Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 3391–3395, May 1988
Biochemistry

A viral cleavage site cassette: Identification of amino acid sequences required for tobacco etch virus polyprotein processing

(potyvirus proteinase/cell-free expression/cleavage site requirements/protein engineering)

JAMES C. CARRINGTON AND WILLIAM G. DOUGHERTY*

Department of Microbiology, Oregon State University, Corvallis, OR 97331

Communicated by Myron K. Brakke, February 1, 1988 (received for review November 27, 1987)

ABSTRACT **Mature viral-encoded proteins of tobacco etch virus (TEV) arise by proteolytic processing of a large precursor. The proteinase responsible for most of these cleavages is a viral-encoded 49-kDa protein. All known or predicted cleavage sites in the TEV polyprotein are flanked by the conserved sequence motif Glu-Xaa-Xaa-Tyr-Xaa-Gln-Ser or Gly, with the scissile bond located between the Gln-Ser or Gly dipeptide. By using cell-free systems to manipulate and express cloned cDNA sequences, a 25-amino acid segment containing a putative proteolytic cleavage site of the TEV polyprotein has been introduced into the TEV capsid protein sequence. This recombinant protein is cleaved by the 49-kDa proteinase at the introduced cleavage site, thus demonstrating portability of a functional cleavage site. The role of the conserved amino acid sequence in determining substrate activity was tested by construction of engineered proteins that contained part or all of this motif. A protein that harbored an insertion of the conserved 7-amino acid segment was cleaved by the 49-kDa TEV proteinase. Cleavage of the synthetic precursor was shown to occur accurately between the expected Gln-Ser dipeptide by microsequence analysis. Proteins containing insertions that generated only the Gln-Ser, or only the serine moiety of the conserved sequence, were insensitive to the 49-kDa proteinase.**

The RNA genome [9.5 kilobases (kb)] of tobacco etch virus (TEV), a plant potyvirus, is encapsidated in long, flexuous rod-shaped particles by capsid protein subunits of 30 kDa (1, 2). The genomic RNA is linked covalently to a small protein (VPg; ref. 3), is polyadenylylated at the 3′ terminus (4) and contains a single open reading frame with the potential to encode a 346-kDa polyprotein (1). The mature viral-encoded proteins arise by proteolytic processing of the large polyprotein precursor (4–6).

The proteinase that functions to catalyze several of the TEV polyprotein cleavages is a viral protein with a molecular mass of 49 kDa (5, 6); this proteinase aggregates with another viral protein (54 kDa) to form inclusion bodies in the nuclei of infected cells (7, 8). The 49-kDa proteinase efficiently autoexcises from polyproteins when synthesized in cell-free translation systems programed with defined SP6 RNA transcripts (6). In addition, the 49-kDa proteinase, synthesized *in vitro* or present in nuclear inclusion bodies purified from infected plants, can function in trans when added to TEV polyprotein precursors (5, 6). Several lines of evidence suggest that the 49-kDa proteinase is a cysteine-type proteinase (ref. 6 and references therein).

The 49-kDa proteinase cleaves at Gln-Gly and Gln-Ser dipeptides in the TEV polyprotein (5, 6); however, not all Gln-Gly and Gln-Ser dipeptides in the TEV polyprotein are substrates for cleavage (see Fig. 1*A*). Information in the polyprotein for substrate specificity can reside at the amino acid sequence level. For example, the proteinase can require a specific sequence motif at the cleavage sites for efficient recognition of the substrate. Such a mechanism has been suggested because glutamic acid, tyrosine, and glutamine are conserved at positions −6, −3, and −1, respectively, at all known or predicted cleavage sites recognized by the 49-kDa proteinase (5, 6). (The numbering convention used in this paper is centered on the scissile bond, where glutamine represents position −1 and serine or glycine represents position +1.) Additionally, amino acid sequence context probably contributes to substrate efficiency. Cleavage sites would be predicted to occur in flexible surface regions in the viral polyprotein precursors; evidence for such a model in poliovirus has been presented by Arnold *et al.* (9) to account for substrate selection by the 3C proteinase.

To identify sequences in the TEV polyprotein sufficient to confer 49-kDa proteinase-mediated substrate activity, amino acid sequences bordering two TEV cleavage sites were introduced, by using recombinant DNA techniques, into proteins at sites not normally processed by the proteinase. The results show that TEV cleavage sites are transferable from one protein to another and that as few as seven amino acid residues (from −6 to +1) surrounding an authentic polyprotein junction can define a functional cleavage site.

