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

JOGIRAJU TANTRAHAHI, MAURICIO ALVIRA, AND ERIK FALCK-PEDERSEN*

*Department of Microbiology, W. Randolph Hearst Foundation, Cornell University
Medical College, 1300 York Avenue, New York, New York 10021*

Received 18 August 1992/Returned for modification 25 September 1992/Accepted 19 October 1992

For the majority of mRNA encoding eukaryotic transcription units, there is little or no knowledge of the elements responsible for transcription termination or how they may interact with RNA polymerase. In this report, we have used recombinant adenovirus reporter vectors to characterize the mouse β^{maj} globin sequence elements that cause transcription termination. Within the globin 3' termination region, we have identified at least three sequence elements which induce significant levels of transcription termination (>50%). The smallest functionally active element (64% termination) is 69 bp in length. The natural arrangement of these elements results in a cumulative termination which is greater than 90%. Recognition of the termination elements by RNA polymerase II depends on the presence of a functional poly(A) signal sequence. We demonstrate that efficient transcription termination depends on appropriate spacing between the poly(A) signal sequence and the termination element.

The genes transcribed by RNA polymerase II (Pol II) exhibit a variety of mechanisms to terminate transcription and form RNA 3' ends. The small nuclear RNAs, including U1 and U2, rely directly on transcription termination as the first step in the formation of their respective 3' ends (1, 5, 20, 21, 25, 44). This direct termination-3' end formation mechanism contrasts with 3' end formation in mRNAs. All mRNAs have their 3' ends generated by an RNA processing event. Histone mRNAs, which are the only nonpolyadenylated mRNA transcripts, undergo endonucleolytic cleavage of the pre-mRNA at a site downstream from an RNA stem-loop and a purine-rich sequence (1, 16, 32, 40; and reviewed in reference 2). The remainder of mRNAs rely on a tightly coupled cleavage/polyadenylation reaction to generate the 3' end of the mRNA transcript (reviewed in references 30, 35, 36, and 41).

Nascent RNA labeling techniques, such as the nuclear run-on assay, have been used to map regions where Pol II elongation occurs in endogenous transcription units, such as in the murine α -amylase, chicken ovalbumin, and human adenosine deaminase genes and several globin genes, including the murine β^{maj} globin gene (6, 19, 24, 28, 34, 37). These experiments also demonstrated that Pol II elongation arrest takes place approximately 1 kb past the poly(A) site in the ovalbumin and globin transcription units and 3 to 4 kb past the main poly(A) signal element in the α -amylase and adenosine deaminase genes. By using elongation arrest in nuclear run-on assays as a working definition of Pol II transcription termination, in each of these cases, Pol II termination has been localized to a region of several hundred to 1,000 bp.

Studies have also demonstrated that for certain histone genes and genes which utilize cleavage and polyadenylation to form 3' ends, the presence of a processing site is required for efficient transcription termination. Deletion of the stem-

loop structure or the purine-rich region of the mouse H2a-614 histone gene results in a loss of both RNA processing and transcription termination (3). Similarly, mutations or deletions of the poly(A) cleavage site signal elements (the consensus AAUAAA or the downstream GU-rich sequence) affect both processing and transcription termination (4, 7, 26, 42). For termination events which require intact 3' processing sites, relatively few downstream termination elements have been identified to the nucleotide level. Using plasmid transfections, Connelly and Manley demonstrated that the inverted CCAAT sequence element of the adenovirus (Ad) major late promoter has the ability to block polymerase elongation when placed downstream of a poly(A) signal sequence in an orientation-dependent manner (8, 9). Chodchoy et al. demonstrated that the termination region for the histone H2a-614 gene is found in the promoter region of the downstream H3-614 transcription unit (3). Promoter elements and the proteins which bind to them may represent one specific class of Pol II termination element. In transcription units like the mouse β^{maj} globin gene, no such promoter element associated with the termination region has been identified. The termination element in the β^{maj} globin region may therefore represent a distinct class of Pol II terminator.

To gain insights into the nature and mechanism of the Pol II termination event, we have studied the mouse β^{maj} globin termination region located downstream from the globin poly(A) signal sequence. The β^{maj} globin termination event is poly(A) site dependent, occurs some 500 to 1,000 nucleotides downstream of the poly(A) site in the globin F (gF) fragment, and causes transcription termination with great efficiency (greater than 90% when measured by nuclear run-on assays) (6, 13, 26). Mapping of steady-state nuclear RNA from induced mouse erythroleukemia cells has previously indicated heterogeneity of globin RNA 3' ends in the gF region (6). Because excised intron RNA and RNA downstream of the processed poly(A) site are subject to very rapid degradation in the nucleus, steady-state RNA analysis

* Corresponding author.

