JOURNAL OF VIROLOGY, May 1993, p. 2832–2843
0022-538X/93/052832-12$02.00/0

Vol. 67, No. 5

Characterization of the Hepatitis C Virus-Encoded Serine Proteinase: Determination of Proteinase-Dependent Polyprotein Cleavage Sites

ARASH GRAKOUI,[1] DAVID W. McCOURT,[2] CZESLAW WYCHOWSKI,[3]†
STEPHEN M. FEINSTONE,[3] AND CHARLES M. RICE[1]*

Department of Molecular Microbiology[1] and Howard Hughes Medical Institute,[2] Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, Missouri 63110-1093, and Division of Virology, Center of Evaluation and Biologics Research, Food and Drug Administration, Bethesda, Maryland 20892[3]

Received 16 December 1992/Accepted 15 February 1993

Processing of the hepatitis C virus (HCV) H strain polyprotein yields at least nine distinct cleavage products: NH_2-C-E1-E2-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH. As described in this report, site-directed mutagenesis and transient expression analyses were used to study the role of a putative serine proteinase domain, located in the N-terminal one-third of the NS3 protein, in proteolytic processing of HCV polyproteins. All four cleavages which occur C terminal to the proteinase domain (3/4A, 4A/4B, 4B/5A, and 5A/5B) were abolished by substitution of alanine for either of two predicted residues (His-1083 and Ser-1165) in the proteinase catalytic triad. However, such substitutions have no observable effect on cleavages in the structural region or at the 2/3 site. Deletion analyses suggest that the structural and NS2 regions of the polyprotein are not required for the HCV NS3 proteinase activity. NS3 proteinase-dependent cleavage sites were localized by N-terminal sequence analysis of NS4A, NS4B, NS5A, and NS5B. Sequence comparison of the residues flanking these cleavage sites for all sequenced HCV strains reveals conserved residues which may play a role in determining HCV NS3 proteinase substrate specificity. These features include an acidic residue (Asp or Glu) at the P6 position, a Cys or Thr residue at the P1 position, and a Ser or Ala residue at the P1′ position.

Hepatitis C virus (HCV) is the recently recognized agent of the parenterally transmitted form of non-A non-B hepatitis (17, 53). It causes the vast majority of transfusion-associated cases of hepatitis and a significant proportion of community-acquired hepatitis worldwide (3, 4, 53). Infection by HCV frequently (perhaps universally) leads to persistent infections that result in a range of clinical conditions (for review, see reference 42) from an inapparent carrier state to severe chronic active hepatitis, cirrhosis, and in some cases hepatocellular carcinoma (9, 22, 80). Identification of HCV rapidly led to the development of specific tests to identify potential HCV carriers among blood donors (53) and has already reduced the frequency of posttransfusion hepatitis (1). Although alpha interferon has been shown to be useful for the treatment of some patients with chronic HCV infections (24, 25), neither a prophylactic immunogen nor a highly effective therapeutic agent has been developed to control this important public health problem. It is hoped that a detailed understanding of HCV replication and pathogenesis will lead to effective means to treat or control infection; however, basic studies have been impeded by the fact that HCV does not replicate appreciably in cell culture systems (48, 82), and the only well-defined animal model is the chimpanzee (2, 43, 86).

At present, the genomes of several strains of HCV have been molecularly cloned and sequenced (15, 18, 23, 47, 49, 68, 69, 87, 88). Comparative analyses using this sequence information (64) combined with viral characterization in the chimpanzee infectivity model (8, 30, 40) indicate that the virus is a member of the flavivirus family, which now includes the classic flaviviruses such as yellow fever virus (YF) (12) and the animal pestiviruses such as bovine viral diarrhea virus (BVDV) (31). These viruses are small enveloped positive-stranded RNA viruses, with genomes ranging between 9.5 (HCV) and 12.5 (BVDV) kb in length (19). Viral genome RNAs contain a single long open reading frame (ORF) which encodes a polyprotein which is cleaved co- or posttranslationally into the mature viral proteins. The structural proteins are encoded by the 5′ one-fourth of the ORF, with the nonstructural proteins encoded in the remainder.

A rough map of the HCV polyprotein cleavage products has been established by cell-free translation (41) and cell culture expression (16, 36, 39, 50, 51, 60, 83) studies. For the HCV H strain (36), the gene order is 5′-C-E1-E2-NS2-NS3-NS4A-NS4B-NS5A-NS5B-3′. C, a basic protein, is believed to be the viral capsid protein; E1 and E2 are possible virion envelope glycoproteins; and NS2 through NS5B are putative nonstructural proteins, at least some of which may be the functional equivalents of homologous proteins encoded by flaviviruses and pestiviruses (reviewed in reference 44). The NS3 and NS5 proteins are presumed enzymatic components of the viral RNA replicase (12). The HCV NS3 protein contains a predicted serine proteinase domain in the N-terminal one-third (7, 64) and motifs characteristic of NTPases and helicases (35, 64) in the C-terminal portion which are shared by the homologous NS3 proteins of flaviviruses (11, 14, 27, 73–75, 92) and the homologous p80 protein of BVDV (94). Studies of the latter two proteins suggest that the serine proteinase domain mediates several cleavages in the respective viral polyproteins (11, 14, 27, 73–75, 92, 94). In the case of the flavivirus West Nile virus, the C-terminal fragment of NS3 has been partially purified and shown to have NTPase

* Corresponding author.
† Present address: Unite de Virologie Moleculaire, Institut Pasteur, 75724 Paris-Cedex 15, France.

activity (92). The Gly-Asp-Asp motif characteristic of RNA-dependent RNA polymerases (72) is found in the C-terminal product encoded in these ORFs which corresponds to the HCV NS5B protein, the NS5 protein of flaviviruses, or p75 of pestiviruses. However, direct evidence that these proteins function as polymerases is lacking. Although the flavivirus NS2B protein appears to be an essential component of the viral serine proteinase (11, 27, 75), the functions of the other hydrophobic proteins encoded by the NS2 and NS4 regions are undefined.

For the flaviviruses and pestiviruses it has been shown that mature viral proteins are produced by a combination of host and viral proteinases located both in the cytosol and in subcellular vesicular compartments (for reviews, see references 19 and 78). Cell-free translation studies indicate that cleavages generating the HCV structural proteins or precursors (C/E1, E1/E2, and perhaps E2/NS2) occur after hydrophobic stretches which are catalyzed in the endoplasmic reticulum lumen by host signalase peptidase (41). Although it has been proposed that HCV NS3 is a serine proteinase responsible for cleavages at specific sites in the HCV polyprotein, this function has not been demonstrated nor have the cleavage sites been defined. In this study, site-directed mutagenesis and deletion analysis were used to test the NS3 serine proteinase model and N-terminal sequencing was used to determine the HCV NS3 proteinase-dependent cleavage sites.

MATERIALS AND METHODS

Cell cultures and virus growth. The BHK-21 cell line, obtained from the American Type Culture Collection, and the A16 subclone of the human hepatoma HepG2 (American Type Culture Collection) cell line (generously provided by Alan Schwartz, Washington University, St. Louis, Mo.) were maintained 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 (33), and various vaccinia virus-HCV recombinants were grown in BSC-40 monolayers and partially purified (45), and titers of infectious progeny were determined by plaque assay on BSC-40 cells (45).

Expression constructs. Standard methods were used for recombinant DNA manipulations (81), and subclones containing regions produced by polymerase chain reaction amplification (79) were verified by nucleotide sequence analysis. The vaccinia virus-T7 expression plasmid, containing the entire HCV H strain ORF, pBRTM/HCV 1-3011, and, beginning at residue 827, pBRTM/HCV 827-3011, have been described (36). Derivatives which express HCV polyproteins beginning at residue 936 or 1007 were constructed by using synthetic oligonucleotides and the polymerase chain reaction (79, 90) to engineer 5′ *Nco*I sites which contained the methionine codon for translation initiation (CCATGG). For pBRTM/HCV 936-3011, this methionine residue corresponds to position 936 of the HCV polyprotein. For the polyprotein expressed by pBRTM/HCV 1007-3011, the engineered *Nco*I site encodes an additional N-terminal methionine which is followed by residue 1007. pBRTM/HCV 1193-3011 was constructed by ligating the *Nco*I-*Not*I (9219) and *Bst*XI (3910)-*Not*I (9219) fragments from pBRTM/HCV 827-3011 (numbers in parentheses refer to the positions of these sites in the full-length HCV H strain sequence [23]). Before ligation, the *Nco*I and *Bst*XI sites were treated with the Klenow fragment of DNA polymerase I and T4 DNA polymerase, respectively, in the presence of deoxynucleoside triphosphates. This results in a construct encoding a polyprotein with an extra N-terminal methionine followed by HCV residues 1193 to 3011.

