The EMBO Journal vol.7 no.5 pp.1281 – 1287, 1988

Biochemical and mutational analysis of a plant virus polyprotein cleavage site

William G.Dougherty, James C.Carrington, Susan M.Cary and T.Dawn Parks

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

Communicated by T.Hohn

The RNA genome of tobacco etch virus (TEV) is organized as a single translational unit coding for a 346 000 (346 kd) mol. wt (M_r) polyprotein. The 346 kd M_r polyprotein is cleaved by a 49 kd M_r virus-encoded proteinase at five different sites between the dipeptides Gln-Ser or Gln-Gly. These cleavage sites or gene product boundaries are defined by the heptapeptide sequence . . . Glu-Xaa-Xaa-Tyr-Xaa-Gln-Ser or Gly . . .. We have used the 54 kd M_r nuclear inclusion protein/30 kd M_r capsid protein junction as a model to examine the role of these conserved amino acids in defining a cleavage site. The 54 kd/30 kd M_r protein cleavage site sequence of 10 TEV isolates from geographically distinct locations has been deduced. The conserved amino acids are present in all isolates. To determine if these four amino acids are an absolute requirement for polyprotein substrate activity, a site-directed mutational analysis has been performed. A recombinant cDNA molecule encoding the TEV 54 kd/30 kd M_r gene product cleavage site was mutated and polyprotein substrates were synthesized and processed in a cell-free system. Single amino acid substitutions made at the different positions reveal a strong preference for the naturally conserved amino acids.

Key words: genome expression/potyvirus/site-directed mutagenesis/tobacco etch virus/virus proteinase

Introduction

A number of plant and animal viruses code for at least one proteinase which processes a primary translation product to a mature product. Viruses possessing a single-stranded RNA genome, such as the potyviridae, comoviridae and picornaviridae, code for a cysteine-like proteinase which catalyzes most of the specific cleavages of the genome-derived polyprotein to functional gene products (Carrington and Dougherty, 1987a; Verver *et al.*, 1987; Hanecak *et al.*, 1982; Palmenberg, 1987 for review).

Tobacco etch virus (TEV), a member of the potato virus Y group (potyviridae), is a flexuous, rod-shaped virus which infects a number of solanaceous hosts (Purcifull and Hiebert, 1982). The single-stranded RNA genome is 9496 nucleotides in length, has a 6000 (6 kd) mol. wt (M_r) protein (VPg) covalently attached to the genome, and a polyadenylated region at the 3′ terminus (Allison *et al.*, 1986; Hari, 1981). The genome is organized as a single translational unit capable of being expressed as a 346 kd M_r polyprotein. A genomic organization has been proposed for TEV; the sequences coding for the 49 kd M_r proteinase, a putative 54 kd M_r replicase and the 30 kd M_r capsid protein have been mapped in this order. Additionally, it is likely that the coding sequence for the VPg is contiguous with the proteinase gene (Allison *et al.*, 1986; Carrington and Dougherty, 1987a). Therefore, TEV is similar to the picornaviridae and comoviridae in its genetic organization.

During TEV genome expression, the 49 kd M_r proteinase autocatalytically releases from the polyprotein and cleaves in *trans* at three other sites to yield five virus gene products. In the polyprotein, these gene product boundaries are delineated by the sequence . . . Glu-Xaa-Xaa-Tyr-Xaa-Gln-Ser or Gly . . .. Cleavage occurs between the Gln-Ser or Gln-Gly dipeptides (Carrington and Dougherty, 1987a,b). Furthermore, insertion at a foreign site of an oligonucleotide which encodes the seven amino acids found at the 54 kd/30 kd M_r protein boundary (. . . Glu-Asn-Leu-Tyr-Phe-Gln-Ser . . .) generates a functional cleavage site (Carrington and Dougherty, in press).

Late in TEV infection, inclusion bodies form within the nucleus of infected cells (Knuhtsen *et al.*, 1974; Christie and Edwardson, 1977). These nuclear inclusion bodies can be purified to near homogeneity and consist of two virus-encoded proteins (54 kd and 49 kd M_r) in approximately equimolar amounts (Knuhtsen *et al.*, 1974; Dougherty and Hiebert, 1980b). The TEV 49 kd M_r nuclear inclusion protein is the virus-specific proteinase and nuclear inclusion bodies are convenient sources of proteolytic activity (Carrington and Dougherty, 1987a).

This study was conducted to test the hypothesis that the amino acid residues at positions −6, −3, −1 and +1 are required to define a functional TEV cleavage site. (The numbering convention is relative to the scissile bond.) We have used the TEV 54 kd M_r nuclear inclusion protein/30 kd M_r capsid protein gene boundary as our prototype. Two studies have been conducted. First, we have deduced the amino acid sequence of this gene product junction in 10 different TEV isolates to determine the extent of sequence conservation. These amino acids are conserved in all isolates examined. Second, this sequence has been altered by site-directed mutagenesis of the TEV recombinant cDNA molecule pTL-8595 which encodes the 54 kd/30 kd M_r protein cleavage site (Carrington and Dougherty, 1987a). Polyprotein substrates were expressed from the altered cDNA sequence in a cell-free system, and their substrate activities using the TEV 49 kd M_r proteinase were quantitated. Substitutions at the −1 position (Gln) significantly reduced or eliminated cleavage. Changes at the other conserved positions on the amino-terminal side of the peptide bond cleaved [−6(Glu) and −3(Tyr)], diminished or eliminated the site as a processing site, while the +1 position (Ser) tolerated a number of different amino acid substitutions.

