Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

VIROLOGY 187, 308–315 (1992)

Characterization of the Potyviral HC-Pro Autoproteolytic Cleavage Site

JAMES C. CARRINGTON[1] AND KERRI L. HERNDON

Department of Biology, Texas A&M University, College Station, Texas 77843

Received October 10, 1991; accepted November 25, 1991

The helper component-proteinase (HC-Pro) encoded by potyviruses functions to cleave the viral polyprotein by an autoproteolytic mechanism at the HC-Pro C-terminus. This protein belongs to a group of viral cysteine-type proteinases and has been shown previously to catalyze proteolysis between a Gly-Gly dipeptide. The amino acid sequence requirements surrounding the HC-Pro C-terminal cleavage site of the tobacco etch virus polyprotein have been investigated using site-directed mutagenesis and *in vitro* expression systems. A total of 51 polyprotein derivatives, each differing by the substitution of a single amino acid between the P5 and P2′ positions, were tested for autoproteolytic activity. Substitutions of Tyr (P4), Val (P2), Gly (P1), and Gly (P1′) were found to eliminate or nearly eliminate proteolysis. Substitutions of Thr (P5), Asn (P3), and Met (P2′), on the other hand, were permissive for proteolysis, although the apparent processing rates of some polyproteins containing these alterations were reduced. These results suggest that auto-recognition by HC-Pro involves the interaction of the enzymatic binding site with four amino acids surrounding the cleavage site. Comparison of the homologous sequences of five potyviral polyproteins revealed that the residues essential for processing are strictly conserved, whereas the nonessential residues are divergent. The relationship between HC-Pro and other viral and cellular cysteine-type proteinases is discussed. © 1992 Academic Press, Inc.

INTRODUCTION

The plant potyviruses belong to the 'picornavirus superfamily' of positive-strand RNA viruses (Mathews, 1991). The potyviruses represent the largest, and perhaps the most destructive, group of plant viral pathogens. Members of this group contain a genome of approximately 10-kilobases encoding a large polyprotein precursor (Fig. 1; Dougherty and Carrington, 1988). A protein termed VPg is attached covalently to the 5′ terminus of genomic RNA, while a poly(A) tail of variable length resides at the 3′ terminus (Hari *et al.*, 1979; Hari, 1981; Siaw *et al.*, 1985).

The activities of three viral proteinases are required for complete processing of the potyviral polyprotein. Cleavage at six positions, each found within the C-terminal two-thirds of the polyprotein, is catalyzed by the NIa proteinase (Carrington and Dougherty, 1987b; Carrington *et al.*, 1988; Garcia *et al.*, 1989; Hellmann *et al.*, 1988). This enzyme functions by auto- or *cis*-proteolysis at some sites and by a *trans*-proteolytic mechanism at others (Carrington and Dougherty, 1987a). The importance for substrate function of a heptapeptide sequence surrounding each of the NIa-mediated cleavage sites in the tobacco etch potyviral (TEV) polyprotein has been well-characterized (Dougherty *et al.*, 1988, 1989; Dougherty and Parks, 1989). Processing reactions near the N-terminus of the polyprotein are catalyzed by two additional proteinases, the helper component-proteinase (HC-Pro) and the "35-kDa protein" (Carrington *et al.*, 1989a; Verchot *et al.*, 1991). Like NIa, the proteolytically active domains of these proteins are found within the C-terminal halves of the respective molecules. The 35-kDa protein and HC-Pro each catalyze a single proteolytic reaction that occurs at their own C-terminal cleavage site. In addition, both processing reactions appear to involve an autoproteolytic mechanism (Carrington *et al.*, 1989b; J. Verchot and J.C.C., unpublished observations).

Based on a systematic analysis of conserved residues by site-directed mutagenesis, HC-Pro appears to resemble the cellular cysteine-type proteinases (Barrett, 1986). Mutations affecting Cys649 and His722 [numbering system beginning with the first position in the TEV polyprotein (Allison *et al.*, 1986)] were found to debilitate proteolytic activity (Oh and Carrington, 1989), suggesting that these residues may form part of the active site of the enzyme. Proteolytic processing in cell-free reactions, and presumably *in vivo*, occurs between Gly763 and Gly764 (Carrington *et al.*, 1989a). Interestingly, proteinases that resemble HC-Pro in terms of sequence similarity, putative active site composition, and cleavage site specificity are encoded by several other viruses and virus-like agents, including the animal alphaviruses and the hypovirulence-associated dsRNA of the chestnut blight fungus, *Cryphonectria parasitica* (Choi *et al.*, 1991a, b; Gorbalenya *et al.*, 1991; Hardy and Strauss, 1989; Koonin *et al.*, 1991; Shapira and Nuss, 1991; Strauss and Strauss, 1990).

[1] To whom reprint requests should be addressed.

0042-6822/92 $3.00
Copyright © 1992 by Academic Press, Inc.
All rights of reproduction in any form reserved.

In this study, we have analyzed the amino acid sequence requirements of the HC-Pro C-terminal cleavage site. The effects of 51 amino acid substitutions spanning seven residues surrounding and including the Gly-Gly cleavage dipeptide have been analyzed using a cell-free expression system. The results indicate that processing depends on the presence of four amino acid residues flanking the cleavage site.

