

**Papain-like proteinase of turnip yellow mosaic virus:  
a prototype of a new viral proteinase group\***

M. N. Rozanov<sup>1,2,3</sup>, G. Drugeon<sup>1</sup>, and A.-L. Haenni<sup>1</sup>

<sup>1</sup>Institut Jacques Monod, Paris Cedex France

<sup>2</sup>Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

<sup>3</sup>Department of Microbiology, University of Pennsylvania, Philadelphia,  
Pennsylvania, U.S.A.

Accepted October 6, 1994

**Summary.** Sequence comparisons predicted a potential papain-like proteinase domain in the N-terminal cleavage product (NRP) of the large nonstructural replicase polyprotein (RP) of turnip yellow mosaic virus (TYMV). Replacement of the predicted catalytic amino acids, Cys-783 by Ser, or of His-869 by Glu, abolished cleavage of the 206K RP into a ~150 K NRP and a ~78 K C-terminal product in reticulocyte lysates, while other substitutions exerted no apparent influence on proteolysis. The proteinase-deficient mutant RPs could not be cleaved in trans by as much as an eight-fold molar excess of wild-type proteinase. Deletion experiments have excluded the possible influence on autoproteolysis of amino acid sequences 1–708 and 982–1204 flanking the proteinase domain. Thus, the proteinase of TYMV with a papain-like dyad of essential amino acids has been mapped just upstream from the putative NTPase domain. Statistically significant sequence similarities with the TYMV proteinase were found for the similarly located domains of the replicase polyproteins of carlaviruses, capilloviruses, apple stem pitting virus and apple chlorotic leaf spot virus as well as for those of other tymoviruses and for the domain located downstream from the putative NTPase domain of the large polyprotein of beet necrotic yellow vein furovirus. All these domains are not significantly similar to other known proteinases, although they conserve papain-like Cys- and His-containing motifs. Thus these domains constitute a compact group of related enzymes, the tymo-like proteinases, within the proposed *papain-like* proteinase supergroup. The resulting alignment of 10 tymo-like proteinase sequences has revealed a third highly conserved residue – Gly (Gly821 in TYMV RP) followed by a hydrophobic residue. We speculate that all the tymo-like

\*Reported in part at the NATO/EEC Course “Regulation of gene expression by animal viruses” Mallorca, May 30–June 8, 1992 and at the IX International Congress of Virology, Glasgow, August 8–13, 1993.

proteinase domains of the viral replicative proteins may share common biochemical and biological features.

### Introduction

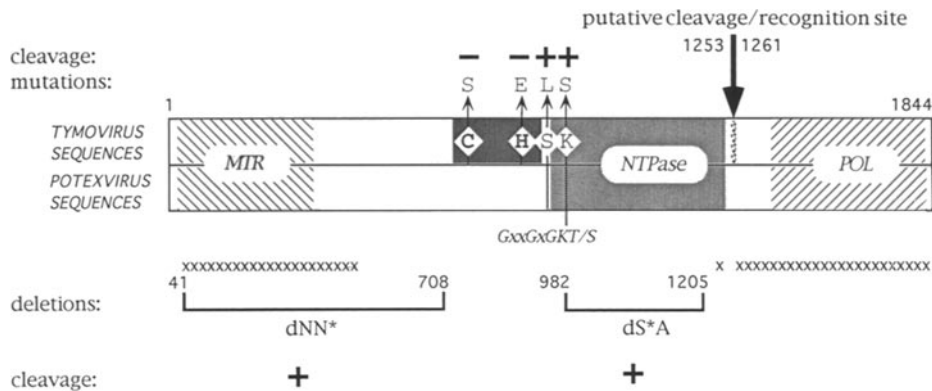
Genomes of all animal positive strand RNA viruses encode proteinases, most of which have been classified as either chymotrypsin-like [1, 2] or papain-like [3, 4] enzymes. The plant "picorna-like viruses" [5], e.g. comoviruses, nepoviruses and potyviruses share this feature with their animal counterparts [6]. Chymotrypsin-like proteinase domains were also predicted for sobemo-like viruses [1, 6]. However, it was not known whether plant viruses belonging to another large virus supergroup, the Sindbis-like viruses [6], contain proteinase genes. The only data suggesting the existence of such genes for viruses of this latter group were obtained for turnip yellow mosaic virus (TYMV).

TYMV is the type member of the tymovirus group. The spherical virion of TYMV contains a single molecule of positive RNA, 6318 nucleotides long, with an m<sup>7</sup>G cap at its 5' end and a tRNA-like structure at its 3' end [7]. TYMV RNA encodes three polypeptides. The larger nonstructural protein of 206 K, involved in RNA replication [8], designated here RP (replicase protein), and the smaller cell-to-cell movement protein of 69 K [9], designated OP (overlapping protein) are translated from overlapping reading frames on the genomic RNA [7, 10]. The 3' adjacent coat protein (CP) gene is expressed via a subgenomic RNA.

Based on sequence comparisons, 3 putative functional domains, a methyltransferase [11], an NTPase/helicase and a polymerase [7], have been highlighted in RP (Fig. 1). These data unequivocally linked tymoviruses with the Sindbis-like supergroup of viruses [5]. In addition, it has come as a surprise that the large replicative proteins of spherical tymoviruses closely resemble those of filamentous potexviruses [12], another distinct group of plant Sindbis-like viruses.

The large nonstructural RP of TYMV undergoes autocleavage [13] yielding at least two products of ~150 K (NRP, or the N-terminal part of RP) and ~78 K (CRP, or the C-terminal part of RP). Similar autocleavages were recently observed for 5 different tymoviruses [14]. The probable proteinase recognition/cleavage site has been mapped between amino acid residues 1253–1261 of the TYMV polyprotein (Fig. 1). The responsible proteinase may be located somewhere in the 688-amino acid stretch upstream from this site [15]. However, attempts to locate the proteinase domain by similarity with known cellular and viral proteinases were unsuccessful ([6, 7, 13, 15]; E. V. Koonin, pers. comm.). This led to the speculation that there are significant differences in the expression strategies of plant and animal viruses [16].

