MOLECULAR AND CELLULAR BIOLOGY, June 1993, p. 3472–3480 Vol. 13, No. 6
0270-7306/93/063472-09$02.00/0

3′ RNA Processing Efficiency Plays a Primary Role in Generating Termination-Competent RNA Polymerase II Elongation Complexes

GRETCHEN EDWALDS-GILBERT, JOHN PRESCOTT, AND ERIK FALCK-PEDERSEN*

Department of Microbiology, W. R. Hearst Research Foundation, Cornell University Medical College, New York, New York 10021

Received 25 January 1993/Returned for modification 17 March 1993/Accepted 29 March 1993

In several mammalian transcription units, a transcription termination mechanism in which efficient termination is dependent on the presence of an intact 3′ RNA processing site has been identified. The mouse β^{maj}- globin transcription unit is one such example, in which an intact poly(A) site is required for efficient transcription termination. It is now evident that 3′ mRNA processing sites are not always processed with the same efficiency. In this study, we characterized several pre-mRNAs as substrates for the 3′ mRNA processing reaction of cleavage and polyadenylation. We then determined whether poly(A) sites which vary in processing efficiency support a poly(A) site-dependent termination event. The level of processing efficiency was determined in vitro by assays measuring the efficiency of the pre-mRNA cleavage event and in vivo by the level of poly(A) site-dependent mRNA and gene product expression generated in transient transfection assays. The β^{maj} globin pre-mRNA is very efficiently processed. This efficient processing correlates with its function in termination assays using recombinant adenovirus termination vectors in nuclear run-on assays. When the β^{maj} globin poly(A) site was replaced by the L1 poly(A) site of the adenovirus major late transcription unit (Ad-ml), which is a poor processing substrate, termination efficiency decreased dramatically. When the β^{maj} globin poly(A) site was replaced by the Ad-ml L3 poly(A) site, which is 10- to 20-fold more efficiently processed than the Ad-ml L1 poly(A) site, termination efficiency remained high. Termination is therefore dependent on the yield of the processing event. We then tested chimeric poly(A) sites containing the L3 core AAUAAA but varied downstream GU-rich elements. The change in downstream GU-rich elements affected processing efficiency in a manner which correlated with termination efficiency. These experiments provide evidence that the efficiency of 3′ processing complex formation is directly correlated to the efficiency of RNA polymerase II termination at the 3′ end of a mammalian transcription unit.

In a number of transcription units, mutations in required 3′ RNA processing signals, such as the consensus AAUAAA or the downstream GU- or U-rich region, which disrupt pre-mRNA cleavage-polyadenylation, also result in loss of RNA polymerase II (polII) termination (7, 20, 35). These findings define a unique relationship between the 3′ RNA processing site or event and the elongating transcription complex. Only after transcription through the poly(A) signal sequence, in a transcription unit such as mouse β^{maj} globin, does the transcription complex become competent for termination. Very little is known about how the presence of the 3′ processing site effects termination. The 3′ site may interact directly with polII at either the DNA or the RNA level to generate the termination-competent elongation complex. Alternatively, the 3′ site may act on the elongation complex indirectly through the processing complex or the RNA cleavage event.

Regardless of how the termination-competent complex is generated, in several instances it has been shown that the transcription complex continues elongation for considerable distances past the 3′ processing site before elongation arrest occurs (6, 15, 19, 22, 28, 30). The termination-competent polII elongation complex undergoes elongation arrest when it addresses specific termination signal elements. Several types of elements capable of causing transcription termination have been identified. DNA-bound protein complexes such as those present in specific promoter elements have been shown to cause poly(A) site-dependent termination (1, 4, 8, 9). In addition, there are several DNA sequence elements which are able to induce arrest of polII elongation without a known protein-DNA interaction (11, 31, 32). It is not known whether these elements interact with polII through a specific primary sequence and/or a secondary structure(s) residing in either the DNA or the nascent RNA.

Understanding how the 3′ processing site influences the transcription complex is an important step in understanding polII termination. On the basis of work done in several laboratories, we have a good understanding of the biochemistry of 3′ processing (reviewed by Manley [24] and more recently by Wahle and Keller [33]). Processing at a poly(A) site depends on formation of a preprocessing complex with as many as nine proteins binding to the pre-mRNA. Within the pre-mRNA, there are two specific sequence elements which are minimally required for cleavage complex formation, the consensus AAUAAA located 15 to 25 nucleotides 5′ to the cleavage site and a downstream GU- and/or U-rich element. The first step in complex formation is binding of cleavage and polyadenylation specificity factor (CPSF) to the AAUAAA consensus sequence present in the pre-mRNA (2, 13, 17). The CPSF–pre-mRNA complex is relatively unstable and is susceptible to disruption by competitor pre-mRNA. Binding of cleavage stimulatory factor (CStF) to the CPSF–pre-mRNA complex has been shown to stabilize the preprocessing complex intermediate and is dependent on the presence of the GU- or U-rich element (13, 25a, 34). Formation of the stabilized preprocessing complex is a good candidate for a rate-limiting step in 3′-end processing. Following stable complex formation, an additional cleavage factor and, in most cases, poly(A) polymerase are required

* Corresponding author.

for cleavage of the pre-mRNA. A poly(A) tract of approximately 200 nucleotides is added to the 5′ cleavage product by poly(A) polymerase in a step which is tightly coupled to the cleavage reaction.

In this investigation, we studied 3′ RNA processing sites and their involvement in the transcription termination event. We used 3′ termination elements from the mouse β^{maj} globin transcription unit which have been shown to function as an efficient termination region in an adenovirus (Ad) reporter vector (12, 20, 32). The mouse β^{maj} globin 3′ termination region is a relatively well-defined complex of polII transcription termination elements. It consists of the β^{maj} globin poly(A) site, which is required for transcription termination, and a termination region, resident in a genomic subclone, globin F (gF). The 800-bp gF termination region is found approximately 500 bp 3′ to the globin poly(A) cleavage site. It consists of multiple sequence elements, each of which has the ability to block polII elongation. When present as a tandem array, these elements can arrest greater than 90% of elongating transcription complexes. We have recently demonstrated that recognition of the termination elements by polII requires the 500-bp spacing sequence between the poly(A) site and the termination elements (32).

We found that pre-mRNAs containing the globin β^{maj} poly(A) site are efficiently used as substrates for the cleavage and polyadenylation reactions both in vivo and in vitro. This efficient use of the globin poly(A) site is essential to the termination reaction. When the globin poly(A) site was replaced with a weak poly(A) substrate site, transcription termination failed to occur. Our interpretation of these findings is that there is a direct correlation between formation of a stable RNA cleavage complex and transcription termination. The RNA processing event therefore dictates formation of the termination-competent elongation complex.

MATERIALS AND METHODS

Cell and tissue culture. Cells of human embryonic kidney line 293 (14), which constitutively express the Ad E1a and E1b genes, were grown in Dulbecco's modified Eagle medium and 5% calf serum. These cells were used for cotransfections, preparation of initial virus lysates, and plaque purifications. Large-scale virus propagation was performed with suspension 293 cells grown in Joklik's minimal essential medium and 5% horse serum. HeLa spinner cells maintained in Joklik's minimal essential medium and 5% horse serum were used for nuclear run-on analyses. HeLa monolayer cells maintained in Dulbecco's modified Eagle medium and 5% calf serum were used in transient transfection assays.

