A novel cleavage site within the potato leafroll virus P1 polyprotein

Xuejun Li,1† Claire Halpin1 and Martin D. Ryan2

1Plant Research Unit, School of Life Sciences, University of Dundee at the Scottish Crop Research Institute (SCRI), Invergowrie, Dundee DD2 5DA, UK
2Centre for Biomolecular Sciences, School of Biology, University of St Andrews, Fife KY16 9ST, UK

To study the proteolytic processing of the potato leafroll virus replicase proteins, the multidomain P1 protein with a c-myc epitope tag attached at the N terminus was expressed in insect cells by using the baculovirus system. Western blotting showed that P1 was cleaved at a site upstream of the serine protease domain, in addition to the cleavage site downstream of the protease domain. Mutational analysis showed that the serine protease domain within P1 was responsible for this cleavage. To characterize this novel cleavage site further, a portion of the P1 protein comprising the protease domain and the two cleavage sites was expressed in Escherichia coli. A similar cleavage event was observed in bacteria and was abolished when the P1 protease was inactivated by mutation. Peptide-sequencing studies indicated that this cleavage occurred at a Glu/Arg junction, separating the N-terminal 204 residues from the serine protease domain of P1.

Potato leafroll virus (PLRV) is now classified as the type member of the genus Polerovirus of the family Luteoviridae, which also contains the genera Luteovirus and Enamovirus (Pringle, 1998). PLRV virions appear to be non-enveloped, isometric particles with a diameter of 24 nm (Harrison, 1984). Viral particles contain a single-stranded (ss), positive-sense genomic RNA of some 5.8 kb in length (1984). Viral particles also contain the genera Polerovirus, Luteovirus and Enamovirus (Pringle, 1998). PLRV virions appear to be non-enveloped, isometric particles with a diameter of 24 nm (Harrison, 1984). Viral particles contain a single-stranded (ss), positive-sense genomic RNA of some 5.8 kb in length (Mayo et al., 1989; van der Wilk et al., 1989; Keese et al., 1990). Like most luteoviruses, PLRV genomic RNA encodes six major open reading frames (ORFs), which are arranged into two blocks separated by a small intergenic region. ORF0 encodes a 28 kDa protein that is important for virus accumulation (Sadowy et al., 2001b). Site-directed mutagenetic analysis has shown that P1 and P2, encoded by ORF1 and ORF2, are required for virus replication (Sadowy et al., 2001a) – consistent with studies on beet western yellow virus (BWYV), another polerovirus (Reutenauer et al., 1993). The 3'-terminal block encodes the capsid protein (ORF3), the putative movement protein (ORF4) and the aphid transmission protein (ORF5; reviewed by Miller et al., 1995; Mayo & Ziegler-Graff, 1996). In addition to these six ORFs, two small ORFs were identified within the 3’-terminal region (Ashoub et al., 1998). Jaag et al. (2003) reported that another small ORF within the P1 region, in an alternative reading frame, was essential for viral multiplication.

Like many positive-strand ssRNA viruses, PLRV has been found to express the products of its genome in a variety of ways, including transcription of subgenomic mRNAs, translational frameshifting, stop-codon readthrough, internal initiation, translation initiation at an internal ribosomal entry site and polyprotein processing (Miller et al., 1995; Mayo & Ziegler-Graff, 1996).

Proteins encoded within the 5' -proximal region are translated from the genomic RNA, whereas 3'-proximal ORFs are expressed from two subgenomic RNAs (Tacke et al., 1990; Miller & Mayo, 1991; Ashoub et al., 1998). ORF2, which encodes the RNA-dependent RNA polymerase, is expressed via −1 frameshifting at a site within ORF1 (Prüfer et al., 1992). The replication-associated protein 1 (Rep1) is translated through internal ribosome entry (Jaag et al., 2003). ORF5 is translated by the readthrough of the amber stop codon of the capsid protein (ORF3).

