JOURNAL OF VIROLOGY, May 1997, p. 3992–3997 Vol. 71, No. 5
0022-538X/97/$04.00+0

Biosynthesis, Purification, and Characterization of the Human Coronavirus 229E 3C-Like Proteinase

JOHN ZIEBUHR,* GERHARD HEUSIPP, AND STUART G. SIDDELL

Institute of Virology, University of Würzburg, 97078 Würzburg, Germany

Received 17 December 1996/Accepted 14 February 1997

Coronavirus gene expression involves proteolytic processing of the gene 1-encoded polyprotein(s), and a key enzyme in this process is the viral 3C-like proteinase. In this report, we describe the biosynthesis of the human coronavirus 229E 3C-like proteinase in *Escherichia coli* and the enzymatic properties, inhibitor profile, and substrate specificity of the purified protein. Furthermore, we have introduced single amino acid substitutions and carboxyl-terminal deletions into the recombinant protein and determined the ability of these mutant 3C-like proteinases to catalyze the cleavage of a peptide substrate. Using this approach, we have identified the residues Cys-3109 and His-3006 as being indispensable for catalytic activity. Our results also support the involvement of His-3127 in substrate recognition, and they confirm the requirement of the carboxyl-terminal extension found in coronavirus 3C-like proteinases for enzymatic activity. These data provide experimental evidence for the relationship of coronavirus 3C-like proteinases to other viral chymotrypsin-like enzymes, but they also show that the coronavirus proteinase has additional, unique properties.

Human coronaviruses (HCV) are a major cause of respiratory disease in both children and adults. HCV infection is usually manifested as a relatively mild illness, the "common cold," but the involvement of HCV in lower respiratory tract illness and gastroenteritis has also been documented (15, 25, 32). To date, two antigenic groups of HCV have been identified, and they are represented by the prototype strains HCV 229E and HCV OC43. HCV 229E is easier to isolate and propagate in tissue culture and, consequently, has been more extensively studied.

The HCV 229E genome is a positive-strand RNA of approximately 27,000 nucleotides. Gene 1, which is located at the 5′ end of the genome, is comprised of two large, overlapping open reading frames (ORFs), ORF 1a and ORF 1b (12). The upstream ORF, ORF 1a, encodes a polyprotein, pp1a, with a calculated molecular mass of 454 kDa. The downstream ORF, ORF 1b, is expressed as a fusion protein with pp1a by a mechanism involving (−1) ribosomal frameshifting (12, 13). The ORF 1ab gene product has a calculated molecular mass of 754 kDa and is referred to as polyprotein 1ab or pp1ab. Analysis of murine coronavirus (MHV) temperature-sensitive mutants (2) indicates that most, if not all, of the functions necessary for coronavirus RNA synthesis are located in pp1a- and pp1ab-related polypeptides.

Proteolytic processing of viral polyproteins is a crucial step in the replication cycle of many positive-strand RNA viruses (5, 16). By and large, these processing events are carried out by virus-encoded proteinases, and in the case of coronaviruses, sequence motifs characteristic of both papain-like cysteine proteinases and a chymotrypsin-like enzyme, the 3C-like proteinase ($3CL^{pro}$), have been identified in the regions of pp1a and pp1ab encoded by ORF 1a (7, 10, 12, 17). The coronavirus 3C-like proteinase derives its name from the observed homology to the 3C cysteine proteinase of picornaviruses, and both active site residues and sequence elements that may be involved in substrate binding have been predicted for the coronavirus protein (10). In vitro-translated and bacterially synthesized MHV and HCV 229E $3CL^{pro}$ have both been used to obtain preliminary data on the substrate specificity of the coronavirus enzyme (11, 20, 21, 28, 33). The sequence analysis of four cleavage products showed that, as had been previously predicted (10, 12, 17), cleavage takes place at Gln┼Ser,Ala dipeptides (crossbar symbol indicates peptide bond cleavage site) (11, 20, 33). This specificity is consistent with features regarded as typical for substrates of 3C proteinases, namely cleavage at Q,E┼G,S,A dipeptides (26).

There are, however, several features that clearly separate the coronavirus $3CL^{pro}$ from other viral proteinases of this type. First, compared to most 3C and 3C-like proteinases, which have a molecular mass of approximately 18 to 22 kDa, the coronavirus $3CL^{pro}$ has a molecular mass of about 30 kDa. The additional mass is mainly due to an extended carboxyl-terminal domain. Second, in the coronavirus $3CL^{pro}$, a conserved Gly (Ala) residue, which is characteristically located close to the His residue of the 3C and 3C-like proteinase substrate-binding pocket, has been replaced with a Tyr residue (10). Third, at least to date, it has not been possible to identify an acidic active site residue in the coronavirus $3CL^{pro}$ (7, 12, 17, 18). This residue is normally required for chymotrypsin-like catalysis, and its spatial equivalent is predicted to be present in, for example, most picornavirus 3C proteinases. In this context, it is interesting to note, however, that recent structural studies on the hepatitis A virus 3C proteinase (HAV-$3C^{pro}$) have shown that the Asp residue which had been predicted to be equivalent to the active site Asp of chymotrypsin (9) is properly located, but because of a different side chain orientation, it does not interact with the catalytic His residue. Therefore, it has been suggested that the thiolate-imidazolium system might be sufficient for catalysis by some 3C proteinases (1). This mechanism is widely accepted for the papain family of cysteine proteinases (6).

