The EMBO Journal vol.13 no.9 pp.2200–2210, 1994

The contribution of AAUAAA and the upstream element UUUGUA to the efficiency of mRNA 3′-end formation in plants

Helen M.Rothnie, Jacqueline Reid and Thomas Hohn

Friedrich Miescher Institute, PO Box 2543, CH-4002 Basel, Switzerland

Communicated by T.Hohn

The requirement for sequence specificity in the AAUAAA motif of the cauliflower mosaic virus (CaMV) polyadenylation signal was examined by saturation mutagenesis. While deletion of AAUAAA almost abolished processing at the CaMV polyadenylation site, none of the 18 possible single base mutations had a dramatic effect on processing efficiency. The effect of replacing all six nucleotides simultaneously varied depending on the sequence used, but some replacements were as detrimental as the deletion mutant. Taken together, these results confirm that AAUAAA is an essential component of the CaMV polyadenylation signal, but indicate that a high degree of sequence variation can be tolerated. A repeated UUUGUA motif was identified as an important upstream accessory element of the CaMV polyadenylation signal. This sequence was able to induce processing at a heterologous polyadenylation site in a sequence-specific and additive manner. The effect of altering the spacing between this upstream element and the AAUAAA was examined; moving these two elements closer together or further apart reduces the processing efficiency. The upstream element does not function to signal processing at the CaMV polyadenylation site if placed downstream of the cleavage site. Analysis of further upstream sequences revealed that almost all of the 200 nt fragment required for maximal processing contributes positively to processing efficiency. Furthermore, isolated far upstream sequences distinct from UUUGUA were also able to induce processing at a heterologous polyadenylation site.

Key words: 3′-end formation/AAUAAA/cauliflower mosaic virus/pararetrovirus/plant mRNA/polyadenylation/upstream elements

Introduction

One of the essential steps in the maturation of mRNA in eukaryotic cells is the formation of the 3′-end. In vertebrate systems, the *cis*-acting signals on the RNA and some of the *trans*-acting factors involved in these reactions are now well characterized (reviewed by Wahle, 1992; Wahle and Keller, 1992). The mature 3′-end of RNA polymerase II transcripts is formed by two tightly coupled reactions: an endonucleolytic cleavage followed by the addition of a poly(A) tail. Correct processing depends on the presence of the highly conserved polyadenylation signal, AAUAAA. This sequence is almost invariably present 10–35 bases upstream from the site of cleavage (Proudfoot and Brownlee, 1976) and is essential for both cleavage and polyadenylation *in vivo* and *in vitro* (Fitzgerald and Shenk, 1981; Wickens and Stephenson, 1984; Manley *et al.*, 1985; Zarkower *et al.*, 1986; Wilusz *et al.*, 1989; Sheets *et al.*, 1990). A second element downstream of the cleavage site is also required. This downstream element (DSE) is less well conserved, but is generally GU- or U-rich (Gil and Proudfoot, 1984, 1987; Hart *et al.*, 1985; McDevitt *et al.*, 1986). The spacing between the AAUAAA and the DSE is important in determining the efficiency of 3′-end formation (Gil and Proudfoot, 1987; Heath *et al.*, 1990; Weiss *et al.*, 1991). A synthetic polyadenylation signal comprising an AAUAAA motif and a DSE flanking the cleavage site has been shown to be sufficient to direct correct 3′-end processing (Levitt *et al.*, 1989). Such an element represents a 'core' polyadenylation signal. In some cases, additional accessory elements upstream of this core polyadenylation signal enhance processing. To date, examples of this phenomenon have been studied almost exclusively in viral systems, e.g. SV40 (Carswell and Alwine, 1989; Schek *et al.*, 1992), adenovirus (DeZazzo and Imperiale, 1989) and in retroviruses and other retroelements which have to differentiate between two apparently identical core polyadenylation sites (reviewed by Imperiale and DeZazzo, 1991; Rothnie *et al.*, 1994).

In the yeast *Saccharomyces cerevisiae*, the *cis*-acting signals required for 3′-end formation are less well defined. The vertebrate AAUAAA signal does not function as a polyadenylation signal in yeast (Butler *et al.*, 1990; Hyman *et al.*, 1991). Several types of *cis*-acting element have been proposed: UUUUUAUA (Henikoff and Cohen, 1984), alternating AU stretches (Abe *et al.*, 1990), sequences related to a bipartite UAG....UAUGUA motif (Zaret and Sherman, 1982; Irniger *et al.*, 1991; Russo *et al.*, 1991; Hou *et al.*, 1994; Irniger and Braus, 1994), and regions containing dyad symmetry (Sadhale and Platt, 1992). It seems likely that a number of different classes of yeast polyadenylation signal exist, with some classes able to function in either orientation (Irniger *et al.*, 1991; Heidmann *et al.*, 1992; Sadhale and Platt, 1992). Yeast transcripts can be accurately processed in yeast whole cell extracts (Butler and Platt, 1988; Abe *et al.*, 1990; Butler *et al.*, 1990; Sadhale *et al.*, 1991). Biochemical studies have characterized at least four chromatographically separable components which are required for the reaction (Chen and Moore, 1992), yeast poly(A) polymerase has been purified (Lingner *et al.*, 1991a) and the corresponding gene cloned (Lingner *et al.*, 1991b).

In contrast to the current state of knowledge in animal and yeast systems, little is known about the process of mRNA 3′-end formation in plants. The finding that mammalian polyadenylation signals are not properly recognized in plants

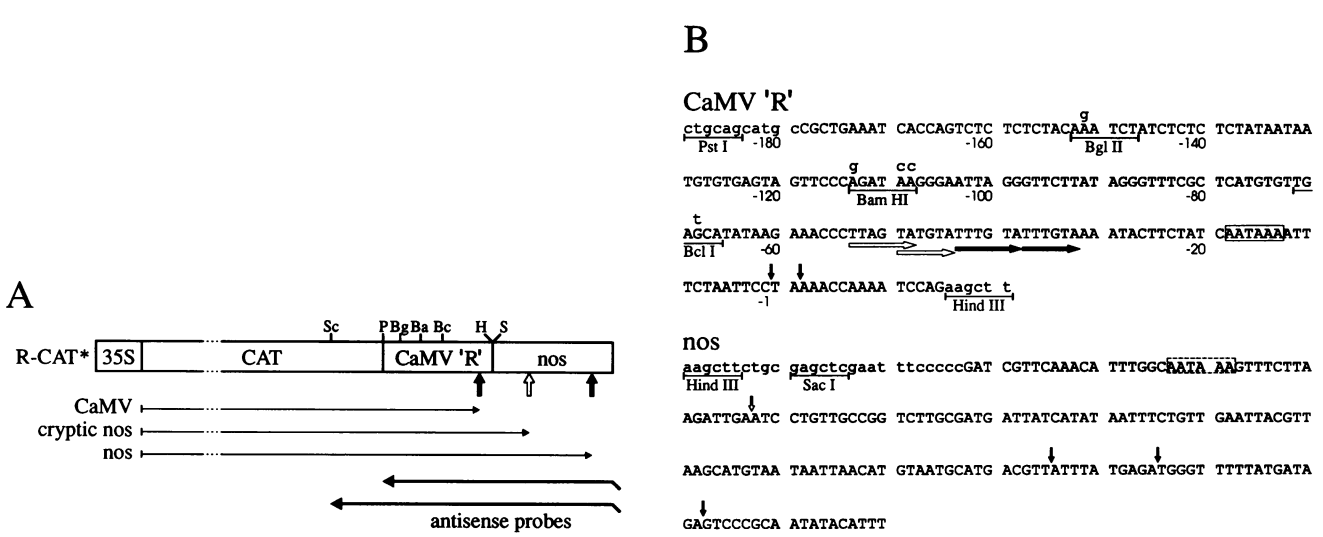


Fig. 1. (**A**) Experimental strategy. Plasmid R-CAT* contains the CAT gene between the CaMV 35S promoter and two polyadenylation signals in tandem: CaMV and *nos*. After transient expression in transfected plant protoplasts, transcripts produced from R-CAT* or mutant derivatives were analysed by RNase A/T1 mapping using homologous antisense probes. Filled arrows indicate the positions of the normal processing sites within the CaMV and *nos* polyadenylation signals. The open arrow indicates the position of the cryptic *nos* site. The thin lines below R-CAT* represent transcripts processed at the CaMV site, the cryptic *nos* site, or the normal *nos* sites. The thick lines represent the extent of the antisense probes used in the protection analysis. Restriction sites relevant to the construction of mutants described here are indicated: Ba (*Bam*HI), Bc (*Bcl*I), Bg (*Bgl*II), H (*Hin*dIII), P (*Pst*I), S (*Sac*I), Sc (*Sca*I). (**B**) Primary DNA sequence of CaMV and *nos* polyadenylation signals in R-CAT*. The original sequence of each region is in upper case letters; lower case letters show polylinker sequences or mutations introduced to create restriction sites at the positions shown. Solid vertical black arrows above the sequence indicate the major processing sites in the CaMV and *nos* fragments as mapped by RNase A/T1 protection analysis in this study. An open vertical arrow indicates the position to which the cryptic *nos* site was mapped. The CaMV sequence is numbered relative to the end point of the shorter of the two major protected fragments (corresponding to processing at the CaMV site), which is designated −1. The AATAAA motifs upstream of the CaMV and the cryptic *nos* processing sites are enclosed by solid and dashed lines, respectively. Solid or open arrows below the sequence from −53 to −32 indicate two perfect and imperfect repeats of the sequence TTTGTA. Restriction sites are as indicated and the same *Hin*dIII site is shown at the 3′-end of the CaMV sequence and the 5′-end of the *nos* sequence.