TABLE 1. DNA fragments used

Fragment	Nucleotides	Derived from:	Inserted into:
Mutants of gF			
F Δ 3' 634	1-634	Exo III ^a digestion	pMLP6 gDE Xba I-Sal I
F Δ 3' 548	1-548	Exo III digestion	pMLP6 gDE Xba I-Sal I
F Δ 3' 404	1-404	Restriction digestion	pMLP6 gDE
F Δ 3' 203	1-203	Exo III digestion	pMLP6 gDE Xba I-Sal I
F ₄₂₂₋₅₁₁	442-511	PCR	pMLP6 gDE Xba I-Sal I
F ₄₂₂₋₆₃₄	442-634	PCR	pMLP6 gDE Xba I-Sal I
F Δ 5' 420	420-809	Restriction digestion	pMLP6 gDE
F Δ 5' 430	430-809	Exo III digestion	pMLP6 gD CAT
F Δ 5' 519	519-809	Exo III digestion	pMLP6 gDE Xba I-Sal I
F Δ 5' 575	575-809	Exo III digestion	pMLP6 gDE Xba I-Sal I
F Δ 5' 619	619-809	Exo III digestion	pMLP6 gDE Xba I-Sal I
CAT spacer	4375-5000	PCR amplification of the bacterial CAT gene	pMLP6 gD

^a Exo III, exonuclease III.

is not the most direct method for identifying and characterizing transcription termination elements. We have chosen to characterize functional β^{maj} globin termination elements by inserting them into the second exon of the Ad E1a transcription unit and measuring their function directly in nuclear run-on termination assays. By working with the entire 3' end of the β^{maj} globin sequence, we have identified specific termination sequence elements present in the gF region as well as parameters required for the interaction between the 3' processing site and the termination elements.

MATERIALS AND METHODS

Cell and tissue culture. Human embryonic 293 monolayer cells (17) constitutively expressing the Ad E1a and E1b genes were grown in Dulbecco's modified Eagle medium and 10% fetal calf serum. These cells were used for cotransfections, the preparation of initial virus lysates, and plaque purifications. Large-scale virus propagation was performed in spinner 293 cells grown in Joklik's minimal essential medium and 5% horse serum. HeLa spinner cells, maintained under the same conditions, were used for nuclear run-on analyses.

Construction of deletion and restriction mutants of gF. Table 1 describes the origin of each gF derivative and how the chloramphenicol acetyltransferase (CAT) spacer was constructed (13, 26, 27). Exonuclease digestion of the gF region was performed according to Maniatis et al. (29). Mutants of gF, whether constructed through exonuclease digestion, restriction digestion, or polymerase chain reaction (PCR), were cloned into vector pMLP6 gD, pMLP6 gDE, pMLP6 gDE Xba I-Sal I, or pMLP6 gD CAT to generate a parent plasmid that would be used for the generation of a recombinant virus. Plasmid pMLP6 gD has been described elsewhere (26). Plasmids pMLP6 gDE, pMLP6 gDE Xba I-Sal I, and pMLP6 gD CAT all contain Ad sequence from nucleotides 1 to 5776 cloned into the pML vector. Both pMLP6 gDE and pMLP6 gDE Xba I-Sal I have the mouse β^{maj} globin gD and gE regions inserted into the unique *Xba*I site at nucleotide 1336 in the second exon of the Ad E1a gene. In addition, pMLP6 gDE Xba I-Sal I has a *Sal*I site directly downstream of the *Xba*I site. Plasmid pMLP6 gD CAT similarly has the mouse β^{maj} globin gD region inserted into the Ad E1a gene's second exon, but the gE region has been replaced with a 624-bp segment from the CAT gene.

The following nomenclature has been devised to name gF region mutants. For deletion mutants, each fragment is

named F Δ (3' or 5') (number). The Δ identifies the fragment as a deletion mutant. The 3' or 5' indicates whether 3' sequences or 5' sequences of gF have been deleted. When 3' is next to a number, the number indicates the length of the fragment. Thus, fragment F Δ 3' 634 is a deletion mutant ending at nucleotide 634. All 3' deletion mutants start at nucleotide 1. When 5' is next to a number, the number indicates the first nucleotide of the fragment. Thus, fragment F Δ 5' 420 is a deletion mutant starting at nucleotide 420 and ending at nucleotide 809. All 5' deletion mutants end at nucleotide 809, since the entire gF region is 809 bp long. The F_{n-x} nomenclature describes internal restriction fragments or PCR-generated mutants of gF. Thus, fragment F₄₄₂₋₅₁₁ encompasses nucleotides 442 to 511 of gF.

Construction of recombinant adenoviruses. Viruses sub360, sub360 gD, sub360 gEF, sub360 gDF, and sub360 gDEF have been described elsewhere (13, 27, 37). New recombinant Ad constructs were made by using the overlap recombination method as previously described (13, 26, 27). Parent plasmids were linearized and cotransfected into 293 cells with the *dl*309 large fragment, using the CaPO₄ 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 according to the parent virus, sub360, and the insertions each virus contains. Thus, virus sub360 gDEF Δ 5' 420 contains the gD, gE, and F mutant F Δ 5' 420 inserted at nucleotide 1336 in the Ad E1a gene.

Nuclear run-on assays. HeLa cells (10⁸) grown in Joklik's minimal essential medium with 5% horse serum were infected with recombinant Ad constructs at 2,000 particles per cell for 4 h. For some infections, cycloheximide was added to 20 μ g/ml at 3 h postinfection to elevate transcription from the E1a gene (10). After infection, the cells were repelleted, and all operations were subsequently done at 4°C. The cells were washed twice in 1 $\times$ phosphate-buffered saline and resuspended in 4 ml of reticulocyte standard buffer (10 mM NaCl, 10 mM Tris-Cl [pH 7.4], 5 mM MgCl₂). The cells were incubated in reticulocyte standard buffer on ice for 5 min and lysed by Dounce homogenization; lysis was checked microscopically. The lysate was centrifuged at 1,500 rpm for 5 min, yielding the nuclear pellet. The nuclear pellet was resuspended in 2 ml of transcription buffer (10% glycerol, 10 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [HEPES; pH 7.9], 140 mM KCl, 10 mM MgCl₂, 1 mM