In the case of HCV 229E, the $3CL^{pro}$ region is located between amino acids 2966 and 3267 of pp1a and pp1ab. In the virus-infected cell, the major form of $3CL^{pro}$ is found as a polypeptide with an apparent molecular mass of 34 kDa (33). Recently, we have shown that bacterially synthesized HCV 229E $3CL^{pro}$ exhibits proteolytic activity in *trans*, using both in

* Corresponding author. Mailing address: Institute of Virology, University of Würzburg, Versbacher Strasse 7, 97078 Würzburg, Germany. Phone: 49-931-2013966. Fax: 49-931-2013934. E-mail: j.ziebuhr@rzbox.uni-wuerzburg.de.

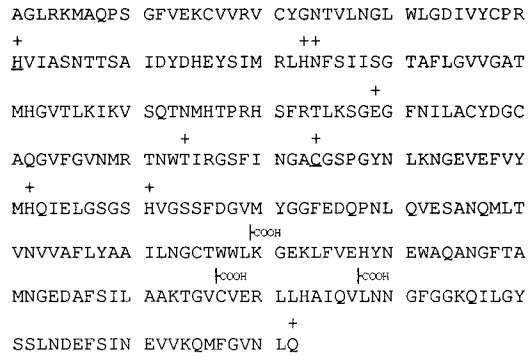

FIG. 1. Amino acid sequence of the HCV 229E 3C-like proteinase domain. The proteinase domain encompasses amino acids 2966 to 3267 of pp1a and pp1ab. Amino acids which have been exchanged in the present study and the termini of the carboxyl-terminally truncated recombinant proteins are indicated by symbols above the amino acid sequence. The predicted catalytic Cys and His residues (10, 12) are underlined.

vitro-translated and bacterially synthesized substrates (11, 33). In this report we describe a more extensive characterization of the HCV 229E 3CL^{pro}, including the enzymatic properties, the inhibitor profile, and the substrate specificity of the purified protein. Our results confirm the relationship of coronavirus 3C-like proteinases to the other viral chymotrypsin-like enzymes, but they also clearly show that the coronavirus 3CL^{pro} has additional unique properties.

MATERIALS AND METHODS

Protein biosynthesis and purification. The construction of plasmid pMalc2-3CL and the bacterial biosynthesis of HCV 229E 3CL^{pro} have been described previously (33). Briefly, a fusion protein, MBP-3CL^{pro}, containing the maltose-binding protein (MBP) of *E. coli* and the predicted 3CL^{pro} domain of HCV 229E (amino acids 2966 to 3267 of pp1a and pp1ab, Fig. 1) was synthesized in bacteria and then purified by amylose-affinity chromatography. The authentic 3CL^{pro} was released from the fusion protein by factor Xa cleavage, and the protein mixture was loaded onto a phenyl-Sepharose HP column (Pharmacia Biotech, Freiburg, Germany) that had been preequilibrated with a solution containing 12.5 mM Bis-Tris-HCl (pH 7.0), 300 mM NaCl, 1 mM dithiothreitol, and 0.1 mM EDTA. After extensive washing of the column, the recombinant 3CL^{pro} was eluted with 12.5 mM Bis-Tris-HCl (pH 7.0)–1 mM dithiothreitol–0.1 mM EDTA. The fractions containing 3CL^{pro} were pooled and concentrated to 3 mg/ml with Centricon-3 concentrators (Amicon, Beverly, Mass.). The concentrated material was purified further by chromatography on a Superdex 75-pg (particle size) column (Pharmacia Biotech) run under isocratic conditions with 10 mM Tris-HCl (pH 7.3)–200 mM NaCl–0.1 mM EDTA–1 mM dithiothreitol. Finally, the purified protein was concentrated to 10 mg/ml with Centricon-3 concentrators and stored at −80°C. The protein concentration was measured by the method of Lowry et al. (19) with bovine serum albumin as a standard.

Site-directed mutagenesis of HCV 229E 3CL^{pro}. Site-directed mutagenesis was done by a recombination-PCR method as described by Yao et al. (31). Briefly, four partially complementary oligonucleotides were used to generate linear fragments from pMalc2-3CL DNA by PCR. These fragments were combined and used to transform competent *E. coli* DH5α cells. Recombination in vivo generated circular plasmids carrying the desired mutation. After sequence analysis, the mutant proteinase-encoding sequences were exchanged with the corresponding sequence of the parental plasmid. Mutant HCV 229E 3CL^{pro} was synthesized and affinity purified as described above but was not purified by hydrophobic interaction and size-exclusion chromatography. The purity and structural integrity of each of the mutant proteins was verified by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis. The control protein for this experiment, wild-type 3CL^{pro}, was purified in an identical manner.

Peptide synthesis. Synthetic peptides were prepared by solid-phase chemistry (24) and purified by high-performance liquid chromatography (HPLC) on a reversed-phase C_{18} silica column (Jerini Bio-Tools, Berlin, Germany). The identity and homogeneity of the peptides were confirmed by mass spectrometry and analytical reversed-phase chromatography.

