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.

VIROLOGY 173, 692–699 (1989)

Identification of Essential Residues in Potyvirus Proteinase HC-Pro by Site-Directed Mutagenesis

CHAN-SEOK OH AND JAMES C. CARRINGTON[1]

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

Received July 13, 1989; accepted August 24, 1989

Two virus-encoded proteinases are responsible for proteolysis of potyvirus polyproteins. One of these, HC-Pro, is a multifunctional protein that autolytically cleaves at its carboxyl-terminus (J. C. Carrington *et al.*, 1989, *EMBO J.* 8, 365–370). To identify the class of proteinase to which HC-Pro belongs, tobacco etch virus (TEV) HC-Pro mutants containing single amino acid substitutions at serine, cysteine, aspartic acid, and histidine positions were synthesized by *in vitro* transcription and translation and were tested for autoproteolytic activity. Combinations of these residues are constituents of the active sites of diverse groups of cellular and viral proteinases. Only those positions that were strictly conserved among four potyvirus HC-Pro proteolytic domains (for which sequences have been deduced) were mutagenized. Of the 19 mutant proteinases synthesized and tested, only those with alterations at Cys-649 and His-722 were defective for HC-Pro autolytic activity. Most of the other mutant proteinases exhibited no impairments in processing kinetics experiments. The spectrum of essential residues, as defined by this genetic analysis, supports the hypothesis that HC-Pro most closely resembles members of the cysteine-type family of proteinases. © 1989 Academic Press, Inc.

INTRODUCTION

Tobacco etch virus (TEV), a member of the potyvirus group, possesses a single-stranded RNA genome (10 kb) that encodes a large polyprotein (Allison *et al.*, 1986). All TEV-derived proteins arise by proteolytic cleavage of the polyprotein precursor (Allison *et al.*, 1985, 1986; Dougherty and Carrington, 1988). The activities of two TEV-encoded proteins, HC-Pro and 49-kDa (or NIa) proteinase, are responsible for most or all proteolytic events associated with polyprotein processing (Carrington and Dougherty, 1987a,b; Carrington *et al.*, 1988, 1989a). The TEV 49-kDa proteinase catalyzes cleavage at five positions, all of which reside within the carboxyl-terminal two-thirds of the polyprotein. The proteolytic domain of the 49-kDa proteinase is highly conserved in structure relative to the analogous NIa proteinases from other potyviruses, to the 24-kDa proteinase of cowpea mosaic virus (CpMV), as well as to the 3C proteinase of picornaviruses (Allison *et al.*, 1986; Domier *et al.*, 1987; Krausslich and Wimmer, 1988). Compelling evidence has been presented that shows the NIa, CpMV 24-kDa, and picornaviral 3C proteinases originated from the trypsin superfamily of serine proteinases by an evolutionary pathway that included conversion of the active-site serine to cysteine (Bazan and Fletterick, 1988; Gorbalenya *et al.*, 1989b).

The second potyvirus proteinase, HC-Pro, appears to catalyze only one proteolytic reaction, that being an autolytic cleavage at its carboxyl-terminus (Carrington *et al.*, 1989a,b). This protein resides near the amino-terminus of the TEV polyprotein, and cleaves between a Gly–Gly dipeptide located at amino acid positions 763–764 in the TEV polyprotein. Processing of the TEV polyprotein by HC-Pro liberates an 87-kDa precursor that is further cleaved *in vivo* to products of 31- and 56-kDa (the latter being HC-Pro) (Thornbury *et al.*, 1985; de Mejia *et al.* 1985). The identity of the proteinase that processes the amino-terminus of HC-Pro has not been identified. In addition to its role as a proteolytic enzyme, HC-Pro functions as a helper component (HC) for virus transmission from plant-to-plant in nature (Thornbury *et al.*, 1985). A limited degree of amino acid sequence similarity has been noted between the amino-terminal region of HC-Pro of potyviruses and the helper component encoded by the cauliflower mosaic virus genome (Domier *et al.*, 1987). Since only the carboxyl-terminal region of HC-Pro is required for proteolytic activity (Carrington *et al.*, 1989a), HC-Pro appears to possess a two-domain structure.

Most proteinases can be grouped into one of four classes according to the chemically reactive groups at their active site: Ser, Cys, Asp, or Zn^{2+} (Neurath, 1984; Barrett, 1986). In most cases, amino acid sequences surrounding the active-site residues, as well as the spatial relationship of active-site component residues to one another, are highly conserved among members of a given class. In theory, therefore, if a proteinase belongs to one of the four classes, its class assignment may be made by generating mutants with substitutions

[1] To whom requests for reprints should be addressed.

0042-6822/89 $3.00

of amino acid residues known to function in proteinases of that class. Serine proteinases would require at least one Ser, His, and Asp residue. Typical cysteine proteinases would show a requirement for at least one Cys and one His residue (Polgar and Halasz, 1982). Aspartic acid proteinases would exhibit a requirement for at least two Asp residues (Tang *et al.*, 1973), while the zinc-dependent endoproteinases would require at least two His residues (Kester and Mathews, 1977; Jongeneel *et al.*, 1989). Additionally, the viral cysteine proteinases should exhibit a requirement for at least one Cys, His, and Asp residue (Bazan and Fletterick, 1988).