FIG. 1. Nuclear run-on mapping of the mouse β^{maj} globin termination region in Ad expression vectors. (A) Configuration of β^{maj} globin 3' termination elements inserted into the Ad E1a transcription unit. The left end of virus sub360 contains the E1a coding sequence [cap site at 498 and poly(A) site at 1630] driven by the Ad major late transcription control region (MLTCR). Globin 3' DNA fragments are inserted at position 1336 of sub360 in the E1a second exon (splice acceptor at 1226). The E1a RNA 13S splice donor is at 1111, and the 12S splice donor is at 976. The relative positions of the E1a region and of β^{maj} globin genomic subclones (1A1, 1A2, 1A3, gD, gE, and gF) are indicated. The virus nomenclature indicates the segment of globin genomic DNA inserted. For sub360 gDEF, +1 indicates the site of the globin poly(A) site. The distance from the poly(A) site of each downstream element is indicated. Viruses were constructed by overlap recombination as indicated in Materials and Methods. (B) Autoradiograph of nuclear run-on [^{32}P]UMP-labeled RNA hybridized to nitrocellulose filters containing genomic subclones. Labeled nuclear run-on RNA isolated from 10^8 HeLa cells infected with 2×10^3 particles of the indicated virus construct per cell was hybridized to a DNA slot blot grid containing the plasmid genomic subclones indicated in the key. Run-on assays, filter preparation, and hybridizations were carried out as described in

dithiothreitol) and incubated on ice for 2 min. The nuclei were spun at 2,000 rpm for 3 min, and the nuclear pellet was resuspended in 100 μl of transcription cocktail (120 μCi of [$\alpha\text{-}^{32}\text{P}$]UTP, 1 mM GTP, 1 mM ATP, 1 mM CTP, and 1 mM dithiothreitol in transcription buffer) by vigorous agitation and incubated on ice for 2 min. The nuclei were then incubated at 30°C for 15 min. The reaction was stopped by the addition of 1 ml of HSB (0.5 M NaCl, 10 mM Tris-Cl [pH 7.4], 5 mM MgCl_2 , 2 mM CaCl_2) and 50 μg of DNase I. After 1 min of DNase treatment, 8 ml of ETS (10 mM EDTA [pH 8.0], 10 mM Tris-Cl [pH 8.0], 0.2% sodium dodecyl sulfate [SDS]) and 200 μl of 0.5 M EDTA were added. The nuclei were extracted successively with hot phenol, phenol-chloroform, and chloroform, and RNA was isolated as described by Nevins (33). The RNA pellet was resuspended in 2 ml of sterile water, and 2 ml of 4 M LiCl was added. After 8 h of incubation at 0°C, the RNA was repelleted by centrifugation at 7,000 rpm for 45 min. The pellet was resuspended in 1.5 ml of sterile water. The RNA was ethanol precipitated once more prior to hybridization and finally resuspended in 210 μl of 0.1 M NETS (0.1 M NaCl, 10 mM EDTA [pH 8.0], 10 mM Tris-Cl [pH 7.4], 0.2% SDS). Total counts per minute for each run-on assay were determined by counting an aliquot in scintillant (AquaSol).

Preparation of nitrocellulose filters for slot blot hybridization. A Schleicher & Schuell slot blot manifold was used for the preparation of all slot blot grids; 5 μg 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 HEPES and loaded onto a nitrocellulose filter in the presence of 5 $\times$ SSC (1 $\times$ 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 (13): 1A1 (Ad2 sequence from 1 to 1010), 1A2 (Ad2 sequence from 1010 to 1336), 1A3 (Ad2 sequence from 1336 to 1671), 1B5' (Ad2 sequence from 1699 to 2153), 1B3' (Ad2 sequence from 3577 to 4030), mouse β^{maj} globin major segment gF_A (nucleotides 1 to 404 from the gF region), gF_B (nucleotides 417 to 809 from the gF region), E2 (Ad2 sequence from 22307 to 228576), and actin (a human actin cDNA cloned into pBlueScript [a gift of D. Catanzaro]). Either pBST SK- or pGEM I was used as a negative control.

Hybridization of nascently labeled RNA to nitrocellulose filters. The RNA was fragmented by treatment with 0.2 N NaOH at 0°C for 10 min prior to hybridization and added to 2 ml of HYB buffer. HYB buffer is 1 $\times$ Denhardt's solution, 100 μg of yeast RNA per ml, and 100 μg of herring sperm DNA per ml in 2 $\times$ TESS [2 $\times$ TESS is 20 mM *N*-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid, 0.3 M NaCl, 10 mM EDTA (pH 8.0), and 0.2% SDS]. Nitrocellulose filters were prehybridized with 2.5 ml of HYB buffer for 2 h at 65°C before the addition of the labeled RNA. RNA was hybridized to the filters for 36 to 40 h at 65°C. The filters were washed successively with 2 $\times$ SSC-0.5% SDS (2 $\times$ SSC is 0.3 M NaCl plus 30 mM sodium citrate) for 30 min at room

Materials and Methods. The 809-bp gF region is represented by two subclones, gF_A and gF_B, which contain the proximal (F₁₋₄₀₄) and distal (F₄₂₀₋₈₀₉) portions of gF, respectively. Termination efficiency is determined by comparing the 1A3 signal intensity with the 1A2 signal intensity as described in the text.