Peptide cleavage. Cleavage reactions were incubated at 25°C and contained 20 mM Bis-Tris-HCl (pH 7.0), varying concentrations of proteinase, and 0.5 mM substrate peptide in a total volume of 20 μl. The reactions were terminated by the addition of 80 μl of 0.1% trifluoroacetic acid, and the mixture was centrifuged for 5 min at 14,000 × *g* prior to analysis by reversed-phase HPLC on a Delta Pak C_{18} column (3.9 by 150 mm; Waters, Milford, Mass.). Cleavage products were resolved using a 22-min, 5 to 90% linear gradient of acetonitrile in 0.1% trifluoroacetic acid. The absorbance was determined at 215 nm. Peak areas were calculated by integration and converted to absolute units by using peptide standards. The peptide substrate used for K_m determination and inhibitor studies was NH_2-V-S-Y-G-S-T-L-Q┼A-G-L-R-K-M-A-COOH, representing amino acids 2958 to 2972 of pp1a and pp1ab. For the calculation of K_m, 0.15 μM purified 3CL^{pro} was used, and reaction aliquots were removed at 0.5, 1, 2, 3, and 5 min. For the inhibitor studies, 0.15 μM 3CL^{pro} in 20 mM Bis-Tris-HCl, pH 7.0, was preincubated with varying concentrations of proteinase inhibitor for 60 min at 25°C. Thereafter, the peptide substrate was added (0.5 mM final concentration), and the extent of cleavage was calculated after 5 min of incubation at 25°C. Cleavage reactions lacking the proteinase inhibitor were defined as 100% cleavage activity.

The enzymatic activity of mutant 3CL^{pro} (1 μg of total protein) was determined by incubation with 0.5 mM substrate peptide (NH_2-V-S-Y-G-S-T-L-Q┼A-G-L-R-K-M-A-COOH) for 2 h at 25°C. The amount of substrate conversion was determined for each reaction by HPLC analysis as described above. The activity of wild-type 3CL^{pro} was taken to be 100%. The substrate specificity of the HCV 229E 3CL^{pro} was tested by the incubation of different peptides (see Table 3, 0.5 mM end concentration) with 1 μM purified enzyme in 20 mM Bis-Tris-HCl, pH 7.0, for 2 h at 25°C. Under these conditions, the conversion of functional substrate peptides was complete. Nonfunctional substrate peptides did not show detectable hydrolysis under the same conditions.

RESULTS

Biosynthesis, purification, and kinetic analysis of recombinant HCV 229E 3CL^{pro}. We have recently reported the synthesis of a biologically active HCV 229E 3CL^{pro} in *E. coli* (14, 33). In the reported experiments, we used partially purified enzyme and were able to demonstrate its *trans* activity using in vitro-translated substrates. In the present study, we have extended our purification scheme to obtain a homogeneous preparation of recombinant HCV 229E 3CL^{pro}, and we have established a peptide cleavage assay that allows us to do quantitative biochemical analyses.

After the induction of synthesis (Fig. 2, lane 1) and cell lysis, the MBP-3CL^{pro} fusion protein was extracted by amylose-affinity chromatography (Fig. 2, lane 2). The fusion protein was then cleaved with factor Xa (Fig. 2, lane 3), and the MBP was quantitatively removed by hydrophobic interaction chromatography on phenyl-Sepharose HP (Fig. 2, lane 4). Finally, residual contaminants were removed by size-exclusion chromatography on a Superdex 75-pg column (Fig. 2, lanes 5 and 6). The purified protein could be concentrated to 12 mg/ml in 10 mM Tris-HCl (pH 7.3)–200 mM NaCl–1 mM dithiothreitol–0.1 mM EDTA, and 5 to 10 mg of recombinant 3CL^{pro} could be obtained from 2 liters of cells. The purified protein reacted specifically with polyclonal 3CL^{pro} antiserum (data not shown [33]) and did not lose enzymatic activity upon storage at −80°C for several months.

For the kinetic analysis of recombinant HCV 229E 3CL^{pro} activity, the synthetic peptide NH_2-V-S-Y-G-S-T-L-Q┼A-G-L-R-K-M-A-COOH was incubated with the purified protein. This 15-mer peptide represents amino acids 2958 to 2972 of pp1a and pp1ab and contains the scissile bond that releases the amino terminus of 3CL^{pro} from the viral polyproteins. Our experiments with in vitro-translated and bacterially synthesized substrates indicate that this substrate is rapidly cleaved by 3CL^{pro}. Initial rate measurements (consuming <25% of the substrate peptide) with three different 3CL^{pro} preparations yielded an apparent K_m value of 0.39 ± 0.07 mM (Fig. 3), assuming the enzyme preparations were 100% active.

Catalytic system of the HCV 229E 3CL^{pro}. Comparative sequence analyses have led to predictions about the importance of specific amino acid residues for the biological activity of 3C and 3C-like proteinases. For example, it has been suggested that His-3006 and Cys-3109 (see Fig. 1) might represent the active site residues of the HCV 229E 3CL^{pro} (9, 10, 12). To test these predictions, we have introduced a series of single