In this article, we have sought to identify the nature of the proteolytic activity of HC-Pro by selective site-directed mutagenesis. We rationalized that the active-site residues within HC-Pro from several potyviruses would be invariant and spacially conserved. Therefore, the amino acid sequences of the proteolytic domain of HC-Pro encoded by four potyviruses were aligned to identify regions that contained strictly conserved Ser, Cys, His, or Asp residues (Fig. 1). These positions were mutagenized in a TEV cDNA construct, and the modified proteinases were expressed and tested for activity using cell-free transcription and translation systems. On the basis of this genetic analysis, we propose that HC-Pro exhibits greatest similarity to well-established members of the cysteine class of proteolytic enzymes.

MATERIALS AND METHODS

Recombinant plasmids and strains

Complementary DNA representing the HC-Pro coding region (from the highly aphid-transmissible strain of TEV) was inserted into the cell-free expression vectors pTL7SN and pTL7S. These vectors resemble pTL-8 (Carrington and Dougherty, 1987b) except that the M13 intergenic region has been added upstream of the bacteriophage SP6 promoter to facilitate production of single-stranded DNA with the aid of helper phage M13K07. A key feature of these vectors is the presence of the TEV 5′-nontranslated region and translational start sequence. DNA sequences lacking their own start codon can be inserted and expressed *in vitro* by cell-free transcription and translation. The difference between pTL7SN and pTL7S is that the former contains a C to A mutation at TEV nucleotide 144 to generate a *Nco*I site around the start codon.

Plasmid pTL7SN-1527 contains an insertion of a *Nde*I–*Hpa*I TEV cDNA fragment (nucleotides 1531–2681) while pTL7S-2027 contains a *Ssp*I–*Hpa*I cDNA fragment (nucleotides 1966–2681; Fig. 2). Similar plasmids have been constructed and described previously (Carrington *et al.*, 1989a). *Escherichia coli* strain TG1 was used as host for all DNA cloning.

In vitro mutagenesis

Nineteen point-mutations were introduced into pTL-7SN-1527 or pTL7S-2027. Oligonucleotides specifying single- or double-base changes were synthesized using a Beckman System 1 Plus DNA Synthesizer. Oligonucleotide-directed mutagenesis was conducted by the procedure of Taylor *et al.* (1985a,b) using reagents purchased from Amersham. Single-stranded DNA production prior to mutagenesis was induced by transfection with the helper phage M13K07 (Vieira and Messing, 1987). Putative mutagenized plasmids were screened by nucleotide sequence analysis (Sanger *et al.*, 1977).

Cell-free transcription, translation, and processing

Transcription with bacteriophage SP6 polymerase and translation in rabbit reticulocyte lysates were conducted using established methods (Carrington and Dougherty, 1987b; Dougherty and Hiebert, 1980; Melton *et al.*, 1984). Plasmids were linearized with *Pvu*II, which cleaves downstream of the TEV insert sequences, prior to transcription. Translation reactions were carried out in the presence of [^{35}S]methionine (>1000 Ci/mmol) for 40 min and terminated by the addition of RNase A (200 μg/ml) and protein dissociation buffer. Translation and processing products were analyzed by SDS–PAGE (Laemmli, 1970) and autoradiography using Kodak XAR film.

Pulse–chase labeling of proteins

Translation products encoded by SP6 transcripts were pulse labeled with [^{35}S]methionine for 7 min, followed by a chase with excess (15 m*M*) nonlabeled methionine. Aliquots were removed after chase periods of 2, 4, 8, 16, and 32 min. Reactions were terminated by the addition of protein dissociation buffer (Laemmli, 1970) and analyzed by SDS–PAGE and autoradiography. Autoradiographs were scanned using a Model 620 Video Densitometer (Bio-Rad Laboratories, Inc.) to measure the intensity of signals representing polyprotein precursor and product. Percentage processing was determined by dividing the quantity of product by the total quantity of precursor plus product.

