Mutational Analysis of Tobacco Etch Virus Polyprotein Processing: *cis* and *trans* Proteolytic Activities of Polyproteins Containing the 49-Kilodalton Proteinase†

JAMES C. CARRINGTON, SUSAN M. CARY, AND WILLIAM G. DOUGHERTY*

Department of Microbiology, Oregon State University, Corvallis, Oregon 97331-3804

Received 21 December 1987/Accepted 15 March 1988

The genome of tobacco etch virus contains a single open reading frame with the potential to encode a 346-kilodalton (kDa) polyprotein. The large polyprotein is cleaved at several positions by a tobacco etch virus genome-encoded, 49-kDa proteinase. The locations of the 49-kDa proteinase-mediated cleavage sites flanking the 71-kDa cytoplasmic pinwheel inclusion protein, 6-kDa protein, 49-kDa proteinase, and 58-kDa putative polymerase have been determined by using cell-free expression, proteolytic processing, and site-directed mutagenesis systems. Each of these sites is characterized by the conserved sequence motif Glu-Xaa-Xaa-Tyr-Xaa-Gln-Ser or Gly (in which cleavage occurs after the Gln residue). The amino acid residue (Gln) predicted to occupy the -1 position relative to the scissile bond has been substituted, by mutagenesis of cloned cDNA, at each of four cleavage sites. The altered sites were not cleaved by the 49-kDa proteinase. A series of synthetic polyproteins that contained the 49-kDa proteinase linked to adjoining proteins via defective cleavage sites were expressed, and their proteolytic activities were analyzed. As part of a polyprotein, the proteinase was found to exhibit *cis* (intramolecular) and *trans* (intermolecular) activity.

The 9.5-kilobase, single-stranded RNA genome of tobacco etch virus (TEV), a plant potyvirus, contains a single open reading frame with the potential to encode a 346-kilodalton (kDa) polyprotein (1). The genome is poly(A)⁺ at the 3' terminus (2, 10), covalently linked to a 6-kDa protein termed VPg (9), and encapsidated in a helical fashion by ~2,000 copies of a 30-kDa capsid protein subunit (11). Mature TEV structural and nonstructural proteins originate by proteolytic processing of the precursor polyprotein (1, 2, 4). Several TEV-encoded proteins have been identified and isolated from infected plants, including the 71-kDa (formerly referred to as the 70-kDa) cytoplasmic pinwheel inclusion protein, VPg, capsid protein, and two proteins (49 and 58 kDa) which form the nuclear inclusion body (see Fig. 1A for genome map). The 49-kDa protein is a TEV polyprotein-specific proteinase (4), whereas the 58-kDa protein (formerly referred to as the 54-kDa protein) has been hypothesized to function as an RNA-dependent, RNA polymerase (1, 5).

Most of the TEV polyprotein cleavage events are thought to be catalyzed by the 49-kDa proteinase. This proteinase functions in *cis* (intramolecular) to autoexcise from the polyprotein and in *trans* (intermolecular) at additional precursor cleavage sites (3, 4). The sites proposed to react with the 49-kDa proteinase in the TEV polyprotein form the putative junctions between the 50-71-, 71-6-, 6-49-, 49-58-, and 58-30-kDa-proteins (see Fig. 1A). These protein junctions are characterized by the sequence motif Glu-Xaa-Xaa-Tyr-Xaa-Gln-Ser or Gly, in which the scissile bond occurs after the Gln residue (-1 position) and the Xaa positions are occupied by hydrophobic or neutral residues (4). Using the 58-30-kDa polyprotein cleavage site as a model, the conserved residues within this motif have been shown to be critical for optimal substrate activity in vitro (W. G. Dougherty, J. C. Carrington, S. M. Cary, and T. D. Parks, EMBO J., in press). In addition, synthetic proteins which contain

insertions (via recombinant DNA techniques) of the conserved seven-amino-acid sequence have been shown to be reactive with the 49-kDa proteinase (J. C. Carrington, and W. G. Dougherty, Proc. Natl. Acad. Sci. USA, in press).

In other viral systems, the sites of cleavage within a polyprotein have been identified by amino- or carboxyl-terminal sequence analysis of the mature (cleaved) proteins isolated from infected cells or virus particles. For TEV, the position of the 58-30-kDa capsid protein junction has been identified in this manner by using capsid protein isolated from virions (2). However, attempts to sequence the amino-terminal residues of the 49-kDa proteinase, the 58-kDa putative polymerase, and the 71-kDa cytoplasmic inclusion protein isolated from infected plants have been unsuccessful presumably due to the presence of a chemical block or modification (W. G. Dougherty, unpublished observations). Therefore, we have pursued alternative routes to identify processing sites in the TEV polyprotein. For example, we have synthesized defined segments of the TEV polyprotein by using SP6 transcription of cloned cDNA and cell-free translation, and we have assayed these synthetic proteins for their reactivity with the 49-kDa proteinase (3, 4). We also have conducted amino-terminal microsequence analyses with radiolabeled protein products generated in the cell-free expression and processing reactions (3; this study). The validity of these approaches has been confirmed by the demonstration that proteolysis in vivo and in vitro at the 58-30-kDa capsid protein junction generates a 30-kDa capsid protein with the same amino-terminal sequence (3).