In this paper we describe the prediction of a papain-like proteinase domain in the TYMV RP polyprotein and of its potential catalytically active amino acids, and confirm this prediction by mutation and deletion analyses. From these data and by sequence comparison we have revealed potential tymo-like



**Fig. 1.** Prediction of the papain-like TYMV proteinase domain and the strategy for its experimental verification. A schematic representation of alignment of amino acid sequences of the RPs of 4 tymoviruses and 5 potexviruses (M.N.R., unpubl.) is shown. The putative methyltransferase (*MTR*), NTPase/helicase (*NTPase*), polymerase (*POL*) and tymovirus-specific domain containing the papain-like Cys (*C*)–His (*H*) dyad are differentially shaded or hatched. Ser (*S*) and Lys (*K*), invariant for both tymo- and potexviruses, are the control targets for mutagenesis. Thin arrows indicate mutations introduced into RP. Numbers indicate amino acid positions in the sequence of the TYMV RP. Thick solid lines show the deletions (dNN\* or dS\*A) made to map the proteinase domain; xxx, the sequences previously shown to be dispensable for autoproteolysis [15]. The effect of each mutation or deletion on autocleavage of the RP is indicated by (–), no cleavage, or (+), apparently unaltered cleavage

proteinase domains in the large nonstructural proteins of carlaviruses, capilloviruses, furoviruses, apple stem pitting virus and apple chlorotic leaf spot virus. These domains constitute a compact group of related enzymes, the tymo-like proteinases within the proposed *papain-like* proteinase supergroup.

## Materials and methods

### Sequence comparisons

The GENESEE package [17] was used for sequence database searches and to look for local similarities between two sequences. Alignments were made by the MULTALIN [18] and OPTAL [19] programs, and then analyzed by visual inspection. The significance of alignments was assessed by the OPTAL program as the number of standard deviation (SD) above the mean score for 25 random permutations of the same sequences. Alignments scoring 6 SD or more were taken into consideration. The sequences of the following viruses, found to possess (putative) proteinase domains, were included in the final alignment: the tymoviruses TYMV [7], kennedy yellow mosaic virus, KYMV [20], ononis yellow mosaic virus, OYMV [21], eggplant mosaic virus, EMV [22] and erysimum latent virus, ELV [23]; potato carlavirus M, PVM, [24]; apple stem pitting virus, ASPV [25]; apple chlorotic leaf spot virus, ACLSV [26]; apple stem grooving capillovirus, ASGV [27]; and beet necrotic yellow vein furovirus, BNYVV [28].

### *Cloning, mutagenesis and deletions*

In all constructs, designations of plasmids and transcripts are preceded by the letter 'p' and 't' respectively. Most DNA manipulations were performed according to standard protocols [29]. Minipreps for the final transcription/translation experiments were prepared using a plasmid mini kit (Qiagen) as indicated by the supplier. Plasmids were screened by restriction analysis. A plasmid construct derived from pEMBL19 and containing a full-length copy of the TYMV cDNA, pTYFL7, was obtained as described [30].

A single-stranded DNA matrix for mutagenesis, MX-wt, was obtained by cloning an *MroI*1846-*XhoI*3282 blunt-ended fragment of pTYFL7 (1436 base pairs) into the *EcoRI*-*SmaI* digested and blunt-ended M13mp18 vector. Mutants were generated according to the USB protocol using the following synthetic oligodeoxynucleotides containing mismatches (bold type letters), and artificially introduced restriction sites (underlined) and codon changes: 5'GGAGAGAAGAGAATTCAAGG3' (*EcoRI*) for the Cys783 to Ser mutation, MX-C783S; 5'GCCGGGTGAGAATTCGGATGGCGGTCC3' (*EcoRI*) for the His869 to Glu mutation, MX-H869E; 5'CTTCATGTTTAAATCAAGTTC3' (*DraI*) for the Ser925 to Leu mutation, MX-S925L; 5'CGGATATGTCGACCCACAGCCAGC3' (*SalI*) for the Lys982 to Ser mutation, MX-K982S; 5'GGAGCCCATGGGGGGTCCGC3' to introduce an additional *NcoI* site, designated *NcoI*\*. Mutations from recombinant phages were transferred to pTYFL7 by replacement of the wild-type *MroI*-*XhoI* fragment by the mutant counterparts yielding the full-length TYMV cDNA mutants, designated pC783S, pH869E, pS925L, pK982S, and p*NcoI*\*, respectively. Selected plasmid constructs were partially sequenced using the T7 DNA polymerase sequencing kit (Pharmacia).

To obtain the polyprotein containing a deletion upstream of the proteinase domain, the full-length mutant p*NcoI*\* was digested with *NcoI*. After heating to destroy the enzymatic activity, the sample was divided into two parts. The first was treated with the Klenow fragment of DNA polymerase I to fill in sticky ends, and the other was left untreated. After electrophoretic separation in an agarose gel and purification, both the sticky (s) and blunt larger DNA fragments (containing the vector and most of the TYMV cDNA sequence) were religated and transformed into competent *Escherichia coli* cells. The resulting plasmids (pdNN\*s and pdNN\*) contained open reading frames for the predicted proteinase domain downstream from the 5'-proximal 41 codons of the OP gene, or from the 5'-proximal 39 codons of the RP gene, respectively.

Polyprotein with a deletion between the proposed proteinase domain and the putative recognition/cleavage site was obtained from pK982S bearing the artificially-introduced *SalI*\* site. Its TYMV RNA-specific *SalI*\*3047-*EcoRI* region of the plasmid was replaced by the corresponding *ApaI*3708-*EcoRI* fragment using a synthetic *SalI*-*ApaI* linker (5'TCGAGTTCTAGACTGGCCAGGGGCC3', (+), upper strand, and 5'CCTGGCCAG-TCTAGAAC3', (-), lower strand).

### *Transcription*

The deleted and mutated plasmids were linearized by *EcoRI* (located in the multiple cloning site) or by *SmaI* (position 6061 in the TYMV cDNA) respectively. Transcription with T7 RNA polymerase (BRL) was performed essentially as recommended by the supplier, using 1 mM each of ATP, CTP and UTP, 0.5 mM GTP and 2 mM m<sup>7</sup>GpppG. After purification on a G-50 Sephadex column, phenol extraction and ethanol precipitation, integrity and size of the resulting transcripts were estimated by native gel electrophoresis, and concentration by A<sub>260</sub> measurements.