(Hunt *et al.*, 1987) suggests important functional differences in the *cis*-acting signals required, and maybe also in the mechanism. The 3′-ends of mRNAs from a single plant gene can be quite heterogeneous, indicating multiple cleavage sites (Dean *et al.*, 1986). This contrasts with animals, where each polyadenylation signal generally directs processing at a single cleavage site. The 'canonical' polyadenylation signal, AAUAAA, is not present near the 3′-end of many plant genes; one analysis revealed the presence of this motif at an appropriate position in only 40% of plant genes (Joshi, 1987). AAUAAA is not necessarily active as part of the polyadenylation signal even when it is present (Sanfaçon, 1994), and variant signals can be recognized (Wu *et al.*, 1993). Another difference from the animal system is the apparent lack of a DSE as part of the polyadenylation signal. Rather, sequences upstream of the cleavage site contain the important *cis*-acting signals (Mogen *et al.*, 1990; Guerineau *et al.*, 1991; MacDonald *et al.*, 1991; Sanfaçon *et al.*, 1991; Wu *et al.*, 1993; Sanfaçon, 1994). Recent studies suggest that plant polyadenylation signals are more complex and diffuse than in animals, with sequences close to the cleavage site (near upstream elements, NUEs) acting as the functional equivalent of an AAUAAA signal and regions further upstream (far upstream elements, FUEs) playing an accessory role in regulating overall processing efficiency (Mogen *et al.*, 1992). Furthermore, the FUE of one polyadenylation signal can act in conjunction with the NUE of another (Mogen *et al.*, 1992; Sanfaçon, 1994). However, as in yeast, it appears that there is no single specific sequence element which is universally required as part of a plant polyadenylation signal.

The polyadenylation signal of the plant pararetrovirus cauliflower mosaic virus (CaMV), is one of the few cases where the *cis*-acting sequences necessary for correct 3′-end processing have been studied in some detail. Although an AAUAAA motif is not found in most plant polyadenylation signals, it occurs at an appropriate position in CaMV and has been shown by deletion to be essential for processing (Sanfaçon and Hohn, 1990; Sanfaçon *et al.*, 1991). However, two mutants (AAGAAA and UAGAAU) were tolerated, indicating a lesser degree of sequence specificity than in vertebrates (Mogen *et al.*, 1990; Sanfaçon *et al.*, 1991). The AAUAAA motif alone is not sufficient to signal 3′-end processing; sequences further upstream are also required (Mogen *et al.*, 1990; Guerineau *et al.*, 1991; Sanfaçon *et al.*, 1991). An important upstream element was localized to a U-rich region situated just upstream of AAUAAA (Sanfaçon *et al.*, 1991).

This report further defines the *cis*-acting elements contributing to efficient processing at the CaMV polyadenylation site. The sequence specificity of the AAUAAA motif was rigorously tested by saturation mutagenesis, an upstream element comprising a repeated UUUGUA motif was identified, and several properties of this and other upstream accessory elements were characterized.

Results

The basic construct used in the analysis of the CaMV polyadenylation signal in this study was R-CAT* (Figure 1A). In this plasmid, the chloramphenicol acetyl transferase (CAT) gene is under the control of the CaMV 35S promoter. A sequence corresponding to the terminal

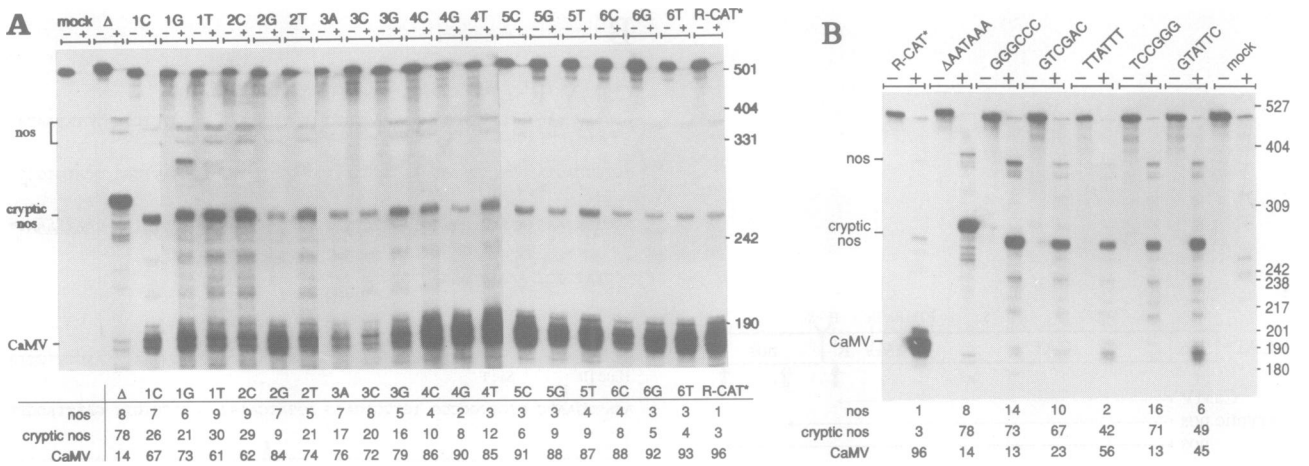


Fig. 2. Mutations in the AATAAA motif. (**A**) RNase A/T1 protection analysis of transcripts produced from constructs carrying either the wild-type CaMV polyadenylation signal (R-CAT*), a mutant in which the AATAAA is deleted (Δ), or one of the 18 possible single base mutations of AATAAA (1C to 6T). (**B**) Mutants in which AATAAA has been replaced with an unrelated sequence (the sequence replacing AATAAA in each case is indicated above the gel). A separate antisense probe was prepared for each mutant (from each plasmid linearized with *Pst*I) and in each case RNA hybrids are shown before (−) and after (+) digestion with RNases A and T1. Protected fragments corresponding to transcripts processed at the CaMV site, the cryptic *nos* site and the *nos* site are indicated. The processing efficiencies at the CaMV, cryptic *nos* and *nos* sites are tabulated beneath each panel, expressed as a percentage of the total for each construct. Transcripts processed at the cryptic *nos* and *nos* sites in lane Δ (A) and ΔAATAAA (B) are longer than in the other lanes due to extra polylinker sequences between the CaMV and the *nos* regions in this construct. Molecular weight markers correspond to ^{32}P-labelled DNA fragments [pUC19/*Msp*I (A), pBR322/*Msp*I (B)].

redundancy (R) of the CaMV 35S RNA (Guilley *et al.*, 1982) plus ~15 nt of downstream sequence follow the CAT gene. This sequence has been previously shown to be sufficient to direct maximally efficient processing at the CaMV polyadenylation site (Sanfaçon and Hohn, 1990). In R-CAT*, several point mutations have been introduced into the wild-type sequence to create three unique restriction sites within the R sequence to facilitate further mutagenesis (Figure 1B). These changes did not affect processing efficiency (data not shown). The CaMV R region is followed by a second polyadenylation signal, from the nopaline synthase (*nos*) gene (Bevan *et al.*, 1982; Depicker *et al.*, 1982) which acts as a trap for transcripts not processed in mutant CaMV polyadenylation sequences. In addition to processing at the normal *nos* polyadenylation sites, some transcripts are processed at a cryptic site in the *nos* region (described in Sanfaçon and Hohn, 1990; Sanfaçon *et al.*, 1991). This site (cryptic *nos*) is ~120 nt upstream of the normal processing site, just downstream of an AATAAA motif (Figure 1B). The proportion of transcripts processed at the CaMV site relative to those processed at the *nos* polyadenylation sites was used as a measure of the processing efficiency at the CaMV site.

Briefly, the experimental strategy consisted of preparing mutant derivatives of the R region in plasmid R-CAT*, and testing these for efficiency of processing by RNase protection analysis of transcripts produced in transiently transfected *Nicotiana plumbaginifolia* protoplasts.

Saturation mutagenesis of the AATAAA motif

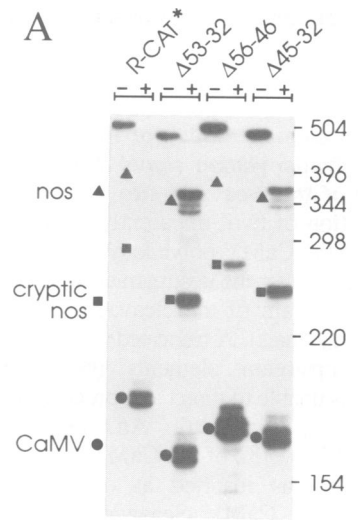
To test the flexibility of the plant polyadenylation machinery to recognize variants of the AATAAA motif, all 18 possible single base mutations of this signal were introduced into the CaMV polyadenylation signal, and were tested for their ability to direct processing at the CaMV site. As shown in Figure 2A, all of these mutations allowed at least 60% of the wild-type processing efficiency at the CaMV site (see also Figure 9). Mutations in the first three bases of the signal had a marginally greater effect than those in the last three positions. Although deletion of AATAAA almost abolishes processing, the point mutation results indicate that all single base variants of the AATAAA motif are recognized by the plant polyadenylation machinery and can direct efficient and accurate processing at the CaMV polyadenylation site.

