



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.

Coronavirus Picornain-like Cysteine Proteinase

DATABANKS

MEROPS name: coronavirus picornain 3C-like peptidase-1

MEROPS classification: clan PA, subclan PA(C), family C30, peptidase C30.001

Species distribution: known only from murine hepatitis virus

Reference sequence from: murine hepatitis virus (UniProt: P19751)

MEROPS name: avian infectious bronchitis coronavirus 3C-like peptidase

MEROPS classification: clan PA, subclan PA(C), family C30, peptidase C30.002

Tertiary structure: Available

Species distribution: family Coronaviridae

Reference sequence from: avian infectious bronchitis virus (UniProt: Q0GNB9)

MEROPS name: human coronavirus 229E main peptidase

MEROPS classification: clan PA, subclan PA(C), family C30, peptidase C30.003

Tertiary structure: Available

Species distribution: known only from human coronavirus
Reference sequence from: human coronavirus (UniProt: Q05002)

MEROPS name: porcine transmissible gastroenteritis virus-type main peptidase

MEROPS classification: clan PA, subclan PA(C), family C30, peptidase C30.004

Tertiary structure: Available

Species distribution: family Coronaviridae

Reference sequence from: transmissible gastroenteritis virus (UniProt: Q9IW06)

MEROPS name: SARS coronavirus picornain 3C-like peptidase

MEROPS classification: clan PA, subclan PA(C), family C30, peptidase C30.005

IUBMB: EC 3.4.22.69 (BRENDA)

Tertiary structure: Available

Species distribution: family Coronaviridae

Reference sequence from: SARS coronavirus (UniProt: P59641)

Name and History

The existence of the coronavirus picornain-like cysteine proteinase was initially predicted by amino acid sequence comparison with chymotrypsin (Chapter 438) and with the picornavirus 3C proteinases (Chapter 403) [1], subsequently the proteinase was named *3C-like proteinase* (**3CLpro** or **3CLP**). To date, 3CLpro has been experimentally confirmed in all coronaviruses studied. Coronavirus taxonomy has identified the non-structural protein (nsp) domain of the coronavirus replicase polyprotein containing 3CLpro activity as nsp5, since the proteinase is the only known activity of the protein. 3CLpro is also referred to as Mpro (main proteinase); however, this is a subjective designation based on the number of cleavage events, rather than on a demonstrated greater importance for replication or pathogenesis. 3CLpro was also infrequently described in early literature by the observed or predicted molecular mass (*p27* or *p29* for MHV, *p34* for HCoV-229E, *p35* for IBV, and *p33* for SARS-CoV). The actual sizes of the 3CL proteinases have not been precisely determined by gel filtration, but have been calculated to be 33 kDa from the sequences. Because of the variation in sizes and evidence for differences in calculated and apparent masses, the 3CLpro, 3CLP, or nsp5 nomenclature is preferred.

Activity and Specificity

Coronavirus 3CLpro proteolytic activity was first confirmed in 1995 [2]. The determinants of 3CLpro cleavage site specificity appear to be highly conserved among coronaviruses [3,4]. 3CLpro almost exclusively cleaves following Gln, with only one known exception of His at a single cleavage site in HCoV-HKU1 [3,5]. There is a strong preference for the small non-charged residues Ser, Gly, Ala at P1', but Cys can also be tolerated [3]. There also is a very strong preference for Leu at P2, but some sites contain Phe, Met and Ile [3]. Much less specificity is found at P3, P4, P2' and P3'. Overall, the most common cleavage site found in all coronaviruses is PLQ↓(S/G/A/V) [2]. 3CLpro catalytic activity has been shown *in vitro* to be most active at the autoproteolytic N-terminal cleavage site [6]. Substitution of the Gln at P1 of IBV 3CLpro abolished proteolytic activity *in vitro* [7]. There are no other known or predicted cleavage site substrates for 3CLpro, and no cleavage of cellular proteins by 3CLpro has been reported.

