The EMBO Journal vol.14 no.15 pp.3809–3819, 1995

Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved

Alexandra Moreira, Matthew Wollerton, Joan Monks and Nick J.Proudfoot[1]

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

[1]Corresponding author

Poly(A) signals of mammalian pre-mRNA have been defined as an AAUAAA sequence 10–30 nt upstream of the cleavage/poly(A) site followed by a GU/U-rich element immediately downstream. However, a number of viral poly(A) signals have been shown to possess additional signals upstream of AAUAAA that increase poly(A) site efficiency. We describe the first non-viral example of such an upstream sequence element (USE) for the poly(A) site of the human C2 complement gene. As this gene is very closely spaced to the related Factor B gene [the C2 poly(A) site is only 421 bp from the transcription start site of Factor B] we have isolated this same intergenic sequence from four other mammals (mouse, cat, rabbit and cow). We show that the USE of the C2 poly(A) site is highly conserved between these five different mammals. Furthermore, extensive mutagenesis of the human USE indicates that most of the 53 nt sequence is required for full activity. The human C2 poly(A) site does not possess any obvious downstream GU/U-rich sequences, although sequences immediately 3′ to AAUAAA as well as 13 nt of sequence following the cleavage site are both required for full activity. Interestingly the other mammalian C2 poly(A) sites do possess significant downstream GU/U-rich sequences. Finally we show that all five mammalian C2 poly(A) signals are immediately followed by conserved signals for transcriptional termination, consistent with the close proximity of the downstream Factor B gene.

Keywords: poly(A) signals/termination/upstream sequence element

Introduction

The sequences that dictate the site of polyadenylation at the 3′-end of mammalian mRNA have been extensively studied. In particular, the near ubiquitous AAUAAA signal is present 15–30 nt upstream of the poly(A) tail of nearly all mRNA (Proudfoot and Brownlee, 1976; Wickens, 1990) and has been shown to play a central role in the selection of poly(A) sites (Proudfoot, 1991), as well as in the biochemistry of the 3′ cleavage and polyadenylation reactions (Manley, 1988; Wahle and Keller, 1992). However, it has been demonstrated in a number of genes that additional signals are required for efficient use of a poly(A) site. In many cases GU- or U-rich sequences have been identified that are positioned immediately 3′ to the site of cleavage (Gil and Proudfoot, 1984, 1987; McDevitt *et al.*, 1984, 1986; McLauchlan *et al.*, 1985). These downstream sequences greatly enhance the efficiency of a poly(A) site both *in vivo* and *in vitro* (Hart *et al.*, 1985) and may act to stabilize the association of 3′ processing factors with AAUAAA (Wilusz *et al.*, 1990; Weiss *et al.*, 1991; Takagaki *et al.*, 1990). It has also been shown *in vivo* that there is a correlation between the length of the GU/U sequences and poly(A) site efficiency (McDevitt *et al.*, 1986; Gil and Proudfoot, 1987). From these observations it seems plausible that genes with poly(A) signals lacking significant GU/U-rich sequence elements may have intentionally weak signals as a means of restricting their expression. For example, in the case of alternatively polyadenylated membrane and secreted immunoglobulin mRNAs which are formed by the alternative use of poly(A) sites in the μ constant gene, the upstream poly(A) site for the secreted form is inefficient, as it lacks a GU/U-rich sequence. This allows a balanced competition between this RNA processing signal and an adjacent splice donor site and so regulates the synthesis of these alternative mRNAs (Peterson, 1994).

More recently a new class of poly(A) site activatory sequence has been discovered in several viral genes: ground squirrel hepatitis virus (GSHV) (Russnak and Ganem, 1990; Russnak, 1991), cauliflower mosaic virus (CaMV) (Sanfaçon *et al.*, 1991; Rothnie *et al.*, 1994), adenovirus L1 (DeZazzo and Imperiale, 1989), SV40 late (Carswell and Alwine, 1989) and HIV-1 (Brown *et al.*, 1991; DeZazzo *et al.*, 1991; Valsamakis *et al.*, 1991; Cherrington and Ganem, 1992). These signals are positioned upstream of the AAUAAA sequence and are termed upstream sequence elements or USEs. In the case of GSHV and HIV, USEs may be involved in the selective use of the poly(A) site at the 3′- but not 5′-end of the duplicated viral transcript. Although there are no clear sequence homologies between these different USEs, they are in general U-rich. Furthermore, some experiments have indicated that the different viral USEs are interchangeable (Russnak and Ganem, 1990; Valsamakis *et al.*, 1991). The mechanism of activation of USEs has not been characterized. However, in the case of HIV-1 and SV40 late mRNAs, their USEs function *in vitro* to activate polyadenylation, suggesting that they are RNA processing rather than transcriptional signals (Schek *et al.*, 1992; Gilmartin *et al.*, 1992; Valsamakis *et al.*, 1992). Additionally, it was shown that CPSF (cleavage polyadenylation specificity factor) interacts with the HIV-1 USE (Gilmartin *et al.*, 1995) and with the U-rich sequence CPE (cytoplasmic polyadenylation element) of c-*mos* RNA in *Xenopus* oocytes (Bilger *et al.*, 1994). Furthermore, the SV40 USE was shown to directly interact with the U1 snRNP protein A, suggesting a direct link between splicing and polyadenylation (Lutz and Alwine, 1994).

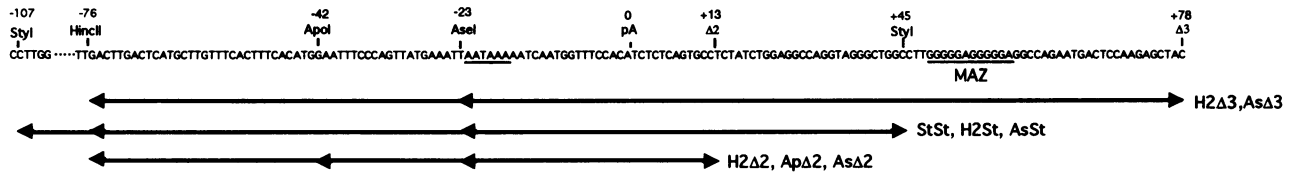

Fig. 1. Sequence of the C2 poly(A) site region. Numbering of nucleotides is with respect to the site of poly(A) addition. The C2 AATAAA is underlined, as is the MAZ site (required for C2 transcriptional temination; Ashfield *et al.*, 1991, 1994). The positions of various restriction sites and the end-points of the two *Bal*31 3′ deletion mutations Δ2 and Δ3 are indicated. The sequence between the 5′ *Sty*I and *Hin*cII sites is denoted by a dotted line. Arrows below the sequence indicate the positions of the various C2 poly(A) site fragments used in subsequent experiments. H2Δ3, DNA fragment from *Hin*cII site to Δ3 end-point; AsΔ3, fragment from *Ase*I site to Δ3 end-point; StSt, fragment between the two *Sty*I sites shown; H2St, *Hin*cII–*Sty*I fragment; AsSt, *Ase*I–*Sty*I fragment; H2Δ2, fragment from *Hin*cII site to Δ2 end-point; ApΔ2, fragment between *Apo*I site and Δ2 end-point; AsΔ2, fragment between *Ase*I site and Δ2 end-point.

We have studied the sequences associated with transcriptional termination and polyadenylation of the human C2 complement gene (Ashfield *et al.*, 1991). This gene is positioned in unusually close proximity to the next downstream gene, Factor B. The transcription start site of the Factor B gene is positioned only 421 bp 3′ to the poly(A) site of the upstream C2 gene (Wu *et al.*, 1987). Our studies on the C2–FB intergenic region revealed that termination of transcription occurs soon after the C2 poly(A) site and is at least in part dictated by the presence of a DNA binding site for a potential termination factor MAZ (Ashfield *et al.*, 1994). This same factor has been shown to regulate c-*myc* gene transcription (Asselin *et al.*, 1989) and the mRNA encoding MAZ has been cloned, revealing that the MAZ protein has zinc finger DNA binding motifs, as well as potential transcriptional activation domains (Bossone *et al.*, 1992).

Since the MAZ binding site is positioned in close proximity to the C2 poly(A) site (~50 nt downstream), we also defined the approximate 3′ boundary of the C2 poly(A) signal and showed that a maximum of 13 nt downstream of the cleavage site was required for efficient polyadenylation (Ashfield *et al.*, 1991). However, there are no obvious GU/U-rich elements in this 13 nt sequence, as would be predicted for an efficient poly(A) signal. The studies presented in this paper characterize in greater detail the sequences, in addition to AAUAAA, that constitute the C2 poly(A) signal. Although we find that sequences immediately downstream of the cleavage/polyadenylation site do have a small effect on the efficiency of C2 polyadenylation, two different sequence elements are more important for full activity. A minor sequence element is placed between the AAUAAA and cleavage site, while the second major element is positioned upstream of the AAUAAA sequence. This latter element is reminiscent of the various USE of viral poly(A) signals, as described above. The C2 poly(A) signal represents the first example of a non-viral poly(A) site to be shown to require such an upstream element. We underline the importance of this USE by showing that it is highly conserved between five different mammals. Furthermore, a potential MAZ binding site immediately follows the C2 poly(A) signal in the termination region of all five mammalian C2 genes.