Table I. Deduced amino acid sequence of TEV polyprotein cleavage sites processed by the TEV 49 kd proteinase[a]

Protein boundary	Amino acid sequence					
	Cleavage site					
	−10	−5	−1	+1	+5	+10
54 kd−30 kd	-I-P-T-T-E-N-L-Y-F-Q			S-G-T-V-D-A-G-A-D-A-		
49 kd−54 kd	-Q-L-M-N-E-L-V-Y-S-Q			G-E-K-R-K-W-V-V-E-A-		
6 kd−49 kd	-D-K-F-N-E-P-V-Y-F-Q			G-K-K-N-Q-K-H-K-L-K-		
70 kd− 6 kd	-F-N-H-L-E-T-I-Y-L-Q			S-D-S-E-V-A-K-H-L-K-		
50 kd−70 kd	-S-T-E-R-E-I-I-Y-T-Q			S-L-D-D-Y-V-T-T-F-D-		

[a]The single letter amino acid code of the cleavage sites of the highly aphid transmitted (HAT) isolate of TEV is presented (Carrington and Dougherty, 1987a; 1988). Boldface letters represent cleavage site conserved amino acids.

Table II. Deduced amino acid sequence at the 54 kd nuclear inclusion/30 kd capsid protein boundary[a]

TEV isolates	Amino acid sequence	
	C-terminal region of the 54 kd nuclear inclusion protein	N-terminal region of the 30 kd capsid protein
	Cleavage site	
	−10 −5 −1	+1 +5 +10
TEV-HAT	-I-P-T-T-E-N-L-Y-F-Q	S-G-T-V-D-A-G-A-D-A-
TEV-NAT	−−−−−T−−−−−	G−−−−−R-S−−−
TEV-PAT	−−−−−T−−−−−	−−−−−−V−−−
TEV-ST1	−−−−−T−−−−−	G−−−−−−S−−−
TEV-Oxnard	−−−−−A−−−−−	−−−−−−−−−−
TEV-Madison	−−−−−T−−−−−	G−−−−−−−−−
TEV-Severe	−−−−−−−−−−	−−−−G−−−−R−−
TEV-NCH10-2	−−−−−T−−−−	G−−−−−S−−V−−
TEV-NCYa-44a	−−−−−T−−−−	−−−−−−−R−−−
TEV-NCgg	−−−−−T−−−−−	−−−−−V−−−−D-

[a]Blank spaces indicate amino acids identical to those shown for the highly aphid transmitted (HAT) isolate of TEV. Boldface letters represent conserved cleavage site amino acids.

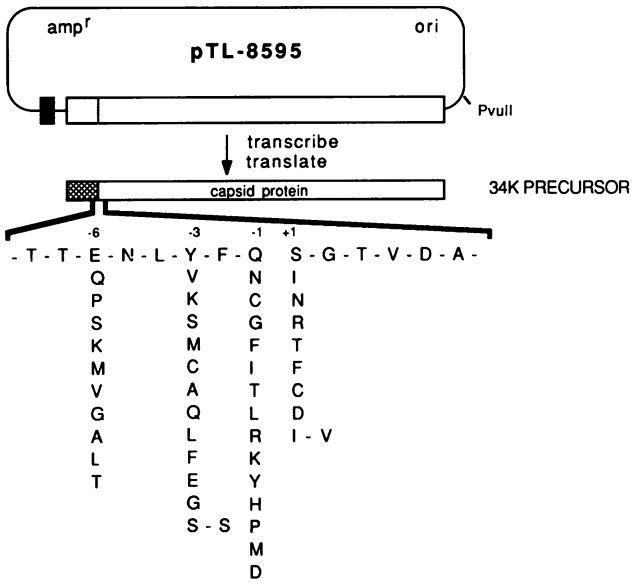

Fig. 1. Schematic of TEV polyprotein expression system and amino acid substitutions at the cleavage site. The expression vector pTL-8595 is presented at the top of the figure. Its essential features are shown: an SP6 promoter sequence (black box) is upstream of the TEV 5′ untranslated leader sequence (thin line). The coding region (box) is comprised of the sequence for the N-terminal 20 amino acids of the TEV polyprotein fused in the same reading frame with the carboxy-terminal 17 amino acids of the 54 kd M_r nuclear inclusion protein (stippled area). Also in the same reading frame is the gene for the entire 30 kd M_r capsid protein (white box). Transcripts of *Pvu*II, digested pTL-8595, when translated in a rabbit reticulocyte lysate, will program the synthesis of a 34 kd M_r polyprotein precursor which is predominantly 30 kd M_r capsid protein. The amino acid sequence (single letter code) of the 54 kd/30 kd M_r polyprotein cleavage site is presented below the 34 kd M_r precursor. Site-directed mutagenesis permitted the individual substitution of amino acids at the four conserved sites. These amino acid replacements are presented below the cleavage site sequence.

Results

To assess the requirement for the conserved amino acid residues at positions −6, −3, −1 and +1, two approaches were taken using the 54 kd/30 kd M_r protein cleavage site as a model. First, we surveyed the extent to which these amino acids were conserved in natural isolates of TEV. Second, the amino acids at these four positions were individually replaced using a site-directed mutagenesis protocol. The ability of these altered polyproteins to serve as substrates for the TEV 49 kd M_r proteinase in *trans* was then tested in a cell-free system.