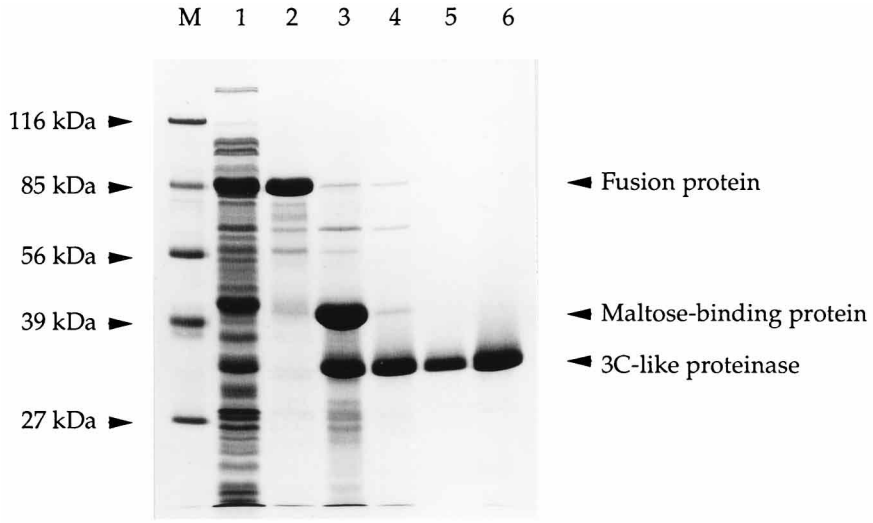

FIG. 2. Purification of recombinant HCV 229E 3C-like proteinase. Aliquots taken at each step of the purification protocol were analyzed on a SDS–12.5% polyacrylamide gel, and the proteins were stained with Coomassie brilliant blue. Lanes: M, protein molecular mass markers; 1, cleared lysate of isopropyl-β-D-thiogalactopyranoside-induced TB1 bacteria with the expression plasmid pMalc2-3CL; 2, pooled peak fractions from the amylose-affinity chromatography column; 3, factor Xa-cleaved MBP-3CL^{pro} fusion protein; 4, pooled peak fractions from the phenyl-Sepharose HP column; 5 and 6, 5 and 20 μg, respectively, of pooled peak fractions from the Superdex 75-pg column after concentration to 10 mg/ml. The MBP-3CL^{pro} fusion protein, MBP, and 3CL^{pro} are indicated.

amino acid substitutions into bacterially synthesized, recombinant HCV 229E 3CL^{pro}. The mutant proteins were partially purified, and their activity was determined in the peptide-based assay system described above. As expected, we have found mutations that do not affect enzymatic activity and mutations that render the proteinases partially or completely inactive. Representative mutations of these three types are shown in Fig. 4. Table 1 gives a summary of all the mutations that have been analyzed.

Previously, we have shown that the replacement of the predicted active site residue His-3006 by Tyr completely abolishes the enzymatic activity of recombinant HCV 229E 3CL^{pro} (33). The data of the present study suggest that His-3006 tolerates no exchange. Substitutions by Gly, Ser, or Thr completely inactivate the proteinase. In contrast, the neighboring His-3028, which is conserved in the 3CL^{pro} of HCV 229E and the related porcine transmissible gastroenteritis virus, can be replaced by Gly or Thr without loss of activity. These data provide genetic evidence that His-3006 most likely represents the general base residue required for catalysis.

Our experiments have also shown that replacement of the predicted nucleophilic Cys-3109 residue with either Pro, Ser, or Val results in a proteolytically inactive molecule. This datum strongly supports the identification of this residue as the nucleophilic amino acid. In this respect, it is interesting to note that we did not observe any residual activity of the C3109S mutant, as has been reported for the avian infectious bronchitis virus 3CL^{pro} (18, 30).

In order to investigate the possible requirement for a catalytic acidic residue in the HCV 229E 3CL^{pro}, we have made a number of substitutions. First, we replaced Asn-3029 (which could be spatially equivalent to the catalytic chymotrypsin Asp residue, according to sequence analyses by Gorbalenya and Koonin [8] and Herold et al. [12]) with either Gly or Pro. Both mutant proteins, N3029G and N3029P, retained proteolytic

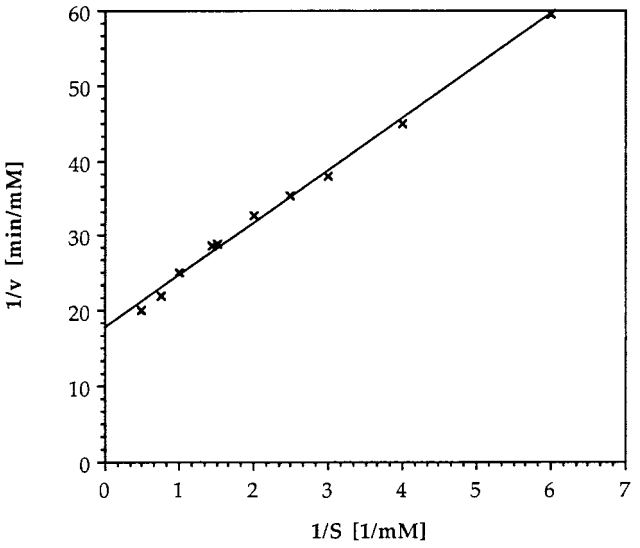

FIG. 3. Lineweaver-Burk plot for the determination of the HCV 3C-like proteinase K_m value. The K_m value was determined by using the peptide substrate NH_2-V-S-Y-G-S-T-L-Q┼A-G-L-R-K-M-A-COOH. The data shown are representative of three experiments.

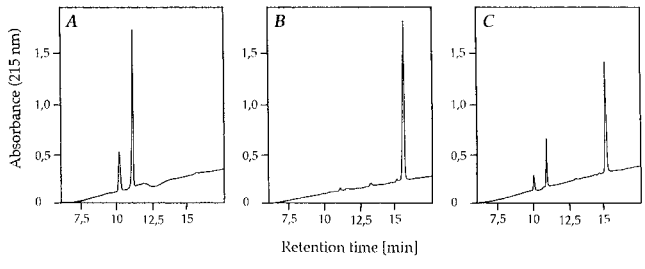

FIG. 4. Peptide-based cleavage assay for the determination of the proteolytic activity of mutant HCV 229E 3C-like proteinases. The extent of 3CL^{pro}-mediated proteolysis of the synthetic peptide NH_2-V-S-Y-G-S-T-L-Q┼A-G-L-R-K-M-A-COOH was analyzed by reversed-phase HPLC as described in Materials and Methods. Proteolytic activities of mutants H3028G (A), H3006G (B), and H3136A (C) are shown.