*Translation in vitro*

The standard translation conditions (total volume 10  $\mu$ l) using a rabbit reticulocyte lysate, and 370 kBq of L-[<sup>35</sup>S]cysteine (~48 TBq/mmol; Amersham) or of L-[<sup>35</sup>S]methionine (~40 TBq/mmol; ICN) were as described [13]. Incubation conditions and final RNA concentrations were as indicated in Fig. 3. The translation products (generally 8  $\mu$ l) were analyzed on 12.5% acrylamide – 0.1% bis-acrylamide-SDS gels [31]. The gels were then fixed and fluorographed.

**Results***Prediction of the proteinase domain*

To develop a hypothesis about the possible location of the TYMV proteinase domain, we analyzed sequences of large proteins of tymo- and potexviruses (Fig. 1). These proteins proved to be closely related in the regions of the putative methyltransferase, NTPase and polymerase domains [11,12]. However, it seemed conceivable that the proteinase domain would be conserved among the tymovirus sequences but not among the potexvirus sequences, because potexviruses did not seem to utilize proteolysis as a strategy for the expression of their proteins (A. Karasev, pers. comm.). After the entire sequences of the large proteins of tymo- and potexviruses were aligned (not shown), stretches of conserved tymovirus sequences that had no counterparts in potexvirus sequences were detected. Such tymovirus-specific portions of the alignment were primary targets of our search.

The tymoviral sequences contained neither chymotrypsin-like nor aspartic proteinase motifs known to occur in polyproteins of many viruses. However, it was also known that the genomes of animal Sindbis-like viruses as well as of several plant and animal viruses belonging to other supergroups encode papain-like proteinases [4]. The updated list of such viruses includes alphaviruses [3, 32], rubiviruses [33], corona- and artheriviruses [34, 35], potyviruses [36], hypovirulence-associated virus (HyAV) of the chestnut blight fungus [37, 28] (Fig. 2A), probably aphthoviruses [4, 39], bymoviruses [40, 41], hepatitis E virus [16], and possibly, pestiviruses [42]. The proteinases of these viruses have little or no sequence similarity with each other or with cellular papain-like proteinases except for two regions, containing (putative) catalytically active Cys and His residues (Fig. 2A). Our analysis of the tymovirus sequences revealed such a conserved amino acid dyad upstream of the putative NTPase domain. In the TYMV sequence, these amino acids, Cys783 and His869, were located within the 688-amino acid region which is reported to be essential for autocleavage [15]. The conserved Cys is followed by Leu (or Met in the case of ELV); this is in contrast to other papain-like proteinases, which have an aromatic amino acid at this position (Figs. 2A, B), with one exception, the Staphylococcal proteinase. Similar to most other papain-like proteinases, both the Cys and His residues are preceded by regions rich in structure breakers, such as Pro, Gly, Asn, Ser, although the conserved amino acid pattern is unique for the tymovirus sequences (Figs 2A, B). There are no other invariant Cys residues



followed by a His residue at an appropriate distance in the alignment of the tymoviral sequences.

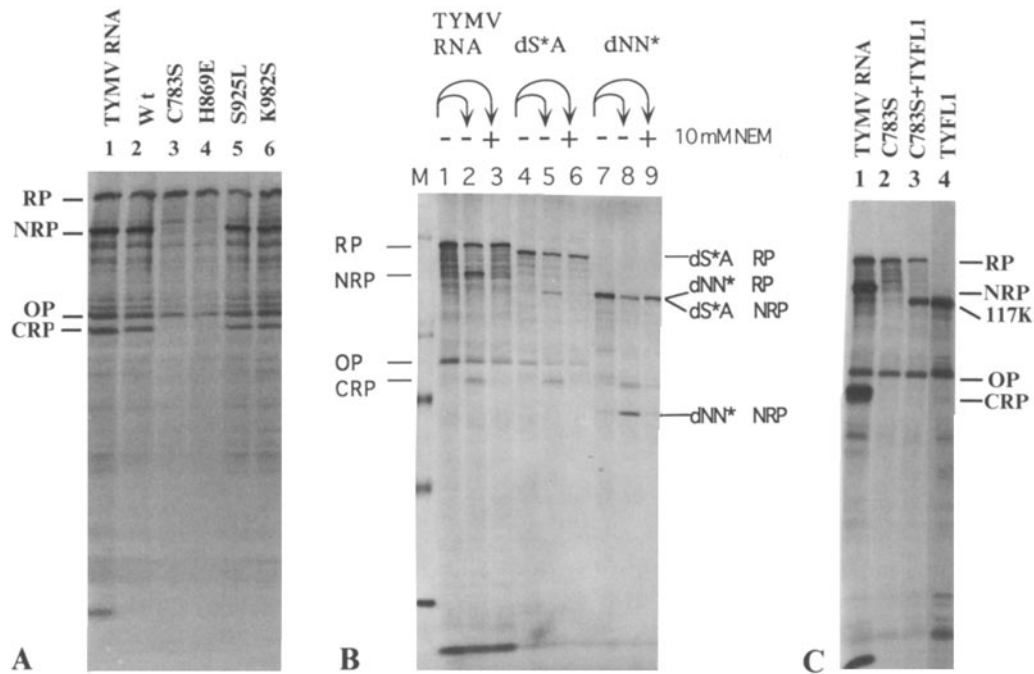
Using the level of similarity between tymovirus sequences as the main criterion, and by analogy with known papain-like proteinases, the N-terminal border of the TYMV proteinase domain was predicted to be not further than 49 amino acid residues upstream of the essential Cys783. The C-terminal border of the proteinase could be located in a fairly short variable stretch of the sequence, between His869 and the putative NTPase/helicase domain. As the borders of the latter were not determined, "motif" A [19], GFAGCGKT, where K is the invariant Lys982, had to be considered its N-terminal border.

Thus, an unusual papain-like proteinase with catalytic Cys783 and His869 was predicted in tymoviral replicase polyproteins close to, and upstream of, the putative NTPase domain.