MATERIALS AND METHODS

Materials. TEV cDNA in recombinant plasmids was synthesized from RNA isolated from virion preparations of the highly aphid transmissible isolate (10). *Escherichia coli* strain HB101 was used for cloning of plasmids. Nuclear inclusion bodies (with 49-kDa proteolytic activity) were isolated from *Datura stramonium* plants infected with the non-aphid transmissible isolate of TEV (11) by the method of Dougherty and Hiebert (8). Enzymes used for DNA subcloning and restriction digests were from New England Biolabs, Bethesda Research Laboratories, and Boehringer Mannheim. SP6 polymerase was purchased from Promega Biotec (Madison, WI) and Bethesda Research Laboratories. Rabbit reticulocyte lysate was from Green Hectares (Oregon, WI). All radioisotopes and EN^{3}HANCE were purchased from New England Nuclear.

Plasmid Constructs. Vector pTL-8 (5) was used in this study for construction of recombinant plasmids. This vector contains cDNA representing the 5′ untranslated and initial coding regions of the TEV genome downstream from an SP6 promoter. A polylinker downstream from the TEV "leader" segment facilitates insertion of foreign DNA sequences that contain open reading frames. SP6 transcripts synthesized from chimeric plasmids are translated efficiently in a rabbit reticulocyte lysate system. pTL-8595, containing DNA representing TEV genome nucleotides 8462–9495 in pTL-8, has been described (5).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: TEV, tobacco etch virus.

*To whom reprint requests should be addressed.

pTL-8595/70-N was assembled by removing a 97-base-pair (bp) *Bal* I–*Xho* I restriction fragment from the capsid protein coding region in pTL-8595 and inserting a 79-bp *Sst* I–*Xho* I restriction fragment representing nucleotides 3608–3686 in the TEV genome (the *Sst* I site was made flush by treatment with mung bean nuclease). This DNA fragment contains the coding sequence surrounding the 50-kDa–70-kDa protein cleavage site in the TEV polyprotein (see Fig. 1A). Due to the position of restriction enzyme sites relative to the reading frame, 2 bp at the 5′ and 3′ ends of the insert contributed to formation of codons not found in the TEV sequence. Hence, the DNA insert encoded the 25 amino acid residues found between positions 1156 and 1180 in the TEV polyprotein. Proper (in frame) insertion of the DNA was verified by nucleotide sequence analysis (data not shown).

Three plasmids (pSCS, pP-6, and pE-5) were constructed by inserting oligonucleotide linkers between the coding sequences for two regions of the 49-kDa proteinase (see Fig. 2*A*). Each plasmid contained (in the 5′ → 3′ direction) DNA coding for amino acid residues 9–77 of the 49-kDa proteinase (*Hin*dIII to *Bam*HI fragment), a linker, and the coding sequence (*Eco*RV to *Acc* I) for proteinase residues 203–354; these sequences were inserted in frame with the TEV leader segment in vector pTL-8. The linker used for pSCS (5′ AGGGAGAATCTTTATTTTCA 3′) coded for the underlined amino acid sequence Glu-Asn-Leu-Tyr-Phe-Gln-Ser. The resulting 7-amino acid sequence is identical to that found between positions −6 and +1 at the 54-kDa–30-kDa protein junction in the TEV polyprotein (amino acids 2786–2792). The pP-6 linker (5′ TGCACTGCAGTGCA 3′) coded for the underlined amino acid sequence Cys-Thr-Ala-Val-Gln-Ser; thus, only the resulting Gln-Ser dipeptide of the 54-kDa–30-kDa protein junction was present in this construction. The pE-5 linker (5′ CCGAATTCGG 3′) coded for the underlined sequence Glu-Phe-Gly-Ser. These three linkers were inserted upstream of the sequence coding for a serine residue (shown above). The sequence at the point of the oligonucleotide insertion was determined by the dideoxynucleotide method for each construct (data not shown).

