Journal of General Virology (2002), **83**, 581–593. Printed in Great Britain

Mutational analysis of the active centre of coronavirus 3C-like proteases

Annette Hegyi,[1] Agnes Friebe,[1] Alexander E. Gorbalenya[2] and John Ziebuhr[1]

[1] Institute of Virology and Immunology, University of Würzburg, Versbacher Straße 7, 97078 Würzburg, Germany
[2] Advanced Biomedical Computing Center, 430 Miller Dr. Rm 228, SAIC/NCI-Frederick, Frederick, MD 21702-1201, USA

Formation of the coronavirus replication–transcription complex involves the synthesis of large polyprotein precursors that are extensively processed by virus-encoded cysteine proteases. In this study, the coding sequence of the feline infectious peritonitis virus (FIPV) main protease, 3CL^{pro}, was determined. Comparative sequence analyses revealed that FIPV 3CL^{pro} and other coronavirus main proteases are related most closely to the 3C-like proteases of potyviruses. The predicted active centre of the coronavirus enzymes has accepted unique replacements that were probed by extensive mutational analysis. The wild-type FIPV 3CL^{pro} domain and 25 mutants were expressed in *Escherichia coli* and tested for proteolytic activity in a peptide-based assay. The data strongly suggest that, first, the FIPV 3CL^{pro} catalytic system employs His^{41} and Cys^{144} as the principal catalytic residues. Second, the amino acids Tyr^{160} and His^{162}, which are part of the conserved sequence signature Tyr^{160}–Met^{161}–His^{162} and are believed to be involved in substrate recognition, were found to be indispensable for proteolytic activity. Third, replacements of Gly^{83} and Asn^{64}, which were candidates to occupy the position spatially equivalent to that of the catalytic Asp residue of chymotrypsin-like proteases, resulted in proteolytically active proteins. Surprisingly, some of the Asn^{64} mutants even exhibited strongly increased activities. Similar results were obtained for human coronavirus (HCoV) 3CL^{pro} mutants in which the equivalent Asn residue (HCoV 3CL^{pro} Asn^{64}) was substituted. These data lead us to conclude that both the catalytic systems and substrate-binding pockets of coronavirus main proteases differ from those of other RNA virus 3C and 3C-like proteases.

Introduction

Coronaviruses are positive-stranded RNA viruses with exceptionally large genome sizes (up to 31 kb). Based on a similar polycistronic genome organization, common transcriptional and (post)-translational strategies and a conserved array of homologous domains in the viral polyprotein, the *Coronaviridae* have been united with the much smaller *Arteriviridae* in the (newly established) order *Nidovirales* (which is composed of these two virus families) (den Boon *et al.*, 1991; Cavanagh, 1997). Another common hallmark of corona- and arteriviruses is the use of a unique discontinuous transcription mechanism, resulting in the production of a nested set of subgenomic mRNAs (Spaan *et al.*, 1983). Available experimental data suggest that fusion of the noncontiguous 5′ and 3′ sequences occurs during negative-strand RNA synthesis (Sawicki & Sawicki, 1998; Sawicki *et al.*, 2001) and involves mechanisms that resemble similarity assisted, copy-choice RNA recombination (van Marle *et al.*, 1999).

Author for correspondence: John Ziebuhr.
Fax +49 931 2013934. e-mail ziebuhr@vim.uni-wuerzburg.de

Feline infectious peritonitis virus (FIPV) strain 79-1146 as well as other feline coronaviruses belong to group I coronaviruses and are related closely to both canine coronavirus and porcine transmissible gastroenteritis virus (TGEV) (Vennema *et al.*, 1992). In fact, all of these viruses are antigenically so similar that they may be regarded as 'host range mutants' rather than as separate species (Horzinek *et al.*, 1982). FIPV infects both domestic and wild cats and generally causes mild enteric diseases. However, in about 5–10% of FIPV infections, the disease progresses and results in a debilitating lethal disease characterized by an exudative fibrinous serositis in the abdominal and thoracic cavities (reviewed by Olsen, 1993; de Groot & Horzinek, 1995). Previous sequence analyses have focussed mainly on the 3′-terminal region of the FIPV genome, which encodes the structural proteins of the virus (Vennema *et al.*, 1992, 1998; Herrewegh *et al.*, 1995, 1998). Although there is nearly no sequence information on the 5′-proximal region of

the FIPV genome, it is reasonable to assume that this sequence will be very similar to the TGEV replicase gene sequence reported previously (Eleouet *et al.*, 1995).

The protein functions encoded by the human coronavirus 229E (HCoV) replicase gene have been shown recently to suffice for both genome replication and transcription (Thiel *et al.*, 2001b) and it is likely that the same applies to other coronaviruses, including FIPV. The replicase gene occupies the 5′-proximal two-thirds of the genome and comprises two open reading frames (ORFs), ORFs 1a and 1b, which are connected by a ribosomal frameshift site (Brierley *et al.*, 1987). Thus, two overlapping polyproteins are translated from the genome RNA: an ORF1a-encoded protein, ~ 450-kDa (pp1a), and a C-terminally extended frameshift protein. The latter protein, called pp1ab, is encoded by ORFs 1a and 1b and has a calculated molecular mass of ~ 750 kDa. The replicative polyproteins are processed extensively by viral proteases to produce the functional subunits of the virus replication/transcription machinery. The N-proximal region of pp1a/pp1ab is cleaved by zinc finger-containing papain-like proteases (so called accessory proteases) at two or three sites (Baker *et al.*, 1989; Gorbalenya *et al.*, 1991; Bonilla *et al.*, 1997; Herold *et al.*, 1998, 1999; Kanjanahaluethai & Baker, 2000; Lim *et al.*, 2000; Ziebuhr *et al.*, 2001). In contrast, the C-proximal region is processed by a 3C-like cysteine protease (3CL^{pro}), which is encoded in the 3′-proximal region of ORF1a and derives its name from a remote similarity to the picornavirus 3C proteases (Gorbalenya *et al.*, 1989b). 3CL^{pro} cleaves the viral polyprotein at 11 conserved interdomain junctions (Ziebuhr *et al.*, 2000), including those that release the core replicative domains, such as the putative RNA-dependent RNA polymerase (RdRp) and the helicase (Gorbalenya *et al.*, 1989b; Liu *et al.*, 1994; Grötzinger *et al.*, 1996; Heusipp *et al.*, 1997b; Denison *et al.*, 1999; Seybert *et al.*, 2000a). Because of its key role in replicase gene expression, 3CL^{pro} has been termed coronavirus 'main protease' (Ziebuhr *et al.*, 2000). Previous studies using recombinant main proteases of avian infectious bronchitis virus (IBV), mouse hepatitis virus (MHV) and HCoV, representing all three major groups of coronaviruses, provided initial insight into the functional and structural properties of this enzyme (reviewed by Ziebuhr *et al.*, 2000). First, the coronavirus main protease was shown to possess a well-defined substrate specificity; all cleavage sites contain bulky hydrophobic residues at the P2 position, Gln at the P1 position and small aliphatic residues at the P1′ position (Liu *et al.*, 1994, 1997, 1998; Lu *et al.*, 1995, 1998; Ziebuhr *et al.*, 1995; Grötzinger *et al.*, 1996; Heusipp *et al.*, 1997a, b; Ziebuhr & Siddell, 1999; Denison *et al.*, 1999). Second, both sequence comparisons and mutation analyses suggested that coronavirus main proteases may employ a catalytic dyad of conserved His and Cys residues. Thus, in contrast to most other viral proteases with (predicted or proven) chymotrypsin-like folds, no clear evidence has been obtained for the involvement of a third (acidic) residue in catalysis (Liu & Brown, 1995; Lu & Denison 1997; Ziebuhr *et al.*, 1997). Third, coronavirus main proteases possess an additional, C-terminal domain with unknown function(s) (Gorbalenya *et al.*, 1989b). Although truncations of this domain reduced significantly or abolished completely the proteolytic activities of the IBV, MHV and HCoV main proteases in both *in vivo* and *in vitro* experiments (Lu & Denison, 1997; Ziebuhr *et al.*, 1997; Ng & Liu, 2000), a direct involvement of the C-terminal domain in catalysis or substrate binding appears unlikely. Thus, it was shown recently for IBV 3CL^{pro} that the removal of as much as 67 amino acids from the C terminus was partially tolerated (Ng & Liu, 2000).

To gain additional insight into the biochemistry of coronavirus 3CL^{pro}-mediated proteolysis, we have cloned and sequenced FIPV 3CL^{pro} and compared the deduced amino acid sequence with those of other coronavirus main proteases and the more distantly related potyviral 3C-like proteases. Functional predictions derived from the sequence comparison were tested subsequently. To this end, a series of recombinant proteins carrying single amino acid replacements was expressed in bacteria and the proteolytic activities of the purified proteins were analysed using a peptide-based assay. Using this approach, we have been able to confirm and extend the results of previous studies on the active site of coronavirus main proteases. The data suggest that coronavirus main proteases have evolved a catalytic system that only resembles very distantly that of the canonical chymotrypsin-like enzymes. Furthermore, we have identified specific amino acid replacements that enhanced significantly the activity of 3CL^{pro}, indicating that the proteolytic activities of coronavirus main proteases may be tuned to meet specific functional constraints.

Methods

■ **Virus and cells.** FIPV strain 79-1146 was propagated in monolayers of *Felis catus* whole foetus (FCWF) cells (Jacobse-Geels & Horzinek, 1983) maintained in Dulbecco's modified Eagle medium containing 10% foetal bovine serum, nonessential amino acids (11140-035; Life Technologies), glutamine and antibiotics.

■ **Determination of the FIPV 3CL^{pro}-coding sequence.** Polyadenylated RNA was prepared from FIPV-infected FCWF cells and reverse transcribed using oligonucleotide AH10 [5′ (A/C)ACAGC-(A/T)CGTGCTTCTTT(A/G)TACAT 3′]. The sequences of both AH10, which was also used as the downstream primer in the subsequent PCR, and the upstream PCR primer AH11 [5′ AATCTTTT(C/T)GAAGG-TGA(C/T)AAATTTG 3′] were derived from replicase gene regions that are highly conserved among group I coronaviruses. The nucleotide sequences of three independently obtained 2·5 kb PCR products were determined and the putative FIPV 3CL^{pro} domain was identified by comparison of the deduced amino acid sequence with coronavirus replicase polyprotein sequences analysed previously (Boursnell *et al.*, 1987; Lee *et al.*, 1991; Herold *et al.*, 1993; Eleouet *et al.*, 1995). The nucleotide sequence of FIPV 3CL^{pro}, together with the flanking sequences, was deposited in GenBank under accession number AF326575.