Results

Definition of minimal sequences required for full activity of the C2 poly(A) signal

As a way to assess the efficiency of a poly(A) signal *in vivo*, it is usual to position it at the end of a gene, adjacent to a reference poly(A) site. The ratio of mRNA species expressed by this gene that use either the reference or test poly(A) signals provides an indirect measure of efficiency for the test poly(A) site. Using this poly(A) site competition strategy, we have previously demonstrated that the poly(A) site of the human α-globin gene is a relatively inefficient RNA processing signal, while the rabbit β-globin poly(A) signal or a synthetic poly(A) site (SPA) derivative is highly efficient (Levitt *et al.*, 1989). The studies described here characterize the human C2 poly(A) signal. Either the intact C2 poly(A) signal or deletions and specific mutations of it were placed in competition with either the weaker α-globin or stronger SPA signals.

We began these studies by establishing the minimal sequence elements required for efficient polyadenylation of the C2 gene. Figure 1 presents the sequence of the C2 poly(A) site region with the positions of the AATAAA and MAZ binding site indicated. Below this sequence arrows denote the extent of DNA fragments containing the C2 poly(A) sequence that were inserted downstream of the α-globin poly(A) site in the clones α/C2 or upstream of the SPA in the clones C2/S. We have previously shown that both the C2 poly(A) site fragments H2Δ2 and H2Δ3 largely out compete the human α-globin poly(A) signal (Ashfield *et al.*, 1991). These data argue that sequences 3′ to the H2Δ2 fragment, beyond +13, are not part of the C2 poly(A) signal. In particular, the immediate downstream termination signals of the C2 gene, including the MAZ protein binding site, do not affect C2 poly(A) site efficiency (Ashfield *et al.*, 1991, 1994). Figure 2 repeats this analysis of clones α/H2Δ2 and α/H2Δ3, as well as testing additional deletions upstream of the C2 AATAAA sequence. Each C2 poly(A) site fragment was inserted 90 bp downstream of the α-globin gene with its relatively inefficient poly(A) site, as shown below the RNA analysis data (Figure 2D). Following transient transfection of these various constructs into HeLa cells, cytoplasmic RNA was isolated and analysed by an S1 nuclease assay. The probe used to detect the relative use of the α or C2 poly(A) sites is the parent α-globin plasmid digested with *Bst*EII and 3′-end-labelled (see Figure 2D). The ratios of C2 to α bands were quantitated by either densitometry or phosphoimager analysis. The C2 3′-end is detected as a mismatch band at the site of insertion of the C2 fragments into the α-globin 3′ sequence. A second plasmid pBR328RβSV was co-transfected with each test plasmid (Grosveld *et al.*, 1982). This plasmid provides dual functions. First, it expresses SV40 T antigen, which allows

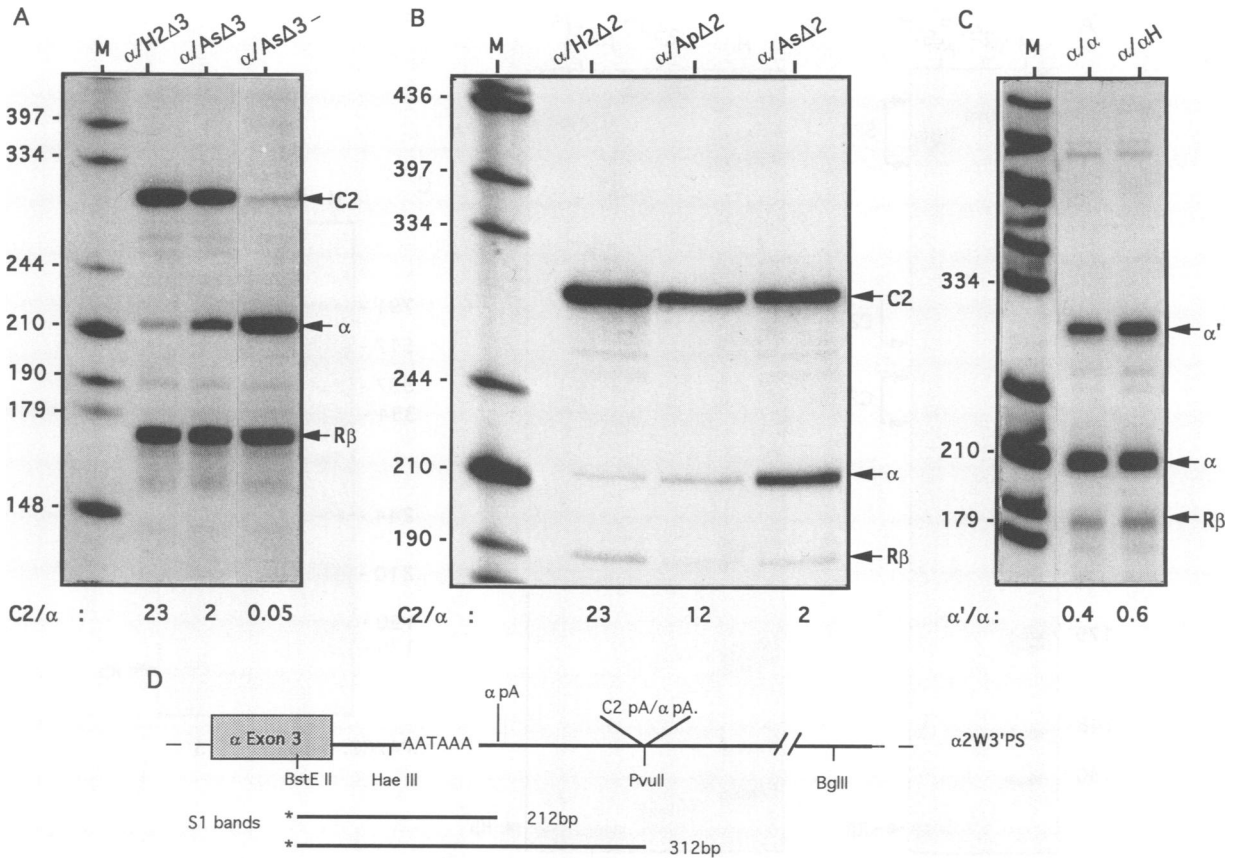

Fig. 2. (**A–C**) S1 nuclease analysis of the RNAs isolated after transient transfection of HeLa cells with the constructs α/HΔ3, α/AsΔ3 and α/AsΔ3– (C2 AsΔ3 fragment inserted in the reverse orientation) (A), α/HΔ2, α/ApΔ2 and α/AsΔ2 (B) and α/α and α/αH (C). The positions of the α and C2 poly(A) S1 bands are indicated. For the downstream C2 and α-globin poly(A) sites (α') a mismatch band of 312 nt is indicated, which corresponds to the downstream *Pvu*II site at which the second poly(A) signal is added. The intensity of this signal measures the level of mRNA utilizing the second C2 or α-globin poly(A) site. The S1 signal for the rabbit β-globin mRNA poly(A) site is also indicated, providing a control for transfection efficiency. (**D**) Diagram showing the 3′-end of the α-globin gene and the position in the 3′ flanking region where the test poly(A) signals are inserted. The α-globin gene S1 probe was generated by digestion at the *Bst*EII site in exon 3 of the α2-globin gene and the expected S1 bands for use of either the α or downstream-positioned poly(A) signals are shown. The C2:α poly(A) ratios, measured by densitometric scanning of the autoradiographs, are indicated beneath the lanes (A–C).

the SV40 origin-containing plasmids used in these experiments to replicate. As described previously, the promoter of the human α-globin gene is activated by DNA replication (Proudfoot *et al.*, 1992). Second, it expresses rabbit β-globin mRNA (Rβ), which provides a co-transfection control for these experiments. A specific S1 probe is used to detect Rβ in each experiment. The Rβ signals for different figures varies, as the S1 probe used was of various specific activities.

From the results presented in Figure 2A we can conclude that sequences upstream of the AATAAA sequence have a significant effect on C2 poly(A) site efficiency. Mutations that delete sequence between the –76 *Hinc*II site and the AATAAA sequence significantly decrease the relative use of C2 to α poly(A) sites (α/AsΔ3). As indicated, clone α/H2Δ3 gives a C2:α ratio of 23, while the α/AsΔ3 clone, which has all sequence upstream of AATAAA removed, gives a ratio of 2. As a control, the C2 poly(A) site in α/AsΔ3 was tested in reverse orientation (α/AsΔ3–), confirming that the α poly(A) site works efficiently in the absence of competition from C2. 5′ deletions were also tested using the α/H2Δ2 clone (Figure 2B). Progressive deletion to first the *Apo*I site (α/ApΔ2) and then the *Ase*I site (α/AsΔ2) upstream of AATAAA caused progressive inactivation of the C2 poly(A) site. These data indicate that C2, in common with several viral poly(A) signals, possesses a T-rich upstream activatory element. As shown before, H2Δ2 and H2Δ3 are both fully efficient C2 poly(A) signals, arguing that sequences downstream of +13 are not involved in C2 poly(A) site efficiency. Apparently there is no requirement for any significant GU-rich sequence element downstream of the C2 poly(A) site, as is the case in other characterized mammalian poly(A) signals (Proudfoot,1991).