TABLE 1. Enzymatic activity of mutant, recombinant HCV 229E 3CL^{pro}

Mutant	% Substrate proteolysis[a]
H3006G	0
H3006S	0
H3006T	0
H3028G	100
H3028T	100
N3029G	100
N3029P	5
E3074H	100
T3099D	0
C3109P	0
C3109S	0
C3109V	0
H3127S	0
H3136A	33
H3136S	23
H3136T	7
Q3267A	100
Δ3185-3267	0
Δ3222-3267	0
Δ3234-3267	0

[a] Substrate proteolysis was determined by incubation of the synthetic peptide NH_2-V-S-Y-G-S-T-L-Q┼A-G-L-R-K-M-A-COOH with mutant HCV 229E 3CL^{pro} for 2 h at 25°C as described in Materials and Methods. The activity of wild-type HCV 229E 3CL^{pro} was taken as 100%.

TABLE 2. Effect of proteinase inhibitors on the activity of recombinant HCV 229E 3CL^{pro}

Inhibitor class	Inhibitor	Highest concentration tested (mM)	Inhibition[a]
Cys/Ser proteinase	Antipain	0.2	Yes
	Leupeptin	0.2	No
	Chymostatin	0.1	No
Cys proteinase	E-64	0.1	No
Ser proteinase	Pefabloc SC	2.5	Yes
	Elastatinal	0.2	No
	PMSF	2.0	Yes
	PMSF plus 10 mM dithiothreitol	2.0	No
	3,4-Dichloroisocoumarin	0.2	Yes
	TPCK	0.1	No
	TLCK	0.1	Yes
Asp proteinase	Pepstatin	0.005	No
Metalloproteinase	1,10-Phenanthroline	5.0	No
	EDTA	10.0	No
General inhibitor	Zinc chloride	1.0	Yes

[a] Inhibition is defined as reduction of enzyme activity by >80% (determined under conditions described in Materials and Methods). No inhibition describes a reduction of enzyme activity of ≤15%.

activity albeit, in the case of N3029P, with a significant reduction. Second, we replaced Glu-3074 with His. Glu-3074 is the only acidic amino acid between the catalytic His-3006 and Cys-3109 residues that is strictly conserved among all coronavirus 3C-like proteinases (4, 7, 12, 17). The mutant protein, E3074H, retained full activity. At the present time, we conclude that there is no experimental evidence for a third catalytic residue in the coronavirus 3CL^{pro}.

To further support the conclusions drawn from the study of mutant proteins, we have challenged the purified HCV 229E 3CL^{pro} with a series of typical serine, cysteine, aspartic acid, and metalloproteinase inhibitors and then tested for peptide cleavage activity. As is shown in Table 2, we found six substances that effectively inhibited the 3CL^{pro}-mediated peptide cleavage. These were 3,4-dichloroisocoumarin, phenylmethylsulfonyl fluoride (PMSF), Pefabloc SC, tosyl lysyl chloromethyl ketone (TLCK), antipain, and $ZnCl_2$. Inhibition was defined as a reduction in cleavage product formation to less than 15% of the inhibitor-free control reaction. In contrast, tosyl phenylalanyl chloromethyl ketone (TPCK), leupeptin, elastatinal, chymostatin, E-64, pepstatin, 1,10-phenanthroline, and EDTA had no significant inhibitory effect. Thus, despite having a nucleophilic Cys residue which cannot be functionally replaced by Ser, the HCV 229E 3CL^{pro} enzyme can be effectively inhibited by several Ser proteinase inhibitors. In this respect, it is also important to note that the inhibitory effect of PMSF can be reversed by the thiol compound dithiothreitol, i.e., supporting the presence of a cysteine active site residue. On the basis of the inhibitor study presented here, we conclude that the classification of the HCV 229E 3CL^{pro} as a conventional serine or cysteine proteinase is not possible.

Substrate binding. Most 3C and 3C-like proteinases have a conserved His residue that has been suggested to form a hydrogen bond with the side chain of the P1 substrate glutamine (3, 9, 10). This hypothesis, which is now also supported by modelling of enzyme-substrate complexes based on the structure of the human rhinovirus 3C proteinase (HRV-3C^{pro}) (23), predicts that the HCV 229E 3CL^{pro} residue His-3127 may be essential for substrate binding. In our experiments, we have found that the substitution of Ser for His-3127 completely abolishes the proteolytic activity of HCV 3CL^{pro}. This datum is the first experimental indication that the HCV 229E 3CL^{pro} His-3127 is required for the proteolytic activity of the enzyme, and it adds genetic support to previous predictions that consider this residue to be equivalent to the conserved His residue present in a variety of 3C and 3C-like proteinases (9).

We have also analyzed the neighboring His-3136 residue with the (among coronaviruses) conserved sequence context G/A-X-H, which is identical to the typical G/A-X-H motif present in the substrate-binding pocket of most 3C and 3C-like enzymes (9, 29). In these experiments, we observed a significant but not complete reduction of activity when the His-3136 residue was replaced with Ala, Ser, or Thr. This leads us to conclude that His-3136 is not likely to represent a residue of central importance for the recognition of the substrate.

It has been suggested recently that a Thr residue in the vicinity of the nucleophilic Ser/Cys residue, which is present in many viral chymotrypsin-like enzymes, might be involved in substrate binding (29). In the coronavirus peptide sequence we could not identify a conserved Thr at this position. However, a nearby threonine (Thr-3099 in the HCV 229E 3CL^{pro}) is highly conserved among coronavirus 3C-like proteinases, and therefore, we decided to replace this residue with Asp. This exchange abolished the proteolytic activity of HCV 3CL^{pro} in the described assay. This result confirms the indispensability of Thr-3099 for the proteolytic activity of the 3CL^{pro}, but additional experiments including structural analyses of the enzyme are required to verify a substrate-binding function of this residue.