This study was initiated for three reasons. First, additional evidence could be acquired to support the proposed functionality of the five conserved amino acid sequence motifs (see above) as polyprotein cleavage sites. Second, the activities of the 49-kDa proteinase could be tested while still part of a polyprotein by inactivating one or both cleavage sites flanking the proteinase. Third, the order or pathway of *cis* cleavage events could be examined in vitro. We introduced site-directed mutations into TEV cDNA sequences, specify-

* Corresponding author.

† Oregon Agricultural Experiment Station Technical Paper 8485.

ing the -1 position at the putative 50-71-, 71-6-, 6-49-, and 49-58-kDa protein cleavage sites (Fig. 1), and assayed their effects by using a cell-free expression system (4). The results have shown that all five sites in the polyprotein which conform to the conserved sequence motif are functional as proteolytic substrate sites, that the TEV 49-kDa proteinase is catalytically active within a polyprotein, and that cleavage in *cis* at the 49-kDa proteinase amino- and carboxyl-terminal borders appears not to require a specific sequential order in vitro.

MATERIALS AND METHODS

Materials. Recombinant plasmids contained cDNA synthesized from RNA of the highly aphid-transmissible isolate of TEV (17). Nuclear inclusion bodies, used as the source of proteinase in several experiments, were isolated (7) from *Datura stramonium* plants infected with the non-aphid-transmissible isolate of TEV (19). Recombinant plasmids were propagated in either *Escherichia coli* HB101 (for routine subcloning) or strains TG1 or MV1190 (for preparation of single-stranded DNA for oligonucleotide-directed mutagenesis). Enzymes and chemicals used for general DNA manipulations were purchased from Bethesda Research Laboratories, Inc., New England BioLabs, Inc., Boehringer Mannheim Biochemicals, Promega-Biotec, and Sigma Chemical Co. Reagents for site-directed mutagenesis were purchased from Amersham Corp. Radioisotopes were obtained from New England Nuclear Corp., and rabbit reticulocyte lysate was purchased from Green Hectares.

Recombinant DNA plasmids. Most TEV cDNA segments used for cell-free expression were inserted into vector pTL-8 (4), which contains cDNA representing the 5'-proximal 205 nucleotides (nt; consisting of a 5'-noncoding region of 145 nt, the ATG start codon for the large polyprotein, and the initial 60 nt of the viral open reading frame) of the TEV genome. In pTL-8, this leader segment is positioned downstream from an SP6 promoter and upstream from a multiple cloning site, allowing insertion of foreign DNA sequences and their subsequent expression via SP6 transcription and cell-free translation (using the TEV start codon). Plasmids pTL-5473 (containing cDNA representing TEV genome nts 5412 to 7285; Fig. 1A), pTL-5495 (nts 5412 to 9495), and pTL-8595 (nts 8462 to 9495) have been described previously (4). Plasmid pTL-3472 was assembled by insertion of a TEV cDNA *ScaI-SalI* (nts 3434 to 7171) fragment into pTL-8 between the *AvaI* (filled-in) and *SalI* sites in the polylinker.

The codon specifying Gln in the -1 position (CAA in each case) at each of the three proposed (4) cleavage sites in pTL-5473 was altered by oligonucleotide-directed mutagenesis, using the method of Taylor et al. (20). At the sequence encoding the proposed 71-6-kDa junction, CAA (Gln) was converted to TTA (Leu; Fig. 1A); pTL-5473 derivatives that contained this mutation were given the suffix 16-7. The CAA (Gln) codon specifying the -1 position at the 6-49-kDa and 49-58-kDa protein boundaries was converted to CCC (Pro), yielding plasmid derivatives with the suffixes 11-N and 12-C, respectively (Fig. 1A). The CAG (Gln) codon specifying the -1 position at the proposed 50-71-kDa protein cleavage site was converted to CGT (Arg) in pTL-3472, resulting in a plasmid with the suffix 17-1. To generate each mutagenesis template, cDNA from the parent plasmid (either pTL-5473 or pTL-3472) was subcloned in pUC118, propagated in *E. coli* TG1 or MV1190, and single-stranded DNA was produced by transfection with the helper bacteriophage M13K07. Details of the mutagenesis protocol have been

FIG. 1. TEV genomic map and expression of pTL-5473 cleavage site mutants. (A) Schematic diagram of the TEV genome and the segment represented as cDNA in plasmid pTL-5473 (enlarged). The TEV genome (9.5 kilobases) is covalently linked to a small protein (VPg, black box). The long open reading frame encoding the 346-kDa polyprotein is indicated by the rectangular box. Delineations within the box indicate the positions of coding regions for the individual TEV proteins. cDNA representing TEV genome nts 5412 to 7285 has been cloned in pTL-8 to generate pTL-5473 (4). The coding sequences for three putative cleavage sites, centered on Gln (Q)-Ser (S) or Gln-Gly (G) dipeptides, in this part of the polyprotein as well as for amino acid residues Cys (C) and His (H) at the putative active center of the 49-kDa proteinase (stippled box) are shown above the cDNA schematic. The Gln residue at the predicted -1 position of the 71-6-, 6-49-, and 49-58-kDa polyprotein cleavage junctions were replaced by Leu (L) or Pro (P) via site-directed mutagenesis of pTL-5473, as indicated below the diagram (mutant codes in parentheses). The Cys-to-Ala (A) mutation at the putative active site of the 49-kDa proteinase has been described previously (4). (B) Expression of pTL-5473 (lane 1) and five mutagenized derivatives of pTL-5473 (lanes 2 to 6) by SP6 transcription and cell-free translation in the presence of [³⁵S]Met. Plasmid DNA was linearized with *PvuII* (cleaves within vector sequences approximately 50 base pairs from the 3' end of the cDNA insert) before the transcription reaction. DNA in pTL-5473 derivatives C-A, 11-N, and 12-C contained single point mutations at the positions indicated in panel A, whereas derivatives 11-N/12-C and 11-N/12-C/16-7 contained two and three mutations, respectively. Radiolabeled products were analyzed by SDS-polyacrylamide gel electrophoresis and fluorography. Molecular sizes (in kilodaltons) are presented at the left.