In the above experiments we argued that the C2 poly(A) signal possesses upstream activatory signals. However, it is formally possible that these activation sequences affect the upstream α-globin poly(A) signal and so increase the ratio of α to C2 poly(A) site usage. Previous studies (McDevitt *et al.*, 1986; Gil and Proudfoot, 1987) demonstrated that poly(A) site efficiency is unaffected by sequences placed >50 bp downstream of the poly(A) site, as in these constructs. To test this possibility we generated two additional constructs, α/α and α/αH with a second copy of the α-globin poly(A) signal in the same downstream position. In the α/αH clone the second α poly(A) site has the C2 *Hinc*II–*Ase*I upstream element positioned downstream in exactly the same position as in the α/H2Δ3

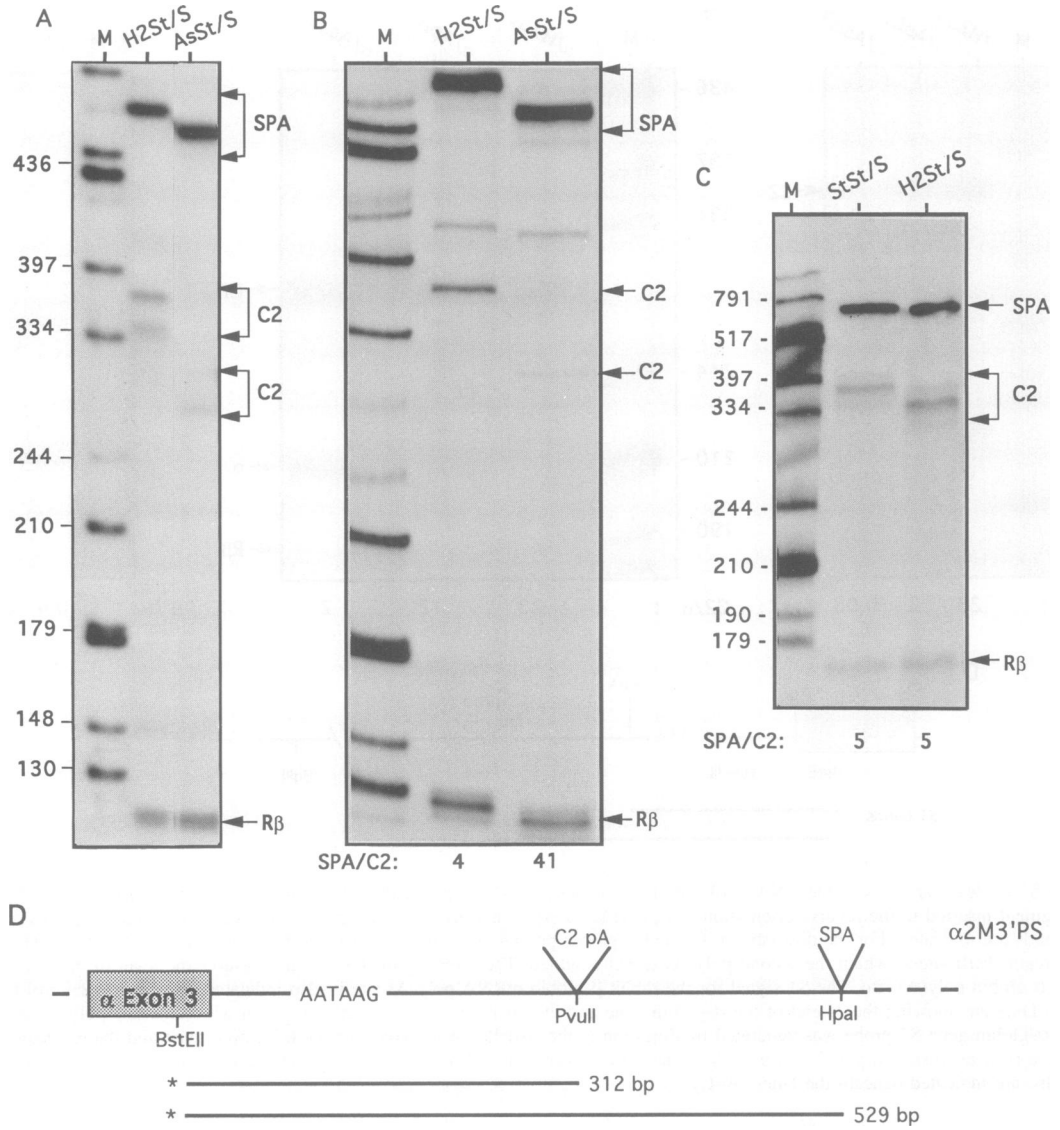

Fig. 3. (**A**) S1 analysis of RNA isolated from HeLa cells transfected with clones H2St/S and AsSt/S. Positions of the C2 and SPA poly(A) site bands are indicated, as well as the rabbit β-globin co-transfection control signal. The C2 bands break up into doublets, due to breathing at the 3′-end RNA:DNA hybrid. (**B**) Exonuclease VII analysis of RNA isolated from HeLa cells transfected with clones H2St/S and AsSt/S. Positions of the C2 and SPA poly(A) site bands are indicated, as well as the rabbit β-globin co-transfection control signal. Quantitation of the ratio of SPA to C2 poly(A) site usage is shown below the lanes. (**C**) S1 analysis of RNA isolated from HeLa cells transfected with clones StSt/S and H2St/S as above. (**D**) Diagram showing arrangement of C2 and SPA poly(A) signals positioned downstream of the α-globin gene with a mutated poly(A) signal. The positions of the S1 and ExoVII probes are also indicated. Homologous probes were used with each different clone.

clone. Figure 2C shows the ratio of mRNAs using the first and second α poly(A) sites. As predicted, the upstream poly(A) signal is used preferentially. However, the ratio of upstream to downstream α poly(A) site usage does not change significantly when the C2 sequence is positioned downstream. These results confirm that the C2 upstream element is not acting as a downstream element to the α poly(A) signal in the α/Δ2 and α/Δ3 clones described before.

We wished to confirm the presence of upstream activatory sequences for the C2 poly(A) site using a different poly(A) site competition construct in which the competing poly(A) site is more efficient than C2. As shown in Figure 3D, C2 poly(A) site fragments (Figure 1) were inserted downstream of a mutated α-globin gene with an inactivating mutation in its poly(A) signal (Whitelaw and Proudfoot, 1986). However, the efficient SPA is positioned further downstream in the 3′ flanking region of the α-globin gene, downstream of the inserted C2 poly(A) sites. Figure 3A demonstrates that the C2 H2St poly(A) signal is less efficient than SPA in clone H2St/S, working at a level of ~25% (this clone has the *Hin*cII–*Sty*I DNA fragment indicated in Figure 1). As predicted, the AsSt insert is less efficient than H2St, dropping to the ~2% level (clone AsSt/S). Since the C2 poly(A) signal region is unusually AU-rich, the S1 products obtained are heterogenous, preventing accurate quantitation. We therefore repeated the RNA analysis using exonuclease VII. This enzyme is a single strand-specific exonuclease and will not cleave at internal RNA:DNA duplexes, as occurs with S1. Figure 3B shows these results, quantitated by phosphoimager analysis. As indicated, the H2St C2 poly(A) site forms a single 3′-end terminus with an SPA:C2 poly(A) site ratio of 4, while the AsSt/S clone