Sequence analyses of naturally occurring TEV isolates

Sequence analysis of the genomic RNA of the highly-aphid-transmissible isolate of TEV (TEV-HAT) revealed a conserved motif of amino acids at the five polyprotein cleavage sites proposed to be processed by the TEV 49 kd M_r proteinase (Table I) (Carrington and Dougherty, 1987a,b). We wanted to address the question: are the amino acids surrounding a particular cleavage site conserved between distinct isolates of TEV collected from different geographical locations? The nucleotide sequence surrounding the 54 kd/30 kd M_r protein boundary coding region was determined for 10 TEV isolates. From this nucleotide sequence, the amino acid sequence was deduced. The 10 residues on either side of the cleavage site are presented in Table II. The amino acid residues at positions −6, −3 and −1 were strictly conserved in all 10 isolates while the residue at the +1 position was either Ser or Gly. Amino acid variability on the amino-terminal side of the cleavage site was observed only at position −5 with Thr the most common amino acid at this position. On the carboxy side of the cleavage site, more sequence heterogeneity was evident. Ser (six isolates) or Gly (four isolates) was found at the +1 position. Considerable sequence divergence was observed between positions +5 and +10. Therefore, the primary sequence analysis of a cleavage site in 10 different TEV isolates supported the hypothesis that the sequence . . . Glu-Xaa-Xaa-Tyr-Xaa-Gln-Ser or Gly . . . is required in defining a TEV polyprotein cleavage site.

Site-directed mutagenesis of the 54 kd/30 kd M_r protein cleavage site

To determine directly if certain amino acids were required to define a TEV polyprotein cleavage site, amino acid substitutions were introduced by site-directed mutagenesis. The altered polyproteins were then processed in a cell-free system. *In vitro* site-directed mutagenesis (Taylor *et al.*, 1985a,b) was performed on the recombinant cDNA molecule pTL-8595. This recombinant molecule contained the coding sequence of the amino-terminal 20 amino acids of

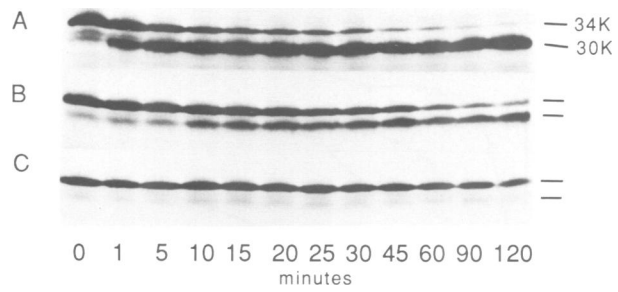

Fig. 2. Time course of processing wild-type and mutant synthetic polyproteins. Transcripts were synthesized from wild-type or mutated pTL-8595 with SP6 polymerase and were used to program the synthesis of a 34 kd M_r polyprotein. TEV proteinase, in the form of nuclear inclusion bodies, was added to the [^{35}S]methionine-labeled precursor and the precursor and processed product were separated by SDS−PAGE. The processing reaction was assayed at the following time points: 0, 1, 5, 10, 15, 20, 25, 30, 45, 60, 90 and 120 min. The autoradiograph in (**A**) shows the conversion of the 34 kd M_r precursor, containing a naturally occurring 54 kd/30 kd M_r polyprotein cleavage sequence (. . . Glu-Asn-Leu-Tyr-Phe-Gln-Ser . . .), to the 30 kd M_r capsid protein product. The autoradiograph in (**B**) presents the results of a similar processing experiment in which Tyr at position −3 had been changed to Val (. . . Glu-Asn-Leu-**Val**-Phe-Gln-Ser . . .). The autoradiograph in (**C**) presents the cleavage reaction of the same 54 kd/30 kd M_r polyprotein cleavage junction after Gly had been substituted for Tyr at position −3 (. . . Glu-Asn-Leu-**Gly**-Phe-Gln-Ser . . .).

the TEV polyprotein fused to the sequence encoding the carboxy-terminal 17 amino acids of the 54 kd M_r TEV nuclear inclusion protein and the 30 kd M_r capsid protein (Carrington and Dougherty, 1987a). A schematic of pTL-8595 is detailed in Figure 1 along with the amino acid sequence of the polyprotein around the authentic 54 kd/30 kd M_r protein cleavage site of the TEV-HAT isolate. Four sets of site-directed mutants were engineered, resulting in codon changes at positions −6, −3, −1 and +1 (Figure 1). All site-directed nucleotide changes were confirmed by nucleotide sequence analysis. Transcripts were synthesized using SP6 polymerase and were translated in a rabbit reticulocyte lysate resulting in a 34 kd M_r [^{35}S]methionine-labeled polyprotein precursor. Polyproteins containing a functional cleavage site were processed to a 30 kd M_r product upon the addition of the 49 kd M_r proteinase (in the form of purified nuclear inclusion bodies). Samples were taken at various time points and analyzed by SDS−polyacrylamide gel electrophoresis (SDS−PAGE). In total, 44 different cleavage site sequences were tested. Figure 2 presents the autoradiograms of three such experiments showing the [^{35}S]methionine-labeled 34 kd M_r polyprotein substrate and its processed 30 kd M_r capsid protein product. Panel A shows the results of a natural cleavage site (Glu-Asn-Leu-Tyr-Phe-Gln-Ser) in which 84% of the precursor was converted to the 30 kd M_r product. The cleavage reaction was rapid, with 50% of this maximum conversion or cleavage of the naturally occurring site occurring 5 min after the addition of proteinase. We refer to this time point as half-time. Panel B shows processing of a polyprotein with a mutated cleavage site (Glu-Asn-Leu-**Val**-Phe-Gln-Ser) in which processing occurred but at a much slower rate, while panel C shows the processing results using a polyprotein with an amino acid substitution (Glu-Asn-Leu-**Gly**-Phe-Gln-Ser) which eliminated cleavage at the site.