Plasmid constructs. The plasmids used for in vitro synthesis of precursor mRNA were pGem2-L1-101 and pGem2-L3-114, which have been previously described (27). The β^{maj} globin pre-mRNA was transcribed from plasmid pBSKgD, which was generated by subcloning the gD element from pmlp6gD (20) by *Nar*I-*Xba*I directional cloning into a pBSK vector. The chimeric pre-mRNAs were generated from plasmids pGEM2-L3 μs, pGEM2-L3μm, pGEM2-L3SV_L, and pGEM2-L3mSV_L, which were a gift from J. Nevins (34). They contain 44 bp of the core L3 sequence, followed by 30 nucleotides of the downstream GT sequence present in each of the indicated poly(A) sites. The downstream GT regions were synthesized as oligonucleotides and cloned into a pGEM2-L3 parent plasmid.

Plasmid pMLSISCAT-pA was generated by deleting the existing pMLSISCAT (16) poly(A) site by *Kpn*I-*Nar*I digestion; overhanging ends were removed by T4 DNA polymerase, and the linear plasmid was circularized with T4 ligase. pMLSISCAT-poly(A) site plasmids were created by excising the poly(A)-containing DNA fragments from the above-described pGEM2 plasmids and inserting the isolated DNA fragments into plasmid pMLSISCAT-PA at the unique *Bam*HI site by blunt-end cloning. All constructs had the orientations of their inserts verified by dideoxy-DNA sequencing.

In a manner similar to that described above, each of the poly(A) sites was inserted into plasmid vector pMLP6-gEF (20), which contains the β^{maj} globin downstream termination sequence inserted into Ad E1a exon 2. For each construct, the pMLP6-gEF vector was linearized with *Bam*HI at the 5′ junction between the E1a and gE sequences. Each poly(A) site was inserted by blunt-end ligation essentially as previously described. The final plasmids constructed in this manner were pMLP6-L1EF, pMLP6-L3EF, pMLP6-L3μmEF, pMLP6-L3SV_LEF, pMLP6-L3mSV_LEF, and pMLP6-L3μsEF.

In vitro processing reactions. Substrate pre-mRNA was synthesized as previously described (27). Briefly, each linearized DNA template was transcribed by the appropriate RNA polymerase in reactions that contained 100 μM CTP, GTP, and UTP; 50 μCi of UTP at 3,000 Ci/mmol; and 0.5 mM 7meG(5′)ppp(5)′G. HeLa cell nuclear extracts were prepared as described by Gilmartin and Nevins (13). Final extracts contained 7 to 11 mg of protein per ml and were stored in aliquots at −70°C. Processing reactions included 1 nM labeled substrate pre-mRNA, 0.5 mM 3′ dATP, 40% nuclear extract, 5 mM creatine phosphate, 2.5% polyvinyl alcohol, and 0.7 mM $MgCl_2$ in a volume of 25 μl. Processing reactions were incubated at 30°C for 30 min and stopped by addition of guanidium thiocyanate phenol. RNA was purified by the method of Chomczynski and Sacchi (5).

Transient transfection, chloramphenicol acetyltransferase (CAT), and luciferase assays. HeLa cells were plated onto 35-mm-diameter six-well plates such that 36 h later they would be 70% confluent (approximately 10^6 cells per well). A $CaPO_4$ DNA precipitate prepared by standard protocols (23) with 1.0 μg of plasmid pMLSISCAT-poly(A), 0.5 μg of pRSVLUC (3), and 3.5 μg of carrier plasmid pGEM2 in a volume of 0.5 ml was added to individual wells. Transfections were always done in duplicate. Cells were harvested 36 h later in 1 ml of scrape buffer (40 mM Tris-HCl [pH 7.4], 1 mM EDTA, 150 mM NaCl). The cell suspension was freeze-thawed three times, cellular debris was removed by 5 min of microcentrifuge centrifugation, and the supernatant was transferred to a new Eppendorf tube. Luciferase assays (3, 10) were done in duplicate with standard assay conditions; 25 μl of extract was added to 435 μl of assay buffer [25 mM glycylglycine, 17 mM $MgSO_4$, 4 mM ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid (EGTA), 1 mM dithiothreitol, 10 mM KPO_4, 2 mM ATP]; assays were initiated by addition of 120 μl of luciferin cocktail (25 mM glycylglycine, 17 mM $MgSO_4$, 4 mM EGTA, 1 mM dithiothreitol, 40 μM luciferin) in an ALL Monolight 2010 luminometer. After calculation of the number of luciferase units per microliter of extract, proportional volumes of extract were used in duplicate CAT assays to determine CAT expression levels as a function of poly(A) site efficiency. CAT assays were done as described by Huang and Gorman (16). Conversion of chloramphenicol to the acetylated form was quantitated with a Molecular Dynamics PhosphorImager scan of reaction products separated by migration on silica gel thin-layer chromatography plates (95:5 $CHCl_3$-MeOH).

RNA characterization by S1 endonuclease hybrid protection. For isolation of RNA, transfections were proportionately scaled up to 100-mm dishes. Cells were harvested exactly as described for the CAT and luciferase protocols; 25% of each transfection was used for luciferase-CAT extract preparation, and RNA was isolated from the remainder. RNA was isolated by the method of Chomczynski and Sacchi as previously described (5). Luciferase and CAT assays were done as described above. CAT RNA analysis was done after normalization with luciferase. Overnight hybridizations [in 80% formamide–300 mM NaCl–40 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES) (pH 6.5)–10 mM EDTA] were done at 55°C. DNA probes were generated from *Eco*RI-digested pMLSISCAT-poly(A) plasmid DNAs. The 3′ recessed ends generated by endonuclease digestion were than elongated with the DNA polI large fragment and α-^{32}P-labeled dATP and dTTP.

Construction of recombinant Ads. Recombinant Ads were constructed by the overlap recombination method as previously described (21). Parent plasmids were linearized and cotransfected into 293 cells with the *dl*309 large fragment and the $CaPO_4$ technique. Individual plaque isolates were used to infect 293 monolayers, and the identity of each recombinant virus was confirmed by analyzing viral DNA isolated from infected 293 cells.

Viruses are named after the parent virus, sub360-EF, and the insertions each virus contains, i.e., sub360-L3EF, sub360-L1EF, sub360-L3μmEF, sub360-L3μsEF, sub360-L3SV$_L$EF, sub360-L3mSV$_L$EF.

Preparation of nitrocellulose filters for slot blot hybridization. A Schleicher & Schuell slot blot manifold was used to prepare all slot blot grids. Five micrograms of double-stranded DNA was loaded for each slot. Each DNA sample was denatured by treating the plasmid with 0.1 N NaOH and boiling the DNA. The DNA samples were neutralized with 20 μl of 2 M *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES) and loaded onto a nitrocellulose filter in the presence of 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). DNAs were affixed to the nitrocellulose filter by baking the filter at 80°C for 2 h under vacuum.

The following double-stranded DNA clones were used in the termination assays and have been described previously (12): 1A1 (Ad type 2 sequence from 1 to 1010), 1A2 (1010 to 1336), 1A3 (1336 to 1671), 1B5′ (1699 to 2153), 1B3′ (3577 to 4030), mouse β^{maj} globin major segments gF$_A$ (contains nucleotides 1 to 404 from the gF region) and gF$_B$ (contains nucleotides 417 to 809 from the gF region). E2 contains the Ad type 2 sequence from 22307 to 228576, and actin is a human actin cDNA cloned into BlueScript (a gift from D. Catanzaro). pBSTSK− or pGEMI was used as a negative control.

