JOURNAL OF VIROLOGY, July 1995, p. 4331–4338 Vol. 69, No. 7
0022-538X/95/$04.00+0

Characterization of a Human Coronavirus (Strain 229E) 3C-Like Proteinase Activity

JOHN ZIEBUHR,* JENS HEROLD, AND STUART G. SIDDELL

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

Received 5 December 1994/Accepted 12 April 1995

The RNA polymerase gene of human coronavirus (HCV) 229E encodes a large polyprotein that contains domains with motifs characteristic of both papain-like cysteine proteinases and proteinases with homology to the 3C proteinase of picornaviruses. In this study, we have, first, expressed the putative HCV 229E 3C-like proteinase domain as part of a β-galactosidase fusion protein in *Escherichia coli* and have shown that the expressed protein has proteolytic activity. The substitution of one amino acid within the predicted proteinase domain (His-3006→Asp-3006) abolishes, or at least significantly reduces, this activity. Amino-terminal sequence analysis of a purified, 34-kDa cleavage product shows that the bacterial fusion protein is cleaved at the dipeptide Gln-2965–Ala-2966, which is the predicted amino-terminal end of the putative 3C-like proteinase domain. Second, we have confirmed the proteolytic activity of a bacterially expressed polypeptide with the amino acid sequence of the predicted HCV 229E 3C-like proteinase by *trans* cleavage of an in vitro translated polypeptide encoded within open reading frame 1b of the RNA polymerase gene. Finally, using fusion protein-specific antiserum, we have identified a 34-kDa, 3C-like proteinase polypeptide in HCV 229E-infected MRC-5 cells. This polypeptide can be detected as early as 3 to 5 h postinfection but is present in the infected cell in very low amounts. These data contribute to the characterization of the 3C-like proteinase activity of HCV 229E.

Human coronaviruses (HCV) are important medical pathogens. They are known to cause respiratory infections in children and adults (26), and their involvement in human enteric disease seems likely (36). It has been established that there are at least two major antigenic groups of HCV, which are represented by the prototypes HCV 229E and HCV OC 43. The two groups are about equally responsible for respiratory infections (19). However, HCV 229E has been more extensively studied because it is easier to isolate and propagate in tissue culture.

The HCV 229E genome is a positive-strand RNA of 27,277 nucleotides. The genome contains at least eight functional open reading frames (ORFs). These ORFs encode the viral RNA-dependent RNA polymerase; the structural proteins S (spike), sM (small membrane), M (membrane), and N (nucleocapsid); and two small, putative, nonstructural proteins, ns4a and ns4b (17). In the infected cell, the viral gene products are expressed from genomic and subgenomic mRNAs (30). In common with other coronaviruses, the subgenomic mRNAs form a 3′ coterminal set and are synthesized by a process that involves discontinuous transcription. Only the 5′ unique region of each mRNA, i.e., the region that is absent from the next smallest mRNA, is thought to be translationally active (for a recent review, see reference 21).

The HCV 229E genomic RNA, which is also known as mRNA 1, encodes the viral RNA polymerase (16). The unique region of the genomic RNA is composed of two large ORFs, ORF 1a and ORF 1b, that overlap by 43 nucleotides. The upstream ORF 1a potentially encodes a polypeptide with a molecular mass of 454 kDa, and the downstream ORF 1b potentially encodes a 300-kDa polypeptide. However, in vitro studies suggest that the downstream ORF 1b is expressed as a fusion protein with the ORF 1a gene product by a mechanism involving (−1) ribosomal frame shifting. This ribosomal frameshifting event is mediated by a slippery sequence and a pseudoknot structure located in the region of the genome encompassing the overlap of ORFs 1a and 1b (18).

Complementation analysis of mouse hepatitis virus (MHV) temperature-sensitive mutants with an RNA minus phenotype has shown that there are at least five distinct viral functions related to RNA synthesis (23, 32). Analysis of these mutants by genetic recombination and sequencing allows the different functions to be located and ordered within the RNA polymerase locus (4, 13). The complementation frequencies of these mutants are indicative of intergenic rather than intragenic complementation (23); therefore, they provide strong evidence for the activity of proteinases that are required to process the polymerase polyprotein(s) into smaller, functional products. Computer-assisted analyses of the MHV, infectious bronchitis virus (IBV), and HCV 229E ORF 1a regions (14, 16, 22) have identified motifs characteristic of both papain-like cysteine proteinases and proteinases with homology to the 3C proteinase of picornaviruses.

The first direct biochemical evidence for co- or posttranslational processing of the coronavirus polymerase polyprotein(s) came from the in vitro translation of MHV genomic RNA (7). In these experiments, virus-specified polypeptides with molecular masses of 28 and 220 kDa were detected. These data were subsequently confirmed by in vitro transcription and translation of mRNA derived from a plasmid cDNA representing the 5′ portion of MHV genomic RNA, which yielded 28- and 160-kDa products (2, 33). The p28 polypeptide, which represents the amino-terminal portion of the polyprotein, and additional cleavage products have also been detected in MHV-infected cells (8, 9).

The location of the putative MHV papain-like proteinase domain, which is responsible for the cleavage of the p28 polypeptide, has been mapped to a region 3.6 to 4.4 kb from

* Corresponding author. Mailing address: Institute of Virology, University of Würzburg, Versbacher Strasse 7, 97078 Würzburg, Germany. Phone: 49-931-201-5954/3896. Fax: 49-931-201-3934. Electronic mail address: siddell@vax.rz.uni-wuerzburg.d400.de.

the 5′ end of the genomic RNA. Site-specific mutation of putative catalytic residues in the proteinase domain identified two amino acids, Cys-1137 and His-1288, which are essential for this activity (3). Recently, the p28 cleavage site has been positioned at the dipeptide Gly-247–Val-248 in the MHV ORF 1a gene product, and the structural role of the amino acids flanking the cleavage site has been investigated by mutational analysis (10).

Moreover, Liu et al. (24) have identified a 100-kDa polypeptide in IBV-infected cells using a polyclonal serum raised against a bacterially expressed fusion protein. This fusion protein contained sequences encoded by the IBV ORF 1b nucleotides 14,492 to 15,520; a region that partially encompasses the IBV polymerase domain originally identified by Gorbalenya et al. (14). Experiments using the vaccinia virus-T7 expression system suggest that the IBV 3C-like proteinase domain, which is located at the 3′ end of ORF 1a, is involved in the processing of the IBV ORF 1b gene product to produce the 100-kDa polypeptide.

In comparison with data about the cleavage products of the coronavirus proteinases, there is very little information about the proteinases themselves. In this paper, we report the detection of a 34-kDa, 3C-like proteinase polypeptide in HCV 229E-infected MRC-5 cells. To do this, we used antiserum raised against a bacterially expressed fusion protein containing sequences encoded by the HCV 229E ORF 1a. In addition, a cDNA encoding the putative 3C-like proteinase domain was cloned into a bacterial vector, and a fusion protein with proteolytic activity was expressed. The fusion protein cleavage products were characterized, and the N-terminal cleavage site of the 3C-like proteinase domain was determined by sequence analysis. Finally, a polypeptide with the amino acid sequence of the predicted 3C-like proteinase was expressed and partially purified from bacteria. This polypeptide was able to cleave an in vitro-synthesized, HCV 229E ORF 1b-specific substrate in *trans.*

MATERIALS AND METHODS

Cells and virus. The HCV 229E isolate used in these studies has been described previously (30). The virus was propagated in monolayers of MRC-5 cells (ECACC. 84101801) grown in minimal essential medium with Earle's salts, 25 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES), Glutamax-1 (42360-024; Life Technologies, Eggenstein, Germany), and 10% fetal bovine serum (MEM10) at 33°C. Virus stocks were concentrated by NaCl-polyethylene glycol precipitation as described by Wege et al. (34).