#### *Testing the prediction of the papain-like proteinase of TYMV*

This hypothesis was tested by mutational and deletion analyses of the predicted proteinase domain (Fig. 1; Fig. 3A, B). Site-directed point mutations were introduced into a TYMV cDNA clone, pTYFL7. Their influence on auto-proteolysis of RP was then investigated by *in vitro* translation of corresponding mutant transcripts using both standard (Fig. 3A) and temperature shift conditions [13] (data not shown) for each of them. In full agreement with our predictions, replacement of Cys783 by Ser (C783S) or of His869 by Glu (H869E) abolished RP to NRP and CRP hydrolysis (Fig. 3A, lanes 3, 4). Other substitutions such as replacement of the invariant Ser925 for Leu (S925L; lane

**Fig. 2.** Alignment of the sequences of the (putative) proteinases of Sindbis-like plant viruses (tymo-like proteinases) and of the sequence fragments around the conserved (catalytic) Cys and His of other viral and cellular papain-like proteinases. **A** Alignment of the sequence segments around the conserved (catalytic) Cys and His of viral and selected cellular papain-like proteinases. The motifs were aligned manually following previously published alignments [4, 35, 38]. M-pro, main proteinase, L-pro, leader proteinase according to [4]. SSP *Streptococcus pneumonia* proteinase, SNBV *Sindbis virus*, RUBV *rubella virus* (a rubivirus), EAV *equine arteris virus* (an arterivirus), MHV *mouse hepatitis coronavirus*, TEV *tobacco etch potyvirus*. Numbers indicate distances between the motifs. **B** Sequence alignment of the (putative) proteinases of tymoviruses. The alignment made use of the program OPTAL [19] as follows. Alignment of 5 tymovirus sequences was produced by adding to the sequence of the TYMV proteinase those of KYMV, EMV, OYMV, and ELM stepwise. The respective alignment scores were 17.5, 17.6, 18.5 and 14.6 SD. *Consensus 1* Consensus pattern for tymovirus proteinases. Positions occupied by bulky hydrophobic residues (I, L, V, M, F, Y or W) in at least 4 sequences are indicated by asterisks. Positions of the sequences in viral polyproteins are shown left. **C** Alignment of the sequences of putative proteinases of carlaviruses, capilloviruses, ASPV, ACLSV and of BNYVV. The sequence fragment of PVM was aligned with that of ASPV and the sequences of ASGV and ACLSV were subsequently added. The scores were 15.6, 6.4 and 7.0 SD, respectively. For further explanations see text. *Consensus 2* is derived using the same rules as *Consensus 1*



**Fig. 3.** Experimental testing of the predicted papain-like proteinase domain of RP of TYMV. **A** Mutation analysis of the TYMV proteinase domain. SDS-PAGE analysis of the in vitro translation products programmed by TYMV RNA (1), wt TYFL7 (2) or the mutated transcripts C783S (3), H869E (4), S925L (5), K982S (6). In this experiment L-[<sup>35</sup>S]methionine was used instead of [<sup>35</sup>S]cysteine. The final concentration of all templates was 10 µg/ml. Incubations were performed at 30 °C. The positions of RP and of its cleavage products, NRP and CRP, as well as of OP, are indicated to the left. **B** Determining the borders of the proteinase domain. SDS-PAGE analysis of the in vitro translation products programmed by TYMV RNA or the deletion transcripts dS\*A and dNN\*. The final template concentration was 50 µg/ml for TYMV RNA or for dNN\*, and 200 µg/ml for dS\*A. Incubations were initiated at 37 °C, and after 30 min an aliquot was removed for analysis (1, 4, 7). To the remainder, puromycin was added without (-) or with (+) N-ethylmaleimide (NEM), and incubation was continued for another 10 min at 37 °C before shifting to 30 °C (2, 3, 5, 6, 8, 9) and incubating for 50 min at that temperature. Samples (8 µl) were analyzed in each case. The positions of RP (1-3), and of NRP (2) and CRP (2, 5, 8) are indicated to the left. To the right are provided the positions of dS\*A-RP (4-6), dNN\*-RP (7-9), dS\*A-NRP (5), and of dNN\*-NRP (8). M corresponds to <sup>14</sup>C-labelled protein markers (Amersham) of 200 K, 97.4 K, 69 K, 46 K and 30 K. **C** SDS-PAGE analysis of the in vitro translation products programmed by transcript C783S in the absence of presence of the transcript produced from TYFL1. The latter construct linearized from pTYFL1 by *Sac*I corresponds to nucleotides 1 to 3283 of the TYMV genome [29] and codes for a polyprotein of 117 K that is terminated upstream of the putative cleavage/recognition site. The final template concentration was 50 µg/ml for TYMV RNA (1), 200 µg/ml for tC783S (2, 3) and 100 µg/ml for tTYFL1 (3, 4), incubation was for 60 min at 30 °C. In 3 the molar ratio of C783S over TYFL1 is approximately 1 to 1. The positions of the major products are indicated to the right. All wells are from the same gel



5), or of the invariant Lys982 (located in the NTPase motif, GXXXXGKT) for Ser (K982S; lane 6) had no influence on proteolysis.

To localize the proteinase domain in a short stretch of sequence containing the essential Cys and His, two deletion variants of pTYFL7 were constructed, pdNN\* and pdS\*A. To delete the 5-proximal half of the RP gene, a single point mutation T2217 to C was introduced into the TYMV cDNA upstream of, and close to, the predicted 5'-border of the proteinase gene so as to create a second *NcoI* site referred to as *NcoI*\*. The transcripts obtained from the *NcoI-NcoI*\* deletion variant of the TYMV cDNA (Fig. 1) gave rise upon translation in reticulocyte lysates to a truncated polypeptide (dNN\*) of 131 K, which underwent autoproteolysis yielding the NRP product and the unaltered CRP product (Fig. 3B, lanes 7–9) under conditions identical to those observed with a full-length transcript (wt; Fig. 3A, lane 2) or with TYMV RNA (Fig. 3B, lanes 1–3). Thus, the deleted amino acid sequences are not required for autoproteolysis of the polyprotein *in vitro*. In addition, the fact that the two transcripts tdNN\* and tdNN\*s, which differ in the first 39 codons upstream of the *NcoI* site, gave identical results (not shown), eliminated the possibility that the N-terminus of RP participates in autoproteolysis. Hence, in agreement with our prediction, the N-terminal border of the proteinase is downstream from amino acid 709 in the polyprotein.