Site-directed mutagenesis of residues in the putative NS3 serine proteinase domain was carried out with synthetic oligonucleotides and uridylated template DNA essentially as described (52). The phagemid construct pH3′2J/HCV 827-1207 was made by subcloning the *Nco*I-*Bam*HI fragment from pET8c/HCV N3.8 into pH3′2J1 (37), which had been digested with the same two restriction enzymes. Uridylated phagemid template DNA was rescued (55) and used to create alanine substitution mutations at positions 1078, 1083, 1164, and 1165 of the HCV polyprotein. The mutant designations and altered nucleotide sequences [shown in boldface type relative to the H strain sequence, with the alanine codon(s) underlined (23)] are as follows: C1078A, 5′-GGGGTC GCGTGG-3′; H1083A, 5′-CTACGCCGGCG-3′; S1164A, 5′-GGCGCCTCG-3′; S1165A, 5′-TCCGCGGGG-3′; S1164A and S1165A, 5′-GGCGCCGCCGGCG-3′. After sequence analysis of these clones to verify that only the desired alterations were present, multiple subcloning steps were used to construct the corresponding pBRTM/HCV 1-3011 derivatives. A triple mutant consisting of H1083A, S1164A, and S1165A (denoted HSS) was also constructed.

The derivation of vaccinia virus-HCV recombinants vHCV 1-1488 and vHCV 827-3011 has been described (36). vHCV 1-3011 HSS, a recombinant expressing the entire HCV polyprotein with alanine substitution mutations for some of the residues in the putative NS3 serine proteinase catalytic triad, was generated by marker rescue on CV-1 cells (56) and identified by the *gpt* selection method (28). The recombinant virus was 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 or vTF7-3 and the indicated vaccinia virus-HCV recombinants by using a multiplicity of infection of 5 PFU per cell (as determined on BSC-40 monolayers) (36). 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 Tran^{35}S-label (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 PFU per cell) for 30 min at 37°C. After removal of the inoculum, cells were transfected with a mixture consisting of 1 μg of plasmid DNA and 12.5 μg of Transfectam (Promega) in 0.5 ml of MEM. After 1.5 to 2.5 h at 37°C, the transfection mixture was replaced with 0.5 ml of MEM containing 1/40 the normal concentration of methionine, 2% FBS, and 50 μCi of Tran^{35}S-label (ICN) per ml; replacement was followed by incubation for 4 h at 37°C.

Cell lysis, immunoprecipitation, and protein analyses. After labeling, monolayers were lysed in sodium dodecyl sulfate (SDS) and cellular DNA was sheared by repeated passage through a 26-gauge needle (36). Immunoprecipitations using HCV region-specific or human sera and SDS-polyacrylamide gel electrophoresis were conducted essentially as previously

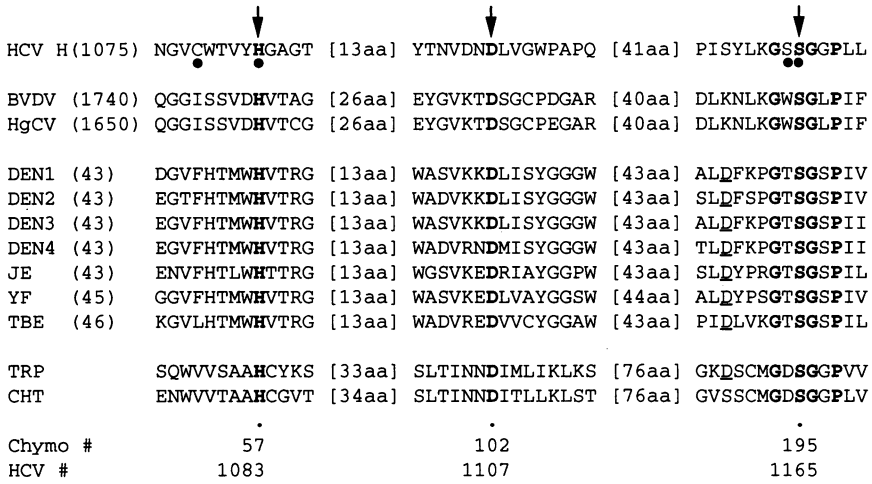

FIG. 1. A trypsin-like serine proteinase domain in flavivirus polyproteins. Aligned sequences of selected members of the flavivirus family in the regions surrounding the residues in the putative catalytic triad (indicated by arrows) of the serine proteinase-like domain. Shown are sequences for the HCV H strain (23); two pestiviruses, BVDV (20) and hog cholera virus (HgCV [61]); seven representative flaviviruses, dengue virus type 1 (DEN1 [32]), dengue virus type 2 (DEN2 [38]), dengue virus type 3 (DEN3 [70]), dengue virus type 4 (DEN4 [57]), Japanese encephalitis virus (JE [85]), YF (77), and tick-borne encephalitis virus (TBE [58]); and cellular enzymes trypsin (TRP) and chymotrypsin (CHT). The numbers in parentheses refer to the distance of the first residue shown from the N termini of the polyproteins (HCV and pestiviruses) or the NS3 protein (flaviviruses). Residues conserved among all viruses are shown in boldface type, and one of the predicted contact residues (6) in the substrate binding pocket is underlined for the flavivirus enzymes and trypsin. Positions of catalytic residues of chymotrypsin (Chymo; numbered according to the chymotrypsinogen scheme) and the putative HCV proteinase are given below the sequences. See references 6, 7, and 34 for further details and additional alignments with cellular proteinases. Residues in the HCV polyprotein which were mutagenized in this study are indicated by black dots. The single-letter code for amino acids is used: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

described (36). ^{14}C-methylated marker proteins were purchased from Amersham.

N-terminal sequence analysis. HepG2 A16 monolayers were coinfected, as described above, with vTF7-3 and vHCV 827-3011. At 2 h postinfection, infected monolayers were washed twice in media lacking the amino acid used for radiolabeling. For preparation of [^{3}H]leucine-labeled NS4A, NS4B, and NS5A, monolayers were labeled for 3 h in MEM containing 1/80 the normal concentration of leucine and 150 μCi of [^{3}H]leucine (147 Ci/mmol; Amersham) per ml. For preparation of NS5B, monolayers were labeled for 2 h in MEM lacking leucine and containing 200 μCi of [^{3}H]leucine (135 Ci/mmol; Amersham) per ml. For preparing [^{3}H]valine-labeled proteins, infected monolayers were labeled for 3 h in MEM lacking valine and containing 200 μCi of [^{3}H]valine (30 Ci/mmol; Amersham) per ml. For preparing [^{3}H]tryptophan-labeled proteins, infected monolayers were labeled for 3 h in MEM lacking tryptophan and containing 200 μCi of [^{3}H]tryptophan (31 Ci/mmol; Amersham) per ml. Metabolic labeling with [^{35}S]methionine was conducted with Tran^{35}S-label (ICN) under the conditions described above for the transient expression assays. After labeling, monolayers were lysed with SDS and the proteins were immunoprecipitated essentially as described (36). In most cases, material from ~10^7 cells was used for isolation of radiolabeled proteins for sequence analysis. NS4A, NS4B, and NS5A were immunoprecipitated with 20 μl of serum from patient JHF (36). [^{3}H]tryptophan-labeled NS5A was immunoprecipitated with 20 μl of polyclonal rabbit antiserum WU123 (36). Radiolabeled NS5B was immunoprecipitated with 20 μl of polyclonal rabbit antiserum WU115 (36).

Radiolabeled HCV-specific proteins were separated by SDS gel electrophoresis, transferred to Immobilon polyvinylidene difluoride membranes (59), and localized by autoradiography. Partial amino acid sequence analyses were performed essentially as described (13).

RESULTS

Evidence for NS3 proteinase-dependent cleavages. On the basis of the alignment of flavivirus (Fig. 1) polyproteins with known serine proteinases (7), HCV polyprotein residues His-1083, Asp-1107, and Ser-1165 in NS3 were predicted to comprise the catalytic triad of a serine proteinase domain. To identify cleavages which might be dependent on this putative HCV serine proteinase, a triple mutant in which His-1083, Ser-1164, and Ser-1165 were substituted with alanine residues was constructed. The change at Ser-1164 was included because of the possibility that the serine residues at either position 1164 or 1165 might be capable of functioning as the nucleophile of the catalytic triad. On the basis of the serine proteinase model, these changes are expected to inactivate the NS3 proteinase and thus provide a means of identifying proteinase-dependent cleavages.