Preparation of antigens. The DNA of plasmid pBS-J28C3 (16) was digested with *Bam*HI and *Hin*dIII. The 1,122-nucleotide restriction fragment was isolated, ligated to *Bam*HI-*Hin*dIII-digested pQE10 DNA (33103; QIAGEN GmbH, Hilden, Germany), and used to transform competent *Escherichia coli* TG1 cells. DNA from the resultant plasmid, pQE-3CL, was digested with *Nsi*I, and the larger restriction fragment was isolated, religated, and used to transform TG1 cells. The sequence of the resultant plasmid, pQE-3CL-*Nsi*I, was determined. The plasmid encodes a fusion protein of 309 amino acids; 14 amino acids at the amino terminus that include 6 consecutive histidine residues, 292 amino acids corresponding to amino acids 2908 to 3046 and 3127 to 3279 in the HCV 229E ORF 1a, and 3 artifactual amino acids at the carboxyl terminus.

Bacterial expression and purification of the recombinant protein were done according to the manufacturer's instructions. Briefly, pQE-3CL-*Nsi*I DNA was used to transform *E. coli* M15(pREP4) cells. Expression of the fusion protein was induced with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) during exponential growth. After 3 h of induction, the bacteria were harvested and stirred for 1 h in lysis buffer (6 M guanidine hydrochloride, 0.1 M NaH_2PO_4, 0.01 M Tris-Cl [pH 8.0]). After centrifugation (10,000 × *g* at 4°C), the supernatant was loaded onto a Ni-nitrilotriacetic acid-agarose column (30230; QIAGEN), and nonspecifically bound proteins were eluted in a gradient of 8 M urea–0.1 M NaH_2PO_4–0.01 M Tris-Cl (pH 8.0 to 6.0). Finally, the recombinant protein was eluted in the same buffer at pH 4.5 and dialyzed for 16 h against 10 mM Tris-Cl (pH 7.5)–150 mM NaCl at 4°C.

In order to produce control reagents, two additional recombinant fusion proteins were expressed and purified. First, the plasmid pQE-HCVN was constructed by subcloning HCV 229E N protein-coding sequences, which were derived from the plasmid pSM-F1 (28), into the *Bam*HI site of pQE9 (33093; QIAGEN). Expression and purification of the recombinant N fusion protein were done as described above. Second, a truncated β-galactosidase protein, corresponding to amino acids 7 to 376, was expressed by using the plasmid pROS (12). In this case, expression was induced in BMH 71-18 cells with 1 mM IPTG during exponential growth. After 3 h of induction, the bacteria were pelleted and lysed for 30 min at 4°C in 50 mM NaH_2PO_4 (pH 7.8)–300 mM NaCl containing lysozyme (1 mg/ml). The lysates were sonicated and separated on preparative sodium dodecyl sulfate (SDS)-polyacrylamide gels, and the β-galactosidase fragment was electroeluted and dialyzed as described previously (31).

Production of antibodies. Fusion protein-specific antisera were obtained after subcutaneous immunization of New Zealand White rabbits. For primary immunizations, 200 μg (200 μl) of antigen was emulsified with an equal volume of Freund's complete adjuvant (F4258; Sigma, Deisenhofen, Germany). The rabbits were given five booster injections at monthly intervals with 100 μg of antigen mixed with an equal volume of Freund's incomplete adjuvant (F5506; Sigma). The pQE-3CL-*Nsi*I fusion protein-specific antiserum was designated K17. The pQE-HCVN fusion protein-specific antiserum was designated K16.

For the production of a β-galactosidase-specific monoclonal antibody (MAb), 3-month-old female BALB/c mice were inoculated five times intraperitoneally with 50 to 100 μg (50 to 100 μl) of antigen mixed with an equal volume of Freund's incomplete adjuvant. Polyethylene glycol fusion of spleen cells with NS1 plasmocytoma cells, screening of tissue culture supernatants by enzyme-linked immunosorbent assaying, and cloning of the hybridomas by limiting dilution were performed by using standard techniques (15). The β-galactosidase-specific MAb was designated 6D2H4.

Metabolic labelling, cell lysis, and immunoprecipitation. MRC-5 cells (7×10^5) were mock infected with MEM10 or were infected at a multiplicity of infection of 10 PFU per cell with HCV 229E in MEM10. After 1 h of virus adsorption at 33°C, the medium was removed and replaced by 3 ml of fresh medium. Radioactive labelling of intracellular proteins was carried out at 1, 3, 5, 7, and 9 h postinfection (p.i.). Prior to labelling, cells were washed twice with methionine- and cysteine-free, supplemented Dulbecco's modified Eagle's medium (11963; Life Technologies) with 2% dialyzed fetal bovine serum. Proteins were labelled with 100 μCi of [^{35}S]methionine and 100 μCi of [^{35}S]cysteine (SJ204 and SJ232, respectively; Amersham, Braunschweig, Germany) per ml of methionine- and cysteine-free Dulbecco's modified Eagle's medium for 2 h at 33°C.

Radiolabelled cells (7×10^5) were washed twice in ice-cold phosphate-buffered saline and lysed in 330 μl of lysis buffer (20 mM Tris-Cl [pH 7.5], 150 mM NaCl, 0.5% sodium deoxycholate, 1% Nonidet P-40 [NP-40], 0.1% SDS) containing proteinase inhibitors (1 mM phenylmethylsulfonyl fluoride, 1 mM benzamidine, 2 μg of leupeptin per ml, 100 μg of aprotinin per ml) for 15 min on ice. After centrifugation (14,000 × *g*, 4°C), EDTA was added to the supernatant to a final concentration of 5 mM. Immunoprecipitation buffer (20 mM Tris-Cl [pH 7.5], 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 0.1% sodium deoxycholate, 0.2% SDS, proteinase inhibitors as described above) was added to between 4 to 100 μl of cell lysate to give a final volume of 300 μl, and the lysates were precleared by the addition of 10 μl of rabbit normal serum for 3 h at 4°C. Subsequently, 25 μl of washed Pansorbin (507861; Calbiochem-Novabiochem, Bad Soden, Germany) was added to the lysates, and the mixture was incubated at 4°C for 1 h. The *Staphylococcus aureus* cells were pelleted, and the supernatant was transferred to a fresh tube. Then, 5 μl of the appropriate preimmune serum, 5 μl of K16 antiserum, or 5 μl of K17 antiserum was added, and the mixture was incubated overnight at 4°C. Following this incubation, 25 μl of washed Pansorbin was added to each reaction mixture for 1 h at 4°C. Finally, the reaction mixtures were centrifuged, the supernatant was discarded, and the pellet fractions were resuspended twice in 500 μl of wash buffer A (20 mM Tris-Cl [pH 7.5], 150 mM NaCl, 5 mM EDTA, 0.1% NP-40) and once in 500 μl of wash buffer B (20 mM Tris-Cl [pH 7.5], 0.1% NP-40). After rinsing, the pellets were resuspended in 25 μl of sample buffer (20), heated at 100°C for 5 min, and centrifuged, and the supernatant was analyzed by polyacrylamide gel electrophoresis (PAGE) on SDS-12.5% polyacrylamide gels.

