The EMBO Journal vol.4 no.1 pp.33 – 37, 1985

Major polyadenylated transcripts of cassava latent virus and location of the gene encoding coat protein

R. Townsend, J. Stanley, S.J. Curson and M.N. Short

John Innes Institute, Colney Lane, Norwich NR4 7UH, UK

Communicated by D.A. Hopwood

The nucleotide sequences of infectious cloned DNAs 1 and 2 of a Kenyan isolate of cassava latent virus (CLV) have been determined. Five virus-specific polyadenylated transcripts have been identified and mapped either to the viral or complementary sense DNAs of both components of the CLV genome, confirming that transcription is bidirectional on both DNAs. A major mRNA has been translated *in vitro* to yield a 30 000 mol. wt. product, which is precipitated by antibodies raised against whole virus, and has been mapped by both the S1 nuclease procedure and hybrid-arrested translation to the long open reading frame (ORF) in the viral sense of DNA 1 which encodes the coat protein. Other transcripts were of sufficient size and appropriate origin to encode at least five potential products.

Key words: geminivirus/transcripts/cassava latent virus/ coat protein gene

Introduction

Cassava latent virus (CLV) (synonym African cassava mosaic virus) is a member of the geminivirus group. Geminiviruses have small twinned or geminate particles containing circular single-stranded (ss) DNAs (Haber *et al.*, 1981). The genome of CLV comprises two similarly sized DNAs, the complete nucleotide sequences of which have been determined (Stanley and Gay, 1983). Computer analysis of the sequences of DNAs 1 and 2 revealed open reading frames (ORFs), with the potential to code for polypeptides of mol. wt. ≥10 000, in both the virion DNAs and their complements. Infectivity studies using recombinant DNA clones of DNAs 1 and 2 have confirmed that both components are required for infection (Stanley, 1983). We report here the ORFs in the sequences of these infectious cloned DNAs, describe the major polyadenylated transcripts isolated from plants infected with cloned DNA and the location of the gene encoding the coat protein. The results are discussed in relation to a recently determined nucleotide sequence (Hamilton *et al.*, 1984) of the infectious cloned DNAs of tomato golden mosaic virus (TGMV) (Hamilton *et al.*, 1983), an evolutionarily related geminivirus (Hamilton *et al.*, 1984).

Results

ORFs in infectious cloned CLV DNAs

Infectious cloned DNAs 1 (pJS092) and 2 (pJS094) of the West Kenyan isolate 844 of CLV (Stanley, 1983) were subcloned in M13 and sequenced by the dideoxy chain termination procedure. Screening the entire nucleotide sequence for ORFs in both the virus DNA sense and its complement revealed a similar arrangement of ORFs to that established by sequencing West Kenyan isolate 844 virion DNAs (Stanley and Gay, 1983) except that the ORF coding for a putative product of 27 K mol. wt. in the complementary sense of DNA 1 was shortened by 156 nucleotides, reducing its coding capacity to mol. wt. 21 K (Figure 1). The sequence of the long ORF in the viral sense of DNA 1, the putative coat protein gene (Stanley and Gay, 1983), showed two substitutions affecting amino acid composition and had a coding capacity of mol. wt. 30 153. Limited variations in sequence affecting amino acid composition are to be expected between a consensus sequence derived from a population of viral DNAs and the sequence of a cloned DNA. However, the shortening of an ORF may indicate that the ORF has arisen fortuitously.

The positions of potential promoters conforming to the consensus sequence TATAT/AA (Breathnach and Chambon, 1981) and polyadenylation signals AATAAA (Proudfoot and Brownlee, 1976) are shown in Figure 1. The majority of ORFs had a preferred purine A, three nucleotides upstream of the 5′-proximal ATG triplet (Kozak, 1984). Conspicuous exceptions were the three ORFs with coding capacities of 10.6 K, 21.0 K and 13.5 K, in the complement-