If AATAAA is indeed an essential *cis*-acting component of the CaMV polyadenylation signal (as suggested by the deletion mutant), despite being relatively insensitive to single base changes (as suggested by the point mutation analysis), it would be predicted that completely replacing AATAAA with another sequence would severely reduce processing efficiency. To test this prediction, a further series of mutants was constructed in which AATAAA was replaced with several different unrelated 6 nt sequences. As shown in Figure 2B, the effect of some of these mutations was comparable with that of the deletion mutant, i.e. processing at the CaMV site was almost abolished. However, certain replacement mutants still supported ~50% of wild-type processing efficiency at the CaMV site. As with the point mutation analysis, these results suggest that AATAAA is required but that a surprisingly high degree of sequence variation can be tolerated by the processing machinery. Although only a few replacement mutants were examined, it appeared that (A+T)-rich sequences were acceptable substitutes for AATAAA, while (G+C)-rich sequences were not recognized at all. The possible implications of these results will be considered further in the Discussion.

With the construct in which the AATAAA motif is deleted (Δ), some transcripts processed at the CaMV site were in fact observed, indicating that even in the absence of AATAAA, the CaMV polyadenylation signalling region can direct a low level of processing at the correct site.

An important upstream activating element is composed of a repeated TTTGTA motif

Deletion analysis of the CaMV polyadenylation signal implicated sequences between −53 and −32 as being part



B

	R-CAT*	Δ53-32	Δ56-46	Δ45-32
nos	1	19	2	13
cryptic nos	3	23	7	20
CaMV	96	58	91	67

Fig. 3. Internal deletions define an active upstream element. **(A)** Protection assay of constructs carrying deletions upstream of the AATAAA. The constructs are named according to the deleted nucleotides. Protected fragments corresponding to transcripts processed at the CaMV, cryptic *nos* and *nos* sites are indicated. Molecular weight markers were ^{32}P-labelled DNA fragments (pBR322/*Hin*fI). **(B)** Processing efficiencies at the CaMV, cryptic *nos* and *nos* sites, expressed as a percentage of the total for each construct.

of an important upstream element (Sanfaçon *et al.*, 1991). This latter analysis showed the importance of this region in the absence of sequences further upstream. The results presented in Figure 3 show the effect of specific deletions within this sequence in the context of the entire R region. Deletion of nucleotides −53 to −32 (Δ53−32) resulted in a drop of processing efficiency to ~60% of wild-type. Two additional deletions, Δ45−32 and Δ56−46, showed that this effect was predominantly due to sequences between positions −45 and −32. This sequence contains a direct repeat of the motif TTTGTA. The deletion mutant Δ45−32 demonstrates that removal of this sequence significantly reduces the efficiency of processing at the CaMV site, although the remaining activity (>60%) indicates that other upstream sequences are important in determining recognition of the CaMV polyadenylation site.

Induction of cryptic nos by TTTGTA

Specific sequences upstream from the CaMV polyadenylation site can induce recognition of a cryptic site within the *nos* polyadenylation signal (Sanfaçon *et al.*, 1991). This 'induction assay' was used to test whether the TTTGTA motif alone was sufficient to induce recognition of cryptic *nos* and whether the effect was additive. A low, but reproducible, level of processing at the cryptic *nos* site was observed upon introduction of a single copy of the TTTGTA motif upstream of the *nos* polyadenylation region (Figure 4). Introduction of two, three or four tandem copies of TTTGTA increased the level, with four copies in tandem giving higher levels of processing at the cryptic *nos* site than an oligonucleotide comprising CaMV R sequences −53 to −32. Comparison with the results presented in Figure 6 demonstrate that this induction is orientation-specific. It could be argued that induction of cryptic *nos* depends simply on the presence of a T-rich sequence or on a stretch of

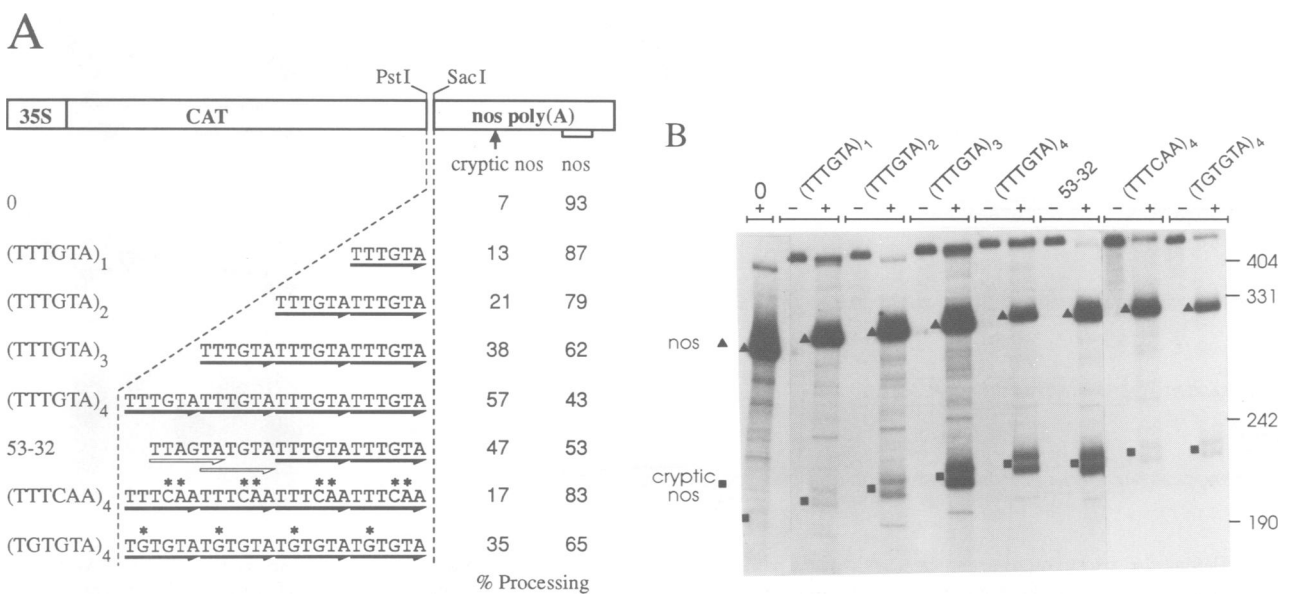


Fig. 4. The sequence TTTGTA specifically induces processing at the cryptic *nos* site. **(A)** Schematic representation of the constructs used to test induction of cryptic *nos* by TTTGTA or mutant derivatives. Synthetic oligonucleotides carrying one, two, three or four copies of TTTGTA, the sequence corresponding to −53 to −32, or four copies of the motifs TTTCAA or TGTGTA were inserted between the *Pst*I and *Sac*I sites as shown. The efficiency of processing at the cryptic *nos* and *nos* sites is indicated for each construct. **(B)** RNase protection analysis of transcripts produced by the constructs depicted in (A). The positions of protected fragments corresponding to transcripts processed at the cryptic *nos* and *nos* sites are indicated. Molecular weight markers were ^{32}P-labelled DNA fragments (pUC19/*Msp*I).

alternating T residues. To test the sequence specificity of the TTTGTA motif, mutations were introduced into the context of the tetramer in the induction assay. Both mutations tested (TTTCAA and TGTGTA) gave rise to only very low levels of processing at the cryptic *nos* site thus showing that, at least in this context, induction specifically depended on a repeated TTTGTA sequence (Figure 4).

Altering the spacing between the upstream element and the AATAAA decreases processing efficiency

If a cleavage/polyadenylation complex forms on plant pre-mRNAs, it seems reasonable to assume that specific *cis*-elements are the recognition and/or binding sites for factors within this complex. In the case of the CaMV polyadenylation site, these specific elements might involve AATAAA and the TTTGTA-containing upstream accessory element defined above. It would thus be predicted that changing the spacing between the AATAAA and the upstream element might interfere with complex formation and thus reduce processing efficiency. To test this prediction, a series of constructs was made in which the distance between the upstream element and the AATAAA was varied from 0 to 25 bases (wild-type spacing = 13; Figure 5A). Moving the AATAAA either closer to, or further away from, the upstream accessory sequences reduced the efficiency of processing at the CaMV site, allowing more readthrough to the *nos* polyadenylation region (Figure 5B). In introducing the *Sal*I site, the A at −32 is changed to a C. As shown above, the A may be important as part of the tandemly repeated TTTGTA motif and the A to C mutation in itself may lower the processing efficiency. However, processing in construct pSX is almost as efficient as in the wild-type (R-CAT*) and since a C is present at this position in plasmids pSX, pDSX and pSX2, the reduction in processing efficiency on altering the distance to the AATAAA is attributable to a 'spacing' effect rather than to this mutation in the primary sequence.