Mutagenesis. Mutagenesis of the predicted catalytic histidine residue of the 3C-like proteinase domain of HCV 229E (16) was done by a recombination-PCR method (35). Four partially complementary oligonucleotides (OL69 [5′ TCT TCA GCA TCT TTT ACT TTC 3′], OL70 [5′ GAA AGT AAA AGA TGC TGA AGA 3′], OL75 [5′ TTG CCC ACG TGA TGT TAT CGC ATC TAA CAC AA 3′], and OL76 [5′ GCG ATA ACA TCA CGT GGG CAA TAA ACA ATA TC 3′]) were used to generate linear fragments from pBS-J28C3 DNA by PCR. These fragments were combined without further purification and were used to transform competent *E. coli* TG1 cells. Recombination in vivo effected the generation of circular plasmids carrying the desired mutation. The introduced mutation, exchanging C for G at nucleotide position 9308, was verified by sequence analysis and substituted Asp for His at amino acid position 3006 in the HCV 229E ORF 1a gene product.

Construction of bacterial expression plasmids pROS-3CL and pROS-3CL/H3006D. The DNA of plasmid pBS-J28C3 was digested with *Bam*HI and *Hin*dIII. The 1.122-nucleotide restriction fragment was isolated, treated with T4 DNA polymerase, ligated to *Eco*RV-digested pROS, and used to transform competent *E. coli* TG1 cells. The sequence of the resultant plasmid, pROS-3CL, was determined. The plasmid encodes a fusion protein of 780 amino acids: 6 artifactual amino acids at the amino terminus, 370 amino acids corresponding to

amino acids 7 to 376 of β-galactosidase, 25 internal artifactual amino acids, 372 amino acids corresponding to amino acids 2908 to 3279 in the HCV 229E ORF 1a, and 7 artifactual amino acids at the carboxyl terminus. In the same way, the *Bam*HI-*Hin*dIII restriction fragment of pBS-J28C3, carrying the H3006D mutation, was inserted into the *Eco*RV restriction site of pROS. The resultant plasmid was designated pROS-3CL/H3006D.

Expression of bacterial fusion proteins and Western blotting (immunoblotting). Plasmids pROS-3CL and pROS-3CL/H3006D were used to express the wild-type or mutated HCV 229E 3C-like proteinase domain as part of a β-galactosidase fusion protein in *E. coli* BMH 71-18. During exponential growth, expression was induced with 1 mM IPTG for 2 h. Thereafter, 1.5 ml of the bacterial culture was pelleted, suspended in 200 μl of sample buffer, and heated at 100°C for 7 min, and aliquots (0.1 to 10 μl) were analyzed by electrophoresis in SDS-11.5% polyacrylamide gels. The separated proteins were transferred electrophoretically onto a nitrocellulose membrane and immunostained with the MAb 6D2H4 or with antiserum K17 by standard techniques (15).

Protein purification and N-terminal protein sequence analysis. Standard procedures were used to purify inclusion bodies from pROS-3CL bacteria that had been induced for 3 h with IPTG (15). After extraction with 6 M urea in 100 mM NaH_2PO_4–10 mM Tris-Cl (pH 8.0), the insoluble fraction of the inclusion bodies was suspended in sample buffer, heated at 100°C for 5 min, and separated by electrophoresis in SDS-15% polyacrylamide gels. After electrophoresis, proteins were transferred electrophoretically onto a Glassybond GB 10 membrane (093-000; Biometra, Göttingen, Germany) according to the manufacturer's instructions. The membrane was stained with 40% (vol/vol) methanol and 10% (vol/vol) acetic acid containing 0.1% Coomassie brilliant blue R-250 and was destained with 30% (vol/vol) methanol–10% (vol/vol) acetic acid. The appropriate area of membrane was isolated, and the N-terminal sequence of the bound protein was determined by standard procedures with a pulsed-liquid protein sequencer (model 476A; Applied Biosystems, Weiterstadt, Germany).

Expression of an HCV 229E 3C-like proteinase and *trans* cleavage assaying. The coding sequence of the putative HCV 229E 3C-like proteinase domain (16) was amplified from plasmid pQE-3CL DNA by PCR. The PCR product was treated with T4 DNA polymerase, phosphorylated with polynucleotide kinase, and ligated with *Xmn*I-digested pMal-c2 DNA (800-64S; New England Biolabs, Schwalbach, Germany). The construct, pMalc2-3CL, which encodes a fusion protein containing the maltose-binding protein (MBP) of *E. coli* and the predicted HCV 229E 3C-like proteinase domain, was used to transform competent *E. coli* TB1 cells. The recombinant protein was expressed and purified according to the manufacturer's instructions, and, by using factor Xa, an authentic 3C-like proteinase containing amino acids 2966 to 3267 of the ORF 1a gene product was produced.

To demonstrate the *trans* cleavage activity of the expressed 3C-like proteinase, a cDNA fragment containing nucleotides 14,599 to 16,069 of the HCV 229E RNA polymerase gene was amplified by PCR. The upper primer contained the bacteriophage T7 promoter, which allowed subsequent in vitro transcription of the purified PCR product by using T7 RNA polymerase (6, 27). The in vitro-synthesized, capped RNA was translated in a rabbit reticulocyte lysate (Promega/Serva, Heidelberg, Germany) in the presence of [^{35}S]methionine. Subsequently, 1 μl of the reaction mixture was incubated for 60 min with 10 μl of buffer CM (20 mM Tris-Cl [pH 7.4], 200 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, 10 mM maltose, 5 μg of factor Xa per ml) containing 3C-like proteinase. As a negative control, 10 μl of buffer CM was used. Following this incubation, the translation and cleavage products were analyzed in SDS-15% polyacrylamide gels.

RESULTS

Production of antiserum K17. Our initial attempts to express the predicted 3C-like proteinase coding region of HCV 229E as a bacterial fusion protein containing a minimal number of additional amino acids failed. Therefore, we constructed the bacterial expression vector pQE-3CL-*Nsi*I. This construct contains an HCV 229E ORF 1a sequence that encodes 58 amino acids upstream of the predicted amino terminus of the 3C-like proteinase domain (16), discontinuous amino- and carboxyl-terminal regions of the predicted 3C-like proteinase domain, and a minimal number of artifactual amino acids at the termini of the fusion protein. As is shown in Fig. 1, the introduction of this deletion, which eliminated amino acids 3047 to 3126 from the wild-type ORF 1a protein, allowed a significant level of fusion protein expression (lanes 1 and 2). Consequently, it was possible to purify the recombinant protein to near homogeneity by using Ni-nitrilotriacetic acid-agarose affinity chromatography (Fig. 1, lane 3) and to generate the fusion protein-specific antiserum K17.