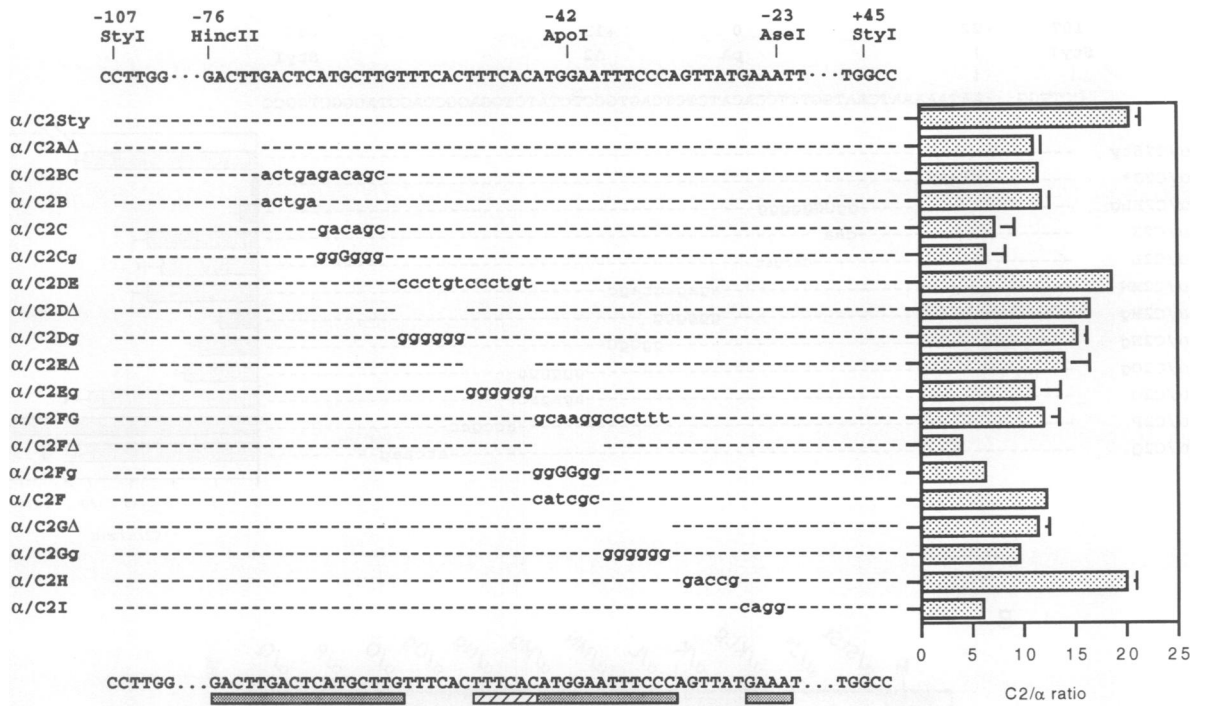

Fig. 4. Sequence of PCR-generated mutations to the C2 *Sty*I–*Sty*I poly(A) signal fragment upstream of AATAAA. Mutant sequence is indicated by lower case letters, while unaltered sequence is denoted by dashes and upper case letters. Deletions are shown by gaps. Each C2 *Sty*I mutation was cloned downstream of the α-globin gene (α/C2) and transfected into HeLa cells. The ratio of the C2 poly(A) site usage to the upstream-positioned α-globin poly(A) site is shown graphically (averaged from up to four different transfections) next to each C2 mutant sequence. Quantitation was done in each case by phosphoimager analysis. Hatched and filled in boxes below the C2 sequence denote minor and major parts of the upstream sequence element.

increases to a ratio of 41 (SPA:C2). These results endorse the observation that the C2 poly(A) site possesses upstream activatory sequence elements. We finally wished to determine if any additional more distant upstream elements exist for the C2 poly(A) site. We therefore tested the *Sty*I–*Sty*I C2 poly(A) site fragment (StSt/S clone), which has a more extensive upstream region to the C2 poly(A) signal in competition with SPA. As shown in Figure 3C, this C2 poly(A) site fragment works with equal efficiency to the H2St C2 poly(A) signal when placed in competition with SPA. We therefore conclude that sequences from the *Hinc*II site at –76 to the 3′-end-point, Δ2 (+13), constitute the entire C2 poly(A) signal. This region of 90 bp was therefore subjected to specific mutagenesis, as described in the next two sections.

Specific mutagenesis of the upstream sequence element of the C2 poly(A) signal

We wished to define the upstream element of the C2 poly(A) site in the context of the complete *Sty*I–*Sty*I fragment (StSt), which contains the whole C2 poly(A) site region (Figure 1). To make these specific mutations, we carried out a PCR-based mutagenesis procedure as described in Materials and methods and generated specific deletions or sequence replacements upstream of the C2 AATAAA sequence. As shown in Figure 4, extensive sequence replacements and deletions were generated from the –76 *Hinc*II to the –23 *Ase*I site (immediately 5′ to the AATAAA sequence). Each mutated C2 poly(A) site was isolated as an intact *Sty*I–*Sty*I fragment and inserted downstream of the poly(A) site of the α-globin gene at the *Pvu*II site, as shown in Figure 2D. These different α/C2 constructs were then transfected into HeLa cells and the ratio of poly(A) site use between the α and mutated C2 poly(A) sites was determined as before. Experiments were repeated up to four times and quantitated by phosphoimager analysis. The ratio of C2 to α poly(A) site usage was averaged between the different experiments and is illustrated adjacent to each C2 mutant sequence (Figure 4).

The results obtained indicate that much of the upstream sequence of the C2 poly(A) site has a significant effect on poly(A) site efficiency. Some mutations decrease the efficiency of the C2 poly(A) signal by >3-fold, such as C2C, C2Cg, C2FΔ, C2Fg and C2I. Only two mutations have no effect, C2DE and C2H. Also, it is apparent that either deletion mutations or substitutions to the oligo(G) sequence (denoted by Δ and g respectively) are in general more inhibitory than other sequence replacements. From detailed analysis of these mutations it is evident that two regions of the C2 USE are especially critical and are indicated by filled boxes below the C2 USE sequence at the bottom of the Figure. Also, mutations C2EΔ and C2Eg both have a significant inhibitory effect, even though the larger mutation C2DE has no effect over the same E region. This suggests that the E region may be important for the full effect of the C2 USE (denoted by the hatched box). Finally, mutation C2I reduces the C2 poly(A) site to 25% efficiency, probably due to a direct effect on recognition of the AAUAAA sequence, to which it is directly adjacent.

Sequences downstream of the C2 AAUAAA also affect poly(A) site efficiency

Figure 5 shows similar mutagenesis of the whole C2 poly(A) site *Sty*I–*Sty*I fragment downstream of AATAAA.

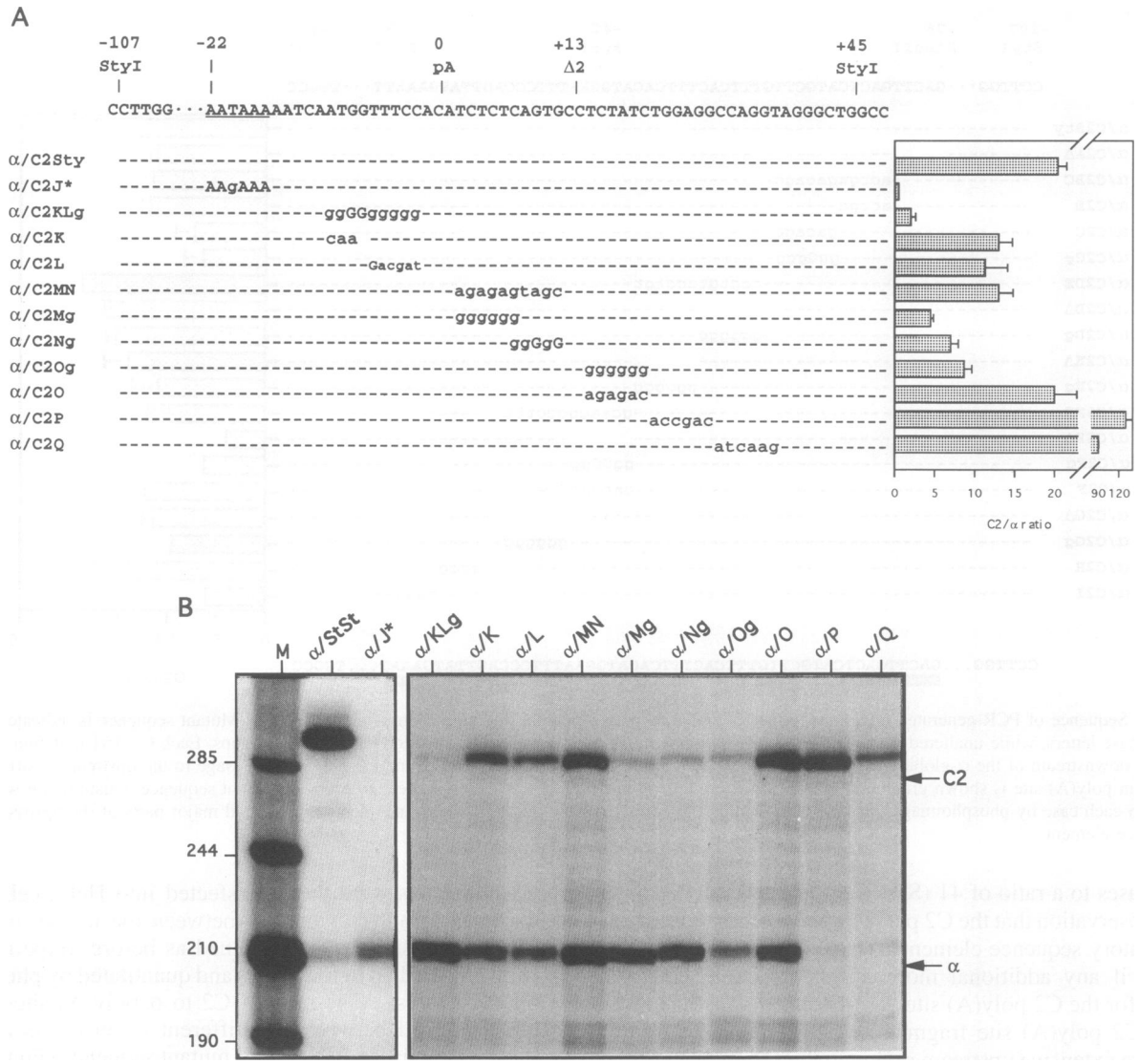

Fig. 5. (**A**) Sequence of PCR-generated mutations to the C2 *Sty*I–*Sty*I poly(A) signal fragment downstream of AATAAA. The data is presented as for Figure 4. Note that the AATAAA mutation α/C2J* causes complete inactivation of the C2 poly(A) signal. (**B**) Representative S1 analysis of the α/C2 clones described in (A). The positions of the C2 and α poly(A) site S1 bands are indicated.

