Volume 280, number 2, 341–343 FEBS 09525 March 1991
© 1991 Federation of European Biochemical Societies 0014-5793/91/$3.50
ADONIS 001457939100235Y

Occurrence of chloroplast ribosome recognition sites within conserved elements of the RNA genomes of carlaviruses

Gary D. Foster[1] and Peter R. Mills[2]

[1]*University of Leicester, Department of Botany, University Road, Leicester LE1 7RH, UK* and [2]*Department of Mycology and Plant Pathology, The Queen's University of Belfast and Plant Pathology Research Division, Department of Agriculture for Northern Ireland, Newforge Lane, Belfast, BT9 5PX, UK*

Received 11 December 1990; revised version received 28 January 1991

The nucleotide sequences upstream from the carlavirus open reading frames were examined for direct sequence homology. Blocks of homology were evident upstream from the 25 K ORFs of potato virus S (PVS), potato virus M (PVM) and lily symptomless virus (LSV), and upstream from the coat protein initiation codons of PVS, PVM, LSV, carnation latent virus and Helenium virus S. These blocks, which correspond to the 5'-terminal regions of the subgenomic RNAs, were shown to contain potential ribosome recognition sequences. The distances between the binding sites and initiation codons ranged from 20 to 40 nucleotides on the viral RNAs. Whilst the majority of chloroplasts mRNAs have a distance of 8 nucleotides between binding site and initiation codon, the remaining have a distance of 23 nucleotides which is similar to that reported here for the carlaviruses.

Subgenomic promoter; Sequence homology; Carlavirus; Ribosome binding site

1. INTRODUCTION

Carlaviruses are a group of filamentous plant viruses consisting of particles 610–700 nm in length, composed of multiple copies of a single coat protein of 33 kDa [1]. Genomic RNA is positive-sense and single-stranded, with a M_r between 2.3 and 2.6×10^6 (7–7.8 kb) which is 3'-polyadenylated and contains a 5' terminal cap structure [2,3]. Genomic RNA appears to encode a large molecular mass product of 190 kDa with the 3'-terminal region expressed through the production of two 3'-coterminal subgenomic RNAs of approximately 2.5 kb and 1.3 kb. These subgenomic RNAs have been detected in RNA from infected leaves and from encapsidated viral RNA [2,4].

Recently the nucleotide sequences of the 3' terminal regions of the carlaviruses carnation latent virus (CLV), potato virus S (PVS), potato virus M (PVM), Helenium virus S and lily symptomless virus (LSV), have been reported [5–9]. The predicted 3' open reading frames (ORFs) for all viruses consisted of a triple gene block of 25 K, 12 K and 7 K, 5' to the coat protein gene and an 11 K ORF 3' to the coat protein gene. The sizes of the subgenomic RNAs would place their 5' termini upstream from the 25 K and the coat protein genes. Indeed nucleic acid sequence homology between the upstream regions of these two genes has been reported for PVM and LSV individually, though it was reported that no homology was observed between the two viruses [7,9].

In this communication we report the first extensive comparison of these homologous nucleotide blocks upstream from the proposed initiation codons (ATG) of the carlavirus 25 K and coat protein genes. These comparisons have enabled us to make suggestions for the functions of the RNA sequences based on the presence of canonical Shine-Dalgarno sequences and possible modes of gene expression.

Correspondence address: G.D. Foster, University of Leicester, Department of Botany, University Road, Leicester LE1 7RH, UK

2. METHODS

The sequences were initially assembled and analysed using the program DNASIS (Pharmacia). For initial homology search, dot matrix comparisons were performed using the computer program DIAGON [10] with further analysis being carried out using ANALYSEQ [11] and CLUSTAL.