The effect of these mutations on HCV polyprotein processing was examined in transient assays with the vaccinia virus-T7 hybrid expression system (26, 33, 65) and a panel of HCV region-specific antisera was used to identify the polyproteins and cleavage products (36). For this analysis, the mutations were incorporated into an expression plasmid encoding the entire HCV polyprotein (pBRTM/HCV 1-3011 HSS), which was then used to rescue the corresponding vaccinia virus recombinant, vHCV 1-3011 HSS. Figure 2 compares the patterns of HCV-specific cleavage products in BHK-21 cells infected with vHCV 1-3011 HSS with those produced in cells infected with either vHCV 1-1488 or vHCV 827-3011 (36). Previous studies demonstrated that vHCV 1-1488 produces C (21 kDa), E1 (31 kDa), E2 (70 kDa), NS2

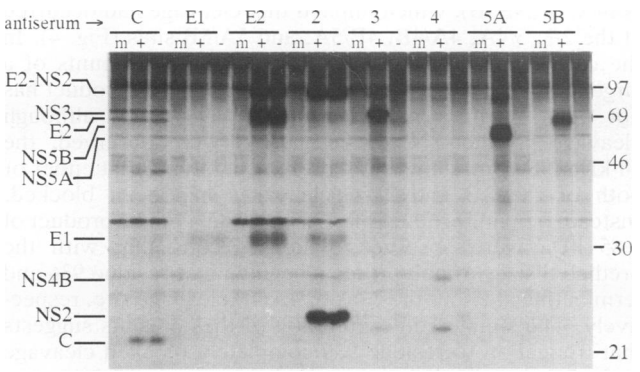

FIG. 2. HCV NS3 proteinase-dependent cleavages. BHK-21 cell monolayers were infected with vTF7-3 alone (m) or coinfected with vTF7-3 and either vHCV 1-1488 (+), vHCV 827-3011 (+), or vHCV 1-3011 HSS (−) and labeled with Tran^{35}S-label as described in Materials and Methods. Cell lysates were prepared and immunoprecipitated with the indicated HCV region-specific antisera (36). Lysates from cells coinfected with vHCV 1-1488 were used for immunoprecipitations with region-specific rabbit antiserum WU120 (C specific), WU122 (E1 specific), WU105 (E2 specific), or WU107 (NS2 specific) (36). Lysates from cells coinfected with vHCV 827-3011 were used for immunoprecipitations with region-specific rabbit antiserum WU110 (NS3 specific), WU111 (NS4 specific), WU123 (NS5A specific), or WU115 (NS5B specific) (36). Samples were separated on an SDS–14% polyacrylamide gel. HCV-specific proteins are identified at the left, with the sizes of protein molecular mass markers indicated at the right. As previously shown (36), some E1 is associated with E2 and E2-NS2 via disulfide linkages and is immunoprecipitated with these antisera. The 22-kDa HCV-specific product which is present in variable amounts in the immunoprecipitates with NS3-, NS4-, and NS5A-specific sera is an N-terminally truncated form of the NS2 protein produced by vHCV 827-3011, which often precipitates nonspecifically.

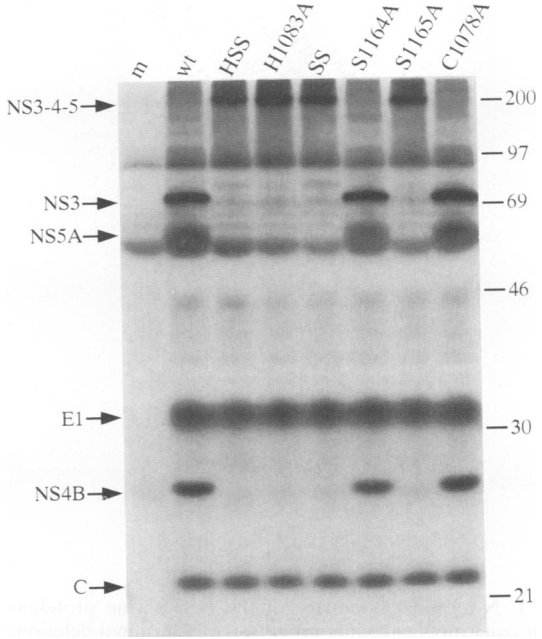

FIG. 3. Effects of substitutions at Cys-1078, His-1083, Ser-1164, and Ser-1165 on processing. BHK-21 cells previously infected with vTF7-3 were transfected with plasmid DNA pTM3 (m) or pBRTM/HCV 1-3011 (wt) or plasmid DNA pBRTM/HCV1-3011 derivatives with alanine substitutions for the following residues: His-1083, Ser-1164, and Ser-1165 (HSS); His-1083 (H1083A); Ser-1164 and Ser-1165 (SS); Ser-1164 (S1164A); Ser-1165 (S1165A); Cys-1078 (C1078A). As described in Materials and Methods, transfected monolayers were labeled with Tran^{35}S-label, cell lysates were prepared, and HCV-specific antigens were immunoprecipitated with human serum from patient F (36). Immunoprecipitated proteins were separated by electrophoresis on an SDS–14% polyacrylamide gel. HCV-specific proteins are identified at the left, with the sizes of protein molecular mass markers indicated at the right.

(23 kDa), and a truncated form of the NS3 protein (36). In addition, a glycosylated E2-NS2 polyprotein of 82 to 88 kDa is observed. vHCV 827-3011 expresses a truncated form of the NS2 protein, NS3 (70 kDa), NS4A (8 kDa), NS4B (27 kDa), NS5A (58 kDa), and NS5B (68 kDa). The mature cleavage products produced by these two overlapping recombinants are therefore identical to those produced by expression of the entire polyprotein (except for the truncated NS3 and NS2 proteins) (36). 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 vHCV 1-3011 or transfected with pBRTM/HCV 1-3011) (36). As shown in Fig. 2, the HCV-specific products immunoprecipitated with antisera specific for HCV C, E1, E2, and NS2 were the same for lysates from vHCV 1-3011 HSS- or vHCV 1-1488-infected cells, suggesting that the NS3 serine proteinase is not required for any of these cleavages. In contrast, instead of the mature cleavage products from the NS3-4-5 region observed in cells infected with vHCV 827-3011, a large ~200-kDa polyprotein reacting with antiserum specific for NS3, NS4, NS5A, or NS5B was observed in the triple mutant (this product is difficult to see in Fig. 2, but see below). The absence of mature products or polyprotein intermediates suggests that inactivation of the NS3 serine proteinase domain blocked cleavage at the 3/4A, 4A/4B, 4B/5A, and 5A/5B sites.

Further support for the NS3 serine proteinase model via site-directed mutagenesis. Given that the polyprotein expressed by vHCV 1-3011 HSS contained three substitution mutations, processing defects due to mutation-induced changes in polyprotein structure, rather than inactivation of the proteinase, cannot be ruled out. To further address this concern and to examine the importance of individual residues for NS3 proteinase activity, the patterns of processed products from a series of mutant polyproteins were examined. pBRTM/HCV 1-3011, a T7 expression plasmid encoding the entire HCV polyprotein, or various mutant derivatives were used to transfect BHK-21 cells which had been previously infected with vTF7-3, a vaccinia virus recombinant which expresses the bacteriophage T7 RNA polymerase (33). Radiolabeled cell lysates were prepared and immunoprecipitated with serum obtained from a patient chronically infected with HCV (patient F). This patient serum immunoprecipitates C, E1, NS3, NS4B, and NS5A (Fig. 3 and unpublished observations). As shown in Fig. 3, results similar to those described above were obtained for the parental construct and the triple mutant pBRTM/HCV 1-3011 HSS. Substitution of His-1083 or Ser-1165 with alanine completely blocked production of NS3, NS4B, and NS5A and, like pBRTM/HCV 1-3011 HSS, resulted in the production of a polyprotein whose size is consistent with its identification as NS3-4A-4B-5A-5B (predicted molecular mass, 215 kDa). This uncleaved polyprotein was also found for a double mutant which contained alanine substitutions

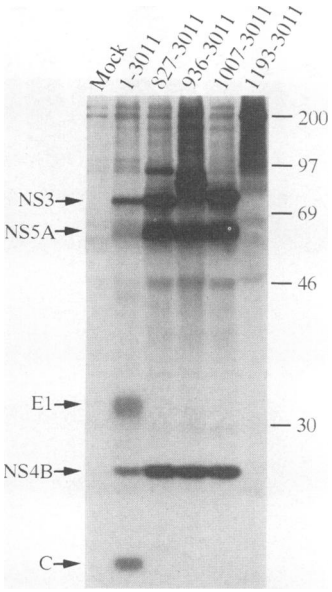

FIG. 4. N-terminal boundary of the NS3 serine proteinase. A series of constructs encoding progressive N-terminal deletions was used to map the N-terminal boundary of the NS3 serine proteinase. BHK-21 cells previously infected with vTF7-3 were transfected with plasmid DNA pBRTM/HCV 1-3011 (1-3011), pBRTM/HCV 827-3011 (827-3011), pBRTM/HCV 936-3011 (936-3011), pBRTM/HCV 1007-3011 (1007-3011), or pBRTM/HCV 1193-3011 (1193-3011) or mock transfected with pTM3 (Mock) and labeled with Tran^{35}S-label as described in Materials and Methods. Cell lysates were prepared, and HCV-specific antigens were immunoprecipitated with human serum from patient F (36). Immunoprecipitated proteins were separated by electrophoresis on an SDS–14% polyacrylamide gel. HCV-specific proteins are identified at the left, with the sizes of protein molecular mass markers indicated at the right.

for both Ser-1164 and Ser-1165 (Fig. 3, lane SS), which was included to address the possibility that either of these serine residues might be capable of functioning as the active site nucleophile. Substitution of alanine for residues not predicted to be important for catalysis, Cys-1078 and Ser-1164, had no effect on processing. These results strengthen our previous suggestion that the N-terminal domain of NS3 encodes a proteinase necessary for processing in the NS3-4-5 region and are consistent with the serine proteinase model in which His-1083 and Ser-1165 are members of the catalytic triad (7, 12).

