Specificity of the Hepatitis C Virus NS3 Serine Protease: Effects of Substitutions at the 3/4A, 4A/4B, 4B/5A, and 5A/5B Cleavage Sites on Polyprotein Processing

ALEXANDER A. KOLYKHALOV, EUGENE V. AGAPOV, AND CHARLES M. RICE*

Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, Missouri 63110-1093

Received 9 June 1994/Accepted 17 August 1994

Cleavage at four sites (3/4A, 4A/4B, 4B/5A, and 5A/5B) in the hepatitis C virus polyprotein requires a viral serine protease activity residing in the N-terminal one-third of the NS3 protein. Sequence comparison of the residues flanking these cleavage sites reveals conserved features including 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. In this study, we used site-directed mutagenesis to assess the importance of these and other residues for NS3 protease-dependent cleavages. Substitutions at the P7 to P2' positions of the 4A/4B site had varied effects on cleavage efficiency. Only Arg at the P1 position or Pro at P1' substantially blocked processing at this site. Leu was tolerated at the P1 position, whereas five other substitutions allowed various degrees of cleavage. Substitutions with positively charged or other hydrophilic residues at the P7, P3, P2, and P2' positions did not reduce cleavage efficiency. Five substitutions examined at the P6 position allowed complete cleavage, demonstrating that an acidic residue at this position is not essential. Parallel results were obtained with substrates containing an active NS3 protease domain in *cis* or when the protease domain was supplied in *trans*. Selected substitutions blocking or inhibiting cleavage at the 4A/4B site were also examined at the 3/4A, 4B/5A, and 5A/5B sites. For a given substitution, a site-dependent gradient in the degree of inhibition was observed, with the 3/4A site being least sensitive to mutagenesis, followed by the 4A/4B, 4B/5A, and 5A/5B sites. In most cases, mutations abolishing cleavage at one site did not affect processing at the other serine protease-dependent sites. However, mutations at the 3/4A site which inhibited cleavage also interfered with processing at the 4B/5A site. Finally, during the course of these studies an additional NS3 protease-dependent cleavage site has been identified in the NS4B region.

Hepatitis C viruses (HCV) comprise a genus in the flavivirus family and are responsible for serious and widespread human liver diseases (reviewed in references 17 and 26). These enveloped viruses are characterized by a positive-sense genome RNA of ~9.4 kb. A relatively long 5' nontranslated region of 341 bases appears to function as an internal ribosome entry site (10, 31–33) for translation of the viral polyprotein, which is typically 3,010 to 3,033 residues long (17, 26). The putative structural proteins (C, E1, and E2) are located in the N-terminal portion of the open reading frame with the nonstructural (NS) proteins in the remainder. Given the absence of an efficient system for replication of HCV in cell culture, cell-free translation and transient expression assays have been employed to study processing of the HCV polyprotein, which appears to be mediated by host and viral proteases. As established by these approaches, the gene order for the HCV-H strain polyprotein is NH₂-C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH (11–13, 22). Host signal peptidase localized in the endoplasmic reticulum lumen is believed to mediate processing at the C/E1, E1/E2, E2/p7, and p7/NS2 sites (12, 15, 22, 27). Assembly of functional HCV RNA replication complexes is believed to require processing at downstream sites by two HCV-encoded proteases. A Zn²⁺-dependent metalloprotease comprising NS2 and the N-termi-

nal one-third of NS3 is responsible for autocatalytic cleavage at the 2/3 site (12, 16). Processing at the downstream 3/4A, 4A/4B, 4B/5A, and 5A/5B sites is dependent on a serine protease domain, also located in the N-terminal one-third of the NS3 protein (3, 7, 11, 16, 25, 30). Although its precise role in proteolysis is not understood, the NS4A protein appears to function as a cofactor which is required for cleavage at the 3/4A and 4B/5A sites and enhances cleavage at the 5A/5B site (8, 23). Processing at these sites can occur either in *cis* or in *trans*. Cleavage at the 3/4A site appears to occur normally in *cis* (3), although *trans* cleavage at this site can be observed with some substrates, such as those which begin downstream of the NS3 serine protease domain (23). Cleavage at the other three sites can occur in *trans* (3, 8, 23, 30), although *cis* processing may also occur during processing of polyproteins containing a functional serine protease.

We previously identified potential cleavage sites for the HCV NS3 serine protease by isolation and N-terminal sequence analysis of NS4A, NS4B, NS5A, and NS5B (11). Alignment of sequences flanking these cleavage sites revealed three conserved positions which might be important in recognition and cleavage by the NS3 serine protease (11). An acidic residue is found at the P6 position for all four sites. At the P1 position, a Thr residue is found at the 3/4A site, whereas Cys residues are present at the other three sites. At the P1' position, Ser residues are found at all four sites, except for Ala at the 4A/4B site of several Asian subtypes. In this study, we have tested the importance of these and other positions for serine protease-dependent cleavage of polyproteins in *cis* and in *trans*. The results demonstrate that while some substitutions

* Corresponding author. Mailing address: Department of Molecular Microbiology, Washington University School of Medicine, Box 8230, 660 S. Euclid Ave., St. Louis, MO 63110-1093. Fax: (314) 362-1232. Electronic mail address: rice@borcim.wustl.edu.

at the P1 and P1' positions could block proteolysis, the acidic residue at P6 is not required for efficient cleavage, at least at the 4A/4B site. Except in one case, mutations blocking cleavage at one site did not affect processing at the other three sites, suggesting the absence of an obligate processing order. Finally, an additional serine protease-dependent cleavage site, which is processed inefficiently, was mapped to the NS4B region.