The effect of amino acid replacements on cleavage was quantified by determining the amount (c.p.m.) of [^{35}S]-methionine in the 34 kd and 30 kd M_r protein band. The results of 44 site-directed substitutions are presented in Figure 3.

Amino acid substitutions at the −1 position (Gln) had the most dramatic effect on processing catalyzed by the 49 kd M_r TEV proteinase (Figure 3a). Independently derived clones replacing Gln with Cys, Gly, Phe or Asn greatly reduced the rate at which cleavage occurred, while substituting the amino acid Thr, Arg, Lys, Tyr, His, Ileu, Leu, Met or Pro for Gln eliminated cleavage. Conversion of the [^{35}S]methionine-labeled 34 kd M_r precursor containing the naturally occurring cleavage site sequence, to the 30 kd M_r product, was rapid with a half-time of 4.5 min. At a similar time point, less than 10% of the 34 kd M_r substrates had been processed in which Cys, Gly, Phe or Asn had been substituted for Gln (Figure 4A).

Replacing the Tyr at position −3 with a variety of amino acids generally preserved this sequence as a site recognized and cleaved by the TEV 49 kd proteinase (Figure 3B). Only the double amino acid substitution of Ser-Ser for Tyr-Phe eliminated cleavage at this site. However, it should be noted that all amino acid substitutions at this site greatly reduced the rate at which cleavage occurred (Figure 4B). The half-time of the 34 kd M_r substrate containing the wild-type cleavage site was 5 min while at the same time only 10−18% of the 34 kd M_r polyprotein with a cleavage site substituting Val, Lys, Met or Ser for Tyr had been processed. Less than 5% conversion was observed for polyprotein substrates containing the remaining substitutions.

At the −6 position (Glu), nine of the 10 amino acid substitutions permitted cleavage to occur but also at greatly reduced rates (Figures 3C and 4C). In this series of experiments, the processing half-time of the 34 kd M_r substrate containing the wild-type cleavage site was 5 min. At this time point, only a 15−20% conversion of the polyprotein substrates to the 30 kd M_r product was observed for synthetic substrates containing Gln, Pro or Ser replacements (Figure 4C). At the same time point, less than a 10% conversion of the polyprotein substrates was observed in which Glu had been replaced with Lys, Met, Val, Ala or Leu. Substituting Thr for Glu eliminated processing of this polyprotein.

Amino acid substitutions at the +1 position (Ser) generally resulted in substrates which could be processed by the 49 kd M_r TEV proteinase (Figure 3D). Polyprotein substrates in which Ser had been replaced with either Ileu, Asn or Arg were cleaved to the same extent (~80%) as the naturally occurring cleavage site. Replacing Ser with Thr, Phe or Cys resulted in substrates that were processed 50−75% to completion. Only substituting Asp for Ser or the double substitution of Ileu-Val for Ser-Gly resulted in this sequence being processed poorly. Unique about the Ser substitutions was the rate at which some of these mutated sites were cleaved by the 49 kd M_r TEV proteinase. Replacing Ser with Ileu or Asn apparently enhanced the rate (half-time = 4 min) compared with the wild-type cleavage sequence (half-time = 5 min) (Figure 4D). The remaining amino acid substitutions were processed at a much slower rate.

Discussion

The TEV 54 kd/30 kd M_r gene product boundary is cleaved by the TEV 49 kd M_r cysteine-like proteinase