Deletion of amino acids 982–1204 removed most of the NTPase domain (dS\*A; Fig. 1). Again, this deletion did not abolish autoproteolysis of the expected 182 K RP precursor to yield the NRP product and the unaltered CRP product (Fig. 3B, lanes 4–6). The possibility that the proteinase domain is located downstream from its cleavage/recognition site was excluded [15] (our unpubl. data). We conclude that the amino acid stretch located between residue 709 and the NTPase domain, displaying some similarity to the papain-like proteinases and bearing the Cys and His residues essential for proteinase activity, indeed harbors the TYMV proteinase.

#### *Does the proteinase work in trans?*

To investigate whether the proteinase can act in *trans*, tC783S was translated *in vitro* in the presence of transcripts encoding the wild-type proteinase domain, either tDNN\* (not shown) or tTYFL1 (Fig. 3C). No cleavage of the undeleted tC783S mutant polyprotein could be detected even with an 8- or 5-fold molar excess of tTYFL1 or tDNN\*, respectively, over tC783S (not shown).

#### *The TYMV proteinase is the prototype of a new distinct group of (putative) viral proteinases*

Figure 2B shows that, as mentioned before, the TYMV-like proteinase domain is strictly conserved among all tymoviruses. To determine whether other Sindbis-like plant viruses have (putative) TYMV-like proteinase domains, we examined their polypeptides for the presence of the tymo-like proteinase sequence motifs (in the case of BNYVV it was kindly done by E. V. Koonin

and V. V. Dolja, pers. comm. and [43]). Stretches of sequences which contained such motifs were analyzed for overall similarity to the tymovirus proteinase.

Figures 2B,C show the results of our analysis. Of the five sequences that were significantly similar to the TYMV proteinase, four (PVM, ASPV, ASGV, and ACLSV) were aligned with each other (Fig. 2C) and then compared with the alignment of the 5 tymoviral proteinase sequences. The resulting alignment score of 6.5 SD supports our hypothesis on the relationship of these sequences of the carlavirus PVM, and the capillovirus ASGV as well as of ASPV and ACLSV, to those of tymoviral proteinases. The best score for the putative BNYVV proteinase (6.7 SD) was obtained when its sequence was compared with the alignment of all the TYMV-like proteinase sequences. Such a score, although the worst one of the respective scores obtained for any of 9 other sequences, reflects a probable relationship.

Consensus sequence 1 derived for the tymoviral proteinases (Fig. 2B) matches well the consensus of the alignment of the other 5 sequences (Consensus 2, Fig. 2C). The resulting consensus (Fig. 2C) consists of the two motifs containing the essential Cys and His residues, respectively, and one more motif located between them, G\*(s/x), where G is Gly, \* is a hydrophobic residue, and s is Ser, which can be replaced by other residues (x). The conserved Gly is not present in the BNYVV sequence. Sequence database searches with the conserved motifs did not reveal additional relatives to the tymo-like proteinases.

*Are there other putative papain-like proteinases encoded by  
Sindbis-like viruses?*

In the sequences of polyproteins of plant Sindbis-like viruses we also observed some additional conserved short stretches that distantly resembled papain-like proteinases, but the similarities were below a statistical significance (not shown). However, taking into account the fact that the viral proteinases are extremely variable in their primary structure [4], one can not exclude the possibility that some of the sequences mentioned belong to unusual proteinase domains only distantly resembling the papain-like domains. Simple deletion/mutagenesis experiments rather than more extensive computer analyses may help to clarify these uncertainties.

### Discussion

The deletion/mutagenesis experiments (Figs. 3A, B) support our prediction of the papain-like thiol proteinase of TYMV with a putative catalytic dyad Cys-783 and His-869 (Fig. 1; Figs. 2A, B). Taking into account that, as mentioned above, there are no other obvious candidates for the role of catalytically active amino acids in the central region of the tymovirus polyproteins, it is likely that Cys783 and His869 play this role in the TYMV proteinase.

Inhibition studies of TYMV RP autocleavage also favor our conclusions. Inhibitors of thiol proteinases such as ZnCl<sub>2</sub> abolish autoproteolysis of RP even at low concentrations, while inhibitors of serine proteinases affect

proteolysis only at high concentration [13]. Failure of cystatin and E64 to influence hydrolysis of the RP [13] does not contradict our hypothesis, because not all genuine papain-like proteinases are sensitive to these inhibitors [44]. The question of whether the TYMV proteinase resembles papain in its 3D structure or at least in the configuration of its active site will be resolved only when X-ray data will become available.

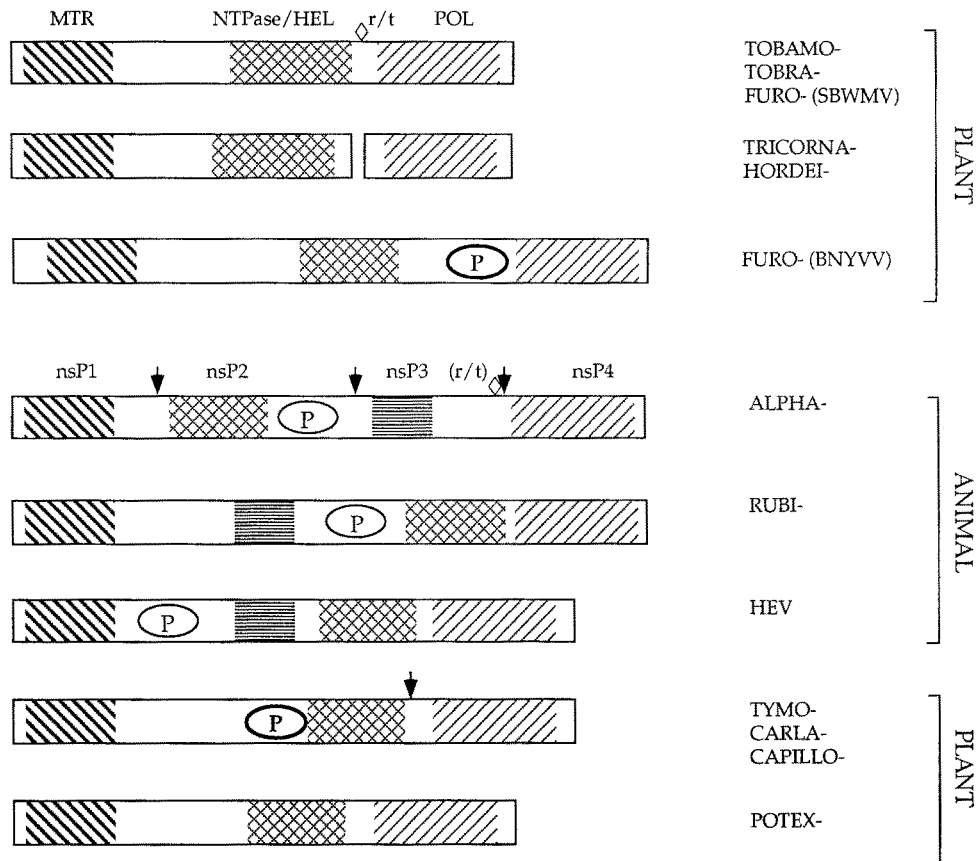
Figure 3C demonstrates that under our conditions the proteinase does not cleave the mutant polyprotein in *trans*, but acts in *cis* in agreement with previously reported data [13, 15]. Indeed, with the exception of the proteinases of alphaviruses [31], most viral papain-like proteinases apparently function only in *cis*.