MATERIALS AND METHODS

Cell cultures and virus stocks. Sources and culture conditions for the BHK-21 and BSC-40 cell lines have been described previously (5). vTF7-3, a vaccinia virus recombinant expressing the T7 DNA-dependent RNA polymerase (9), was amplified, and titers were determined on BSC-40 cells (18).

Plasmid constructions and site-directed mutagenesis. Standard recombinant DNA techniques were used for plasmid construction, propagation, and analysis (1). Site-directed mutagenesis was performed with uridylylated pH3'2J/HCV1193-2623 DNA as previously described (19, 21). Mutations were identified by dideoxy sequence analysis of single- and double-stranded DNA. Sequences of all regions in the expression constructs derived by mutagenesis or PCR were verified.

(i) **cis-cleavage substrates.** Substrates containing the NS3 protease in *cis* were derivatives of pBRTM/HCV827-3011 (13), which contains the HCV-H strain coding region for the majority of the nonstructural region (from residue 827 to the end of the polyprotein). The polyprotein expressed from this construct includes a truncated NS2 protein (lacking 17 N-terminal residues), followed by NS3, NS4A, NS4B, NS5A, and NS5B (Fig. 1). Derivatives with substitutions at each of the cleavage sites were constructed by subcloning appropriate fragments from the mutagenized pH3'2J/HCV1193-2623 plasmids (described below). These restriction fragments were *NsiI-XhoI* (3/4A site), *NsiI-SnaBI* (4A/4B), *Bsu36I-Sse8387I* (4B/5A), and *BamHI-HindIII* (5A/5B). For the 4A/4B site mutants, fragments were subcloned into the recipient pBRTM/HCV827-3011ΔXB in which the sequence between the *XhoI* (nucleotide position 5610 in the HCV cDNA sequence [6]) and *Bsu36I* (6209) sites had been deleted. Similarly for the 5A/5B site mutants, fragments were subcloned into pBRTM/HCV827-3011ΔSH in which the sequence between the *SalI* (7556) and *HindIII* (7681) sites was deleted (after filling in the recessed ends with Klenow fragment).

(ii) **trans-cleavage substrates.** pH3'2J/HCV1193-2623 was constructed by subcloning the *ClaI-EcoRI* fragment of pBRTM/HCV1193-3011 (11) into *NarI*- and *EcoRI*-digested pH3'2J1 which contained the f1 origin (14). The resulting phagemid contains the T7 promoter, the encephalomyocarditis virus internal ribosome entry site, and the HCV-H cDNA encoding polyprotein residues 1193 to 2623. This truncated polyprotein (hereafter referred to as 1193-2623), used as the substrate in *trans*-cleavage assays, lacks the serine protease domain but includes the C-terminal 465 residues of NS3, NS4A, NS4B, NS5A and the 203 N-terminal residues of NS5B.

Synthetic oligonucleotides were used to produce pH3'2J/HCV1193-2623 derivatives with the indicated substitution mutations (altered codons are shown in boldface type, nucleotide substitutions are in uppercase letters, and restriction sites introduced to facilitate mutant screening are underlined): 3/4A site, P1 position: Arg (**AGGTCGac**), Gly (**GGGTCGac**), Asn (**aACagTacT**), or Asp (**GACagTacT**); 3/4A site, P1' position: Pro (**CCcagGtg**) or Arg (**CgcagGtg**); 4A/4B site, P7 position: His (**Cac**) or Arg (**CGc**); 4A/4B site, P6 position: Ala (**gCA**), Glu (**gaA**), Pro (**CCA**), Thr (**ACA**), or Gln (**CaA**); 4A/4B site, P5 position: Lys (**Aag**); 4A/4B site, P3 position: Asn (**AaT**) or

Lys (**AaG**); 4A/4B site, P2 position: Lys (**Aag**) or Asn (**AaT**); 4A/4B site, P1 position: Arg (**GCA**), Gly (**GgA**), Ser (**Agc**), Asn (**AAc**), Asp (**GAc**), Thr (**ACc**), or Leu (**CTc**); 4A/4B site, P1' position: Ile (**ATt**), Thr (**AcT**), Arg (**CGt**), Ala (**Gct**), Asp (**GAt**), His (**CAt**), or Pro (**CCt**); 4A/4B site, P2' position: Lys (**Aag**); 4B/5A site, P1 position: Arg (**AgGtccggAtcc**), Gly (**GgGtccggAtcc**), Asn (**AActccggAtcc**), or Asp (**GActccggAtcc**); 4B/5A site, P1' position: Pro (**CccggAtcc**) or Arg (**CGcggAtcc**); 5A/5B site, P1 position: Arg (**AgAtcT**), Gly (**GgAtcT**), Asn (**AAcAGCatgt**), or Asp (**GAcAGCatgt**); and 5A/5B site, P1' position: Pro (**CcC**) or Arg (**CGCatgt**). To produce the plasmids used for assaying *trans* cleavage at the 4A/4B site, the *NsiI-XhoI* fragment from each mutant derivative (which had been sequenced) was ligated with a *XhoI-Bsu36I* fragment and *NsiI-Bsu36I*-digested deletion recipient pH3'2J/HCV1193-2623ΔNN (which contained a deletion between *NaeI* [5414] and *NaeI* [6049]).