3. RESULTS AND DISCUSSION

The 100 nucleotides upstream from the initiation codons of the carlaviruses 25 K and 33 K (coat protein) ORFs were aligned for maximum homology using DNASIS and CLUSTAL. The aligned sequences, as shown in Fig. 1, revealed high homology between the non-coding regions upstream of the 25 K ORFs of LSV, PVS and PVM and between the upstream regions of the coat proteins (CP) of LSV, PVS, PVM, CLV and

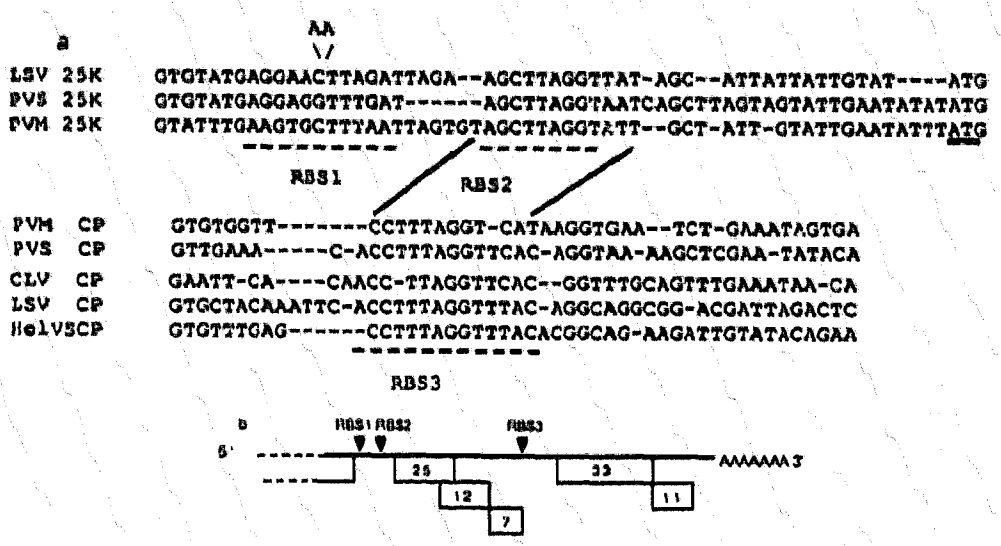

Fig. 1 (A) Alignment of the nucleotide sequences 5′ to the coat protein and 25 K ORFs of carlaviruses. Gaps have been introduced for maximum alignment. ATG start codons of the 25 K ORFs are underlined with a solid bar, potential ribosome binding sites, as discussed in the text, are underlined with hatched lines. RBS1 is further analysed in Fig. 2 and RBS2 and RBS3 in Fig. 3. Areas of homology between the 25 K and CP upstream regions are indicated by sloping bars. (B) Genetic organisation of the 3′-terminal region of a carlavirus RNA. The sizes and positions of the open reading frames (sizes given in kDa) are based on the sequence information previously reported [5–9]. The positions of the ribosome bindind sites on the carlavirus RNAs are indicated.

HelVS. Within the aligned sequences were blocks with highly conserved homology common to both the 25 K and CP sequences. These regions, which correspond approximately to the termini of the identified subgenomics, were examined further for homologies with other known sequences via data base searches.

These searches revealed that the conserved blocks upstream from the AUG of the 25 K ORF contained canonical Shine–Dalgarno sequences (S-D) AGGAGGT [12], which have been shown to be present in, and required for efficient expression of, nearly all *E. coli* mRNAs studied to date [13]. This classical purine-rich S-D sequence AGGAGGT, which is present 5′ to the initiation codon in prokaryotic mRNA, has been shown to base pair with a pyrimidine rich sequence near the 3′ end of the 16 S rRNA, thus allowing ribosome binding and subsequent initiation and elongation. The extent of the ribosome binding potential of the regions upstream from the carlavirus 25 K ORF can be seen in Fig. 2A where the sequences are aligned with the 3′-terminal sequence of 16 S rRNA [14,15]. As can be seen from the figure, PVS contains a perfect Shine–Dalgarno sequence of AGGAGGT which could base pair with the 3′ sequence of rRNA, with six other bases also exhibiting base pairing potential, suggesting a strong ribosome binding region. Similar sequences are also evident at the equivalent positions upstream from the AUG of the 25 K genes of LSV and PVM, at approximately 40 nucleotides from their respective initiation codons. These blocks are most striking when compared with the well defined ribosome binding sites of *E. coli* RNA polymerase β and the A protein of the *E. coli* phage MS2 as shown in Figs 2B and 2C [16].

