POLYPROTEIN PROCESSING IN THE EXPRESSION OF THE COWPEA MOSAIC VIRUS GENOME

Joan Wellink

ISN 299628

Promotoren: Dr. A. van Kammen, hoogleraar in de moleculaire biologie
Dr. R.W. Goldbach, hoogleraar in de virologie.

Joan Wellink

POLYPROTEIN PROCESSING IN THE EXPRESSION OF THE COWPEA MOSAIC VIRUS GENOME

Proefschrift
ter verkrijging van de graad van
doctor in de landbouwwetenschappen,
op gezag van de rector magnificus,
dr. C.C. Oosterlee,
in het openbaar te verdedigen
op vrijdag 13 november 1987
des namiddags te vier uur in de aula
van de Landbouwuniversiteit te Wageningen.

ISN 299628

De elektronenmicroscopische opname op de omslag toont cowpea mozaiek virus deeltjes 2 x 10^5 keer vergroot.
(Met dank aan Jan van Lent).

CONTENTS

NN0270, 1177

STELLINGEN

1. De hypothese van Pearl en Taylor dat het retrovirus protease actief is als dimeer valt moeilijk te rijmen met het proces van autokatalytische klieving waarmee dit protease zich vrij maakt.

   Pearl and Taylor (1987) Nature 329, 351-354.

2. De genen voor de 2A en 3C eiwitten van poliovirus zijn ontstaan door duplicatie van een enkel protease gen.

   Dit proefschrift.

3. Het onderzoek aan proteases van poliovirus en cowpeamozaïek virus toont aan dat de evolutie van virusonderzoek ten minste zo interessant is als de evolutie van virussen.

   Franssen et al. (1984) J. Virol. 50, 183-190.
   Toyoda et al. (1986) Cell 45, 761-770.
   Vos (1987) Proefschrift LU Wageningen.

4. Uit de homologie in aminozuurvolgorden tussen het 250 kilodalton eiwit van beet necrotic yellow vein virus en het nsP1 eiwit van sindbisvirus kan geconcludeerd worden dat de furovirussen behoren tot de supergroep van "sindbis-like" plantevirussen.

   Goldbach (1987) Microbiological Sciences 4, 197-202.

5. Hoewel het tot nu toe niet gelukt is een door het cowpea mozaïek virus M-RNA gecodeerd 58 kilodalton eiwit te detecteren in met cowpea mozaïek virus geïnfecteerde planten, wordt dit eiwit waarschijnlijk wel door het virus geproduceerd.

   Dit proefschrift.

6. De eiwitten die Flore bij gekoppelde in vitro transcriptie en translatie van het sonchus yellow net virus na 22 uur incubatie in een tarwekiemextract gedetecteerd heeft, zijn het gevolg van een bacteriële en/of schimmelinfectie.

   Flore (1986) Proefschrift LU Wageningen.

7. Het toenemend aantal aanwijzingen dat recombinatie een belangrijk mechanisme is in de evolutie van virussen pleit voor het vermijden van het gebruik van termen als soort, geslacht en familie in de taxonomie van virussen.

4009s1

8. Kleren maken niet de mens.

9. De Olympische Spelen dienen in het vervolg onder de naam Wereldspelen georganiseerd te worden.

10. Wetenschappelijke doorbraken op medisch gebied hebben niet vanzelfsprekend een positieve invloed op het welzijn van de mens.

11. Toerisme leidt tot culturele armoe.

Stellingen behorend bij het proefschrift:

"Polyprotein processing in the expression of the cowpea mosaic virus genome".

13 november 1987

Joan Wellink

VOORWOORD

Dit proefschrift is tot stand gekomen met de hulp en inzet van velen. Met name wil ik bedanken:

Ab van Kammen en Rob Goldbach voor hun inspirerende begeleiding en betrokkenheid bij het onderzoek;

de CPMV groep: Geertje Rezelman, Jan Verver, Rik Eggen, Martine Jaegle, Pieter Vos en Juan Garcia voor de wijze waarop ze een groep vorm(d)en waarin het goed werken en vooral ook goed sámenwerken was (is);

Jan van Lent voor het maken van uitstekende E.M. opnames;

Louis Schrijvers, Agnes Tan, Eus Arts en Ves Jacobs voor hun bijdrage aan het onderzoek tijdens hun doctoraalstudie;

Konrad Beyreuther voor het bepalen van gedeeltelijke aminozuurvolgordes van virale eiwitten en voor het bieden van de mogelijkheid om een maand in zijn laboratorium te werken;

Heinrich Prinz voor het maken van een synthetische peptide en voor het aan mij overbrengen van de kennis om zulke peptides zelf te kunnen maken;

Piet Madern, Peter van Druten, Job Tielens en Reindert de Fluiter voor het teken- en fotowerk;

Marie-José van Neerven en Gré Heitkönig voor het typewerk;

Alle Molbiollers voor de uitstekende sfeer.

CHAPTER 1

SCOPE OF THE INVESTIGATION

Cowpea mosaic virus (CPMV) is a plant virus with a genome consisting of two positive strand RNAs which are separately encapsidated in icosahedral particles denoted Bottom (B) component and Middle (M) component. While both particles (or RNAs) are necessary to infect plants, B component RNA (B-RNA) is able to replicate independently in isolated cowpea protoplasts. Both RNAs contain a small protein (VPg) at their 5' end and have a poly(A) tail at their 3' end. Sequence studies have revealed that each RNA contains a single long open reading frame, strongly supporting the concept that the RNAs are translated into so-called polyproteins that are subsequently cleaved into several functional proteins. This polyprotein processing takes an important position in the expression of the CPMV RNAs because the production of proteins involved in replication of the viral RNAs and virus assembly is regulated by this process.

Several other groups of RNA viruses e.g. the animal picorna- and alphaviruses and, as was more recently shown, the plant poty- and nepoviruses use a strategy for gene expression similar to that of CPMV. A common feature of these viruses is that they encode the enzyme(s) for the proteolytic activity involved in the specific processing of the polyproteins. In chapter 2 different groups of viruses and their specific proteases are reviewed.

In chapters 3, 4, 5 and 6 of this thesis the processing of the CPMV polyproteins is studied. Chapter 3 describes the determination of the proteolytic cleavage sites in the B-RNA-encoded 200K polyprotein. This was established by partial sequencing of isolated cleavage products.

It was shown that the B-RNA-encoded 24K protein is a protease, able to cleave at specific sites by expressing cDNA clones, containing both M-RNA and B-RNA-specific sequences, in Escherichia coli (Chapter 5). Furthermore, by employing antibodies raised against a synthetic peptide of 25 residues it was shown that this protease is found in infected cells both as intermediate cleavage products (170K, 110K, 84K) and as a final cleavage product (24K) (Chapter 4).

A study on the expression of M-RNA in vivo is described in chapter 6. The detection of the M-RNA-encoded 48K protein and 60K precursor to both capsid proteins in infected protoplasts definitely proves that, similar as has been observed in vitro, M-RNA is also in vivo expressed via proteolytic processing of a polyprotein (Chapter 6).

Finally we have localized viral proteins in an infected cell by staining ultrathin sections of CPMV infected protoplasts with specific antibodies and gold-labeled protein A. It was found that the electron-dense material, which is a major cytophatic structure induced in a cell by a CPMV infection, contains viral non-structural proteins (Chapter 7).

CHAPTER 2

PROTEASES INVOLVED IN THE PROCESSING OF VIRAL POLYPROTEINS

Joan Wellink and Ab van Kammen

Department of Molecular Biology, Agricultural University,
De Dreijen 11, 6703 BC Wageningen, The Netherlands

Archives of Virology, in press

1. INTRODUCTION

Various RNA viruses express their genomic RNA by synthesis of a high molecular weight polyprotein which is post-translationally cleaved by proteases into functional viral gene products. This mode of gene expression has first been demonstrated for poliovirus (69), but has then been found with several other animal viruses as well as plant viruses.

The proteases catalyzing the specific cleavages of a polyprotein can either be encoded by the virus or occur in the host prior to infection. As pointed out by Rice and Strauss (55) there is a clear difference in the location in the cell of polyproteins cleaved by virus-encoded and host-encoded proteases. Polyproteins that reside in the cytoplasm are cleaved by virus-encoded proteases. These polyproteins are usually primary translation products. On the other hand viral glycoproteins which occur associated with membranes are cleaved by host proteases during maturation and/or transport.

In this review we will focus on the processing of cytoplasmic polyproteins catalyzed by virus-encoded proteases. Some viruses which encode protease(s) for polyprotein processing use in addition host proteases for the processing of their structural (glyco)proteins. This is for example the case with alphaviruses and flaviviruses (see sections 2.2. and 2.3.).

Cleavage of polyproteins by virus-encoded proteases into several active polypeptides has sofar only been found with some RNA viruses and not with DNA viruses. Some DNA viruses like adenovirus and bacteriophage T4 use proteases in the maturation of their capsid structures, but these proteases fall outside the scope of our review since in their case it is a question of releasing a single mature protein from its precursor. By definition these precursors are not considered polyproteins.

In the following section we shall describe different groups of viruses for which polyprotein processing by virus encoded proteases has been studied in some detail. For each virus group the current model for the expression of the genetic information will be presented. Some details will be given about the identification and characterization of the protease(s) such as sensitivity to certain protease inhibitors and the cleavage sites recognized by the protease(s). Some viruses for which processing by a virus-encoded protease has not yet firmly been established, but is very likely to occur, will also be described briefly. Finally the different virus-encoded proteases and their specificities will be compared with each other and with cellular proteases, and the cleavage sites recognized by these proteases will be discussed. We conclude the review by discussing possible mechanisms used by viruses to regulate this processing.

2. VIRUSES USING VIRUS-ENCODED PROTEASES IN THE EXPRESSION OF THEIR GENOMES

2.1. Retroviruses

Retroviruses contain a positive strand RNA of mol.wt. 3.2 x 10^6 which is replicated via DNA intermediates by the action of a virus-encoded reverse transcriptase. The retrovirus genome contains three regions which encode proteins as illustrated in Fig. 1 with the genetic map of moloney murine leukemia virus (Mo-MuLV) RNA. The gag region encodes a polyprotein that is cleaved to generate four structural proteins, the pol region encodes the reverse transcriptase and the env region glycoproteins found at the surface of the virion. The genomes of some other retroviruses possess additional genes. The gag polyprotein is cleaved by a virus-encoded protease that can be expressed in rather different ways. Rous sarcoma virus (RSV) and related retroviruses have a fifth gag protein, p15. Actually the first virus-encoded protease that was purified and characterized has been the p15 protein. It was isolated from virus particles and identified as the protease able to cleave in vitro the gag polyprotein $Pr76^{gag}$ synthesized by translation of RSV RNA in a cell-free system of ascites cells (76). Through the use of several inhibitors, the P15 protease of avian myeloblastosis virus (AMV), a retrovirus related to RSV, was characterized as a cysteine protease showing papain like activity (14). This protease was purified and

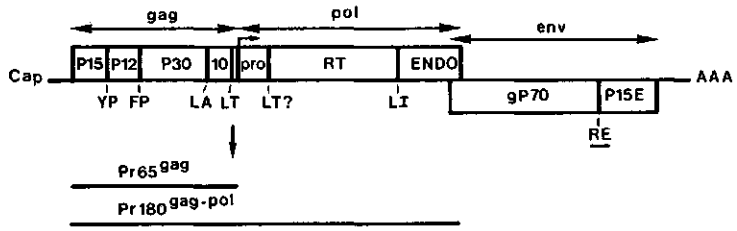


Fig. 1. Genetic map of Mo-MulV RNA. Some other retroviruses possess additional genes. The non-coding regions of the RNA are shown as single lines and the open reading frames as white bars. The coding regions of the individual proteins are shown inside the open reading frame and the cleavage sites used to produce these proteins are indicated underneath. Sites which are underlined are cleaved by host proteases. A question mark indicates that the origin of the protease catalyzing this cleavage is unknown. Readthrough of the stop codon between p10 and p20 which results in Pr180 gag-pol is indicated by Polyproteins translated from the RNA are indicated by single lines. Pro = protease, RT = reverse transcriptase, endo = endonuclease.

shown to cleave $Pr76^{gag}$ and the readthrough translation product $Pr180^{gag-pol}$, obtained by immunoprecipitation from infected cells, into defined viral proteins. The protease was not able to cleave the env precursor protein (45). The cleavages in the env polyprotein are probably achieved by a host protease during transport to the surface membrane of the infected cell (63). When the nucleotide sequence of RSV was determined, the cleavage sites in the gag polyprotein could be identified by sequencing the amino (N)-termini and carboxy (C)-termini of the gag proteins (p19-p10-p27-p12-p15). This revealed that small peptides were removed between p19 and p10, and p27 and p12 and that cleavages occurred between different amino acid pairs (Tyr His, Gly Ser, Met Pro, Met Ser, Met Ala, Ser Leu, and Leu Thr). No homology was present between the sequences surrounding the cleavage sites (cleavage occurs between the amino acid residues on positions -1 and +1), but on positions -2, -3, and -4 only amino acids with an aliphatic group (Pro, Ala, Val, Leu, Ile) were found (61, 63).

For retroviruses related to the murine (Mo-MulV) retroviruses the protease is not one of the gag proteins but is rather a part of the pol region produced after readthrough of the gag termination codon (Fig. 1) (83). In the Mo-Mulv gag precursor protein some homology around the cleavage sites is apparent, e.g. -4 Ser/Thr, -2 Ala/Leu, -1 Tyr/Leu and +1 Pro/Ala/Thr (65, 84). Homology around the cleavage sites also exists between different viruses in this group and some proteases are able to cleave the gag polyprotein of several other viruses (84, 85).

A third group of retroviruses encode their proteases in a reading frame independent of both the gag and pol regions e.g. human T cell leukemia virus 1 (HTLVI) and bovine leukemia virus (BLV) (62). The BLV protease has been purified and sequenced and based on this information it was proposed that this protease was expressed via a frameshift suppression of the gag termination codon (85).

Although the proteases of retroviruses can be expressed in different ways, their relative position in the viral genomes is similar and they show considerable sequence homology with each other. One of these conserved regions shows homology with the active site of acid proteases like pepsin (62, 71; see section 3). Furthermore, the processing of the gag polyprotein seems to be identical in all retroviruses.

It remains to be seen how the active retrovirus protease is generated from its precursor. Autocatalysis or an initial cleavage by a cellular enzyme may be involved.

2.2. Alphaviruses

Alphaviruses are enveloped viruses with a plus strand RNA genome of mol. wt. 4.3 x 10^6. The RNA is capped and contains a poly (A) tail at its 3' end. The genome of several alphaviruses has been sequenced (67, 70). Genomic RNA is translated into a high molecular weight protein which is the precursor of the four non-structural proteins. During an infection a subgenomic RNA is produced which is translated into a polyprotein from which

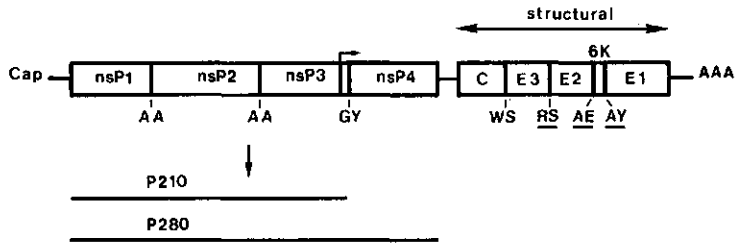


Fig. 2. Genetic map of SV RNA. SFV RNA does not code for a stop codon between nsP3 and nsP4. The structural proteins are translated from a subgenomic messenger (not shown). C is the capsid protein, E1, E2 and E3 are the envelope glycoproteins. For explanation of the symbols see legend to Fig. 1.

the structural proteins are released (see Fig. 2). Since processing of the alphavirus polyproteins has recently been reviewed by Strauss et al. (68), only a short summary will be presented here. Sindbis virus (SV) RNA and middelburg virus (MV) RNA are translated into two polyproteins, the larger being the result of readthrough of an opal termination codon located behind nsP3 (66). Thus the nsP4 protein is produced in lower amounts compared to the other proteins (42). The sequence of semliki forest virus (SFV) RNA does not contain a stopcodon at a similar position (70), but cells infected with SFV yet contain nsP4 in lower amounts than the other proteins, due presumably to a shorter half life of the nsP4 protein (36; see section 4). The nsP4 protein shows sequence homology with proteins implicated in RNA replication (1, 35), but it is not known why the amount of this protein has to be regulated in this way.

The cleavage sites in the polyproteins of several alphaviruses have been established by sequencing the cleavage products. A clear consensus is found for these sites, -3 Val/Ile/Ala, -2 Gly, -1 Ala/Gly/Cys, +1 Ala/Gly/Tyr (68). In vitro, the cleavages are inhibited by $ZnCl_2$ and TPCK (chymotrypsin inhibitor) and insensitive to PMSF (serine protease inhibitor) and TLCK (trypsin inhibitor) but this does not determine which type of protease is involved (12). One of the non-structural proteins probably is the virus-encoded

protease. By using an in vitro transcription translation system, evidence has been obtained that nsP2 is involved in the processing (E. Strauss, VII International Congress of Virology, Edmonton, August 1987).

The structural proteins probably are translated from a subgenomic RNA as a polyprotein (see Fig. 2). The capsid protein located at the N-terminus of the polyprotein is released by an autoprotease activity of the capsid protein. This was first postulated on the basis of in vitro translation experiments (2) and further evidence was obtained from sequence studies of temperature sensitive mutants defective in cleaving the capsid protein from the polyprotein precursor. The mutants mapped into a region of the capsid protein that showed considerable homology with the active site of serine proteases (27). Moreover this cleavage occurs behind a tryptophan residue which is characteristic of chymotrypsin-like serine proteases (68). The other structural proteins are thought to be released from their common precursor by host proteases such as signalases and golgi apparatus associated proteases (24, 55, 68).

2.3. Flaviviruses

Flaviviruses contain a single positive stranded RNA genome of mol. wt. 3.9 x 10^6. The RNA is capped and contains no poly (A) tail. Only after the complete nucleotide sequence of yellow fever virus (YFV) had been determined, revealing the presence of a single large open reading frame on the RNA, it was established that the flavivirus-encoded proteins were produced upon processing of a single polyprotein (54). The present knowledge on processing of flaviviral proteins has recently been reviewed by Strauss et al. (68). A model for the expression of flaviviral RNA is shown in Fig 3. Sequence data on the structural proteins and some non-structural proteins have revealed that the primary large-sized translation product is probably processed by both viral and host proteases (54, 68).

Only limited data are available for the amino acid sequences of the non-structural

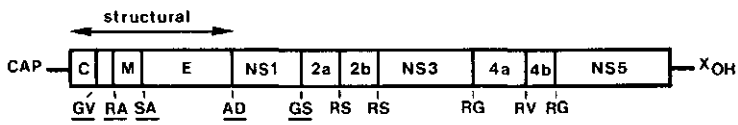


Fig. 3. Genetic map of YFV RNA. The non-structural proteins 2a, 2b, 4a and 4b have not yet been identified in infected cells. NS1 is a glycosylated non-structural protein, with an unknown function, C is the capsid protein, M is a glycoprotein and E is the envelope glycoprotein. For explanation of the symbols see legend to Fig. 1.

proteins of different flaviviruses (53) but a consistent set of amino acids is found at

known and putative cleavage sites; -2 Lys/Arg, -1 Arg, -1 Gly/Ser/Thr/Ala. Most flavivirus proteins are associated with membranes (10), but the cleavages are carried out in the cytosol by a protease which is probably virus-encoded. Furthermore, it was found that incubation of flavivirus infected cells with $ZnCl_2$ or TPCK resulted in the accumulation of larger flavivirus specific proteins compared to control infections (11), indicating that the flavivirus protease probably is a serine protease.

The organisation of the flavivirus structural proteins shows some similarity with the organisation of the alphavirus structural proteins, and probably in both cases the same type of proteases are used i.e. signalases and golgi apparatus associated proteases. The cleavages between the envelope glycoprotein E and glycoprotein NS1, and NS1 and NS2a are probably also carried out by signalases since these cleavage sites show homology to sites cleaved by signalases (54, 68). The occurrence of a conserved Cys Trp amino acid pair, which is part of the active site of thiol proteases, in glycoprotein precursor prM sequences from several flaviviruses, suggests that prM may function as a protease in the cleavage of glycoprotein precursor prM to the M protein (13). This cleavage occurs after two basic residues and consequently it has also been proposed that a trypsin-like golgi apparatus associated protease catalyses this cleavage (68).

2.4. Picornaviruses

The genome of picornaviruses consists of a single positive strand RNA molecule with a mol. wt. of about 2.5×10^6 with a small protein, VPg, covalently linked to the 5' end and a poly(A) tail at the 3' end. The proteolytic processing of the polyproteins encoded by these viruses has been reviewed by Palmenberg (56) and Nicklin et al. (52) and we refer to these reviews for additional information. The nucleotide sequence of the genome of various picornaviruses has been determined including members of each of the picornavirus subgroups i.e. the enteroviruses, poliovirus (PV), coxsackievirus (COX), and hepatitis A virus (HAV), the human rhinoviruses, HRV2 and HRV14, the cardiovirus, encephalomyocarditis virus (EMCV) and the aphthovirus, foot and mouth disease virus (FMDV). The genomes of the picornaviruses have a similar organisation. As an example the genetic map of PV is shown in Fig. 4. The RNA contains a single open reading frame encoding for a large polyprotein which has been divided into three regions, P1 (1ABCD), P2 (2ABC), and P3 (3ABCD) (see Fig. 4) from which proteins are derived with similar functions in all picornaviruses.

For PV the N-termini of most of the virus-encoded proteins found in infected cells have been sequenced and for some of these proteins the C-terminal ends as well. In this way it

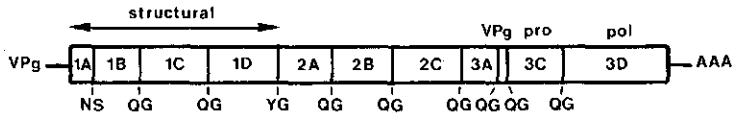


Fig. 4. Genetic map of PV RNA. The genome can be divided in three regions: P1 (1ABCD), P2 (2ABCD) and P3 (3ABCD). FMDV and EMCV RNA code for one additional protein (L) located at the 5' end. FMDV possesses three tandemly arranged VPgs and almost completely lacks the 2A coding region. For explanation of the symbols see legend to Fig. 1. Pol = RNA polymerase.

has been firmly established that the processing of the PV polyprotein happens by eight cleavages between Gln Gly amino acid pairs, one cleavage between 1D and 2A between a Tyr Pro pair and one cleavage between 1A and 1B between a Asn Ser pair (38; see Fig. 4).

For FMDV cleavage sites in the polyprotein have also been determined by protein sequencing but the cleavage sites in the polyproteins of most of the other picornaviruses have been determined by computer-assisted alignment of sequences using the high degree of sequence homology shared by all the viruses. There are three types of cleavage sites in picornavirus polyproteins.

1. Most of the cleavages occur between Gln Gly pairs or between pairs of related amino acids like Gln Ser, Glu Ser or Glu Gly.
2. The cleavage between 1A and 1B occurs at Ala Asp or Asn Ser pairs.
3. The cleavage between 1D and 2A is different from the other two types for most of the picornaviruses.

Three different proteolytic activities are involved in these cleavages, probably all virus-encoded.

Protein 3C is responsible for the first type of cleavage in all picornaviruses. In PV, protein 3C was identified as a protease by Hanecak et al. (28) using a cell free extract of PV infected Hela cells and antibodies raised against protein 3C. The antibodies were able to inhibit all cleavages at Gln Gly sites in the PV polyprotein produced in the extract, but not at the Tyr Gly site between 1D and 2A. Furthermore the part of the PV genome coding for the 3C protein and some of the flanking sequences were expressed in E. coli and it was found that the 3C protein was active as a protease in E. coli cells capable of removing the flanking sequences by cleavage at the normal Gln Gly sites (29). Introduction of an insertion in the 3C coding region abolished this activity, confirming that the 3C protein was responsible for the cleavages. Also for EMCV and FMDV, the 3C protein has been shown to possess proteolytic activity (8, 39, 58, 59). For several years it was thought that

the Tyr Gly cleavage between 1D and 2A in the PV polyprotein was achieved by a host-encoded activity, but recently, following the same approach as described for PV 3C protease, Toyoda et al. (72) showed that the 2A protein is the protease responsible for this cleavage.

The Asn Ser cleavage between 1A and 1B is different from the previous two types. This cleavage is linked with the maturation of the virus particles and occurs upon completion of the assembly of the RNA and capsids into virions. For all picornaviruses this cleavage is probably an autocatalytic event mediated by a serine residue, located at the N-terminus of 1B, and the RNA (5). All picornaviruses have the following pattern of processing. The first cleavage is between the P1 and P2 regions i.e. between protein 1D and 2A and happens in the nascent polyprotein chain. For several picornaviruses this cleavage is catalyzed by 2A but in the cardioviruses and aphthoviruses represented by EMCV and FMDV respectively this cleavage occurs between 2A and 2B and is mediated by a sofar unknown agent (33). The large majority of the remaining cleavages are catalyzed by protein 3C in all picornaviruses. The final cleavage occurs during maturation of the virion between 1A and 1B and appears to be an autocatalytic reaction.

There exists a considerable amount of amino acid homology between the picornaviral 3C proteases. Especially conserved are a histidine and cysteine residue in the C-terminal part of the proteins. These residues are possibly part of the active site of the protease (4). Furthermore Ivanoff et al. (32) have shown that the cysteine is essential since mutation of this cysteine to a serine residue abolished the proteolytic activity of PV 3C tested by expression of the corresponding sequence in E. coli. There also is some homology between the PV, HRV and COX 2A proteins and the 3C proteases, especially with regard to the sequence around the conserved cysteine residue (40; see section 3). This cysteine motive is not present in the HAV and EMCV 2A or 2B proteins and it remains an open question if in these viruses 2A (or 2B) is a protease.

All known cleavages carried out by the 3C proteases occur between Gln or Glu and Gly or Ser (and less frequently Thr/Ala/Met/Val) amino acid pairs. The surrounding sequences of these sites do not show much homology. Only at position -4 there is usually (but not always) an amino acid with an aliphatic side chain (52): preferentially Ala in the case of PV and COX, Val/Ile/Ala in the case of HRV, Leu/Val/Ile in the case of HAV and Pro in the case of FMDV. For EMCV Phe or Leu are usually present at position -4, and for 8 out of 9 sites a Pro is found at position -2 or +2 (56, 57). The above described characteristics are usually not present around amino acid pairs that are not cleaved (see section 4).

2.5. Comoviruses

Cowpea mosaic virus (CPMV), the type member of the comoviruses, is a plant virus with a bipartite positive stranded RNA genome of mol.wt. 2.2 x 10^6 (B RNA) and 1.2 x 10^6 (M RNA). Both RNAs have a VPg (3K) at their 5' ends and poly (A) tail at their 3' ends.

The nucleotide sequences of B-RNA and M-RNA contain each a single large open reading frame (41, 74). Both RNAs are translated into large polyproteins, which are subsequently cleaved through a number of successive steps into functional proteins (Fig. 5). The cleavage sites in the primary translation products were established by determining partial N-terminal sequences of various virus specific proteins found in CPMV infected cells and locating the coding regions of these proteins on each of the CPMV RNAs. Cleavages were found to occur between Gln Ser, Gln Gly and Gln Met amino acid pairs (Fig. 5). The cleavage sites have an Ala at position -4, and an Ala or Pro at position -2, but further show little amino acid homology (81). There are quite a number of Gln Ser, Gln Gly and Gln Met pairs in the polyprotein which are not cleaved, indicating that secondary and tertiary structures are probably important in determining the sites used for proteolytic processing (see section 4).

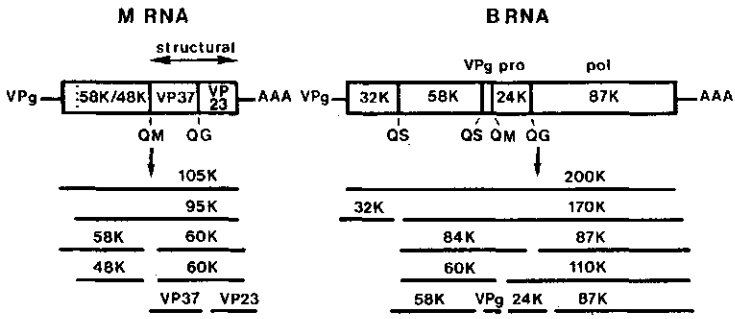


Fig. 5. Genetic map of CPMV RNAs. All proteins indicated in the figure have been identified in infected cells, with the exception of the M-RNA-encoded 105K, 95K and 58K proteins which only have been observed upon in vitro translation of M-RNA. For explanation of the symbols see legend to Fig. 1 and 4.

Results from in vitro translation studies with both CPMV RNAs suggested that two B-RNA-encoded proteins namely the 32K and 24K proteins were involved in the proteolytic processing of the CPMV polyproteins (19, 21, 60). The involvement of the 24K protein in the proteolytic processing was strongly supported by a strong amino acid sequence homology between the 24K protein and the picornaviral 3C proteases (20). Direct evidence that the B-RNA-encoded 24K polypeptide is the protease catalyzing the cleavages at the three types

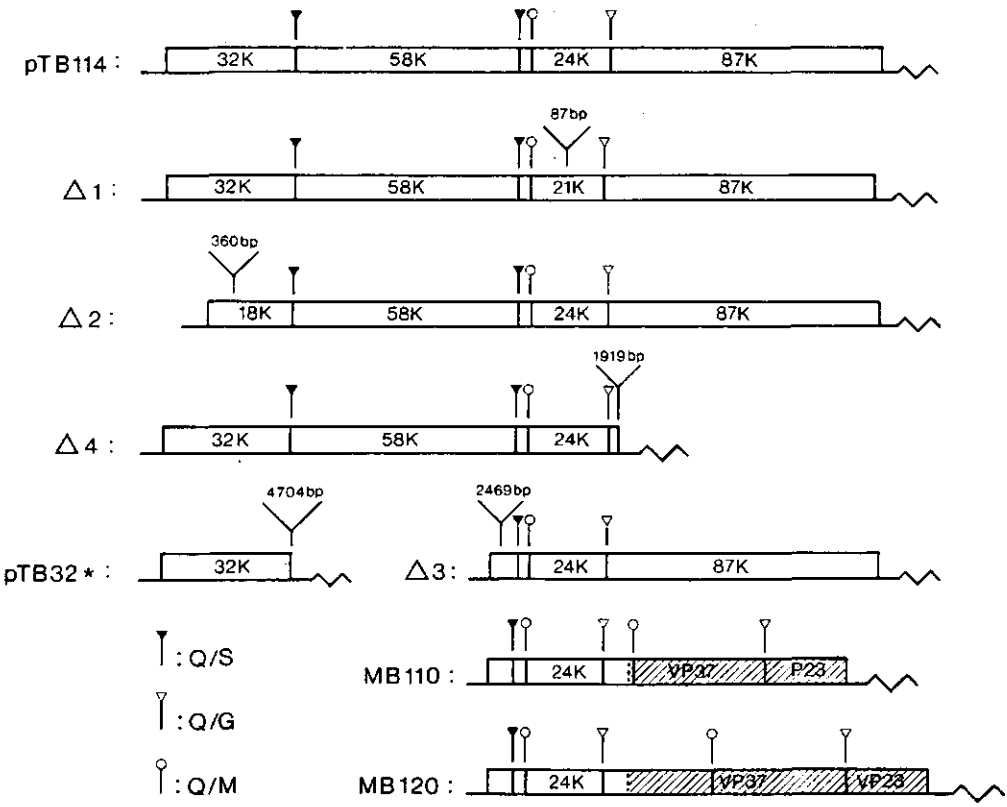


Fig. 6. Schematic representation of transcripts derived from complete and mutagenized cDNA clones of CPMV B- and M-RNAs. The open reading frames are indicated by white bars and the non-coding sequences by single lines. The open reading frames derived from M-RNA are hatched. The construction of these mutants is described by Vos et al. (1987). The positions and sizes of the deletions are indicated above each RNA. The cleavage sites used to release the individual proteins from the polyprotein are indicated by symbols which are explained in the figure.

of sites in the polyproteins encoded by B- and M-RNA was obtained by experiments using cDNA clones of both M- and B-RNA placed downstream of the T7 or SP6 promoter. RNA, transcribed in vitro from these clones produced, upon in vitro translation in reticulocyte lysates, proteins which were faithfully processed into final cleavage products identical to the proteins translated from the viral RNAs (75, 79, 80). Starting with full-length clones, several deletion mutants were constructed as illustrated in Fig. 6.