described previously (3). Incorporation of the correct mutant sequence was verified by dideoxynucleotide sequence analysis. The mutagenized DNA was then subcloned back into pTL-5473 or pTL-3472 to allow cell-free expression. Plasmid pTL-5473/C-A, harboring the mutation which resulted in a Cys-to-Ala substitution at the putative active site of the 49-kDa proteinase, has been described previously (3).

Cell-free transcription, translation, and proteolytic processing. Reaction conditions for transcription with SP6 polymerase, translation of the synthetic transcripts in an mRNA-dependent reticulocyte lysate, and proteolytic processing of the translation products, using the 49-kDa proteinase activity in either nuclear inclusion bodies or synthesized *in vitro*, have all been described previously (3, 4, 8, 16). Unless stated otherwise, plasmid DNA was linearized with *PvuII* (which cleaves within vector sequences) before SP6 transcription. Proteins were analyzed by electrophoresis in sodium dodecyl sulfate (SDS)-polyacrylamide (12.5% resolving) gels, using a discontinuous buffer system (14) followed by fluorography.

Immunoprecipitation and protein microsequence analysis. Immunoprecipitation of cell-free translation products with immunoglobulin G (IgG) (purified by chromatography on protein A-Sepharose) monospecific for the 71-kDa cytoplasmic inclusion protein or the 49-kDa proteinase was by the method of Kessler (13) as modified by Dougherty and Hiebert (8). Antibody preparations used in this study were produced in rabbits and have been described elsewhere (4, 7).

Protein microsequence analysis was conducted as previously described (3). Radiolabeled protein products were electroeluted from SDS-polyacrylamide preparative gels (12) and subjected to automated Edman degradation by using an Applied Biosystems Sequenator. Radioactivity released at each cycle was quantitated with a Beckman scintillation counter.

RESULTS

Mutations affecting the 49-kDa proteinase and 6-kDa protein cleavage sites. Plasmid pTL-5473 contains a TEV cDNA insert which spans the 49-kDa coding sequence as well as flanking regions on both the 5' and 3' sides (Fig. 1A). The pTL-5473 insert has the potential to encode a synthetic polypeptide of 75 kDa; however, translation of pTL-5473 transcripts results in the accumulation of a 49-kDa species which (i) forms as the result of autoproteolysis, (ii) exhibits proteolytic activity, and (iii) is indistinguishable by several criteria from the authentic TEV 49-kDa proteinase isolated from infected plants (3, 4). The cleavage site at the amino terminus of the 49-kDa proteinase, inferred from microsequence analysis of protein synthesized and processed *in vitro*, occurs between a Gln-Gly dipeptide (position 1849-1850 in the TEV polyprotein) that resides within the context of the conserved sequence motif Glu-Xaa-Xaa-Tyr-Xaa-Gln-Ser or Gly (see Introduction). In addition, cleavage of the pTL-5473-derived polypeptide has been predicted to occur between two other dipeptides (Gln-Ser at position 1796-1797 and Gln-Gly at position 2279-2280) on the basis of their appearance within the conserved motif (4). The Gln-Gly (2279-2280) site may form the junction between the 49-kDa proteinase and the 58-kDa protein, whereas the Gln-Ser (1796-1797) dipeptide may be cleaved to generate the amino terminus of a 6-kDa protein (Fig. 1A). To date, attempts to sequence the amino-terminal residues of the 58-kDa protein isolated from plants or synthesized and processed *in vitro* have been unsuccessful.

To provide evidence that all three predicted cleavage sites in the pTL-5473-derived polypeptide are functional and to test the activities of the 49-kDa proteinase within a variety of noncleaved precursors, mutations were introduced to convert the codon specifying Gln at the predicted -1 position of all three sites to either Leu or Pro. Transcripts were generated from the mutagenized plasmid DNA and translated in the reticulocyte lysate, and the resulting proteins were analyzed relative to products synthesized from transcripts of pTL-5473 and pTL-5473/C-A. Unaltered pTL-5473 transcripts yielded the 49-kDa proteinase on cell-free translation and subsequent autoproteolysis (Fig. 1B, lane 1) as well as products which were not resolved in our gel system due to their small size. Translation of transcripts from pTL-5473/C-A, containing a mutation affecting the putative active site of the proteinase (3), resulted in accumulation of the 75-kDa precursor (Fig. 1B, lane 2).