As before, Figure 5A shows the sequences of each of the C2 poly(A) signal mutations, with a graphical representation of their effect on poly(A) site efficiency alongside. Figure 5B then shows representative S1 analysis data for each of these mutations. As indicated, sequences between the AAUAAA and polyadenylation site significantly influence poly(A) site efficiency. The mutation C2KLg causes a 10-fold reduction in C2 poly(A) site efficiency. Even though G replacement mutations, such as in C2KLg, generally have a more drastic effect than other sequence replacements, it is clear that the sequence between the AAUAAA and cleavage site is critical to poly(A) site efficiency. The smaller, non-oligo(G) replacement C2K and C2L mutations each have a 2-fold effect, which may indicate some degree of redundancy in these sequence elements. As a control for these mutations a single base mutation, AATAAA→AAGAAA was also tested (C2J*), which, as predicted, completely inactivates the C2 poly(A) site. This result illustrates that none of the other mutations around the C2 poly(A) site are as critical as mutations to the canonical AATAAA sequence.

Sequences downstream of the C2 poly(A) site appear to have only a limited role in influencing its efficiency, consistent with the absence of any GU/U-rich elements in this region. The replacement of most of the sequence (10 nt) between the poly(A) site and the Δ2 deletion endpoint (mutant C2MN) causes an ~2-fold reduction in efficiency. As before, G replacements in this sequence have more drastic effects, confirming that oligo(G) sequences close to the poly(A) signals significantly reduce efficiency. Interestingly, as these G mutations are moved further away from the C2 poly(A) site their inhibitory effect diminishes (compare C2Mg, C2Ng and C2Og). The final non-G mutations C2O, C2P and C2Q have no inhibitory affect on the C2 poly(A) site. Indeed, both C2P and C2Q have significant activatory effects, possibly because G-rich sequences normally present at these positions are mutated.

In summary, the results presented in Figure 5 suggest that the sequences both immediately upstream and downstream of the cleavage/poly(A) site have a significant role in determining the efficiency of the C2 poly(A) signal. However,

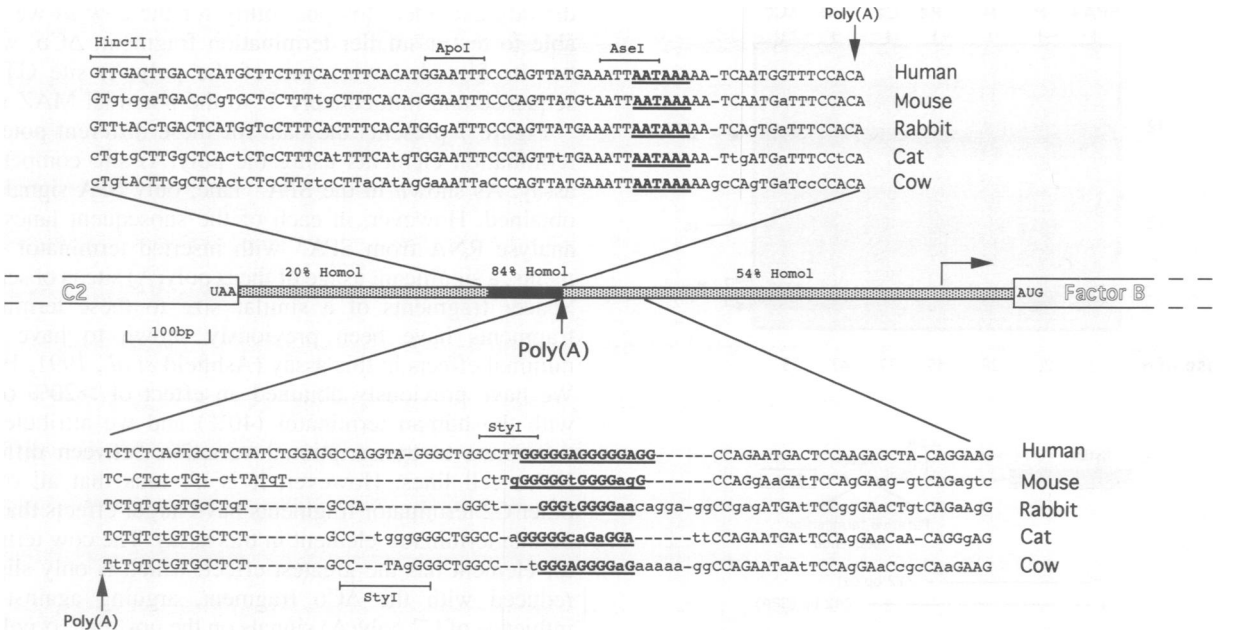

Fig. 6. Comparison of C2–FB sequences between five different mammals. The sequences upstream and downstream of the cleavage/poly(A) sites are compared. Sequences that diverge from human are in lower case letters and the AATAAA and G-rich MAZ binding elements are in bold and underlined. The GT-rich sequences following the poly(A) site are underlined. Restriction sites used in subcloning are also indicated. Between these two blocks of sequence alignments the C2–FB region is diagrammed. Open boxes denote C2 or Factor B coding regions. Stippled boxes denote either the C2 3′ untranslated region or the C2–FB intergenic sequence and the Factor B 5′ untranslated region. The per cent homology between the five sequences is indicated. These figures were calculated by scoring more than one sequence change between the five mammals as a difference (Pileup and Prettyplot, plurality 4; computer programs of the GCG package). The black box denotes the region of the C2 poly(A) site that is most homologous. The transcription start site of the Factor B gene is indicated, as is the C2 poly(A) site.

replacement of these sequences with G nucleotides has a more severe inhibitory effect, presumably due to direct blocking of the poly(A) factors CPSF upstream or CstF downstream of the poly(A) site (Wahle and Keller, 1992).

Phylogenetic analysis of the C2 poly(A) and termination signals

The short intergenic region between the C2 and Factor B genes in humans has also been characterized in mice (Nonaka *et al.*, 1989). However, since we wished to obtain more extensive phylogenetic data on this region, we isolated the same C2–FB intergenic region from three further mammals, rabbit, cat and cow, as well as from mouse. PCR primers were designed that matched conserved regions of the C2 C-terminal and Factor B N-terminal gene regions (based on a sequence comparison of the human and mouse DNA sequences; Nonaka *et al.*, 1989). Successful amplification of these different C2–FB intergenic sequences was achieved and the PCR products obtained were subcloned into a plasmid vector (See Materials and methods for details of these PCR DNA isolations).

Each intergenic region was sequenced and Figure 6 shows a comparison of all five mammalian C2 poly(A) site regions. As indicated, the C2–FB intergenic region can be divided into three sections, based on sequence conservation. The C2 3′ untranslated sequence between the C2 termination codon and C2 poly(A) site region is highly divergent between the five species (~20% sequence identity). The sequence downstream of the C2 poly(A) site up to the Factor B initiation codon is more conserved in sequence (~50%). Since this region includes enhancer and promoter sequences of the Factor B gene (Wu *et al.*, 1987; Ashfield *et al.*, 1991), such a level of sequence homology is to be expected. However, we were intrigued to observe that the sequence immediately upstream of the C2 AAUAAA sequence (corresponding to the USE element) and through to the downstream cleavage/poly(A) site is 84% conserved between the five mammals. Such a high degree of sequence conservation greatly strengthens the view that these sequences have an important role, as suggested by our mutagenesis of the human C2 poly(A) site region. It is particularly striking that the high level of homology between these five C2 poly(A) site regions diverges to the lower 20% level exactly at the 5′ boundary of the C2 USE, as defined in humans.