MATERIALS AND METHODS

Plasmids, strains, and site-directed mutagenesis

Several recombinant plasmids were assembled using complementary DNA derived from the highly aphid transmissible strain of TEV (obtained from W. G. Dougherty, Oregon State University). The vectors used for these constructs (pTL7SN and pTL7SN.3) each contain a bacteriophage SP6 promoter for transcription *in vitro,* the TEV 5′ nontranslated region to facilitate efficient translation initiation of transcripts during cell-free protein synthesis, and a multiple cloning site for insertion of coding sequences that lack their own start codon (Carrington and Freed, 1990). Both vectors also contain the M13 IG region to permit production of single-stranded DNA in *Escherichia coli.* The vectors differ from one another in that pTL7SN.3 contains cDNA representing the TEV 3′ noncoding region and polyadenylate tail.

Complementary DNA corresponding to TEV nucleotides 1531–2681 was inserted into pTL7SN, generating pTL7SN-1527 (Fig. 1). Except for the presence of the M13 IG region, this plasmid is similar to pTL-1527 described previously (Carrington *et al.,* 1989a,b). The plasmid pTL7SN.3-0027 contains TEV cDNA representing nucleotides 1–2681 (Verchot *et al.,* 1991). The plasmids pTL7SN.3-0027/H10 and pTL7SN.3-0027/Pst were derived by site-directed mutagenesis of pTL7SN.3-0027. The former contains a mutation that converts the codon for His722 to one that specifies Ser, while the latter contains a replacement of the sequence AACCGA with CTGCAG (a *Pst*I site) at nucleotides 2440–2445.

The chloramphenicol acetyltransferase (CAT) gene (bounded by *Eco*RI restriction sites) was inserted into the *Eco*RI site of pTL7SN. The 5′ and 3′ ends of the CAT coding sequence were modified by site-directed mutagenesis to introduce *Pst*I restriction sites. The CAT coding sequence was then excised by *Pst*I digestion and inserted at the introduced *Pst*I site of pTL7SN.3-0027/Pst, forming pTL7SN.3-0024/CAT (Fig. 1). This plasmid encodes a polyprotein consisting of the 35-kDa protein, HC-Pro, the HC-Pro C-terminal cleavage site, and CAT.

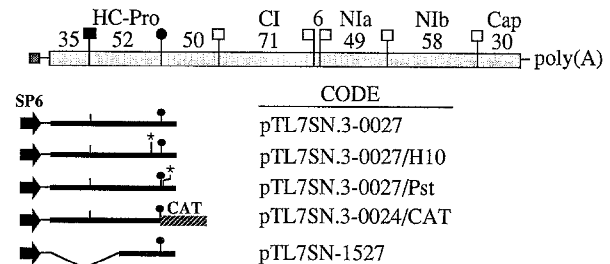

FIG. 1. Genetic map of tobacco etch virus (TEV), and relevant features of recombinant plasmids used in this study. (Top) The position of the single open reading frame in the TEV genome is indicated by the shaded rectangular box, while the genome-linked protein attached to the 5′ terminus is illustrated by the dark shaded box. Sequences encoding proteolytic cleavage sites are shown by vertical lines. Proteolysis at the cleavage sites is mediated by the NIa proteinase (open boxes), HC-Pro (solid circle), and the 35-kDa proteinase (solid box). The approximate molecular weights (in kilodaltons) of TEV-encoded proteins are shown above the genetic map; some proteins, such as NIa (Dougherty and Parks, 1991), may undergo further proteolytic processing. Abbreviations: HC-Pro, helper component-proteinase; CI, cylindrical inclusion protein; NIa, nuclear inclusion protein "a;" NIb, nuclear inclusion protein "b;" Cap, capsid protein; poly(A), polyadenylate tail. (Bottom) The TEV sequence represented as complementary DNA in each recombinant plasmid is shown to scale below the genetic map. Each plasmid contains a bacteriophage SP6 transcriptional promoter and the TEV 5′ nontranslated region. Asterisks indicate the positions of site-directed mutations. The gene for chloramphenicol acetyltransferase (CAT) was fused in-frame to the TEV sequence after nucleotide 2439, which is six nucleotides from the 5′ end of the "50-kDa protein" coding region. The HC-Pro-mediated cleavage site encoded by these plasmids is indicated by the vertical line attached to the solid circle.

Saturation mutagenesis using pTL7SN-1527 was targeted to codons for seven amino acid residues surrounding the HC-Pro C-terminal cleavage site. Oligonucleotides containing multiple codon possibilities for residues between the P5 and P2′ cleavage site positions were used for mutagenesis. [The nomenclature for cleavage site positions will follow that of Schechter and Berger (1967), where the positions to the N-terminal side of the scissile bond are indicated by P1, P2, etc., and the positions to the C-terminal side are indicated by P1′, P2′, etc.] Mutagenesis was conducted by the method of Kunkel *et al.,* (1987). Plasmids were screened by dideoxynucleotide sequence analysis. Mutagenized plasmids and the resulting polyproteins were named according to both the amino acid residue affected by the substitution and the identity of the substituting residue. For example, a mutagenized derivative of pTL7SN-1527 that results in a substitution of Gly with Ala at position 763 in the TEV polyprotein is given the designation G763A.

In vitro transcription and translation

Plasmids were linearized by digestion with *Pvu*II, which cleaves within vector sequences, and tran-

scribed using bacteriophage SP6 RNA polymerase as described (Carrington and Dougherty, 1987b; Melton *et al.*, 1984). Synthetic RNA transcripts were translated in a rabbit reticulocyte lysate (Promega) in the presence of [^{35}S]methionine (New England Nuclear).