Substrate structure. There are very few experimental data on even the primary structure of substrates that are recognized and cleaved by coronavirus 3C-like proteinases. Therefore, we studied the proteolytic activity of purified, recombinant HCV 229E 3CL^{pro} using synthetic peptides containing predicted 3CL^{pro} cleavage sites and peptides that resemble these sequences. These data, combined with published results on the proteolytic activity of recombinant HCV 229E 3CL^{pro} with

TABLE 3. Substrate specificity of recombinant HCV 229E 3C-like proteinase

Substrate[a]	Sequence[e]															Virus	pp1ab amino acids	Cleavage[d]
	P8	P7	P6	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′	P6′	P7′			
P	Asp	Ser	Phe	Cys	Lys	Thr	Ile	Gln	Ser	Ala	Leu	Ser	Val	Val	Ser	HCV 229E	947–961	No
B[b]/P	Val	Ser	Tyr	Gly	Ser	Thr	Leu	Gln	Ala	Gly	Leu	Arg	Lys	Met	Ala	HCV 229E	2958–2972	Yes
P	Gln	Met	Phe	Gly	Val	Asn	Leu	Gln	Ser	Gly	Lys	Thr	Thr	Ser	Met	HCV 229E	3260–3274	Yes
B[c]	Thr	Cys	Asp	Arg	Thr	Ala	Ile	Gln	Ser	Phe	Asp	Asn	Ser	Tyr	Leu	HCV 229E	4061–4075	Yes
B[c]	Tyr	Glu	Lys	Ser	Thr	Val	Leu	Gln	Ala	Ala	Gly	Leu	Cys	Val	Val	HCV 229E	4088–5002	Yes
P	Lys	Gln	Arg	Ile	Thr	Thr	Ile	Gln	Gly	Pro	Pro	Gly	Ser	Gly	Lys	HCV 229E	5270–5284	No
P	Glu	Ile	Thr	Met	Thr	Asp	Leu	Gln	Ser	Glu	Ser	Ser	Cys	Gly	Leu	HCV 229E	5585–5599	Yes
P	Leu	Cys	Thr	Thr	Ser	Phe	Leu	Gln	Ser	Gly	Ile	Val	Lys	Met	Val	MHV-JHM	3343–3357	Yes

[a] P, synthetic peptide; B, bacterially synthesized fusion protein.
[b] Fusion protein described by Ziebuhr et al. (33).
[c] Fusion protein as described by Grötzinger et al. (11).
[d] Cleavage was analyzed by incubation of the substrates with recombinant HCV 229E 3CL^{pro}. Proteolysis of the substrates was verified by reversed-phase HPLC (peptides) as described in Materials and Methods or by SDS-PAGE (for bacterially synthesized fusion protein substrates [11, 33]).
[e] Amino acid residues flanking the proteinase cleavage sites are designated according to the scheme introduced by Schechter and Berger (27). The residues are numbered toward the carboxyl terminus as follows: -P3-P2-P1 ┼P1′-P2′-P3′-.

regard to bacterially synthesized protein substrates (11, 33), are shown in Table 3. The following conclusions can be made. The Gln residue in the P1 position is common to all HCV 229E 3CL^{pro} substrates identified to date, and the P2 and P1′ positions are occupied by Leu/Ile and Ser/Ala, respectively.

The data we present above are consistent with a substrate "core" represented by the amino acid sequence Ile/Leu-Gln-Ala/Ser. However, at the present time, we do not wish to exclude the possibility that other amino acids might be tolerated at the positions indicated above. For example, the finding that the substrate analog of trypsin, TLCK, inhibits the HCV 229E 3C-like proteinase activity (Table 2) indicates that a wider range of amino acids, e.g., Lys at position P1, could also be potential substrates of the HCV 3CL^{pro}, as indeed has been predicted for the MHV 3CL^{pro} (17). Also, our data show that a peptide containing a Lys residue at the P4 position (pp1a/1ab 947 to 961) is not a substrate for the HCV 3CL^{pro}, although the core sequence is present. Clearly, cleavage is not exclusively determined by the P2 through P1′ positions. Finally, we also note that an MHV-derived peptide (position pp1a/1ab 3343 to 3357), representing the amino-terminal cleavage site of the MHV 3CL^{pro} (20), is cleaved very efficiently by the HCV 229E 3CL^{pro}. This result indicates that, despite the modest amino acid identity of coronavirus 3C-like proteinases (i.e., <30%), the substrate specificity is, at least to some extent, conserved among these enzymes.

Carboxyl-terminal extension. In order to address the question of whether the extended carboxyl-terminal domain of the HCV 229E 3CL^{pro} is required for proteolytic activity, we generated carboxyl-terminal deletions in the recombinant proteins Δ3185-3267, Δ3222-3267, and Δ3234-3267 (Fig. 1). None of these mutant proteins retained proteolytic activity (Table 1). The minor substitution introduced into the protein Q3267A was, however, tolerated. We conclude that the HCV 229E 3CL^{pro} carboxyl-terminal region has to be present, in an essentially unmodified form, in order to retain enzymatic activity.

DISCUSSION

This paper reports on the biosynthesis and purification of the HCV 229E 3CL^{pro} and the development of a peptide-cleavage assay for the characterization of its enzymatic activity. The high level of bacterial fusion protein synthesis and the solubility of the 3C-like proteinase has greatly facilitated purification of the recombinant protein, and the material we have obtained is suitable for both functional and structural analyses.