■ **Construction of plasmid pMALc2-FIPV 3CL^{pro}.** The coding sequence of the putative FIPV 3CL^{pro} domain was amplified by PCR from the cDNA template described above using oligonucleotides AH12 (5′

TCCGGATTGAGAAAAATGGCA 3′) and AH17 (5′ CGCGGATCC-TTACTGAAGATTAACACCATACATTTG 3′). The 918 bp PCR product was digested with *Bam*HI and ligated with *Xmn*I/*Bam*HI-digested pMAL-c2 DNA (New England Biolabs). The resulting plasmid, pMALc2-FIPV 3CL^{pro}, encodes a 75 kDa fusion protein consisting of the *Escherichia coli* maltose-binding protein (MBP) and the putative FIPV 3CL^{pro} domain.

■ **Bacterial expression and purification of FIPV 3CL^{pro}.** Expression of recombinant FIPV 3CL^{pro} in *E. coli* TB1(pMALc2-FIPV 3CL^{pro}) and purification of the recombinant fusion protein were carried out essentially as described previously for HCoV 3CL^{pro} (Herold *et al*., 1996; Ziebuhr *et al*., 1997). The MBP–FIPV 3CL^{pro} fusion protein was purified by amylose-affinity chromatography and cleaved with factor Xa to release the FIPV 3CL^{pro} domain. The protein mixture was then loaded onto a phenyl–Sepharose HP column (Pharmacia Biotech) that had been preequilibrated with a solution containing 15 mM Bis–Tris–HCl (pH 7·0), 600 mM NaCl, 1 mM DTT and 0·1 mM EDTA. Recombinant FIPV 3CL^{pro} was eluted with 15 mM Bis–Tris–HCl (pH 7·0), 1 mM DTT and 0·1 mM EDTA. The protein was concentrated (Centricon-3, Millipore) and dialysed with buffer I containing 20 mM Tris–HCl (pH 8·8), 55 mM NaCl, 1 mM DTT and 0·1 mM EDTA. Then, the protein solution was loaded onto an anion-exchange column (UNO Q-1, Bio-Rad Laboratories) that had been preequilibrated with buffer I. The flow-through fractions containing 3CL^{pro} were pooled, concentrated and loaded onto a Superdex 75 column (Pharmacia Biotech) run under isocratic conditions with 11 mM Tris–HCl (pH 7·5), 100 mM NaCl, 1 mM DTT and 0·1 mM EDTA. The purified protein was concentrated to 15 mg/ml and stored at −80 °C.

■ **Peptide synthesis.** The synthetic 15-mer peptides F1, SP7 and SP8 were prepared by solid-phase chemistry (Merrifield, 1965) and purified by high-performance liquid chromatography (HPLC) on a reverse-phase C18 silica column (Jerini Bio-Tools). Identity and homogeneity of the peptides were confirmed by mass spectrometry and analytical reverse-phase chromatography. Peptide F1 represents the FIPV replicase polyprotein sequence H_2N–VSVNSTL**QS**GLRKMA–COOH (boldface indicates the cleaved dipeptide bond). Peptides P7 and P8 have the sequences H_2N–VDYGSDTVTYKSTAC–COOH and H_2N–NKDASF-IGKNLKSNC–COOH, respectively.

■ **Peptide cleavage.** The enzymatic activity of mutant and wild-type FIPV 3CL^{pro} (1 μg total protein) was determined by incubation with 0·5 mM substrate peptide F1 at 25 °C in 20 mM Tris–HCl (pH 7·5), 200 mM NaCl, 1 mM DTT and 1 mM EDTA. Reaction aliquots were mixed with equal volumes of 2% trifluoroacetic acid and stored at −20 °C prior to analysis by reverse-phase HPLC on a Delta Pak C18 column (3·9 × 150 mm; Waters). Cleavage products were resolved using a 5–90% linear gradient of acetonitrile in 0·1% trifluoroacetic acid, as described previously (Ziebuhr *et al*., 1997). Quantification of peak areas was used to determine the extent of substrate conversion.

■ **Site-directed mutagenesis of FIPV 3CL^{pro}.** Site-directed mutagenesis was done by a recombination PCR method, as described by Yao *et al*. (1992). The nucleotide sequences of the primers used for site-directed mutagenesis are given in Table 1. The pMALc2-FIPV 3CL^{pro}-derived plasmids encoding mutant forms of FIPV 3CL^{pro} were then transformed into *E. coli* TB1 cells and the recombinant proteins were synthesized, affinity purified and cleaved with factor Xa, as described above. The purity and structural integrity of the mutant proteins were analysed by SDS–PAGE. The control protein for this experiment, wild-type FIPV 3CL^{pro}, was purified in an identical manner.

■ **Computer-aided comparative sequence analyses.** Coronavirus amino acid sequences were derived from the Genpeptides database. 3CL^{pro} sequence alignments were produced using CLUSTAL X, version 1.81 (Thompson *et al*., 1997) and the Blossum series of scoring interresidue tables (Henikoff & Henikoff, 1994). The virus interfamily alignments were generated in the profile mode separately for the N- and C-terminal halves and then fused in an overall alignment. The alignments obtained were sent as input for the PhD program (Rost *et al*., 1995; Rost, 1996) to predict secondary structures and used also to build profiles using PROFILEWEIGHT (Thompson *et al*., 1994). These profiles were compared in pairs using PROPLOT (Thompson *et al*., 1994). Two profiles were compared by sliding windows of variable sizes along each possible register. Matches between two profiles that were within the top 0·1% and between the top 0·1 and 0·05% were marked by two different types of dots. Cluster phylogenetic trees were reconstructed using the neighbour-joining (NJ) algorithm (Saitou & Nei, 1987) with the Kimura correction (Kimura, 1983) and evaluated with 1000 bootstrap trials, as implemented in CLUSTAL X, version 1.81. Parsimonious trees were generated through exhaustive search and evaluated with bootstrap analysis using a UNIX version of PAUP* (Swofford, 2000) that is included in the GCG-Wisconsin Package of programs (Genetics Computer Group). Trees were prepared and modified using TREEVIEW (Page, 1996).

Results and Discussion

Comparative sequence analysis of FIPV 3CL^{pro} and related coronavirus proteases

Using RT–PCR, we have amplified the complete coding region of FIPV 3CL^{pro}. The sequences of the primers used in these experiments were derived from 3CL^{pro} flanking regions that are highly conserved among group I coronaviruses. The nucleotide sequences of three independent RT–PCR products were determined and the deduced amino acid sequence was compared to the replicase polyprotein sequences of other coronaviruses (Figs 1 and 2). On the basis of the overall sequence similarity with other coronavirus main proteases and the flanking sites of the MHV, IBV, and HCoV main proteases characterized previously, it is safe to predict that mature FIPV 3CL^{pro} is released from the replicase polyproteins at flanking Gln–Ser cleavage sites and encompasses 302 amino acids (Fig. 2). With 94% amino acid sequence identity, the 3CL^{pro} domain of FIPV is related most closely to the TGEV main protease, which, as yet, remains to be characterized. Overall, the phylogeny of the 3CL^{pro} domains (Fig. 1) closely parallels that obtained for the structural proteins (Siddell, 1995), the core of the RdRp (Stephensen *et al*., 1999) and the ORF1b-encoded region of pp1ab (Chouljenko *et al*., 2001), indicating a concerted evolution of all coronavirus components.

The 3CL^{pro} residues for which functional assignments have been made previously for other coronaviruses were found to be conserved also in the FIPV enzyme (Fig. 2), supporting these earlier predictions; these predictions were inferred originally from a comparison with a limited group of 3C(-like) proteases (Gorbalenya *et al*., 1989b). We have decided to extend these studies and therefore compared the coronavirus main proteases with other viral chymotrypsin-like proteases in a systematic way. To this end, profiles were generated using alignments of serine and cysteine chymotrypsin-like proteases of different virus families. These profiles were then compared

Table 1. Oligonucleotides used for the amplification or mutagenesis of FIPV and HCoV sequences

Underlined residues in the oligonucleotide sequence indicate mutant codons. The base code is as follows: R = A/G, Y = C/T, M = A/C, K = G/T, S = C/G, W = A/T, B = C/G/T, D = A/G/T, H = A/C/T, V = A/C/G and N = A/C/G/T. FIPV 3CL^{pro} amino acids are numbered continuously from the presumed N-terminal residue, Ser^{1}, to the presumed C-terminal residue, Gln^{302} (AF326575); HCoV 3CL^{pro} amino acids are numbered continuously from the presumed N-terminal residue, Ala^{1}, to the presumed C-terminal residue, Gln^{302}. The sequence corresponds to the HCoV pp1a/pp1ab residues 2966–3267 (X69721; Herold *et al.*, 1993).

