



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.

# The Autocatalytic Release of a Putative RNA Virus Transcription Factor from Its Polyprotein Precursor Involves Two Paralogous Papain-like Proteases That Cleave the Same Peptide Bond\*

Received for publication, May 7, 2001, and in revised form, June 26, 2001  
Published, JBC Papers in Press, June 28, 2001, DOI 10.1074/jbc.M104097200

John Ziebuhr<sup>‡§</sup>, Volker Thiel<sup>‡</sup>, and Alexander E. Gorbalenya<sup>¶||</sup>

From the <sup>‡</sup>Institute of Virology and Immunology, University of Würzburg, Versbacher Straße 7, 97078 Würzburg, Germany and <sup>¶</sup>Advanced Biomedical Computing Center, Science Application International Corporation/NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201

The largest replicative protein of coronaviruses is known as p195 in the avian infectious bronchitis virus (IBV) and p210 (p240) in the mouse hepatitis virus. It is autocatalytically released from the precursors pp1a and pp1ab by one zinc finger-containing papain-like protease (PLpro) in IBV and by two paralogous PLpros, PL1pro and PL2pro, in mouse hepatitis virus. The PLpro-containing proteins have been recently implicated in the control of coronavirus subgenomic mRNA synthesis (transcription). By using comparative sequence analysis, we now show that the respective proteins of all sequenced coronaviruses are flanked by two conserved PLpro cleavage sites and share a complex (multi)domain organization with PL1pro being inactivated in IBV. Based upon these predictions, the processing of the human coronavirus 229E p195/p210 N terminus was studied in detail. First, an 87-kDa protein (p87), which is derived from a pp1a/pp1ab region immediately upstream of p195/p210, was identified in human coronavirus 229E-infected cells. Second, *in vitro* synthesized proteins representing different parts of pp1a were autocatalytically processed at the predicted site. Surprisingly, both PL1pro and PL2pro cleaved between p87 and p195/p210. The PL1pro-mediated cleavage was slow and significantly suppressed by a non-proteolytic activity of PL2pro. In contrast, PL2pro, whose proteolytic activity and specificity were established in this study, cleaved the same site efficiently in the presence of the upstream domains. Third, a correlation was observed between the overlapping substrate specificities and the parallel evolution of PL1pro and PL2pro. Collectively, our results imply that the p195/p210 autoprocessing mechanisms may be conserved among coronaviruses to an extent not appreciated previously, with PL2pro playing a major role. A large subset of coronaviruses may employ two proteases to cleave the same site(s) and thus regulate the expression of the viral genome in a unique way.

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.

\* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Supported by Deutsche Forschungsgemeinschaft Grants SI 357/2-2, ZI 618/2-1, and ZI 618/2-2. To whom correspondence may be addressed. Tel.: 49-931-2013966; Fax: 49-931-2013934; E-mail: ziebuhr@vim.uni-wuerzburg.de.

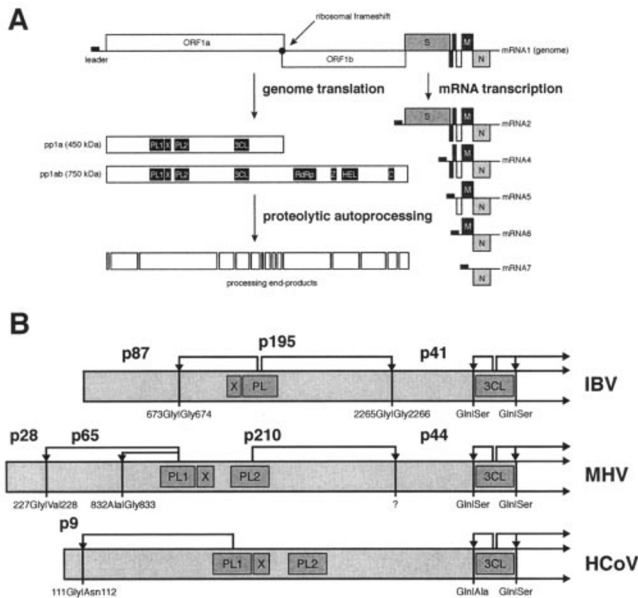
|| Supported by Grant NO1-CO-56000 from the NCI, National Institutes of Health. To whom correspondence may be addressed: Advanced Biomedical Computing Center, 430 Miller Dr., Rm. 228, SAIC/NCI-Frederick, Frederick, MD 21702-1201. Tel.: 301-846-1991; Fax: 301-846-5762; E-mail: gorbalen@ncifcrf.gov.

All positive-stranded RNA viruses infecting vertebrates, but also many other RNA viruses, employ proteolytic processing as the major regulation mechanism in virus genome expression. The virion RNA enters ribosomes and directs the synthesis of one or two multidomain protein precursors (polyproteins) that, in a controlled fashion, are proteolytically processed by viral and, sometimes, cellular proteases to produce intermediate and mature products. This processing proceeds in *cis* and in *trans* at interdomain junctions that contain the specific signals recognized by proteases (1–3).

We have been studying the protease-mediated regulation of viral gene expression using human coronavirus strain 229E (HCoV),<sup>1</sup> which belongs to the *Coronaviridae* family. Based on a similar polycistronic genome organization, common transcriptional and (post)-translational strategies, and a conserved array of nonstructural domains, the *Coronaviridae* have been united with the *Arteriviridae* in the order Nidovirales (Fig. 1A) (4, 5). With genome sizes of up to 32 kilobases, coronaviruses have the largest genomes among RNA viruses, whereas the related arteriviruses are much smaller (13–16 kilobases). The positive-stranded genomic RNA of nidoviruses contains 5'- and 3'-nontranslated regions as well as 6–12 ORFs that, in some cases, partially overlap each other. One of the most striking features of the nidovirus life cycle is the specific mode of genome transcription, which results in the synthesis of a 3'-coterminal, nested set of 4–8 sg mRNAs. Except for the smallest transcript, these sg mRNAs are structurally polycistronic (Fig. 1A), but generally, only the most 5'-proximal ORF is translated (reviewed in Ref. 6).

The two largest ORFs (ORF1a and ORF1b), which encompass the 5'-proximal two-thirds of the genome, are believed to encode all the protein functions required for nidovirus RNA synthesis (7, 8). The ORF1b-encoded polyprotein, which includes the putative RdRp activity and a recently established RNA helicase activity that is associated with a unique zinc finger structure (9, 10), is only produced if a ribosomal frameshift from ORF1a into ORF1b takes place during translation (11). This translational strategy is expected to yield two extremely large polyproteins, pp1a and pp1ab, of about 450 and 750 kDa, respectively. To date, pp1a and pp1ab have not been detected *in vivo*, most probably because they are cotranslationally and autocatalytically processed into numerous processing intermediates and mature nonstructural proteins. Both the

<sup>1</sup> The abbreviations used are: HCoV, human coronavirus 229E; sg, subgenomic; BCoV, bovine coronavirus; IBV, avian infectious bronchitis virus; MHV, murine hepatitis virus; TGEV, porcine transmissible gastroenteritis virus; MBP, maltose-binding protein; NJ, neighbor-joining; ORF, open reading frame; PLpro, papain-like protease; PVDF, polyvinylidene difluoride; PCR, polymerase chain reaction.



**FIG. 1. Outline of the HCoV life cycle and proteolytic processing of the N-terminal regions of coronavirus replicative polyproteins.** A, ORFs in the polycistronic genome are indicated as boxes. The replicase gene, encompassing ORFs 1a and 1b, the gene for the surface glycoprotein protein, *S*, the triple-spanning membrane protein, *M*, and the nucleocapsid protein, *N*, are shown. The filled rectangle at the 5' end of the genome represents the common leader sequence that is also present at the 5' end of the subgenomic mRNAs that are shown below the genome. The conserved domains/functions encoded by the replicase gene are shown in the boxes depicting the two replicative polyproteins (pp1a and pp1ab). B, the N-terminal regions of the IBV, MHV, and HCoV replicative polyproteins pp1a/pp1ab are shown with the previously identified processing products and the corresponding cleavage sites (P1 and P1' residues indicated). The following abbreviations are used: *PL*, papain-like protease; *PL1*, papain-like protease 1; *X*, domain conserved in coronaviruses, alphaviruses, rubiviruses, and hepatitis E virus (56); *PL2*, papain-like protease 2; *3CL*, 3C-like protease; *RdRp*, RNA-dependent RNA polymerase; *Z*, putative zinc finger; *HEL*, NTPase/RNA helicase; *C*, conserved domain specific for nidoviruses (4).

number and the origin of most of these proteins remain to be determined for coronaviruses (reviewed in Ref. 12).

The coronavirus pp1ab can be divided into an N-terminal region that is processed by one or two accessory papain-like proteases (Fig. 1B) and a C-terminal region that is processed by the main 3C-like cysteine protease (3CLpro) (12). The N-terminal region of pp1a/pp1ab spans from the initiator Met to the N terminus of the 3CLpro and consists of ~2800–3300 amino acids (Fig. 1B). Although IBV contains only one papain-like protease (PLpro), which is preceded by a conserved X domain, all other coronaviruses encode two paralogous and sequentially positioned papain-like proteases (PL1pro and PL2pro) that flank the X domain from both sides (13, 14). The IBV PLpro is part of an ~1550-amino acid protein (p195) that is autocatalytically released at flanking sites (15, 16). Cleavage at the N terminus of p195 produces p87, which is the N-terminal processing product of the IBV pp1a/pp1ab. In contrast, at least three proteins, p28, p65, and p210 (also known as p240), are produced from this region of pp1a/pp1ab in mouse hepatitis virus (MHV) (17–19). The MHV p210 protein, which is an ortholog of IBV p195, is autocatalytically released through cleavages mediated by PL1pro at the N-terminal site (20, 21) and PL2pro at the C-terminal site (22). PL1pro also cleaves the p28 ↓ p65 junction (23, 24) which, except for IBV, is conserved in all coronaviruses (25). Accordingly, a PL1pro-mediated cleavage at this site, resulting in the production of a small N-terminal protein (p9, p28 equivalent), was also detected in HCoV (25). The IBV PLpro cleavage sites flanking p195, the

MHV p28 ↓ p65, and p65 ↓ p210 PL1pro cleavage sites and the HCoV PL1pro cleavage site producing p9 were verified by site-directed mutagenesis and/or N-terminal protein sequencing (15, 16, 21, 25–27). Irrespective of the virus studied, the position in pp1a and the protease identity, all established and predicted coronavirus PLpro/PL1pro cleavage sites, contain a small amino acid (commonly Gly) at the P1 or P1' position, or at both positions.

Technically, coronavirus PLpros have only been characterized in surrogate systems, since the extreme size of the coronavirus RNA genome proved to be a serious obstacle to the development of straightforward reverse genetic approaches (28–30). In most cases, the *in vitro* results have been corroborated by the identification of corresponding cleavage products in coronavirus-infected cells, and thus, they are biologically relevant. Although the function of the N-terminal region of pp1a/pp1ab is not known, both the transcription-negative phenotype of an alphavirus X domain mutant (31) and the conservation of a transcription factor-like zinc finger in coronavirus PLpros (32) indicated that p195/p210 might be involved in coronavirus RNA synthesis. This hypothesis is strongly supported by a recent report in which the equine arteritis virus nonstructural protein 1, which, most probably, is a distant homolog of the coronavirus PLpros, is shown to be a transcriptional factor that is indispensable for sg mRNA synthesis (33).

In this study, we analyzed the mechanism of the coronavirus p195/p210 processing. We updated our previous alignment for the poorly conserved p195/p210 region (14) and found that p195/p210 (i) has a uniform domain organization and (ii) is flanked by cleavage sites that are conserved in all coronaviruses, including IBV. Contrary to the current belief that IBV encodes only one PLpro, we show here that IBV, like other coronaviruses, may in fact encode two PLpro domains as follows: a proteolytically defective remnant of PL1pro and an active PL2pro, currently known as PLpro. We then confirmed the identity of the N-terminal site in HCoV and demonstrated that this site to be cleaved by either of the two PLpros, indicating that these proteases may be (partly) redundant. The ability of PL1pro to cleave the cognate site was found to be considerably down-regulated by flanking sequences that included PL2pro. In contrast, PL2pro cleaved the same site more efficiently in the presence of the upstream sequences. The combined data suggest that the regulation of coronavirus genome expression may include a unique autoproteolytic mechanism that recruits two paralogous proteases to cleave the same site.