In a first series of experiments, we have analyzed the active site and substrate-binding residues of the HCV 229E 3CL^{pro}. To do this, we synthesized mutant 3C-like proteinases and determined their enzymatic activity with a peptide substrate that is normally cleaved very efficiently. Using this approach, we have been able to identify His-3006 and Cys-3109 as residues that are absolutely indispensable for enzymatic activity. These data are consistent with the predictions on putative catalytic residues in coronavirus 3C-like proteinases by Gorbalenya et al. (10) and Herold et al. (12).

Using the same approach, we have also addressed the question of whether a third acidic amino acid residue is required for catalysis by the HCV 229E 3CL^{pro}. The data we have obtained support previous predictions (7, 12, 17) which conclude that a conserved Asp/Glu residue between the catalytic His and Cys residues (which would be equivalent to the active site Asp of chymotrypsin-like enzymes) is not present. Also, the Asn-3029 residue which has been suggested to be involved in catalysis (8) could be replaced by Gly without loss of activity in our assay system. Theoretically, a third active site residue might be important for proteolysis at some, but not all, cleavage sites, or it might influence the enzyme kinetics. Studies to address these possibilities are still in progress.

With respect to the possibility of a third (acidic) catalytic residue, it is also relevant to note that, in our experiments, the substitution of Ser for Cys at position 3109 completely abolished the proteolytic activity of the HCV 229E 3CL^{pro}. In our opinion, this result adds some support to the speculation by Allaire et al. (1) that differences in the nucleophilic properties of Cys, compared to Ser, could make an acidic residue dispensable for catalysis by some 3C and 3C-like proteinases.

A second important result of this study is the experimental data which indicate that His-3127 may represent a key residue for substrate recognition, as is the case for the HRV-3C^{pro} (23). A structural model for the coronavirus 3C-like proteinase has not yet been reported, so, at present, it is impossible to predict additional residues that might form, for example, the shallow hydrophobic substrate-binding pocket that has been modelled for HAV-3C^{pro} and HRV-3C^{pro} (22, 23). When a structural model of a coronavirus 3CL^{pro} becomes available, a more detailed analysis of substrate-binding residues can be undertaken.

Another interesting finding of this study is that the HCV 229E 3CL^{pro} appears to require an authentic carboxyl-terminal region in order to retain enzymatic activity. All carboxyl-ter-

minal deletions yielded inactive proteinases. In addition, we have observed that even minor changes at the carboxyl terminus, e.g., the addition of six His residues or seven vector-derived amino acids (34), strongly inhibit the enzymatic activity of the coronavirus proteinase. The only exception we have found is the replacement of the carboxyl-terminal Gln residue with Ala (Q3267A), which retained full activity. We believe that the carboxyl-terminal region of coronavirus 3C-like proteinases, which has no counterpart in other 3C-like enzymes, is necessary for proteolytic activity. Several functions could be related to this domain including (i) maintenance of the overall folding of the enzyme, (ii) catalysis or substrate recognition (by virtue of the presence of the additional residues), and (iii) enzymatic activity (e.g., the domain may represent a required additional cofactor). Further experiments and, possibly, structural analyses will be necessary to address these alternatives.

In relation to our studies on the structure of substrates that are recognized and cleaved by the HCV 229E $3CL^{pro}$, the picture that emerges is still incomplete. There is, obviously, a core sequence (Ile/Leu-Gln-Ala/Ser) that is present in all substrates of the HCV 229E $3CL^{pro}$ that have been identified so far. However, the question of whether, and to what extent, peptides with substitutions at these positions are cleaved has not yet been addressed. Also, the inability of the HCV $3CL^{pro}$ to cleave one peptide (pp1a/1ab amino acids 947 to 961), which contains the core sequence Ile-Gln-Ser, demonstrates that substrate specificity is also defined by the adjacent amino acids. Additional substitutions in the enzyme and the peptide substrates and, ideally, the use of combinatorial peptide libraries or phage display systems should help to define the parameters of substrate recognition and cleavage.

In summary, the production of milligram quantities of recombinant HCV 229E $3CL^{pro}$ and the introduction of a peptide-based cleavage assay has allowed for an extensive biochemical study of this enzyme. We hope the recombinant protein can also be used for the determination of its three-dimensional structure by X-ray crystallography. The HCV 3C-like proteinase is a key enzyme in the polyprotein processing required to produce a functional viral RNA polymerase complex, and potentially, it is an ideal target for the design of drugs that will disrupt the coronavirus replication cycle.

ACKNOWLEDGMENTS

We thank Alexander Gorbalenya for helpful discussions and critical reading of the manuscript.

This work was supported by grants from the Deutsche Forschungsgemeinschaft (SFB 165/B1, GK Infektiologie) and the BMBF (01 KI 8838/0).

