Bipartite structure of the downstream element of the mouse beta globin (major) poly(A) signal

Jing S.Chen and Jeffrey L.Nordstrom*

Department of Biological Sciences, Fordham University, Bronx, NY 10458, USA

Received December 12, 1991; Revised and Accepted April 13, 1992

ABSTRACT

The downstream region of the mouse beta (major) globin poly(A) signal was mutated and analyzed for function in transfected COS cells. From analysis of unidirectional Bal31 deletions, the 3' boundary of the downstream element was defined as +22 (22 nucleotides downstream from the cleavage site). Analysis of cluster mutations, in which 5 or 6 adjacent bases were replaced with a random CA-containing sequence in a manner that did not alter spacing, confirmed +22 as the 3' boundary of the downstream element. The analysis also revealed two short UG-rich sequences, located from +5 to +10 and from +17 to +22, as major functional components. In contrast, a more refined series of mutations, in which clusters of 3 bases were replaced, failed to cause loss of function. We conclude that the downstream element of the mouse beta globin poly(A) signal is bipartite in structure, and that portions of its sequence are functionally redundant.

INTRODUCTION

A poly(A) signal is required for the formation of the 3' end of most eukaryotic mRNA [See references 1–3 for reviews]. The signal directs two RNA processing reactions: site-specific endonucleolytic cleavage of the RNA transcript, and stepwise addition of adenylates (approx. 250 in vertebrates) to the newly generated 3' end to form the poly(A) tail. The cleavage and poly(A) addition reactions are coupled and involve a multiplicity of protein-containing factors, including AAUAAA-specificity factor, cleavage factors and poly(A) polymerase, that assemble into a large (500–1000 kDa) complex.

A poly(A) signal has at least three parts: hexanucleotide, cleavage site and downstream element. The hexanucleotide is typically AAUAAA and cleavage sites are most frequently 3' to the dinucleotide CA [4–6]. Downstream elements are required for optimal poly(A) signal function and are located downstream of the cleavage site [1]. Mutations in downstream regions primarily affect cleavage efficiency, but some have caused shifts in cleavage site [7–9]. Although various consensus sequences have been proposed [10–12], sequence requirements for

downstream elements remain vague and are currently viewed as diffuse UG- or U-rich sequences [2,3]. Lengths of downstream elements and their distance from cleavage sites also are not well defined; most important sequences appear to be located within 40 nt of the cleavage site [3,11,13–16], but some are located significantly further downstream [17–20]. Refinements in the definition of downstream elements have been thwarted by the common occurrence of UG- or U-rich regions in the sequences used to replace them [1], and several poly(A) signals, based on the failure of most small substitutions to inhibit function, appear to have multiple downstream components that are functionally redundant [7,21,22]. However, the inhibition of poly(A) signal function by certain small substitutions (22) suggests that certain downstream element positions are not redundant. For some poly(A) signals, upstream elements also have been identified [23–26].

In this report, we have begun to define the critical portions of the downstream element of the mouse β^{maj} -globin poly(A) signal. First, from an analysis of Bal31 deletions, the 3' boundary of the downstream element was localized to 22 nucleotides beyond the cleavage site. Second, using a series of cluster mutations generated by PCR with degenerate primers, we replaced 5–6 base regions of the downstream element with various C-A-containing sequences in a manner that did not alter spacing. The cluster mutation experiments confirmed the 3' boundary of the downstream element region, and revealed two short UG-rich sequences that are separated by approximately 6 nucleotides as its major functional components. Thus, the downstream element of the mouse β^{maj} -globin poly(A) signal is located within a small region and has a bipartite structure. Analysis of a series of 3-base cluster mutations failed to show inhibition of poly(A) signal function, thus demonstrating that portions of the downstream sequence are functionally redundant.

METHODS

Plasmids

Balb/c mouse β^{maj} -globin sequences were derived from λ gtWES.M β G2 [27]. A 2.7 kb PstI-MboI fragment (1150 bp before the transcription start site to 147 bp beyond the poly(A) addition site) was subcloned between the PstI and BamHI sites

* To whom correspondence should be addressed

of pUC8 [28] to generate p β 2.7. The 3' portion of the gene, from the BamHI site in exon 2 to 156 bp beyond the poly(A) addition site, was removed from p β 2.7 as a 1088 bp BamHI-EcoRI fragment and ligated to the 2958 bp BglII-EcoRI fragment of pSV2neo [29] to generate pDH2. pMBG1 was constructed in four steps: First, p β 2.7 was modified by addition of a Sall linker (GGTCGACC) to the MscI site 160 bp upstream from the poly(A) addition site, and a XhoI linker (CCTCGAGG) to the SmaI site in the pUC8 portion of the plasmid. Second, pSV2neo was modified by addition of a BamHI linker (GGGATCCC) to its unique SmaI site. Third, the 1100 bp BamHI-EcoRI fragment of the Sal/XhoI-modified p β 2.7 was ligated to the 3960 bp BamHI-EcoRI fragment of the Bam-modified pSV2neo to generate an intermediate clone named pMBG1- Δ SVL, so named because it lacks the SV40 late poly(A) signal. Fourth, an EcoRI linker (GG-AATTCC) was added to the RsaI site of pSV2neo located at 4676, and the 1066 bp EcoRI fragment from this EcoRI-modified plasmid was inserted into the single EcoRI site of pMBG1- Δ SVL. The resultant plasmid, pMBG1, has the SV40 late poly(A) site located 1058 nts downstream of the mouse β -globin poly(A) site (Fig. 1A). pMBG2, constructed from pMBG1 by deletion of the 489 bp XhoI-ApaI fragment, has the mouse β -globin and SV40 late poly(A) sites separated by 570 nts. pMBG1kpn Δ +27 was constructed by adding a KpnI linker (GGGTACCC) to the XhoI site (blunted by Klenow DNA polymerase) of pMBG1 Δ +27.