RESULTS

Conserved amino acid sequences among four potyvirus HC-Pro proteolytic domains

We have employed a genetic approach to identify the class of proteinase to which potyvirus HC-Pro belongs.

```
                 S        H     S D   D                C
                 1        8     215   16               12
TEV      609     YSDLKHPTKRHLVIGNSGDSKYLDLPVLNEEKMYIANEGYC     649
TVMV     559     YSELKMPTKNHLVIGNSGDPKYLDLPGEISNLMYIAKEGYC     599
PPV      612     ESEIISPTKNHLVVGNSGDSKYVDLPTAKGGAMFIAKAGYC     652
PVY      586     ESTFYPPTKKHLVIGNSGDQKFVDLPKGDSEMLYIAKQGYC     626
                  *    *** *** ***** *  ***        **  ***
                                        D             D
                                       17            18
TEV      650     YMNIFFALLVNVKEEDAKDFTKFIRDTIVPKLGAWPTMQDV     690
TVMV     600     YINIFLAMLVNVDEANAKDFTKRVRDESVQKLGKWPSLIDV     640
PPV      653     YINIFLAMLININEDEAKSFTKTVRDTLVPKLGTWPSMMDL     693
PVY      627     YINVFLAMLINISEEDAKDFTKKVRDMCVPKLGTWPTMMDL     667
                 * * * * * *  *  ** ***  **  * *** **   *
                   C   S               DH     H  DS  S
                  13   3              199    10 204  5
TEV      691     ATACYLLSILYPDVLSAELPRILVDHDNKTMHVLDSYGSRT     731
TVMV     641     ATECALLSTYYPAAASAELPRLLVDHAQKTIHVVDSYGSLN     681
PPV      694     ATACHFLAVLYPETRNAELPRILVDHEAKIFHVVDSFGSLS     734
PVY      668     ATTCAQMRIFYPDVHDAELPRILVDHDTQTCHVVDSFGSQT     708
                 ** *      **    ***** ****     ** ** **
                   H       S           S
                  11       6           7
TEV      732     TGYHMLKMNTTSQLIEFVHSGLESEMKTYNVG G            764
TVMV     682     TGYHILKANTVSQLEKFASNTLESPMAQYKVG G            714
PPV      735     TGMHVLKANTINQLISFASDTLDSNMKTYLVG G            767
PVY      709     TGYHILKASSVSQLILFANDELESDIKHYRVG G            741
                 ** * **     **  *    * *    * ** *
```

FIG. 1. Alignment of the amino acid sequences of the HC-Pro proteolytically active domain from tobacco etch virus (TEV) (Allison *et al.*, 1986), tobacco vein mosaic virus (TVMV) (Domier *et al.*, 1986), plum pox virus (PPV) (Maiss *et al.*, 1989), and potato virus Y (PVY) (Robaglia *et al.*, 1989). All conserved residues are shown by an asterisk (*). The codons for the amino acid residues shown in boldface type were mutagenized in a TEV cDNA clone. Each mutation was given a code name (S1, S2, etc.) as shown in Table 1; these codes are shown directly above each amino acid residue that was mutagenized. The arrow shows the position of the HC-Pro carboxyl-terminal cleavage site that is processed *in vitro* (Carrington *et al.*, 1989a).

By identifying essential Ser, His, Cys, or Asp residues within the proteinase, and assuming that HC-Pro functions by a mechanism described for one of the four major proteinase classes, the catalytic nature of the enzyme may be deduced. Working on the assumption that HC-Pro from several potyviruses will possess a highly conserved three-dimensional structure and catalytic mechanism, we aligned the sequences of the proteolytic domain of HC-Pro of TEV, tobacco vein mottling virus (TVMV), potato virus Y (PVY), and plum pox virus (PPV) to identify those positions with invariant residues among all four viruses (Fig. 1).

The proteolytic domains of each virus were highly conserved in sequence; nearly half of the positions (75 of 155) were invariant, and no insertions or deletions were detected. The TEV sequence was most similar to that of TVMV (66%), although extensive similarity was evident with PVY (62%) and PPV (60%). Given that the entire HC-Pro sequences of TEV and TVMV display 54% similarity (Domier *et al.*, 1987), the proteolytic domain (carboxyl-terminal) appears to be more conserved than the helper component domain (amino-terminal). The Gly–Gly cleavage site recognized by TEV HC-Pro *in vitro* is constant in each viral sequence, allowing the prediction that the carboxyl-terminus of HC-Pro will originate from position 713 in the TVMV polyprotein, 766 in PPV, and 740 in PVY. Of the 29 TEV positions occupied by Ser, His, Cys, or Asp, only 17 were conserved among all four viruses.

Mutagenesis of conserved amino acid residues

Oligonucleotides were synthesized to generate amino acid substitutions at the 17 conserved positions containing Ser, His, Cys, or Asp, as well as 2 nonconserved Ser positions (designated S3 and S6; Table 1). Except for His positions, substitutions that maintained the chemical nature of the original amino acid residue were selected, since these would be predicted to disrupt less the overall structure of the protein but would still eliminate proteolytic activity if the substitution were imposed at an active site residue. Thus, Ser residues were converted to Thr in mutants S1 to S7, Cys residues were changed to Ser in mutants C12 and C13, and Asp residues were converted to Glu in mutants D15 to D20. The choice of substituting His residues with Ser in mutants H8 to H11 was made not on the basis of high chemical similarity, but because His to Ser substitutions are most common at positions in HC-Pro where one of the potyviral sequences contains a nonconserved His. We reasoned, therefore, that imposition of Ser at His positions could maintain the general structure of the proteinase, although the weakly basic character of His would be lost. All mutations with the