Oligonucleotide	Oligonucleotide sequence (5′ → 3′)	Polarity	Mutation introduced
AF1	CTGCCCTAGATACGTCATTGCTAGTGATACATC	Forward	FIPV 3CL^{pro} His^{41}
AF2	TAGCAATGACGTATCTAGGGCAGATGACTTCAT	Reverse	FIPV 3CL^{pro} His^{41}
AF3	CTGCCCTAGACKCGTCATTGCTAGTGATACATC	Forward	FIPV 3CL^{pro} His^{41}
AF4	TAGCAATGACGCGTCTAGGGCAGATGACTTCAT	Reverse	FIPV 3CL^{pro} His^{41}
AF5	TGCTGGTACCKCTGGGTCAGTAGGTTATGTATT	Forward	FIPV 3CL^{pro} Cys^{144}
AF6A	CTACTGACCCAGCGGTACCAGCAATAAATGAAC	Reverse	FIPV 3CL^{pro} Cys^{144}
AF6S	CTACTGACCCAGAGGTACCAGCAATAAATGAAC	Reverse	FIPV 3CL^{pro} Cys^{144}
AF7	CGTGTACATGCTCCACTTGGAATTAGGTAATGG	Forward	FIPV 3CL^{pro} His^{162}
AF8	ATTCCAAGTGGAGCATGTACACGAAATAGAGCG	Reverse	FIPV 3CL^{pro} His^{162}
AF9	CGTGTACATGGCCCACTTGGAATTAGGTAATGG	Forward	FIPV 3CL^{pro} His^{162}
AF10	ATTCCAAGTGGGCCATGTACACGAAATAGAGCG	Reverse	FIPV 3CL^{pro} His^{162}
AF11	CTATTTCGTGTTCATGCACCACTTGGAATTAGG	Forward	FIPV 3CL^{pro} Tyr^{160}
AF12	AGTGGTGCATGAACACGAAATAGAGCGTTCCAT	Reverse	FIPV 3CL^{pro} Tyr^{160}
AF13	CTATTTCGTGGGCATGCACCACTTGGAATTAGG	Forward	FIPV 3CL^{pro} Tyr^{160}
AF14	AGTGGTGCATGCCCACGAAATAGAGCGTTCCAT	Reverse	FIPV 3CL^{pro} Tyr^{160}
AF15	GCGTTTACATGCCTTTTCTATAGCCAAAAATAA	Forward	FIPV 3CL^{pro} Asn^{64}
AF16	CTATAGAAAAGGCATGTAAACGCACACTAGACA	Reverse	FIPV 3CL^{pro} Asn^{64}
AF17	GCGTTTACATGACTTTTCTATAGCCAAAAATAA	Forward	FIPV 3CL^{pro} Asn^{64}
AF18	CTATAGAAAAGTCATGTAAACGCACACTAGACA	Reverse	FIPV 3CL^{pro} Asn^{64}
AF19	CAAATATAAGGATGTAAATCTTGTGCTTAAAGT	Forward	FIPV 3CL^{pro} Gly^{83}
AF20	CAAGATTTACATCCTTATATTTGGCAGACACAA	Reverse	FIPV 3CL^{pro} Gly^{83}
AF21	CAAATATAAGGCTGTAAATCTTGTGCTTAAAGT	Forward	FIPV 3CL^{pro} Gly^{83}
AF22	CAAGATTTACAGCCTTATATTTGGCAGACACAA	Reverse	FIPV 3CL^{pro} Gly^{83}
AF23	TATCAAAGGTGCATTTATTGCTGGTACCTGTGG	Forward	FIPV 3CL^{pro} Ser^{138}
AF24	CAGCAATAAATGCACCTTTGATAGTACCCTGAC	Reverse	FIPV 3CL^{pro} Ser^{138}
AF25	TATCAAAGGTACATTTATTGCTGGTACCTGTGG	Forward	FIPV 3CL^{pro} Ser^{138}
AF26	CAGCAATAAATGTACCTTTGATAGTACCCTGAC	Reverse	FIPV 3CL^{pro} Ser^{138}
AF27	CTATTTCGTGGCCATGCACCACTTGGAATTAGG	Forward	FIPV 3CL^{pro} Tyr^{160}
AF28	AGTGGTGCATGGCCACGAAATAGAGCGTTCCAT	Reverse	FIPV 3CL^{pro} Tyr^{160}
AF29	CTATTTCGTGACCATGCACCACTTGGAATTAGG	Forward	FIPV 3CL^{pro} Tyr^{160}
AF30	AGTGGTGCATGGTCACGAAATAGAGCGTTCCAT	Reverse	FIPV 3CL^{pro} Tyr^{160}
AF31	TTTCGTGTACGCGCACCACTTGGAATTAGGTAA	Forward	FIPV 3CL^{pro} Met^{161}
AF32	CCAAGTGGTGCGCGTACACGAAATAGAGCGTTC	Reverse	FIPV 3CL^{pro} Met^{161}
AH-10*	MACAGCWCGTGCTTCTTTRTACAT	Reverse	–
AH-11†	AATCTTTTTGAAGGTGAYAAATTTG	Forward	–
AH-12†	TCCGGATTGAGAAAAATGGCA	Forward	–
AH-17†	CGCGGATCCTTACTGAAGATTAACACCATACATTTG	Reverse	–
JZ142	TGCGGCTTCATGATTTTTCTATAATATCTGGTACAGCATTTC	Forward	HCoV 3CL^{pro} Asn^{64}
JZ143	TTATAGAAAAATCATGAAGCCGCATAATACTATATTCGTG	Reverse	HCoV 3CL^{pro} Asn^{64}
JZ144	TGCGGTTGCATGAATTCTCTATAATATCTGGTACAGCATTTC	Forward	HCoV 3CL^{pro} Asn^{64}
JZ145	TTATAGAGAATTCATGCAACCGCATAATACTATATTCGTG	Reverse	HCoV 3CL^{pro} Asn^{64}
JZ146	TGCGGTTGCACCAGTTTTCTATAATATCTGGTACAGCATTTC	Forward	HCoV 3CL^{pro} Asn^{64}
JZ147	TTATAGAAAACTGGTGCAACCGCATAATACTATATTCGTG	Reverse	HCoV 3CL^{pro} Asn^{64}
JZ148	TGCGGTTGCATGCATTTTCTATAATATCTGGTACAGCATTTC	Forward	HCoV 3CL^{pro} Asn^{64}
JZ149	TTATAGAAAATGCATGCAACCGCATAATACTATATTCGTG	Reverse	HCoV 3CL^{pro} Asn^{64}
JZ234	CAAATATAAGCCTGTAAATCTTGTGCTTAAAGTGAA	Forward	FIPV 3CL^{pro} Gly^{83}
JZ235	CAAGATTTACAGGCTTATATTTGGCAGACACAACAC	Reverse	FIPV 3CL^{pro} Gly^{83}
JZ236	CAAATATAAGCGTGTAAATCTTGTGCTTAAAGTGAA	Forward	FIPV 3CL^{pro} Gly^{83}
JZ237	CAAGATTTACACGCTTATATTTGGCAGACACAACAC	Reverse	FIPV 3CL^{pro} Gly^{83}
JZ238	CAAATATAAGTGGGTAAATCTTGTGCTTAAAGTGAA	Forward	FIPV 3CL^{pro} Gly^{83}
JZ239	CAAGATTTACCCACTTATATTTGGCAGACACAACAC	Reverse	FIPV 3CL^{pro} Gly^{83}
JZ240	CAAATATAAGGAGGTAAATCTTGTGCTTAAAGTGAA	Forward	FIPV 3CL^{pro} Gly^{83}
JZ241	CAAGATTTACCTCCTTATATTTGGCAGACACAACAC	Reverse	FIPV 3CL^{pro} Gly^{83}
JZ242	CAAATATAAGGTTGTAAATCTTGTGCTTAAAGTGAA	Forward	FIPV 3CL^{pro} Gly^{83}
JZ243	CAAGATTTACAACCTTATATTTGGCAGACACAACAC	Reverse	FIPV 3CL^{pro} Gly^{83}
JZ244	CAAATATAAGACTGTAAATCTTGTGCTTAAAGTGAA	Forward	FIPV 3CL^{pro} Gly^{83}
JZ245	CAAGATTTACAGTCTTATATTTGGCAGACACAACAC	Reverse	FIPV 3CL^{pro} Gly^{83}

* Used for reverse transcription.
† Used for PCR.

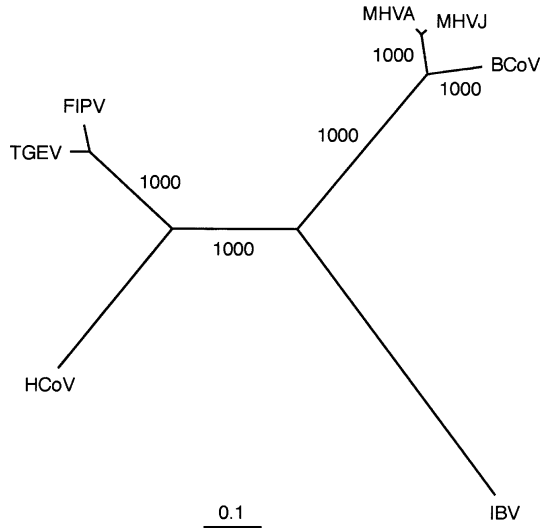

Fig. 1. An unrooted tree of coronavirus 3C-like proteases. The tree was generated using a multiple alignment of coronavirus 3C-like proteases (see Fig. 2B) by the NJ algorithm, as implemented in the CLUSTAL X, version 1.81 program. The same topology was inferred by exhaustive tree searches using parsimony criteria (data not shown). The two bovine coronavirus (BCoV) branches collapsed to a single branch. The number of trees, in which a particular bifurcation sustained in the course of 1000 bootstrap simulations, is given at each node. The scale indicates 10% amino acid replacements.

with the coronavirus 3CL^{pro} profile in a dot-plot fashion. Among all proteases analysed, the potyvirus 3C-like proteases proved to be most similar to the coronavirus enzymes (data not shown), as they revealed a similarity in the regions that flank the catalytic His and Asp residues (Fig. 2A; Ziebuhr *et al.*, 2000). This sequence affinity between corona- and potyvirus 3C-like proteases is compatible with other similarities between these viruses (Gorbalenya *et al.*, 1989b) and with the similarity between the main proteases of potyviruses and the unclassified nidovirus gill-associated virus (Cowley *et al.*, 2000). When this plot (Fig. 2A) was converted to an alignment (Fig. 2B), it became evident that the catalytic Asp residue of the potyvirus proteases may be replaced by a conserved Gly residue in the coronavirus enzymes (Gly^{83} in FIPV 3CL^{pro}) (Fig. 2B). This unusual substitution was also corroborated by an alignment of the predicted β-strands that encompass the whole catalytic domain, including a region around the catalytic Cys residue (Fig. 2B). In the latter region, no conservation was evident in the presented (Fig. 2A) and other dot-plots (data not shown) due to unique mutations in the coronavirus proteases. To characterize the coronavirus 3C-like proteases further, some of these and other positions were probed by site-directed mutagenesis in our subsequent experiments.