(iii) **Other plasmids.** pTM3/HCV/NS3₁₈₁ was used to express the N-terminal 181 residues of NS3, which constitute a functional serine protease domain used in *trans*-cleavage assays (23). pTM3/HCV/NS4A-4B expresses an NS4A-4B polyprotein with two additional N-terminal residues (Met-Ala) (23). pTM3/Ubi/HCV/NS4A_{T>R} was used to express a chimeric protein consisting of a ubiquitin monomer fused to HCV NS4A with Arg substituted for the N-terminal Thr residue of NS4A. PCR using primers 5'-CGTACCGCGGTGGACGTACGTGGGTGCTCGTTGGC-3' (sense) and 5'-CCTACTGCAGTAGCACTCTTCCATCTC-3' (antisense), followed by digestion with *SacII* and *PstI*, was used to produce the NS4A cassette, which was then subcloned into pTM3/ubiquitin (23) which had been digested with the same two enzymes. pBRTM/HCV827-1971-myc {4A/4B P1 Arg} was used to express a polyprotein which is processed to yield truncated NS2, NS3, and uncleaved NS4A-4B fused in frame to a 10-amino-acid sequence derived from *c-myc* (24). This plasmid was constructed by subcloning the *Bsu36I-NotI* fragment from SIN-rep5/HCV1658-1971-myc (24) into pBRTM/HCV827-3011 {4A/4B P1 Arg}.

Transient expression assay. Subconfluent monolayers of BHK-21 cells in 35-mm six-well plates (approximately 10⁶ cells per well) were infected with vTF7-3 using a multiplicity of infection of 10 PFU per cell for 1 h at 37°C. After removal of the inoculum, cells were washed two times with prewarmed minimal essential medium (MEM) and transfected for 3 h at 37°C with a mixture consisting of 1 to 2 μg of plasmid DNA, 15 μg of Lipofectamine (GIBCO BRL) (or in initial experiments, 12.5 μg of Transfectam from Promega), and 0.5 ml of MEM. After incubation, cells were washed two times with MEM lacking methionine prior to incubation in the same medium for 30 min. Proteins were metabolically labeled by replacing this medium with 0.6 ml of MEM lacking methionine which contained 2% fetal bovine serum and Tran ³⁵S-label (70 μCi/ml; ICN). Cells were labeled for either 30 min or 4 h (in long labeling experiments) at 37°C. After labeling, cells were washed twice with MEM and lysed with a solution of 0.5% sodium dodecyl sulfate (SDS) containing 20 mM phenylmethylsulfonyl fluoride. Genomic DNA was sheared by repeated passage through a 26-gauge needle, and samples were either used immediately for immunoprecipitation or stored in aliquots at -70°C.

Immunoprecipitation and protein analyses. Prior to immunoprecipitation, a portion of each lysate was heated at 70°C for 15 min, diluted five times with solution containing 50 mM Tris-Cl (pH 7.6), 200 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, and bovine serum albumin (1 mg/ml), and clarified by centrifugation at 16,000 × g for 15 min. HCV-specific proteins

FIG. 1. Diagram of the HCV polyprotein and expression constructs. The HCV-H strain polyprotein and its cleavage products are shown as boxes. The identities of the mature proteins, including putative structural proteins (C, E1, and E2), p7, and the nonstructural proteins (NS2, NS3, NS4A, NS4B, NS5A, and NS5B), are indicated with the NS3 serine protease domain shaded. The number at the bottom of each cleavage junction indicates the position of the N-terminal residue of the cleavage product in the polyprotein sequence. The expression constructs used in this study are shown below the polyprotein diagram (and described in Materials and Methods and Results). Ubi, ubiquitin. Open circle, c-myc tag.

were immunoprecipitated by incubation with 3 μ l of patient JHF serum (13) or 5 μ l of polyclonal rabbit sera specific for NS2-3 (rabbit WU43), NS2 (WU107), NS3 (WU117), NS5A (WU123), NS5A-5B (WU113), or NS5B (WU115) (13). One μ l ($\sim 1 \mu$ g) of monoclonal antibody M5 specific for NS4A (kindly provided by Sushil Devare, Abbott Laboratories) was used for some immunoprecipitations. Incubations were usually for 12 h at 4°C with agitation. Immune complexes were collected by using *Staphylococcus aureus* Cowan strain I (Pansorbin; Calbiochem) as described in reference 29 and analyzed by SDS-12% polyacrylamide gel electrophoresis (PAGE) unless otherwise indicated (20). Following electrophoresis, gels were fixed with 10% acetic acid-ethanol, dried on filter paper (Whatman) and exposed to X-ray film for 12 to 24 h.

RESULTS

trans-cleavage substrate and assay. The vaccinia virus/T7 hybrid expression system was used to test the effects of mutations at each of the serine protease-dependent cleavage sites. The substrate used for *trans*-cleavage assays consisted of a truncated HCV polyprotein spanning residues 1193 to 2623 (Fig. 1). This substrate lacks the serine protease domain, contains the 465 C-terminal residues of NS3, and extends through the first 203 residues of NS5A. Coexpression of this polyprotein with the NS3 serine protease domain (designated 3₁₈₁) led to the production of four cleavage products identified as truncated NS3 (designated 3*), NS4A, NS4B, and NS5A on the basis of their sizes (Fig. 2) and immunoreactivity with region-specific antisera (data not shown). The patient serum used for the immunoprecipitation shown in Fig. 2 did not react with 3₁₈₁ or NS5B products (data not shown). The mobilities of the NS4A, NS4B, and NS5A species produced by *trans* cleavage of 1193-2623 were identical to those produced from the residue 827-3011 polyprotein, which contains both HCV proteases and spans nearly the entire HCV nonstructural region. The 1193-2623 substrate expressed by itself was not cleaved, nor was cleavage observed during coexpression with an inactive 3₁₈₁ derivative which contained an Ala substitution for Ser-1165 (data not shown), the putative nucleophile of the serine

FIG. 2. *trans*-processing assay. vTF7-3-infected BHK-21 cell monolayers were transfected with the plasmid DNAs (+) indicated above the gel or pGEM 3Zf(+) as a negative control (–) and labeled for 4 h with ³⁵S-protein labeling mixture as described in Materials and Methods. Cell lysates were prepared and immunoprecipitated with patient JHF serum, and the products were separated by SDS-12% PAGE. HCV-specific proteins and polyproteins (without the NS prefixes) are identified at the left, and the sizes of ¹⁴C-labeled protein molecular mass markers (in kilodaltons) are indicated at the right.

protease (3, 7, 11, 16, 25, 30). These results demonstrate that serine protease-dependent *trans* cleavage can occur at the 3/4A, 4A/4B, 4B/5A, and 5A/5B sites of the 1193-2623 substrate.