Coronavirus 3CLpro activity *in vitro* occurs in *trans*, while *cis* autoproteolytic cleavage of a polyprotein precursor has not been detected [8]. Autoproteolytic cleavage of 3CLpro from the polyprotein occurs at an N-terminal cleavage site in IBV, MHV and HCoV-229E [7–9]. *In vitro* analyses of expressed and purified 3CLpro have

resulted in the conclusion that protease exists as a mixture of monomer and dimer forms with continual monomeric-dimeric exchange [10]. Experiments altering or deleting residues that result in solution or crystallographic obligate monomers incapable of dimerization are catalytically inactive *in vitro* [11]. Whether one or both monomers in a dimer are functional is controversial, and may be dependent on the termini, presence of flanking proteins and pH. The loss of function in a monomer has been associated with collapse of the active site cavity in one monomer and maintenance of an open conformation in the other [12]. Several groups have concluded from *in vitro* experiments that dimerization is absolutely required for 3CLpro activity [10]. However, a recent study has suggested that disruption of the dimerization determinants may not prevent N-terminal autoproteolytic cleavage [13].

Modification of the N-terminus or C-terminus by the addition of sequences or tags significantly reduces *in vitro* 3CLpro activity [5]. The calculated pI values for coronavirus 3CLpro range from 5.81 to 6.08. The experimental pH optimum for MHV 3CLpro activity is 7.0, although the enzyme remains active across a broad pH range from pH 5.5 and 8.0. Reducing agents are not required for cleavage by *E. coli*-expressed recombinant 3CLpro (r3CLpro), but also do not interfere with proteolytic activity; activity is maintained in the presence of 0.2–25 mM 2-ME and 0.1–100 mM DTT [14]. The MHV 3CLpro is sensitive to inhibition by inhibitors of both serine and cysteine proteinases, but is not affected by aspartic proteinase inhibitors [15]. MHV 3CLpro activity is inhibited in MHV-infected cells by leupeptin, PMSF, *N*-ethylmaleimide and E-64d [7]. MHV 3CLpro is also inhibited by high concentrations of ZnCl₂ (2 mM) and 1,10-phenanthroline (10 mM), but is unaffected by EDTA even at high concentrations (10 mM) [15]. Thus any role of metal ions in 3CLpro activity remains to be established.

Structural Chemistry

More than 17 solved X-ray structures of 3CLpro are available for at least five different coronaviruses (HCoV-229E, TGEV, HCoV-HKU1, SARS-CoV, and IBV) and with resolutions ranging from 2.8 Å to 1.6 Å. In addition, structures of SARS-CoV 3CLpro have been solved for deletion and substitution mutants, as well as for 3CLpro complexed with peptide substrates and potential inhibitors [10,11,16–18]. Crystals for all solved 3CLpro molecules are asymmetric units consisting of 1-to-3 3CLpro homodimers (Figure 546.1B). Structural data indicate greater conservation of tertiary and quaternary structures than of primary amino acid structure, even between distantly related coronaviruses. 3CLpro also contains a C-terminal

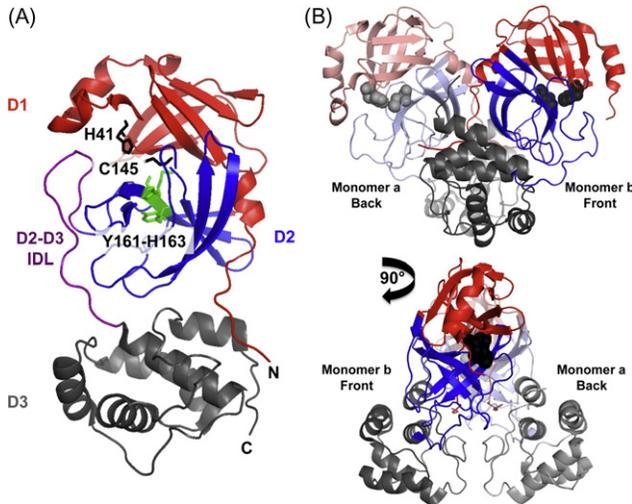


FIGURE 546.1 Three-dimensional structures of SARS 3CLpro monomer and dimers. A. Monomeric 3CLpro showing the three domains and interdomain loop region (D1, D2, D3, and D2-D3 IDL), amino and carboxy termini (N and C), the catalytic dyad residues (H41 and C145, black) and substrate binding residues (Y161–H163, green) are shown. The image was prepared from the Protein Data Bank entry (2H2Z) [25]. B. SARS 3CLpro dimers with domain 3 (gray) of monomer b oriented out of the page and monomer a oriented into the page. The image above shows the catalytic residues (black/gray) exposed distal from the dimerization interface. The image below exhibits the dimer above turned 90° out of the page. The images were prepared from the Protein Data Bank entry (2GT7) [37].

third domain (D3) of unclear function. Truncations of D3 reduced or abolished proteolytic activity of MHV and HCoV 3CLpro, while 67 amino acid truncations of the C-terminal domain were tolerated by IBV 3CLpro [15,19,20].

