3, with more than 35% Gln and 10% Asp (Serrano et al., 2007). Because of the non-conserved amino-acid sequence, this region is also designated as “hypervariable region (HVR)” (Neuman, 2016). The HVR region is intrinsically disordered in SARS-CoV and in MHV (Serrano et al., 2007; Keane and Giedroc, 2013) and does not affect the conformation of the globular Ubl1 domain in SARS-CoV (Serrano et al., 2007). Currently, the function of HVR in CoVs is unknown. Gln/Asp-rich proteins are often involved in many biological roles, such as DNA/RNA mimicry, metal-ion binding, and protein–protein interactions (Chou and Wang, 2015). The Ubl1 + HVR region has been demonstrated via a yeast-two-hybrid (Y2H) assay to interact with SARS-CoV Nsp6, whereas a GST pull-down study identified Nsp8, Nsp9, and NAB–JSM–TM1 of Nsp3 (NAB: nucleic-acid binding domain; JSM: betacoronavirus-specific marker; TM1: transmembrane region 1; see below) as binding partners (Imbert et al., 2008). Does the HVR play any role in these protein–protein interactions? This question is yet to be answered. Furthermore, the acidic region is dispensable for MHV replication (Hurst et al., 2013). On the other hand, this region does exist in all CoVs. It is conceivable that it may have regulatory rather than essential roles in the coronavirus replication process. However, the exact role(s) of the acidic region in CoVs should be further investigated.

3. Papain-like protease 1 domain

The papain-like protease domain(s) is/are responsible for releasing Nsp1, Nsp2, and Nsp3 from the N-terminal region of polyproteins 1a/1ab in CoVs (Harcourt et al., 2004; Barretto et al., 2005). The papain-like protease 1 domain (PL1pro) follows the HVR region (see Fig. 1A) in the alpha-CoVs and in clade A of beta-CoVs (Graham and Denison, 2006; Ziebuhr et al., 2001; Chen et al., 2007; Wojdyla et al., 2010; Neuman, 2016). Interestingly, the PL1pro is not complete in the gamma-CoV infectious bronchitis virus (IBV; Ziebuhr et al., 2001) and in Hipposideros pratti bat CoV, a virus relating to clade B of the beta-CoVs (Genbank code NC_025217.1; Neuman, 2016). In these latter viruses, some parts (such as the zinc-finger motif; see below) and the residues of the catalytic triad of the PL1pros are missing. Furthermore, the PL1pro is totally absent in beta-CoV clades B, C, and D as well as in delta-CoVs. Both the two highly human-pathogenic SARS-CoV (Fig. 1A) and MERS-CoV thus do not have a PL1pro domain; they only possess the other papain-like protease, the PL2pro domain that is conserved in all coronaviruses (see below). It is still not clear why certain CoVs encode two PLpros.

Thus far, only one structure of a PL1pro domain has been determined, that from the alpha-CoV Transmissible Gastroenteritis Virus
As mentioned above, for reasons unknown so far, many CoVs contain two PLpro's. Both PL1pro and PL2pro are involved in releasing Nsp1, Nsp2, and Nsp3 in these CoVs. However, the two PLpro's in different CoVs show varying substrate specificity. The PL1pro of MHV cleaves Nsp1[2 and Nsp2][3, while the PL2pro cleaves Nsp3[1 (Table 2; Bonilla et al., 1997; Kanjanahalueeth and Baker, 2000). Human coronavirus NL63 (HCoV-NL63) PL1pro processes Nsp1[2 while the PL2pro processes the other two cleavage sites, Nsp2[3 and Nsp3[4 (Table 2; Chen et al., 2007). Both PL1pro and PL2pro of HCoV 229E can cleave Nsp1[2 and Nsp2[3 (Table 2); however, the PL1pro is more efficient in cleaving Nsp1[2 while the PL2pro is more efficient with respect to the latter site (Ziebuhr et al., 2007). Some viruses, such as SARS-CoV, MERS-CoV, and IBV, comprise only one functional PL2pro to process all three cleavage sites (Table 2). The residues (P5–P2') of the three cleavage sites are diversified in MHV, HCoV NL63, and HCoV 229E, although the P1 is conserved as a small residue (Gly or Ala) (Table 2). In contrast, the P1 and P2 residues (Gly–Gly or Ala–Gly) are absolutely identical in all the cleavage sites of SARS-CoV, MERS-CoV, and IBV; furthermore, the P4–P1 residues are to a certain extent conserved in each of these three viruses (Table 2). Therefore, the presence of two PLpro's with slightly different substrate specificity in some CoVs may be required to cleave native substrates that deviate from the uniform ones processed by SARS-CoV, MERS-CoV, or IBV PLpro's. Unfortunately, studies on the details of recognition of different substrates by PL1pro and PL2pro are hampered by the fact that no crystal structures of the two enzymes from the same virus are available.

4. Macrodinains and the "Domain Preceding Ub12 and PL2pro (DPUP)"

4.1. Macrodinain 1 (Mac1, X domain)

A conserved macrodomain (also called "X domain", Nsp3b) follows the HVR or the PL1pro domain in all coronaviruses (Fig. 1A; Gorbalenya et al., 1991; Neuman et al., 2008; Neuman, 2016). Macrodomains widely exist in bacteria, archaea, and eukaryotes (Han et al., 2011). In addition, these conserved domains are also present in several positive-sense ssRNA (+ ssRNA) viruses of the families Hepeviridae, Togaviridae, and Coronaviridae, such as hepatitis E virus (HEV), alphavirus, rubella virus, and all coronaviruses (Koonin et al., 1992; Snijder et al., 2003). Our group has shown that the X domain (Mac1) is dispensable for RNA replication in the context of a SARS-CoV replicon (Kusov et al., 2015). Recently, evidence accumulated showing that the X domain plays a role in counteracting the host innate immune response (Erikkson et al., 2008; Kuri et al., 2011; Fehr et al., 2015, 2016).

The first crystal structure of an Nsp3 domain of any coronavirus was the unliganded X domain of SARS-CoV (Table 1; Saikatenda et al., 2005). A little later, the structure of the SARS-CoV X domain in complex with ADP-ribose (ADPr) was determined (Table 1; Egloff et al., 2006). Subsequently, structures of the unliganded X domain and/or its complex with ADPPr from HCoV 229E, IBV, HCoV NL63, Feline CoV (FCoV), and MERS-CoV were reported (Table 1; Piotrowski et al., 2009; Xu et al., 2009; Woydyla et al., 2009; Cho et al., 2016). All structures show that the X domain adopts a conserved three-layered α/β/α sandwich fold (Fig. 4). The domain with this fold is called a macrodomain because of its similarity to the extra domain in the MacroH2A variant of human histone 2A (Pehrson and Fried, 1992; Saikatenda et al., 2005). Typically, the X domain includes a central α helix (P1) followed by a central β sheet with seven β strands in the order (P2) β1–β2–β3–β6–β3–β5–β4, with β1 and β4 being antiparallel to the rest (Fig. 4). Only the X domain of IBV is an exception, since it lacks the first strand, β1 (Piotrowski et al., 2009; Xu et al., 2009). Six helices are located on the two sides of this β sheet, with helices α1, α2, and α3 on one side and α4, α5, and α6 on the other (Fig. 4).

One function of the conserved macrodomain is the binding of ADP-ribose or poly(ADP-ribose) (Han et al., 2011). The binding...
Table 2

| Nsp1|2 | Nsp2|3 | Nsp3|4 | Reference |
|-----|---|-----|---|-------|------------------|
| TGEV | RTGRG|AI | KNMGG|GD | PKSGS|GF | Putics et al. (2006) |
| HCoV NL63 | GHAGA|SV | TKLAG|BK | AKQGA|GF | Chen et al. (2007) |
| HCoV 229E | KGKG|NV | TKAAG|GK | AKQGA|GG | Ziebuhr et al. (2007) |
| MIV | KGRRG|K | RFGCA|GK | SLKGG|AY | Bonilla et al. (1997) |
| SARS-CoV* | ELLGG|AV | RLKGG|AP | SLLGG|KI | Kanjahanalathai and Baker (2000) |
| RSARS-CoV* | LIKGG|DV | RLKGG|AP | KIVGG|AP | Harcourt et al. (2004) |
| IBV* | / | VCKAG|GK | EKKAG|G | Yang et al. (2014) |
| / | / | PLZGG|G | PLZGG|G | Lim et al. (2000) |

: cleavage site; n. d., not determined; *: absence of PL2pro; #: partial presence of PL1pro; /: absence of the cleavage site.

Characteristics of the cleavage sites of PL1pro and PL2pro in CoVs and the P5-P2 residues for each cleavage site.