Expression, purification and proteolytic activity of FIPV 3CL^{pro}

We have reported previously the production and purification of biologically active main proteases of both MHV and HCoV using *E. coli* expression systems (Ziebuhr *et al.*, 1995, 1997; Seybert *et al.*, 1997). In the present study, we have adapted this strategy to produce recombinant FIPV 3CL^{pro} for biochemical analyses. FIPV 3CL^{pro} was expressed as a fusion protein with the maltose-binding protein of *E. coli* and first purified using amylose-affinity chromatography (Fig. 3, lanes 1 and 2). The fusion protein was cleaved with factor Xa to release mature (302 aa) FIPV 3CL^{pro} (Fig. 3, lane 3), which was then purified to near homogeneity by hydrophobic interaction, anion-exchange and size-exclusion chromatography (Fig. 3, lanes 4, 5, and 6).

The proteolytic activity of the recombinant FIPV 3CL^{pro} was analysed in peptide-based cleavage assays. In these experiments, 1 μg FIPV 3CL^{pro} was incubated with a range of 15-mer peptides. Peptide F1 represented the predicted N-terminal autoprocessing site of FIPV 3CL^{pro} and peptides P7 and P8 contained unrelated sequences. Analysis of the reaction products by reverse-phase chromatography revealed that FIPV 3CL^{pro} hydrolysed the peptide F1 effectively, whereas the control peptides P7 and P8 remained uncleaved under the same reaction conditions, indicating the specificity of the cleavage reaction (Fig. 4). The assay thus seemed to be suitable to test the activities of FIPV 3CL^{pro} mutants in subsequent experiments.

Mutation analysis of the catalytic system of coronavirus main proteases

In a first set of experiments, we substituted two residues, His^{41} and Cys^{144}, which, based on sequence alignments (Gorbalenya *et al.*, 1989b; Lee *et al.*, 1991; Herold *et al.*, 1993; Eleouet *et al.*, 1995) (Fig. 2B) and previous mutagenesis studies with other coronavirus main proteases, are predicted to be the principal catalytic residues of FIPV 3CL^{pro}. The structural integrity of the partially purified mutant proteins was confirmed by SDS–PAGE (Fig. 5A) and the proteolytic activities of the mutant proteins were measured in the peptide assay described above. As Table 2 shows, the proteolytic activities of the FIPV 3CL^{pro} H^{41}Y, H^{41}R, C^{144}S and C^{144}A mutants were below the detection limit of the peptide assay. These data strongly support a catalytic function for His^{41} and Cys^{144} and are fully consistent with mutagenesis data published previously on the IBV, HCoV, MHV-A59 and MHV-JHM main proteases (Liu & Brown, 1995; Lu *et al.*, 1995; Lu & Denison, 1997; Seybert *et al.*, 1997; Ziebuhr *et al.*, 1997). It should be noted that, in contrast to many other 3C and 3C-like proteases in which Cys → Ser substitutions retained partial or even full activities (Dessens & Lomonossoff, 1991; Hellen *et al.*, 1991; Lawson & Semler, 1991; Grubman *et al.*, 1995), the equivalent Cys → Ser replacement in FIPV 3CL^{pro} reduced proteolytic activity to undetectable levels. Similar data have also been reported for recombinant MHV and HCoV main proteases (Seybert *et al.*, 1997; Ziebuhr *et al.*, 1997).

The inactive phenotype of the Cys → Ser mutant may be linked to the unique active site organization of coronavirus main proteases. Obviously, these enzymes have accepted

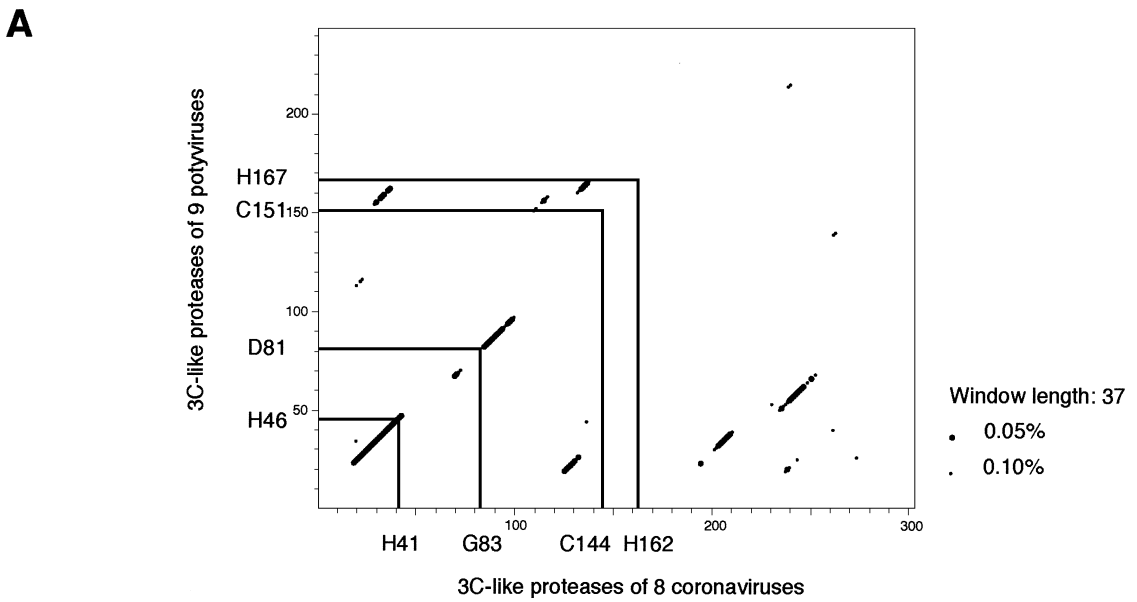

Fig. 2. Sequence conservation between coronavirus and potyvirus 3C-like proteases. (A) Dot-plot cross-comparison of corona- and potyvirus 3C-like proteases. Multiple alignments of corona- and potyvirus 3C-like proteases (see below) were converted into profiles and compared in a dot-plot fashion, as described in Methods. Shown is the dot-plot generated using a window of 37 amino acid residues. The projected positions of the catalytic residues (H^{46} versus H^{41}, D^{81} versus G^{83} and C^{151} versus C^{144}), as well as the substrate-binding H^{167} versus H^{162} residues, are shown at each axis. Those dots, which lay at any of the four possible crosses of projections of two functionally equivalent residues (e.g. H^{46} and H^{41}) or close to a nonvisible diagonal passing these crosses, belong, or may belong, to the true matches between two profiles. The rest of the dots are background hits (false positives). (B) Sequence alignment & corona- and potyvirus 3C-like proteases. A multiple sequence alignment was produced using CLUSTAL X, version 1.81, as explained in Methods. The sequence of FIPV $3CL^{pro}$ was determined in this study (AF326575) and the $3CL^{pro}$ sequences of other coronaviruses, HCoV (strain 229E), TGEV (strain Purdue 115), MHV (strains JHM and A59), BCoV (isolates ENT and LUN) and IBV (strain Beaudette), were derived from the replicative polyproteins of the respective viruses whose sequences are deposited in GenBank under the following accession numbers: HCoV, X69721 (Herold *et al.*, 1993), Z34093 (Eleouet *et al.*, 1995), M55148 (Lee *et al.*, 1991) and M95169 (Boursnell *et al.*, 1987); and BCoV, AAK83364 (Chouljenko *et al.*, 2001). The amino acid sequences of potyvirus 3C-like proteases were derived from polyproteins deposited in SWISS-PROT under the following accession numbers: TVMV, tobacco vein mottling virus (P09814); TUMVQ, turnip mosaic virus (strain Quebec) (Q02597); TEV, tobacco etch virus (P04517); PVY, potato virus Y (strain N) (P18247); PSBMV, pea seed-borne mosaic virus (strain DPD1) (P29152); PPVRA, plum pox virus (strain Rankovic) (P17767); PRSVH, papaya ringspot virus (strain P/mutant HA) (Q01901); PEMVC, pepper mottle virus (California isolate) (Q01500); and BSMRV, brome streak mosaic rymovirus (strain 11-Cal) (Q65730). The secondary structures of the two protease groups, as predicted by the PhD program, are shown at the top of the alignment: A, a, α-helix and B, b, β-strand; predictions in upper- and lower-case letters indicate prediction reliabilities of > 5 and $\leqslant 5$, respectively (Rost, 1996). The positions of the putative catalytic residues H^{41} and C^{144} (#) as well as other conserved residues characterized in this study (N^{64}, G^{83}, S^{138}, Y^{160}, M^{161} and H^{162}) are highlighted by bold, italic letters.

replacements in a number of key positions that are not observed in other chymotrypsin-like serine/cysteine proteases (Fig. 2) (Gorbalenya *et al.*, 1989b; Ziebuhr *et al.*, 2000). These replacements may be incompatible with the conversion of coronavirus main proteases from a Cys- to a Ser-based protease. Specifically, our prediction that coronavirus main proteases do not possess a third catalytic residue at the canonical position may be a critical factor. This residue promotes the polarization of the catalytic serine–histidine pair in serine proteases (the charge-relay system) and ensures the correct orientation of the thiolate–imidazolium ion pair in cysteine proteases. To validate the prediction of our alignment with regard to the third catalytic residue, we have characterized a set of mutations at two positions in the FIPV and HCoV main proteases.

First, we probed both the HCoV pp1a/pp1ab Asn^{3029} residue (which corresponds to Asn^{64} of the HCoV $3CL^{pro}$ sequence) and the homologous FIPV $3CL^{pro}$ Asn^{64} residue. These residues occupy the only position between the catalytic His and Cys residues whose conservation profile (Asn, Asp and Glu) could be reconciled with the role of the third catalytic residue (Gorbalenya & Koonin, 1993). Although this position has a limited variability in coronavirus main proteases, a number of nonconservative substitutions were tolerated in the main proteases of MHV, HCoV and IBV, without significant loss of proteolytic activity in different *in vitro* assays (Liu & Brown, 1995; Lu & Denison, 1997; Ziebuhr *et al.*, 1997). Only in one case, in which the HCoV Asn^{3029} residue was replaced by Pro, was a substantial decrease of activity measured (Ziebuhr *et al.*, 1997). In the present study, we extended the list of mutations and tested Ala, Asp, Glu and Gln replacements (Fig. 2B and 5B) at this position of HCoV $3CL^{pro}$. As shown in Table 2, all these mutations resulted in proteins with significantly increased proteolytic activities (up to threefold).