Two other features of this sequence analysis are important. First, in all species except humans the sequence immediately downstream of the poly(A) site, although divergent, is significantly GT-rich (these GT sequences are underlined). It would appear that in these four mammals the efficiency of the C2 poly(A) signal may be further enhanced by possessing the usual GT-rich activating elements. Immediately downstream of these GT-rich elements there is significant sequence divergence. However, in each case a G-rich element is present (underlined and in bold), which, based on our previous studies, is likely to be a binding site for MAZ (Ashfield *et al.*, 1994). It is also interesting to note that in these non-human C2 sequences the MAZ binding site is brought closer to the poly(A) site. In the case of mouse and rabbit the MAZ site is positioned immediately downstream of the GT-rich element. As described in Figure 5, we have shown that G-rich sequences at this position in the human C2 gene

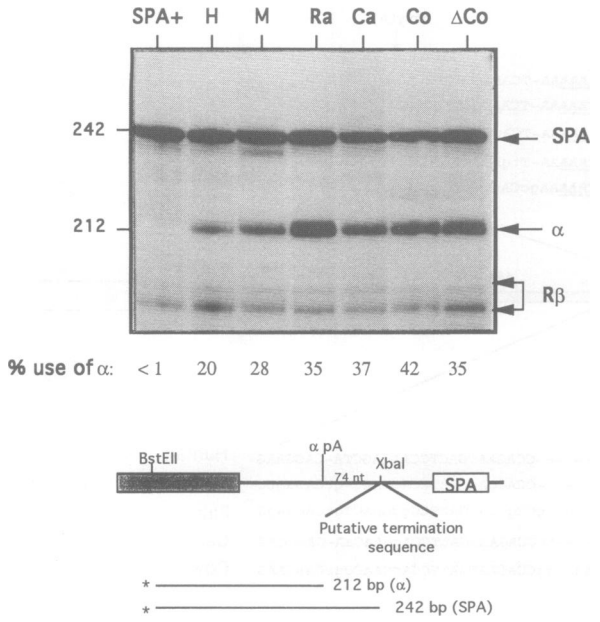

Fig. 7. Termination data obtained using the poly(A) site competition assay. A diagram of the SPA$^+$ vector is shown below, indicating the site at which putative terminator fragments are added, as well as the sizes of S1 signals obtained using the *BstE*II site to make a 3′-end-labelled probe (Ashfield *et al.*, 1991). This probe mismatches with the RNA transcripts 5′ of the *Xba*I site, so the C2 S1 band is 242 bp. The distance between the upstream α poly(A) site and *Xba*I site at which the terminator fragments are inserted is 74 nt. Finally the per cent use the of α poly(A) site is quantitated below each lane and is an indirect measure of termination efficiency.

have an inhibitory effect on C2 poly(A) site efficiency (see mutants C2O, C2Og, C2P and C2Q). Since MAZ sites are G-rich sequences, it is possible that the juxtapositioning of C2 poly(A) and termination signals in the non-human sequences may account for why in these cases a more GT-rich downstream poly(A) signal is also present.

Evidence for termination elements immediately downstream of the C2 poly(A) signal in five different mammals

We wished to demonstrate that the other mammalian C2 downstream sequences are associated with Pol II termination. We have previously employed a poly(A) site competition assay to identify Pol II terminator elements (Ashfield *et al.*, 1991, 1994). We therefore tested the potential termination sequences downstream of the mouse, rabbit, cat and cow poly(A) signals using this assay. As indicated in Figure 7, potential termination elements are inserted between the weak α-globin and strong SPA poly(A) signals in the poly(A) site competition vector SPA$^+$. In the absence of termination or pausing between these two poly(A) sites, SPA is used exclusively. Fragments were isolated from the four non-human C2–FB subclones using either PCR primers to define specific ends or, where available, restriction sites (see Materials and methods). In each case the 5′-end of the terminator fragment (150–300 bp in length) is positioned immediately downstream of the AAUAAA sequence. Even though these fragments include 3′ parts of the C2 poly(A) signal, we reasoned that they would be placed too far from the α-globin poly(A) site in SPA$^+$ (74 nt away) to directly affect its efficiency as an RNA processing signal. However, we directly excluded this possibility for the cow as we were able to test a smaller termination fragment, ΔCo, which has its 5′-end downstream of the poly(A) site GT-rich sequence and immediately 5′ to the potential MAZ site.

Figure 7 presents the data for these different potential termination elements using the poly(A) site competition assay. As shown in the SPA$^+$ lane, only SPA signal was obtained. However, in each of the subsequent lanes that analyse RNA from SPA$^+$ with inserted terminator fragments, a significant usage of the α poly(A) site is observed. Spacer fragments of a similar size to these terminator fragments have been previously shown to have only minimal effects in this assay (Ashfield *et al.*, 1991, 1994). We have previously obtained an effect of >20% α use with the human terminator (40%) and we attribute this difference to transcriptional variation between different HeLa cell lines. However, it is obvious that all of the other C2 terminator fragments have larger effects than the human terminator element in this assay. The cow terminator element has the greatest effect, which is only slightly reduced with the ΔCo fragment, arguing against any influence of C2 poly(A) signals on the upstream α poly(A) site. We conclude that all five mammalian C2 sequences possess efficient Pol II termination elements immediately proceeding their poly(A) signals.

C2 upstream sequence elements and GT-rich downstream elements cooperate to form a very efficient poly(A) signal

As a final experiment in these studies we mutated the human C2 poly(A) signal sequence 3′ to the cleavage/poly(A) site into the more GT-rich sequence found in the rabbit C2 poly(A) signal. As before, we made a specific mutation in the whole human C2 *Sty*I–*Sty*I fragment using the same PCR procedure. This mutated *Sty*I–*Sty*I fragment, as well as a truncated *Ase*I–*Sty*I fragment that lacks the USE, are diagrammed in Figure 8 and were inserted upstream of SPA in the α-globin gene vector used in Figure 3. These two constructs, StRa/S and AsRa/S, were transfected into HeLa cells and the SPA:C2 usage ratio compared with StSt/S. As before, the intact human C2 poly(A) signal works 5-fold less well than SPA. However, the modified human C2 poly(A) signal with the rabbit GT-rich elements is much more efficient, largely out competing SPA. This result indicates that the GT-rich sequence present downstream of the rabbit C2 cleavage site is a strong activator of poly(A) signal efficiency. This is in agreement with previous work (Gil and Proudfoot, 1984, 1987; McDevitt *et al.*, 1986, Levitt *et al.*, 1989). Interestingly, the truncated construct lacking the USE but having the rabbit GT-rich element works with intermediate efficiency, giving nearly equal use of C2 and SPA poly(A) signals (SPA:C2 ratio 1.2).

These results indicate that the C2 USE and downstream GT-rich elements have similar activatory effects on poly(A) site efficiency. They therefore underline the importance of the C2 USE as a component of the C2 poly(A) signal.

Discussion

It has been apparent for some years that sequences in addition to AAUAAA are required to encode a fully

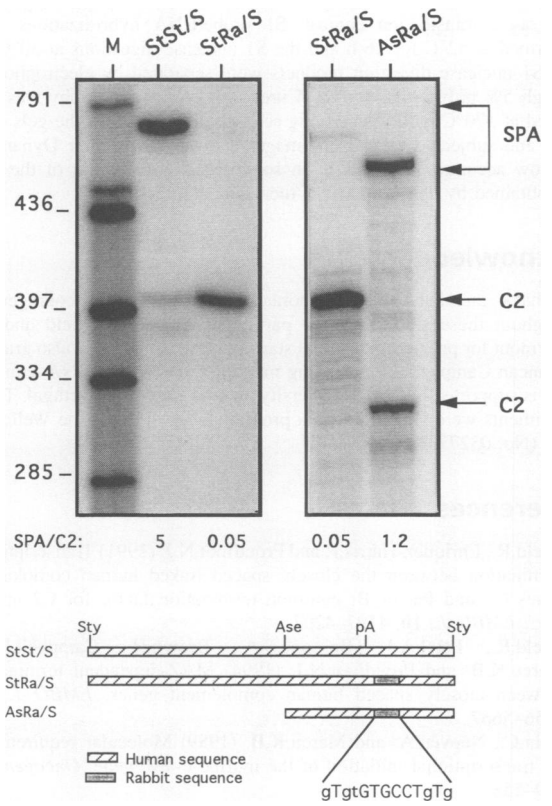

Fig. 8. S1 analysis of hybrid human–rabbit C2 poly(A) signals cloned into the vector described in Figure 3 upstream of SPA (C2/S). Homologous S1 probes were used for each different construct and the positions of the C2 and SPA S1 bands are indicated. The diagram below depicts each of the three C2 inserts used in the above S1 analysis.

efficient poly(A) site. In many cases GU- or U-rich sequence elements positioned immediately downstream of the site of poly(A) addition significantly increase the efficiency of polyadenylation. Indeed, we have designed a highly efficient SPA based solely on these two elements with appropriate spacing between them (Levitt *et al.*, 1989). However, the recent findings that a number of viral poly(A) signals also possess USE, often in addition to GU/U-rich elements, demonstrates that a more complex structure may exist. It remains a possibility that poly(A) signals in viruses may be specialized RNA processing elements that possess features not generally present in chromosomal genes. For example, the special need of retro- and pararetroviruses to by-pass promoter-proximal poly(A) sites is not likely to be common in chromosomal genes (Imperiale and DeZazzo, 1991; Proudfoot, 1991) Thus the presence of USEs in HIV-1 and GSHV may reflect the need for a level of control not required by most other poly(A) signals.

The C2 poly(A) signal is the first non-viral example to show dependence on a USE. This USE extends over ~50 bp and appears to be bipartite in structure. Interestingly, as in the case of viral upstream elements, this USE is a very U-rich element (42% U), although it does not share any other significant homology with viral USEs. Our results indicate that much of the 53 nt USE region appears to have some effect on C2 poly(A) site efficiency. This may argue that the USE has an overall RNA secondary or tertiary structure required for its activity. The greater severity of either deletions or oligo(G) replacements in the USE is consistant with this notion, as these two types of mutation would be more likely to disrupt RNA structure. The fact that this USE is so highly conserved between the five different mammals further emphasises the importance of this sequence element.