#### EXPERIMENTAL PROCEDURES

**Computer-aided Comparative Sequence Analyses**—Amino acid sequences were derived from the Genpeptides data base. Sequence alignments were produced using the ClustalX program (34), the Dialign2 program (35), and the Macaw workbench (36). Non-redundant sequence data bases were searched with single sequences (37), and with Hidden Markov models trained on multiple sequence alignments using the HMMER 2.0.1 package (38). Upon protein comparisons, the Blossum62 (39) was used as the scoring inter-residues table. The obtained alignments were also sent as inputs for the PhD program (40, 41) to predict secondary structures and transmembrane helices. Cluster phylogenetic trees were reconstructed using the neighbor-joining (NJ) algorithm of Saitou and Nei (42) with the Kimura correction (43) and were evaluated with 1000 bootstrap trials, as implemented in the ClustalX program. Parsimonious trees were generated through exhaustive search and evaluated with bootstrap analysis using a UNIX version of the PAUP\* 4.0.0d55 program (44) that is included in the GCG-Wisconsin Package programs (Genetics Computer Group, Madison, WI). Trees were prepared and modified using the TreeView program (45).

**Virus and Cells**—The methods for HCoV propagation in MRC-5 cells (ECACC 84101801) and concentration of virus with polyethylene glycol have been described previously (46).

**Preparation of Antiserum  $\alpha$ -H2**—The HCoV ORF 1a nucleotide sequence coding for the pp1a/pp1ab amino acids 112–322 was amplified

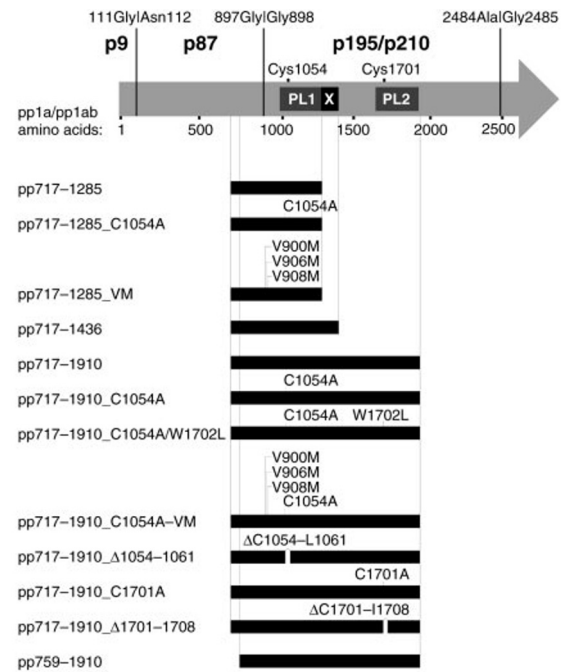


by PCR from pBS-J12E6 plasmid DNA (47) using primers 134 and 135. The upstream primer contained a *Bam*HI restriction site, and the downstream primer contained a translation stop codon followed by a *Pst*I restriction site. The PCR product was digested with *Bam*HI and *Pst*I and ligated with *Bam*HI/*Pst*I-digested pMal-c2 DNA (New England Biolabs, Frankfurt, Germany). The resulting plasmid, pMal-H2, encoded the specified ORF1a amino acids fused to the *Escherichia coli* maltose-binding protein (MBP). The plasmid was used to transform competent *E. coli* TB1 cells, and the bacterial fusion protein was expressed and purified as described previously (46, 48). The HCoV-specific polypeptide, which contained 211 amino acids of pp1a/pp1ab and is preceded by six N-terminal vector-derived amino acids, was released from MBP by cleavage with endoprotease Xa (Amersham Pharmacia Biotech) and used to immunize rabbits as described previously (46). The resulting antiserum was designated  $\alpha$ -H2.

**Metabolic Labeling, Cell Lysis, and Immunoprecipitation**—Infection or mock infection of MRC-5 cells was done essentially as described previously (49). Briefly,  $3 \times 10^6$  MRC-5 cells were mock-infected or infected with HCoV at a multiplicity of 10 plaque-forming units per cell. Radioactive labeling of newly synthesized proteins was done for 2.5 h, between 7 and 9.5 h postinfection, with 100  $\mu$ Ci of L-[ $^{35}$ S]methionine per ml. Before labeling, the cells were washed twice with methionine-free minimal essential medium supplemented with 2% dialyzed fetal bovine serum. The cells were lysed in 1 ml of lysis buffer (46). One hundred microliters of cell lysate was mixed with 400  $\mu$ l of immunoprecipitation buffer (46) and 5  $\mu$ l of preimmune serum or 5  $\mu$ l of  $\alpha$ -H2 serum. After 60 min at 4 °C, 25  $\mu$ l of protein A-Sepharose CL-4B (P9424; Sigma) was added to isolate the immune complexes, which were washed and eluted as described previously (46). The immunoprecipitated proteins were analyzed by SDS-polyacrylamide gel electrophoresis in a 10–17% gradient gel and autoradiography.

**Expression of pp1a/pp1ab Amino Acids by *in Vitro* Translation**—Previously, the HCoV PL2pro coding sequence has been found to be non-clonable in *E. coli* (47). We therefore used PCR-based methods to express this region of the HCoV genome. If not otherwise specified, we used a DNA template that had been isolated from a recombinant vaccinia virus, vHCoV-inf-1, carrying a complete cDNA copy of the HCoV genome (30). The nucleotide sequences of all PCR products used for *in vitro* RNA synthesis were determined to exclude any PCR-derived nucleotide misincorporations. The amino acid sequences of the proteins analyzed in this study are summarized in Fig. 2, and the primers used to generate appropriate DNA templates for *in vitro* RNA synthesis are given in Table I. To produce proteins pp717–1285, pp717–1436, pp717–1910, and pp759–1910, the coding sequences of the HCoV pp1a/pp1ab amino acids 717–1285, 717–1436, 717–1910, and 759–1910 were amplified by PCR using the primer pairs 111/103, 111/105, 111/107, and 110/107, respectively. The upstream primers (110 and 111, respectively) contained a T7 RNA polymerase promoter followed by an initiator Met codon and the downstream primers (103, 105, and 107, respectively) contained a translation stop codon. By using the purified PCR products as templates, capped RNAs were synthesized *in vitro* by use of a Riboprobe T7 system (P1440 and P1711; Promega, Mannheim, Germany) and subsequently translated in reticulocyte lysate (L4960, Promega) in the presence of [ $^{35}$ S]methionine as described previously (48). After 40 min, the translation reactions (15- $\mu$ l mixtures) were stopped by the addition of 1.7  $\mu$ l of 10 $\times$  translation stop mix (0.1 mg of RNase A per ml, 10 mg of cycloheximide per ml, 5 mM [ $^{35}$ S]methionine), and the mixtures were divided into 2 aliquots. One of the aliquots was stored at –80 °C, and the other one was further incubated at 30 °C for 120 min. Finally, 0.2  $\mu$ l of each reaction aliquot was analyzed by SDS-polyacrylamide gel electrophoresis and autoradiography. To obtain quantitative data on the extent of substrate conversion, the radioactivities incorporated into the full-length substrate and the C-terminal cleavage product were determined using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA) equipped with ImageQuant 1.1 software. The data obtained were adjusted to the number of methionines present in the respective proteins, and the calculations were done essentially as described by Teng *et al.* (50).

**Codon and Deletion Mutagenesis**—To generate pp717–1285\_C1054A, the coding sequence of pp1a/pp1ab amino acids 717–1285 was amplified by PCR using primers 111 and 103. The PCR product was digested with *Nco*I and *Eco*RI and ligated into the *Nco*I-*Eco*RI site of plasmid pBST (51). The resulting plasmid, pBST-111–103, was subjected to site-directed mutagenesis by *in vivo* recombination PCR (48, 52) using primers 112 and 113. The resulting plasmid was designated pBST-111–103\_C1054A. Both pBST-111–103\_C1054A DNA and the parental pBST-111–103 DNA were linearized with *Eco*RI and used as templates for RNA synthesis.



**FIG. 2. Schematic representation of the N-terminal HCoV pp1a/pp1ab region and the proteins tested for proteolytic activity.** The numbering of the amino acids is according to Ref. 47. The putative nucleophilic cysteine residues of PL1pro and PL2pro are indicated. The positions of the PL1pro p9 ↓ p87 cleavage site (25) and two additional PLpro cleavage sites (this study) are given. The proteins tested for proteolytic activity are depicted on the left, and the black lines designate these proteins in relation to their positions in pp1a and pp1ab. See “Experimental Procedures” for a description of the generation of each expression construct.

To generate pp717–1285\_VM, an *in vivo* recombination PCR was done using primers 188 and 189 and pBST-111–103 DNA as a template. The resulting plasmid, pBST-111–103\_VM, was linearized with *Eco*RI and used as a template for RNA synthesis. pp717–1285\_VM contained the pp1a/pp1ab amino acids 717–1285 in which each of the three valine residues, Val<sup>900</sup>, Val<sup>906</sup>, and Val<sup>908</sup>, was replaced with methionine (V900M, V906M, and V908M).

To generate pp717–1910\_C1054A-VM, an *in vivo* recombination PCR was done using primers 112 and 113 and pBST-111–103\_VM plasmid DNA as a template. The resulting plasmid, pBST-111–103\_C1054A-VM, served then as a template to amplify nucleotides 2441–3912 using primers 111 and 139. In a separate reaction, nucleotides 3881–6022 were amplified from vHCoV-inf-1 genomic DNA by using primers 165 and 107. The two PCR products were digested with *Bsa*I, purified, and ligated together by using T4 DNA ligase. The ligated product was then used as a template for second round PCR amplification with outside primers 111 and 107. The resulting 3,614-base pair PCR product was used as a template for RNA synthesis.

To generate pp717–1910\_C1054A/W1702L, nucleotides 2441–3912 were amplified from pBST-111–103\_C1054A plasmid DNA by using primers 111 and 139, and nucleotides 3881–6022 were amplified from vaccinia virus vF10 DNA<sup>2</sup> by using primers 165 and 107. The recombinant vaccinia virus vF10 contained a cDNA copy of the HCoV ORF1a in which codon 1702, TGG, has been changed to TTG. The two PCR products were digested with *Bsa*I and ligated together with T4 DNA ligase. The ligated product was then used as a template for second round PCR amplification with primers 111 and 107. The RNA derived from the purified PCR template encoded the pp1a/pp1ab amino acids 717–1910 in which active-site residues of both PL1pro and PL2pro have been replaced (C1054A and W1702L, respectively).

To generate pp717–1910\_C1054A, nucleotides 2006–3451 and 3454–6022 were amplified by PCR from vHCoV-inf-1 genomic DNA in two separate reactions by using the primer pairs 137/212 and 213/39, respectively. The two products were digested with *Bsa*I and ligated together with T4 DNA ligase. The ligated product was then used as a template for second round PCR amplification with primers 111 and 107.

<sup>2</sup> V. Thiel, unpublished observations.