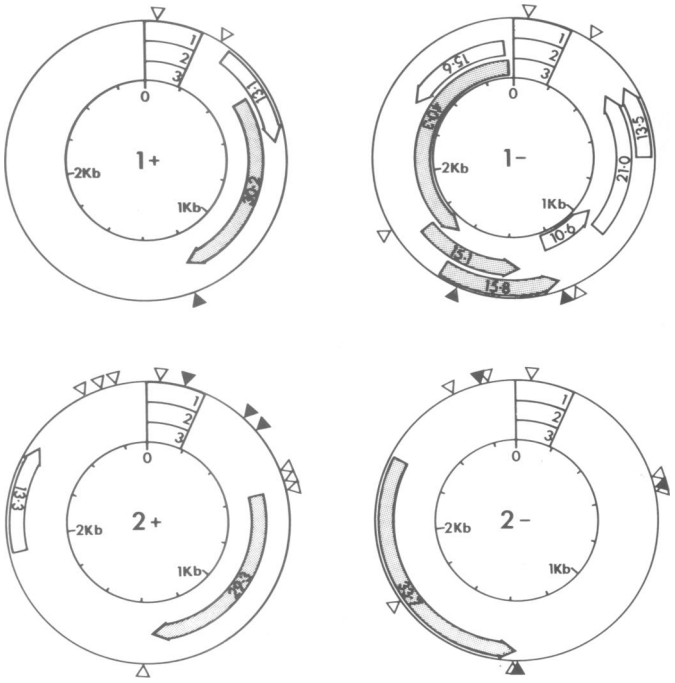

Fig. 1. Potential ORFs within cloned CLV DNAs 1 and 2 in both the virion (+) sense and its complement (−) assuming that the first in-phase ATG triplet of each ORF initiates protein synthesis. Only those ORFs with a coding capacity of mol. wt. ≥10 000 are indicated. Nucleotides 1 – 195 represent the non-coding regions common to both DNAs. TATA boxes are indicated by open triangles and AATAAA sequences by closed triangles. [The positions of the overlapping ORFs with coding capacities of 15.1 K and 15.8 K within DNA 1 (−) erroneously located in Figure 4 of Stanley and Gay (1983), have been corrected.] The ORFs (stippled) with coding capacities of 30.2 K (1+); 40.3 K, 15.1 K and 15.8 K (1−); 29.3 K (2+) and 33.7 K (2−) are conserved in terms of approximate size and location within the nucleotide sequences of CLV and TGMV (Hamilton *et al.*, 1984).

Table I. Experimentally determined amino acid composition of CLV coat protein.

Amino acid	Residues per molecule	
	Experimental	Predicted
Lysine	17	18
Histidine	10	10
Arginine	23	26
Aspartic acid	29	30
Threonine	16	16
Serine	11	11
Glutamic acid	17	16
Proline	15	14
Glycine	16	18
Alanine	13	11
Valine	15	14
Methionine	6	7
Isoleucine	15	17
Leucine	15	16
Tyrosine	15	15
Phenylalanine	11	11
Cysteine	6	6
Total residues	250	256

The determination based on a mol. wt. of 29 272 (tryptophan was not determined but was assumed to be two residues per molecule on the basis of predicted composition), compared with the composition of the putative product (mol. wt. 30 153) of the long ORF in the viral sense of DNA 1 (Figure 1).

ary sense of DNA 1 which overlap with the putative coat protein gene in the viral sense.

Amino acid composition of CLV coat protein

Purified CLV virus analysed by SDS-PAGE migrated as a single band of mol. wt. 30 000 ± 500 (results not shown). The empirical amino acid composition of CLV coat protein based on a mol. wt. of 29 272 was in good agreement with the predicted composition of the longer ORF in the viral sense of DNA 1 (Table I). The putative products of all other ORFs failed to show a similar correlation. The compositions of 18 of the 34 peptides generated by a tryptic digest of purified coat protein were determined. All had counterparts in the predicted amino acid sequence and two corresponded to the amino- and carboxy-terminal peptides.

Virus-specific RNAs in plants infected with cloned CLV DNAs

Northern blots of poly(A)$^+$ RNA probed with nick-translated full-length cloned dsDNA 1 showed a major RNA of 1 kb (Figure 2, band i) and two other RNAs of 1.7 kb (ii) and 0.7 kb (iii). Probing with dsDNA 2 revealed two major RNAs, one of 1.1 kb (iv), and a second of 0.9 kb (v). A minor species of 1.35 kb (vi) was also detectable. Further minor bands corresponding to transcripts of 2.0 kb and 2.2 kb were visible on the original autoradiograph. None of the RNAs hybridised to both probes, suggesting that none of them is transcribed across the common region. No RNAs in healthy tissue hybridised to either of the cloned probes (results not shown).

