Acta virol. **37**: 523-528, 1993

NUCLEOTIDE SEQUENCE OF THE 7 K GENE OF *HELENIUM* VIRUS S

R. L. TURNER[1], P. R. MILLS[2], G. D. FOSTER[1]*

[1]Botany Department, University of Leicester, Leicester LE1 7RH; and [2]Plant Pathology, The Queens University of Belfast, Newforge Lane, Belfast BP9 5PX, U. K.

Received September 15, 1992; revised February 19, 1993

Summary. - The sequence of 380 nucleotides upstream of the start codon of the open reading frame (ORF) of the coat protein of the carlavirus *Helenium* virus (HelVS) was determined from cloned cDNA. This portion of the viral RNA contained one complete ORF encoding a protein of 7 K which had homology with similar sized proteins from other members of the carla- and potexvirus groups. Sequence data was also obtained beyond the start codon of the 7 K ORF into what was presumed to be the 12 K protein gene of the previously identified triple gene block of carla- and potexviruses. Significant sequence similarity of 46.3 % and 41.6 % was evident at the amino acid level with the equivalent regions of the 12 K proteins of potato virus S, Andean strain (PVS^A) and potato virus M (PVM), respectively, with 25.4 % similarity detected with the equivalent region of potato virus X (PVX).

Key words: HelVS; carlavirus; 7 K gene; nucleotide sequence

HelVS belongs to the carlavirus group of plant viruses (Wetter and Milne, 1981). It has been shown to have a single stranded positive sense genomic RNA molecule, 7.4 kb in size, which is encapsidated in a 31 K coat protein to produce flexuous rod-shaped particles (Kuschi *et al.*, 1978). As with some other members of the carlavirus group, in addition to full genomic 650 nm viral particles smaller fragments, 320 nm and 180 nm size have been detected at a low concentration (Foster and Mills, 1990*a*). It is thought that these smaller fragments are encapsidated subgenomic RNAs, the larger containing a 3.3 kb and the smaller containing a 1.5 kb RNA molecule. The 1.5 kb RNA molecule is an efficient message for coat protein synthesis *in vitro* (Foster and Mills, 1990*a*). The 3' coding region of several members of the group such as PVM (Rupasov *et al.*, 1989), PVS^A (MacKenzie *et al.*, 1989), and lily symptomless virus (LSV) (Memelink *et al.*, 1990), have been recently determined revealing a triple gene protein block of 25 K, 12 K and 7 K upstream from the coat protein open reading frame. A similar gene block has been identified in the potexvirus PVX (Huisman *et al.*, 1988).

*Whom requests for reprints should be addressed to

This short communication will analyse the nucleotide sequence immediately upstream from the HelVS coat protein gene which has previously been reported (Foster *et al.*, 1990), compare the deduced amino acid sequence of these regions with those of carla- and potexviruses, and, finally, will raise some questions regarding the mode of expression of the viral subgenomic RNAs, in particular coat protein subgenomic RNAs.

Two clones, pHEL19 and pHEL23, were available for analysis, both of which were constructed following double stranded cDNA synthesis of HelVS RNA and ligation into *Sma*I digested vector pUC13. The two clones were selected by immunoscreening using HelVS polyclonal antiserum for their ability to express the HelVS coat protein presumably fused to the alpha peptide of the *lac*Z gene after induction of the *lac*I promoter using IPTG. Restriction enzyme analysis revealed insert sizes of 1.4 kb and 1.6 kb for pHEL19 and pHEL23, respectively. It was determined that the difference in nucleotide length was at the 5' end due to the detection of a polyA tail in both clones (Foster and Mills, 1992). Both insert fragments were subcloned from pUC13 into *Sac*I, *Hind*III digested vector pBluescript SK+ (Stratagene) for sequence analysis and generation of RNA transcripts. When the pHEL19 clone was sequenced it was shown to contain the entire HelVS coat protein ORF with a downstream 12.6 K protein gene (Foster *et al.*, 1990), but one aspect of the clone was of interest in respect to its expression after its induction with IPTG. Between the initiating methionine of the viral coat protein and the initiating methionine of the *lac*Z of the vector two stop codons could be detected. Indeed, stop codons were present in all three reading frames between the two methionines thus raising the question as to how the coat protein ORF was being efficiently expressed in *E. coli* as it was clearly not using the Shine-Dalgarno sequence and initiating methionine of the *lac*Z protein. When the nucleotide sequence upstream from the coat protein was examined more closely a potential ribosome binding site (RBS) was identified based around the sequence CCTTTAGGTTTA, as previously reported by Foster and Mills (1991). This RBS present in pHEL19 may be binding to the 3'-terminal region of the 16 S rRNA of *E. coli* allowing subsequent initiation and elongation from the same methionine of the coat protein ORF. In the case of pHEL23 the evidence for internal ribosome binding was perhaps stronger as pHel23 contained more sequence extended 5' to the ATG of the coat protein ORF. Indeed, enough sequence was present to encode an entire ORF, the 7 K ORF which has previously been identified upstream from the coat protein ORF.