FIG. 2. Quantitation of termination function for terminator virus constructs. (A) Quantitation of termination efficiency was determined by comparing the signal intensities corresponding to run-on RNA which hybridized to E1A subclones 1A2 and 1A3, which precede or follow the inserted globin elements. Autoradiographs were scanned with an LKB densitometer, or slot blot signals were directly measured with the Molecular Dynamics PhosphorImager system. The ratio of 1A3 to 1A2 after compensation for U content in the homologous transcript and subtraction of background hybridization (pBST SK-) plasmid DNA is defined as percent termination. (B) Determination of polymerase density in the globin termination region. As a relative measure of where a block to RNA polymerase elongation occurs, the intensity of labeled RNA complementary to the globin genomic subclones gD, gE, gF_A and gF_B was tabulated. The proximal globin element was used as a standard for distal elements. All measurements are shown after compensation for U content and subtraction of background.

temperature, $2\times$ SSC-0.1% SDS for 45 min at 65°C , $2\times$ SSC only for 15 min at 65°C , $2\times$ SSC-5 μg of pancreatic RNase per ml for 30 min at room temperature, $2\times$ SSC only for 15 min at 65°C , and $2\times$ SSC only for 15 min at room temperature. The filters were dried under a heating lamp and autoradiographed. Termination efficiencies were quantified by using a Molecular Dynamics PhosphorImager.

Chemicals, enzymes, and reagents. Nucleoside triphosphates were purchased from Pharmacia P-L. Radiolabeled [α - ^{32}P]UTP and [α - ^{32}P]dATP were purchased from DuPont/New England Nuclear. Sequenase enzyme and sequencing reagents were purchased from United States Biochemical. Oligonucleotide primers were made by Maija Skangalis (Department of Microbiology, Cornell University Medical College) or bought from Oligos Etc. or Operon, Inc. Restriction enzymes, modifying enzymes, and *Taq* DNA polymerase were obtained from many sources, including Promega Biotec, Bethesda Research Laboratories, New England Biolabs, Boehringer Mannheim, and Perkin-Elmer Cetus. RNAzol was either purchased from Cinna/Biotecx or made as described by Chomczynski and Sacchi (4). Tissue culture supplies were bought from GIBCO, and all other chemicals were from either Fisher or Sigma.

RESULTS

Previous work has defined two elements, the murine β^{maj} globin poly(A) signal sequence and a downstream 809-bp termination region (gF), that are necessary and sufficient to induce Pol II termination (13, 26). This determination was made by using two forms of nascent RNA labeling, in vivo [^3H]uridine pulse-labeling and the nuclear run-on assay, to measure termination efficiency. Termination efficiency was determined by comparing RNA polymerase density in regions upstream and downstream of the globin termination elements. We have used the nuclear run-on assay with the previously described recombinant Ad system to identify where in the gF element termination occurs. Nuclear run-on assays were carried out by using HeLa cells infected with

one of the six viruses diagrammed in Fig. 1A. The globin 3' sequences were inserted in the second exon of the E1A transcription unit. Hybridization of labeled RNA to a DNA slot blot (Fig. 1B) yields a measurement of termination efficiency by comparing the Pol II density upstream of the termination element (1A2 signal) with the Pol II density downstream of the termination element (1A3 signal). Cellular and viral DNA slots included as hybridization controls are indicated and described in Materials and Methods. As expected, the parent virus, sub360, does not show a change in polymerase density between 1A2 and 1A3 (Fig. 2A). Likewise, virus sub360 gD shows no difference in polymerase density between 1A2 and 1A3. In sub360 gD, the globin poly(A) cleavage site is at position 159 of the 308-bp gD insert (Fig. 1A). When we extend the distance between the poly(A) cleavage site and the 1A3 fragment to 600 bp in virus sub360 gDE, we see that the additional spacer sequence does not significantly increase termination efficiency. These data provide evidence that in the absence of a termination element, Pol II can elongate well past the globin poly(A) site.

When the gF element is present 3' to the gDE and gD regions (viruses sub360 gDEF and sub360 gDF, respectively), we clearly see the effect of this element on RNA polymerase elongation into the 1A3 region of the E1A reporter gene (Fig. 1B). Quantitation of termination efficiency (Fig. 2A) yields levels of 90% termination with sub360 gDEF and 71% termination with sub360 gDF, in agreement with previous results (11). By quantitatively measuring the hybridization signal among the globin subclones, we can approximate where transcription declines within the globin sequence. Such a measurement made with sub360 gDEF shows that the gD and gE fragments are transcribed at comparable levels (Fig. 2B). The next contiguous segment of globin sequence is a gF subclone (gF_A) containing the first 404 bp of the gF region. This region has Pol II density at levels equivalent to those of the gD and gE regions (Fig. 2B). Transcription levels strongly decline in the gF_B region,

FIG. 3. Termination efficiency measured with the 3'→5' gF deletion virus constructs. (A) Schematic representation of termination virus constructs which contain 3'→5'-deleted gF sequences. All constructs contain intact gDE elements and proximal gF sequence. The 3' endpoints of deletions generated by either restriction digestion or exonuclease III (as indicated in Materials and Methods and Table 1) are denoted by the construct nomenclature (i.e., sub360 gDEFA3' 634 contains gF sequence from nucleotides 1 through 634). MLTCR, Ad major late transcription control region. Percent termination is the average of the three separate experiments (values shown in parentheses). (B) Nuclear run-on slot blot for each gF deletion construct. Experimental conditions were as described in the legend to Fig. 1 and Materials and Methods.

which contains gF sequence from nucleotides 420 to 809. In virus construct sub360 gDF, which lacks the 443-bp gE region, we also see high levels of labeled RNA hybridizing to subclone gFA and a lower level of labeled transcript hybridizing to subclone gFB. Virus sub360 gEF, which lacks the major globin poly(A) signal sequence (present in gD) but contains weak cryptic poly(A) signal elements (26) in the gE region, shows decreased termination within the E1A gene (Fig. 1 and 2). We used this virus as a globin hybridization control showing roughly equal levels of labeled transcript hybridizing to the gE, gFA, and gFB fragments. We conclude that in the termination-competent constructs sub360 gDEF and sub360 gDF, the majority of RNA polymerases transcribe well into the gF region, terminating somewhere in the gFB region.