Coronavirus 3CLpro is synthesized as a domain within an 800 kDa polyprotein that is autoproteolytically cleaved to generate the mature 3CLpro and other viral replicase proteins. Mutagenesis of 3CLpro of MHV, IBV, HCoV-229E and FIPV has confirmed that His and Cys are essential active site residues [9,15,21,22]. In contrast, substitutions at residues consistent with catalytic Asp residues of other RNA virus 3Cpro or 3CLpro did not alter 3CLpro-mediated cleavage [9,15,21,22]. Thus coronavirus 3CLpro functions with a catalytic dyad of His and Cys. The substrate-binding S1-subsite domains for coronavirus 3CLpro may differ from those of other RNA viruses, with Tyr-X-His substituted for Gly-X-His and strongly constrained (Figure 546.1A) [3]. A conditional temperature-sensitive mutation (V148A) was identified behind the active site pocket of MHV 3CLpro. Predictive structural modeling suggested alterations in the rotameric position of a Met side-chain of the Tyr-Met-His sequence as a possible mechanism for temperature sensitivity [34].

Preparation

MHV 3CLpro has been isolated in an active form from virus-infected cells [8]. Active 3CLpro from MHV and IBV have also been expressed from multiple types of vectors including rabbit reticulocyte lysates and a variety of *in vitro* translation systems. MHV 3CLpro/maltose-binding protein (MBP/3CLpro) and His-tagged fusion products has also been expressed in *E. coli* [14,23]. However, studies have recently shown that modification of the N- and C-termini may affect protease activity [6]. The recombinant proteinase is active in *trans* after separation of the MBP from the 3CLpro at a factor Xa cleavage site [10,24]. More recent studies have engineered 3CLpro to have N- and C-terminal modifications, which may be removed before *in vitro* analysis [6,25]. Similarly, an HCoV-229E 3CLpro/ β -galactosidase fusion protein expressed in *E. coli* has proteolytic activity in an *in vitro*, *trans*-cleavage assay [9]. The FIPV 3CLpro has been expressed as a fusion with maltose-binding protein [4]. Rabbit polyclonal antibodies have been produced against 3CLpro from MHV, IBV, HCoV-229E, and SARS-CoV [8,9,21,26], and have been used for immunoprecipitation, Western blotting and purification of 3CLpro. There are no commercially available antibodies to these proteinases.

Biological Aspects

Coronaviruses contain a 27–32 kb, single-stranded, positive-sense RNA genome. During viral replication, the viral proteinases, polymerase, and other replicase proteins are translated from the input genome as a large (~800 kDa) polyprotein that contains multiple protein domains and 11 confirmed 3CLpro cleavage sites. The replicase polyprotein undergoes post-translational processing to liberate the active forms of the viral replicase proteins (Figure 546.2). One or two papain-like proteinases also are expressed as part of the coronavirus replicase polyprotein and mediate one, two, or three maturation cleavages [27–29] (see Chapter 494). Processing of the coronavirus polyprotein occurs continuously throughout the infection cycle, and 3CLpro is required for viral replication [30]. Within the polyprotein, 3CLpro (nsp5) is flanked nsp4 and nsp6, hydrophobic proteins with multiple transmembrane domains that are involved in membrane association and presentation of the proteinase and formation of replication complexes. Microsomal membranes are not required for activity of the mature 3CLpro from cells or *in vitro* translates from *E. coli*, or for activity of 3CLpro expressed with only small portions of nsp4 and nsp6 [8,9]. In contrast, expression of nsp5 with nsp4 and/or nsp6 confers a requirement for membranes for cleavage of nsp5 [7,31].

