JOURNAL OF VIROLOGY, Mar. 1993, p. 1385–1395
0022-538X/93/031385-11$02.00/0

Vol. 67, No. 3

Expression and Identification of Hepatitis C Virus Polyprotein Cleavage Products

ARASH GRAKOUI,[1] CZESLAW WYCHOWSKI,[2]† CHAO LIN,[1] STEPHEN M. FEINSTONE,[2] AND CHARLES M. RICE[1]*

Department of Molecular Microbiology, Washington University School of Medicine, Box 8230, 660 South Euclid Avenue, St. Louis, Missouri 63110-1093,[1] and Division of Virology, CBER, Food and Drug Administration, Bethesda, Maryland 20892[2]

Received 12 October 1992/Accepted 13 November 1992

Hepatitis C virus (HCV) is the major cause of transfusion-acquired non-A, non-B hepatitis. HCV is an enveloped positive-sense RNA virus which has been classified as a new genus in the flavivirus family. Like the other two genera in this family, the flaviviruses and the pestiviruses, HCV polypeptides appear to be produced by translation of a long open reading frame and subsequent proteolytic processing of this polyprotein. In this study, a cDNA clone encompassing the long open reading frame of the HCV H strain (3,011 amino acid residues) has been assembled and sequenced. This clone and various truncated derivatives were used in vaccinia virus transient-expression assays to map HCV-encoded polypeptides and to study HCV polyprotein processing. HCV polyproteins and cleavage products were identified by using convalescent human sera and a panel of region-specific polyclonal rabbit antisera. Similar results were obtained for several mammalian cell lines examined, including the human HepG2 hepatoma line. The data indicate that at least nine polypeptides are produced by cleavage of the HCV H strain polyprotein. Putative structural proteins, located in the N-terminal one-fourth of the polyprotein, include the capsid protein C (21 kDa) followed by two possible virion envelope proteins, E1 (31 kDa) and E2 (70 kDa), which are heavily modified by N-linked glycosylation. The remainder of the polyprotein probably encodes nonstructural proteins including NS2 (23 kDa), NS3 (70 kDa), NS4A (8 kDa), NS4B (27 kDa), NS5A (58 kDa), and NS5B (68 kDa). An 82- to 88-kDa glycoprotein which reacted with both E2 and NS2-specific HCV antisera was also identified (called E2-NS2). Preliminary results suggest that a fraction of E1 is associated with E2 and E2-NS2 via disulfide linkages.

Prospective and retrospective serologic studies indicate that even before the implementation of screening tests for hepatitis B surface antigen, non-B hepatitis accounted for the majority of transfusion-associated hepatitis in the United States (21, 22, 54). Recently, the major etiologic agent of non-A, non-B hepatitis (NANBH), hepatitis C virus (HCV), was cloned and sequenced (7, 32). This breakthrough has led to the development of immunological and nucleic acid-based methods for detecting HCV infection. HCV infection can result in various clinical outcomes, including acute hepatitis, chronic hepatitis, cirrhosis, or the establishment of an asymptomatic carrier state which may persist for life (for a review, see reference 31). Recent studies have also uncovered a strong association between chronic HCV infection and the development of hepatocellular carcinoma (12, 13, 61).

Since the initial molecular cloning of this agent and implementation of first-generation diagnostics, sequence data for a number of independent HCV isolates have been reported, immunoassays for detection of antibody have been improved, and our knowledge of HCV molecular biology is advancing rapidly (for a review, see reference 33). On the basis of their genome organizations and virion properties, HCV (9, 33), the pestiviruses (11), and the flaviviruses (4) have been classified as three genera in the family *Flaviviridae* (19). Properties shared by these three groups include a lipid envelope, conferring sensitivity to organic solvents, and a single-stranded, positive-polarity RNA genome containing a long open reading frame (ORF) which encodes the viral polypeptides. Polyproteins encoded by the HCV, flavivirus, and pestivirus ORFs are ~3,000, ~3,400, and ~4,000 amino acids long, respectively. The structural proteins are located in the N-terminal portion of the polyprotein and are followed by the putative nonstructural replicase components. Mature proteins, at least as shown for the flaviviruses and pestiviruses (for a review, see references 10 and 58), are produced by a combination of host and viral proteinases located in both the cytosol and the subcellular vesicular compartments.

Although the cleavage products and proteolytic processing schemes of the flaviviruses and pestiviruses have been extensively characterized, similar information has been reported only for the structural protein coding region of HCV (30). In this report, a hybrid vaccinia virus-T7 transient expression system (16, 20, 48) has been used to study processing of the entire HCV ORF. HCV-specific cleavage products were identified by using a collection of region-specific polyclonal rabbit antisera. These results provide a preliminary picture of HCV processing and a map of the polyprotein cleavage products.

MATERIALS AND METHODS

Cell cultures and virus growth. The BHK-21 and CV-1 cell lines were obtained from the American Type Culture Collection (ATCC), Rockville, Md., the BSC-40 cell line (3) was obtained from D. Hruby (Oregon State University), and the A16 subclone of the human hepatoma HepG2 cell line (ATCC) was generously provided by Alan Schwartz (Wash-

* Corresponding author.
† Present address: Unité de Virologie Moléculaire, Institut Pasteur, 75724 Paris Cedex 15, France.

TABLE 1. HCV immunogens expressed in *E. coli*

Construct[a]	HCV restriction site[b]	Vector site[c]	Rabbit no.	Specificity
pET-3xa/HCV 1-142	***Apa*LI** (335), ***Nar*I** (762)	*Bam*HI[d]	WU120	C
pET-3xa/HCV 236-382	***Xho*I** (1046), ***Sal*I** (1482)	*Bam*HI[d]	WU122	E1
pMALc/HCV 393-670	*Nae*I (1515), ***Bsp*MI** (2357)	*Stu*I	WU105	E2
pET-8c/HCV 936-1032	*Nco*I[e], ***Spl*I** (3433)	*Nco*I, ***Bam*HI**	WU107	NS2
pMALc/HCV 1039-1207	***Stu*I** (3452), ***Bam*HI**[e]	***Eco*RI**	WU110	NS3
pET-8c/HCV 936-1207	*Nco*I[e], *Bam*HI[e]	*Nco*I, *Bam*HI	WU43	NS2, NS3
pMALc/HCV 1240-1488	***Eco*0109** (4057), *Bam*HI[e]	*Stu*I, *Bam*HI	WU117	NS3
pMALc/HCV 1651-1976	***Eag*I** (5290), ***Eco*NI** (6265)	***Eco*RI**	WU111	NS4A, NS4B
pET-3xa/HCV 1977-2313	***Eco*NI** (6265), ***Nco*I** (7274)	*Bam*HI[d]	WU123	NS5A
pMALc/HCV 2312-2623	***Nco*I** (7274), *Eco*RI (8205)	*Stu*I, *Eco*RI	WU113	NS5A, NS5B
pMALc/HCV 2622-2872	*Eco*RI (8205), *Xba*I[f]	*Eco*RI, *Xba*I	WU115	NS5B

[a] Constructs producing fusion proteins are shown in bold type. The numbers refer to the amino acid sequence of the HCV ORF which is included in each construct.

[b] Restriction sites in the HCV cDNA used for the plasmid constructs. Nucleotide numbers in parentheses refer to the full-length HCV H-strain sequence (14), assuming that the 5′ noncoding region of HCV H and HCV type 1 (8, 28) are the same length (341 nucleotides). Restriction sites shown in bold type indicate that protruding ends were treated with either the Klenow fragment of DNA polymerase I or T4 DNA polymerase before ligation to produce blunt ends.

[c] Restriction sites in plasmid vectors used for cloning. Sites shown in bold type indicate that protruding ends were treated with either the Klenow fragment of DNA polymerase I or T4 DNA polymerase, prior to ligation to produce blunt ends.

[d] In the case of pET-3xa/HCV 1-142, pET-3xa/HCV 236-382, and pET-3xa/HCV 1977-2313, the indicated HCV cDNA fragments were first cloned into the *Stu*I site of pMALc. The *Bam*HI fragments from these pMALc constructs (containing the properly oriented HCV insert) were then subcloned into *Bam*HI-digested pET-3xa.

[e] For constructs beginning with amino acid residue 936 or ending with residue 1207 or 1488, the restriction sites used for cloning, *Nco*I and *Bam*HI, respectively, were derived by PCR.

[f] This *Xba*I site is in the pGEM-3Zf(+) HCV(H) 17-8958 polylinker.

ington University, St. Louis, Mo.). Cell monolayers were grown in Eagle's minimal essential medium (MEM) supplemented with 2 mM L-glutamine, nonessential amino acids, penicillin, streptomycin, and 10% fetal bovine serum (FBS).

Stocks of vTF7-3, a vaccinia virus recombinant expressing the T7 DNA-dependent RNA polymerase (20), and various vaccinia virus-HCV recombinants were grown in BSC-40 monolayers and partially purified (34), and titers of infectious progeny were determined by plaque assay on BSC-40 cells (34).

