MOLECULAR AND CELLULAR BIOLOGY, Feb. 2011, p. 639–651 Vol. 31, No. 4
0270-7306/11/$12.00 doi:10.1128/MCB.00919-10

Poly(A) Signals Located near the 5′ End of Genes Are Silenced by a General Mechanism That Prevents Premature 3′-End Processing▽†‡

Jiannan Guo,[1] Matthew Garrett,[2] Gos Micklem,[2] and Saverio Brogna[1]*

School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom,[1] and Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, United Kingdom[2]

Received 9 August 2010/Returned for modification 28 September 2010/Accepted 29 November 2010

Poly(A) signals located at the 3′ end of eukaryotic genes drive cleavage and polyadenylation at the same end of pre-mRNA. Although these sequences are expected only at the 3′ end of genes, we found that strong poly(A) signals are also predicted within the 5′ untranslated regions (UTRs) of many *Drosophila melanogaster* mRNAs. Most of these 5′ poly(A) signals have little influence on the processing of the endogenous transcripts, but they are very active when placed at the 3′ end of reporter genes. In investigating these unexpected observations, we discovered that both these novel poly(A) signals and standard poly(A) signals become functionally silent when they are positioned close to transcription start sites in either *Drosophila* or human cells. This indicates that the stage when the poly(A) signal emerges from the polymerase II (Pol II) transcription complex determines whether a putative poly(A) signal is recognized as functional. The data suggest that this mechanism, which probably prevents cryptic poly(A) signals from causing premature transcription termination, depends on low Ser2 phosphorylation of the C-terminal domain of Pol II and inefficient recruitment of processing factors.

The 3′ ends of mRNAs are generated by a cleavage/polyadenylation reaction that is coupled to transcription (29). A complex multiprotein machine, made up of about 85 proteins in human cells, carries out the reaction (33). Several of these proteins also function in transcription, splicing, and other mRNA processing events and may mediate the cross talk that occurs *in vivo* between these processes (33). The coupling between transcription and 3′ polyadenylation is primarily mediated by interactions between core components of the 3′-end processing complex and the C-terminal domain (CTD) of the largest subunit of polymerase II (Pol II) (1, 20). The CTD consists of several YSPTSPS 7-amino-acid repeats that can bind distinct proteins depending on which serine residues and how many repeats are phosphorylated (7). Phosphorylation of Ser-5 on many CTD repeats signals the transition from initiation to early transcription elongation, whereas Ser-2 hyperphosphorylation, which peaks at the 3′ end of genes, is required for later stages of elongation, efficient 3′-end processing, and transcription termination (7, 31).

The key trigger for 3′-end processing is the poly(A) signal located at the 3′ end of the nascent pre-mRNA. In mammalian systems, this signal consists of a bipartite sequence made of an AAUAAA hexamer located 10 to 30 nucleotides (nt) upstream of the cleavage and polyadenylation site and a GU-rich region downstream (13, 39). The poly(A) signal is recognized cotranscriptionally by components of the cleavage and polyadenylation complex, some of which are probably preloaded on the Pol II CTD (1, 17, 20, 22). The cleavage and polyadenylation specificity factor (CPSF) binds the AAUAAA hexamer, while the cleavage stimulation factor (CstF) binds the downstream GU-rich element (13). Not only are recognition of the poly(A) signal and 3′-end processing essential for polyadenylation, but they are also key determinants for Pol II termination (31). One key protein that might be involved in this linkage with termination is Pcf11, a conserved component of the cleavage and polyadenylation complex which *in vitro* bridges the CTD to the RNA at Pol II pause sites and dismantles the transcription elongation complex, leading to transcription termination (14, 32, 41, 43, 44).