Pulse-chase analysis of proteolytic processing

Pulse-chase analysis of cell-free translation products encoded by pTL7SN-1527 transcripts was conducted as described previously (Oh and Carrington, 1989). Briefly, transcripts were translated under normal conditions for 5 min in the presence of [^{35}S]-methionine, after which nonradiolabeled methionine (20 m*M*) was added to inhibit further radiolabel incorporation. Samples were removed to 5 vol protein dissociation buffer (Laemmli, 1970) at various times after addition of methionine and analyzed by SDS–polyacrylamide gel electrophoresis and autoradiography. Transcripts from pTL7SN-1527 encode a ~47-kDa polyprotein (with 21 methionine residues) composed of the HC-Pro proteolytic domain, the HC-Pro C-terminal cleavage site, and a short segment of the adjacent 50-kDa protein (Fig. 1). Processing of the polyprotein yields products of ~37 kDa (the HC-Pro fragment; 14 methionine residues) and 10 kDa. The levels of polyprotein precursor and HC-Pro-related cleavage product were assessed by video-based image analysis (densitometry) of autoradiographs. To normalize for differences in methionine content that would affect autoradiographic signal intensity, values for the cleavage product were multiplied by 1.5. Percentage proteolytic processing at each time point was determined by the following formula:

$$\frac{[\text{product}] \times 1.5}{[\text{precursor}] + [\text{product}] \times 1.5} \times 100.$$

RESULTS

Substitution of sequences downstream of the P2′ position of the HC-Pro C-terminal cleavage site

The transcripts from pTL7SN.3-0027 encode a ~97-kDa polyprotein that undergoes HC-Pro-mediated proteolytic processing between Gly763 and Gly764 to yield products of 87 and 10 kDa (Carrington *et al.*, 1989a; Fig. 2, lane 1). The 87-kDa product is stable in the rabbit reticulocyte system, whereas in the wheat germ extract it is processed further to yield the 35-kDa protein and HC-Pro (Verchot *et al.*, 1991). To determine if amino acid residues to the C-terminal side of the P2′ position around the HC-Pro cleavage site are necessary for HC-Pro activity or substrate recognition, we have analyzed the proteolytic activity of the pTL7SN.3-0027/Pst-derived polyprotein, which contains a substitution of the Asn766-Arg767 dipeptide with Leu-Gln at the P3′-P4′ positions. This change had little or no effect on formation of the 87-kDa proteolytic product *in vitro* (lane 3). For comparison, processing of the polyprotein derived from pTL7SN.3-0027/H10, which contains a mutation affecting His722 at the putative active site of the enzyme, was completely inhibited (lane 2).

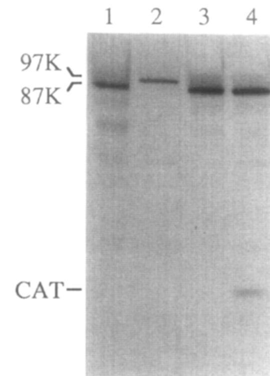

FIG. 2. Synthesis and processing of HC-Pro-containing polyproteins *in vitro*. Synthetic mRNA transcripts were translated in a rabbit reticulocyte lysate in the presence of [^{35}S]methionine. Radiolabeled products were separated by SDS–PAGE and detected by autoradiography. Translation products were encoded by transcripts from pTL7SN.3-0027 (lane 1), pTL7SN.3-0027/H10 (lane 2), pTL7SN.3-0027/Pst (lane 3), and pTL7SN.3-0024/CAT (lane 4). The 87-kDa product has been shown to result from HC-Pro-mediated proteolysis at its C-terminus and is a polyprotein consisting of the 35-kDa protein and HC-Pro (Carrington *et al.*, 1989a); further processing of this polyprotein by the 35-kDa proteinase is inhibited in the rabbit reticulocyte lysate system. The 97-kDa product represents a full-length, nonprocessed polyprotein. Note that the pTL7SN.3-0027/Pst- and pTL7SN.3-0024/CAT-derived polyproteins undergo proteolysis at the HC-Pro C-terminus.

The mutation that resulted in amino acid substitutions at the P3′ and P4′ positions also resulted in the incorporation of a *Pst*I restriction site. This site was utilized to replace the coding sequence for the N-terminal region of the "50-kDa protein" with the sequence encoding CAT. Cell free translation of pTL7SN.3-0024/CAT transcripts resulted in the accumulation of the 87-kDa product and CAT (Fig. 2, lane 4), indicating that proteolytic processing by HC-Pro was not affected. These results suggest that all sequences necessary for proteolytic activity reside to the N-terminal side of the P3′ position.

Amino acid substitutions between the P5 and P2′ positions of the HC-Pro C-terminal cleavage site

To examine the roles of specific amino acid residues surrounding the HC-Pro C-terminal cleavage site, polyproteins were synthesized from transcripts derived from pTL7SN-1527 containing mutations at codons for

P6	*P5*	*P4*	*P3*	*P2*	*P1* ↓	*P1′*	*P2′*	*P3′*	*P4′*
K	**T**	**Y**	**N**	**V**	**G**	**G**	**M**	**N**	**R**
	V	L	C	A	S	E	V		
	L	G	K	F	P	H	I		
	A	S	Y	R	V	A	Y		
	G	R	G	G	A	V	L		
	Y	C	I	I	D	I	S		
		F	D	L	F	M	D		
			L		K	L	K		
			F		M		R		
			V		T				
			H						