***In Vitro* Transcription, Translation, and Proteolytic Processing.** Plasmid DNA used for SP6 transcription was linearized by restriction enzyme digestion, purified by extraction with phenol and ether, and concentrated by ethanol precipitation. SP6 transcription reactions were conducted as described (5, 12) with ≈2 μg of DNA per 30-μl reaction volume. RNA transcription products were precipitated with ethanol in the presence of 0.66 M ammonium acetate and resuspended in 40 μl of deionized H_2O. A mRNA-dependent, rabbit reticulocyte lysate system containing [^{35}S]methionine was programed with 5 μl of transcription products by using the reaction conditions described by Dougherty and Hiebert (13). Proteolytic processing reactions using TEV nuclear inclusion bodies as the proteinase source were conducted as in previous studies (5, 6). All processing reactions, containing 3 mM phenylmethylsulfonyl fluoride and 400–800 μg of partially purified nuclear inclusion bodies per ml, were performed immediately after synthesis of the substrates in the reticulocyte lysate at 30°C. [^{35}S]labeled products were analyzed by $NaDodSO_4$/PAGE (12.5% resolving gel; ref. 14) and fluorography using EN^{3}HANCE.

Protein Microsequencing. Amino-terminal microsequence analysis was conducted on the 18-kDa protein product released by proteolytic processing of the pSCS-derived 29-kDa precursor. Approximately 10 μg of pSCS transcripts was translated in the reticulocyte lysate (250 μl) in the presence of either [^{35}S]cysteine or [^{3}H]leucine. The radiolabeled proteins were subjected to proteolysis in the presence of TEV nuclear inclusions, and the products were resolved by preparative gel electrophoresis. The 18-kDa protein product was eluted from the gels by using the method of Kelly *et al.* (15) and was subjected to automated Edman degradation using a gas-phase sequenator (Applied Biosystems, Foster City, CA). The radioactivity released at each cycle was quantitated by using a Beckman scintillation counter.

RESULTS

Transfer of the TEV 50-kDa–70-kDa Cleavage Site. A map of the TEV genome and the deduced positions of known or proposed cleavage sites in the polyprotein are shown in Fig. 1*A*. By protein microsequence analysis, we have determined that cleavage to generate the amino terminus of the 70-kDa cytoplasmic pinwheel inclusion protein occurs between the Gln-Ser dipeptide at position 1163–1164 in the polyprotein (J.C.C., Susan Cary, and W.G.D., unpublished data). The amino acid residues flanking this site (glutamic acid at −6, tyrosine at −3, glutamine at −1) are conserved at all known or predicted cleavage sites recognized by the 49-kDa proteinase (4, 5, 6, 16). To determine if a short amino acid sequence surrounding this TEV polyprotein cleavage site could be transferred into another protein and maintain functional substrate activity, a restriction fragment coding for 25 amino acid residues that contained the 50-kDa–70-kDa protein cleavage site was inserted in the same reading frame as the capsid protein coding region in pTL-8595. The transferred fragment encoded 8 residues toward the amino terminus from the Gln-Ser dipeptide and 17 residues toward the carboxyl terminus. The resulting plasmid was termed pTL-8595/70-N (Fig. 1*A*). Since a 97-bp fragment was eliminated from pTL-8595 and replaced by the 79-bp insert in pTL-8595/70-N, the sizes of the proteins expressed from each construct should be approximately the same.

As shown in Fig. 1*B* (lanes 1 and 2) and in a previous study (5), translation of pTL-8595 transcripts resulted in synthesis of a 34-kDa protein that was processed to 30-kDa (capsid) and 4-kDa (not seen on the gel fluorograph) products after incubation with the 49-kDa proteinase; partial amino acid sequence analysis of the 30-kDa cleavage product has indicated that the cell-free processing reaction occurs at the authentic capsid protein cleavage site (6). Translation of pTL-8595/70-N transcripts also resulted in production of a 34-kDa precursor (Fig. 1*B*, lane 3). If the cleavage site transferred into the capsid protein in the pTL-8595/70-N product was functional, processing with the 49-kDa proteinase should yield three products of 23 kDa, 7 kDa, and 4 kDa. Reaction of the 34-kDa synthetic polyprotein with the 49-kDa proteinase activity resulted in accumulation of a 23-kDa product (Fig. 1*B*, lane 4; the two smaller products would not have been resolved in our gel system). In addition, a product of 30 kDa was observed and corresponded to a partially digested protein cleaved only at the capsid protein amino terminus. The 25-amino acid segment appeared to contain information necessary to direct substrate activity using the 49-kDa proteinase.