A deletion in the 24K coding region prevented proteolytic processing of the primary translation products completely, whereas deletions in other regions did not block proteolytic cleavages at the specific cleavage sites. This proves that the 24K protein, or proteins containing the 24K polypeptide, is able to catalyze all cleavages in the B-RNA-encoded polyprotein. It was also demonstrated that the 32K polypeptide previously thought to be a second B-RNA-encoded protease, did not have any protease activity.

To determine whether the 24K protein was also responsible for the Gln Met and Gln Gly cleavages in the M-RNA-encoded polyproteins, hybrid DNA copies were constructed which contained in their open reading frame the coding sequence of the B-RNA-encoded 24K protein and the M-RNA-encoded capsid proteins including the Gln Met and Gln Gly cleavage sites (Fig. 6). The results from in vitro translation experiments with transcripts from these hybrid clones demonstrated that the 24K protein was able to cleave at the Gln Met and Gln Gly sites occurring in the M-polyprotein part of the hybrid protein as well as the Gln Met and Gln Gly cleavage sites bordering the 24K sequence and occurring in the B-polyprotein part of the hybrid protein. The processing of the M-RNA-encoded polyproteins by the 24K protein is however more complicated than it appeared. Processing of these polyproteins was also tested by adding the translation products from RNA transcripts of full-size B-cDNA and of the deletion mutants thereof. If translation products obtained with mutant B-cDNA lacking an intact 32K polypeptide but having an intact 24K protein were added to M-RNA-encoded polyproteins no cleavage at the Gln Met site in the M-polypeptide occurred. Thus cleavage under native conditions requires not only the 24K protease but in addition an intact 32K protein. There is no specific proteolytic activity associated with the 32K polypeptide, but this protein is required as a cofactor or helper to achieve the Gln Met cleavage in the M-RNA-encoded polyprotein (79).

2.6. Nepoviruses

The genome of tobacco ringspot virus (TobRV), type member of the plant nepoviruses, consists of two species of plus strand RNA of mol. wt. 2.8 x 10^6 (RNA 1) and 1.3 x 10^6 (RNA 2). Both RNAs contain a VPg (4K) at their 5' end and a poly (A) tail at their 3'

end. In vitro TobRV RNA 1 is translated into a 207K protein which upon prolonged incubation in the presence of DTT is cleaved into 37K and 180K proteins whereupon the 180K protein is further processed into 128K and 65K proteins (34). RNA 2 has been translated into a 116K protein, which was processed upon addition of the translation products of RNA2 into smaller products among which the viral coat protein (18, 34). From these experiments it was concluded that RNA 1 coded for a protease able to cleave the RNA 2 translation product and probably also the RNA 1 products. A tentative model for the expression of the TobRV RNAs is shown in Fig. 7.

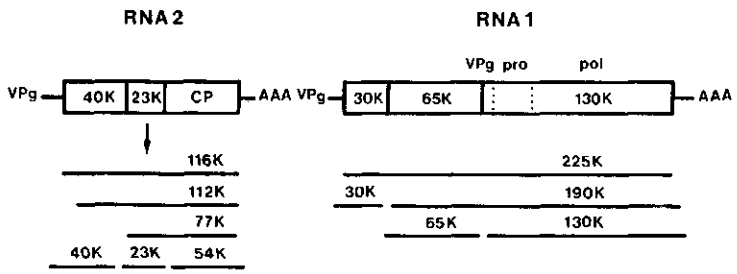


Fig. 7. Genetic map of TobRV RNAs. CP is the capsid protein. The dashed lines bordering the protease (pro) indicate that until now such a protein has not been identified. The presence of VPg, pro and pol in the TobRV 130K protein is based on TBRV RNA 1 sequence data (C. Fritsch, personal communication). For explanation of the symbols see legend to Fig. 1 and 4.

Translation of the RNA 1 of another nepovirus, grapevine fanleaf virus (GFLV), resulted in 220K, 195K and 190K proteins (49). In the presence of amino acid analogs only the 220K protein was synthesized. Purified RNA 2 could be translated into a 125K protein. Addition of RNA 1 and RNA 2 translation products resulted in processing of the 125K protein into a 68K protein and the 58K coat protein. A 62K protein was also detected but its origin was not clear.

Recently, the nucleotide sequence of the nepovirus tomato blackring virus (TBRV) RNA 2 was determined (44). The sequence contained one large open reading frame which could code for a 150K protein. This is in agreement with in vitro translation studies with RNA 2 in which a 160K protein was produced (22). The 160K protein could not be processed into smaller products, but antiserum against the coat protein recognized this protein (22). The coat protein was tentatively located at the C-terminal part of the 150K protein and is possibly

released from its precursor by cleavage at a Glu Ser or a Glu Thr site (Fig. 7; 44). This cleavage is probably catalyzed by a TBRV RNA 1 encoded protease similar to what has been observed for TobRV. The assumption of a protease encoded by RNA 1 is supported by recent sequence information obtained for TBRV RNA 1 (C. Fritsch, personal communication). A considerable amount of amino acid homology with the CPMV B-RNA-encoded 24K protease was found at a position in the TBRV RNA 1 sequence corresponding to the position of the 24K coding sequence on the CPMV B-RNA (Fig. 6) (C. Fritsch, personal communication, see section 3).

2.7. Potyviruses

The potyvirus group is the largest among plant viruses and is characterized by rod shaped virions with a length of 720-800 nm. The genome of these viruses consists of plus strand RNA of mol. wt. 3,2 x 10^6. The RNA contains a VPg (6K) at the 5' end and a poly (A) tail at the 3' end. In vitro translation of potyvirus RNA has usually resulted in a large set of proteins. Using antisera raised against several purified potyvirus proteins, and peptide mapping, a genetic map of the potyvirus genome has been constructed and it was proposed that potyvirus proteins are produced by proteolytic processing of a polyprotein (17). By translating soybean mosaic virus RNA in the presence of N-formyl-(^{35}S)methionyl -tRNAi^{met} and amino acid analogs, additional evidence was obtained that polyproteins were produced (73). When translation of papaya ringspot virus (PRV) RNA was carried out without a reducing agent (DTT), a large protein (330K) was produced clearly showing that PRV RNA can be translated into a single large polyprotein which is subsequently processed (82). Using the hybrid arrested translation technique the genetic map of tobacco vein mottling virus (TVMV) RNA was further refined and some evidence was obtained for a proteolytic activity correlated with a defined region on the genetic map (31). Recently the complete nucleotide sequence of the RNAs of two potyviruses, tobacco etch virus (TEV) and TVMV were determined. It was found that the genome of both viruses contained a single large open reading frame (3, 15).

Protein sequence information was only obtained for the capsid proteins while for TVMV it was shown that the capsid protein was released from the C-terminal end of the polyprotein by cleavage at a Gln Ser site. Furthermore, by placing the genetic map on the protein sequence, it was possible to propose the cleavage sites most probably used for the release of the other proteins from the polyprotein (see Fig. 8). Five of these sites show a high degree of homology (-4 Val, -3 Arg/Lys, -2 Phe, -1 Gln, +1 Gly/Ser) (15). In TEV the capsid protein is also released by cleavage at a Gln Ser site while other sites have also been

tentatively identified. The five sites corresponding to those in TVMV showed a consensus (-6 Glu, -4 Ile/Val/Leu, -3 Tyr, -1 Gln, +1 Gly/Ser) (3).

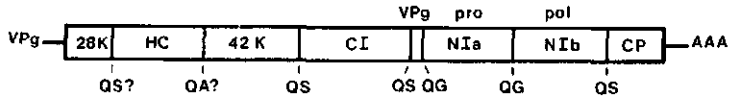


Fig. 8. Genetic map of TVMV RNA. CP is the capsid protein, NIa and NIb are proteins that in the case of TEV form nuclear inclusions, CI forms the cytoplasmic inclusions and HC is the helper component involved in vector transmission. VPg has not yet been mapped on the TVMV genome and its position is based on the analogy with the picorna- and comoviruses and the presence of suitable cleavage sites (see Fig. 4 and 5). For explanation of the symbols see legend to Fig. 1 and 4.

Amino acid homology was detected between the NIb protein of TEV and TVMV and the CPMV 87K protein involved in RNA replication and between the NIa protein and the CPMV 24K protease (3, 16). By in vitro translation of in vitro transcripts of TEV cDNA it was shown that the NIa protein is the protease responsible for at least three cleavages in the polyprotein (9). Also some evidence has been obtained that the NIa protease is released from its precursor by autocatalytic cleavages. Furthermore, the proteolytic activity resided only in the C-terminal 25K region of the NIa protein (9). It appears likely that the potyvirus NIa protease cleaves at the 5 homologous cleavage sites. Probably a second viral protease, like the 2A protease in PV or a host-encoded protease is responsible for the other cleavages. There is some homology between the 42K protein of TEV and the 2A protein of PV suggesting that this 42K protein might be a second protease. On the other hand, such homology has not been found with the TVMV 42K protein (16). Therefore it is possible that the homology of TEV 42K and PV 2A is fortuitous and it seems premature to assume a protease function for the 42K protein.

2.8. Tymoviruses

Turnip yellow mosaic virus (TYMV), the type member of the tymoviruses, possesses a monopartite RNA genome of mol. wt. 2.10^6. The RNA contains a cap at its 5' end and a t-RNA-like structure at its 3' end (for a recent review see 30). The coat protein (20K) is translated from a subgenomic messenger corresponding to the 3' region of the genomic RNA. The genomic RNA is translated in vitro into 150K and 195K proteins, the 150K protein being

the result of a so-called "stop" signal which is probably not a normal termination codon (47). Translation into the 195K protein stops at a UAG codon which can be read through resulting in the production of a 210K protein which partially overlaps the coat protein gene coding region (47). Upon translation in different in vitro systems the 195K and 210K proteins are proteolytically cleaved into 120K and 78K and 93K proteins respectively whereas in each system the 150K protein was found to be a stable product (46, 48, 87). On the basis of these studies it was proposed that the C-terminal part of the 195K protein encodes a proteolytic activity which cleaves the 195K polypeptide autocatalytically (48). Until now nothing is known about the sequence of this putative protease and the nature of the cleavage sites.

2.9. Other plant viruses

For some other plant viruses there is circumstantial evidence that they use proteolytic processing in their expression and that this activity is virus encoded. Turnip rosette virus (TRosV) contains a single RNA species of mol. wt. 1.4 x 10^6. The RNA contains a VPg (12K) linked to its 5' end. In vitro translation of this RNA resulted in four products with m.w. 105K, 67K, 35K and 30K (coat protein) and it was shown that the 67K protein was probably a cleavage product of the 105K protein (50). The protease catalyzing this cleavage seems to be virus encoded and is of the thiol type, as deduced from the effect of the inhibitor N-ethylmaleimide on the processing (50).

Pea enation mosaic virus (PEMV) contains two RNAs with positive polarity. The RNAs possess a VPg (17.5K) and no poly (A) tail. In vitro, RNA 1 is translated into 147K, 88K and 36K proteins. Tryptic peptide comparisons reveal that the latter two proteins are related to the 147K protein. Furthermore it has been shown that the 88K protein contains sequences related to the capsid protein. The capsid protein has a mol. wt. of 22K and therefore it is very likely that proteolytic processing in involved in the expression of this virus (23).

This is probably also the case for the monopartite luteoviruses. The genomic RNA of potato leaf roll virus (PLRV) possesses a VPg (7K) and no poly (A). Translation of the RNA in vitro yields 71K and 125K proteins, but no capsid protein (24K) can be detected under these conditions (43).

Recently the RNA 1 of beet necrotic yellow vein virus (BNYVV), type member of the furoviruses, was sequenced to reveal one large open reading frame capable of coding for a protein of 240K (7). In vitro translation of this RNA yields 50K and 150K proteins and a small amount of a 200K protein (88). Probably these proteins are derived from a 240K precursor by proteolytic processing but there are no data available about the protease(s)

involved in this process.

3. COMPARISON OF THE VIRAL PROTEASES WITH EACH OTHER AND WITH CELLULAR PROTEASES

In this section we will discuss the implications of the amino acid homology between the different viral proteases on the one hand and between viral proteases and cellular proteases on the other hand (4, 16, 20, 26, 71). The cellular proteases can be divided into four main families, the serine-, cysteine-, acid- and metallo-proteases (51).

Table 1. Families of proteolytic enzymes.

family	characteristic active site residues*	representative protease
serine proteases I	His (57), Asp (102), Ser (195)	chymotrypsin
serine proteases II	Asp (32), His (64), Ser (221)	subtilisin
cysteine proteases	Cys (25), Asp (158), His (159)	papain
acid proteases	Asp (32), Asp (215)	pepsin
metallo proteases I	Zn, Tyr (248), Glu (270)	carboxypeptidase A
metallo proteases II	Zn, Glu (143), His (231)	thermolysin

*The numbers behind each residue correspond to the position in the amino acid sequence of the enzymes listed in column 3.

A representative example of each family is given in Table 1 (modified from 51), together with the amino acids constituting the active site. In the serine and metallo protease families two groups can be distinguished which have similar active site configurations but bear no other structural or evolutionary relation to each other. They can therefore be considered examples of convergent evolution. At the other hand a protease with amino acid sequence homology and an active site configuration similar to one of the six proteases shown in Table 1 is probably derived from a common ancestor protein. There are still other families of proteases for which the mechanism of action remains to be elucidated but it seems likely that a peptide bond can be cleaved in only a few ways. The four families listed in Table 1 probably represent the most important mechanisms (51).

The survey of virus-encoded proteases presented in the previous section leads us to distinguish three different groups. First there is a large group of picornavirus 3C-related (P3CR) proteases which comprises also the virus-encoded proteases of comoviruses, potyviruses and nepoviruses. The proteases encoded by these viruses show homology in amino acid sequence and have similar configurations of amino acids presumably involved in the active site (4, 16, 20; see also Table 2). Probably the protease of the insect (picorna) virus cricket paralysis virus (CrPV) should also be included in this group (37). These different groups of viruses have moreover a very similar genome organisation and use the same strategy for the expression of their genetic information. The sequence homology is not limited to the proteases encoded by the viruses but also includes significant amino acid sequence homology between other proteins with presumably similar functions in viral RNA replication (for a more elaborate discussion about the implications of these homologies see 25). Gorbalenya et al. (26) noticed an homology between the active sites of the cellular serine proteases and the P3CR proteases, and also some homology with cellular cysteine proteases, and on basis thereof proposed that the P3CR proteases might form an evolutionary link between the serine and cysteine proteases. In our opinion it is equally possible that these homologies are the result of convergent evolution since the location in the protein of the amino acids forming the putative active site of the P3CR proteases namely a cysteine and histidine residue (4), is quite different from the location of the cysteine (or serine) and histidine forming the active site in the cellular serine and cysteine proteases (Tables 1 and 2). The active center of the serine and cysteine proteases moreover contains an aspartic acid residue (see Table 1). Although there is less amino acid homology between the N-termini of the P3CR proteases, the sequence (Lys/Arg, Asp, aliphatic amino acid) is found in all P3CR proteases around amino acid position 80 (see Table 2). It seems plausible that this aspartic acid residue is also part of the active center of the P3CR proteases.

Table 2. Viral proteases.

group	viral protease	type of protease as inferred from protease inhibitor studies (1)	putative active site residues based on homology with other proteases (2)	consensus cleavage site (3) -1	+1	consensus amino acids surrounding the cleavage site
retroviruses	RSV p15	cysteine	Asp (30), Asp (95)	-	-	-2, -3, -4, Pro, Ala Val, Leu, Ile
	MoMuLV p13	-	Asp (30), Asp (95)	Tyr, Leu	Pro, Ala, Thr	-4 Ser, Thr -2 Ala, Leu
alphaviruses	nsP2	serine	-	Ala, Gly, Cys	Ala, Gly Tyr	-3 Ala, Val, Ile -2 Gly
	capsid	?	His (140), Asp (145), Ser (215)	Trp	Ser	-5, -3 Glu -2 Glu, Gln
flaviviruses	?	serine	-	Arg	Gly, Ser Thr, Ala	-2 Lys, Arg
picornaviruses	2A (4)	-	Asp (40), Cys (110), His (118)	Tyr, Thr, Ala	Gly	-4 Leu, Ile, Met
	3C	cysteine	Asp (80), Cys (150), His (165)	Gln, Glu	Gly, Ser	-4 Ala, Val, Leu, Ile, Pro, Phe
comoviruses	CPMV 24 K	cysteine	Asp (80), Cys (150), His (165)	Gln	Gly, Ser, Met	-4 Ala -2 Ala, Pro
nepoviruses	TBRV pro (5)	cysteine	Asp (90 ?), Cys (170 ?), His (185 ?)	?	?	-
potyviruses	*TEV NIa*	*cysteine*	*Asp (255), Cys (340),* His (355)	*Gln*	*Gly, Ser*	-6 Glu, -4 Ile, Leu, Val -3 Tyr

(1) A question mark indicates that the use of inhibitors did not reveal the nature of the protease.

(2) The position of the active site residue in the protease is shown in brackets.

(3) A question mark indicates that the cleavage sites are not known.

(4) Only PV, HRV and COX possess a 2A protease.

(5) The nepovirus protease is not yet identified, and its position is only based on amino acid homologies (see section 2.6).

The sequence around a conserved cysteine residue in the picornaviral 2A proteases shows homology with the P3CR proteases (40). A histidine is found to be conserved in the 2A proteases at about a similar position relative to the cysteine as the histidine of the P3CR proteases. The 2A proteases contain furthermore a conserved Ser/Arg, Asp, Leu sequence around amino acid position 40 which also may be a constituent of the active site of these proteases just as was proposed for the P3CR proteases. In view of the similarities between the 2A protein of picornaviruses, for which this protein is a protease, and the 3C protein, it is possible that the 2A gene is the result of a duplication of the 3C gene.

A second group of virus-encoded proteases is formed by the retrovirus proteases which contain several conserved regions if compared among each other but show no homology with the other viral proteases. One of the conserved regions shows homology with the sequence around the aspartic acid present in the active site of cellular acid proteases (Leu, Val, Asp, Thr, Gly, Ala; 71). As there are two aspartic acids involved in the active site of cellular acid proteases (Table 1) it is of interest that in the retrovirus protease a second conserved sequence can be observed around an aspartic acid near position 150 suggesting that these two aspartic acids may constitute the active site of the retroviral proteases. The sequence around amino acid position 150 (Ile, Leu, Gly, Arg, Asp) bears no resemblance to the acid protease sequence. If these two aspartic acid residues indeed constitute the active site, the homology of the sequence around one of the aspartic acids of the retroviral proteases and that of the acid proteases might be another example of convergent evolution. Based on inhibition studies the retroviral protease has previously been characterized as a cysteine protease, but a conserved cysteine has not been found in these enzymes (see 62, 71). The retrovirus protease is therefore probably not a cysteine protease, but in view of the occurrence of the aspartic acid residues as described above, an acid protease. This is an example of how cautious one has to be with the interpretation of the results of studies with protease inhibitors (see Table 2). In caulimoviruses (plant DNA viruses) a protease domain related to the retrovirus protease is present (71). This suggests that protein processing plays a role in the expression of the caulimovirus genome, but at this time it is not clear whether processing of caulimovirus proteins by this putative protease really takes place.

A third group of virus-encoded proteases are the alphavirus and flavivirus proteases involved in the cleavages of the non-structural proteins of these viruses. The proteases have sofar not been characterized, but since none of the non-structural proteins of these viruses show homology to known viral or cellular proteases it seems probable that these proteases constitute a new group within one of the four main families of proteases.

The alphavirus capsid protease was characterized as a serine protease and was shown to

have a substantial amount of homology with the trypsin-like serine proteases (27). This homology is not limited to the amino acids forming the active site and those surrounding the active site but also includes the relative position of these amino acids in the two proteins. It is therefore possible that the capsid protease and the trypsin-like proteases are derived from a common ancestor.

For the other virus groups described in this review, proteases still have to be identified and sequenced. It is, however, tempting to assume, based on similarities in genome organisation and expression, that the TYMV and BNYVV protease may be related to the animal alphavirus protease and that the proteases of the VPg containing plant viruses (PEMV, TRosV, PLRV) are related to the P3CR proteases.

From the sequence information for cellular and viral proteases available at this moment, it may be concluded that the virus-encoded proteases of most viruses have evolved rather independently and that homology with cellular protases is the result of convergent evolution. A possible exception may be the capsid protease of the alphaviruses.

4. THE CLEAVAGE SITES

The amino acid pairs that are cleaved by proteases encoded by different virus groups are shown in Table 2. Almost all viral proteases cleave their substrates between a limited set of amino acid pairs. The polypeptide chain usually also contains a number of the same amino acid pairs that are not cleaved, which implies that there are additional factors involved in determining the sites that are used for proteolytic pocessing.

When the cleavage sites shown in Table 2 are compared with each other, it can be concluded that the P3CR proteases not only show amino acid sequence homology with each other (see section 3) but also catalyze cleavages between related amino acid pairs (Gln or Glu at -1 and Gly or Ser or less frequently Ala, Met, Thr, Val at +1). Each of the P3CR proteases is however highly specific for its own polyprotein and not able to cleave polyproteins from other viruses. The amino acid sequences surrounding the sites cleaved by the picornavirus 3C proteases show no homology. At position -4 there is a preference for an amino acid with a hydrophobic group (Ala, Val, Leu, Ile, Pro, Phe) the identity of which can be different for each group (see section 2.4). The sequences surrounding the sites cleaved by the CPMV and potyvirus proteases show a higher degree of homology (Table 2). These proteases also prefer an amino acid with a hydrophobic group at position -4. The

picornavirus 2A proteases, which show homology with the picornaviral 3C proteases, only perform a single autocatalytic cleavage at their own N-termini. These proteases cleave between a different subset of amino acid pairs (Tyr Gly, Ala Gly or Thr Gly) but these cleavage sites possess again an amino acid with a hydrophobic group at position -4.
From this comparison it appears probable that the hydrophobic residue at position -4 plays a role in determining whether a site is used for proteolytic processing by the P3CR protease.
The cleavage sites in the polyprotein between the non-structural proteins of the alphaviruses and flaviviruses are very well conserved. Also the amino acids at position -2 (flaviviruses) and -2 and -3 (alphaviruses) are very well conserved (Table 2) which implies that sites that are substrates for processing by these proteases are clearly defined. The cleavage sites of the retrovirus proteases show some more variation. The only consensus found for the RSV gag protein cleavage sites is that only aliphatic amino acids are present at positions -2, -3 and -4 (Table 2 and section 2.1). On the other hand, the cleavage sites in the MoMuLV gag protein are rather similar (Table 2), and also show homology with cleavage sites in the gag protein of other retroviruses. Furthermore, contrary to the picornaviruses proteases, some retroviral proteases are able to cleave the gag precursor of related viruses.
In summary, cleavage sites of viral proteases are usually characterized by a preference for certain amino acids at one or more of the positions -2, -3, -4, -5. It seems that amino acids at the (+) positions (with the exception of +1) do not have an important role in determining whether a site is used for proteolytic processing. This conclusion is also supported by experiments described by Vos et al. (79). By using site directed mutagenesis on cDNA clones of M-RNA of CPMV, the cleavage site between the two capsid proteins (Gln Gly) was modified. Insertion of an amino acid at the +2 position (Gln Gly Pro → Gln Gly Arg Pro) did not change the efficiency of cleavage of this site by the 24K protease in an in vitro assay (79). However, when the Gly in position +1 was replaced by an Ala, Met or Ser (amino acids which are found at the +1 position in other cleavage sites, see above), the cleavages at these sites proceeded with a very low efficiency. Semler et al. (64) modified the cleavage site between PV 3C and 3D by inserting a Ser in the -3 position. This site was no longer cleaved by the 3C protease indicating the importance of the amino acids at the (-) positions in determining a cleavage site.
For cellular proteases, amino acids in the (-) positions are equally important in determining a cleavage site. Trypsin cleaves after Lys or Arg in the -1 position and V8 protease of Staphylococcus aureus after Glu in the -1 position. Signalases only cleave at sites where certain subsets of amino acids are present at position -1 and -3 (77, 78) and

substilisin has a specificity towards certain amino acids at position -4. However the great specificity of most viral proteases for their substrates indicates that factors like secondary and tertiary structure are also important in determining which sites are cleaved by these proteases.

Interestingly, almost all new N-termini created by viral proteases consist of Gly, Ser, Ala, Thr, Met or Val (see Table 2). Bachmair et al. (6) recently proposed that proteins present in the cytoplasm which have at their N-terminus one of these amino acids have a long half-life ($t_{½}$ > 20 h). It is tempting to speculate that viruses have selected for proteins having one of these amino acids at their N-termini and therefore also for proteases having specificity towards these amino acids at the +1 position of cleavage sites. A notable exception is nsP4 of alphaviruses which has a Tyr at its N-terminus. These proteins have indeed a short half-life ($t_{½}$ ~ 10 m) (36 ; see section 2.2) as predicted by the N-terminal rule of Bachmair et al. (6).

5. CONCLUSIONS

Many proteins are synthesized as precursor proteins which still require a proteolytic processing to produce mature active proteins. The final processing reaction is usually an important step in the regulation of the production of the active protein. Likewise a virus can regulate the processing of a polyprotein, and thus the production of active proteins in several ways. 1) Some cleavages only happen under certain conditions as for example cleavage of PV 1AB into 1A and 1B which can only take place during the last steps in virus assembly when the RNA is present in the capsid. This is essentially a maturation reaction (5). 2) Cleavage products may have very different half lives as in the case of the alphavirus nsP4 which has a much shorter half-life than the other non-structural proteins (see sections 2.2 and 4). The functional significance of this difference is not yet clear. 3) A polyprotein can be cleaved via alternative pathways resulting in intermediates with different activities. The CPMV 170K protein can be cleaved in 110K and 60K products or 87K and 84K products (see Fig. 5). Presumably the different intermediates possess different activities or have different roles in viral RNA replication. 4) In most cases the protease catalyzing the cleavages in a polyprotein is itself part of this protein. This offers the possibility to regulate the processing, as the protease can cleave intramolecularly (in cis) and/or intermolecularly (in trans). Usually the intramolecular cleavages are rapid for example the 1D 2A cleavage in the PV polyprotein catalyzed by 2A. Most viral proteases

are able to release themselves by autocatalytic cleavages at both borders. However some viral proteases are also able to cleave intramolecularly at other sites, for example the 32K 170K cleavage in the CPMV B-RNA-encoded polyprotein catalyzed by the 24K protein as soon as during translation the growing polypeptide chain has reached the 24K protein coding region. 5) Another way to regulate occurs if cleavages take place in a specific order, i.e. a certain cleavage can only take place after another cleavage has been carried out. For example, during replication of CPMV (and picornavirus) RNA, VPg becomes attached at the 5' end of newly synthesized RNA. In the case of CPMV, VPg is probably released from a 60K precursor which itself first has to be released from a 170K precursor (see Fig. 5). In that case the activity of the protease is closely linked up with the replication of viral RNA. This connection is also apparent in the remarkably conserved order of the VPg, protease and polymerase proteins on the genomes of picorna-, como-, poty-and nepoviruses (see Fig. 4, 5, 7, 8 and ref. 25).

Recently Ypma-Wong and Semler (86) used an in vitro transcription translation system to show that the P1 region, containing the capsid proteins 1A, B, C and D, first has to be released from the polyprotein by action of 2A before the other cleavages in 1ABCD can take place. Furthermore they showed that an intact 3C coding region was sufficient to support cleavages between 2A 2B, 2B 2C and 2C 3A, while the cleavages between the structural proteins 1B, 1C and 1D only took place in the presence of an intact 3CD coding region (see Fig. 4).

The activity of a viral protease can again be regulated by other viral proteins. This has recently been described for the 24K protease of CPMV which has to be associated with the 32K protein for cleaving the Gln Met site in the M-polyprotein (see Fig. 5; 79).

In conclusion it is apparent that RNA viruses that express their viral genomic RNA or part of it by polyprotein synthesis and proteolytic processing encode their own proteases. The proteolytic processing proceeds in successive steps and appears to have a regulatory role in the expression of viral encoded functions. There are a number of different viral encoded proteases, which appear to have evolved independently from each other and from existing cellular proteases. The viral encoded proteases are highly specific in recognizing only the polyprotein of its own virus. Usually the proteases are part of the polyprotein that is proteolytically cleaved by them and presumably the proteases have evolved as a part of the polyprotein. It provides the virus with a method of regulating the production of proteins involved in viral RNA replication and virus assembly.

ACKNOWLEDGEMENTS

We thank William Dougherty, Christiane Fritsch, Anne-Lise Haenni, James Strauss and Ann Palmenberg for providing us with manuscripts and results prior to publication, Piet Madern for artwork and Gré Heitkönig for typing the manuscript.

