Synthesis of the tomato golden mosaic virus AL1, AL2, AL3 and AL4 proteins in vitro

P. A. Thömmes† and K. W. Buck*

Department of Biology, Imperial College of Science, Technology and Medicine, Prince Consort Road, London SW7 2BB, U.K.

Transcripts derived from the leftward region of tomato golden mosaic virus DNA A were translated in wheat-germ and rabbit reticulocyte lysate systems. The largest protein (Mr 40K) produced from transcripts encompassing open reading frame (ORF) AL1 was identified as the AL1 protein by immunoprecipitation with AL1-specific antibodies. However the main product was a protein of Mr 10K, that was shown by in vitro mutagenesis to be the product of AL4, an ORF contained within AL1 DNA in a different reading frame. Translation of transcripts containing ORF AL2 or ORF AL3 gave the AL2 and AL3 proteins respectively; both proteins were also efficiently produced from transcripts containing both ORFs which overlap over about two-thirds of their length. Translation of a transcript containing the four ORFs gave all four proteins, consistent with a previous report that three of these (AL1, AL2 and AL3) can be translated from a single polycistronic RNA in transgenic tobacco plants. It is suggested that the leftward region of DNA A of the whitefly-transmitted geminiviruses may be expressed by two principal messenger RNAs, one encoding the AL1 and AL4 proteins and the other encoding the AL2 and AL3 proteins, and that the AL4 and AL3 proteins may be translated from these messenger RNAs by a leaky scanning mechanism.

Introduction

Tomato golden mosaic virus (TGMV) belongs to a subgroup of the geminiviruses, members of which are transmitted by whiteflies and infect dicotyledonous plants (reviewed by Lazarowitz, 1992). Its genome consists of two circular single-stranded DNA components, A and B, both of which are required for infection and symptom development in plants (Hamilton et al., 1983). However, DNA A is sufficient for virus replication in single cells (Rogers et al., 1986). DNA A has five open reading frames (ORFs) with the potential to code for proteins of Mr > 9500 (Hamilton et al., 1984; Elmer et al., 1988) which are transcribed in a rightwards (clockwise) (AR1) or leftwards (anticlockwise) (AL1, AL2, AL3, AL4) direction. Genetic analysis indicates that the AL1 protein is absolutely required for DNA replication (Elmer et al., 1988; Sunter et al., 1990). Introduction of an amber mutation near the start of the AL4 ORF did not affect virus replication or symptom development in Nicotiana benthamiana plants (Elmer et al., 1988). However in N. benthamiana protoplasts, expression of ORF AL4 from a plasmid resulted in the suppression of expression from the AL1 promoter on another plasmid (Gröning et al., 1994). Furthermore there is an equivalent ORF in several geminiviruses (Stanley et al., 1992) and mutations in the equivalent C4 ORF of beet curly top virus cause symptom alterations (Stanley & Latham, 1992) suggesting that it is expressed during virus infection of plants. Of these five ORFs in TGMV DNA A, the protein products of only two, the coat protein (Kallender et al., 1988) and the AL3 protein (Hanley-Bowdoin et al., 1990), have so far been detected in virus-infected plants. ORFs AL1 and AR1 are separated by an intergenic region (nt 14 to 326) that contains transcriptional promoters for these genes (Hamilton et al., 1984; Hayes et al., 1989; Sunter et al., 1993). Analysis of TGMV-specific RNAs in infected plants has identified putative mRNAs for the AL1 and AR1 genes starting at nucleotides 61 or 62 and 319 or 320 respectively (Petty et al., 1988; Sunter et al., 1989; Sunter & Bisaro, 1989; Hanley-Bowdoin et al., 1989). The transcript starting at

† Present address: Marie Curie Cancer Research Institute, The Chart, Oxted, Surrey RH8 0TL, U.K.