In vivo mapping of the globin termination element by

FIG. 4. Termination efficiency measured with 5'→3' gF deletion virus constructs. (A) Schematic representation of termination virus constructs which contain 5'→3' gF deletions. All constructs contain the gDE sequence, which is followed by the gF 5'-deleted element. The nomenclature indicates the 5' gF nucleotide present in each construct; i.e., sub360 gDEFA5' 420 contains gF sequence from nucleotides 420 to 809, which is the 3' end of the intact gF element. The thin line represents the deleted gF sequence in each construct. MLTCR, Ad major late transcription control region. Percent termination is indicated as in Fig. 3A. (B) Nuclear run-on slot blot for each gF deletion construct. Experimental conditions were described in the legend to Fig. 1 and Materials and Methods.

deletion analysis. We set out to identify functional elements present in the 809-bp gF sequence by constructing a library of gF deletion viruses and using the nuclear run-on assay to define sequences required for termination. Two sets of plasmid exonuclease III deletions of the gF element consisting of 3'→5' deletions and 5'→3' deletions were made. From each plasmid deletion library, specific clones were selected, sequenced, and constructed into recombinant Ad vectors as described in Materials and Methods. For all termination assays carried out in this study, we have averaged the termination measurements obtained from at least three separate experiments for each virus. All termination measurements were normalized against values for the control virus, sub360 gDE (as in Fig. 1).

The data generated from gF deletion viruses indicate the complexity underlying termination in gF. Figure 3A shows

FIG. 5. Characterization of the 69-bp termination element $F_{442-511}$. (A) Nucleotide sequence of gF surrounding and including the 69-bp element $F_{442-511}$. The 69-bp sequence is presented in bold type. (B) Nuclear run-on characterization of the minimal termination construct sub360 gDEF₄₄₂₋₅₁₁. Assays were carried out exactly as described in the text.

diagrams of four viruses which have distal regions of gF deleted. Percent termination represents the average of three separate experiments (values shown in parentheses in Fig. 3A). Quantitation was done by using a Molecular Dynamics PhosphorImager. Slot blots are shown in Fig. 3B. The 3'→5' deletion virus sub360 gDEFΔ3' 634 causes a level of termination (88%) equivalent to that seen with sub360 gDEF (90%). In sub360 gDEFΔ3' 634, the distance between the globin poly(A) cleavage site and the 1A3 segment is 1,137 bp. Each successive virus which contains gF deletions from the 3' end demonstrates a decrease in termination efficiency. Virus sub360 gDEFΔ3' 404, which contains sequence equivalent to that of subclone gF_A (Fig. 1A), prevents 70% of the RNA polymerase molecules from transcribing into the 1A3 segment. Removing an additional 200 bp of sequence (sub360 gDEFΔ3' 203) yields a lower yet still significant level of transcription termination (57%).

The virus constructs containing the gF 5'→3' deletions complement the gF 3'→5' deletions in defining termination elements present in the gF element. It should be noted that unlike the 3'→5' deletions, the 5'→3' deletions alter the spacing between the poly(A) site and the termination elements. This alteration may affect the overall efficiency of termination by these distal elements. Deletion of the first 420 bp of the gF sequence had little impact on termination function (82%), as seen in virus sub360 gDEFΔ5' 420 (Fig. 4). Combining this result with those taken from Fig. 3, we conclude that multiple termination elements are present within the gF sequence. The next deletion construct, sub360 gDEFΔ5' 519, which contains the distal 290 bp of gF, demonstrated a sharp decline in termination function (to 28%). Removal of additional gF sequence (virus sub360 gDEF Δ5' 575) reduces termination to 11%. When we assayed the next deletion virus, sub360 gDEFΔ5' 619, we observed that termination function increased to 37%. Our best explanation for this result is that we have revealed a minor cryptic termination sequence which is inactive in

sub360 gDEFΔ5' 575 and sub360 gDEFΔ5' 519, possibly as a result of altered local RNA or DNA secondary structure.

The threefold difference between sub360 gDEFΔ5' 420 and sub360 gDEFΔ5' 519 identifies a segment of gF which may contain significant termination activity. We used PCR to generate a globin fragment corresponding to globin $F_{442-511}$ (Fig. 5A). This fragment was subcloned into the parent plasmid pMLP6 gDE and used to construct sub360 gDEF₄₄₂₋₅₁₁. Having verified the presence of the $F_{442-511}$ fragment in the final virus construct, we used the nuclear run-on assay test whether this element could induce a strong block to elongation (Fig. 5B). Construct sub360 gDEF₄₄₂₋₅₁₁ reproducibly yields an efficient termination function. The average of three assays gave 64% termination. This 69-bp gF subclone is the smallest functional termination sequence that we have identified. The sequence is 70% purine rich (Fig. 4) and, with the exception of a TATA-like sequence, does not contain a DNA binding site for transcription factors present in the Faisst and Meyer compilation (12). Using the RNA fold program (14, 45; Genetics Computer Group UNIX-7 version), a very weak ($\Delta G = -3.5$ kcal [ca. -14.6 kJ]) stem-loop configuration can be generated from the $F_{442-511}$ sequence, which is immediately followed by the A-rich sequence CAAGAAAAAATAC. Interestingly, the AGAA AAAAATA sequence was identified in earlier studies as a consensus element in the 3' region of the mouse β^{maj} globin gene and in the human A_γ and G_γ globin genes (6).