1. Ahlquist P, Strauss EG, Rice CM, Strauss JH, Haseloff J, Zimmern D (1985) Sindbis virus proteins nsP1 and nsP2 contain homology to nonstructural proteins from several RNA plant viruses. J Virol 53: 536-542
2. Aliperti G, Schlesinger MJ (1978) Evidence for an autoprotease activity of sindbis virus capsid protein. Virology 90: 366-369
3. Allison R, Johnston RE and Dougherty WG (1986) The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for the synthesis of a single polyprotein. Virology 154: 9-20
4. Argos P, Kamer G, Nicklin MJH and Wimmer E (1984) Similarity in gene organisation and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. Nucleic Acids Res 12: 7251-7267
5. Arnold D, Luo M, Vriend G, Rossmann MG, Palmenberg AC, Parks GD, Nicklin MJH and Wimmer E (1987) Implications of the picornavirus capsid structure for polyprotein processing. Proc Natl Acad Sci USA 84: 21-25
6. Bachmair A, Finley D and Varshavsky A (1986) In vivo half-life of a protein is a function of its amino-terminal residue. Science 234: 179-196
7. Bouzoubaa S, Quillet L, Guilley H, Jonard G and Richards K (1987) Nucleotide sequence of Beet Necrotic Yellow Vein virus RNA-1. J Gen Virol 68: 615-626
8. Burroughs JN, Sangar DV, Clarke BE, Rowlands DJ, Billiau A and Collen D (1984) Multiple proteases in foot-and-mouth disease virus replication. J Virol 50: 878-883
9. Carrington JC and Dougherty WG (1987) Small nuclear inclusion protein encoded by a plant potyvirus is a protease. J Virol 61: 2540-2548
10. Castle E, Leidner U, Nowak T, Wengler G (1986) Primary structure of the west nile Flavivirus genome region coding for all nonstructural proteins. Virology 149: 10-26
11. Cleaves GR (1985) Identification of dengue type 2 virus-specific high molecular weight proteins in virus-infected BHK cells. J Gen Virol 66: 2767-2771
12. Collins PL, Fuller FJ, Marcus PI, Hightower LE and Ball LA (1982) Synthesis and processing of sindbis virus nonstructural proteins in vitro. Virology 118: 363-379
13. Dalgarno L, Trent, DW, Strauss JH and Rice CM (1986) Partial nucleotide sequence of the murray valley encephalitis virus genome. J Mol Biol 187: 309-323
14. Dittmar KJ and Moelling K (1978) Biochemical properties of p15 associated protease in an avian RNA tumor virus. J Virol 28: 106-118
15. Domier LL, Franklin KM, Shahabuddin M, Hellmann GM, Overmeyer JH, Hiremath ST, Siaw MFE, Lomonossoff GP, Shaw JG and Rhoads RE (1986) The nucleotide sequence of tobacco vein mottling virus RNA. Nucleic Acids Res 14: 5417-5430
16. Domier LL, Shaw JG and Rhoads RE (1987) Potyviral proteins share amino acid sequence homolgoy with picorna-, como- and caulimoviral proteins. Virology 158: 20-27
17. Dougherty WG and Hiebert E (1980) Translation of potyvirus RNA in a rabbit reticulocyte lysate: Cell-free translation strategy and a genetic map of the potyviral genome. Virology 104: 183-194
18. Forster RLS and Morris-Krisinich BAM (1985) Synthesis and processing of the translation products of tobacco ringspot virus in rabbit reticulocyte lysates. Virology 144: 516-519
19. Franssen H, Goldbach R and Van Kammen A (1984) Translation of bottomcomponent RNA of cowpea mosaic virus in reticulocyte lysate: faithful proteolytic processing of the primary translation product. Virus Res 1: 39-49
20. Franssen H, Leunissen J, Goldbach R, Lomonossoff G and Zimmern D (1984) Homologous sequences in nonstructural proteins from cowpea mosaic virus and picornaviruses. EMBO J 3: 855-861

21. Franssen H, Moerman M, Rezelman G and Goldbach R (1984) Evidence that the 32,000-dalton protein encoded by bottom-component RNA of cowpea mosaic virus is a proteolytic processing enzyme. J Virol 50: 183-190
22. Fritsch C, Mayo MA and Murant AF (1980) Translation products of genome and satellite RNAs of tomato black ring virus. J Gen Virol 46: 381-389
23. Gabriel CJ and De Zoeten GA (1984) The in vitro translation of pea enation mosaic virus RNA. Virology 139: 223-230
24. Garoff H, Frischauf A, Simons K, Lehrach H and Delius H (1980) Nucleotide sequence of cDNA coding for Semliki Forest virus membrane glycoproteins. Nature 288: 236-241
25. Goldbach R (1987) Genome similarities between plant and animal RNA viruses. Microbiological Sciences 4: 197-202
26. Gorbalenya AE, Blinov VM and Donchenko AP (1986) Poliovirus-encoded proteinase 3C: a posible evolutionary link between cellular serine and cysteine proteinase families. FEBS Lett 194: 253-256
27. Hahn CS, Strauss EG and Strauss JH (1985) Sequence analysis of three Sindbis virus mutants temperature-sensitive in the capsid protein autoprotease. Proc Natl Acad Sci USA 82: 4648-4652
28. Hanecak R, Semler BL, Anderson CW and Wimmer E (1982) Proteolytic processing of poliovirus polypeptides: Antibodies to polypeptide P3-7c inhibit cleavage at glutamine-glycine pairs. Proc Natl Acad Sci USA 79: 3973-3977
29. Hanecak R, Semler BL, Ariga H, Anderson CW and Wimmer E (1984) Expression of a cloned segment of poliovirus in E. coli: Evidence for autocatalytic production of the viral proteinase. Cell 37: 1063-1073
30. Heanni A, Morch M, Drugeon G, Valle R, Joshi RL and Denial T (1987) Gene expression in turnip yellow mosaic virus. In: Briton MA, Rueckert RR (eds) Positive strand RNA viruses. Alan R. Liss Inc, New York, pp 149-158
31. Hellman GM, Hiremath ST, Shaw JG and Rhoads RE (1986) Cistron mapping of tobacco Vein mottling virus. Virology 151: 159-171
32. Ivanoff LA, Towatari T, Ray J, Korant BD and Petteway Jr SR (1986) Expression and site-specific mutagenesis of the poliovirus 3C protease in Escherichia coli. Proc Natl Acad Sci USA 83: 5392-5396
33. Jackson RJ (1986) A detailed kinetic analysis of the in vitro synthesis and processing of encephalomyocarditis virus products. Virology 149: 114-127
34. Jobling SA and Wood KR (1985) Translation of Tobacco Ringspot virus RNA in reticulocyte lysate: proteolytic processing of the primary translation products. J Gen Virol 66: 2589-2596
35. Kamer G and Argos P (1984) Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. Nucleic Acids Res 18: 7269-7282
36. Keränen S and Ruohonen L (1983) Nonstructural proteins of semliki forest virus: synthesis, processing and stability in infected cells. J Virol 47: 505-515
37. King LA, Pullin JSK, Stanway G, Almond JW and Moore NF (1987) Cloning of the genome of criket paralysis virus: sequence of the 3' end. Virus Res 6: 331-334
38. Kitamura N, Semler B, Rothberg PG, Larsen GR, Adler CJ, Dorner AJ, Emini EA, Hanecak R, Lee JJ, Van der Werf S, Anderson CW and Wimmer E (1981) Primary structure, gene organisation and polypeptide expression of poliovirus RNA. Nature (London) 291: 547-553
39. Klump W, Marquardt O and Hofschneider PH (1984) Biologically active protease of foot and mouth disease virus is expressed from cloned viral cDNA in Escherichia coli. Proc Natl Acad Sci USA 81:3351-3355
40. Lloyd RE, Toyoda H, Etchison D, Wimmer E and Ehrenfeld E (1986) Cleavage of the cap binding protein complex polypeptide P220 is not affected by the

second poliovirus protease 2A. Virology 50: 229-303
41. Lomonossoff GP and Shanks M (1983) The nucleotide sequence of cowpea mosaic virus B RNA. EMBO J 2: 2253-2258
42. Lopez S, Bell JR, Strauss EG and Strauss JH (1985) The nonstructural proteins of sindbis virus as studied with an antibody specific for the C terminus of the nonstructural readthrough polyprotein. Virology 141: 235-247
43. Mayo MA, Barker H, Robinson DJ, Tamada T and Harrison BD (1982) Evidence that potato leafroll virus RNA is positive-stranded, is linked to a small protein and does not contain polyadenylate. J Gen Virol 59: 163-167
44. Meyer M, Hemmer O, Mayo MA and Fritsch C (1986) The nucleotide sequence of tomato black ring virus RNA-2. J Gen Virol 67: 1257-1271
45. Moelling K, Scott A, Dittmar KEJ and Ouada M (1980) Effect of p15associated protease from an avian RNA tumor virus on avian virus-specific polyprotein precursors. J Virol 33, 680-688
46. Morch MD and Benicourt C (1980) Post translational proteolytic cleavage of in vitro synthesized turnip yellow mosaic virus RNA-coded high molecular-weight proteins. J Virol 34: 85-94
47. Morch MD, Drugeon G and Benicourt C (1982) Analysis of the in vitro coded properties of the 3' region of turnip yellow mosaic virus genome RNA. Virology 119: 193-198
48. Morch MD, Zagorski W and Haenni AL (1982) Proteolytic maturation of the turnip yellow mosaic virus polyprotein coded in vitro occurs by internal catalysis. Eur J Biochem 127: 259-265
49. Morris-Krisinich BAM, Forster RLS and Mossop DW (1983) The synthesis and processing of the nepovirus grape vine fanleaf virus proteins in rabbit reticulocyte lysate. Virology 130: 523-526
50. Morris-Krisinich BAM and Hull R (1981) Translation of turnip rosette virus RNA in rabbit reticulocyte lysates. Virology 114: 98-112
51. Neurath H (1984) Evolution of proteolytic enzymes. Science 224: 350-357
52. Nicklin MJH, Toyoda H, Murray MG and Wimmer E (1986) Proteolytic processing in the replication of polio and related viruses. Biotechnology 4: 33-42
53. Rice CM, Aebersold R, Teplow DB, Pata J, Bell JR, Vorndam AV, Trent DW, Brandiss MW, Schlesinger JJ and Strauss JH (1986) Partial N-termini amino acid sequences of three nonstructural proteins of two flaviviruses. Virology 151: 1-9
54. Rice CM, Lenches EM, Eddy SR, Shin SJ, Sheets RL and Strauss JH (1985) Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science 229: 726-733
55. Rice CM and Strauss JH (1981) Nucleotide sequence of the 26S mRNA of sindbis virus and deduced sequence of the encoded virus structural proteins. Proc Natl Acad Sci USA 78: 2062-2066
56. Palmenberg AC (1987) Picornaviral processing: some new ideas. J Cell Bioch 33: 191-198
57. Palmenberg AC, Kirby EM, Janda MR, Drake NL, Duke GM, Potratz KF and Collett MS (1984) The nucleotide and deduced amino acid sequences of the encephalomyocarditis viral polyprotein coding region. Nucleic Acids Res 12: 2969-2983
58. Palmenberg AC, Pallansch MA and Rueckert RR (1979) Protease required for processing coat protein resides in the viral replicase gene. J Virol 32: 770-778
59. Parks GD, Duke GM and Palmenberg AC (1986) Encephalomyocarditis virus 3C protease: efficient cell-free expression from clones which link viral 5' noncoding sequences to the P3 region. J Virol 60: 376-384
60. Pelham HRB (1979) Synthesis and proteolytic processing of cowpea mosaic virus proteins in reticulocyte lysates. Virology 96: 463-477
61. Pepinsky RB, Mattaliano RJ and Vogt VM (1986) Structure and processing of the P2 region of avian sarcoma and leukemia virus gag precursor polyproteins. J

Virol 58: 50-58
62. Sagata N, Yasunaga T and Ikawa Y (1984) Identification of a potential protease-coding gene in the genomes of bovine leukemia and human T-cell leukemia viruses. FEBS Lett 178: 79-82
63. Schwartz DE, Tirard R and Gilbert W (1983) Nucleotide sequence of Rous sarcoma virus. Cell 32: 853-869
64. Semler BL, Johnson VH, Dewalt PG and Ypma-Wong MF (1987) Site-specific mutagenesis of cDNA clones expressing a poliovirus proteinase. J Cell Bioch 33, 39-51
65. Shinnick TM, Lerner RA and Sutcliffe JG (1981) Nucleotide sequence of moloney murine leukaemia virus. Nature 293: 543-548
66. Strauss EG, Rice CM and Strauss JH (1983) Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon. Proc Natl Acad Sci USA 80: 5271-5275
67. Strauss EG, Rice CM and Strauss JH (1984) Complete nucleotide sequence of the genomic RNA of Sindbis virus. Virology 133: 92-110
68. Strauss JH, Strauss EG, Hahn CS, Hahn YS, Galler R, Hardy WR and Rice CM (1987) Replication of alphaviruses and flaviviruses: proteolytic processing of polyproteins. In: Briton MA, Rueckert RR (eds) Positive strand RNA viruses. Alan R Liss, Inc, New York, pp 209-226
69. Summer DF and Maizel JV (1968) Evdence for large precursor proteins in poliovirus synthesis. Proc Natl Acad Sci USA 59: 966-971
70. Takkinen K (1986) Complete nucleotide sequence of the nonstructural protein genes of semliki forest virus. Nucleic Acids Res 14: 5667-5682
71. Toh H, Kikkuno R, Hayashida H, Miyata T, Kugimiya W, Inouye S, Yuki S and Saigo K (1985) Close structural resemblance between putative polymerase of a drosophila transposable genetic element 17.6 and pol gene product of moloney murine leukaemia virus. EMBO J 4: 1267-1272
72. Toyoda H, Nicklin MJH, Murray MG, Anderson CW, Dunn JJ, Studier FW and Wimmer E (1986) A second virus-encoded proteinase involved in proteolytic processing of poliovirus polyprotein. Cell 45: 761-770
73. Vance VB and Beachy RN (1984) Translation of soybean mosaic virus RNA in vitro; evidence of protein processing. Virology 132: 271-281
74. Van Wezenbeek P, Verver J, Harmsen J, Vos P and Van Kammen A (1983) Primary structure and gene organisation of the middle-component RNA of cowpea mosaic virus. EMBO J 2: 941-946
75. Verver J, Goldbach R, Garcia JA and Vos P (1987) In vitro expression of a full-length DNA copy of cowpea mosaic virus B-RNA: Identification of the B-RNA encoded 24-kilodalton protein as a viral protease. EMBO J 6: 549-554
76. Von der Helm K (1977) Cleavage of Rous Sarcoma viral polypeptide precursor into internal structural proteins in vitro involves viral protein p15. Proc Natl Acad Sci USA 74: 911-915
77. Von Heyne G (1983) Patterns of amino acids near signal-sequence cleavage sites. Eur J Biochem 133: 17-21
78. Von Heyne G (1986) A new method for predicting signal sequence cleavage sites. Nucleic Acids Res 14: 4683-4690
79. Vos P, Verver J, Jaegle M, Wellink J, Van Kammen A and Goldbach R (1987) Studies on the proteolytic processing of the cowpea mosaic virus polyproteins by in vitro expression of cDNA clones. Submitted
80. Vos P, Verver J, Van Wezenbeek P, Van Kammen A and Goldbach R (1984) Study of the genetic organisation of a plant viral RNA genome by in vitro expression of a full-length DNA copy. EMBO J 3: 3049-3053
81. Wellink J, Rezelman G, Goldbach R and Beyreuther K (1986) Determination of the proteolytic processing sites in the polyprotein encoded by the B-RNA of cowpea mosaic virus. J Virol 59: 50-58

82. Yeh SD and Consalves D (1985) Translation of papaya ringspot virus RNA in vitro: Detection of a possible polyprotein that is processed for capsid protein, cylindrical-inclusion protein and amorphous-inclusion protein. Virology 143: 260-271
83. Yoshinaka Y, Katoh I, Copeland TD and Oroszlan S (1985) Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. Proc Natl Acad Sci USA 82: 1618-1622.
84. Yoshinaka Y, Katoh I, Copeland TD and Oroszlan S (1985) Translation readthrough of an amber termination codon during synthesis of feline leukemia virus protease. J Virol 55: 870-873
85. Yoshinaka Y, Katoh I, Copeland TD, Smythers GW and Oroszlan S (1986) Bovine leukemia virus protease: purification, chemical analysis and in vitro processing of gag precursor poyproteins. J Virol 57: 826-832
86. Ypma-Wong MF and Semler BL (1987) In vitro molecular genetics as a tool for determining the differential cleavage specificities of the poliovirus 3C proteinase. Nucleic Acids Res 15: 2069-2088
87. Zagorski W, Morch MD and Haenni A (1983) Comparison of three different cell-free systems for turnip yellow mosaic virus RNA translation. Biochimie 55: 127-133
88. Ziegler V, Richards K, Guiley H, Jonard G and Putz C (1985) Cell-free translation of beet necrotic yellow vein virus: Readthrough of the coat protein cistron. J Gen Virol 66: 2079-2087

CHAPTER 3

DETERMINATION OF THE PROTEOLYTIC PROCESSING SITES IN THE POLYPROTEIN ENCODED BY THE BOTTOM-COMPONENT RNA OF COWPEA MOSAIC VIRUS

Joan Wellink, Geertje Rezelman, Rob Goldbach and Konrad Beyreuther[1]

Department of Molecular Biology, Agricultural University,
De Dreijen 11, 6703 BC Wageningen, The Netherlands and
[1]Institut für Genetik, Universität Köln,
D-5000 Cologne 41, FRG

Journal of Virology, 59, 50-58 (1986)

JOURNAL OF VIROLOGY, July 1986, p. 50–58
0022-538X/86/070050-09$02.00/0

Vol. 59, No. 1

Determination of the Proteolytic Processing Sites in the Polyprotein Encoded by the Bottom-Component RNA of Cowpea Mosaic Virus

JOAN WELLINK,[1] GEERTJE REZELMAN,[1] ROB GOLDBACH,[1]* AND KONRAD BEYREUTHER[2]

Department of Molecular Biology, Agricultural University, 6703 BC Wageningen, The Netherlands,[1] and Institut für Genetik, Universität Köln, D-5000 Cologne 41, Federal Republic of Germany[2]

Received 3 December 1985/Accepted 24 March 1986

The bottom-component RNA (B-RNA) of cowpea mosaic virus is expressed by the production of a ~200,000-dalton polyprotein (200K polyprotein), from which the functional proteins are formed by specific proteolytic cleavages. Partial amino-terminal sequences of the various B-RNA-encoded proteins have now been determined. Comparison of the information obtained with the B-RNA sequence allowed the localization of the coding regions for these proteins on B-RNA, the calculation of their precise molecular weights, and the determination of the cleavage sites at which they are released from the polyprotein precursor. Sequence analysis of the 32K protein, which is derived from the amino-terminal end of the 200K polyprotein, indicated that the AUG codon at nucleotide position 207 of the RNA sequence is the translation initiation codon. Sequence analysis of the 170K, 110K, 87K, 84K, 60K, and 58K proteins revealed the existence of three types of cleavage site in the 200K polyprotein: glutamine-serine (two sites), glutamine-methionine (one site), and glutamine-glycine (one site) amino acid pairs. The nature of these cleavage sites suggested that two different viral proteases are involved in the processing of the B-RNA-encoded polyprotein.

The genetic information of cowpea mosaic virus (CPMV) is organized in two separately encapsidated plus-strand RNA molecules which are characterized by a small protein, denoted VPg, covalently bound to their 5' ends and a poly(A) tail at their 3' ends (for a review, see references 4 and 14a). The expression of both RNAs involves the production of large polyproteins that are subsequently processed by proteolytic cleavage into smaller, functional proteins (7, 8, 11, 25). The nucleotide sequences of the smaller middle-component RNA (M-RNA) (3,481 nucleotides) and the larger bottom-component RNA (B-RNA) (5,889 nucleotides) reveal a single, long open reading frame in each of them, confirming polyprotein processing as their mode of expression (20, 30). M-RNA has been shown to possess two translation initiation sites and is translated in vitro into two overlapping primary translation products of ~105,000 and ~95,000 daltons which differ only in their amino-terminal sequences (Fig. 1; 7, 14, 22). Proteolytic cleavage of these proteins occurs at two different sites. The first cleavage is at a glutamine-methionine dipeptide sequence, resulting in products of 60,000, 58,000, and 48,000 daltons (7, 30; H. Franssen, Ph.D. thesis, Agricultural University, Wageningen, The Netherlands, 1984). Subsequent cleavage of the 60,000-dalton intermediate (60K intermediate) at a glutamine-glycine sequence releases the two capsid proteins VP37 (37,000 daltons) and VP23 (23,000 daltons) (30; Franssen, Ph.D. thesis). B-RNA produces a single 200K polyprotein from which, in addition to 170K, 110K, 84K, and 60K intermediates, five final cleavage products are derived in the order NH_2-32K-58K-4K (VPg)-24K-87K-COOH (Fig. 1; 11).

Information about the nature of the various cleavage sites in the B-RNA-encoded polyprotein (B-polyprotein) is still rather poor. Mapping of the amino-terminal end of VPg on the open reading frame of B-RNA demonstrated that this protein is released from its precursor at this terminus by cleavage at a glutamine-serine site (31). Inspection of the theoretical sequence of the 200K polyprotein has led to the suggestion that all cleavages may occur at glutamine-serine sites (31).

To understand in more detail the proteolytic processing pathway of the B-polyprotein, we have now determined partial amino-terminal sequences of its cleavage products by automated Edman degradation with radiolabeled protein samples (26). Comparison of the amino acid sequence data obtained with the nucleotide sequence of B-RNA allowed the determination of the cleavage sites in the B-polyprotein and predicted the molecular sizes of all cleavage products. The data presented in this paper demonstrate the existence of three different types of cleavage site and suggest that two different proteases are involved in the processing of the B-polyprotein.

MATERIALS AND METHODS

Virus and plants. CPMV (Sb isolate) was propagated in cowpea plants (*Vigna unguiculata* L., California blackeye) as described previously (18, 29). Purified B-components were obtained by repeated sucrose gradient centrifugation as described in detail before (12).

Infection of cowpea protoplasts and labeling of proteins. Mesophyll protoplasts were prepared from 10-day-old primary cowpea leaves and inoculated with purified CPMV B-components as previously described (15, 25). To label viral proteins, we incubated protoplast suspensions (5 ml, 5 × 10^5 cells per ml) in the presence of [^{35}S]methionine (1,000 Ci/mmol; New England Nuclear Corp., Boston, Mass.) or [^{3}H]leucine (150 Ci/mmol; Amersham Corp., Arlington Heights, Ill.). Radioactive amino acid was supplied in two portions ([^{35}S]methionine, 150 μCi; [^{3}H]leucine, 750 μCi) at 18 h and 25 h after inoculation. At 42 h after inoculation, cells were harvested by centrifugation (2 min, 600 × *g*) and disrupted by suspension in 0.5 ml of TKEDP buffer (50 mM Tris-acetate [pH 7.4], 10 mM potassium acetate, 1 mM EDTA, 10 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride). The lysate obtained was centrifuged for 30 min at

* Corresponding author.

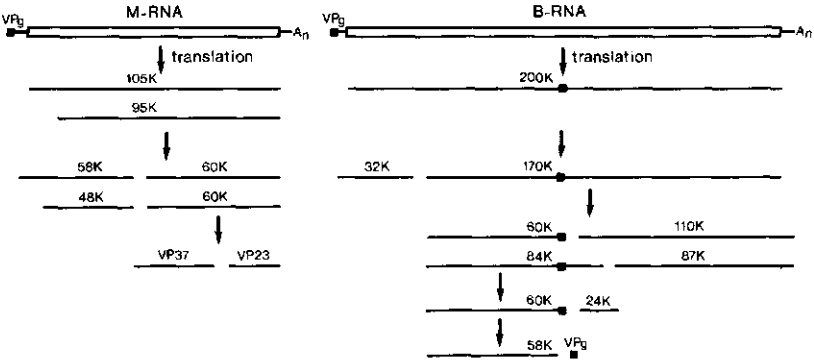


FIG. 1. Model of the proteolytic processing of the primary translation products of CPMV M-RNA and B-RNA. Open reading frames in both RNAs are indicated by double-lined bars, VPg is indicated by a black box, and all other proteins are indicated by single lines. For further details, see the text.

30,000 × *g* and 4°C, yielding the 30,000 × *g* supernatant, which served as the source of radiolabeled viral proteins.

Immunoprecipitation of viral proteins. As a first step in the purification of the 32K, 58K, 60K, 84K, and 170K proteins, these proteins were coprecipitated from the 30,000 × *g* supernatant of B-component-inoculated protoplasts with anti-32K serum under nondenaturing conditions (10). For this purpose, 140 μl of anti-32K serum was mixed with 500 μl of the 30,000 × *g* supernatant, and the mixture was diluted with TKE buffer (50 mM Tris-acetate [pH 7.4], 10 mM potassium acetate, 1 mM EDTA) to a final volume of 2,000 μl. After incubation for 4 h at 4°C, 560 μl of a 10% (wt/vol) suspension of *Staphylococcus aureus* cells (IgGsorb; Enzyme Center Inc., Boston, Mass.) in TKE buffer was added, and incubation was continued for 1 h at 4°C. *S. aureus* cells with bound immunocomplexes were precipitated by centrifugation through a sucrose cushion as previously described (7). The 32K, 58K, 60K, 84K, and 170K viral proteins present in the precipitate were separated by electrophoresis in a 7.5 or 10% sodium dodecyl sulfate–polyacrylamide gel as previously described (25). The supernatant left after removal of the immunoprecipitate was used for the isolation of the 87K protein, free of the 84K protein.

Isolation of viral proteins from polyacrylamide gels. Viral proteins separated in preparative polyacrylamide gels and localized by autoradiography were eluted from nonstained gel strips by using an ISCO 1759 sample concentrator (3) at 180 V for 6 h at 4°C. The sample compartment buffer contained 25 mM Tris, 192 mM glycine, 0.1% sodium dodecyl sulfate, and 50 μg of horse heart cytochrome *c* (Calbiochem-Behring, La Jolla, Calif.) per ml as the carrier protein. Eluates were concentrated by using a Centricon-30 microconcentrator (Amicon Corp., Danvers, Mass.) and brought to 10 mM Tris hydrochloride (pH 7.2)–0.02% sodium dodecyl sulfate.

Protein sequence analysis. Sequencing of radiolabeled proteins was performed by the method of Beyreuther et al. (2) with some modifications (31). Control experiments were performed to minimize the nonspecific release of radioactivity, especially at cycle 1 of Edman degradation. This was achieved with a single complete nondegradative washing cycle by omitting the delivery of the coupling agent phenylisothiocyanate. A spinning-cup sequencer (Beckman Instruments model 890) and a gas-liquid solid-phase sequencer (Applied Systems model 470A) were both used. The performance of the sequencer was monitored with 1% of the sample for each degradation cycle by high-performance liquid chromatography of the phenylthiohydantoin derivatives released from a second carrier protein. To do this, we used performic acid-oxidized lysozyme (1.45 mg for the spinning-cup sequencer and 0.145 mg for the gas-liquid solid-phase sequencer), since the carrier cytochrome *c* added prior to gel elution of the viral proteins is N terminally blocked. Polybrene (1.5 mg for both sequencers) was added as the nonprotein carrier. Identification of phenylthiohydantoin derivatives by high-performance liquid chromatography was done as described previously (1). The radioactivity from 90% of the sample released during each degradation cycle was dissolved in 0.1 ml of methanol and determined in 10 ml of Aquasol 2 (New England Nuclear Corp.) by using a Packard Tri-Carb liquid scintillation counter.

RESULTS

Preparation of radiolabeled viral proteins. In previous experiments we showed that upon inoculation of cowpea protoplasts with B-components, B-RNA is efficiently replicated in the absence of M-RNA, resulting in the overproduction of B-RNA-encoded proteins (12, 25). This finding was now used to obtain these viral proteins in quantities sufficient for radiochemical sequencing. Cowpea protoplasts were inoculated with purified B-components and labeled with [^{3}H]leucine or [^{35}S]methionine. A typical 30,000 × *g* supernatant fraction obtained from such protoplasts incubated with [^{35}S]methionine is shown in Fig. 2, lane B. While the 170K, 110K, 87K, 60K, 58K, and 32K proteins were directly detectable in such extracts, the 84K protein was mostly marked by host protein bands or present in amounts too low to allow its detection. This protein could, however, readily be detected in immunoprecipitates of 30,000 × *g* supernatant fractions obtained with antiserum raised against the 32K protein (Fig. 2, lane J). As a first step in the purification of the 170K, 84K, 60K, 58K, and 32K proteins, larger quantities of 30,000 × *g* supernatant fractions from [^{3}H]leucine- or [^{35}S]methionine-labeled cells were subjected to immunoprecipitation with anti-32K serum, and the precipitates obtained were electrophoresed in preparative polyacrylamide gels (data not shown). The viral proteins were

subsequently isolated from gel slices by electroelution. The 87K protein was simultaneously obtained from the supernatant fraction of such immunoprecipitation reactions. The 110K protein, finally, was directly excised and eluted from preparative gels of an unfractioned 30,000 × *g* supernatant fraction. To verify whether the viral proteins thus obtained were pure and intact, we analyzed aliquots by gel electrophoresis (Fig. 2, lanes C to I).

Amino acid sequence analysis of the 32K protein. Since translation studies have already shown that the 32K protein is derived from the amino-terminal part of the 200K polyprotein (11), sequencing of this protein would allow the precise determination of the translation initiation site of B-RNA. Radiochemical sequencing revealed that this protein contains leucine residues at positions 2, 12, 13, 16, 29, 30, and 33 (Fig. 3A) and a methionine residue at position 24 (Fig. 3B). Comparison of these results with the nucleotide sequence of B-RNA (20) determined a unique amino-terminal coding location for this protein starting at nucleotide position 210, i.e., one codon downstream of the AUG codon which heads the long open reading frame in B-RNA.

Thus, the amino-terminal sequence of the 32K protein demonstrated that this AUG codon at nucleotide position 207 represents the (unique) translation initiation site of B-RNA and that the first methionine residue of the translation product is posttranslationally removed. There seems to be one discrepancy between the partial amino acid sequence of the 32K protein and the nucleotide sequence of B-RNA. From the triplet sequence in the open reading frame in B-RNA (20) a valine residue is expected at amino acid position 33 instead of the leucine residue found (Fig. 3A). However, a single point mutation at nucleotide position 306 could have occurred, converting the valine triplet into a leucine triplet (GUC → CUC).

Amino acid sequence analysis of the 170K, 84K, 60K, and 58K proteins. Primary cleavage of the 200K polyprotein yielded a 32K protein and a 170K protein. For localization of the coding region for the 170K protein on B-RNA, purified samples of this protein, labeled with either [^{3}H]leucine or [^{35}S]methionine, were subjected to sequence analysis. The results showed that leucine is present at positions 6, 7, 10, and 22 (Fig. 4A) and that methionine is present at position 14 (Fig. 4B). This amino acid arrangement is consistent with a coding region for the 170K protein beginning at nucleotide position 1185 in the B-RNA sequence and reveals that the amino-terminal residue of this protein is serine (Fig. 4). Further inspection of the B-RNA sequence (20) revealed that this amino-terminal residue is preceded by a glutamine in the precursor protein. These data strongly suggest that the 170K protein is generated by proteolytic cleavage at a glutamine-serine dipeptide sequence. According to the scheme in Fig. 1, further proteolytic cleavage of the 170K protein would yield 84K, 60K, and 58K proteins, all sharing their amino-terminal sequences with the 170K protein. To verify this hypothesis, we carried out additional sequence analyses on these cleavage products, labeled at either their leucine or methionine residues. Leucine and methionine residues were indeed detected in all three proteins at the same locations as in the 170K protein (Fig. 4C to H). These results confirm the positions of these proteins in the processing map of Fig. 1 and, moreover, demonstrate that their amino termini are completely identical to that of the 170K protein and not subject to terminal trimming.

Amino acid sequence analysis of the 110K and 87K proteins. The 110K and 87K proteins are expected to be derived from the carboxy-terminal half of the 200K polyprotein, both

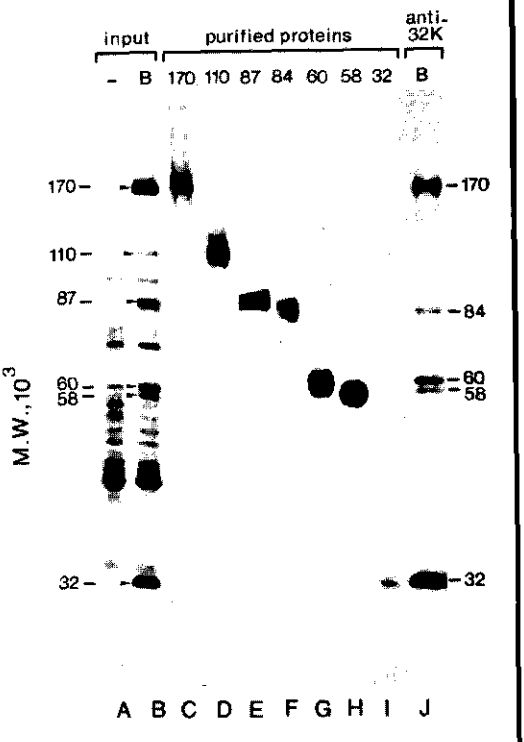


FIG. 2. Electrophoretic analysis of purified viral proteins used for sequence determinations. Viral proteins labeled in vivo with [^{35}S]methionine were purified from a 30,000 × *g* supernatant fraction of B-component-inoculated cowpea protoplasts (lane B). As a first step in the purification of the 170K, 84K, 60K, 58K, and 32K proteins, this supernatant fraction was subjected to immunoprecipitation with antibodies raised against the 32K protein (lane J). For further details, see the text. The proteins obtained were reelectrophoresed to test their quality (lanes C to I). Lane A contains the 30,000 × *g* supernatant fraction from [^{35}S]methionine-labeled noninoculated protoplasts. M.W., Molecular weight.

having different amino termini. For the 110K protein, samples labeled with [^{3}H]leucine or [^{35}S]methionine were subjected to sequence analysis. The results showed that in this protein, leucine is present at positions 3 and 18 (Fig. 5A) and methionine is present at positions 1, 11, and 30 (Fig. 5B and C). These data are consistent with a unique coding region for the 110K protein beginning at nucleotide position 3048 in the B-RNA sequence (20) and identify the amino-terminal residue of this protein as a methionine (Fig. 5). It should be noted that different preparations of this protein showed some heterogeneity, which could easily be explained by variable demethionylation in vivo. For instance, the sequence analysis shown in Fig. 5A reveals that some [^{3}H]leucine label was released in cycles 2 and 17. Equally, in the analysis shown in Fig. 5B, some [^{35}S]methionine label was released in cycle 10, whereas in a second [^{35}S]methionine-labeled protein sample (Fig. 5C), most of this label was found in cycle 10, with decreased label release in cycles 1 and 11. Inspection of the B-RNA sequence (20) demonstrated that the amino-terminal methionine of the 110K protein is preceded

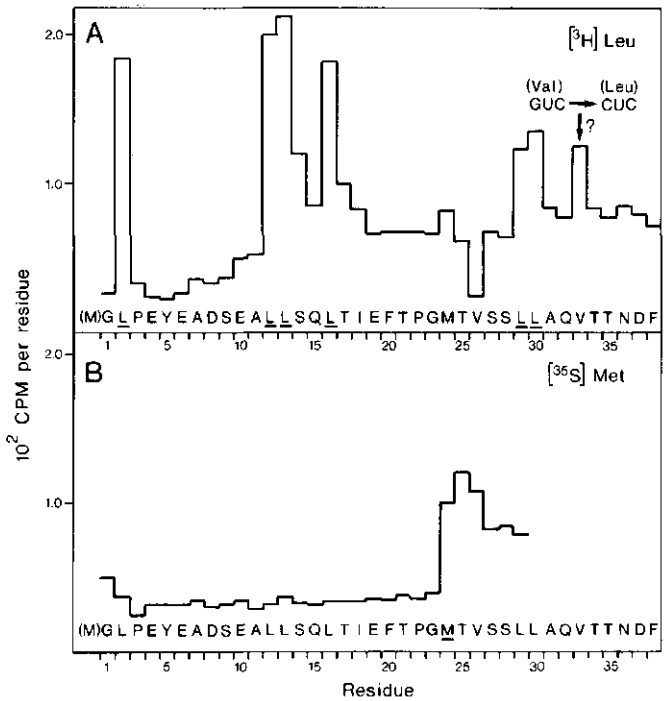


FIG. 3. Radiochemical amino-terminal sequence analysis of the 32K protein encoded by CPMV B-RNA. Purified 32K protein labeled with [^{3}H]leucine (A) or [^{35}S]methionine (B) was subjected to automated Edman degradation as described in detail in Materials and Methods. The radioactivity released during each Edman degradation cycle is plotted versus the cycle number. The predicted amino-terminal sequence of the 32K protein is given at the bottom of each panel, with the amino acids whose positions are considered proven by the presented data being underlined. (M) refers to the methionine residue encoded by the initiation codon, which is posttranslationally removed. The amounts of radioactivity applied to the sequencer were as follows: A, 39,000 cpm of [^{3}H]leucine; B, 30,000 cpm of [^{35}S]methionine.