Poly(A) signals consist of short redundant sequences that are not restricted to mapped gene 3′ ends. This is particularly apparent in the yeast *Saccharomyces cerevisiae*, in which the poly(A) signals are made of poorly conserved sequence motifs flanking the cleavage site (28). Therefore, these sequences cannot fully account for the specificity of 3′-end processing. Putative poly(A) signals are frequently found within AU-rich sequences, such as introns and intergenic regions (21, 37). In addition, bioinformatical studies have determined that the sequence motifs that make the poly(A) signal are, contrary to earlier predictions, not well conserved, even in mammalian genes. In human expressed sequence tag (EST) data sets, only 53% of 3′ untranslated regions (UTRs) include the AAUAAA hexamer, with polyadenylation events also associated with up to 12 other variants of the hexamer, such as AUUAAA, which is found in 17% of the transcript 3′ ends (36). Recently, it was reported that in a noncanonical poly(A) signal, which may represent one-third of human poly(A) signals, efficient 3′-end processing only requires a potent downstream sequence element (DSE) and an A-rich upstream sequence, but the closest match of the AAUAAA found in this gene 3′-UTR has little effect on the reaction (25).

Because polyadenylation is coupled to transcription, cryptic poly(A) signals early in the gene sequence have the potential to

* Corresponding author. Mailing address: School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom. Phone: 44 121 414 5569. Fax: 44 121 414 5925. E-mail: s.brogna@bham.ac.uk.

† Supplemental material for this article may be found at http://mcb.asm.org/.

▽ Published ahead of print on 6 December 2010.

‡ The authors have paid a fee to allow immediate free access to this article.

terminate transcription prematurely, implying that evolution should have selected against the presence of putative poly(A) signals in places other than the 3′ end of the transcription units. Surprisingly, we report here that the 5′-UTRs of over 3,000 *Drosophila melanogaster* transcripts encode sequences that look like strong poly(A) signals. Moreover, we have demonstrated that several of these can function as strong poly(A) signals when they are placed at the 3′ end of reporter genes. However, they show little or no activity when placed close to transcription start sites (TSS). We also observed this effect, in both *Drosophila* and human cells, with strong canonical poly(A) signals, like that derived from simian virus 40 (SV40) late transcription unit. The data suggest the existence of a general mechanism that prevents poly(A) signals from initiating 3′-end processing if they are situated near the 5′ end of the pre-mRNA.

MATERIALS AND METHODS

Bioinformatical poly(A) signal prediction. The programs Polya_svm and Polyadq were used for the identification of putative polyadenylation sites (10, 34). Polyadq scans sequences for AATAAA or ATTAAA and downstream elements which are typical of mammalian poly(A) signals (34). Polya_svm searches for poly(A) signals by using a window-based scoring scheme to evaluate the fitness of 15 *cis* elements identified from known human poly(A) signals (10). The whole data set of *D. melanogaster* 3′-UTRs and 5′-UTRs is available in Flybase (ftp://flybase.net/genomes/, under folder Drosophila_melanogaster/; release dmel_r4.3_20060303). The UTR sequences were extended by 150 nt at the 3′ ends with the corresponding genomic sequence, because both programs rely on downstream elements. The E-value represents the probability of being a poly(A) signal, and the higher the probability, the lower the E-value.

Construction of reporter genes. The *Adh-Luc* plasmid constructs were generated in pAc5.1/V5-His A (Invitrogen). The 5′-UTR sequences were PCR amplified from adult fly genomic DNA and cloned into an AvrII site of the previously described *Adh-Luc* reporter (30). Poly(A) signals were inserted at positions P1, P2, and P3 of the *Adh* coding region by cloning them into a BglII site previously introduced at codon positions 64, 126, and 203 by ligation of BglII-flanked fragments followed by PCR of the joined product. The *lacZ* reporter was generated by inserting a PCR fragment of the complete *lacZ* coding region (amplified from plasmid pAc5.1/V5-His/LacZ [Invitrogen]) in the KpnI and XhoI sites of pAc5.1. In the *lacZ* reporter the bovine growth hormone gene (BGH) poly(A) signal (PCR amplified from pCDNA3.1 [Invitrogen]) was inserted into the KpnI site (position P1) or into an AvrII site introduced at codon 49 (P2) or 149 (P3) of *lacZ*. The firefly luciferase (*Luc*) reporter was PCR amplified from the *Adh-Luc* reporter (shown below in Fig. 2) and inserted into the KpnI and XhoI sites in pAc5.1. In the *Luc* reporters the UTR-4 sequence was cloned into the KpnI site (position P1) or into a previously introduced Avr II site at codon 64 (P2) or 203 (P3). Cytomegalovirus (CMV)-regulated and ecodysone-inducible reporters (ERE) carry the same gene cassette as the *Adh-Luc* constructs (with UTR-9 at P1, P2, or P3), except that the *Luc* gene is truncated by deleting the EcoRI-XhoI fragment at the 3′ end of the coding region.