Multiple elements from F_{1-809} possess termination function.

The results from the deletion constructs indicate the presence of multiple termination elements which contribute to the overall efficiency of gF termination activity. To verify this conclusion, we used four constructs, each containing approximately 200 bp of gF sequence, in the nuclear run-on termination assay (Fig. 6). The constructs chosen for this assay divide the entire 809-bp gF sequence into quarters. Three of the four virus constructs demonstrate termination function in the 55- to 65% range. The construct containing

the distal quarter of gF showed a low level of activity as previously indicated. Although we have demonstrated independent function present in the three distinct elements of gF, it is difficult to assess from these studies how effective each element is when they are presented as a contiguous unit. The run-on assays of sub360 gDEF (Fig. 1) led us to believe that an element such as F₄₄₂₋₅₁₁ would act as the dominant termination sequence in the entire gF domain. It is apparent from these studies that a single termination site does not account for the 90% level of termination seen in sub360 gDEF. When we combine the accumulated termination function found in each of the elements, there is a strong correlation with the function measured in the entire gF region. An RNA polymerase which has read through the first site will be stopped by one of the downstream sites.

A required spacing between the termination element and the poly(A) site signal sequence. Coupling between the 3' processing event and downstream termination elements has been shown in several RNA Pol II transcription units (7, 26, 42). When a nonfunctional polyadenylation signal is present, RNA Pol II is capable of transcribing through termination elements such as gF (26). This observation raises the question of whether or not the termination element by itself induces a transient pause or a block to RNA polymerase elongation. We reasoned that if the termination element independently exerts a block to elongation, then the distance between the poly(A) signal sequence and the termination element should not effect termination efficiency. On the other hand, if there is a comparatively short block to elongation, or if recognition of the termination element absolutely depends upon prior recognition and/or cleavage at the 3' processing site, altering the spacing between the two required elements should affect termination efficiency.

We constructed virus sub360 gDEFΔ5' 420, which has the previously described termination element positioned 150 nucleotides downstream of the globin poly(A) cleavage site (Fig. 7). When termination was measured by nuclear run-on assays (Fig. 7B), the termination efficiency was 2.5%, comparable to that seen with sub360 gD or sub360 gDE (Fig. 7C). The same gF sequence present in sub360 gDEFΔ5' 420 yields 82% termination. Thus, the most likely explanation for the loss of termination function in virus sub360 gDΔ5' 420 is the reduced spacing between the poly(A) signal sequence and the termination regions present in FΔ5' 420.

To demonstrate a spacing requirement, we placed this element downstream of a sequence taken from the bacterial CAT gene in virus sub360 gD CAT FΔ5' 430. The presence of the 624-bp CAT spacer between the globin poly(A) site and the gF termination sequence resulted in the recovery of termination efficiency to 70% (Fig. 7B). These experiments indicate that there is a minimum distance between the poly(A) cleavage site and the 3' termination element which is required for recognition of termination sequences by RNA Pol II. Although the gE sequence can be replaced by the CAT spacer, we have noticed that invariably when the gE element is present, a slightly more efficient termination event occurs. This may be appreciated by comparing the termination efficiency in sub360 gDEFΔ5' 420 with that in sub360 gD CAT FΔ5' 420. Further experiments will be required to determine whether the gE sequence is contributing something more than a spacing function to the termination event.

DISCUSSION

Progress in understanding the molecular basis of RNA Pol II termination has been hampered by the technical difficul-

FIG. 6. Termination efficiency for quarter segments of the gF region. (A) Schematic of terminator constructs which dissect the gF region into quarters. Constructs sub360 gDEFΔ3' 203 and sub360 gDEFΔ5' 619 are depicted in Fig. 3A and 4A, respectively. Constructs sub360 gDEF₁₈₅₋₄₀₄ and sub360 gDEF₄₄₂₋₆₃₄ were made by use of endonuclease restriction sites combined with existing deletion constructs described in detail in Materials and Methods. Percent termination is indicated as in Fig. 3A. (B) Representative nuclear run-on assays for the virus constructs depicted in panel A. Experimental conditions were as described in the legend to Fig. 1 and Materials and Methods.

ties encountered in characterizing inherently low levels of unstable nuclear RNA. In this study, we have used recombinant Ad constructs as reporter vectors for our characterization of the mouse β^{major} globin 3' termination elements. We have used the nuclear run-on assay to measure RNA Pol II density in regions upstream and downstream of an inserted putative termination element. This assay does not indicate direct RNA polymerase displacement from the template, nor does it measure changes in RNA polymerase elongation rate. Taking these caveats into consideration, our use of this system has provided new insight into basic features of RNA Pol II transcription termination: (i) identification of a 69-bp functional termination element, (ii) the presence of multiple functional elements in the termination region which are additive in their effect on RNA polymerase elongation, and (iii) determination that a minimum distance is required

FIG. 7. Evidence that altered spacing between the globin poly(A) site and termination element affects termination efficiency. (A) Schematic of virus constructs which differ in the position of the F₄₂₀₋₈₀₉ termination element relative to the globin poly(A) site. sub360 gDE, sub360 gD, and sub360 gDCAT are control virus constructs. MLTCR, Ad major late transcription control region; CAT, a DNA element excised from the bacterial CAT gene as described in Materials and Methods. All viruses were constructed as previously described. (B) Autoradiograph of nuclear run-on RNA hybridized to the indicated slot blot grid. All run-on assays were carried out as described in the text. (C) Graph of average termination efficiencies for each virus construct indicated in panel A. Average termination efficiency was calculated as described in the legend to Fig. 1.

between the poly(A) signal sequence and the termination sequence for termination to occur.