The ORF1 encodes a 70 kDa protein comprising multiple domains. The central third contains a serine protease motif (Koonin & Dolja, 1993; Miller et al., 1995), downstream of which is the VPg domain (van der Wilk et al., 1997). Prüfer et al. (1999) showed that P1 is processed in plants infected with PLRV, giving rise to a 25 kDa protein representing the C terminus of P1, cleavage occurring between the protease and VPg domains. Previously, we have demonstrated that the P1 protein expressed in insect cells is processed in a similar way and that the protease domain within P1 is responsible for this proteolytic activity (Li et al., 2000). Here, we report that the P1 protease domain mediates another cleavage event to release its own N terminus. This processing event was also observed in Escherichia coli expressing a part of the P1 protein containing the protease

1Present address: Gatty Marine Laboratory, School of Biology, University of St Andrews, Fife KY16 8LB, UK.
domain and small extensions at both the N and C termini. The novel cleavage site was determined to be glutamate 204/arginine 205 by peptide sequencing.

Using the baculovirus system, we reported previously that the C-terminal region of the P1 polyprotein, containing the VPg domain, was released by the proteolytic activity of the upstream protease domain (Li et al., 2000). We also provided evidence that this processing could be carried out in trans. As the antisera that we have used were raised against the region downstream of the cleavage, it is not known whether the N-terminal region upstream of the protease domain is processed. To answer this question, we constructed a recombinant baculovirus that expressed the wild-type P1 polyprotein with an epitope tag, c-myc, fused to its N terminus (Fig. 1; pBac-mycP1). Western blot analyses using a monoclonal anti-c-myc antibody detected a cleavage product of approximately 23 kDa (Fig. 2). This band was not detected in the extract of insect cells expressing the P1 polyprotein lacking the c-myc epitope (Fig. 2; Bac-P1). This indicated that the N-terminal region of the P1 polyprotein is processed proteolytically in insect cells. Based upon the estimated molecular mass of the cleavage product, this cleavage site is positioned between the N-terminal hydrophobic region (anchor-M) and the protease domain (Fig. 1).

We have shown that the protease domain was inactive when glycine 355 (immediately downstream of serine 354 of the catalytic triad) was mutated to valine. When the P1 protein containing this mutation was expressed in insect cells, only the full-length 70 kDa polyprotein was detected by Western blotting (Li et al., 2000) (Fig. 2; Bac-mycP1GV in this study). It is highly unlikely, therefore, that the processing that we observed at the N-terminal region of P1 was mediated by a cellular or baculovirus-specific protease.

To confirm that the protease motif within P1 is responsible for this cleavage, we co-infected insect cells with baculovirus Bac-mycP1GV, which directed the synthesis of the P1 protein containing the valine 355 mutation and the N-terminal c-myc tag as the substrate, and baculovirus BacP1, which provided the active P1 protease (untagged). Western blot analyses with monoclonal anti-c-myc (Roche) showed that the c-myc–P1GV was cleaved upon co-infection with BacP1 (Fig. 2). These data showed that the N-terminal region of the P1 was indeed processed by the P1 protease in a trans reaction.

As the amount of P1 processing product in insect cells was too low to permit the cleavage product to be purified for further analyses, we tried expressing the P1 polyprotein in E. coli. The full-length P1 protein is expressed poorly in E. coli (data not shown), so we attempted to express part of P1. The 764 bp BssHII–HindIII fragment, encoding the protease domain with small extensions at both the N and C termini of the domain, was fused to glutathione S-transferase (GST). E. coli harbouring this plasmid (pGEX-P1BsH) was induced with IPTG (0.2 mM final concentration) at 25 °C. Cell lysates were analysed by SDS-PAGE (Fig. 3). Three hours after adding IPTG, the recombinant protein produced was mainly the full-length GST–P1BsH (55 kDa). Upon extended expression (a further 3 h), the majority of the full-length GST–P1BsH disappeared and two smaller proteins of 29 and 25 kDa accumulated. The 29 kDa protein was bound by glutathione beads (the GST moiety of the recombination protein), whereas the 25 kDa protein was not.
The identity of the 25 kDa protein was confirmed to contain the P1 protease domain by peptide-mass fingerprinting. Protein samples excised from an SDS–polyacrylamide gel were digested in gel with trypsin (Promega) by using an Investigator Progest digestion robot (Genomic Solutions). Mass spectra were obtained on a Micromass TofSpec 2E instrument, equipped with a 337 nm laser and operated in reflectron mode. The peak list was subjected to a MASCOT search, which identified the PLRV P1 protein as the top hit with a score of 132, with 13 peptide matches covering 20 % of the P1 sequences, in the region where the protease domain resides.