TABLE I  
Oligonucleotides used in this study for amplification or mutagenesis of HCoV sequences

Oligonucleotide	Oligonucleotide sequence (5' to 3') <sup>a</sup>	Nucleotide no. <sup>b</sup>	Polarity	Use
134	<u>AAAAGGATCCA</u> TGTGACATACACTGACCAGTATCTC	626–652	Forward	PCR
135	<u>AAA</u> ACTGCAGTTACTGAAGAGCAATCACTCTCCAACACAGAG	1231–1258	Reverse	PCR
110	<b><u>TATAATACGACTCACTATAGGG</u></b> CACCATGAAACGTTGGCTTGGCATTITTTGCTAG	2567–2591	Forward	PCR
111	<b><u>TATAATACGACTCACTATAGGG</u></b> CACCATGGTTATAGATGGAGGCCAAAATATGGAAAT	2441–2467	Forward	PCR
103	<u>TGAGGAATTC</u> TATTGGTAGAAAGCTACATTGTCATG	4124–4147	Reverse	PCR
105	<u>TCCCGAATTC</u> TACTACTAAACCAGACACAAAATCCTT	4577–4600	Reverse	PCR
107	<u>TTATGGTTAGG</u> TTTCACTGTATTAACAGG	5926–6022	Reverse	PCR
112	TTGGATAACA <u>ATGC</u> ATGGGTTAACTCAGTTATGTTACA	3440–3477	Forward	Codon mutagenesis
113	TGAGTTAACCC <u>ATGC</u> ATGGTTATCCAATGTTTGAGTATC	3427–3466	Reverse	Codon mutagenesis
188	<u>ATGTCATTTTCTGATGACAT</u> GGAAATGAAAGACATTGAAC	2990–3029	Forward	Codon mutagenesis
189	<u>CATTTCCATGTCATCAGAAAATGACATTTTACCACCAGC</u>	2978–3016	Reverse	Codon mutagenesis
139	TGGTAATGACCACAAGAGACCCGACCACGAAAATTTGATG	3873–3912	Reverse	PCR
165	TTTCGTGGTGC <u>GGTCTCT</u> TGTGGTCATTACCAGACT	3881–3916	Forward	PCR
210	<u>ATATGGTCTCATGG</u> TTGTTATCCAATGTTTGAGTATCT	3426–3451	Reverse	Deletion mutagenesis
211	<u>ATATGGTCTCAAC</u> CAAAATACAATTAACCTGGTATACTTGA	3476–3501	Forward	Deletion mutagenesis
212	<u>TATAGGTCTCAGGC</u> GTTGTTATCCAATGTTTGAGTATC	3427–3451	Reverse	Codon mutagenesis
213	<u>TATAGGTCTCAGC</u> CTGGGTTAACTCAGTTATGTTACAA	3454–3478	Forward	Codon mutagenesis
214	<u>TATAGGTCTCAATGC</u> ATTATTAATCAGTITTTTAAACAC	5369–5392	Reverse	Codon mutagenesis
215	<u>TATAGGTCTCAGCAT</u> GGGTGAATGCTGTTGATTGCAC	5396–5420	Forward	Codon mutagenesis
167	<u>ACCCGGTCTCA</u> ATTATTATCACTAGTTTTTAACTCTAATACC	5360–5393	Reverse	Deletion mutagenesis
168	<u>ACCCGGTCTCT</u> TAATGCACTACAGTATTCGAAACCCCATTTTATTTTC	426–5448	Forward	Deletion mutagenesis
137	<u>TTTTTCCATGG</u> GCTACTTCCGCTTATGGCCAGTCCAA	2006–2033	Forward	PCR
39	AATCAAAGA <u>ACTTCT</u> GTGCC	6046–6065	Reverse	PCR

<sup>a</sup> The underlined sequences were required for cloning, mutagenesis, and expression of HCoV sequences. The boldface sequences present in oligonucleotides 110 and 111 indicate the T7 RNA polymerase promoter.

<sup>b</sup> HCoV nucleotide numbers correspond to the sequence published by Herold *et al.* (47).

The RNA derived from the purified PCR template encoded the pp1a/pp1ab amino acids 717–1910 in which the PL1pro catalytic Cys residue has been replaced with Ala (C1054A).

To generate pp717–1910\_Δ1054–1061, nucleotides 2006–3451 and 3476–6022 were amplified by PCR from vHCoV-inf-1 genomic DNA in two separate reactions by using the primer pairs 137/210 and 211/39, respectively. The two products were digested with *Bsa*I and ligated together with T4 DNA ligase. The ligated product was then used as a template for second round PCR amplification with primers 111 and 107. The RNA derived from the purified PCR template encoded the pp1a/pp1ab amino acids 717–1053 and 1062–1910, *i.e.* residues 1054–1061 have been removed from PL1pro.

To generate pp717–1910\_C1701A, nucleotides 2006–5392 and 5396–6022 were amplified by PCR from vHCoV-inf-1 genomic DNA in two separate reactions by using the primer pairs 137/214 and 215/39, respectively. The two products were digested with *Bsa*I and ligated together with T4 DNA ligase. The ligated product was then used as a template for second round PCR amplification with primers 111 and 107. The RNA derived from the purified PCR template encoded the pp1a/pp1ab amino acids 717–1910 in which the PL2pro catalytic Cys residue has been replaced with Ala (C1701A).

To generate pp717–1910\_Δ1701–1708, nucleotides 2006–5393 and 5426–6022 were amplified by PCR from vHCoV-inf-1 genomic DNA in two separate reactions by using the primer pairs 137/167 and 168/39, respectively. The two products were digested with *Bsa*I and ligated together with T4 DNA ligase. The ligated product was then used as a template for second round PCR amplification with primers 111 and 107. The RNA derived from the purified PCR template encoded the pp1a/pp1ab amino acids 717–1700 and 1709–1910, *i.e.* residues 1701–1708 have been removed from PL2pro.

**N-terminal Protein Sequence Analysis**—The proteins pp717–1285\_VM and pp717–1910\_C1054-VM were produced by *in vitro* translation in the presence of [<sup>35</sup>S]methionine as described above. After incubation of the translation reactions for 160 min, the products were separated by electrophoresis in SDS-polyacrylamide gels and transferred electrophoretically to polyvinylidene difluoride (PVDF) membranes (162–0180, Bio-Rad). The areas of the membranes containing the C-terminal cleavage products were identified by autoradiography and isolated. The bound proteins were then subjected to 16 cycles of Edman degradation by use of a pulsed liquid protein sequencer (ABI 467A, Applied Biosystems, Inc., Weiterstadt, Germany). The eluate from each cycle was mixed with scintillation mixture, and the radioactivity was measured.

## RESULTS

*The Coronavirus p195/p210 Proteins Are Flanked by Conserved PLpro Cleavage Sites and Share a Conserved Five-*

*domain Organization*—The N-terminal part of the replicative polyproteins pp1a/pp1ab of coronaviruses is poorly conserved (14, 47). The proteolytic domains responsible for the processing of this region are part of the largest pp1a/pp1ab cleavage product, known as p195 in IBV and p210 in MHV. These proteins are autocatalytically processed by non-identical mechanisms that, in the case of p195, involve a single PLpro activity and, in the case of p210, both the PL1pro and PL2pro activities (15, 16, 20, 22). Because of these differences, it remained unknown whether the N and C termini of p195/p210 are conserved in coronaviruses. Also, although we suspected that the IBV PLpro may be an ortholog to PL2pro (13, 14), its relationship with the pair of PLpros conserved in all other coronaviruses remained unresolved. By using software for generating global alignments (ClustalX) and local alignments (Dialign2 and Macaw), we have produced a coronavirus-wide multiple sequence alignment which included p195/p210 together with flanking sequences (Fig. 3). The conserved features identified by this alignment include two cleavage sites at the N and C termini and five domains in the order Ac, PL1pro, X, PL2pro and Y, where Ac is the N-terminal domain and Y is the C-terminal domain as defined in this study (see below).

Block A is the only conserved sequence block that was identified upstream of PL1pro. This block is part of a newly recognized domain that varies in size among the different coronaviruses (~150–240 amino acids). Because this domain is highly enriched in acidic Asp and Glu residues, it was named Ac (acidic domain). PL1pro was previously identified in all coronaviruses except IBV. Our alignment shows that IBV may in fact encode a deviant form of PL1pro, which, like its counterparts in other coronaviruses, is located between the Ac and X domains and contains a conserved sequence around the catalytic Cys residue (block B). We consider this domain to be enzymatically defective because, in contrast to other coronaviruses, the other conserved sequences essential for proteolytic activity (*e.g.* catalytic His residue and zinc finger) are missing in the IBV sequence. The assignment of this domain as a PL1pro remnant is consistent with the PL2pro-like features observed for the IBV PLpro. Thus, both the coronavirus PL2pros and the IBV PLpro occupy similar positions in pp1a/



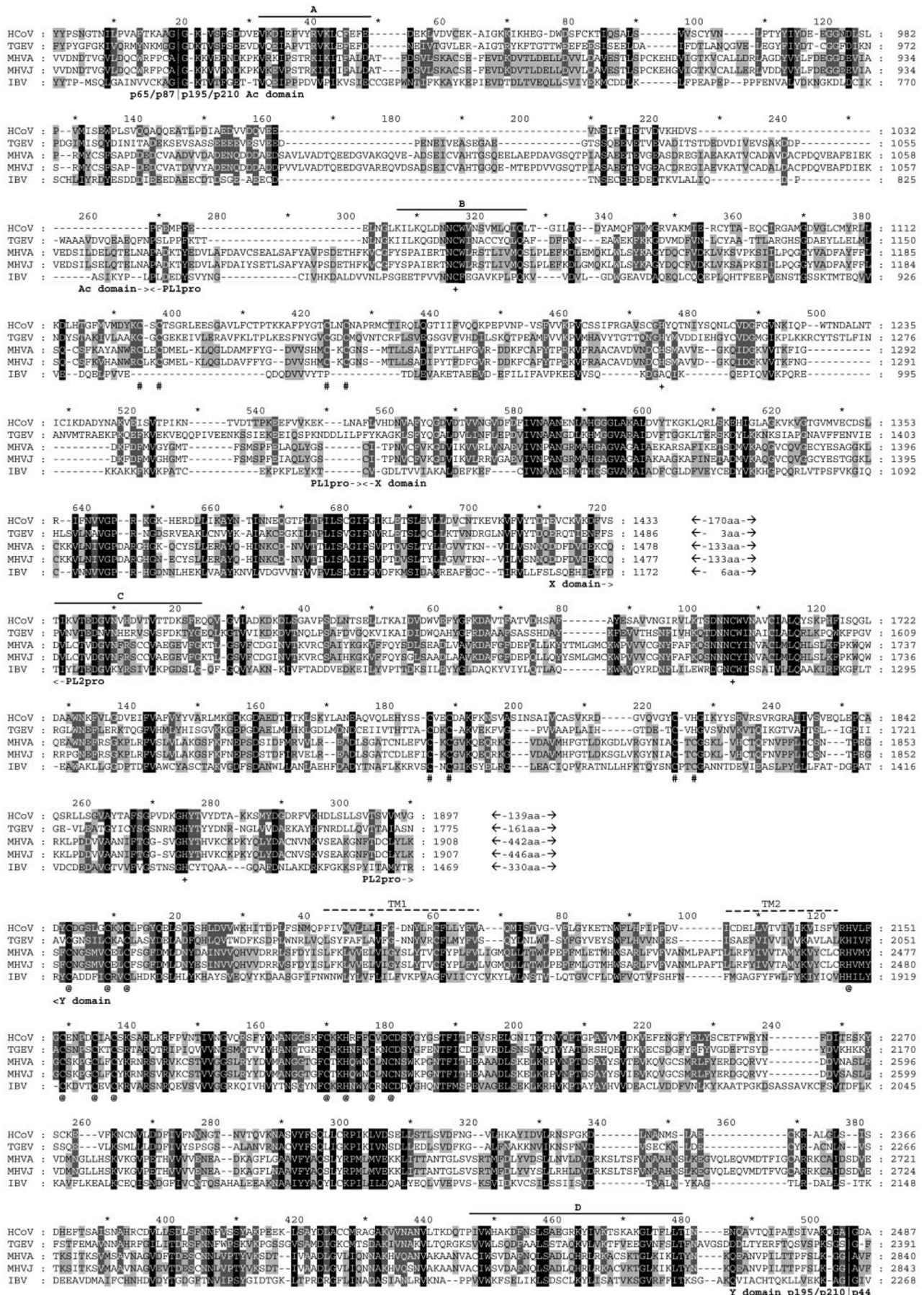


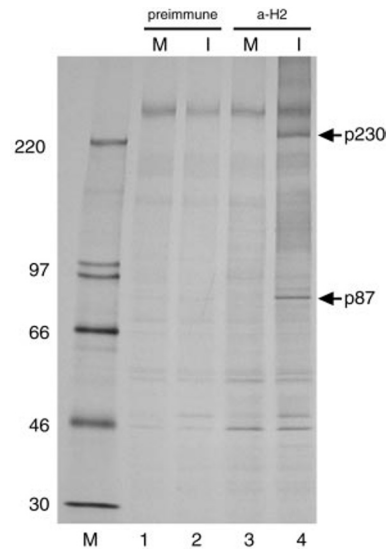
FIG. 3. Multiple sequence alignment of the p195/p210 regions of coronavirus replicase polypeptides. An initial draft of this alignment was generated using the Dialign2 program (35) and subsequently improved with the ClustalX program (34). The alignment was further checked

pp1ab; they are embedded between nonconserved regions of variable sizes that, on the upstream side, separate the X domain from PL2pro/PLpro and, more downstream, PL2pro/PLpro from the Y domain. Furthermore, immediately upstream of the catalytic Cys residue, both the PL2pros and PLpro (but not the PL1pros), share a moderately conserved region of ~80 amino acids, which includes the particularly conserved block C. These observations suggest that the p195/p210 proteins of all coronaviruses have a uniform domain organization and include two PLpros. In IBV, one of these proteases (PL1pro) is proteolytically defective and the other one (PL2pro) is proteolytically active. (Henceforth, we use PL2pro rather than PLpro to refer to the proteolytically active papain-like protease of IBV).