TABLE 1

OLIGONUCLEOTIDES USED FOR SITE-DIRECTED MUTAGENESIS OF TEV HC-Pro

Mutant code	Oligonucleotide (5′ to 3′)[a]	Amino acid change
S1	CTCAATATACGGATCTAAAG	Ser610 to Thr
S2	ATTGGCAACACTGGCGATTC	Ser625 to Thr
S3	TACTTACTTACCATTCTTTACCC	Ser698 to Thr
S4	GTTTTGGATACGTATGGGTC	Ser726 to Thr
S5	CGTATGGGACTAGAACGAC	Ser729 to Thr
S6	GAACACAACAACCCAGCTAATTG	Ser743 to Thr
S7	GTTTGGAAACCGAAATGAAAAC	Ser755 to Thr
H8	CGAAGAGATCTCTGGTCATTG	His619 to Ser
H9	GGTTGATTCTGACAACAAAAC	His716 to Ser
H10	CAAAACAATGTCTGTTTTGG	His722 to Ser
H11	CAGGATACTCCATGTTGAAAATG	His735 to Ser
C12	GAAGGTTATTCCTACATGAAC	Cys649 to Ser
C13	CAACTGCATCCTACTTACTTTC	Cys694 to Ser
D15	CTCTGGCGAATCAAAGTAC	Asp627 to Glu
D16	GTACCTAGAACTTCCAGTTC	Asp632 to Glu
D17	GTTTATAAGGGAAACAATTGTTC	Asp675 to Glu
D18	CAATGCAAGAAGTTGCAAC	Asp689 to Glu
D19	GAATTTTGGTTGAACATGACAAC	Asp715 to Glu
D20	CATGTTTTGGAATCGTATGG	Asp725 to Glu

[a] Codons affected by mutagenesis are underlined.

exception of S1 were placed in pTL7S-2027, which harbors TEV cDNA representing a genome region encoding a 30-kDa segment of the polyprotein (Fig. 2). This segment encompasses the proteolytic domain (20-kDa) of HC-Pro, the Gly–Gly cleavage site, and a 10-kDa fragment of the protein adjacent to HC-Pro within the polyprotein. Expression of a similar construct *in vitro* yields products of 20 and 10 kDa, which result from autoproteolytic cleavage of the 30-kDa synthetic precursor polyprotein (Carrington *et al.*, 1989a,b). Mutation S1 was generated in pTL7SN-1527, which is similar to pTL7S-2027 except that it encodes a 37-kDa (rather than 20-kDa) segment of HC-Pro (Fig. 2).

Synthetic transcripts were generated from wild-type and mutagenized pTL7SN-1527 and pTL7S-2027 and subjected to cell-free translation in the message-dependent, rabbit reticulocyte lysate system. As shown previously, translation resulted in proteolytic products of 37 and 10 kDa using wild-type pTL7SN-1527 transcripts and in proteolytic products of 20 and 10 kDa using wild-type pTL7S-2027 transcripts (Fig. 3). Of the 19 mutants tested, only 2 were deficient in proteolytic activity. Alterations of pTL7S-2027 affecting Cys-649 (mutant C12) and His-722 (mutant H10) completely inhibited formation of the 20- and 10-kDa products and promoted accumulation of the intact 30-kDa precursor. Little or no effect on proteolytic activity was detected in the other 17 mutant HC-Pro derivatives tested. Although it appears that mutant D19 was partially defective in proteolytic activity, this result was not reproducible in subsequent experiments (for example, see below).

To further characterize these HC-Pro derivatives, and to identify mutants that exhibited relatively subtle effects on the rate of proteolytic cleavage, reaction progress curves were generated by pulse–chase analysis for each mutant proteinase. Translation products were synthesized during a 7-min period in the presence of [^{35}S]methionine, followed by addition of excess nonlabeled methionine to terminate incorporation. Samples were withdrawn at various times and analyzed by SDS–PAGE. As shown previously (Carrington *et al.*, 1989b), proteolysis of the synthetic polyproteins encoded by wild-type pTL7SN-1527 and pTL7S-2027 transcripts proceeded rapidly, with greater than 50% conversion of polyprotein precursor to product immediately after the pulse period (Fig. 4). Consistent with results shown in Fig. 3, HC-Pro mutants with substitutions of Cys-649 (C12) and His-722 (H10) exhibited no proteolytic activity over the time-course examined. The processing kinetics measured for all other mutant proteinases were nearly indistinguishable from those of the wild-type enzyme, although a few reacted slightly slower (e.g., mutant C13 at Cys-694) and a few slightly faster (e.g., mutant S5 at Ser-729). The percentage processing determined at most time-points for each mutant, other than C12 and H10, was within 20% of wild-type in all cases.

DISCUSSION

We have conducted a genetic analysis of the proteolytic domain of potyvirus proteinase HC-Pro. Sequences encoding Ser, His, Cys, and Asp residues that were conserved within the HC-Pro molecule from four separate potyviruses were mutagenized *in vitro* and expressed using a cell-free translation and proteolytic processing assay. The results unambiguously show that only 2 of 19 mutant proteinases, containing amino acid substitutions at Cys-649 and His-722, were proteolytically nonfunctional. Kinetic analysis of processing reactions has demonstrated that the targeted amino acid residues in the remaining 17 mutant proteinases were nonessential to the catalytic mechanism, even though they were strictly conserved among the potyviruses TEV, TVMV, PVY, and PPV.