Cleavage Site Insertion Using a Synthetic Oligonucleotide Linker. A motif of conserved amino acid residues (Glu-Xaa-Xaa-Tyr-Xaa-Gln-Ser or Gly, where the scissile bond is after the Gln residue) flanks all known cleavage sites recognized by the 49-kDa proteinase (4, 5, 6, 16). To test the hypothesis that this 7-amino acid sequence within a polyprotein is sufficient to direct cleavage by the 49-kDa proteinase, a synthetic oligonucleotide linker encoding the underlined sequence Glu-Asn-Leu-Tyr-Phe-Gln-Ser was inserted into a modified 49-kDa proteinase gene (the serine residue was present in the 49-kDa protein) to generate the plasmid pSCS (Fig. 2*A*). The resulting 7-amino acid sequence was identical to that existing between positions −6 to +1 at the authentic 54-kDa–30-kDa protein cleavage site in the TEV polyprotein. Two additional recombinant plasmids were assembled that contained insertions of oligonucleotide linkers coding for the

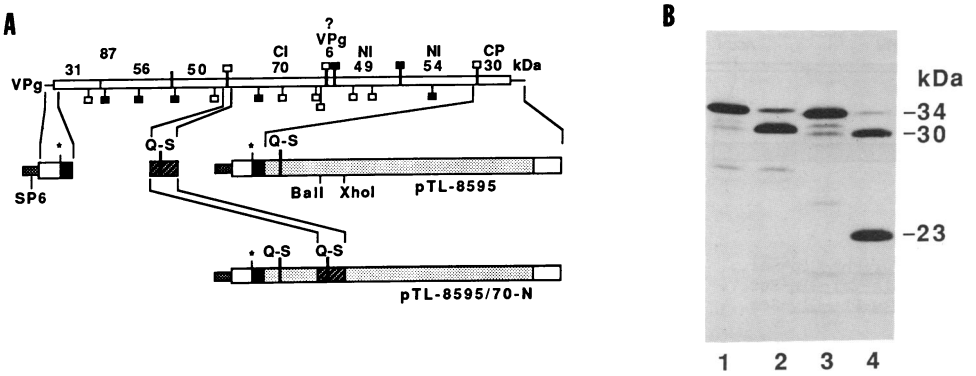

FIG. 1. Transfer of the TEV 50-kDa–70-kDa protein proteolytic cleavage site into the TEV capsid protein. (*A*) TEV genetic map (top) and schematic diagrams of DNA in constructs pTL-8595 and pTL-8595/70-N. The open reading frame in the TEV genome map is indicated by the open box, whereas the known cleavage sites in the polyprotein are shown by thick vertical lines above the map. Open squares or filled-in squares attached to the cleavage site markings indicate Gln-Ser or Gln-Gly dipeptides, respectively. The thin vertical lines below the genetic map indicate Gln-Ser (with open squares) or Gln-Gly (filled-in squares) dipeptides that apparently are not sites for proteolysis in the polyprotein (refs. 5 and 6; and unpublished data). The cleavage sites delineating the 87-kDa–50-kDa protein boundary, as well as the boundary between the 31-kDa and 56-kDa proteins, are speculative. Each plasmid construct (pTL-8595 and pTL-8595/70-N) contains an SP6 promoter (shaded box) linked to the TEV leader (from vector pTL-8; see ref. 5); the 5′ untranslated and initial coding sequences of the leader are indicated by open and filled-in boxes, respectively. The translational start site is shown by the asterisk. pTL-8595 contains DNA encoding the carboxyl-terminal region of the 54-kDa protein, a polyprotein cleavage site (Q-S), and the entire 30-kDa capsid sequence (stippled box) as well as the 3′ untranslated region (open box). pTL-8595/70-N contains a 79-bp DNA insertion, coding for 25 amino acid residues surrounding the 50-kDa–70-kDa cleavage site (striped box), in place of a 97-bp *Bal* I–*Xho* I fragment in the 30-kDa capsid gene. CI, cytoplasmic pinwheel inclusion protein; NI, nuclear inclusion protein; CP, capsid protein; VPg, viral protein, genome linked. (*B*) ^{35}S-labeled synthetic polyproteins, synthesized from pTL-8595 (lanes 1 and 2) or pTL-8595/70-N (lanes 3 and 4) transcripts, were treated with either H_2O (without proteinase, lanes 1 and 3) or the proteolytic activity present in TEV nuclear inclusion bodies (with proteinase, lanes 2 and 4) as described (5, 6). Products were analyzed by $NaDodSO_4$/PAGE and fluorography.