Bal31 Deletions

Sequences flanking the mouse β -globin poly(A) site were deleted to various extents by Bal31 digestion of EcoRI-digested pDH2 DNA (the EcoRI site is 156 bp downstream of the poly(A) site) for various times [30], followed by religation in the presence of XhoI linker. The resultant bidirectional deletions were screened by restriction enzyme analysis and endpoints determined by DNA sequencing using Sequenase 2.0 (United States Biochemicals). The pMBG1-series of unidirectional deletions were constructed by replacing the 799 bp EcoRV-XhoI fragment of pMBG1 with the EcoRV-XhoI fragments of the various deleted pDH2 constructs. Endpoints of the pMBG1-series of deletions were confirmed by DNA sequencing. The pMBG2-series of deletions were constructed by deletion of the 489 bp XhoI-ApaI fragment from each member of the pMBG1-series.

Cluster mutations

Fragments of DNA with clusters of consecutive point mutations in targeted portions of the mouse β -globin poly(A) signal were obtained by using the polymerase chain reaction (PCR) with degenerate primers. Primer C1 (5' CCACTGCCTTGGCTCA-CAAGTACCA 3') spans the region from -160 to -136, with its 3' end at -136 (numbers are relative to the cleavage site which occurs between -1 and +1). The degenerate primers (where K = G or T) span the region from +29 to +1, with their 3' ends at +1. Primer C2 (5' GGKKKKKGACAAATAATTTA-AAACATCATT 3') targets mutations to +23/+27; C3 (5' GGACACAKKKKKKTAATTTAAAACATCATT 3') to +17/+22; C4 (5' GGACACAGACAAAKKKKKKAAAACATCATT 3') to +11/+16; C5 (5' GGACACAGACAAATAATTTKKKKKKTTCATT 3') to +5/+10; C6 (5' GGACAKK-KACAAATAATTTAAAACATCATT 3') to +23/+25; C7 (5' GGACACAGKKKAATAATTTAAAACATCATT 3') to +20/+22; C8 (5' GGACACAGACAKKKAATTTAAAACATCA-TT 3') to +17/+19; and C9 (5' GGACACAGACAAATKKK-

TTAAAACATCATT 3') to +14/+16. PCR reactions contained 18 μ g of pMBG1 Δ +27 DNA, 60 pmoles of primer C1, 60 pmoles of one of the other primers, and two units of Taq DNA polymerase (Promega Biochemicals) in 100 μ l of 20 mM Tris-HCl (pH 8.3), 1.5 mM MgCl₂, 25 mM KCl, 0.05% Tween 20, 0.1 μ g/ μ l bovine serum albumin, and 50 μ M dNTP's [31]. A Hybaid thermoreactor was used for temperature programming, which was 2 min at 93°; then 5 cycles of 93°/20sec, 55°/20sec, 60°/30sec, 72°/20sec; then 45 cycles of 93°/20sec, 55°/20sec, 72°/20sec. The amplified DNA products (190 bp in length) were directly cloned into the blunt-ended SmaI site of pGEM3Z (Promega). Plasmids with inserts were screened for orientation by restriction enzyme analysis and sequenced. The Sall-KpnI fragments containing the clustered mutations were then substituted for the Sall-KpnI fragment of pMBG1kpn Δ +27.

Probes

To monitor cleavage at the globin and SV40 late poly(A) sites, uniformly labeled riboprobe was prepared by T7 RNA polymerase transcription of MscI-digested pMOB2 (M. Bordonaro, unpublished). This probe consists of a 306 nt region spanning the wild type mouse β -globin poly(A) site that is linked to a 314 nt region spanning the SV40 late poly(A) site (Fig. 1B). For homologous probes, the smaller BamHI fragment from each of the pMBG1 deletions was cloned into the BamHI site of pGEM3Z, such that the BamHI site of globin exon 2 was proximal to the T7 promoter. After deletion of the BsmI-XbaI fragment, homologous probes were prepared by RsaI digestion and transcription by SP6 RNA polymerase. The BamHI fragment for pMBG1 Δ +10 was cloned into pGEM3Z in opposite orientation; following deletion of the BsmI-SmaI fragment, homologous probe was prepared by PstI digestion, blunt-ending with Klenow DNA polymerase, and transcription by T7 RNA polymerase. Homologous probes for each of the +5/+10 cluster mutations in pGEM3Z was prepared by SP6 RNA polymerase transcription of RsaI digested DNA. All riboprobes were prepared with [α -³²P]UTP (New England Nuclear) according to the procedures of Promega Biochemicals.