FIG. 3. Amino acid sequence and substitutions around the HC-Pro C-terminal cleavage site. The wild-type sequence between the P6 and P4′ positions is shown in boldface. The amino acid substitutions that were analyzed in this study are listed directly below the wild-type residues. Mutations were introduced into pTL7SN-1527 (see Fig. 1). Arrow indicates the scissile bond.

residues between the P5 to P2′ positions. The ~47-kDa polyprotein encoded by pTL7SN-1527 contains a ~37-kDa segment from the C-terminus of HC-Pro, the cleavage site, and a short segment from the 50-kDa protein; as demonstrated previously, HC-Pro-mediated proteolysis yields products of 37- and 10-kDa (Carrington *et al.,* 1989a). A total of 51 plasmids encoding polyprotein derivatives, with a minimum of five substitutions at each of the seven cleavage site positions, were generated (Fig. 3).

Representative examples of translation products derived from each of the seven sets of mutagenized plasmids are shown in Fig. 4. The amino acid substitutions at the P5 (Thr in wild type) and p2′ (Met) positions appeared to have relatively little effect on processing (compare to the nonmutagenized control lanes labeled wt). The precursor polyprotein encoded by each of the 13 plasmids with alterations affecting these two positions underwent proteolysis to generate the 37-kDa product. Similarly, nine of 10 polyproteins with substitutions at the P3 (Asn) position exhibited proteolytic processing competence similar to the wild-type polyprotein. Only one substitution, Asp for Asn, at the P3 site appeared to be significantly inhibitory to proteolysis [P3 (N), lane D].

In contrast to the effects of substitutions at the three sites described above, substitutions at the P4 (Tyr), P2 (Val), P1 (Gly), and P1′ (Gly) positions dramatically inhibited proteolysis (representative examples are shown in Fig. 4). Significant levels of processing occurred only with polyproteins containing Phe in place of Tyr and Leu in place of Val at the P4 and P2 sites, respectively, in which cases approximately 50% of the precursor was cleaved after the 60 min translation period [Fig. 4, P4 (Y) and P2 (V)]. Trace amounts of processing product were detected after synthesis of some altered polyproteins containing substitutions at the P2 and P1′ positions [Fig. 4, P2 (V) and P1′ (G)], although these minor levels of proteolysis were not observed consistently. Alterations affecting the P1 position appeared to completely inhibit processing of each of the nine polyproteins tested [Fig. 4, P1 (G)].

To determine the relative autoproteolytic processing rates of the various altered polyproteins, we have conducted a pulse-chase analysis in which cleavage reaction progress over time was monitored. *In vitro* synthesized proteins were pulse labeled during the initial 5 min of the translation reaction and then chased for up to 160 min. The time courses for protein processing of the wild-type polyprotein, as well as five altered polyproteins that exhibited varying degrees of proteolytic efficiency, are shown in Fig. 5. Full-size, radiolabeled products were first detected 7.5 min after translation initiation (5 min pulse plus 2.5 min chase). Approximately 50% of the wild-type polyprotein was found to process within the first 2.5 min of the chase period (Fig. 5A). Several of the mutant-derived polyproteins with substitutions at the P5, P3, and P2′ positions, exemplified by the M765L polyprotein, were shown to process with rates comparable to the wild-type polyprotein (Fig. 5B). Several other polyproteins were found to cleave at considerably slower rates (panels C-F).

The processing time-courses for wild-type and all 51 altered polyproteins were quantitated by video image densitometry of autoradiographs. Relative to the polyprotein encoded by the wild-type sequence, polyproteins containing substitutions at the P5, P3, and P2′

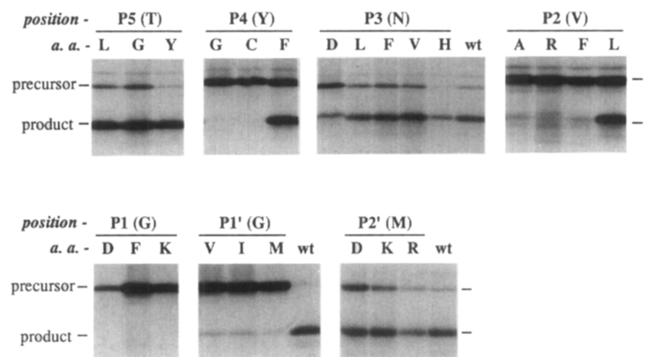

FIG. 4. Cell-free synthesis and processing of polyproteins containing wild-type and altered HC-Pro C-terminal cleavage sites. Polyproteins were synthesized using pTL7SN-1527 (mutagenized) transcripts and analyzed as described under Materials and Methods and in the Fig. 2 legend. Wild-type transcripts from pTL7SN-1527 encode a 47-kDa polyprotein precursor that can undergo proteolysis to yield a 37-kDa product. The cleavage site position, and the wild-type residue at that position (in parentheses), are shown at the top of each panel. The amino acid substitution is given above each lane. Data are shown for representative examples of each set of polyproteins. The single-letter code for amino acids is used. Abbreviations: a.a., amino acid; wt, wild-type.