The upstream element does not function as part of the CaMV polyadenylation signal if placed downstream of the cleavage site

The identification of two, apparently separable, *cis*-acting components of the CaMV polyadenylation signal prompted the question of whether the upstream element would be active if placed downstream of the cleavage site (cf. vertebrate polyadenylation signals). A truncated CaMV polyadenylation site lacking upstream elements (but still containing AATAAA) was unable to direct any processing at the CaMV site [US/DS(−), Figure 6]. An oligo encompassing nucleotides −53 to −32 of the CaMV upstream sequences (oligo 53−32) was inserted at one of two positions downstream of the CaMV cleavage site in both orientations [US/DSH(→), US/DSH(←), US/DSS(→), and US/DSS(←), Figure 6]. RNA from these constructs was not processed at the CaMV site. Processing occurs predominantly at the cryptic *nos* site if oligo 53−32 is in its original orientation, and predominantly at the *nos* site if the oligo is reversed or absent. A series of derivatives of these five constructs in which 22 bp between the CaMV cleavage site and the *Hin*dIII site was removed (thus bringing oligo 53−32 closer to the cleavage site and to the CaMV AATAAA) gave the same result (data not shown), i.e. none of the constructs resulted in processing at the CaMV site, and in all cases there was orientation-dependent induction of the cryptic *nos* site. These results indicate that this element must be present upstream of the cleavage site for correct recognition in the context of the CaMV polyadenylation site, although the possibility remains that positioning the element between two polyadenylation sites (those of the CaMV and

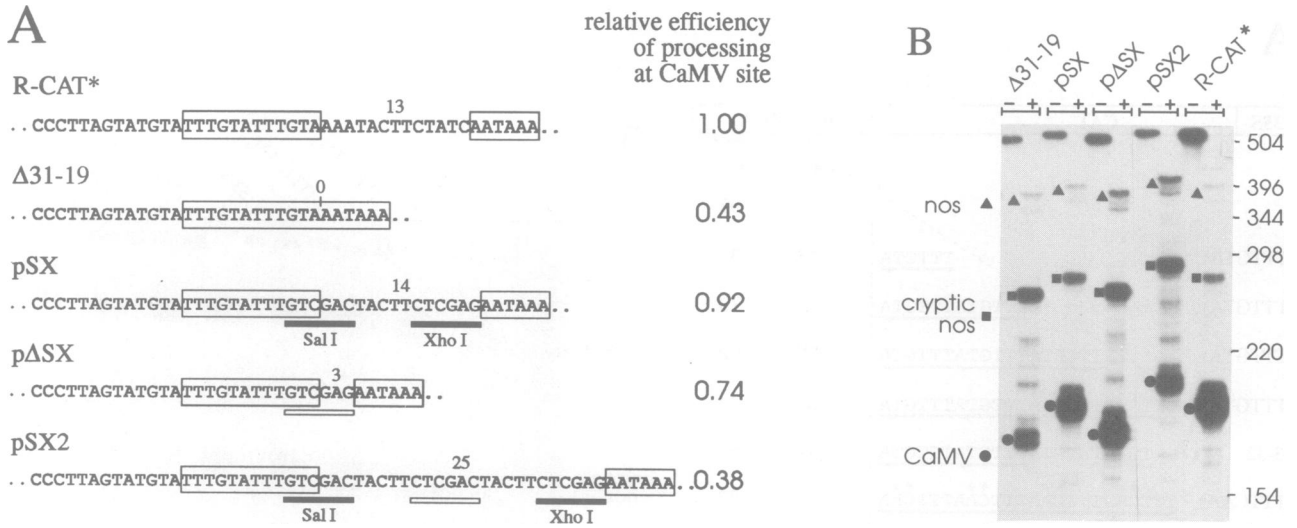


Fig. 5. Altering the spacing of *cis*-acting elements reduces processing efficiency. (**A**) The effect of altering the distance between the AATAAA and the upstream element was tested in the context of the whole R region. Only the relevant parts of the sequence are shown in the figure. The AATAAA and the repeated TTTGTA motif are boxed. *Sal*I and *Xho*I sites which were introduced into the R-CAT* construct are underlined with thick black lines. The junctions formed when *Sal*I and *Xho*I ends were ligated together are underlined with open lines. The distance in nucleotides between the upstream element and the AATAAA is indicated. The relative efficiency of processing at the CaMV site is shown. (**B**) RNase protection analysis of transcripts produced by the constructs shown in (A). Homologous antisense probes linearized at the *Pst*I site of each construct were used to map the transcripts. RNA hybrids are shown before (−) and after (+) digestion with RNases A and T1. Protected fragments corresponding to transcripts processed at the CaMV, cryptic *nos*, and *nos* sites are indicated. Molecular weight markers were ^{32}P-labelled DNA fragments (pBR322/*Hin*fI).

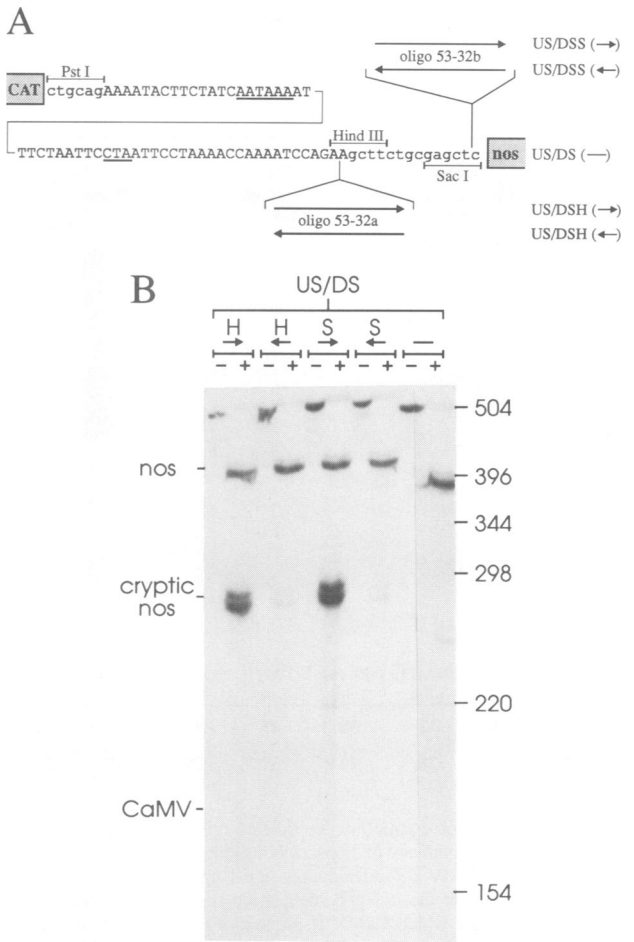


Fig. 6. The upstream element does not function downstream of the CaMV cleavage site. (**A**) Constructs to test the effect of moving the upstream activating element to a position downstream of the cleavage site. US/DS(−) is a truncated CaMV polyadenylation signal construct lacking all upstream activating elements. US/DSH(→) and US/DSH(←) have oligo 53−32 inserted into the *Hin*dIII site 25 bp downstream of the CaMV cleavage site in the sense and antisense orientation, respectively. Similarly, US/DSS(→) and US/DSS(←) have the same oligo inserted into the *Sac*I site a further 14 bp downstream. (**B**) Protection assay of RNAs produced from the constructs shown in A. The expected positions of protected fragments corresponding to transcripts processed at the CaMV, cryptic *nos* and *nos* sites are shown. Transcripts were mapped using homologous antisense probes linearized at the *Sca*I site in the CAT gene (see Figure 1A). RNA hybrids are shown before (−) and after (+) digestion with RNase T1. RNase T1 only was used in this experiment as RNase A gave rise to additional bands due to 'breathing' of the RNA hybrids in (A + U)-rich regions. Molecular weight markers correspond to ^{32}P-labelled DNA fragments (pBR322/*Hin*fII).

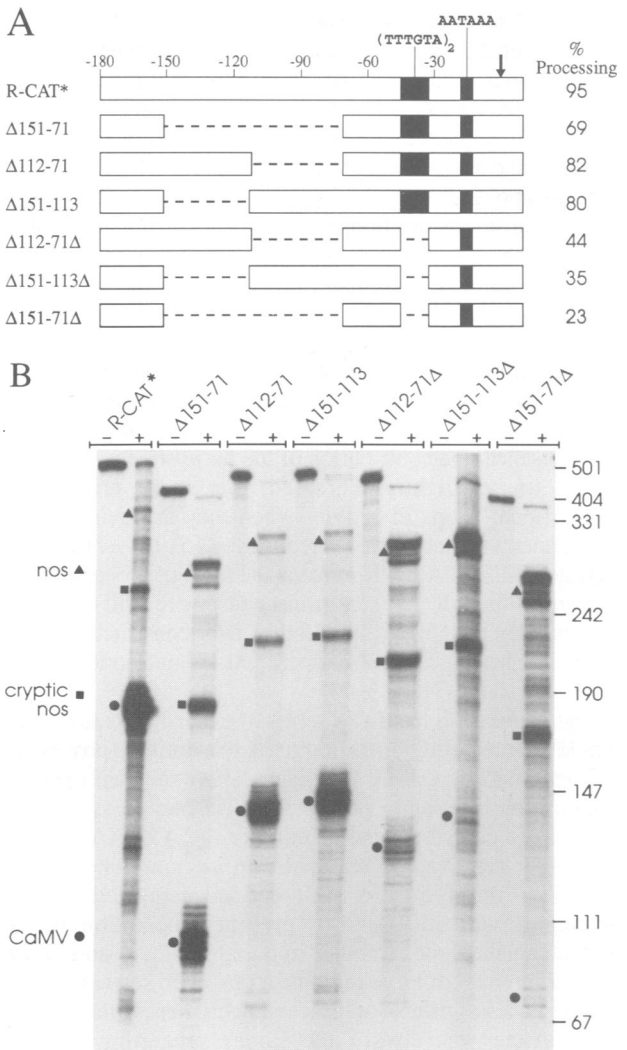


Fig. 7. Contribution of upstream sequences other than TTTGTA. (**A**) Diagrammatic representations of constructs designed to test the contribution of upstream sequences to processing efficiency at the CaMV site (vertical arrow). The regions deleted from each construct are depicted by dashed lines. The repeated TTTGTA motif and the AATAAA are shown as shaded blocks. Processing efficiencies at the CaMV site are as indicated. (**B**) RNase protection analysis of constructs shown in A. RNA hybrids are shown before (−) and after (+) RNase A/T1 digestion and protected fragments corresponding to transcripts processed at the CaMV, cryptic *nos* and *nos* sites are indicated. Molecular weight markers were ^{32}P-labelled DNA fragments (pUC19/*Msp*I).

cryptic *nos* sites), favours the use of the downstream site over the upstream one.