Substitutions at the 4A/4B cleavage site. We began by examining a number of substitutions for residues flanking the 4A/4B cleavage site (P7 to P2'), using the *trans*-cleavage assay (Fig. 3A). As mentioned earlier, conserved positions at the four cleavage sites include an acidic residue at P6, Cys or Thr at P1, and Ser or Ala at P1'. Furthermore, basic residues are not found between P7 and P2', nor are His or Asn. On the basis of these observations, we focused mainly on substitutions at the P6, P1, and P1' positions but also examined a limited number of substitutions at the other positions. Since mutations blocking cleavage at the 4A/4B site do not affect *trans* processing at other sites (as discussed below), the degree of inhibition of cleavage at the 4A/4B site could be monitored by the appearance of the 31-kDa NS4A-4B polyprotein relative to the NS4A and NS4B cleavage products.

As shown in Fig. 3B, none of the substitutions at the P7 (Phe→His or Arg), P5 (Glu→Lys), P3 (Glu→Asn or Lys), P2 (Glu→Asn or Lys), or P2' (Gln→Lys) position had observable effects on cleavage at the 4A/4B site during a 4-h labeling period. This was also true for each of five substitutions examined at P6 (Asp→Thr, Ala, Asn, Glu, or Pro) (Fig. 3C). In this analysis, it should be noted that some of the P6 substitutions appeared to inhibit or abolish reactivity of the NS4A product with the patient serum used for immunoprecipitation. At the P1' position, six substitutions permitted efficient cleavage (Ser→Ile, Thr, Arg, Ala, Asp, or His), whereas substitution with Pro dramatically inhibited cleavage (Fig. 3D). Instead of the NS4A and NS4B cleavage products, uncleaved NS4A-4B (31 kDa) was observed, as were additional minor species (p24.5, p21, and p10). As shown later, at least two of these polypeptides, p21 and p10, appear to be produced by an alternative cleavage in the NS4B region which is favored when the 4A/4B site is blocked. Similar species were also observed for many of the P1 (Cys) substitutions, which had varied effects on cleavage at the 4A/4B site (Fig. 3E). Thr and Leu were tolerated, whereas substitution with Arg almost completely blocked processing. For the remaining substitutions, Asp

FIG. 3. Effects of substitutions at the 4A/4B site on *trans* processing. (A) Diagram of the 1193-2623 *trans*-processing substrate and the sequence (from P7 to P2') flanking the 4A/4B cleavage site. Below the sequence, the single substitution mutations examined in this study are listed (using the single-letter code). (B) Substitutions at the P7, P5, P3, P2, and P2' positions. (C) Substitutions at the P6 position. (D) Substitutions at the P1' position. (E) Substitutions at the P1 position. The *trans*-processing assays, immunoprecipitation and SDS-PAGE were conducted as described in Materials and Methods and the Fig. 2 legend. Transfected plasmids are indicated at the top of each panel. Plasmids expressing the parental substrate (wt) or the substitution mutations (in the single-letter code) are indicated immediately above the lanes. HCV-specific proteins and polyproteins are identified at the left, and the sizes of ^{14}C -labeled protein molecular mass markers (in kilodaltons) are indicated at the right of each panel.

showed the greatest inhibitory effect, followed by Asn > Gly > Ser.

Since these initial experiments were conducted with a 4-h labeling period, we examined processing patterns after only a 30-min pulse to determine if substitutions at the P6, P1, and P1' positions had more subtle effects on cleavage efficiency (Fig. 4). Under these conditions, uncleaved NS4A-4B (p31) can still be observed for the parental substrate. Similar to the longer labeling results, substitution mutants at the P6 position were essentially indistinguishable from the parental substrate (Fig. 4A). For P1', no cleavage was observed for the Pro substitution, whereas the other substitutions had little effect, except for perhaps a slight decrease in cleavage efficiency for the Ile and Arg substitutions (Fig. 4B). The most dramatic inhibition was again seen for the P1 substitutions (Fig. 4C). No

cleavage was observed for the Arg and Asp substitutions, and a gradient of inhibition was observed for four other substitutions (Asn > Gly > Ser > Thr). Remarkably, the P1 Cys→Leu substitution appeared to enhance cleavage at the 4A/4B site.

Substitutions at the 3/4A, 4A/4B, 4B/5A, and 5A/5B sites assayed using the 827-3011 polyprotein. Substitutions which abolished or inhibited cleavage at the 4A/4B site in the *trans*-cleavage assay were also created at the 3/4A, 4B/5A, and 5A/5B sites of the 827-3011 polyprotein. This protein begins 17 residues downstream from the NS2 N terminus and encodes essentially the entire HCV nonstructural region, including the two functional HCV proteases. As mentioned in the introduction, cleavages at the 2/3 and 3/4A sites of this polyprotein are believed to occur primarily in *cis*; and, although *trans* cleavage has been observed at the other three sites, *cis* cleavage may

FIG. 4. *trans* processing of P6, P1, and P1' 4A/4B site substitutions during a short pulse-label. *trans*-processing assays, immunoprecipitation, and SDS-PAGE were conducted as described for Figs. 2 and 3 and in Materials and Methods, except that labeling lasted for 30 min instead of 4 h. (A) Substitutions at the P6 position. (B) Substitutions at the P1' position. (C) Substitutions at the P1 position. Labeling is the same as that for Fig. 2 and 3.

also occur. As shown in Fig. 5 and reported previously (11), processing occurs at the 2/3 site as well as at the downstream serine protease-dependent sites to produce NS3, NS4A, NS4B, and NS5A. Minor amounts of the NS2*-3 polyprotein were also observed for the parental polyprotein. Truncated NS2* and NS5B were not seen, since the patient serum used for these immunoprecipitations does not recognize these antigens (data not shown).