Each of the major groups of proteinases contains a unique assemblage of amino acid residues that constitute the active site (see Introduction). Given the spectrum of essential residues we have identified among those mutagenized in this study, we can rule out the possibility that HC-Pro belongs to the serine-type or

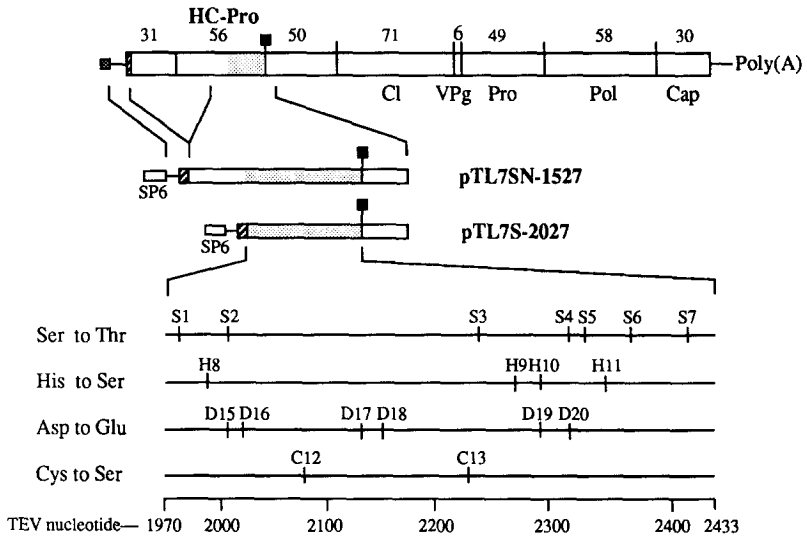

FIG. 2. Genetic map of the tobacco etch virus (TEV) genome, structure of plasmids pTL7SN-1527 and pTL7S-2027, and positions of mutagenized sequences. The 5′- and 3′-nontranslated regions in the genome diagram are indicated by straight lines, whereas the long open reading frame spanning most of the RNA is shown by the rectangular box. The genome-linked protein (VPg) is shown as the shaded square box, and is attached to the 5′-terminus of the viral RNA (Hari, 1981). The vertical lines show the positions of sequences encoding cleavage sites, while the black square indicates the coding sequence for the Gly–Gly dipeptide cleaved by the HC-Pro (Carrington *et al.*, 1989a). The sizes (in kilodaltons) of the processed TEV proteins are shown above the diagram. The proteolytic domain of HC-Pro is shown by the shaded region within the HC-Pro segment. DNA fragments (enlarged diagram) representing segments of the TEV genome were inserted into vectors pTL7SN (to generate pTL7SN-1527) and pTL7S (to generate pTL7S-2027), which carry cDNA representing the 5′-nontranslated region and initial coding sequence (striped box) of the TEV genome, and a bacteriophage SP6 promoter. Mutation S1 was introduced into pTL7SN-1527, whereas the remaining 18 mutations were placed within pTL7S-2027. TEV nucleotide numbers are shown on the scale at the bottom. Abbreviations: HC-Pro, helper component-proteinase; Cl, cylindrical inclusion protein; VPg, genome-linked protein; Pro, proteinase, Pol, polymerase; Cap, capsid protein; Poly(A), polyadenylated tail; Ser, serine; Thr, threonine; His, histidine; Cys, cysteine; Asp, aspartic acid; Glu, glutamic acid.

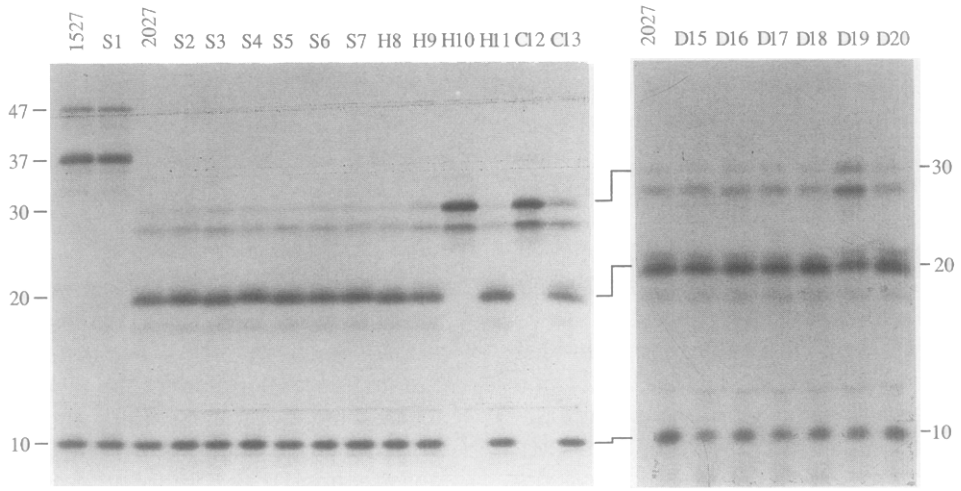

FIG. 3. Expression and proteolytic activities of wild-type and mutagenized TEV HC-Pro derivatives. Synthetic transcripts were generated from wild-type plasmids pTL7SN-1527 (designated 1527) and pTL7S-2027 (designated 2027) and from plasmids carrying point mutations affecting single amino acid substitutions in the HC-Pro proteolytic domain (designated S1, S2, . . . , D19, D20). The exact position mutagenized in each construct is given in Table 1 and Figs. 1 and 2. Transcripts were translated in the rabbit reticulocyte lysate system in the presence of [^{35}S]methionine, and the radiolabeled products were detected by SDS–PAGE and autoradiography. Note that mutants H10 (affecting His-722) and C12 (affecting Cys-649) are defective in proteolytic activity and thus accumulate the precursor polyprotein. Sizes (in kilodaltons) of the radiolabeled proteins are given at the left and right sides.