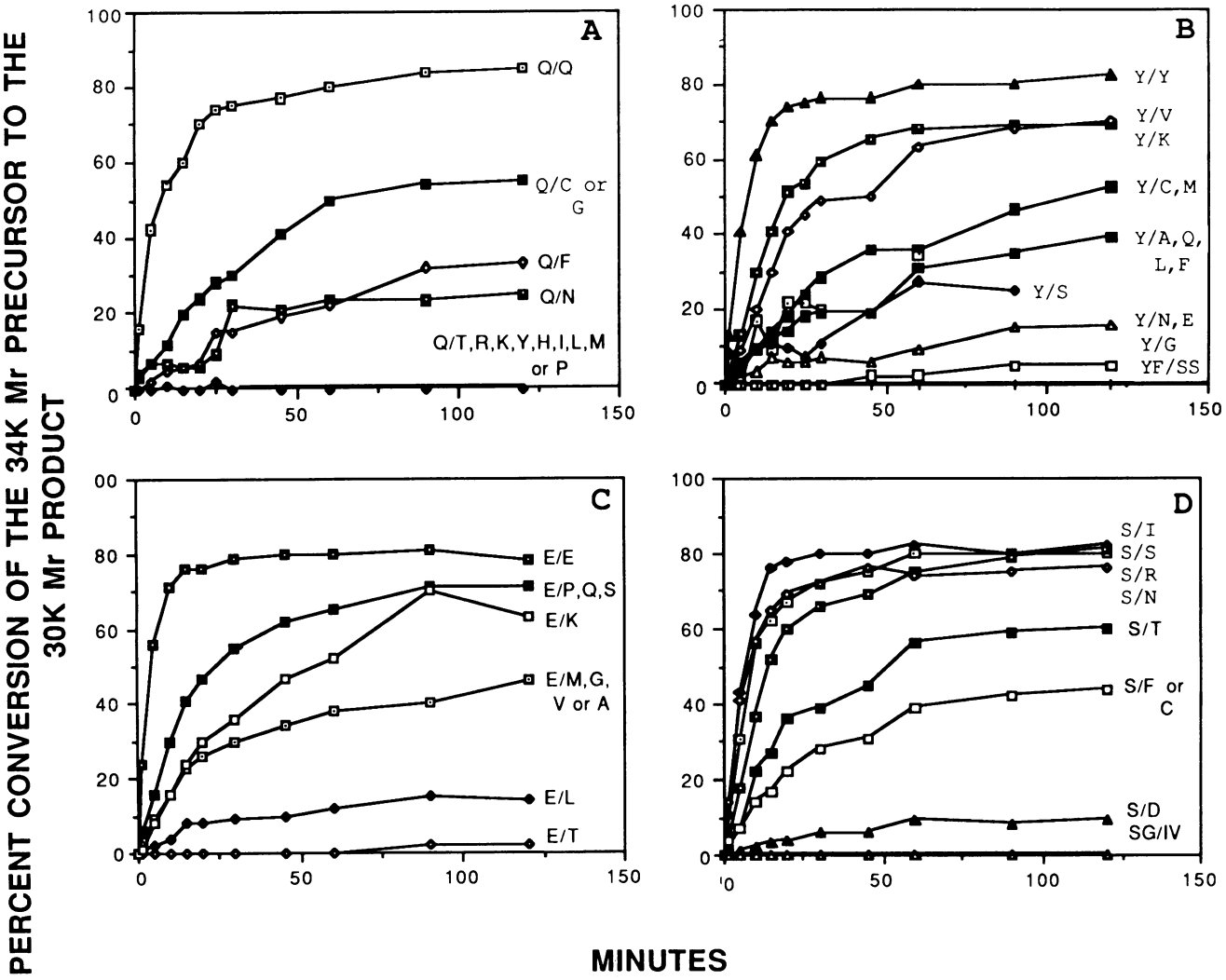

Fig. 3. Processing of natural and mutated TEV polyprotein cleavage sites. The [^{35}S]methionine-labeled 34 kd M_r polyprotein substate was reacted with TEV proteinase at 30°C. In this study, 25 μl of TEV nuclear inclusion bodies (3 mg/ml) were added as the proteinase source. Samples were taken at 0, 1, 5, 10, 15, 20, 25, 30, 45, 60, 90 and 120 min after the addition of the TEV proteinase. The 34 kd M_r precursor was separated from the 30 kd M_r product by SDS–PAGE. The gels were dried and the amount of [^{35}S]methionine in each band was quantitated using an AMBIS Beta Scanning System. The results are presented as the percentage of the 34 kd M_r precursor converted to the 30 kd M_r product versus time (min). Results are presented for amino acid substitutions made at the Gln or −1 position (**A**), at the Tyr or −3 position (**B**), at the Glu or −6 position (**C**) or at the Ser or +1 position (**D**). The amino acid substitutions (single letter amino acid code) are indicated at the end of the line graph. For amino acid substitutions which gave a similar processing profile, only a single reaction curve is presented.

(Carrington and Dougherty, 1987a). Analysis of different isolates of TEV reveals the amino acid residues comprising this polyprotein processing site are highly conserved on the N-terminal side of the scissile bond (Table II). Similar sequence conservation is evident at other TEV-HAT polyprotein cleavage sites (Table I) and the conserved cleavage sequence . . . Glu-Xaa-Xaa-Tyr-Xaa-Gln-Ser or Gly . . . has emerged. The results of this study indicate that for TEV polyprotein processing, mediated by the TEV 49 kd M_r proteinase, a strong preference exists for the conserved amino acids on the N-terminal side of the peptide bond cleaved. Amino acid substitutions in this region always decreased the rate of cleavage and usually the amount of precursor cleavage.

While viral proteinases involved in polyprotein processing may be 'mechanistically related' to cellular proteinases, there are a number of properties which distinguish them (Argos *et al.*, 1984; Barrett, 1986). In the case of the TEV 49 kd M_r proteinase, the Cys and His residues which have been proposed to function at the active site are separated by only 15 amino acids (amino acids 339 and 355 of the TEV 49 kd M_r proteinase). Their role in proteolysis has been inferred from amino acid homology and site-directed mutagenesis studies (Allison *et al.*, 1986; Carrington and Dougherty, 1987b). Cys and His have been shown to constitute the active site of papain and other cysteine proteinases. However, in cellular cysteine proteinases these two amino acids are separated by over 100 amino acids (e.g. in papain, the Cys at position 25 and the His at position 159 are involved) and are found in two distinct domains. These two domains are folded in such a manner that a cleft is formed and the Cys and His are brought proximal to one another in the active enzyme. It is believed that the substrate preference or specificity is determined in this cleft region. For papain, substrates containing Phe in the P2 position (or −2 relative to the cleavage site) are favored along with substrates that are