HCV cloning and sequence analysis. Cloning and sequence analysis of the HCV Hutchinson (H) strain (18) are only briefly described here. The HCV H strain, a human isolate from an American with posttransfusion NANBH, was passaged twice in chimpanzees. Both of these animals developed elevated serum alanine amino transferase levels and acute hepatitis. Liver tissue from the second chimpanzee passage was used for preparation of crude RNA suitable for cDNA synthesis, nested polymerase chain reaction (PCR) amplification (60, 72), and cloning. Synthetic oligonucleotide primers for amplification of specific regions of the HCV genome were originally synthesized on the basis of the published HCV type 1 cDNA sequence (8 [and references therein]). PCR-amplified cDNA was cloned into bacterial plasmid vectors, and several independent clones were isolated and used for sequence analysis, expression studies, and reconstruction of longer cDNA clones. Utilizing partial sequence data and restriction enzyme mapping, a clone containing the entire ORF, called pGEM-3Zf(+) HCV(H) 17-9389F (see below), has been assembled. The clone has been completely sequenced (14) by the Sanger method (62) with a set of synthetic oligonucleotide primers whose sequences were based on preliminary H strain sequence data. The sequence of this clone is colinear and is >98.5% homologous (at the nucleotide level) to recently published full-length (35) and partial (50) HCV H strain sequences.

Bacterial expression constructs. Constructs were made by using standard methodology (62), and regions amplified by PCR (60) were verified by sequence analysis (62). *Escherichia coli* expression systems included the pET-3x series, which produce N-terminal fusions with the T7 gene *10* product (66), pMALc derivatives, which produce N-terminal fusions with *E. coli* maltose-binding protein (New England Biolabs), and pET-8c (also called pET-3d), which was used to produce unfused HCV proteins by using the T7 expression system (66). The expression constructs and subcloning strategies are summarized in Table 1.

HCV-specific antisera. For production of HCV region-specific antisera, HCV polypeptides or fusion proteins expressed in *E. coli* were obtained from the insoluble fraction (5) or total cell extracts and purified by preparative sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) (5, 39). Gel slices containing the antigens were stored frozen or lyophilized and then emulsified with complete Freund's adjuvant prior to immunization of rabbits (5). Serum samples were collected after multiple booster injections with incomplete Freund's adjuvant. Reactivity and specificity of the antisera were assessed by immunoprecipitation assays with a set of radiolabeled HCV-specific cell-free translation products (data not shown; see also Results and Table 1).

Serum samples from human patients chronically infected with HCV were generously provided by Henry Hsu and Harry Greenberg (Stanford University). Coded patient designations are R, F, RJ, DS, and JHF.

Mammalian expression constructs. As described above, pGEM-3Zf(+) HCV(H) 17-9389F was used as the parent for mammalian expression constructs. Plasmid expression vectors were derivatives of pTM3 (provided by B. Moss) (48). pTM3 contains a unique *Nco*I site and a polycloning region immediately 3′ to the T7 promoter and the encephalomyocarditis virus (EMC) internal ribosome entry site (IRES) (16). This system yields high levels of protein expression by using vTF7-3, a vaccinia virus recombinant which expresses the T7 DNA-dependent RNA polymerase (20). In addition, pTM3 contains flanking vaccinia virus DNA and a dominant selectable marker which readily allow rescue of the corresponding vaccinia virus recombinants (see below).

TABLE 2. HCV mammalian expression constructs

Construct[a]	HCV restriction site[b]	Vector site[c]	C-terminal residue (no.)[d]
pTM3/HCV 1-3011	***Apa*LI** (335), *Xba*I[e]	***Nco*I, *Spe*I**	0
pBRTM/HCV 1-3011	***Apa*LI** (335), *Xba*I[e]	***Nco*I, *Spe*I**	0
pBRTM/HCV 827-3011	*Nco*I[g], *Xba*I[e]	*Nco*I, *Spe*I	0
pBRTM/HCV 1-2940[f]	***Apa*LI** (335), ***Nde*I** (9158)	***Nco*I, *Pac*I**	6
pBRTM/HCV 1-2813	***Apa*LI** (335), *Nru*I (8776)	***Nco*I, *Stu*I**	0
pBRTM/HCV 1-2508	***Apa*LI** (335), ***Hind*III** (7861)	***Nco*I, *Pac*I**	0
pBRTM/HCV 1-2398	***Apa*LI** (335), ***Bam*HI** (7529)	***Nco*I, *Stu*I**	0
pBRTM/HCV 1-2205	***Apa*LI** (335), *Pvu*II (6954)	***Nco*I, *Pac*I**	0
pBRTM/HCV 1-2101	***Apa*LI** (335), *Sna*BI (6642)	***Nco*I, *Pac*I**	0
pBRTM/HCV 1-2051	***Apa*LI** (335), ***Sse*8387I** (6493)	***Nco*I, *Pac*I**	0
pBRTM/HCV 1-1957	***Apa*LI** (335), ***Bsu*36I** (6209)	***Nco*I, *Stu*I**	1
pBRTM/HCV 1-1864	***Apa*LI** (335), ***Bsm*I** (5934)	***Nco*I, *Stu*I**	3
pBRTM/HCV 1-1773	***Apa*LI** (335), *Ssp*I (5659)	***Nco*I, *Stu*I**	1
pBRTM/HCV 1-1692	***Apa*LI** (335), *Nae*I (5414)	***Nco*I, *Stu*I**	3
pBRTM/HCV 1-1676[f]	***Apa*LI** (335), *Hinc*II (5366)	***Nco*I, *Stu*I**	1
pBRTM/HCV 1-1546	***Apa*LI** (335), *Sma*I (4976)	***Nco*I, *Stu*I**	3
pTM3/HCV 1-1488	***Apa*LI** (335), *Bam*HI[g,h]	***Nco*I, *Pst*I**	0

[a] The numbers refer to the portion of the HCV polyprotein encoded by each construct. Flanking residues present in the polyproteins are not included. For all of the constructs except pBRTM/HCV 827-3011, three additional N-terminal residues (Met-Cys-Thr) are predicted to be present prior to the Met residue initiating the HCV polyprotein (see Materials and Methods).

[b] Restriction sites in the HCV cDNA used for the plasmid constructs. Nucleotide numbers in parentheses refer to the positions of these sites in the full-length HCV H-strain sequence (14), assuming that the 5′ noncoding region of HCV H and HCV type 1 (8, 28) are the same length. Restriction sites shown in bold type indicate that protruding ends were treated with either the Klenow fragment of DNA polymerase I or T4 DNA polymerase prior to ligation to produce blunt ends. All constructs contained the expected sequences at the ligation junctions except as noted above.

[c] Restriction sites in plasmid vectors used for cloning. Sites shown in bold type indicate that protruding ends were treated with either the Klenow fragment of DNA polymerase I or T4 DNA polymerase prior to ligation to produce blunt ends.

[d] Number of predicted non-HCV C-terminal residues prior to the first termination codon.

[e] This *Xba*I site is in the pGEM-3Zf(+) HCV(H) 17-8958 polylinker.

[f] In both pBRTM/HCV 1-2940 and pBRTM/HCV 1-1676, two nucleotides were found to be deleted at the 3′ ligation junction during subcloning.

[g] For constructs beginning with residue 827 or ending with residue 1488, the restriction sites used for cloning, *Nco*I and *Bam*HI, respectively, were derived by PCR.

[h] In the case of pTM3/HCV 1-1488 the HCV cDNA fragment containing the PCR-engineered termination codon and *Bam*HI site was obtained from pH2JC1/HCV N3.8C2 as an *Eco*O47III(2848)-*Nsi*I fragment which was subcloned into pTM3/HCV 1-966 (*Eco*O47III-*Pst*I) (27).

pTM3/HCV 1-3011 was constructed by several subcloning steps and contains the entire HCV ORF and short 5′ and 3′ flanking sequences (cDNA from nucleotide 336 to 9389). The HCV cDNA sequence is located immediately 3′ to the *Nco*I site of pTM3 which had been filled in with the Klenow fragment of *E. coli* DNA polymerase I. The 5′ DNA sequence (5′-. . . CC**ATGTGCACCATGA**. . . -3′) contains the ATG corresponding to the preferred translation initiation site of the EMC IRES followed, in frame, by the ATG which initiates the long HCV ORF. This results in the addition of three non-HCV amino acid residues to the N terminus of the predicted translation product. The 3′ flanking sequence is 5′-. . . **TGA**ACGGGGAGCTAGGGGATCCTCTAGT. . . -3′, where the termination codon of the HCV ORF is shown in boldface and the underlined nucleotides correspond to the remnants of the pTM3 *Spe*I restriction site inactivated during subcloning. pTM3/HCV 1-1488 contains an amber termination codon, engineered by PCR, following residue 1488 of the HCV ORF.