0001-2310 © 1994 SGM
nucleotides 61 or 62 encompasses ORFs AL1, AL2, AL3 and AL4 (Fig. 1) and may act as a polycistronic mRNA for all these ORFs. Indeed, an equivalent polycistronic RNA expressed in transgenic plants was able to complement mutants in the AL1, AL2 or AL3 genes in trans, implying that expression of these three proteins can occur from a single RNA (Hanley-Bowdoin et al., 1989). Nevertheless, several other transcripts, which initiate downstream of the AL1 translation initiation codon (nt 13 to 11) and might serve as mRNAs for the AL2, AL3 or AL4 proteins, have been identified in virus-infected plants (Sunter & Bisaro, 1989; Hanley-Bowdoin et al., 1989). Transcripts starting at nucleotides 2548, 2540 or 2515 might serve as mRNA for the AL4 protein (translation initiation codon at nucleotides 2447 to 2445 or 2441 to 2439) and those starting at nucleotides 1935 or 1828 might act as mRNA for the AL2 or AL3 proteins (translation initiation codons at nucleotides 1601 to 1599 and 1465 to 1463, respectively). However, it is not known if all these transcripts are functional in vivo and there are no reports of the synthesis of the AL1, AL2, AL3 or AL4 proteins in vivo. Here we report the results of in vitro translation of RNA transcribed from different parts of the leftward region of TGMV DNA A.

Methods

Construction of expression plasmids containing ORFs of TGMV DNA A. NdeI and BglII sites were introduced by in vitro mutagenesis (Kunkel et al., 1987) at the ATG translation initiation codon and just downstream of the termination codon of the AL1, AL2 and AL3 ORFs respectively in pAX, a clone of TGMV DNA in the xhoI site of pEMBL9X (Brough et al., 1988), to give pAX1, pAX2 and pAX3 respectively (Gröning et al., 1994). The AL1, AL2 and AL3 ORFs were excised from pAX1, pAX2 and pAX3 with NdeI and BglII and cloned between the NdeI and BamHI sites of pGEMEX-2*174 to give pTAL1/4, pTAL2 and pTAL3 respectively. The vector pGEMEX-2 is a modified version of pGEMEX-2 (Promega) in which a HindIII site (which contained TGMV DNA from position 50 to 174) is a leftwards or anticlockwise direction including ORFs ALl, AL2, AL3 and AL4 (translation initiation codon at nucleotides 2447 to 2445 or 2441 to 2439) and those starting at nucleotides 1935 or 1989 might act as mRNA for the AL2 or AL3 proteins (translation initiation codons at nucleotides 1601 to 1599 and 1465 to 1463, respectively). However, it is not known if all these transcripts are functional in vivo and there are no reports of the synthesis of the AL1, AL2, AL3 or AL4 proteins in vivo. Here we report the results of in vitro translation of RNA transcribed from different parts of the leftward region of TGMV DNA A.

Results and Discussion

Translation of the AL1 and AL4 ORFs

The region of TGMV DNA containing ORF AL1, which includes ORF AL4 (Fig. 1), was cloned into a transcription vector under the control of the T7 promoter. Transcripts from the resultant clone, pTAL1/4, were synthesized in vitro with T7 RNA polymerase and then translated in wheatgerm extract and reticulocyte lysate systems. The products were analysed by SDS-PAGE. In the wheatgerm extract, the largest product had an Mr of 40 K, as expected for the AL1 protein (Fig. 2, lane 1).
Several minor bands were also produced which may have arisen by initiation of translation at an internal methionine codon in the AL1 ORF (Fig. 1), premature termination of translation of the AL1 ORF or degradation of the full-length AL1 protein. The most prominent of the minor bands ($M_r$, 18K) may result from internal initiation at the second methionine codon of the AL1 ORF and would correspond to the predicted helicase domain of AL1 (Gorbalenya et al., 1990), to L1" of maize streak virus and similar multipartite geminiviruses (Lazarowitz, 1992) and to the postulated PRO1 protein of Abutilon mosaic virus (Frischmuth et al., 1990). However the major translation product had an $M_r$ of about 10K, which is the value for the AL4 protein predicted from the sequence data. Essentially similar results were found in the reticulocyte lysate, but the presence of a large amount of globin in this system distorted the migration of proteins in the lower region of the gel (not shown).