None of the substitutions at the 3/4A site completely blocked processing, although significant inhibition was observed for Thr→Arg at P1 and Ser→Pro at P1' (Fig. 5A). Inhibited processing resulted in the accumulation of uncleaved NS3-4A and the disappearance of NS4A and NS3. Interestingly, mutations which inhibited processing at the 3/4A site also seemed to result in increased accumulation of NS4B-5A (as discussed below). In the polyprotein with the Arg substitution at P1', the NS4A polypeptide exhibited significantly slower mobility than the parental and other constructs. This might have resulted from utilization of an alternative upstream processing site or a direct effect of the Arg substitution on NS4A mobility. To address this issue, we constructed a plasmid which could be used to express a ubiquitin-NS4A fusion protein. This protein consisted of a ubiquitin monomer fused in frame to NS4A with the Arg substitution for the N-terminal Ser residue. This fusion protein should be efficiently processed by ubiquitin carboxy-terminal hydrolase (2) at the sequence Arg-Gly-Gly↓Arg to produce ubiquitin and NS4A with an N-terminal Arg residue. As shown in Fig. 5B, this NS4A product migrates more slowly than parental NS4A and comigrates with the NS4A produced by serine protease-dependent cleavage of the polyprotein with the P1' Arg substitution. Although not verified by sequence analysis, these results strongly argue that cleavage has occurred at the authentic 3/4A site.

At the P1' and P1 positions of the 4A/4B site, results similar to those of the *trans*-cleavage assay were obtained. At P1', only the Pro substitution significantly blocked cleavage, resulting in uncleaved NS4A-4B, a trace of NS4B, and enhanced levels of p24.5, p21 and p10 (Fig. 5C). At P1 (Fig. 5D), the degree of inhibition was greatest for the Arg substitution, followed by Asp > Asn > Gly > Ser > Thr. The P1 Leu substitution was indistinguishable from the parental construct.

Substitutions at the 4B/5A and 5A/5B sites had more dramatic effects on processing. For the 4B/5A site, blocked or inhibited cleavage resulted in the disappearance of NS4B and NS5A and the accumulation of an NS4B-5A polyprotein (Fig. 5E). Although interpretation is complicated by an apparent background doublet migrating in the NS4B region, Cys→Gly, Arg, or Asp at P1 and Ser→Pro at P1' appeared to block detectable processing at the 4B/5A site. Asn at P1 or Arg at P1' dramatically reduced or slightly inhibited cleavage, respectively. Several HCV-specific polypeptides migrating between 38 and 48 kDa were enhanced when processing at the 4B/5A site was inhibited; these did not react with any of our region-specific antisera (data not shown) and have not been characterized further. At the 5A/5B site (Fig. 5F), processing was not detected for any of the P1 substitutions (Cys→Gly, Arg, Asn, and Asp) or the P1' Ser→Pro substitution and rather than NS5A, only uncleaved NS5A-5B was observed. Partial processing for the P1' Arg substitution occurred.

Effects of cleavage site mutations on processing at other sites. For mutations inhibiting or blocking cleavage at the 4A/4B, 4B/5A, and 5A/5B sites of the 827-3011 polyprotein, no obvious effects on processing at other sites were noticed. In contrast, mutations at the 3/4A site, none of which completely blocked cleavage, seemed to result in the accumulation of NS4B-5A (Fig. 5A). This was examined more carefully in a 30-min pulse-labeling experiment in which the samples were separated on a lower (9%)-concentration polyacrylamide gel (Fig. 6). Polyproteins and cleavage products were identified by size and reactivity with region-specific antisera (data not shown). Mutations inhibiting cleavage at the 3/4A site resulted in enhanced accumulation of NS3- and NS4A-containing polyproteins NS3-4A, NS2-3-4A, and NS3-4A-4B. In addition, NS4B-5A also accumulated and the level of this polyprotein correlated with the degree of cleavage inhibition at the 3/4A site. This was surprising given that cleavage at the 4B/5A site can occur in *trans* and suggests that proteases consisting of unprocessed NS3-4A may be less efficient for *trans* cleavage at this site (see Discussion).

Additional cleavage site in the NS4B region. During the mutagenesis studies of the 4A/4B cleavage site, increased levels of a 21-kDa protein (p21) and a 10-kDa protein (p10) were observed when cleavage at the 4A/4B site was inhibited.

FIG. 5. Processing of 827-3011 polyproteins containing substitutions at the 3/4A, 4A/4B, 4B/5A, and 5A/5B sites. vTF7-3-infected BHK-21 cell monolayers were transfected with plasmids expressing the indicated parental (wt) or mutant (in the single-letter code) 827-3011 polyproteins (+) or pGEM 3Zf(+) as a negative control (-) and labeled for 4 h with 35 S-protein labeling mixture as described in Materials and Methods. Cell lysates were prepared and immunoprecipitated with patient JHF serum, and the products were separated by SDS-12% PAGE. (A) Substitutions at the P1 and P1' positions of the 3/4A site. (B) Comparison of the NS4A species produced by 827-3011 (wt), 827-3011 containing the Arg substitution at the P1' position of the 3/4A site (R), and the ubiquitin-NS4A (Ubi-4A) fusion with Arg at P1' (see Fig. 1 and text). (C) Substitutions at the P1' position of the 4A/4B site. (D) Substitutions at the P1 position of the 4A/4B site. (E) Substitutions at the P1 and P1' positions of the 4B/5A site. (F) Substitutions at the P1 and P1' positions of the 5A/5B site. Labeling is the same as that for Fig. 2 and 3.