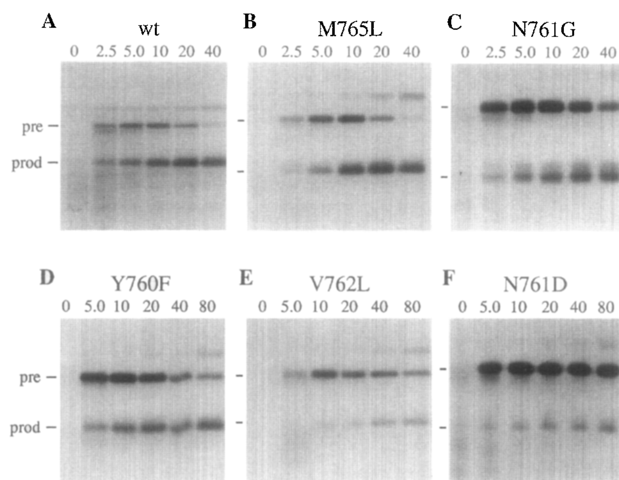

FIG. 5. Pulse-chase analysis of polyproteins containing substitutions around the HC-Pro C-terminal cleavage site. [^{35}S]Methionine labeling of cell-free translation products was conducted for 5 min, after which excess nonlabeled methionine was added. Aliquots were removed at the times indicated (in min) after initiation of the chase and analyzed by SDS–PAGE and autoradiography. The polyprotein designation, which indicates the nature of the amino acid substitution around the cleavage site (see Materials and Methods), is given above each panel. Proteolytic processing is shown for wild-type and five "mutagenized" polyproteins that exhibited a range of processing efficiencies.

positions underwent proteolysis at slightly slower rates, although the final levels of product after the 40-min chase generally were within 25% of the wild-type product (Figs. 6A, 6C, and 6G). As also indicated in Figs. 4 and 5, proteolysis of a few polyproteins (such as N761D) containing exchanges at these three positions was impaired more dramatically. The processing time courses for polyproteins altered at the P4, P2, P1, and P1′ positions confirmed that most of these polyproteins were inactive. The Y760F and V762L polyproteins displayed proteolytic competence whereby approximately 75% of each precursor was processed after 160 min; the apparent cleavage rates, however, were much slower than wild type (Figs. 5D and 5E; Figs. 6B and 6D). Some polyproteins affected at the P1′ position were marginally active (such as G764M and G764A), with approximately 10% product formation after the 160-min incubation (Fig. 6F). No processing was measured at any time point using polyproteins containing P1 substitutions (Fig. 6E).

DISCUSSION

We have employed an *in vitro* expression system in combination with site-directed mutagenesis to investigate the roles of amino acid residues surrounding the C-terminal cleavage site of HC-Pro during autoproteolysis. Four residues (occupying the P4, P2, P1, and P1′ positions) were found to be crucial for proteolytic function. On the other hand, amino acid residues at the P5, P3, and P2′ positions were found to play relatively minor roles. All substitutions at these sites, including those that involved exchange of chemically nonconservative residues, permitted proteolysis although generally at slightly reduced rates. As revealed by insertion of chloramphenicol acetyltransferase, protein sequence to the C-terminal side of the P2′ position also appeared to have little effect on processing. Significantly, the four amino acid residues defined as critical in this study are strictly conserved among the homologous cleavage sites from polyproteins encoded by five potyviruses, while each of the noncritical residues occupies a variable position (Fig. 7). We propose that the sequence Tyr-Xaa-Val-Gly-Gly between positions P4 and P1′ constitutes an essential, highly conserved motif required for HC-Pro-mediated cleavage site function in potyviral polyproteins.

The majority of the mutations affecting Tyr at the P4 position and Val at P2 eliminated most or all proteolytic processing. However, conversion of Tyr (P4) to Phe, a structurally conservative substitution, was tolerated. This suggests that the presence of Tyr in the wild-type sequence reflects the necessity of an R group with an aromatic ring in the P4 position. Similarly, polyproteins containing Leu rather than Val in the P2 position retained a significant level of proteolytic activity, indicating the possible requirement for a residue containing a hydrophobic, branched side chain. The intolerance of any substitutions at the P1 and P1′ positions suggests a stringent requirement for a Gly-Gly dipeptide framing the scissile bond.

Due to the presence of essential Cys and His residues within the proteolytic domain, HC-Pro has been postulated to belong to the cysteine proteinase superfamily (Oh and Carrington, 1989), which has as members the well-characterized enzymes papain and actinidin. Using synthetic peptides, papain has been demonstrated to contain an active site that can accommodate or interact with substrate residues between the P4 and P3′ positions and to bind preferentially to substrates containing Phe in the P2 position (Schechter and Berger, 1967, 1968). Our data support the notion of an extended substrate recognition pocket within HC-Pro, but they also suggest that residues in addition to that in the P2 position participate in cleavage site/active site interactions. The most straightforward interpretation of our results is that residues at the P4, P2, P1, and P1′ positions interact with residues at subsites [termed S4, S2, S1, and S1′ according to Schechter and Berger (1967)] within a binding pocket near the active center of the molecule. Alternatively, one or more of the residues occupying the four "es-