Nuclear run-on assays and hybridization of nascently labeled RNA to a nitrocellulose filter. Nuclear run-on assays were done as described by Tantravahi et al. (32). Radiolabeled nascent RNA was fragmented by treatment with 0.2 N NaOH at 0°C for 10 min. Fragmented RNA was brought to 2 ml by addition of HYB buffer (1× Denhardt's solution, 100 μg of yeast RNA per ml, 100 μg of herring sperm DNA per ml in 2× TESS [2× TESS is 20 mM *N*-tris(hydroxymethyl) methyl-2-aminoethanesulfonic acid, 0.3 M NaCl, 10 mM EDTA (pH 8.0), and 0.2% sodium dodecyl sulfate (SDS)]. Nitrocellulose filters were prehybridized with 2.5 ml of HYB buffer for 2 h at 65°C before addition of the labeled RNA. RNA was hybridized to the filters for 36 to 40 h at 65°C. Following hybridization, the filters were washed successively with 2× SSC–0.5% SDS for 30 min at room temperature, 2× SSC–0.1% SDS for 45 min at 65°C, 2× SSC only for 15 min at 65°C, 2× SSC–5 μg of pancreatic RNase per ml for 30 min at room temperature, 2× SSC only for 15 min at 65°C, and 2× SSC only for 15 min at room temperature. The filters were dried under a heating lamp and autoradiographed. Termination efficiencies were quantified with a Molecular Dynamics PhosphorImager.

RESULTS

Characterization of the mouse β^{maj} globin 3′ RNA processing site. It has been taken for granted that all transcripts which contain 3′ processing signals are efficiently processed and converted into mature mRNA products. With recent in vitro and in vivo characterizations of the 3′ processing event, it has become clear that conversion of pre-mRNA substrates into the final mRNA product is dependent on many factors and is not necessarily (or possibly ever) 100% efficient. Because the mouse β^{maj} globin poly(A) site is required for efficient transcription termination, characterization of the β^{maj} globin pre-mRNA as a substrate for the 3′ processing event is an important step in defining the relationship between the 3′ processing event and transcription termination. To demonstrate its function as a substrate for the processing reaction, we compared the β^{maj} globin poly(A) site to well-characterized poly(A) sites L1 and L3 from the Ad major late transcription unit. On the basis of in vitro RNA processing assays, the L1 site has previously been defined as a low-efficiency processing substrate, whereas the L3 site functions as a comparatively high-efficiency substrate (27). These studies have also shown that the difference in processing efficiency can be attributed to differences in the stability of preprocessing complex formation.

The ability to carry out the 3′ processing reaction in vitro gives us a simple and direct means to determine the relative strength of the β^{maj} globin 3′ processing site. The pre-mRNA substrates (core sequences shown in Fig. 1A) are transcribed in vitro by T7 RNA polymerase in the presence of a 7-meG(5′)ppp(5′)G cap analog to produce transcripts which contain the 5′ cap structure characteristic of polII mRNAs. Each pre-mRNA was incubated for 30 min in a 40% HeLa Dignam nuclear extract mixture (as indicated in Materials and Methods). In these assays, we added the ATP analog cordycepin triphosphate, which acts as a chain terminator of poly(A) tail addition by poly(A) polymerase. Use of the poly(A) chain terminator allows RNA cleavage products to be identified as discrete species when they are electrophoresed on a 4.5% bipolyacrylamide–urea gel. After autoradiography, the processing efficiency for each substrate was determined by calculating the percentage of the cleavage product as a function of the input RNA. Past assays have indicated that 30 to 60% of L3 pre-mRNAs are processed, whereas 1 to 10% of L1 pre-mRNAs undergo cleavage in this assay. By using approximately 25 fmol each of the L1, β^{maj} globin, and L3 pre-mRNAs in 25-μl reaction mixtures (Fig. 1B), we determined how efficiently each pre-mRNA supports formation of the processing cleavage complex. We found that the β^{maj} globin pre-mRNA functioned as a high-efficiency processing substrate (69% cleavage) and was more efficient than the L3 pre-mRNA (37% cleavage) in this in vitro processing assay (Fig. 1B; compare lanes 4 and 5). As expected, the L1 pre-mRNA (2.7% cleavage) was not processed efficiently. This pattern of processing efficiency was also obtained when different nuclear extract preparations were used. While there were variations in the amount of

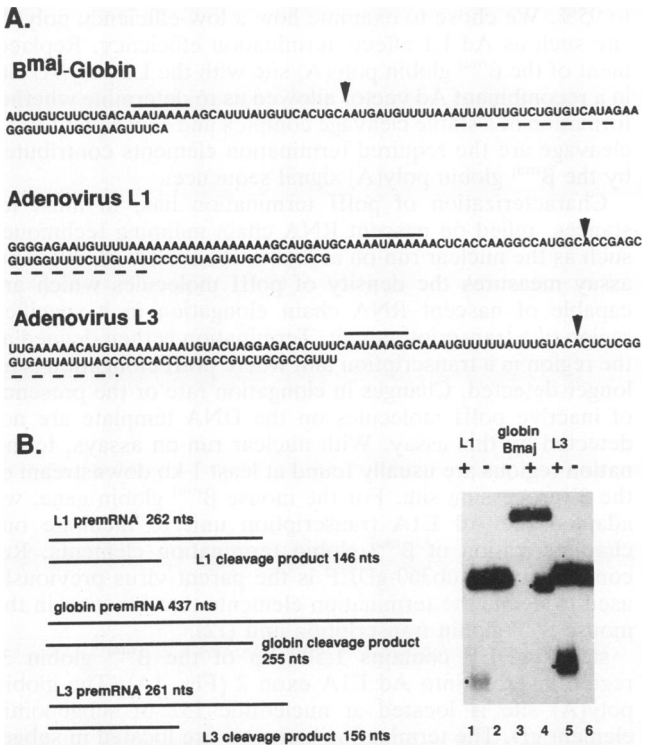

FIG. 1. Processing efficiency of β^{maj} globin pre-mRNA in vitro. (A) Ribonucleotide sequences surrounding the poly(A) cleavage sites present in mouse β^{maj} globin and Ad L1 and L3 pre-mRNAs. A solid line is positioned over the required consensus sequence (AAUAAA), the arrowheads show the sites of actual cleavage, and the dashed lines indicate the locations of the GU/U-rich elements also required for efficient processing. (B) The lengths of each pre-mRNA and the corresponding 5′ cleavage product are indicated on the left. Lanes 1, 4, and 5 contained products isolated from cleavage reactions (+) for the L1, β^{maj} globin, and L3 pre-mRNAs following gel electrophoresis as described in Materials and Methods. Lanes 2, 3, and 6 contained the unprocessed (−) pre-mRNA substrate. The cleavage products are indicated by dots.

