The primary structure of the 24K protease from red clover mottle virus: implications for the mode of action of comovirus proteases

M. Shanks and G. P. Lomonossoff*

Department of Virus Research, John Innes Institute and AFRC Institute of Plant Science Research, Colney Lane, Norwich NR4 7UH, U.K.

We have determined the nucleotide sequence of the region of red clover mottle virus (RCMV) bottom component RNA which encodes the RCMV equivalent of the cowpea mosaic virus (CPMV) 24K protease. From the alignment of the deduced amino acid sequence of the RCMV 24K protein with that of the homologous protein from CPMV, we speculate on the relative importance of the various amino acid residues which have been implicated in the catalytic mechanism of comovirus proteases.

Comoviruses are a group of bipartite, positive-strand RNA plant viruses which express their genomes through the synthesis and subsequent cleavage of large polyproteins by a virus-encoded protease (for a review, see Goldbach & van Kammen, 1985). In the case of the type member of the group, cowpea mosaic virus (CPMV), the enzyme responsible for most, if not all, the cleavages has been identified as a 24K protein derived from the larger, bottom component (B) RNA (Franssen et al., 1984a; Verver et al., 1987; Vos et al., 1988). A combination of nucleotide and protein sequencing (Lomonossoff & Shanks, 1983; Wellink et al., 1986) has allowed the amino acid sequence of this protein to be deduced.

A striking property of comovirus proteases is their specificity, since they recognize only a very limited number of cleavage sites on the substrate polypeptides. In addition, it has been demonstrated that the proteolytic activity associated with one comovirus can cleave only the middle component (M) RNA-encoded proteins from the homologous virus (Gabriel et al., 1982; Goldbach & Krijt, 1982). However, it is possible that the specificity for this reaction may lie with the B RNA-encoded 32K protein which, though not possessing proteolytic activity itself, appears to have a role in the cleavage of the M RNA-encoded polyproteins by the 24K protease (Vos et al., 1988).

Amino acid sequence alignments suggest that comovirus proteases are homologous to the 3C proteases of picornaviruses (Franssen et al., 1984b; Argos et al., 1984). Studies with inhibitors have provided evidence that both sets of proteases have cysteine residues at their active sites (for a review, see Kräusslich & Wimmer, 1988). Despite this, Bazan & Fletterick (1988) and Gorbunova et al. (1989) have suggested that both the comovirus 24K and the picornavirus 3C proteases are related to the trypsin-like family of serine proteases rather than the cellular thiol proteases. Both papers attempted to identify the active site residues of comovirus and picornavirus proteases by alignment of their amino acid sequences with those of serine proteases. In the case of the CPMV 24K protease, the two papers arrived at different conclusions regarding the active site residues. In an attempt to resolve the differences in the alignments and to provide further insight into the mechanism and specificity of comovirus proteases, we have determined the nucleotide sequence of the region of red clover mottle virus (RCMV) B RNA encoding the RCMV equivalent of the CPMV 24K protease and compared the amino acid sequence of the encoded protein with that from CPMV.

RCMV was propagated and purified as previously described (Shanks et al., 1986). B components were isolated by centrifugation once in Nycodenz gradients and the RNA was extracted as described by Zimmern (1975). First-strand cDNA was synthesized using p(dT)$_{12-18}$ as a primer and converted into a double-stranded form exactly as previously described (Shanks et al., 1986). The double-stranded cDNA was digested with a variety of restriction enzymes and the resulting fragments were cloned into appropriately linearized bacteriophage M13 vectors. Sequence analysis and restriction mapping of a number of clones established that one clone, Taq-7, contained a 1.4 kb TaqI fragment in the AccI site of M13mp701 (D. R. Bentley, unpublished), the right-hand end of which corresponded to a site 1.5 kb from the 3' end of B RNA (Fig. 1). By comparison with the genetic map
clones derived from both Taq-7 and Sma/Pst-2. The protease was sequenced in both directions using sub-clones occurring in Taq-7. The region encompassing the 24K protease was sequenced in both directions using sub-clones occurring in Taq-7. The region encompassing the 24K protease from its precursor polyprotein. The 3' terminal 666 nucleotides of this fragment also identify the probable cleavage sites used to release the RCMV and CPMV proteins (see below) we could sequence from the equivalent region of CPMV B RNA. Because of the high degree of homology between the two proteins is shown in Fig. 3. It is clear that the two proteins are very similar, though the degree of homology does vary markedly with position. For example, in the region between residues 85 and 118, the proteins are 82% identical, a figure dropping to only 32% when the next 34 residues (119 to 152) are compared. The region of low homology coincides approximately with the part of the polypeptide chain which would link the two six-stranded β-barrel domains in a trypsin-like protease. This observation is consistent with the idea that comovirus 24K proteases may have a two-domain structure similar to that of serine proteases (Bazan & Fletterick, 1988). Comparison of the amino acid sequences of the VPg of RCMV (Fig. 2) with that of CPMV shows that both are 28 amino acids long and have 21 amino acids in common, including the N-terminal serine which is involved in the linkage of the protein to the RNA in CPMV (Jaegle et al., 1987).