Cell cultures, transfections and RNA analyses

COS-1 cells [32], obtained from the American Type Culture Collection (ATCC CRL 1650), were cultured and transfected with plasmids using DEAE-dextran as described previously [33]. Total cellular RNA was isolated 48 hours later by acidic phenol extraction of cells lysed in sodium guanidine isothiocyanate [34] in the presence of 1 mM aurintricarboxylic acid, a nonspecific nuclease inhibitor [35]. Mouse erythroleukemia (MEL) cells were obtained from the Human Genetic Mutant Cell Repository (GM0086E, clone #745) and total RNA isolated after dimethylsulfoxide induction [36]. RNase T2 mapping was performed according to the guidelines of Promega Biochemicals with hybridizations at 42°C for 24 hours and digestions at 30°C for 1 hour with 33 units/ml RNase T2 (GIBCO/BRL). Samples were electrophoresed through 6% denaturing polyacrylamide gels with wedge spacers for 1–3 hours at 84 watts. Autoradiography was performed by exposing Kodak X-Omat AR Film to each gel at -80°C with and without intensifying screens. Densitometry was performed on autoradiographs that were within the range of sensitivity of the densitometer (Model 620, Bio-Rad Laboratories, Richmond, CA), which had been calibrated as

described by Swillens et al.[37]. Correction curves for autoradiographs exposed with and without intensifying screens were used, and the corrected intensity of each band was normalized for the number of ^{32}P -labeled uridine residues.

RESULTS

Definition of the 3' boundary of the downstream element

pMBG1 is a chimeric construct in which the 3' half of the mouse β -globin gene, including its second intron and poly(A) signal, is placed downstream of the SV40 promoter and neo region and upstream from the SV40 late poly(A) signal (Fig. 1A). In transfected COS cells, plasmids of this type are replicated and transcribed efficiently [38]. Since cleavage and polyadenylation are required for mRNA stabilization and accumulation, the distal SV40 late poly(A) site serves to stabilize those transcripts that fail to be processed at the β -globin poly(A) site. The percentage of transcripts processed at the β -globin site, determined by quantitative RNase T2 mapping, is used as the measure of β -globin poly(A) signal function; this value is approximately 90% for the wild type signal (data not shown).

Various portions of sequence on the 3' side of the globin poly(A) site were deleted by Bal31 exonuclease digestion, beginning at a site located 156 bp downstream from the globin poly(A) site. A XhoI linker was added to each deletion endpoint,

and deleted globin sequences were replaced by a portion of SV40 late sequence (see Fig. 1 legend). Transcripts from these unidirectional deletions were analyzed with a uniformly labeled riboprobe, derived from pMOB2, that is complementary to the SalI-XhoI and BamHI-EcoRI fragments that span the β -globin and SV40 late poly(A) sites, respectively (Fig. 1B). Transcripts polyadenylated at the β -globin site protect a band of 160 nt (labeled 'MBG pA'; from MscI to MBG pA). Transcripts not polyadenylated at the MBG pA site protect two bands: one of variable length (labeled 'no MBG pA'; from MscI to the deletion endpoint), and one of 142 nt (labeled 'SVL pA', from BamHI to SVL pA). These data permit globin poly(A) signal function to be calculated in two equivalent ways: MBG pA vs. no MBG pA, or MBG pA vs. SVL pA.

Analysis of transcripts from the pMBG1-series of deletions is shown in Fig. 2A (lanes 1–6). Deletion of globin sequences up to +22 caused little change in globin poly(A) site usage (lanes 1–3), but a significant transition, noticeable mainly as increased SV40 late poly(A) site usage, occurred when sequences between +22 and +16 were deleted (lane 4). Deletion to +10 had little further effect (lane 5), and deletion to –1 was ambiguous, due to convergence in the sizes of the 'MBG pA' and 'no MBG pA' bands (lane 6). The RNA analyses were repeated 5 times, and band intensities were quantitated by densitometry. These data, summarized in Fig. 2B, show that globin poly(A) signal function decreased by approximately 37% after deletion of sequences

Figure 1. A. Structure of the expression vector, pMBG1: The plasmid is 6124 bp in length and contains DNA sequences from SV40 (white boxes), neo (shaded box), mouse β^{maj} -globin (cross-hatched box), and pBR322 (not shown). Locations of the SV40 early transcription start site, globin splice sites, and globin (MBG pA) and SV40 late (SVL pA) poly(A) sites are marked. Structures of the two major RNA products are depicted. The 489 bp XhoI-ApaI fragment that was deleted from pMBG1 to make pMBG2 is indicated. In pMBG1, the MBG and SVL poly(A) sites are separated by 1058 nts; in pMBG2, these sites are separated by 570 nts. B. Structure of the probe used to assay MBG and SVL poly(A) site usage: In pMOB2 two separate fragments from pMBG1 (the 315 bp MscI-XhoI fragment with the MBG pA site, and the 371 bp BamHI-EcoRI fragment with the SVL pA site) have been fused together; 'XhoI/BamHI' marks the junction. The probe is a uniformly labeled T7 RNA polymerase transcript of pMOB2 digested with MscI. In RNase T2 protection assays, transcripts processed at the MBG pA site protect a band of 160 nt. Transcripts not processed at the MBG pA site protect two bands: One band, 315 nt in length and labeled as 'no MBG pA', represents transcripts that extend beyond the MBG site. Its exact length depends on the location of the deletion endpoint in the transcripts being assayed. The other band, 142 nt in length, represents transcripts processed at the SVL pA site. The unlabeled arrowhead marks the position of a cryptic poly(A) site that is located approximately 114 nt downstream from the normal mouse β -globin poly(A) site; all of the Bal31 deletion endpoints were located upstream from this site. C. Sequence of the mouse β^{maj} -globin poly(A) signal and locations of Bal31 deletion endpoints: The sequence of a cloned mouse β^{maj} -globin cDNA ends with GCAA(A)_n [52], and thus the exact location of the cleavage/polyadenylation site is ambiguous. Our assignment of the cleavage site is based on experiments with the adenovirus L3 and SV40 late poly(A) signals that demonstrate cleavage on the 3' side of CA dinucleotides [5,6]. Nucleotides downstream from the assigned cleavage site are given positive values. The unidirectional Bal31 deletions initiated from a site located 156 bp downstream from the globin poly(A) site. Each number designates the last remaining base of globin sequence present in a particular Bal31 deletion. Replacement sequences for the pMBG1-series, which are originate from SV40 late DNA and begin with a XhoI linker (underlined), are: CCUCGAGGGGAAUCCUUGCCUAAUUUAAAUGAGGACUUAACCUGUGGAA... To vary the replacement sequences, the pMBG2-series of deletions were constructed by deletion of the XhoI-ApaI fragments from each of the pMBG1-series deletions. These replacement sequences, which include a portion of the XhoI linker (underlined), are: CCUCGACUUGUGCAAAGCUGACAGCUUGUAUGUUUCUGCUGUUGACAUUUGU...