Plasmid pTL-5473/11-N contained a mutation that converted Gln to Pro at the -1 position relative to the 6-49-kDa proteinase junction (Fig. 1A). Expression of this construct in the cell-free system resulted predominantly in the accumulation of a 55-kDa product with the apparent size of a 6-49-kDa noncleaved precursor (Fig. 1B, lane 3). In addition, a minor amount of a 63-kDa product was detected which possessed the size of a precursor containing the entire predicted polypeptide except the segment at the carboxyl terminus derived from the 58-kDa protein. Plasmid pTL-5473/12-C contained a mutation that converted Gln to Pro at the 49-kDa proteinase-58-kDa protein cleavage site. Expression of this DNA by transcription and translation yielded a protein product with an apparent molecular size of 62 kDa, corresponding to the predicted size of the 49-kDa proteinase linked to the 58-kDa amino-terminal sequence (Fig. 1B, lane 4).

The mutations altering the amino- and carboxyl-terminal cleavage sites of the 49-kDa proteinase were linked in the same cDNA molecule to generate the double mutant pTL-5473/11-N/12-C (Fig. 1A). Expression of this clone in the cell-free system resulted in the accumulation of a 68-kDa species, apparently representing a polypeptide consisting of the 6-49-58-kDa fragment formed by cleavage only at the 6-kDa amino-terminal processing site in the 75-kDa polypeptide precursor (Fig. 1B, lane 5). A triple mutant, pTL-5473/11-N/12-C/16-7, which bore mutations affecting both 49-kDa proteinase-terminal cleavage sites as well as the 6-kDa amino-terminal site was assembled (Fig. 1A). Cell-free translation of transcripts derived from this DNA resulted in the accumulation of the complete 75-kDa precursor. Two mutations (11-N and 12-C) also were incorporated into pTL-5495, which contains cDNA starting at the same 5' position as pTL-5473 but continues through to the genome 3' terminus (i.e., it contains cDNA representing the entire 6-, 49-, 58-, and 30-kDa coding regions; Fig. 1A). Expression of pTL-5495/11-N/12-C yielded products which were the apparent result of cleavage at the 6-kDa amino terminus and at the 58-30-kDa cleavage site (data not shown). In addition, mutation 16-7 by itself was inserted into pTL-5473. Translation of transcripts from this plasmid resulted in the accumulation of the 49-kDa proteinase (data not shown), indicating that a defect at the 71-6-kDa protein cleavage junction has little if any effect on processing at the 49-kDa proteinase boundaries. Most or all of the minor radiolabeled proteins shown in Fig. 1B appeared to be the result of premature translational termination events.

To test *in trans* the proteolytic activities of the 49-kDa proteinase-containing polypeptides formed as the result of

FIG. 2. *trans*-proteolytic activities of the cleavage site-defective polyproteins containing the 49-kDa proteinase. The 34-kDa, [35 S]Met-labeled substrate was synthesized by translation of pTL-8595 transcripts. The substrate represents the carboxyl terminus of the TEV polyprotein, consisting of a short segment of the 58-kDa protein, a cleavage site, and the entire 30-kDa capsid protein (see Fig. 1A). Radiolabeled 34-kDa substrate and nonradiolabeled proteinase sources were synthesized individually in the cell-free translation system. Translations were terminated by the addition of RNase A (100 μ g/ml). Phenylmethylsulfonyl fluoride (3 mM final concentration) was added to both proteinase and substrate sources to reduce nonspecific degradation by potential reticulocyte proteinases. Substrate was added to each of the proteinase sources at a molar ratio of approximately 5 to 1, and the mixtures were incubated at 30°C for 1.5 h. Proteinase sources were H₂O (lane 1) and products synthesized from transcripts of *Pvu*II-linearized pTL-5473 (lane 2) or from mutagenized pTL-5473 derivatives C-A (lane 3), 11-N (lane 4), 12-C (lane 5), 11-N/12-C (lane 6), and 11-N/12-C/16-7 (lane 7). Substrate precursor and product were analyzed by SDS-polyacrylamide gel electrophoresis and fluorography.

altered cleavage sites, nonradioactive translation products expressed from each construct (described above) were reacted with the [35 S]Met-labeled, 34-kDa precursor substrate synthesized from pTL-8595 transcripts (4). This synthetic substrate contains sequences from the carboxyl-terminal region of the TEV 346-kDa polyprotein, consisting of a short segment of the 58-kDa protein, a cleavage site, and the entire 30-kDa capsid protein. As in previous studies, reaction of the 34-kDa substrate with the 49-kDa proteinase synthesized by translation of pTL-5473 transcripts generated the 30-kDa capsid protein (Fig. 2, lane 2) (3, 4). The 34-kDa precursor remained intact after reacting with either H₂O (lane 1) or the translation products from pTL-5473/C-A (proteinase-active site mutant) transcripts (lane 3). Reaction of the 34-kDa substrate with the products derived from the cleavage site mutants described above in each case resulted in the accumulation of the 30-kDa species (Fig. 2, lanes 4 to 7), indicating that each non- or partially processed proteinase-containing polyprotein was active in *trans*.