B

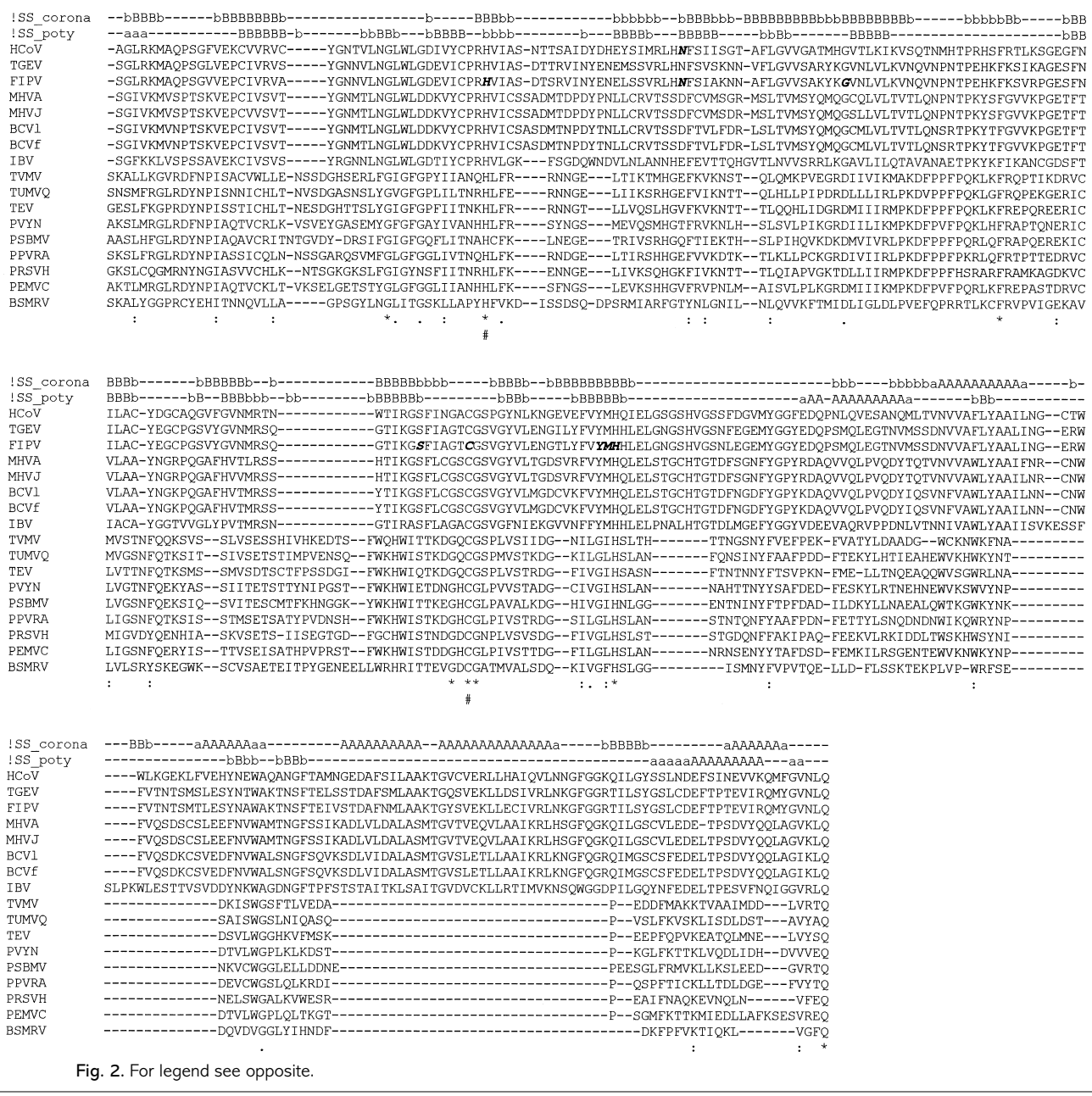

Fig. 2. For legend see opposite.

Notably, residues lacking acidic side chains also increased significantly the proteolytic activity of 3CL^{pro}, with Gln being the most active replacement. Likewise, the Ala and Asp mutations at the equivalent Asn^{64} position of FIPV (Fig. 2B) had activities that not only retained but even surpassed the activities of the wild-type enzyme.

These data, in combination with results from previous studies (Liu & Brown, 1995; Lu & Denison, 1997; Ziebuhr *et al*., 1997), strongly contradict a catalytic function of the FIPV Asn^{64} residue and the equivalent residues of other coronavirus main proteases. Upon substitution of another candidate residue, Gly^{83}, with Ala and Asp, only slight effects on the proteolytic activity of FIPV 3CL^{pro} were observed. Even nonconservative replacements like Arg, Trp, Thr, Glu, Val and Pro resulted in active enzymes, albeit with significantly reduced activities (Table 2). These results indicate that other assays and approaches must be employed to rationalize the conservation of the Gly residue at this position and the unique evolution of the catalytic acidic residue in coronavirus main proteases.

The lack of the canonical acidic catalytic residue in coronavirus main proteases can be viewed as another manifestation of the plasticity of the catalytic system of (viral)

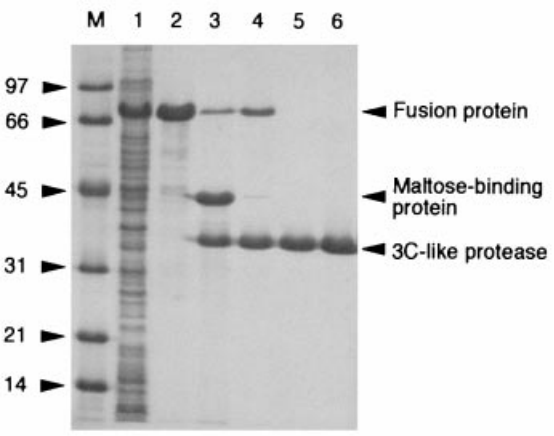

Fig. 3. Bacterial expression and purification of recombinant FIPV $3CL^{pro}$. Aliquots taken at each step of the purification protocol were analysed by SDS–PAGE on a 12·5% polyacrylamide gel and the proteins were stained with Coomassie brilliant blue. Lanes: M, protein molecular mass markers (kDa); 1, total cell lysate from IPTG-induced *E. coli* TB1 (pMALc2-FIPV $3CL^{pro}$); 2, pooled peak fractions from the amylose-affinity chromatography column; 3, factor Xa-cleaved fusion protein; 4, pooled peak fractions from the phenyl–Sepharose HP column; 5, pooled peak fractions from the Uno Q column; 6, pooled peak fractions from the Superdex 75-pg column.

chymotrypsin-like cysteine proteases. Thus, for example, as is also the case for the 3C-(like) proteases of picorna-like viruses, a limited variability of the third catalytic residue has been documented (Kean *et al.*, 1991; Yu & Lloyd, 1991; Boniotti *et al.*, 1994; Grubman *et al.*, 1995). Furthermore, structural data have indicated that the hepatitis A virus (HAV) $3C^{pro}$ Asp^{84} residue, although it occupies a position in the main chain that is topologically equivalent to that of the catalytic Asp of chymotrypsin, is not, evidently, hydrogen-bonded to the catalytic His^{44} residue (Allaire *et al.*, 1994; Bergmann *et al.*, 1997). The unique Asp^{84} side-chain orientation contradicts the predicted catalytic role of this residue (Gorbalenya *et al.*, 1989a) and leaves the conservation of this residue unexplained.

Our data also provide experimental support for previous studies in which (based on sequence comparisons) a large evolutionary distance between coronaviruses and other positive-stranded RNA viruses was observed (Boursnell *et al.*, 1987; Gorbalenya *et al.*, 1989b). In addition to $3CL^{pro}$, coronavirus RdRps, helicases and papain-like proteases also have unique structural and/or biochemical properties that separate them clearly from the homologous proteins of other positive-stranded RNA viruses, again indicating the special phylogenetic position of coronaviruses among all other positive-stranded RNA viruses (Gorbalenya *et al.*, 1989b; Koonin & Dolja, 1993; Herold *et al.*, 1999; Seybert *et al.*, 2000a, b; Ziebuhr *et al.*, 2001).

Mutation analysis of putative substrate-binding residues

An important part of the $3CL^{pro}$ active site is formed by the characteristic sequence signature Gly–X–His, which plays a