REFERENCES

1. **Allaire, M., M. M. Chernaia, B. A. Malcolm, and M. N. G. James.** 1994. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. Nature **369:**72–76.
2. **Baric, R. S., K. Fu, M. C. Schaad, and S. A. Stohlman.** 1990. Establishing a genetic recombination map for murine coronavirus strain A59 complementation groups. Virology **177:**646–656.
3. **Bazan, J. F., and R. J. Fletterick.** 1988. Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. Proc. Natl. Acad. Sci. USA **85:**7872–7876.
4. **Boursnell, M. E. G., T. D. K. Brown, I. J. Foulds, P. F. Green, F. M. Tomley, and M. M. Binns.** 1987. Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. J. Gen. Virol. **68:**57–77.
5. **Dougherty, W. G., and B. L. Semler.** 1993. Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. Microbiol. Rev. **57:**781–822.
6. **Drenth, J., K. Kalk, and H. M. Swen.** 1976. Binding of chloromethyl ketone substrate analogues to crystalline papain. Biochemistry **15:**3731–3738.
7. **Eleouet, J.-F., D. Rasschaert, P. Lambert, L. Levy, P. Vende, and H. Laude.** 1995. Complete sequence (20 kilobases) of the polyprotein-encoding gene 1 of transmissible gastroenteritis virus. Virology **206:**817–822.
8. **Gorbalenya, A. E., and E. V. Koonin.** 1993. Comparative analysis of amino acid sequences of key enzymes of replication and expression of positive-strand RNA viruses: validity of approach and functional and evolutionary implications. Sov. Sci. Rev. Sect. D. **11:**1–84.
9. **Gorbalenya, A. E., A. P. Donchenko, V. M. Blinov, and E. V. Koonin.** 1989. Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. FEBS Lett. **243:**103–114.
10. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov.** 1989. Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. Nucleic Acids Res. **17:**4847–4861.
11. **Grötzinger, C., G. Heusipp, J. Ziebuhr, U. Harms, J. Süss, and S. G. Siddell.** 1996. Characterization of a 105-kDa polypeptide encoded in gene 1 of the human coronavirus HCV 229E. Virology **222:**227–235.
12. **Herold, J., T. Raabe, B. Schelle-Prinz, and S. G. Siddell.** 1993. Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. Virology **195:**680–691.
13. **Herold, J., and S. G. Siddell.** 1993. An 'elaborated' pseudoknot is required for high frequency frameshifting during translation of HCV 229E polymerase mRNA. Nucleic Acids Res. **21:**5838–5842.
14. **Herold, J., S. G. Siddell, and J. Ziebuhr.** 1996. Characterization of coronavirus RNA polymerase gene products. Methods Enzymol. **275:**68–89.
15. **Johnston, S., and S. Holgate.** 1996. Epidemiology of viral respiratory tract infections, p. 1–38. *In* S. Myint and D. Taylor-Robinson (ed.), Viral and other infections of the human respiratory tract. Chapman and Hall, London, United Kingdom.
16. **Kräusslich, H.-G., and E. Wimmer.** 1988. Viral proteinases. Annu. Rev. Biochem. **57:**701–754.
17. **Lee, H. J., C.-K. Shieh, A. E. Gorbalenya, E. V. Koonin, N. la Monica, J. Tuler, A. Bagdzhadzhyan, and M. M. C. Lai.** 1991. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. Virology **180:**567–582.
18. **Liu, D. X., and T. D. K. Brown.** 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.
19. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193:**265–275.
20. **Lu, Y., X. Lu, and M. R. Denison.** 1995. Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. J. Virol. **69:**3554–3559.
21. **Lu, X., Y. Lu, and M. R. Denison.** 1996. Intracellular and in vitro-translated 27-kDa proteins contain the 3C-like proteinase activity of the coronavirus MHV-A59. Virology **222:**375–382.
22. **Malcolm, B. A.** 1995. The picornaviral 3C proteinases: cysteine nucleophiles in serine proteinase folds. Protein Sci. **4:**1439–1445.
23. **Matthews, D. A., W. W. Smith, R. A. Ferre, B. Condon, G. Budahazi, W. Sisson, J. E. Villafranca, C. A. Janson, H. E. McElroy, C. L. Gribskov, and S. Worland.** 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.
24. **Merrifield, R. B.** 1965. Automated synthesis of peptides. Science **150:**178–185.
25. **Myint, S. H.** 1995. Human coronavirus infections, p. 389–401. *In* S. G. Siddell (ed.) The Coronaviridae. Plenum Press, New York, N.Y.
26. **Palmenberg, A. C.** 1990. Proteolytic processing of picornaviral polyprotein. Annu. Rev. Microbiol. **44:**603–623.
27. **Schechter, I., and A. Berger.** 1967. On the size of the active site in proteases. Biochem. Biophys. Res. Commun. **27:**157–162.
28. **Seybert, A., J. Ziebuhr, and S. G. Siddell.** 1997. Expression and characterization of a recombinant murine coronavirus 3C-like proteinase. J. Gen. Virol. **78:**71–75.
29. **Snijder, E. J., A. L. M. Wassenaar, L. C. van Dinten, W. J. M. Spaan, and A. E. Gorbalenya.** 1996. The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases. J. Biol. Chem. **271:**4864–4871.
30. **Tibbles, K. W., I. Brierley, D. Cavanagh, and T. D. K. Brown.** 1996. Characterization in vitro of an autocatalytic processing activity associated with the predicted 3C-like proteinase domain of the coronavirus avian infectious bronchitis virus. J. Virol. **70:**1923–1930.
31. **Yao, Z., D. H. Jones, and C. Grose.** 1992. Site-directed mutagenesis of herpesvirus glycoprotein phosphorylation sites by recombination polymerase chain reaction. PCR Methods Appl. **1:**205–207.
32. **Zhang, X. M., W. Herbst, K. G. Kousoulas, and J. Storz.** 1994. Biological and genetic characterization of hemagglutinating coronavirus isolated from a diarrhoeic child. J. Med. Virol. **44:**152–161.
33. **Ziebuhr, J., J. Herold, and S. G. Siddell.** 1995. Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. J. Virol. **69:**4331–4338.
34. **Ziebuhr, J.** Unpublished data.