An example, in which the 1193-2623 polyprotein with the Pro substitution at the P1' position of the 4A/4B site was processed *in trans* by 3₁₈₁, is shown in Fig. 7A. Besides p21 and p10, a minor 24.5-kDa band migrating slightly faster than NS4B, which has not been characterized further, was also observed. Smaller amounts of p21 were also observed by *trans* processing of an NS4A-4B substrate (Fig. 7A), which suggested that p21 production was serine protease dependent and that p21 was derived from the NS4A-4B region. p10 also mapped to this region since it, but not p21, reacted with an NS4A-specific monoclonal antibody (data not shown). On the basis of these results, we believe that p21 and p10 result from an additional cleavage in the NS4B region. To map this cleavage site, a 10-residue sequence from human *c-myc* was fused in frame to the C terminus of NS4B (lacking the C-terminal Cys residue), to produce 827-1971-myc {4A/4B P1 Arg}. NS4A-4B and p21 were absent from the products expressed by this construct and were instead replaced by two slower-mobility forms, which we believe to be NS4A-4B-myc and p21-myc, respectively (Fig.

7B). The migration of p10 was unaffected. This suggests that the additional cleavage site lies near the N terminus of NS4B and that p21 is a C-terminal fragment of NS4B.

DISCUSSION

To begin to understand more about the substrate requirements for polyprotein processing by the HCV NS3 serine protease, we have examined the effects of a number of substitutions for residues flanking the four serine protease-dependent cleavage sites. The most extensive analysis was conducted for the 4A/4B cleavage site, where the effects of substitutions at the P7 to P2' positions were tested. Substitutions at the P7 to P2 and P2' positions had no observable effect on cleavage efficiency. We previously noted that acidic residues were found at P6 positions of all four cleavage sites and suggested a role for these residues in cleavage site recognition by the HCV serine protease (11). At least for the 4A/4B site (and the cleavage assays employed in our study), an acidic

FIG. 6. Mutations inhibiting cleavage at the 3/4A site affect processing at the 4B/5A site. vTF7-3-infected BHK-21 cell monolayers were transfected with plasmids expressing the indicated parental (wt) or mutant (in the single-letter code) 827-3011 polyproteins (+) or pGEM 3Zf(+) as a negative control (-) and labeled for 30 min with 35 S-protein labeling mixture. Cell lysates were prepared and immunoprecipitated with patient JHF serum, and the products were separated by SDS-9% PAGE. Labeling is the same as that for Fig. 2 and 3.

residue at P6 does not appear to be required for efficient cleavage. It is still possible that an acidic residue at this position is important for optimal cleavage efficiency at other sites or under more stringent conditions, such as in HCV-infected cells which may contain lower levels of protease than those obtained using the vaccinia virus/T7 hybrid expression system. In contrast, more dramatic effects on cleavage efficiency were observed for substitutions at the P1 and P1' positions. At the P1' Ser, only substitution with Pro dramatically reduced cleavage. Even during a short pulse, slight inhibition was observed for only two of the P1' substitutions (Ile and Arg). Other substitutions such as Ala, which is found at the P1' position of several HCV strains, did not inhibit cleavage. The most important position was clearly P1, which is Cys at the 4A/4B, 4B/5A, and 5A/5B sites and Thr at the 3/4A site (11). The degree of inhibition was greatest for the Arg substitution, followed by Asp > Asn > Gly > Ser > Thr. Similar results were obtained in *trans*-cleavage assays or when the mutations were examined in the protease-containing 827-3011 polyprotein, where cleavage could also occur in *cis*, at least in theory (this has not been examined experimentally). Interestingly, substitution with Leu actually increased cleavage efficiency, suggesting that the parental Cys at this position may be suboptimal for cleavage at the 4A/4B site. This result appears to conflict with the recent modeling study of the HCV serine protease substrate binding pocket which predicted a preference for Cys at P1 (28). In any case, the importance of the P1 residue is certainly not unexpected since this position plays a critical role in substrate recognition for many other serine proteases (see reference 4).

Selected substitutions which blocked or inhibited cleavage at the 4A/4B site were also examined at the other three serine protease-dependent sites of the 827-3011 polyprotein. These included Arg, Asp, Asn, and Gly substitutions at the P1 positions and Pro and Arg substitutions at the P1' positions. From the limited data obtained thus far, there appears to be a site-dependent gradient in the degree of inhibition, with the 3/4A site being least sensitive to mutagenesis, followed by the 4A/4B, 4B/5A, and 5A/5B sites. For example, all four P1 substitutions partially inhibited processing at the 3/4A site (Arg > Asp > Asn ≈ Gly), while significantly greater inhibi-

FIG. 7. An alternative NS3 serine protease-dependent cleavage site in NS4B. (A) Longer exposure of the last four lanes of Fig. 3D showing the effect of the P1' Pro substitution at the 4A/4B site on the appearance of alternative cleavage products p21 and p10. (See the legend of Fig. 3 for experimental details.) (B) Mapping the alternative cleavage site. vTF7-3-infected BHK-21 cell monolayers were transfected with plasmids expressing 827-3011 {4A/4B P1 Arg}, 827-1971-myc {4A/4B P1 Arg} (+) or pGEM 3Zf(+) as a negative control (-) and labeled for 4 h with 35 S-protein labeling mixture. Cell lysates were prepared and immunoprecipitated with patient JHF serum, and products were separated by SDS-12% PAGE. The origin of the doublet migrating between p21 and NS4A-4B, which appears to increase in size with fusion of the *c-myc* sequence, is unclear. The top band is believed to represent NS4B or NS4B-myc produced by partial cleavage at the 4A/4B site. The lower band in each doublet may represent p24.5 and the corresponding *myc*-tagged derivative. The main caveat to this interpretation is that the shifts in the mobilities of the bands in the doublet appear to be smaller than would be expected given the difference in mobilities between p21 and NS4A-4B and their putative *myc*-tagged derivatives. Labeling is the same as that for Fig. 2 and 3.