Two methods were used to determine whether the 10K protein is the product of ORF AL4, or is derived from ORF AL1, e.g. by internal initiation, premature termination or degradation. In the first method, the products of the translation of pTAL1/4 transcripts were immunoprecipitated with antibodies to the AL1 protein. The 40K protein was precipitated by AL1 antibodies (Fig. 3, lane 2), but not by preimmune antibodies (Fig. 3, lane 1), confirming its identity as the AL1 protein. The 10K protein was precipitated by both the preimmune and the immune AL1 antibodies, precluding the use of
In vitro translation of transcripts containing wild-type or mutant ORFs AL1 and AL4. RNA transcripts from pTAL1/4 (lane 1), pTAL4 (lane 2) and pTALX (lane 3) were translated in a wheatgerm system in the presence of [35S]methionine. The translation products were separated by SDS-PAGE and visualized by autoradiography. The M₀ values of marker proteins are shown on the side of the autoradiograph.

Fig. 2. In vitro translation of transcripts containing wild-type or mutant ORFs AL1 and AL4. RNA transcripts from pTAL1/4 (lane 1), pTAL4 (lane 2) and pTALX (lane 3) were translated in a wheatgerm system in the presence of [35S]methionine. The translation products were separated by SDS-PAGE and visualized by autoradiography. The M₀ values of marker proteins are shown on the side of the autoradiograph.

In the second method, two mutants were constructed. The first mutant, pTAL4, had an intact AL4 ORF, but contained a mutation which would result in a translational frameshift two amino acids after the first ATG of the AL1 ORF, culminating in termination after a further 36 amino acids. In vitro translation of pTAL4 transcripts produced an amount of the 10K protein similar to that produced by pTAL1/4 transcripts (Fig. 2, lanes 1 and 2). This shows that the 10K protein is not produced by translational initiation at the first ATG of the AL1 ORF followed by premature termination, or degradation of the full-length AL1 protein, but does not exclude the (unlikely) possibility of initiation at an internal methionine codon followed by premature termination or degradation.

The second mutant, pTALX, contained a mutation which would result in a translational frameshift one amino acid after the first ATG codon of the AL4 ORF culminating in termination after a further 41 amino acids. The second ATG codon, two codons downstream of the first ATG codon, was also destroyed. This mutation also frameshifted the AL1 ORF into the AL4 ORF and would be predicted to produce an AL1-AL4 fusion protein of M₀ about 16K (Fig. 1). In vitro translation of pTALX transcripts produced no detectable 40K AL1 protein, although a band of the predicted size for the AL1-AL4 fusion protein, not produced from pTAL1 or pTAL4 transcripts, was detected (Fig. 2, lane 3). Furthermore, no 10K protein could be detected. Since the internal methionine codons of ORF AL1 were unchanged in pTALX, it is clear that production of the 10K protein did not result from initiation at an internal methionine codon of the AL1 ORF, followed by premature termination or degradation. Additionally, since the ability of pTAL4 transcripts to direct the synthesis of normal amounts of the 10K protein showed that this protein did not result from initiation at the first methionine of the AL1 ORF, the failure of pTALX transcripts to direct the synthesis of the 10K protein proves conclusively that the 10K protein is the product of ORF AL4.