The largest conserved domain identified in this study encompasses a region of ~450–490 amino acids at the C terminus of all coronavirus p195/p210 proteins. It was named Y domain. This domain contains two highly hydrophobic stretches and 11 conserved Cys/His residues (14), all in the N-terminal ~180 amino acids. These structural features lead us to predict that the Y domain may anchor p195/p210 into membranes and bind  $Zn^{2+}$  or similar metal ions. Overall, the multidomain organization of p195/p210 implies that the protein is multifunctional.

The conservation of the cleavage sites flanking p195/p210 was recognized on the basis of the immediate proximity of these sites to sequence blocks whose identification proved to be statistically rigorous. The cleavage sites previously identified at the p65 ↓ p210 junction in MHV (Ala ↓ Gly) and the p87 ↓ p195 junction in IBV (Gly ↓ Gly) matched one another in a region upstream of the conserved sequence block A (Fig. 3). In HCoV and porcine transmissible gastroenteritis virus (TGEV), previously uncharacterized Gly-Gly dipeptides with similar positions were identified as putative cleavage sites. Likewise, all coronaviruses have putative cleavage sites of similar composition (Gly ↓ Gly, Gly ↓ Ala, Ala ↓ Gly, or Ser ↓ Gly) in the vicinity of the C terminus of block D. Strikingly, after this analysis was performed, one of these predicted sites, Gly ↓ Gly, proved to be cleaved by PL2pro at the p195 ↓ p41 junction in IBV (16). Similarly, the preliminary mapping data obtained recently for the MHV PL2pro p210 ↓ p44 cleavage site (22) are compatible with the location of the predicted Gly ↓ Ala scissile bond.

In this study, we were specifically interested in relating the implications of the sequence analysis to the understanding of the autocatalytic release mechanisms of p195/p210 in HCoV, which have not been characterized to date. In particular, we were intrigued by the observation that the cleavage at the N terminus of p195/p210 is apparently mediated by different paralogous proteases in two coronaviruses, namely by PL2pro in IBV and PL1pro in MHV (see above; Fig. 1B). Furthermore, this site was not cleaved by the HCoV PL1pro in different *in vitro* assays, although the enzyme was shown to be active at the p9 ↓ 87 junction in mono- and bimolecular assays (25, 32). In light of this striking variation, we reasoned that the compre-



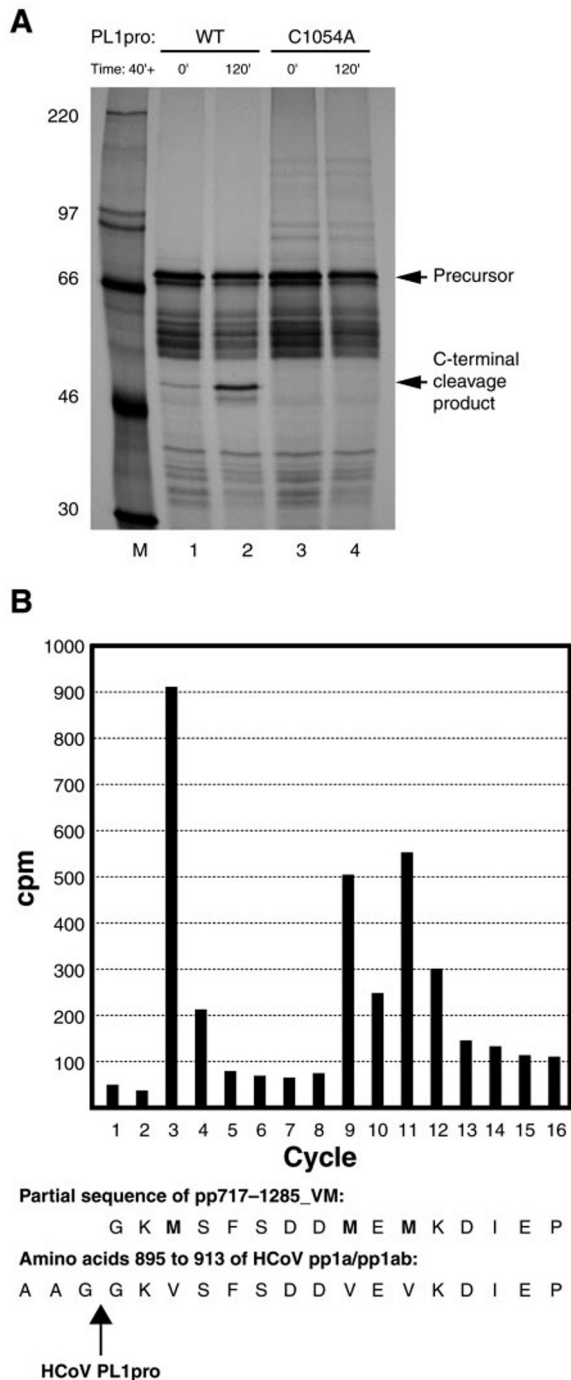
**FIG. 4. Detection of an ORF1a-encoded 87-kDa cleavage product in HCoV-infected cells.** Metabolically labeled lysates from  $3 \times 10^5$  mock-infected (*M*) (lanes 1 and 3) or HCoV-infected (*I*) (lanes 2 and 4) MRC-5 cells were analyzed by SDS-polyacrylamide gel electrophoresis in a 10–17% acrylamide gradient gel after immunoprecipitation with  $\alpha$ -H2 antiserum (lanes 3 and 4) or the corresponding preimmune serum (lanes 1 and 2). Antiserum  $\alpha$ -H2 recognizes the HCoV ORF1a-encoded amino acids 112–322. The cells were labeled from 7 to 9.5-h postinfection with  $100 \mu\text{Ci}$  of [ $^{35}\text{S}$ ]methionine per ml. Sizes of molecular mass markers (CFA 626; Amersham Pharmacia Biotech) with masses in kilodaltons as well as the 230- and 87-kDa processing products, p230 and p87, respectively, are indicated.

hensive characterization of this cleavage in HCoV should be especially informative, and we did the following experiments.

**Size and Origin of a 87-kDa Protein Identified in HCoV-infected Cells Are Compatible with Cleavage at the Predicted N Terminus of p195/p210**—The results of the comparative sequence analysis described above led us to predict a second processing product in the N-terminal proximal region of HCoV pp1a/pp1ab. We expected this protein to be derived from a pp1a/pp1ab region immediately downstream to the previously identified HCoV p9 polypeptide (25) and upstream of the putative HCoV p195/p210. This protein is predicted to be released through cleavages at Gly<sup>111</sup> ↓ Asn<sup>112</sup> and Gly<sup>897</sup> ↓ Gly<sup>898</sup> and would have a calculated molecular mass of 87,345. To test this prediction, we first generated a polyclonal antiserum,  $\alpha$ -H2, specific for the N-terminal region of the predicted protein (pp1a/pp1ab amino acids Asn<sup>112</sup> to Gln<sup>322</sup>). The antiserum was used to immunoprecipitate ORF1a-encoded polypeptides from HCoV-infected MRC-5 cells. The results of this experiment, shown in Fig. 4, revealed two major proteins that had apparent molecular masses of 87 (p87) and 230 kDa (p230). The proteins were specifically precipitated by immune serum from metabo-

and corrected using results of a Macaw-mediated (36) analysis that involved all coronaviruses except MHVJ, which was excluded due to its closeness to MHVA. Five domains were recognized in the alignment, and their positions were indicated with ><. The borders of the domains are tentative. The alignments of the PL1pro and PL2pro regions were based on results of our previous analysis (32). For two regions that are located between domains X and PL2pro, and PL2pro and Y, respectively, no consistent alignments have been produced. Therefore, only the sizes of these regions are indicated. The pp1a position of the rightmost residue in an alignment row is indicated at the right side. The shading of individual residues in the alignment was done according to a four-level conservation; *black background and white letters*, *gray background and white letters*, *gray background and black letters*, respectively, indicate residues that are conserved in 100, 80, and 60% of the sequences. Groups of conserved amino acids are as follows: IVLM; FYW; KRH; DNQE; ST; AG. According to the Macaw, four blocks, which are labeled with letters from A to D above the alignment and are discussed in the main text are statistically significant for the entire pp1a searching space: A,  $p = 1.1e^{-002}$ ; B,  $p = 2.1e^{-002}$ ; C,  $p = 4.2e^{-006}$ ; and D,  $p = 3.9e^{-015}$ . Two hydrophobic regions predicted to be trans-membrane domains (40) are marked with *dashed lines* and denoted with TM1 and TM2, respectively. Other highlights are as follows: +, catalytic Cys and His residues of PLpros; #, postulated metal-chelating Cys and His residues of the PLpro zinc fingers; @, conserved Cys and His residues of domain Y; |, cleavage sites of PLpros. MHVA and MHVJ, MHV strains A59 and JHM. The National Center for Biotechnology Information sequence ID: IBV, 138147; HCoV, 464694; TGEV, 872319; MHVA, 453423 (nucleotide); MHVJ, 266958 (corrected according to Ref. 57); all sequences are for proteins unless otherwise specified.





**FIG. 5. Evidence for a second PL1pro cleavage site in the HCoV pp1a and pp1ab polyproteins.** *A*, *in vitro* translation reactions of capped RNAs derived from *EcoRI*-linearized plasmids pBST-111–103 and pBST-111–103\_C1054A. The respective RNAs encode the pp1a/pp1ab amino acids 717–1285 (WT, lanes 1 and 2) or the same sequence with an active-site replacement of the catalytic Cys<sup>1054</sup> (C1054A, lanes 3 and 4). The translation reactions were done as described under “Experimental Procedures,” and the reaction products were either analyzed directly (lanes 1 and 3) or after further incubation for 120 min (lanes 2 and 4). The positions of full-length precursor proteins and cleavage products are indicated. *B*, a protein called pp717–1285\_VM was translated in a reticulocyte lysate in the presence of [<sup>35</sup>S]methionine. Except for three amino acid substitutions (V900M, V906M, and V908M), which had been introduced downstream to the presumed cleavage site, this protein contained the HCoV pp1a/pp1ab wild-type sequence from residues 717 to 1285. The translation reaction was incubated for 160 min at 30 °C, and the reaction products were separated on an SDS-12.5% polyacrylamide gel. After electrophoretic transfer to PVDF membranes, the position of the C-terminal cleavage product was determined by autoradiography. The isolated protein was subjected to 16 cycles of Edman degradation, and the distribution of

ically labeled lysates of HCoV-infected cells (Fig. 4, lane 4). They did not react with preimmune serum and were not present in mock-infected cells (Fig. 4, lanes 1–3). Taking into account the specificity of antiserum  $\alpha$ -H2 and the size of the protein, the data are fully consistent with the initial model that p87 represents a pp1a/pp1ab processing product that is released from the polyprotein by cleavages at (or near) the predicted sites, Gly<sup>111</sup> ↓ Asn<sup>112</sup> and Gly<sup>897</sup> ↓ Gly<sup>898</sup>.