A summary of the *cis*- and *trans*-proteolytic activities associated with the mutant polyproteins and a schematic of the processed products are shown in Fig. 3. All three predicted cleavage sites within the pTL-5473-derived polyprotein were found to exhibit substrate activity in the cell-free reactions. Each of these cleavage events was inhibited by substituting Pro or Leu for Gln at the proposed -1 position relative to the scissile bond. However, proteolytic activity in *cis* at the unmodified sites and in *trans* was not

FIG. 3. Summary of the *cis*- and *trans*-proteolytic activities associated with the 49-kDa proteinase-containing, defective polyproteins described in Fig. 1. Schematic diagrams of the coding regions (expanded region) in seven plasmid constructs are shown as straight lines. Positions of sequences coding for potential cleavage sites are indicated by short vertical dashes; mutagenized sites are indicated by the asterisks or by the arrow (for pTL-5473/C-A). Apparent products formed by cell-free translation of transcripts and subsequent proteolytic processing are shown by the open rectangular boxes beneath the plasmid construct diagrams. The proteolytic activities in *cis* at the potential cleavage sites encoded by pTL-5473 derivatives or by pTL-5495/11-N/12-C, and in *trans* at the cleavage site within the 34-kDa substrate encoded by the pTL-8595 transcripts, are presented in tabular form at the left.

FIG. 4. Immunoprecipitation of products expressed from pTL-5473/C-A and pTL-5473/11-N/12-C. [35 S]Met-labeled products programmed with C-A (lanes 1, 3, and 4) or 11-N/12-C (lanes 2, 5, and 6) transcripts were either untreated (lanes 1 and 2) or immunoprecipitated with IgG specific for the 49-kDa proteinase (lanes 3 and 5) or the 71-kDa cytoplasmic pinwheel inclusion body protein (lanes 4 and 6). Note that the 68-kDa product processed from the 11-N/12-C precursor (see Fig. 1A and 3) was not precipitated by the anti-71-kDa protein IgG. Radiolabeled proteins were analyzed by SDS-polyacrylamide gel electrophoresis and fluorography.

affected by linkage of amino acid sequences to the amino and carboxyl termini of the 49-kDa proteinase.

Mapping the carboxyl terminus of the 71-kDa cytoplasmic pinwheel inclusion protein. The cleavage site mutants described above have been employed to map the carboxyl terminus of the 71-kDa cytoplasmic inclusion protein. Transcripts from pTL-5473/C-A and pTL-5473/11-N/12-C were translated to give rise to products of 75 kDa (noncleaved polyprotein) and 68 kDa (once-cleaved polyprotein, at the 6-kDa amino terminus), respectively. These polyproteins were immunoprecipitated with IgG monospecific for the 49-kDa proteinase or the 71-kDa pinwheel inclusion protein. The 75-kDa polyprotein (from pTL-5473/C-A transcripts) was precipitated with both IgG preparations (Fig. 4, lanes 3 and 4). However, the 68-kDa polyprotein (from pTL-5473/11-N/12-C transcripts) was precipitated selectively by anti-49-kDa proteinase and not by anti-71-kDa protein IgG (Fig. 4, lanes 5 and 6), indicating that all epitopes recognized by the anti-71-kDa protein IgG were removed by cleavage at the proposed 6-kDa amino terminus. This suggests that the carboxyl terminus of the 71-kDa protein and the amino terminus of the proposed 6-kDa protein are linked by the Gln-Ser dipeptide at position 1796-1797 in the TEV polyprotein.

Mapping the amino terminus of the 71-kDa cytoplasmic pinwheel inclusion protein. The deduced amino acid sequence at positions 1158 to 1164 in the TEV polyprotein is Glu-Ile-Ile-Tyr-Thr-Gln-Ser, and contains the conserved residues (underlined) found at the other 49-kDa proteinase-mediated cleavage sites. Given its position in the TEV polyprotein relative to the proposed 71-kDa carboxyl-terminal cleavage site (see above), the Gln-Ser peptide bond in this sequence was a likely candidate to be cleaved to form the 71-kDa protein amino terminus. To address this, we assembled the plasmid pTL-3472 by inserting cDNA predicted to encode part of the 50-kDa protein, and the entire 71-, 6-, and 49-kDa proteins, into pTL-8. When transcripts encompassing this entire region were translated in the reticulocyte lysate, polypeptide synthesis was found to terminate abruptly at a position approximately three-quarters through the 71-kDa coding sequence (unpublished data). This termination occurred during translation of transcripts from two independent cDNA clones spanning this region. Since the

FIG. 5. Cell-free expression of *Ball*-linearized pTL-3472 and the mutagenized derivative pTL-3472/17-1. (A) The TEV genome segment expressed by using *Ball*-linearized pTL-3472 is shown in the enlarged diagram. The position of the sequence coding for the putative dipeptide delineating the 50-71-kDa proteins [Gln-Ser (Q-S)] is shown above, and the amino acid sequence alteration [Gln to Arg (R) in plasmid pTL-3472/17-1] induced by site-directed mutagenesis is shown below the diagram. The shaded region is predicted to form the amino-terminal portion (41 kDa) of the 71-kDa protein. (B) Cell-free synthesis and processing of synthetic polyproteins encoded by pTL-3472 (lanes 1 and 2) and 3472/17-1 (lanes 3 and 4). [35 S]Met-labeled proteins were synthesized from *Ball*-linearized transcripts and reacted with H_2O (lanes 1 and 3) or partially purified nuclear inclusion bodies (1 mg/ml; lanes 2 and 4) containing the 49-kDa proteinase activity. Radiolabeled proteins were analyzed by SDS-polyacrylamide gel electrophoresis and fluorography.

nucleotide sequence through this portion of the TEV genome contains an uninterrupted open reading frame devoid of in-frame stop codons, the reason for premature termination was not clear and has not been investigated further. Therefore, a shorter segment of the TEV polyprotein, predicted to encompass the 71-kDa protein amino-terminal cleavage site, was synthesized by the translation of SP6 transcripts from *Ball*-linearized pTL-3472 (Fig. 5A). The synthetic polyprotein specified by these transcripts was predicted to contain domains of 10 and 41 kDa on the amino and carboxyl sides of the proposed cleavage site, respectively.