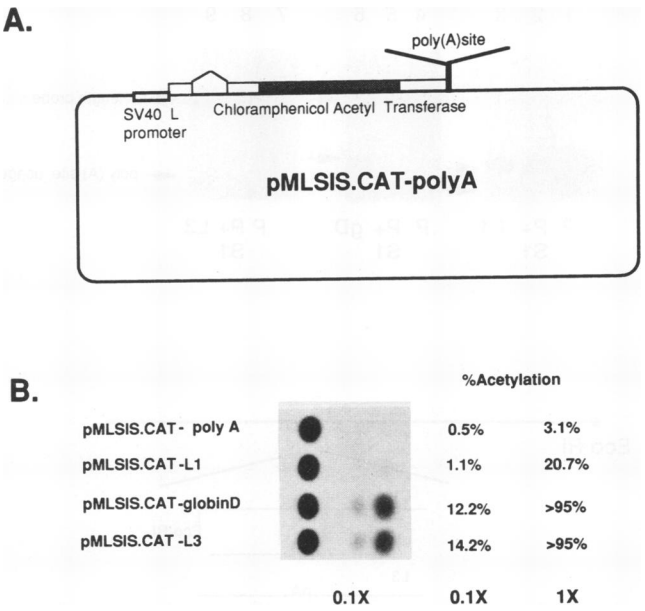

FIG. 2. Functional characterization of poly(A) site processing in vivo. (A) Schematic of the functional elements in the core pMLSIS-CAT plasmid and the locations of inserted poly(A) sites. SV40, simian virus 40. (B) Following plasmid $CaPO_4$ transfection into HeLa cells, cell extracts generated as described in Materials and Methods were normalized to each other by using luciferase expression from the cotransfection control plasmid. A 1× normalized volume of extract from each pMLSISCAT plasmid transfection used in a CAT assay gave the indicated amount of substrate acetylation. The 0.1× value represents the amount of chloramphenicol which was acetylated with 10% of the normalized volume. Product migration obtained with silica gel thin-layer chromatography is shown for each extract at the 0.1× volume of normalized extract.

general cleavage activity present in different extract preparations, the product ratio of L1 to L3 to β^{maj} globin (gD) was constant. Keeping the limitations of the nuclear extract in vitro assay in mind, we feel that the data presented in Fig. 1B give us an initial indication that the globin β^{maj} poly(A) site functions as a comparatively high-efficiency substrate for the 3′ RNA processing reaction.

Since there are a limited number of examples in which poly(A) site strength has been characterized, it is important to demonstrate that in vitro measurement of poly(A) site strength be supported by an independent in vivo measurement of poly(A) site strength. For our ongoing studies of the Ad L1 and L3 poly(A) sites (26, 27), a transient transfection assay was designed in which expression of the gene product CAT was dependent on efficiency of poly(A) site utilization. Plasmid vector pMLSISCAT-PA contains all of the elements required for CAT mRNA expression when transfected into HeLa cells, except for a functional poly(A) signal sequence. A CAT assay done after transfection of plasmids pMLSIS-CAT-PA, pMLSISCAT-L1, pMLSISCAT-globinD, and pMLSISCAT-L3, containing the L1, gD, and L3 poly(A) signal elements, respectively (Fig. 2A), reflects CAT levels proportional to the efficiency of poly(A) site utilization.

Results of such an experiment are shown in Fig. 2B. In this experiment, all transfections were normalized by cotransfection of a plasmid that codes for luciferase. An assay for CAT following transfection of pMLSISCAT-PA into HeLa cells indicated background levels of CAT activity. In this experiment, two different cell extract volumes were used to measure CAT expression. At the 1× level, pMLSIS-CAT-L1 showed a modest but significant increase in CAT expression compared with pMLSISCAT-PA, indicating L1 poly(A) site function. Expression from pMLSISCAT-L1 contrasted with CAT expression generated from plasmids containing the L3 poly(A) site and the β^{maj} globin poly(A) site. At the 1× cellular extract level, both the L3- and the β^{maj} globin-containing plasmids were outside the linear range of the CAT assay. At a 0.1× extract concentration, we made an accurate estimate of CAT activity for these transfections. The presence of the L3 or β^{maj} globin 3′ RNA processing site resulted in at least a 10-fold increase in expression over pMLSISCAT-L1. The in vivo utilization of each poly(A) site agreed with the results obtained with the in vitro cleavage assay.

To verify that CAT expression levels represent steady-state mRNA levels, we performed an S1 endonuclease hybrid protection assay with RNAs isolated from cells transfected with each of the pMLSISCAT plasmids (Fig. 3). The hybrid protection assay showed levels of steady-state mRNA consistent with the CAT measurements. Previous experiments have shown that the half-life of the L1-contain-

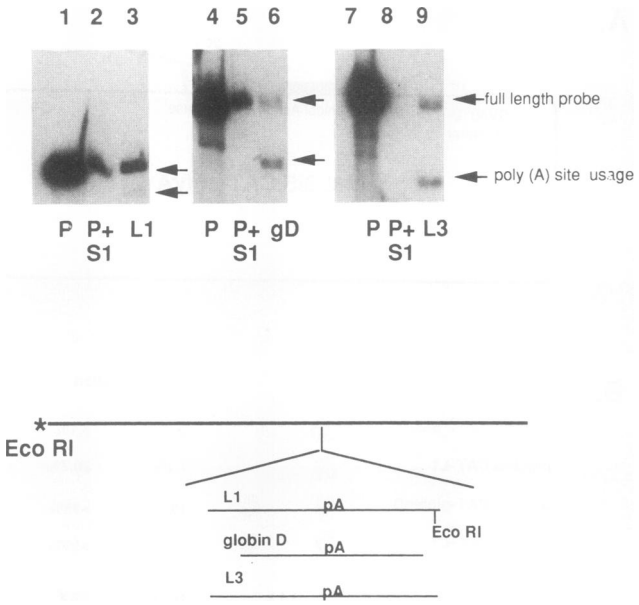

FIG. 3. 3′ end mapping of pMLSISCAT mRNAs by S1 nuclease hybrid protection assays. The pMLSISCAT-L1, gD, and L3 plasmids were digested with restriction endonuclease *Eco*RI, which cleaves at a single site 5′ to the poly(A) site (and at a second site in pMLSISCAT-L1 downstream of the L1 poly(A) site, as indicated). With ^{32}P-labeled deoxynucleoside triphosphate, each template was labeled by filling in recessed ends with the large fragment of DNA polI. Following normalization to luciferase RNA, the total RNA isolated from each transfection was hybridized to end-labeled plasmid DNA. S1 endonuclease treatment and gel electrophoresis were done as described in Materials and Methods. Lanes 1, 4, and 7 show the full-length DNA probes (P), lanes 2, 5, and 8 show digestion of each DNA probe after hybridization to nonspecific yeast RNA, and lanes 3, 6, and 9 show the products generated after digestion of each DNA probe hybridized to isolated RNA from the indicated transfection. The top arrowhead beside each lane indicates the position of each probe DNA, and the lower arrowhead indicates the position of the RNA hybrid protected species.

ing mRNA is slightly longer than that of the L3-containing mRNA (26). With pMLSISCAT-L1, assuming that transcription through the L1 poly(A) site is the same as in the L3- or β^{maj} globin-containing constructs, greater than 90% of the cleavage site-containing pre-mRNAs are not processed and are presumably degraded in the nucleus. Taken together, these data indicate that poly(A) site strength measured in vitro reflects in vivo usage and confirms that the β^{maj} globin site is an effective substrate for the 3′ processing machinery.