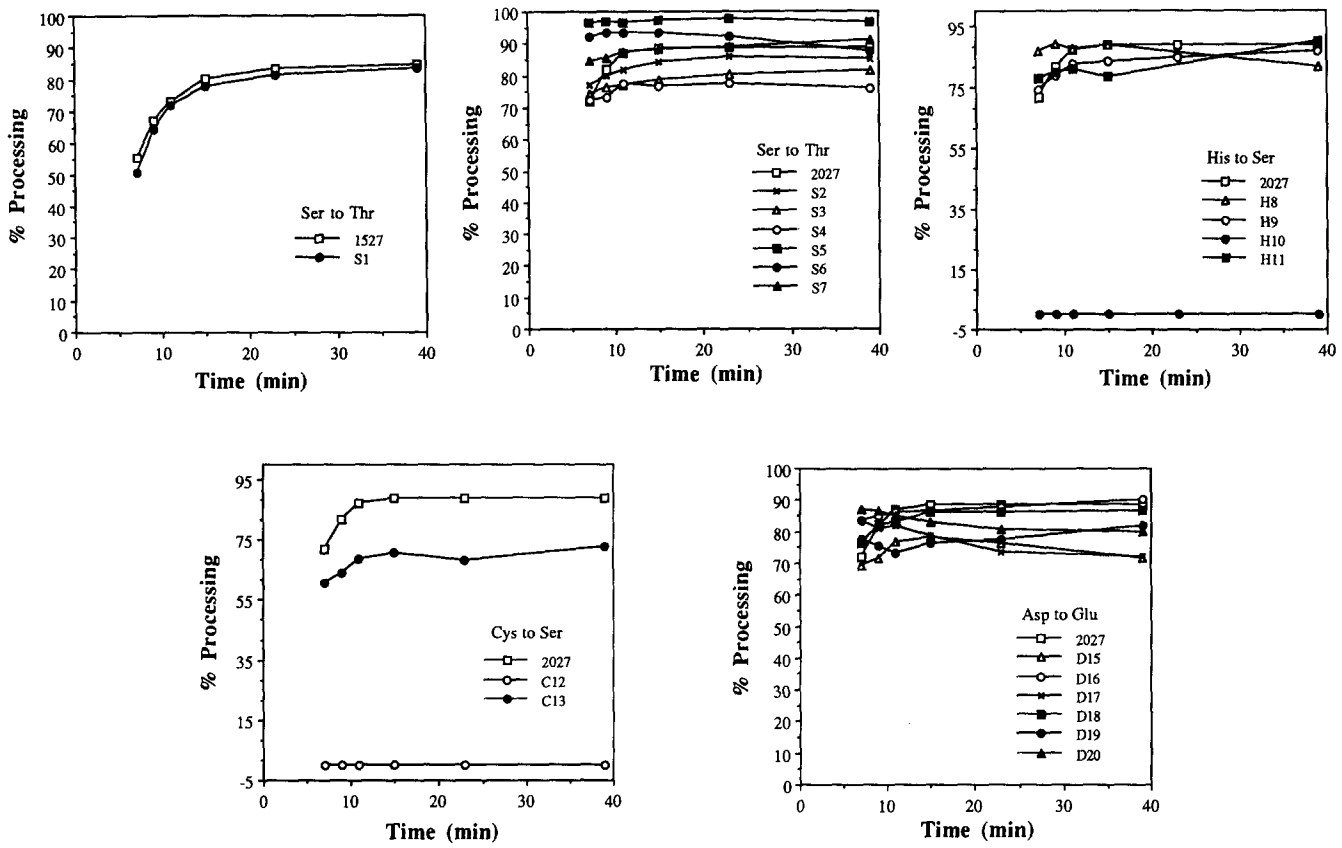

FIG. 4. Pulse–chase analysis of wild-type and mutant TEV HC-Pro-containing polyproteins. [^{35}S]Methionine was added at the beginning of the translation reaction. After 7 min, nonlabeled methionine was added to a final concentration of 15 m*M*. Samples were withdrawn at 7 min (immediately after addition of methionine) and at 9, 11, 15, 23, and 39 min. Proteins were analyzed by SDS–PAGE and autoradiography. The intensities of precursor and product bands in the autoradiographs were quantified by scanning densitometry. The ratio of proteolytic products to the sum of the precursor plus products at each time-point was determined and plotted as percentage processing.

aspartic acid-type proteinase families, since no Ser or Asp residues were required for proteolytic activity of HC-Pro. If the conserved Asp residue identified by Bazan and Fletterick (1988) in the 3C- and 2A-type viral proteinases functions as an active-site component in a manner analogous to Asp-102 in serine proteinases, we can likewise eliminate the possibility that HC-Pro belongs to the "viral cysteine proteinase" group. Since only one His residue (His-722) was necessary for proteolytic activity, it also appears unlikely that HC-Pro belongs to the metalloendoproteinase group. To the contrary, the finding that only Cys and one His residue were required for proteolytic activity fits precisely with the expectation for a cysteine-type proteinase. Indeed, the cysteine proteinase family is the only group not excluded by our data.