Samples of nonpolyadenylated RNA sometimes showed faint bands of 2.7 kb and occasionally 5.6 kb. They also occurred in poly(A)$^+$ RNA, particularly if the DNase I digestion step following oligo(dT)-cellulose chromatography was omitted. These bands could also be detected on Southern blots of alkaline gels and are presumably composed, at least

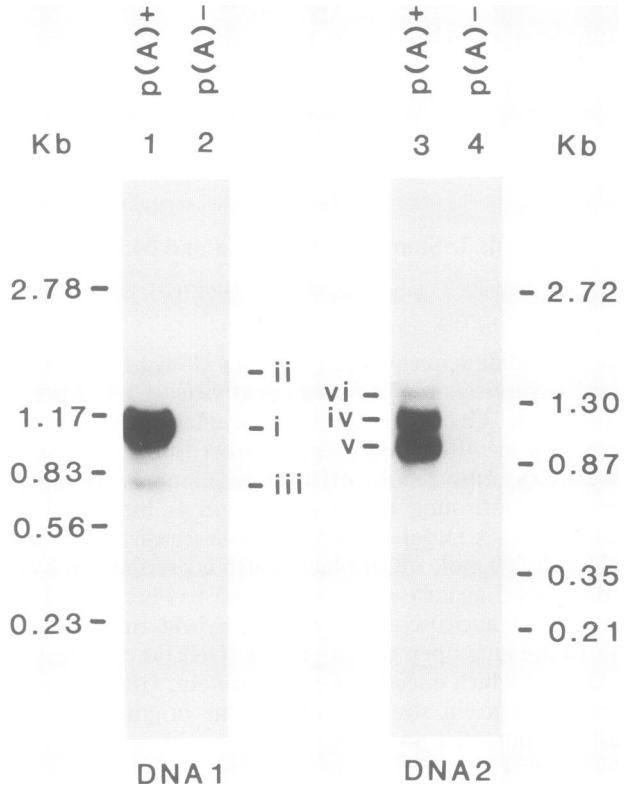

Fig. 2. Northern blot of 5 μg of total cellular poly(A)$^+$ RNA (**1** and **3**) and 10 μg of non-polyadenylated RNA (**2** and **4**) probed with nick-translated full length cloned DNAs 1 (**1** and **2**) or 2 (**3** and **4**) of CLV.

in part, of virion ssDNA, which also occurs as dimers (Stanley, 1983), and binds to oligo(dT)-cellulose.

Mapping of virus-specific poly(A)$^+$ RNAs to CLV DNAs

To determine the origin of viral transcripts, polyadenylated RNA was hybridised to full-length ssDNA clones of CLV DNAs 1 (pJS073 and pJS074) and 2 (pJS023 and pJS093) in either orientation and then digested with S1 nuclease. Protected DNA fragments were resolved on alkaline gels or glyoxal gels following alkali hydrolysis of the RNA/DNA duplexes. In the presence of excess DNA, duplex formation is quantitative with respect to the abundance of individual RNAs but if the protecting RNA is polyadenylated the resulting S1 resistant duplex will be correspondingly smaller.

In the absence of RNA, the M13 ssDNA clones were completely digested by S1 nuclease. However, in view of the small quantities of virion ssDNA contaminating some RNA preparations, RNA samples hydrolysed with alkali or digested with DNase-free RNase were also included as additional controls. These preparations sometimes protected very small amounts of complementary sense DNA clones pJS074 and pJS093 to yield fragments of 2.7 kb, confirming that the contaminating nucleic acid is virion ssDNA. Virion DNA was not present in concentrations which interfered with the mapping of sub-genomic length RNAs.

Polyadenylated RNA protected two pieces of virion sense DNA 1 (pJS073) of 0.6 kb and 1.5 kb (Figure 3) and a large quantity of complementary sense DNA 1 (pJS074) ~0.5 kb in size. This broad band resolved into two bands representing protected fragments of 0.45 kb and 0.5 kb on glyoxal gels (Figure 3), indicating that the abundant 1-kb poly(A)$^+$ RNA is transcribed from the complementary sense DNA of circle 1