![Figure 4](http://www.pymol.org/)

**Figure 4.** Structure of the MERS-CoV macrodomain I (Mac1, X domain) in complex with ADP-ribose (ADPr) (PDB entry: 5HOL). The protein features an α/β/α sandwich fold. The central β sheet with the strand order β1 – β2 – β3 – β6 – β3 – β5 – β4 is shown in purple, β1 and β4 are labeled. An F – F omit difference map of ADPr is shown in black (contoured at 4.0 σ). The ADPr itself is displayed as brown sticks. The five regions (blue) relating to ADPr binding are marked by Roman numbers I – V. Fixing the two ends of the ADPr, Asp21 and Asn39 are displayed by thicker red sticks. The O2′ of ADPr forms a hydrogen bond with a water molecule (H2O 308; green sphere) being stabilized by the side-chain of Asn155 (region VI). The two phosphate groups accept a total of four hydrogen bonds from Ile48 (region III) and Gly129, Ile130 as well as Phe131 (region IV). The distal ribose is in contact with regions II and III; The O1′ and O2′ of this ribose form hydrogen bonds with the amides of Gly47 and Gly45 (region III), respectively. The O3′ forms a hydrogen bond with the side-chain amide of Asn39 (region II). Thus, Asp21 and Asn39 appear to fix the two ends of the ADPr-ribose, thereby stabilizing its binding to the cleft (Fig. 4). Surprisingly, the orientation of the corresponding Asp in the HCoV-229E X domain is different; this Asp does not directly bind ADP-ribose but is in contact with its neighboring residue Thr-22, and not with the N6 atom of adenine (Piotrowski et al., 2009; Xu et al., 2009). This difference could explain why the binding affinity between the X domain of HCoV-229E and ADPr is about 10-fold lower than that of the MERS-CoV homologue (Piotrowski et al., 2009; Cho et al., 2016). Interestingly, the X domain from IBV strain M41 but not of IBV strain Beaudette can bind ADPr (Xu et al., 2009; Piotrowski et al., 2009). The important “Gly – Gly – Gly” motif of the M41 X domain, involved in binding the distal ribose, is mutated to “Gly – Ser – Gly” in the Beaudette virus, thus preventing ADPr interaction with the X domain (Piotrowski et al., 2009). The virulence of IBV strain Beaudette is attenuated compared to that of IBV strain M41 (Geilhhausen et al., 1973). It is an interesting hypothesis that the loss of the ability to bind ADPr may be one of the reasons for the lower pathogenicity of the former IBV.
CoV inhibits the expression of innate-immunity genes (such as IFN-β, interleukin 6 (IL-6)) in *vitro* and thereby blocks the host immune response. At variance with this, Eriksson et al. (2008) and Fehr et al. (2015) reported that the Asn-to-Ala mutation in the MHV (strains A59 and JHM, resp.) X domain reduces the production of inflammatory cytokines (e.g., IL-6) in *vitro* and *in vivo*. Eriksson et al. (2008) hypothesized that the X domain aggravates MHV-induced severe liver pathology, likely by inducing the expression of inflammatory cytokines. These results suggest that the main function of the X domain may differ in different CoVs. On the other hand, the expression level of type-I IFN (α or β) is increased in cells infected with SARS-CoV or MHV carrying the Asn-to-Ala mutation in the X domain (Eriksson et al., 2008; Kuri et al., 2011; Fehr et al., 2016). This indicates that suppression of innate immunity by the X domain may be a feature conserved across the coronaviruses.

Recently, it was demonstrated that macrodomains from several +ssRNA viruses (such as HEV, SARS-CoV, HCoV 229E, Venezuelan equine encephalitis virus (VEEV), and Chikungunya virus (CHIKV)) act as hydrolases, removing mono- or poly(ADP-ribose) from mono- or poly(ADP-ribosylated) proteins, activities designated as de-mono-ADP-ribosylation (de-MARylation) and de-poly-ADP-ribosylation (de-PARylation), respectively (Li et al., 2016a). Interestingly, the macrodomain of VEEV and SARS-CoV can also re-accept ADPr(s) (Laing et al., 2011; Leidecker et al., 2016), but the acidic residue in the X domain (Eriksson et al., 2008; Kuri et al., 2011; Fehr et al., 2016) has been suggested to be an acceptor site for ADP-ribosylation; this still seems to be a matter of some debate. Arg and Ser have certainly been shown to accept ADPr(s) (Laing et al., 2011; Leidecker et al., 2016), but the acidic residues are also thought to be important sites of ADP-ribosylation (Feijs et al., 2013). PARP7, 10, and 12 can act as type-I IFN-stimulated genes (ISGs) and inhibit VEEV replication (Atasheva et al., 2014). Moreover, the RNA and protein synthesis of PARP14 (ARTD8) and PARP10 (ARTD10), without releasing free monomeric ADPr (Li et al., 2016a).

The MARylation activity is a common feature of the X domain (i.e., the central macrodomain fold (Fig. 5A and B)). The three subdomains were named SUD-N, SUD-M, and SUD-C, respectively. A region corresponding to parts of SUD was found to exist in other coronaviruses, mostly of clades B, C, and D of the genus Betacoronavirus (Neuman, 2016). For example, domains similar to SUD-C, SUD-M, and SUD-N are also encoded by the MERS-CoV genome (Kusov et al., 2015; Ma-Lauer et al., 2016). Thus, it is no longer appropriate to call this domain “SARS-unique”. Recently, the Nsp3 of MHV was shown by X-ray crystallography to contain a SUD-C-like fold (Chen et al., 2015). These authors renamed this region into “Domain Preceding Ubl2 and PL2αmo” (DUP). In this review, we follow the nomenclature proposed by Chen et al. (2015) and Neuman (2016), and use the designations macrodomain II (Mac2), macrodomain III (Mac3), and Domain Preceding Ubl2 and PL2αmo (DUP) for SUD-N, SUD-M, and SUD-C, respectively.

Mac2 (SUD-N) has been shown to be dispensable for the SARS-CoV replication/transcription complex within the context of a SARS-CoV replicon, but surprisingly, Mac3 (SUD-M) is essential, even though it is not conserved throughout the coronaviruses (Kusov et al., 2015). Mac2 and Mac3 each display a typical α/β/α macrodomain fold (Fig. 5A and B). The central β sheet with six β strands in the order β1 – β6 – β5 – β2 – β4 – β3 is flanked by two (or three) helices on either side. Only the last strand, β3, is antiparallel to the other strands. Interestingly, Mac2 and Mac3 have the same number of β strands in the central β sheet as the X domain of IBV (see above for X domain of IBV). The R.M.S.D. values are 2.5 Å - 2.6 Å (for 119/171 Co atoms) between Mac2, Mac3, and the X domain of SARS-CoV, according to the Dali server (Holm and Rosenstrom, 2010). The corresponding values are 2.6 Å - 2.7 Å (for 120/165 Ca atoms) when comparing SARS-CoV Mac2 and Mac3 with the X domain of IBV. Although the X-domain and Mac2/
Mac2-3 (SUD-NM) preferentially binds oligo(G), which activity (Kusov et al., 2015). Also, Mac3 can bind (GGGA)2 and (GGGA)5 as well as (GGGA)2GG (Johnson et al., 2010). In contrast, Mac2-3 cannot bind ADP-ribose (Tan et al., 2009; Chatterjee et al., 2010). These data indicate that the DPUP subdomain may fine-tune the specificity of RNA binding by Mac3 (Johnson et al., 2010).

The SARS-CoV genome contains three G6-stretches and two G5-stretches (Tan et al., 2009; Johnson et al., 2010), but none of them is conserved in all SARS-CoV strains. However, two GGAGGGUGAGG nucleotide segments, located in the Nsp2 and Nsp12 coding sequences, are highly conserved in various SARS-CoV strains (Johnson et al., 2010). These two nucleotide segments differ by only one base from the sequence favored by Mac3 – DPUP, (GGGA)2GG. Johnson et al. (2010) therefore proposed that these two sequences could be potential physiological substrates of Mac3 – DPUP. Besides specific elements in the genome of SARS-CoV, Mac3-2-3 might bind G-rich stretches in host mRNAs. In fact, Mac2-3 prefers to bind longer G-stretches, such as (G)10 to (G)14 (Tan et al., 2007). Such long G-stretches exist in several 3′ untranslated regions of host mRNAs, such as the NF-κB signaling pathway-related protein TAB3 mRNA and apoptotic signaling pathway protein Bbc3 mRNA (Tan et al., 2007, 2009). Mac2-3 may regulate the expression of these genes by binding to the poly(G) stretches in the corresponding mRNAs, thereby leading to disruption of the host antiviral response as well as of apoptotic signals. Mac3 has also been reported to bind oligo(A) (Chatterjee et al., 2009; Johnson et al., 2010). This observation (which is not in agreement with the results reported by Tan et al. (2007, 2009)) might suggest that Mac3 binds the poly(A) tail of the viral genome, or of subgenomic mRNAs, or of host mRNAs. Poly(A)-binding protein (PABP) binds the genomic poly(A) tails of BCoV (bovine coronavirus), MHV,
and TGEV, thereby enhancing the replication of these viruses (Spagnolo and Hogue, 2000; Galán et al., 2009). Is it possible that Mac3 binding to oligo(A) competes with the binding between PABP and the poly(A) tail? The question is yet to be answered.