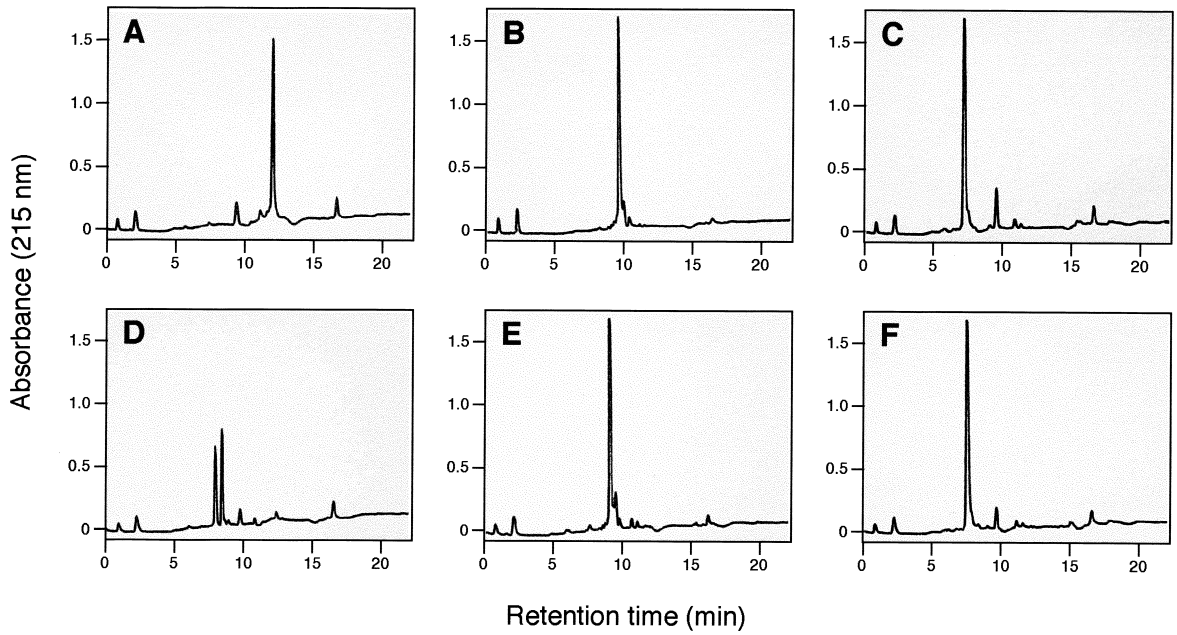

Fig. 4. Peptide-based assay for the determination of the proteolytic activity of FIPV $3CL^{pro}$. Cleavage of the FIPV-specific synthetic 15-mer peptide F1 (H_2N–VSVNSTLQSGLRKMA–COOH) by purified, recombinant FIPV $3CL^{pro}$ was analysed by reverse-phase HPLC, as described in Methods. To assess the specificity of the reaction, the extent of proteolysis of two 15-mer control peptides, P7 (H_2N–VDYGSDTVTYKSTAC–COOH) and P8 (H_2N–ESGKAKPPLNRNSVC–COOH), by recombinant FIPV $3CL^{pro}$ was determined under identical reaction conditions. A, Peptide F1 in the absence of FIPV $3CL^{pro}$; B, peptide P7 in the absence of FIPV $3CL^{pro}$; C, peptide P8 in the absence of FIPV $3CL^{pro}$; D, peptide F1 in the presence of FIPV $3CL^{pro}$; E, peptide P7 in the presence of FIPV $3CL^{pro}$; F, peptide P8 in the presence of FIPV $3CL^{pro}$.

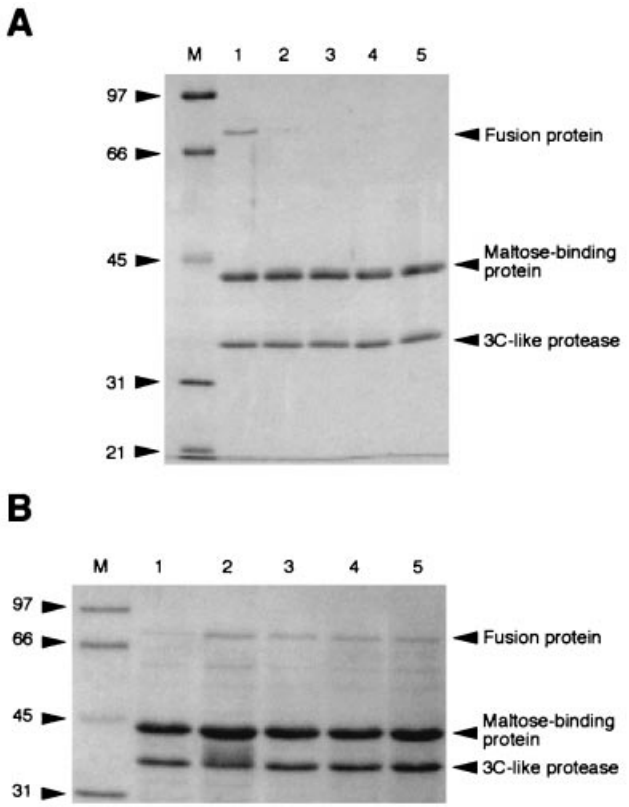

Fig. 5. Purification of recombinant FIPV and HCoV 3C-like proteases. (A) Purification and factor Xa cleavage of MBP–FIPV 3CL^{pro} fusion proteins carrying mutations of the predicted active-site residues His^{41} and Cys^{144}. The indicated fusion proteins were partially purified by amylose-affinity chromatography, treated with endoprotease Xa and analysed by SDS–PAGE on a 12·5% polyacrylamide gel stained with Coomassie brilliant blue. Lanes M, protein molecular mass markers (kDa); 1, MBP–FIPV 3CL^{pro} (wild-type); 2, MBP–FIPV 3CL^{pro} H^{41}Y; 3, MBP–FIPV 3CL^{pro} H^{41}R; 4, MBP–FIPV 3CL^{pro} C^{144}A; 5, MBP–FIPV 3CL^{pro} C^{144}S. (B) Purification and factor Xa cleavage of MBP–HCoV 3CL^{pro} fusion proteins in which Asn^{64} was replaced by Ala, Glu, Gln and Asp, respectively. The indicated fusion proteins were partially purified by amylose-affinity chromatography, treated with endoprotease Xa and analysed by SDS–PAGE on a 12·5% polyacrylamide gel stained with Coomassie brilliant blue. Lanes M, protein molecular mass markers (kDa); 1, MBP–HCoV 3CL^{pro} (wild-type); 2, MBP–HCoV 3CL^{pro} N^{64}A; 3, MBP–HCoV 3CL^{pro} N^{64}E; 4, MBP–HCoV 3CL^{pro} N^{64}Q; 5, MBP–HCoV 3CL^{pro} N^{64}D.

central role in substrate binding (Bazan & Fletterick, 1988; Gorbalenya *et al.*, 1989a; Bergmann *et al.*, 1997; Mosimann *et al.*, 1997). It has been predicted that coronavirus main proteases employ a deviant form of this signature, Tyr–X–His (Gorbalenya *et al.*, 1989b) (Fig. 2B). To date, this theoretical functional assignment has not been tested experimentally for any coronavirus 3CL^{pro}. The need for experimental data becomes even more evident when the unusual nature of the Gly → Tyr replacement and the very low overall similarity between coronavirus main proteases and other chymotrypsin-like enzymes (see above) are taken into account. We thus decided to target the FIPV 3CL^{pro} Tyr^{160} residue by site-specific mutagenesis.

Table 2. Enzymatic activities of mutant, recombinant FIPV and HCoV 3C-like proteases

Proteolytic activities were determined in parallel experiments using a peptide-based cleavage assay (see Methods). The sequences of the 15-mer substrate peptides were derived from the N-terminal autocatalytic cleavage sites of the FIPV and HCoV 3C-like proteases, respectively. For FIPV 3CL^{pro}, H_2N–VSVNSTL**QS**GLRKMA–COOH was used; for HCoV 3CL^{pro}, H_2N–VSYGSTL**QA**GLRKMA–COOH was used (Ziebuhr *et al.*, 1997). Residues shown in bold-face indicate the cleaved dipeptide bond. The activities of wild-type FIPV and HCoV 3C-like proteases were taken as 100% and the mean value of three experiments is shown.

Protein	Activity (%)
FIPV 3CL^{pro}	100
FIPV 3CL^{pro} H^{41}Y	< 1
FIPV 3CL^{pro} H^{41}R	< 1
FIPV 3CL^{pro} C^{144}A	< 1
FIPV 3CL^{pro} C^{144}S	< 1
FIPV 3CL^{pro} N^{64}A	135
FIPV 3CL^{pro} N^{64}D	108
FIPV 3CL^{pro} G^{83}P	5
FIPV 3CL^{pro} G^{83}V	7
FIPV 3CL^{pro} G^{83}W	32
FIPV 3CL^{pro} G^{83}R	25
FIPV 3CL^{pro} G^{83}T	32
FIPV 3CL^{pro} G^{83}E	58
FIPV 3CL^{pro} G^{83}A	95
FIPV 3CL^{pro} G^{83}D	113
FIPV 3CL^{pro} S^{137}A	22
FIPV 3CL^{pro} S^{137}T	64
FIPV 3CL^{pro} Y^{160}G	4
FIPV 3CL^{pro} Y^{160}F	3
FIPV 3CL^{pro} Y^{160}A	1
FIPV 3CL^{pro} Y^{160}T	< 1
FIPV 3CL^{pro} M^{161}A	114
FIPV 3CL^{pro} H^{162}A	< 1
FIPV 3CL^{pro} H^{162}L	< 1
FIPV 3CL^{pro} Y^{160}G–G^{83}D	< 1
FIPV 3CL^{pro} Y^{160}G–G^{83}A	< 1
HCoV 3CL^{pro}	100
HCoV 3CL^{pro} N^{64}A	140
HCoV 3CL^{pro} N^{64}D	120
HCoV 3CL^{pro} N^{64}E	276
HCoV 3CL^{pro} N^{64}Q	299

First, Tyr–X–His was converted into the canonical Gly–X–His constellation by the Tyr → Gly substitution. The mutant exhibited a strongly reduced activity towards the peptide substrate, indicating that the aromatic side chain of Tyr is involved (either directly or indirectly) in proteolytic activity. To test this hypothesis, we analysed the effect of a structurally conservative mutation by substituting Phe → Tyr (mutant Y^{160}F). Interestingly, as Table 2 shows, proteolytic activity was, again, strongly reduced to a level similar to that of the Y^{160}G mutant, suggesting a specific function of the polar hydroxyl group. We then constructed mutant Y^{160}T in order

to answer the question of whether the postulated function of the hydroxyl group is retained when attached to an aliphatic side chain. Obviously, the complete loss of activity in this mutant (Table 2) indicates both structural and functional constraints for the FIPV 3CL^{pro} amino acid residue at position 160 (and probably also the equivalent positions in other coronavirus main proteases). The specificity of this function is supported further by the fact that the neighbouring (and also conserved) Met^{161} residue can be substituted (M^{161}A) without significant effect on proteolytic activity. Our data thus lead us to propose that both the aromatic and hydroxyl properties of Tyr have been selected at this position. These properties, which cannot be delivered by the Gly residue conserved in other 3C-(like) proteases or another residue, may be involved in catalysis and/or substrate binding in a unique way.

One plausible role for the Tyr^{160} residue would be to take over the function of the third catalytic residue, which is predicted to be replaced by Gly^{83} (see above). On the basis of structural data, a similar model has been developed recently for another Tyr residue of a related 3C protease, the Tyr^{143} residue of HAV 3C^{pro} (Bergmann *et al.*, 1997). To test our hypothesis, we have characterized the double mutant Y^{160}G–G^{83}D, in which two unique positions of the coronavirus main proteases were converted into the canonical form, and a negative control to this mutant, Y^{160}G–G^{83}A. Both mutants proved to be inactive (Table 2), indicating that either our hypothesis is not correct or other mutations are required to make this conversion functional.