underlined peptides Cys-Thr-Ala-Val-Gln-Ser (pP-6) and Glu-Phe-Gly-Ser (pE-5); the pP-6 linker regenerated the Gln-Ser dipeptide found at positions −1 to +1 in the authentic cleavage site, whereas the pE-5 linker preserved only the −2 (phenylalanine) and +1 (serine) coding positions. Each linker was inserted between two noncontiguous regions of the 49-kDa proteinase coding sequence (see *Materials and Methods* and Fig. 2*A*). SP6 transcripts that terminated at the *Acc* I site 228 bp from the 3′ end of the 49-kDa coding region were used to synthesize protein in a rabbit reticulocyte lysate. Since synthetic proteins that lack the carboxyl-terminal 76 amino acid residues of the 49-kDa proteinase are enzymatically inactive (5), the pSCS-, pP-6-, and pE-5-derived proteins were used only as proteinase substrates, whereas nuclear inclusion bodies served as the source of proteinase. Substrate proteins of 28 kDa were predicted (from the deduced amino acid sequence) to result from translation of all three transcripts. Accurate processing at the amino acids introduced by the oligonucleotide insertions was predicted to yield products of 17 kDa and 11 kDa.

Translation of SP6 transcripts derived from pSCS, pP-6, and pE-5 each resulted in accumulation of an ≈29-kDa protein (Fig. 2*B*, lanes 1, 3, and 5). Reaction of the pSCS derived protein (containing the −6 to +1 cleavage site sequence) with proteinase resulted in accumulation of a product with an apparent molecular mass of ≈18 kDa (Fig. 2*B*, lane 2). In addition, a smaller product that migrated slightly slower than the dye front in our gel system was formed during proteolysis. Processing at the synthetic cleavage site could account for this pattern of reaction products. In contrast, incubation of the TEV proteinase with the 29-kDa proteins derived from pP-6 and pE-5 transcripts resulted in no apparent cleavage (Fig. 2*B*, lanes 4 and 6).

To identify the site at which the 49-kDa proteinase was cleaving the pSCS-derived substrate protein, amino-terminal microsequence analysis was conducted on the ≈18-kDa proteolytic reaction product. The pSCS transcripts were translated in the presence of either [^{35}S]cysteine or [^{3}H]leucine, and the resulting 29-kDa synthetic precursor was cleaved by the TEV 49-kDa proteinase. The radiolabeled ≈18-kDa product was isolated and subjected to automated Edman degradation. Based on the amino acid residue position at which radioactivity was released, a cysteine residue occupied position 5 from the amino terminus (Fig. 3). In addition, leucine was found at position 7 from the amino terminus (Fig. 3). Within the deduced amino acid sequence of the protein expressed from pSCS transcripts, Cys-Xaa-Leu appeared only at position +5 to +7 relative to the peptide specified by the oligonucleotide insert. This result suggested that cleavage occurred between the Gln-Ser pair of the synthetic site to generate the ≈18-kDa protein product.

DISCUSSION

In experiments reported here, we have identified amino acid sequences that function as substrate recognition signals for proteolysis mediated by the TEV 49-kDa proteinase. As the proteinase can accumulate to high levels within infected cells in the absence of cell death, and since only a specific subset of Gln-Ser and Gln-Gly dipeptides in the TEV polyprotein is cleaved, interaction of the 49-kDa proteinase with the correct substrate sites must involve some level(s) of specificity. To gain insight into the nature of possible amino acid sequences in the TEV polyprotein involved in substrate activity, amino acid segments flanking two cleavage sites have been transferred into proteins at positions not normally recognized by the viral proteinase. In one case, a peptide containing 8 residues upstream (amino-terminal side) and 17 residues downstream (carboxyl-terminal side) from a cleavage site confers substrate activity (Fig. 1) with an efficiency approximately equal the same cleavage site in its natural environment (between the 50-kDa and 70-kDa proteins; data not shown). Therefore, all sequence and structural requirements for substrate activity are apparently met within this 25-amino acid region.