Other upstream sequences also contribute to processing efficiency

In the context of the whole R region, deletion of the repeated TTTGTA motif still allows processing of >60% of the transcripts. Obviously, other upstream sequences are also exerting some positive effect. In order to study these further, deletions were introduced into R by removing sequences between combinations of the *Bgl*II, *Bam*HI and *Bcl*I sites as described in Materials and methods. A construct in which the extreme 5′-end of the R region is deleted (up to position −149; results in a processing efficiency at the CaMV site of ~87% of wild-type (construct Δ149; Sanfaçon *et al.*, 1991). Results shown in Figure 7 illustrate the result of deleting sequences between positions −151 to −71, −112 to −71 and −151 to −113. All these deletions reduce efficiency of processing at the CaMV site slightly, with the smaller two deletions dropping to ~80% of wild-type and the larger deletion, combining the other two, a further 10% to ~70%. These deletions arbitrarily define three upstream regions of ~40 nt each, which together contribute to the wild-type processing efficiency to the extent of ~40% [a construct in which this whole upstream region is missing gives rise to a processing efficiency of ~60% (Sanfaçon *et al.*, 1991)]. Taken individually, positions −184 to −152 contribute the least (~10%), while regions −151 to −113

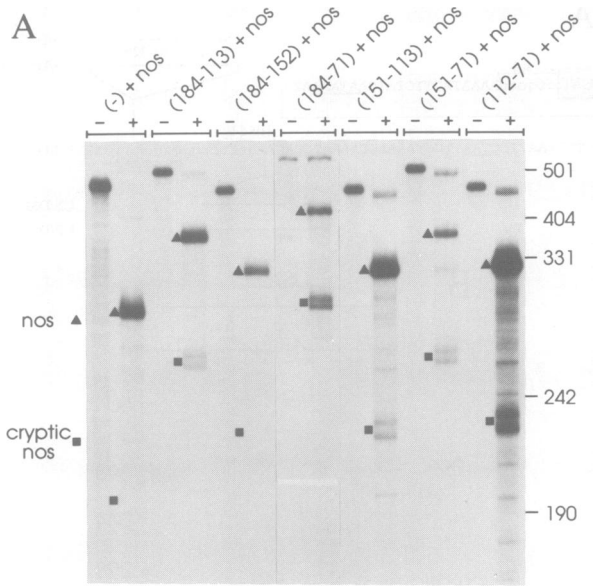
and −112 to −71 both result in a 20% loss of efficiency when deleted separately, and a 30% drop when deleted together (but it has to be borne in mind that these specific deletions may also interrupt *cis*-acting elements). There are no obvious sequence motifs common to each of these three regions, and no elements which match the TTTGTA motif defined above.

These upstream deletions were further tested in the context of the Δ45−32 deletion to assess the ability of different upstream sequences to support processing at the CaMV site in the absence of the TTTGTA motif. Again, individual removal of regions −151 to −113 and −112 to −71 further reduced processing efficiency, suggesting that both these regions have an overall positive effect on processing and that their action is independent of the TTTGTA motif. In fact, in its absence, the contribution of the far upstream sequences seems to be slightly increased, i.e. there are greater proportional drops in activity when they are deleted. The simultaneous deletion of nucleotides −151 to −71 and the repeated TTTGTA motif reduced the level of processing still further, although 23% of transcripts were still correctly processed at the CaMV site. In this latter construct, the only remaining sequences upstream of −31 are nucleotides −184 to −152 and −71 to −46.

The conclusion from this series of experiments is that the whole of the R region is required for optimal processing efficiency at the CaMV polyadenylation site, with several regions contributing to processing efficiency in an additive manner. In contrast to the deletion of the AATAAA motif, no single removal of an upstream region, whether the TTTGTA motif or less well defined sequences further upstream, was sufficient to prevent recognition of the polyadenylation site. Another important point to note is that despite removal of large regions of upstream sequence, with greater or lesser effects on processing efficiency, the position of cleavage was always the same, suggesting that the upstream elements are contributing to the efficiency of the processing reaction, without having an influence on the site of processing itself.

Induction of cryptic nos by isolated upstream sequences from the CaMV R region

Sequences from the CaMV R upstream region were tested for the ability to induce processing at the cryptic *nos* site. The absence of other sequences from the CaMV R region in this assay allows the ability of defined fragments to induce processing at the cryptic *nos* site to be compared directly. A *Bcl*I−*Sac*I fragment (containing the TTTGTA upstream element, the AATAAA and the CaMV polyadenylation site) was deleted from R-CAT*, allowing remaining regions of R (delineated by the *Pst*I, *Bgl*II, *Bam*HI and *Bcl*I sites) to be tested either alone or in combination for their ability to induce processing at the cryptic *nos* site. In agreement with the results presented above, all of the regions of upstream sequence tested (184−152, 151−113 and 112−71) were able to induce processing to some extent (Figure 8). The higher levels of processing at the cryptic *nos* site induced by combinations of two or three of these sequence elements (184−113, 151−71 and 184−71) were approximately additive, when compared with each fragment in isolation. The longest fragment used in the assay (184−71) gave a level of induction higher than the −53 to −32 element alone, and is equivalent to the effect of four perfect tandem copies



B

	(-) + nos	(184-113) + nos	(184-152) + nos	(184-71) + nos	(151-113) + nos	(151-71) + nos	(112-71) + nos
nos	93	72	88	43	80	53	74
cryptic nos	6	28	12	57	20	47	26

Fig. 8. Induction of cryptic *nos* by isolated upstream sequences. (**A**) RNase protection analysis of transcripts produced from constructs in which various CaMV polyadenylation signal upstream regions are present upstream of the *nos* polyadenylation region. The names of the constructs indicate which CaMV R sequences are upstream of *nos*. RNA hybrids are shown before (−) and after (+) RNase T1 digestion and protected fragments corresponding to transcripts processed at the cryptic *nos* and *nos* sites are indicated. RNase T1 only was used in this experiment as RNase A gave rise to additional bands due to 'breathing' of the RNA hybrids in (A + U)-rich regions. Molecular weight markers correspond to ^{32}P-labelled DNA fragments (pUC19/*Msp*I). (**B**) Processing efficiencies at the CaMV, cryptic *nos* and *nos* sites, expressed as a percentage of the total for each construct.

of the TTTGTA motif (Figure 4). This result demonstrates again that the upstream elements from one plant polyadenylation signal can function as upstream elements in a heterologous signal, and that several different sequence elements can act independently or in tandem on a given polyadenylation site.

Discussion

Despite frequently being referred to as 'the' eukaryotic polyadenylation signal, the AAUAAA motif is clearly not a ubiquitous component of polyadenylation signals in yeast and plant genes. In plants, as in yeast, it remains possible that there are several classes of polyadenylation signal and that AAUAAA may be an important component in some, but not all, of these. This study presents the first systematic mutational analysis of a functional AAUAAA motif in a plant polyadenylation signal. All single base mutants of AAUAAA are well tolerated by the processing machinery, giving rise to processing efficiencies in excess of 60% of wild-type. This is in complete contrast to the effects of the same mutations in an animal polyadenylation signal (Figure 9). Although none of the CaMV AAUAAA mutations had a drastic effect, there were slight differences depending on the position of

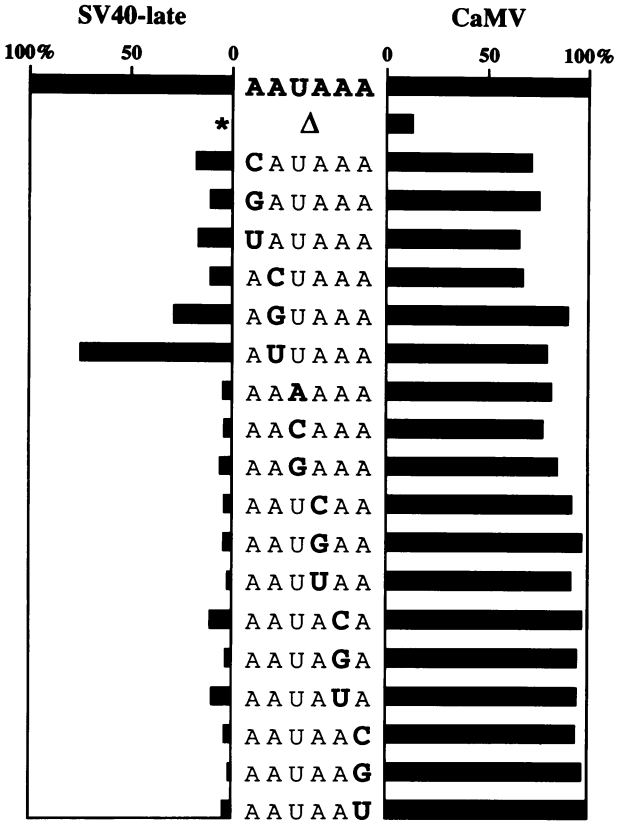


Fig. 9. A comparison of the effects of mutations in the AAUAAA motif in animal and plant polyadenylation signals. This figure summarizes the results of the effect on processing efficiency of mutations in the AAUAAA motif in the SV40 late and the CaMV polyadenylation sites. Data for SV40 refer to the efficiency of polyadenylation *in vitro* and are taken from Sheets *et al.* (1990). *: although not included in the study of Sheets *et al.*, deletion of the AAUAAA from this polyadenylation signal renders it non-functional, (Fitzgerald and Shenk, 1981). Processing efficiencies are expressed relative to the wild-type AAUAAA (100%).

the mutation. Examination of the analysis of putative plant polyadenylation signals compiled by Joshi (1987) reveals that, in the AAUAAA consensus sequences considered, there were more natural variants in the 3′-half of the motif than in the 5′-half. Our data are consistent with that distribution, in that mutations in the 5′-half were slightly more detrimental.