In order to obtain the nucleotide sequence to the 5' upstream region of the coat protein ORF the pHEL23 clone was sequenced using the dideoxynucleotide termination method of Sanger *et al.* (1977), and compiled using the computer program DNASIS (Pharmacia). Sequence data obtained extended about 380 nucleotides upstream from the ATG of the coat protein gene (Fig. 1). Within this region one complete ORF was identified to the 5'region of the coat protein gene, consisting of 183 nucleotides, with the deduced peptide containing 61 amino acids, equivalent to a molecular mass of 6723.64 (7 K).

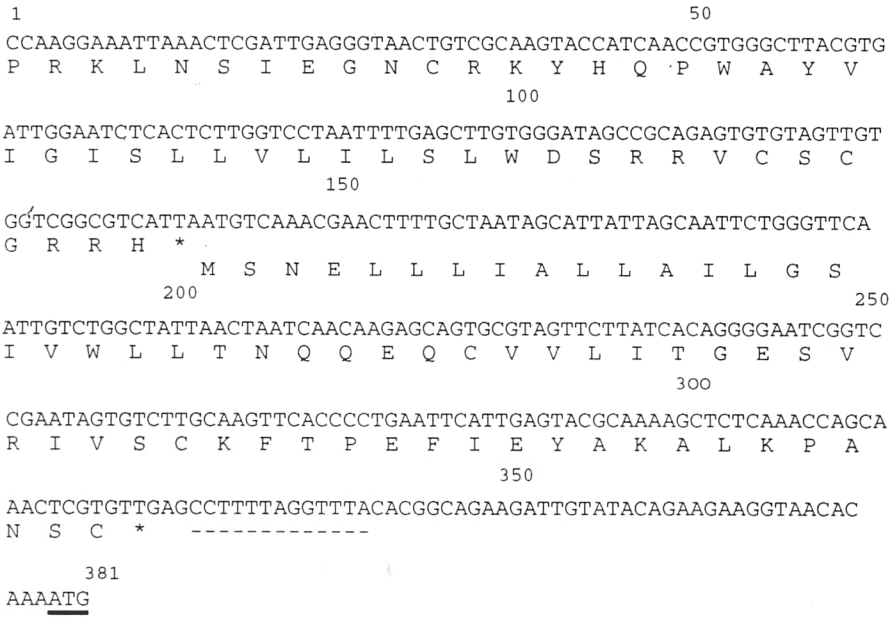

Fig. 1
The nucleotide sequence corresponding to the 7 K gene and upstream partial 12 K gene of HelVS The DNA sequence is shown as the equivalent of the viral plus strand and the amino acid sequence is presented below the nucleotide sequence. Termination codons are indicated as (*). The initiation codon of the HelVS coat protein is underlined with a bold line. Nucleotides thought to be involved in base pairing in HelVS are underlined with a hatched line.

The predicted amino acid sequence of the HelVS 7 K peptide was compared with equivalent peptides of other carla-and potexviruses. Amino acid homologies are given in Fig. 2, with 30.8 % and 31.8 % similarity at the amino acid level with the 7 K peptides of the carlaviruses PVS^{A} and PVM, respectively (MacKenzie *et al.*, 1989; Rupasov *et al.*, 1989). A slightly lower though a significant homology of 28.6 % was detected with the equivalent peptide of the potexvirus, PVX (Huisman *et al.*, 1988).

Sequence data obtained extended beyond the start codon of the 7 K ORF into what was presumed to be the 12 K protein gene of the previously identified triple gene block of carla- and potexviruses. This sequence extended 141 nucleotides to the 5' side of the 7 K gene ATG start codon with a deduced partial peptide containing 47 amino acids presumably corresponding to the C-terminal region of the HelVS 12 K protein. This sequence was compared with the predicted C-terminal regions of equivalent peptides from other carla- and potexviruses (Fig. 3). Significant sequence similarity of 46.3 % and 41.6 % was evident at the

amino acid level with the equivalent regions of the 12 K proteins of PVS[A] and PVM, respectively, with 25.4 % similarity detected with the equivalent region of PVX.