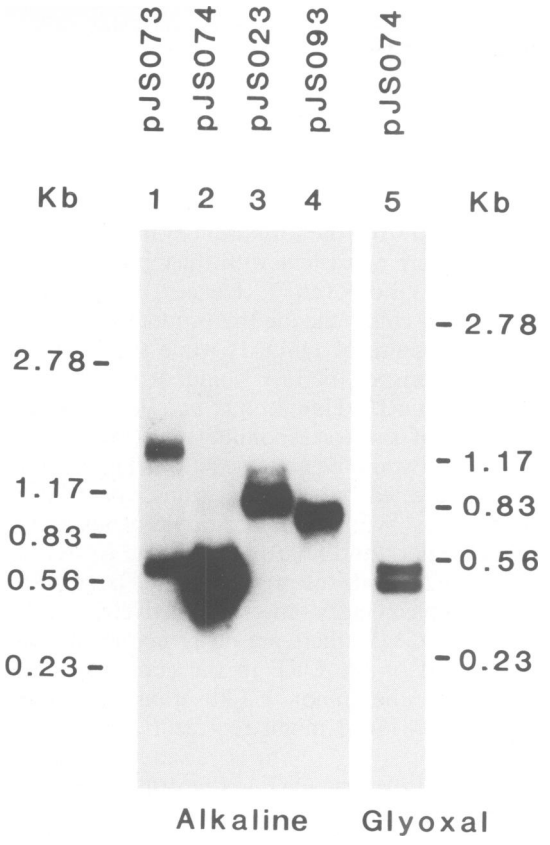

Fig. 3. Southern blot of S1 nuclease-resistant fragments of cloned CLV ssDNAs 1 (**1** and **2**) and 2 (**3** and **4**) in both orientations (+, **1** and **3**; −, **2** and **4**) protected by virus-specific RNAs and analysed on an alkaline gel (**1**–**4**). **Track 5** is an equivalent sample to that in **track 2** but analysed on a glyoxal gel following alkali denaturation.

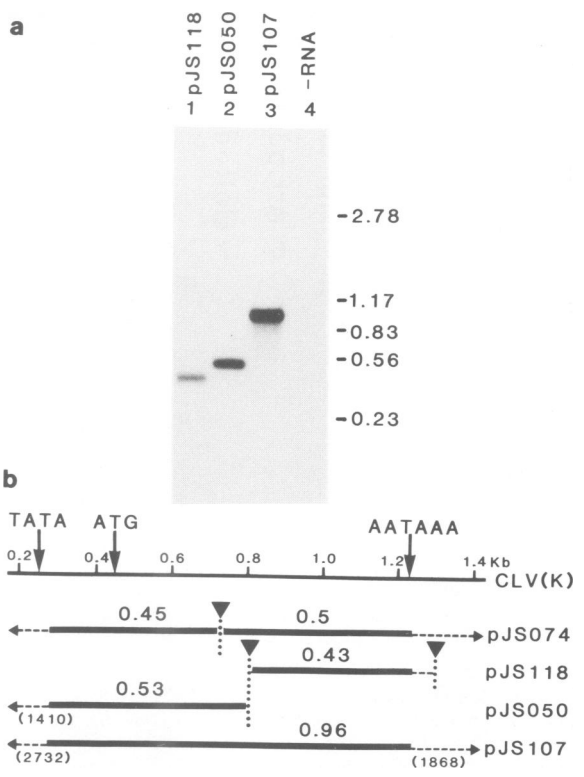

Fig. 4. (a) S1 nuclease-resistant fragments of complementary sense (−) M13 ssDNA clones protected by the 1-kb transcript, analysed on an alkaline gel. **(b)** protected fragments (solid lines) of ssDNA clones (broken lines) aligned with part of CLV DNA 1. Solid triangles delineate the termini of the clones where these fall within the mapped region. Nucleotide numbers in brackets delineate the clone termini where these fall outside the mapped region.

and traverses the pJS074 *Mlu*I cloning site at nucleotide 734. Hybridisation of poly(A)$^+$ RNA to virion sense DNA 2 (pJS023) resulted in the protection of one fragment of 1.0 kb and a small quantity of a 1.3 kb fragment while hybridisation to complementary sense of DNA 2 (pJS093) protected a fragment of 0.85 kb (Figure 3).

Mapping of the 1-kb transcript on DNA 1

Poly(A)$^+$ RNA was hybridised to M13 ssDNA clones containing defined fragments of complementary sense DNA 1. *Eco*RI clone pJS107 (2732–1868) which traverses the *Mlu*I site and is complementary to the entire coding region of the long ORF in the viral sense of DNA 1 protected a single fragment of 0.96 kb (Figure 4a). *Bgl*II clone pJS050 (1410–804) and *Sau*3A clone pJS118 (805–1293), which include sequences complementary to the 5′ and 3′ ends of that ORF, gave fragments of 0.53 kb and 0.43 kb, respectively. In Figure 4b the protected fragments are aligned with part of circle 1 of the CLV genome, showing that the abundant 1-kb virus-specific polyadenylated transcript corresponds to the entire sequence of the longer ORF on viral sense DNA 1. The 5′ terminus of the transcript maps near nucleotide 280 and the 3′ terminus is located in the region of nucleotide 1240.