Considering the extreme tolerance to point mutations in the AAUAAA sequence, there are at least two possible explanations for the drastic effect on processing efficiency upon its deletion. On the one hand, it could be argued that the presence of the AAUAAA at this position is simply fortuitous and that the effect of the deletion can be explained by disruption of the local secondary structure, or changes in the distance between other important upstream elements and the cleavage site itself. That this is unlikely to be the only explanation is indicated by other constructs (e.g. those shown in Figures 3 and 5) in which deletions and insertions might also be predicted to alter the secondary structure in the vicinity of the AAUAAA. The effect on processing efficiency of these mutations is very much less than when the AAUAAA is deleted. A second explanation is that AAUAAA is specifically recognized by a *trans*-acting factor involved in the processing reaction. If this is the case, it would be predicted that replacement mutants would not be active. From the results presented in Figure 2B, it can be seen that in some cases, this prediction holds; replacing AAUAAA with GGGCCC, GUCGAC or UCCGGG drastically reduced processing at the CaMV site. In other constructs, the partial use of UUAUUU and GUAUUC as acceptable substitutes for AAUAAA suggests that the latter may not be part of a specific signal, and that (A+U)-rich motifs can generally replace AAUAAA. However, AAUAAA is clearly the optimal sequence at this position and the relatively high processing efficiencies seen with the (A+U)-rich replacement mutants may simply be indicative of a high degree of flexibility in the plant polyadenylation machinery which allows recognition of AAUAAA, or an acceptable alternative, providing the motif is present at an appropriate location relative to other *cis*-acting signals. The essential role of upstream accessory sequences is confirmed by the inability of a 5′-truncated polyadenylation signal which still contains AAUAAA to support 3′-end processing. Given the degree of flexibility allowed in the AAUAAA sequence, what is the basis of specific recognition of the CaMV polyadenylation signal? If the AAUAAA itself is recognized by a *trans*-acting factor, the specificity of the interaction must be very much lower than the equivalent interaction in the animal system. In vertebrates, the reason for the strict conservation of the AAUAAA motif has become clear from *in vitro* studies. A key initial step in the 3′-end processing reaction is the recognition of the pre-mRNA substrate by a multi-component factor called CPSF [cleavage and polyadenylation specificity factor, previously referred to as CPF (Christofori and Keller, 1988), PF2 (Gilmartin and Nevins, 1989), or SF (Takagaki *et al.*, 1989)]. CPSF specifically binds to RNAs containing AAUAAA (Bardwell *et al.*, 1991; Bienroth *et al.*, 1991; Keller *et al.*, 1991). RNA modification interference analysis showed that AAUAAA is sufficient for this binding and that all six nucleotide positions are involved in the interaction (Keller *et al.*, 1991). Clearly, the situation is very different in plants. In the case of the CaMV polyadenylation signal, it seems that assembly of a specific and efficient processing complex is dependent on the presence of both the AAUAAA motif and upstream sequences. Since plant polyadenylation signals do not seem to require specific sequences downstream of the cleavage site, these upstream sequences may play the equivalent role of the vertebrate DSE, i.e. they might be the binding sites for factors which stabilize, and add specificity to, the putative processing factor−AAUAAA interaction. Alternatively, interaction between the different components of the *cis*-acting signals themselves might form specific secondary structure elements which can then act as the target for recognition by processing factors.

A repeated UUUGUA motif present upstream of the AAUAAA in the CaMV polyadenylation signal is important for efficient processing; deletion of just this sequence alone accounted for a drop in efficiency of >30%. In the cryptic *nos* induction assay, this motif was able to allow processing at the cryptic *nos* site, with induction being sequence-specific for UUUGUA (Figure 4). Even a single copy of UUUGUA allowed recognition of the cryptic *nos* site and this induction increased with the number of copies of the motif inserted. Additive effects of upstream sequences on the efficiency of polyadenylation site use have also been reported in the case of the SV40 late polyadenylation site (Schek *et al.*, 1992)

and in ground squirrel hepatitis virus (GSHV; Russnak, 1991), and upstream elements from heterologous plant polyadenylation signals can act together to augment processing efficiencies (Mogen *et al.*, 1992; Sanfaçon, 1994).

The repeated UUUGUA motif from the CaMV polyadenylation signal is strikingly similar to motifs found in the upstream elements of SV40 and GSHV. In the SV40 late polyadenylation signal, three elements with the consensus sequence AUUUGURA are important for maximal processing efficiency, with the relative positive contribution of each of the three motifs increasing with increasing proximity to the AAUAAA (Schek *et al.*, 1992). One of the GSHV upstream activating elements, PS1B, includes the sequence GCUGAAAUUAUUUGUAUAGGA (Russnak, 1991). PS1B can functionally substitute for positions −53 to −32 in the CaMV polyadenylation signal (H.Rothnie, unpublished results). A recent study on the figwort mosaic virus (FMV) polyadenylation signal also implied that several related motifs (CUUGUAA, UCUGUA and UUUGAA) were important for efficient processing (Sanfaçon, 1994). Similar sequences can be found upstream of the polyadenylation sites of other plant viruses related to CaMV and FMV [carnation etched ring virus (Hull *et al.*, 1986), rice tungro bacilliform virus (Hay *et al.*, 1991), and commelina yellow mottle virus (Medberry *et al.*, 1990)]. Since all these examples are viral in origin, it could be argued that the UUUGUA-like motif is a specific feature of the upstream elements of viral polyadenylation sites. However, similar motifs have been identified in the upstream regions of the octopine synthase (*ocs*) and the pea Rubisco small subunit (*rbcS*) polyadenylation sites (MacDonald *et al.*, 1991; Mogen *et al.*, 1992), leading to the suggestion that this element may have a more general role in plant polyadenylation signals. The overall consensus sequence UUGUA has emerged from comparison of all these upstream elements. This sequence may represent one class of upstream accessory element which exists in plant polyadenylation signals.

Having identified an upstream accessory element in the CaMV polyadenylation signal, several constructs were made to test the importance of the relative position of the components of the *cis*-acting signals. The upstream activating element is unable to direct processing at the CaMV site if placed downstream of the cleavage site (Figure 6). Moving the upstream sequences closer to, or further away from, the AAUAAA and cleavage site affected the efficiency of processing (Figure 5). This result is consistent with the idea that processing-complex formation is made sub-optimal by imposing spatial constraints on the interaction of factors with their binding sites, but does not distinguish whether recognition of these binding sites depends on linear sequence motifs or on a particular secondary structure. Although the wild-type spacing is optimal (Figure 5), some flexibility is allowed in the distance between the upstream elements and either the AAUAAA, the cleavage site, or both. In the cryptic *nos* induction assays, the upstream elements tested for inducing ability were always at least 30 nt upstream of the cryptic *nos* AAUAAA, and in the construct in which the CaMV AAUAAA is deleted, the upstream element is even further removed from the cryptic *nos* site and the latter is still activated.

Hunt and his colleagues (Mogen *et al.*, 1992) have proposed that plant polyadenylation signals are composed of several distinct types of element, with NUEs, FUEs and the cleavage site itself all contributing to correct 3′-end processing. We would suggest that the NUE of the CaMV polyadenylation signal consists of the AAUAAA together with the repeated UUUGUA upstream element. Full processing activity is only achieved in the presence of further upstream sequences. These correspond to the FUE of the CaMV polyadenylation signal. In an attempt to dissect the FUE to determine if specific sequence elements are responsible for this enhancing effect, a series of internal deletions was constructed in this region. The results shown in Figures 7 and 8 taken together illustrate that this entire region makes some positive contribution to processing efficiency. No specific sequences responsible for this effect could be discerned, and there are no occurrences of UUGUA in any of these upstream elements. These upstream regions also differ in their base composition, although the region −151 to −71 is relatively U-rich, and we are left to conclude that we still do not know what it is about these sequences which allows them to enhance processing efficiency.