tion was observed at the 4A/4B site (Arg > Asp > Asn > Gly). At the 4B/5A site, only the Asn substitution allowed detectable processing; and at the 5A/5B site, all four substitutions blocked detectable cleavage. We also observed differences in the effects of particular substitutions which did not fit with this general trend. For instance, the P1' Arg substitution was more effective at inhibiting cleavage at the 3/4A site than at the 4A/4B site. Several factors, including polyprotein context and cleavage mechanism, may account for these differences. During normal processing of the HCV polyprotein, cleavage at the 3/4A site is believed to occur in *cis* (3). Efficient intramolecular cleavage at this site may explain why substitutions are more readily tolerated. The mechanism(s) (*cis* or *trans*) by which the other sites in the HCV polyprotein are cleaved is not known and may vary depending on the concentration of *trans*-acting protease. At least in *trans*-cleavage reactions, cleavage at the 4B/5A site requires both the NS3 serine protease and the NS4A protein (8, 23). In contrast, NS4A stimulates *trans* cleavage at the 5A/5B site but is not required (8, 23). Hence, one explanation for the greater sensitivity of the 5A/5B site to mutagenesis is that cleavage efficiency is determined primarily by the interaction between the protease substrate binding pocket and the residues at the cleavage site. In the case of 4B/5A or the other sites, the protease-substrate interaction may also be stabilized by interactions with NS4A such that mutations at the cleavage site are less deleterious.

FIG. 8. Model for processing in the NS4A-4B region. A diagram of the HCV-H polyprotein is shown at the top, with the cleavage products indicated. Also shown are the four major cleavages (small arrows) believed to be mediated by the NS3 serine protease domain (hatched region). Processing at the 4A/4B, 4B/5A, and 5A/5B sites can occur in *trans* but may also occur in *cis*, which is believed to be the mechanism for cleavage at the 3/4A site. Besides the predominant processing pathways (solid arrows), an alternative serine protease-dependent processing site (wavy line) in the NS4B region is cleaved inefficiently (shaded arrow) to produce p21 and p10. p10 can be further processed at the 4A/4B site; mature NS4B, however, does not appear to be a substrate for processing at the alternative site (large $\times$). See the text for further discussion of the model.

In the context of the 827-3011 polyprotein, mutations blocking or inhibiting cleavage at one site did not usually affect processing at other sites. Exceptions were mutations at the 3/4A site, which appeared to also inhibit processing at the 4B/5A site. This may result from an effect of the mutations on the NS3 protease or the NS4A cofactor, both of which are required for *trans* cleavage at the 4B/5A site (8, 23). Alternatively, cleavage at 3/4A may be required for efficient *cis* or *trans* cleavage at the 4B/5A site. Further experiments are needed to explore these possibilities. Although the majority of our data might be cited as evidence against an obligatory processing order, the interpretation of these results is complicated. By the time our analyses were performed, expression of the HCV polyprotein using the vaccinia virus system had led to accumulation of NS3 serine protease which can act in *trans*. Since processing at the 4A/4B, 4B/5A, and 5A/5B sites can occur efficiently in *trans*, substitutions blocking cleavage at one site would not be expected to inhibit cleavage at other sites (unless they do so by altering the conformation of the polyprotein substrate). This situation is likely to be quite different from the initial stages of HCV infection. *cis* cleavages catalyzed by the viral proteases are expected to dominate early processing of polyprotein translation products, and such cleavages may be important for assembly of RNA replication complexes. In this regard, it will be of interest to establish a cell-free system which will allow us to determine which, if any, of the serine protease-dependent cleavages besides 3/4A can occur in *cis* and then use this system to determine if there is an obligate or preferred processing order for these cleavages.

During analysis of mutations at the 4A/4B site, at least one alternative cleavage site in the NS4B region has been identified. This alternative site also appears to be inefficiently utilized in the parental polyprotein (see Fig. 5C and 7A). Substitutions inhibiting cleavage at the 4A/4B site enhanced

the production of at least three species: p24.5, p21, and p10. We have not been able to identify p24.5, but p10 (on the basis of its reactivity with an NS4A-specific monoclonal antibody) contains NS4A sequences and p21 was shown to be a C-terminal fragment of NS4B. On the basis of these data and the sizes of these products, the following model can be proposed (Fig. 8). *cis* processing at the 3/4A site and *cis* or *trans* processing at the 4B/5A site can lead to the production of NS4A-4B or mature cleavage products NS4A and NS4B. Inefficient cleavage can also occur at an alternative site in the NS4B region 20 to 30 residues downstream from the 4A/4B site to produce p10 and p21. p10 is observed only when cleavage at the 4A/4B site is inhibited. Thus, in the wild-type polyprotein, p10 must be a substrate for cleavage at the 4A/4B site to produce NS4A and a small peptide (p4 in Fig. 8). The observation that NS4B is a stable cleavage product suggests that processing at the alternative site in NS4B cannot occur if prior cleavage at the 4A/4B site occurs. This would explain the enhanced accumulation of p21 in polyproteins with mutations inhibiting cleavage at the 4A/4B site. Although it has not been shown that the HCV serine protease is directly responsible for this alternative cleavage, processing at this site has been observed only in the presence of active NS3 protease. Further work is needed to determine if similar complex processing in the NS4A-4B region occurs with other HCV types and whether these alternative products play any role in HCV replication.