Translation of virus-specific poly(A)$^+$ RNAs

To characterise the products of viral mRNAs, total cellular poly(A)$^+$ RNA from healthy *Nicotiana benthamiana* and plants infected with cloned CLV was translated in a rabbit reticulocyte lysate. A large number of ^{35}S-labelled polypeptides were produced but only one of mol. wt. 30 000 ± 500 was unique to lysate programmed with polyadenylated RNA from CLV-infected plants. This was produced in rather small amounts and was partially obscured on autoradiographs by host products (results not shown).

The 30 000 mol. wt. product was immunoprecipitated by antiserum raised against purified CLV virions (Figure 5, track 3) but despite cross-adsorbing the antiserum against an acetone powder of healthy *N. benthamiana*, several host polypeptides, notably one of 20 K, were also precipitated.

Mapping the coat protein mRNA by hybrid arrest translation

To confirm the identity of the 1-kb transcript as the coat protein mRNA, poly(A)$^+$ RNA was hybridised to either an M13 clone pJS069 (485–804), which contained a ss insert of DNA 1 complementary to approximately half the coding region of the coat protein gene, or to the full length ssDNA clone of viral DNA 1 (pJS073) of the same sense as the transcript. Programming the lysate with RNA hybridised to pJS069 (Figure 5, track 4) resulted in almost total inhibition of 30 000 mol. wt. polypeptide synthesis which was completely restored upon melting of the hybrid (track 5) whereas synthesis of this polypeptide was unaffected by hybridisation to pJS073 (tracks 8 and 9).

Discussion

The hybridisation of virus-specific transcripts to both senses of CLV DNAs 1 and 2 is consistent with the bi-directional transcription of both DNAs which is suggested by the arrangement of ORFs, and associated TATA and AATAAA boxes in the nucleotide sequences of CLV (Stanley and Gay,

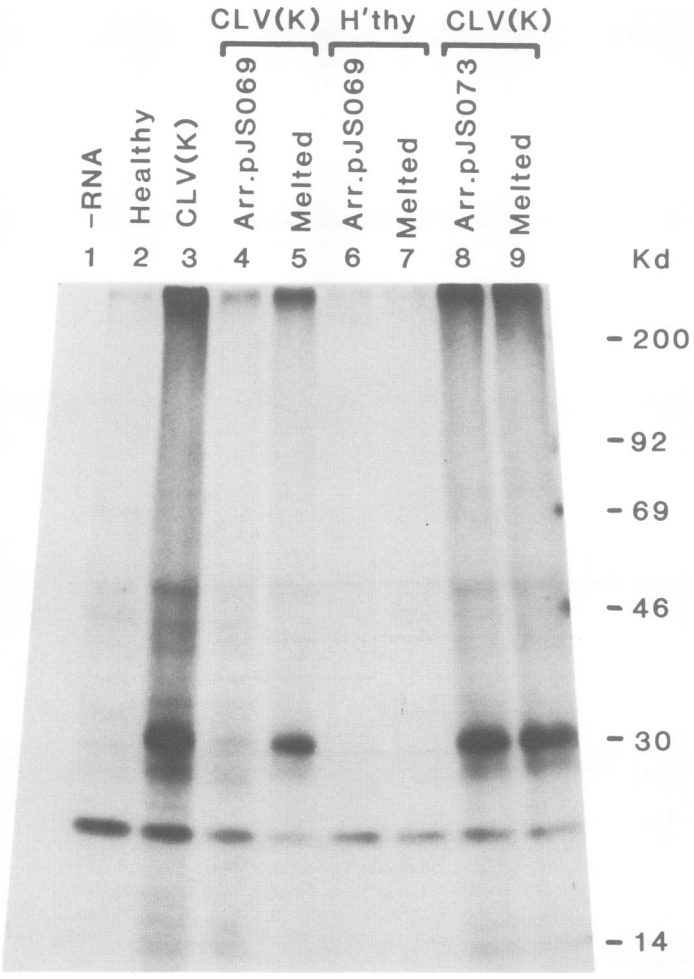

Fig. 5. Fluorograph of [^{35}S]methionine-labelled immunoprecipitated polypeptides from *in vitro* translations of total cellular poly(A)$^+$ RNA from healthy (**2, 6** and **7**) and CLV infected (**3–5, 8** and **9**) plants after hybridisation to ssDNA clones pJS069 (**4** and **6**) and pJS073 (**8**) and after heat melting of hybrids (**5, 7** and **9**).