Whilst the function of these proteins remains unknown at this stage, all 7 K and 8 K proteins that have been sequenced to date contain blocks of hydrophobic amino acids common to membrane spanning domains. It has been suggested that these domains may interact with the plasmodesmata thereby facilitating cell-to-cell movement. The 7 K protein may act individually or as a complex with other proteins, with a high degree of variation at the N- and C-termini implicated in host specificity.

Analysis of these clones revealed that the coat protein of HelVS was not only being produced in a prokaryotic system from a clone with stop codons between the ATG of the coat protein and the ATG of the vector in the case of pHEL19, but also from a clone with a complete ORF between these two initiating methionines in pHEL23. This findings suggested that some form of internal ribosome entry was occuring in the upstream region from the coat protein ORF, at least in a prokaryotic system. To test whether coat protein could be produced from this internal ORF from this clone in an eukaryotic system, the insert was

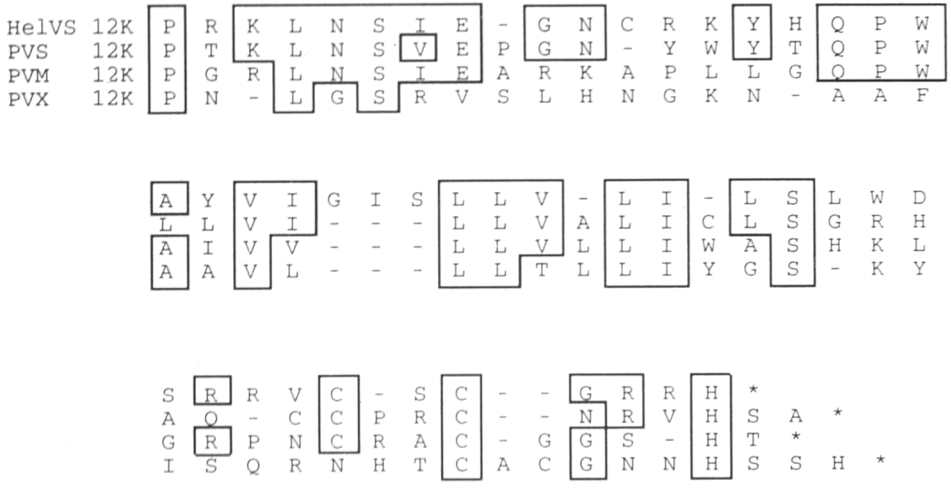

Fig. 2
Alignment of a portion of the amino acid sequence of the HelVS ORF upstream of the 7 K protein with the 12 K sized proteins of PVS, PVM, PVX
Gaps (-) have been introduced for maximum alignment.

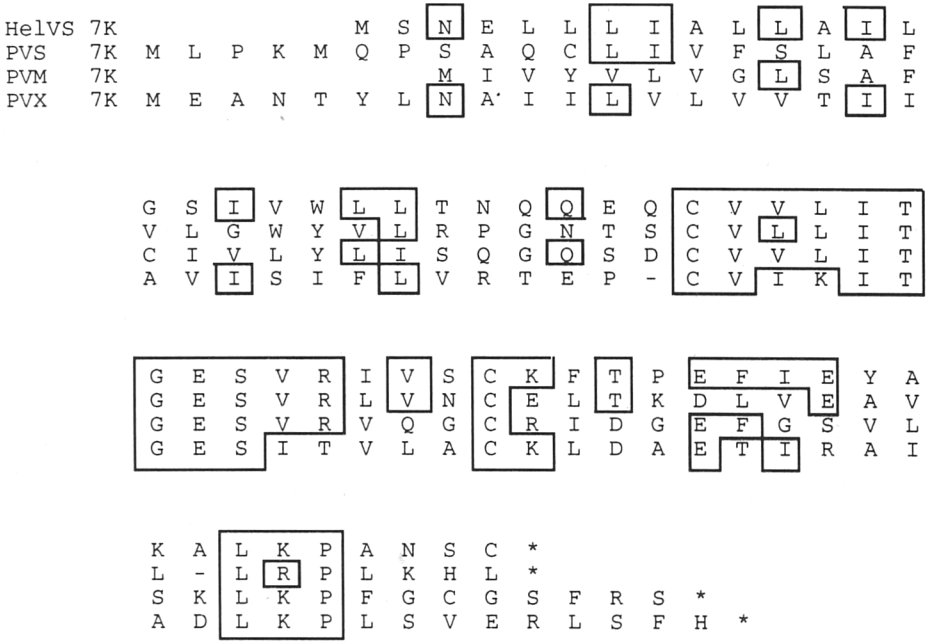

Fig. 3
Alignment of the 7 K protein amino acid sequence of HelVS with the equivalent ORFs from PVS, PVM and PVX