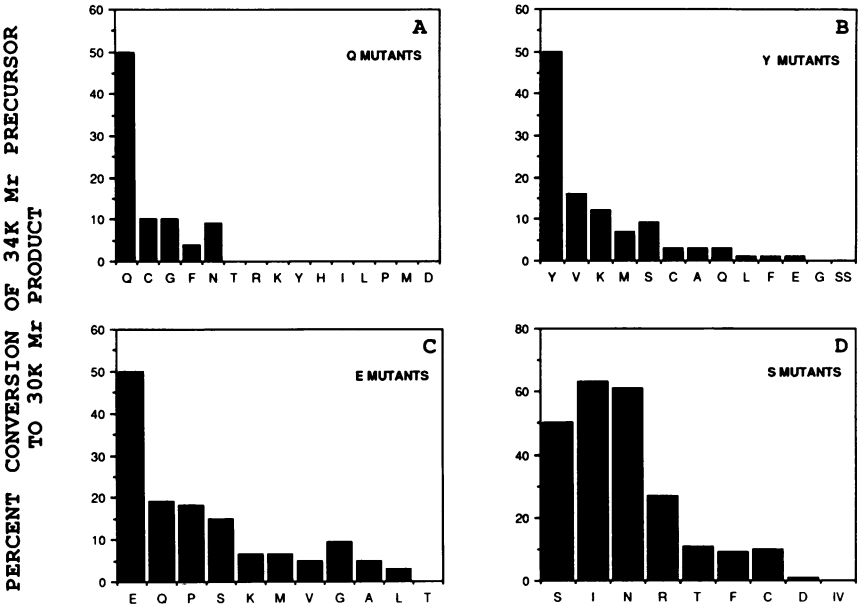

Fig. 4. Comparative analysis of processing. The amount of 34 kd M_r precursor converted to the 30 kd M_r product is presented at a time point at which 50% of the substrate containing the naturally occurring cleavage site had been processed. In the text this time point is referred to as half-time and occurred 4–5 min after the addition of the TEV 49 kd M_r proteinase. The amino acid replacements are presented on the Y-axis. Substitutions are shown for (**A**) the −1 position (Gln), (**B**) the −3 position (Tyr), (**C**) the −6 position (Glu) and (**D**) the +1 position (Ser).

generally hydrophobic in nature (Schechter and Berger, 1967, 1968; for review see Lowe, 1976; Rich, 1986). Therefore, these cellular and viral enzymes are quite distinct structurally, and the domain which determines specificity may be quite different. In the case of the TEV proteinase, it is possible that there are three substrate binding sites or pockets at the active site corresponding to the conserved substrate residues on the N-terminal side of the polyprotein cleavage site at positions −1, −3 and −6. In addition, the active site may favor hydrophobic or neutral amino acids from substrate residues −5, −4 and −2. The conserved amino acids and the uncharged nature in this region of the substrate appear critical for TEV cleavage as all of the amino acid substitutions made in this region greatly reduced or eliminated substrate function.

A similar theme is not evident on the C-terminal side of the 54 kd/30 kd M_r polyprotein peptide bond cleaved or at other TEV-HAT polyprotein cleavage sites. Considerable sequence variability is evident between TEV isolates (Table II) and the five cleavage sites on the TEV-HAT polyprotein (Table I) (Carrington and Dougherty, 1987a). Two or three charged amino acids are present in a region within six amino acids of the scissile bond. Only the Gly or Ser at position +1 is conserved. Changing the Ser of the 54 kd/30 kd cleavage site to Asp was one of two mutants which eliminated cleavage. The carboxy side chain group of Asp may interfere with the formation of the cleavage intermediate or the negative charge may inhibit enzyme and substrate interaction. A cleavage site containing the double mutant Ileu-Val, which replaced Ser-Gly, also was not processed. Such amino acid changes may alter the secondary structure and render the site inaccessible to the proteinase. The other amino acid substitutions at this position permitted cleavage. While most of the substrates were processed, it was at a greatly reduced rate with three mutants and at an equal or greater rate with two other mutants. Therefore, the proteinase appears to be more tolerant in accepting the amino acid occupying the +1 substrate position.

A number of published plant and animal virus genome sequences were examined for conserved features between the different virus polyprotein cleavage sites. Tobacco vein mottling virus (TVMV) is a potyvirus whose nucleic acid sequence has been determined (Domier *et al.*, 1986). The predicted amino acid sequence of the 54 kd M_r-like protein [N1b]/32 kd M_r capsid protein gene junction (. . . Glu-Thr-Val-Arg-Phe-Gln-Ser. . .) is different from the TEV sequence. While the amino acids at positions +1, −1, −2, −5 and −6 are common to TEV and TVMV at this gene product boundary, quite a different consensus sequence can be proposed when the other polyprotein cleavage sites of TVMV are considered. Although there appears to be uncertainty regarding the N-terminal cleavage site of the 49 kd-like protein [N1a] of TVMV (Domier *et al.*, 1986, 1987) the sequence which has evolved as a putative TVMV conserved cleavage sequence is . . . Val-Arg or Lys-Phe or Thr-Gln-Ser or Gly . . .. Therefore, TVMV cleavage site characteristics may be analogous to TEV sites in that (i) cleavage is between Gln-Ser or Gln-Gly dipeptides; (ii) the amino acids flanking the N-terminal side of the scissile bond appear to be conserved although distinct from TEV; and (iii) there is little similarity on the C-terminal side of the TVMV polyprotein cleavage sites. TVMV cleavage sites differ from those of TEV in their lack of a hydrophobic region flanking the scissile bond on the N-terminal side.

Comparison of the TEV polyprotein cleavage sites (positions −6 through +1) with those contained in the primary translation products of cowpea mosaic virus (CPMV) or picornaviridae has not revealed a similar group of conserved amino acids. Cleavage of the CPMV polyproteins occurs between certain Gln-Ser, Gln-Gly or Gln-Ala dipeptides (Wellink *et al.*, 1986). Ala can be found in the −4 position at most of the CPMV cleavage sites processed by the 24 kd

M_r proteinase. The neutral charge nature of the TEV cleavage sequence is not evident in CPMV. The cleavage sites of a number of picornaviruses have been determined. Cleavages mediated by the 3C proteinase are frequently between Gln-Gly or Gln-Ser dipeptides and an Ala or a branched β carbon amino acid is present at the -4 position (Toyoda *et al.*, 1986; Nicklin *et al.*, 1986). Taken together, the amino acid in the -4 position appears to be an Ala or a neutral amino acid with a branched side chain and may play a role in determining a virus polyprotein cleavage site. However, while encephalomyocarditis virus cleavage sites are usually between Gln-Gly or Gln-Ser dipeptides, they lack this characteristic at the -4 position. Instead, the presence of Pro adjacent to (position -2 or $+2$) most of the dipeptides cleaved has been noted and may play a role in defining cleavage sites in this virus system (Palmenberg *et al.*, 1984).