The difficulty in defining precise elements which function as polyadenylation signals in plants is reminiscent of the situation in yeast, where several different classes of polyadenylation site seem to exist. Interestingly, the CaMV polyadenylation site is accurately processed in yeast, suggesting that some of the *cis*-acting signals may be shared between plants and yeast. However, dissection of the sequence revealed that the important sequence determinants of the polyadenylation signal in yeast are distinct from those in plants (Irniger *et al.*, 1992). Processing in yeast depends on a UAGUAUGUA (−52 to −44) motif which is dispensible in plants (see Figure 3) and saturation mutagenesis of this motif has suggested that UAYRUA is an important signal in yeast (Irniger and Braus, 1994). This motif also occurs upstream of the polyadenylation site of the yeast Ty element (Hou *et al.*, 1994). Whether the presence of this yeast processing signal in the CaMV polyadenylation site is simply fortuitous, or whether it reflects a common evolutionary origin of the pararetrovirus CaMV and the retrotransposon Ty, is unclear.

The analysis of the CaMV polyadenylation signal presented here has provided some important clues as to what determines the *cis*-acting requirements for mRNA 3′-end formation in plants. However, the immediate challenge in the field of 3′-end processing in plants is the development of a functional *in vitro* assay, which can specifically reproduce the 3′-end processing patterns observed *in vivo*. Until this is achieved, and *trans*-acting factors and their cognate recognition sites can be studied in isolation, what gives plant polyadenylation signals their accuracy and specificity will remain a difficult question to resolve.

Materials and methods

Plasmid constructions

The basic plasmid used in this study, R-CAT* (Figure 1A), was constructed as follows: the CaMV R region from plasmid R-CAT (Sanfaçon and Hohn, 1990) was replaced by two pairs of synthetic oligonucleotides, which reconstruct the wild-type sequence with the introduction of three unique restriction sites (*Bgl*II, *Bam*HI and *Bcl*I) to facilitate further mutation (see Figure 1B). Transfer of the 35S promoter, CAT gene, synthetic CaMV R and the *nos* terminator from this latter construct into pGEM1 (Promega) resulted in plasmid R-CAT*. R-CAT* behaves as wild-type in the standard

processing assay. The presence of the T7 promoter in the pGEM vector allowed the same plasmid to be used for plant protoplast transfections and for *in vitro* transcription to produce homologous antisense RNA probes for RNase protection analysis.

Point mutations within the AATAAA motif were introduced using synthetic oligonucleotides. Pairs of DNA oligonucleotides containing a random mixture of the three possible variant nucleotides at each position of AATAAA were cloned between the *Bcl*I and the *Hin*dIII sites of R-CAT*. Potential mutants were screened by DNA sequencing using a Sequenase 2.0 kit from United States Biochemicals. All 18 possible mutants were identified. Similarly, mutations in which AATAAA was replaced with 6 nt of unrelated sequence were constructed using pairs of oligonucleotides reconstructing the sequence between the *Bcl*I and *Hin*dIII sites of R-CAT*. In these pairs of oligonucleotides, AATAAA was replaced with either an *Apa*I site (GGGCCC), a *Sal*I site (GTCGAC), the sequence TTATTT, or a random mix of oligonucleotides in which every nucleotide of the AATAAA motif was replaced with one of the three others. Two clones derived from cloning of this latter pair were tested, with the sequences TCCGGG and GTATTC replacing AATAAA. In plasmid Δ, the CaMV R region of R-CAT* is replaced with that of plasmid ΔAATAAA, where the AATAAA motif is deleted (Sanfaçon and Hohn, 1990). This plasmid carries more polylinker sequences between the CaMV and *nos* sequences, thus giving rise to slightly longer protected fragments corresponding to transcripts processed at the cryptic *nos* and *nos* sites (see Figure 2). Homologous antisense RNA probes for this series of mutants were prepared by *in vitro* transcription from each of these plasmids (digested with *Pst*I) using T7 RNA polymerase.

A relatively strong protected fragment reproducibly appears between the cryptic *nos* and *nos* sites with construct 1G (Figure 2). While we cannot rule out that this fragment represents a specific processing event, we rather suspect that it is an artefact arising from a point mutation introduced during cloning and plasmid preparation.

The US/DS series of constructs was assembled first by replacing the CaMV R *Pst*I−*Hin*dIII fragment from R-CAT with a synthetic fragment containing a minimal CaMV polyadenylation site sequence containing the AATAAA and the processing site. The resulting *Pst*I−*Hin*dIII fragment has the sequence 5′-ctgcagAAAATACTTCTATC**AATAAA**ATTTCTAATTCC**TA**ATT-CCTAAACCAAATCCAGaagctt-3′, where the restriction sites are in lower case letters and the AATAAA motif and the processing site marked in bold. This construct was called US/DS(−). Insertion of a double-stranded oligonucleotide (oligo 53−32a; annealed oligos 5′-agcTTATGTATTTG-TATTTGT-3′ and 5′-agcTACAAATACAAATACATA-3′) comprising CaMV 'R' sequences −53 to −32 with *Hin*dIII-compatible ends into the *Hin*dIII site of US/DS(−) created plasmids US/DSH(→) and US/DSH(←), with the inserted oligo in the sense and antisense orientations, respectively (see Figure 6A). A *Sac*I site 14 bp downstream from the *Hin*dIII site (see Figure 1B) allowed the insertion of the same oligo (but with *Sac*I-compatible ends; oligo 53-32b) to be inserted slightly further downstream from the CaMV processing site to make constructs US/DSS(→) and US/DSS(←). This series of constructs was digested with *Sca*I for the preparation of RNA probes.

Plasmid Δ31-19 was made by introducing a double-stranded oligonucleotide, carrying a precise deletion of nucleotides −31 to −19, between the *Bcl*I and *Hin*dIII sites of R-CAT*. In plasmid pSX, the *Bcl*I−*Hin*dIII fragment of R-CAT* was replaced with an oligo with the wild-type sequence except that the sequence between −34 and −19 was mutated to create a *Sal*I site and an *Xho*I site (**GTCGACTACTTCTCGAG**). In pSX the spacing between the AATAAA and the upstream element is increased from the wild-type (13 nt) to 14 nt. Plasmid ΔSX was constructed by cutting plasmid pSX with *Sal*I and *Xho*I and religating the resulting large fragment to itself. This results in a 3 nt spacing between −32 and the AATAAA. Plasmid pSX2 was made by ligating together the larger fragment obtained after cutting pSX with *Sal*I and *Xba*I to the smaller fragment resulting from cleavage of pSX with *Xho*I and *Xba*I (a unique *Xba*I site lies between the 35S promoter and the CAT gene in these constructs). The spacing between −32 and the AATAAA in pSX2 is 25 nt. This series of constructs was digested with *Pst*I for probe preparation.

Plasmids Δ53-32 and Δ45-32 were constructed by inserting a synthetic oligonucleotide carrying deletion of nucleotides −53 to −32 and −45 to −32, respectively, between the *Bcl*I and *Hin*dIII sites in R-CAT*. Plasmid Δ56-46 has been described previously (Sanfaçon *et al.*, 1991).

Plasmids Δ151-113, Δ151-71 and Δ112-71 were constructed by digesting R-CAT* with *Bgl*II and *Bam*HI, *Bgl*II and *Bcl*I, or *Bam*HI and *Bcl*I, respectively (see Figure 1B) and religating the largest resulting fragment in each case. Δ151-113Δ, Δ151-71Δ and Δ112-71Δ were constructed by performing the same set of digests on plasmid Δ45-32. Probes were prepared from *Pst*I-digested plasmids.

Plasmid '0' was constructed by digesting R-CAT with *Pst*I and *Sac*I, polishing the ends with T4 DNA polymerase and religating the large fragment to itself. Plasmids $(TTTGTA)_1$, $(TTTGTA)_2$, $(TTTGTA)_3$, $(TTTGTA)_4$, 53-32, $(TTTCAA)_4$ and $(TGTGTA)_4$ were constructed by replacing the *Pst*I−*Sac*I R fragment of R-CAT by synthetic oligonucleotides carrying one, two, three or four copies of TTTGTA (or four copies of the mutated motifs TTTCAA or TGTGTA), or the sequence −53 to −32 (TTATGATGTATTTGTATTTGTA), flanked by a 5′-*Pst*I site and a 3′-*Sac*I site. Plasmids (184−152)+*nos*, (184−113)+*nos*, and (184−71)+*nos* were made by digesting R-CAT* with *Sac*I and either *Bgl*II, *Bam*HI or *Bcl*I, respectively, creating blunt ends and religating the large fragment to itself. Plasmids (151−113)+*nos* and (151−71)+*nos* were made by further deletion of sequences between the *Pst*I and *Bgl*II sites of (184−113)+*nos* and (151−71)+*nos*, respectively. Similarly, (112−71)+*nos* was created by further deletion between the *Pst*I and *Bam*HI sites of (184−71)+*nos*. Probes for this series of plasmids were prepared from *Sca*I-digested plasmids.