Besides binding to nucleic acids, Mac2-3 of SARS-CoV has been shown to interact directly with host proteins, e.g. the E3 ubiquitin ligase RCHY1 (Ma-Lauer et al., 2015). RCHY1 and several other host proteins, Paip1, MKRN2, and MKRN3 etc. were reported to interact with Nsp3 (Pfefferle et al., 2011). However, the detailed binding region(s) on Nsp3 have not been identified. Ma-Lauer et al. (2016) demonstrated that Mac2-3 and the PL2pro of Nsp3 bind RCHY1, thus resulting in down-regulation of the antiviral protein p53 (see below). It is an interesting hypothesis that such interactions, which are absent from other CoVs because they lack Mac2-3, might account for a unique pathogenicity-related pathway utilized by SARS-CoV.

The DPUP (SUD-C) follows the Mac3 domain in SARS-CoV (Fig. 1A). Deletion of the domain within the context of a SARS-CoV replicon leads to a large reduction of RNA synthesis, but some basal RTC activity remains, indicating that the DPUP is not absolutely essential for replication (Kusov et al., 2015). Currently, three DPUP structures are available, one each from SARS-CoV and MHV (Table 1; Fig. 5C and D; Johnson et al., 2010; Chen et al., 2015), and the third one from bat coronavirus HKU9 (Table 1; Hammond et al., 2017). All DPUPs adopt a similar topology and overall structure. The R.M.S.D values between SARS-CoV DPUP and that of MHV or HKU9 are 2.1 Å (for 62 out of 74 Ca atoms; Z-score: 7.1) or 2.0 Å (for 62 out of 77 Ca atoms; Z-score: 7.0), respectively, according to the Dali server (Holm and Rosenström, 2010). The DPUP consists of an anti-parallel β sheet with two α helices located N- and C-terminal to this β sheet (Johnson et al., 2010; Chen et al., 2015). The two α helices form one plane while the β sheet forms the other; this resembles a typical frataxin-like fold (Bencze et al., 2006). Proteins featuring the frataxin-like fold are commonly involved in controlling cellular oxidative stress by binding iron to maintain the iron homeostasis (Bencze et al., 2006). In case of the yeast frataxin homologue Yfh1, cells lacking this gene were demonstrated to be highly sensitive to H2O2 and elevated metal ion levels (such as iron and copper) (Foury and Cazzalini, 1997). Several Glu and Asp residues in the N-terminal α helix of Yfh1 are possibly involved in binding metal ions (Fig. 5E; He et al., 2004; Bencze et al., 2006). Interestingly, “EXXXE” and “DDD” motifs exist in the first helix of the SARS-CoV and MHV DPUP, respectively, even though the sequence identity of DPUP is only 13% between these two viruses. Neumann et al. (2008) found that SARS-CoV Mac2−Mac3−DPUP can bind cobalt ions, while Mac3 alone and Mac2−Mac2*−Mac3 (2*: C-terminal half of Mac2) cannot. According to these observations, it is conceivable that the DPUP region binds metal ions. Furthermore, infection with SARS-CoV can induce transcription of oxygen stress-related genes of the host (Hu et al., 2012). Any involvement of DPUP in this biological process is speculative at this time.

The Mac2-3−DPUP oligodomain (SUD) has been shown to interact with Nsp9, Nsp12, and NAB−JSM−TM1 (see below) of Nsp3 by using a GST pull-down assay (Imbert et al., 2008). Using Y2H and co-immunoprecipitation (CoIP) assays, the oligoprotein Ubl1−Nsp2, ORF3a, and ORF9b (von Brunn et al., 2007); However, 2-3* (3*, N-terminal third of Mac3) of SARS-CoV Nsp3 has been found to bind Nsp2, ORF3a, and ORF9b (von Brunn et al., 2007); However, the slightly larger region Ubl1−HVR−Mac1-2-3 (see below) of Nsp3 has been found to bind Nsp2, ORF3a, and ORF9b (von Brunn et al., 2007); However, with the slightly larger region Ubl1−HVR−Mac1-2-3−DPUP, these interactions were not confirmed in an Y2H assay (Pan et al., 2008). It seems that DPUP might modulate the various binding processes. Furthermore, the DPUP subdomain could also regulate the sequence specificity of RNA binding by Mac3 as mentioned above (Johnson et al., 2010).

The relative orientation of SARS-CoV Mac2 and Mac3 is fixed by an artificial disulfide bond and dimer formation in the crystal (Tan et al., 2009). The NMR structure shows that Mac2 and Mac3 as well as Mac3 and DPUP have no preferred relative orientations to one another (Johnson et al., 2010). However, Mac2, Mac3, and DPUP are surrounded by other domains within Nsp3; it is unclear whether these other domains affect the relative orientation among the three. More multi-domain structures will be needed to answer this question and to elucidate the structural basis of mutual influences of these modules onto each other (see, e.g., above for the influence of the HEV helicase on the macromdomain of this virus).

5. Ubiquitin-like domain 2 and papain-like protease 2

Besides the Mac1 (X) domain, the largest number of crystal structures for any Nsp3 domain have been determined for the ubiquitin-like domain 2 (UbII) plus the papain-like protease 2 (PL2pro). So far, structures of this region are available for SARS-CoV, MERS-CoV, IBV, and MHV (Table 1; Ratia et al., 2006; Lei et al., 2014; Kong et al., 2015; Chen et al., 2015). UbII and PL2pro are conserved in all CoVs (Neuman et al., 2008; Neuman, 2016). The exact functional role of the UbII domain is not clear so far, while the PL2pro was reported to possess proteolytic, deubiquitinating, and delSGylating activities (Barretto et al., 2005; Lindner et al., 2005; Yang et al., 2014; Mielech et al., 2014).

5.1. Ubiquitin-like domain 2 (UbII)

The UbII is the second ubiquitin-like subdomain located within Nsp3 (Figs. 2C and 6). The structures of UbII in different CoVs are more conserved compared to the UbII. For example, the R.M.S.D between the UbIIs of SARS-CoV and MHV is 1.2 Å (for 58 out of 68 Ca atoms; Z-score: 11.1) according to the Dali server (Holm and Rosenström, 2010), whereas the corresponding value for the UbIIIs of the two viruses is 2.8 Å (for 85 out of 93 Ca atoms; Z-score: 7.5).

Some host USPs (with a fold similar to the CoV PL2pro) also include one or more Ub-like-domain(s), which is/are used to regulate the catalytic activity as well as to interact with partners (Komander et al., 2009; Faesen et al., 2012; Ploh et al., 2015). For example, the N-terminal UbII domain of USP14 is critical for its recruitment at the proteasome, thereby enhancing its catalytic activity (Hu et al., 2005; Faesen et al., 2012). USP7 (also named “HAUSP”: Herpesvirus-associated USP) includes five UbII domains (UbI 1–5), which are located at...
the C-terminus of the protease domain. Ub4-5 promote Ub binding and enhance the DUB activity of USP7 by about 100-fold via interacting with the “switching loop” (Trp285-Phe291) in the USP7 catalytic domain (Faesen et al., 2012). Ub2 of USP7 interacts with the HSV-1 immediate-early protein ICP0 to antagonize the host antiviral response (Pfoh et al., 2015). In contrast to the variable relative orientations of the Ub domains and the catalytic domain of USP7, the Ub2 domain is anchored to the CoV PL2pro by two salt-bridges in MERS-CoV and SARS-CoV (Lei et al., 2014), so it is unlikely to regulate the catalytic activity of PL2pro. In agreement with this conclusion, the presence or absence of the Ub2 of SARS-CoV or MERS-CoV shows almost no effect on the PL2pro activities (Friedman et al., 2009; Clasman et al., 2017).

Currently, several inconsistent roles of Ub2 are reported. Friedman et al. (2009) demonstrated that the Ub2 of SARS-CoV is necessary to antagonize the host innate immune response via blocking IRF3 or the NF-kB pathway. In contrast, Clementz et al. (2010) reported that the Ub2 of SARS-CoV is not necessary for antagonizing IFN production. Also, Mielech et al. (2015) showed that the Val787Ser mutation (Nsp3 numbering) in the MHV Ub2 reduces the thermal stability of the PL2pro, whereas Clasman et al. (2017) reported that the deletion of Ub2 in MERS-CoV did not affect PL2pro thermal stability. The former Val residue of MHV is conserved in SARS-CoV and MERS-CoV. It is located in the first strand (β1) and contributes to the hydrophobic core of Ub2; therefore, the Val-to-Ser change might disrupt the global Ub2 structure, leading to a decrease in the stability of the PL2pro domain (Mielech et al., 2015).

On the basis of molecular dynamics simulations, the MERS-CoV Ub2 has recently been proposed to display more molecular flexibility when the PL2pro binds ubiquitin, compared to the situation in the free enzyme. The authors speculate that the difference in flexibility of the Ub2 might regulate the interaction with downstream targets, thereby modulating the innate immune response (Alfuwaires et al., 2017). Ubiquitination and deubiquitination cannot only regulate the catalytic activity of PL2pro, with the goal of facilitating error correction in MERS-CoV (see above). The catalytic Cys is located in the thumb subdomain (at the N terminus of helix 4 of SARS-CoV and MERS-CoV (Lei et al., 2014; Lei and Hilgenfeld, 2016). Interestingly, this Asp164 is unique among CoV PL2pros, and the Asp164Ala replacement leads to an about 4.5-fold and 3.5-fold reduction of the proteolytic and DUB activities (Frieman et al., 2009; Clasman et al., 2017). Meanwhile, the corresponding activity of the Leu-to-Trp mutation in MERS-CoV PL2pro is about 50-fold higher than that of the wild-type enzyme, using the same substrate (Lei et al., 2014). As we mentioned before (Lei and Hilgenfeld, 2016), the efficiency of viral proteases does not always have to be optimized during virus evolution. Rather, the creation of temporary intermediates of polyprotein cleavage, in the right temporal order, is necessary for correct virus replication (Kanjanahaluethai and Baker, 2000; Gosert et al., 2002; Harcourt et al., 2004); thus, the proper but not necessarily the highest protease activity is beneficial for viral survival.