We also observe that the sequence between the AAUAAA and poly(A) site is highly conserved and, based on mutagenesis, is also required for full poly(A) site efficiency; G replacements have a more drastic effect than other nucleotide changes. Since cleavage/poly(A) factors must closely interact with this region of the poly(A) signal, it seems likely that oligo(G) sequences block these interactions. Downstream of the C2 cleavage/poly(A) site the sequence diverges significantly between the five different mammals. For the human poly(A) site this region appears to be relatively unimportant for efficiency. However, the four other mammalian sequences have a clear GU-rich element at this position, which we show, at least for rabbit, has a significant activatory effect. Indeed, it is apparent from the data in Figure 8 that the USE and GU-rich elements of the C2 poly(A) site have equivalent activatory effects. We speculate that this extra activatory element in mouse, rabbit, cat and cow may be required to compensate for the closer proximity of the G-rich MAZ binding sequences in these four species. This may be necessary because we also show that when the human C2 sequence is mutated to a G-rich sequence at this position it is significantly inhibited.

We finally demonstrate that all five C2–FB sequences have Pol II termination signals positioned close to their poly(A) sites. Obviously the proximity of the factor B gene promoter requires that C2 transcription is abruptly halted before it invades downstream transcriptional initiation signals. Although the close spacing of the C2 and Factor B genes is unusual in mammals, such a gene arrangement may well be the norm in more compact eukaryotic genomes, such as viruses and yeasts. We have recently shown that the poly(A) signals and site of Pol II termination are very closely spaced in the *ura4* gene of *Schizosaccharomyces pombe* (Humphrey *et al.*, 1994; C.Birse, personal communication). It therefore follows that the C2–FB mRNA 3′-end formation signals may share common features with lower eukaryotes such as yeasts. The presence of USE elements associated with poly(A) signals may be a general feature of closely spaced genes, such as those found in viruses and yeasts (Wahle, 1995). Furthermore, in the few cases so far investigated plant poly(A) signals also require USEs for full activity (MacDonald *et al.*, 1991; Mogen *et al.*, 1992). This may indicate that the general arrangement of mRNA 3′-end formation signals in eukaryotes is a poly(A) signal with its associated USE, closely followed by termination signals. Mammalian mRNAs may deviate from this arrangement because their genes are in general more widely separated. The exception that proves the rule would therefore be the rare cases where mammalian genes are closely spaced, such as C2 and Factor B.

Materials and methods

Plasmid constructions

α/C2 constructs. C2 poly(A) site fragments were isolated using the restriction sites indicated above the C2 sequence in Figure 1. The Δ2

and Δ3 sites correspond to *Bal*31 deletion end-points obtained from a deletion series made downstream of the C2 poly(A) signal in pMLC2.B. A *Bgl*II linker was added to each deletion end-point to allow cleavage at that position (Ashfield *et al.*, 1991). All these C2 fragments were filled in with DNA polymerase and inserted into the *Pvu*II site of the human α2-globin gene-containing plasmid α2W3'PSpSVed (Whitelaw and Proudfoot, 1986) to generate the α/C2 constructs. α/α and α/αH constructs were made by inserting either an α-globin poly(A) site fragment (150 bp *Apa*I–*Pvu*1II fragment) or the equivalent fragment isolated from α/H2Δ3 (200 bp *Apa*I–*Ase*I fragment), including the C2 USE region, into the same α2-globin gene *Pvu*II site.

C2/S constructs. The SPA site was isolated from the polylinker sequence of plasmid pySPA– (Enriquez-Harris *et al.*, 1991) as a filled in *Bam*HI–*Asp*718 fragment and inserted into the *Hpa*I site of α2M3'PS pSVed (Whitelaw and Proudfoot, 1986). The various C2 poly(A) site fragments were then inserted into the 5'-positioned *Pvu*II site as above. The C2–FB gene *Sty*I–*Sty*I fragment was isolated from a subclone of the C2–FB sequence (pGemC2.B) in which the intergenic 890 bp *Nco*I fragment (extending from near the termination codon of C2 to the initiation codon of Factor B) was inserted into pGem-72f at the *Sma*I site in the sense orientation.

C2 poly(A) site mutations generated by PCR. Specific mutations to the C2 poly(A) site were generated as follows. Pairs of C2 oligonucleotide primers were synthesized with specific mutations at their 5'-ends. Oligonucleotide pairs were positioned next to one another on the C2 sequence, but reading in opposite polarity. Each primer was then combined with either forward or reverse primers hybridizing outside the pGem polylinker sequence and used to generate either 5' or 3' C2–FB PCR products. These PCR DNAs were cut at restriction sites in the polylinker region and then combinations of 5' and 3' PCR products were ligated into pUC18 at homologous restriction sites. Such a procedure generated specific sequence replacements or deletions to the C2 poly(A) site region sequence. Each C2 mutation was sequenced and the *Sty*I–*Sty*I fragment containing the mutant C2 poly(A) site was then isolated and inserted into the *Pvu*II site of the α-globin gene as above. The sequences of each C2 poly(A) site mutation are shown in Figures 4 and 5.

Isolation of mouse, rabbit, cat and cow C2–FB sequences by PCR

PCR primers were synthesized that correspond to the C-terminus of the human C2 (TGCAGCCCTGGCTGAGGCAGCACCT) or the N-terminus of the Factor B gene (AGGGCATCAGGCAGAGTTGGGGGCT). PCR reactions were carried out using this primer pair and ~1 μg genomic DNA purified from mouse, rabbit, cat and cow. Low stringency annealing conditions were used to allow potential mismatching between the primers (based on the human sequence) and other mammalian C2 or Factor B sequences. PCR products were fractionated on 2% agarose gels blotted onto nylon filters (Hybond N; Amersham) and hybridized with a human C2–FB probe. Positive bands were obtained (0.8–1.0 kb) for each genomic DNA and these were then purified and cloned into the commercial vector pCRII, which has T residues added to the insert site (TA cloning vector; Invitrogen). The four C2–FB clones were then sequenced using a combination of Sequenase sequencing with either vector primers or internal primers (US Biochemical) and by automated ABI sequencing.

Definition of C2 terminator sequences using a poly(A) site competition assay

Fragments of the mouse, rabbit, cat and cow C2–FB intergenic region were isolated using PCR primers to define the 5' boundary [corresponding to the sequence that immediately follows the C2 poly(A) signal AAUAAA sequence] and restriction sites for the 3' boundary: mouse (M), primer–*Dra*I (140 bp); rabbit (Ra), primer–*Apa*I (320 bp); cat (Ca), primer–*Apa*I (290 bp); cow (Co), primer–*Apa*I (300 bp); deleted cow (ΔCo) *Sty*I–*Bss*HII (140 bp). The ΔCo *Sty*I site is immediately 5' of the cow C2 MAZ site (see Figure 6). The human terminator fragment (H), *Sty*I–*Bam*HI (160 bp), has been previously described. Each of these C2 terminator fragments were cloned into the poly(A) site competition vector SPA^+ (Ashfield *et al.*, 1991) and transfected into HeLa cells as below.

Cell transfections and RNA analysis

The methods used for HeLa cells transfection and RNA analysis are described elsewhere (Levitt *et al.*, 1989). All plasmids were co-transfected with the plasmid pBR328RβSV (Grosveld *et al.*, 1982), containing both the SV40 T antigen gene, to provide SV40 T antigen necessary for the replication of the transfected plasmids, and the rabbit β-globin gene, to serve as a transfection control. S1 probe/RNA hybridizations were performed at 52°C for 16 h and the S1 nuclease digestions at 30°C for 1 h. S1 nuclease digestion products were separated by electrophoresis through 5% polyacrylamide–7 M urea gels. Autoradiography was performed at –70°C with intensifying screens. In most cases the gels were dried and subjected to phosphoimager analysis (Molecular Dynamics) to allow accurate quantitation. In some cases quantitation of the data was obtained by densitometry of the radioautographs.

Acknowledgements

We thank members of our laboratory for help and encouragement throughout these studies and in particular Rebecca Ashfield and Jan Eggermont for providing essential starting materials. We are also grateful to Duncan Campbell for providing mammalian genomic DNA samples. A.M. is supported by Porto University and the JNICT of Portugal. These experiments were supported by a programme grant from the Wellcome Trust (No. 032773).

References

Ashfield,R., Enriquez-Harris,P. and Proudfoot,N.J. (1991) Transcriptional termination between the closely spaced linked human complement genes C2 and Factor B: common termination factor for C2 and c-*myc*? *EMBO J.*, **10**, 4197–4207.

Ashfield,R., Patel,A.J., Bossone,S.A., Brown,H., Campbell,R.D., Marcu,K.B. and Proudfoot,N.J. (1994) MAZ-dependent termination between closely spaced human complement genes. *EMBO J.*, **13**, 5656–5667.

Asselin,C., Nepveu,A. and Marcu,K.B. (1989) Molecular requirements for transcriptional initiation of the murine c-*myc* gene. *Oncogene*, **4**, 549–558.

Bilger,A., Fox,C.A., Wahle,E. and Wickens,M. (1994) Nuclear polyadenylation factors recognize cytoplasmic polyadenylation elements. *Genes Dev.*, **8**, 1106–1116.