between +16 and +22. Deletion of adjacent sequence regions, from +22 to +27 and from +10 to +16, decreased globin poly(A) signal function minimally, by 2% and 6%, respectively. It appears, therefore, that a major component of the downstream element is located between +16 and +22.

To verify these data and to permit unambiguous evaluation of the function of the -1 deletion, RNAs also were analyzed with homologous probes. Portions of each deletion mutant were subcloned into pGEM3Z and transcribed into uniformly labeled riboprobes. The analysis with the $\Delta -1$ homologous probe shows that only 30% of the $\Delta -1$ transcripts were processed at the globin site (130 nt band), the remainder appearing as 'no MBG pA' RNA (257 nt band) (Fig. 2A, lane 8). The size of the MBG pA band for $\Delta -1$ transcripts (lane 8) coincides with that of wild type pMBG1 transcripts (lane 7); thus, no shift in cleavage site was evident even when all sequences downstream of the normal cleavage site were deleted. Similarly, no shifts in cleavage site were observed when transcripts from the other deletions were

analyzed with homologous probes (data not shown). Quantitation of the relative intensities of the cleaved and non-cleaved bands yields values that closely match those obtained with the pMOB2 probe (Fig. 2B). The homologous probe data confirm that the major component of the downstream element is located between +16 and +22, and that regions between +22 and +27, and between +10 and +16, are relatively unimportant. In addition, these data demonstrate that deletion of sequences between -1 and +10 further decreases globin poly(A) function by approximately 15%. Thus, an additional component may be located between -1 and +10.

Even after deleting all sequences downstream of the cleavage site, the mouse β -globin poly(A) signal retained 30% of its activity. Thus, the replacement sequences in the pMBG1-series may be restoring partial function in a serendipitous manner. In an attempt to evaluate the influence of the replacement sequences, the pMBG2-series of deletions were constructed. These have the same deletion endpoints as the pMBG1-series, but different

Figure 2. Assay of mouse β -globin poly(A) signal function in the Bal31 deletions: A. RNase protection assays of transcripts from the pMBG1-series of deletions: Lanes 1–6: Approximately 10 μ g total RNA from COS cells transfected with pMBG1 Δ +47 (lane 1), pMBG1 Δ +27 (lane 2), pMBG1 Δ +22 (lane 3), pMBG1 Δ +16 (lane 4), pMBG1 Δ +10 (lane 5), pMBG1 Δ -1 (lane 6) were analyzed with the MOB2 probe. In lane 2, the faint pair of 'no MBG pA' bands results from incomplete RNase T2 digestion of an imperfect duplex (approximately 9 bp) that happened to form between the riboprobe and RNA adjacent to the deletion endpoint; both bands represent transcripts not cleaved at the globin poly(A) site. Lane 3 is from a separate experiment than the others. In lane 4, the 'no MBG pA' band is slightly truncated, because the terminus of the riboprobe/RNA hybrid is A-rich and susceptible to over digestion by RNase T2 [60]. In lane 6, due to the deletion to -1 , the 'no MBG pA' and 'MBG pA' bands have same sizes. Lanes 7–8: Approximately 10 μ g total RNA from COS cells transfected with pMBG1 (lane 7) and pMBG1 Δ -1 (lane 8) were analyzed with a homologous probe prepared from a portion of pMBG1 -1 that was subcloned into pGEM3Z. For RNAs cleaved, or not cleaved at the globin poly(A) site, the homologous probe yields bands of 130 and 257 nt, respectively. In lane 7, the appearance of the 257 nt 'no MBG pA' band is an artifact due to incomplete RNase T2 digestion of the phosphodiester bond in the riboprobe opposite a looped out portion of pMBG1 RNA. B. Summary of the functional properties of the pMBG1-series of deletions: Relative band intensities within a lane were analyzed densitometrically and the percentage of transcripts cleaved at the MBG pA site was plotted versus deletion endpoint. The percentage of cleavage at the MBG pA site was calculated as either (100)(MBG pA)/(MBG pA + SVL pA) or as (100)(MBG pA)/(MBG pA + no MBG pA). Open squares connected by a black line are data obtained with the MOB2 probe; bars indicate the standard deviation from 5 experiments. Diamonds connected by a grey line are data obtained with homologous probes. C. RNase protection assays of transcripts from the pMBG2-series of deletions: Approximately 10 μ g of total RNA from COS cells transfected with pMBG2 Δ +74 (lane 1), pMBG2 Δ +47 (lane 2), pMBG2 Δ +27 (lane 3), pMBG2 Δ +22 (lane 4), pMBG2 Δ +16 (lane 5), pMBG2 Δ +10 (lane 6) and pMBG2 Δ -1 (lane 7) were analyzed with the pMOB2 probe. D. Summary of the functional properties of the pMBG2-series of deletions.