In order to investigate the mechanism of the DUB and deISGylating activities of CoV PL2pro’s, the complex of the enzyme with ubiquitin (or ISG15) is important. Until now, structures of SARS-CoV and MERS-CoV PL2pro with mono-Ub as well as of SARS-CoV PL2pro with di-Ub have been obtained (Chou et al., 2014; Ratia et al., 2014; Békes et al., 2016; Bailey-Elkin et al., 2014; Lei and Hilgenfeld, 2016). Very recently, the structures of both SARS-CoV and MERS-CoV PL2pro in complex with the C-terminal Ub domain of hISG15 or mISG15 have been reported (Daczkowski et al., 2017a,b). These structures show that the PL2pro of SARS-CoV possesses two ubiquitin-binding sites (named Ub1 and Ub2 sites here; Ratia et al., 2014; Békes et al., 2016). From the prior structure of USP14 in complex with ubiquitin, it is known that two blocking loops (BL1 and BL2) regulate substrate binding (Hu et al., 2005). Different from that, only the BL2 exists in CoV PL2pro’s and is involved in substrate binding (Fig. 6) (Chou et al., 2014; Ratia et al., 2014; Bailey-Elkin et al., 2014; Lei and Hilgenfeld, 2016), whereas BL1 is absent in CoV PL2pro’s (Ratia et al., 2006; Lei et al., 2014).

The proximal Ub binding site (Ub1) is, to a certain degree, conserved between the PL2pro’s of SARS-CoV and MERS-CoV. The region includes the narrow substrate channel between the thumb and the palm subdomains, as well as a hydrophobic patch in the fingers subdomain (Fig. 6). The narrow substrate channel binds the C-terminal RLRGG residues of ubiquitin (Chou et al., 2014; Ratia et al., 2014; Bailey-Elkin et al., 2014; Lei and Hilgenfeld, 2016; in order to be clear, Ub residues appear in italics here). The C-terminal RLRGG of ubiquitin is similar to the unprimed side of the polyprotein substrates, (R/K)(L/I)XGG in the two viruses. The S1, S2, and S4 sites are well conserved to accommodate the two small glycines (P1, P2) and the hydrophobic P4 residue (Leu or Ile). In contrast, the flexible side-chains in P3 and P5 feature binding patterns that are slightly different between SARS-CoV and MERS-CoV PL2pro. In the SARS-CoV PL2pro (Cys112Ser) – Ub complex, P3-Arg forms a weak salt-bridge with Glu162 (Chou et al., 2014), whereas the corresponding P3-Arg is exposed to solvent in the MERS-CoV complex (Lei and Hilgenfeld, 2016). On the other hand, the P5-Arg is exposed to solvent in the SARS-CoV complex (Chou et al., 2014) but forms a strong salt-bridge with Asp164 in MERS-CoV (Bailey-Elkin et al., 2014; Lei and Hilgenfeld, 2016). Interestingly, this Asp164 is unique among CoV PL2pro’s, and the Asp164Ala replacement leads to an about 4.5-fold and 3.5-fold reduction of the proteolytic and DUB activities, respectively (Lei and Hilgenfeld, 2016). As just mentioned, the proteolytic activity of the MERS-CoV PL2pro is not optimized due to the deficient oxyanion hole. On the other hand, the virus requires a strong DUB activity to counteract the host immune response. The suboptimal enzymatic activities may be partly compensated by the unique Asp164 (Lei and Hilgenfeld, 2016).
In addition to the binding of the Ub C-terminus to the substrate channel, there is an interaction between a hydrophobic region of the SARS-CoV and MERS-CoV PL2pro's in the fingers subdomain and a hydrophobic patch (Ile44, Ala46, Gly47) of Ub (Chou et al., 2014; Ratia et al., 2014; Bailey-Elkin et al., 2014; Lei and Hilgenfeld, 2016; Békés et al., 2016). This hydrophobic patch of Ub is commonly used to interact with Ub-binding proteins (Dikic et al., 2009). The fingers subdomain residues involved are Tyr208 and Met209 in SARS-CoV, and Tyr209 and Val210 in MERS-CoV (Chou et al., 2014; Ratia et al., 2014; Bailey-Elkin et al., 2014; Lei and Hilgenfeld, 2016; Békés et al., 2016). Moreover, these hydrophobic interactions between the PL2pro and Ub are important for the DUB activity of the enzyme, because disrupting them via a Val210Arg mutation dramatically diminishes the DUB activity in MERS-CoV PL2pro (Bailey-Elkin et al., 2014).

Near the hydrophobic patch of Ub, Arg42 forms a salt-bridge with Glu168 of PL2pro in two structures of the SARS-CoV PL2pro in complex with mono-ubiquitin or Lys48-linked di-Ub (Chou et al., 2014; Ratia et al., 2014; Békés et al., 2016). However, this Glu is replaced by Arg in MERS-CoV PL2pro, resulting in Arg42 instead forming a salt-bridge with Asp165 in the MERS-CoV PL2pro-ubiquitin complex (Lei and Hilgenfeld, 2016). This illustrates that various fine-tuned binding patterns exist between Ub and PL2pro's in different CoVs.

Besides the Ub1 binding site, the Ub2 binding site is mapped by the complex of SARS-CoV PL2pro with Lys48-linked di-Ub (Fig. 6; Békés et al., 2016). The Ub2 binding site is located at the first α helix of the thumb subdomain. Phe70 interacts with the common hydrophobic patch (Ile44, Ala46, Gly47) of Ub. Interestingly, MERS-CoV PL2pro seems to lack the corresponding Ub2 binding site. Phe70 of SARS-CoV PL2pro is changed to Lys69 in MERS-CoV (Békés et al., 2016). In addition, Békés et al. (2016) predicted that Trp107 and Ala108 could constitute the Ub1 binding site in SARS-CoV PL2pro. The Trp107Leu/Ala108Ser double mutation reduces the enzyme's activity towards this year (Basters et al., 2017). Surprisingly, the N-terminal Ubl domain of mISG15 shows almost no interaction with mUSP18. Does ISG15 behave similarly when binding to the CoV PLpro? How does the N-terminal domain of ISG15 of different species recognize the cognate CoV PLpro? It would be of interest to determine not only the structure of a full-length hISG15–HCoV PLpro complex but also that of mISG15 with MHV PLpro.

The DUB and deISGylating activities of CoV PLpro's are well established, but the detailed mechanism of the PLpro antagonism of the host innate immune response is still ambiguous (see Lei and Hilgenfeld, 2017, for a recent review). Various cytokines (including interferons (IFNs) and tumor necrosis factors (TNFs)) are produced to inhibit virus replication by two main pathways, the IRF3 pathway and the NF-κB pathway (Seth et al., 2006; Hiscott et al., 2006). For more information on the host innate immune system signaling pathways, the reader should consult other reviews (e.g., Mogensen, 2009; Lei and Hilgenfeld, 2017). Devaraj et al. (2007) found that the SARS-CoV PL2pro can directly bind IRF3 to block its phosphorylation, dimerization, and nuclear translocation, thereby inhibiting IFN-β induction. Furthermore, the PL2pro was found not to block the NF-κB signaling pathway and the protease activity was described as dispensable for antagonizing the IFN response (Devaraj et al., 2007). Clementz et al. (2010) also confirmed that the enzyme activity of HCoV-NL63 PL2pro is not essential for counteracting the antiviral IFN production. In contrast, Friesan et al. (2009) reported that the SARS-CoV PL2pro does not directly bind IRF3 or disrupt its phosphorylation. Instead, the PL2pro was proposed to inhibit the NF-κB signaling pathway by stabilizing its inhibitor, IκBα (Friesan et al., 2009). Furthermore, the protease activity of SARS-CoV PL2pro is important for blocking the TNF-α/NF-κB signaling pathway (Friesan et al., 2009). In addition, the HCoV-NL63 but not the MHV PL2pro has the ability to impede the IRF3 and NF-κB pathways, indicating that the functions of the PL2pro are specific for different CoVs (Friesan et al., 2009). Later, a protein comprising the SARS-CoV PL2pro and the TM (transmembrane region of Nsp3) was demonstrated to inhibit the STING/TBK1/IKKε-mediated signaling pathway (upstream regulators of IRF3; Chen et al., 2014), thereby disrupting IRF3 phosphorylation and dimerization, and blocking the type-I IFN response. SARS-CoV PL2pro plus TM can also physically interact with the STING-TRAF3-TBK1 complex and remove the ubiquitins from ubiquitinated RIG-I, STING, TRAF3, TBK1, as well as IRF3 (Chen et al., 2014). In 2016, it was reported that the SARS-CoV PL2pro can inhibit the Toll-like receptor 7 (TLR7)–mediated type-I IFN response and the NF-κB pathway by removing the Lys63-linked polyUb chain from TRAF3 and TRAF6 (upstream regulators of IRF3 and NF-κB; Li et al., 2016b). Interestingly, the SARS-CoV PL2pro only removes the Lys63 but not the Lys48-linked polyUb chain from TRAF3 and TRAF6 in vivo (Li et al., 2016b). On the other hand, Báez-Santos et al. (2014) and Békés et al. (2015, 2016) have shown that SARS-CoV PL2pro prefers to digest Lys48-over Lys63-linked polyUb chains in vitro (see above). Why does the substrate specificity of PL2pro seem to be different in vivo and in vitro? Does any other factor influence the substrate specificity of PL2pro in vivo when counteracting the cellular innate immune response? These questions are yet to be answered.