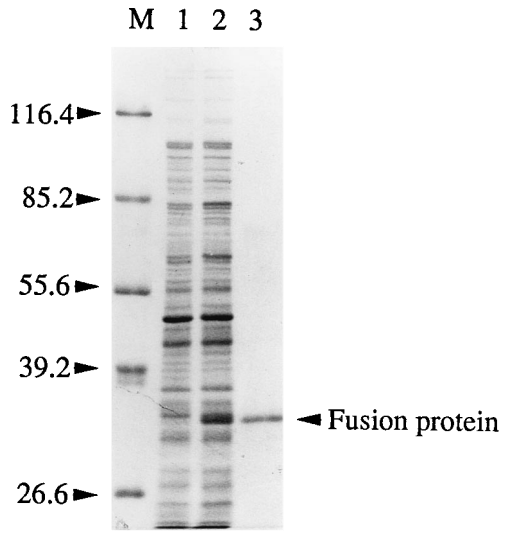

FIG. 1. Expression and purification of a bacterial fusion protein containing discontinuous regions of the HCV 229E 3C-like proteinase domain. Cell lysates from noninduced or IPTG-induced *E. coli* M15(pREP4) cells transformed with the plasmid pQE-3CL-*Nsi*I (lanes 1 and 2, respectively), as well as the purified fusion protein (lane 3), were separated by SDS-PAGE in a 12.5% polyacrylamide gel and stained with Coomassie brilliant blue R-250. Protein molecular weight markers (Combithek 1317 474; Boehringer Mannheim, Mannheim, Germany) are shown in lane M.

Expression of an active HCV 229E 3C-like proteinase in *E. coli*. To express a fusion protein that contained the complete 3C-like proteinase domain of HCV 229E, we constructed the bacterial expression vector pROS-3CL. This construct encodes a fusion protein that is essentially composed of the amino-terminal third of β-galactosidase, fused to the HCV 229E ORF 1a amino acids 2908 to 3279. This region of the HCV 229E ORF 1a gene product includes the entire putative 3C-like proteinase domain flanked by 58 amino-proximal and 12 carboxyl-proximal, virus-encoded amino acids.

Following induction, bacteria with the construct pROS-3CL expressed a fusion protein with an apparent molecular mass of 102 kDa. This primary translation product was detected in Western blots using either the K17 fusion protein-specific serum or the β-galactosidase-specific MAb 6D2H4 (Fig. 2, lanes 3 and 6). However, in the same lysates, 34- and 66-kDa polypeptides were also detected by immunoblotting. Both the 34- and the 66-kDa polypeptides were detected by the K17 serum (Fig. 2, lane 3), while MAb 6D2H4 detected only the 66-kDa polypeptide (Fig. 2, lane 6).

These results can be interpreted as follows. First, the expressed fusion protein has a proteolytic activity that cleaves the primary translation product at the predicted amino end of the putative 3C-like protease domain (Gln-Ala [amino acid positions 2965 to 2966]). Thus, the fusion protein-specific antiserum K17, which has components specific for the 3C-like proteinase domain, as well as components specific for the amino-proximal flanking sequences, will detect the primary translation product (predicted molecular mass, 87 kDa; observed molecular mass, 102 kDa), as well as the amino-terminal (predicted molecular mass, 52 kDa; observed molecular mass, 66 kDa) and carboxyl-terminal (predicted molecular mass, 33 or 35 kDa; observed molecular mass, 34 kDa) cleavage products. The β-galactosidase-specific MAb 6D2H4 detects only the primary translation product and the amino-terminal cleavage product. From the data shown in Fig. 2, it is not possible to deduce whether the predicted carboxyl-terminal cleavage site

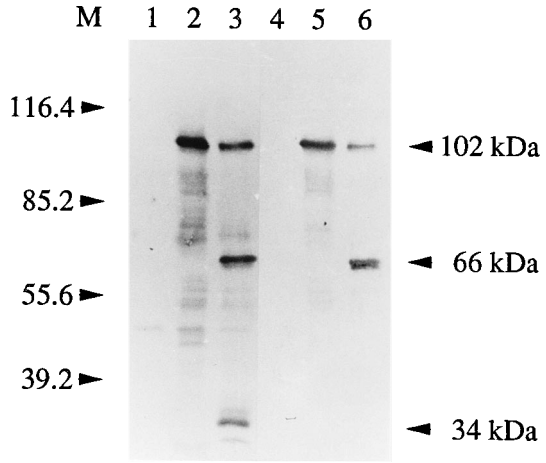

FIG. 2. Expression of the HCV 229E 3C-like proteinase domain as part of a β-galactosidase fusion protein. Cell lysates from BMH 71-18 cells transformed with pROS-3CL/H3006D (lanes 1, 2, 4, and 5) or pROS-3CL (lanes 3 and 6) were separated by SDS-PAGE in an 11.5% polyacrylamide gel, transferred to a nitrocellulose membrane, and immunostained with the fusion protein-specific antiserum K17 (lanes 1, 2, and 3) or the β-galactosidase-specific MAb 6D2H4 (lanes 4, 5, and 6). The bacteria were mock induced (lanes 1 and 4) or induced with IPTG (lanes 2, 3, 5, and 6). The positions of protein molecular weight markers are indicated.

of the 3C-like protease domain (Gln-Ser [amino acid positions 3267 to 3268]) is used. The predicted cleavage products would differ in length by only 19 amino acids.

To provide further evidence that the proteolytic activity that we observed is mediated by the 3C-like proteinase domain rather than, for example, by a bacterial proteinase, a single amino acid substitution was introduced into the HCV 229E ORF 1a sequence. This substitution exchanged the putative catalytic histidine residue of the proteinase domain (16) for aspartic acid. Thus, following induction, bacteria with pROS-3CL/H3006D expressed a fusion protein with an apparent molecular mass of 102 kDa. However, in this case, no specific cleavage products could be detected by either the K17 serum or MAb 6D2H4 (Fig. 2, lanes 2 and 5).

Partial purification of the 34-kDa cleavage product and N-terminal sequence analysis. The interpretation presented above supposes that the amino-terminal cleavage site of the HCV 229E 3C-like proteinase domain is the Gln-Ala dipeptide at positions 2965 to 2966 in the ORF 1a gene product. However, because of the inaccuracy of molecular mass estimations in SDS-polyacrylamide gels, we decided to confirm the position of the cleavage site by partial purification of the 34-kDa cleavage product and subsequent N-terminal sequencing.

The 34-kDa cleavage product was partially purified as described in Materials and Methods. Essentially, the purified material consists of an inclusion body fraction resistant to solubilization in 6 M urea. As can be seen in Fig. 3 (lane 1), this material consists mainly of the fusion protein cleavage products with molecular masses of 66 and 34 kDa. However, there are many contaminating bacterial proteins, one of which migrates only slightly faster than the 34-kDa protein. Nevertheless, the 34-kDa fusion protein cleavage product could be unambiguously identified by immunodetection using the K17 antiserum (Fig. 3, lane 3). After electrophoretic transfer to a polyvinylidene difluoride membrane, approximately 10 pmol of the protein was isolated for N-terminal sequence analysis.