Unlike p87, the origin of p230 remains uncertain from the obtained data. To resolve whether p230 is a precursor protein of p87 or whether it was coprecipitated by the  $\alpha$ -H2 antiserum due to specific interactions with p87, more experiments involving antisera with new specificities are to be performed.

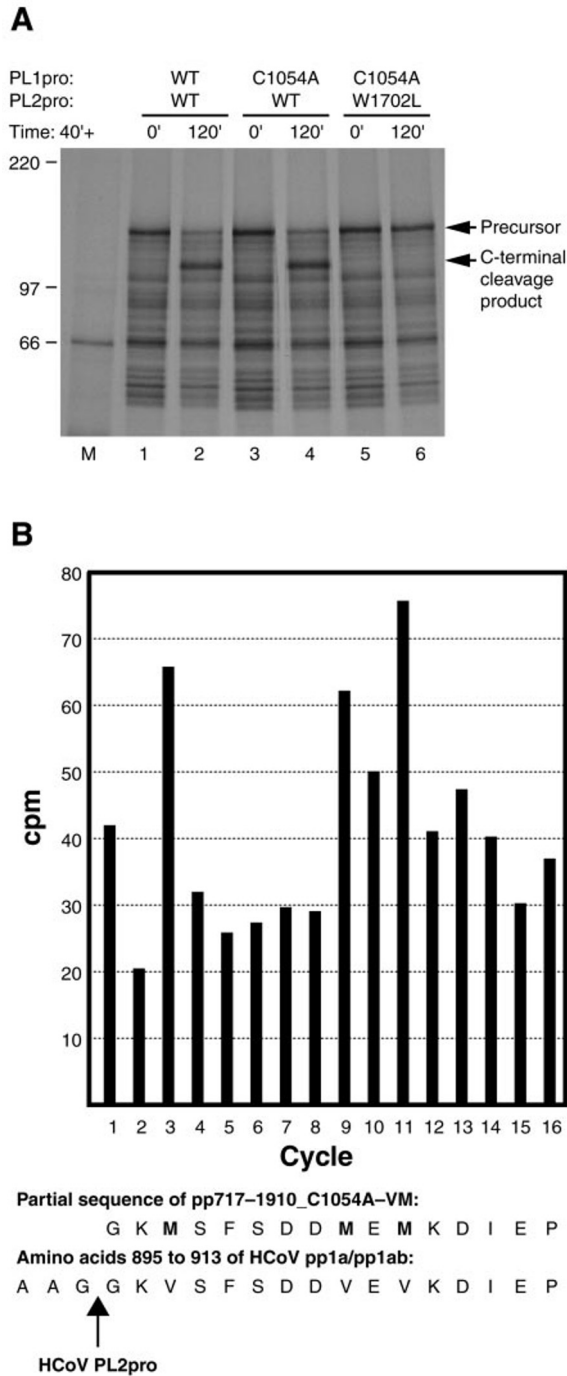
**Approach to Analyze the Processing at the Predicted p87 ↓ p195/p210 Junction *In Vitro***—Next, we analyzed the processing at the p87 ↓ p195/p210 junction *in vitro*. In the past, proteins encompassing the entire p87 or a large portion of it at the N terminus and PL1pro at the C terminus were not processed at the predicted p87 ↓ p195/p210 cleavage site in HCoV *in vitro* (25, 32). The observed stability of the precursors might be due to the inability of PL1pro to process this site or an incorrect assignment of this HCoV site by the computer-assisted analysis (Fig. 3). Alternatively, the PL1pro-mediated cleavage at this site could be inhibited by p87, as was previously observed for another coronavirus, MHV (20). (It should be noted that these aspects have not been discussed in the original studies (25, 32) as they are brought to light by the computer-based analysis described in Fig. 3).

To exclude the potential negative effect of p87 on the p87 ↓ p195/p210 processing in this study, we expressed proteins that contained only small fragments of p87 (~140–180 amino acids) immediately adjacent to the predicted C terminus of this protein. By PCR, we produced four DNAs that contained a T7 RNA polymerase promoter, a Met initiator codon, and different 3′-extensions encoding the pp1a/pp1ab amino acids Val<sup>717</sup> to Pro<sup>1910</sup> or truncated versions of it (Fig. 2). In total, four basic constructs and eight mutated variants were designed. All constructs shared the predicted HCoV pp1a/pp1ab Gly<sup>897</sup> ↓ Gly<sup>898</sup> cleavage site, preceded by a small domain and followed by the Ac-PL1pro domains. In three of the basic constructs, this minimal sequence was extended to include either the X domain alone or both the X and PL2pro domains. Furthermore, two basic constructs, pp717–1285 and pp717–1910, were subjected to site-directed mutagenesis, for example, to inactivate the PL1pro and PL2pro domains or to allow radiosequence analyses of [<sup>35</sup>S]methionine-labeled cleavage products (for details on the constructs, see “Experimental Procedures” and Fig. 2). By using these PCR templates, capped RNAs were generated and translated in reticulocyte lysates to characterize the cleavage at the predicted junction by SDS-polyacrylamide gel electrophoresis and N-terminal protein sequence analysis.

**HCoV PL1pro Can Cleave the Conserved Site at the Predicted p87 ↓ p195/p210 Junction**—First, the involvement of PL1pro in the cleavage of the p87 ↓ p195/p210 junction was addressed. To this end, the proteolytic processing of pp717–1285, a protein that contained PL1pro but lacked the X and PL2pro domains, was characterized. We analyzed the processing of the wild-type protein and a mutant, pp717–1285\_C1054A, in which the catalytic Cys-1054 of PL1pro was replaced with Ala. Previously, this mutation was proved to block the PL1pro-mediated cleav-

radiolabeled amino acids was determined by scintillation counting. The amino acid sequence of pp1a and pp1ab from positions 895 to 913 is shown. The amino acids Met<sup>900</sup>, Met<sup>906</sup>, and Met<sup>908</sup> present in pp717–1285\_VM are shown in **boldface type**, and the newly identified PL1pro cleavage site is indicated by an arrow.





**FIG. 6. Proteolytic activity of the PL2pro domain and identification of its cleavage site.** A, the HCoV ORF1a-encoded amino acids 717-1910 were translated in rabbit reticulocyte lysates in the presence of [<sup>35</sup>S]methionine. The proteins to be tested for proteolytic activity contained wild-type sequence (*lanes 1 and 2*) or the same sequence with active-site replacements in PL1pro (C1054A, *lanes 3 and 4*) and both PL1pro and PL2pro (C1054A and W1702L), respectively. The proteins were translated at 30 °C for 40 min, and after the termination of translation, the reaction products were either analyzed directly (*lanes 1, 3, and 5*) or after further incubation for 120 min (*lanes 2, 4, and 6*). The analysis was done by SDS-polyacrylamide gel electrophoresis in a 10-17% polyacrylamide gradient gel. The positions of full-length precursor proteins and cleavage products are indicated. B, a protein called pp717-1910\_C1054A-VM was translated in a reticulocyte lysate in the presence of [<sup>35</sup>S]methionine. Except for a PL1pro-inactivating amino acid replacement (C1054A) and three additional substitutions (V900M, V906M, and V908M), which had been introduced downstream to the predicted cleavage site, this protein contained the HCoV pp1a/1ab wild-type sequence from residues 717 to 1910. The translation reaction was incubated for 160 min at 30 °C, and the reaction products were separated on an SDS-10% polyacrylamide gel. After electro-

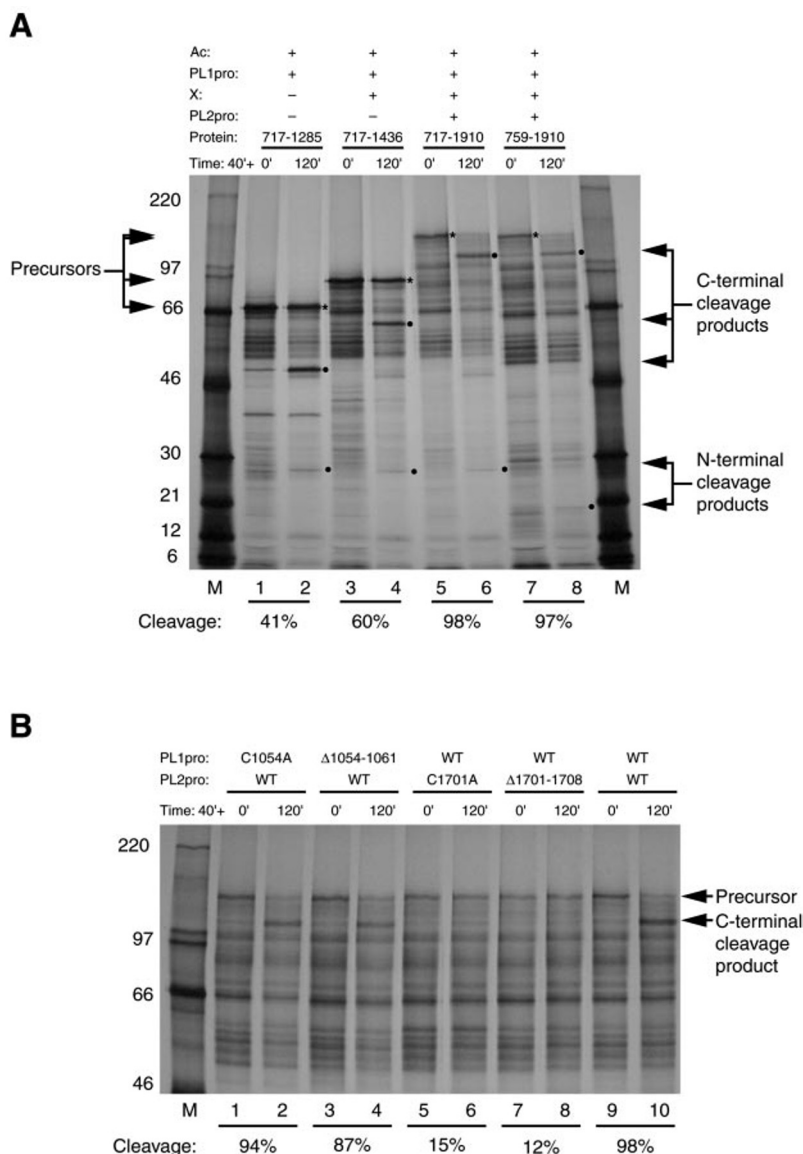
age of the p9 ↓ p87 junction (25). Upon *in vitro* translation of RNAs encoding pp717-1285 and pp717-1285\_C1054A, respectively, numerous products were detected after 40 min. The most prominent protein had an apparent molecular mass of ~70 kDa, which corresponded well to the expected size of the primary translation product (Fig. 5A, *lanes 1 and 3*). In the pp717-1285 sample, another protein of ~51 kDa was clearly detectable. This protein was not identified in the pp717-1285\_C1054A translation reaction (Fig. 5A, compare *lanes 1 and 2* with *lanes 3 and 4*). The 51-kDa protein was detectable as early as 40 min after translation initiation and became increasingly prominent after translation termination and further incubation of the translation products for 120 min at 30 °C (Fig. 5A, *lanes 1 and 2*). The data suggest that the 51-kDa protein may represent a processing product of pp717-1285, and the size of the protein is consistent with the expected size of the C-terminal pp717-1285 processing product if cleavage occurred at Gly<sup>897</sup> ↓ Gly<sup>898</sup>. Subsequently, also the N-terminal processing product of pp717-1285 was identified (see Fig. 7A). Taken together, we concluded from this experiment that PL1pro can cleave the Gly<sup>897</sup> ↓ Gly<sup>898</sup> bond or a nearby site.