replacement sequences. In addition, the changes reduce the spacing between the globin and SV40 late poly(A) sites (Fig. 1A). The data for the RNAs from the pMBG2-series of deletions are shown in Fig. 2C and summarized in Fig. 2D. Although very similar, the data differ from those of the pMBG1-series in two ways. First, the maximal activity of the globin poly(A) signal in the pMBG2-series is reduced from 90% to 84%. This small, but reproducible, decrease may reflect an increase in the cis-

competition between the globin and SV40 poly(A) signals when the distance between them is shortened by 589 nt. Alternatively, it may reflect the presence in the replacement sequences of an inhibitory sequence (see below). Second, although a major decline in globin poly(A) signal function occurred when sequences between +16 and +22 were deleted, a decline of similar magnitude also occurred when sequences between +22 and +27 were deleted. These data suggest that sequences between +22 and +27 are important. However, due to the nature of their construction, the +22 deletions of pMBG1 and pMBG2 have identical sequences (ccucg) in the region between +22 and +27 (Fig. 2). Thus, the observed difference in poly(A) signal function must be due to replacement sequences that are further downstream. A possible explanation is a secondary structure that inhibits the accessibility of a portion of the downstream element. Based on this line of reasoning, +22 remains the best estimate for the 3' boundary of the mouse beta globin downstream element.

The downstream element is bipartite

To refine its boundaries and structure, a series of cluster mutations were constructed in which 5–6 bp portions of the downstream element region were replaced with differing, random C-A-containing sequences. This approach was developed so as to avoid the ambiguities associated with non-neutral replacement sequences and altered spacing arrangements. The cluster mutations were generated by PCR with degenerate primers. Randomized CA-containing sequences were chosen for the replacement sequences because they lack the G's and U's normally associated with downstream elements, and because C-A-containing sequences have been shown to be ineffective as downstream elements [22,39]. The nucleotide sequences of the individual cluster mutants are shown in Fig. 3. Four different CA-rich sequences, labeled A through D, were generated for each 5–6 nt cluster to control for anomalies that might result from a particular sequence. Analysis of transcripts from representative cluster mutants with the pMOB2 probe is shown in Fig. 4A (lanes 2–10). Because of their close proximity to the cleavage site, homologous probes were needed to analyze the function of the

		pA		+30
Wild type		CAAUGAUGUUUAAAUAUUUGUCUGUGUCAU		
+23/+27	A	CAAUGAUGUUUAAAUAUUUGUC <u>cccc</u> CAU		
	B	CAAUGAUGUUUAAAUAUUUGUC <u>acccc</u> CAU		
	C	CAAUGAUGUUUAAAUAUUUGUC <u>ccccc</u> CAU		
	D	CAAUGAUGUUUAAAUAUUUGUC <u>accac</u> CAU		
+17/+22	A	CAAUGAUGUUUAAAUA <u>cccaac</u> UGUGUCAU		
	B	CAAUGAUGUUUAAAUA <u>acccc</u> UGUGUCAU		
	C	CAAUGAUGUUUAAAUA <u>caacc</u> -UGUGUCAU		
	D	CAAUGAUGUUUAAAUA <u>cccacc</u> UGUGUCAU		
+11/+16	A	CAAUGAUGUUU <u>cccccc</u> UUUGUCUGUGUCAU		
	B	CAAUGAUGUUU <u>ccccca</u> UUUGUCUGUGUCAU		
	C	CAAUGAUGUUU <u>ccacca</u> UUUGUCUGUGUCAU		
	D	CAAUGAUGUUU <u>caacca</u> UUUGUCUGUGUCAU		
+5/+10	A	CAAUGA <u>caaac</u> CAAAUAUUUGUCUGUGUCAU		
	B	CAAUGA <u>ccaac</u> CAAAUAUUUGUCUGUGUCAU		
	C	CAAUGA <u>caaaa</u> CAAAUAUUUGUCUGUGUCAU		
	D	CAAUGA <u>cccaac</u> CAAAUAUUUGUCUGUGUCAU		
+23/+25	A	CAAUGAUGUUUAAAUAUUUGUC <u>cc</u> GUCAU		
+20/+22	A	CAAUGAUGUUUAAAUAUUU <u>ac</u> UGUGUCAU		
	B	CAAUGAUGUUUAAAUAUUU <u>cc</u> UGUGUCAU		
+17/+19	A	CAAUGAUGUUUAAAUA <u>cca</u> GUCUGUGUCAU		
+14/+16	A	CAAUGAUGUUUAAA <u>ccc</u> UUUGUCUGUGUCAU		
	B	CAAUGAUGUUUAAA <u>cac</u> UUUGUCUGUGUCAU		

Figure 3. Cluster mutations of targeted regions of the mouse β -globin poly(A) signal: Different mutations for each targeted region are labeled A–D. The nucleotide sequence of each cluster of mutations is highlighted by underlined lowercase letters.