Since near-full-length HCV clones in pTM3 were found to be difficult to propagate, a plasmid derivative with a lower copy number and tetracycline resistance was constructed by using pBR322 (designated the pBRTM series). The *Xba*I-*Pvu*I fragment of pTM3/HCV 336-9389F, which contains the HCV cDNA insert as well as the flanking vaccinia virus sequences, was inserted into pBR322 (2) which had been digested with *Dra*I and *Eco*RI. Prior to ligation, both DNA fragments were treated with T4 DNA polymerase in the presence of deoxynucleoside triphosphates. In the parental plasmid, pBRTM/HCV 1-3011, the HCV cDNA coding sense is oriented in the same direction as *tet* transcription. The subcloning strategies for other pBRTM expression constructs used in these studies are summarized in Table 2.

Two expression constructs, pTM3/HCV 1-1488 and pBRTM/HCV 827-3011, were used to construct vaccinia virus recombinants. The corresponding vaccinia virus-HCV recombinants, vHCV 1-1488 and vHCV 837-3011, were generated by marker rescue on CV-1 cells (42) and identified by using the *gpt* selection method (17). Recombinant viruses were plaque purified three times under selective conditions prior to growth of large-scale stocks.

Vaccinia virus transient-expression assays. For expression assays utilizing vaccinia virus-HCV recombinants, the indicated cell types were infected with vTF7-3 alone or in combination with vHCV1-1488 or vHCV 827-3011 by using a multiplicity of infection of 5 PFU per cell of each recombinant (as determined on BSC-40 monolayers). After 30 min at room temperature, the inoculum was removed and replaced with MEM containing 2% FBS. At 2 h postinfection, monolayers were washed once with prewarmed MEM lacking methionine and labeled by incubation for 4 h at 37°C with MEM containing 1/40 the normal concentration of methionine, 2% FBS, and 50 μCi of ^{35}S-translabel (ICN) per ml.

Expression assays of transfected plasmid constructs utilized subconfluent monolayers of BHK-21 cells in 35-mm dishes (approximately 10^6 cells) which had been previously infected with vTF7-3 (5 to 10 PFU per cell) in 0.2 ml MEM for 30 min at 37°C. After removal of the inoculum, cells were transfected at 37°C by using a mixture consisting of 1 μg of plasmid DNA and 12.5 μg of Transfectam (Promega) in 0.5 ml of MEM. After 2.5 h, the transfection mixture was removed and the cells were incubated for 4 h at 37°C in 0.5

ml of MEM containing 1/40 the normal concentration of methionine, 2% FBS, and 40 μCi of ^{35}S-translabel (ICN) per ml.

Cell lysis, immunoprecipitation, and protein analyses. Labeled monolayers were washed with phosphate-buffered saline, lysed with a solution of 0.5% SDS containing 20 μg of phenylmethylsulfonyl fluoride per ml (~0.3 ml per 10^6 cells), and sheared by repeated passage through a 26-gauge needle. If the lysates were not used immediately, aliquots were stored frozen at −70°C. Before use, samples were heated at 70°C for 15 min, diluted into the immunoprecipitation buffer containing Triton X-100 and carrier bovine serum albumin (57), and clarified by centrifugation at 16,000 × *g* for 15 min. Portions of each lysate were incubated with the indicated HCV region-specific antisera (usually 5 μl), and the immune complexes were collected by using *Staphylococcus aureus* Cowan strain I (Calbiochem) as described previously (57). Immunoprecipitates were solubilized and analyzed by SDS-PAGE (39, 63). After treatment for fluorography by using either diphenyloxazole (40) or En^{3}Hance (Du Pont), gels were dried and exposed at −70°C by using prefogged (40) X-ray film (Kodak). The apparent molecular weights of HCV-specific antigens were estimated by comparison with ^{14}C-methylated marker proteins (Amersham).

RESULTS

Identification of HCV polyprotein processing products. To examine the synthesis and processing of HCV proteins in mammalian cell cultures, a series of expression plasmids were assembled from HCV H strain cDNA clones (Fig. 1C). These constructs were designed for expression of HCV polyproteins by using the vaccinia virus-T7 system, either by using a plasmid transfection protocol or by rescue of vaccinia virus-HCV recombinants (vHCV). Uncapped mRNA transcripts were made by the T7 DNA-dependent RNA polymerase encoded by vaccinia virus recombinant vTF7-3 and contained the EMC 5′ IRES in order to achieve efficient cap-independent translation of HCV coding regions. To identify HCV-specific polyproteins and cleavage products, subregions of the HCV polyprotein were expressed in *E. coli*, purified, and used to produce a panel of polyclonal rabbit antisera suitable for immunoprecipitation of SDS-denatured HCV antigens (Table 1; Fig. 1B).

Two vaccinia virus recombinants encoding overlapping polyproteins were used to produce the [^{35}S]Met-labeled HCV-specific products shown in Fig. 2. vHCV1-1488 is predicted to express an HCV polyprotein which initiates with 3 extra N-terminal residues followed by the first 1,488 residues of the HCV ORF (Fig. 1). The second recombinant, vHCV827-3011, begins at Met-827 in the putative NS2 region and extends to the end of the HCV ORF. These two recombinants were used since preliminary experiments indicated that C-terminal HCV polypeptides were underproduced in cells expressing the entire HCV polyprotein (either infected with vHCV1-3011 or transfected with pBRTM/HCV 1-3011). However, products of identical immunoreactivity and size identical to those shown in Fig. 2 are observed in expression studies with constructs expressing the full-length polyprotein (26).

Figure 2A shows the products immunoprecipitated from SDS-denatured extracts from vHCV-infected BHK-21 cells. Antiserum WU120, directed against a fusion protein containing HCV polyprotein residues 1 to 142, immunoprecipitated a protein of 21 kDa which is thought to represent the HCV capsid protein. Antiserum WU122, which was directed against the putative HCV E1 envelope protein sequences, weakly immunoprecipitated a diffuse 31-kDa protein. Surprisingly, this polypeptide was also immunoprecipitated by antisera WU105 and WU107, which are directed against the

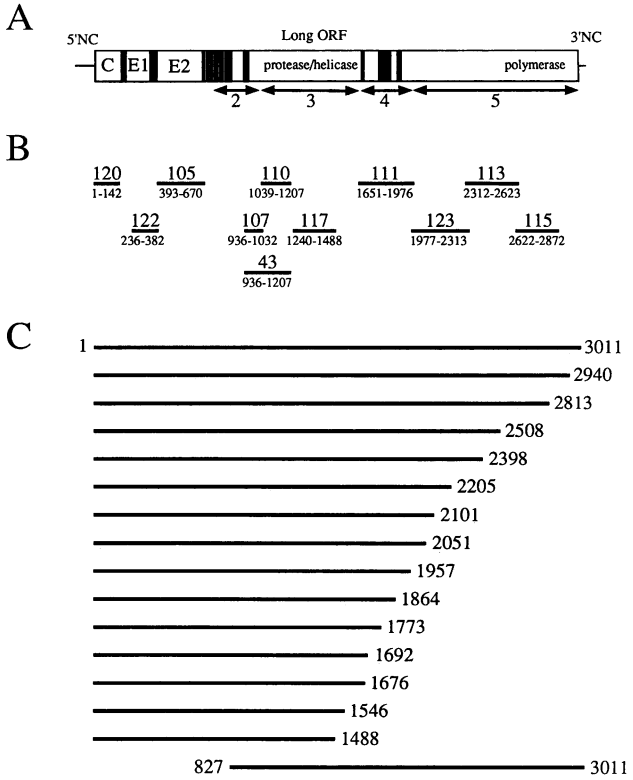

FIG. 1. HCV genome structure, region-specific antisera, and expression constructs. (A) Diagram of the HCV H strain genome RNA is shown with the 5′ and 3′ noncoding regions (NC) indicated by lines and the long ORF denoted as a box. It is not known if the H strain genome RNA contains a 5′-terminal cap structure or a 3′-terminal poly(A) (28) or poly(U) (36, 52, 67) tract. The locations of the putative structural proteins (30), the basic capsid protein (C) and two envelope glycoproteins, E1 and E2, are shown. Regions of the polyprotein containing predominantly uncharged amino acids are indicated as black bars. In this report, the nomenclature used to describe the remaining regions of the HCV polyprotein is based on that of the flaviviruses (for a review, see reference 4) and assumes similar functional organization (8, 47, 67). It appears that HCV may not encode a protein analogous to the secreted nonstructural NS1 glycoprotein of flaviviruses (64). Following E2, the HCV polyprotein contains a hydrophobic portion, like the NS2 region of flaviviruses, which precedes a putative serine proteinase domain (1) and NTPase and helicase motifs (23) which are present in the flavivirus NS3 protein (58, 74). Following the NS3 region is another hydrophobic region called NS4. The remaining portion of the ORF is referred to as the NS5 region, and the C-terminal part of this coding sequence contains the Gly-Asp-Asp motif characteristic of RNA-dependent-RNA polymerases (53). (B) The portions of the HCV polyprotein used as immunogens for production of polyclonal rabbit antisera are indicated as black lines. Above each line is the designation for each antiserum as used in this report; below each line is the region of the polyprotein present in each expression construct (numbered from the first Met residue in the long HCV ORF). See Table 1 and Materials and Methods for further details. (C) Summary of the HCV polyprotein expression constructs used in this study (Table 2). Polyprotein sequences present in each construct are indicated by black lines which are drawn to scale and oriented with respect to the diagram of the HCV genome shown in panel A.