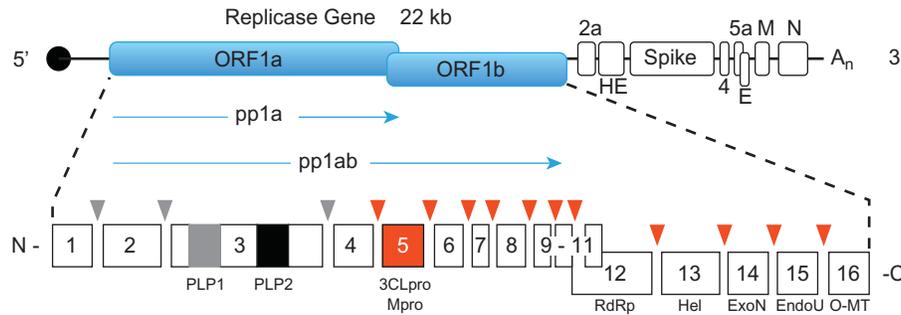


FIGURE 546.2 Schematic of the genome of Murine Hepatitis Virus (MHV) showing translated products and proteinase cleavage sites. The genome of MHV is 32 kb in size and consists of seven genes. The replicase gene (blue) consists of two open reading frames (ORF1a and ORF1b) offset by a -1 ribosomal frameshift and is translated to yield two polyproteins, pp1a and pp1ab, encoding non-structural proteins (nsps) 1 through 16. Maturation cleavages (arrows) are mediated by three viral proteinases, coronavirus papain-like proteases (PLP) 1 and 2 (gray and black, respectively) and the coronavirus 3C-like proteinase (nsp5, Mpro) (red). 3CLpro is responsible for 11 separate cleavage events throughout the replicase polyprotein. RdRp, polymerase; Hel, helicase; ExoN, exoribonuclease; EndoU, endoribonuclease; O-MT, O-methyl transferase.

Recently, several studies have shown that interactions within the coronavirus replication complex may be critical for 3CLpro activity. Several studies have reported that mutations in other replicase proteins have directly altered and impaired 3CLpro activity [32,33]. 3CLpro mutations also have been identified that significantly alter 3CLpro activity *in vitro* and during the course of an infection despite physical distance from the substrate binding pocket, catalytic residues, and known dimerization determinants [34,35]. While the global regulation of coronavirus replication and the significance of these findings remain to be determined, the development and use of robust reverse genetics systems and *in vitro* 3CLpro expression systems promises rapid progress in understanding these important proteinases.

Distinguishing Features

The coronavirus 3CLpro functions with a catalytic dyad of His and Cis. Whether the 3C endopeptidases of family C3 contain a catalytic triad has been controversial, but if so, then the coronavirus 3CLpro may have a different structure and catalytic mechanism from other 3C-like proteinases and may provide the nucleophile activity of a third residue in novel ways. The coronavirus 3CL proteinases contain a significantly larger number of amino acids C-terminal to the putative substrate-binding residues than other 3C or 3C-like proteinases. The extended region contains several predicted β sheet and α helical motifs and appears to be important for proteinase activity and dimerization. The helical C-terminal domain (D3) is unique to the protein databases and is known to contain key dimerization determinants. However, other functions of this domain have yet to be explored. Finally, the conservation of hydrophobic domains flanking 3CLpro in the coronavirus replicase polyproteins suggests they may serve a novel role in regulation of 3CLpro activity.

Further Reading

The articles of Ziebuhr *et al.* [3], and Hegyi *et al.* [4] provide extensive reviews of all aspects of the enzyme. The review article by Perlman *et al.* [36] expands on the role of 3CLpro in the context of general coronavirus replication and pathogenesis.