Bossone,S.A., Asselin,C., Patel,A.J. and Marcu,K.B. (1992) MAZ, a zinc finger protein, binds to C-MYC and C2 gene sequences regulating transcriptional initiation and termination. *Proc. Natl Acad. Sci. USA*, **89**, 7452–7456.

Brown,P.H., Tiley,L.S. and Cullen,B.R. (1991) Efficient polyadenylation within the human immunodeficiency virus type 1 long terminal repeat requires flanking U3-specific sequences. *J. Virol.*, **65**, 3340–3343.

Carswell,S. and Alwine,J.C. (1989) Efficiency of utilisation of the simian virus 40 late polyadenylation site: effects of upstream sequences. *Mol. Cell. Biol.*, **9**, 4248–4258.

Cherrington,J. and Ganem,D. (1992) Regulation of polyadenylation in human immunodeficiency virus (HIV): contributions of promoter proximity and upstream sequences. *EMBO J.*, **11**, 1513–1524.

DeZazzo,J.D. and Imperiale,M.J. (1989) Sequences upstream of AAUAAA influence poly(A) site selection in a complex transcription unit. *Mol. Cell. Biol.*, **9**, 4951–4961.

DeZazzo,J.K., Kilpatrick,J.E. and Imperiale,M.J. (1991) Involvement of a long terminal repeat U3 sequence overlapping the transcription control region in human immunodeficiency virus type 1 mRNA 3' end formation. *Mol. Cell. Biol.*, **11**, 1624–1630.

Enriquez-Harris,P., Levitt,N., Briggs,D. and Proudfoot,N.J. (1991) A pause site for RNA polymerase II is associated with termination of transcription. *EMBO J.*, **10**, 1833–1842.

Gil,A. and Proudfoot,N.J. (1984) A sequence downstream of AAUAAA is required for rabbit β globin mRNA 3' end formation. *Nature*, **312**, 473–474.

Gil,A. and Proudfoot,N.J. (1987) Position-dependent sequence elements downstream of AAUAAA are required for efficient mRNA 3' end formation. *Cell*, **49**, 399–406.

Gilmartin,G.M., Fleming,E.S. and Oetjen,J. (1992) Activation of HIV-1 3' processing *in vitro* requires both an upstream element and TAR. *EMBO J.*, **11**, 4419–4428.

Gilmartin,G.M., Fleming,E.S., Oetjen,J. and Graveley,B.R. (1995) CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition. *Genes Dev.*, **9**, 72–83.

Grosveld,G.G., de Boer,E., Shewmaker,C.K. and Flavell,R.A. (1982) DNA sequences necessary for transcription of the rabbit β-globin gene *in vivo*. *Nature*, **295**, 120–126.

Hart,R.P., McDevitt,M.A. and Nevins,J.R. (1985) Poly(A) site cleavage

in HeLa nuclear extract is dependent on downstream sequences. *Cell*, **43**, 677–683.
Humphrey,T., Birse,C.E. and Proudfoot,N.J. (1994) RNA 3′ end signals of the *S.pombe ura4* gene comprise a site determining and efficiency element. *EMBO J.*, **13**, 2441–2451.
Imperiale,M.J. and DeZazzo,J.D. (1991) Poly(A) site choice in retroelements—deja-vu all over again. *New Biologist*, **3**, 531–537.
Levitt,N., Briggs,D., Gil,A. and Proudfoot,N.J. (1989) Definition of an efficient synthetic poly(A) site. *Genes Dev.*, **3**, 1019–1025.
Lutz,C.S. and Alwine,J.C. (1994) Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 polyadenylation signal. *Genes Dev.*, **8**, 576–586.
Manley,J.L. (1988) Polyadenylation of messenger-RNA precursors. *Biochim. Biophys. Acta*, **950**, 1–12.
MacDonald,M.H., Mogen,B.D. and Hunt,A.G. (1991) Characterisation of the polyadenylation signal from the T-DNA encoded octopine synthase gene. *Nucleic Acids Res.*, **19**, 5575–5581.
McDevitt,M.A., Imperiale,M.J., Ali,H. and Nevins,J.R. (1984) Requirement of a downstream sequence for generation of a poly(A) addition site. *Cell*, **37**, 993–999.
McDevitt,M.A., Hart,R.P., Wong,W.W. and Nevins,J.R. (1986) Sequences capable of restoring poly(A) site function define two distinct downstream elements. *EMBO J.*, **5**, 2907–2913.
McLauchlan,J., Gaffney,J.L., Whitton,J.L. and Clements,J.B. (1985) The consensus sequence YGTGTTYY located downstream from the AAUAAA signal is required for efficient formation of mRNA 3′ termini. *Nucleic Acids Res.*, **13**, 1347–1368.
Mogen,B.D., MacDonald,M.H., Leggewie,G. and Hunt,A.G. (1992) Several distinctive types of sequence elements are required for efficient mRNA 3′ end formation in a pea *rbcS* gene. *Mol. Cell. Biol.*, **12**, 5406–5414.
Nonaka,M., Gitlin,J.D. and Colten,H.R. (1989) Regulation of human and murine complement: comparison of 5′ structural and functional elements regulating human and murine complement factor B gene expression. *Mol. Cell. Biochem.*, **89**, 1–14.
Peterson,M.L. (1994) Regulated immunoglobulin (Ig) RNA processing does not require specific *cis*-acting sequences: non-Ig RNA can be alternatively processed in B cells and plasma cells. *Mol. Cell. Biol.*, **14**, 7891–7898.
Proudfoot,N.J. (1991) Poly(A) signals. *Cell*, **64**, 671–674.
Proudfoot,N.J. and Brownlee,G.G. (1976) 3′ Non coding region sequences in eukaryotic messenger RNA. *Nature*, **263**, 211–214.
Proudfoot,N.J., Lee,B.A. and Monks,J. (1992) Multiple SP1 binding sites confer enhancer-independent, replication-activated transcription of HIV-1 and globin gene promoters. *New Biologist*, **4**, 369–381.
Rothnie,H.M., Reid,J. and Hohn,T. (1994) The contribution of AAUAAA and the upstream element UUUGUA to the efficiency of messenger-RNA 3′-end formation in plants. *EMBO J.*, **13**, 2200–2210.
Russnak,R. (1991) Regulation of polyadenylation in hepatitis B viruses: stimulation by the upstream activating signal PS1 is orientation-dependent, distance-dependent, and additive. *Nucleic Acids Res.*, **19**, 6449–6456.
Russnak,R. and Ganem,D. (1990) Sequences 5′ to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses. *Genes Dev.*, **4**, 764–776.
Sanfaçon,H., Brodmann,P. and Hohn,T. (1991) A dissection of the cauliflower mosaic virus polyadenylation signal. *Genes Dev.*, **5**, 141–149.
Schek,N., Cooke,C. and Alwine,J. (1992) Definition of the upstream efficiency element of the Simian Virus 40 late polyadenylation signal by using *in vitro* analyses. *Mol. Cell. Biol.*, **12**, 5386–5393.
Takagaki,Y., Manley,J.L., MacDonald,C.C., Wilusz,J. and Schenk,T. (1990) A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. *Genes Dev.*, **4**, 2112–2120.
Valsamakis,A., Zeichner,S., Carswell,S. and Alwine,J.C. (1991) The human immunodeficiency virus type 1 polyadenylation signal: a 3′ long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylation. *Proc. Natl Acad. Sci. USA*, **88**, 2108–2112.
Valsamakis,A., Schek,N. and Alwine,J.C. (1992) Elements upstream of the AAUAAA within the human immunodeficiency virus polyadenylation signal are required for efficient polyadenylation *in vitro*. *Mol. Cell. Biol.*, **12**, 3699–3705.
Wahle,E. (1995) 3′-end cleavage and polyadenylation of mRNA precursors. *Biochim. Biophys. Acta*, **1261**, 183–194.
Wahle,E. and Keller,W. (1992) The biochemistry of 3′-end cleavage and polyadenylation of messenger RNA precursors. *Annu. Rev. Biochem.*, **61**, 419–440.
Weiss,E.A., Gilmartin,G.M. and Nevins,J.R. (1991) Poly(A) site efficiency reflects the stability of complex formation involving the downstream element. *EMBO J.*, **10**, 215–219.
Whitelaw,E. and Proudfoot,N.J. (1986) Alpha thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3′ end processing in the human α2 globin gene. *EMBO J.*, **5**, 2915–2922.
Wickens,M. (1990) How the messenger got its tail: addition of poly(A) in the nucleus. *Trends Biochem. Sci.*, **15**, 271–281.
Wilusz,J., Shenk,T., Takagaki,Y. and Manley,J.L. (1990) A multi-component complex is required for the AAUAAA-dependent cross-linking of a 64-kilodalton protein to polyadenylation substrates. *Mol. Cell. Biol.*, **10**, 1244–1248.
Wu,L.C., Morley,B.J. and Campbell,R.D. (1987) Cell specific expression of the human complement protein Factor B gene: evidence for the role of two distinct 5′ flanking elements. *Cell*, **48**, 331–342.

Received on April 3, 1995; revised on May 11, 1995