When examined further, another potential ribosome binding site (RBS) is present downstream from the first 25 K ORF RBS and 10–12 nucleotides upstream from the 25 K AUG (Fig. 1). This region, based around the sequence AGCTTAGGTTT, has strong homology to

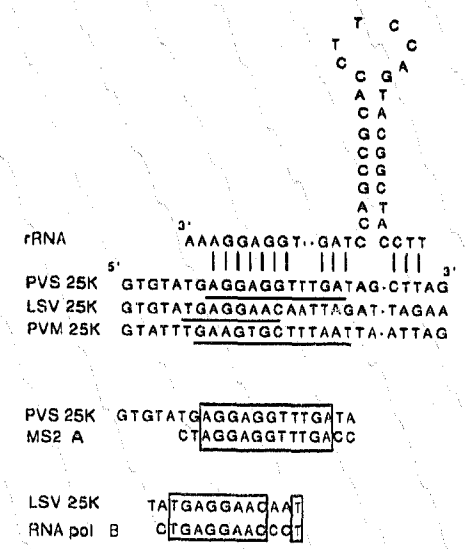

Fig. 2 (A) Complementarity of the 3′-terminal sequence from rRNA (with stem/loop) with the non-coding region upstream from the 25 K ORF (RBS1). The sequence of the rRNA 3′ end is shown as the complementary strand for direct comparison with viral sequences. Potential recognition sequences are marked by underlining, with base pairing indicated by vertical bars. (B) Homology between the potential ribosome binding site RNS1 of the PVS 25 K ORF and the ribosome binding site of the MS2 A protein of the RNA bacteriophage. (C) Homology between the potential ribosome site RBS1 of LSV 25 K ORF and the ribosome binding site of B-RNA polymerase of *E. coli*.

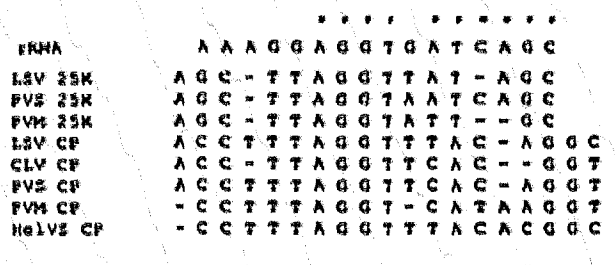

Fig. 3. Alignment of the highly conserved blocks RBS2 from the 25 K ORF, RBS3 from the CP ORF and rRNA (relaxed) shown in the complementary sense for direct comparison. Asterisks(*) above the sequences indicate nucleotides which may be involved in base pairing.

the sequence, ACCTTTAGGTTT, which is present 25-37 nucleotides upstream from the coat protein AUGs of PVM, PVS, CLV, LSV and HelVS (Fig. 1). These two regions also have the ability to act as ribosome binding sites via base pairing with ribosomal RNA in the relaxed state, as shown in Fig. 3.

The question which must be asked at this stage is why such sequences are present upstream from the AUG of a viral RNA from a eukaryotic host, which, in contrast to prokaryotes, conspicuously lacks the key anti S-D sequences CCTCC from the otherwise conserved 3′ end of the small 18 S rRNA of eukaryotes.

The model proposed by Kozak, for eukaryotic initiation requires the 40 S ribosomal subunit to bind to the 5′ end of a mRNA facilitated by the 5′ m7G cap structure. The ribosomal subunit subsequently migrates to the first AUG codon in a favourable context for initiation of translation [17]. The observation that the genomic RNA of carlaviruses appear to be capped at the 5′ terminus but that the subgenomic RNAs are not [3], suggests that an alternative mechanism may operate for ribosome binding and initiation for carlavirus subgenomic RNAs.

The basis of an alternative mechanism may lie in ribosome binding via base pairing, with one possible source of suitable ribosomes being the small subunit rRNA from chloroplasts. These have been shown to be highly homologous to the bacterial 16 S rRNA with strong conservation of their 3′-terminal sequences [18]. In Figs 2A and 3 the 3′-terminal sequences of the depicted rRNA is identical between *E. coli* rRNA and chloroplast 16 S rRNA [14,15]. This had led to the proposal that binding of plastidic mRNA to the small ribosomal subunit is mediated by base pairing, as has been demonstrated within prokaryotes. Indeed, Ruf and Kossel reported that 90% of chloroplast mRNA species examined contained ribosome recognition sequences upstream of the respective initiation codons [19].

