The primary structure of the virion protein gene and encoded protein of erysimum latent tymovirus

Pattana Srifah,1 Paul Keese,2 Dharma Shukla3 and Adrian Gibbs1*

1Research School of Biological Sciences, Australian National University, P.O. Box 475, Canberra, A.C.T. 2601, 2Division of Plant Industry, C.S.I.R.O., Canberra, A.C.T. 2601 and 3Division of Biomolecular Engineering, C.S.I.R.O., 343 Royal Parade, Parkville, Victoria 3052, Australia

The nucleotide sequence of the virion protein (VP) gene of erysimum latent tymovirus (ELV) has been determined and the amino acid sequence of the VP deduced and confirmed by peptide analysis. The ELV VP is larger than the VPs of other tymoviruses because it has, unexpectedly, 11 more amino acid residues at its N terminus. The amino acid sequences of the VPs of ELV and four other tymoviruses align unequivocally and their relationships, as assessed from the percentage of identical residues, correlate well with previously reported serological tests which have shown ELV to be distant from other tymoviruses.

Erysimum latent tymovirus (ELV) (Shukla & Schmelzer, 1972) causes diseases of wild and crop brassica plants in Europe and has been isolated there from naturally infected plants of several Erysimum species and Barbarea vulgaris. Its properties, described below, have been well studied (Shukla & Gough, 1980; Shukla et al., 1980) and show it to be a member of the tymovirus group. However, although its host range and symptoms are closely similar to those of turnip yellow mosaic virus (TYMV), its particles were found to be serologically unrelated to those of TYMV and only distantly related to those of three other tymoviruses, i.e. okra mosaic, Andean potato latent (APLV) and ononis yellow mosaic (OYMV) viruses (Shukla & Gough, 1980; Shukla et al., 1980).

The viroins of tymoviruses are isometric and about 28 nm in diameter; their structure is currently being studied by X-ray crystallography methods (J. Varghese, personal communication). They sediment as two components that are serologically indistinguishable; those with a sedimentation coefficient of approximately 120 s contain the ssRNA genome, and those which sediment at approximately 60 s are genome-free. The ELV genome, like that of other tymoviruses, is about 6000 nucleotides long and has three main open reading frames (ORFs) (P. Srifah, unpublished data) of which the 3' proximal one encodes the virion protein (VP). The ELV VP is a single species and was estimated by SDS-PAGE to have an Mr of approximately 21700 (Gough et al., 1982) and, by FITMOL analysis of its amino acid composition, to contain 208 amino acid residues (Shukla et al., 1980). Thus, these indirect estimates of the size of the VP of ELV indicate that it may be somewhat larger than those of most other tymoviruses, which are between 188 and 192 residues long.

In this paper we report the nucleotide sequence of the ELV VP gene and the deduced amino acid sequence of the encoded protein, which we have confirmed by direct, but partial, amino acid sequencing. We also report comparisons of these sequences with those of the VP genes and encoded proteins of other tymoviruses, namely eggplant mosaic virus (EMV; Osorio-Keese et al., 1989), kennedya yellow mosaic virus (KYMV; Ding et al., 1990), OYMV (Ding et al., 1989) and TYMV (Keese et al., 1989).

For this study we used the ELV isolate described by Shukla & Gough (1980) and propagated it in Brassica campestris ssp. pekinensis L. Three weeks after plants were inoculated, the virions were extracted and purified as described by Shukla et al. (1973), using an extraction buffer of 100 mM-Na2HPO4 and 50 mM-ascorbic acid pH 7-0. The final preparation was suspended in 0·1 × SSC. Genomic RNA was extracted from purified virions as described by Blok et al. (1987). This RNA was polyadenylated, transcribed into cDNA, hydrolysed into fragments of a convenient size, ligated into M13 phage and used to transform Escherichia coli as described by Keese et al. (1989).

The cloned DNAs were sequenced (Sanger et al., 1980) using Sequenase, as advised by the suppliers (United States Biochemical). Their sequences were compiled by the Staden (1982) library of computer programs for shotgun sequencing and analysed using the SEQ program package of the Research School of Biological Sciences, Australian National University, Canberra,
Fig. 1. Nucleotide sequence of the VP gene of ELV and amino acid sequence of the encoded VP. Bold letters and underlined type show the peptides whose sequences were chemically determined.