Termination is dependent on the presence of an efficient processing site. Having defined the β^{maj} globin poly(A) site as an efficient 3′ RNA processing site, we next determined whether altered processing efficiency can directly affect transcription termination efficiency. Previous experiments which indicated coupling of 3′ RNA processing and termination relied on mutations of the core AAUAAA or the downstream GU-rich element, which eliminated the ability to form a cleavage complex. A study by Lanoix and Acheson (18) indicated that replacement of the polyomavirus poly(A) site (which was known to be less than 100% efficient) with the rabbit β-globin poly(A) site decreased the readthrough transcript level and increased the steady-state mRNA level. By their estimate of in vivo uridine pulse-labeled RNA, transcription termination increased from an estimated 80% to 95%. We chose to examine how a low-efficiency poly(A) site such as Ad L1 affects termination efficiency. Replacement of the β^{maj} globin poly(A) site with the L1 poly(A) site in a recombinant Ad vector allowed us to determine whether formation of a stable cleavage complex and subsequent RNA cleavage are the required termination elements contributed by the β^{maj} globin poly(A) signal sequence.

Characterization of polII termination has, in most instances, relied on nascent RNA chain mapping techniques such as the nuclear run-on assay (6, 15, 19, 22, 28, 30). This assay measures the density of polII molecules which are capable of nascent RNA chain elongation in a specified region of a transcription unit. Termination is then defined as the region in a transcription unit where polII elongation is no longer detected. Changes in elongation rate or the presence of inactive polII molecules on the DNA template are not detected by this assay. With nuclear run-on assays, termination regions are usually found at least 1 kb downstream of the 3′ processing site. For the mouse β^{maj} globin gene, we adapted the Ad E1A transcription unit to facilitate our characterization of β^{maj} globin termination elements. Recombinant Ad sub360-gDEF is the parent virus previously used to define the termination elements that function in the mouse β^{maj} globin transcription unit (12).

sub360-gDEF contains 1,500 bp of the β^{maj} globin 3′ region inserted into Ad E1A exon 2 (Fig. 4A). The globin poly(A) site is located at nucleotide 152 of subgenomic element gD. The termination elements are located in subgenomic element gF. In sub360-gEF, the 308-bp gD element, which contains the globin poly(A) site, is deleted. We created two viral constructs which have replaced the globin poly(A) signal sequence with 200-bp DNA sequence elements containing the L1 poly(A) site (sub360-L1EF) or the L3 poly(A) site (sub360-L3EF) (Fig. 4A). In these constructs, the poly(A) signal sequence is placed immediately upstream of the gE and gF DNA segments. The 458-bp gE element is not required directly for termination but does function as a spacer element between the poly(A) signal (present in gD) and downstream termination element gF. Our determination of termination efficiency depends on detection of changes in polII density immediately upstream and downstream of the inserted poly(A) site and termination element. As indicated in Fig. 4A, E1A genomic subclones 1A2 and 1A3 are perfectly situated for this purpose.

For the termination assay, HeLa cells were infected with one of the four Ad constructs shown in Fig. 4A. After 4 h, cells were harvested and nuclei were isolated as described in Materials and Methods. The infected nuclei were suspended in a solution containing a final concentration of 100 μM each ATP, CTP, and GTP and 120 μCi of [^{32}P]UTP (3,000 Ci/mmol). Nuclei were incubated at 30°C for 15 min to allow brief nascent RNA chain elongation. RNA was then isolated and prepared for slot blot hybridization as described in Materials and Methods. Nitrocellulose filters containing the plasmid subclones depicted in Fig. 4B were hybridized for 36 to 40 h with approximately 2×10^7 cpm of labeled RNA. After the hybridization and filter wash procedures were completed, the nitrocellulose filters were exposed first to a Molecular Dynamics PhosphorImager cassette and then to Kodak XAR-5 film for autoradiography (Fig. 4B). Pixel values for each slot were measured by the PhosphorImager, and the termination efficiency was determined for each construct. Termination efficiency was calculated by subtracting the pGem background and compensating for varied U content in the 1A2 and 1A3 E1A subclones. The 1A3-to-1A2 ratio was equal to the percentage of readthrough tran-

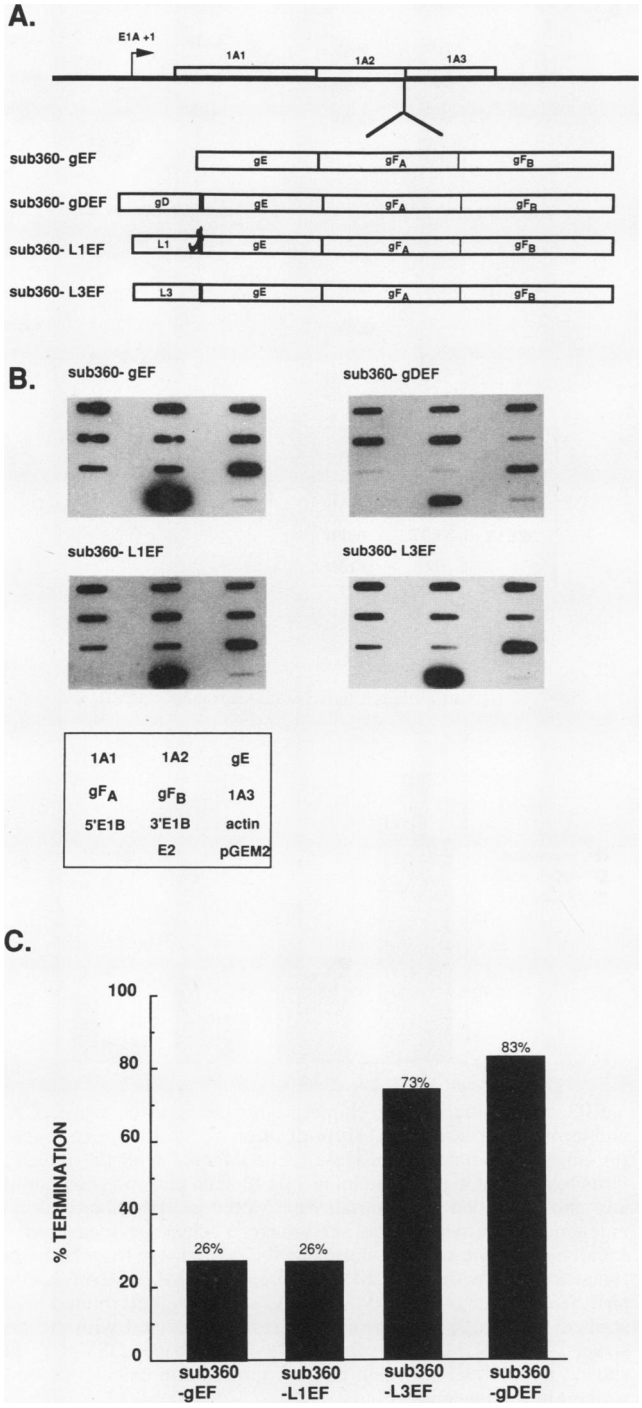

FIG. 4. Effect of poly(A) site replacements on transcription termination. (A) Schematic of the E1A region of four recombinant Ads, sub360-gEF, sub350-gDEF, sub360-L1EF, and sub360-L3EF. The relative position of the poly(A) site and the 3′ globin termination region (gEF) inserted into exon 2 of the E1A gene is indicated with respect to E1A genomic subclones 1A2 and 1A3. The level of termination was measured by determining RNA polII density in the 1A3 region versus that in the 1A2 region, as described in the text. (B) Hybridization of 2×10^7 cpm of nuclear run-on RNA isolated from HeLa cells infected with the indicated virus to a DNA slot blot grid containing 7 μg of each plasmid shown in the key. (C) Graphic indication of termination efficiency for each run-on, determined as described in Materials and Methods and elsewhere in the text.

scription. Percent termination efficiency equals 100 minus the percentage of readthrough.