We then addressed the function of two additional residues proposed previously to be involved in substrate recognition. First, we exchanged Leu and Ala for the conserved FIPV 3CL^{pro} His^{162} residue. In both cases, the proteolytic activities dropped below the detection limit of the assay used in this study. The obvious indispensability of His^{162} for proteolytic activity supports the conclusions drawn from sequence comparisons (Gorbalenya *et al.*, 1989b; Lee *et al.*, 1991; Herold *et al.*, 1993; Eleouet *et al.*, 1995) and is also consistent with mutagenesis data obtained previously for the equivalent His residue in HCoV 3CL^{pro} (Ziebuhr *et al.*, 1997). It is thus reasonable to believe that the His^{162} residue is the functional equivalent to the His^{191} residue shown to be present in the S1-specificity pocket of HAV 3C^{pro} (Bergmann *et al.*, 1997). Second, we mutagenized the FIPV 3CL^{pro} Ser^{138} residue. This residue has been predicted to be functionally equivalent to the Thr^{141} residue of human rhinovirus type 14 3C^{pro} (Gorbalenya & Snijder, 1996) and the Thr^{142} residue of poliovirus 3C^{pro}, both of which have been proposed to be involved in the formation of the S1 subsite of the substrate-binding pocket by hydrogen-bonding to the carboxamide side chain of the P1 Gln residue (together with the conserved His residue mentioned above) (Matthews *et al.*, 1994; Mosimann *et al.*, 1997). Both mutants, FIPV 3CL^{pro} S^{138}A and S^{138}T, had reduced proteolytic activities. Importantly, the Ser → Thr mutant retaining a hydroxyl group was significantly more active than the Ser → Ala mutant retaining the side-chain size (Table 2). Similar results were obtained in a previous mutagenesis study in which the equivalent Thr^{1179} residue of equine arteritis virus 3C-like nsp4 protease was shown to tolerate substitutions without loss of proteolytic activity in most cases (Snijder *et al.*, 1996). Some of these mutants, however, proved to be unable to cleave some but not other cleavage sites, which strongly suggested an involvement of the Thr^{1179} residue in substrate recognition (rather than catalysis). But, clearly, more experiments are needed to determine conclusively the functions of His^{162} and Ser^{138}. Specifically, genetic data, including a wide range of substrates, and structural data need to be obtained.

Concluding remarks

In this study, we have presented further theoretical and experimental evidence that coronavirus main proteases have evolved unique structural and functional characteristics in the framework of a chymotrypsin-like fold. Using an *in vitro* assay, we have characterized more than two dozen FIPV and HCoV 3CL^{pro} point mutations at eight very conserved positions thought to be part of the active centre. The observed effects on the proteolytic activities of the respective enzymes were rationalized within a model derived from a broad comparison of 3C-(like) proteases of coronaviruses and other viruses, especially potyviruses. The model, which is consistent with the obtained data, suggests that coronavirus main proteases use a catalytic Cys–His dyad that is not assisted by an acidic residue in the canonical sequence position. In addition to the Ser (Thr) and His residues present commonly in the substrate pocket of 3C-(like) proteases, coronavirus enzymes also employ the unique properties of a Tyr residue that replaces an otherwise conserved Gly residue. Finally, specific substitutions have been identified that confer significantly increased proteolytic activities on coronavirus main proteases. It will be interesting to test the *in vivo* effects of such 3CL^{pro} mutants using the reverse-genetic systems established recently (Almazán *et al.*, 2000; Yount *et al.*, 2000; Thiel *et al.*, 2001a). Future studies, including structure analyses, are required to elucidate the peculiarities of the coronavirus 3CL^{pro} active site in greater detail.

We thank Andreas Kolb for providing FIPV strain 79-1146. The work was supported by grants from the Deutsche Forschungsgemeinschaft (GK Infektiologie, Zi 618/2-1, 2-2) and the Bayerische Forschungsstiftung (Neue Antiinfektiva). A.E.G. was supported with funds from the National Cancer Institute, National Institutes of Health, under contract no. NO1-CO-56000. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services nor does mention of trade names, commercial products or organization imply endorsement by the US Government.

References

Allaire, M., Chernaia, M. M., Malcolm, B. A. & James, M. N. (1994). Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* **369**, 72–76.

Almazán, F., González, J. M., Pénzes, Z., Izeta, A., Calvo, E., Plana-Durán, J. & Enjuanes, L. (2000). Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proceedings of the National Academy of Sciences, USA* **97**, 5516–5521.

Baker, S. C., Shieh, C. K., Soe, L. H., Chang, M. F., Vannier, D. M. & Lai, M. M. (1989). Identification of a domain required for autoproteolytic cleavage of murine coronavirus gene A polyprotein. *Journal of Virology* **63**, 3693–3699.

Bazan, J. F. & Fletterick, R. J. (1988). Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proceedings of the National Academy of Sciences, USA* **85**, 7872–7876.

Bergmann, E. M., Mosimann, S. C., Chernaia, M. M., Malcolm, B. A. & James, M. N. (1997). The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. *Journal of Virology* **71**, 2436–2448.

Bonilla, P. J., Hughes, S. A. & Weiss, S. R. (1997). Characterization of a second cleavage site and demonstration of activity *in trans* by the papain-like proteinase of the murine coronavirus mouse hepatitis virus strain A59. *Journal of Virology* **71**, 900–909.

Boniotti, B., Wirblich, C., Sibilia, M., Meyers, G., Thiel, H. J. & Rossi, C. (1994). Identification and characterization of a 3C-like protease from rabbit hemorrhagic disease virus, a calicivirus. *Journal of Virology* **68**, 6487–6495.

Boursnell, M. E. G., Brown, T. D. K., Foulds, I. J., Green, P. F., Tomley, F. M. & Binns, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *Journal of General Virology* **68**, 57–77.

Brierley, I., Boursnell, M. E., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D. & Inglis, S. C. (1987). An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO Journal* **6**, 3779–3785.

Cavanagh, D. (1997). *Nidovirales*: a new order comprising *Coronaviridae* and *Arteriviridae*. *Archives of Virology* **142**, 629–633.

Chouljenko, V. N., Lin, X. Q., Storz, J., Kousoulas, K. G. & Gorbalenya, A. E. (2001). Comparison of genomic and predicted amino acid sequences of respiratory and enteric bovine coronaviruses isolated from the same animal with fatal shipping pneumonia. *Journal of General Virology* **82**, 2927–2933.

Cowley, J. A., Dimmock, C. M., Spann, K. M. & Walker, P. J. (2000). Gill-associated virus of *Penaeus monodon* prawns: an invertebrate virus with ORF1a and ORF1b genes related to arteri- and coronaviruses. *Journal of General Virology* **81**, 1473–1484.

de Groot, R. J. & Horzinek, M. C. (1995). Feline infectious peritonitis. In *The Coronaviridae*, pp. 293–315. Edited by S. G. Siddell. New York: Plenum Press.

den Boon, J. A., Snijder, E. J., Chirnside, E. D., de Vries, A. A., Horzinek, M. C. & Spaan, W. J. (1991). Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. *Journal of Virology* **65**, 2910–2920.

Denison, M. R., Spaan, W. J., van der Meer, Y., Gibson, C. A., Sims, A. C., Prentice, E. & Lu, X. T. (1999). The putative helicase of the coronavirus mouse hepatitis virus is processed from the replicase gene polyprotein and localizes in complexes that are active in viral RNA synthesis. *Journal of Virology* **73**, 6862–6871.

Dessens, J. T. & Lomonossoff, G. P. (1991). Mutational analysis of the putative catalytic triad of the cowpea mosaic virus 24K protease. *Virology* **184**, 738–746.

Eleouet, J. F., Rasschaert, D., Lambert, P., Levy, L., Vende, P. & Laude, H. (1995). Complete sequence (20 kilobases). of the polyprotein-encoding gene 1 of transmissible gastroenteritis virus. *Virology* **206**, 817–822.

Gorbalenya, A. E. & Koonin, E. V. (1993). Comparative analysis of the amino acid sequences of the key enzymes of the replication and expression of positive-strand RNA viruses. Validity of the approach and functional and evolutionary implications. *Soviet Scientific Reviews Section D Physicochemical Biology Reviews* **11**, 1–84.

Gorbalenya, A. E. & Snijder, E. J. (1996). Viral cysteine proteinases. *Perspectives in Drug Discovery and Design* **6**, 64–86.

Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M. & Koonin, E. V. (1989a). Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. *FEBS Letters* **243**, 103–114.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1989b). Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Research* **17**, 4847–4861.

Gorbalenya, A. E., Koonin, E. V. & Lai, M. M. (1991). Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and apthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. *FEBS Letters* **288**, 201–205.

Grötzinger, C., Heusipp, G., Ziebuhr, J., Harms, U., Süss, J. & Siddell, S. G. (1996). Characterization of a 105-kDa polypeptide encoded in gene 1 of the human coronavirus HCV 229E. *Virology* **222**, 227–235.

Grubman, M. J., Zellner, M., Bablanian, G., Mason, P. W. & Piccone, M. E. (1995). Identification of the active-site residues of the 3C proteinase of foot-and-mouth disease virus. *Virology* **213**, 581–589.

Hellen, C. U., Facke, M., Kräusslich, H. G., Lee, C. K. & Wimmer, E. (1991). Characterization of poliovirus 2A proteinase by mutational analysis: residues required for autocatalytic activity are essential for induction of cleavage of eukaryotic initiation factor 4F polypeptide p220. *Journal of Virology* **65**, 4226–4231.

Henikoff, S. & Henikoff, J. G. (1994). Position-based sequence weights. *Journal of Molecular Biology* **243**, 574–578.

Herold, J., Raabe, T., Schelle-Prinz, B. & Siddell, S. G. (1993). Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. *Virology* **195**, 680–691.

Herold, J., Siddell, S. & Ziebuhr, J. (1996). Characterization of coronavirus RNA polymerase gene products. *Methods in Enzymology* **275**, 68–89.

Herold, J., Gorbalenya, A. E., Thiel, V., Schelle, B. & Siddell, S. G. (1998). Proteolytic processing at the amino terminus of human coronavirus 229E gene 1-encoded polyproteins: identification of a papain-like proteinase and its substrate. *Journal of Virology* **72**, 910–918.