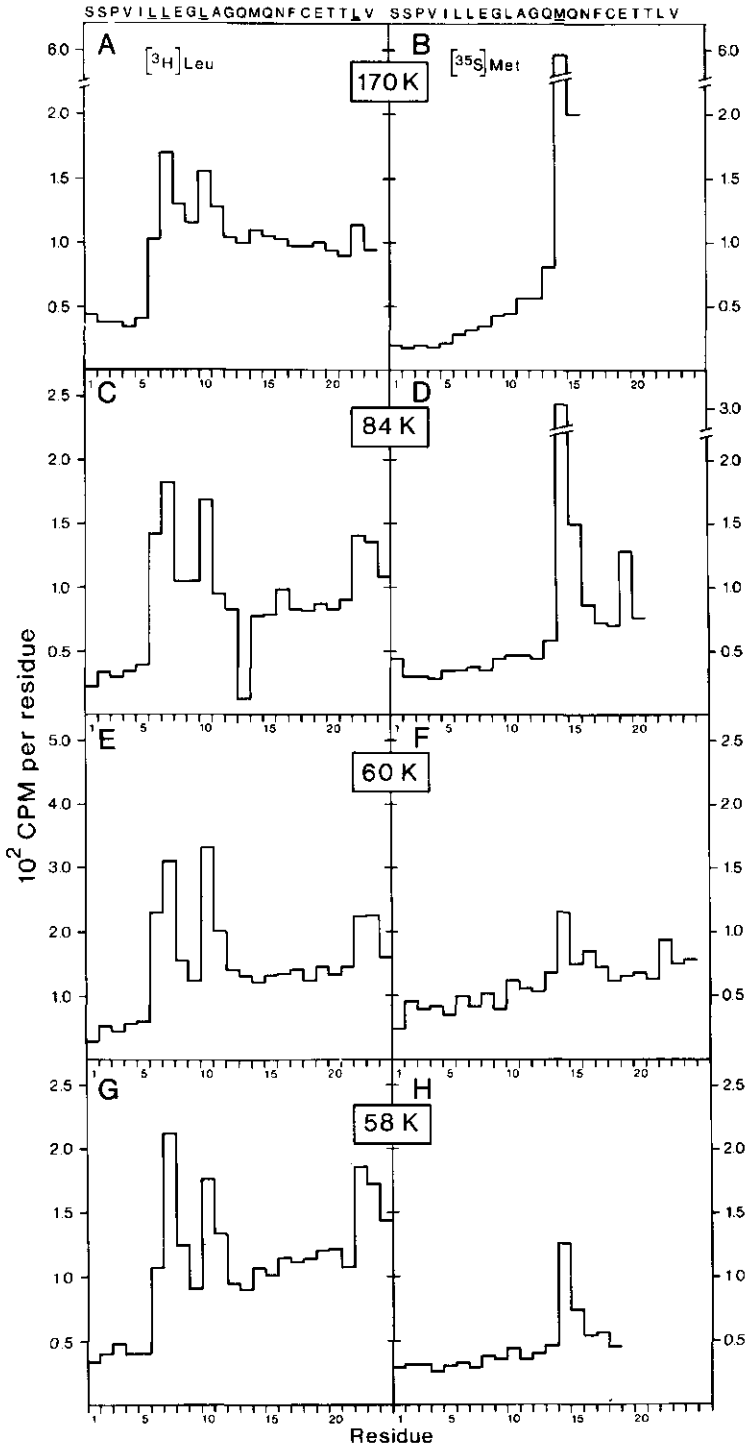
by a glutamine in the 200K polyprotein, indicating that the 110K protein is released from its direct 170K precursor by proteolytic cleavage at a glutamine-methionine dipeptide sequence. Sequence analysis of purified, radiolabeled 87K protein samples revealed the presence of a leucine at position 9 and a methionine at position 21 (Fig. 6). Combined with the proteolytic processing map of the 200K polyprotein (Fig. 1), these analyses allowed the localization of the coding region for the 87K protein beginning at nucleotide position 3672 in the B-RNA sequence. The mapping of the 87K protein on the open reading frame of B-RNA indicated further that the amino-terminal residue of this protein is a glycine and that it must be derived from longer precursors by cleavage at a glutamine-glycine dipeptide sequence.

DISCUSSION

Initiation site of B-RNA translation. A summary of the protein sequence analyses presented in this paper is schematically illustrated in Fig. 7. The exact location of the amino-terminus of each of the proteins in the B-polyprotein is indicated by different symbols, depending on the nature of the cleavage sites used for their release. Analysis of the 32K protein allowed the identification of the AUG codon at position 207, the first AUG codon in the open reading frame of B-RNA (20), as the initiation codon for the synthesis of the 200K polyprotein. This initiation codon is also the AUG codon nearest to the 5′ end of the RNA, while the surrounding nucleotide sequence CCAACAUGG conforms almost perfectly to the consensus sequence around eucaryotic initiation codons ($CC^{A}/_{G}CCAUGG$), as described by Kozak (19). The 32K protein lacks the amino-terminal methionine derived from this AUG codon, apparently because of posttranslational demethionylation, a widely occurring cellular process. As a consequence, the first residue of the 32K protein is glycine (Fig. 3).

Proteolytic cleavage sites in the B-polyprotein. Additionally, the protein sequence analyses revealed the presence of three different types of processing signal in the 200K precursor: a glutamine-serine pair (two sites), a glutamine-methionine pair (one site), and a glutamine-glycine pair (one site) (Fig. 7). While the last two types of cleavage site have also been found in the M-RNA polyproteins (30; Franssen, Ph.D. thesis), the glutamine-serine cleavage sites appear to be unique to the B-polyprotein.

Cleavage at the glutamine-glycine dipeptide sequence (Fig. 7) is consistent with computer comparisons which showed that the two regions within the 200K polyprotein of CPMV which are homologous to the polioviral protease and polymerase are separated at this glutamine-glycine pair (9). However, the glutamine-methionine cleavage site from which the 110K protein is released was totally unexpected.


SSPVILLEGLAGQMQNFCETTLV SSPVILLEGLAGQMQNFCETTLV
A
[3H]Leu
B
[35S]Met
170 K
C
D
84 K
E
F
60 K
G
H
58 K
10^2 CPM per residue
Residue

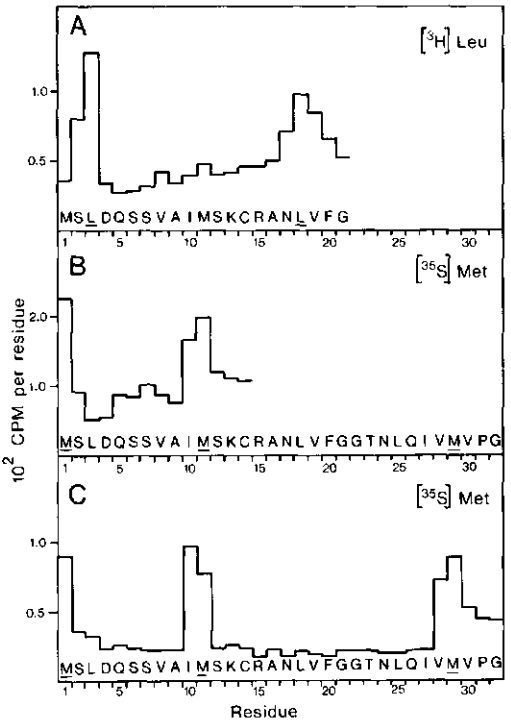


FIG. 5. Partial sequence analysis of the 110K protein encoded by CPMV B-RNA. Purified 110K protein labeled with [^{3}H]leucine (A) or [^{35}S]methionine (B and C) was subjected to automated Edman degradation. For further explanation of the graphs, see the legend to Fig. 3. The amounts of radioactivity applied to the sequencer were as follows: A, 74,000 cpm of [^{3}H]leucine; B, 31,000 cpm of [^{35}S]methionine; C, 41,000 cpm of [^{35}S]methionine.

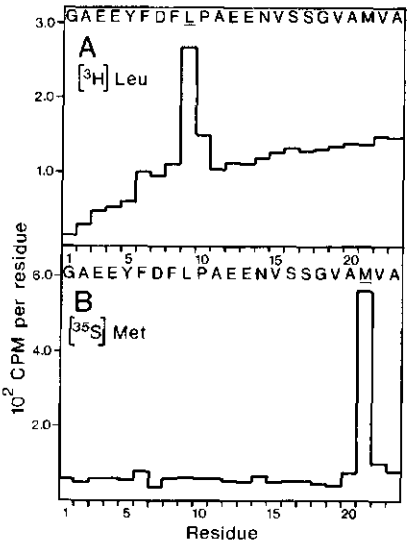


FIG. 6. Partial amino-terminal sequence analysis of the 87K protein encoded by CPMV B-RNA. Purified 87K protein labeled with [^{3}H]leucine (A) or [^{35}S]methionine (B) was subjected to automated Edman degradation. For further explanation of the graphs, see the legend to Fig. 3. The amounts of radioactivity applied to the sequencer were as follows: A, 140,000 cpm of [^{3}H]leucine; B, 40,000 cpm of [^{35}S]methionine.

As a consequence, VPg, the sequence of which preceeds that of the 110K protein within the 200K polyprotein (11, 13), is not released by cleavages at identical sites (glutamine-serine pairs), as proposed by us in a previous paper (31), but at two different sites: a glutamine-serine pair (at amino acid positions 919 and 920; Fig. 7) is used to generate its amino terminus, and a glutamine-methionine pair (at positions 947 and 948; Fig. 7) is now predicted to generate its carboxy terminus. VPg would thus have a length of 28 residues, instead of 33 residues, as previously proposed (31), which indeed is more consistent with the length of VPg found for the related picornaviruses (21 to 24 residues) (5, 6, 16, 17, 21, 24, 27, 28).

Comparison of the amino acid sequences surrounding all cleavage sites in the polyproteins encoded by both CPMV RNAs revealed that these sequences are rather variable, with similar amino acid residues at only two positions (Table 1). In five of six cleavage sites, alanine is found at position −4, while either alanine or proline is found at position −2. It is noteworthy that alanine at position −4 is also found in most (five of nine) glutamine-glycine cleavage sites in the polyprotein of poliovirus (17). Furthermore, inspection of the predicted amino acid sequence of the polyproteins encoded by B-RNA and M-RNA revealed the presence of 11 glutamine-glycine pairs, 7 glutamine-serine pairs, and 4 glutamine-methionine pairs which are not used for proteolytic processing. For only 3 of these 22 sites, alanine is found at position −4 (data not shown), further supporting the idea that the occurrence of such a residue at position −4 may represent an important signal. Besides, it is very likely that, more than simply the surrounding sequences, factors like secondary structure, such as β-turns, and hydrophilicity are involved in determining the sites used for proteolytic processing.

Molecular weights of the B-RNA-encoded proteins. Assuming that the cleavage products of the 200K polyprotein do not undergo any trimming at their carboxy termini, as is the case for picornaviral proteins, the mapping of the amino-terminal

FIG. 4. Partial amino acid sequence analysis of the CPMV B-RNA-encoded 170K, 84K, 60K, and 58K proteins. Purified proteins labeled *in vivo* with [^{3}H]leucine (*A, C, E,* and *G*) or [^{35}S]methionine (*B, D, F,* and *H*) were subjected to automated Edman degradation. The radioactivity released during each Edman degradation cycle is plotted versus the cycle number. The predicted amino-terminal sequence of the 170K, 84K, 60K, and 58K proteins is given at the top of panels A and B, with the amino acids whose positions are considered proven by the presented data being underlined. The amounts of radioactivity applied to the sequencer were as follows: A, 148,000 cpm of [^{3}H]leucine; B, 145,000 cpm of [^{35}S]methionine; C, 85,000 cpm of [^{3}H]leucine; D, 37,500 cpm of [^{35}S]methionine; E, 103,000 cpm of [^{3}H]leucine; F, 17,000 cpm of [^{35}S]methionine; G, 122,000 cpm of [^{3}H]leucine; H, 25,000 cpm of [^{35}S]methionine.

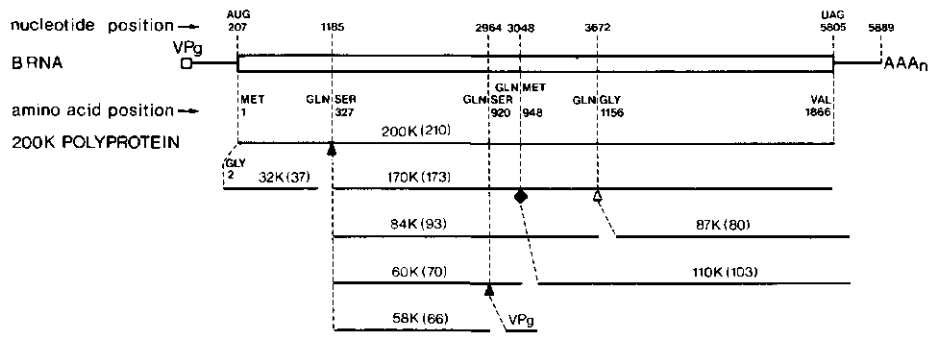


FIG. 7. Scheme of the proteolytic processing of the 200K polyprotein encoded by CPMV B-RNA. Nucleotide positions in the open reading frame in the RNA refer to the first nucleotide of the codon specifying the amino-terminal residue of the various proteins released from the 200K polyprotein. The proteins are indicated by single lines and are not drawn to scale. The numbers in parentheses represent the molecular weights (10^3) of the proteins determined from the data presented in this paper (see also Table 2). The UAG codon at position 5805 of the RNA is proposed to represent the translation termination codon. Cleavage sites are indicated as follows: ▲, Gln-Ser; △, Gln-Gly; ◆, Gln-Met.

ends of these products on the B-RNA sequence allows the calculation of their molecular weights (Table 2). For some proteins, the difference between their apparent molecular weights in polyacrylamide gels (on which the nomenclature of the CPMV proteins has been based) and their calculated molecular weights was relatively small, whereas for other proteins, this difference was more significant. The most striking discrepancy was found for the 87K and 84K proteins, which had calculated molecular weights of approximately 80,000 and 93,000, respectively (Table 2), suggesting that the 87K protein would run faster in sodium dodecyl sulfate-polyacrylamide gels than the 84K protein, which is not the case. Although additional cleavage or trimming at the carboxy-terminal end of the 84K protein cannot be ruled out, it is more likely that one (or both) of these proteins has a very deviant electrophoretic behavior in polyacrylamide gels. In this context, it is worthwhile mentioning that the proteins derived from the carboxy-terminal part of the 170K protein (110K and 87K) had calculated molecular weights approximately 7,000 lower than their apparent molecular weights; on the other hand, the proteins derived from the amino-terminal part of the 170K protein (84K, 60K, and 58K) all had calculated molecular weights approximately 9,000 higher than their apparent molecular weights. The similar deviations in apparent molecular weights found for these two sets of overlapping proteins indeed provide evidence that the reversed order in apparent sizes of the 87K and 84K proteins is due to specific conformations of these products.

Proteolytic activities involved in the cleavages. Another intriguing question to be resolved is which protease(s) is involved in the various cleavage steps during processing of the 200K polyprotein. It has now become clear that this polyprotein contains three different types of cleavage site. Although all cleavage sites share a glutamine residue in the first position, the residues in the second position have different properties. Serine and glycine both have small polar side chains, whereas methionine has a rather large, nonpolar side chain. It is therefore very unlikely that a single proteolytic enzyme is responsible for all these cleavages. Indeed, in vitro translation and proteolytic processing studies (8, 23) and in vivo inhibition studies with zinc ions (J. Wellink, unpublished results) suggest that the activity involved in the generation of the 87K and 84K protein pair (recognizing the glutamine-glycine cleavage site) and the activity involved in the generation of the 110K and 60K protein pair (recognizing the glutamine-methionine cleavage site) have different identities. In a previous report (10), we presented evidence that the 32K protein possesses the proteolytic activity involved in cleavage at the glutamine-methionine cleavage site in the M-RNA-encoded polyproteins. It is therefore very tempting to assume that this protein is also responsible for cleavage at an identical site in the B-RNA-encoded polyprotein. This remains to be

TABLE 1. Amino acids surrounding the proteolytic cleavage sites in the polyproteins encoded by CPMV B-RNA and M-RNA

Cleavage sites[a] in polyproteins encoded by:	Amino acid at position:							
	−4	−3	−2	−1	1	2	3	4
B-RNA								
32K-58K	Asp	Asn	Ala	Gln	Ser	Ser	Pro	Val
58K-VPg	Ala	Glu	Pro	Gln	Ser	Arg	Lys	Pro
VPg-24K	Ala	Asp	Ala	Gln	Met	Ser	Leu	Asp
24K-87K	Ala	Gln	Ala	Gln	Gly	Ala	Glu	Glu
M-RNA								
58K-VP37	Ala	Phe	Pro	Gln	Met	Glu	Gln	Asn
VP37-VP23	Ala	Ile	Ala	Gln	Gly	Pro	Val	Cys

[a] See Fig. 1.

TABLE 2. Molecular weights of the CPMV proteins

Proteins	Mol wt	
	Apparent (from electrophoresis)	Calculated (from mapping on RNAs)
B-RNA encoded	200,000	209,667
	170,000	173,065
	110,000	103,140[a]
	87,000	79,970
	84,000	93,113
	60,000	69,812
	58,000	66,292
	32,000	36,449[a]
	24,000	23,319
	4,000 (VPg)	3,538
M-RNA encoded	105,000	116,219
	95,000	102,329
	58,000	51,847
	48,000	37,957
	37,000 (VP37)	40,700
	23,000 (VP23)	23,708

[a] Corrected for removal of the first methionine residue.

verified. Furthermore, since the 24K protein derived from the amino-terminal part of the 110K protein (Fig. 1) exhibits sequence homology to protease 3C of picornaviruses (9), it is the logical candidate for a second protease. Although its activity still must be determined, we propose that this 24K protein is involved in cleavages at both glutamine-serine and glutamine-glycine dipeptide sequences.

ACKNOWLEDGMENTS

We thank K. Neifer and S. Pinto for excellent technical assistance in the sequence analysis, Piet Madern for photography and artwork, and Gré Heitkönig for typing the manuscript.

This work was supported by The Netherlands Foundation for Chemical Research (S.O.N.) with financial aid from The Netherlands Organisation for the Advancement of Pure Research (Z.W.O.) to R.G. and by grants from the Deutsche Forschungsgemeinschaft through SFB74, the Bundesministerium für Forschung und Technologie, and the Fonds der Chemischen Industrie to K.B.

LITERATURE CITED

1. **Beyreuther, K., B. Bieseler, J. Bovens, R. Dildrop, K. Neifer, K. Stüber, S. Zaiss, R. Ehring, and P. Zabel.** 1983. Microsequence analysis of peptides and proteins, p. 303–325. *In* H. Tscheche (ed.), Modern methods in protein chemistry. Walter de Gruyter & Co., Berlin.
2. **Beyreuther, K., R. Ehring, P. Overath, and J. K. Wight.** 1980. Microsequence analysis of lactose permease of *Escherichia coli*, p. 199–212. *In* C. Birr (ed.), Methods in peptide and protein sequence analysis. Elsevier Biomedical Press, Amsterdam.
3. **Bhown, A. S., J. E. Mole, F. Hunter, and C. J. Bennett.** 1980. High sensitivity sequence determination of proteins quantitatively recovered from sodium-dodecyl sulfate gels using an improved electrodialysis procedure. Anal. Biochem. **103:** 184–190.
4. **Bruening, G.** 1981. Biochemistry of plant viruses, p. 571–631. *In* A. Marcus (ed.), The biochemistry of plants, vol. 6. Proteins and nucleic acids. Academic Press, Inc., New York.
5. **Forss, S., and H. Schaller.** 1982. A tandem repeat gene in a picornavirus. Nucleic Acids Res. **10:**6441–6450.
6. **Forss, S., K. Strebel, E. Beck, and H. Schaller.** 1984. Nucleotide sequence and genome organisation of foot-and-mouth disease virus. Nucleic Acids Res. **12:**6587–6601.
7. **Franssen, H., R. Goldbach, M. Broekhuijsen, M. Moerman, and A. van Kammen.** 1982. Expression of middle-component RNA of cowpea mosaic virus: in vitro generation of a precursor to both capsid proteins by a bottom-component RNA-encoded protease from infected cells. J. Virol. **41:**8–17.
8. **Franssen, H., R. Goldbach, and A. van Kammen.** 1984. Translation of bottom component RNA of cowpea mosaic virus in reticulocyte lysate: faithful proteolytic processing of the primary translation product. Virus Res. **1:**39–49.
9. **Franssen, H., J. Leunissen, R. Goldbach, G. Lomonossoff, and D. Zimmern.** 1984. Homologous sequences in non-structural proteins from cowpea mosaic virus and picornaviruses. EMBO J. **3:**855–861.
10. **Franssen, H., M. Moerman, G. Rezelman, and R. Goldbach.** 1984. Evidence that the 32,000-dalton protein encoded by bottom-component RNA of cowpea mosaic virus is a proteolytic processing enzyme. J. Virol. **50:**183–190.
11. **Goldbach, R., and G. Rezelman.** 1983. Orientation of the cleavage map of the 200-kilodalton polypeptide encoded by the bottom-component RNA of cowpea mosaic virus. J. Virol. **46:**614–619.
12. **Goldbach, R., G. Rezelman, and A. van Kammen.** 1980. Independent replication and expression of B-component RNA of cowpea mosaic virus. Nature (London) **286:**297–300.
13. **Goldbach, R., G. Rezelman, P. Zabel, and A. van Kammen.** 1982. Expression of the bottom-component RNA of cowpea mosaic virus: evidence that the 60-kilodalton VPg precursor is cleaved into single VPg and a 58-kilodalton polypeptide. J. Virol. **42:**630–635.
14. **Goldbach, R. W., J. G. Schilthuis, and G. Rezelman.** 1981. Comparison of *in vivo* and *in vitro* translation of cowpea mosaic virus RNAs. Biochem. Biophys. Res. Commun. **99:**89–94.

14a. **Goldbach, R., and A. van Kammen.** 1986. Structure, replication and expression of the bipartite genome of cowpea mosaic virus, p. 83–120. *In* J. W. Davies (ed.), Molecular plant virology, vol. 2. CRC Press, Boca Raton, Fla.

15. **Hibi, T., G. Rezelman, and A. van Kammen.** 1975. Infection of cowpea mesophyll protoplasts with cowpea mosaic virus. Virology **64:**308–318.
16. **Kitamura, N., C. J. Adler, P. G. Rothberg, J. Martinko, S. G. Nathenson, and E. Wimmer.** 1980. The genome-linked protein of picornaviruses. VII. Genetic mapping of poliovirus VPg by protein and RNA sequence studies. Cell **21:**295–302.
17. **Kitamura, N., B. Semler, P. G. Rothberg, G. R. Larson, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanecak, J. J. Lee, S. van der Werf, C. W. Anderson, and E. Wimmer.** 1981. Primary structure, gene organisation and polypeptide expression of poliovirus RNA. Nature (London) **291:**547–553.
18. **Klootwijk, J., I. Klein, P. Zabel, and A. van Kammen.** 1977. Cowpea mosaic virus RNAs have neither m^7GpppN . . . nor mono-, di-, or triphosphates at their 5'-ends. Cell **11:**73–82.
19. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12:**857–872.
20. **Lomonossoff, G. P., and M. Shanks.** 1983. The nucleotide sequence of cowpea mosaic virus B-RNA. EMBO J. **2:**2253–2258.
21. **Palmenberg, A. C., E. M. Kirby, M. R. Janda, N. L. Drake, G. M. Duke, K. F. Potratz, and M. S. Collett.** 1984. The nucleotide and deduced amino acid sequences of the encephalomyocarditis viral polyprotein coding region. Nucleic Acids Res. **12:**2969–2983.
22. **Pelham, H. R. B.** 1979. Synthesis and proteolytic processing of cowpea mosaic virus proteins in reticulocyte lysates. Virology **96:**463–477.
23. **Peng, X. X., and D. S. Shih.** 1984. Proteolytic processing of the proteins translated from the bottom-component RNA of cowpea mosaic virus. J. Biol. Chem. **259:**3197–3201.
24. **Racaniello, V. R., and D. Baltimore.** 1981. Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. Proc. Natl. Acad. Sci. USA **78:**4887–4891.
25. **Rezelman, G., R. Goldbach, and A. van Kammen.** 1980. Expression of bottom-component RNA of cowpea mosaic virus in cowpea protoplasts. J. Virol. **36:**366–373.

26. **Semler, B. L., C. W. Anderson, N. Kitamura, P. G. Rothberg, W. L. Wishart, and E. Wimmer.** 1981. Poliovirus replication proteins: RNA sequence encoding P3-1b and the sites of proteolytic processing. Proc. Natl. Acad. Sci. USA **78:** 3464–3468.
27. **Skern, T., W. Sommergruber, D. Blaas, P. Gruendler, F. Fraundorfer, C. Pieler, I. Fogy, and E. Kuechler.** 1985. Human rhinovirus 2: complete nucleotide sequence and proteolytic processing signals in the capsid protein region. Nucleic Acids Res. **13:**2111–2126.
28. **Stanway, G., P. J. Hughes, R. C. Mountford, P. D. Minor, and J. W. Almond.** 1984. The complete nucleotide sequence of a common cold virus: human rhinovirus 14. Nucleic Acids Res. **12:**7859–7877.
29. **Van Kammen, A.** 1967. Purification and properties of the components of cowpea mosaic virus. Virology **31:**633–642.
30. **Van Wezenbeek, P., J. Verver, J. Harmsen, P. Vos, and A. van Kammen.** 1983. Primary structure and gene organisation of the middle-component RNA of cowpea mosaic virus. EMBO J. **2:**941–946.
31. **Zabel, P., M. Moerman, G. Lomonossoff, M. Shanks, and K. Beyreuther.** 1984. Cowpea mosaic virus VPg: sequencing of radiochemically modified protein allows mapping of the gene on B-RNA. EMBO J. **3:**1629–1634.

CHAPTER 4

DETECTION OF A NOVEL PROTEIN ENCODED BY THE BOTTOM-COMPONENT RNA OF COWPEA MOSAIC VIRUS, USING ANTIBODIES RAISED AGAINST A SYNTHETIC PEPTIDE

Joan Wellink, Martine Jaegle and Rob Goldbach

Department of Molecular Biology, Agricultural University, De Dreijen 11, 6703 BC Wageningen, The Netherlands

Journal of Virology, 61, 236-238 (1987)

JOURNAL OF VIROLOGY, Jan. 1987, p. 236–238
0022-538X/87/010236-03$02.00/0

Vol. 61, No. 1

Detection of a Novel Protein Encoded by the Bottom-Component RNA of Cowpea Mosaic Virus, Using Antibodies Raised against a Synthetic Peptide

JOAN WELLINK,* MARTINE JAEGLE, AND ROB GOLDBACH

Department of Molecular Biology, Agricultural University, De Dreyen 11, 6703 BC Wageningen, The Netherlands

Received 4 June 1986/Accepted 22 September 1986

A peptide was synthesized that corresponded to a sequence in the cowpea mosaic virus bottom-component RNA-encoded 200-kilodalton polyprotein showing homology to the picornaviral 3C proteases. By injecting a rabbit with this peptide, antibodies were obtained that allowed the detection of a novel viral protein derived from the 200-kilodalton polyprotein. This protein, which had a size of 24 kilodaltons was found in both infected cowpea leaves and cowpea protoplasts.

Cowpea mosaic virus (CPMV) is a positive-strand RNA virus with a bipartite genome. Both RNAs are polyadenylated and have a small protein (VPg) linked to their 5′ ends. The RNAs are translated into so-called polyproteins, which are posttranslationally cleaved to give the mature functional proteins (for a recent review, see reference 11). A detailed processing pathway of the 200-kilodalton (200K) polyprotein encoded by the bottom-component RNA (B-RNA) has been established (Fig. 1) by both in vitro translation studies (5, 10, 17) and in vivo observations (8, 9, 18). Except for the 24K polypeptide, all proteins included in Fig. 1 have been detected in B-component-inoculated cowpea protoplasts by radioactive labeling and immunological methods (9, 18, 19). The 24K protein was included in the model since it could easily be generated by further cleavage of the 84K or 110K proteins (18). After the discovery of sequence homology between nonstructural proteins from CPMV and picornaviruses (6), the sequence of the putative 24K protein was found to exhibit homology to the picornaviral 3C proteases. These proteases have been found in infected host cells as stable products (12, 13, 15, 16). Hence, the CPMV 24K protein could likewise be expected to occur as a stable protein, the existence of which had always escaped our attention. Moreover, the homology suggests that this protein may possess proteolytic activity.

To demonstrate the existence of such a 24K protein in vivo, we decided to synthesize a peptide corresponding to part of this protein. A region was chosen that was well-conserved among CPMV and the picornaviruses and represented the putative active site of the picornaviral proteases (1, 6). The peptide (Fig. 2) was synthesized by the solid-phase method of Barany and Merrifield (2) and after purification was characterized by amino acid analysis and high-performance liquid chromatography (data not shown). It was injected in free form into a New Zealand white rabbit at 14-day intervals (14a). Serum collected 8 days after the fifth injection showed a high titer against the peptide, as measured in an enzyme-linked immunosorbent assay (data not shown) and will be referred to in this paper as anti-24K serum. In a first experiment, cowpea protoplasts inoculated with only B components were analyzed for the presence of a 24K protein, since B-RNA-encoded proteins are known to be overproduced in these cells by the absence of middle-component RNA (M-RNA) (20). The protoplasts were labeled with [^{35}S]methionine from 16 to 40 h after inoculation (18), and a 30,000 × *g* supernatant fraction (S30[7]) was prepared from these cells which was used for immunological analyses. With the anti-24K serum, a protein of approximately 24,000 daltons was precipitated (Fig. 3, lane 4) together with the B-RNA-encoded 170K, 110K, and 84K proteins, which are all known to contain the 24K protein sequence (Fig. 1). The preimmune serum only showed a faint reaction with the 170K and 110K proteins, probably due to nonspecific aggregation of these proteins (Fig. 3, lane 1). The specificity of the immunoreactions was confirmed by adding excess peptide (100 μg) to the anti-24K serum prior to the immunoprecipitation. This pretreatment almost completely inhibited the precipitation of the 170K, 110K, 84K, and 24K proteins (Fig. 3, lane 5). To test whether the 24K protein was produced at the same time as the other viral proteins and whether it was a stable cleavage product, S30 fractions of protoplasts inoculated with B component or with complete virus (M + B) were prepared at different times after inoculation. With these fractions, a protein blot was prepared which was subsequently incubated with anti-24K serum and ^{125}I-protein A as described previously (21). It was found (Fig. 4) that the 24K protein appeared at the same time after inoculation as the other viral proteins and that its concen-

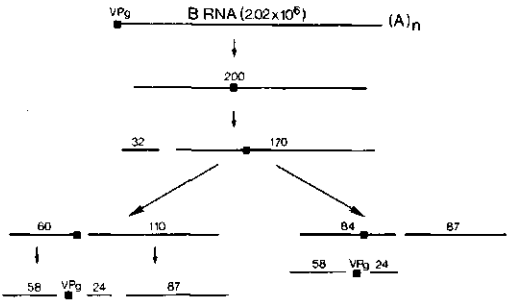


FIG. 1. Model for the proteolytic processing of the primary translation product of CPMV B-RNA. The RNA consists of 5,889 nucleotides, excluding the poly(A) tail, and has a molecular weight of 2.02 × 10⁶. VPg is indicated by a black box and all other polypeptides as lines.