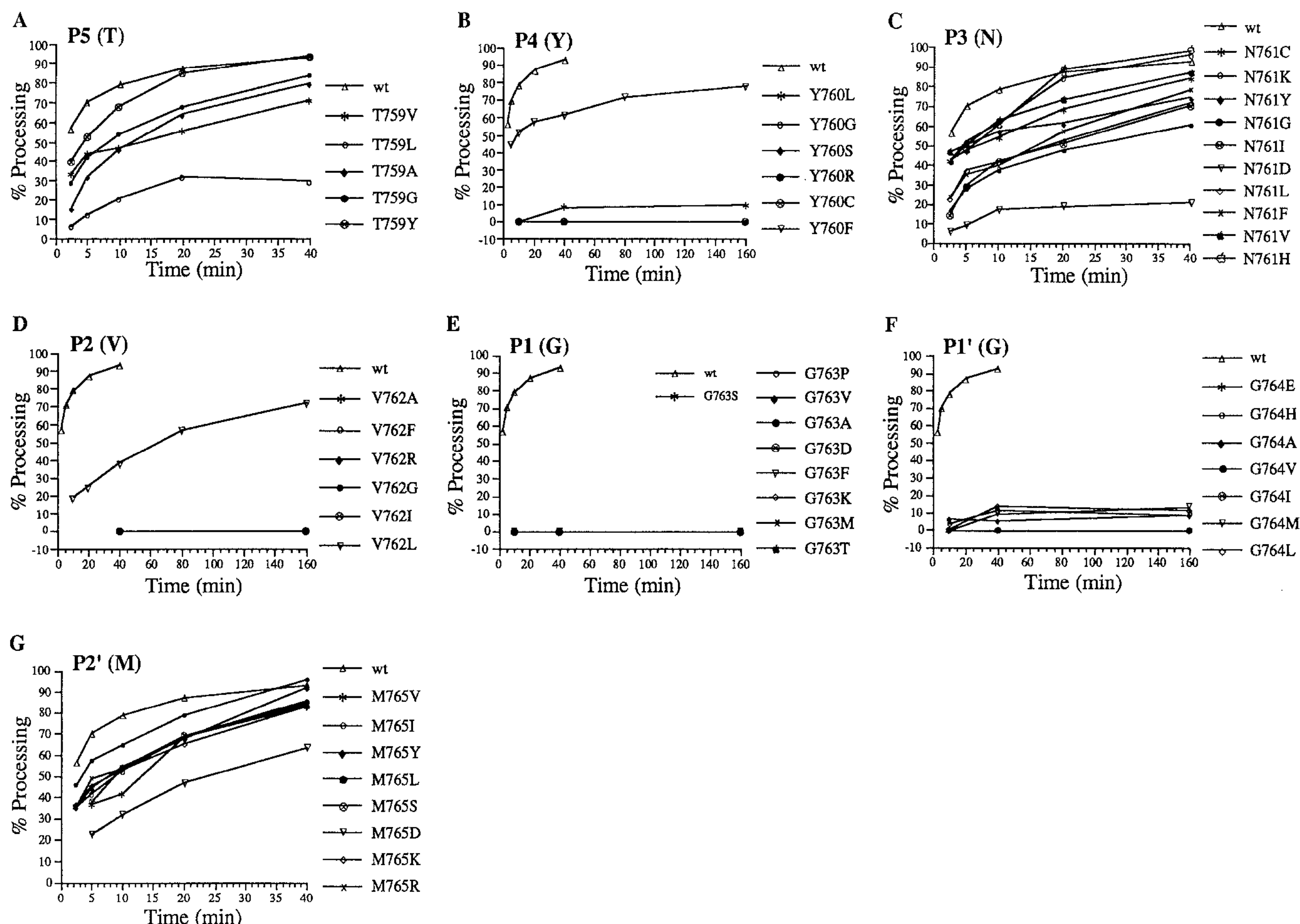

FIG. 6. Pulse-chase analysis of polyproteins containing substitutions around the HC-Pro C-terminal cleavage site. A processing time-course analysis was conducted for each of the modified polyproteins as described in the text and in the Fig. 5 legend. Percentage processing at each time point was determined by video densitometry of autoradiographs. Due to no detectable proteolytic processing of most of the polyproteins containing substitutions at the P4, P2, P1, and P1′ positions, several of the data points in (B, D, E, and F) are superimposed.

sential" positions may be involved in proper folding of the enzyme or secondary structure formation around the cleavage site. Both possibilities should be considered seriously in view of the fact that the P4-P1 residues constitute the C-terminus of the proteolytic enzyme itself, and that proteolysis at this site appears to proceed through an autolytic mechanism.

	P6	*P5*	*P4*	*P3*	*P2*	*P1*	*P1'*	*P2'*	*P3'*
TEV	**K**	**T**	**Y**	N	**V**	**G**	**G**	M	N
PPV	**K**	**T**	**Y**	L	**V**	**G**	**G**	S	E
PVY	**K**	H	**Y**	**R**	**V**	**G**	**G**	V	P
TVMV	A	Q	**Y**	K	**V**	**G**	**G**	L	V
PSBMV	**K**	M	**Y**	**R**	**V**	**G**	**G**	T	Q

FIG. 7. Comparison of HC-Pro C-terminal cleavage sites from polyproteins encoded by five potyviruses. Identical residues at each position are indicated in boldface. The single-letter amino acid designation is used. Abbreviations and published references for sequences: TEV, tobacco etch virus (Allison *et al.*, 1986); PPV, plum pox virus (Maiss *et al.*, 1989); PVY, potato virus Y (Robaglia *et al.*, 1989); TVMV, tobacco vein mottling virus (Domier *et al.*, 1986); PSBMV, pea seedborne mosaic virus (Johansen, 1990).