Cell-free translation of *Ball*-linearized, pTL-3472 transcripts yielded a product with an apparent molecular size of 51 kDa (Fig. 5B, lane 1), as predicted from the deduced amino acid sequence. Incubation of the 51-kDa synthetic

product with partially purified TEV nuclear inclusion bodies (as the source of the 49-kDa proteinase activity) resulted in processed products with approximate molecular sizes of 41 and 10 kDa (Fig. 5B, lane 2). The 41-kDa product was immunoprecipitated with anti-71-kDa protein IgG, whereas the 10-kDa product was not (data not shown). A site-directed mutation, changing the codon specifying Gln (CAG) to Arg (CGT) at the -1 position relative to the proposed cleavage site was introduced into pTL-3472 to form pTL-3472/17-1. Translation of pTL-3472/17-1 transcripts resulted in the formation of a protein with a size similar to that of the 51-kDa product derived from the parent clone (Fig. 5B, lane 3). However, the mutant protein was completely insensitive to cleavage by the 49-kDa proteinase (Fig. 5B, lane 4). These data suggested that 49-kDa proteinase-mediated cleavage of the TEV polyprotein occurred between amino acid residues 1163 and 1164.

To map definitively the position of cleavage of the pTL-3472-derived polyprotein, amino-terminal microsequence analysis was conducted on the large cleavage product released from [35 S]Met- or [3 H]Leu-labeled precursor generated from *Nhe*I-linearized pTL-3472 transcripts. Restriction enzyme *Nhe*I cleaves pTL-3472 at a position 336 base pairs upstream from the *Bal*I site in the cDNA insert (Fig. 5A). Translation of *Nhe*I-linearized transcripts generated a 38-kDa precursor from which products of 28 and 10 kDa were formed by proteolytic processing, using partially purified nuclear inclusion bodies. On the basis of the release of radioactivity at each cycle of the microsequence analysis, [35 S]Met was found to occupy residue positions 13 and 21 relative to the amino terminus of the 28-kDa protein, whereas [3 H]Leu was identified at positions 2, 17, 19, and 24 (Fig. 6). This combination of Met and Leu residues from the amino terminus of the cleavage product could be accounted for only if proteolysis occurred between amino acids 1163 and 1164 in the TEV polyprotein.

DISCUSSION

Proteolytic activities of polyproteins containing the 49-kDa proteinase. Cleavage at the 49-kDa proteinase boundaries within a polyprotein appears to require an efficient autoproteolytic mechanism *in vitro*; these sites are refractory to proteolysis when the 49-kDa proteinase is supplied *in trans* (3, 4). One possible explanation for this observation is that the autoproteolytic cleavage sites are accurately and sequentially presented to the active site by a precise protein-folding arrangement such that exposure of the autoproteolytic sites to exogenous proteinase is restricted. By altering the amino- and carboxyl-terminal cleavage sites flanking the 49-kDa proteinase, the *cis*- and *trans*-cleavage activities of the proteinase linked within a series of polyproteins were tested. Prevention of cleavage at the amino terminus of the proteinase had little effect on cleavage *in cis* at the carboxyl terminus. Likewise, disruption of cleavage at the proteinase carboxyl terminus had little effect on *cis* proteolysis at the amino terminus. It appears that autoproteolytic cleavage at one terminus does not necessarily occur via a pathway dependent on prior cleavage at the other terminus. This may indicate that the positions of the 49-kDa proteinase cleavage sites relative to the enzyme-active center are not locked in a rigid conformation before autoproteolysis; a fixed orientation of one cleavage site at the active center would have resulted in exclusion of other sites.

The cleavage site-defective 49-kDa proteinase-containing polyproteins exhibited *trans* proteolytic activity, using an

FIG. 6. Amino-terminal microsequence analysis of the 28-kDa cleavage product released by proteolysis of the translation product programmed by *Nhe*I-linearized pTL-3472 transcripts. Transcripts were translated in the presence of [35 S]Met or [3 H]Leu. The resulting synthetic polyproteins were reacted with the 49-kDa proteinase present in a partially purified nuclear inclusion body preparation. The cleavage products were resolved by preparative SDS-polyacrylamide gel electrophoresis, and the large product was eluted and subjected to automated Edman degradation. Radioactivity released at each cycle was monitored with a scintillation counter. The amino acid sequence presented above the plots is the deduced sequence (from the nucleotide sequence) within the TEV polyprotein (residues 1164 to 1188) starting at the Ser residue predicted to form the 71-kDa protein amino terminus (see Fig. 5A).