* Corresponding author.

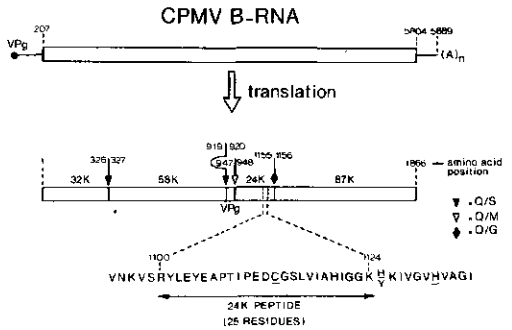


FIG. 2. Location and amino acid sequence of the synthetic peptide. The amino acids, as predicted from the B-RNA sequence (14), are denoted by the single-letter code. Amino acids in this sequence corresponding to important constituents of the active site of the picornaviral 3C proteases (1) are underlined. The open reading frame of the B-RNA and the 200K (K) polyprotein are both indicated by white bands. The positions of the amino acids surrounding the cleavage sites are indicated.

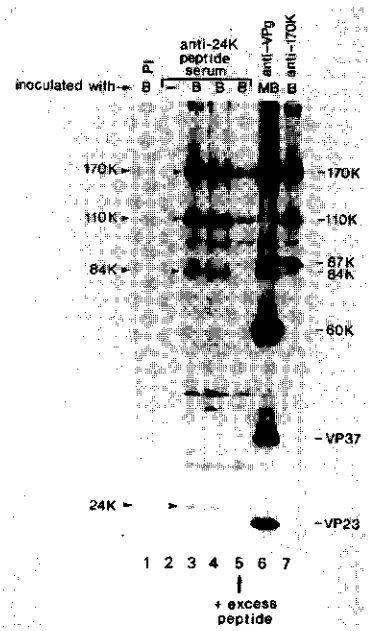


FIG. 3. Detection of the viral 24K protein in CPMV-infected cowpea protoplasts. The S30 supernatant fractions were prepared from protoplasts inoculated with B components (B) or complete virus (B + M) or uninoculated (7). Immunoprecipitations of these fractions were carried out with preimmune serum (PI) (lane 1), anti-24K serum (lanes 2 to 5), anti-VPg serum (lane 6), or anti-170K serum (lane 7) as described previously (4, 7). The immunoprecipitates were analyzed on a 12.5% polyacrylamide gel. For the immunoprecipitation shown in lane 5, 5 μl of anti-24K serum was pretreated with 100 μg of peptide for 3 h at 4°C in phosphate-buffered saline–TDS (4). Antiserum against VPg was prepared as described previously (21). This serum also reacted with both capsid proteins, which contaminated the VPg preparation used for immunization of rabbits (21). Antiserum against the 170K protein was prepared as described previously (7) and showed affinity for the 170K, 110K, and 87K proteins. Sizes are indicated in kilodaltons (K).

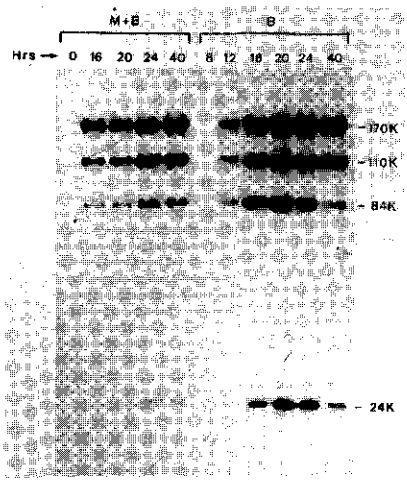


FIG. 4. Time course experiment with cowpea protoplasts inoculated with CPMV components. S30 fractions of cowpea protoplasts inoculated with complete virus (M + B) or with B components alone (B) were prepared at different times after inoculation as indicated at the top. After electrophoresis in a polyacrylamide gel, the fractions were blotted onto nitrocellulose, and the resulting blot was subsequently incubated with anti-24K serum and ^{125}I-protein A. Sizes are indicated in kilodaltons (K).

tration remained low. The stability of the 24K protein was approximately similar to that of the 84K protein, and it did not accumulate with time.

To test whether this 24K protein was also detectable in intact infected cowpea leaves, a protein blot was made containing cytoplasmic and membrane fractions from infected and noninfected leaves prepared as described in detail previously (3). Treatment of this blot with anti-24K serum and ^{125}I-protein A revealed that the 24K protein was detectably synthesized in cowpea leaves, being present mainly in the cytoplasmic fraction (Fig. 5, lane 5). A competition experiment with excess peptide again demonstrated the specificity of the reaction of the 24K, 84K, 110K, and 170K proteins with the anti-24K serum (data not shown).

Clearly, by using antibodies raised against a synthetic peptide, we were able to detect a novel viral protein in CPMV-infected protoplasts and leaves. This protein, which has a size of 24,000 daltons, has not been detected before and formed a "gap" in our processing model of the 200K polyprotein (Fig. 1). This protein is of particular interest since it shows sequence homology to the picornaviral 3C proteases (1, 6) and therefore probably also represents a protease. Indeed, in vitro translation experiments in which transcripts of modified cDNA clones of B-RNA were used, indicate that the 24K protein is involved in the cleavage of at least one glutamine-serine and two glutamine-glycine proteolytic cleavage sites in the CPMV polyproteins (J. Verver, R. Goldbach, J. A. Garcia and P. Vos, manuscript in preparation). It will be of interest to test whether the 24K

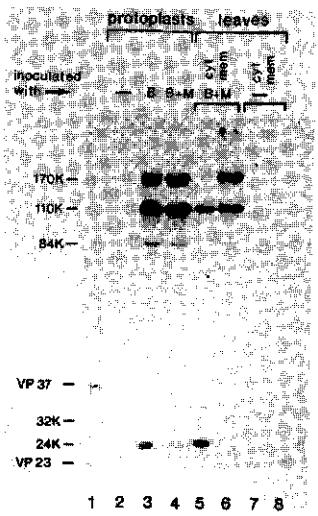


FIG. 5. Detection of the 24K protein in CPMV-infected cowpea leaves. Cytoplasmic (cyt) and membrane (mem) fractions from cowpea leaves were isolated as described before (3). These fractions, together with S30 fractions from cowpea protoplasts, were run on a polyacrylamide gel, blotted onto nitrocellulose, and probed with anti-24K serum (see the legend to Fig. 4). Lanes 2, 3, and 4 contain S30 fractions from uninoculated, B-, and B+M-inoculated protoplasts, respectively. Lanes 5 and 6 contain cytoplasmic and membrane fractions from infected leaves; lanes 7 and 8 contain similar fractions from uninfected leaves. Lane 1 contains an S30 fraction of [^{35}S]methionine-labeled CPMV-infected protoplasts. Sizes are indicated in kilodaltons (K).

protein in its free form has enzymatic activity or whether only precursor proteins containing its sequence (i.e., the 200K, 170K, 110K, and 84K proteins, see Fig. 1) are involved in proteolytic cleavage reactions.

We thank Rijndert de Fluiter and Piet Madern for photography and artwork and Gré Heitkönig for typing the manuscript.

This work was supported by the Netherlands Foundation for Chemical Research with financial aid from the Netherlands Organisation for the Advancement of Pure Research to R.G. M.J. was the recipient of a fellowship from the Commission of the European Communities in the molecular biology and radiobiology program.

LITERATURE CITED

1. **Argos, P., G. Kamer, M. J. H. Nicklin, and E. Wimmer.** 1984. Similarity in gene organisation and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. Nucleic Acids Res. **12**:7251–7267.
2. **Barany, G., and R. B. Merrifield.** 1980. Solid-phase peptide synthesis, p. 1–284. *In* E. Gross and J. Merenhofer (ed.), The peptides, vol. 2. Academic Press, Inc., New York.
3. **Dorssers, L., S. van der Krol, J. van der Meer, A. van Kammen, and P. Zabel.** 1984. Purification of cowpea mosaic virus RNA replication complex: identification of a virus-encoded 110,000-dalton polypeptide responsible for RNA chain elongation. Proc. Natl. Acad. Sci. USA **81**:1951–1955.
4. **Franssen, H., R. Goldbach, M. Broekhuysen, M. Moerman, and A. van Kammen.** 1982. Expression of middle-component RNA of cowpea mosaic virus: in vitro generation of a precursor to both capsid proteins by a bottom-component RNA-encoded protease from infected cells. J. Virol. **41**:8–17.
5. **Franssen, H., R. Goldbach, and A. van Kammen.** 1984. Translation of bottom-component RNA of cowpea mosaic virus in reticulocyte lysate: faithful proteolytic processing of the primary translation product. Virus Res. **1**:39–49.
6. **Franssen, H., J. Leunissen, R. Goldbach, G. Lomonossof, and D. Zimmern.** 1984. Homologous sequences in non-structural proteins from cowpea mosaic virus and picornaviruses. EMBO J. **3**:855–861.
7. **Franssen, H., M. Moerman, G. Rezelman, and R. Goldbach.** 1984. Evidence that the 32,000-dalton protein encoded by bottom-component RNA of cowpea mosaic virus is a proteolytic processing enzyme. J. Virol. **50**:183–190.
8. **Goldbach, R., and G. Rezelman.** 1983. Orientation of the cleavage map of the 200-kilodalton polypeptide encoded by the bottom-component RNA of cowpea mosaic virus. J. Virol. **46**:614–619.
9. **Goldbach, R., G. Rezelman, P. Zabel, and A. van Kammen.** 1982. Expression of the bottom-component RNA of cowpea mosaic virus: evidence that the 60-kilodalton VPg-precursor is cleaved into single VPg and a 58-kilodalton polypeptide. J. Virol. **42**:630–635.
10. **Goldbach, R. W., J. G. Schilthuis, and G. Rezelman.** 1981. Comparison of *in vivo* and *in vitro* translation of cowpea mosaic virus RNAs. Biochem. Biophys. Res. Commun. **99**:89–94.
11. **Goldbach, R., and A. van Kammen.** 1985. *Structure, replication* and expression of the bipartite genome of cowpea mosaic virus, p. 83–120. *In* J. W. Daires (ed.), Molecular plant virology, vol. 2. CRC Press, Boca Raton, Fla.
12. **Gorbalenya, A., Y. Svitkin, Y. Kazachkov, and V. Agol.** 1979. Encephalomyocarditis virus-specific polypeptide p22 is involved in the processing of the viral precursor polypeptides. FEBS Lett. **108**:1–5.
13. **Hanecak, R., B. L. Semler, C. W. Anderson, and E. Wimmer.** 1982. Proteolytic processing of poliovirus polypeptides: antibodies to polypeptide P3-7C inhibit cleavage at glutamine-glycine pairs. Proc. Natl. Acad. Sci. USA **79**:3973–3977.
14. **Lomonossoff, G. P., and M. Shanks.** 1983. The nucleotide sequence of cowpea mosaic virus B RNA. EMBO J. **2**:2253–2258.

14a. **Muller, S., S. Plaue, M. Couppez, and M. H. V. van Regenmortel.** 1986. Comparison of different methods for localizing antigenic regions in histone H2A. Molec. Immun. **23**:593–601.

15. **Palmenberg, A. C., M. A. Pallansch, and R. R. Rueckert.** 1979. Protease required for processing picornavirus coat protein resides in the viral replicase gene. J. Virol. **32**:770–778.
16. **Palmenberg, A. C., and R. R. Rueckert.** 1982. Evidence for intramolecular self-cleavage of picornavirus replicase precursors. J. Virol. **41**:244–249.
17. **Pelham, H. R. B.** 1979. Synthesis and proteolytic processing of cowpea mosaic virus proteins in reticulocyte lysates. Virology **96**:463–477.
18. **Rezelman, G., R. Goldbach, and A. van Kammen.** 1980. Expression of bottom-component RNA of cowpea mosaic virus in cowpea protoplasts. J. Virol. **36**:366–373.
19. **Rottier, P. J. M., G. Rezelman, and A. van Kammen.** 1980. Protein synthesis in cowpea mosaic virus-infected cowpea protoplasts: further characterization of virus-related protein synthesis. J. Gen. Virol. **51**:373–383.
20. **Wellink, J., G. Rezelman, R. Goldbach, and K. Beyreuther.** 1986. Determination of the proteolytic processing sites in the *polyprotein encoded by the bottom-component RNA of cowpea* mosaic virus. J. Virol. **59**:50–58.
21. **Zabel, P., M. Moerman, F. van Straaten, R. Goldbach, and A. van Kammen.** 1982. Antibodies against the genome-linked protein VPg of cowpea mosaic virus recognize a 60,000-dalton precursor polypeptide. J. Virol. **41**:1083–1088.

CHAPTER 5

PROTEOLYTIC ACTIVITY OF THE COWPEA MOSAIC VIRUS ENCODED 24K PROTEIN SYNTHESIZED IN ESCHERICHIA COLI

Juan A. Garcia, Louis Schrijvers, Agnes Tan, Pieter Vos, Joan Wellink and Rob Goldbach[1]

Department of Molecular Biology and [1]Virology, Agricultural University, De Dreijen 11, 6703 BC Wageningen, The Netherlands

Virology, 159, 67-75 (1987)

VIROLOGY 159, 67–75 (1987)

Proteolytic Activity of the Cowpea Mosaic Virus Encoded 24K Protein Synthesized in *Escherichia coli*

JUAN A. GARCIA,[1] LOUIS SCHRIJVERS, AGNES TAN, PIETER VOS, JOAN WELLINK,[2] AND ROB GOLDBACH*

*Departments of Molecular Biology and *Virology, Agricultural University, De Dreijen 11, 6703 BC Wageningen, The Netherlands*

Received November 19, 1986; accepted February 10, 1987

The function of the 24-kilodalton (24K) protein encoded by cowpea mosaic virus (CPMV) has been studied by constructing a bacterial expression plasmid that contained a cloned chimeric segment consisting of partial DNA copies of CPMV M-RNA (including sequences coding for both capsid proteins) and B-RNA (including sequences coding for the 24K protein). Viral sequences were transcribed from the phage T7 promoter ϕ10 of plasmid pT7-6 using T7-RNA polymerase expressed from plasmid pGP1-2 present in the same cells. Upon inducing the synthesis of T7-RNA polymerase several new polypeptides that contained CPMV-specific sequences were expressed, as demonstrated by immunoprecipitation and immunoblotting. Furthermore a proteolytic activity was detected in induced cells which cleaved the viral protein sequences specifically at two glutamine–glycine sites. One of the cleavage products represented capsid protein VP23. The proteolytic activity was absent when an 87-bp deletion was introduced in the coding region for the 24K protein, indicating that this protein represented the protease involved in the proteolytic processing at those specific sites.

INTRODUCTION

Cowpea mosaic virus (CPMV), the type member of the comoviruses, has a bipartite genome consisting of two separately encapsidated plus-stranded RNA molecules which possess a genome-linked protein (VPg) at their 5′-terminus and are polyadenylated (El Manna and Bruening, 1973; Daubert *et al.*, 1978; Stanley *et al.*, 1978). Both RNAs are translated into large polypeptides from which the functional proteins are derived by proteolytic cleavages (for a recent review, see Goldbach and Van Kammen, 1985). The larger of the two genomic RNAs, the B-RNA (5889 nucleotides; Lomonossoff and Shanks, 1983) is translated into a 200-kilodalton (200K) polyprotein which is first cleaved to give 32K and 170K polypeptides (Pelham, 1979; Goldbach *et al.*, 1981). The 170K polypeptide is then further cleaved following two different pathways to give four final cleavage products in the order NH_2-58K-VPg-24K-87K-COOH (see Fig. 1 and Rezelman *et al.*, 1980; Goldbach *et al.*, 1982; Wellink *et al.*, 1987). The smaller of the two RNAs, the M-RNA (3481 nucleotides; Van Wezenbeek *et al.*, 1983) is translated *in vitro* into two polypeptides of 95K and 105K with overlapping amino acid sequences (Pelham, 1979; Franssen *et al.*, 1982). These polypeptides are proteolytically cleaved at the same position to give two polypeptides of 58K and 48K, derived from the amino-terminal parts of the polyproteins, together with a 60K polypeptide which represents the precursor to both capsid proteins VP37 and VP23 (Fig. 1).

The proteolytic cleavages by which the M- and B-RNA-encoded polyproteins are processed occur between three different dipeptide sequences, i.e., between glutamine–methionine, glutamine–serine, and glutamine–glycine residues (Fig. 1 and Wellink *et al.*, 1986). The 32K protein, the amino-terminal cleavage product from the 200K polyprotein encoded by the B-RNA, has been proposed by Franssen *et al.* (1984b) to represent the protease involved in the cleavage at the glutamine–methionine site in the M-RNA-encoded 105K and 95K polyproteins (Fig. 1). Recently, evidence has been obtained that the 24K protein, also encoded by the B-RNA, represents a second protease, responsible for the primary cleavage (at a glutamine–serine site) in the 200K polyprotein which gives rise to the 32K and 170K proteins (Verver *et al.*, 1987). This conclusion is supported by the observation that the 24K protein shows amino acid sequence homology to the picornaviral 3C proteins (Franssen *et al.*, 1984a) involved in the proteolytic processing of the polyproteins encoded by these viruses.

In this report we describe the expression in *Escherichia coli* of a chimeric cDNA segment consisting of both CPMV M- and B-RNA-specific sequences, by using the phage T7 promoter-RNA polymerase binary system described by Tabor and Richardson (1985). Following this approach we have been able to show that the CPMV B-RNA-encoded 24K protein can be

[1] Present address: Centro de Biologia Molecular (CSIC-UAM), Universidad Autonoma, Canto Blanco, 28049 Madrid, Spain.

[2] To whom requests for reprints should be addressed.

0042-6822/87 $3.00

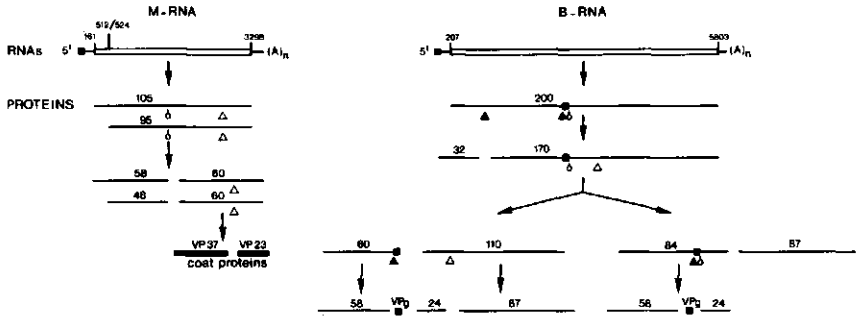


FIG. 1. Model of the expression of CPMV M- and B-RNA. The open reading frames in M- and B-RNA are indicated with open bars. VPg is indicated by a black box, other proteins by single lines. Cleavage sites are indicated as follows: ⚷, Gln–Met, △, Gln–Gly, and ▲, Gln–Ser.

produced in *E. coli* as an active protease. The experiments presented show that this protease is responsible for the proteolytic cleavages at the two glutamine–glycine sites present in the CPMV polyproteins. One of the cleavage products released by the protease represents the capsid protein VP23.

MATERIALS AND METHODS

Bacterial strains and plasmids

E. coli strain JM109 (Yanish-Perron *et al.*, 1985) was used during the construction of the expression plasmids. *E. coli* strain K38 (Russel and Model, 1984) was used for all expression experiments. Plasmid pT7-6, related to pT7-1 (Tabor and Richardson, 1985), was a kind gift of S. Tabor. Plasmid pGP1-2 has been described by Tabor and Richardson (1985), plasmids pSPMB401 and pSPMB373 by Vos *et al.* (manuscript in preparation), plasmid pINIII^{103}ompA2 by Ghrayeb *et al.* (1984), and pUC8 by Vieira and Messing (1982). The construction of the plasmids pTMBA1 and pTMBA1d is described under Results. All DNA manipulations were essentially as described by Maniatis *et al.* (1982).

Labeling of proteins

E. coli K38 cells, harboring the different expression plasmids, were grown at 30° in Luria broth containing 50 μg/ml ampicillin and kanamycin to A_{590} = 0.5. A sample of 0.2 ml was collected by centrifugation and the cell pellet was washed once with 1 ml of M9 medium and resuspended in 1 ml of the same medium supplemented with 20 μg/ml thiamine and 0.01% amino acid mixture (minus cysteine and methionine). After a 60-min incubation at 30° the temperature was shifted to 43° for 15 min (at this temperature the cI857 repressor, which controls the expression of the T7 RNA polymerase, is inactivated). Rifampicin (Sigma, St. Louis) was added to a final concentration of 400 μg/ml and the incubation was continued for an additional 10 min. The temperature was then lowered to 30° and 20 min later 10 μCi of [^{35}S]methionine (1050 Ci/mmol, NEN, Boston) was added. After 5 min of labeling the cells were collected by centrifugation, resuspended in 120 μl of sample buffer (10 m*M* Tris–HCl, pH 8.0, 1 m*M* EDTA, 10% glycerol, 2% SDS, 5% β-mercaptoethanol and 0.001% bromphenol blue), and boiled for 2 min. Aliquots were either directly analyzed by SDS–polyacrylamide gel electrophoresis or first subjected to immunoprecipitation.

Pulse–chase labeling of proteins

For pulse–chase analysis *E. coli* cells were treated as described above, except that cells were pulse labeled with [^{35}S]methionine for only 1 min, followed by a chase with excess nonlabeled methionine (1 mg/ml). Samples were collected at different times, adjusted to 20 m*M* sodium phosphate, pH 7.0, 1% formaldehyde, and 1 mg/ml methionine, and placed on ice. The preparation of bacterial extracts suitable for immunological analysis was as described above.

Accumulation of proteins in *E. coli*

Cells were grown in enriched medium (2% tryptone, 1% yeast extract, 0.5% NaCl, 0.2% glycerol, 50 m*M* potassium phosphate, pH 7.2, 50 μg/ml ampicillin and kanamycin) at 30°. At an A_{590} of 0.6, the temperature was raised to 42° for 25 min. After this period rifampicin was added to a final concentration of 100 μg/ml and the temperature was shifted to 37°. Cells were harvested at desired times and prepared for analysis on SDS–polyacrylamide gels as described above.

Immunological methods

Cell pellets were prepared for immunological analysis by boiling in 120 μl of sample buffer. Aliquots of 30 μl were diluted to 500 μl in TNETDS (50 m*M* Tris–HCl, pH 7.5, 150 m*M* NaCl, 5 m*M* EDTA, 1% Triton X-100, 0.5% DOC, 0.1% SDS) and immunoprecipitations were carried out on these samples as described previously (Franssen *et al.*, 1984b). Precipitates obtained were analyzed on SDS–polyacrylamide gels. Western blots of bacterial extracts, electrophoresed in 12.5% polyacrylamide gels, were prepared and subsequently incubated with antiserum and ^{125}I-protein A as has been described before (Zabel *et al.*, 1982).

RESULTS

Construction of plasmids pTMBA1 and pTMBA1d

The construction of the plasmids pTMBA1 and pTMBA1d is schematically shown in Fig. 2. A fragment containing both an M-RNA-specific cDNA sequence (from nucleotide 1446 to nucleotide 3496 of M-RNA) and a B-RNA-specific cDNA sequence (from nucleotide 2820 to nucleotide 3865 of B-RNA) was obtained by digestion of plasmid pSPMB401 (Fig. 2 and Vos *et al.*, manuscript in preparation) with *Bam*HI (partially) and *Eco*RI. After the sticky ends were filled in using Klenow fragment of *E. coli* DNA polymerase I, this 3-kb fragment was inserted in the *Sma*I site of pUC8, resulting in pU410. This plasmid was digested with *Bam*HI (partially) and *Eco*RI and the 3-kb fragment thus released inserted in pINIII^{113}ompA2 (Ghrayeb *et al.*, 1984), previously linearized with *Bam*HI and *Eco*RI. Thus, plasmid pompA2MB349 was obtained, in which the chimeric virial cDNA segment, fused to the sequence coding for the signal peptide of the *E. coli* OmpA protein, was under the control of Lpp/Lac promoters (Grahyeb *et al.*, 1984). As synthesis of CPMV-induced polypeptides could not be detected in *E. coli* harboring this plasmid either by labeling with [^{35}S]methionine or Western blotting (results not shown), we decided to make use of the T7-RNA polymerase/promoter system (Tabor and Richardson, 1985) which combines high level expression from the T7 ϕ10 promoter with very low background synthesis of bacterial proteins. Hence, plasmid pompA2MB349 was digested with *Xba*I, which cleaved between the Lpp/Lac promoters and the OmpA signal sequence (Ghrayeb *et al.*, 1984), and with *Bam*HI (partial), and the 3.1-kb fragment containing the chimeric cDNA sequence was inserted in the polylinker region of plasmid pT7-6 (S. Tabor, unpublished). The resulting plasmid was called pTMBA1. Plasmid pTMBA1d was constructed by exchanging the 740-bp *Kpn*I–*Bam*HI fragment (from nucleotide 3138 to 3857 in the B-RNA sequence) of pTMBA1 by the equivalent *Kpn*I–*Bam*HI

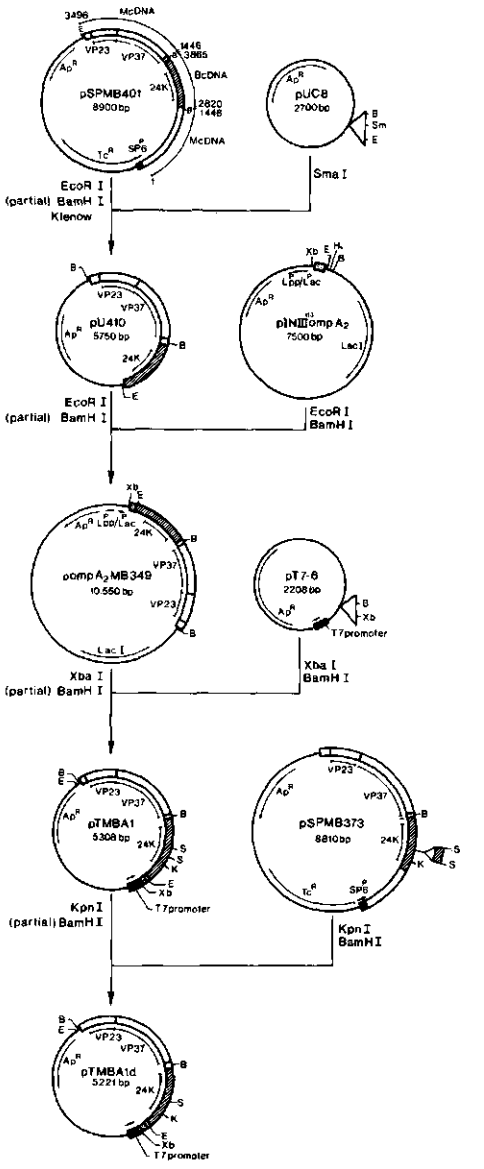


FIG. 2. Construction of the expression plasmids pTMBA1 and pTMBA1d. For details of the construction see the text. Complementary DNA sequences derived from viral M- and B-RNA are indicated by white bars and cDNA sequences derived from B-RNA by hatched bars. The 24K protein is encoded by the B-RNA from nucleotide 3048 to nucleotide 3672. The SP6 promoter and the T7 ϕ10 promoter are indicated with black boxes and the OmpA signal sequence with a dotted box. Next to plasmid pSPBM373 the *Sau*3A fragment, deleted in this plasmid, is shown. Only relevant restriction sites are shown: B = *Bam*HI, E = *Eco*RI, H = *Hin*dIII, K = *Kpn*I, S = *Sau*3A, and Xb = *Xba*I.

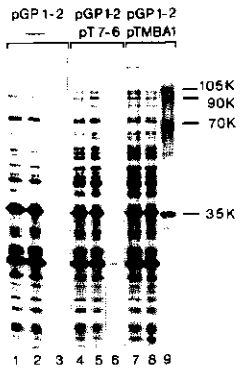


FIG. 3. Expression of plasmid pTMBA1 in *E. coli*. Bacteria harboring different plasmid combinations were either not induced (lanes 1, 4, 7) or induced in the absence (lanes 2, 5, 8) or presence (lanes 3, 6, 9) of rifampicin and labeled with [^{35}S]methionine for 5 min. Bacterial pellets were boiled in sample buffer and the extracts were analyzed on a 12.5% polyacrylamide gel containing SDS. Lanes 1, 2, and 3 contain extracts of cells harboring only pGP1-2, lanes 4, 5, and 6 extracts of cells harboring pGP1-2 and pT7-6, and lanes 7, 8, and 9 extracts of cells harboring pGP1-2 and pTMBA1.

fragment from pSPMB373 (Vos *et al.*, manuscript in preparation) in which an 87-bp *Sau*3A fragment (from nucleotide 3240 to 3327 in the B-RNA sequence) was deleted as has been described in detail elsewhere (Fig. 2, Verver *et al.*, 1987).

Expression of CPMV-specific sequences from plasmid pTMBA1

Plasmid pTMBA1 was introduced into the *E. coli* strain K38 harboring plasmid pGP1-2 which contained the gene for the T7 RNA polymerase under the control of the λPL promoter (Tabor and Richardson, 1985). After induction of the T7 RNA polymerase by raising the temperature to 43° and addition of rifampicin, the bacterial proteins were labeled with [^{35}S]methionine for 5 min (see Materials and Methods). Bacterial extracts were prepared, and labeled proteins were subjected to analysis on a SDS–polyacrylamide gel (Fig. 3). In the presence of rifampicin, the synthesis of proteins by heat-induced bacteria containing pGP1-2 or pGP1-2 and pT7-6 decreased dramatically (Fig. 3, lanes 3 and 6). On the other hand, under these conditions several labeled polypeptides were synthesized in cells containing both pTMBA1 and pGP1-2 (Fig. 3, lane 9). Immunoprecipitation of the induced proteins demonstrated their viral nature. Anti-VP23 serum precipitated polypeptides of 105K, 90K, and 70K and a polypeptide with the same mobility as the mature capsid protein VP23 (Fig. 4, lanes 3 and 7). Anti-VP37 serum also immunoreacted with the 105K, 90K, and 70K proteins but a polypeptide with the size of capsid protein VP37 was not detected (Fig. 4, lanes 4 and 8). Antiserum against the B-RNA-encoded 24K protein precipitated the 105K and 90K products as well as 35K, 33K, and 20K polypeptides, whereas anti-VPg serum recognized the 105K, 35K, and 33K proteins (Fig. 4, lanes 5 and 6).

Pulse–chase analysis

The results presented above suggested that, in cells harboring plasmid pTMBA1, viral polypeptides were synthesized, which were proteolytically processed to produce a series of smaller products. To verify this assumption, pulse–chase experiments were performed. Heat-induced cultures of bacteria containing the plasmids pGP1-2 and pTMBA1 were labeled for 1 min with [^{35}S]methionine in the presence of rifampicin and then chased by adding an excess of unlabeled methionine. Samples removed after different chase periods were subjected to immunoprecipitation using several antisera and analyzed on SDS–polyacrylamide gels.

