Expression of the turnip yellow mosaic virus proteinase in *Escherichia coli* and determination of the cleavage site within the 206 kDa protein

Gress Kadaré,* Mikhail Rozanov† and Anne-Lise Haenni

Institut Jacques Monod, 2 Place Jussieu-Tour 43, 75251 Paris Cedex 05, France

The large non-structural polyprotein (206 kDa) of turnip yellow mosaic tymovirus (TYMV) undergoes autocleavage, producing N- and C-terminal proteins. Here we show that the viral proteinase responsible for this event is active when produced in *Escherichia coli*, as monitored in Western blots by examining the generation of the C-terminal cleavage product after induction by IPTG. The outer boundaries and critical amino acids of the proteinase domain were characterized by deletion analysis and site-directed mutagenesis. A miniproteinase of 273 residues resulting from combined N- and C-terminal deletions still performed efficient cleavage. Sequence analysis of the bacterially-purified C-terminal cleavage product indicated that cleavage occurs between Ala1259 and Thr1260 of the non-structural protein.

Turnip yellow mosaic tymovirus (TYMV) contains a single plus-strand RNA that directs the synthesis of the virus non-structural 69 kDa and 206 kDa proteins. The latter is a multifunctional protein. Computer-based sequence analysis suggests that it includes a methyltransferase (Rozanov et al., 1992), NTPase/helicase (Gorbalenya & Koonin, 1989) and polymerase domain (Kamer & Argos, 1984). The 20 kDa capsid is expressed from a subgenomic RNA.

In *vitro* translations first showed that the 206 kDa protein undergoes autocatalytic cleavage, yielding N- and C-terminal proteins of approximately 150 kDa and 78 kDa, respectively (Morch et al., 1989). It was proposed that the papain-like proteinase was located between the methyltransferase and NTPase/helicase domains, and that cleavage occurred between the NTPase/helicase and polymerase domains (Bransom & Dreher, 1994; Rozanov et al., 1995).

To define certain characteristics of the proteinase *in vivo*, and identify the cleavage site, the proteinase was expressed in *Escherichia coli*. Two regions of the TYMV genome encoding the 206 kDa protein and containing the proteinase domain were expressed *in vivo* (Fig. 1): an expanded [nucleotides (nt) 1754-4085] and a shortened (nt 2218-4085) version. Plasmid pTYFL7 (Boyer et al., 1993) containing the full-length genomic cDNA was used to construct chimaeric genes with the chloramphenicol acetyltransferase (CAT) gene as a reporter. Plasmid p204* (with a mutated initiator ATG) served as a source of the CAT gene. The CAT SalIcat–SmalMCS fragment (MCS; multiple cloning site) replaced the TYMV SaII4085–SmalMCS fragment in pTYFL7. In the resulting plasmid pEMBL-TYCAT, the CAT gene is located downstream of the cleavage site and in-frame with the virus sequence.

A BamHI17sa BamHICAT fragment excised from pEMBL-TYCAT was subcloned into the BamHI site of pET-3a (Novagen) yielding pTYCAT, harbouring the expanded version of the cDNA for the putative proteinase domain (Fig. 1). To produce the shortened version, the p204* CAT SalIcat–BamHICAT fragment replaced the SalI4085–BamHIMCS fragment of pTYFL7 with an artificially introduced Ncol site (Ncol*; Rozanov et al., 1995) yielding pEMBL-ΔTYCAT. The Ncol*2218–BamHIMCS fragment obtained by partial digestion of pEMBL-ΔTYCAT was cloned into pET-3d (Novagen) digested with Ncol–BamHI, yielding pΔTYCAT.

Cell growth, IPTG-induced expression and preparation and analysis of bacterial lysates were as described (Studier et al., 1990). CAT synthesis was monitored by Western blot using anti-CAT antibodies (Tebu). It was expected that if the bacterially-expressed protein was active, autocleavage should produce two polypeptides, of which only the C-terminal fragment of 34 kDa would be detected. If the proteinase was inactive, an uncleaved protein would be detected, of 113 kDa or 94 kDa for the expanded or shortened version, respectively.
In the induced bacteria, expression of pTYCAT and pΔTYCAT yielded the 34 kDa product (Fig. 2a, b). This product was absent or present in low amounts in uninduced bacteria and absent from bacteria containing the pET vectors (Fig. 2a, b; lanes 1 and 2). Thus, active protease was produced and its N-terminal boundary is located downstream of amino acid 709, corresponding to the NcoI* site.