If HC-Pro derives from the cysteine-type proteinase family, the enzyme has evolved substantially since divergence from its cellular counterparts. The overall length of the proteolytic domain of HC-Pro is no longer than 155 residues as defined by deletion studies (Carrington *et al.*, 1989a), whereas the cellular cysteine proteinases possess over 200 residues (Kamphius *et al.*, 1985). The distance between the putative active-site Cys-649 and His-722 is 73 residues, while the active-site residues of papain (type member of the cysteine-type proteinases) are separated by 134 positions. However, the relative placements of putative active-site residues along the polypeptide chain are similar. In papain, the reactive Cys residue occupies position 25, while in HC-Pro Cys-649 resides within 30 amino acid residues from the amino-terminal boundary of the proteolytic domain. Comparisons of the sequence environments surrounding the essential Cys and His residues of cellular cysteine-type proteinases and HC-Pro suggest a distant relationship. A strictly conserved Gly residue is located two positions to the amino-side of the catalytic Cys of several cellular cysteine proteinases, as well as Cys-649 of HC-Pro from the four sequenced potyviruses. Additionally, the consensus sequence Lys–Xaa–Xaa–His–Ala–Val–(3 Xaa)–Gly, which centers on the active site His of cellular cysteine proteinases, bears resemblance to the consensus sequence Lys–Xaa–Xaa–His–Val–Val–(3 Xaa)–Gly,

which surrounds the essential His residue of HC-Pro from the four potyviruses. Direct proof that HC-Pro is assigned correctly to the cysteine proteinase family will require high-resolution structural analysis using X-ray crystallography, although protein modeling using a cysteine proteinase of known structure (e.g., papain) as a guide may yield clues as to the validity of this hypothesis.

Although a cysteine-type proteinase comparable to HC-Pro is not encoded by the genomes of picornaviruses or comoviruses, a possible analog has been identified as a product of the coronavirus genome. Translation of the 5′-proximal cistron of coronaviral genomic RNA yields a large polyprotein precursor termed F1–F2 (Boursnell *et al.*, 1987; Brierley *et al.*, 1987). Two proteinase domains within the polyprotein have been identified by computer-assisted sequence similarity searches (Gorbalenya *et al.*, 1989b). Remarkably, one exhibits similarity to the 3C family of viral proteinases, which includes the potyviral NIa (49-kDa proteinase of TEV). The other, located near the amino-terminus of the polyprotein, strongly resembles a cysteine-type proteinase from *Streptococcus pneumoniae.* The sequence similarity is particularly noteworthy surrounding the putative active-site Cys residue. Compared to potyvirus HC-Pro, the coronavirus avian infectious bronchitis virus (IBV) protein contains the sequence Gly–His–Cys–Tyr, surrounding the putative active-site Cys residue (underlined), which is nearly identical to the sequence Gly–Tyr–Cys–Tyr surrounding the essential Cys-649 of TEV HC-Pro. Interestingly, the relative positions of these proteinase domains within their respective polyproteins are similar. Further comparisons between these proteins will be possible once the biochemical activity and substrate specificity of the IBV putative proteinase are established and its role in polyprotein processing is understood.

It is important to point out that these experiments have identified Cys-649 and His-722 of HC-Pro as residues important for autolytic activity at its carboxyl-terminus. When a polypeptide representing the initial 97-kDa of the TEV polyprotein is synthesized *in vitro,* proteolysis at the amino-terminus of HC-Pro (to yield products of 31 and 56 kDa) fails to occur, even though HC-Pro-mediated cleavage at its carboxyl-terminus proceeds efficiently to yield products of 87 and 10 kDa (Carrington *et al.*, 1989a). This has led to speculation that a third proteolytic activity is required for complete HC-Pro maturation, or that auxiliary factors are necessary to complement the activity of HC-Pro or the 49-kDa proteinase. It is possible that HC-Pro possess two distinct proteolytic activities, one of which catalyzes cleavage at the carboxyl-terminus (the focus of this study) while the other processes at the amino-terminus. This might explain the presence of the sequence motif, Gly–Asn–Ser–Gly (positions 623–626 in the TEV polyprotein), which is invariant in all potyviral HC-Pro sequences determined to date, and which is very similar to the highly conserved sequence surrounding the active-site Ser residue in cellular serine-type proteinases. However, one still must concomitantly postulate the involvement of a host factor or separate viral protein to affect proteolysis by this putative activity, since cleavage at the amino-terminus of HC-Pro fails to occur *in vitro.*

ACKNOWLEDGMENTS

We thank members of our lab, including Deon Freed, Maria Restrepo, and Dale Dreyer, for valuable assistance and discussions relating to this project, and Ray Vonderhaar for rapid and reliable synthesis of oligonucleotides. This work was supported in part by USDA Competitive Grant 89-37263-4503 and 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.