References

- [1] Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., Blinov, V.M. (1989). Coronavirus genome: prediction of putative functional domains in the nonstructural polyprotein K by comparative amino acid sequence analysis. *Nucleic Acids Res.* 17, 4847–4861.
- [2] Lu, Y., Lu, X., Denison, M.R. (1995). Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. *J. Virol.* 69, 3554–3559.
- [3] Ziebuhr, J., Snijder, E.J., Gorbalenya, A.E. (2000). Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. *J. Gen. Virol.* 81, 853–879.
- [4] Hegyi, A., Ziebuhr, J. (2002). Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.* 83, 595–599.
- [5] Woo, P.C., Huang, Y., Lau, S.K., Tsoi, H.W., Yuen, K.Y. (2005). In silico analysis of ORF1ab in coronavirus HKU1 genome reveals a unique putative cleavage site of coronavirus HKU1 3C-like protease. *Microbiol. Immunol.* 49, 899–908.
- [6] Grum-Tokars, V., Ratia, K., Begaye, A., Baker, S.C., Mesecar, A.D. (2008). Evaluating the 3C-like protease activity of SARS-Coronavirus: Recommendations for standardized assays for drug discovery. *Virus Res.* 133, 63–73.
- [7] Tibbles, K.W., Brierley, I., Cavanaugh, D., Brown, T.D.K. (1996). Characterization *in vitro* of an autocatalytic processing activity associated with the predicted 3C-like proteinase domain of the coronavirus avian infectious bronchitis virus. *J. Virol.* 70, 1923–1930.
- [8] Lu, X., Lu, Y., Denison, M.R. (1996). Intracellular and *in vitro* translated 27-kDa proteins contain the 3C-like proteinase activity of the coronavirus MHV-A59. *Virology* 222, 375–382.