Herold, J., Siddell, S. G. & Gorbalenya, A. E. (1999). A human RNA viral cysteine proteinase that depends upon a unique Zn^{2+}-binding finger connecting the two domains of a papain-like fold. *Journal of Biological Chemistry* **274**, 14918–14925.

Herrewegh, A. A., Vennema, H., Horzinek, M. C., Rottier, P. J. & de Groot, R. J. (1995). The molecular genetics of feline coronaviruses: comparative sequence analysis of the ORF7a/7b transcription unit of different biotypes. *Virology* **212**, 622–631.

Herrewegh, A. A., Smeenk, I., Horzinek, M. C., Rottier, P. J. & de Groot, R. J. (1998). Feline coronavirus type II strains 79-1683 and 79-1146 originate from a double recombination between feline coronavirus type I and canine coronavirus. *Journal of Virology* **72**, 4508–4514.

Heusipp, G., Grötzinger, C., Herold, J., Siddell, S. G. & Ziebuhr, J. (1997a). Identification and subcellular localization of a 41 kDa, poly-

protein 1ab processing product in human coronavirus 229E-infected cells. *Journal of General Virology* **78**, 2789–2794.

Heusipp, G., Harms, U., Siddell, S. G. & Ziebuhr, J. (1997b). Identification of an ATPase activity associated with a 71-kilodalton polypeptide encoded in gene 1 of the human coronavirus 229E. *Journal of Virology* **71**, 5631–5634.

Horzinek, M. C., Lutz, H. & Pedersen, N. C. (1982). Antigenic relationships among homologous structural polypeptides of porcine, feline, and canine coronaviruses. *Infection and Immunity* **37**, 1148–1155.

Jacobse-Geels, H. E. L. & Horzinek, M. C. (1983). Expression of feline infectious peritonitis coronavirus antigens on the surface of feline macrophage-like cells. *Journal of General Virology* **64**, 1859–1866.

Kanjanahaluethai, A. & Baker, S. C. (2000). Identification of mouse hepatitis virus papain-like proteinase 2 activity. *Journal of Virology* **74**, 7911–7921.

Kean, K. M., Teterina, N. L., Marc, D. & Girard, M. (1991). Analysis of putative active site residues of the poliovirus 3C protease. *Virology* **181**, 609–619.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*, p. 367. Cambridge, NY: Cambridge University Press.

Koonin, E. V. & Dolja, V. V. (1993). Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Critical Reviews in Biochemistry and Molecular Biology* **28**, 375–430.

Lawson, M. A. & Semler, B. L. (1991). Poliovirus thiol proteinase 3C can utilize a serine nucleophile within the putative catalytic triad. *Proceedings of the National Academy of Sciences, USA* **88**, 9919–9923.

Lee, H. J., Shieh, C. K., Gorbalenya, A. E., Koonin, E. V., La Monica, N., Tuler, J., Bagdzhadzhyan, A. & Lai, M. M. (1991). The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. *Virology* **180**, 567–582.

Lim, K. P., Ng, L. F. & Liu, D. X. (2000). Identification of a novel cleavage activity of the first papain-like proteinase domain encoded by open reading frame 1a of the coronavirus avian infectious bronchitis virus and characterization of the cleavage products. *Journal of Virology* **74**, 1674–1685.

Liu, D. X. & Brown, T. D. (1995). Characterisation and mutational analysis of an ORF 1a-encoding proteinase domain responsible for proteolytic processing of the infectious bronchitis virus 1a/1b polyprotein. *Virology* **209**, 420–427.

Liu, D. X., Brierley, I., Tibbles, K. W. & Brown, T. D. (1994). A 100-kilodalton polypeptide encoded by open reading frame (ORF) 1b of the coronavirus infectious bronchitis virus is processed by ORF 1a products. *Journal of Virology* **68**, 5772–5780.

Liu, D. X., Shen, S., Xu, H. Y. & Wang, S. F. (1998). Proteolytic mapping of the coronavirus infectious bronchitis virus 1b polyprotein: evidence for the presence of four cleavage sites of the 3C-like proteinase and identification of two novel cleavage products. *Virology* **246**, 288–297.

Lu, Y. & Denison, M. R. (1997). Determinants of mouse hepatitis virus 3C-like proteinase activity. *Virology* **230**, 335–342.

Lu, Y., Lu, X. & Denison, M. R. (1995). Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. *Journal of Virology* **69**, 3554–3559.

Lu, X. T., Sims, A. C. & 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. *Journal of Virology* **72**, 2265–2271.

Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L. and others (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell* **77**, 761–771.

Merrifield, R. B. (1965). Automated synthesis of peptides. *Science* **150**, 178–185.

Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S. & James, M. N. (1997). Refined X-ray crystallographic structure of the poliovirus 3C gene product. *Journal of Molecular Biology* **273**, 1032–1047.

Ng, L. F. & Liu, D. X. (2000). Further characterization of the coronavirus infectious bronchitis virus 3C-like proteinase and determination of a new cleavage site. *Virology* **272**, 27–39.

Olsen, C. W. (1993). A review of feline infectious peritonitis virus: molecular biology, immunopathogenesis, clinical aspects, and vaccination. *Veterinary Microbiology* **36**, 1–37.

Page, R. D. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357–358.

Rost, B. (1996). PhD: predicting one-dimensional protein structure by profile-based neural networks. *Methods in Enzymology* **266**, 525–539.

Rost, B., Casadio, R., Fariselli, P. & Sander, C. (1995). Transmembrane helices predicted at 95% accuracy. *Protein Science* **4**, 521–533.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

Sawicki, S. G. & Sawicki, D. L. (1998). A new model for coronavirus transcription. *Advances in Experimental Medicine and Biology* **440**, 215–219.

Sawicki, D. L, Wang, T. & Sawicki, S. G. (2001). The RNA structures engaged in replication and transcription of the A59 strain of mouse hepatitis virus. *Journal of General Virology* **82**, 385–396.

Seybert, A., Ziebuhr, J. & Siddell, S. G. (1997). Expression and characterization of a recombinant murine coronavirus 3C-like proteinase. *Journal of General Virology* **78**, 71–75.

Seybert, A., Hegyi, A., Siddell, S. G. & Ziebuhr, J. (2000a). The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5′-to-3′ polarity. *RNA* **6**, 1056–1068.

Seybert, A., van Dinten, L. C., Snijder, E. J. & Ziebuhr, J. (2000b). Biochemical characterization of the equine arteritis virus helicase suggests a close functional relationship between arterivirus and coronavirus helicases. *Journal of Virology* **74**, 9586–9593.

Siddell, S. G. (1995). The *Coronaviridae*: an introduction. In *The Coronaviridae*, pp. 1–10. Edited by S. G. Siddell. New York: Plenum Press.

Snijder, E. J., Wassenaar, A. L., van Dinten, L. C., Spaan, W. J. & Gorbalenya, A. E. (1996). The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases. *Journal of Biological Chemistry* **271**, 4864–4871.

Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B. A. & Siddell, S. G. (1983). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO Journal* **2**, 1839–1844.

Stephensen, C. B., Casebolt, D. B. & Gangopadhyay, N. N. (1999). Phylogenetic analysis of a highly conserved region of the polymerase gene from 11 coronaviruses and development of a consensus polymerase chain reaction assay. *Virus Research* **60**, 181–189.

Swofford, D. L. (2000). PAUP*: Phylogenetic Analysis using Parsimony, version 4. Sunderland: Sinauer Associates.

Thiel, V., Herold, J., Schelle, B. & Siddell, S. G. (2001a). Infectious

RNA transcribed *in vitro* from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *Journal of General Virology* **82**, 1273–1281.

Thiel, V., Herold, J., Schelle, B. & Siddell, S. G. (2001b). Viral replicase gene products suffice for coronavirus discontinuous transcription. *Journal of Virology* **75**, 6676–6681.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Computer Applications in the Biosciences* **10**, 19–29.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876–4882.

van Marle, G., Dobbe, J. C., Gultyaev, A. P., Luytjes, W., Spaan, W. J. & Snijder, E. J. (1999). Arterivirus discontinuous mRNA transcription is guided by base pairing between sense and antisense transcription-regulating sequences. *Proceedings of the National Academy of Sciences, USA* **96**, 12056–12061.

Vennema, H., Rossen, J. W., Wesseling, J., Horzinek, M. C. & Rottier, P. J. (1992). Genomic organization and expression of the 3′ end of the canine and feline enteric coronaviruses. *Virology* **191**, 134–140.

Vennema, H., Poland, A., Foley, J. & Pedersen, N. C. (1998). Feline infectious peritonitis viruses arise by mutation from endemic feline enteric coronaviruses. *Virology* **243**, 150–157.

Yao, Z., Jones, D. H. & Grose, C. (1992). Site-directed mutagenesis of herpesvirus glycoprotein phosphorylation sites by recombination polymerase chain reaction. *PCR Methods & Applications* **1**, 205–207.

Yount, B., Curtis, K. M. & Baric, R. S. (2000). Strategy for systematic assembly of large RNA and DNA genomes: transmissible gastroenteritis virus model. *Journal of Virology* **74**, 10600–10611.

Yu, S. F. & Lloyd, R. E. (1991). Identification of essential amino acid residues in the functional activity of poliovirus 2A protease. *Virology* **182**, 615–625.

Ziebuhr, J. & Siddell, S. G. (1999). Processing of the human coronavirus 229E replicase polyproteins by the virus-encoded 3C-like proteinase: identification of proteolytic products and cleavage sites common to pp1a and pp1ab. *Journal of Virology* **73**, 177–185.

Ziebuhr, J., Herold, J. & Siddell, S. G. (1995). Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *Journal of Virology* **69**, 4331–4338.

Ziebuhr, J., Heusipp, G. & Siddell, S. G. (1997). Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *Journal of Virology* **71**, 3992–3997.

Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. (2000). Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. *Journal of General Virology* **81**, 853–879.

Ziebuhr, J., Thiel, V. & Gorbalenya, A. E. (2001). 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. *Journal of Biological Chemistry* **276**, 33220–33232.

Received 18 September 2001; Accepted 9 November 2001