The N-terminal sequence of the 34-kDa cleavage product was shown to be Ala-Gly-Leu-Arg-Lys-Met-Ala (Fig. 4, panels b to h). This sequence is present at amino acid positions 2966 to 2972 in the HCV 229E ORF 1a gene product, indicating that cleavage takes place at the predicted amino terminus of the 3C-like proteinase domain, between Gln-2965 and Ala-2966.

***trans* cleavage activity of the HCV 229E 3C-like proteinase.** To provide direct evidence for a proteolytic activity of the HCV 229E 3C-like proteinase in *trans*, the proteinase domain was expressed as part of an *E. coli* MBP fusion protein, and, after affinity chromatographic purification, a polypeptide with the amino acid sequence of the 3C-like proteinase was released by factor Xa treatment. As a substrate for the proteinase, a polypeptide was translated in vitro from a synthetic mRNA. This polypeptide represents amino acids 4818 to 5259 of the HCV 229E polymerase polyprotein and contains a predicted cleavage site (dipetide Gln-4995–Ala-4996 [16, 22, 24]) within the coronavirus ORF 1ab gene product. After 60 min of incubation with the 3C-like proteinase, the primary translation product (predicted molecular mass, 49.5 kDa; observed molecular mass, 50 kDa) was cleaved into two polypeptides (predicted and observed molecular masses of the carboxyl-terminal product, 29.3 and 27 kDa, respectively; predicted and observed molecular masses of the amino-terminal product, 20.2 and 23 kDa, respectively) (Fig. 5, lane 4). In a control reaction with identical buffer conditions but lacking the 3C-like proteinase, no cleavage was observed (Fig. 5, lane 3).

Identification of ORF 1 gene products in HCV 229E-infected cells. The expression of HCV 229E ORF 1a gene products, and in particular the putative 3C-like proteinase, in HCV 229E-infected MRC-5 cells was studied by the immunoprecipitation of metabolically labelled proteins. To do this, cells infected at a multiplicity of infection of 10 PFU per cell were labelled with [^{35}S]methionine and [^{35}S]cysteine for 2 h at different time points p.i., and the resulting cell lysates were analyzed directly by electrophoresis in SDS-polyacrylamide gels or after immunoprecipitation with antiserum as described in Materials and Methods.

First, we used the fusion protein-specific antiserum K17 to immunoprecipitate cell lysates. Despite a high background of

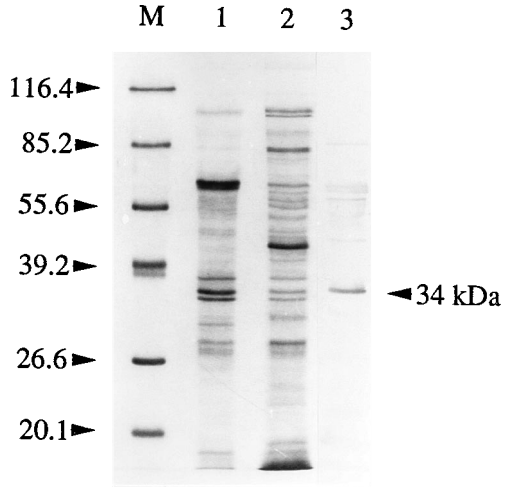

FIG. 3. Partial purification of the 34-kDa fusion protein cleavage product. Fractionated lysates from BMH 71-18 cells transformed with pROS-3CL were separated by SDS-PAGE in a 15% polyacrylamide gel. The insoluble material after treatment with 6 M urea (lane 1) and the soluble fraction after cell lysis (lane 2) were stained with Coomassie brilliant blue R-250. In parallel, a duplicate of lane 2 was transferred to a nitrocellulose membrane and immunostained with the fusion protein-specific antiserum K17 (lane 3). Protein molecular weight markers are shown in lane M.

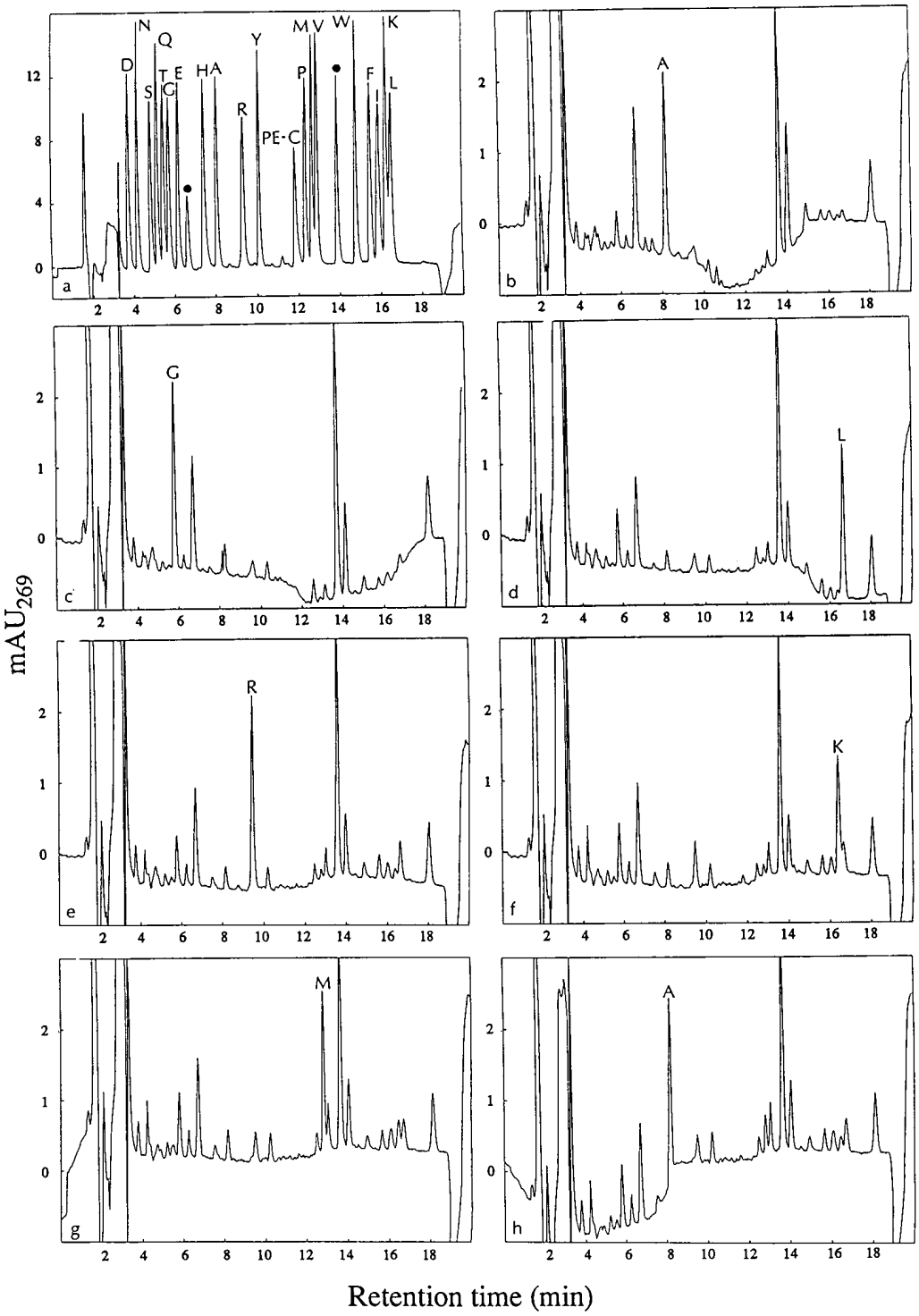

FIG. 4. N-terminal sequence analysis of the 34-kDa cleavage product. The purified 34-kDa polypeptide was subjected to Edman degradation, and phenylthiohydantoin (PTH)-amino acids generated during each reaction cycle were detected by their UV A_{269} (mAU_{269}) and identified by their characteristic retention time on a reverse-phase high-pressure liquid chromatography support. (a) Chromatogram of PTH-amino acid standards (20 pmol); (b to h) chromatograms of PTH-amino acids from reaction cycles 1 to 7. Specific peaks of PTH-amino acids are indicated by the single-letter code. Typical side products of the Edman degradation (DMPTU and DPTU) are indicated by filled circles in the standards chromatogram.