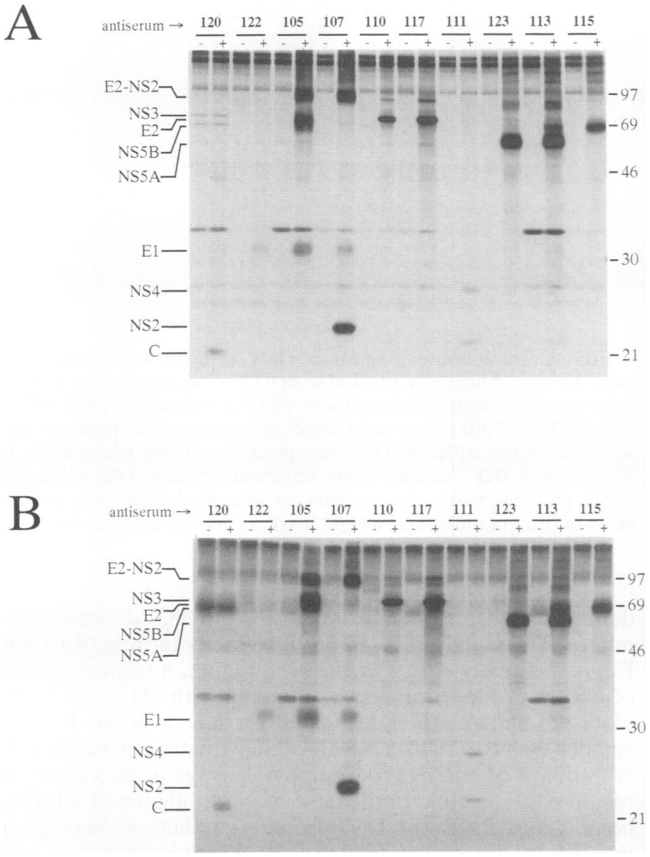

FIG. 2. Identification of HCV polyprotein cleavage products. BHK-21 monolayers were infected with vTF7-3 alone (−) or coinfected with vTF7-3 and either vHCV 1-1488 or vHCV 827-3011 (+) and labeled with ^{35}S-translabel as described in Materials and Methods. Cell lysates were prepared and immunoprecipitated by using the indicated HCV region-specific antisera, as described in Table 1 and Fig. 1. As discussed in Results, lysates from cells coinfected with vHCV 1-1488 were used for immunoprecipitations with region-specific antisera WU120, WU122, WU105, and WU107. Lysates from cells coinfected with vHCV 827-3011 were used for immunoprecipitations with region-specific antisera WU110, WU117, WU111, WU123, WU113, and WU115. Samples were separated on SDS–14% polyacrylamide gels. (A) Lysates prepared from BHK-21 cells; (B) lysates prepared from HepG2 A16 cells. HCV-specific proteins are identified at the left of each panel with the sizes of protein molecular weight markers indicated at the right.

E2 and NS2 regions of the polyprotein, respectively (see Discussion). WU105 antiserum also immunoprecipitated diffuse products of 88 and 70 kDa (the smaller product being the putative E2 glycoprotein). The 88-kDa product was also immunoprecipitated by WU107 antiserum, which is directed against the NS2 region, and is therefore called E2-NS2. A predominant 23-kDa product also reacted with the NS2 region-specific antiserum and is referred to as NS2. Antiserum directed against either the putative serine proteinase domain (WU110) or the helicase-NTPase domain (WU117) specifically immunoprecipitated a 70-kDa protein, NS3, which is nearly identical in size to the homologous flavivirus protein. The HCV NS4 region-specific antiserum (WU111) reacted weakly with a 27-kDa product. (The 22-kDa HCV-specific product which is also present in this lane [as well as some of the other lanes] is an N terminally truncated form of the NS2 protein produced by vHCV 827-3011 which often precipitates nonspecifically.) Antiserum directed against the N-terminal portion of the NS5 region (WU123) precipitated a predominant species of 58 kDa, with additional minor slower mobility forms (up to 68 kDa, but difficult to see in the exposures shown in Fig. 2), which are collectively called NS5A. Antiserum to the middle portion of the NS5 region (WU113) reacted with a 68-kDa polypeptide in addition to NS5A. This species, called NS5B, was also immunoprecipitated by antiserum directed against the C-terminal portion of the NS5 region (WU115). Besides these major species, larger polypeptides, consistent with uncleaved polyproteins, were identified with several of the antisera (in particular, see WU110, WU117, WU123, WU113, and WU115). We are currently studying the kinetics of HCV polyprotein processing (41) and the identification of these possible processing intermediates will be discussed in detail elsewhere.

Since hepatocytes are believed to be permissive for HCV infection and replication (49), we also examined the HCV-specific proteins expressed in the human hepatoma cell line, HepG2 A16. Dramatic host-specific differences in processing were not observed, and results essentially identical to those discussed above for BHK-21 cells were obtained (Fig. 2B). Similar patterns of processed products were also found in CV-1 (monkey kidney) and CHO (hamster ovary) cells (data not shown).

N-linked glycosylation of HCV polypeptides. To determine whether any of the HCV-specific proteins expressed by using this system contained asparagine-linked carbohydrate, lysates of ^{35}S-labeled BHK-21 cells were immunoprecipitated with each region-specific antiserum and digested with endoglycosidase F, which removes both high-mannose and complex glycans (15). The only HCV-specific polypeptides converted to faster migrating forms by endoglycosidase F digestion were E1, E2, and E2-NS2, suggesting that they contain N-linked glycans (Fig. 3). E1 was converted from a 31-kDa species to a 21-kDa deglycosylated form. The pattern of E2-specific products was more complex, and at least three E2-specific endoglycosidase F digestion products were observed. The largest product (62 kDa) was also present in the sample immunoprecipitated with NS2 region-specific antiserum (WU107), and hence probably represents deglycosylated E2-NS2. The two other discrete species, of 41 and 36 kDa, presumably represent deglycosylated forms of E2. Whether these multiple forms reflect different E2 polypeptide backbones or result from other posttranslational modifications is unknown (see Discussion). As mentioned above, although SDS-denatured lysates are heated before immunoprecipitation, E1 coprecipitates with the E2 and NS2 region-specific sera, and deglycosylated forms of E1 can be observed in these samples. These results are consistent with previous expression studies which have indicated that HCV E1 and E2 produced in cell-free translation systems (30) or mammalian (37, 45, 64) or insect cells (45) are both heavily modified by N-linked glycosylation. The predicted sequences of the HCV H strain E1 (polyprotein residues 192 to 383 [30]) and E2 (polyprotein residues 384 to ~750 [30]) proteins contain 5 and 9 potential acceptor sites, respectively, and our data indicate that the majority of these sites are utilized in mammalian cells (assuming 2 to 3 kDa per oligosaccharide unit).

HCV antigens recognized by human sera. Although the majority of the HCV polyprotein was represented in the antigens used for production of region-specific sera, immunogens from some regions have not be obtained (note the NS2 region in Fig. 1), and the immunodominant epitopes

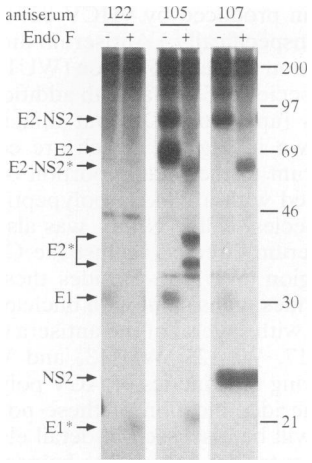

FIG. 3. Endoglycosidase F digestion of HCV glycoproteins. Cell monolayers were coinfected with vTF7-3 and vHCV 1-1488 and labeled with ^{35}S-translabel as described in Materials and Methods. Equivalent portions of the cell lysate were immunoprecipitated with WU122 (E1 specific), WU105 (E2 specific), or WU107 (NS2 specific). Immunoprecipitates were resuspended and incubated overnight in either the absence (−) or presence (+) of endoglycosidase F. Digestions were conducted essentially as previously described (44 [and references therein]). Samples were separated on SDS–14% polyacrylamide gels. HCV-specific proteins are identified at the left, and the sizes of protein molecular weight markers are indicated at the right. The positions of endoglycosidase F-digested forms (indicated by asterisks) are also indicated. Although not shown, parallel samples were analyzed from vTF7-3-infected monolayers to unambiguously allow identification of the HCV-specific products (see also Fig. 2).