ACKNOWLEDGMENTS

We thank Chao Lin and Brett Lindenbach for plasmids; Henry Hsu and Harry Greenberg for patient sera; Sushil Devare for the NS4A-specific anti-peptide monoclonal antibody M5; and Bernie Moss for the vaccinia virus expression system. We are also grateful to many colleagues for helpful discussions during the course of this work and to Sean Amberg, Chao Lin, and Karen Reed for critical reading of the manuscript.

This work was supported in part by a grant from the Public Health Service (CA57973).

REFERENCES

1. Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.). 1993. Current protocols in molecular biology. Greene Publishing Associates, New York.
2. Bachmair, A., D. Finley, and A. Varshavsky. 1986. *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* 234:179-186.
3. Bartenschlager, R., L. Ahlborn-Laake, J. Mous, and H. Jacobsen. 1993. Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. *J. Virol.* 67:3835-3844.
4. Bazan, J. F., and R. J. Fletterick. 1990. Structural and catalytic models of trypsin-like viral proteases. *Semin. Virol.* 1:311-322.
5. 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.
6. Daemer, R., C. Wychowski, A. Grakoui, C. M. Rice, and S. M. Feinstone. Unpublished data.
7. Eckart, M. R., M. Selby, F. Masiares, C. Lee, K. Berger, K. Crawford, C. Kuo, G. Kuo, M. Houghton, and Q.-L. Choo. 1993. The hepatitis C virus encodes a serine protease involved in processing of the putative nonstructural proteins from the viral polyprotein precursor. *Biochem. Biophys. Res. Commun.* 192:399-406.
8. Failla, C., L. Tomei, and R. De Francesco. 1994. Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. *J. Virol.* 68:3753-3760.
9. 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.
10. Fukushi, S., K. Katayama, C. Kurihara, N. Ishiyama, F. B. Hoshino, T. Ando, and A. Oya. 1994. Complete 5' noncoding region is necessary for the efficient internal initiation of hepatitis C virus RNA. *Biochem. Biophys. Res. Commun.* **199**:425–432.
 11. Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice. 1993. Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. *J. Virol.* **67**:2832–2843.
 12. Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice. 1993. A second hepatitis C virus-encoded proteinase. *Proc. Natl. Acad. Sci. USA* **90**:10583–10587.
 13. 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.
 14. 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.
 15. 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.
 16. Hijikata, M., H. Mizushima, T. Akagi, S. Mori, N. Kakiuchi, N. Kato, T. Tanaka, K. Kimura, and K. Shimotohno. 1993. Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. *J. Virol.* **67**:4665–4675.
 17. 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.
 18. 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.
 19. Kunkel, T. A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**:488–492.
 20. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**:680–685.
 21. Lin, C., T. J. Chambers, and C. M. Rice. 1993. Mutagenesis of conserved residues at the yellow fever virus 3/4A and 4B/5 dibasic cleavage sites: effects on cleavage efficiency and polyprotein processing. *Virology* **192**:596–604.
 22. Lin, C., B. D. Lindenbach, B. Pragai, D. W. McCourt, and C. M. Rice. 1994. Processing of the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini. *J. Virol.* **68**:5063–5073.
 23. Lin, C., B. Pragai, A. Grakoui, J. Xu, and C. M. Rice. 1994. Hepatitis C virus NS3 serine proteinase: *trans*-cleavage requirements and processing kinetics. *J. Virol.*, in press.
 24. Lindenbach, B., S. M. Feinstone, and C. M. Rice. Unpublished data.
 25. Manabe, S., I. Fuke, O. Tanishita, C. Kaji, Y. Gomi, S. Yoshida, C. Mori, A. Takamizawa, I. Yoshida, and H. Okayama. 1994. Production of nonstructural proteins of hepatitis C virus requires a putative viral protease encoded by NS3. *Virology* **198**:636–644.
 26. Matsuura, Y., and T. Miyamura. 1993. The molecular biology of hepatitis C virus. *Semin. Virol.* **4**:297–304.
 27. Mizushima, H., M. Hijikata, Y. Tanji, K. Kimura, and K. Shimotohno. 1994. Analysis of N-terminal processing of hepatitis C virus nonstructural protein 2. *J. Virol.* **68**:2731–2734.
 28. Pizzi, E., A. Tramontano, L. Tomei, N. La Monica, C. Failla, M. Sardana, T. Wood, and R. De Francesco. 1994. Molecular-model of the specificity pocket of the hepatitis C virus protease: implications for substrate recognition. *Proc. Natl. Acad. Sci. USA* **91**:888–892.
 29. Rice, C. M., and J. H. Strauss. 1982. Association of Sindbis virion glycoproteins and their precursors. *J. Mol. Biol.* **154**:325–348.
 30. Tomei, L., C. Failla, E. Santolini, R. De Francesco, and N. La Monica. 1993. NS3 is a serine protease required for processing of hepatitis C virus polyprotein. *J. Virol.* **67**:4017–4026.
 31. Tsukiyama-Kohara, K., N. Iizuka, M. Kohara, and A. Nomoto. 1992. Internal ribosome entry site within hepatitis C virus RNA. *J. Virol.* **66**:1476–1483.
 32. Wang, C., P. Sarnow, and A. Siddiqui. 1993. Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism. *J. Virol.* **67**:3338–3344.
 33. Yoo, B. J., R. R. Spaete, A. P. Geballe, M. Selby, M. Houghton, and J. H. Han. 1992. 5' end-dependent translation initiation of hepatitis C viral RNA and the presence of putative positive and negative translational control elements within the 5' untranslated region. *Virology* **191**:889–899.