In the second case, oligonucleotide linkers have been inserted between two artificial protein domains. In one construction, the resulting amino acid sequence is identical to

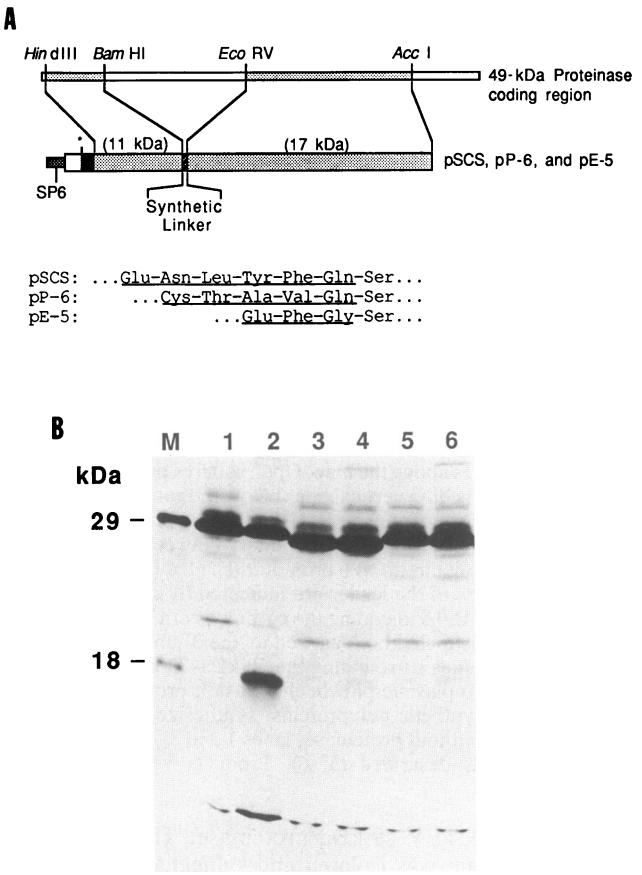

FIG. 2. Substrate activities of engineered proteins. (*A*) Schematic diagrams of DNA constructs containing oligonucleotide linkers encoding the underlined amino acid sequence Glu-Asn-Leu-Tyr-Phe-Gln-Ser (in pSCS), Cys-Thr-Ala-Val-Gln-Ser (in pP-6), or Glu-Phe-Gly-Ser (in pE-5). In each construct, the linker was positioned adjacent to DNA encoding a serine residue. The amino acid sequence specified by the pSCS linker was identical to the sequence present at the TEV 54-kDa–30-kDa polyprotein cleavage site from positions −6 to +1 (proteolysis occurs between the Gln-Ser dipeptide). The pP-6 linker-encoded peptide regenerates cleavage site positions −1 to +1 (Gln-Ser); the pE-5 linker-encoded peptide regenerates residue positions −2 and +1 (phenylalanine and serine). Each linker was positioned upstream from sequences encoding a 17-kDa polypeptide segment of the 49-kDa proteinase (this fragment is enzymatically inactive). Sequences encoding an 8-kDa segment of the 49-kDa proteinase, and the TEV leader sequence (from vector pTL-8) encoding a 3-kDa peptide, were positioned upstream from the linkers. The deduced molecular masses of the polypeptides encoded by sequences flanking the oligonucleotides are shown in parentheses. (*B*) ^{35}S-labeled proteins synthesized *in vitro* from SP6 transcripts of *Acc* I-linearized pSCS (lanes 1 and 2), pP-6 (lanes 3 and 4), and pE-5 (lanes 5 and 6). Aliquots of proteins synthesized in the cell-free system were mixed with either H_2O (lanes 1, 3, and 5) or the 49-kDa proteinase (lanes 2, 4, and 6) for 1 hr at 30°C. Protein products were analyzed by $NaDodSO_4$/PAGE and fluorography. Lane M, protein molecular mass markers; 29 kDa, carbonic anhydrase; 18 kDa, lactoglobulin A.

the sequence at the 54-kDa–30-kDa cleavage site from positions −6 to +1. This insertion is sufficient to confer susceptibility to the 49-kDa proteinase (Fig. 2). Insertion of peptides that generate the Gln-Ser dipeptide found at positions −1 to +1 at several authentic cleavage sites in the TEV polyprotein, or the phenylalanine and serine found at positions −2 and +1, respectively, at the 54-kDa–30-kDa cleavage site, are insufficient for substrate activity. Apparently, substrate activity of the pSCS-derived protein is not the result of activation of a cryptic cleavage site (by perturbation of the protein structure) since proteins engineered with the two