Figure 4. Representative assays of mouse β -globin poly(A) signal function in the 5–6 base cluster mutations: A. RNase protection assays: Lane 1: markers (end-labeled MspI-digested pBR322 DNA). Lanes 2–10: Approximately 10 μ g of total RNA from COS cells transfected with +23/+27 C (lane 2), +23/+27 D (lane 3), +17/+22 C (lane 4), +17/+22 D (lane 5), +11/+16 C (lane 6), +11/+16 D (lane 7), +5/+10 C (lane 8), +5/+10 D (lane 9) and wild type construct, pMBG1kpnΔ+27 (lane 10) were analyzed with the pMOB2 probe. Lanes 11–14: Approximately 10 μ g of total RNA from COS cells transfected with +10/+5 A (lane 11), B (lane 12), C (lane 13) and D (lane 14) were analyzed with homologous probes prepared from portions of each mutant subcloned into pGEM3Z. RNAs cleaved, or not cleaved at the MBG poly(A) site, yield bands of 130 nt, or 200 nt, respectively. Bands slightly above or below the 130 nt 'MBG pA' band (lanes 13 and 14) implicate the presence of additional cleavage sites. B. Summary of the functional properties of the 5–6 base cluster mutations: Percentage of cleavage at the MBG pA site was calculated as described in Fig. 2B. Bars indicate the standard deviation from 2–3 experiments.

Figure 5. Summary of the functional properties of the 3 base cluster mutations: RNA from COS cells transfected with each of the 3-base cluster mutations described in Fig. 3 were analyzed with the pMOB2 probe. Percentage of cleavage at the MBG pA site was calculated as described in Fig. 2B.

+5/+10 cluster mutations (Fig. 4A, lanes 11–14). Quantitative data for all of the cluster mutants are summarized in Fig. 4B. The largest decline in globin poly(A) signal function occurred in the +17/+22 cluster mutants, where approximately 50% inhibition was observed. This result is completely consistent with the data from the Bal31 exonuclease deletion experiments, and confirms the +17/+22 region as the most critical component of the downstream element. Cluster mutations in the adjacent regions, however, exhibited relatively minor effects: the +11/+16 cluster was inhibited by approximately 7%, and the +23/+27 was inhibited by approximately 10%. The small contribution from the +11/+16 region is consistent with the Bal31 exonuclease deletion data, and the small effect of the mutations in the +23/+27 region confirms that the 3' boundary of the downstream element is best defined by nucleotide +22. Cluster mutations in the +5/+10 region exhibited a decline in poly(A) signal function of approximately 27%. This decline is also consistent with the Bal31 exonuclease deletion data, and demonstrates that the downstream element has two major components: one from +5 to +10 and the other from +17 to +22. Of the two, the +17/+22 component is functionally more important. None of the cluster mutations caused noticeable shifts in the cleavage site, except for two of the +5/+10 clusters, where minor bands with shifted mobilities were observed (Fig. 4A, lanes 13 and 14). Thus, the bipartite downstream element primarily serves to influence poly(A) signal efficiency.

The downstream element is functionally redundant

In an attempt to define the sequence requirements of the +17/+22 component more precisely, various 3-base portions of the region between +14 and +25 were replaced with random CA-containing sequences. The nucleotide sequences of these 3-base cluster mutations are shown in the bottom portion of Fig. 3, and the quantitative data are summarized in Fig. 5. None of the 3-base cluster mutations inhibited poly(A) signal function by more than 10%, including the ones from +17 to +19, and +20 to +22. Thus, neither half of the +17 to +22 region is obligatory. One possibility is that either half of the 6-base sequence between +17 and +22 is sufficient. Another possibility is that adjacent sequences, perhaps the U-rich sequences between +14 and +16 or between +22 and +25, can compensate for the half of the +17/+22 region that was mutated. In either case, the data demonstrate the existence of sequences that are functionally redundant.

Figure 6. Structural and functional components of the mouse β^{maj} -globin poly(A) signal: The hexanucleotide (Hex), cleavage/polyadenylation site (pA), and the two downstream element components (I and II) that are major contributors to poly(A) signal function are double-underlined.

DISCUSSION

The locations of the functional components of the mouse β^{maj} -globin poly(A) signal are summarized in Fig. 6. The 3' boundary of the sequences required for an effective poly(A) signal is +22. Because sequences between +23 and +27 contribute slightly, the 3' boundary for a maximally effective poly(A) signal is +27. No evidence for the existence of functional elements located more distal than +27 was detected. Major components of the downstream element are located between +5 and +10 and between +17 and +22; the one located between +17 and +22 appears to be the most important. The sequences in these two regions consist almost entirely of U's and G's; it is therefore appropriate to describe both as UG-rich components. None of the mutations caused a major shift in cleavage site, including the -1 deletion of pMBG1, in which the entire downstream region was replaced. Thus, the downstream element primarily affects cleavage efficiency, and appears to have little influence on cleavage site determination.