The analyses indicated that the polypeptides of 105K and 90K, immunoprecipitable with both anti-VP23 and anti-24K serum, represented primary translation products from pTMBA1. The amount of these products rap-

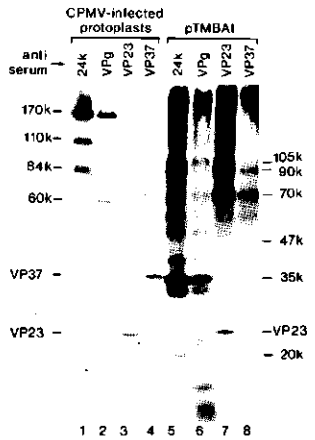


FIG. 4. Immunological analysis of polypeptides expressed from plasmid pTMBA1. Extracts of cells harboring both pGP1-2 and pTMBA1 were prepared after induction in the presence of rifampicin and labeling for 5 min with [^{35}S]methionine. After immunoprecipitating aliquots with different antisera the precipitates obtained were analyzed on a 12.5% polyacrylamide gel containing SDS (lanes 5 to 8). As a control, immunoprecipitations were carried out on [^{35}S]methionine-labeled, CPMV-infected cowpea protoplast fractions using the same antisera (lanes 1 to 4). Antisera used are indicated above each lane.

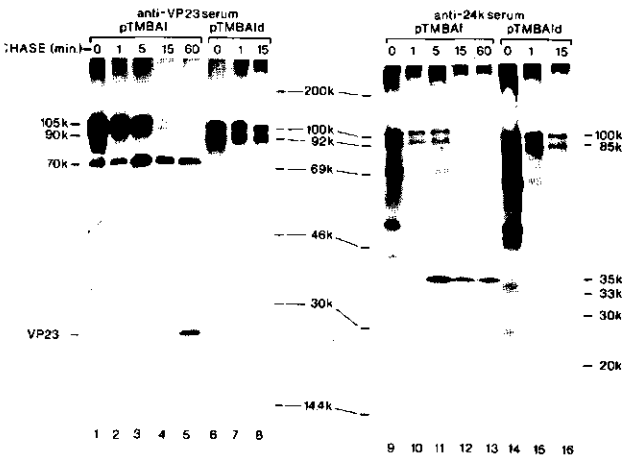


FIG. 5. Pulse–chase analysis of polypeptides expressed from pTMBA1 or pTMBA1d. *E. coli* cells were induced in the presence of rifampicin and labeled with [^{35}S]methionine for 1 min followed by a chase with unlabeled methionine. Extracts from bacteria containing pTMBA1 and pGP1-2 (lanes 1 to 5 and lanes 9 to 13) or pTMBA1d and pGP1-2 (lanes 6 to 8 and lanes 14 to 16) were subjected to immunoprecipitation using anti-VP23 serum (lanes 1 to 8) or anti-24K serum (lanes 9 to 16) and the precipitates were analyzed on a 12.5% polyacrylamide gel containing SDS. The chase times are indicated at the top of the figure. The positions of molecular weight markers in the gels are indicated in the center of the figure. The origin of the 30K protein that appeared in these experiments as a minor processing product from the polyproteins expressed from pTMBA1 was not determined.

idly decreased after 15 min of chase, while at the same time the amounts of 70K protein, reactive with anti-VP23 serum, and the 35K protein, reactive with anti-24K serum, increased (Fig. 5, lanes 1–4 and 9–13), indicating that proteolytic processing occurred in these cells. After longer chase periods the amount of free VP23 increased considerably (Fig. 5, lanes 1–4), whereas some decrease in intensity of the 70K band after longer chase periods was observed. The 33K, 30K, and 20K proteins, which all immunoprecipitated with anti-24K serum, appeared after 15 min of chase (Fig. 5, lanes 9–13).

The disappearance of the 105K and 90K proteins and the simultaneous appearance of the smaller products unequivocally demonstrate that two primary translation products (105K and 90K), which undergo proteolytic processing, are expressed from pTMBA1.

Effect of a deletion of 87 nucleotides in the coding sequence of the 24K protein on proteolytic processing

Previous experiments have indicated that the 24K polypeptide is the protease which cleaves at a glutamine–serine site in the 200K primary translation product of B-RNA, thus giving rise to the 32K and 170K polypeptides (Verver *et al.*, 1987; see also Fig. 1). To test whether this 24K polypeptide is also responsible for the proteolytic activity detected in *E. coli* cells harboring plasmid pTMBA1, we constructed plasmid pTMBA1d which was similar to pTMBA1 but contained a small deletion of 87 nucleotides in the coding region for this protein. Bacteria containing this plasmid and pGP1-2 were subjected to pulse–chase analysis as described above. This resulted in the detection of two polypeptides of 100K and 85K which run slightly faster in SDS–polyacrylamide gels than the 105K and 90K proteins, respectively, produced in cells harboring plasmid pTMBA1 (Fig. 5, lanes 6–8 and 14–16). These differences in size are in agreement with the deletion introduced in the open reading frame from pTMBA1d. After longer chase periods the amounts of both polypeptides decreased, indicating some degradation (Fig. 5, lanes 8 and 16). However, no distinct smaller products appeared, demonstrating that the specific proteolytic activity was eliminated by the small deletion in the 24K protein coding region.

Accumulation of CPMV-specific polypeptides in *E. coli* as detected by immunoblotting

To analyze the stability and possible accumulation of the CPMV-specific polypeptides detected in *E. coli*, bacteria harboring plasmids pGP1-2 and pTMBA1 or pTMBA1d were grown in enriched medium, heat-induced, and subsequently cultured in the presence of rifampicin. At different times after heat induction, samples were collected and subjected to SDS–polyacryl-

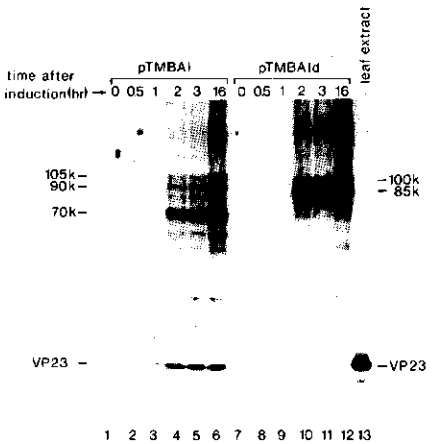


FIG. 6. Accumulation of CPMV-specific polypeptides in *E. coli*. Bacteria were grown in enriched medium, induced in the presence of rifampicin, and collected at different times after induction, as indicated at the top of the figure. Extracts from cells containing pTMBA1 and pGP1-2 (lanes 1 to 6) or pTMBA1d and pGP1-2 (lanes 7 to 12) were separated in a 12.5% polyacrylamide gel containing SDS, blotted onto nitrocellulose, and probed with anti-VP23 serum and ^{125}I-protein A. A cytoplasmic fraction of CPMV-infected cowpea leaves was run as a marker in the same gel (lane 13).

amide gel electrophoresis. The separated proteins were blotted onto nitrocellulose paper and screened with anti-VP23 serum for the presence of mature capsid protein VP23. Protein VP23 could be detected 30 min after heat induction and increasing amounts were found after longer periods of incubation (Fig. 6). Two and three hours after heat induction the major CPMV products were the 70K protein (see Fig. 6) and 35K and 33K proteins (as detected with anti-24K serum, data not shown). Although these data indicate a reasonable stability of the various polypeptides derived upon proteolytic processing from the primary translation product from pTMBA1, their amounts were too low to be visualized by Coomassie brilliant blue staining (data not shown). In extracts of cells containing the plasmid pTMBA1d only the 100K and 85K proteins could be detected, confirming the effect of the deletion in the 24K protein in abolishing the proteolytic activity (Fig. 6, lanes 7–12).

DISCUSSION

Using the binary T7 expression system (Tabor and Richardson, 1985) we have been able to express in *E. coli* cells a chimeric cDNA construct (pTMBA1) containing sequences derived from both CPMV B- and M-RNA under control of the T7 promoter ϕ10. Upon induction of the T7 RNA polymerase, long precursor proteins were produced which enabled us to study the activity and specificity of the 24K protease of CPMV, residing in the precursor.

After heat induction, bacteria harboring plasmid pGP1-2 (supplying the T7 RNA polymerase) and plasmid pTMBA1 produced a 105K polypeptide that corresponded to the full coding capacity of the CPMV-specific cDNA segment in pTMBA1 (Fig. 7; this polypeptide should not be confused with the 105K protein encoded by CPMV M-RNA that is shown in Fig. 1). In addition a second large polypeptide of 90K was detected which in contrast to the 105K protein did not react with anti-VPg serum. Since the region coding for VPg is located at the 5′ end of the chimeric cDNA insert of pTMBA1 (see Fig. 7), and since it appeared concurrently with the 105K protein in pulse–chase experiments, the 90K polypeptide may have originated from an internal intiation site, possibly at nucleotide 3210 of the B-RNA sequence, where a Shine- and Dalgarno-like sequence is found between nucleotides 3201 and 3205 (GGA, see Lomonossoff and Shanks, 1983). This idea is further supported by the finding that expression of pTMBA1d resulted in the production of a similar second product (of 80K) which, in view of the presence of a deletion in the 24K protease coding region in this plasmid, was very unlikely to be a proteolytic processing product (see later). Pulse–chase experiments and immunological analysis showed that the primary products from pTBMA1 were cleaved giving rise to a 70K protein, corresponding to the carboxy-terminal part of the 90K and 105K polypeptides (since it precipitated with anti-VP23 and anti-VP37 serum) and to several small peptides corresponding to the amino-terminal portion of these proteins. The 35K polypeptide which reacted with both anti-24K and anti-VPg serum probably corresponded to the amino-terminus of the 105K polypeptide, while the 20K polypeptide, reactive only to anti-24K serum, most likely corresponded to the amino-terminal part of the 90K protein (see Fig. 7). Furthermore a protein comigrating with the capsid protein VP23 and reactive with anti-VP23 serum was found in bacteria harboring pTMBA1.

In *E. coli* cells harboring plasmid pTMBA1d containing a small deletion in the 24K protein coding region, the proteolytic activity was completely abolished and only high-molecular-weight polypeptides (of 100K and 85K) could be detected (Fig. 5, lanes 6–8 and 14–16). This finding demonstrates that the 24K protein represents the protease responsible for the cleavages in the viral proteins produced in cells carrying pTMBA1.

The occurrence in bacteria harboring pTMBA1 of the 70K, 35K, 33K, and 20K proteins indicates that the 105K and 90K proteins are proteolytically cleaved at

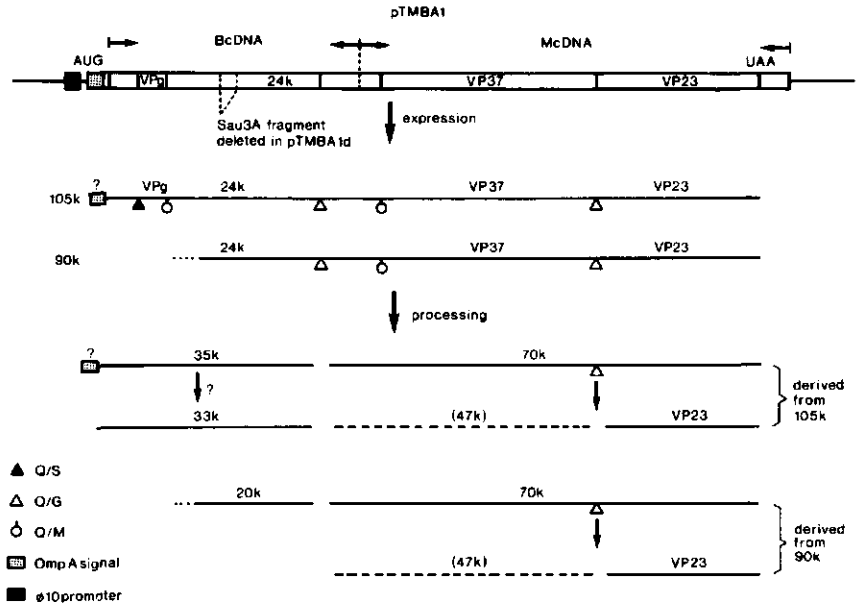


FIG. 7. Model for the processing of polypeptides expressed from pTMBA1 in *E. coli*. The CPMV-specific cDNA fragment in pTMBA1 is represented by a white bar with the relevant viral protein-coding regions and the position of the *Sau*3A fragment deleted in pTMBA1d indicated. Polypeptides are indicated by single lines. The dotted line at the N-terminus of the 90K protein illustrates that this N-terminus is not precisely known. The presence of the OmpA signal sequence at the N-termini of the 105K and 35K proteins was not analyzed. The 47K protein is represented as a broken line, since this protein was found in very low amounts. Cleavage sites are indicated by ▲, Gln–Ser, △, Gln–Gly, and ♁, Gln–Met.

the glutamine–glycine dipeptide cleavage site at the right border of the 24K protein (see Fig. 7). We have not been able to determine whether this cleavage occurred intra- or intermolecularly. *In vitro* translation experiments with B-RNA seem to indicate that the corresponding cleavage reaction in the 170K protein (see Fig. 1) is intramolecular (Peng and Shih, 1984). In view of the efficiency of the reaction this could also be the case in the artificial protein expressed from pTMBA1. The fate of the OmpA protein signal peptide in the chimeric viral polypeptides has not been analyzed, but the detection of a 33K protein immunoprecipitable with both anti-24K and anti-VPg serum suggests that probably this signal peptide sequence (of 22 amino acids) was removed from some of the 35K polypeptides. This would also agree with the observation that the 33K protein is the only protein detectable in the periplasmic space, the signal sequence being removed during transport (data not shown).

We have obtained direct evidence of cleavage at the glutamine–glycine site that is used to separate the capsid proteins from each other by detecting free VP23 both by immunoprecipitation and Western blotting. The results do not allow us to conclude whether VP23 is directly cleaved from the primary translation products from pTMBA1 or from the 70K intermediate (Fig. 7). The observation that the 70K protein appeared in pulse–chase experiments prior to VP23 and the increase of VP23 at chase times in which the amount of the 70K product remained constant suggests that the primary translation proteins are first cleaved into the 35K and 70K polypeptides and that VP23 is subsequently released from the 70K product. Thus this cleavage most likely occurs intermolecularly as is also the case upon expression of the natural RNAs of CPMV (Goldbach and Van Kammen, 1985). Immunological detection of the amino-terminal counterparts of this cleavage reaction proved to be difficult. With anti-VP37 serum only a small amount of a 47K protein was detected (Fig. 4, lane 8) that, possibly because of its hybrid nature, was highly unstable in *E. coli*.

In a previous report we have presented evidence that the 32K protein encoded by B-RNA also represents a protease that cleaves at a glutamine–methionine pair in the two M-RNA-encoded polyproteins thus giving rise to the 58K and 48K polypeptides together with a

60K precursor to the capsid proteins VP23 and VP37 (Fig. 1, Franssen *et al.*, 1984b). The chimeric protein expressed from the cDNA segment present in the plasmid pTMBA1 contains this glutamine–methionine dipeptide and, additionally, a similar cleavage site at the left border of the 24K protein (Fig. 7). Cleavage at both these sites could not be detected in the experiments described in this paper, which is in agreement with our previous results.

It has been reported that the 24K protease cleaves the B-RNA-encoded 200K polyprotein at the glutamine–serine site by which the 32K and 170K proteins are released (Verver *et al.*, 1987). Cleavage at the glutamine–serine site bordering the left side of the VPg sequence in the CPMV fusion protein expressed from pTMBA1 was not observed. This result is in agreement with the fact that synthesis of free VPg has never been detected *in vivo* or *in vitro*. A possible explanation could be that the cleavage at this glutamine–serine site (probably catalyzed by the 24K protease) can only take place in close association with the linkage of VPg to the 5′ end of the RNA during RNA replication (Jaegle *et al.*, 1987).

For poliovirus two virus-encoded proteases have also been described (Hanecak *et al.*, 1982; Toyoda *et al.*, 1986). Both proteins showed specific proteolytic activity when they were produced by expression of their coding regions in *E. coli* (Hanecak *et al.*, 1984; Toyoda *et al.*, 1986). Since protease 3C of poliovirus shows sequence homology to the 24K protease of CPMV (Franssen *et al.*, 1984) and both recognize glutamine–glycine sites (Hanecak *et al.*, 1982; this paper), these two proteases may be functionally closely related.

Thirteen glutamine–glycine and 9 glutamine–serine sites are present in the CPMV polyproteins, while only two of each group are recognized by the 24K protease. The ability of this protein to be functional in a bacterial system and to be able to recognize genuine cleavage sites in artificial precursor proteins very different in size and structure from the natural substrates will permit the study of the structural requirements around the cleavage sites by mutagenesis of these protein precursors.

ACKNOWLEDGMENTS

The authors are grateful to Dr. S. Tabor for the gift of plasmids pT7-6 and pGP1-2 and to Dr. M. Inouye for the gift of plasmid pINIII^{113}ompA2, to R. de Fluiter and P. Madern for artwork, and to G. Heitkönig for typing of the manuscript. J.A.G. was supported by an EMBO long-term fellowship.

REFERENCES

Daubert, S. D., Bruening, G., and Najarian, R. C. (1978). Protein bound to the genome RNAs of cowpea mosaic virus. *Eur. J. Biochem.* **92,** 45–51.

El Manna, M. M., and Bruening, G. (1973). Polyadenylate sequences in the ribonucleic acids of cowpea mosaic virus. *Virology* **56,** 198–206.

Franssen, H., Goldbach, R., Broekhuysen, M., Moerman, M., and Van Kammen, A. (1982). Expression of middle-component RNA of cowpea mosaic virus: *In vitro* generation of a precursor to both capsid proteins by a bottom-component RNA-encoded protease from infected cells. *J. Virol.* **41,** 8–17.

Franssen, H., Leunissen, J., Goldbach, R., Lomonossoff, G., and Zimmern, D. (1984a). Homologous sequences in non-structural proteins from cowpea mosaic virus and picornaviruses. *EMBO J.* **3,** 855–861.

Franssen, H., Moerman, M., Rezelman, G., and Goldbach, R. (1984b). Evidence that the 32,000-dalton protein encoded by bottom-component RNA of cowpea mosaic virus is a proteolytic processing enzyme. *J. Virol.* **50,** 183–190.

Ghrayeb, J., Kimura, H., Takahara, M., Hsiung, H., Masai, Y., and Inouye, M. (1984). Secretion cloning vectors in *Escherichia coli*. *EMBO J.* **3,** 2437–2442.

Goldbach, R., Rezelman, G., Zabel, P., and Van Kammen, A. (1982). Expression of the bottom-component RNA of cowpea mosaic virus: Evidence that the 60-kilodalton VPg-precursor is cleaved into single VPg and a 58-kilodalton polypeptide. *J. Virol.* **42,** 630–635.

Goldbach, R., and Van Kammen, A. (1985). Structure, replication and expression of the bipartite genome of cowpea mosaic virus. *In* "Molecular Plant Virology" (J. W. Daires, Ed.), Vol. 2, pp. 83–120. CRC Press, Boca Raton, FL.

Goldbach, R. W., Schilthuis, J. G., and Rezelman, G. (1981). Comparison of *in vivo* and *in vitro* translation of cowpea mosaic virus RNAs. *Biochem. Biophys. Res. Commun.* **99,** 89–94.

Hanecak, R., Semler, B. L., Anderson, C. W., and Wimmer, E. (1982). Proteolytic processing of poliovirus polypeptides: Antibodies to polypeptide P3-7c inhibit cleavage of glutamine–glycine pairs. *Proc. Natl. Acad. Sci. USA* **79,** 3973–3977.

Hanecak, R., Semler, B. L., Ariga, H., Anderson, C. W., and Wimmer, E. (1984). Expression of a cloned segment of poliovirus in *E. coli*: Evidence for autocatalytic production of the viral proteinase. *Cell* **37,** 1063–1073.

Jaegle, M., Wellink, J., and Goldbach, R. (1987). The genome-linked protein of cowpea mosaic virus is linked to the 5′-terminus of B and M RNA by a phosphodiester linkage to serine. *J. Gen. Virol.* **68,** 627–632.

Lomonossoff, G. P., and Shanks, M. (1983). The nucleotide sequence of cowpea mosaic virus B RNA. *EMBO J.* **2,** 2253–2258.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Pelham, H. R. B. (1979). Synthesis and proteolytic processing of cowpea mosaic virus proteins in reticulocyte lysates. *Virology* **96,** 463–477.

Peng, X. X., and Shih, D. S. (1984). Proteolytic processing of the proteins translated from the bottom-component RNA of cowpea mosaic virus. *J. Biol. Chem.* **259,** 3197–3201.

Rezelman, G., Goldbach, R., and Van Kammen, A. (1980). Expression of bottom-component RNA of cowpea mosaic virus in cowpea protoplasts. *J. Virol.* **36,** 366–373.

Russel, M., and Model, P. (1984). Replacement of the fip gene of *Escherichia coli* by an inactive gene cloned on a plasmid. *J. Bacteriol.* **159,** 1034–1039.

Stanley, J., Rottier, R., Davies, J. W., Zabel, P., and Van Kammen, A. (1978). A protein linked to the 5′ termini of both RNA components of the cowpea mosaic virus genome. *Nucleic Acids Res.* **5,** 4505–4522.

TABOR, S., and RICHARDSON, C. C. (1985). A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. USA* **82**, 1074–1078.

TOYODA, H., NICKLIN, M. J. H., MURRAY, M. G., ANDERSON, C. W., DUNN, J. J., STUDIER, F. W., and WIMMER, E. (1986). A second-virus-encoded proteinase involved in proteolytic processing of poliovirus polyprotein. *Cell* **45**, 761–770.

VAN WEZENBEEK, P., VERVER, J., HARMSEN, J., VOS, P., and VAN KAMMEN, A. (1983). Primary structure and gene organisation of the middle-component RNA of cowpea mosaic virus. *EMBO J.* **2**, 941–946.

VERVER, J., GOLDBACH, R., GARCIA, J. A., and VOS, P. (1987). In vitro expression of a full-length DNA copy of cowpea mosaic virus B-RNA: Identification of the B-RNA encoded 24-kilodalton protein as a viral protease. *EMBO J.* **6**, 549–554.

VIEIRA, J., and MESSING, J. (1982). The pUC plasmids, an M13 mp7-derived system for insertion mutagenisis and sequencing with synthetic universal primers. *Gene* **19**, 259–268.

WELLINK, J., JAEGLE, M., and GOLDBACH, R. (1987). Detection of a novel protein encoded by the B-RNA of cowpea mosaic virus, using antibodies raised against a synthetic peptide. *J. Virol.* **61**, 236–238.

WELLINK, J., REZELMAN, G., GOLDBACH, R., and BEYREUTHER, K. (1986). Determination of the proteolytic processing sites in the polyprotein encoded by the B-RNA of cowpea mosaic virus. *J. Virol.* **59**, 50–58.

YANISH-PERRON, C., VIEIRA, J., and MESSING, J. (1985). Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.

ZABEL, P., MOERMAN, M., VAN STRAATEN, F., GOLDBACH, R., and VAN KAMMEN, A. (1982). Antibodies against the genome-linked protein VPg of cowpea mosaic virus recognize a 60,000-dalton precursor polypeptide. *J. Virol.* **41**, 1083–1088.

CHAPTER 6

EXPRESSION OF THE MIDDLE COMPONENT RNA OF COWPEA MOSAIC VIRUS IN VIVO

Joan Wellink, Martine Jaegle, Heinrich Prinz[1],
Ab van Kammen and Rob Goldbach[2]

Department of Molecular Biology, Agricultural University,
De Dreijen 11, 6703 BC Wageningen, The Netherlands,
[1]Institut für Genetik, Universität Köln,
D-5000 Köln 41, FRG and [2]Department of Virology,
Binnenhaven 11, 6709 PD Wageningen, The Netherlands

Journal of General Virology, in press

J. gen. Virol. (1987), 68, 000–000. *Printed in Great Britain*

Key words: *CPMV/M RNA/capsid protein precursor*

Expression of the Middle Component RNA of Cowpea Mosaic Virus *in vivo*

By JOAN WELLINK,[1]* MARTINE JAEGLE,[1]† HEINRICH PRINZ,[2] AB VAN KAMMEN[1] AND ROB GOLDBACH[3]

[1]*Department of Molecular Biology, Agricultural University, De Dreijen 11, 6703 BC Wageningen, The Netherlands,* [2]*Institut für Genetik, Universität Köln, D-5000 Köln 41, F.R.G. and* [3]*Department of Virology, Binnenhaven 11, 6709 PD Wageningen, The Netherlands*

(*Accepted 25 June 1987*)

SUMMARY

Upon infection of cowpea mesophyll protoplasts with cowpea mosaic virus (CPMV), the only M RNA-encoded proteins detected so far have been the two capsid proteins VP37 and VP23. We now report the detection of a M_r 60000 (60K) precursor to both capsid proteins in infected protoplasts cultured in the pressence of zinc ions. Furthermore a M RNA-encoded 48K protein was detected in the membrane fraction of infected cells using antisera raised against synthetic peptides. The results obtained indicate that, just like the bottom component RNA, the M RNA of CPMV is translated *in vivo* into a polyprotein from which proteins are derived by proteolytic cleavage.

INTRODUCTION

The genome of cowpea mosaic virus (CPMV) consists of two polyadenylated plus-strand RNAs which are separately encapsidated. Previous studies on the viral proteins encoded by the larger bottom (B) component RNA, 5889 nucleotides long excluding the poly(A) tail (Lomonossoff & Shanks, 1983), have revealed that the expression of this RNA involves translation into a M_r 200000 (200K) polyprotein which upon proteolytic processing gives rise to six final cleavage products (for a recent review, see Goldbach & Van Kammen, 1985). The expression of the smaller middle (M) component RNA, 3481 nucleotides long excluding the poly(A) tail (Van Wezenbeek *et al.*, 1983), is less well understood. Although a translation model for this RNA has been proposed, this model (Fig. 1) is almost entirely based on translation studies *in vitro*.

In both the wheat germ system and rabbit reticulocyte lysates, M RNA is translated into two overlapping proteins of 105K and 95K (Pelham, 1979; Franssen *et al.*, 1982), the 105K protein being initiated at an AUG codon at position 161 and the 95K protein at an AUG codon at position 512 and/or 524 (Vos *et al.*, 1984; Van Wezenbeek *et al.*, 1983). Both primary translation products are processed by a B RNA-encoded protease to give 58K and 48K proteins respectively, together with a 60K precursor to both capsid proteins VP37 and VP23 (Fig. 1; Pelham, 1979; Franssen *et al.*, 1982, 1984). All attempts to achieve cleavage of the 60K precursor into mature capsid proteins have so far failed. Since the mature capsid proteins are the only M RNA-encoded proteins detected in infected cells (Goldbach *et al.*, 1980) it is not known whether the translation model presented in Fig. 1 is also valid *in vivo*. In this paper we report the occurrence of a 60K capsid precursor protein and an M RNA-encoded 48K protein in CPMV-infected cells. These results definitely show that the M RNA is also expressed *in vivo* by polyprotein processing.

METHODS

Virus and plants. CPMV (Sb isolate) was propagated in cowpea plants (*Vigna unguiculata* L. 'California Blackeye') as described by Klootwijk *et al.* (1977) and Van Kammen (1967), and B and M components were purified as described by Rezelman *et al.* (1980).

† Present address: Plant Breeding Institute, Maris Lane, Trumpington, Cambridge CB2 2LQ, U.K.

Inoculation of protoplasts and labelling of proteins. Mesophyll protoplasts were prepared from 10-day-old primary cowpea leaves and inoculated and incubated as described by Rezelman *et al.* (1980). When labelled proteins were required, protoplasts were supplied with [^{35}S]methionine (1000 Ci/mmol: New England Nuclear) as described by Franssen *et al.* (1984). Protoplasts were collected by centrifugation and frozen. Frozen pellets, corresponding to portions of 10^6 protoplasts, were resuspended in 100 µl of HB buffer (50 mM-Tris–acetate pH 8·2, 10 mM-potassium acetate, 1 mM-EDTA, 5 mM-dithioerythritol, 1 mM-phenylmethylsulphonyl fluoride, 10% sucrose) and centrifuged at 4 °C for 30 min at 30000 *g* yielding a supernatant (S30) fraction and a pellet fraction. In some cases the medium in which the protoplasts had been incubated was also analysed.

Fractionation of cowpea leaves. CPMV-infected and non-infected cowpea leaves were fractionated into nuclear, membrane and cytoplasmic fractions, as described by Zabel *et al.* (1982).

Preparation of antiserum against synthetic peptides. Two peptides (of nine and 30 amino acids) were synthesized by the solid-phase method (Barany & Merrifield, 1980). The peptides were characterized by amino acid analyses and high performance liquid chromatography. The shorter peptide, peptide S, was coupled to keyhole limpet haemocyanin using glutaraldehyde (Baron & Baltimore, 1982), and the longer peptide, peptide L, was not coupled to a carrier protein. To obtain monospecific antibodies the coupled peptide S and free peptide L were injected into New Zealand white rabbits at 14 day intervals (Muller *et al.*, 1986). Serum was collected 8 to 10 days after the third, fourth and fifth injections.

Translation in vitro. CPMV M RNA was translated in rabbit reticulocyte lysates (a generous gift from Dr R. J. Jackson) at 30 °C for 1 h as described by Pelham (1979). The 105K and 95K translation products were proteolytically processed into 58K, 48K and 60K proteins by the addition of 1 vol. of an S30 fraction from B RNA-inoculated protoplasts as described by Franssen *et al.* (1982).

Immunological methods. For immunoprecipitation, samples were either directly adjusted to 1 × RIA buffer (50 mM-Tris–HCl pH 7·5, 150 mM-NaCl, 5 mM-EDTA, 1% Triton X-100, 0·5% sodium deoxycholate, 0·1% SDS), or first brought to 2% SDS, 5% 2-mercaptoethanol, boiled for 1 min, diluted at least 10-fold and then adjusted to 1 × RIA buffer. Immunoprecipitation was as described by Franssen *et al.* (1982). For protein blot analysis, fractions from cowpea protoplasts or leaves were separated in 12·5% SDS–polyacrylamide gels, blotted onto nitrocellulose filters and probed with antisera (1500-fold dilution in RIA buffer) and ^{125}I-labelled Protein A as described by Zabel *et al.* (1982).

RESULTS

Detection of a M_r *60K precursor to the capsid proteins*

According to the translation model for M RNA *in vitro* (Fig. 1) the two capsid proteins VP37 and VP23 are generated from a 60K precursor protein. To verify the existence of such a precursor *in vivo*, immunoprecipitations using anti-VP23 serum were carried out in extracts from CPMV-infectred [^{35}S]methionine-labelled cowpea protoplasts. However, only mature VP23 was detected in such extracts (Fig. 2, lane 3). In an attempt to accumulate possible precursor proteins by blocking or slowing down the proteolytic processing, CPMV-infected protoplasts were labelled in the presence of $ZnCl_2$, since zinc ions are known to inhibit *in vitro* the processing of both the M RNA-encoded polyproteins (Pelham, 1979; Franssen *et al.*, 1982) and the B RNA-encoded 170K protein (Peng & Shih, 1984). Sixteen h after inoculation, a solution of $ZnCl_2$ was added to infected protoplasts to a final concentration of 2 mM-Zn^{2+} and 1 h later [^{35}S]methionine (15 µCi/ml) was added. After a further 7 h (i.e. 24 h after inoculation) the protoplasts were collected by centrifugation and extracts were prepared for immunological analysis. In protoplasts incubated in the presence of $ZnCl_2$, but not in untreated ones, a 60K protein was precipitated with anti-VP23 serum (Fig. 2, lanes 4 and 5). This protein comigrated with the 60K capsid precursor protein generated *in vitro* by processing of the translation products from M RNA *in vitro* with viral protease (Fig. 1) (Franssen *et al.*, 1982). The 60K protein was also precipitated with anti-VP37 serum (data not shown) and was only detectable in protoplasts inoculated with B and M components and not in protoplasts inoculated with B component alone (Fig. 2, lanes 7 and 9). These data are consistent with the idea that the 60K protein detected *in vivo* contains the sequences of both M RNA-encoded capsid proteins and represents a precursor to these products.