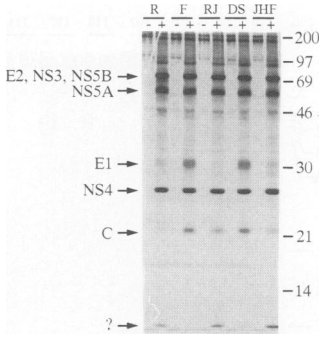

FIG. 4. Immunoprecipitation of HCV antigens with human antisera. Lysates from cells infected with vTF7-3 alone (−) or a mixture of lysates from cells coinfected with vTF7-3 and either vHCV 1-1488 or vHCV 827-3011 (+) were used for immunoprecipitation with serum from five different HCV seropositive patients (denoted R, F, RJ, DS, and JHF. Samples were separated on SDS–14% polyacrylamide gels. HCV-specific proteins are identified at the left, and the sizes of protein molecular weight markers are indicated at the right.

recognized by our panel of sera have not been defined. Hence, some HCV-encoded polypeptides may have been missed in our analyses. In the hope of identifying additional HCV-specific cleavage products, we examined the reactivity of serum samples from five HCV-infected patients. Radiolabeled SDS-denatured lysates were prepared from BHK-21 cells infected with vTF7-3 alone or coinfected with either vHCV 1-1488 or vHCV 827-3011. Coinfected lysates which had been pooled were used for the immunoprecipitation analyses shown in Fig. 4. All patient sera showed strong reactivity with NS5A and the 27-kDa NS4 product and various degrees of reactivity with E1 and C. None of the sera showed detectable reactivity with NS2 or its truncated form. A strong band migrating at ~70 kDa was immunoprecipitated by all patient sera, but since E2, NS3, and NS5B all migrate in this size range it is difficult to interpret these results. In addition to these species, a small HCV-specific polypeptide of ~8 kDa, which had not been previously identified, was immunoprecipitated by serum from patients R, RJ, DS (weak reaction), and JHF. A similarly sized species was also observed in longer exposures of immunoprecipitations of vHCV 827-3011-infected lysates with the NS4 region-specific antiserum (WU111), suggesting that this product is derived from additional processing of the NS4 region (see below).

Fine mapping the positions of HCV nonstructural protein by deletion analyses. The reactivity of the HCV proteins with region-specific antisera can be used to roughly map their locations in the HCV polyprotein. The sizes and immunoreactivities of C, E1, and E2 from our studies are consistent with previous results from cell-free translation studies which defined cleavages after residues 191 and 383, dependent on microsomal membranes, to produce the N termini of E1 and E2, respectively (30). For the NS2 region, a truncated form of NS2 is produced in lysates infected with vHCV 827-3011 (data not shown). This truncated product is about 1.5 kDa smaller than the 23-kDa NS2 protein, which suggests that the N terminus of NS2 is produced by cleavage in the vicinity of residues 805 to 815. In the case of NS3, alignment with the homologous flavivirus NS3 proteins predicts a cleavage in the vicinity of residues 1020 to 1030, which is consistent with the observed reactivity of HCV NS3 with antiserum directed against the fusion protein encompassing residues 1039 to 1207 but not the antiserum directed against the NS2 region (residues 936 to 1032). The locations of the remaining cleavage sites generating the putative HCV nonstructural proteins are less well defined. As mentioned above, NS4 region antiserum (WU111) recognized two proteins of 27 and 8 kDa, but their order in the polyprotein cannot be established from these experiments. Finally, while the order of NS5A and NS5B in the polyprotein is clear from the experiments presented in Fig. 2, the location of the 4-5A and 5A-5B cleavage sites can only be roughly localized on the basis of the apparent sizes and immunoreactivity of these products.

A more precise map of the locations of the putative HCV nonstructural proteins was obtained by examining the cleavage products from a series of polyproteins with C-terminal truncations (diagrammed in Fig. 1). It should be noted that all of these constructs contained the putative HCV serine proteinase domain (approximately residues 1020 to 1207), which has been shown to be required for downstream proteolytic processing (25). Hence, if a C-terminal deletion does not affect processing at the normal cleavage sites, the sizes of truncated products should allow rough mapping of the C-terminal boundaries of HCV proteins. Polyproteins terminating at residues 2940 or 2813 produced truncated NS5B-specific products of 63 or 43 kDa, respectively, compared with the 68-kDa NS5B species produced by the full-length polyprotein (Fig. 5A). Normal NS3, NS4 region products (data not shown), and NS5A (Fig. 5B) were produced by these constructs, which suggests that NS5B is the C-terminal HCV nonstructural protein with its C terminus located at or very near the end of the HCV ORF. Polypro-

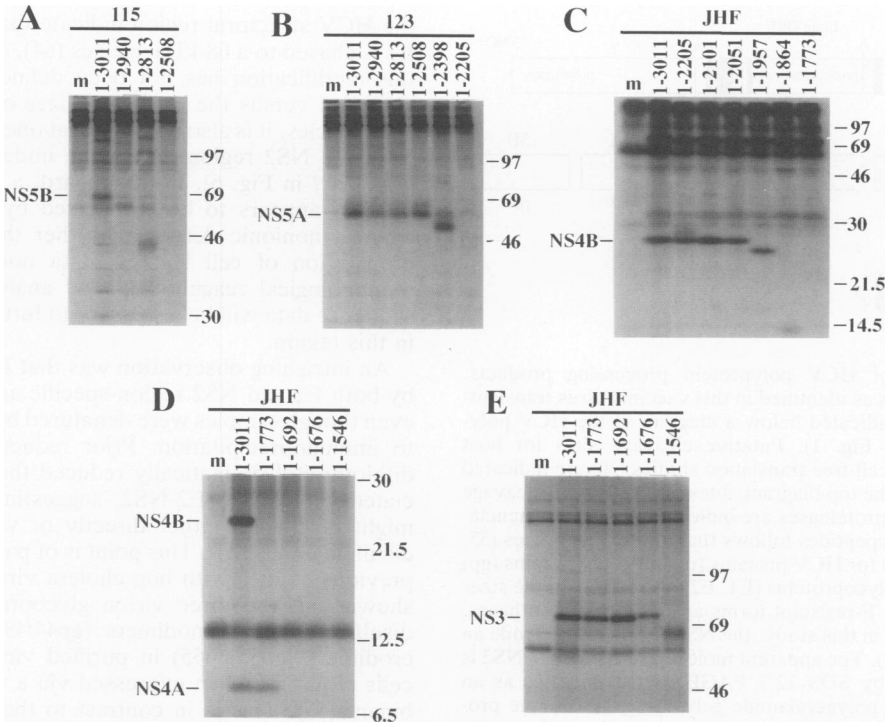

FIG. 5. C-terminal boundaries of the HCV nonstructural proteins. A series of constructs encoding progressive C-terminal deletions (Table 2 and Fig. 1) were used to map the C-terminal boundaries of NS5B (A), NS5A (B), NS4B (C), NS4A (D), and NS3 (E). BHK-21 cells previously infected with vTF7-3 were transfected with the indicated plasmid DNAs or mock transfected (m) and labeled with ^{35}S-translabel as described in Materials and Methods. Cell lysates were prepared, and HCV-specific antigens were immunoprecipitated by using rabbit antiserum WU115 (A), WU123 (B), or human serum from patient JHF (C, D, and E). Immunoprecipitated proteins were separated by electrophoresis on 10% (A and B), 14% (C), or 8% (E) polyacrylamide-SDS gels or a 14% polyacrylamide tricine gel (D). In the case of the NS5A-specific (WU123) immunoprecipitation products (B), it should be noted that a nonspecific protein comigrates with the predominant form of the NS5A protein. HCV-specific proteins are identified at the left of each panel, and the sizes of protein molecular weight markers are indicated at the right.

teins terminating at residues 2398, 2205, 2101, or 2051 produced NS3 and NS4 region species identical to the full-length polyprotein (data not shown; also see Fig. 5C). However, the construct terminating at residue 2398 produced an NS5A-specific species of 53 kDa, truncated by about 5 kDa (Fig. 5B). This suggests that the 5A-5B cleavage site lies between residues 2398 and 2508 (near residue 2440, on the basis of the size of the truncated NS5A product). Polyproteins terminating at residues 1957 and 1864 produced truncated NS4 forms of 25 and 14 kDa (Fig. 5C), respectively, mapping the cleavage site producing the C terminus of the 27-kDa NS4 region species to between residues 1957 and 2051 (near residue 1975). This protein is subsequently referred to as NS4B. Analyses to define the C-terminal boundaries of the 8-kDa NS4 region species and NS3 are shown in Fig. 5D and E. The polyprotein terminating at residue 1773 produced the 8-kDa NS4 region species, but this product disappeared after truncation to residue 1692 (Fig. 5D). This species is subsequently referred to as NS4A, and the data suggest that the 4A-4B cleavage site lies between residues 1692 and 1773. The polyprotein terminating at residue 1692 produced normal NS3; however, the construct truncated to residue 1676 produced a slightly larger (~1-kDa) form, suggesting that 3-4A cleavage had been blocked (Fig. 5E). The 1-1546 polyprotein produced a 63-kDa form of the NS3 protein, truncated by about 7 to 8 kDa. These data are consistent with NS3-NS4A cleavage between residues 1546 and 1676, probably near residue 1665.