Since the positions of the 5–6 bp cluster mutations were chosen rather arbitrarily, the boundaries and critical portions of the two components cannot yet be precisely described. Our attempt to achieve further definition of the +17 to +22 component with various 3 nt cluster mutations failed, since none caused notable loss of function. These data clearly show that neither half of the +17 to +22 region is essential. Yet, mutation of both halves caused a large decrease in poly(A) signal function. Thus, redundancies must exist. The sequences immediately adjacent to the +17 to +22 region are U- or UG-rich. Although these sequences also are not obligatory, they may serve a compensatory role when portions, but not all, of the +17 to +22 region is mutated. It appears, therefore, that the sequence and location requirements for this downstream component are somewhat flexible.

The downstream region of the rabbit β -globin poly(A) signal has a structure that resembles that of the mouse β -globin signal. Two segments of close sequence homology are evident: the +5 to +10 sequence of the rabbit and mouse signals match in 5 of 6 positions, and the +18 to +24 region of the rabbit signal and the +17 to +23 region of the mouse signal match in 6 of 7 positions. Other segments of the downstream region are more variable in sequence. Thus, the positions of the most highly conserved sequences in the downstream regions of these related poly(A) signals closely correlate with the positions of the two components delineated in our study. Furthermore, the rabbit β -globin positions are located within segments that have been shown by mutagenesis and oligonucleotide replacement to contain functionally important sequences [13,41,42]. Although the boundaries of the rabbit sequences were defined at lower resolution, the combined data clearly suggest that the important

components in the downstream regions of the mouse and rabbit β -globin poly(A) signals are nearly identical in sequence and location.

Other poly(A) signals also have downstream elements that appear to consist of two separated UG-rich components. A biphasic pattern of loss of function was observed when closely spaced 3' deletions of the SV40 late [14,21] or HSV thymidine kinase signals [7,15] were analyzed, and the locations of the deletion endpoints that caused transitions in the activity of these signals are consistent with UG-rich sequences, located at approximately +5/+10 and +17/+22, as the major functional components. Inspection of the downstream sequences of other poly(A) signals shows that UG-rich sequences are often localized in or near the +5/+10 and +17/+22 regions [Refer to reference 11 for a compendium of poly(A) signal sequences]. Thus, bipartite downstream elements of this type could be a common feature of poly(A) signal structure. However, a better understanding of the sequence and spatial requirements of the downstream components is needed before generalizations can be made. This is highlighted by the finding that important downstream sequences have been localized to more distal sites for at least two poly(A) signals [16–20].

UG-rich downstream elements are required for poly(A) signal function in unfractionated nuclear extracts [42–44], and contribute to the formation of pre-cleavage complexes [43,45]. The two UG-rich downstream components identified by our analysis may represent, therefore, binding sites for factors involved in poly(A) signal recognition and cleavage. The probable nature of the proteins that interact with these downstream components is uncertain, however. Nevins and coworkers report that a downstream element is required for the cleavage of RNA substrates by purified factors [46], for polyadenylation complex stability [47], and for the crosslinking of a 64 kDa polypeptide [48]; these data suggest that downstream elements interact directly with cleavage factors. By contrast, Manley and coworkers [49,50] report that cleavage factors lose much of their dependence on downstream sequences as they are further fractionated, and Wilusz and coworkers [19,51] have identified stretches of contiguous U's that crosslink to core ribonucleoprotein C and are capable of serving as functional downstream elements. These observations suggest an alternative viewpoint that downstream elements interact, not with cleavage factors per se, but with more general factors that influence poly(A) signal accessibility [50,51]. Finally, based on the recent discovery of a protein (DSEF-1) that binds to a G-rich sequence found so far only in the downstream region of the SV40 late poly(A) signal [20], the contributions of proteins that interact with selected poly(A) signals need to be considered. Elimination of sequence redundancies and identification of the critical bases that comprise poly(A) signal downstream elements should aid in the differentiation of these possibilities.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Deborah Harvazinski (while in the Department of Biochemistry and Biophysics at Texas A&M University) for constructing the pMBG1 series of constructs, including the Bal31 deletions, and Michael Bordonaro for constructing pMOB2. We thank Kuo Chuang for his technical assistance, Lida Burney for her photographic skills, and Colette

Saccomanno for her critical reading of the manuscript. This work was supported by Grant GM33410 from the National Institutes of Health.