exogenous substrate even when the proteinase was linked within a 113-kDa species consisting of the 6-, 49-, and 58-kDa proteins. This is consistent with results obtained with the 49-kDa proteinase *cis*-cleavage site mutants and further supports the notion that the noncleaved autoproteolytic sites do not mask the active site via a rigid conformation. Also, these results raise the possibility that functional 49-kDa proteolytic activity *in vivo* may be associated with a polyprotein (perhaps consisting of the 49- and 58-kDa proteins) at some point in the cleavage cascade. Proteolytically active, proteinase-containing polyproteins also may function in the poliovirus and cowpea mosaic virus systems. It has been suggested that most or all of the P3 region of the poliovirus polyprotein (containing proteinase 3C and polymerase 3D) may be required to effect proteolysis at sites within the P1 precursor (23). Also, a proteinase-polymerase precursor has been characterized from cowpea mosaic virus-infected cells (6). This issue could be addressed by examination of the proteolytically active form(s) of the 49-kDa proteinase *in vivo* at various stages postinoculation.

Cleavage sites within the TEV polyprotein. Ideally, we would like to obtain protein sequence data for the termini of each TEV protein product isolated from infected plants so that the positions of cleavage within the polyprotein could be inferred. Such a strategy has worked well for the identification of cleavage sites within the polyprotein precursors for

poliovirus (15, 18) and cowpea mosaic virus (22, 24). But since most of the TEV proteins isolated from infected plants are refractory to sequence analysis (see Introduction), our efforts have been diverted to a molecular genetic approach. We have identified and characterized several TEV polyprotein cleavage sites in at least three of four ways. First, synthetic substrate polyproteins have been synthesized from defined transcripts in cell-free reactions, and their substrate activities have been evaluated by using the TEV 49-kDa proteinase. Five cleavage sites in the TEV polyprotein have been identified with this strategy (3, 4; present study). Second, the conserved sequence motif Glu-Xaa-Xaa-Tyr-Xaa-Gln-Ser or Gly is found at each site processed by the 49-kDa proteinase in which cleavage is predicted to occur after the Gln residue. Third, the amino acid residue (Gln) predicted to occupy the -1 position relative to the scissile bond at all five sites has been substituted via site-directed mutagenesis of cloned cDNA. Since each mutation rendered the respective synthetic polyproteins nonreactive with the proteinase, the altered amino acid residue must have been at or very near the point of cleavage. And finally, protein microsequence analysis has been conducted on three products (formed by cleavage at the amino termini of the 30-kDa capsid, 49-kDa proteinase, and 71-kDa protein) generated by proteolysis in the cell-free systems, allowing the precise positions of cleavage to be inferred. Thus, bonds cleaved by the 49-kDa proteinase have been located between the dipeptides at positions 1163-1164 (50-71-kDa protein junction), 1796-1797 (71-6-kDa junction), 1849-1850 (6-49-kDa junction), 2279-2280 (49-58-kDa junction), and 2791-2792 (58-30-kDa junction; see Fig. 1A for relative positions) in the TEV polyprotein.

The presence of a 6-kDa protein encoded by sequences upstream from the 49-kDa proteinase coding region has been postulated in earlier studies, even though its existence has not been definitively linked to any known TEV gene product isolated from infected plants. However, two lines of circumstantial evidence support the notion that this predicted product is formed *in vivo*. First, functional cleavage sites flanking the 6-kDa segment of the polyprotein have been characterized *in vitro* (3; this study). Second, this predicted polypeptide appears not to reside within the 49-kDa proteinase (3) or the 71-kDa protein (this study) sequence on the basis of protein microsequence analysis or immunological reactivity with polyclonal antisera. Given the apparent size (6 kDa) of the VPg linked to the TEV genome, there is a strong possibility that the predicted 6-kDa protein functions as the VPg. Experiments using specific antisera raised against a portion of the predicted 6-kDa protein are currently underway to test this hypothesis.

Proteolysis *in trans* at the 71-6-kDa protein junction could not be detected in prior studies (3, 4) using various synthetic polyprotein substrates and exogenously supplied 49-kDa proteinase. However, this site was processed in the polyproteins synthesized from transcripts of pTL-5473 derivatives, suggesting that the 71-6-kDa boundary may be cleaved by an intramolecular or *cis* mechanism. Efficient *cis* and relatively inefficient *trans* substrate activity is characteristic of the autoproteolytic cleavage sites flanking the 49-kDa proteinase sequence (3, 4). Thus, we propose that three junctions (71-6, 6-49, and 49-58 kDa) in the TEV polyprotein are cleaved *in cis* and two junctions (50-71 and 58-30 kDa) are cleaved by a *trans* proteolytic mechanism.

To date, only five positions in the TEV polyprotein have been identified as cleavage sites recognized by the 49-kDa proteinase. These have been characterized by using a series

of polyprotein substrates synthesized and processed *in vitro*. All five sites reside within the carboxyl-terminal two-thirds of the TEV polyprotein. We have been unable to find additional 49-kDa proteinase-mediated processing sites situated near the amino terminus of the TEV polyprotein, although cleavage is predicted to occur to generate the putative 31- and 56-kDa products (see Fig. 1A). At least one of these uncharacterized processing events may be catalyzed by the TEV 50-kDa protein (encoded by sequences 5' to the 71-kDa cytoplasmic inclusion protein coding region). A limited degree of amino acid sequence similarity has been detected between this protein and the poliovirus proteinase 2A (5, 21). Experiments currently are under way to identify the activity and substrate sites recognized by this potential proteinase.