Although N- and C-terminal sequence analyses will be needed to define the precise boundaries of the HCV polyprotein cleavage products, these results establish a preliminary map of the HCV H strain-encoded polypeptides and cleavage sites. The HCV polyprotein organization defined by these expression studies is C(p21)-E1(gp31)-E2(gp70)-?-NS2(p23)-NS3(p70)-NS4A(p8)-NS4B(p27)-NS5A(p58)-NS5B(p68). These results are summarized in the diagram shown in Fig. 6, in which the sizes and locations of the cleavage products are drawn to scale on the basis of the data presented in this paper and elsewhere (25, 30).

DISCUSSION

The vaccinia virus transient-expression system has been used for numerous studies examining processing of RNA virus polyproteins. In general, the results from these studies mimic the authentic processing reactions observed in virus-infected cells. However, since an efficient cell culture replication system is lacking for HCV, such a comparison is not yet possible. With this caveat in mind, several points have emerged from our studies.

Consistent with previous studies (6, 29, 37, 38, 45, 64), expression of the full-length HCV polyprotein or truncated derivatives containing the putative structural region led to the production of a 21-kDa N-terminal product believed to represent the HCV capsid protein and two glycoproteins which were heavily modified by N-linked glycosylation,

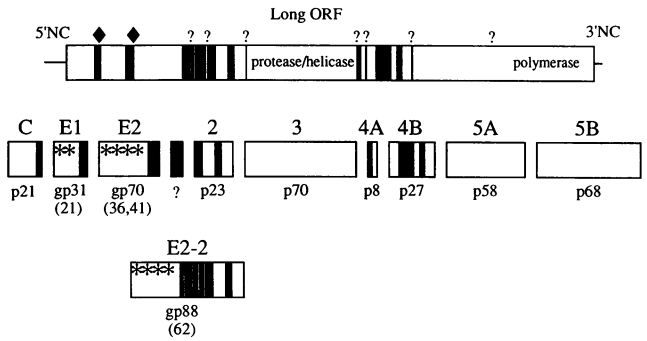

FIG. 6. Summary of HCV polyprotein processing products. HCV cleavage products as identified in this vaccinia virus transient-expression study are indicated below a diagram of the HCV polyprotein (labeled as in Fig. 1). Putative cleavage sites for host signalase identified by cell-free translation studies (30) are indicated by filled diamonds. In the top diagram, sites of polyprotein cleavage mediated by unknown proteinases are indicated (?). The nomenclature used for HCV polypeptides follows that of the flaviviruses (55, 56). The observed sizes for HCV proteins (p) and glycoproteins (gp) are indicated. For the glycoproteins (E1, E2, and E2-NS2), the sizes of the endoglycosidase F-resistant forms are given in parentheses. Although not identified in this study, the NS2 region may encode an additional product(s) (?). The apparent molecular mass of E2-NS2 is 88 kDa, as measured by SDS–12% PAGE, but it migrates as an 82-kDa species on 8% polyacrylamide gels. Asterisks denote proteins with N-linked glycans but do not necessarily indicate the position or number of sites utilized. See the text for further discussion.

designated E1 (31 kDa) and E2 (70 kDa). In addition to these products, an 88-kDa glycoprotein reacting with both E2 and NS2 region-specific antisera was identified. Preliminary studies indicate that this E2-NS2 protein may represent a precursor to E2 (24). However, several observations suggest that processing in the E2-NS2 region is complex. Endoglycosidase F digestion of E2, which migrates as a broad band on our SDS gels, produces at least two species of 36 and 41 kDa (obvious heterogeneity is also apparent in the 41-kDa product). The relationship of these two species is not yet clear. In addition, it is difficult to propose a simple processing model for this region on the basis of the apparent sizes of the E2 species, NS2, and E2-NS2. The NS2 protein is 23 kDa, with its C terminus predicted in the vicinity of residues 1020 to 1030. Given that cell-free translation studies have defined the putative signalase cleavage site generating the E2 N terminus after residue 383 (30), this predicts that the polypeptide backbone of an E2-NS2 precursor should be ~71 kDa. However, the endoglycosidase F digestion product derived from the 88-kDa E2-NS2 glycoprotein is only 62 kDa. If the N-terminal residue of E2-NS2 is 384, then a polypeptide with a predicted size of 62 kDa would barely overlap with the NS2 region used to produce our NS2 region-specific antiserum. Although these products may simply exhibit aberrant migration on SDS-polyacrylamide gels, it is also possible that alternative proteolytic processing or other posttranslational modifications are occurring in the E2-NS2 region, leading to the production of multiple forms of E2 with possibly distinct biological functions in HCV replication. Alternatively, the two endoglycosidase F-resistant forms of E2 could reflect a delayed cleavage in the maturation of HCV E2, similar to those observed for the spike glycoproteins of many enveloped viruses (65). Interestingly, pulse-chase studies with a CHO cell line expressing the HCV structural region indicate that the 70-kDa form of E2 is chased to a 68-kDa species (64), although the nature of this modification has not been defined. Finally, from the observed versus the predicted sizes of the various E2 and NS2 species, it is also possible that one or more polypeptides from the NS2 region have gone undetected in our studies (labeled ? in Fig. 6). In this regard, a 10-kDa HCV-specific product appears to be recognized by several human sera when a nonionic detergent, rather than SDS, is used for preparation of cell lysates (data not shown). Additional immunological reagents, kinetic analyses, and N-terminal sequence data will be necessary to further clarify processing in this region.

An intriguing observation was that E1 was coprecipitated by both E2 and NS2 region-specific antisera (Fig. 2 and 4), even though samples were denatured by heating in SDS prior to immunoprecipitation. Prior reduction of samples with dithiothreitol dramatically reduced the amount of E1 associated with E2 or E2-NS2, suggesting that these proteins might be linked, either directly or via other proteins, by disulfide bonds (24). This point is of particular interest, since previous studies with hog cholera virus, a pestivirus, have shown that the three virion glycoproteins are present as disulfide-linked homodimers (gp44/48 and gp55) and heterodimers (gp35-gp55) in purified virions (69), in infected cells (70), and when expressed via a vaccinia virus recombinant (59). This is in contrast to the flavivirus West Nile virus (WN), in which the envelope protein precursors, prM and E, are associated as a stable heterodimer in nonionic detergent, but not by disulfide bridges (73). Although further studies are needed to clarify the situation with HCV, preliminary analyses with the vaccinia virus expression system indicate that some but not all of the putative virion envelope proteins are associated as discrete disulfide-linked oligomers which are formed soon after synthesis (24).

The HCV-specific proteins produced by processing of the remainder of the polyprotein are remarkably similar to those of the pestiviruses and flaviviruses (for reviews, see references 4, 10, and 58). The 70-kDa HCV NS3 protein is nearly identical in size to the homologous NS3 protein of flaviviruses (p80 in the case of some pestiviruses). For flaviviruses (58) and pestiviruses (75), the N-terminal one-third of this protein has been shown to function as a serine proteinase mediating several cleavages in the viral polyproteins, and similar results have now been obtained for HCV (25). The remainder of the protein contains motifs characteristic of NTPases and helicases (23), and for WN virus this domain has been shown to possess NTPase activity (74). For strains of the pestivirus bovine viral diarrhea virus (BVDV), a fascinating correlation has been made between production of p80 and virus-induced cytopathic effect and fatal mucosal disease (46 [and citations therein]). In noncytopathic strains, the cleavage producing the N terminus of p80 does not occur, resulting in the production of p125 (p54-p80 polyprotein). In cytopathic strains isolated from animals with mucosal disease, which sometimes occurs after congenital transmission of BVDV, insertion of host sequences and/or duplications in the BVDV genome RNA allow production of both p125 and p80 (46). Thus far, it appears that processing at the equivalent cleavage site (2-3) occurs efficiently in flavivirus-infected cells (58) and for the HCV H strain (at least as assayed by transient expression using vaccinia virus). Whether the efficiency of the 2-3 cleavage in these viruses will also correlate with the severity of cytopathic effects and pathogenesis remains to be determined. Given the growing number of divergent HCV isolates (68; see

reference 33 for a review) and the high mutation rate and evolution of this virus during chronic infection of primate hosts (43, 50, 51, 71), it will be of interest to see if strain-specific differences in processing which correlate with clinical disease can be found.

Immediately C terminal to the NS3 protein, the relatively hydrophobic NS4 region is processed to yield two proteins, NS4A (8 kDa) and NS4B (27 kDa). Similar processing events occur in this region of the flavivirus and pestivirus polyproteins, but the function of these small proteins in virus replication is unknown. For most flaviviruses, NS4A has been difficult to identify because of a lack of immune reagents and possible instability of the protein, but this region of the polyprotein encompasses about ~16 kDa of protein coding sequence. In pestiviruses, the corresponding region is processed to yield the p10 protein. The following protein, NS4B or p30 for the pestiviruses (46), appears to be similar in size for all three flavivirus genera. Finally, two proteins are derived from the HCV NS5 region, NS5A (58 kDa) and the C-terminal NS5B product (66 kDa). The HCV NS5B protein is predicted to contain the RNA-dependent RNA polymerase activity on the basis of the presence of the characteristic Gly-Asp-Asp sequence (residues 2737 to 2739) and surrounding conserved motifs (53). This is similar to the case for pestiviruses, which also produce two cleavage products, called p58 and p75. In flaviviruses, the NS5 region is not further processed but remains as a single polypeptide of ~100 kDa (905 residues in the case of yellow fever virus).