Detection of an M RNA-encoded 48K protein in infected protoplasts

To enable the detection of M RNA-encoded 48K and/or 58K proteins in infected cells, antisera were prepared against two peptides corresponding to the theoretical carboxy-terminal

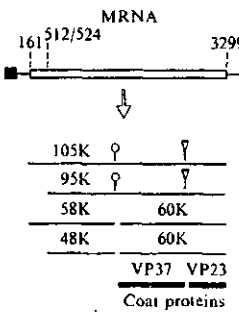


Fig. 1

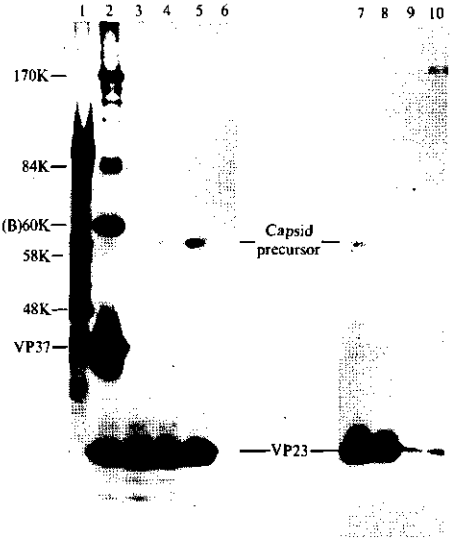


Fig. 2

Fig. 1. Model for the expression of CPMV M RNA. The long open reading frame of M RNA is indicated with an open bar on which the positions of the initiation codons used for translation are shown. The proteins are represented by single lines and the cleavage sites indicated by ○, glutamine-methionine and ▽, glutamine-glycine.

Fig. 2. Detection of the 60K precursor to both capsid proteins in CPMV-infected protoplasts. Protoplasts inoculated with whole virus (lanes 2 to 5, 7 and 8) or B component alone (lanes 9 and 10) or mock-inoculated (lane 6) were labelled with [^{35}S]methionine in the presence (lanes 4, 5, 7 and 9) or absence (lanes 2, 3, 6, 8 and 10) of 2 mM-$ZnCl_2$. Immunoprecipitations were performed using 5 µl anti-VPg serum (lane 2, this serum also reacted with both capsid proteins, see Zabel *et al.*, 1982) or 5 µl anti-VP23 serum (lanes 4 to 10). The precipitates were analysed in a 12·5% polyacrylamide gel and the proteins visualized by fluorography. Lane 1 shows the products of translation *in vitro* of CPMV M RNA. The immunoprecipitation with anti-VPg serum was performed to show the difference between the B RNA-encoded 60K protein [(B)60K, VPg precursor] and the M RNA-encoded 60K capsid protein precursor.

amino acid sequence of these proteins, as deduced from the M RNA nucleotide sequence (Van Wezenbeek *et al.*, 1983). The primary structure of the peptides S and L is schematically shown in Fig. 3. The reactivity of the antisera obtained after the fourth injection was tested by immunoprecipitation of the products of M RNA translation *in vitro*. Both antisera appeared to be able to recognize the 58K and 48K proteins as well as the 105K and 95K primary translation products (Fig. 4, lanes 1 and 2). The binding to the latter products was less pronounced probably because the antibodies bind to an internal region of these proteins. Preimmune sera did not react with any of the translation products of M RNA (data not shown). After verification of their reactivity and specificity, the two antisera were used to investigate the possible occurrence of M RNA-encoded 48K and 58K proteins in infected cells.

At first, antibodies against peptide S were used to detect these proteins. Two identical protein blots were prepared containing both cytoplasmic (S30) and pellet fractions from protoplasts inoculated with whole virus or with B component alone and from uninoculated protoplasts. The antiserum appeared to react with various proteins, mainly present in the cytoplasmic fraction

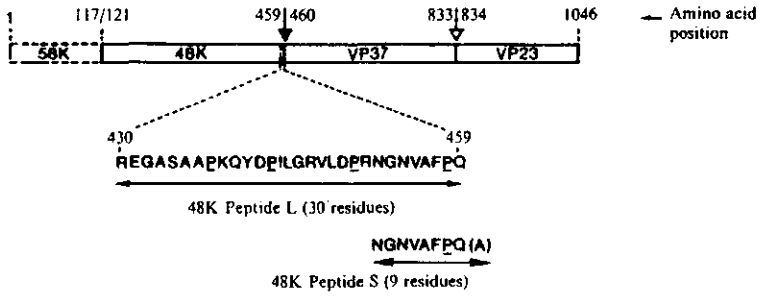


Fig. 3. Structure of the synthetic peptides and their position in the M RNA-encoded proteins. The primary structure of the peptides is shown using the single letter code for amino acids. Prolines, thought to make a large contribution to the structure of the peptides, are underlined. The alanine at the carboxy terminus of peptide S is in parentheses since it is not present at the putative carboxy-termini of the 48K and 58K proteins. ▽, glutamine-glycine; ▼, glutamine-methionine.

(Fig. 5, lanes 10 to 15). The reaction with a number of these proteins was non-specific as these proteins also reacted with antiserum pretreated with excess (100 μg) peptide S (Fig. 5, lanes 2 to 7). Further inspection of these blots revealed a protein reacting specifically with anti-peptides S serum and comigrating with the 48K product produced *in vitro* by M RNA (compare lanes 7, 15 and 16 in Fig. 5). This 48K protein was only found in the pellet fraction of protoplasts inoculated with whole virus and was missing in a similar fraction from protoplasts inoculated with the B component alone (Fig. 5, lane 14). Furthermore a 90K protein, present in low amounts in the same fraction as the 48K protein, reacted specifically with the antiserum. Surprisingly, three B RNA-encoded proteins (170K, 110K and 87K) present in the cytoplasmic fraction also appeared to react specifically with the anti-peptides S serum (Fig. 5, lanes 3 and 11) although there is no amino acid homology between these proteins and the peptide used to prepare the antiserum.

The association of the 48K protein with the pellet fraction was rather strong as it was not possible to solubilize this protein by treatment of the pellet fraction with 1% Triton X-100 (data not shown). Since one of the plant proteins that reacted non-specifically with the anti-peptide S serum had a size of approximately 60K (see Fig. 5) it was not possible to determine whether an M RNA-encoded 58K protein as predicted by the model of translation *in vitro* (Fig. 1) was present in CPMV-infected protoplasts as well. However, with the antiserum raised against peptide L, which did not show the non-specific reaction with a 60K plant protein, only a 48K protein could be detected in the pellet fraction of CMPV-infected protoplasts (see Fig. 6). Again a 90K protein also reacted with this antiserum (Fig. 6, lane 1). The anti-peptide L serum did not react with the B RNA-encoded proteins (data not shown).

Detection of the 48K protein in the membrane fraction of CPMV-infected cowpea leaves

Infected cowpea leaves were also tested for the presence of 48K and 58K proteins.CMPV-infected and uninfected leaves were fractionated into cytoplasmic and membrane fractions (Zabel *et al.*, 1982) which were incubated with anti-peptide L serum. The immunoprecipitates were separated in a 12·5% SDS–polyacrylamide gel and subsequently blotted onto nitrocellulose filters. When the blot was treated with anti-peptide L serum and ^{125}I-labelled Protein A (Fig. 7) a 48K protein was detected in the membrane fraction of infected leaves in agreement with the results obtained with protoplasts, but the 58K protein was again not found (Fig. 7, lanes 4 and 5). Proteins of about 90K and 130K, present in the same fractions as the 48K protein, also appeared to react with this antiserum. Neither the 48K nor the 58K protein could be detected in low-speed pellet fractions (nuclear fraction) or cell wall fractions prepared as described by Godefroy-Colburn *et al.* (1986) of infected leaves (data not shown).

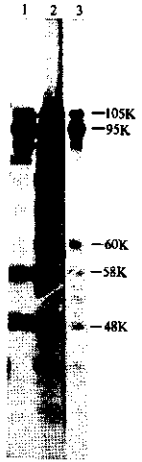


Fig. 4

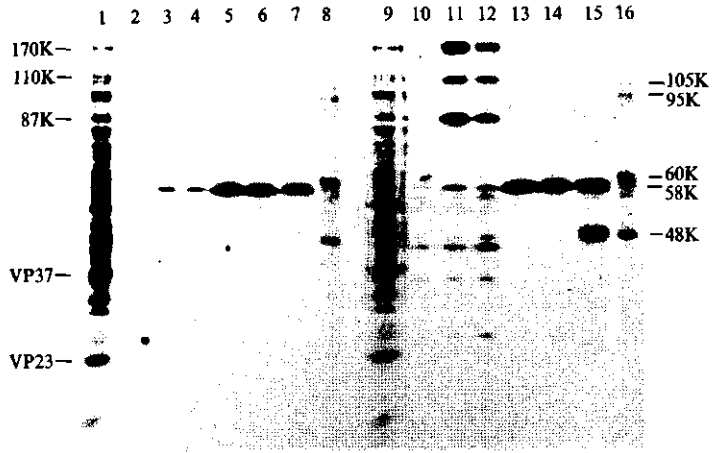


Fig. 5

Fig. 4. Immunoprecipitation of *in vitro* translation products of M RNA with anti-peptide sera. Immunoprecipitation with 10 µl anti-peptide L serum (lane 1) and 10 µl anti-peptide S serum (lane 2) were performed on products translated and processed *in vitro* of CPMV M RNA (lane 3). Immunoprecipitates were separated in a 12·5% SDS–polyacrylamide gel and the proteins visualized by fluorography.

Fig. 5. Detection of an M RNA-encoded 48K protein in CPMV-infected protoplasts. Protoplasts were inoculated with B and M components (lanes 4, 7, 12, 15), B-component alone (lanes 3, 6, 11, 14) or were left uninoculated (lanes 2, 5, 10, 13). Forty h after inoculation, protoplasts were collected and fractionated into pellet fractions (lanes 5, 6, 7, 12, 14, 15) and cytoplasmic (S30) fractions (lanes 2, 3, 4, 10, 11, 12). Two identical protein blots were prepared containing these fractions (each isolated from 2×10^5 protoplasts) separated in a 12·5% SDS–polyacrylamide gel. Lanes 10 to 15 were incubated with anti-peptide S serum and ^{125}I-labelled Protein A. Lanes 2 to 7 were incubated with the same antiserum, pretreated with excess (100 µg) peptide S, and ^{125}I-labelled Protein A. Lanes 1 and 9 contain [^{35}S]methionine-labelled CPMV-infected protoplasts and lanes 8 and 16 [^{35}S]methionine-labelled products of CPMV M RNA translated *in vitro*.

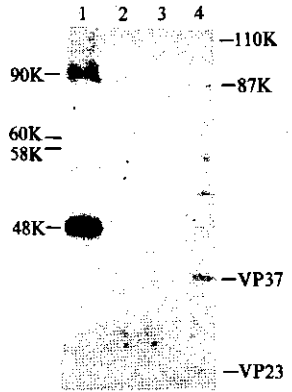


Fig. 6

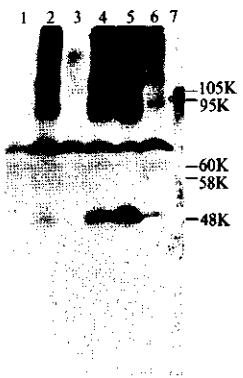


Fig. 7

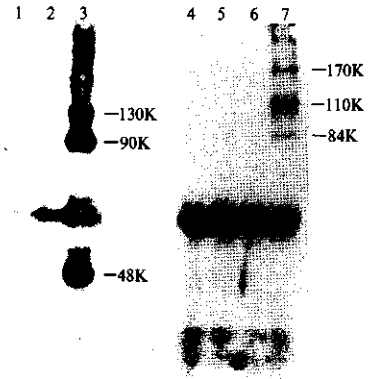


Fig. 8

Fig. 6. Protoplasts were inoculated with B and M component (lane 1) or B component only (lane 2) or were left uninoculated (lane 3). Forty h after inoculation protoplasts were collected and pellet fractions were isolated. A protein blot was prepared containing these fractions separated in a 12·5% SDS–polyacrylamide gel. The blot was incubated with anti-peptide L serum and ^{125}I-labelled Protein A. Lane 4 contains [^{35}S]methionine-labelled CPMV-infected protoplasts.

Fig. 7. Detection of the 48K protein in the membrane fraction of CPMV-infected cowpea leaves. A total leaf extract (lane 2) was fractionated into cytoplasmic (lane 3), membrane (lane 4), washed membrane (lane 5) and membrane wash fractions (lane 6). Immunoprecipitations with 10 µl of anti-peptide L serum were performed on these fractions which each contained approximately 50 µg of protein. The precipitates were separated in a 12·5% SDS-polyacrylamide gel and subsequently blotted onto nitrocellulose filters. The gel was incubated with anti-peptide L serum and ^{125}I-labelled Protein A. Lane 1 contains the precipitate of the anti-peptide L serum and a membrane fraction of uninfected leaves. Lane 7 contains [^{35}S]methionine-labelled products of CPMV M RNA translated *in vitro*.

Fig. 8. Detection of the 48K protein in the culture medium of infected protoplasts. Protoplasts were inoculated with complete virus (B + M) (lanes 3, 5 and 7), B components only (lanes 2 and 4), or were left uninoculated (lanes 1 and 6) and collected 40 h later. Immunoprecipitations were performed on 300 µl of the culture medium that had contained about 3×10^5 protoplasts (lanes 1 to 5) and on a cytoplasmic fraction isolated from about 2×10^5 protoplasts (lanes 6 and 7) using 10 µl anti-peptide L serum (lanes 1, 2 and 3) or 5 µl anti-24K serum (lanes 4 to 7). The precipitates were separated in a 12·5% SDS–polyacrylamide gel and subsequently blotted onto nitrocellulose filters. The blot containing lanes 1 to 3 was incubated with anti-peptide L serum and ^{125}I-labelled Protein A while that containing lanes 4 to 7 was incubated with anti-24K serum and ^{125}I-labelled Protein A.

Possible excretion of the 48K protein from cowpea protoplasts

It has been suggested that the 48K protein may have a function in the cell-to-cell transport of virus (or viral RNA) in cowpea leaves (Rezelman *et al.*, 1982). Therefore we examined whether this protein was excreted by infected protoplasts. Considerable amounts of the 48K protein could indeed be precipitated by anti-peptide L serum from the culture medium (Fig. 8, lane 3). Some larger products of about 90K and 130K were also detected in this medium. To test for proteins in the medium originating from damaged protoplasts, the culture medium was tested for the presence of other viral proteins. No proteins were detected using the anti-24K serum which specifically reacts with B RNA-encoded 170K, 110K, 84K and 24K proteins (Wellink *et al.*, 1987) (Fig. 8, lanes 4 and 5). It is possible that the 48K protein was much more stable in medium than other viral proteins, but it appears more plausible that the 48K protein was specifically excreted by the protoplasts into the medium.

DISCUSSION

Of the two RNAs that make up the genome of CPMV, the expression of B RNA has been extensively studied *in vivo* as well as *in vitro* (Goldbach & Van Kammen, 1985). Analyses of the translation *in vivo* of M RNA have been hampered because the only M RNA-specified products found *in vivo* were the two mature capsid proteins VP37 and VP23. As a result the expression of M RNA has almost exclusively been studied in translation systems *in vitro*. The *in vitro* data (Fig. 1) have now gained substantial support by the detection in infected cells of a CPMV M RNA-encoded 48K protein and a 60K precursor to both capsid proteins similar to the proteins produced *in vitro* by proteolytic processing of the M RNA-encoded 95K primary translation product.

The 60K precursor was only detectable in CPMV-infected cowpea protoplasts when they were incubated in the presence of 2 mM-$ZnCl_2$. It is possible that normally the 60K capsid precursor is rapidly cleaved *in vivo* into the mature capsid proteins VP23 and VP37, while in the presence of Zn^{2+} ions the protease (or alternatively the production of the protease) responsible for this cleavage is inhibited, thus causing the accumulation of the unstable precursor. Zn^{2+} ions are indeed known to interfere with the processing of the CPMV polyproteins by inhibiting the cleavage of the B RNA-encoded 170K precursor into 110K and 60K proteins (Peng & Shih, 1984; J. Wellink, unpublished results) as well as the cleavage of the M RNA-encoded polyproteins *in vitro* (Pelham, 1979; Franssen *et al.*, 1982).

Two initiation sites are used during the translation of M RNA *in vitro* resulting in primary translation products of 105K and 95K (Pelham, 1979; Franssen *et al.*, 1982; Vos *et al.*, 1984). These proteins are cleaved by a B RNA-encoded protease into 58K and 48K proteins together with the 60K capsid precursor protein discussed above (see Fig. 1; Pelham, 1979; Franssen *et al.*, 1982, 1984). By employing antibodies raised against synthetic peptides the 48K but not the 58K protein has now been detected in infected cowpea leaves and protoplasts. Since the 58K and 48K protein have identical carboxy termini it is not very likely that only the carboxy terminus of the 58K protein is processed (this would make it impossible for our antisera to react with this protein). Therefore such hypothetical modification does not provide the reason for our failure to detect this protein in infected cells. This raises the question whether the AUG codon at position 161 at the beginning of the long open reading frame of M RNA is actually used for translation *in vivo*. When M RNA is translated *in vitro*, the 95K protein is always produced in greater quantities than the 105K protein, thus indicating that, *in vitro*, there is a preference for the internal initiation codons (at positions 512 and/or 524) over the 5′ proximal AUG at position 161. Both internal initiation codons indeed conform better to the Kozak rule (Kozak, 1984; Van Wezenbeek *et al.*, 1983). It is therefore possible that *in vivo* the AUG codon at position 161 is rarely or never used. Alternatively the 105K and 95K proteins may be produced in more or less the same proportion as *in vitro*, but upon processing the 58K protein is rapidly degraded, whereas the 48K protein is accumulated. This idea is supported by the amino-terminal rule described by Bachmair *et al.* (1986), if one assumes that the amino-terminal methionine, coded for by the initiation codon, is removed from the 58K and 48K proteins [as is the case with the amino-terminal methionine of the CPMV B RNA-encoded 200K polyprotein (Wellink *et al.*,

1986)]. This would result in a 58K protein with an amino-terminal phenylalanine (predicted to be unstable; Bachmair *et al.*, 1986) and a 48K protein with an amino-terminal serine (predicted to be stable. Bachmair *et al.*, 1986) when the AUG codon at position 524 is used as the initiation codon. Since all comoviruses tested so far produce two polyproteins upon translation *in vitro* of their M RNAs (Goldbach & Krijt, 1982), it is possible that they all follow this strategy. This idea is further supported by the sequence of the M RNA of red clover mottle virus (RCMV), another comovirus (Shanks *et al.*, 1986). A considerable amount of amino acid homology was detected between the proteins encoded by the M RNAs of CPMV and RCMV, except for the region coding for the amino terminus of the CPMV 105K protein. In fact, the homology with the RCMV proteins starts just behind the AUG codon of the 95K protein (Shanks *et al.*, 1986). Also for RCMV, the amino-terminal rule predicts an unstable 58K protein and a stable 48K protein. But why do comoviruses use two initiation sites on their M RNA resulting in a stable and an unstable protein with partially overlapping sequences? It may be possible that by the use of several initiation sites, the M RNA of these viruses is translated more efficiently, thus producing the large amounts of capsid proteins needed to encapsidate the viral RNA.

A 95K primary translation product has not been detected in cowpea leaves or in protoplasts, either with antisera against the capsid proteins or with antisera against the synthetic peptides, presumably because it is rapidly cleaved into the 48K protein and the 60K capsid precursor. In fractions that contained the 48K protein, proteins of 90K and 130K were also found to react specifically with the anti-peptide L serum. Although their origin is unknown, their sizes suggest that they are dimers and trimers of the 48K protein.

The 48K protein was detected by antisera directed against both peptide S and peptide L. Both antisera reacted non-specifically with several plant-encoded proteins. The anti-peptide S serum also reacted specifically with three CPMV B RNA-encoded proteins, but at the moment we cannot explain this reaction.

The 48K protein was found to be present in the membrane fraction of both infected leaves and protoplasts. The association with membranes was strong since treatment of the membranes with 1% Triton X-100 did not solubilize the protein. Additionally the 48K protein and its presumptive multimers were detected in medium in which infected protoplasts had been cultured, whereas no other non-structural proteins could be detected in this medium. The presence of free 48K protein seems not to be due simply to leakage of damaged protoplasts, but may be the result of specific excretion. It has been suggested that the 48K protein is involved in transport of virus particles or RNA throughout leaves (Rezelman *et al.*, 1982). The presence of this protein in membranes of infected cells as well as in culture medium is consistent with this idea. Furthermore the 48K protein shows limited homology to the 30K protein of tobacco mosaic virus (Meyer *et al.*, 1986) which is thought to be involved in cell-to-cell transport (Leonard & Zaitlin, 1982; Ohno *et al.*, 1983; Zimmern & Hunter, 1983).

We thank Job Tielens for photography and artwork, and Gré Heitkönig for typing the manuscript. This work was supported by the Netherlands Foundation for Chemical Research (S.O.N.) with financial aid from the Netherlands Organization for the advancement of Pure Research (Z.W.O.). M.J. was the recipient of a fellowship from the Commission of the European Communities in the molecular biology and radiobiology program.

REFERENCES

BACHMAIR, A., FINLEY, D. & VARSHAVSKY, A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179–186.

BARANY, G. & MERRIFIELD, R. B. (1980). Solid-phase peptide synthesis. In *The Peptides*, vol. 2, pp. 1–24. Edited by E. Gross & J. Meienhofer. New York: Academic Press.

BARON, M. H. & BALTIMORE, D. (1982). Antibodies against the chemically synthesized genome-linked protein of poliovirus react with native virus-specific proteins. *Cell* **28**, 395–404.

FRANSSEN, H., GOLDBACH, R., BROEKHUYSEN, M., MOERMAN, M. & VAN KAMMEN, A. (1982). Expression of middle-component RNA of cowpea mosaic virus: in vitro generation of a precursor to both capsid proteins by a bottom-component RNA-encoded protease from infected cells. *Journal of Virology* **41**, 8–17.

FRANSEN, H., MOERMAN, M., REZELMAN, G. & GOLDBACH, R. (1984). Evidence that the 32,000-dalton protein encoded by bottom-component RNA of cowpea mosaic virus is a proteolytic processing enzyme. *Journal of Virology* **50**, 183–190.

GODEFROY-COLBURN, T., GAGEY, M.-J., BERNA, A. & STUSSI-GARAUD, C. (1986). A non-structural protein of alfalfa mosaic virus in the walls of infected tobacco cells. *Journal of General Virology* **67**, 2233–2239.

GOLDBACH, R. & KRIJT, J. (1982). Cowpea mosaic virus-encoded protease does not recognize primary translation products of M RNAs from other comoviruses. *Journal of Virology* **43**, 1151–1154.

GOLDBACH, R. & VAN KAMMEN, A. (1985). Structure, replication and expression of the bipartite genome of cowpea mosaic virus. In *Molecular Plant Virology*, vol. 2, pp. 83–120. Edited by J. W. Davies, Boca Raton: CRC Press.

GOLDBACH, R., REZELMAN, G. & VAN KAMMEN, A. (1980). Independent replication and expression of B-component RNA of cowpea mosaic virus. *Nature, London* **286**, 297–300.

KLOOTWIJK, J., KLEIN, I., ZABEL, P. & VAN KAMMEN, A. (1977). Cowpea mosaic virus RNAs have neither m^7GpppN . . . nor mono-, di-, or triphosphates at their 5'-ends. *Cell* **11**, 73–82.

KOZAK, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Research* **12**, 857–872.

LEONARD, D. A. & ZAITLIN, M. (1982). A temperature-sensitive strain of tobacco mosaic virus defective in cell-to-cell movement generates an altered viral-coded protein. *Virology* **117**, 416–424.

LOMONOSSOFF, G. P. & SHANKS, M. (1983). The nucleotide sequence of cowpea mosaic virus B RNA. *EMBO Journal* **2**, 2253–2258.

MEYER, M., HEMMER, O., MAYO, M. A. & FRITSCH, C. (1986). The nucleotide sequence of tomato black ring virus RNA-2. *Journal of General Virology* **67**, 1257–1271.

MULLER, S., PLAUE, S., COUPPEZ, M. & VAN REGENMORTAL, M. H. V. (1986). Comparison of different methods for localizing antigenic regions in histone H2A. *Molecular Immunology* **23**, 593–601.

OHNO, T., TAKAMATSU, T., MESHI, T., OKADA, Y., NISHIGUCHI, M. & KIHO, Y. (1983). Single amino acid substitution in 30K protein of TMV defective in virus transport function. *Virology* **131**, 255–258.

PELHAM, H. R. B. (1979). Synthesis and proteolytic processing of cowpea mosaic virus proteins in reticulocyte lysates. *Virology* **96**, 463–477.

PENG, X. X. & SHIH, D. S. (1984). Proteolytic processing of the proteins translated from the bottom-component RNA of cowpea mosaic virus. *Journal of Biological Chemistry* **259**, 3197–3201.

REZELMAN, G., GOLDBACH, R. & VAN KAMMEN, A. (1980). Expression of bottom component RNA of cowpea mosaic virus in cowpea protoplasts. *Journal of Virology* **36**, 366–373.

REZELMAN, G., FRANSSEN, H. J., GOLDBACH, R. W., IE, T. S. & VAN KAMMEN, A. (1982). Limits to the independence of bottom component RNA of cowpea mosaic virus. *Journal of General Virology* **60**, 335–342.

SHANKS, M., STANLEY, J. & LOMONOSSOFF, G. P. (1986). The primary structure of red clover mottle virus middle component RNA. *Virology* **155**, 697–706.

VAN KAMMEN, A. (1967). Purification and properties of the components of cowpea mosaic virus. *Virology* **31**, 635–642.

VAN WEZENBEEK, P., VERVER, J., HARMSEN, J., VOS, P. & VAN KAMMEN, A. (1983). Primary structure and gene organisation of the middle-component RNA of cowpea mosaic virus. *EMBO Journal* **2**, 941–946.

VOS, P., VERVER, J., VAN WEZENBEEK, P., VAN KAMMEN A. & GOLDBACH, R. (1984). Study of the genetic organisation of a plant viral RNA genome by in vitro expression of a full-length DNA copy. *EMBO Journal* **3**, 3049–3053.

WELLINK, J., REZELMAN, G., GOLDBACH, R. & BEYREUTHER, K. (1986). Determination of the proteolytic processing sites in the polyprotein encoded by the B-RNA of cowpea mosaic virus. *Journal of Virology* **59**, 50–58.

WELLINK, J., JAEGLE, M. & GOLDBACH, R. (1987). Detection of a novel protein encoded by the B-RNA of cowpea mosaic virus, using antibodies raised against a synthetic peptide. *Journal of Virology* **61**, 236–238.

ZABEL, P., MOERMAN, M., VAN STRAATEN, F., GOLDBACH, R. & VAN KAMMEN, A. (1982). Antibodies against the genome-linked protein VPg of cowpea mosaic virus recognize a 60,000-dalton precursor polypeptide. *Journal of Virology* **41**, 1083–1088.

ZIMMERN, D. & HUNTER, T. (1983). Point mutation in the 30K open reading frame of TMV implicated in temperature-sensitive assembly and local lesion spreading of mutant Ni 2519. *EMBO Journal* **2**, 1893–1900.

(*Received 27 April 1987*)

CHAPTER 7

DETECTION OF VIRAL PROTEINS IN THE CYTOPHATIC STRUCTURES OF COWPEA MOSAIC VIRUS INFECTED COWPEA PROTOPLASTS

Joan Wellink, Jan van Lent[1] and Rob Goldbach[1]

Department of Molecular Biology, Agricultural University,
De Dreijen 11, 6703 BC Wageningen, The Netherlands and
[1]Department of Virology, Binnenhaven 11,
6709 PD Wageningen, The Netherlands

SUMMARY

Infection of cowpea (Vigna unguiculata) cells with cowpea mosaic virus (CPMV) is accompanied by the appearance of characteristic cytophatic structures. A major constituent of these cytophatic structures represented by the electron-dense material, was shown to contain viral non-structural proteins by an immunocytochemical technique using colloidal gold labelled protein A as a marker. With this same technique virus particles were found to accumulate mainly in the cytoplasm.

The genome of cowpea mosaic virus (CPMV) consists of two positive strand RNAs which are separately encapsidated in identical protein shells. Both RNAs (denoted B and M RNA) have a small protein (VPg) linked at their 5' end, are polyadenylated and possess a single long open reading frame (for a recent review see Goldbach and Van Kammen, 1985). Figure 1 schematically shows how the large polyproteins encoded by these RNAs are processed into functional polypeptides. Infection with CPMV is accompanied by the appearance of characteristic cytopathic structures (Assink et al., 1973; De Zoeten et al., 1974). These structures consist of amorphous electron-dense material and numerous membranous vesicles. Autoradiography performed on sections of isolated cytophatic structures treated with [^{3}H] uridine provided evidence that the replication of the CPMV RNAs was associated with these vesicles (De Zoeten et al., 1974), but the nature of the electrondense material has remained unknown.

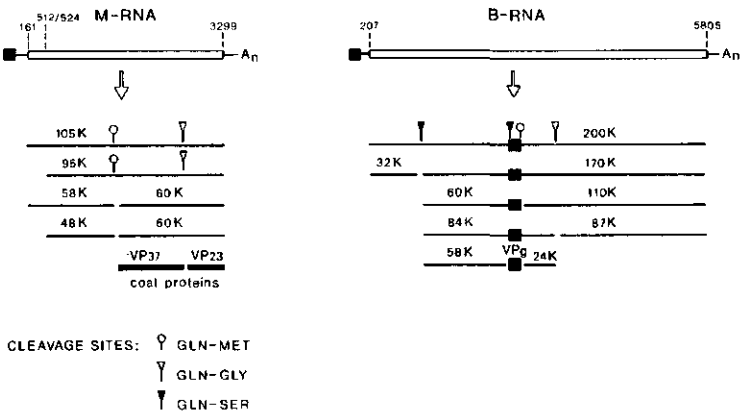


Fig. 1. Model for the expression of the CPMV RNAs. The open reading frames in both RNAs are indicated by white bars. Proteins are represented as single lines and VPg as a black square. The positions of the translational start and stop codons are indicated.

In CPMV infected cells large amounts of virus particles are produced (up to 1 mg per g leaves). Besides, most of the non-structural proteins are also easily detectable in infected protoplasts by labelling with [^{35}S] methionine (Rezelman et al., 1980). In an effort to localize sites within the infected cell where these structural and non-structural viral proteins accumulate we used an immunocytochemical technique based on colloidal gold labelled protein A.

Preparation of protein A-gold complexes with gold particle diameter of 7 nm and immunogold labelling were as described by Van Lent and Verduin (1986). The antisera used for the experiments described in this paper have been characterized for their reactivity with CPMV specific proteins presented on protein blots. The anti-24K serum recognized the 24K protein and all proteins containing the 24K sequence namely 170K, 110K and 84K proteins (fig. 1; Wellink et al., 1987). Anti-170K serum recognized the 170K, 110K and 87K proteins (Dorssers et al., 1984). Anti-VPg serum, obtained after injection of a rabbit with synthetic VPg (Jaegle et al., 1987), reacted with 170K, 84K and 60K proteins on a protein blot (J. Wellink, unpublished results). The anti-CPMV serum was obtained by injecting a rabbit with purified CPMV components. Gamma globulins (IgG) were purified from the antisera by affinity-chromatography on columns of sepharose CL-4B-protein A.

Cowpea mesophyll protoplasts were isolated, inoculated with CPMV components or RNA and cultured under continuous light as described by Rezelman et al. (1980) and Maule et al. (1980). At different times after inoculation (t = 0, 12, 24 and 48 h) samples of these cells were fixed, dehydrated and embedded at low teperature in lowicryl K4M as described by Van Lent and Verduin (1986).

The sections of embedded protoplasts showed that cell organization and organelles were preserved, however, preservation of membranes was poor due to extraction during the embedding procedure. Infection of protoplasts was first recognized at t = 12 h after inoculation by the presence of electron-dense material in the cytoplasm. No vesicles were found at this time. At t = 24 and 48 h cytopathic structures had developed consisting of electron-dense material and an increasing number of vesicles. Similar results were reported by Rezelman et al. (1982).