In addition, the HCoV-NL63 PL2pro was shown to block the p53-IRF7-IFNβ signaling pathway (Yuan et al., 2015). p53 can induce type-I interferon production via IRF7 (interferon regulatory factor 7; Yuan et al., 2015). Meanwhile, p53 can be degraded via the MDM2- (an E3 ubiquitin ligase) mediated ubiquitin-proteasome system (Haupt et al., 2009). Furthermore, the SARS-CoV PL2pro cleaves the polyUb chain by removing di-Ub, not mono-Ub units as in MERS-CoV (Békés et al., 2015). This strongly suggests that MERS-CoV PL2pro cleaves the polyUb chain by removing di-Ub, not mono-Ub units as in MERS-CoV (Békés et al., 2015). This strongly suggests that MERS-CoV PL2pro prefers to digest Lys48- and Lys63-linked polyUb chains with similar efficiency (Báez-Santos et al., 2014). Lys48-linked Ub chains mainly cause target protein degradation via the 26S proteasome, while Lys63-linked polyUb is mainly related to DNA repair and signal transduction (Ikeda and Dikic, 2008), in particular, in the signal transduction cascades of the host innate immune system (Dikic and Dötsch, 2009). However, the biological significance of the CoV PLpro's showing different cleavage activities on Lys48- and Lys63-linked polyUb is still unclear. Furthermore, the SARS-CoV PL2pro cleaves the polyUb chain by removing di-Ub, not mono-Ub units as in MERS-CoV (Békés et al., 2015). This strongly suggests that MERS-CoV PL2pro possesses the Ub1 and Ub1' binding sites but not a Ub2 site, consistent with the Phe70 to Lys mutation in MERS-CoV PL2pro as just mentioned.

At the same time, ISG15 utilizes a different Ub2 binding site of SARS-CoV PL2pro, compared to Lys48-linked di-Ub (Békés et al., 2016), but no structure for a full-length ISG15–CoV PL2pro complex is available so far. Daczkowski et al. (2017a) reported that the C-terminal domains of ISG15s (similar to Ub1 mentioned above) from different species have different binding characteristics with SARS-CoV PL2pro according to two structures, the PL2pro in complex with the C-terminal domain of hISG15 and mISG15, respectively. In addition, the structure of mouse USP18 in complex with full-length mISG15 became available this year (Basters et al., 2017). Surprisingly, the N-terminal UbI domain of mISG15 shows almost no interaction with mUSP18. Does ISG15...
coronavirus replication (Ma-Lauer et al., 2016). While the HCoV-NL63 PL2<sup>pro</sup> stabilizes MDM2 by deubiquitinating it (Yuan et al., 2015), the SARS-CoV PL2<sup>pro</sup> surprisingly does not deubiquitinate RCHY1 (Ma-Lauer et al., 2016). How does the PL2<sup>pro</sup> stabilize RCHY1? The mechanism has yet to be elucidated.

Besides the functions of PL2<sup>pro</sup> discussed above, the enzyme was shown to interact with other viral proteins. The region from PL2<sup>pro</sup> to the C-terminus of Nsp3 in SARS-CoV can interact with the Nsp2, ORF3a, and ORF9b proteins, as identified by Y2H and CoIP assays (von Brunn et al., 2007). Through similar assays, the region PL2<sup>pro</sup>−NAB−βSM was found to interact with Nsp4 as well as Nsp12 (Pan et al., 2008). The SARS-CoV PL2<sup>pro</sup> was further shown to bind ORF7a and Nsp6 by using proteomics analysis (Neuman et al., 2008).

Coronavirus PL2<sup>pro</sup> is an important target for developing antiviral drugs. This aspect has been well reviewed by Báez-Santos et al. (2015) within this series; hence, we mention only inhibitors here that have been described since. Two big challenges exist when designing PL2<sup>pro</sup> inhibitors: 1) the S1 and S2 binding sites are tailor-made to accommodate glycan residues and hence they are small; therefore, identifying suitable peptidomimetic chemical structures is difficult; 2) many host USPs feature folds and active sites similar to the PL2<sup>pro</sup>’s, so specificity of the inhibitors could be an issue. However, there is a good chance that the BL2 loop (mentioned above) of CoV PL2<sup>pro</sup> could provide sufficient uniqueness to solve the specificity problem. This loop is involved in substrate binding and is different not only between USPs and CoV PL2<sup>pro</sup>s but also among different CoVs (Hu et al., 2005; Ratia et al., 2006; Lei et al., 2014; Báez-Santos et al., 2014, 2015; Lee et al., 2015). For example, this loop comprises 6 amino-acid residues (GNYQCG) in SARS-CoV PL2<sup>pro</sup> but 7 (GIETAVG) in the enzyme of MERS-CoV, leading to the inability of SARS-CoV PL<sup>pro</sup> inhibitors to act on MERS-CoV PL<sup>pro</sup> (Báez-Santos et al., 2014; Hilgenfeld, 2014; Lee et al., 2015). Using a high-throughput assay, the purine derivative 8-(trifluoromethyl)-9H-purin-6-amine (compound 4; Fig. 7A) was identified as a competitive MERS-CoV PL2<sup>pro</sup> inhibitor, with an IC<sub>50</sub> of about 2 μM in vitro (Lee et al., 2015). Furthermore, the authors also reported that this inhibitor shows very high selectivity against human ubiquitin C-terminal hydrolase (hUCH-L1; IC<sub>50</sub> > 100 μM), which is one of the host proteins most closely related to the CoV PL<sup>pro</sup> (Lee et al., 2015). In contrast, Clasman et al. (2017) reported that compound 4 features no selective inhibition of CoV PL<sup>pro</sup>s nor host USPs; therefore, this compound could be a pan-assay interference inhibitor (or PAIN).

Recently, nine alkylated chalcones (1−9) and four coumarins (10−13), which were isolated from the perennial plant Angelica keiskei, had their inhibitory activities against both the SARS-CoV M<sup>pro</sup> (3CL<sup>pro</sup>, chymotrypsin-like protease) and the PL2<sup>pro</sup> tested (Park et al., 2016). One of the chalcones, compound 6 (Fig. 7B), exhibited relatively strong inhibition of both the 3CL<sup>pro</sup> and the PL2<sup>pro</sup> in vitro, with IC<sub>50</sub> values of 11.4 and 1.2 μM respectively (Park et al., 2016). Chalcone 6 uses different inhibition mechanisms for 3CL<sup>pro</sup> and PL2<sup>pro</sup>. It is a competitive inhibitor for the former enzyme but a non-competitive one for the latter (Park et al., 2016). Clearly, the large body of structural information available for the CoV PL<sup>pro</sup>s and host DUBs should enable more design of inhibitors specific for the viral enzyme.

6. Nucleic acid-binding (NAB) domain and betacoronavirus-specific marker (βSM) domain

The nucleic-acid binding (NAB) and betacoronavirus-specific marker (βSM) domains together are also named “Nsp3e” (Neuman et al., 2008). The latter domain alone was previously called “group 2-specific marker” (G2M) (Neuman et al., 2008). The NAB and βSM domain exist in the genus Betacoronavirus. The corresponding region is absent in alphacoronaviruses and deltacoronaviruses (Neuman, 2016). In gammacoronaviruses, there is a gammacoronavirus-specific marker (γSM) domain at this position (Neuman, 2016).

Structural information on this region is very limited for all coronaviruses. Thus far, only an NMR structure of the NAB domain of SARS-CoV is available (Table 1; Fig. 8; Serrano et al., 2009). This structure comprises two antiparallel β sheets (β1 + β6; β2 + β8) and one parallel β sheet (β3 − β4 − β5 − β7) as well as two α helices and two 3₁₀ helices (α1 and α2) in the order β1 − β2 − β3 − α1 − β4 − β5 − α1 − β2 − β6 − β7 − α2 − β8. Four β strands (β3 − β4 − β5 − β7) and two helices (α1, α2) form a “half-barrel”. The structure of the NAB represents a unique fold (Serrano et al., 2009). The domain has been shown to bind ssRNA as well as to unwind dsDNA (Neuman et al., 2008). When binding to ssRNA, the NAB prefers sequences with repeats of three consecutive Gs (Serrano et al., 2009), such as (GGGA), and (GGGA)<sub>2</sub>. A positively charged surface patch (Lys75, Lys76, Lys99, and Arg106) is involved in RNA binding (Fig. 8). These residues are located in the loop between α2 and β6 as well as in helix α2 (Serrano et al., 2009). The RNA binding behavior of the NAB appears to be similar to that of SARS-CoV Mac3 (SUD-M), which has a specificity for oligo(G) (Tan et al., 2007, 2009), although the latter is also reported to bind oligo(A) (Chatterjee et al., 2009; Johnson et al., 2010, mentioned above). Whether there is a functional relation between Mac3 and NAB, remains to be investigated.