high-molecular-weight proteins and nucleocapsid protein (due to the extended autoradiography), the results clearly show that the K17 antiserum specifically precipitated a protein with an apparent molecular mass of 34 kDa from HCV 229E-infected cells (Fig. 6A, lanes 14 to 18). The amount of immunoprecipitated 34-kDa protein varied during the experiment and was maximal between 3 to 5 and 5 to 7 h p.i. The 34-kDa protein was not precipitated by the corresponding preimmune serum

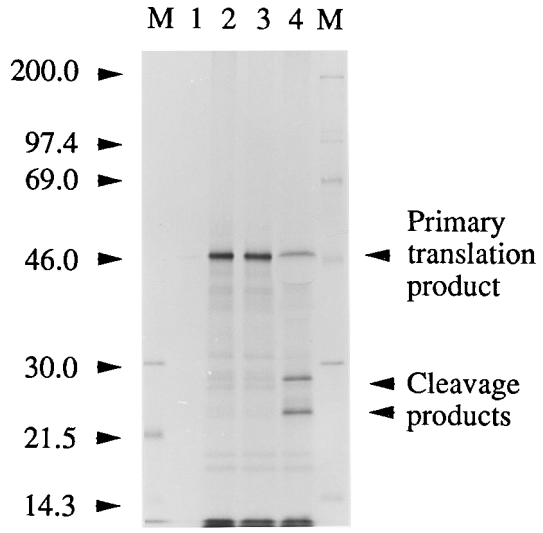

FIG. 5. *trans* cleavage activity of a bacterially expressed HCV 229E 3C-like proteinase. A polypeptide encoded in the HCV 229E ORF 1b was translated in vitro from a synthetic RNA and incubated with the bacterially expressed 3C-like proteinase as described in Materials and Methods. Lanes: 1, translation reaction without RNA; 2, translation reaction with RNA, without incubation; 3, translation reaction with RNA, incubation with CM buffer; 4, translation reaction with RNA, incubation with CM buffer containing proteinase. Protein molecular weight markers (CFA 626 and CFA 645; Amersham, Braunschweig, Germany) are shown in lanes M, and the primary translation product and cleavage products are indicated.

(Fig. 6A, lanes 8 to 12), nor was it detected in mock-infected cells (Fig. 6A, lanes 7 and 13). The electrophoretic mobility of the 34-kDa protein detected in virus-infected cells was identical to that of the carboxyl-terminal cleavage product of the proteolytic bacterial fusion protein and that of the bacterially expressed 3C-like proteinase described above (data not shown).

The earliest time point at which the intracellular 34-kDa protein could be detected was 3 to 5 h p.i. To compare these kinetics with that of a viral structural protein, the same radiolabelled cell lysates were analyzed after immunoprecipitation with K16 antiserum, i.e., the HCV N fusion protein-specific antiserum. In this experiment, synthesis of the N protein could also be detected as early as 3 to 5 h p.i. (Fig. 6B, lanes 14 to 18). However, in contrast to the 34-kDa protein, the amount of immunoprecipitated N protein rose to a plateau at 5 to 7 h p.i. and then remained constant. As expected, the N protein was not precipitated by the corresponding preimmune serum (Fig. 6B, lanes 8 to 12), nor was it detected in mock-infected cells (Fig. 6B, lanes 7 and 13).

DISCUSSION

Bacterial expression vectors have been successfully used to produce viral proteinases for functional and structural analyses. Typically, the expressed proteins can be used to investigate the catalytic activity and specificity of the enzyme, and, often, they can also be purified to allow inhibitor and X-ray diffraction studies. To date, the structures of several positive-strand RNA virus proteinases have been elucidated by this approach (1, 5, 25).

In the experiments reported here, the use of bacterial expression systems has led us to the following conclusions. First, using a β-galactosidase bacterial fusion protein, we have shown that the predicted HCV 229E 3C-like proteinase domain, located in ORF 1a, encodes a polypeptide with proteolytic activity. In addition, we have been able to define, for the first time, a cleavage site of a coronavirus-encoded 3C-like proteinase. The cleavage site (Q/A) is consistent with the features regarded as typical for the substrate of 3C-like proteinases, namely, cleavage at Q,E/G,S,A (11, 29). These data, therefore, add experimental support to the theoretical predictions made by Gorbalenya and others (14, 16, 22). Additionally, we have shown that the introduction of a mutation that substitutes aspartic acid for histidine at position 3006 in the HCV 229E ORF 1a gene product (which corresponds to amino acid 41 in the 3C-like proteinase domain) abolished the proteolytic activity. This result supports but does not prove the idea, that this amino acid constitutes a catalytic residue of the HCV 229E 3C-like proteinase.

Second, also using a bacterial system, we have expressed a polypeptide with the amino acid sequence of the HCV 229E 3C-like proteinase domain (amino acids 2966 to 3267 of the ORF 1 gene product) and have shown that it has proteolytic activity in *trans*. The results that we have obtained are consistent with the use of a predicted cleavage site at the dipetide Gln-4995–Ala-4996 in the HCV 229E ORF 1ab translation product. However, this conclusion has yet to be confirmed by sequence data.

We hope to develop the bacterial expression systems described above to identify the catalytic residues of the HCV 229E 3C-like proteinase and, possibly, define the sequence requirements for specific *cis* and *trans* cleavage by the enzyme. This in turn would allow for more-refined predictions of processing sites in the HCV 229E ORF 1a and ORF 1ab polyproteins. Eventually, it may be possible to purify an active proteinase for structural and functional analysis.

The second important finding in this study is the identification of a 34-kDa polypeptide as the major intracellular protein containing the HCV 229E 3C-like proteinase domain. The predicted 3C-like proteinase domain of HCV 229E is located in the ORF 1a gene product between Ala-2966 and Gln-3267 (16). The results that we have obtained with the bacterial expression system suggest that, indeed, Ala-2966 represents the amino-terminal residue of the intracellular polypeptide. In addition, it seems likely that the predicted Gln-Ser dipeptide at positions 3267 to 3268 is used as the carboxyl-terminal cleavage site. This would be the only putative cleavage site that is consistent with the size of the 3C-like proteinase polypeptide that we have detected in the infected cell. Furthermore, the experiments done in the bacterial system lead us to predict that the 34-kDa polypeptide detected in virus-infected cells represents an active form of the HCV 229E 3C-like proteinase. They indicate that there is no requirement for additional cofactors to attain enzymatic activity, at least, in vitro. However, we cannot exclude that there are cellular or viral factors that modify the proteinase activity in vivo.