Although both Bazan & Fletterick (1988) and Gorbalenya et al. (1989) agreed that Cys-166 is part of the active site of CPMV 24K protease, they identified different residues as the other two members of the catalytic triad characteristic of serine proteases [for a review of the mechanism of action of serine proteases, see Kraut (1977)]. Bazan & Fletterick (1988) suggest His-44 and Asp-93 fulfil this role but Gorbalenya et al. (1989) identify His-40 and Gly-75 as the critical residues. The alignment shown in Fig. 3 shows that, whereas His-40 is conserved in the RCMV 24K protein, His-44 becomes an Asn residue. The fact that His-40 is conserved in the CPMV and RCMV proteases suggests that His-40 is more likely to be an active site residue than His-44, a fact which favours the alignment of Gorbalenya et al. (1989), at least in the N-terminal part of the protein. Since both Glu-75 and Asp-93 are conserved in RCMV, the alignment in Fig. 3 cannot distinguish between them as proposed active site residues. The alignments of the CPMV 24K protease with the trypsin-like proteases by Bazan & Fletterick (1988) and Gorbalenya et al. (1989) differ considerably in this region, a problem exacerbated by the occurrence of a repeated three amino acid sequence, Cys-Trp-Asp, in the CPMV sequence. Though the data obtained from the alignment in Fig. 3 are useful in suggesting the identity of active site residues, the unambiguous assignment of residues to the active site of

Fig. 1. Schematic representation of the M13 clones used to obtain the nucleotide sequence of the region of RCMV B RNA which encodes the 24K protease. The top line represents a genetic map of CPMV B RNA, the 24K protein being the protease. The box at the 5' end of the RNA represents the covalently linked VPg. The inserts in the two clones used, Taq-7 and Sma/Pst-2, were aligned against the genetic map of CPMV B RNA by restriction mapping and sequence analysis.

of CPMV (Wellink et al., 1986), it was reasoned that this clone should encompass at least part of the region of RCMV B RNA which encodes the 24K protease.

A second, independent clone was isolated using d(AATATCGTTGCCTTGACT) as a primer, this sequence being complementary to a region of RCMV B RNA approximately 1.5 kb from the 3' terminus. Double-stranded cDNA synthesis was carried out as described above, the DNA was digested with PstI and the fragments were cloned into PstI/SmaI-digested M13mp19. One recombinant (Sma/Pst-2) was further analysed and found to contain a 2.8 kb insert, the 3' end of which contained the sequence of the primer. Inserts from both Taq-7 and Sma/Pst-2 were excised from the double-stranded replicative forms of the M13 clones, digested with several different restriction enzymes either singly or in pairs and subcloned into M13 for sequence analysis by the dideoxynucleotide method (Sanger et al., 1977). It proved relatively easy to identify the region encoding the RCMV 24K protease by comparison with the sequence from the equivalent region of CPMV B RNA. Because of the high degree of homology between the RCMV and CPMV proteins (see below) we could also identify the probable cleavage sites used to release the 24K protease from its precursor polyprotein. The entire region encoding the RCMV 24K protease is in a 771 bp PvulI fragment contained within Pst/Sma-2, the 3' terminal 666 nucleotides of this fragment also occurring in Taq-7. The region encompassing the 24K protease was sequenced in both directions using sub-clones derived from both Taq-7 and Sma/Pst-2. The nucleotide sequence, together with the derived protein sequence is shown in Fig. 2. The PvulI fragment contained not only the region encoding for the RCMV 24K protease but also the region encoding the genome-linked protein, VPg.

Assuming the cleavage sites have been correctly identified, the RCMV 24K protease consists of 208 residues, exactly the same number as its CPMV equivalent. An alignment of the RCMV 24K protease with the trypsin-like proteases by Bazan & Fletterick (1987). Comparison of the amino acid sequences of the VPg of RCMV (Fig. 2) with that of CPMV shows that both are 28 amino acids long and have 21 amino acids in common, including the N-terminal serine which is involved in the linkage of the protein to the RNA in CPMV (Jaegle et al., 1987).
The comovirus 24K protease will ultimately depend on a mutational analysis of such residues.

In addition to identifying putative active site residues, Bazan & Fletterick (1988) also identify seven residues as potentially being part of a substrate-binding pocket (see Fig. 3). All seven residues identified in CPMV protease are conserved in the RCMV sequence. If the residues involved in substrate binding have been correctly identified, this observation implies that the specificity of comovirus proteases for their homologous substrate does not depend on the enzymes recognizing a target sequence of amino acids near the cleavage sites. This idea is supported by the observation that the sequences around the cleavage sites vary as much within a given comovirus polyprotein as between the polyproteins of different viruses (Wellink et al., 1986; Shanks et al., 1986; Chen et al., 1989; this communication). It is therefore probable that the specificity of the comovirus proteases depends on some three-dimensional structural feature of the substrates. Such features have recently been shown to be important for correct processing of picornavirus polyproteins by the 3C proteases (Ympa-Wong et al., 1988; Dewalt et al., 1989).

We thank Professor J. Wooton of Leeds University for interesting discussions about the protein alignments, Drs J. W. Davies and A. J. Maule for critical reading of the manuscript and M. Harvey for help in the preparation of the figures.
Short communication

References


(Received 8 August 1989; Accepted 3 November 1989)