The family of viral cysteine-type proteinases to which HC-Pro appears to belong includes the nsP2 protein of alphaviruses (Sindbis virus) and p29 and p48 encoded by the dsRNA genetic element associated with hypovirulence of *Cryphonectria parasitica* strain EP713 (see Introduction). Analyses of cleavage sites processed by these proteinases now provides further confirmation of their relatedness. The cleavage sites recognized by HC-Pro, p29, and p48 are characterized by dipeptides containing small, noncharged R groups (Gly and Ala) in the P1 and P1′ positions (Carrington *et al.*, 1989a; Choi *et al.*, 1991b; Shapira and Nuss, 1991). With the exception of the site between the nsP3 and nsP4 proteins, which contains Tyr in the P1′ position, the nsP2-mediated cleavage sites in the alphavirus polyprotein are composed of Gly, Ala, and Cys residues (Strauss and Strauss, 1990). For cleavage sites recognized by

HC-Pro, p48, and nsP2, the residues occupying the P2 positions have been shown to be necessary for substrate function (Shapira and Nuss, 1991; Shirako and Strauss, 1990; this study). Like the cellular cysteine-type proteinases, the interaction of residues at the substrate P2 and binding pocket S2 sites appears to play a dominant role in promoting the processing reaction. In addition, the P4 positions of both HC-Pro and p48 cleavage sites are critical (Shapira and Nuss, 1991; this study).

Information concerning the cleavage site requirements for viral proteinases such as HC-Pro will be useful for the development of potential inhibitors that may have antiviral activity. One can envision transgenic plants that produce peptide competitors that reduce the activity of viral proteinases. The possibility also exists to modify known proteinase inhibitors to facilitate binding to the viral enzymes. In addition, knowledge of essential cleavage site features will be necessary for the identification of residues and structures that facilitate specificity within the binding site region of these proteinases.

ACKNOWLEDGMENTS

We thank Valerian Dolja, Deon Freed, Xiao Hua Li, María Restrepo-Hartwig, and Jeanmarie Verchot for their helpful comments on this manuscript; Larry Harris-Haller for synthesis of oligonucleotides; Vince Cassone for assistance with the video image analysis of autoradiographs; and members of the Friday Institute for their advice. This work was supported in part by a USDA Competitive Grant (91-37303-6435) and a Public Health Service Grant (AI27842) from the National Institute of Allergy and Infectious Disease.

REFERENCES

ALLISON, R. F., JOHNSTON, R. E., and DOUGHERTY, W. G. (1986). The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: Evidence for the synthesis of a single polyprotein. *Virology* **154,** 9–20.

BARRETT, A. J. (1986). An introduction to the proteinases. *In* "Proteinase Inhibitors" (A. J. Barrett and G. Salvesen, Eds.), pp. 3–22. Elsevier, New York.

CARRINGTON, J. C., CARY, S. M., and DOUGHERTY, W. G. (1988). Mutational analysis of tobacco etch virus polyprotein processing: cis and trans proteolytic activity of polyproteins containing the 49-kilodalton proteinase. *J. Virol.* **62,** 2313–2320.

CARRINGTON, J. C., and DOUGHERTY, W. G. (1987a). Processing of the tobacco etch virus 49K protease requires autoproteolysis. *Virology* **160,** 355–362.

CARRINGTON, J. C., and DOUGHERTY, W. G. (1987b). Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease. *J. Virol.* **61,** 2540–2548.

CARRINGTON, J. C. and FREED, D. D. (1990). Cap-independent enhancement of translation by a plant potyvirus 5′ nontranslated region. *J. Virol.* **64,** 1590–1597.

CARRINGTON, J. C., FREED, D. D., and OH, C.-S. (1990). Expression of potyviral polyproteins in transgenic plants reveals three proteolytic activities required for complete processing. *EMBO J.* **9,** 1347–1353.

CARRINGTON, J. C., CARY, S. M., PARKS, T. D., and DOUGHERTY, W. G. (1989a). A second proteinase encoded by a plant potyviral genome. *EMBO J.* **8,** 365–370.

CARRINGTON, J. C., FREED, D. D., and SANDERS, T. C. (1989b). Autocatalytic processing of the potyvirus helper component proteinase in *Escherichia coli* and in vitro. *J. Virol.* **63,** 4459–4463.

CHOI, G. H., PAWLYK, D. M., and NUSS, D. L. (1991a). The autocatalytic protease p29 encoded by the hypovirulence-associated virus of the chestnut blight fungus resembles the potyvirus-encoded protease HC-Pro. *Virology* **183,** 747–752.

CHOI, G. H., SHAPIRA, R., and NUSS, D. L. (1991b). Cotranslational autoproteolysis involved in gene expression from a double-stranded RNA genetic element associated with hypovirulence of the chestnut blight fungus. *Proc. Natl. Acad. Sci. USA* **88,** 1167–1171.

DOMIER, L. L., FRANKLIN, K. M., SHAHABUDDIN, M., HELLMANN, G. M., OVERMEYER, J. H., HIREMATH, S. T., SIAW, M. F. E., LOMONOSSOFF, G. P., SHAW, J. G., and RHOADS, R. E. (1986). The nucleotide sequence of tobacco vein mottling virus RNA. *Nucleic Acids Res.* **14,** 5417–5430.

DOUGHERTY, W. G. and CARRINGTON, J. C. (1988). Expression and function of potyviral gene products. *Annu. Rev. Phytopathol.* **26,** 123–143.

DOUGHERTY, W. G., CARRINGTON, J. C., CARY, S. M., and PARKS, T. D. (1988). Biochemical and mutational analysis of a plant virus polyprotein cleavage site. *EMBO J.* **7,** 1281–1287.