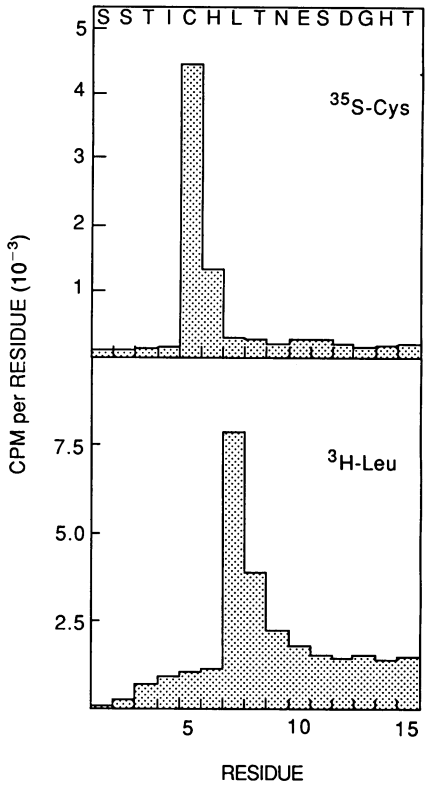

FIG. 3. Amino-terminal microsequence analysis of the ≈18-kDa product released by proteolysis of the pSCS-derived synthetic precursor. [^{35}S]Cysteine- or [^{3}H]leucine-labeled proteins were synthesized from SP6 transcripts of the plasmid pSCS and reacted with the 49-kDa proteinase. The ≈18-kDa product was isolated and subjected to automated Edman degradation. The radioactivity released at each cycle was quantitated by using a scintillation counter. The amino acid sequence residing adjacent to the carboxyl end of the inserted peptide is given above the plots. The serine shown in position 1 of the given sequence is the serine that occupies position +1 in the pSCS synthetic cleavage site. The single-letter amino acid code is used.

other peptides at the same position fail to process when reacted with the proteinase. Furthermore, protein microsequence data show that cleavage occurs accurately between the Gln-Ser dipeptide generated by the oligonucleotide insertion. Therefore, we conclude that amino acid residues integral to cleavage site function are contained within positions −6 to +1. It is important to point out, however, that the synthetic cleavage site in pSCS-derived protein is processed less efficiently than the cleavage site within its authentic context (54-kDa–30-kDa junction; data not shown). This implies that additional amino acid sequence requirements affecting the efficiency of cleavage exist outside of the −6 to +1 region or that the context within which the synthetic site has been inserted is suboptimal. Although the synthetic cleavage site has been engineered between two noncontiguous polypeptides with the aim of creating a flexible, accessible region, the cleavage site could have been partially masked such that its presentation to proteinase is restricted. This would not be surprising, since even authentic cleavage sites in various viral polyproteins are refractory to proteolysis under certain conditions. For example, the (autocatalytic) processing sites at the termini of the 49-kDa proteinase in the TEV polyprotein are processed in trans with very poor efficiency or not at all when supplied with exogenous 49-kDa proteinase (6). Additionally, in poliovirus, cleavage mediated by proteinase P3C at P1 precursor sites in the polyprotein is restricted in synthetic polyproteins lacking the carboxyl end of the P1 region (17–19).

The decision to insert an oligonucleotide that results in a site consisting of seven amino acid residues (pSCS) is based on the length of the conserved sequence flanking each known or proposed 49-kDa proteinase-mediated cleavage site. At the five protein junctions in the TEV polyprotein (see Fig. 1A), positions −6 to +1 contain three strictly conserved residues (glutamic acid, −6; tyrosine, −3; glutamine, −1); furthermore, position +1 is occupied by either serine or glycine, and positions −5, −4, and −2 contain exclusively neutral or hydrophobic residues (1, 4, 5). Except for the presence of a polar residue at position +3, no obvious amino acid conservation exists between positions +2 to +6. In addition, a conserved set of amino acid residues is predicted to occupy positions −4 to +1 at the cleavage sites of the tobacco vein-mottling virus (a related potyvirus) polyprotein, although the residue identities are different from those of TEV (20). The presence of similar sequence motifs surrounding each TEV processing site may be indicative of a primary sequence requirement for efficient proteolysis. One possible role for these amino acids is to position the scissile bond next to the nucleophilic cysteine in the proteinase active center by means of interaction with nearby subsites in the proteinase. Such a mechanism for substrate selection and specificity has been shown to operate for the well-characterized cysteine-type proteinase papain (21–25). The requirement for the conserved amino acid residues flanking the TEV cleavage sites during efficient proteolysis recently has been shown by using site-directed mutagenesis (26). In other virus groups that utilize a polyprotein processing strategy for gene expression (e.g., picornaviruses and comoviruses), strict conservation of an extended sequence motif is not found at the cleavage sites (27, 28).