We have developed a working hypothesis about the requirements of a TEV polyprotein site processed by the 49 kd M_r proteinase. First, the site in the polyprotein must be accessible to the proteinase (Arnold *et al.*, 1987). Computer predicted secondary structure suggests that TEV cleavage sites reside in or proximal to turn regions. Once exposed to the proteinase, binding specificity and the rate of cleavage are determined by the amino acids between -6 and -1. The conserved amino acids and the hydrophobic nature of this region may directly influence substrate−proteinase interaction. The Gln at position -1 may be the most critical amino acid in this cleavage domain. This is suggested by two observations. The dipeptide cleaved in most virus polyproteins almost always contains a Gln in the -1 position. Amino acid replacements at this position in our studies often eliminated cleavage. Gln may be required in the -1 position to stabilize or promote the formation of a cleavage intermediate. The amino acid substitutions on the C-terminal side of the scissile bond can promote or inhibit the cleavage reaction. While a number of amino acids could fulfil the role of the $+1$ position, we believe amino acids which are resistant to degradation have been selected (Bachmair *et al.*, 1986). This would explain why only Ser or Gly is found in the $+1$ position of TEV products, yet the proteinase is capable of cleaving polyprotein substrates containing one of several amino acids in the $+1$ position. The conserved amino acids at positions -6, -3 and -1, while not an absolute requirement for cleavage, do play a major role in defining an 'optimal' cleavage site. It may be important to note that while changing one of these amino acids results in a reduced rate and extent of cleavage, the 49 kd M_r proteinase is still able to specifically recognize and cleave these altered sites. The ability to recognize a similar yet different sequence may be a manner in which viruses that express their genomes as polyproteins regulate virus gene function post-translationally. Virus proteins may be cleaved at 'suboptimal' cleavage sequences later in the life cycle to inactivate an early function and/or to turn on a late function.

Materials and methods

TEV isolates and virus purification

TEV isolates were maintained in *Nicotiana tabacum* Havana 425 or Burley 21. Virions were purified as described (Dougherty and Hiebert, 1980a). RNA was purified using the procedures described by Brakke and Van Pelt (1970). The TEV isolates used in this study and their geographic origins were: TEV-HAT, TEV-PAT (Kentucky, USA), TEV-NAT, TEV-Madison (Florida, USA), TEV-ST1, TEV-Severe, TEV-NCH10-2, TEV-NCYa-44a, TEV-NCgg (North Carolina, USA), TEV-Oxnard (California, USA). TEV-HAT and TEV-PAT were obtained from Dr Thomas Pirone, University of Kentucky; TEV-NAT, TEV-Madison and TEV-Oxnard were obtained from Dr Dan Purcifull, University of Florida; and TEV-Severe, TEV-NCH10-2, TEV-NCYa-44a and TEV-NCgg were obtained from Dr Guy Gooding Jr, North Carolina State University.

Nuclear inclusion preparation

Nuclear inclusion bodies were purified from *Datura stramonium* leaves which had been inoculated 3 weeks earlier with the NAT isolate of TEV (Dougherty and Hiebert, 1980b). These inclusion bodies served as the source of TEV 49 kd M_r proteinase activity. The nuclear inclusion body stock preparation used throughout this study had an estimated protein concentration of 3.0 mg/ml.

Site-directed mutagenesis

Site-directed mutagenesis was performed on the cDNA insert contained in pTL-8595. This cDNA recombinant molecule, containing TEV genomic sequences from nucleotides 8462 to 9495 (Carrington and Dougherty, 1987a), was subcloned into the cloning vector pUC118. A single-stranded form of this cDNA sequence was synthesized using the helper phage M13K07. Using this single-stranded cDNA, a number of oligonucleotides, containing mixtures of all four nucleotides at specific codons, were hybridized individually and the method of Taylor *et al.* (1985a,b) was used to synthesize double-stranded cDNA sequences. This strategy allowed us to generate altered codons at positions -6, -3, -1 or $+1$ relative to the scissile bond. The mutated cDNA sequence was ligated back into pTL-8595. A detailed description of the procedure used in our laboratory has been published (Carrington and Dougherty, 1987b). Codon changes were screened by nucleotide sequence analysis.

Nucleotide sequence analysis

Sequence determination of the coding region around the 54 kd/30 kd M_r gene cleavage site was determined from viral RNA (for the 10 TEV isolates) using a method described by Zimmern and Kaesberg (1978) or from a single-stranded DNA template (for the 44 site-directed mutants) using the protocol of Sanger *et al.* (1977). The oligonucleotide primer used in both cases was 5′-T-G-A-A-G-T-T-C-C-A-G-C-A-T-3′ which hybridized to the TEV sequence corresponding to genome nucleotides 8612 through 8625 (Allison *et al.*, 1986).