To express a 'minimum' active protease independent of other portions of the polyprotein, and to study the importance of the spacer region between the catalytic and cleavage sites, the N-terminal deletion was combined with an internal deletion (amino acids 982-1204; Fig. 1) in pdΔTYCAT. The resulting miniprotease exerted full proteolytic activity, since the 34 kDa product was detected (data not shown). Thus, the protease, confined to 273 residues, functions efficiently and independently. In the 206 kDa protein, the C terminus of the protease, as defined in pdΔTYCAT (amino acid 982), is separated from the cleavage site by about 270 residues. However, even when the spacer sequence is reduced by 80% (to 53 amino acids), cleavage remains efficient. This contrasts with results concerning the spacer region (approximately 1000 amino acids) between the papain-like protease and cleavage site in mouse hepatitis coronavirus, where decreasing the spacer region by 33% or 66% decreases or abolishes cleavage, respectively (Baker et al., 1993).

Previous studies suggested that the TYMV protease possesses a characteristic catalytic Cys and His dyad (Bransom & Dreher, 1994; Rozanov et al., 1995). To determine whether these residues are essential for protease activity in vivo and confirm the virus-specific
Short communication

Fig. 2. Expression of the TYMV proteinase constructs in *E. coli* analysed by Western blot using anti-CAT antibodies. Total bacterial proteins were resolved by SDS-PAGE, transferred to a nitrocellulose membrane and the blot developed by enhanced chemiluminescence. Bacteria were induced with IPTG or left uninduced as indicated. (a) Lanes 1 and 2, bacteria transformed with pET3a; lanes 3 and 4, bacteria harbouring pTYCAT. (b) Lanes 1 and 2, bacteria transformed with pET3d; lanes 3 and 4, bacteria transformed with pΔTYCAT used for the introduction of site-specific mutations; lanes 5–12, bacteria harbouring pΔTYCAT-C783S, pΔTYCAT-S925L, pΔTYCAT-H869E and pΔTYCAT-K982S (designated pC783S, pS925L, pH869E and pK982S), respectively. Positions of the uncleaved (94 kDa) and cleaved (34 kDa) proteins are indicated. The products of approximately 23 and 30 kDa originate from the polyclonal antibodies produced from a bacterially-expressed CAT-containing transposon (Tebu). The low level of 34 kDa protein visible in lanes 3 and 11 results from basal expression before induction (Studier et al., 1990). Molecular mass standards run in parallel are indicated on the left.

origin of cleavage, the putative catalytic Cys<sup>783</sup> and His<sup>869</sup> were mutated singly, as well as the conserved Ser<sup>925</sup> and Lys<sup>982</sup> located beyond the presumed proteinase domain. These mutants (pΔTYCATmut in Fig. 1) were constructed by replacing the virus sequence between *KpnI*<sub>2561</sub>–*KpnI*<sub>2726</sub> in pΔTYCAT by the corresponding *KpnI*–*KpnI* fragment from pC783S, pH869E, pS925L or pK982S (Rozanov et al., 1995).

Induction and expression of pΔTYCAT-C783S and pΔTYCAT-H869E yielded the 94 kDa protein, whereas pΔTYCAT-S925L and pΔTYCAT-K982S yielded the 34 kDa cleavage product (Fig. 2). These results demonstrate that Cys<sup>783</sup> and His<sup>869</sup> are essential for autocleavage *in vivo* and support the classification of the TYMV proteinase among the papain-like virus proteinases.