The termination values obtained for each virus are shown in Fig. 4C. Virus sub360-gEF, which lacks the authentic β^{maj} globin poly(A) site, has previously been shown to have a reduced level of transcription termination. In this assay, we used the sub360-gEF virus as a poly(A) site-negative control. The difference between sub360-gEF (26% termination) and sub360-gDEF (83% termination) showed the requirement of a functional poly(A) site for transcription termination. When we performed termination assays with sub360-L1EF, we consistently found a termination level indistinguishable from that of the sub360-gEF construct lacking an authentic poly(A) site. We attribute the low level of transcription termination seen with this construct to inefficient processing complex formation and RNA cleavage. We cannot distinguish between these two events by available in vivo assays. Virus sub360-L3EF, which has the L3 poly(A) site in place of the globin sequence, functioned well (73% termination) in the termination assay. This experiment showed that loss of the termination function with sub360-L1EF was not due to a specific requirement for the globin poly(A) site but was dependent on the presence of a poly(A) site which is processed efficiently.

Chimeric poly(A) sites varied only in the downstream element shown varied processing efficiency, which correlated directly with termination efficiency. Because the L1 poly(A) site is inefficient in both processing and termination, we feel that there is a direct correlation between these two events. From previous characterization of the L1 pre-mRNA substrate, we know that it does not form a stable preprocessing complex. The basis for its low activity is the subject of ongoing investigation but appears to be the result of a lack of cooperation between the core AAUAAA (and those sequences immediately surrounding the consensus element) and the downstream GU region (26). We therefore wanted to demonstrate that the termination defect of sub360-L1EF was simply due to inefficient use of the poly(A) site and not the result of a unique feature of the L1 RNA processing site. To demonstrate that inefficient termination is the result of inefficient processing, we obtained from the laboratory of J. Nevins a series of chimeric poly(A) sites which use the core AAUAAA from the L3 poly(A) site combined with different downstream GU/U regions (34). These constructs (Fig. 5A) include downstream sequences from the μs and μm immunoglobulin heavy-chain poly(A) sites as well as from the simian virus 40 late poly(A) site. We also obtained a chimeric L3-simian virus 40 late poly(A) site which contains a 2-bp substitution in the downstream GU/U region. This alteration yields a decrease in cleavage efficiency and complex stability in vitro, as well as a reduced level of the cleavage product formed in vivo (34). Each of these poly(A) sites was used in the assays described previously for the L1, β^{maj} globin, and L3 poly(A) sites. The in vitro cleavage assay with the chimeric poly(A) sites (shown in Fig. 5B and graphically in Fig. 6) demonstrated that the order of strength of these pre-mRNA processing substrates is μm > SV_L > μs > mSV_L.

We next determined the function of the chimeric 3′ RNA processing sites by in vivo assays that measure CAT expression and transcription termination (Table 1 and Fig. 6). Of the chimeric poly(A) sites, L3μm was the strongest processing substrate and demonstrated the most efficient termination function (69%). The simian virus 40 poly(A) site was slightly less effective in the CAT assays and also showed a diminished termination function (45%). The constructs

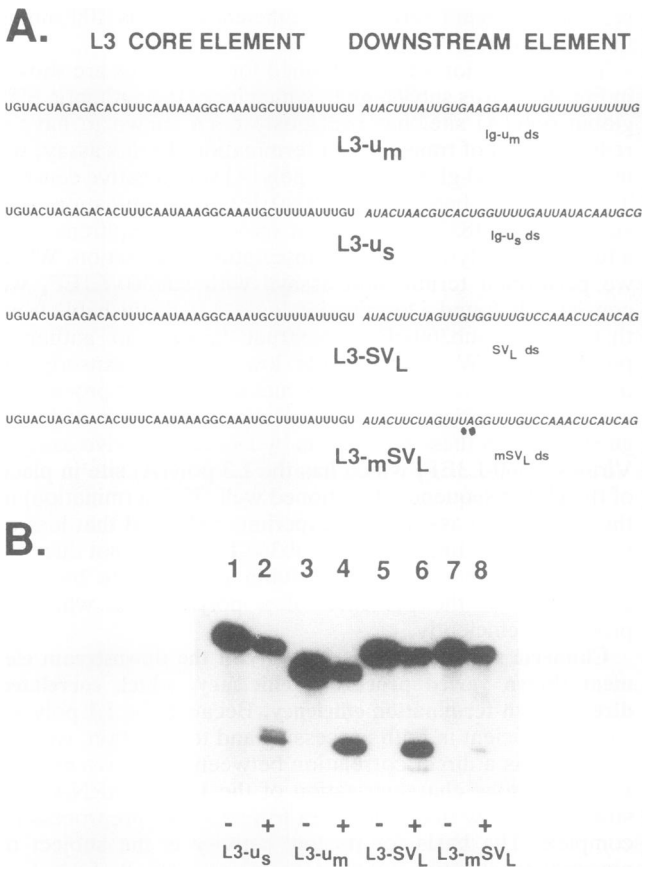

FIG. 5. Cleavage assays using chimeric pre-mRNAs containing the L3 core AAUAAA combined with various GU-rich regions. (A) RNA sequences corresponding to four chimeric RNAs, L3μm, L3μs, L3SV$_L$, and L3mSV$_L$, which were generated by T7 RNA polymerase transcription of the corresponding pGem plasmid as described in Materials and Methods. Run-off pre-RNAs were 150 nucleotides long, and all of the pre-mRNAs yielded the same final cleavage product. ds, downstream. (B) Cleavage assays done with each of the chimeric pre-mRNA substrates as described in the legend to Fig. 1 and in Materials and Methods. Lanes 2, 4, 6, and 8 contained cleavage reactions for the indicated pre-mRNAs (+). Lanes 1, 3, 5, and 7 contained input pre-mRNA substrates (−).