NS2 region sequences are not required for NS3 serine proteinase-dependent cleavages. In the case of several flaviviruses, NS2B expressed either in *cis* or in *trans* is required for NS3 proteinase activity (11, 27, 75). To map the N-terminal boundary of the HCV proteinase necessary for processing at downstream sites, a series of plasmids encoding polyproteins with N-terminal truncations was produced. In HCV, the N terminus of the 23-kDa NS2 protein is in the vicinity of residues 805 to 815, with the 2/3 cleavage occurring near residues 1020 to 1030 (36). The parental construct, pBRTM/HCV 1-3011, or derivatives expressing truncated polyproteins 827-3011, 936-3011, 1007-3011, and 1193-3011 were used to transfect vTF7-3-infected BHK-21 cells. Radiolabeled lysates were prepared and immunoprecipitated with patient F serum. As described above, the parental 1-3011 and 827-3011 polyproteins were processed to yield NS3, NS4B, and NS5A (this antiserum does not react with NS2, NS4A, or NS5B), which implied that cleavage had occurred at the 2/3, 3/4A, 4A/4B, 4B/5A, and 5A/5B sites (Fig. 4). In the case of the 827-3011 polyprotein, minor amounts of a larger ~90-kDa species were observed, but this product has not been further characterized. For 936-3011, although cleavage products NS4B and NS5A were observed, the 70-kDa NS3 protein was not present, implying that either or both of the 2/3 and 3/4A cleavages had been blocked. Instead, a major product of ~79 kDa and a minor product of ~85 kDa were observed, which is consistent with the predicted sizes of polyproteins beginning at residue 936 and terminating at either the 3/4A or 4A/4B cleavage site, respectively. The simplest interpretation of these results suggests that truncation to residue 936 completely blocked cleavage at the 2/3 site and slightly impaired cleavage at the 3/4A site. Further truncation to residue 1007 still allowed efficient cleavage at the downstream sites, but cleavage at the 2/3 site appeared to be blocked since a 72-kDa protein but not the 70-kDa NS3 protein was observed. As expected, truncation to residue 1193, which deletes the majority of the predicted serine proteinase domain, completely abolished the production of NS4B and NS5A and led to the accumulation of a predominant ~200-kDa product and a smear of lower-molecular-mass species. These results indicate that the majority of the NS2 region is dispensable for HCV NS3 serine proteinase activity and that the N-terminal boundary of the proteinase is C terminal to residue 1007.

N-terminal sequence analysis of NS4A, NS4B, NS5A, and NS5B. To further refine the map of the HCV nonstructural proteins and to determine the sequence at the four potential NS3 proteinase cleavage sites, the N-terminal sequences of NS4A, NS4B, NS5A, and NS5B were determined. HCV-specific proteins metabolically labeled with either [^{3}H]leucine, [^{3}H]valine, [^{35}S]methionine, or [^{3}H]tryptophan were immunoprecipitated with region-specific or human patient antisera, separated by electrophoresis on SDS-polyacrylamide gels, and electroblotted onto Immobilon P membranes (see Materials and Methods for details). The appropriate bands were localized and sequenced for 10 to 30 cycles of Edman degradation. Some of these results are shown in Fig. 5. For each protein, the deduced sequence was sufficient to unambiguously localize its N terminus in the HCV polyprotein. For NS4A, leucine residues were recovered at positions 5, 10, 13, and 18 and valine residues were recovered at positions 4, 6, and 9. These data establish the N terminus of NS4A as residue 1658. For NS4B, leucine residues were recovered at positions 4 and 13 and methionine residues were recovered at positions 11 and 12. These data establish the N terminus of NS4B as residue 1712. Similar data for NS5A and NS5B indicate that these proteins begin at residues 1973 and 2421, respectively. The locations of these cleavage sites, as determined by N-terminal sequence analyses, agree remarkably well with our previous map, which was determined by expressing a series of polyproteins with C-terminal truncations (36). Assuming that C-terminal trimming or additional cleavages do not occur in this region, NS4A is 54 amino acid residues in length (predicted molecular mass, 5,781 kDa), NS4B is 261 residues (predicted molecular mass, 27,181 kDa), NS5A is 448 residues (predicted molecular mass, 49,074 kDa), and NS5B is 591 residues (predicted molecular mass, 65,384 kDa). With the exception of NS5A (and perhaps NS4A), the predicted sizes of these proteins agree reasonably well with their apparent molecular masses as estimated by SDS-polyacrylamide gel electrophoresis. As noted previously (36), NS5A migrates as a heterogeneous collection of species on SDS-polyacrylamide

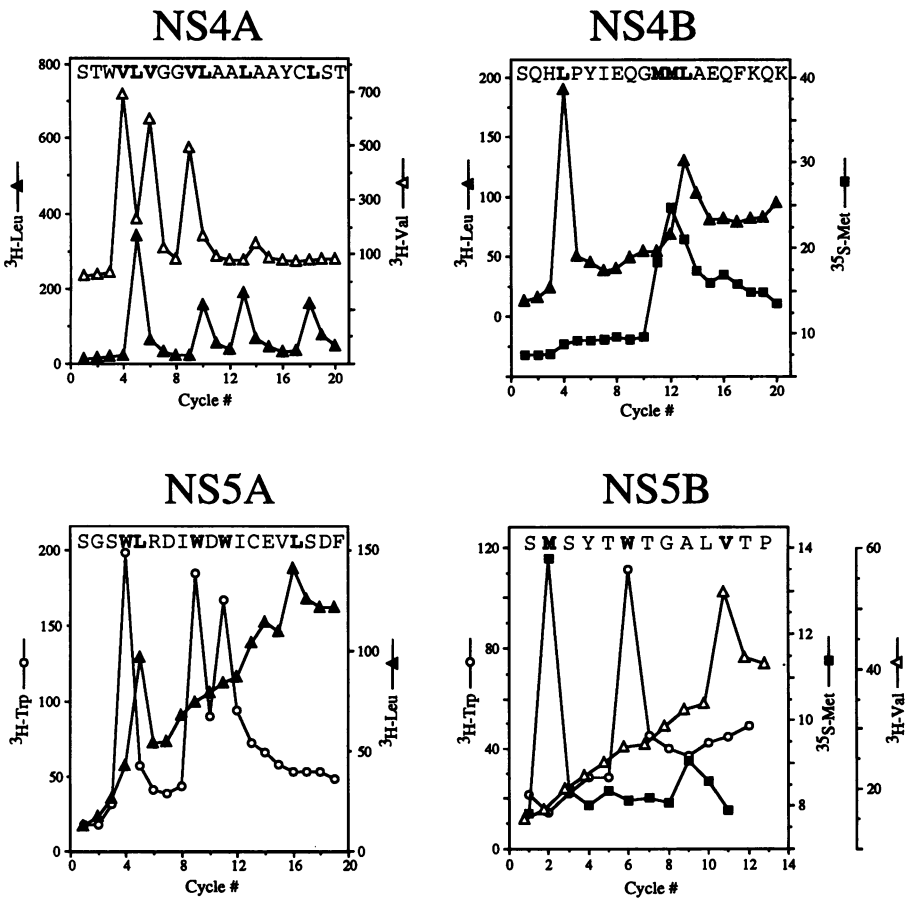

FIG. 5. N-terminal sequence analysis of HCV NS4A, NS4B, NS5A, and NS5B proteins. The graphs show uncorrected counts per minute released per sequencing cycle. From these data, the sequences for NS4A, NS4B, NS5A, and NS5B begin at HCV polyprotein residues 1658, 1712, 1973, and 2421, respectively, and are shown at the top of each graph. In these sequences, the amino acid residues shown in boldface type indicate the residues determined by N-terminal sequence analysis of radiolabeled HCV-specific proteins; the remaining residues are deduced from HCV H strain nucleotide sequence data (23, 47). See the legend to Fig. 1 for the single-letter code.

gels, with a predominant band at 58 kDa and minor species ranging up to 68 kDa. It is possible that NS5A exhibits aberrant migration under these conditions. However, the 9-kDa discrepancy between the observed size of the predominant form and the size predicted from sequence data (58 versus 49 kDa) could also reflect an as yet undefined posttranslational modification(s) or alternative proteolytic processing.