The results shown in Fig. 6 appear to suggest that the HCV 229E proteinase polypeptide is synthesized maximally between 3 and 7 h p.i. However, this conclusion still has to be considered preliminary. First, we have to show that our immunoprecipitation reactions are quantitative for each time point. Second, it is possible that in the course of the infection, the 34-kDa polypeptide becomes sequestered by protein-protein interactions that compete for its reactivity with antibody. Third, the 34-kDa polypeptide is a product of proteolytic processing, and we have no information on the kinetics of this process.

One conclusion that however can be made from our studies is that the HCV 229E 3C-like proteinase is present in the infected cell in very low amounts. Thus, to analyze the genesis of this proteinase, as well as the other RNA polymerase locus gene products, a panel of specific and sensitive immunological

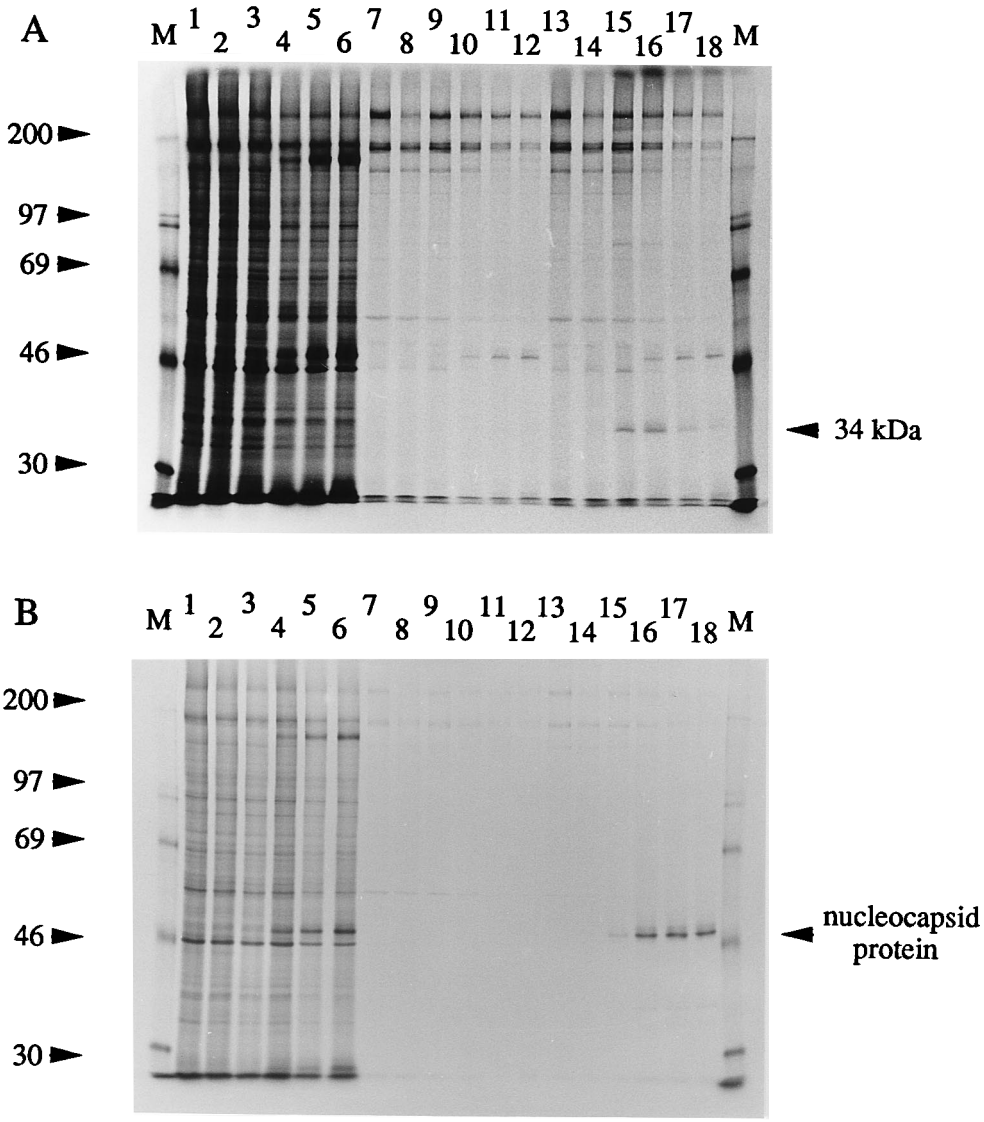

FIG. 6. Detection of a 34-kDa ORF 1a polypeptide in HCV 229E-infected cells. (A) Radiolabelled cell lysates from mock-infected and HCV 229E-infected MRC-5 cells were analyzed directly by SDS-PAGE in a 12.5% polyacrylamide gel or after immunoprecipitation with either preimmune serum or the fusion protein-specific antiserum K17. The following were analyzed: a 0.5-μl lysate that was mock infected and labelled 1 to 3 h p.i. (lane 1); a 0.5-μl lysate that was HCV 229E infected and labelled 1 to 3, 3 to 5, 5 to 7, 7 to 9, and 9 to 11 h p.i. (lanes 2, 3, 4, 5, and 6, respectively); a 100-μl lysate that was immunoprecipitated with preimmune serum, mock infected, and labelled 1 to 3 h p.i. (lane 7); a 100-μl lysate that was immunoprecipitated with preimmune serum, HCV 229E infected, and labelled 1 to 3, 3 to 5, 5 to 7, 7 to 9, and 9 to 11 h p.i. (lanes 8, 9, 10, 11, and 12, respectively); a 100-μl lysate that was immunoprecipitated with K17 antiserum, mock infected, and labelled 1 to 3 h p.i. (lane 13); a 100-μl lysate that was immunoprecipitated with K17 antiserum, HCV 229E infected, labelled 1 to 3, 3 to 5, 5 to 7, 7 to 9, and 9 to 11 h p.i. (lanes 14, 15, 16, 17, and 18, respectively). Protein molecular weight markers (CFA 626; Amersham) are shown in lane M, and the 34-kDa polypeptide is indicated. Autoradiography was for 28 days. (B) Radiolabelled cell lysates from mock-infected and HCV 229E-infected MRC-5 cells were analyzed directly by SDS-PAGE in a 12.5% polyacrylamide gel or after immunoprecipitation with either preimmune serum or the nucleocapsid protein-specific antiserum K16. The following were analyzed: a 1.0-μl lysate that was mock infected and labelled 1 to 3 h p.i. (lane 1); a 1.0-μl lysate that was HCV 229E infected and labelled 1 to 3, 3 to 5, 5 to 7, 7 to 9, and 9 to 11 h p.i. (lanes 2, 3, 4, 5, and 6, respectively); a 4.0-μl lysate that was immunoprecipitated with preimmune serum, mock infected, and labelled 1 to 3 h p.i. (lane 7); a 4.0-μl lysate that was immunoprecipitated with preimmune serum, HCV 229E infected, and labelled 1 to 3, 3 to 5, 5 to 7, 7 to 9, and 9 to 11 h p.i. (lanes 8, 9, 10, 11, and 12, respectively); a 4.0-μl lysate that was immunoprecipitated with K16 antiserum, mock infected, and labelled 1 to 3 h p.i. (lane 13); a 4.0-μl lysate that was immunoprecipitated with K16 antiserum, HCV 229E infected, and labelled 1 to 3, 3 to 5, 5 to 7, 7 to 9, and 9 to 11 h p.i. (lanes 14, 15, 16, 17, and 18, respectively). Protein molecular weight markers are shown in lane M, and the nucleocapsid protein is indicated. Autoradiography was for 3 days.

reagents will be required. These reagents may sometimes be difficult to generate; however, the results presented in this paper show that this approach is feasible. Another approach, which we are also pursuing, is to (over)express the HCV RNA polymerase locus by using a vaccinia virus-T7 expression system. This system should facilitate the analysis of polymerase locus expression and, more importantly, provide a means to investigate the function of individual polymerase gene products by reverse genetics.