Table II. Comparison of sequences immediately upstream of AUG triplets in the untranslated 5′ leader of the mRNA encoding coat protein with those upstream of the initiation codon.

Sequence 5′- -3′	Location of AUG triplet nucleotide no.
GCGCAAUG	286[a]
GGUGGAUG	305
GGUGCAUG	326
GGUGUAUG	340
UAAAUAUG	446 Initiation codon

[a]May be upstream of the 5′ terminus of the transcript

1983) and TGMV (Hamilton *et al.*, 1984). Six ORFs are conserved between CLV and TGMV (Figure 1). They show a marked symmetry of distribution on both DNAs, converging on polyadenylation signal sequences which are separated by a few nucleotides. Although there is no AATAAA sequence at the 3′ end of the long ORF in the viral sense of CLV DNA 2, there is an AT-rich region, including the sequence TATAAA (nucleotides 1367–1372). This arrangement suggests that overlapping transcription may be confined to short stretches downstream of the polyadenylation signals at the 3′ terminus of each transcript. These features invite comparison with the bidirectional transcription of circular dsDNA viruses, such as SV40 (reviewed by Tooze, 1980).

Although we do not discount the possibility that CLV may have additional properties, the virus-specific RNAs which we have identified are of appropriate sizes to encode the products of the six ORFs which are conserved between CLV and TGMV. Moreover, the sizes, origins and orientations of RNAs as deduced from the formation of hybrids with ssDNA clones of CLV are consistent with their potential to encode the products of those ORFs. Hence, the polyadenylated RNAs of 1.7 kb could encode the product of 40.3 K in the complementary sense of DNA 1, while the 0.7-kb RNA is large enough to encode the 15.1-K and/or 15.8-K products of the overlapping ORFs. Hamilton *et al.* (1984) suggested that the expression of the corresponding two small overlapping ORFs in TGMV would be possible by splicing but analysis of the ssDNA protected by the small transcript on denaturing gels did not reveal evidence of additional bands indicative of spliced RNAs (Berk and Sharp, 1977). The RNAs of 0.9 kb and 1.1 kb could encode the products of the large ORFs in the virion and complementary senses, respectively, of DNA 2. In common with TGMV, there are no promoter-like sequences near the 5′ end of the ORF in the complementary DNA sense; however, the minor 1.3-kb transcript detected by hybridisation to pJS023 might represent a precursor of the 1.1-kb RNA.

The identity of the long ORF in the viral sense of DNA 1 as the coat protein gene, which was tentatively identified on the basis of amino acid analysis of virus coat protein (Stanley and Gay, 1983), has been confirmed by the isolation of the 1-kb mRNA, which encodes the coat protein, and which maps to that ORF. This mRNA has an untranslated leader of ~160 nucleotides and its 5′ terminus maps just downstream of a TATA box (251). A consensus transcription modulation sequence or CAAT box (Benoist *et al.*, 1980) (GGCCAATC) is located 48 nucleotides upstream of this promoter. The same sequence occurs 65 nucleotides upstream of the TATA box proximal to the first ATG in the putative coat protein gene on DNA A of TGMV (Hamilton *et al.*, 1984). Within the 5′-untranslated leader there are four apparently non-functional AUGs upstream of the initiation codon. Two are immediately upstream of a termination codon but two are in-phase with the smaller overlapping ORF. None, including the initiation codon, conforms to the consensus eukaryotic initiation sequence CCA/GCCAUG (Kozak, 1984) but they exhibit an interesting conservation of sequence in positions −2 to −5, relative to the AUG, which is completely different from the sequence context about the initiation codon (Table II). The initiation codon is the only one of the five with the preferred purine A in the −3 position (Kozak, 1984) and may be the only site at which polypeptide synthesis initiates. A minority of other mRNAs (reviewed by Kozak, 1983, 1984) have one or more AUGs upstream of the initiation site. The 5′ region of *Phaseolus vulgaris* lectin mRNA contains four AUG codons and is potentially dicistronic since one could initiate a 58 amino acid peptide out-of-phase with lectin translation (Hoffman *et al.*, 1982; Hoffman, 1984). In some instances, for example the mRNA encoding the thymidine kinase gene of herpes simplex virus type 1, such RNAs are functionally dicistronic, and direct synthesis of two polypeptides (Preston and McGeoch, 1981). In view of the high level of host polypeptide synthesis *in vitro*, it is unlikely that we would have detected other minor virus-specific products, particular-

ly if these were not antigenic components of intact virions. It is hoped that current studies using fine mapping of transcript termini and translation of hybrid-selected RNAs will resolve this point.