Translation of most eukaryotic mRNAs initiates at the 5' proximal AUG (Kozak, 1989). However the translation initiation efficiency is affected by the flanking sequences, the most common in plants being A/GNNAUGG (Lütcke et al., 1987; Cavener & Ray, 1991). If the 5'-proximal AUG does not match this consensus sequence, a proportion of ribosomes may
In vitro synthesis of TGMV proteins

1 2 3 4 5

Fig. 4. In vitro synthesis of the AL2 and AL3 proteins. RNA transcripts from pTAL1/4 (lane 1), pTAL2 (lane 2), pTAL3 (lane 3), pTAL2/3 (lane 4) and pTAL1/2/3/4a (lane 5) were translated, and the products were analysed, as in Fig. 2.

bypass the first AUG and initiate at a downstream AUG with the correct sequence context, a process called leaky scanning (Kozak, 1986). The sequence surrounding the initiation codon of the AL1 ORF is AAAAUGC which is suboptimal. The next AUG with the optimal sequence context is the second AUG of ORF AL4 (AAGAUGG). It is therefore possible that ORF AL4 is translated by a leaky scanning mechanism, starting at its second AUG. The first AUG of ORF AL4 (two codons upstream of the second AUG) has the suboptimal sequence context UUCAUGA and is not conserved in the equivalent region of other geminiviruses (Stanley et al., 1992).

It is noteworthy that in vitro translation of pTAL1/4 transcripts produced a protein of \( M_r \) 36K and, in a smaller amount, one of \( M_r \) 35K (Fig. 2, lane 1). The 36K protein was not produced from transcripts of either pTAL4 or pTALX (Fig. 2, lanes 2 and 3). The production of the 36K protein therefore depends on translation initiation from the first methionine codon of the AL1 ORF and this protein is probably a product of premature termination of translation of the AL1 ORF or degradation of the AL1 protein. The 35K protein was not produced from pTALX transcripts, but it was produced from pTAL4 transcripts and in greater amounts than from pTAL1/4 transcripts (Fig. 2, lanes 1 and 2). Production of the 35K protein therefore depends on initiation of the AL4 ORF. As this protein is larger than the AL4 protein, its formation must depend on translational readthrough or frameshift. Frameshift into the AL1 ORF is the only possibility to produce a protein of \( M_r \) 35K (Fig. 1). Evidence that the 35K in vitro translation product is the postulated AL4-AL1 fusion protein was provided by its specific precipitation by the AL1 antiserum (Fig. 3, lanes 3 and 4). Gröning et al. (1994) reported that expression of ORF AL4 in N. benthamiana protoplasts caused a 60% reduction of expression of the AL1 gene on another plasmid. If small amounts of an AL4-AL1 fusion protein were produced in the N. benthamiana protoplasts, this suppressive effect could have been caused by the fusion protein, rather than by the AL4 protein itself, since the AL1 protein is known to bind to a region of TGMV DNA that includes the AL1 core promoter (Fontes et al., 1992) and to suppress strongly transcription of the AL1 gene (Sunter et al., 1993; Gröning et al., 1994).

Translation of the AL2 and AL3 ORFs

In vitro translation of transcripts from pTAL2 which contains ORF AL2 (Fig. 1) produced a protein which migrated in SDS-PAGE as a doublet with apparent \( M_r \) of about 21K (Fig. 4, lane 2). This \( M_r \) is greater than the value predicted for the AL2 protein (15K). Since transcripts of pTAL2 cannot in theory produce a protein larger than the AL2 protein, this suggests that the AL2 protein migrates anomalously in SDS-PAGE. To test this, the AL2 protein was produced in bacteria transformed with pTAL2. The bacterially produced AL2 protein also migrated in SDS-PAGE with an apparent \( M_r \) around 21K (not shown), confirming its anomalous migration. The reason why the AL2 protein migrates as a doublet is not known. The slower migrating component of the doublet was always present in a lower amount than the faster migrating component and was sometimes barely detectable. One possibility is that one of the components is a phosphorylated form of the other. However translation of pTAL2 transcripts in the presence of \([\gamma-32P]ATP\) as the only labelled substrate did not result in labelling of either of the bands of the doublet (not shown).