An additional indirect indication of the similarity of the tymovirus proteinase to other viral papain-like domains is the putative recognition/cleavage site. This site has been tentatively located by site-directed mutagenesis within a 9 amino acid stretch, between amino acids 1253 and 1261 [15], at the border of the NTPase and polymerase domains. Alignment of five tymovirus sequences shows that this region of the tymoviral polyprotein is not highly conserved. But it does contain only small amino acid residues, Gly, Ala or Ser, in positions which correspond to amino acids 1258 and 1259 of the RP polyprotein of TYMV (M.N.R. unpubl.). Notably, a Gly residue is present in one of the corresponding positions in each sequence. Cleavage of the tymovirus polyproteins may take place at this site by analogy with those of other viral papain-like proteinases known to hydrolyze the peptide bond between two small amino acid residues (alphaviruses – nsP1/nsP2 and nsP2/nsP3 sites [45], potyviruses [46], HyAV [37, 38], artheriviruses [35]), or when at least P1 position (i.e. the position to the left of the scissile bond) is occupied by Gly or Ala (alphaviruses – nsP3/nsP4 site [45], coronaviruses – p28 site [47, 48]). Similar potential cleavage/recognition sites are found at corresponding positions in polyproteins of other viruses containing the TYMV-like proteinase domain (M.N.R., unpubl.).

Does the putative NTPase/helicase domain influence (regulate) autoproteolysis of RP? Previous studies have not excluded this possibility [15]. Deletion of the putative TYMV NTPase domain or mutation of the invariant Lys982, an essential amino acid residue in the vast majority of NTP-binding enzymes [19, 49] did not affect autoproteolysis (Fig. 3A, B). Consequently, the putative NTPase is not required for cleavage of RP in reticulocyte lysates.

Our results were recently confirmed by experiments [50] following our prediction of the TYMV papain-like proteinase domain and our mutagenesis data of the predicted catalytic Cys783 and His869 [51]. Extensive deletion analysis has shown that the proteinase can tolerate deletions upstream from residue 731 and downstream from residue 885 *in vitro* [50].

Computer-assisted sequence analysis predicts that proteinase domains related to that of TYMV are encoded upstream from the putative NTPase gene of tymoviruses, carlaviruses, capilloviruses, ASPV and ACLSV. We suggest to name the new group the TYMO-LIKE PROTEINASE GROUP, since TYMV



**Fig. 4.** Domainal organization of the replicative proteins of Sindbis-like viruses. Conserved domains are hatched. *MTR* (Putative) methyltransferase; *NTPase/Hel* putative NTPase/helicase; *pol* (putative) polymerase; *P* or **P** papain-like proteinase predicted/characterized previously or in this study, respectively; *r/t* read-through; *nsP1*, *nsP2*, *nsP3* and *nsP4* are the alphavirus non-structural proteins; diamonds: stop codons; arrows: cleavage sites; *SBWMV* soilborne wheat mosaic furovirus

is the prototype of this new group. In addition to the invariant papain-like Cys-His dyad, the tymo-like proteinases contain a third highly conserved amino acid residue, Gly (Gly821 in TYMV RP), followed by a hydrophobic residue. This motif is apparently missing from other papain-like proteinases and is thus an additional hallmark of this group. After the Cys- and His-containing motifs, the conserved Gly seems to be the third interesting target for mutagenesis experiments. The putative proteinase of BNYVV does not conserve the Gly residue, although it resembles the tymo-like proteinases at a statistically significant level. This is not surprising considering different location of this domain (between the putative NTPase and polymerase domains, Fig. 4) in the polyprotein and the fact that other BNYVV replicative proteins are not closely related to those of any plant RNA virus [11, 16].

Our attempts to produce significant sequence alignments between tymo-like and other proteinases were unsuccessful. We conclude that tymo-like proteinases

constitute a distinct group (Fig. 2) within the majority of viral papain-like enzymes [4], and speculate that the members of this group may share biochemical and biological features. Taking into account the indispensability of the essential papain-like Cys-His dyad for replication of TYMV [50], (K. Séron, G. Kadaré, A.-L. H., unpubl.) and of blueberry scorch carlavirus (D. M. Lawrence, M. N. R., B. I. Hillman, manuscript in prep.), and similar processing patterns observed for 5 different tymoviruses [13–15] and 3 carlaviruses [52], it is tempting to speculate that all tymo-like proteinases cleave the large viral polyproteins yielding a functional “methyltransferase-helicase” block [11, 53] as an N-terminal product and an RNA polymerase domain as a C-terminal product. Notably, the alphavirus counterparts of the TYMV cleavage products, P123 and P34 are essential to form the viral replicative complex [54]. The other six groups of Sindbis-viruses shown on Fig. 4 do not seem to utilize the strategy of proteolysis of their proteins. Nevertheless 5 of them, i.e. tobamo-, tobra-, furo (SBWMV), tricorna- and hordeiviruses, do produce the “methyltransferase-helicase” block as a separate protein, but use two different mechanisms [5]. Thus, the strategy of processing of the large nonstructural viral polyprotein by the papain-like domain seems to be one of several possible solutions to the problem of organizing and regulating production of the viral replicase complex.