Materials and methods

Cloning and sequencing

Full length infectious cloned DNAs 1 and 2 of CLV (Stanley, 1983) were subcloned into appropriate M13 vectors after restriction with various endonucleases. Sequences of inserts were established using the dideoxy chain termination procedure (Sanger *et al.*, 1980) employing a synthetic deoxynucleotide 17 nucleotides in length (Duckworth *et al.*, 1981). Full length inserts were excised from RFs of M13 clones with the appropriate restriction endonuclease and purified by sucrose density gradient centrifugation.

Infection of plants and virus purification

N. benthamiana plants were inoculated with purified inserts from RFs of M13 clones (1 μg of cloned insert DNA per plant). Plants were maintained at 25°C and systemically infected leaves were harvested after 14 days. Virus was purified by the method of Sequeira and Harrison (1982).

Amino acid analysis of coat protein

Purified CLV virus (5 mg) was hydrolysed by the HCl method and the amino acid composition of the dried hydrolysate was determined as described by Spackman *et al.* (1958) using a Beckman 121 amino acid analyser. Methionine and cysteine were determined as methionine sulphone and cysteic acid, respectively, after performic acid oxidation (Moore, 1963). Threonine and serine values were corrected for hydrolytic loss (Rees, 1946).

RNA extraction

Total cellular nucleic acid was extracted from systemically infected leaves, 9 days after infection with purified virus derived from plants infected with cloned DNAs. The DNA was removed from the preparation according to the method of Covey and Hull (1981). Aliquots of heat-denatured total cellular RNA were fractionated by two cycles of oligo(dT)-cellulose chromatography. Poly(A)$^+$ RNA was resuspended in 50 mM Tris-HCl, pH 7.6, 5 mM $MgCl_2$ and incubated with RNase-free DNase I (10 μg/ml) for 30 min at 20°C. Protein was removed by extraction with phenol/chloroform and the RNA was ethanol precipitated.

Northern hybridisations

Samples of RNA were glyoxalated and electrophoresed through 1.4% agarose slab gels in a buffer containing 25 mM Tris-acetate pH 7.9, 1 mM EDTA and 5 mM Na acetate (glyoxal gels) as described by Covey *et al.* (1983). Gel-fractionated nucleic acid was blotted from glyoxal gels onto nitrocellulose (Thomas, 1980). Blots were probed with cloned DNAs labelled to high specific activity by nick-translation synthesis (Rigby *et al.*, 1977) in the presence of [α-^{32}P]dATP (Amersham International). Size markers were cloned purified full length CLV dsDNA 1 or 2 restricted with *Hae*II/*Acc*I and *Acc*I respectively.

Nuclease S1 mapping

Poly(A)$^+$ RNAs were mapped using the S1 nuclease procedure of Berk and Sharp (1977) as modified by Favaloro *et al.* (1980). In the hybridisation mix, cloned CLV ssDNA was in 10-fold molar excess, based on insert size, over the estimated amount of virus-specific RNA present in each 200 ng aliquot of total cellular poly(A)$^+$ RNA. Hybridisation was at 48°C for 3 h and digestion was with 50 units of S1 nuclease (Sigma) per reaction mixture. S1-resistant DNA fragments were electrophoresed in alkaline 1.2% agarose slab gels (Favaloro *et al.*, 1980). Alternatively RNA/DNA hybrids were hydrolysed at 60°C for 15 min in alkaline mix containing 0.15 M NaOH and 10 mM EDTA. The mixture was neutralised and resistant ssDNA precipitated with isopropanol before being analysed on glyoxal gels (Covey *et al.*, 1983). Gels were blotted onto nitrocellulose as described by Southern (1975) and probed with nick-translated purified full-length CLV DNAs.

In vitro translation of poly(A)$^+$ RNAs

Aliquots (2 μg) of total cellular poly(A)$^+$ RNA from healthy and infected plants were translated *in vitro* using a polypeptide synthesising system derived from lysed rabbit reticulocytes (Hunt and Jackson, 1974) rendered messenger dependent (Pelham and Jackson, 1976). [^{35}S]methionine was incorporated into the reaction mixture to label synthesised polypeptides. Antiserum raised in rabbits against purified CLV virions was produced according to the schedule described by Archer and Best (1980). Polypeptide products were immunoprecipitated from the rabbit reticulocyte lysate according to Martin and Northcote (1982) using a 1/10 volume of antiserum and 60 μl of a freshly prepared suspension of protein A-Sepharose CL-4B (Pharmacia) (62.5 mg/ml).