It is interesting to note that the distance between the RBS and AUG within chloroplast RNAs peaked at 8 nucleotides, similar to the spacing of 7 ± 2 observed for the classical S-D sequences of bacterial mRNA. However, an additional peak was also observed at position −23 for the RBS, which is similar to the distances observed for the carlavirus RNAs, thus suggesting that chloroplast ribosomes can accommodate larger distances between the recognition sites and AUGs.

Whether these viral RNAs are translated within chloroplasts remains unknown at this stage. However it has long been established that viral infection does affect photosynthesis, with chloroplasts being structurally and biochemically affected by infection [20-22], though the exact mechanism remains unclear at this stage. If these viral blocks are involved in ribosome binding at the 25 K and CP ORFs then it has implications on the mode of translation of the other open reading frames, as these blocks were not present upstream from their AUGs. The expression of the other ORFs could therefore be by a 'relay race' mechanism of translation or translational frameshifting [23,24].

REFERENCES

[1] Koenig, R. (1982) Carlavirus Group CMI/AAB. Descriptions of Plant Viruses no 259.
[2] Foster, G.D. and Mills, P.R. (1991) Acta Virologica, in press.
[3] Foster, G.D. and Mills, P.R. (1991) Virus Genes, in press.
[4] Foster, G.D. and Mills, P.R. (1990) J. Gen. Virol. 71, 1247-1249.
[5] Foster, G.D., Meehan, B.M. and Mills, P.R. (1991) Plant Mol. Biol. in press.
[6] Mackenzie, D.J., Tremaine, J.H. and Stace-Smith, R. (1989) J. Gen. Virol. 70, 1053-1063.
[7] Rupasov, V.V., Morozov, S. Yu, Kanyuka, K.V. and Zavriev, S.K. (1989) J. Gen. Virol. 70, 1861-1869.
[8] Foster, G.D. and Mills, P.R. (1990) J. Gen. Virol. 71, 1877-1880.
[9] Memelink, J., van der Vlugt, C.I.M., Linthorst, H.J.M., Derks, A.F.L., Asjes, C.J. and Bol, J.F. (1990) J. Gen. Virol. 71, 917-924.
[10] Staden, R. (1982) Nucleic Acids Res. 10, 2951-2961.
[11] Staden, R. (1984) Nucleic Acids Res. 12, 521-538.
[12] Shine, J. and Dalgarno, L. (1975) Nature 254, 34-38.
[13] Dahlberg, A.E. (1989) Cell 57, 525-529.
[14] Noller, H.F., Asine, M., Barta, A., Douthwaite, S., Goldstein, T., Gutell, R., Moazed, D., Normanly, J., Prince, J.B., Stern, S., Triman, K., Turner, S., Van Stolk, B., Wheaton, V., Weiser, B. and Wulse, C.R. (1986) in: Structure, Function and Genetics of Ribosomes (B. Hardesty and G. Kramer, eds) pp. 143-163, Springer, New York.
[15] Schwartz, Z. and Kossel, H. (1979) Nature, 279, 520-522.
[16] Adams, R.C.P., Knowler, J.T. and Leader, D.P. (1986) The Biochemistry of the Nucleic Acids, Chapman and Hall, New York.
[17] Kozak, M. (1978) Cell 15, 1109-1123.
[18] Kuntzel, H., Heidrich, H. and Piechulla, B. (1981) Nucleic Acids Res. 9, 1451.
[19] Ruf, M. and Kossel, H. (1988) FEBS Lett. 240, 41-44.
[20] Goffeau, A. and Bove, J.M. (1965) Virology 27, 243-252.
[21] Bedbrook, J.R. and Matthews, R.E.F. (1972) Virology 48, 255-258.
[22] Bedbrook, J.R. and Matthews, R.E.F. (1973) Virology 53, 84-91.
[23] Gordon, K., Pfeiffer, P., Futterer, J. and Hohn, T. (1988) EMBO J. 7, 309-317.
[24] Valle, R.P.C. and Morch, M.D. (1988) FEBS Lett. 235, 1-15.