Polyprotein synthesis and cleavage

The transcription vector pTL-8595 (Carrington and Dougherty, 1987a) containing the naturally occurring or an altered cleavage site, was linearized with the restriction endonuclease *Pvu*II. Transcripts were synthesized using SP6 polymerase as described by Melton *et al.* (1984). Transcripts were added to a rabbit reticulocyte lysate (obtained from Green Hectares, Madison, WI) which contained [^{35}S]methionine. The total translation reaction volume was 180 μl. Translation conditions were as described by Dougherty and Hiebert (1980a). After 60 min at 30°C, 15 μl was removed for a zero time sample ($t = 0$). TEV proteinase (25 μl of a nuclear inclusion preparation) was then added to the remaining translation mix (165 μl) and the reaction was incubated at 30°C. Samples (15 μl) were taken at the following time points: 1, 5, 10, 15, 20, 25, 30, 45, 60, 90 and 120 min. The proteolytic processing reaction was stopped by adding the 15 μl sample to 35 μl of Laemmli dissociation solution. The precursor and processed product were separated by SDS−PAGE on 12.5% polyacrylamide gels using the buffer system of Laemmli (1970).

Quantification of the processing reaction

Polyacrylamide gels containing the separated 34 kd M_r precursor and 30 kd M_r product were fixed in an acetic acid/methanol (10%:50%) solution and dried. The dried gel was scanned using an Ambis Beta Scanning System (Automated Microbiology Systems,Inc.). The radioactivity in the 34 kd and 30 kd M_r [^{35}S]methionine-labeled proteins was determined as c.p.m.

Acknowledgements

We would like to thank Ronnie Johnson for excellent technical assistance and Dennis Hruby for his helpful comments. 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 is Oregon Agriculture Experiment Station Technical Paper Number 8431.

References

Allison,R.F., Johnston,R.E. and Dougherty,W.G. (1986) *Virology*, **154**, 9–20.

Argos,P., Kamer,G., Nicklin,M.J.H. and Wimmer,E. (1984) *Nucleic Acids Res.*, **12**, 7251–7267.

Arnold,E., Luo,M., Vriend,G., Rossmann,M.G., Palmenberg,A.C., Parks, G., Nicklin,M.J.H. and Wimmer,E. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 21–25.

Bachmair,A., Finley,D. and Varshavsky,A. (1986) *Science*, **234**, 179–186.

Barrett,A.J. (1986) In Barrett,A.J. and Salvesen,G. (eds), *Proteinase Inhibitors*. Elsevier, New York, pp. 3–22.

Brakke,M. and Van Pelt,N. (1974) *Virology*, **42**, 699–706.

Carrington,J.C and Dougherty,W.G. (1987b) *Virology*, **160**, 355–362.

Carrington,J.C. and Dougherty,W.G. (1988) *Proc. Natl. Acad. Sci. USA.*, in press.

Christie,R. and Edwardson,J.R. (1977) *Fla. Agric. Exp. Sta. Mono. Ser.*, 4.

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) *Nucleic Acids Res.*, **14**, 5417–5430.

Domier,L.L., Shaw,J.G. and Rhoads,R.E. (1987) *Virology*, **158**, 20–27.

Dougherty,W.G. and Hiebert,E. (1980a) *Virology*, **101**, 466–474.

Dougherty,W.G. and Hiebert,E. (1980b) *Virology*, **104**, 174–182.

Hanecak,R., Semler,B.L., Anderson,C.W. and Wimmer,E. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 3973–3977.

Hari,V. (1981) *Virology*, **112**, 391–399.

Knuhtsen,H., Hiebert,E. and Purcifull,D. (1974) *Virology*, **61**, 200–209.

Laemmli,U.K. (1970) *Nature*, **227**, 680–685.

Lowe,G. (1976) *Tetrahedron*, **32**, 291–302.

Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.

Nicklin,M.J.H., Toyoda,H., Murray,M.G. and Wimmer,E. (1986) *Bio/Technology*, **4**, 33–44.

Palmenberg,A.C. (1987) *J. Cell. Biochem.*, **33**, 191–198.

Palmenberg,A.C., Kirby,E.M., Janda,M.R., Drake,N.L., Duke,G.M., Potratz,K.F. and Collett,M.S. (1984) *Nucleic Acids Res.*, **12**, 2696–2983.

Purcifull,D. and Hiebert,E. (1982) *CMI/AAB Description of Plant Viruses*, 258.

Rich,A. (1986) In Barrett,A.J. and Salvesen,G. (eds), *Proteinase Inhibitors*. Elsevier, New York, pp. 153–178.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Schechter,I. and Berger,A. (1967) *Biochem. Biophys. Res. Commun.*, **27**, 157–162.

Schechter,I. and Berger,A. (1968) *Biochem. Biophys. Res. Commun.*, **32**, 898–902.

Taylor,J.W., Ott,J. and Eckstein,F. (1985a) *Nucleic Acids Res.*, **13**, 8765–8785.

Taylor,J.W., Schmidt,W., Cosstick,R., Okruszek,A. and Eckstein,F. (1985b) *Nucleic Acids Res.*, **13**, 8749–8764.

Toyoda,H., Nicklin,M.J.H., Murray,M.G. and Wimmer,E. (1986) In Inouye,M. and Sarma,R. (eds), *Protein Engineering*. Academic Press, New York, pp. 319–337.

Verver,J., Goldbach,R., Garcia,J.A. and Vos,P. (1987) *EMBO J.*, **6**, 549–554.

Wellink,J., Rezelman,G., Goldbach,R. and Beyreuther,K. (1986) *J. Virol.*, **59**, 50–58.

Zimmern,D. and Kaesberg,P. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 4257–4261.

Received on October 27, 1987; revised on January 26, 1987