Protoplast transfection and RNA analysis

Nicotiana plumbaginifolia leaf protoplasts were transfected with 5 μg plasmid DNA in the presence of polyethylene glycol as described by Goodall *et al.* (1990). Total RNA was isolated from the protoplasts 6 h after transfection and subjected to RNase protection analysis according to published protocols (Goodall *et al.*, 1990). For each mutant tested, a specific, homologous antisense RNA probe was used. Radioactively-labelled probes were synthesized by *in vitro* transcription (in the presence of [α-^{32}P]UTP) from the T7 or SP6 promoter present on the pGEM vector. Plasmids were digested with either *Pst*I or *Sca*I prior to *in vitro* transcription. Protected fragments were resolved on 6% polyacrylamide denaturing gels and visualized by autoradiography. Fragments corresponding to transcripts processed at the CaMV, cryptic *nos* and *nos* sites were identified based on their size relative to radioactively-labelled single-stranded DNA fragments. Quantification of the amount of radioactivity in the protected fragments was obtained by cutting out the relevant bands from the dried gel, adding scintillation fluid and measuring the amount of radioactivity in a scintillation counter. The percentage of transcripts processed at each site for each mutant was calculated, taking into account the number of labelled nucleotides in each fragment. Processing efficiencies expressed as percentages represent the mean values from at least three separate transfections, rounded to the nearest whole per cent. Standard deviations did not exceed 8% of the mean.

Acknowledgements

We are grateful to Gerhard Braus, Witold Filipowicz and Hélène Sanfaçon for critical reading of the manuscript, and to colleagues in the laboratory (especially Yvan Chapdelaine and Johannes Fütterer) for helpful discussions. We would like to thank Gerhard Braus, Terry Platt and Hélène Sanfaçon for making information available prior to publication. We very much appreciate the technical assistance of Matthias Müller. Thanks also to Mike Rothnie for preparation of the figures.

References

Abe,A., Hiraoke,Y. and Fukasawa,T. (1990) *EMBO J.*, **9**, 3691−3697.
Bardwell,V.J., Wickens,M., Bienroth,S., Keller,W., Sproat,B.S. and Lamond,A.I. (1991) *Cell*, **65**, 125−133.
Bevan,M., Barnes,W.M. and Chilton,M.-D. (1982) *Nucleic Acids Res.*, **11**, 369−385.
Bienroth,S., Wahle,E., Suter-Crazzolara,C. and Keller,W. (1991) *J. Biol. Chem.*, **266**, 19768−19776.
Butler,J.S. and Platt,T. (1988) *Science*, **242**, 1270−1274.
Butler,J.S., Sadhale,P.P. and Platt,T. (1990) *Mol. Cell. Biol.*, **10**, 2599−2605.
Carswell,S. and Alwine,J.C. (1989) *Mol. Cell. Biol.*, **9**, 4248−4258.
Chen,J. and Moore,C. (1992) *Mol. Cell. Biol.*, **12**, 3470−3481.
Christofori,G. and Keller,W. (1988) *Cell*, **54**, 875−889.
Dean,C., Tamaki,S., Dunsmuir,P., Favreau,M., Katayama,C., Dooner,H. and Bedbrook,J. (1986) *Nucleic Acids Res.*, **14**, 2229−2240.
Depicker,A., Stachel,S., Dhaese,P., Zambryski,P. and Goodman,H.M. (1982) *J. Mol. Appl. Genet.*, **1**, 561−573.
DeZazzo,J.D. and Imperiale,M.J. (1989) *Mol. Cell. Biol.*, **9**, 4951−4961.
Fitzgerald,M. and Shenk,T. (1981) *Cell*, **24**, 251−260.
Gil,A. and Proudfoot,N.J. (1984) *Nature*, **312**, 473−474.
Gil,A. and Proudfoot,N.J. (1987) *Cell*, **49**, 399−406.
Gilmartin,G.M. and Nevins,J.R. (1989) *Genes Dev.*, **3**, 2180−2189.
Goodall,G.J., Wiebauer,K. and Filipowicz,W. (1990) *Methods Enzymol.*, **181**, 148−161.

Guerineau,F., Brooks,L. and Mullineaux,P. (1991) *Mol. Gen. Genet.*, **226**, 141–144.
Guilley,H., Dudley,R.K., Jonard,G., Balázs,E. and Richards,K.E. (1982) *Cell*, **30**, 763–773.
Hart,R.P., McDevitt,M.A. and Nevins,J.R. (1985) *Cell*, **43**, 677–683.
Hay,J.M., Jones,M.C., Blakebrough,M.L., Dasgupta,I., Davies,J.W. and Hull,R. (1991) *Nucleic Acids Res.*, **19**, 2615–2621.
Heath,C.V., Denome,R.M. and Cole,C.N. (1990) *J. Biol. Chem.*, **265**, 9098–9104.
Heidmann,S., Obermaier,B., Vogel,K. and Domdey,H. (1992) *Mol. Cell. Biol.*, **12**, 4215–4229.
Henikoff,S. and Cohen,E.H. (1984) *Mol. Cell. Biol.*, **4**, 1515–1520.
Hou,W., Russnak,R. and Platt,T. (1994) *EMBO J.*, **13**, 446–452.
Hull,R., Sadler,J. and Longstaff,M. (1986) *EMBO J.*, **5**, 3083–3090.
Hunt,A.G., Chu,N.M., Odell,J.T., Nagy,F. and Chua,N.-H. (1987) *Plant Mol. Biol.*, **8**, 23–35.
Hyman,L.E., Seiler,S.H., Whoriskey,J. and Moore,C.L. (1991) *Mol. Cell. Biol.*, **11**, 2004–2012.
Imperiale,M.J. and DeZazzo,J.D. (1991) *New Biol.*, **3**, 531–537.
Irniger,S. and Braus,G.H. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 257–261.
Irniger,S., Egli,C.M. and Braus,G.H. (1991) *Mol. Cell. Biol.*, **11**, 3060–3069.
Irniger,S., Sanfaçon,H., Egli,C.M. and Braus,G.H. (1992) *Mol. Cell. Biol.*, **12**, 2322–2330.
Joshi,C.P. (1987) *Nucleic Acids Res.*, **15**, 9627–9640.
Keller,W., Bienroth,S., Lang,K.M. and Christofori,G. (1991) *EMBO J.*, **10**, 4241–4249.
Levitt,N., Briggs,D., Gil,A. and Proudfoot,N.J. (1989) *Genes Dev.*, **3**, 1019–1025.
Lingner,J., Kellerman,J. and Keller,W. (1991b) *Nature*, **354**, 496–498.
Lingner,J., Radke,I., Wahle,E. and Keller,W. (1991a) *J. Biol. Chem.*, **266**, 8741–8746.
MacDonald,M.H., Mogen,B.D. and Hunt,A.G. (1991) *Nucleic Acids Res.*, **19**, 5575–5581.
Manley,J.L., Yu,H. and Ryner,L. (1985) *Mol. Cell. Biol.*, **5**, 373–379.
McDevitt,M.A., Hart,R.P., Wong,W.W. and Nevins,J.R. (1986) *EMBO J.*, **5**, 2907–2913.
Medberry,S.L., Lockhart,B.E.L. and Olszewski,N.E. (1990) *Nucleic Acids Res.*, **18**, 5505–5512.
Mogen,B.D., MacDonald,M.H., Graybosch,R. and Hunt,A.G. (1990) *Plant Cell*, **2**, 1261–1272.
Mogen,B.D., MacDonald,M.H., Leggewie,G. and Hunt,A.G. (1992) *Mol. Cell. Biol.*, **12**, 5406–5414.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211–214.
Rothnie,H.M., Chapdelaine,Y. and Hohn,T. (1994) *Adv. Virus Res.*, **44**, in press.
Russnak,R.H. (1991) *Nucleic Acids Res.*, **19**, 6449–6456.
Russo,P., Li,W.-Z., Hampsey,D.M., Zaret,K.S. and Sherman,F. (1991) *EMBO J.*, **10**, 563–571.
Sadhale,P.P. and Platt,T. (1992) *Mol. Cell. Biol.*, **12**, 4262–4270.
Sadhale,P.P., Sapolsky,R., Davis,R.W., Butler,J.S. and Platt,T. (1991) *Nucleic Acids Res.*, **19**, 3683–3688.
Sanfaçon,H. (1994) *Virology*, **198**, 39–49.
Sanfaçon,H. and Hohn,T. (1990) *Nature*, **346**, 81–84.
Sanfaçon,H., Brodmann,P. and Hohn,T. (1991) *Genes Dev.*, **5**, 141–149.
Schek,N., Cooke,C. and Alwine,J.C. (1992) *Mol. Cell. Biol.*, **12**, 5386–5393.
Sheets,M.D., Ogg,S.C. and Wickens,M.P. (1990) *Nucleic Acids Res.*, **18**, 5799–5805.
Takagaki,Y., Ryner,L.C. and Manley,J.L. (1989) *Genes Dev.*, **3**, 1711–1724.
Wahle,E. (1992) *BioEssays*, **14**, 113–117.
Wahle,E. and Keller,W. (1992) *Annu. Rev. Biochem.*, **61**, 419–440.
Weiss,E.A., Gilmartin,G.M. and Nevins,J.R. (1991) *EMBO J.*, **10**, 215–219.
Wickens,M. and Stephenson,P. (1984) *Science*, **226**, 1045–1051.
Wilusz,J., Pettine,S.M. and Shenk,T. (1989) *Nucleic Acids Res.*, **17**, 3899–3908.
Wu,L., Ueda,U. and Messing,J. (1993) *Plant J.*, **4**, 535–544.
Zaret,K. and Sherman,F. (1982) *Cell*, **28**, 563–573.
Zarkower,D., Stephenson,P., Sheets,M. and Wickens,M. (1986) *Mol. Cell. Biol.*, **6**, 2317–2323.

Received on November 24, 1993; revised on February 2, 1994