The experiments reported here have given us a preliminary picture of HCV polyprotein organization and processing. However, this view is far from complete, and additional studies are needed to define polyprotein cleavage sites and the responsible proteinases and to verify that the products observed in these expression studies are similar to those produced in authentic HCV infections. Such information should prove valuable for expression and characterization of HCV-encoded enzymes as potential targets for antiviral therapy and will allow future studies to be undertaken to assess the involvement of individual HCV polypeptides in the establishment of chronic infections, virus-induced cytopathic effects (if this is the case), and evasion of immunological surveillance and to determine if these proteins play a direct role in the association of HCV with hepatocellular carcinoma.

ACKNOWLEDGMENTS

We thank Dorota Paluszynska for expert technical assistance; Bernard Moss and F. William Studier for plasmid vectors; Henry Hsu and Harry Greenberg for human sera; Alan Schwartz for the HepG2 A16 cell line; many colleagues, especially John Majors, for helpful discussions during the course of this work; and Peter Bredenbeek, Jean Dubuisson, Mark Heise, Julie Lemm, and Margaret MacDonald for critical reading of the manuscript.

This work was supported by grants from the Pew Memorial Trusts and the Public Health Service (CA57973). C.L. is a predoctoral candidate and was supported in part by the Division of Biology and Biomedical Sciences at Washington University.

REFERENCES

1. **Bazan, J. F., and R. J. Fletterick.** 1990. Structural and catalytic models of typsin-like viral proteases. Sem. Virol. **1**:311–322.
2. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow.** 1977. Construction and characterization of new cloning vehicles: a multipurpose cloning system. Gene **2**:95–113.
3. **Brockman, W. W., and D. Nathans.** 1974. The isolation of simian virus 40 variants with specifically altered genomes. Proc. Natl. Acad. Sci. USA **71**:942–946.
4. **Chambers, T. J., C. S. Hahn, R. Galler, and C. M. Rice.** 1990. Flavivirus genome organization, expression, and replication. Annu. Rev. Microbiol. **44**:649–688.
5. **Chambers, T. J., D. W. McCourt, and C. M. Rice.** 1989. Yellow fever virus proteins NS2A, NS2B, and NS4B: identification and partial N-terminal amino acid sequence analysis. Virology **169**: 100–109.
6. **Chiba, J., H. Ohba, Y. Matsuura, Y. Watanabe, T. Katayama, S. Kikuchi, I. Saito, and T. Miyamura.** 1991. Serodiagnosis of hepatitis C virus (HCV) infection with an HCV core protein molecularly expressed by a recombinant baculovirus. Proc. Natl. Acad. Sci. USA **88**:4641–4645.
7. **Choo, Q.-L., G. Kuo, A. Weiner, L. R. Overby, D. W. Bradley, and M. Houghton.** 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science **244**:359–362.
8. **Choo, Q.-L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, A. Medina-Selby, P. J. Barr, A. J. Weiner, D. W. Bradley, G. Kuo, and M. Houghton.** 1991. Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA **88**:2451–2455.
9. **Choo, Q.-L., A. J. Weiner, L. R. Overby, G. Kuo, M. Houghton, and D. W. Bradley.** 1990. Hepatitis C virus: the major causative agent of viral non-A, non-B hepatitis. Br. Med. Bull. **46**:423–441.
10. **Collett, M. S.** 1992. Molecular genetics of pestiviruses. Comp. Immunol. Microbiol. Infect. Dis. **15**:145–154.
11. **Collett, M. S., D. K. Anderson, and E. Retzel.** 1988. Comparisons of the pestivirus bovine viral diarrhoea virus with members of the Flaviviridae. J. Gen. Virol. **69**:2637–2643.
12. **Colombo, M., G. Kuo, Q.-L. Choo, M. F. Donato, E. D. Ninno, M. A. Tommasini, N. Dioguardi, and M. Houghton.** 1989. Prevalence of antibodies to hepatitis C virus in Italian patients with hepatocellular carcinoma. Lancet **ii**:1006–1008.
13. **Colombo, M. G., and M. G. Rumi.** 1991. HCV and hepatocellular-carcinoma: direct or indirect connection?, p. 30. *In* 3rd International Symposium on Hepatitis C Virus. Advanced Therapeutics Communications, Strasbourg, France.
14. **Daemer, R., C. Wychowski, A. Grakoui, C. M. Rice, and S. M. Feinstone.** Unpublished data.
15. **Elder, J. H., and S. Alexander.** 1982. *endo*-β-*N*-Acetylglucosaminidase F: endoglycosidase from *Flavobacterium meningosepticum* that cleaves both high-mannose and complex glycoproteins. Proc. Natl. Acad. Sci. USA **79**:4540–4544.
16. **Elroy-Stein, O., T. R. Fuerst, and B. Moss.** 1989. Cap-independent translation of mRNA conferred by encephalomyocarditis virus 5′ sequence improves the performance of the vaccinia virus/bacteriophage T7 hybrid expression system. Proc. Natl. Acad. Sci. USA **86**:6126–6130.
17. **Falkner, F., and B. Moss.** 1988. *Escherichia coli gpt* gene provides dominant selection for vaccinia virus open reading frame expression vectors. J. Virol. **62**:1849–1854.
18. **Feinstone, S., H. J. Alter, H. P. Dienes, Y. Shimizu, H. Popper, D. Blackmore, D. Sly, W. T. London, and R. H. Purcell.** 1981. Non-A, non-B hepatitis in chimpanzees and marmosets. J. Infect. Dis. **144**:588–598.
19. **Francki, R. I. B., C. M. Fauquet, D. L. Knudson, and F. Brown (ed.).** 1991. Classification and nomenclature of viruses. Fifth report of the International Committee on Taxonomy of Viruses. Arch. Virol. **1991**(Suppl. 2):223.
20. **Fuerst, T. R., E. G. Niles, F. W. Studier, and B. Moss.** 1986. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. Proc. Natl. Acad. Sci. USA **83**:8122–8126.
21. **Gocke, D. J.** 1972. A prospective study of posttransfusion hepatitis: the role of the Australia antigen. JAMA **219**:1165–1170.
22. **Gocke, D. J., H. B. Greenberg, and N. B. Kavey.** 1970. Correlation of Australia antigen with posttransfusion hepatitis. JAMA **212**:877–879.
23. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov.** 1989. Two related superfamilies of putative helicases

involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. **17**:4713–4729.
24. **Grakoui, A., C. Lin, and C. Rice.** Unpublished data.
25. **Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice.** Characterizaton of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. Submitted for publication.
26. **Grakoui, A., B. Prágai, and C. Rice.** Unpublished data.
27. **Grakoui, A., and C. Rice.** Unpublished data.
28. **Han, J. H., V. Shyamala, K. H. Richman, M. J. Brauer, B. Irvine, M. S. Urdea, P. Tekamp-Olson, G. Kuo, Q.-L. Choo, and M. Houghton.** 1991. Characterization of the terminal regions of hepatitis C viral RNA: identification of conserved sequences in the 5′ untranslated region and poly(A) tails at the 3′ end. Proc. Natl. Acad. Sci. USA **88**:1711–1715.
29. **Harada, S., Y. Watanabe, K. Takeuchi, T. Suzuki, T. Katayama, Y. Takebe, I. Saito, and T. Miyamura.** 1991. Expression of processed core protein of hepatitis C virus in mammalian cells. J. Virol. **65**:3015–3021.
30. **Hijikata, M., N. Kato, Y. Ootsuyama, M. Nakagawa, and K. Shimotohno.** 1991. Gene mapping of the putative structural region of the hepatitis C virus genome by *in vitro* processing analysis. Proc. Natl. Acad. Sci. USA **88**:5547–5551.
31. **Hollinger, F. B.** 1990. Non-A, non-B hepatitis viruses, p. 2239–2273. *In* B. N. Fields (ed.), Virology. Raven Press, Ltd., New York.
32. **Houghton, M., Q. Choo, and G. Kuo.** November 1988. European patent application number 88310922.5. Publication number 0318216.
33. **Houghton, M., A. Weiner, J. Han, G. Kuo, and Q.-L. Choo.** 1991. Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. Hepatology **14**:381–388.
34. **Hruby, D. E., L. A. Guarino, and J. R. Kates.** 1979. Vaccinia virus replication. I. Requirement for the host-cell nucleus. J. Virol. **29**:705–715.
35. **Inchauspe, G., S. Zebedee, D.-H. Lee, M. Sugitani, M. Nasoff, and A. M. Prince.** 1991. Genomic structure of the human prototype strain H of hepatitis C virus: comparison with the American and Japanese isolates. Proc. Natl. Acad. Sci. USA **88**:10292–10296.
36. **Kato, N., M. Hijikata, Y. Ootsuyama, M. Nakagawa, S. Ohkoshi, T. Sugimura, and K. Shimotohno.** 1990. Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. Proc. Natl. Acad. Sci. USA **87**:9524–9528.
37. **Kohara, M., K. Tsukiyama-Kohara, N. Maki, K. Asano, K. Yamaguchi, K. Miki, S. Tanaka, N. Hattori, Y. Matsuura, I. Saito, T. Miyamura, and A. Nomoto.** 1992. Expression and characterization of glycoprotein gp35 of hepatitis C virus using recombinant vaccinia virus. J. Gen. Virol. **73**:2313–2318.
38. **Kumar, U., D. Cheng, H. Thomas, and J. Monjardino.** 1992. Cloning and sequencing of the structural region and expression of putative core gene of hepatitis C virus from a British case of chronic sporadic hepatitis. J. Gen. Virol. **73**:1521–1525.
39. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
40. **Laskey, R. A., and A. D. Mills.** 1975. Quantitative film detection of ^{3}H and ^{14}C in polyacrylamide gels by fluorography. Eur. J. Biochem. **56**:335–341.
41. **Lin, C., A. Grakoui, and C. Rice.** Unpublished data.
42. **Mackett, M., and G. L. Smith.** 1986. Vaccinia virus expression vectors. J. Gen. Virol. **67**:2067–2082.
43. **Martell, M., J. I. Esteban, J. Quer, J. Genesca, A. Weiner, R. Esteban, J. Guardia, and J. Gomez.** 1992. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of the HCV genome distribution. J. Virol. **66**:3225–3229.
44. **Mason, P. W.** 1989. Maturation of Japanese encephalitis virus glycoproteins produced by infected mammalian and mosquito cells. Virology **169**:354–364.
45. **Matsuura, Y., S. Harada, R. Suzuki, Y. Watanabe, Y. Inoue, I. Saito, and T. Miyamura.** 1992. Expression of processed envelope protein of hepatitis C virus in mammalian and insect cells. J. Virol. **66**:1425–1431.
46. **Meyers, G., N. Tautz, R. Stark, J. Brownlie, E. Dubovi, M. S. Collett, and H.-J. Thiel.** 1992. Rearrangement of viral sequences leads to a cytopathogenic pestivirus. Virology **191**:368–386.
47. **Miller, R. H., and R. H. Purcell.** 1990. Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. Proc. Natl. Acad. Sci. USA **87**:2057–2061.
48. **Moss, B., O. Elroy-Stein, T. Mizukami, W. A. Alexander, and T. R. Fuerst.** 1990. New mammalian expression vectors. Nature (London) **348**:91.
49. **Negro, F., D. Pacchioni, Y. Shimizu, R. H. Miller, G. Bussolati, R. H. Purcell, and F. Bonino.** 1992. Detection of intrahepatic replication of hepatitis C virus RNA by *in situ* hybridization and comparison with histopathology. Proc. Natl. Acad. Sci. USA **89**:2247–2251.
50. **Ogata, N., H. J. Alter, R. H. Miller, and R. H. Purcell.** 1991. Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. Proc. Natl. Acad. Sci. USA **88**:3392–3396.
51. **Okamoto, H., M. Kojima, S.-I. Okada, H. Yoshizawa, H. Iizuka, T. Tanaka, E. E. Muchmore, D. A. Peterson, Y. Ito, and S. Mishiro.** 1992. Genetic drift of hepatitis C virus during an 8.2 year infection in a chimpanzee: variability and stability. Virology **190**:894–899.
52. **Okamoto, H., S. Okada, Y. Sugiyama, K. Kurai, H. Iizuka, A. Machida, Y. Miyakawa, and M. Mayumi.** 1991. Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions. J. Gen. Virol. **72**:2697–2704.
53. **Poch, O., I. Sauvaget, M. Delarue, and N. Tordo.** 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. **8**:3867–3874.
54. **Purcell, R. H., J. H. Walsh, P. V. Holland, A. G. Morrow, S. Wood, and R. M. Chanock.** 1971. Seroepidemiological studies of transfusion-associated hepatitis. J. Infect. Dis. **123**:406–413.
55. **Rice, C. M., E. M. Lenches, S. R. Eddy, S. J. Shin, R. L. Sheets, and J. H. Strauss.** 1985. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science **229**:726–733.
56. **Rice, C. M., E. G. Strauss, and J. H. Strauss.** 1986. Structure of the flavivirus genome, p. 279–326. *In* S. Schlesinger and M. J. Schlesinger (ed.), The Togaviridae and Flaviviridae. Plenum Press, New York.
57. **Rice, C. M., and J. H. Strauss.** 1982. Association of Sindbis virion glycoproteins and their precursors. J. Mol. Biol. **154**:325–348.
58. **Rice, C. M., and J. H. Strauss.** 1990. Production of flavivirus polypeptides by proteolytic processing. Sem. Virol. **1**:357–367.
59. **Rümenapf, T., R. Stark, G. Meyers, and H.-J. Thiel.** 1991. Structural proteins of hog cholera virus expressed by vaccinia virus: further characterization and induction of protective immunity. J. Virol. **65**:589–597.
60. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–491.
61. **Saito, I., T. Miyamura, A. Ohbayashi, H. Harada, T. Katayama, S. Kikuchi, Y. Watanabe, S. Koi, M. Onji, Y. Ohta, Q.-L. Choo, M. Houghton, and G. Kuo.** 1990. Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. Proc. Natl. Acad. Sci. USA **87**:6547–6549.
62. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
63. **Schägger, H., and G. vonJagow.** 1987. Tricine-sodium dodecyl sulfate polyacrylamide gel electrophoresis for the separation of proteins in the range of 1 to 100 kDa. Anal. Biochem. **166**:368–379.
64. **Spaete, R. R., D. Alexander, M. E. Rugroden, Q.-L. Choo, K. Berger, K. Crawford, C. Kuo, S. Leng, C. Lee, R. Ralston, K. Thudium, J. W. Tung, G. Kuo, and M. Houghton.** 1992. Char-

acterization of the hepatitis E2/NS1 gene product expressed in mammalian cells. Virology **188:**819–830.

65. **Strauss, E. G., and J. H. Strauss.** 1985. Assembly of enveloped animal viruses, p. 205–234. *In* S. J. Casjens (ed.), Virus structure and assembly. Jones and Bartlett, Portola Valley, Calif.
66. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185:**60–89.
67. **Takamizawa, A., C. Mori, I. Fuke, S. Manabe, S. Murakami, J. Fujita, E. Onishi, T. Andoh, I. Yoshida, and H. Okayama.** 1991. Structure and organization of the hepatitis C virus genome isolated from human carriers. J. Virol. **65:**1105–1113.
68. **Tanaka, T., N. Kato, M. Nakagawa, Y. Ootsuyama, M.-J. Cho, T. Nakazawa, M. Hijikata, Y. Ishimura, and K. Shimotohno.** 1992. Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals. Virus Res. **23:**39–53.
69. **Thiel, H.-J., R. Stark, E. Weiland, T. Rumenapf, and G. Meyers.** 1991. Hog cholera virus: molecular composition of virions from a pestivirus. J. Virol. **65:**4705–4712.
70. **Weiland, E., R. Stark, B. Haas, T. Rümenapf, G. Meyers, and H.-J. Thiel.** 1990. Pestivirus glycoprotein which induces neutralizing antibodies forms part of a disulfide-linked heterodimer. J. Virol. **64:**3563–3569.
71. **Weiner, A. J., M. J. Brauer, J. Rosenblatt, K. H. Richman, J. Tung, K. Crawford, F. Bonino, G. Saracco, Q.-L. Choo, M. Houghton, and J. H. Han.** 1991. Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. Virology **180:**842–848.
72. **Weiner, A. J., G. Kuo, D. W. Bradley, F. Bonino, G. Saracco, C. Lee, J. Rosenblatt, Q.-L. Choo, and M. Houghton.** 1991. Detection of hepatitis C viral sequences in non-A, non-B hepatitis. Lancet **335:**1–3.
73. **Wengler, G., and G. Wengler.** 1989. Cell-associated West Nile flavivirus is covered with E+pre-M protein heterodimers which are destroyed and reorganized by proteolytic cleavage during virus release. J. Virol. **63:**2521–2526.
74. **Wengler, G., and G. Wengler.** 1991. The carboxy-terminal part of the NS3 protein of the West Nile flavivirus can be isolated as a soluble protein after proteolytic cleavage and represents an RNA-stimulated NTPase. Virology **184:**707–715.
75. **Wiskerchen, M., and M. S. Collett.** 1991. Pestivirus gene expression: protein p80 of bovine viral diarrhea virus is a proteinase involved in polyprotein processing. Virology **184:**341–350.