- [9] Ziebuhr, J., Herold, J., Siddell, S.G. (1995). Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *J. Virol.* 69, 4331–4338.
- [10] Shi, J., Sivaraman, J., Song, J. (2008). Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. *J. Virol.* 82, 4620–4629.
- [11] Chen, S., Hu, T., Zhang, J., Chen, J., Chen, K., Ding, J., Jiang, H., Shen, X. (2008). Mutation of Gly-11 on the dimer interface results in the complete crystallographic dimer dissociation of severe acute respiratory syndrome coronavirus 3C-like protease: crystal structure with molecular dynamics simulations. *J. Biol. Chem.* 283, 554–564.
- [12] Chen, H., Wei, P., Huang, C., Tan, L., Liu, Y., Lai, L. (2006). Only one protomer is active in the dimer of SARS 3C-like proteinase. *J. Biol. Chem.* 281, 13894–13898.
- [13] Chen, S., Jonas, H., Shen, C., Hilgenfeld, R. (2010). Liberation of SARS-CoV main protease from the viral polyprotein: N-terminal autocleavage does not depend on the mature dimerization mode. *Protein Cell* 1, 59–74.
- [14] Sims, A.C., Lu, X.T., Denison, M.R. (1998). Expression, purification and activity of recombinant MHV-A59 3CLpro. *Adv. Exp. Med. Biol.* 440, 129–134.
- [15] Lu, Y.Q., Denison, M.R. (1997). Determinants of mouse hepatitis virus 3C-like proteinase activity. *Virology* 230, 335–342.
- [16] Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G.F., Anand, K., Bartlam, M., Hilgenfeld, R., Rao, Z. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. USA* 100, 13190–13195.
- [17] Tan, J., Verschueren, K.H., Anand, K., Shen, J., Yang, M., Xu, Y., Rao, Z., Bigalke, J., Heisen, B., Mesters, J.R., Chen, K., Shen, X., Jiang, H., Hilgenfeld, R. (2005). pH dependent conformational flexibility of the SARS-CoV main proteinase (M(pro)) dimer: molecular dynamics simulations and multiple X-ray structure analyses. *J. Mol. Biol.* 354, 25–40.
- [18] Zhong, N., Zhang, S., Zou, P., Chen, J., Kang, X., Li, X., Liang, C., Jin, C., Xia, B. (2008). Without its N-finger, the main protease of severe acute respiratory syndrome coronavirus can form a novel dimer through its C-terminal domain. *J. Virol.* 82, 4227–4234.
- [19] Ziebuhr, J., Heussipp, G., Siddell, S.G. (1997). Biosynthesis, purification, and characterization of the human coronavirus 229E 3CLike proteinase. *J. Virol.* 71, 3992–3997.
- [20] Ng, L.F. (2000). Further characterization of the coronavirus infectious bronchitis virus 3C-like proteinase and determination of a new cleavage site. *Virology* 272, 27–39.
- [21] Liu, D.X., Brown, T.D.K. (1995). Characterisation and mutational analysis of an ORF1a-encoding proteinase domain responsible for proteolytic processing of the infectious bronchitis virus 1a/1b polyprotein. *Virology* 209, 420–427.
- [22] Hegyi, A., Friebe, A., Gorbalenya, A.E., Ziebuhr, J. (2002). Mutational analysis of the active center of coronavirus 3C-like proteases. *J. Gen. Virol.* 83, 581–593.
- [23] Seybert, A., Ziebuhr, J., Siddell, S.G. (1997). Expression and characterization of a recombinant murine coronavirus 3C-like proteinase. *J. Gen. Virol.* 78, 71–75.
- [24] Lu, X., Sims, A., Denison, M.R. (1998). Mouse hepatitis virus 3C-like protease cleaves a 22-kilodalton protein from the open reading frame 1a polyprotein in virus-infected cells and *in vitro*. *J. Virol.* 72, 2265–2271.
- [25] Xue, X., Yang, H., Shen, W., Zhao, Q., Li, Q., Yang, K., Chen, C., Jin, Y., Bartlam, M., Rao, Z. (2007). Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J. Mol. Biol.* 366, 965–975.
- [26] Prentice, E., McAuliffe, J., Lu, X., Subbarao, K., Denison, M.R. (2004). Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins. *J. Virol.* 78, 9977–9986.
- [27] Baker, S.C., Yokomori, K., Dong, S., Carlisle, R., Gorbalenya, A.E., Koonin, E.V., Lai, M.M. (1993). Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus. *J. Virol.* 67, 6056–6063.
- [28] Dong, S., Baker, S.C. (1994). Determinants of the p28 cleavage site recognized by the first papain-like cysteine proteinase of murine coronavirus. *Virology* 204, 541–549.
- [29] Kanjanahaluethai, A., Baker, S.C. (2000). Identification of mouse hepatitis virus papain-like proteinase 2 activity. *J. Virol.* 74, 7911–7921.
- [30] Kim, J.C., Spence, R.A., Currier, P.F., Lu, X., Denison, M.R. (1995). Coronavirus protein processing and RNA synthesis is inhibited by the cysteine proteinase inhibitor, E64d. *Virology* 208, 1–8.
- [31] Pinon, J.D., Turner, J.D., Khan, F.S., Bonilla, P.J., Weiss, S.R. (1997). Efficient autoproteolytic processing of the MHV-A59 3CLike proteinase from the flanking hydrophobic domains requires membranes. *Virology* 230, 309–322.
- [32] Donaldson, E.F., Sims, A.C., Graham, R.L., Denison, M.R., Baric, R.S. (2007). Murine hepatitis virus replicase protein nsp10 is a critical regulator of viral RNA synthesis. *J. Virol.* 81, 6356–6368.
- [33] Fang, S., Shen, H., Wang, J., Tay, F.P., Liu, D.X. (2010). Functional and genetic studies of the substrate specificity of coronavirus infectious bronchitis virus 3C-like proteinase. *J. Virol.* 84, 7325–7336.
- [34] Sparks, J.S., Donaldson, E.F., Lu, X., Baric, R.S., Denison, M.R. (2008). A novel mutation in murine hepatitis virus nsp5, the viral 3C-like proteinase, causes temperature-sensitive defects in viral growth and protein processing. *J. Virol.* 82, 5999–6008.
- [35] Barrila, J., Bacha, U., Freire, E. (2006). Long-range cooperative interactions modulate dimerization in SARS 3CLpro. *Biochemistry* 45, 14908–14916.
- [36] Perlman, S., Netland, J. (2009). Coronavirus post-SARS; update on replication and pathogenesis. *Nat. Rev. Microbiol.* 7, 439–450.
- [37] Lee, T.W., Cherney, M.M., Liu, J., James, K.E., Powers, J.C., Eltis, L.D., James, M.N.G. (2007). Crystal structures reveal an induced-fit binding of a substrate-like Aza-peptide epoxide to SARS coronavirus main proteinase. *J. Mol. Biol.* 366, 916–932.

Christopher C. Stobart

*Department of Microbiology and Immunology, Vanderbilt University Medical Center, D6221, Medical Center North, Nashville, TN 37232-2581, USA.
Email: chris.stobart@vanderbilt.edu*

Mark R. Denison

Department of Pediatrics and Microbiology and Immunology, Elizabeth B. Lamb Center for Pediatric Research, Vanderbilt University Medical Center, D6217, Medical Center North, Nashville, TN 37232, USA. Email: mark.denison@vanderbilt.edu

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

© 2013 Elsevier Ltd. All rights reserved.
DOI: <http://dx.doi.org/10.1016/B978-0-12-382219-2.00544-5>