Translation of transcripts of pTAL3, which contains ORF AL3 (Fig. 1), produced a protein migrating with an \( M_r \) of about 16K (Fig. 4, lane 3), the value predicted for the AL3 protein. The AL3 protein produced in bacteria transformed with pTAL3 also migrated with an apparent \( M_r \) of 16K (not shown). Interestingly translation of transcripts of pTAL2/3, which contains both of the overlapping AL2 and AL3 ORFs, gave both a doublet which comigrated with the 21K doublet produced by the pTAL2 transcripts and a band which comigrated with the 16K protein produced by the pTAL3 transcripts. It appears therefore that the AL2 and AL3 proteins can be produced together from a single RNA. The sequence context of the AL2 initiation codon (AUAAUGC) is suboptimal, whereas that of the AL3 initiation codon
(ACAAUGG) is optimal for translational initiation (Lütcke et al., 1987). Hence the AL3 protein may be produced by a leaky scanning mechanism, as postulated for the AL4 protein.

It is noteworthy that translation of pTAL2 transcripts produced, in addition to the 21K doublet, a major protein migrating with an apparent Mr of 14-5K (Fig. 4, lane 2), which was immunoprecipitated by an AL3 antiserum, but not by an AL2 antiserum (not shown). Since pTAL2 contains the 5'-terminal two-thirds of ORF AL3, it is therefore likely that the 14-5K protein represents this AL3 fragment, although its predicted Mr (10K) is lower than the apparent value observed from its migration rate. This interpretation is supported by the observation that the 14-5K protein was not produced from the pTAL2/3 transcripts (Fig. 4, lane 4), which contain the complete AL3 ORF.

The in vitro translation products from pTAL1/2/3/4a transcripts which contained ORFs AL1, AL2, AL3 and AL4 (Fig. 1) appeared to contain all the products obtained with the pTAL1/4 (Fig. 4, lane 1) and pTAL2/3 transcripts (Fig. 4, lane 4). The 40K and 10K products comigrated with the AL1 and AL4 proteins produced by the pTAL1/4 transcripts, and the 21K and 16K products comigrated with the AL2 and AL3 proteins produced by the pTAL2/3 transcripts. Analysis of the pTAL1/2/3/4a transcripts by denaturing formaldehyde–agarose gel electrophoresis indicated a single sharp band of the expected size (2-5 kb) with no evidence of degradation (not shown). Thus all four ORFs can be translated in vitro from a single RNA. Hanley-Bowdoin et al. (1989) showed that expression of an RNA equivalent to pTAL1/2/3/4a in transgenic tobacco plants complemented mutants in the AL1, AL2 or AL3 genes in trans, indicating expression of these ORFs from a single RNA.

The efficiency of translation of eukaryotic mRNAs can depend on, amongst other things, the presence or absence of a cap structure at the 5' terminus and the nature of the 5' untranslated sequence (Gallie, 1993). Although the mRNAs of many viruses are capped, those of a substantial number of viruses are not capped. It is not known whether TGMV transcripts are capped in vivo, but in vitro translation of transcripts of pTAL1, pTAL2, pTAL3, pTAL2/3, pTAL1/2/3/4a and pTAL1/2/3/4b was not significantly affected by the presence or absence of a 7mGpppG cap at their 5' ends (Fig. 5, compare lanes 1, 3, 5, 7, 9 and 11 with lanes 2, 4, 6, 8, 10 and 12). There were small differences in the amounts of the AL2 and AL3 proteins produced by pTAL1/2/3/4a or b transcripts in different experiments, but these were not related to capping.