Currently, no structural information is available concerning the βSM or γSM, and nothing is known about the function of these modules either. A gene encoding the βSM domain of SARS-CoV could not be expressed in E. coli; this module has been predicted to be a non-enzymatic domain (Neuman et al., 2008). In the absence of sequence similarity to any domain of known function, we performed an ab-initio protein structure prediction using the sequence of the SARS-CoV βSM domain and the QUARK online server (Xu and Zhang, 2012). The result indicates that most of this region is intrinsically disordered. This does not preclude that it might adopt a defined structure upon interaction with another Nsp or RNA, or a host protein.

7. Transmembrane regions (TM1 and TM2), Nsp3 ectodomain, Y1 domain, and CoV-Y domain

This part of Nsp3 includes two transmembrane regions as well as three soluble domains, which together constitute about one third of the multidomain protein. The two transmembrane regions are TM1 and TM2, while the three domains are the Nsp3 ectodomain (3Ecto), Y1, and CoV-Y. The sequential order of this part is TM1 − 3Ecto − TM2 − Y1 − CoV-Y (Fig. 1A and B). Even though this part

---

Fig. 7. Recently described inhibitors of the CoV PL2<sup>pro</sup>. (A) Structural formula of the purine derivative 8-(trifluoromethyl)-9H-purin-6-amine (compound 4). This compound is a competitive MERS-CoV PL2<sup>pro</sup> inhibitor (Lee et al., 2015). It is also active against SARS-CoV PL2<sup>pro</sup> but acts as an allosteric inhibitor in this case. (B) A natural-product chalcone, compound 6 from the perennial plant Angelica keiskei, inhibits the SARS-CoV M<sup>pro</sup> (3CL<sup>pro</sup>) and PL2<sup>pro</sup> in vitro (Park et al., 2016).
and has also been designated as a zinc-Nsp3 (Table 1; Fig. 1B). Oostra et al. (2008) proposed that the transmembrane regions plus the 3Ecto are important for the PL2pro to process the Nsp3/4 cleavage site in SARS-CoV and MHV (Harcourt et al., 2004; Kanjanahaluethai et al., 2007); a possible reason is that the transmembrane part could bring the PL2pro close to the cleavage site between the membrane-associated proteins Nsp3 and Nsp4. Asparagine (N)-linked glycosylation has been found in the 3Ecto domains of SARS-CoV and MHV (Figs. 1B and 9; Harcourt et al., 2004; Kanjanahaluethai et al., 2007). It is unclear if the N-glycan modification affects the 3Ecto conformation or stability. Frequently, N-linked glycans serve as recognition points for partner molecules (Aebi, 2013). It has been shown that interaction of the 3Ecto with the luminal loop of Nsp4 is essential for the ER rearrangements occurring in cells infected by SARS-CoV or MHV (the 3Ecto is named “luminal loop of Nsp3” in this paper; Hagemeijer et al., 2014).

The Y1 and CoV-Y domains are located at the cytosolic side of the ER. The Y1 domain is conserved in all viruses of the order Nidovirales, while CoV-Y is only conserved in all coronaviruses (Neuman, 2016). Since no three-dimensional structure is available for this part, the domain assignment of Y1 and CoV-Y is ambiguous (Neuman, 2016). We found that the sequence identity of Y1 + CoV-Y between different CoV genera is above 25% and two Cys-His clusters are present in the N-terminal part of the Y1 domain, possibly binding zinc ions (Fig. 9). However, it is still unclear if the fold and function in this region are conserved. Currently, functional information on this part is limited. It has been shown that the C-terminal third of Nsp3 (βSM (partial) − TM1 − 3Ecto − TM2− AH1 − Y1 + CoV-Y) of Nsp3 binds less efficiently to Nsp4 without the Y1 and CoV-Y domains (Hagemeijer et al., 2014), although these two domains are not as important for this process as the 3Ecto.

According to a Y2H screen, CoIP, as well as GST pull-down assays, different constructs of Nsp3 with different C-terminal regions were identified to interact with various viral non-structural proteins of SARS-CoV (von Brunn et al., 2007; Imbert et al., 2008; Pan et al., 2008). For example, a construct comprising the domains from PL2pro to the end of Nsp3 can bind Nsp2, ORF3a, and ORF9b (see above; von Brunn et al., 2007); the NAB + βSM−TM1 of Nsp3 can interact with Nsp5, Nsp7 ~ 8, as well as Nps 12–16, and Y1 plus CoV-Y interacts with Nsp9 and Nsp12 (Imbert et al., 2008); in addition, the NAB−βSM−TM1 of Nsp3 can also interact with other domains within Nsp3, except for Mac1 (X domain) (Imbert et al., 2008); a PL2pro−NAB−βSM−TM1 construct of Nsp3 can bind Nsp4 and Nsp12, while the region from TM1 to the end of Nsp3 only binds Nsp8 (Pan et al., 2008). It has been found that the interaction between the C-terminal region of Nsp3 and Nsp4 is essential for the formation of CMs and DMVs derived from the ER in CoV-infected cells (Angelini et al., 2013; Hagemeijer et al., 2014). The viral RNA and replicase proteins (Nps) need to be associated with these modified membranes to form the replicative organelles (see Neuman, 2016, for review). In addition, these membranes can protect the viral RNA and Nps against nucleases and proteases in vitro (van Hemert et al., 2008). Besides the Nsp3−Nsp4 interaction, it is still unclear whether all other interactions really exist or how these interactions affect the viral life cycle. At least, it seems that the membrane-associated region of Nsp3 may regulate the interactions with other viral proteins. It is definitely necessary to put more effort into the structural and functional characterization of this region.

8. Conclusions

Overall, the multi-domain Nsp3 plays various roles in coronavirus infection. It releases Nsp1, Nsp2, and itself from the polyproteins and interacts with other viral Nps as well as RNA to form the replication/transcription complex. It acts on posttranslational modifications of host proteins to antagonize the host innate immune response (by de-MARYlation, de-PARYlation (possibly), deubiquitination, or

Fig. 8. NMR structure of the nucleic acid-binding (NAB) domain in SARS-CoV (cartoon style; PDB entry: 2K87; Serrano et al., 2009). The order of secondary-structure elements is β1−β2−β3−α1−β4−β5−α2−β6−β7−α2−β8. The overall structure of NAB represents a unique fold. The residues involved in RNA binding (Lys75, Lys76, Lys99, and Arg106) are displayed in blue. The N and C termini of the NAB domain are labeled.

exists in all coronaviruses (Neuman et al., 2008; Neuman, 2016), thus far, no three-dimensional structure is available for the entire region nor for a part of it.

Nsp3 of CoVs is thought to pass the ER membrane twice, since there are two predicted transmembrane regions, TM1 and TM2 (Harcourt et al., 2004; Kanjanahaluethai et al., 2007; Oostra et al., 2008). According to the transmembrane region prediction server TMHMM (Krogh et al., 2001), there is a total of three hydrophobic regions in SARS-CoV Nsp3 (Table 1; Fig. 1B). Oostra et al. (2008) proposed that the first two of the three hydrophobic regions span the membrane while the last one (AH1), which has more amphiphatic character, does not (Fig. 1B). Thus, the 3Ecto would be the only domain located on the luminal side of the ER in SARS-CoV Nsp3 (Fig. 1B). The 3Ecto is thought to bind metal ions and has also been designated as a zinc-finger (ZF) domain before (Neuman et al., 2008). Neuman (2016) found that the metal binding Cys-His cluster is not conserved in all CoVs and has renamed this domain into “3Ecto”. In fact, only two cysteine residues are conserved in the CoV 3Ecto domain (Fig. 9A), hence this domain is unlikely to be a zinc-finger domain. The transmembrane regions plus the 3Ecto are important for the PL2pro to process the Nsp3/4 cleavage site in SARS-CoV and MHV (Harcourt et al., 2004; Kanjanahaluethai et al., 2007); a possible reason is that the transmembrane part could bring the PL2pro close to the cleavage site between the membrane-associated proteins Nsp3 and Nsp4. Asparagine (N)-linked glycosylation has been found in the 3Ecto domains of SARS-CoV and MHV (Figs. 1B and 9; Harcourt et al., 2004; Kanjanahaluethai et al., 2007). It is unclear if the N-glycan modification affects the 3Ecto conformation or stability. Frequently, N-linked glycans serve as recognition points for partner molecules (Aebi, 2013). It has been shown that interaction of the 3Ecto with the luminal loop of Nsp4 is essential for the ER rearrangements occurring in cells infected by SARS-CoV or MHV (the 3Ecto is named “luminal loop of Nsp3” in this paper; Hagemeijer et al., 2014).