ACKNOWLEDGMENTS

We thank Barbara Schelle for technical help and Eric Snijder and Raoul de Groot for advice on radioimmunoprecipitation. We also thank J. Hoppe and T. Brugger for protein sequence data.

This work was supported by grants from the DFG (SFB 165/B1 and GK Infektiologie) and the BMFT (01 KI 8838/0).

REFERENCES

1. **Allaire, M., M. M. Chernaia, B. A. Malcom, and M. N. G. James.** 1994. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. Nature (London) **369:**72–76.
2. **Baker, S. C., C.-K. Shieh, L. H. Soe, M.-F. Chang, D. M. Vannier, and M. M. C. Lai.** 1989. Identification of a domain required for autoproteolytic cleavage of murine coronavirus gene A polyprotein. J. Virol. **63:**3693–3699.
3. **Baker, S. C., K. Yokomori, S. Dong, R. Carlisle, A. E. Gorbalenya, E. V. Koonin, and M. M. C. Lai.** 1993. Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus. J. Virol. **67:**6056–6063.
4. **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.
5. **Choi, H.-K., L. Tong, W. Minor, P. Dumas, U. Boege, M. G. Rossmann, and G. Wengler.** 1991. Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. Nature (London) **354:**37–43.
6. **Contreras, R., H. Cheroutre, W. Degrave, and W. Fiers.** 1982. Simple, efficient in vitro synthesis of capped RNA useful for direct expression of cloned eucaryotic genes. Nucleic Acids Res. **10:**6353–6362.
7. **Denison, M. R., and S. Perlman.** 1986. Translation and processing of mouse hepatitis virus virion RNA in a cell-free system. J. Virol. **60:**12–18.
8. **Denison, M. R., and S. Perlman.** 1987. Identification of putative polymerase gene product in cells infected with murine coronavirus A59. Virology **157:**565–568.
9. **Denison, M. R., P. W. Zoltick, S. A. Hughes, B. Giangreco, A. L. Olson, S. Perlman, J. L. Leibowitz, and S. R. Weiss.** 1992. Intracellular processing of the N-terminal ORF 1a proteins of the coronavirus MHV-A59 requires multiple proteolytic events. Virology **189:**274–284.
10. **Dong, S., and S. C. Baker.** 1994. Determinants of the p28 cleavage site recognized by the first papain-like cysteine proteinase of murine coronavirus. Virology **204:**541–549.
11. **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.
12. **Ellinger, S., R. Glockshuber, G. Jahn, and A. Plückthun.** 1989. Cleavage of procaryotically expressed human immunodeficiency virus fusion proteins by factor Xa and application in Western blot (immunoblot) assays. J. Clin. Microbiol. **27:**971–976.
13. **Fu, K., and R. S. Baric.** 1994. Map locations of mouse hepatitis virus temperature-sensitive mutants: confirmation of variable rates of recombination. J. Virol. **68:**7458–7466.
14. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov.** 1989. Coronavirus genome: prediction of putative functional domains in the nonstructural polyprotein by comparative amino acid sequence analysis. Nucleic Acids Res. **17:**4847–4861.
15. **Harlow, E., and D. Lane.** 1988. Antibodies: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **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.
17. **Herold, J., T. Raabe, and S. G. Siddell.** 1993. Molecular analysis of the human coronavirus (strain 229E) genome. Arch. Virol. **7**(Suppl.)**:**63–74.
18. **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.
19. **Isaacs, D., D. Flowers, J. R. Clarke, H. B. Valman, and M. R. Macnaughton.** 1983. Epidemiology of coronavirus respiratory infections. Arch. Dis. Child. **58:**500–503.
20. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
21. **Lai, M. M. C.** 1990. Coronavirus: organization, replication and expression of genome. Annu. Rev. Microbiol. **44:**303–333.
22. **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.
23. **Leibowitz, J. L., J. R. DeVries, and M. V. Haspel.** 1982. Genetic analysis of murine hepatitis virus strain JHM. J. Virol. **42:**1080–1087.
24. **Liu, D. X., I. Brierley, K. W. Tibbles, and T. D. K. Brown.** 1994. A 100-kilodalton polypeptide encoded by open reading frame (ORF) 1b of the coronavirus infectious bronchitis virus is processed by ORF 1a products. J. Virol. **68:**5772–5780.
25. **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.
26. **McIntosh, K.** 1990. Coronaviruses, p. 857–864. *In* B. N. Fields et al. (ed.), Virology, 2nd ed. Raven Press, New York.
27. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridisation probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12:**7035–7056.
28. **Myint, S., D. Harmsen, T. Raabe, and S. G. Siddell.** 1990. Characterization of a nucleic acid probe for the diagnosis of human coronavirus 229E infections. J. Med. Virol. **31:**165–172.
29. **Palmenberg, A. C.** 1990. Proteolytic processing of picornaviral polyprotein. Annu. Rev. Microbiol. **44:**603–623.
30. **Raabe, T., B. Schelle-Prinz, and S. G. Siddell.** 1990. Nucleotide sequence of the gene encoding the spike glycoprotein of human coronavirus HCV 229E. J. Gen. Virol. **71:**1065–1073.
31. **Routledge, E., R. Stauber, M. Pfleiderer, and S. G. Siddell.** 1991. Analysis of murine coronavirus surface glycoprotein functions by using monoclonal antibodies. J. Virol. **65:**254–262.
32. **Schaad, M. C., S. A. Stohlman, J. Egbert, K. Lum, K. Fu, T. Wei, and R. S. Baric.** 1990. Genetics of mouse hepatitis virus transcription: identification of cistrons which may function in positive and negative strand RNA synthesis. Virology **177:**634–645.
33. **Soe, L. H., C.-K. Shieh, S. C. Baker, M.-F. Chang, and M. M. C. Lai.** 1987. Sequence and translation of the murine coronavirus 5′-end genomic RNA reveals the N-terminal structure of the putative RNA polymerase. J. Virol. **61:**3968–3976.
34. **Wege, H., H. Wege, K. Nagashima, and V. ter Meulen.** 1979. Structural polypeptides of the murine coronavirus JHM. J. Gen. Virol. **42:**37–47.
35. **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.
36. **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.