Even in the absence of a cap, the 5' leader can still influence the efficiency of translational initiation as shown by the enhancing effect of the tobacco mosaic virus Ω sequence on the in vitro translation of uncapped RNAs (Sleat et al., 1987). To determine if this was the case for the initiation of translation of the AL1 ORF, two leader sequences were compared. The leader sequence of pTAL1/2/3/4b transcripts contained about
50 nucleotides of the TGMV sequence upstream of the AL1 translational initiation codon, whereas pTAL1/2/3/4a transcripts contained an additional 150 upstream nucleotides of the TGMV sequence, including a sequence potentially capable of forming a stable stem–loop structure. The results of in vitro translation in wheatgerm extract (Fig. 5, lanes 3, 4, 5 and 6) indicated that there was not a large difference between the amounts of the AL1 protein produced by the pTAL1/2/3/4b transcripts and the pTAL1/2/3/4a transcripts. However, in the rabbit reticulocyte lysate, the amount of AL1 protein produced by the pTAL1/2/3/4b transcripts was at least five times as great as that produced by the pTAL1/2/3/4a transcripts (not shown). The lower translational efficiency of the transcripts with the longer leader sequence may reflect the ability of the leader to form a stable secondary structure. The difference between the wheatgerm and reticulocyte lysate in this respect could be due to different contents of translational initiation factors that have the ability to unwind RNA secondary structure (Jaramillo et al., 1990; Gallie, 1993). The nature of the leader did not appear to have a great effect on the translation of the internal ORFs.

Further work will be required to determine the extent to which the in vitro translation results reported here reflect the translational events that occur in virus-infected plants. Our in vitro translation results, together with those employing transgenic tobacco plants (Hanley-Bowdoin et al., 1989), suggest that the longest transcript detected in extracts of TGMV-infected plants (initiating at nucleotides 61 or 62; Sunter & Bisaro, 1989; Hanley-Bowdoin et al., 1989) is capable of directing the synthesis, not only of the AL1 protein, but also of the AL2, AL3 and AL4 proteins. Shorter transcripts which might act as mRNAs for the AL2, AL3 or AL4 proteins have also been detected in extracts of TGMV-infected plants (see the Introduction), but the extent to which some of these may be artefacts of the procedures used to map them remains unclear (Lazarowitz, 1992). The high efficiency with which the AL4 protein was translated from the pAL1/4 transcripts and with which the AL3 protein was translated from the pAL2/3 transcripts suggests that the transcripts initiating at nucleotides 61 or 62 (AL62) and 1629 (AL1629) respectively may be the mRNAs responsible for the synthesis of the AL1 and AL4, and the AL2 and AL3 proteins respectively in virus-infected plants. It is noteworthy that potential core promoter sequences are found 25 to 30 nucleotides upstream of the 5' ends of these two transcripts (Hamilton et al., 1984). Furthermore transcripts equivalent to AL62 and AL1629 (but not to the other leftward TGMV transcripts) were detected in extracts of plants infected with the related geminiviruses, Abutilon mosaic virus (Frischmuth et al., 1991) and tomato leaf curl virus (Mullineaux et al., 1993). Two transcripts which mapped to the leftward region of DNA A of another related geminivirus, African cassava mosaic virus, may also be equivalent (Townsend et al., 1985). Hence translation of the ORFs of the leftward region of DNA A by messenger RNAs equivalent to AL62 and AL1629 may be typical of the whitefly-transmitted geminiviruses.

It is noteworthy that the AL4 protein was synthesized in a large excess over the AL1 protein by both the pTAL1/4 and pTAL1/2/3/4 transcripts. If a similarly large amount is produced in infected plants, it would be anticipated that the protein should play an important role in the virus replication cycle. Although the AL4 ORF did not appear to be required for TGMV replication or symptom development in N. benthamiana (Elmer et al., 1989), a permissive host for many plant viruses (Dawson & Hilf, 1992), it is possible that it has important functions in other hosts, such as tomato.

We thank the European Molecular Biology Organization for the award of a long term fellowship to P. A. Thümmes.

References


In vitro synthesis of TGMV proteins 1833


(Received 16 December 1993; Accepted 1 March 1994)