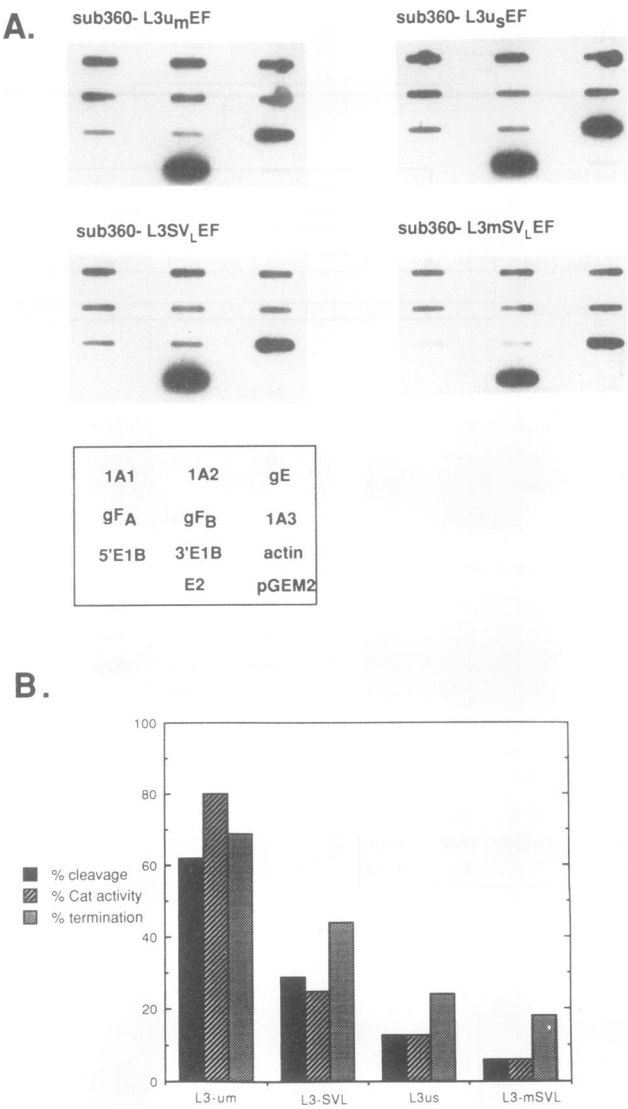

FIG. 6. Comparison of chimeric substrates in cleavage, CAT, and termination assays. (A) Hybridization of 2×10^7 cpm of nuclear run-on RNA isolated from HeLa cells infected with the indicated virus to a slot blot grid containing 7 μg of each plasmid shown in the key and described in Materials and Methods. (B) Quantitation of independent assays for the percentage of cleavage measured for each chimeric pre-mRNA; normalized CAT activity from HeLa cell transfections with pMLSISCAT-L3μm, pMLSISCAT-L3SV$_L$, pMLSISCAT-L3μs, and pMLSISCAT-L3mSV$_L$ determined as described for Fig. 2; and termination results obtained with Ad constructs sub360-L3μmEF, sub360-L3μsEF, sub360-L3SV$_L$EF, and sub360-L3mSV$_L$EF given in percent termination calculated as described in the legend to Fig. 4.

which contain the immunoglobulin μs downstream GU/U region functioned at a level similar to that of the L1 substrate seen in the previous experiments. Finally, the L3mSV$_L$ constructs yielded processing, CAT, and termination values at background levels. In the termination assay, we actually observed less termination than with the sub360-gEF virus. One possible explanation for the very low level of termination seen with this construct is that the mutated site may mask use of downstream cryptic poly(A) sites which have been previously identified in the gE region (20). The results obtained from these experiments indicate a general consistency between the level of observed 3′ processing and the levels of observed termination. These results support the concept that altered efficiencies of poly(A) site utilization directly correlate with the altered termination efficiencies.

DISCUSSION

The objective of the present series of experiments was to expand our understanding of 3′-end processing as it relates to transcription termination. Our results indicate that the primary determinant of 3′ transcription termination for the β^{maj} globin gene is efficient use of its 3′ processing site. Formation of the preprocessing complex and, presumably, cleavage of the pre-mRNA are required first steps which initialize the series of events leading to RNA polII transcription termination. The experimental data also reinforce the notion that poly(A) sites can vary considerably as substrates for the processing machinery and that an inefficient processing event leads to an inefficient termination event. We also

TABLE 1. In vivo CAT and transcription termination assays using L3 chimeric pre-mRNA vectors[a]

Chimeric construct	% CAT activity	% Termination
L3μm	80, 70	75, 63
L3μs	12, 11	17, 30
$L3SV_L$	25, 23	45, 42
$L3mSV_L$	6, 11	20, 16

[a] Results of duplicate assays are shown.

found that poly(A) site cleavage efficiency can be altered by varying the downstream GU-rich region.

The data presented here and in our previous report (32) indicate that when RNA polII transcribes through a poly(A) site there is no direct interaction between the pre-mRNA cleavage site and polII, which converts the elongation complex into a termination complex. This conclusion is based on two experimental findings. (i) 3′ processing sites that contain functional required elements, such as L1 or the chimeric constructs, did not function equally in the termination assay. If AAUAAA or GU-rich pre-mRNA sequences did interact directly with RNA polymerase, we would not expect to see a difference in termination activity with the various poly(A) sites used in these studies. (ii) We previously demonstrated that a construct which has a termination element positioned 200 bp downstream of the β^{maj} globin poly(A) site (as opposed to the 600 bp downstream seen in sub360-gDEF) had lost the ability to elicit an efficient termination event. If the sequences of the poly(A) site were required to act directly on RNA polymerase at the point of transcription, we would not expect spacing between the poly(A) site and the termination site to effect termination. These same arguments diminish the likelihood that polymerase pausing at a processing site is important to termination or that a termination or antitermination factor is involved directly through the processing site as it is being generated.

The accumulation of evidence from this work indicates that 3′ processing precedes termination and is necessary for termination to occur. We are making the assumption that processing complex formation is indistinguishable from cleavage and that cleavage of the pre-mRNA generates the effector responsible for conversion of a stable polII elongation complex into a termination-competent complex. A second model for creation of the termination-competent complex would rely on some form of the uncleaved preprocessing complex as the effector for generation of termination-competent polII. This model would require either direct preprocessing complex contact with the polymerase through an RNA loop, possible displacement of an associated antitermination factor (possibly a subunit of CPSF) from the RNA polymerase, or loading of a termination factor via the preprocessing complex and scanning 5′ to 3′ along the RNA until contact with the elongation complex occurs.

In the first model, in which the cleavage event converts the elongation complex into a termination-competent complex, there is no evidence which indicates how the cleaved pre-mRNA transmits information to the elongating polymerase. A looping interaction between the newly generated 3′ cleavage product, which does not contain a 5′ cap structure, and the elongation complex may be the simplest model which depends on the cleavage event. Another prominent model, proposed by J. Manley and separately by N. Proudfoot and their respective colleagues (7, 24, 29), invokes a 5′→3′ exonuclease which attacks the unprotected 5′ end of the downstream cleavage product. Eventual contact between the nuclease and the polymerase in some way causes polymerase destabilization, pausing, and/or displacement. Because of the technical difficulties associated with characterization of short-lived, low-abundance nuclear RNA, experiments which prove or disprove this hypothesis are not available. Resolution of the mechanism involved in creating the termination-competent complex requires very novel in vivo techniques or development of a suitable in vitro system which successfully couples the elongation, processing, and termination machineries.

Finally, we ask how, once the termination-competent complex is generated, does the termination event take place? Concrete evidence which addresses this issue is not available. An appropriate in vitro system is required to work out the biochemistry of this event. What we know about the poly(A) site-dependent termination event is that multiple heterogeneous elements which are able to arrest polII elongation 3′ to an efficient poly(A) site exist. As we have previously explained, we feel that the spacing requirement between the poly(A) site and the termination elements in gF allows time for the cleavage reaction to occur; this then effects RNA polymerase conversion into a termination-competent complex. The spacing requirement is therefore not directly involved in termination once the termination-competent elongation complex has been formed. In the β^{maj} gF termination region, we have identified at least three elements which can induce termination in the 70% range by the nuclear run-on assay. The placement of these sites implies that the natural efficiency we have originally observed with the β^{maj} globin 3′ termination region is the result of redundant termination elements which act in concert. The smallest of the β^{maj} globin elements is a 69-bp sequence which is A rich and has similarities to a rabbit β-globin element identified by Enriquez-Harris et al. as a transcription pause element which also acted as a site of transcription termination (11). As mentioned earlier, Connelly and Manley (7) identified a poly(A) site-dependent termination event which involved a CAAT-binding protein as a presumable block to polymerase elongation. It is our impression that once a termination-competent complex is formed, there is a variety of RNA-DNA or DNA-protein structures which may act to arrest polymerase elongation. Therefore, the efficiency of 3′ processing complex formation is the key determinant of transcription termination and presumably equally important to recycling of RNA polII to another initiation event.