Of these observations, the most intriguing from a mechanistic perspective is the spatial requirement between the poly(A) site and termination elements. A spacing requirement of approximately 500 bp between the poly(A) site and the distal termination region imposes several conditions on the termination event. First, simple transcription through the poly(A) site by RNA polymerase is not sufficient to create a termination-competent transcription complex. The notion that a termination factor directly loads onto, or an antitermination factor dissociates from, RNA Pol II at the processing site is not supported by this observation. We also know from the required spacing of elements that if the termination elements arrest Pol II elongation, independently of poly(A) site recognition, the duration of arrest is short compared with the amount of time required to recognize the poly(A)

site and transmit the information to the Pol II elongation complex. We feel that the data indicate that the nascent RNA strand is altered at the poly(A) site and that this alteration is in some way transmitted to the transcription complex, creating a termination-competent complex. Formation of a processing complex and subsequent cleavage of the pre-mRNA are the most logical steps under consideration.

Placing a spacer sequence between the poly(A) site and the termination element may allow enough time for the processing complex to form before the polymerase has read through the termination site. Indirect support for the importance of a spacer between the poly(A) site and the termination region can be taken from previous studies. In the majority of endogenous transcription units that have been characterized by the nuclear run-on assay, the termination region is positioned 1 kb or more downstream of the poly(A)

signal sequence (6, 19, 24, 28, 34, 37). A second form of experimental evidence that supports coupling between these elements is found in the work of Lanoix and Acheson (22). When they replaced the inefficient polyomavirus polyadenylation site with the stronger rabbit β globin poly(A) site, they observed a decrease in read-through transcription past the polyomavirus promoter, implying that efficient processing was accompanied by an increase in termination efficiency. Work in our laboratory has also shown that a weak poly(A) site results in a loss of termination efficiency (10a).

The work presented in this report demonstrates the first example of RNA Pol II termination resulting from multiple termination elements acting in concert. A population of termination sites such as those found in the gF region may be required to ensure an efficient termination event. Use of redundant elements which probably rely on some form of RNA or DNA secondary structure to terminate transcription has proven to be an effective strategy found in procaryotic transcription units, whereby a combination of rho-dependent and rho-independent termination elements can be used to efficiently arrest RNA polymerase (23, 31, 38, 43; discussed in reference 15). The other example of redundant termination elements is found in the mouse ribosomal genes transcribed by RNA Pol I (18). In contrast to the β^{maj} 3' termination region, the Pol I termination element is a discrete DNA sequence which binds a termination factor (a mechanism more comparable to the Pol II CCAAT terminator). Since ours is the first characterization of such an extensive Pol II termination region, it will be interesting to determine whether this type of a termination region is characteristic of mammalian transcription units.

At the simplest level, the block to elongation may be the result of polymerase interactions with RNA or DNA secondary structures. In contrast to what has been found for promoter complex-mediated termination, when we examined the 69bp β^{maj} globin termination element sequence for known transcription factor binding sites, we found only a TATA-like motif. Although we feel that there is a sequence-specific requirement to the gF termination elements, an obvious consensus sequence is not immediately apparent. A homology search between the 69-bp element and the sequence present in the proximal gF segments showed relatively minor regions of homology which arise out of the general AT richness of the entire gF sequence. As noted earlier, the sequence AGAAAAAATA has been found in other globin 3' regions. Of other elements described as Pol II terminators, one has primary sequence compatibility with the globin F₄₄₂₋₅₁₁ sequence. Enriquez-Harris et al. (11) have mapped a pause/termination element in the human $\alpha 2$ globin gene, 350 bp 3' to the $\alpha 2$ poly(A) site. This sequence, by virtue of its A-rich nature, is similar to the F₄₄₂₋₅₁₁ sequence. There are regions within gF which are T rich as well. In the human gastrin gene, a T-rich element has been identified as a termination element. It is found 3' to the gastrin poly(A) site and contains a U-rich region in the RNA coding strand (39). In contrast to termination identified with the β^{maj} globin element, the gastrin element is not functionally dependent on the use of an upstream poly(A) signal sequence. From analysis of the termination elements present in the globin region or those identified in other systems, it is becoming evident that a termination-competent RNA polymerase may respond to a variety of heterogeneous termination elements.

ACKNOWLEDGMENTS

We thank John Prescott and Stewart Shuman for critical reading of the manuscript.

This work was supported by NIH grant 5R01 GM41967 awarded to Erik Falck-Pedersen. JogiRaju Tantravahi was the recipient of a Cornell M.D./Ph.D. fellowship.