As the sequence alignment in Fig. 3 shows, the predicted HCoV p87 ↓ p195/p210 cleavage site, Gly<sup>897</sup> ↓ Gly<sup>898</sup>, is preceded by two alanine residues and is not well conserved in other coronaviruses. Thus, alternative assignments of the scissile bond within the Ala-Ala-Gly-Gly sequence, which would be compatible with our current understanding of the coronavirus PLpro substrate specificities and the results of the above experiment, could not be ruled out. Consequently, an N-terminal radiosequence of the C-terminal cleavage product was performed to determine the scissile bond precisely. Because the predicted N-terminal Gly<sup>898</sup> is immediately followed by three Val residues at positions 900, 906, and 908, we initially attempted, by using either [<sup>3</sup>H]valine or [<sup>14</sup>C]valine in the translation reactions, to incorporate radiolabel into pp717-1285. However, these efforts failed to incorporate sufficient label for radiosequence analyses, and therefore, we decided to analyze a cleavage product in which these three Val residues were replaced with Met (V900M, V906M, and V908M). *In vitro* synthesis of the mutated precursor, pp717-1285\_VM, in the presence of [<sup>35</sup>S]methionine revealed that the Met-for-Val substitutions were compatible with PL1pro-mediated autoprocessing (data not shown), and thus, we were able to isolate a sufficiently labeled C-terminal cleavage product. The data we obtained in the subsequent sequence analysis conclusively showed that the PL1pro-mediated cleavage occurs at the Gly<sup>897</sup> ↓ Gly<sup>898</sup> peptide bond (Fig. 5B), confirming our previous prediction.

**HCoV PL2pro Can Cleave the Conserved Site at the Predicted p87 ↓ p195/p210 Junction**—To address a possible role of PL2pro in the cleavage of the Gly<sup>897</sup> ↓ Gly<sup>898</sup> bond, two pp717-1910 mutants, which contained PL1pro and PL2pro, were characterized. In the first mutant, pp717-1910\_C1054A, the PL1pro active-site nucleophile, Cys<sup>1054</sup>, was replaced with Ala. This substitution completely inactivated the HCoV PL1pro activity toward the p9 ↓ p87 (25) and p87 ↓ p195/p210 (Fig. 5A) junctions. Surprisingly, we found that the pp717-1910\_C1054A protein was as efficiently processed as its wild-

phoretic transfer to PVDF membranes, the position of the C-terminal cleavage product was determined by autoradiography. The isolated protein was subjected to 16 cycles of Edman degradation, and the distribution of radiolabeled amino acids was determined by scintillation counting. The amino acid sequence of pp1a and pp1ab from positions 895 to 913 is shown. The amino acids Met<sup>900</sup>, Met<sup>906</sup>, and Met<sup>908</sup> present in pp717-1910\_C1054A-VM are shown in *boldface type*, and the deduced PL2pro cleavage site is indicated by an arrow.

**FIG. 7. Relationship between PL1pro and PL2pro proteolytic activities.** *A*, *in vitro* translation reactions of capped RNAs encoding the pp1a/1ab amino acids 717–1285 (*lanes 1 and 2*), amino acids 717–1436 (*lanes 3 and 4*), amino acids 717–1910 (*lanes 5 and 6*), and amino acids 759–1910 (*lanes 7 and 8*). The proteins to be tested for proteolytic activity were translated in rabbit reticulocyte lysates in the presence of [<sup>35</sup>S]methionine at 30 °C for 40 min. After the termination of translation, the reaction products were either analyzed directly (*lanes 1, 3, 5, and 7*) or after further incubation for 120 min (*lanes 2, 4, 6, and 8*). The analysis was done by SDS-polyacrylamide gel electrophoresis in a 10–17% polyacrylamide gradient gel. Full-length precursor proteins and major processing products are indicated (\*, precursor protein; ●, processing product). Also, the calculated cleavage activities of the full-length precursor proteins are given (see “Experimental Procedures” for details). *B*, proteolytic activities of pp717–1910-derived proteins carrying active-site mutations in the two HCoV PLpro domains. The proteins to be tested for proteolytic activity were translated in rabbit reticulocyte lysates in the presence of [<sup>35</sup>S]methionine at 30 °C for 40 min. After termination of translation, the reaction products were separated by SDS-polyacrylamide gel electrophoresis. They were either analyzed directly (*lanes 1, 3, 5, 7, and 9*) or after further incubation for 120 min (*lanes 2, 4, 6, 8, and 10*). The proteins, which all encompassed the HCoV pp1a/pp1ab amino acids 717–1910, contained Cys-to-Ala replacements of the putative nucleophilic residues of PL1pro (C1054A, *lanes 1 and 2*) and PL2pro (C1701A, *lanes 5 and 6*), 8-amino acid deletions including the putative nucleophilic residues of PL1pro ( $\Delta$ 1054–1061, *lanes 3 and 4*) and PL2pro ( $\Delta$ 1701–1708, *lanes 7 and 8*) or wild-type sequence (*lanes 9 and 10*). The positions of full-length precursor proteins and cleavage products are indicated. Also the calculated cleavage activities of the full-length precursor proteins are given.



type parent pp717–1910. Both precursors were cleaved to produce proteins with apparent molecular masses of ~120 kDa (Fig. 6A, *lanes 1–4*). This result indicated that another (non-PL1pro-mediated) activity may be responsible for cleavage of the pp717–1910\_C1054A protein at the p87 ↓ p195/p210 junction. To verify that PL2pro is the protease that mediates this cleavage, the double mutant pp717–1910\_C1054A/W1702L was analyzed. In this mutant, both the PL1pro and PL2pro domains were inactivated by active-site replacements: PL1pro by a replacement of the active-site nucleophile Cys<sup>1054</sup> with Ala, and PL2pro by a Leu substitution for the highly conserved Trp<sup>1702</sup>, which is immediately adjacent to the active-site nucleophile Cys<sup>1701</sup>. The mutated protein proved to be proteolytically inactive (Fig. 6A, *lanes 5 and 6*), confirming that the observed cleavage is indeed associated with the activity of PL2pro.

Based on the data described above, it was reasonable to believe that PL2pro, like PL1pro, cleaves the (same) Gly<sup>897</sup>–Gly<sup>898</sup> bond. Alternatively, PL1pro and PL2pro might use partly overlapping sites (for example, in the Ala<sup>895</sup>–Ala–Gly–Gly<sup>898</sup> sequence) or the two proteases might cleave separate but adjacent sites in the viral polyprotein. To establish the specificity of PL2pro unequivocally, we determined the newly identified HCoV PL2pro cleavage site by protein sequencing,

using the same approach as described above for the determination of the PL1pro cleavage site structure. We produced a [<sup>35</sup>S]methionine-labeled derivative of pp717–1910\_C1054A, in which each of the Val<sup>900</sup>, Val<sup>906</sup>, and Val<sup>908</sup> residues was replaced with Met (pp717–1910\_C1054A-VM). These replacements did not affect the PL2pro-mediated processing pattern of the primary translation product (data not shown). The radiosequence analysis of the C-terminal pp717–1910\_C1054A-VM processing product revealed that PL2pro cleaves the pp1a/pp1ab Gly<sup>897</sup> ↓ Gly<sup>898</sup> peptide bond (Fig. 6B). Hence, our combined data show that the HCoV PL1pro and PL2pro domains cleave the same site in the viral polyprotein *in vitro*.

**PL2pro Dominates Over PL1pro in the Cleavage of the Gly<sup>897</sup>–Gly<sup>898</sup> Peptide Bond**—The above findings imply that, by cleavage of the same site, both PL1pro and PL2pro are able to mediate the autoproteolytic release of the protein of which they are part. To gain initial insight into how these activities might be coordinated, the effects of the individual domains in the PL1pro-X-PL2pro constellation on the efficiency of the Gly<sup>897</sup> ↓ Gly<sup>898</sup> cleavage were analyzed.

We initially investigated how the X domain or the combination of X and PL2pro affect the PL1pro-mediated cleavage at the p87 ↓ p195/p210 junction. The data shown in Fig. 7A indicate that, irrespective of whether or not the X domain was



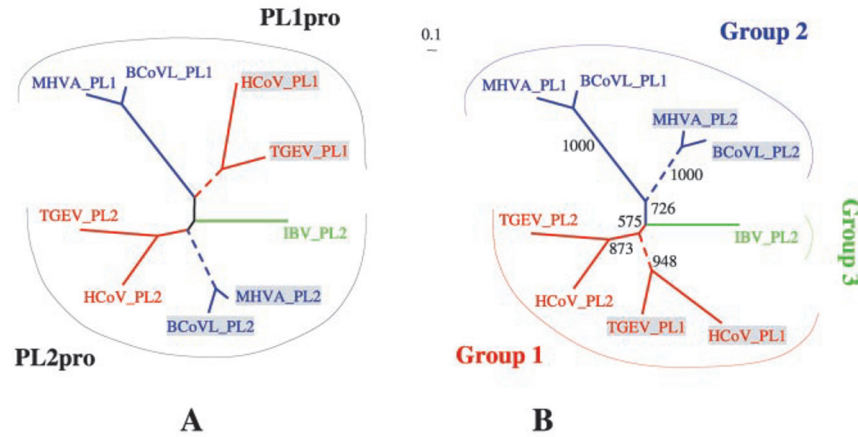


FIG. 8. **Phylogenetic relationships among coronavirus PLpros.** A, expected, and B, reconstructed and unrooted trees for coronavirus PLpros. The genetic groups recognized in coronaviruses are uniquely colored. A, the topology of this tree of PLpros was approximated on the basis of sequence comparisons of other coronavirus replicative enzymes<sup>3</sup> that result in a tree topology identical to the one shown here for the PL1pro and PL2pro subgroups. This representation is called the “expected” tree. The two lineages that were reshuffled in B, MHVA/BCoVL\_PL2 and TGEV/HCoV\_PL1, are shown with a *gray background*. B, the “reconstructed” tree was generated using an alignment of PLpros (Fig. 3; see also Fig. 2A in (32)) and the NJ algorithm with the Kimura correction as implemented in the ClustalX program. The alignment included the following sequences: HCoV\_PL1pro (1043–1227), HCoV\_PL2pro (1690–1885), TGEV\_PL1pro (1082–1266), TGEV\_PL2pro (1577–1763), MHVA\_PL1 (1110–1292), MHVA\_PL2 (1705–1896), BCoV\_PL1 (1063–1245), BCoV\_PL2 (1660–1851) and IBV\_PL2 (1263–1457). BCoV is from Footnote 3; for the sources of the other sequences see Fig. 3. The number of trees, in which a particular bifurcation was sustained in the course of 1000 bootstrap simulations, is given at each node.

present, the PL1pro-mediated cleavage progressed slowly, that is even after 160 min, significant amounts of the pp717–1285 and pp717–1436 primary translation products remained uncleaved (Fig. 7A, lanes 2 and 4). In contrast, the C-terminally extended pp717–1910 and pp759–1910 proteins, which both contained PL1pro, X, and PL2pro, were (almost) completely cleaved after the same incubation time (Fig. 7A, lanes 6 and 8). This observation suggests that either (i) PL1pro and PL2pro act in concert or (ii) PL2pro takes over the activity toward the Gly<sup>897</sup> ↓ Gly<sup>898</sup> site from PL1pro to cleave this site more rapidly, which then leads to nearly complete substrate conversion within the given period.

To address the latter issue, we characterized a series of pp717–1910 derivatives in which either PL1pro or PL2pro was selectively inactivated. As shown in Fig. 7B (lanes 1–4), inactivation of the proteolytic activity of PL1pro (either by substitution of the catalytic Cys residue alone (pp759–1910\_C1054A) or by deletion of the highly conserved predicted  $\alpha$ -helix of which the Cys nucleophile is part (pp759–1910\_Δ1054–1061)) did not significantly affect the processing at the Gly<sup>897</sup> ↓ Gly<sup>898</sup> site. In contrast, the analogous PL2pro mutants (pp759–1910\_C1701A; pp759–1910\_Δ1701–1708) were markedly inhibited in the cleavage of the same bond (Fig. 7B, lanes 5–8). Importantly, the Gly<sup>897</sup> ↓ Gly<sup>898</sup> site cleavage by PL1pro was more pronounced in the absence rather than in the presence of the inactivated PL2pro (compare Fig. 7A, lanes 1–4, to Fig. 7B, lanes 5–8). The combined results presented in Fig. 7 indicate that (i) PL2pro cleaves the Gly<sup>897</sup> ↓ Gly<sup>898</sup> site more efficiently than PL1pro and (ii) PL2pro suppresses the PL1pro proteolytic activity. Since the active-site deletion mutant of PL2pro retained the dominant-negative effect on PL1pro, it is not likely that PL2pro simply outcompetes PL1pro for the cleavage site. Further experiments need to be performed to elucidate the details of this mechanism.