In sections of protoplasts, isolated at t = 12, 24 and 48 h after inoculation and treated with anti-24K IgG and protein A-gold, gold label was exclusively located on the electron-dense material of the cytopathic complex (see Fig. 2). Similar results were obtained with anti-VPg serum (data not shown) and anti-170K serum (Fig. 3). No gold label could be detected in or near the vesicles of the cytopathic complex, the cytoplasm and cell organelles when antisera against the non-structural proteins were used (Fig. 2,3). However, with anti-CPMV IgG and protein A-gold, label was found throughout the cytoplasm of

Fig. 2. Section of cowpea mesophyll protoplasts, isolated 48 h after inoculation with CPMV RNA, and incubated with anti-24K IgG and protein A-gold. Gold label (black dots) is mainly present over the electron-dense material (ED); C, chloroplast; V, vacuole. Bar marker represents 200 nm.

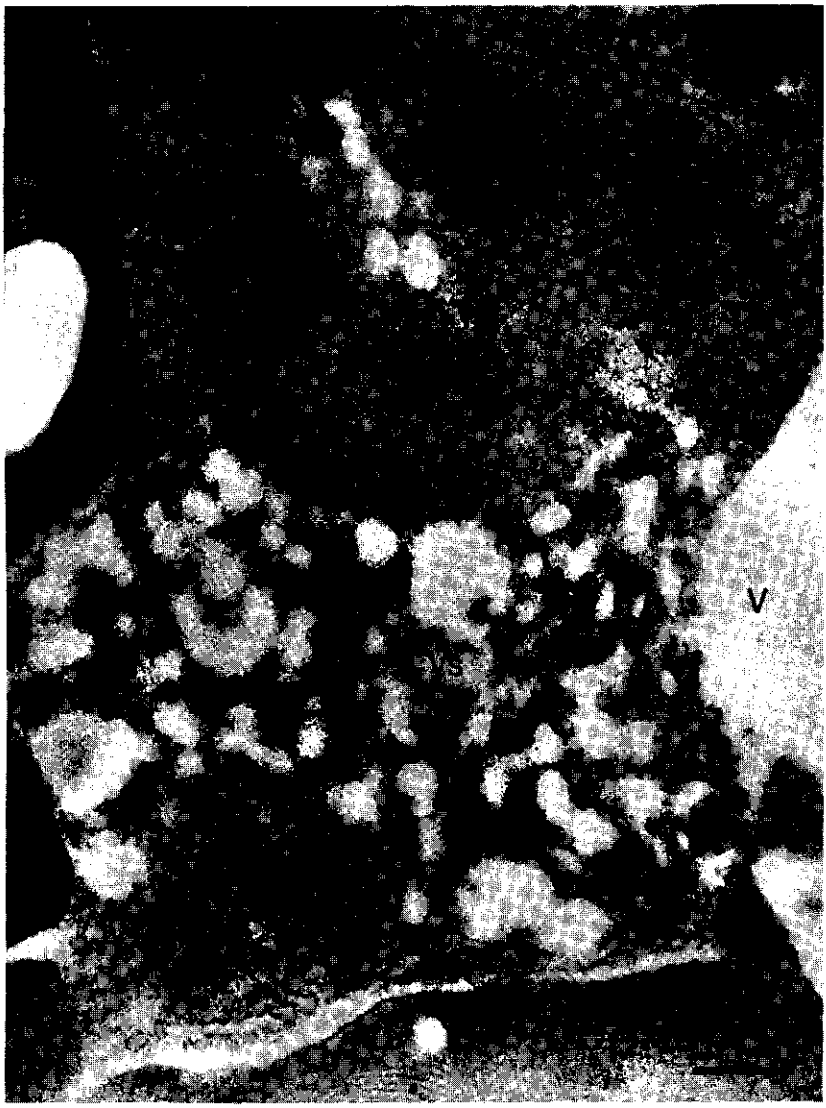


Fig. 3. Section of cowpea mesophyll protoplast, isolated 48 h after inoculation with CPMV, and incubated with anti-170K IgG and protein A-gold. Gold label is mainly present over the electron-dense material (ED) but not over and between the vesicles (Vs). C, chloroplast; V, vacuole. Bar marker represents 400 nm.

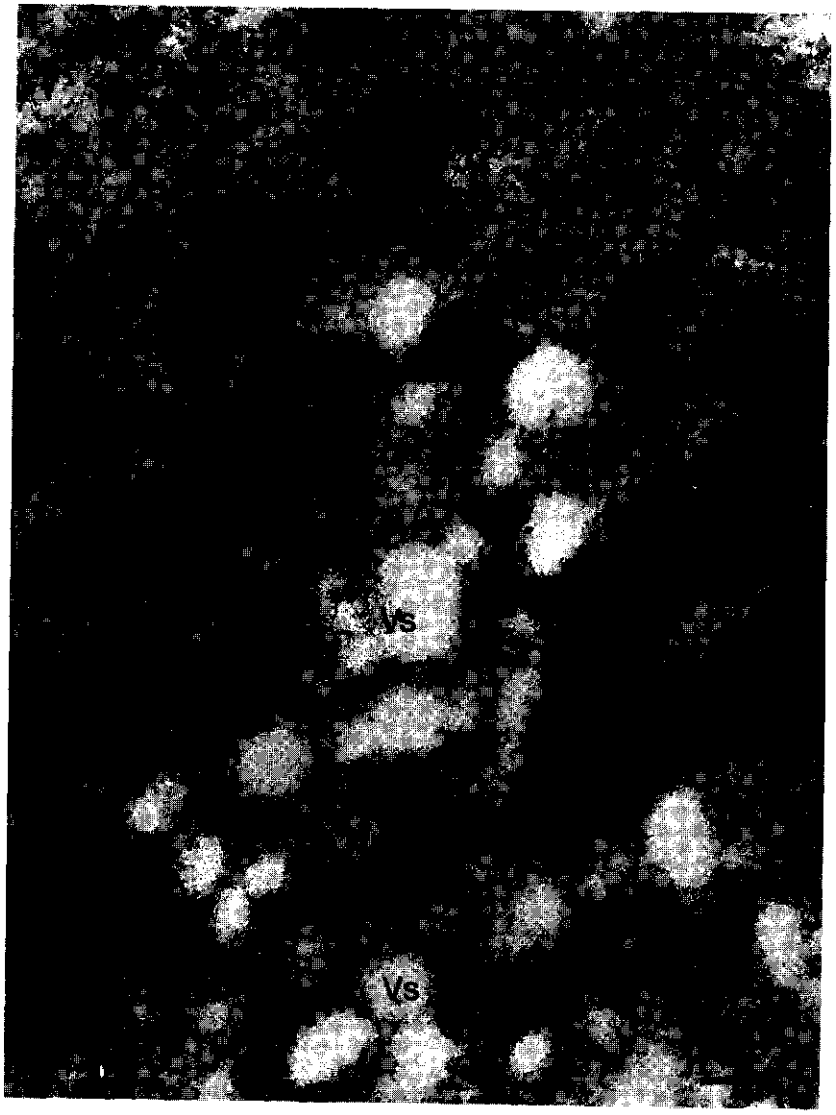


Fig. 4. Section of cowpea mesophyll protoplast, isolated 48 h after inoculation with CPMV RNA, and incubated with anti-CPMV IgG and protein A-gold. Gold label is present over the cytoplasm and between the vesicles (Vs) but is almost absent over the electron-dense material (ED); N, nucleus. Bar marker represents 200 nm.

protoplasts isolated 48 h after inoculation (Fig. 4). Similar occurrence of virus particles was reported for another comovirus, red clover mottle virus, in pea leaf cells (Tomenius et al., 1983).

The results obtained with the antisera against the 24K and 170K proteins and VPg show that the electron-dense material from the CPMV-induced cytopathic structure contains CPMV B RNA-encoded non-structural proteins. So far we have not been able to determine whether all B RNA-encoded proteins are present in this structure, since the three antisera used for these experiments each react with a set of overlapping proteins (Fig. 1). The vesicles that are also present in the cytopathic structure have been implicated in viral RNA replication (De Zoeten et al., 1974), but it remains to be established whether the electron dense material also fulfils a function in this process. It has been proposed that the viral proteins active in the replication of the CPMV RNAs are only able to synthesize one RNA strand (Van Kammen and Eggen, 1986). It is therefore possible that these used up proteins accumulate in the electron-dense material.

The authors thank Gré Heitkönig for typing of the manuscript. This work was financially supported by a grant from the Netherlands Organisation for Biological Research (BION) to J.v.L.

REFERENCES

Assink, A.M., Swaans, H. and Van Kammen, A. (1973) The localisation of virus-specific double-stranded RNA of cowpea mosaic virus in subcellular fractions of infected Vigna leaves. Virology 53, 384-391.

De Zoeten, G.A., Assink, A.M. and Van Kammen, A. (1974) Association of cowpea mosaic virus-induced double-stranded RNA with a cytopathological structure in infected cells. Virology 59, 341-355.

Dorssers, L., Van der Krol, S., Van der Meer, J., Van Kammen, A. and Zabel, P. (1984) Purification of cowpea mosaic virus RNA replication complex: Identification of a virus-encoded 110,000 dalton polypeptide responsible for RNA chain elongation. Proceedings of the National Academic of Sciences USA 81, 1951-1955.

Goldbach, R. and Van Kammen, A. (1985) Structure, replication and expression of the bipartite genome of cowpea mosaic virus. In Molecular Plant Virology, vol. 2, pp. 83-120. Edited by J.W Davies, Boca Raton, CRC Press.

Jaegle, M., Wellink, J. and Goldbach, R. (1987) The genome-linked protein of cowpea mosaic virus is bound to the 5' terminus of virus RNA by a phosphodiester linkage to serine. Journal of General Virology 68, 627-632.

Maule, A.J., Boulton, M.I., Edmunds, C. and Wood, K.R. (1980) Polyethylene glycol-mediated infection of cucumber protoplasts by cucumber mosaic virus and virus RNA. Journal of General Virology 47, 199-203.

Rezelman, G., Franssen, H.J., Goldbach, R.W., Ie, T.S. and Van Kammen, A. (1982) Limits to the independence of bottom component RNA of cowpea mosaic virus. Journal of General Virology 60, 335-342.

Rezelman, G., Goldbach, R. and Van Kammen, A. (1980) Expression of bottom component RNA of cowpea mosaic virus in cowpea protoplasts. Journal of Virology 36, 366-373.

Tomenius, K., Clapham, D. and Oxelfelt, P. (1983) Localization by immunogold cytochemistry of viral antigen in sections of plant cells infected with red clover mottle virus. Journal of General Virology 64, 2669-2678.

Van Kammen, A. and Eggen, H.I.L. (1986) The replication of cowpea mosaic virus. BioEssays 5, 261-266.

Van Lent, J.W.M. and Verduin, B.J.M. (1986) Detection of viral protein and particles in thin sections of infected plant tissue using immunogold labeling. In Developments in Applied Biology I. Developments and applications in virus testing, pp. 193-211. Edited by R.A.C. Jones and L. Torrance. Association of Applied Biologist, Wellsbourne, U.K.

Wellink, J., Jaegle, M. and Goldbach, R. (1987) Detection of a novel protein encoded by the B RNA of cowpea mosaic virus, using antibodies raised against a synthetic peptide. Journal of Virology 61, 263-238.

CHAPTER 8

SUMMARY AND CONCLUSIONS

In our study on the proteolytic processing of the 200K polyprotein encoded by CPMV B-RNA we first examined the types of cleavage sites present in this polyprotein. Previously Zabel *et al.* (1984) had shown that VPg is released from its 60K precursor by cleavage between a glutamine-serine dipeptide sequence, and the question was whether all cleavages in the 200K polyprotein occurred at glutamine-serine sites or not. The determination of partial amino acid sequences of isolated B-RNA-encoded proteins and alignment of these sequences with the open reading frame in B-RNA revealed that three types of cleavage sites are used to process the 200K polyprotein namely glutamine-serine (2x), glutamine-glycine and glutamine-methionine amino acid pairs (Chapter 3). A glutamine-methionine and glutamine-glycine site are also present in the M-RNA-encoded polyprotein, as revealed by partial amino acid sequencing of the capsid proteins (Franssen *et al.*, 1986). A common feature of the sequences surrounding the different cleavage sites is that they have alanine at position -4 and alanine or proline at position -2, but beyond that there is no obvious homology among the cleavage sites. Since there occur several glutamine-glycine, glutamine-serine and glutamine-methionine dipeptide sequences in the polyproteins which are not cleaved, probably the secondary and tertiary structure of the polypeptide chains are also important factors in determining the cleavage sites involved in the processing.

The processing model of the 200K polyprotein, based on analysis of B-RNA-encoded proteins found *in vivo* and *in vitro* translation studies, postulated the formation of a 24K protein as a final cleavage product (Rezelman *et al.*, 1980; Franssen *et al.*, 1984a). To demonstrate the actual presence of such protein in infected cells we have used antibodies raised against a synthetic peptide with an amino acid sequence corresponding to part of this hypothetical protein. The antibodies indeed reacted with a 24K protein in CPMV infected protoplasts and also with the 84K, 110K and 170K precursors which contain the sequence of the 24K protein (Chapter 4). In view of the results of processing of the *in vitro* translation products of B-RNA and its homology to the picornavirus encoded proteases it was previously suggested that the 24K protein possesses proteolytic activity. The protease activity of the 24K protein, was examined by expressing a hybrid cDNA construct, containing the 24K coding region linked to the coding region of both capsid proteins, in *E. coli* using the T7 promoter/polymerase system of Tabor and Richardson (1985) (Chapter 5). This resulted in the production of several virus-specific proteins which were characterized using specific antibodies. Pulse-chase experiments showed that two primary products were pro-

duced (the smaller one was probably the result of internal initiation of translation) which underwent faithful cleavage at two glutamine-glycine sites. One of the cleavage products represented the small capsid protein VP23. When a construct was used in which the 24K coding sequence contained a small deletion only two large proteins could be detected. These results unequivocally indicate that the 24K protein catalyzes the cleavages at the glutamine-glycine sites in the CPMV polyproteins (Chapter 5).

Recent experiments described by Verver *et al.* (1987) have shown that the 24K protein (or proteins containing the 24K sequence) is also able to cleave one of the glutamine-serine sites in the 200K polyprotein. Furthermore, Franssen *et al.* (1984b) had obtained evidence that the B-RNA-encoded 32K protein was involved in the cleavage at the glutamine-methionine site in the M-RNA-encoded polyprotein. Therefore it was proposed that the 24K protein is probably catalyzing the cleavages at all glutamine-serine and glutamine-glycine sites, while on the other hand the 32K protein would be involved in the cleavage of both glutamine-methionine sites in the CPMV polyproteins. Indeed, serine and glycine are similar amino acids with respect to their side groups (small polar) whereas methionine is very different (large non-polar side group), further supporting the idea that two different proteases would be necessary to cleave all cleavage sites. However, recently Vos *et al.* (1987a) have definitely shown that the 24K protein is able to catalyze all cleavages in the CPMV polyproteins but that for the cleavage of the glutamine-methionine site in the M-polyprotein the 32K protein is essential as a cofactor (see Vos *et al.*, 1987a).

Another question we have addressed in our study on the expression of the CPMV RNAs is the expression of M-RNA *in vivo*. *In vitro*, M-RNA is translated into two carboxy-terminal overlapping polyproteins (105K and 95K) the smaller one as a result of initiation at an internal AUG codon (Vos *et al.*, 1984). These proteins are cleaved by a B-RNA-encoded activity into 58K and 48K proteins and a 60K capsid protein precursor (Franssen *et al.*, 1982). At the other hand in CPMV-infected cells the two capsid proteins VP23 and VP37 are the only M-RNA-encoded proteins readily detectable. To elucidate the expression mechanism of M-RNA *in vivo* we have searched for the 60K capsid precursor and for the 58K and 48K proteins in CPMV-infected cells. Using antibodies against the capsid proteins and $ZnCl_2$ the 60K capsid precursor protein was detected in CPMV-inoculated protoplasts incubated in the presence of $ZnCl_2$ (Chapter 6). Zn^{++} ions are known to inhibit the proteolytic processing of several viral polypro-

teins and probably this has caused the 60K protein to accumulate in these cells.

Using antibodies against synthetic peptides, corresponding to the common carboxy-terminus of the 48K and 58K proteins, a 48K protein was detected in the membrane fractions of infected cells. A viral 58K protein could not be detected (Chapter 6). The presence of the 48K and 60K proteins in infected cells links the in vitro translation results with the in vivo situation and shows that also in vivo CPMV M-RNA is expressed via proteolytic processing of a polyprotein. As sofar direct evidence for the occurrence of a M-RNA-encoded 58K polypeptide in vivo is lacking it remains unknown whether in vivo M-RNA, besides being translated into a 95K polypeptide starting at the initiation codon at position 524, is also expressed by translation, starting at the AUG codon at position 161, into a 105K polypeptide. In vitro the 105K protein is usually produced in considerable smaller amounts than the 95K protein and it seems plausible that this also occurs in vivo and perhaps the amount of 58K protein is very low and remains below the present level of detection. Another possibility is that the 58K protein is unstable and rapidly degraded in infected cells. It is a striking fact that the M-RNAs of all comoviruses studied sofar produce two polyproteins upon in vitro translation (Goldbach and Krijt, 1982), supporting the idea that the presence of two AUG initiation codons in the same reading frame has some functional significance. In our opinion it seems therefore likely that in vivo some 58K protein will be produced.

Previously it was suggested that the M-RNA-encoded 48K (and 58K) protein has a function in cell to cell transport of the virus (Rezelman et al., 1982). We have now shown that the 48K protein is present in the membrane fraction of infected cells and is excreted in the incubation medium of CPMV infected protoplasts (Chapter 6). These findings are consistent with a possible function of the 48K protein in virus transport.

Infection of cells with CPMV is accompanied with the production of membranous vesicles and electron-dense amorphous material. We have demonstrated that the electron-dense material contains (if not represents) non-structural B-RNA-encoded proteins. This was established by immunocytochemical labeling of sections of CPMV infected protoplasts with specific antibodies and protein A-gold (Chapter 7). The data collected correlate well with the observation that during a CPMV infection a considerable amount of non-structural proteins is produced (Goldbach and Van Kammen, 1985). Apparently these proteins are located in

the cytophatic structure as a part of the electron-dense material. The membranous vesicles have been implicated with viral RNA replication (De Zoeten et al., 1974) but the question whether the electron-dense material also has a function in virus multiplication remains to be answered.

With the detection of the 60K, 48K and 24K proteins in infected cells probably all major proteins expressed from the CPMV RNAs have now been identified. It is also clear now that all these proteins are produced by proteolytic processing of two polyproteins between three different pairs of amino acids and that the 24K protein is the protease catalyzing these cleavages. Future research will concentrate on the functions of these proteins. The synthesis of infectious transcripts from complete cDNA clones of both M- and B-RNA as recently described by Vos et al. (1987) opens the way to introduce specific mutations in the RNAs and this will, together with the techniques which have been described in this thesis, form the basis for further research on the elucidation of the replication mechanism of CPMV and its pathogenic action in the host plant.

REFERENCES

De Zoeten, G.A., Assink, A.M. and Van Kammen, A. 1974. Virology 59, 341-355.

Franssen, H., Goldbach, R., Broekhuysen, M., Moerman, M. and Van Kammen, A. 1982. J. Virol. 41, 8-17.

Franssen, H., Goldbach, R. and Van Kammen, A. 1984a. Virus Res. 1, 39-49.

Franssen, H., Moerman, M., Rezelman, G. and Goldbach, R. 1984b. J. Virol. 50, 183-190.

Franssen, H., Roelofsen, W. and Goldbach, R. 1986. Chin. J. Virol. 2, 46-59.

Goldbach, R. and Krijt, J. 1982. J. Virol. 43, 1151-1154.

Goldbach, R. and Van Kammen, A. 1986. Molecular Plant Virology vol. 2, 83-120, CRC Press, Boca Raton, Fla.

Rezelman, G., Franssen, H.J., Goldbach, R.W., Ie, T.S. and Van Kammen, A. 1982. J. Gen. Virol. 60, 335-342.

Rezelman, G., Goldbach, R. and Van Kammen, A. 1980. J. Virol. 36, 366-373.

Tabor, S. and Richardson, C.C. 1985. Proc. Natl. Acad. Sci. USA, 82, 1074-1078.

Verver, J., Goldbach, R., Garcia, J.A. and Vos, P. 1987. EMBO J. 6, 549-554.

Vos, P., Verver, J., Van Wezenbeek, P., Van Kammen, A. and Goldbach, R. 1984. EMBO J. 3, 3049-3053.

Vos, P., Jaegle, M., Wellink, J., Verver, J, Van Kammen, A. and Goldbach, R. 1987a. submitted.

Vos, P., Verver, J., Jaegle, M., Wellink, J., Van Kammen, A. and Goldbach, R. 1987b. submitted.

Zabel, P., Moerman, M., Lomonossoff, G., Shanks, M. and Beyreuther, K. 1984. EMBO J. 3, 1629-1634.

CHAPTER 9

SAMENVATTING

Cowpea mozaïek virus (CPMV) is een virus dat als voornaamste gastheer de cowpea boon (Vigna unguiculata L., in Nederland beter bekend als kouseband) infecteert. De symptomen die op een cowpea plant ontstaan door een infectie met dit virus kunnen afhankelijk van de virusstam variëren van mild mozaïek, hevig mozaïek en bladvervorming tot afstervingsverschijnselen. Het genoom van het virus bestaat uit twee lange RNA moleculen die in een cel direct als boodschapper RNA kunnen dienen. De RNA's zijn elk afzonderlijk verpakt in identieke eiwitmantels en vormen aldus de zg. Bodem (B) en Midden (M) component. Hoewel beide deeltjes noodzakelijk zijn om planten succesvol te infecteren, kan het RNA uit de B component (B-RNA) zich zelfstandig vermenigvuldigen in geïsoleerde cowpea bladcellen (protoplasten). De RNA's bezitten ieder een klein eiwit (VPg) aan het 5' uiteinde, een poly A staart aan het 3' uiteinde, en één enkel open leesraam waaruit zg. poly-eiwitten vertaald worden. Deze poly-eiwitten worden direct, of zelfs tijdens hun synthese door proteolytische klievingen in verscheidene functionele eiwitten geknipt.

Eén van de voornaamste doelen van het in dit proefschrift beschreven onderzoek was meer inzicht te krijgen in het proces van deze proteolytische klievingen. Hierbij ging de meeste aandacht uit naar het 200K poly-eiwit dat door het B-RNA gecodeerd wordt. Als een eerste stap is uitgezocht welke klievingsplaatsen gebruikt worden in dit 200K eiwit. Door het bepalen van gedeeltelijke aminozuurvolgorden van het amino-uiteinde van gezuiverde, door het B-RNA gecodeerde eiwitten, en het vergelijken ervan met de aminozuurvolgorde die afgeleid is van de nucleotidenvolgorde van het B-RNA, werd vastgesteld dat er 3 soorten klievingsplaatsen zijn in het 200K eiwit, namelijk tussen de aminozuurparen glutamine-serine (op twee plaatsen), glutamine-glycine en glutamine-methionine (hoofdstuk 3). Eerder was al vastgesteld dat glutamine-methionine en glutamine-glycine klievingsplaatsen ook voorkomen in het poly-eiwit dat gecodeerd wordt door het M-RNA.

Het klievingsmodel van het 200K eiwit, dat gebaseerd is op analyses van de door het B-RNA gecodeerde eiwitten, voorspelde dat er ook een 24K eiwit gevormd kon worden. Om het 24K eiwit aan te tonen in geïnfecteerde cellen, hebben we antilichamen opgewekt tegen een synthetisch peptide met een aminozuurvolgorde die overeenkomt met een deel van dit hypothetische 24K eiwit. Het voorkomen van het 24K eiwit in CPMV-geïnfecteerde protoplasten kon daadwerkelijk vastgesteld worden met dit antiserum. Het antiserum reageerde ook met de al langer bekende 84K,

110K en 170K eiwitten die allen de 24K sequentie bevatten (hoofdstuk 4).
Op grond van de resultaten van *in vitro* vertalingsexperimenten met B-RNA en op grond van de homologie die het 24K eiwit vertoont met een protease van picornavirussen, was het vermoeden gerezen dat dit 24K eiwit een protease is. De mogelijke proteolytische activiteit ervan werd onderzocht door een cDNA fragment, coderend voor dit eiwit, te koppelen aan een cDNA fragment dat codeerde voor de beide manteleiwitten en dit geheel tot expressie te brengen in de bacterie *Escherichia coli* (hoofdstuk 5). Dit resulteerde in de aanmaak van een aantal virusspecifieke eiwitten die gekarakteriseerd werden met verschillende antisera. Uit de resultaten van een "pulse-chase" experiment bleek dat er twee grote eiwitten werden gesynthetiseerd, waarvan het kleinste waarschijnlijk het resultaat van interne initiatie was. Beide eiwitten bleken vervolgens op twee glutamine-glycine plaatsen gekliefd te worden in kleinere producten, waaronder het kleine manteleiwit VP23. In *E. coli* cellen die een vergelijkbaar construct met een deletie in de sequentie, die codeert voor het 24K eiwit, bevatten werden alleen de twee grote virusspecifieke eiwitten gevormd en vonden geen verdere klievingen plaats. Deze resultaten toonden aan dat het 24K eiwit verantwoordelijk is voor beide glutamine-glycine klievingen.
Tegelijkertijd werd in ons laboratorium aangetoond dat het 24K eiwit (of 24K bevattende eiwitten) ook verantwoordelijk is voor één van de glutamine-serine klievingen in het 200K eiwit. Verder zijn er al eerder proeven beschreven waaruit bleek dat het 32K eiwit, dat gecodeerd wordt door het B-RNA, betrokken is bij de klieving van een glutamine-methionine dipeptide sequentie in het poly-eiwit dat gecodeerd wordt door het M-RNA. Al deze resultaten maakten het waarschijnlijk dat het 24K eiwit verantwoordelijk was voor alle glutamine-glycine en glutamine-serine klievingen en het 32K eiwit voor de beide glutamine-methionine klievingen. Omdat serine en glycine aminozuren zijn met een kleine polaire zijgroep terwijl methionine een grote apolaire zijgroep bezit, was het inderdaad niet onaannemelijk dat twee verschillende proteases betrokken zouden zijn bij de diverse klievingen. Deze aanvankelijke veronderstelling bleek echter een misvatting te zijn omdat later is aangetoond dat het 24K eiwit (en/of 24K bevattende eiwitten) verantwoordelijk is voor alle klievingen in de M- en B-poly-eiwitten (P. Vos, J. Verver, M. Jaegle, J. Wellink, A. van Kammen en R. Goldbach, manuscript in voorbereiding). Alleen bij de glutamine-methionine klieving in het M-poly-eiwit is ook het 32K eiwit betrokken, echter niet als protease maar als cofactor.

Een ander interessant probleem bij de expressie van de CPMV-RNAs is de expressie van M-RNA *in vivo*. De nucleotidenvolgorde van het M-RNA bevat één lang open leesraam, maar *in vitro* wordt het M-RNA vertaald in twee eiwitten, van 105K en 95K, die alleen verschillen in hun amino-terminale uiteinden. Het kleinste eiwit is het resultaat van initiatie van de vertaling op een intern AUG codon. Beide eiwitten worden gekliefd in 58K en 48K eiwitten en een 60K precursor voor de beide manteleiwitten. In geïnfecteerde cellen zijn echter de beide manteleiwitten de enig bekende eiwitten die door het M-RNA gecodeerd zijn. Dit heeft ons er toe gezet om op zoek te gaan naar de 60K, 58K en 48K eiwitten in geïnfecteerde cellen. Het bleek mogelijk om met behulp van antiserum tegen de manteleiwitten de 60K precursor voor de beide manteleiwitten ook *in vivo* te detecteren door de geïnfecteerde protoplasten te behandelden met $ZnCl_2$ (hoofdstuk 6). Zink-ionen staan bekend als remmers van de klieving van virale poly-eiwitten en hierdoor kan de 60K precursor zich kennelijk ophopen. Door gebruik te maken van antilichamen gericht tegen synthetische peptiden, corresponderend met het gemeenschappelijke carboxyl-uiteinde van de 58K en 48K eiwitten, was het mogelijk het 48K eiwit te detecteren in de membraanfractie van CPMV-geïnfecteerde cellen. Een 58K eiwit werd niet aangetroffen. Met de detectie van de 48K en 60K eiwitten in de geïnfecteerde cel zijn nu de resultaten van de *in vitro* vertalings-experimenten grotendeels in overeenstemming te brengen met waarnemingen *in vivo* en is aangetoond dat het CPMV M-RNA ook in de geïnfecteerde cel via proteolytische klievingen van een poly-eiwit tot expressie komt (hoofdstuk 6).

Omdat tot nu toe het voorkomen van een door het M-RNA gecodeerd 58K eiwit in geïnfecteerde cellen niet aangetoond kan worden, is het nog steeds een open vraag of het M-RNA *in vivo* al dan niet vertaald wordt in een 105K eiwit. In *in vitro* vertalingen wordt meestal minder 105K dan 95K eiwit gemaakt en als dit ook *in vivo* het geval is dan zal dit product met deze methode niet detecteerbaar zijn. Het is daarnaast ook mogelijk dat het 58K eiwit minder stabiel is dan het 48K eiwit en snel wordt afgebroken. Het is opvallend dat de M-RNA's van alle tot nu toe onderzochte comovirussen *in vitro* vertaald worden in twee poly-eiwitten. Dit is een aanwijzing dat het gebruik van twee startcodons in hetzelfde leesraam een functionele betekenis heeft en het is dan ook waarschijnlijk dat er *in vivo* 58K eiwit wordt gemaakt.

Uit eerdere experimenten is geconcludeerd dat de door het M-RNA gecodeerde 48K (en 58K) eiwit mogelijk een functie heeft bij het transport van het virus (of virale RNA) van cel naar cel. We hebben aangetoond dat het 48K eiwit voor-

namelijk aangetroffen wordt in de membraanfractie van de geïnfecteerde cel en dat dit eiwit uitgescheiden wordt in het voedingsmedium van protoplasten geïnfecteerd met CPMV (hoofdstuk 6). Deze waarnemingen zijn in overeenstemming met het idee dat het 48K eiwit betrokken is met het transport van het virus.

Verschillende virussen induceren zg. cytopathologische structuren in de cellen die ze hebben geïnfecteerd. Zo veroorzaakt een CPMV infectie de productie van vesiculaire membraanzakjes (vesicles) en electronendicht, amorf materiaal. De electronendichte structuren blijken nu niet-structurele door het B-RNA gecodeerde eiwitten te bevatten. Dit is met behulp van elektronenmicroscopie vastgesteld door coupes van met CPMV geïnfecteerde protoplasten te incuberen met specifieke antilichamen en deze vervolgens te voorzien van goudbolletjes gekoppeld aan proteïne A (hoofdstuk 7). Gedurende een CPMV infectie worden vrij grote hoeveelheden niet-structurele eiwitten geproduceerd en het is dus zeer goed mogelijk dat deze eiwitten zich ophopen in deze electronendichte structuren. Verder onderzoek is nodig om te bepalen of deze electronendichte structuren een functie hebben in de virusvermenigvuldiging, of dat deze structuren het gevolg zijn van ophopen van "verbruikte" en/of niet-functionele virale eiwitten.

CURRICULUM VITAE

Joan Wellink werd op 18 september 1959 te Oldenzaal geboren. In 1977 behaalde hij het diploma VWO-B aan het Carmel Lyceum te Oldenzaal. In hetzelfde jaar begon hij te studeren aan de Landbouwuniversiteit te Wageningen waar hij in januari 1984 slaagde voor het doctoraalexamen Moleculaire Wetenschappen met als hoofdvakken Moleculaire Biologie en Celbiologie. Van 1 februari 1984 tot 1 februari 1987 was hij in dienst van de Landbouwuniversiteit te Wageningen en werkzaam op de vakgroep Moleculaire Biologie waar hij onderzoek verrichtte aan de expressie van cowpea mosaic virus zoals in dit proefschrift beschreven wordt. Sinds 1 februari 1987 is hij in tijdelijke dienst van de Landbouwuniversiteit verbonden aan dezelfde vakgroep.