We concluded that the 29 and 25 kDa proteins were the cleavage products of the 55 kDa full-length fusion protein. As the mass of GST itself is approximately 26 kDa, the cleavage site was estimated to be at or near the cleavage site that we observed in insect cells, suggesting that this cleavage was mediated by the P1 protease. Further evidence was obtained by expressing the GST–P1BsH protein containing the valine 355 mutation (GST–P1GVBsH). When E. coli carrying pGEX-P1GVBsH was induced with IPTG, only full-length fusion protein was produced at 25 °C (Fig. 3).

To locate the exact site of cleavage, protein samples resolved by SDS-PAGE (10 % gel) were transferred onto a PVDF membrane and stained with Coomassie brilliant blue R-250. The 25 kDa protein band was cut out and sequenced by Edman degradation. The first six residues of this protein were determined as NH₂–R–A–V–E–G–Y, indicating that cleavage occurs between glutamate 204 and arginine 205 of the P1 protein.

Viral proteases play a key role in the replication of many viruses, particularly the positive-strand ssRNA viruses. The precursors of replicative proteins, e.g. polymerase, are invariably processed by viral proteases – sometimes in combination with host cellular proteases (reviewed by Dougherty & Semler, 1993). The PLRV P1 protein contains multiple domains (Fig. 1), which comprise (N- to C-terminally) a hydrophobic membrane-anchoring domain, a protease domain (Koonin & Dolja, 1993), a VPg domain (van der Wilk et al., 1997) and a strongly basic region (-KxKxKKRxRRxRxK-), which has been found to exhibit nucleic acid-binding properties (Prüfer et al., 1999). Cleavage between the protease domain and the VPg domain was observed in infected plants (Prüfer et al., 1999) and in insect cells (Li et al., 2000). Here, we report a novel cleavage site between the N-terminal hydrophobic domain and the protease domain, cleaved by the P1 protease to release its own N terminus. Hydrophilic/hydrophobic profiling shows that this novel cleavage site (E204/R205) resides at a very hydrophilic region, which has high surface probability and possibly forms a connection loop between the two domains (data not shown).

No information is available concerning the temporal or spatial pattern of the two processing events within cells. As we rarely observed the intermediate processing products produced by a single cleavage, it seems that both cleavages are carried out efficiently within insect cells. The result of these cleavages is the release of the protease from the putative membrane-anchoring domain. Many positive-strand ssRNA viruses use membranous surfaces or perturb...
membrane traffic to form special membrane structures for replication. Additionally, for many viruses, the final processing products and their precursors/intermediates have different functions – part of the regulatory mechanism of virus replication. It is still unknown whether PLRV uses this strategy.

It also should be noted that the scheme that we propose is not complete. For example, the mass of VPg found in virus particles is approximately 8 kDa. The peptide-sequencing data of van der Wilk et al. (1997) suggest that there should be another cleavage site between the VPg domain and the nucleic acid-binding domain. However, such a cleavage site was not found in plants (Prüfer et al., 1999) or in insect cells.

Our data suggest that the cleavage site N-terminal to the protease domain is present at a glutamate–arginine scissile pair. The cleavage site C-terminal to the protease domain is assumed to be a glutamate–serine pair. As few cleavage sites of the PLRV P1 protease have been identified, no consensus of cleavage sites can be determined. It is interesting to note that the cleavage sites of the equine arteritis virus serine protease (EAV SP) were predicted to have a glutamate (or aspartate) at the P1 position, a glycine (or serine) at the P1′ position and a bulky, hydrophobic residue (valine, phenylalanine or leucine) at the P3′ position (Godeny et al., 1993). The first cleavage site in PLRV has glutamate at P1, arginine at P1′ and valine at P3′; the second site (putative; van der Wilk et al., 1997) has glutamate at P1 and serine at P1′, but alanine at P3, showing a similar site specificity. Both PLRV P1 protease and EAV SP belong to the 3C-like serine proteases (Snijder et al., 1994). Among the 21 members of this group listed by Snijder et al. (1994), all others have a conserved threonine and histidine in their substrate-binding pocket apart from PLRV P1, which has the conserved threonine, but a leucine in place of the histidine. It is unknown whether this leucine will affect the site specificity.

Acknowledgements

We thank Dr B. Reavy (SCRI) for assistance in cell culture, Dr C. Botting for peptide-mass fingerprinting and Dr N. Morrice (University of Dundee, UK) for carrying out Edman sequencing.

References