Common features of potential HCV NS3 proteinase cleavage sites. The N-terminal sequence data obtained in this study have defined potential cleavage sites for the HCV NS3 proteinase in the NS3-4-5 region. Amino acid sequences flanking these cleavage sites were aligned for a number of HCV isolates and are shown in Fig. 6. Given the high degree of amino acid sequence homology observed among HCV isolates, conserved features at a given cleavage site, possibly important for NS3 proteinase specificity, are difficult to discern. However, when the sequences flanking all four potential cleavage sites are aligned, several common features emerge. A Ser residue is usually found at the P1′ position except for in the Japanese and Taiwanese isolates in which Ala is found at the 4A/4B site. A Cys residue is invariant at the P1 position of the 4A/4B, 4B/5A, and 5A/5B sites, but Thr is found at this position of the 3/4A site. Finally, all four cleavage sites contain one or more upstream acidic residues, with an Asp (3/4A, 4A/4B, and 4B/5A) or Glu (5A/5B) residue always found at the P6 position. These common features suggest that the same enzyme, presumably the HCV NS3 proteinase, could be directly responsible for cleavage at these sites. If this is the case, a possible cleavage site recognition motif for the HCV NS3 proteinase would appear to be Asp/Glu-X-X-X-Cys/Thr ↓ Ser/Ala (where X is variable). Examination of the HCV H strain polyprotein for additional sequences conforming to this motif has identified five additional sites: Asp-2167 Thr-2172 Ser-2173, Asp-2177 Thr-2182 Ala-2183, Glu-2337 Thr-2342 Ala-2343, Asp-2779 Thr-2784 Ser-2785, and Glu-2781 Cys-2786 Ser-2787. Thus far, no products consistent with cleavage at these alternative sites have been observed, suggesting that additional factors, such as polyprotein conformation, may play a role in determining cleavage site preferences of the HCV NS3 proteinase.

DISCUSSION

The experiments described in this report provide evidence that the predicted serine proteinase domain located in the N-terminal one-third of the HCV NS3 protein is required for efficient cleavage at four sites in the HCV polyprotein. These serine proteinase-dependent cleavages are all C terminal to

H-FDA	CMSADLEVVT STWVLVGGVL	3/4A (1658)
H-AP	CMSADLEVVT STWVLVGGVL	
HCV-1	CMSADLEVVT STWVLVGGVL	
HCV-J	CMSADLEVVT STWVLVGGVL	
HCV-BK	CMSADLEVVT STWVLVGGVL	
HC-J6	CMQADLEVMT STWVLAGGVL	
HCV-T	CMSADLEVVT STWVLVGGVL	
HC-J8	CMQADLEIMT SSWVLAGGVL	
HCV-JT,JT'	CMSADLEVVT STWVLVGGVL	
H-FDA	YQEFDEMEEC SQHLPYIEQG	4A/4B (1712)
H-AP	YQEFDEMEEC SQHLPYIEQG	
HCV-1	YREFDEMEEC SQHLPYIEQG	
HCV-J	YQEFDEMEEC ASHLPYIEQG	
HCV-BK	YQEFDEMEEC ASHLPYIEQG	
HC-J1,4	YEAFDEMEEC ASRAALIEEG	
HCV-T	YQEFDEMEEC ASHLPYIEQG	
HC-J8	YQAFDEMEEC ASKAALIEEG	
HCV-JT,JT'	YREFDEMEEC ASHLPYIEQG	
H-FDA	WISSECTTPC SGSWLRDIWD	4B/5A (1973)
H-AP	WISSECTTPC SGSWLRDIWD	
HCV-1	WISSECTTPC SGSWLRDIWD	
HCV-J	WINEDCSTPC SGSWLKDVWD	
HCV-BK	WINEDCSTPC SGSWLRDVWD	
HC-J6	WITEDCPIPC SGSWLRDVWD	
HCV-T	WINEDCSTPC SGSWLRDVWD	
HC-J8	WITEDCPVPC SGSWLQDIWD	
HCV-JT	WINEDCSTPC SGSWLKDVWD	
HCV-JT'	WINEDCSTPC SGSWLRDVWD	
H-FDA	GADTEDVVCC SMSYTWTGAL	5A/5B (2421)
H-AP	GADTEDVVCC SMSYSWTGAL	
HCV-1	EANAEDVVCC SMSYSWTGAL	
HCV-J	GEAGEDVVCC SMSYTWTGAL	
HCV-BK	EEASEDVVCC SMSYTWTGAL	
HC-J6	SEEDDSVVCC SMSYSWTGAL	
HCV-T	EEDGEGVICC SMSYTWTGAL	
HC-J8	SDQEDSVICC SMSYSWTGAL	
HCV-JT,JT'	GEASDDIVCC SMSYTWTGAL	
CONSENSUS	D C S E T A	

FIG. 6. HCV NS3 serine proteinase-dependent cleavage sites. As determined from the N-terminal sequence data shown in Fig. 5, the 10 amino acid residues surrounding the 3/4A, 4A/4B, 4B/5A, and 5A/5B cleavage sites are aligned for the H strain as well as other published HCV sequences. The P1 and P1′ positions are shown in boldface type, and the conserved acidic residue at the P6 position is underlined. The single-letter code for amino acids is used (see the legend to Fig. 1). The HCV sequences are from the following sources: H-FDA (23), H-AP (47), HCV-1 (18), HCV-J (49), HCV-BK (87), HC-J6 (69), HCV-T (15), HC-J8 (68), and HCV-JT and HCV-JT′ (88).

this domain and produce the N termini of nonstructural proteins NS4A, NS4B, NS5A, and NS5B. Alignment of the sequences flanking these four cleavage sites has revealed conserved features, which suggests that the HCV serine proteinase may be directly responsible for proteolytic processing at all of these sites, although studies using purified proteinase and defined substrates will be necessary to directly test this hypothesis.

The polyprotein processing scheme which is emerging for HCV is remarkably similar to those of the pestiviruses and flaviviruses (Fig. 7) and further supports their classification as three genera in the flavivirus family (31). For both HCV and BVDV, the NS3 and p80 (94) serine proteinases appear to be responsible for four downstream cleavages which generate proteins which are similar in size and hydropathic profiles and contain the same conserved motifs (64). In the case of flaviviruses such as YF, previous studies have shown that the 3/4A and 4B/5 cleavages are also mediated by the NS3 proteinase but that the cleavage generating the N terminus of NS4B is more likely to be mediated by host signal peptidase (10, 13, 74, 84). Recent studies, however, have shown that an additional cleavage site for the YF NS3 proteinase exists just upstream from the 4A/4B signalase site and that cleavage at this site is required for subsequent processing at the downstream signalase site (54). Thus, the involvement of the viral serine proteinases in processing in the NS3-4-5 region is quite similar for all three groups, with the exception being the NS5 region, which is not further processed for flaviviruses.

N terminal to the serine proteinase domain, the processing strategies of the three groups appear to diverge. From the results reported here, the HCV NS3 serine proteinase is not required for cleavage at the 2/3 site and the responsible proteinase, whether host or virus encoded, has not yet been elucidated. The situation for the pestiviruses is complex and varies with different isolates. The homologous cleavage does not occur in noncytopathic strains of BVDV, resulting in accumulation of p125 (Fig. 7). For the cytopathic NADL strain of BVDV, partial processing at this site leads to the production of three species, uncleaved p125 and two cleavage products, p54 and p80 (Fig. 7). While transient expression assays of truncated BVDV polyproteins in mammalian cells indicate that this cleavage is dependent on the p80 proteinase (94), the situation is somewhat unclear since cleavage at this site was not observed when the baculovirus system was used to express BVDV nonstructural proteins in *Spodoptera frugiperda* Sf9 cells (71). Other cytopathic BVDV isolates appear to have evolved different strategies for processing at this site, including recombination with host sequences to provide cleavage sites for host enzymes (such as ubiquitin isopeptidase [62]) or incorporation of a second copy of the p20 proteinase (93) which can then produce the N terminus of p80, presumably via autocatalytic cleavage (63). In contrast, numerous studies have shown that the flavivirus NS3 proteinase is responsible for cleavage at both the 2A/2B and 2B/3 polyprotein cleavage sites (14, 75, 91) and at an additional site in the YF NS2A region (66).

In the case of the flaviviruses, but not the pestiviruses (93, 94), cleavage at an additional site in the structural region is believed to be catalyzed by the NS3 proteinase (12). For the flavivirus West Nile virus, analysis of cell-free translation products made in the presence of microsomal membranes indicates that a precursor of the capsid protein with a C-terminal hydrophobic segment (called anchored C) is produced by host signal peptidase cleavage (67). This C-terminal segment is not present on the C protein isolated from mature virions (67), and recent studies using YF have shown that this cleavage can be catalyzed, at least in vitro, by the NS2B-3 serine proteinase complex (5). For HCV, cell-free translation studies also indicate that host signal peptidase mediates the initial cleavage separating C and E1 (41), with the prediction that this form of the C protein would contain a hydrophobic C-terminal segment. The observation that mutations inactivating the HCV NS3 serine proteinase made no difference in the apparent migration of the putative 21-kDa C protein suggests that a proteinase-dependent cleavage similar to the anchored C cleavage of flaviviruses does not occur for HCV. However, it remains possible that additional host factors or viral sequences are required or that such a cleavage does occur in the vaccinia virus transient expression system but does not affect the migration of the protein on SDS-polyacrylamide gels.