*The Partial Substrate Redundancy Is Accompanied by Parallel Evolution of the Paralogous PLpros*—The ubiquitous occurrence of PL1pro and PL2pro in all coronaviruses sequenced to date (Fig. 3)<sup>3</sup> indicates that these enzymes most probably

originated from the duplication of a papain-like protease in one of the ancestors of the contemporary coronaviruses. PL1pro and PL2pro have subsequently evolved as part of the pp1a/pp1ab polyproteins that automatically determines the branching of their phylogeny (Fig. 8A). However, when the radial tree was reconstructed using a multiple alignment of coronavirus PLpros, its topology markedly deviated from the one determined for other replicative proteins (compare Fig. 8, A and B). Only the orthologous enzymes of the most closely related viruses, TGEV and HCoV, and MHV and bovine coronavirus (BCoVL), respectively, were clustered together (Fig. 8B). More deeply rooted branches of PL1pro and PL2pro proteins were interleaved (Fig. 8B) rather than forming two separate divisions, one for PL1pros and another for PL2pros (Fig. 8A). The tree topology of Fig. 8B was inferred using an NJ algorithm (42). It was supported by results of a bootstrap analysis (Fig. 8B) and was also observed in one of the five most parsimonious trees with the best score after an exhaustive search of the entire tree-space using PAUP\*4.0.0d55 (44). The four other most parsimonious trees also deviated from the expected tree presented in Fig. 8A (data not shown). These observations indicate that either our initial assumption of the one-time duplication of a PLpro domain in the coronavirus ancestral lineage is not correct or the evolution of the paralogous PLpros was complicated by homoplasy events that fooled the reconstruction of the genuine topology presented in Fig. 8A. Remarkably, in the inferred PLpro tree shown in Fig. 8B, the PLpros cluster according to the coronavirus genetic grouping (53), bringing together the paralogous enzymes of the same virus (Fig. 8B). This topology makes the second evolutionary scenario most probable as it is compatible with a parallel evolution of the paralogous PL1pro and PL2pro under the pressure of the common substrate that these proteases cleave in HCoV and in other coronaviruses. Thus, the phylogenetic analysis of PLpros supports the results of other analyses (see above) and indicates that the substrate pressure had a significant impact on the structure of the coronavirus PLpros.

#### DISCUSSION

The life cycle of many RNA viruses is driven by the concerted action of several proteases. The proteolytic enzymes mediate

<sup>3</sup> Chouljenko, V. N., Lin, X. Q., Storz, J., Kousoulas, K. G., and Gorbalenya, A. E. (2001) *J. Gen. Virol.* **82**, in press.

the production of diverse functional subunits and thus couple (and regulate) the replication, expression, and encapsidation of the virus genome in a timely and spatially coordinated manner. To do this, proteases cleave non-overlapping sets of few sites in the virus-encoded polyprotein(s) (1–3). In this study, we have characterized two sequentially positioned and paralogous proteases of a human coronavirus with an unusual structural organization featuring a  $Zn^{2+}$  finger embedded between the two domains of a papain-like fold (32). We now show that these proteases also possess unique functional properties as they have overlapping substrate specificities. The coordination of these protease activities may require an extent of complexity not observed elsewhere.

*PL2pro Is the Dominant Force and PL1pro Is Tightly Regulated to Release p195/p210 in Coronaviruses*—To gain insight into the autocatalytic release mechanisms of the largest coronavirus replicative protein, p195/p210, from its polyprotein precursor, we updated a coronavirus-wide multiple alignment of this protein. The results we obtained revealed that p195/p210 has a conserved domain organization and is flanked by conserved cleavage sites. Furthermore, IBV (like other coronaviruses) is predicted to have two PLpro domains (rather than only one PLpro as thought before). The previously characterized IBV papain-like protease, known as PLpro, was shown to be an ortholog of the coronavirus PL2pro domains, and an inactivated remnant of PL1pro was identified at a more upstream position in the viral polyprotein. We then sought to connect the revised domain organization of p195/p210 with the available experimental data and found that, strikingly, the processing mechanisms at the N terminus of p195/p210 vary among different coronaviruses. Thus, it emerged that the conserved peptide bond at the N terminus of p195/p210 is cleaved by different proteases in two coronaviruses, by *PL1pro* in MHV (20) but by *PL2pro* in IBV (15). Furthermore, this bond was apparently not cleaved by PL1pro *in vitro* in HCoV (25). We attempted to reconcile the results of our theoretical analysis and the published data and, to this end, performed a comprehensive characterization of this cleavage reaction in HCoV. By using *in vitro* translation of synthetic RNAs in reticulocyte lysates, it was established that both PL1pro and PL2pro cleave the predicted Gly<sup>897</sup> ↓ Gly<sup>898</sup> bond at the N terminus of p195/p210. Although p195/p210 itself remains to be identified in HCoV-infected cells, its upstream neighbor in the polyprotein, p87, which is released by cleavage of the same bond, was detected in this study. This result strongly suggests that our observations do not reflect an *in vitro* artifact.

We then characterized mutants of PL1pro and PL2pro and found specific conditions under which the proteolytic activities of PL1pro and PL2pro, respectively, were evident. In a precursor containing PL1pro but lacking PL2pro, the p195/p210 N-terminal cleavage was mediated by PL1pro. If the HCoV PL1pro was expressed in combination with PL2pro from the same RNA template, the N-terminal cleavage of p195/p210 was significantly stimulated (Fig. 7A). Both results are consistent with similar phenomena reported recently for MHV trans-cleavage assays (50) (see below). Subsequently, a more detailed analysis of the enhancement of the HCoV PL2pro activity led us to conclude that PL2pro is able to cleave the p195/p210 site on its own. Moreover, it became evident that PL2pro is capable of silencing the PL1pro activity. These conclusions were not reached in a similar study of MHV (50), which failed to positively identify the proteolytic activity of PL2pro. We believe that the apparent discrepancy between the HCoV and MHV data results from technical reasons and does not reflect virus-specific differences in the p195/p210 processing mechanism.

Thus, in the MHV study, the proteolytic activities of proteins

containing both the PL1pro and PL2pro domains were demonstrated in respect to the equivalents of the p9 ↓ p87 and p87 ↓ p195/p210 sites in *bimolecular reactions*. Furthermore, the design of these previously tested proteins differed from that of the proteins characterized in our study in *monomolecular reactions*. Teng *et al.* (50) observed a significant stimulation of the cleavages in the presence of PL2pro, and in a separate experiment (Fig. 7A in (50)), the cleavages were blocked in a nonconservative (His-to-Pro) PL1pro active-site mutant. Although the latter result was interpreted (50) to argue against an involvement of PL2pro in the cleavage, we consider this conclusion premature since no data on the corresponding PL2pro active-site mutant(s) were presented in that paper. Likewise, in two MHV mutants in which PL1pro was deleted, the C terminus of the deletion was placed downstream of Block C, that is *within* the predicted N-terminal region of PL2pro (see Fig. 3). As a result, PL2pro was unintentionally truncated, which, according to our model of p195/p210, predetermined the processing-negative phenotype of those mutants. The ability of PL1pro and PL2pro to cleave the p87 ↓ p195/p210 junction *in trans* has yet to be characterized for HCoV.

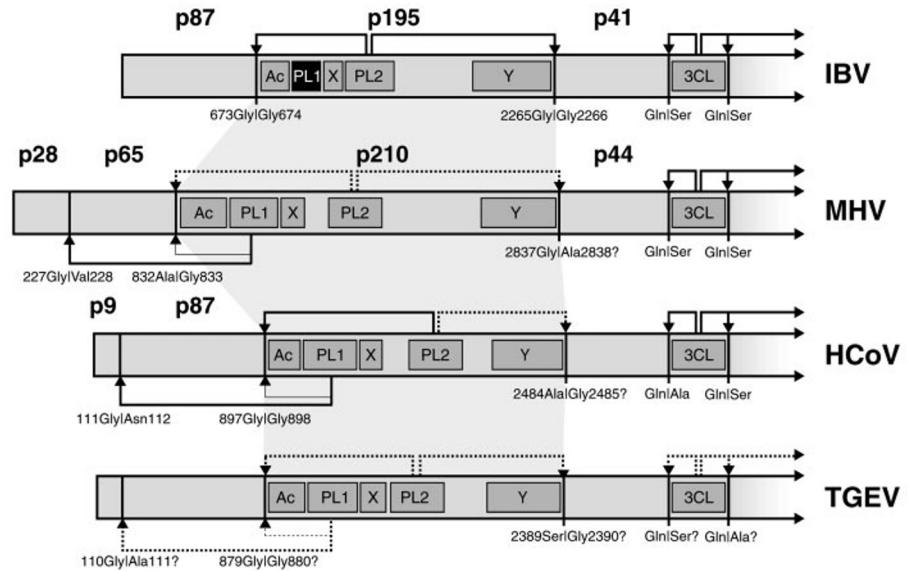
The results discussed above and the other published data (20, 25, 32) suggest that the PL1pro activity at the N terminus of p195/p210 is tightly down-regulated by upstream and downstream domains in both MHV and HCoV. Regardless of the mechanisms of these effects, which remain to be elucidated, these observations indicate that PL1pro may have a very short time frame to exert its proteolytic activity *in cis*. We therefore suggest that PL2pro releases the N terminus of the p195/p210 proteins in HCoV and other coronaviruses and dominates over PL1pro in this cleavage reaction (Fig. 9), although we acknowledge that the PL2pro activity at this site remains to be formally proved for MHV (and some other coronaviruses). In contrast, the processing at the p195/p210 C terminus may be mediated by PL2pro alone (Fig. 9). This site was shown to be effectively processed by PL2pro in IBV (16) and MHV (22), although the ability of PL1pro to cleave this site was not yet rigorously tested for any coronavirus. Future studies on the p195/p210 C-terminal cleavage site structure should also resolve a slight uncertainty of our computer-assisted prediction, which was due to the low complexity and weak conservation in this region, about the precise location of the scissile bond (Fig. 3). This advance would allow us to correlate the structure of the three cleavage sites with the type of the cognate protease(s) in the N-terminal part of the coronavirus pp1a/pp1ab proteins (Fig. 9).

The characterization of the HCoV p87 ↓ p195/p210 cleavage *in vitro* proved to be a significant technical challenge, since template DNAs containing the “non-clonable” PL2pro coding sequence had to be produced. *In vitro* ligation and PCR approaches (combined with extensive nucleotide sequencing) finally allowed us to analyze the processing of large size precursors containing both HCoV papain-like proteases. The spectrum of constructs (Fig. 2) allowed us to discriminate between the activities of PL1pro and PL2pro. However, more experiments are yet to be done to address a possible involvement of other conserved and non-conserved domains of p195/p210 as well as other replicative proteins in the modulation of the PL1pro and PL2pro activities. Also, since the cellular environment may be involved in the control of specific proteolytic activities, future studies of the p195/p210 autoprocessing in the natural setting using reverse genetics are required to understand the full complexity of these processes. Also, these studies might reveal variations among coronaviruses.