subcloned into Bluescript (Stratagene), allowing *in vitro* generated RNA transcripts from the clone to be produced, as previously described by Foster and Mills (1990*b*). When RNA transcripts were generated from pHEL19 and pHEL23 and translated *in vitro* both in wheat germ and rabbit reticulocyte lysate systems, high levels of coat protein were equally produced from both clones (results not presented).

It has been observed that the immediate upstream non-coding region of the HelVS coat protein gene has a canonical Shine-Dalgarno sequence (Foster and Mills, 1991). Such sequences are thought to be involved in prokaryotic ribosome binding due to base pair complementarity with a pyrimidine rich sequence near the 3' end of the 16 S rRNA. These conserved homologous blocks have been observed throughout the carlavirus group upstream from the coat protein and triple gene block ORFs (Foster and Mills, 1991). Nucleotides thought to be involved in base pairing in HelVS are underlined with a hatched line in Fig. 1. At present it is unclear why prokaryotic RBS should be present upstream from the ATG of a viral RNA from an eukaryotic host, but *in vivo* prokaryotic and *in vitro*

eukaryotic expression results suggest that some form of internal ribosome binding and subsequent initiation of translation is occurring in the upstream non-coding region of the carlavirus coat protein.

References

Foster, G. D., and Mills, P. R. (1990*a*): Evidence for the role of subgenomic RNA species in the production of *Helenium* virus S coat protein during *in vitro* translation. *Virus Res.* **17**, 61–70.

Foster, G. D., and Mills, P. R. (1990*b*): Investigation of the 5' terminal structures of genomic and subgenomic RNAs of potato virus S. *Virus Genes* **4**, 359–366.

Foster, G. D., Millar, A. W., Meehan, B. M., and Mills, P. R. (1990): Nucleotide sequence of the 3'-terminal region of *Helenium* virus S RNA. *J. gen. Virol.* **71**, 1877–1880.

Foster, G. D., and Mills, P. R. (1991): Occurrence of chloroplast ribosome recognition sites within conserved elements of the RNA genomes of carlaviruses. *FEBS Lett.* **280**, 341–343.

Foster, G. D., Scott, R., Draper, J., and Mills, P. R. (1992): Expression of *Helenium* virus S coat protein in *Escherichia coli, in vitro* in rabbit reticulocyte lysate and transgenic tobacco. *Acta virol.* **36**, 567–575.

Huisman, N. J., Linthorst, H. J. M., Bol, J. F., and Cornelissen, B. J. C. (1988): The complete nucleotide sequence of potato virus X and its homologies at the amino acid level with various plus-stranded RNA viruses. *J. gen. Virol.* **69**, 1789–1798.

Kuschki, G., Koenig, R., Duvel, D., and Kuhne, H. (1978): *Helenium* virus S and Y: two new viruses from commercially grown helenium hybrids. *Phytopathology* **68**, 1407–1411.

MacKenzie, D. J., Tremaine, J. H., and Stace-Smith, R. (1989): Organization and interviral homologies of the 3' terminal portion of potato virus S RNA. *J. gen. Virol.* **70**, 1053–1063.

Memelink, J., van der Vlught, C. I. M., Linthorst, H. J. M., Derks, A. L. F., Asjes, C. J., and Bol, J. F. (1990): Homologies between the genomes of a carlavirus (lily symptomless virus) and a potexvirus (lily virus X) from lily plants. *J. gen. Virol.* **71**, 917–924.

Rupasov, V. V., Morozov, S. Yu., Kanyuka, K. V., and Zavriev, S. K. (1989): Partial nucleotide sequence of potato virus M RNA shows similarities to potexviruses in gene arrangement and the encoded amino acid sequences. *J. gen. Virol.* **70**, 1861–1869.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977): DNA sequencing with chain- terminating inhibitors. *Proc. natn. Acad. Sci. U. S. A.* **74**, 5463–5546.

Wetter, C., and Milne, R. G. (1981): Carlaviruses, pp. 695–730. In E. Kurstak (Ed.): *Handbook of Plant Virus Infections Comparative Diagnosis,* Elsevier/North Holland Biomedical Press, Amsterdam.