ACKNOWLEDGMENTS

We thank J. Nevins for making the chimeric L3 poly(A) sites available to us.

This work was supported by NIH grant GM41967 awarded to E.F.-P.

REFERENCES

1. **Ashfield, R., P. Enriquez-Harris, and N. J. Proudfoot.** 1991. Transcriptional termination between closely linked human complement genes C2 and factor B: common termination factor for C2 and c-myc. EMBO J. **10**:4197–4207.
2. **Bardwell, V. J., M. Wickens, S. Bienroth, W. Keller, B. S. Sproat, and A. I. Lamond.** 1991. Site directed ribose methylation identifies 2′ OH groups in polyadenylation substrates critical for AAUAAA recognition and poly(A) addition. Cell **65**:125–133.
3. **Brasier, A. B., J. E. Tate, and J. F. Habener.** 1989. Optimized use of the firefly luciferase assay as a reporter gene in mamma-

lian cell lines. BioTechniques **7**:1116–1122.
4. **Chodchoy, N., N. B. Pandey, and W. F. Marzluff.** 1991. An intact histone 3′ processing site is required for transcription termination in a mouse histone H2A gene. Mol. Cell. Biol. **11**:497–509.
5. **Chomczynski, P., and N. Sacchi.** 1987. Single step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extractions. Anal. Biochem. **612**:156–159.
6. **Citron, B., E. Falck-Pedersen, M. Salditt-Georgieff, and J. E. Darnell, Jr.** 1984. Transcription termination occurs within a 1000 base pair region downstream from the poly(A) site of the mouse β^{maj}-globin gene. Nucleic Acids Res. **12**:8723–8731.
7. **Connelly, S., and J. L. Manley.** 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev. **2**:440–452.
8. **Connelly, S., and J. L. Manley.** 1989. RNA polymerase II transcription termination is mediated specifically by protein binding to a CCAAT box sequence. Mol. Cell. Biol. **9**:5254–5259.
9. **Connelly, S., and J. L. Manley.** 1989. A CCAAT box in the adenovirus major late promoter functions as part of an RNA polymerase II termination signal. Cell **57**:561–571.
10. **de Wet, J. R., K. V. Wood, M. DeLuca, D. R. Helinski, and S. Subramani.** 1987. Firefly luciferase gene: structure and expression in mammalian cells. Mol. Cell. Biol. **7**:725–737.
11. **Enriquez-Harris, P., N. Levitt, D. Briggs, and N. J. Proudfoot.** 1992. A pause site for RNA polymerase II is associated with termination of transcription. EMBO. J. **10**:1833–1842.
12. **Falck-Pedersen, E., J. Logan, T. Shenk, and J. E. Darnell, Jr.** 1985. Transcription termination within the E1A gene of adenovirus induced by insertion of the mouse β^{maj}-globin terminator element. Cell **40**:897–905.
13. **Gilmartin, G. M., and J. R. Nevins.** 1989. An ordered pathway of assembly of components required for polyadenylation site recognition and processing. Genes Dev. **3**:2180–2189.
14. **Graham, F. R., J. Smiley, W. Russell, and R. Naim.** 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J. Gen. Virol. **36**:59–72.
15. **Hagenbüchle, O., P. K. Wellauer, D. L. Cribbs, and U. Schibler.** 1984. Termination of transcription in the mouse α-amylase gene *Amy 2a* occurs at multiple sites downstream of the polyadenylation site. Cell **38**:737–744.
16. **Huang, M. T. F., and C. M. Gorman.** 1990. Intervening sequences increase efficiency of RNA 3′ processing and accumulation of cytoplasmic RNA. Nucleic Acids Res. **18**:937–947.
17. **Keller, W., S. Bienroth, K. M. Lang, and G. Christofori.** 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3′ processing signal AAUAAA. EMBO J. **10**:4241–4249.
18. **Lanoix, J., and N. H. Acheson.** 1988. A rabbit β-globin polyadenylation signal directs efficient termination of transcription of polyoma virus DNA. EMBO. J. **7**:2515–2522.
19. **LeMeur, M. A., B. Galliot, and P. Gerlinger.** 1984. Termination of the ovalbumin gene transcription. EMBO J. **3**:2779–2786.
20. **Logan, J., E. Falck-Pedersen, J. E. Darnell, Jr., and T. Shenk.** 1987. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse β^{maj}-globin gene. Proc. Natl. Acad. Sci. USA **84**:8306–8310.
21. **Logan, J., and T. Shenk.** 1984. Adenovirus tripartite leader sequence enhances translation of mRNAs late after infection. Proc. Natl. Acad. Sci. USA **81**:3655–3659.
22. **Maa, M.-C., J. M. Chinsky, V. Ramamurthy, B. D. Martin, and R. E. Kellems.** 1990. Identification of transcription stop sites at the 5′ and 3′ ends of the murine adenosine deaminase gene. J. Biol. Chem. **265**:12513–12519.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Manley, J. L.** 1988. Polyadenylation of mRNA precursers. Biochim. Biophys. Acta **950**:1–12.
25. **Moore, C. L., and P. A. Sharp.** 1985. Accurate cleavage and polyadenylation of exogenous RNA substrate. Cell **41**:845–855.

25a. **Murthy, K. G. K., and J. L. Manley.** 1992. Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. J. Biol. Chem. **267**:14804–14811.

26. **Prescott, J. C.** Unpublished data.
27. **Prescott, J. C., and E. Falck-Pedersen.** 1992. Varied poly(A) site efficiency in the adenovirus major late transcription unit. J. Biol. Chem. **267**:8175–8181.
28. **Pribyl, T. M., and H. G. Martinson.** 1988. Transcription termination at the chicken βH-globin gene. Mol. Cell. Biol. **8**:5369–5377.
29. **Proudfoot, N. J.** 1989. How RNA polymerase II terminates transcription in higher eukaryotes. Trends Biochem. Sci. **14**: 105–110.
30. **Rohrbaugh, M. L., J. E. Johnson III, M. D. James, and R. C. Hardison.** 1985. Transcription unit of the rabbit β1 globin gene. Mol. Cell. Biol. **5**:147–160.
31. **Sato, K., R. Ito, K.-H. Baek, and K. Agarwal.** 1986. A specific DNA sequence controls termination of transcription in the gastrin gene. Mol. Cell. Biol. **6**:1032–1043.
32. **Tantravahi, J., M. Alvira, and E. Falck-Pedersen.** 1993. Characterization of the mouse β^{maj} globin transcription termination region: a spacing sequence is required between the poly(A) signal sequence and multiple downstream termination elements. Mol. Cell. Biol. **13**:578–587.
33. **Wahle, E., and W. Keller.** 1992. The biochemistry of 3′ end cleavage and polyadenylation of messenger RNA precursors. Annu. Rev. Biochem. **61**:419–440.
34. **Weiss, E. A., G. M. Gilmartin, and J. R. Nevins.** 1991. Poly(A) site efficiency reflects the stability of complex formation involving the downstream element. EMBO J. **10**:215–219.
35. **Whitelaw, E., and N. J. Proudfoot.** 1986. Alpha thalassaemia caused by poly(A) site mutation reveals that transcription termination is linked to 3′ end processing in the human alpha-2 globin gene. EMBO J. **5**:2915–2922.