DOUGHERTY, W. G., CARY, S. M., and PARKS, T. D. (1989). Molecular genetic analysis of a plant virus polyprotein cleavage site: A model. *Virology* **171,** 356–364.

DOUGHERTY, W. G., and PARKS, T. D. (1989). Molecular genetic and biochemical evidence for the involvement of the heptapeptide cleavage sequence in determining the reaction profile at two tobacco etch virus cleavage sites in cell-free assays. *Virology* **172,** 145–155.

DOUGHERTY, W. G., and PARKS, T. D. (1991). Post-translational processing of the tobacco etch virus 49-kDa small nuclear inclusion polyprotein: Identification of an internal cleavage site and delimitation of VPg and proteinase domains. *Virology* **183,** 449–456.

GARCIA, J. A., RIECHMANN, J. L., and LAIN, S. (1989). Proteolytic activity of the plum pox potyvirus NIa-like protein in *Escherichia coli. Virology* **170,** 362–369.

GORBALENYA, A. E., KOONIN, E. V., and LAI, M. M.-C. (1991). Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, α- and coronaviruses. *FEBS Lett.* **288,** 201–205.

HARDY, W. R., and STRAUSS, J. H. (1989). Processing the nonstructural proteins of Sindbis virus: Nonstructural proteinase is in the C-terminal half of nsP2 and functions both in *cis* and *trans. J. Virol.* **63,** 4653–4664.

HARI, V. (1981). The RNA of tobacco etch virus: further characterization and detection of protein linked to RNA. *Virology* **112,** 391–399.

HARI, V., SIEGLE, A., ROZEK, C., and TIMBERLAKE, W. E. (1979). The RNA of tobacco etch virus contains poly(A). *Virology* **92,** 568–571.

HELLMANN, G. M., SHAW, J. G., and RHOADS, R. E. (1988). *In vitro* analysis of tobacco vein mottling virus NIa cistron: Evidence for a virus-encoded protease. *Virology* **163,** 554–562.

JOHANSEN, E. (1990). Molekylaerbiologiske Metoder til Plantebeskyttelse. Diagnostik og Virusresistens. Ph. D. Thesis, Institut for Okologi og Molekylaer Biologi Den Kongelige Veterinaer og Landbohojskole, Copenhagen, Denmark.

KOONIN, E. V., CHOI, G. H., NUSS, D. L., SHAPIRA, R., and CARRINGTON, J. C. (1991). Evidence for common ancestry of a chestnut blight

hypovirulence-associated double-stranded RNA and a group of positive-strand RNA plant viruses. *Proc. Natl. Acad. Sci. USA* **88,** 10647–10651.

KUNKEL, T. A., ROBERTS, J. D., and ZAKOUR, R. A. (1987). Rapid and efficient site-directed mutagenesis without phenotypic selection. *In* "Methods in Enzymology" (R. Wu and L. Grossman, Eds.), Vol. 154, pp. 367–382. Academic Press, San Diego.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature (London)* **227,** 680–685.

MAISS, E., TIMPE, U., BRISSKE, A., JELKMANN, W., CASPER, R., HIMMLER, G., MATTANOVICH, D. and KATINGER, H. W. D. (1989). The complete nucleotide sequence of plum pox virus RNA. *J. Gen. Virol.* **70,** 513–524.

MATHEWS, R. E. F. (1991). "Plant Virology," (3rd ed.). Academic Press, San Diego.

MELTON, D. A., KRIEG, P. A., REBAGLIATI, M. R., MANIATIS, T., ZINN, K. and GREEN, M. R. (1984). Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12,** 7035–7056.

OH, C.-S., and CARRINGTON, J. C. (1989). Identification of essential residues in potyvirus proteinase HC-Pro by site-directed mutagenesis. *Virology* **173,** 692–699.

ROBAGLIA, C., DURAND-TARDIFF, M., TRONCHET, M., BOUDAZIN, G., ASTIER-MANIFACIER, S., CASSE-DELBART, F. (1989). Nucleotide sequence of potato virus Y (N strain) genomic RNA. *J. Gen. Virol.* **70,** 935–947.

SCHECHTER, I., and BERGER, A. (1967). On the active site of proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27,** 157–162.

SCHECHTER, I., and BERGER, A. (1968). On the active site of proteases. III. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochem. Biophys. Res. Commun.* **32,** 898–902.

SHAPIRA, R., and NUSS, D. L. (1991). Gene expression by a hypovirulence-associated virus of the chestnut blight fungus involves two papain-like protease activities. *J. Biol. Chem.* **266,** 19419–19425.

SHIRAKO, Y., and STRAUSS, J. H. (1990). Cleavage between nsP1 and nsP2 initiates the processing pathway of Sindbis virus nonstructural polyprotein P123. *Virology* **177,** 54–64.

SIAW, M. F. E., SHAHABUDDIN, M., BALLARD, S., SHAW, J. G., and RHOADS, R. E. (1985). Identification of a protein covalently linked to the 5′ terminus of tobacco vein mottling virus RNA. *Virology* **142,** 134–143.

STRAUSS, J. H. (1990). Introduction: Viral proteinases. *Semin. Virol.* **1,** 307–309.

STRAUSS, J. H., and STRAUSS, E. G. (1990). Alphavirus proteinases. *Semin. Virol.* **1,** 347–356.

VERCHOT, J., KOONIN, E. V., and CARRINGTON, J. C. (1991). The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. *Virology* **185,** 527–535.