*Evolution of PL1pro and PL2pro and Their Substrates in Coronaviruses*—The PL1pro-inactive/PL2pro-active organi-



**FIG. 9. Proposed scheme for the proteolytic processing of coronavirus replicative polyproteins by the accessory proteases PL1pro and PL2pro.** Cleavage sites (P1 and P1' residues indicated) identified in the pp1a/pp1ab proteins of IBV, MHV, HCoV, and TGEV and the corresponding processing products identified in virus-infected cells are shown. Putative cleavage sites, which are predicted on the basis of the results of the present study, are indicated by ?. Also, the protease domains responsible for specific cleavages are given, with *solid lines* indicating experimentally characterized cleavages and *dotted lines* indicating predicted cleavages. The proteolytically inactive IBV PL1pro is marked by a *black background color*. For other abbreviations see Figs. 1B and 3.



zation of IBV was imitated in two of the HCoV mutants (pp717–1910\_C1054A; pp717–1910\_Δ1054–1061) tested in this study (Fig. 7B). Remarkably, the mutated precursors were processed in an IBV-like (that is PL2pro-controlled) fashion. This result connects IBV with other coronaviruses that have two active PLpros. It allows us to reconstruct a plausible scenario of the evolutionary events that might have led to the present day diversity of the N-terminal region of the coronavirus replicative polyproteins. We speculate that an immediate ancestor of the contemporary coronaviruses already encoded a pair of PLpros. It is likely that PL2pro, probably assisted by PL1pro, mediated the p195/p210 auto-processing, whereas PL1pro could have been responsible for the release of the small N-terminal protein. (It should be noted that the ability of PL2pro to release the N-terminal protein has not been tested rigorously in coronaviruses with two active PLpros.) Three coronavirus lineages, known as group 1 (prototyped by HCoV), group 2 (prototyped by MHV), and group 3 (prototyped by IBV), have evolved from the common coronavirus ancestor (53).<sup>3</sup> The individual lineages display a considerable sequence variability that also includes the N terminus of the replicative polyproteins (Fig. 9). Groups 1 and 2 encode very specific (and possibly unrelated) versions of the N-terminal protein, p28 in MHV and p9 in HCoV, that significantly differ in size. The activities of these proteins are unknown but, because of their unique structural characteristics, they must be lineage-specific. The IBV lineage does not encode a counterpart to the N-terminal proteins of other coronaviruses (Fig. 9); most likely, it was deleted or, after fusion with the upstream protein, diverged beyond recognition. In the absence of its major substrate, the proteolytic activity of PL1pro was no longer essential for the IBV ancestor and, as a result, PL1pro was inactivated by accumulating mutations. Since the IBV PL1pro was not deleted, it must possess another (nonproteolytic) activity, which remains to be determined.

The PL1pro and PL2pro domains of coronaviruses have probably evolved by the duplication of an ancestral papain-like protease. Since then, they have diverged substantially and share less than ~25% of identical residues in every coronavirus pair. The evolution of paralogous proteases is commonly driven by the need to process novel substrates. There are numerous paralogous proteases with different specificities among cellular

enzymes; and the entero-/rhinovirus 2A and 3C proteases, employing similar chymotrypsin-like folds and recognizing different sites, illustrate this trend in viruses (reviewed in Refs. 3 and 54). Surprisingly, PL1pro and PL2pro of HCoV (and presumably other coronaviruses) retained overlapping substrate specificities despite a profound divergent evolution elsewhere in the genome. This conservation involves yet-to-be-identified determinants in PL1pro and PL2pro, although it can already be noted that all proteolytically active coronavirus PLpros share a unique zinc finger that was shown to be essential for the proteolytic activity of the HCoV PL1pro (32). The PL1pro and PL2pro alignments (Fig. 3) revealed that only few positions are occupied by lineage-specific amino acid residues,<sup>4</sup> and this unusual pattern could be linked to the selective pressure of a common substrate, driving the parallel evolution of PL1pro and PL2pro. Accordingly, the expected topology of the coronavirus PLpro tree (Fig. 8A) was not readily reconstructed using the conventional algorithms (Fig. 8B).

*p195/p210 Is a Multifunctional Protein with a Unique Regulation of Expression*—The cleavage of the same site by two proteases may provide a specific selective advantage since it creates an additional level of regulation in processes that involve (and consume) p195/p210. The unique character of this regulation might be dictated by the exceptional complexity of the domain organization of p195/p210. It is conceivable that the p195/p210 processing may have different kinetics if either PL1pro/PL2pro or PL2pro alone mediates this reaction. Kinetic parameters could affect the localization of specific products in the cell and/or their interactions with other partners. The p195/p210 product may be anchored in membranes through hydrophobic regions of the C-terminal Y domain or other mechanisms (55) and, furthermore, may be involved in the regulation of transcription through the PLpro-associated zinc finger domains. Finally, the fact that field isolates of BCoV obtained from two different tissues of the same animal were recently shown to selectively accumulate non-synonymous mutations in the p195/p210 protein<sup>3</sup> points to yet another functional aspect of this multidomain protein. Taken together, these data suggest that the above-mentioned or other activities of this multifunctional protein may have profound effects on the host cell or

<sup>4</sup> J. Ziebuhr, V. Thiel, and A. E. Gorbalenya, unpublished data.

even the entire organism. It is thus tempting to speculate that the sophisticated two-protease regulation at the N terminus of p195/p210 might be involved in specific coronavirus-host interactions.

**Acknowledgments**—We are grateful to V. Chouljenko and K. Kousoulas for access to the BCoV sequence and permission to cite unpublished observations. We also thank V. Hoppe for protein sequence data and the staff of the Advanced Biomedical Computing Center for the assistance with computer resources.

## REFERENCES

- Kräusslich, H. G., and Wimmer, E. (1988) *Annu. Rev. Biochem.* **57**, 701–754
- Dougherty, W. G., and Semler, B. L. (1993) *Microbiol. Rev.* **57**, 781–822
- Gorbalenya, A. E., and Snijder, E. J. (1996) *Perspect. Drug Discov. Des.* **6**, 64–86
- den Boon, J. A., Snijder, E. J., Chirnside, E. D., de Vries, A. A., Horzinek, M. C., and Spaan, W. J. (1991) *J. Virol.* **65**, 2910–2920
- Cavanagh, D. (1997) *Arch. Virol.* **142**, 629–633
- Lai, M. M. C., and Cavanagh, D. (1997) *Adv. Virus Res.* **48**, 1–100
- Molenkamp, R., van Tol, H., Rozier, B. C., van der Meer, Y., Spaan, W. J., and Snijder, E. J. (2000) *J. Gen. Virol.* **81**, 2491–2496
- Thiel, V., Herold, J., Schelle, B., and Siddell, S. G. (2001) *J. Virol.* **75**, 6676–6681
- Seybert, A., van Dinten, L. C., Snijder, E. J., and Ziebuhr, J. (2000) *J. Virol.* **74**, 9586–9593
- Seybert, A., Hegyi, A., Siddell, S. G., and Ziebuhr, J. (2000) *RNA* **6**, 1056–1068
- Brierley, I., Boursnell, M. E., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D., and Inglis, S. C. (1987) *EMBO J.* **6**, 3779–3785
- Ziebuhr, J., Snijder, E. J., and Gorbalenya, A. E. (2000) *J. Gen. Virol.* **81**, 853–879
- Gorbalenya, A. E., Koonin, E. V., and Lai, M. M. C. (1991) *FEBS Lett.* **288**, 201–205
- Lee, H. J., Shieh, C. K., Gorbalenya, A. E., Koonin, E. V., La Monica, N., Tuler, J., Bagdzhadzhyan, A., and Lai, M. M. (1991) *Virology* **180**, 567–582
- Lim, K. P., and Liu, D. X. (1998) *Virology* **245**, 303–312
- Lim, K. P., Ng, L. F., and Liu, D. X. (2000) *J. Virol.* **74**, 1674–1685
- Denison, M., and Perlman, S. (1987) *Virology* **157**, 565–568
- Denison, M. R., Hughes, S. A., and Weiss, S. R. (1995) *Virology* **207**, 316–320
- Schiller, J. J., Kanjanahaluethai, A., and Baker, S. C. (1998) *Virology* **242**, 288–302
- Bonilla, P. J., Hughes, S. A., Piñón, J. D., and Weiss, S. R. (1995) *Virology* **209**, 489–497
- Bonilla, P. J., Hughes, S. A., and Weiss, S. R. (1997) *J. Virol.* **71**, 900–909
- Kanjanahaluethai, A., and Baker, S. C. (2000) *J. Virol.* **74**, 7911–7921
- Baker, S. C., Shieh, C. K., Soe, L. H., Chang, M. F., Vannier, D. M., and Lai, M. M. (1989) *J. Virol.* **63**, 3693–3699
- Baker, S. C., Yokomori, K., Dong, S., Carlisle, R., Gorbalenya, A. E., Koonin, E. V., and Lai, M. M. (1993) *J. Virol.* **67**, 6056–6063
- Herold, J., Gorbalenya, A. E., Thiel, V., Schelle, B., and Siddell, S. G. (1998) *J. Virol.* **72**, 910–918
- Dong, S., and Baker, S. C. (1994) *Virology* **204**, 541–549
- Hughes, S. A., Bonilla, P. J., and Weiss, S. R. (1995) *J. Virol.* **69**, 809–813
- Almazan, F., González, J. M., Pénzes, Z., Izeta, A., Calvo, E., Plana-Durán, J., and Enjuanes, L. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5516–5521
- Yount, B., Curtis, K. M., and Baric, R. S. (2000) *J. Virol.* **74**, 10600–10611
- Thiel, V., Herold, J., Schelle, B., and Siddell, S. G. (2001) *J. Gen. Virol.* **82**, 1273–1281
- LaStarza, M. W., Lemm, J. A., and Rice, C. M. (1994) *J. Virol.* **68**, 5781–5791
- Herold, J., Siddell, S. G., and Gorbalenya, A. E. (1999) *J. Biol. Chem.* **274**, 14918–14925
- Tijms, M. A., van Dinten, L. C., Gorbalenya, A. E., and Snijder, E. J. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1889–1894
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res.* **25**, 4876–4882
- Morgenstern, B. (1999) *Bioinformatics* **15**, 211–218
- Schuler, G. D., Altschul, S. F., and Lipman, D. J. (1991) *Proteins* **9**, 180–190
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402
- Eddy, S. R. (1996) *Curr. Opin. Struct. Biol.* **6**, 361–365
- Henikoff, S., and Henikoff, J. G. (1994) *J. Mol. Biol.* **243**, 574–578
- Rost, B., Casadio, R., Fariselli, P., and Sander, C. (1995) *Protein Sci.* **4**, 521–533
- Rost, B. (1996) *Methods Enzymol.* **266**, 525–539
- Saitou, N., and Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425
- Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, p. 367, Cambridge University Press, Cambridge, NY
- Swofford, D. L. (2000) *PAUP\* Phylogenetic Analysis Using Parsimony*, Version 4, Sinauer Associates, Inc., Sunderland, MA
- Page, R. D. (1996) *Comput. Appl. Biosci.* **12**, 357–358
- Ziebuhr, J., Herold, J., and Siddell, S. G. (1995) *J. Virol.* **69**, 4331–4338
- Herold, J., Raabe, T., Schelle-Prinz, B., and Siddell, S. G. (1993) *Virology* **195**, 680–691
- Herold, J., Siddell, S. G., and Ziebuhr, J. (1996) *Methods Enzymol.* **275**, 68–89
- Ziebuhr, J., and Siddell, S. G. (1999) *J. Virol.* **73**, 177–185
- Teng, H., Piñón, J. D., and Weiss, S. R. (1999) *J. Virol.* **73**, 2658–2666
- Grötzinger, C., Heussipp, G., Ziebuhr, J., Harms, U., Süß, J., and Siddell, S. G. (1996) *Virology* **222**, 227–235
- Yao, Z., Jones, D. H., and Grose, C. (1992) *PCR Methods Appl.* **1**, 205–207
- Siddell, S. G. (1995) in *The Coronaviridae* (Siddell, S. G., ed) pp. 1–10, Plenum Press, New York
- Ryan, M. D., and Flint, M. (1997) *J. Gen. Virol.* **78**, 699–723
- Bi, W., Piñón, J. D., Hughes, S., Bonilla, P. J., Holmes, K. V., Weiss, S. R., and Leibowitz, J. L. (1998) *J. Neurovirol.* **4**, 594–605
- Gorbalenya, A. E., and Koonin, E. V. (1993) *Sov. Sci. Rev. Sect. D Physicochem. Biol.* **11**, 1–84
- Bonilla, P. J., Gorbalenya, A. E., and Weiss, S. R. (1994) *Virology* **198**, 736–740