REFERENCES

- Birchmeier, C., D. Schumperli, G. Sconzo, and M. L. Birnstiel. 1984. 3' editing of mRNAs: sequence requirements and involvement of a 60 nucleotide RNA in maturation of histone mRNA precursors. *Proc. Natl. Acad. Sci. USA* **81**:1057-1061.
- Birnstiel, M. L., M. Busslinger, and K. Strub. 1985. Transcription termination and 3' processing: the end is in site! *Cell* **41**:349-359.
- 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.
- Chomczynski, P., and N. Sacchi. 1987. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extractions. *Anal. Biochem.* **162**:156-159.
- Ciliberto, G., N. Dathan, R. Frank, L. Philipson, and I. W. Mattaj. 1986. Formation of the 3' end on U snRNAs requires at least three sequence elements. *EMBO J.* **5**:2931-2937.
- 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.
- 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.
- 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.
- 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.
- Cross, F. R., and J. E. Darnell, Jr. 1983. Cycloheximide stimulates early adenovirus transcription if early gene expression is allowed before treatment. *J. Virol.* **45**:683-692.
- Edwards-Gilbert, G., and E. Falck-Pedersen. Unpublished data.
- 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.
- Faisst, S., and S. Meyer. 1992. Compilation of vertebrate-encoded transcription factors. *Nucleic Acids Res.* **20**:3-26.
- 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.
- Freir, S. M., R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Nelson, and D. H. Turner. 1986. Improved free energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. USA* **83**:9373-9377.
- Friedman, D. I., M. J. Imperiale, and S. Adhya. 1987. RNA 3' end formation in the control of gene expression. *Annu. Rev. Genet.* **21**:453-488.
- Georgiev, O., and M. L. Birnstiel. 1985. The conserved CAA GAAAGA spacer sequence is an essential element for the formation of 3' termini of the sea urchin H3 histone mRNA by RNA processing. *EMBO J.* **4**:481-489.
- 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.
- Grummt, I., U. Maier, A. Ohrlain, N. Hassouna, and J. P. Bachellerie. 1985. Transcription of mouse rDNA terminates downstream of the 3' end of 28S RNA and involves the interaction of factors with repeated sequences in the 3' spacer. *Cell* **43**:801-810.
- 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.
- Hernandez, N. 1985. Formation of the 3' end of U1 snRNA is

- directed by a conserved sequence located downstream of the coding region. *EMBO J.* **4**:1827–1837.
21. Kunkel, G. R., and T. Pederson. 1985. Transcription boundaries of U1 small nuclear RNA. *Mol. Cell. Biol.* **5**:2332–2340.
 22. 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.
 23. Lau, L. F., J. W. Roberts, and R. Wu. 1983. RNA polymerase pausing and transcript release at the λt_{R1} terminator in vitro. *J. Biol. Chem.* **258**:9391–9397.
 24. LeMeur, M. A., B. Galliot, and P. Gerlinger. 1984. Termination of the ovalbumin gene transcription. *EMBO J.* **3**:2779–2786.
 25. Lobo, S. M., and W. F. Marzluff. 1987. Synthesis of U1 RNA in isolated mouse cell nuclei: initiation and 3' end formation. *Mol. Cell. Biol.* **7**:4290–4296.
 26. 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.
 27. 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.
 28. 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.
 29. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 30. Manley, J. L. 1988. Polyadenylation of mRNA precursors. *Biochim. Biophys. Acta* **950**:1–12.
 31. Morgan, W. D., D. G. Bear, and P. H. von Hippel. 1983. Rho-dependent termination of transcription: identification and characterization of termination sites for transcription from the bacteriophage lambda P_R promoter. *J. Biol. Chem.* **258**:9565–9574.
 32. Mowry, K. L., R. Oh, and J. A. Steitz. 1989. Each of the conserved sequence elements flanking the cleavage site of mammalian histone pre-mRNA has a distinct role in the 3' end processing reaction. *Mol. Cell. Biol.* **9**:3105–3108.
 33. Nevins, J. R. 1981. Definition and mapping of adenovirus 2 nuclear transcription. *Methods Enzymol.* **70**:768–784.
 34. Pribyl, T. M., and H. G. Martinson. 1988. Transcription termination at the chicken β -H-globin gene. *Mol. Cell. Biol.* **8**:5369–5377.
 35. Proudfoot, N. 1991. Poly(A) signals. *Cell* **64**:671–674.
 36. Proudfoot, N. J. 1989. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem. Sci.* **14**:105–110.
 37. Rohrbach, M. L., J. E. Johnson III, M. D. James, and R. C. Hardison. 1985. Transcription unit of the rabbit beta-1 globin gene. *Mol. Cell. Biol.* **5**:147–160.
 38. Rosenthal, E. R., and J. M. Calvo. 1987. Transcription termination sites at the distal end of the leu operator of *Salmonella typhimurium*. *J. Mol. Biol.* **194**:443–552.
 39. 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.
 40. Vasserot, A. P., F. J. Schaufele, and M. L. Birnstiel. 1989. Conserved terminal hairpin sequences of histone mRNA precursors are not involved in duplex formation with U7 RNA but act as a target site for a distinct processing factor. *Proc. Natl. Acad. Sci. USA* **86**:4345–4349.
 41. Wahle, E., and W. Keller. 1992. The biochemistry of 3' end cleavage and polyadenylation of messenger RNA precursors. *Annu. Rev. Biochem.* **61**:419–440.
 42. 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.
 43. Wu, A. M., G. E. Christie, and T. Platt. 1981. Tandem termination sites in the tryptophan operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **78**:2913–2917.
 44. Yuo, C., M. Ares, Jr., and A. M. Weiner. 1985. Sequences required for 3' end formation of human U2 small nuclear RNA. *Cell* **42**:193–202.
 45. Zuker, M., and P. Stiegler. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**:133–148.