This study shows (see also Fig. 4) that the large replicase polyproteins of not only all animal Sindbis-like viruses but also of many of their plant counterparts such as tymo-, carlo-, capilloviruses, ACLSV, ASPV, and BNYVV possess (putative) papain-like proteinase domains in addition to conserved (putative) methyltransferase, helicase and polymerase domains [11]. We believe that this reflects similar functional and/or evolutionary constraints on the organization of the replicative machineries of animal and plant viruses. However, precise roles of the TYMV-like proteinases in virus replication and other potential functions remain to be elucidated.

### Acknowledgements

We thank Drs. A. Gorbalenya for useful discussions and critical reading of the manuscript; J. Gombold for critical reading of the manuscript; N. Teterina, F. Bernardi, M. Girondot, K. Séron, A. Kujawa, G. Kadaré and J.-C. Boyer for discussions and various help; and E. Koonin for checking the sequence of BNYVV for the presence of the TYMV-like proteinase motifs. M.N.R. is also grateful to T. Candresse and J. Wellink for discussions, to Dr. M. Kirpichnikov for the opportunity to perform some of experiments in the Institute of Molecular Biology (Moscow), and to Dr. S. Weiss for her interest in this work, and the opportunity to complete computer analysis and preparation of this manuscript in the University of Pennsylvania. M.N.R. was supported by a ‘Poste Rouge’ from the ‘Centre National de la Recherche Scientifique’. This work benefited from grants from the Indo-French Centre for the Promotion of Advanced Research, and from: the ‘Ligue Nationale Française Contre le Cancer’. The Institut Jacques Monod is an ‘Institut Mixte, CNRS-Université, Paris 7’.

### References

1. Gorbalenya AE, Donchenko AP, Blinov VM, Koonin EV (1989) Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases: a distinct protein superfamily with a common structural fold. *FEBS Lett* 243: 103–114
2. Bazan JF, Fletterick RJ (1988) Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proc Natl Acad Sci USA* 85: 7872–7876
3. Hardy WR, Strauss JH (1989) Processing of nonstructural polyproteins of Sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in *cis* and in *trans*. *J Virol* 63: 4653–4664
4. Gorbalenya AE, Koonin EV, Lai MM-C (1991) Putative papain-related thiol proteases of positive-strand RNA viruses: identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-,  $\alpha$ - and coronaviruses. *FEBS Lett* 288: 201–205
5. Goldbach R (1987) Genome similarities between plant and animal RNA viruses. *Microbiol Sci* 4: 197–202
6. Goldbach R (1990) Plant viral proteinases. *Semin Virol* 1: 335–346
7. Morch M-D, Boyer J-C, Haenni A-L (1988) Overlapping open reading frames revealed by complete nucleotide sequencing of turnip yellow mosaic virus genomic RNA. *Nucleic Acids Res* 16: 6157–6173
8. Weiland JJ, Dreher TW (1993) Cis-preferential replication of the turnip yellow mosaic virus RNA genome. *Proc Natl Acad Sci USA* 90: 6095–6099
9. Bozarth CS, Weiland JJ, Dreher TW (1992) Expression of ORF-69 of turnip yellow mosaic virus is necessary for viral spread in plants. *Virology* 187: 124–130
10. Weiland JJ, Dreher TW (1989) Infectious TYMV RNA from cloned cDNA: effects in vitro and in vivo of point substitutions in the initiation codons of two extensively overlapping ORFs. *Nucleic Acids Res* 17: 4675–4687
11. Rozanov MN, Koonin EV, Gorbalenya AE (1992) Conservation of the putative methyltransferase domain: a hallmark of the 'Sindbis-like' supergroup of positive-strand RNA viruses. *J Gen Virol* 73: 2129–2134
12. Rozanov MN, Morozov SYu, Skryabin KG (1990) Unexpected close relationship between the large nonvirion proteins of filamentous potexviruses and spherical tymoviruses. *Virus Genes* 3: 373–379
13. Morch M-D, Drugeon G, Szafranski P, Haenni A-L (1989) Proteolytic origin of the 150-kilodalton protein encoded by turnip yellow mosaic virus genomic RNA. *J Virol* 63: 5153–5158
14. Kadaré G, Drugeon G, Savithri HS, Haenni A-L (1992) Comparison of the strategies of expression of five tymovirus RNAs by in vitro translation studies. *J Gen Virol* 73: 493–498
15. Bransom KL, Weiland JJ, Dreher TW (1991) Proteolytic maturation of the 206-kDa nonstructural protein encoded by turnip yellow mosaic virus RNA. *Virology* 184: 351–358
16. Koonin EV, Gorbalenya AE, Purdy MA, Rozanov MN, Reyes GR, Bradley DW (1992) Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: delineation of an additional group of positive-strand RNA plant and animal viruses. *Proc Natl Acad Sci USA* 89: 8259–8263
17. Leontovich AM, Brodsky LI, Gorbalenya AE (1990) Generation of a complete local similarity map for two biopolymer sequences (DOTHELIX program of the GENESEE package) [in Russian]. *Biopolimery i Kletka* 6: 12–19

18. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16: 10881–10890
19. Gorbalenya AE, Blinov VM, Donchenko AP, Koonin EV (1989) An NTP-binding motif is the most conserved sequence in a highly diverged monophyletic group of proteins involved in positive strand RNA viral replication. *J Mol Evol* 28: 256–268
20. Ding S, Keese P, Gibbs A (1990) The nucleotide sequence of the genomic RNA of kennedy yellow mosaic tymovirus-Jervis Bay isolate: relationships with potex- and carlaviruses. *J Gen Virol* 71: 925–931
21. Ding S-W, Keese P, Gibbs A (1989) Nucleotide sequence of the ononis yellow mosaic tymovirus genome. *Virology* 172: 555–563
22. Osorio-Keese ME, Keese P, Gibbs A (1989) Nucleotide sequence of the genome of eggplant mosaic tymovirus. *Virology* 172: 547–554
23. Srifah P, Keese P, Weiller G, Gibbs A (1992) Comparisons of the genomic sequences of *erysimum latent virus* and other tymoviruses: a search for the molecular basis of their host specificities. *J Gen Virol* 73: 1437–1447
24. Zavriev SK, Kanyuka KV, Levay KE (1991) The genome organization of potato virus M RNA. *J Gen Virol* 72: 9–14
25. Jelkmann W (1994) Nucleotide sequences of apple stem pitting virus and of the coat protein gene of a similar virus from pear associated with vein yellows disease and their relationship to potex- and carlaviruses. *J Gen Virol* 75: 1535–1542
26. German S, Candresse T, Lanneau M, Huet JC, Pernollet JC, Dunez J (1990) Nucleotide sequence and genomic organization of apple chlorotic leaf spot closterovirus. *Virology* 179: 104–112
27. Yoshikawa N, Sasaki E, Kato M, Takahashi T (1992) The nucleotide sequence of apple stem grooving capillovirus genome. *Virology* 191: 98–105
28. Bouzoubaa S, Quillet L, Guilley H, Jonard G, Richards K (1987) Nucleotide sequence of beet necrotic yellow vein virus RNA-1. *J Gen Virol* 68: 615–626
29. Ausubel FM, Brent R, Kingston ER, Moore DD, Seidman JG, Smith JA, Struhl K (1990) *Curr Prot Mol Biol* 1: 1.1–1.7
30. Boyer J-C, Drugeon G, Séron K, Morch-Devignes M-D, Agnès F, Haenni A-L (1993) *in vitro* transcripts of turnip yellow mosaic virus encompassing a long 3' extension or produced from a full-length cDNA clone harbouring a 2 kb-long PCR-amplified segment are infectious. *Res Virol* 144: 339–348
31. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680–685
32. Strauss JH, Strauss EG (1990) Alphavirus proteinases. *Semin Virol* 1: 347–356
33. Marr LD, Wang C-Y, Frey TK (1994) Expression of the rubella virus nonstructural protein ORF and demonstration of proteolytic processing. *Virology* 198: 586–592
34. Baker SC, Yokomori K, Dong S, Carlisle R, Gorbalenya AE, Koonin EV, Lai MMC (1993) Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus. *J Virol* 67: 6056–6063
35. Snijder EJ, Wassenaar ALM, Spaan WJM (1992) The 5' end of the equine arteritis virus replicase gene encodes a papainlike cysteine protease. *J Virol* 66: 7040–7048
36. Oh C-S, Carrington JC (1989) Identification of essential residues in potyvirus proteinase HC-Pro by site-directed mutagenesis. *Virology* 173: 692–699
37. Choi GH, Pawlyk DM, Nuss DL (1991) The autocatalytic protease p29 encoded by a hypovirulence-associated virus of the chestnut blight fungus resembles the potyvirus-encoded protease HC-Pro. *Virology* 183: 747–752
38. Shapira R, Nuss DL (1991) Gene expression by a hypovirulence-associated virus of the chestnut blight fungus involves two papain-like proteinase activities. *Essential*

- residues and cleavage site requirements for p48 autoproteolysis. *J Biol Chem* 266: 19419–19425
39. Kleina LG, Grubman MJ (1992) Antiviral effects of a thiol protease inhibitor on foot-and-mouth disease virus. *J Virol* 66: 7168–7175
  40. Davidson AD, Pröls M, Schell J, Steinbiss H-H (1991) The nucleotide sequence of RNA 2 of barley yellow mosaic virus. *J Gen Virol* 72: 989–993
  41. Kashiwazaki S, Minobe Y, Hibino H (1991) Nucleotide sequence of barley yellow mosaic virus RNA 2. *J Gen Virol* 72: 995–999
  42. Stark R, Meyers G, Rumenapf T, Thiel H-J (1993) Processing of pestivirus polyprotein: cleavage site between autoprotease and nucleocapsid protein of classical swine fever virus. *J Virol* 67: 7088–7095
  43. Koonin EV, Dolja VV (1993) Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol Biol* 28: 375–430
  44. Ritonja A, Rowan AD, Buttle DJ, Rawlings ND, Turk V, Barrett AJ (1989) Stem bromelain: amino acid sequence and implications for weak binding of cystatin. *FEBS Lett* 247: 419–424
  45. Shirako Y, Strauss JH (1990) Cleavage between nsP1 and nsP2 initiates the processing pathway of sindbis virus nonstructural polyprotein P123. *Virology* 177: 54–64
  46. Carrington JC, Herndon KL (1992) Characterization of the potyviral HC-Pro autoproteolytic cleavage site. *Virology* 187: 308–315
  47. Hughes SA, Bonilla PJ, Weiss SR (1994) Analysis of MHV-A59 p28 cleavage site. 13th Annual Meeting of the American Society for Virology. Madison, Wisconsin, July 9–13. Abstract W45-2
  48. Dong S, Baker S (1994) Identification of the cleavage site of the first papain-like cysteine proteinase of murine coronavirus. 13th Annual Meeting of the American Society for Virology. Madison, Wisconsin, July 9–13 Abstract W45-3
  49. Schulz GE (1992) Binding of nucleotides by proteins. *Curr Opin Struct Biol* 2: 61–67
  50. Bransom KL, Dreher TW (1994) Identification of the essential cysteine and histidine residues of the turnip yellow mosaic virus protease. *Virology* 198: 148–154
  51. Rozanov M, Drugeon G, Séron K, Haenni AL (1992) Papain-related proteinase of turnip yellow mosaic virus. NATO/EEC Course “Regulation of gene expression by animal viruses, Mallorca, May 30-June 8 Abstract S16
  52. Foster GD (1992) The structure and expression of the genome of carlaviruses. *Res Virol* 143: 103–113
  53. Kao CC, Ahlquist P (1982) Identification of the domains required for direct interaction of the helicase-like and polymerase-like RNA replication proteins of brome mosaic virus. *J Virol* 66: 7293–7302
  54. Lemm JA, Rice CM (1993) Assembly of functional Sindbis virus RNA replication complexes: requirement for coexpression of P123 and P34. *J Virol* 67: 1905–1915

Authors' address: Dr. M. N. Rozanov, Department of Microbiology, University of Pennsylvania, 209 Johnson Pavilion, Philadelphia, PA 19104-6076, U.S.A.

Received June 28, 1994