The site-specific mutations which abolish HCV NS3 proteinase activity support previous sequence alignments with members of the trypsin superfamily (7, 12), which predicted that His-1083 and Ser-1165 would compose two of the three

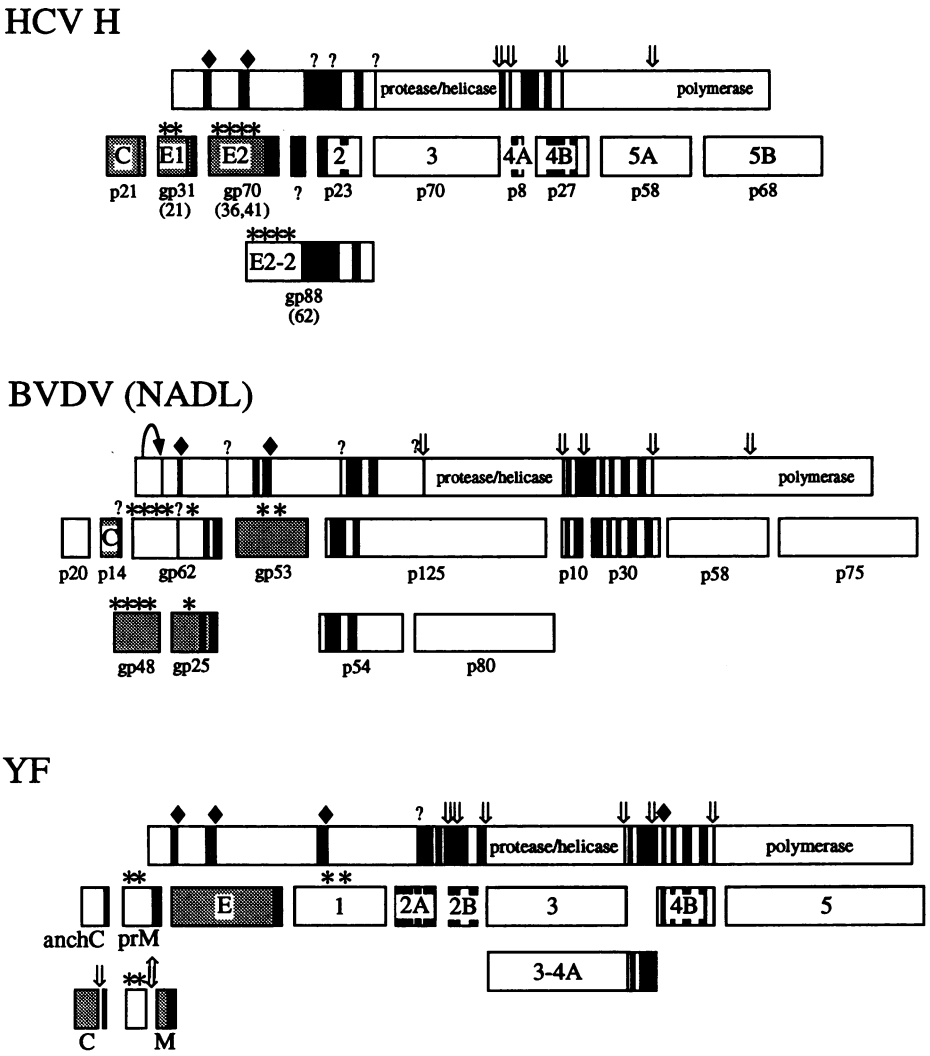

FIG. 7. Polyprotein processing schemes of HCV, BVDV, and YF. Diagrams of the HCV H strain (36), BVDV NADL strain (19, 21, 63, 94), and the YF 17D strain (5, 11, 12, 54, 78) polyproteins and cleavage products are shown as boxes. Regions of the polyproteins containing predominantly uncharged amino acids are indicated as black bars. Putative cleavage sites for host signalase (◆), viral NS3 or p80 serine proteinase (⇓), putative Golgi proteinases (⇳), or unknown proteinases (?) are indicated. The proposed autocatalytic cleavage releasing the N-terminal BVDV p20 protein from the polyprotein (93) is indicated by the curved arrow. The observed sizes for HCV (36) and BVDV (63) proteins (p) and glycoproteins (gp) are indicated. For the HCV glycoproteins (E1, E2, and precursor E2-NS2), the sizes of the endoglycosidase F-resistant forms are given in parentheses. Demonstrated pestivirus (89) and flavivirus (reviewed in reference 12) structural proteins and putative HCV structural proteins (41, 44) are shown as shaded boxes. 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.

residues in the catalytic triad. Mutagenesis studies with similar results have also been reported for the pestivirus p80 proteinase (94) and the flavivirus NS3 proteinases (14, 91). The linear alignments of these residues in these viral polyproteins are remarkably similar, especially for HCV and the flaviviruses, but the pestiviruses contain 13 additional amino acids in between the putative active site His and Asp residues (Fig. 1). For the flaviviruses, NS2B and NS3 are both required for proteinase activity (11, 27, 75). The NS2B sequence can be supplied either as part of the polyprotein or in *trans*, and the current model suggests that NS2B interacts with NS3 to form the functional flavivirus proteinase complex (78). In contrast, the BVDV p80 proteinase appears to be capable of functioning with only limited upstream sequences (~7 kDa) (94). This is similar to the results obtained here for HCV, in which deleting the majority of the NS2 region sequences did not abolish processing at the 3/4A, 4A/4B, 4B/5A, and 5A/5B sites.

Finally, the definition of potential cleavage sites for the HCV NS3 proteinase allows a comparison between the apparent specificities of the HCV and flavivirus NS3 proteinases. Similar data are not yet available for the pestiviruses. As discussed above, the NS2B-3 proteinases of members of the flavivirus genus cleave their respective viral polyproteins at probably six or seven specific sites (Fig. 7). At these cleavage sites, only the residues at positions P2, P1, and P1′ are highly conserved among all flaviviruses (12, 78). Basic residues (Arg or Lys) are usually found at the P2 and P1 positions, with the exception being Gln at the P2 position of the 2B/3 site of the four dengue virus serotypes. The

residue at the P1′ position of the scissile bond is usually Gly or Ser, except for Ala (dengue virus type 2) or Thr (Langat virus) at the 2B/3 site (46). In the case of YF, the importance of these conserved residues for recognition and cleavage by the NS2B-3 proteinase has recently been verified via site-directed mutagenesis (55, 66). As suggested from our studies, if the HCV NS3 proteinase does indeed cleave at the 3/4A, 4A/4B, 4B/5A, and 5A/5B sites, then the cleavage specificity of this enzyme is distinct from that of the flavivirus enzymes. Similar to the flavivirus enzymes, residues with small side chains, Ser or Ala, seem to be preferred at the P1′ position. However, polar residues (Cys or Thr) are found at the P1 position, and acidic residues (Glu or Asp) are found at the P6 position. Although the importance of these residues, and possibly other residues flanking the cleavage sites, in determining cleavage efficiency remains to be tested, these observations suggest that residues in the proteinase substrate binding pocket which determine substrate specificity may differ between these two genera. The proposed framework for these viral enzymes, based on structural and biochemical studies with members of the trypsin superfamily, has already allowed predictions about critical residues in the substrate binding pocket important for cleavage site specificity (6, 7). Although not experimentally tested, Asp-189 (chymotrypsinogen numbering; underlined in Fig. 1), which is conserved among the flavivirus NS3 enzymes and trypsin, is believed to be important for their preference for cleavage after basic residues (6). A Ser residue is found at this position of the HCV NS3 proteinase and a Lys residue is found at this position of the pestivirus p80 enzymes (Fig. 1). Given these as well as other differences in proposed contact residues in the substrate binding pockets (6), it is not surprising that the cleavage site preferences of the HCV and flavivirus NS3 proteinases appear to be different. Further, it is expected that the pestiviruses will exhibit yet another cleavage site specificity distinct from those of the other two genera.