The Y1 and CoV-Y domains are located at the cytosolic side of the ER. The Y1 domain is conserved in all viruses of the order Nidovirales, while CoV-Y is only conserved in all coronaviruses (Neuman, 2016). Since no three-dimensional structure is available for this part, the domain assignment of Y1 and CoV-Y is ambiguous (Neuman, 2016). We found that the sequence identity of Y1 + CoV-Y between different CoV genera is above 25% and two Cys-His clusters are present in the N-terminal part of the Y1 domain, possibly binding zinc ions (Fig. 9). However, it is still unclear if the fold and function in this region are conserved. Currently, functional information on this part is limited. It has been shown that the C-terminal third of Nsp3 (βSM (partial) − TM1 − 3Ecto − TM2− AH1 − Y1 + CoV-Y) of Nsp3 binds less efficiently to Nsp4 without the Y1 and CoV-Y domains (Hagemeijer et al., 2014), although these two domains are not as important for this process as the 3Ecto.

According to a Y2H screen, CoIP, as well as GST pull-down assays, different constructs of Nsp3 with different C-terminal regions were identified to interact with various viral non-structural proteins of SARS-CoV (von Brunn et al., 2007; Imbert et al., 2008; Pan et al., 2008). For example, a construct comprising the domains from PL2pro to the end of Nsp3 can bind Nsp2, ORF3a, and ORF9b (see above; von Brunn et al., 2007); the NAB−βSM−TM1 of Nsp3 can interact with Nsp5, Nsp7 ~ 8, as well as Nps 12–16, and Y1 plus CoV-Y interacts with Nsp9 and Nsp12 (Imbert et al., 2008); in addition, the NAB−βSM−TM1 of Nsp3 can also interact with other domains within Nsp3, except for Mac1 (X domain) (Imbert et al., 2008); a PL2pro−NAB−βSM−TM1 construct of Nsp3 can bind Nsp4 and Nsp12, while the region from TM1 to the end of Nsp3 only binds Nsp8 (Pan et al., 2008). It has been found that the interaction between the C-terminal region of Nsp3 and Nsp4 is essential for the formation of CMs and DMVs derived from the ER in CoV-infected cells (Angelini et al., 2013; Hagemeijer et al., 2014). The viral RNA and replicase proteins (Nps) need to be associated with these modified membranes to form the replicative organelles (see Neuman, 2016, for review). In addition, these membranes can protect the viral RNA and Nps against nucleases and proteases in vitro (van Hemert et al., 2008). Besides the Nsp3−Nsp4 interaction, it is still unclear whether all other interactions really exist or how these interactions affect the viral life cycle. At least, it seems that the membrane-associated region of Nsp3 may regulate the interactions with other viral proteins. It is definitely necessary to put more effort into the structural and functional characterization of this region.

8. Conclusions

Overall, the multi-domain Nsp3 plays various roles in coronavirus infection. It releases Nsp1, Nsp2, and itself from the polyproteins and interacts with other viral Nps as well as RNA to form the replication/transcription complex. It acts on posttranslational modifications of host proteins to antagonize the host innate immune response (by de-MARYlation, de-PARYlation (possibly), deubiquitination, or
As the largest non-structural protein of CoVs, Nsp3 has also been identified as the major selective target for driving evolution in lineage C betaCoVs on the basis of a high rate of positively selected mutation sites (Forni et al., 2016). Furthermore, the adaptive evolution of Nsp3 of MERS-CoV is still ongoing (Forni et al., 2016). For example, the Arg911Cys mutation (located in the palm subdomain of the PL2 domain, corresponding to Arg283 in Lei et al., 2014) of Nsp3 exists in the viral strain KOR/KNIH responsible for the 2015 South Korean outbreak but not in the ancestral strain EMC/2012 (Forni et al., 2016). It is interesting to speculate why coronaviruses keep many essential functions in one protein, while this protein shows high-rate genetic diversity during CoV evolution. In the end, increased research efforts into the structure and function of Nsp3 are needed to achieve a more complete understanding of this protein.

References

Adams, M.J., Carstens, E.B., 2012. Ratification vote on taxonomic proposals to the international committee on taxonomy of viruses. Arch. Virol. 157, 1411–1422. Aeby, M., 2013. N-linked protein glycosylation in the ER. Biochim. Biophys. Acta 1833, 2430–2437.

Alfaureix, M., Altaher, A., Kandeel, M., 2017. Molecular dynamic studies of interferon-γ sensitization in signaling and disease. Cells 4, 569–595.

Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., Mesecar, A.D., Baker, S.C., 2005. The SARS coronavirus papain-like protease. Mol. Cell 62, 572–582.

Bastiaens, P.T.M., Bredenbeek, P.J., van Kasteren, P.B., Bredenbeek, P.J., Snijder, E.J., Kikkert, M., Mark, B.L, 2014. Crystal structure of the Middle East respiratory syndrome coronavirus (MERS-CoV) papain-like protease bound to ubiquitin facilitates targeted disruption of deubiquitinating activity to demonstrate its role in innate immune suppression. J. Biol. Chem. 289, 308–317.

Báez-Santos, Y.M., Mielec, A.M., Deng, X., Baker, S., Mesecar, A.D., 2014. Catalytic function and substrate specificity of the papain-like protease domain of np3 from the Middle East respiratory syndrome coronavirus. J. Virol. 88, 12511–12527.

Báez-Santos, Y.M., St John, S.E., Mesecar, A.D., 2015. The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. Antivir. Res. 115, 21–38.

Bailey-Elkin, B.A., Knapp, R.C., Johnson, G.G., Dalebout, T.J., Ninaber, D.K., van Rateren, P.B., Bredenbeek, P.J., Snijder, E.J., Kikkert, M., Mark, B.L., 2014. Crystal structure of the Middle East respiratory syndrome coronavirus (MERS-CoV) papain-like protease bound to ubiquitin facilitates targeted disruption of deubiquitinating activity to demonstrate its role in innate immune suppression. J. Biol. Chem. 289, 34667–34682.

Bakshi, S., Holzer, B., Bridgen, A., McMullan, G., Quinn, D.G., Baron, M.D., 2013. Duplex virus ovarian tumour disorder interferes with ubiquitin/ISG15-regulated innate immune cell signalling. J. Gen. Virol. 94, 298–307.

Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., Mesecar, A.D., Baker, S.C., 2005. The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. J. Virol. 79, 15189–15198.

Baxters, A., Geurink, P.P., Röcker, A., Witting, K.F., Tadayon, R., Hess, S., Semrau, M.S., Briers, A., Geurink, P.P., Röcker, A., Witting, K.F., Tadayon, R., Hess, S., Semrau, M.S., Stori, P., Osva, H., Knoebloch, K.F., Fritz, G., 2017. Structural basis of the specificity of USP18 toward ISG15. Nat. Struct. Mol. Biol. 24, 270–278.

Bekes, M., Rut, W., Kasperkiewicz, P., Mulder, M.P., Osva, H., Drag, M., Lima, C.D., Huang, T.T., 2015. SARS-CoV papain-like protease is a unique lys48 linkage-specific di-distributive deubiquitinating enzyme. Biochem. J. 468, 215–226.

Bekes, M., van der Heiden van Noort, G.J., Ekkebus, R., Osva, H., Huang, T.T., 2016. Recognition of Lys48-linked di-ubiquitin and deubiquitinating activities of the SARS coronavirus papain-like protease. Mol. Cell 62, 572–585.

Bencez, K.Z., Kondapalli, K.C., Cook, J.D., McMahon, S., Millán-Pacheco, C., Pastor, N., Bécés, M., Rut, W., Kasperkiewicz, P., Mulder, M.P., Osva, H., Drag, M., Lima, C.D., Huang, T.T., 2015. SARS-CoV papain-like protease is a unique lys48 linkage-specific di-distributive deubiquitinating enzyme. Biochem. J. 468, 215–226.

Béz-Santos, Y.M., Mielec, A.M., Deng, X., Baker, S., Mesecar, A.D., 2014. Catalytic function and substrate specificity of the papain-like protease domain of np3 from the Middle East respiratory syndrome coronavirus. J. Virol. 88, 12511–12527.
Kuri, T., Eriksson, K.K., Putics, A., Züst, R., Snijder, E.J., Davidson, A.D., Siddell, S.G., Kuiken, T., Fouchier, R.A., Schutten, M., Rimmelzwaan, G.F., van Amerongen, G., van Li, C., Debing, Y., Jankevicius, G., Neyts, J., Ahel, I., Coutard, B., Canard, B., 2016a. Viral...

Koopes, K., Kikkt, M., den van Worm, S.H.E., Zevenhoven-Dobbe, J.C., van der Meer, Laing, S., Unger, M., Koch-Nolte, F., Haag, F., 2011. ADP-ribosylation of arginine.

Koziaczek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peric, T., Emery, S., Tong, U., Sugiyama, C., Comer, J.A., Lim, W., Rodin, P.E., Dowell, S.F., Ling, A.E., Humphrey, C.D., Shiue, W.J., Guarnier, J., Paddock, C.D., Rota, P., Fields, B., DeRisi, J., Yang, J.Y., Cox, N., Hughes, J.M., LeDuc, J.W., Bellini, W.J., Anderson, L.J., SARS Working Group, 2003. A novel coronavirus associated with severe acute respiratory syndrome. J. Virol. 77, 1665–1676.