Products of *in vitro* translations and purified virus preparations were solubilised in SDS and analysed on SDS/linear gradient polyacrylamide (7.5 – 25%), bisacrylamide (0.2 – 0.125%) slab gels with a 4.5% acrylamide, 0.09% bisacrylamide stacking gel using the buffer system of Laemmli (1970). Gels were stained with Coomassie brilliant blue or fluorographed (Jen and Thach, 1982).

Hybrid-arrested translation

Samples of total cellular poly(A)$^+$ (4 μg) were hybridised to 1 μg of cloned CLV ssDNA (pJS069 or pJS073) for 2 h at 50°C. Hybrid and heat melted portions of the sample were used to programme rabbit reticulocyte lysate according to the method of Paterson *et al.* (1977). Polypeptide products were immunoprecipitated and analysed by PAGE as described above.

Acknowledgements

We wish to thank Drs. S. Covey, J.W. Davies, G.P. Lomonossoff and P.M. Mullineaux for their interest and advice. This work was carried out under MAFF licence No. 49/152.

References

Archer,D.B. and Best,J. (1980) *J. Gen. Microbiol.*, **119**, 413-422.
Benoist,C., O'Hare,K., Breathnach,R. and Chambon,P. (1980) *Nucleic Acids Res.*, **8**, 127-142.
Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
Covey,S.N. and Hull,R. (1981) *Virology*, **111**, 463-474.
Covey,S.N., Turner,D. and Mulder,G. (1983) *Nucleic Acids Res.*, **111**, 251-264.
Duckworth,M.L., Gait,M.J., Goelet,P., Hong,G.F., Singh,M. and Titmas, R.C. (1981) *Nucleic Acids Res.*, **9**, 1691-1706.
Favaloro,J.M., Treisman,R.H. and Kamen,R.I. (1980) *Methods Enzymol.*, **65**, 718-749.
Haber,S., Ikegami,M., Bajet,N.B. and Goodman,R.M. (1981) *Nature*, **289**, 324-326.
Hamilton,W.D.O., Bisaro,D.M., Coutts,R.H.A. and Buck,K.W. (1983) *Nucleic Acids Res.*, **11**, 7387-7391.
Hamilton,W.D.O., Stein,V.E., Coutts,R.H.A. and Buck,K.W. (1984) *EMBO J.*, **3**, 2197-2205.
Hoffman,L.M. (1984) *J. Mol. Appl. Genet.*, **2**, 447-453.
Hoffman,L.M., Ma,Y. and Barker,R.F. (1982) *Nucleic Acids Res.*, **10**, 7819-7828.
Hunt,T. and Jackson,R.J. (1974) in Neth,R., Gallo,R.C., Spiegelman,S. and Stohlman,F. (eds.), *Modern Trends in Human Leukaemia*, J. F. Lehmans, Verlag, Munich, pp. 300-307.
Jen,G. and Thach,R.F. (1982) *J. Virol.*, **43**, 250-261.
Kozak,M. (1983) *Microbiol. Rev.*, **47**, 1-45.
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857-872.
Laemmli,U.K. (1970) *Nature*, **227**, 680-685.
Martin,C. and Northcote,D.H. (1982) *Planta*, **154**, 174-183.
Moore,S. (1963) *J. Biol. Chem.*, **238**, 235-237.
Paterson,B.M., Roberts,B.E. and Kuff,E.L. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 4370-4374.
Pelham,H.R.B. and Jackson,R.J. (1976) *Eur. J. Biochem.*, **67**, 247-256.
Preston,C.M. and McGeoch,D.G. (1981) *J. Virol.*, **38**, 593-605.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211-214.
Rees,M.W. (1946) *Biochem. J.*, **40**, 632-640.
Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237-251.
Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161-178.
Sequeira,J.C. and Harrison,B.D. (1982) *Ann. Appl. Biol.*, **101**, 33-42.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Spackman,D.H., Stein,W.H. and Moore,S. (1958) *Anal. Chem.*, **30**, 1190-1206.
Stanley,J. (1983) *Nature*, **305**, 643-645.
Stanley,J. and Gay,M.R. (1983) *Nature*, **301**, 260-262.
Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201-5205.
Tooze,J. (1980) *Molecular Biology of Tumour Viruses Part 2. DNA Tumour Viruses*, 2nd ed., published by Cold Spring Harbor Laboratory Press, NY.

Received on 2 October 1984; revised on 12 November 1984