Earlier attempts to identify the site of cleavage of the TYMV proteinase by sequencing of the *in vitro*-translated C-terminal cleavage product failed (Morch et al., 1989; Bransom et al., 1991). pTYCAT was used to purify the C-terminal cleavage product, whose sequence was modified to include six histidines at the C terminus. To this end the following linker:

```
5’ CAGGGCTCCCCGGG (CAC)<sub>6</sub>TAAACGGG 3’
3’ CGAGGGCCC (GTG)<sub>6</sub>ATTG 5’
```

was ligated into pTYCAT digested by *NcoI* and *SacII*, generating pTYCAT-6His. Recombinants were screened with *SmaI*. The corresponding C-terminal cleavage product lacking the last 40 amino acids of CAT was approximately 30 kDa.

Lysis of bacteria and solubilization of membrane fractions containing the 30 kDa protein were exactly as described (Das & Banerjee, 1993) except for an additional extraction step after lysis with 6 M-urea pH 8.0. The insoluble 30 kDa protein was then solubilized with 6 M-guanidinium hydrochloride pH 8.0, clarified by centri-
proteinase cleavage site among plant viruses of the Sindbis-like (or alpha-like) supergroup. This is in contrast to animal viruses of this supergroup, all of which possess proteinase-encoding genes; among plant viruses, TYMV and blueberry scarlet carlavirus (BBSv; Lawrence et al., 1995) are the only members for which such a gene has been demonstrated unambiguously.

Alignment of the cleavage sites (nsP1/nsP2, nsP2/nsP3 and nsP3/nsP4) of the nsP2 papain-like proteinase of six alphaviruses reveals that the residue at position P2 is Gly, whose importance has been demonstrated by site-specific mutagenesis (for review see Strauss & Strauss, 1990). When mutated to Ala, a conservative replacement, cleavage in vitro was less efficient, whereas mutation to Val or Glu abolished cleavage. A similar situation exists for the papain-like proteinase of BBSv (Lawrence et al., 1995). Similarly, in TYMV a Gly residue occupies position P2 (Fig. 1). Inspection of the amino acid sequences surrounding the putative cleavage site of three other tymoviruses (Fig. 1) indicates that although this region is not highly conserved, only small amino acids (Gly or Ala) would occupy P2, consistent with the replaceability of Gly by Ala in alphaviruses.

Alignment of the cleavage site-containing regions of six alphaviruses and four tymoviruses reveals that P1 is always occupied by an amino acid with a short side chain, Ala, Gly, Ser or Cys. The residues downstream of the cleavage site show little conservation between the four tymoviruses. In TYMV a Thr is present in P1. This is unusual, since hydrolysis by most papain-like viral proteinases occurs between small amino acids (Choi et al., 1991; Shapira & Nuss, 1991; Carrington & Herndon, 1992; Snijder et al., 1992; Dong & Baker, 1994; for review see Strauss & Strauss, 1990) except for nsP2 of alphaviruses, which cleaves before a Tyr at the nsP3/P4 site. The insensitivity of the latter proteinase to the residue in P1 was confirmed by various mutations in this position (see Strauss & Strauss, 1990). Several other papain-like proteinases exist that accept various amino acids in P1 (Shapira & Nuss, 1991; Snijder et al., 1992; Dong & Baker, 1994).

Considering the importance of Cys and His for TYMV replication in protoplasts (Bransom & Dreher, 1994; G. Kadaré, K. Seron and A.-L. Haenni, unpublished results) and the similarity in the processing pattern observed for five tymoviruses in vitro (Kadaré et al., 1992), the tymovirus protease is undoubtedly important in regulating the virus life-cycle. Indeed, the Sindbis virus proteinase nsP2 (which shares features with the TYMV protease) plays a pivotal role in the temporal regulation of minus- and plus-strand RNA synthesis (Lemm et al., 1994). The precise role and possible additional functions of the TYMV proteinase in virus replication remain to be elucidated.

We are grateful to J.-M. Camadro for helpful discussions and expert technical advice. We thank R. P. C. Valle for plasmid p204*, J.-P. Le Caer and H. S. Savitri for determining the amino-terminal sequence and F. Bernardi for many fruitful suggestions. The project was initiated when G. K. was the recipient of a Fellowship from the MRES and from the IFCPAR. The Institut Jacques Monod is an "Institut mixte CNRS-Université Paris 7".

References


(Received 5 May 1995, Accepted 14 July 1995)