Australia. The compiled 3'-terminal sequence has a single ORF which encodes a protein that shows clear sequence similarities with the VPs of other tymoviruses. The nucleotide sequence of the ORF and the amino acid sequence it encodes are shown in Fig. 1. It was compiled from the overlapping sequences of 53 clones; each part of the sequence was obtained from between three and nine clones, of which about half were in each orientation. Some of the sequences differ at single nucleotide positions but none of these differences have been confirmed by sequencing additional clones; they are therefore likely to represent infrequent point mutations of the viral genome or to have resulted from transcriptional errors during cDNA synthesis.

Fig. 2 shows the sequences of the VPs of ELV and four other tymoviruses aligned by the computer program of Feng & Doolittle (1987). It can be seen that the VPs align unequivocally over most of their length but that the N terminus of the ELV VP has an extra 11 amino acid residues, which are not present in the other VPs. This is unexpected for the following reasons. The VP of ELV probably has an eight-stranded anti-parallel $\beta$-barrel
structure (J. Varghese, personal communication) like the VPs of most other simple RNA-genome viruses with isometric virions. Within such virions the N-terminal portions of the VPs form a network on the inner surface of the virion shell. This network interacts with the encapsidated genomic RNA and seems to direct and stabilize virion assembly (Carrington et al., 1987; Hogle et al., 1990). Thus, one would expect closely related viruses, such as different members of the tymovirus group, to have VPs of closely similar size and structure, especially in their N-terminal regions.

In order to check that the N-terminal part of the deduced VP sequence is expressed in vivo, major peptides from the ELV VP were sequenced directly. The virion protein was isolated (Shukla et al., 1980) and digested with either chymotrypsin, trypsin, Staphylococcus aureus protease V8 or cyanogen bromide. The resulting peptides were separated using reversed-phase HPLC and sequenced both manually and in an Applied Biosystems Model 470A protein sequencer (Shukla et al., 1986, 1987); the sequenced peptides are shown in bold letters and underlined type in Fig. 1. This analysis directly confirms the presence of the unusually long N-terminal region of the native protein and also confirms earlier indications that the VP of ELV is larger than that of other tymoviruses. The reason, if any, for these differences will probably only become evident when the virion structures have been determined.

The nucleotide composition of the VP gene is given in Table 1. It is close to that reported for the whole genome (Shukla et al., 1980) and, like that of other tymoviruses (Keese et al., 1989), has an unusually large cytosine content. Thus, as might be expected, the third codon position is dominated by cytosines (44%).

The composition of the encoded amino acid sequence of the ELV VP gene is given in Table 2. The amino acid composition of the VP is also close to that reported by Shukla et al. (1980), except that they reported a larger glycine content (8.4% compared with 4.95%) which is perhaps an arithmetic error as the total percentage composition they report is 103.1%. The composition of ELV VP, like that of the TYMV-CL VP (Keese et al., 1989), is distinctive and, although rich in amino acids with cytosine-dominated codons such as leucine, proline and serine and correspondingly short of those with purine-dominated codons, differs greatly from the composition obtained by repeatedly randomizing and translating the VP gene (Table 2).

The relationships of the tymovirus VPs whose sequences have been reported were assessed by determining the percentage of non-identical residues in each pair of aligned sequences. Fig. 3 shows a dendrogram computed from these differences by the ‘neighbour-joining’ method (Saitou & Nei, 1987; Studier & Keppler, 1985).
1988). It can be seen that the relationships of the tymovirus VPs illustrated in this way closely correlate with their reported serological relationships (Koenig, 1976; Shukla & Schmelzer, 1972; Shukla & Gough, 1980); TYMV and KYMV are most closely related, EMV and OYMV are also related though more distantly. ELV is serologically the most distinct of all tymoviruses examined so far, having only a very distant serological relationship with OYMV and APLV (Shukla et al., 1980). The tymovirus relationships also correlate well with their relationships as assessed from the amino acid composition of their VPs (Paul et al., 1980).

References


(Received 1 June 1990; Accepted 28 August 1990)