In addition to yielding information regarding the TEV cleavage site–proteinase recognition system, this study may have relevance to the field of protein engineering. The 49-kDa proteinase appears to possess a high affinity toward the TEV polyprotein cleavage sites. Since we have demonstrated the transferability of short peptides with high specificity normally found at the cleavage sites, it may be possible to engineer proteins (of academic or commercial interest) harboring proteolytic cleavage sites that may be activated by addition of the TEV proteinase.

We thank Susan Cary for excellent assistance during this project, Reg McParland for conducting the amino acid sequence analysis, and T. Dawn Parks for helpful comments on the manuscript. This work was supported by a grant from the National Science Foundation (to W.G.D.) and a postdoctoral fellowship from the National Institutes of Health (to J.C.C.). This manuscript is Oregon Agricultural Experiment Station Technical Paper No. 8432.

1. Allison, R. F., Johnston, R. E. & Dougherty, W. G. (1986) *Virology* **154**, 9–20.
2. Hollings, M. & Brunt, A. A. (1981) in *Handbook of Plant Virus Infection and Comparative Diagnosis*, ed. Kurstak, E. (Elsevier/North–Holland, Amsterdam), pp. 732–777.
3. Hari, V. (1981) *Virology* **112**, 391–399.
4. Allison, R. F., Sorenson, J. C., Kelly, M. E., Armstrong, F. B. & Dougherty, W. G. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3969–3972.
5. Carrington, J. C. & Dougherty, W. G. (1987) *J. Virol.* **61**, 2540–2548.
6. Carrington, J. C. & Dougherty, W. G. (1987) *Virology* **160**, 355–362.
7. Knuhtsen, H., Hiebert, E. & Purcifull, D. E. (1974) *Virology* **61**, 200–209.
8. Dougherty, W. G. & Hiebert, E. (1980) *Virology* **104**, 174–182.
9. Arnold, E., Luo, M., Vriend, G., Rossman, M. G., Palmenberg, A. C., Parks, G. D., Nicklin, M. J. H. & Wimmer, E. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 21–25.
10. Pirone, T. P. & Thornbury, D. W. (1983) *Phytopathology* **73**, 872–875.
11. Simons, J. N. (1976) *Phytopathology* **66**, 652–654.
12. Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. & Green, M. R. (1984) *Nucleic Acids Res.* **12**, 7035–7056.
13. Dougherty, W. G. & Hiebert, E. (1980) *Virology* **101**, 466–474.
14. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
15. Kelly, C., Totty, N. F., Waterfield, M. D. & Crumpton, M. J. (1983) *Biochem. Int.* **6**, 525–544.
16. Allison, R. F., Dougherty, W. G., Parks, T. D., Willis, L., Johnston, R. E., Kelly, M. & Armstrong, F. B. (1985) *Virology* **147**, 309–316.
17. Nicklin, M. J. H., Krausslich, H. G., Toyoda, H., Dunn, J. J. & Wimmer, E. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4002–4006.
18. Ypma-Wong, M. F. & Semler, B. L. (1987) *Nucleic Acids Res.* **15**, 2069–2088.
19. Ypma-Wong, M. F. & Semler, B. L. (1987) *J. Virol.* **61**, 3181–3189.
20. 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. & Rhoads, R. E. (1986) *Nucleic Acids Res.* **14**, 5417–5430.
21. Berger, A. & Schechter, I. (1970) *Philos. Trans. R. Soc. London Ser. B* **257**, 249–264.
22. Drenth, J., Kalk, K. H. & Swen, H. M. (1976) *Biochemistry* **15**, 3731–3738.
23. Lowe, G. & Yuthavong, Y. (1971) *Biochem. J.* **124**, 107–115.
24. Schechter, I. & Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27**, 157–162.
25. Schechter, I. & Berger, A. (1968) *Biochem. Biophys. Res. Commun.* **32**, 898–902.
26. Dougherty, W. G., Carrington, J. C., Cary, S. M. & Parks, T. D. (1988) *EMBO J.*, in press.
27. Nicklin, M. J. H., Toyoda, H., Murray, M. G. & Wimmer, E. (1986) *Bio/Technology* **4**, 33–44.
28. Wellink, J., Rezelman, G., Goldbach, R. & Beyreuther, K. (1986) *J. Virol.* **59**, 50–58.