Besides improving diagnostic assays for detecting HCV infection, other important goals in HCV research include the development of effective vaccines to prevent virus infection and therapeutic agents to eliminate or control virus replication in infected individuals. The prospects for vaccination seem distant given recent studies with chimpanzees which indicate that infection with HCV does not prevent reinfection with heterologous or even homologous HCV strains (29, 76). In terms of developing inhibitors of HCV replication, the N-terminal sequences determined in this study and the preliminary characterization of the HCV NS3 proteinase and its possible cleavage sites could be important for future work aimed at developing effective HCV therapy. Future expression studies, utilizing individual proteins or different combinations of HCV cleavage products, may be useful for characterizing essential viral enzymes, such as the HCV-encoded RNA-dependent RNA polymerase, and for determining the possible effects of viral proteins on host cell biology and immune responses. Regarding the NS3 serine proteinase, studies with YF have shown that mutations which abolish NS3 proteinase activity (14) or block cleavage at specific sites (66) are lethal for virus replication, suggesting that the HCV NS3 proteinase may be an attractive target for development of antiviral agents. Further studies are now needed to characterize the structure and activity of this proteinase and to develop assays which can be used to identify inhibitors and assess their efficacy at blocking HCV replication.

ACKNOWLEDGMENTS

We thank Dorota Paluszynska for expert technical assistance, Henry Hsu and Harry Greenberg for human sera, and Alan Schwartz for the HepG2 A16 cell line. We are also grateful to many colleagues for helpful discussions during the course of this work and to Jean Dubuisson, Julie Lemm, and Chao Lin for critical reading of the manuscript.

This work was supported by grants from the Pew Memorial Trusts and the Public Health Service (CA57973).

REFERENCES

1. **Alter, H. J., B. W. Jett, A. J. Polito, P. Farci, J. C. Melpolder, J. W. Shih, Y. Shimizu, and R. H. Purcell.** 1991. Analysis of the role of hepatitis C virus in transfusion-associated hepatitis, p. 396–402. *In* F. B. Hollinger, S. M. Lemon, and H. S. Margolis (ed.), Viral hepatitis and liver disease. The Williams & Wilkins Co., Baltimore.
2. **Alter, H. J., R. H. Purcell, P. V. Holland, and H. Popper.** 1978. Transmissible agent in non-A, non-B hepatitis. Lancet **i**:459–463.
3. **Alter, H. J., R. H. Purcell, J. W. Shih, J. C. Melpolder, M. Houghton, Q.-L. Choo, and G. Kuo.** 1989. Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A non-B hepatitis. N. Engl. J. Med. **321**:1494–1500.
4. **Alter, M. J., S. C. Hadler, F. N. Judson, A. Mares, W. J. Alexander, P. Y. Hu, J. K. Miller, L. A. Moyer, H. A. Fields, D. W. Bradley, and H. S. Margolis.** 1990. Risk factors for acute non-A, non-B hepatitis in the United States and association with hepatitis C virus infection. JAMA **264**:2231–2235.
5. **Amberg, S. M., and C. M. Rice.** Unpublished data.
6. **Bazan, J. F., and R. J. Fletterick.** 1989. Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses. Virology **171**:637–639.
7. **Bazan, J. F., and R. J. Fletterick.** 1990. Structural and catalytic models of trypsin-like viral proteases. Semin. Virol. **1**:311–322.
8. **Bradley, D. W., J. E. Maynard, H. Popper, E. H. Cook, J. W. Ebert, K. A. McCaustland, C. A. Schable, and H. A. Fields.** 1983. Posttransfusion non-A, non-B hepatitis: physicochemical properties of two distinct agents. J. Infect. Dis. **148**:254–265.
9. **Bruix, J., X. Calvet, J. Costa, M. Ventura, M. Bruguera, R. Castillo, J. M. Barrera, G. Ercilla, J. M. Sanchez-Tapias, M. Vall, C. Bru, and J. Rodes.** 1989. Prevalence of antibodies to hepatitis C virus in Spanish patients with hepatocellular carcinoma and hepatic cirrhosis. Lancet **ii**:1004–1006.
10. **Cahour, A., B. Falgout, and C.-J. Lai.** 1992. Cleavage of the dengue virus polyprotein at the NS3/NS4A and NS4B/NS5 junctions is mediated by viral protease NS2B-NS3, whereas NS4A/NS4B may be processed by a cellular protease. J. Virol. **66**:1535–1542.
11. **Chambers, T. J., A. Grakoui, and C. M. Rice.** 1991. Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. J. Virol. **65**:6042–6050.
12. **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.
13. **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.
14. **Chambers, T. J., R. C. Weir, A. Grakoui, D. W. McCourt, J. F. Bazan, R. J. Fletterick, and C. M. Rice.** 1990. Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. Proc. Natl. Acad. Sci. USA **87**:8898–8902.
15. **Chen, P.-J., M.-H. Lin, K.-F. Tai, P.-C. Liu, C.-J. Lin, and D.-S. Chen.** 1992. The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5′ termini of viral genomic and antigenomic RNA. Virology **188**:102–113.
16. **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.

17. **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.
18. **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.
19. **Collett, M. S.** 1992. Molecular genetics of pestiviruses. Comp. Immunol. Microbiol. Infect. Dis. **15:**145–154.
20. **Collett, M. S., R. Larson, C. Gold, D. Strick, D. K. Anderson, and A. F. Purchio.** 1988. Molecular cloning and nucleotide sequence of the pestivirus bovine viral diarrhea virus. Virology **165:**191–199.
21. **Collett, M. S., M. A. Wiskerchen, E. Welniak, and S. K. Belzer.** 1991. Bovine viral diarrhea virus genomic organization. Arch. Virol. **3**(Suppl.)**:**19–27.
22. **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.
23. **Daemer, R., C. Wychowski, A. Grakoui, C. M. Rice, and S. M. Feinstone.** Unpublished data.
24. **Davis, G. L., L. A. Balart, E. R. Schiff, K. Lindsay, H. C. Bodenheimer, Jr., R. P. Perrillo, W. Carey, I. M. Jacobson, J. Payne, J. L. Dienstag, D. H. VanThiel, C. Tamburro, J. Lefkowitch, J. Albrecht, C. Maschievitz, T. J. Ortego, A. Gibas, and the Hepatitis Interventional Therapy. Group.** 1989. Treatment of chronic hepatitis C with recombinant interferon alpha: a multicenter randomized, controlled trial. N. Engl. J. Med. **321:**1501–1506.
25. **DiBisceglie, A. M., P. Martin, C. Kassianides, M. Lisker-Melman, L. Murray, J. G. Waggoner, Z. Goodman, S. M. Banks, and J. H. Hoofnagle.** 1989. Recombinant interferon alpha therapy for chronic hepatitis C: a randomized, double-blind, placebo-controlled trial. N. Engl. J. Med. **321:**1506–1510.
26. **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.
27. **Falgout, B., M. Pethel, Y.-M. Zhang, and C.-J. Lai.** 1991. Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. J. Virol. **65:**2467–2475.
28. **Falkner, F. G., and B. Moss.** 1988. *Escherichia coli gpt* gene provides dominant selection for vaccinia virus open reading frame expression vectors. J. Virol. **62:**1849–1854.
29. **Farci, P., H. J. Alter, S. Govindarajan, D. C. Wong, R. Engle, R. R. Lesniewski, I. K. Mushahwar, S. M. Desai, R. H. Miller, N. Ogata, and R. H. Purcell.** 1992. Lack of protective immunity against reinfection with hepatitis C virus. Science **258:**135–140.
30. **Feinstone, S. M., K. B. Mihalik, T. Kamimura, H. J. Alter, W. T. London, and R. H. Purcell.** 1983. Inactivation of hepatitis B virus and non-A, non-B hepatitis by chloroform. Infect. Immun. **41:**816–821.
31. **Francki, R. I. B., C. M. Fauquet, D. L. Knudson, and F. Brown.** 1991. Classification and nomenclature of viruses: fifth report of the International Committee on Taxonomy of Viruses. Arch. Virol. **2**(Suppl.)**:**223.
32. **Fu, J., B.-H. Tan, E.-H. Yap, Y.-C. Chan, and Y. H. Tan.** 1992. Full-length cDNA sequence of dengue type 1 virus (Singapore strain S275/90). Virology **188:**953–958.
33. **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.
34. **Gorbalenya, A. E., A. P. Donchenko, E. V. Koonin, and V. M. Blinov.** 1989. N-terminal domains of putative helicases of flavi- and pestiviruses may be serine proteases. Nucleic Acids Res. **17:**3889–3897.
35. **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.
36. **Grakoui, A., C. Wychowski, C. Lin, S. M. Feinstone, and C. M. Rice.** 1993. Expression and identification of hepatitis C virus polyprotein cleavage products. J. Virol. **67:**1385–1395.
37. **Hahn, C. S., Y. S. Hahn, T. J. Braciale, and C. M. Rice.** 1992. Infectious Sindbis virus transient expression vectors for studying antigen processing and presentation. Proc. Natl. Acad. Sci. USA **89:**2679–2683.
38. **Hahn, Y. S., R. Galler, T. Hunkapiller, J. M. Dalrymple, J. H. Strauss, and E. G. Strauss.** 1988. Nucleotide sequence of dengue 2 RNA and comparison of the encoded proteins with those of other flaviviruses. Virology **162:**167–180.
39. **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.
40. **He, L. F., D. Alling, T. Popkin, M. Shapiro, H. J. Alter, and R. H. Purcell.** 1987. Determining the size of non-A, non-B hepatitis by filtration. J. Infect. Dis. **156:**636–640.
41. **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.
42. **Hollinger, F. B.** 1990. Non-A, non-B hepatitis viruses, p. 2239–2273. *In* B. N. Fields (ed.), Virology. Raven Press, Ltd., New York.
43. **Hollinger, F. B., G. Gitnick, R. D. Aach, W. Szmuness, J. W. Mosley, C. E. Stevens, R. L. Peters, J. M. Weiner, J. B. Werch, and J. J. Lander.** 1978. Non-A, non-B hepatitis transmission in chimpanzees: a project of the transfusion-transmitted viruses study group. Intervirology **10:**60–68.
44. **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.
45. **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.
46. **Iacono-Conners, L. C., and C. S. Schmaljohn.** 1992. Cloning and sequence analysis of the genes encoding the nonstructural proteins of Langat virus and comparative analysis with other flaviviruses. Virology **188:**875–880.
47. **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.
48. **Jacob, J. R., K. H. Burk, J. W. Eichberg, G. R. Dreesman, and R. E. Lanford.** 1990. Expression of infectious viral particles by primary hepatocytes isolated during the acute phase of non-A, non-B hepatitis. J. Infect. Dis. **161:**1121–1127.
49. **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.
50. **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.
51. **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.
52. **Kunkel, T. A.** 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA

82:488–492.
53. **Kuo, G., Q.-L. Choo, H. J. Alter, G. L. Gitnick, A. G. Redeker, R. H. Purcell, T. Miyamura, J. L. Dienstag, M. J. Alter, C. E. Stevens, G. E. Tegtmeier, F. Bonino, M. Colombo, W.-S. Lee, C. Kuo, K. Berger, J. R. Shuster, L. R. Overby, D. W. Bradley, and M. Houghton.** 1989. An assay for circulating antibodies to a major etiologic virus of human non-A non-B hepatitis. Science **244:**362–364.
54. **Lin, C., S. M. Amberg, T. J. Chambers, and C. M. Rice.** 1973. Cleavage at a novel site in the NS4A region by the yellow fever virus NS2B-3 proteinase is a prerequisite for processing at the downstream 4A/4B signalase site. J. Virol. **67:**2327–2335.
55. **Lin, C., T. J. Chambers, and C. M. Rice.** 1993. Mutagenesis of conserved residues at the yellow fever NS3/4A and NS4B/5 dibasic cleavage sites: effects on cleavage efficiency and polyprotein processing. Virology **192:**596–604.
56. **Mackett, M., and G. L. Smith.** 1986. Vaccinia virus expression vectors. J. Gen. Virol. **67:**2067–2082.
57. **Mackow, E., Y. Makino, B. Zhao, Y.-M. Zhang, L. Markoff, A. Buckler-White, M. Guiler, R. Chanock, and C.-J. Lai.** 1987. The nucleotide sequence of dengue type 4 virus: analysis of genes coding for nonstructural proteins. Virology **159:**217–228.
58. **Mandl, C. W., F. X. Heinz, E. Stöckl, and C. Kunz.** 1989. Genome sequence of tick-borne encephalitis virus (Western subtype) and comparative analysis of nonstructural proteins with other flaviviruses. Virology **173:**291–301.
59. **Matsudaira, P.** 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. **262:**10035–10038.
60. **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.
61. **Meyers, G., T. Rumenapf, and H. J. Thiel.** 1989. Molecular cloning and nucleotide sequence of the genome of hog cholera virus. Virology **171:**555–567.
62. **Meyers, G., N. Tautz, E. J. Dubovi, and H. J. Thiel.** 1991. Viral cytopathogenicity correlated with integration of ubiquitin-coding sequences. Virology **180:**602–616.
63. **Meyers, G., N. Tautz, R. Stark, J. Brownlie, E. Dubovi, M. S. Collett, and H.-J. Thiel.** 1992. Rearrangement of viral sequences in cytopathogenic pestiviruses. Virology **191:**368–386.
64. **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.
65. **Moss, B., O. Elroy-Stein, T. Mizukami, W. A. Alexander, and T. R. Fuerst.** 1990. New mammalian expression vectors. Nature (London) **348:**91–92.
66. **Nestorowicz, A., T. J. Chambers, and C. M. Rice.** Unpublished data.
67. **Nowak, T., P. M. Farber, G. Wengler, and G. Wengler.** 1989. Analyses of the terminal sequences of West Nile virus structural proteins and of the *in vitro* translation of these proteins allow the proposal of a complete scheme of the proteolytic cleavages involved in their synthesis. Virology **169:**365–376.
68. **Okamoto, H., K. Kurai, S.-I. Okada, K. Yamamoto, H. Iizuka, T. Tanaka, S. Fukuda, F. Tsuda, and S. Mishiro.** 1992. Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes. Virology **188:**331–341.
69. **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.
70. **Osatomi, K., and H. Sumiyoshi.** 1990. Complete nucleotide sequence of dengue type 3 virus genome RNA. Virology **176:** 643–647.
71. **Petric, M., R. H. Yolken, E. J. Dubovi, M. Wiskerchen, and M. S. Collett.** 1992. Baculovirus expression of pestivirus nonstructural proteins. J. Gen. Virol. **73:**1867–1871.
72. **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.
73. **Preugschat, F., E. M. Lenches, and J. H. Strauss.** 1991. Flavivirus enzyme-substrate interactions studied with chimeric proteinases: identification of an intragenic locus important for substrate recognition. J. Virol. **65:**4749–4758.
74. **Preugschat, F., and J. H. Strauss.** 1991. Processing of nonstructural proteins NS4A and NS4B of dengue 2 virus *in vitro* and *in vivo*. Virology **185:**689–697.
75. **Preugschat, F., C.-W. Yao, and J. H. Strauss.** 1990. In vitro processing of dengue virus type 2 nonstructural proteins NS2A, NS2B, and NS3. J. Virol. **64:**4364–4374.
76. **Prince, A., B. Brotman, T. Huima, D. Pascual, M. Jaffery, and G. Inchauspé.** 1992. Immunity in hepatitis C infection. J. Infect. Dis. **165:**438–443.
77. **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.
78. **Rice, C. M., and J. H. Strauss.** 1990. Production of flavivirus polypeptides by proteolytic processing. Semin. Virol. **1:**357–367.
79. **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.
80. **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.
81. **Sambrook, J., E. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
82. **Shimizu, Y. K., A. Iwamoto, M. Hijikata, R. H. Purcell, and H. Yoshikura.** 1992. Evidence for in vitro replication of hepatitis C virus genome in a human T-cell line. Proc. Natl. Acad. Sci. USA **89:**5477–5481.
83. **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. Characterization of the hepatitis E2/NS1 gene product expressed in mammalian cells. Virology **188:**819–830.
84. **Speight, G., G. Coia, M. D. Parker, and E. G. Westaway.** 1988. Gene mapping and positive identification of the non-structural proteins NS2A, NS2B, NS3, NS4B and NS5 of the flavivirus Kunjin and their cleavage sites. J. Gen. Virol. **69:**23–34.
85. **Sumiyoshi, H., C. Mori, I. Fuke, K. Morita, S. Kuhara, J. Kondou, Y. Kikuchi, H. Nagamatu, and A. Igarashi.** 1987. Complete nucleotide sequence of the Japanese encephalitis virus genome RNA. Virology **161:**497–510.
86. **Tabor, E., R. J. Garety, J. A. Drucker, L. B. Seeff, J. F. Hoofnagle, D. R. Jackson, M. April, L. F. Barker, and G. Pineda-Tamondong.** 1978. Transmission of non-A, non-B hepatitis from man to chimpanzee. Lancet **i:**463–466.
87. **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.
88. **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.
89. **Thiel, H.-J., R. Stark, E. Weiland, T. Rümenapf, and G. Meyers.** 1991. Hog cholera virus: molecular composition of virions from a pestivirus. J. Virol. **65:**4705–4712.
90. **Weiner, A. J., G. Kuo, D. W. Bradley, F. Bonino, G. Saracco, C. Lee, J. Rosenblatt, Q.-L. Choo, and M. Houghton.** 1990. Detection of hepatitis C viral sequences in non-A, non-B hepatitis. Lancet **335:**1–3.
91. **Wengler, G., G. Czaya, P. M. Farder, and J. H. Hegemann.** 1991. *In vitro* synthesis of West Nile virus proteins indicates

that the amino terminal segment of the NS3 protein contains the active centre of the protease which cleaves the viral polyprotein after multiple basic amino acids. J. Gen. Virol. **72:**851–858.

92. **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.
93. **Wiskerchen, M., S. K. Belzer, and M. S. Collett.** 1991. Pestivirus gene expression: the first protein product of the bovine viral diarrhea virus large open reading frame, p20, possesses proteolytic activity. J. Virol. **65:**4508–4514.
94. **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.