Koonin, E.V., Gorbalenya, A.E., Purdy, M.A., Rozanov, M.N., Reyes, G.R., Bradley, D.W. 1992. Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis 

Koetse, A.J., Hansen, G., Boomaars-van der Zanden, A.L., Snijder, E.J., Gorbalenya, A.E., Enjuanes, L., Hilgenfeld, R., 2015. A G-quadruplex-binding motif from the MERS coronavirus reveals unusual, poten-

Koppes, P.H., de Jong, T., van Doornum, G., Lim, W., Ling, A.E., Chan, P.K., Group, 2003. A novel coronavirus associated with severe acute respiratory syndrome (SARS). Lancet 362, 263–270.

Kothe, S., Esselen, K.E., Putsch, A., Ziot, R., Snijder, E.J., Davidson, A.D., Siddell, S.G., Thiel, V., Ziebuhr, J., Weber, F., 2011. The ADP-ribose-1'-monophosphatase domains of severe acute respiratory syndrome coronavirus and human coronavirus 229E mediate resistance to antiviral interferon responses. J. Gen. Virol. 92, 1899–1905. Kusov, Y., Ikeda, F., Dikic, I., 2008. Atypical ubiquitin chains: new molecular signals. EMBO Rep. 9, 542–547.

Kusov, Y., Alvarez, E., Enjuanes, L., Hilgenfeld, R., 2015. A G-quadruplex-binding motif from the MERS coronavirus reveals unusual, poten-

Kvietys, P., Carnevale, E., Laﬀer, G., Thiel, H.I., Schwemmle-Wessels, C., Pohlmann, S., Haas, D., Drosten, C., van Brunn, A., 2011. The SARS-coronavirus-host interface: identifi-

Kusov, Y., von Brunn, B., Bairad, D.R., Hünten, S., Drosten, C., Hermeking, H., von Brunn, B., 2005. p53 down-regulates via E3 ubiquitin ligase RCHY1. Proc. Natl. Acad. Sci. U. S. A. 113, E5192–E5201.

Kusov, Y., von Brunn, B., Bairad, D.R., Hünten, S., Drosten, C., Hermeking, H., von Brunn, A., 2015. p53 down-regulates via E3 ubiquitin ligase RCHY1. Proc. Natl. Acad. Sci. U. S. A. 113, E5192–E5201.

Kuroda, M., Inoue, M., Iwama, K., Kondo, T., Nishigori, C., Tsuruo, T., 2008. The common interaction domains of SARS-CoV, H1N1 and influenza A virus nucleocapsid protein.

Kuiken, T., Fouchier, R.A., Schutten, M., Rimmelzwaan, G.F., van Amerongen, G., van Li, C., Debing, Y., Jankevicius, G., Neyts, J., Ahel, I., Coutard, B., Canard, B., 2016a. Viral...
protease: structure of a viral deubiquitinating enzyme. Proc. Natl. Acad. Sci. U. S. A. 103, 5717–5722.

Ratia, K., Kilianski, A., Baez-Santos, Y.M., Baker, S.C., Mesecar, A., 2014. Structural basis for the ubiquitin-linkage specificity and deubiquitination activity of SARS-CoV papain-like protease. PLoS Pathog. 10, e1004113.

Saikatendu, K.S., Joseph, B.W., Buchmeier, M.J., Stevens, R.C., Kuhn, P., 2005. Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1'-phosphate dephosphorylation by a conserved domain of nsp3. Structure 13, 1665–1675.

Serrano, P., Johnson, M.A., Almeida, M.S., Horst, R., Herrmann, T., Joseph, J.S., Neuman, B.W., Subramanian, V., Saikatendu, K.S., Buchmeier, M.J., Stevens, R.C., Kuhn, P., Wiithrich, K., 2007. Nuclear magnetic resonance structure of the N-terminal domain of nonstructural protein 3 from the severe acute respiratory syndrome coronavirus. J. Virol. 81, 12049–12060.

Serrano, P., Johnson, M.A., Chatterjee, A., Neuman, B.W., Joseph, B.W., Buchmeier, M.J., Kuhn, P., Wiithrich, K., 2009. Nuclear magnetic resonance structure of the nucleic acid-binding domain of severe acute respiratory syndrome coronavirus nonstructural protein 3. J. Virol. 83, 12998–13008.

Seth, R.B., Sun, L., Chen, Z.J., 2006. Antiviral innate immunity pathways. Cell Res. 16, 141–147.

Snijder, E.J., Breidenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, I.L., Guan, Y., Rozanov, M., Spaan, W.J., Borrelli, B., 2004. The 'SARS-unique domain' (SUD) of SARS coronavirus is an oligo(G)-binding protein. J. Mol. Biol. 331, 991–1004.

Snijder, E.J., van der Meer, Y., Zevenhoven-Dobbe, J., Onderwater, J.J., van de Meulen, J., Korten, H.K., Mommaas, A.M., 2006. Ultrastructure and origin of membrane vesicles associated with the severe acute respiratory syndrome coronavirus replication complex. J. Virol. 80, 5927–5940.

Spagnolo, J.F., Hogue, B.G., 2000. Host protein interactions with the 3' end of bovine coronavirus RNA and the requirement of the poly(A) tail for coronavirus defective genome replication. J. Virol. 74, 5053–5065.

Tan, J., Kusov, Y., Mutschall, D., Tech, S., Nagarajan, K., Hilgenfeld, R., Schmidt, C.L., 2007. The 'SARS-unique domain' (SUD) of SARS coronavirus is an oligo(G)-binding protein. Biochem. Biophys. Res. Commun. 364, 877–882.

Tan, J., Vornhein, C., Smart, D.S., Bricogne, G., Bollati, M., Kusov, Y., Hansen, G., Mesters, J.R., Schmidt, C.L., Hilgenfeld, R., 2009. The SARS unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. Proc. Natl. Acad. Sci. U. S. A. e1000428.

Tatar, G., Tok, T.T., 2016. Clarification of interaction mechanism of mouse hepatitis virus (MHV) N and nsp3 protein with homology modeling and protein–protein docking analysis. Curr. Comput. Aided Drug Des. 12, 98–106.

Tyrell, D.A., Byrose, M.L., 1965. Cultivation of a novel type of common-cold virus in organ cultures. Br. Med. J. 1, 1467–1470.

van der Hoek, L., Pyrc, K., Jebbink, M.F., Vermeulen-Oost, W., Arts, B.W., Wolthers, C.K., Wertheim-van Dillen, P.M., Kaandorp, J., Spaargaren, J., Berkhout, B., 2004. Identification of a new human coronavirus. Nat. Med. 10, 368–373.

van Hemert, M.J., von den Worm, S.H., Knoops, K., Mommaas, A.M., Borrelli, B., 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J. Virol. 79, 884–895.

Xu, Y., Cong, L., Chen, C., Wei, L., Zhao, Q., Xu, X., Ma, Y., Bartlam, M., Rao, Z., 2009. Crystal structures of two coronavirus ADP-ribose-1'-monophosphatases and their complexes with ADP-ribose: a systematic structural analysis of the viral ADRP domain. J. Virol. 83, 1083–1092.

Xu, D., Zhang, Y., 2012. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins 80, 1733–1747.

Yang, X., Chen, X., Bian, G., Tu, J., Xing, Y., Wang, Y., Chen, Z., 2014. Proteolytic processing, deubiquitinase and interferon antagonistic activities of Middle East respiratory syndrome coronavirus papain-like protease. J. Gen. Virol. 95, 614–628.

Yuan, W., Krug, R.M., 2001. Inhibition of type I interferon signaling by a coronavirus papain-like protease. J. Virol. 75, 33232–33236.

Ziebuhr, J., Thiel, V., 2007. Human coronavirus 229E papain-like proteases have overlapping proteolytic activities. J. Virol. 81, 12049–12060.

Ziebuhr, J., Manolaridis, I., van Kasteren, P.B., Kikkert, M., Snijder, E.J., Borrelli, B., 2010. Papain-like protease 1 from transmissible gastroenteritis virus: crystal structure and enzymatic activity toward viral and cellular substrates. J. Virol. 84, 10063–10073.

Woo, P.C., Lau, S.K., Cha, C.M., Chan, K.H., Tooi, H.W., Huang, Y., Wong, B.H., Poon, R.W., Cai, J.J., Luk, W.K., Poon, I.L., Wong, S.S., Guan, Y., Peiris, J.S., Yuan, K.Y., 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J. Virol. 79, 884–895.

Ziebuhr, J., Thiel, V., 2007. Human coronavirus 229E papain-like proteases have overlapping proteolytic activities. J. Virol. 81, 1813–1820.

Ziebuhr, J., Thiel, V., 2007. Human coronavirus 229E papain-like proteases have overlapping proteolytic activities. J. Virol. 81, 1813–1820.

Ziebuhr, J., Thiel, V., 2007. Human coronavirus 229E papain-like proteases have overlapping proteolytic activities. J. Virol. 81, 1813–1820.