ALLISON, R. F., SORENSON, J. G., KELLY, M. E., ARMSTRONG, F. B., and DOUGHERTY, W. G. (1985). Sequence determination of the capsid protein gene and flanking regions of tobacco etch virus: Evidence for the synthesis and processing of a polyprotein in potyvirus genome expression. *Proc. Natl. Acad. Sci. USA* **82,** 3969–3972.

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.

BAZAN, J. F., and FLETTERICK, R. J. (1988). Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: Structural and functional implications. *Proc. Natl. Acad. Sci. USA* **85,** 7872–7876.

BOURSNELL, M. E. G., BROWN, T. D. K., FOULDS, I. J., GREEN, P. F., TOMLEY, F. M., and BINNS, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *J. Gen. Virol.* **68,** 57–77.

BRIERLEY, I., BOURSNELL, M. E. G., BINNS, M. M., BILIMORIA, B., BLOK, V. C., BROWN, T. D. K., and INGLIS, S. C. (1987). An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J.* **6,** 3779–3785.

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

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., FREED, D. D., and SANDERS, T. S. (1989b). Autoproteolytic processing of potyvirus proteinase HC-Pro in *Escherichia coli* and in vitro. *J. Virol.* **63,** 4459–4463.

DE MEJIA, M. V. G., HIEBERT, E., and PURCIFUL, D. E. (1985). Identification of potyvirus amorphous inclusion protein as a nonstructural virus-specific protein related to helper component. *Virology* **142,** 34–43.

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.

DOMIER, L. L., SHAW, J. G., and RHOADS, R. E. (1987). Potyviral proteins share amino acid sequence homology with picorna-, como-, and caulimoviral proteins. *Virology* **158,** 20–27.

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

DOUGHERTY, W. G., and HIEBERT, E. (1980). Translation of potyvirus RNA in a reticulocyte lysate: Reaction conditions and identification of capsid protein as one of the products of *in vitro* translation of tobacco etch and pepper mottle viral RNAs. *Virology* **101,** 466–474.

GORBALENYA, A. E., DONCHENKO, A. P., BLINOV, V. M., and KOONIN, E. V. (1989a). Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. *FEBS Lett.* **243,** 103–114.

GORBALENYA, A. E., KOONIN, E. V., DONCHENKO, A. P., and BLINOV, V. M. (1989b). Coronavirus genome: Prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res.* **12,** 4847–4861.

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

JONGENEEL, C. V., BOUVIER, J., and BAIROCH, A. (1989). A unique signature identifies a family of zinc-dependent metallopeptidases. *FEBS Lett.* **242,** 211–214.

KAMPHUIS, I. G., DRENTH, J., and BAKER, E. N. (1985). Thiol proteases: Comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain. *J. Mol. Biol.* **182,** 317–329.

KESTER, W. R., and MATHEWS, B. W. (1977). Crystallographic study of the binding of dipeptide inhibitors to thermolysin: Implications for the catalytic mechanism. *Biochemistry* **16,** 2506–2516.

KRAUSSLICH, H. G., and WIMMER, E. (1988). Viral proteinases. *Annu. Rev. Biochem.* **57,** 701–754.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the 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. Viol.* **70,** 513–524.

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

NEURATH, H. (1984). Evolution of proteolytic enzymes. *Science* **224,** 350–357.

POLGAR, L., and HALASZ, P. (1982). Current problems in mechanistic studies of serine and cysteine proteinase. *Biochem. J.* **207,** 1–10.

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

SANGER, F., NICKLEN, S., and COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5469.

TANG, J., SEPULVEDA, P., MARCINISZYN, J., CHEN, K. C. S., HUANG, W. Y., TAO, N., LIU, D., and LANIER, J. P. (1973). Amino acid sequence of porcine pepsin. *Proc. Natl. Acad. Sci. USA* **70,** 3437–3439.

TAYLOR, J. W., OTT, J., and ECKSTEIN, F. (1985a). The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA. *Nucleic Acids Res.* **13,** 8765–8785.

TAYLOR, J. W., SCHMIDT, W., COSSTICK, R., OKRUSZEK, A., and ECKSTEIN, F. (1985b). The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA. *Nucleic Acids Res.* **13,** 8749–8764.

THORNBURY, D. W., HELLMANN, G. M., RHOADS, R. E., and PIRONE, T. P. (1985). Purification and characterization of potyvirus helper component. *Virology* **144,** 260–267.

VIEIRA, J., and MESSING, J. (1987). Production of single-stranded plasmid DNA. *In* "Methods in Enzymology" (R. Wu, Ed.), Vol. **153,** pp. 3–11. Academic Press, New York.