REFERENCES

- Manley, J. L. (1988) *Biochim. Biophys. Acta*, **950**, 1–12.
- Wickens, M. (1990) *Trends Biochem. Sci.*, **15**, 277–281.
- Proudfoot, N. (1991) *Cell*, **64**, 671–674.
- Sheets, M., Ogg, S. C. and Wickens, M. (1990) *Nucleic Acids Res.*, **18**, 5799–5805.
- Moore, C. L., Skolnik-David, H. and Sharp, P. A. (1986) *EMBO J.*, **5**, 1929–1938.
- Sheets, M. D., Stephenson, P. and Wickens, M. (1987) *Mol. Cell. Biol.*, **7**, 1518–1529.
- Zhang, F., Denome, R. M. and Cole, C. N. (1986) *Mol. Cell. Biol.*, **6**, 4611–4623.
- Woychik, R. P., Lyons, R. H., Post, L. and Rottman, F. M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3944–3948.
- Mason, P. J., Elkington, J. A., Lloyd, M. M., Jones, M. B. and Williams, J. G. (1986) *Cell*, **46**, 263–270.
- Taya, Y., Devos, R., Tavernier, J., Cheroutre, M. H. and Berk, A. J. (1982) *EMBO J.*, **1**, 953–958.
- McLaughlin, J., Gaffney, D., Whitton, J. L. and Clements, J. B. (1985) *Nucleic Acids Res.*, **13**, 1347–1368.
- Renan, M. J. (1987) *Gene*, **60**, 245–254.
- Gil, A. and Proudfoot, N. J. (1987) *Cell*, **49**, 399–406.
- Conway, L. and Wickens, M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3949–3953.
- Cole, C. N. and Stacy, T. P. (1985) *Mol. Cell. Biol.*, **5**, 2104–2113.
- Hart, R. P., McDevitt, M. A., Ali, H. and Nevins, J. R. (1985) *Mol. Cell. Biol.*, **5**, 2975–2983.
- Sadofsky, M. and Alwine, J. C. (1984) *Mol. Cell. Biol.*, **4**, 1460–1468.
- Sadofsky, M., Connelly, S., Manley, J. L. and Alwine, J. C. (1985) *Mol. Cell. Biol.*, **5**, 2713–2719.
- Wilusz, J. and Shenk, T. (1990) *Mol. Cell. Biol.*, **10**, 6397–6407.
- Qian, Z. and Wilusz, J. (1991) *Mol. Cell. Biol.*, **11**, 5312–5320.
- Zarkower, D. and Wickens, M. (1988) *J. Biol. Chem.*, **263**, 5780–5788.
- McDevitt, M. A., Hart, R. P., Wong, W. W. and Nevins, J. R. (1986) *EMBO J.*, **5**, 2907–2913.
- Carswell, S. and Alwine, J. C. (1989) *Mol. Cell. Biol.*, **9**, 4248–4258.
- DeZazzo, J. D. and Imperiale, M. J. (1989) *Mol. Cell. Biol.*, **9**, 4951–4961.
- Russnak, R. and Ganem, D. (1990) *Genes Dev.*, **4**, 764–776.
- Russnak, R. (1991) *Nucl. Acids Res.*, **19**, 6449–6456.
- Tilghman, S. M., Tiemeier, D. C., Polsky, F., Edgell, M. H., Seidman, J. G., Leder, A., Enquist, L. W., Norman, B. and Leder, P. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 4406–4410.
- Messing, J. and Vieira, J. (1982) *Gene*, **19**, 269–276.
- Southern, P. J. and Berg, P. (1982) *J. Mol. Appl. Genet.*, **1**, 327–341.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Innis, M. A. and Gelfand, D. H. (1990) In Innis, M. A., Gelfand, D. H., Sninsky, J. J. and White, T. J. (eds.), *PCR Protocols: A Guide to Methods and Applications*. Academic Press, NY, pp. 3–12.
- Gluzman, Y. (1981) *Cell*, **23**, 175–182.
- Nordstrom, J. L., Hall, S. L. and Kessler, M. M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1094–1098.
- Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.*, **162**, 156–159.
- Williams, C. A. and Krawisz, B. R. (1989) *Meth. Molec. Cell. Biol.*, **1**, 35–40.
- Gopalakrishnan, T. V. and Anderson, W. F. (1979) *Methods Enzymol.*, **58**, 506–511.
- Swilens, S., Cochaux, P. and Lecocq, R. (1989) *Trends Biochem. Sci.*, **14**, 440–441.
- Mellon, P., Parker, V., Gluzman, Y. and Maniatis, T. (1981) *Cell*, **27**, 279–288.
- Stefano, J. E. and Adams, D. E. (1988) *Mol. Cell. Biol.*, **8**, 2052–2062.
- Gil, A. and Proudfoot, N. J. (1984) *Nature (London)*, **312**, 473–474.
- Levitt, N., Briggs, D., Gil, A. and Proudfoot, N. J. (1989) *Genes Dev.*, **3**, 1019–1025.
- Skolnik-David, H., Moore, C. L. and Sharp, P. A. (1987) *Genes Dev.*, **1**, 672–682.

43. Zarkower, D. and Wickens, M. (1988). *J. Biol. Chem.*, **263**, 5780–5788.
44. Zhang, F. and Cole, C. N. (1987). *Mol. Cell. Biol.*, **7**, 3277–3286.
45. Moore, C. L., Skolnik-David, H. and Sharp, P. A. (1988). *EMBO J.*, **5**, 1929–1938.
46. Gilmartin, G. M. and Nevins, J. R. (1989). *Genes Dev.*, **3**, 2180–2189.
47. Weiss, E. A., Gilmartin, G. M. and Nevins, J. R. (1991). *EMBO J.*, **10**, 215–219.
48. Gilmartin, G. M. and Nevins, J. R. (1991). *Mol. Cell. Biol.*, **11**, 2432–2438.
49. Takagaki, Y., Ryner, L. C. and Manley, J. L. (1989). *Genes Dev.*, **3**, 1711–1724.
50. Ryner, L. C., Takagaki, Y. and Manley, J. L. (1989). *Mol. Cell. Biol.*, **9**, 1759–1771.
51. Wiluz, J., Feig, D. I. and Shenk, T. (1988). *Mol. Cell. Biol.*, **8**, 4477–4483.
52. Rougeon, F. and Mach, B. (1977) *Gene*, **1**, 229–239.