ACKNOWLEDGMENTS

We thank Reg McParland for conducting the amino acid sequence analysis and T. Dawn Parks and Donna Knauber for comments and advice.

This project was supported by grant DBM-8601939 from the National Science Foundation (to W.G.D.) and by postdoctoral fellowship GM 12492-01 from the National Institutes of Health (to J.C.C.).

LITERATURE CITED

- Allison, R. F., R. E. Johnston, and W. G. Dougherty. 1986. The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for synthesis of a single polyprotein. *Virology* **154**:9-20.
- Allison, R. F., J. C. Sorenson, M. E. Kelly, F. B. Armstrong, and W. G. Dougherty. 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.
- Carrington, J. C., and W. G. Dougherty. 1987. Processing of the tobacco etch virus 49K protease requires autoproteolysis. *Virology* **160**:355-362.
- Carrington, J. C., and W. G. Dougherty. 1987. Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease. *J. Virol.* **61**:2540-2548.
- Domier, L. L., J. G. Shaw, and R. E. Rhoads. 1987. Potyviral proteins share amino acid sequence homology with picorna-, como-, and caulimoviral proteins. *Virology* **158**:20-27.
- Dorssers, L., S. Van der Krol, J. Van der Meer, A. van Kamen, and P. Zabel. 1984. The cowpea mosaic virus B-RNA-encoded 110K polypeptide is involved in viral RNA replication. *Proc. Natl. Acad. Sci. USA* **81**:1951-1955.
- Dougherty, W. G., and E. Hiebert. 1980. Translation of potyvirus RNA in a rabbit reticulocyte lysate: identification of nuclear inclusion proteins as products of tobacco etch virus RNA translation and cylindrical inclusion protein as a product of the potyvirus genome. *Virology* **104**:174-182.
- Dougherty, W. G., and E. Hiebert. 1980. Translation of potyvirus RNA in a rabbit reticulocyte lysate: reaction conditions and identification of the capsid protein as one of the products of *in vitro* translation of tobacco etch and pepper mottle viral RNAs. *Virology* **101**:466-474.
- Hari, V. 1981. The RNA of tobacco etch virus: further characterization and detection of a protein linked to RNA. *Virology* **112**:391-399.
- Hari, V., A. Siegel, C. Rozek, and W. E. Timberlake. 1979. The RNA of tobacco etch virus contains poly(A). *Virology* **92**:568-571.
- Hollings, M., and A. A. Brunt. 1981. Potyviruses, p. 731-807. *In* E. Kurstak (ed.), *Handbook of plant virus infection and comparative diagnosis*. Elsevier/North-Holland Publishing Co., Amsterdam.
- Kelly, C., N. F. Totty, M. D. Waterfield, and M. J. Krumpton.

1983. Amino acid sequencing of polypeptides eluted from 2-D polyacrylamide gels. *Biochem. Int.* **6**:535-544.
13. Kessler, S. 1975. Rapid isolation of antigens from cells with a staphylococcal protein A-antibody adsorbant: parameters of the interaction of antibody-antigen complexes with protein A. *J. Immunol.* **115**:1617-1624.
14. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**:680-685.
15. Larsen, G. R., C. W. Anderson, A. J. Dorner, B. L. Semler, and E. Wimmer. 1982. Cleavage sites within the poliovirus capsid protein precursors. *J. Virol.* **41**:340-344.
16. Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**:7145-7156.
17. Pirone, T. P., and D. W. Thornbury. 1983. Role of virion and helper component in regulating aphid transmission of tobacco etch virus. *Phytopathology* **73**:872-875.
18. Semler, B. L., R. Hanecak, C. W. Anderson, and E. Wimmer. 1981. Cleavage sites in the polyprotein precursors of poliovirus protein P2-X. *Virology* **114**:589-594.
19. Simons, J. N. 1976. Aphid transmission of a nonaphid-transmissible strain of tobacco etch virus. *Phytopathology* **66**:652-654.
20. Taylor, J. W., J. Ott, and F. Eckstein. 1985. The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA. *Nucleic Acids Res.* **13**:8765-8785.
21. Toyoda, H., M. J. H. Nicklin, M. G. Murray, C. W. Anderson, J. J. Dunn, F. W. Studier, and E. Wimmer. 1986. A second virus-encoded protease involved in proteolytic processing of poliovirus polyprotein. *Cell* **45**:761-770.
22. Wellink, J., G. Rezelman, R. Goldbach, and K. Beyreuther. 1986. Determination of the proteolytic processing sites in the polyprotein encoded by the bottom-component RNA of cowpea mosaic virus. *J. Virol.* **59**:50-58.
23. Ypma-Wong, M. F., and B. L. Semler. 1987. *In vitro* molecular genetics as a tool for determining the differential cleavage specificities of the poliovirus 3C proteinase. *Nucleic Acids Res.* **15**:2069-2088.
24. Zabel, P., M. Moerman, G. Lomonosoff, M. Shanks, and K. Beyreuther. 1984. Cowpea mosaic virus VPg: sequencing of radiochemically modified protein allows mapping of the gene on B-RNA. *EMBO J.* **3**:1629-1634.