Cell culture, transfection, and RNA purification. *Drosophila* S2 cells were grown in Insect XPRESS medium with 4% fetal bovine serum and 1× penicillin-streptomycin-glutamine mix (Lonza), at 27°C with no CO_2. Transfection of plasmids was performed with dimethyldioctadecylammonium bromide (DDAB) as previously described (30). Typically, cells were transfected in six-well plates and grown for 24 to 48 before harvesting. HEK 293T cells were cultured as adherent cells in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and L-glutamine (Lonza). Transfections into 293T cells were carried out using FuGENE HD transfection reagent (Roche) following the manufacturer's instructions.

Total RNA was extracted from transfected cells by using TRI reagent (Molecular Research Center) according to the manufacturer's instructions. Poly(A) selection was carried out using the PolyATract kit (Promega) according to the manufacturer's instructions.

RNA analysis. For Northern analysis, typically 5 μg of total RNA was separated on a 1% agarose gel in the presence of formaldehyde and blotted onto nylon membranes (Hybond-N; Amersham) by capillary transfer followed by UV light cross-linking. The membranes were hybridized with PCR-amplified probes labeled by random hexamer priming using [^{32}P]dCTP. Adh, Luc, and Egfp probes were the same as those previously described (30). The UTR-9 probe was PCR amplified from genomic DNA with the primers 5′-GGGCCTAGGTAACAATGAAGTTTAAGCGCA-3′ (UTR-9_FW) and 5′-GGGCCTAGGTAAATTTGGTTTTCGGTGTTC-3′ (UTR-9_RV). The UTR-4 probe was similarly amplified with the primers 5′-GGGCCTAGGTAAAATCGGTGCGGTTCAGTT-3′ (UTR-4_FW) and 5′-GGGCCTAGGTAACGCTCAAATCTGATCGCA-3′ (UTR-4_RV).

Real-time PCRs were performed on reverse transcription (RT) products by using SYBR premix *Ex Taq* II (TAKARA) on an ABI Prism 7000 real-time PCR system. Reverse transcription was carried out with SuperScript III reverse transcriptase (Invitrogen) from 5 μg of total RNA and oligo(dT). The following mRNA-specific primers were used: for Rrp6 mRNA, 5′-GCCCTTTACCTAAGCTATCC-3′ (Rrp6_Val_FW) and 5′-ACCATTAGTTCGGTTTCTGC-3′ (Rrp6_Val_RV); for Pcf11, 5′-ATCGCTATGTTCGCAATGGA-3′ (Pcf11_Val_FW) and 5′-TCGTGGGATTTGAGTTGAGC-3′ (Pcf11_Val_RV); for Cdk9, 5′-CTCCAGCAGCCTTCGGGGTCG-3′ (Cdk9_Val_FW) and 5′-GCCAAGCCAAAGTCAGCCAGC-3′ (Cdk9_Val_RV); for CycT, 5′-AGCCAGTGCCTCAGTCTCAGC-3′ (CycT_Val_FW) and 5′-ATGGACACAGACTCTCCTTTA-3′ (CycT_Val_RV); for Fcp1, 5′-CGCTACAGAAGCACCCAAAG-3′ (Fcp1_Val_FW) and 5′-ACCGCCACTAGATGCGTTAT-3′ (Fcp1_Val_RV). Quantitations were obtained with the ABI Prism 7000 SDS software. Levels of mRNA were normalized to the level of Rpl32 mRNA, which was amplified by primers 5′-CGCCGCTTCAAGGGACAGTAT-3′ (Rpl32_Val_FW) and 5′-TCTTGAGAACGCAGGCGACCG-3′ (Rpl32_Val_RV).

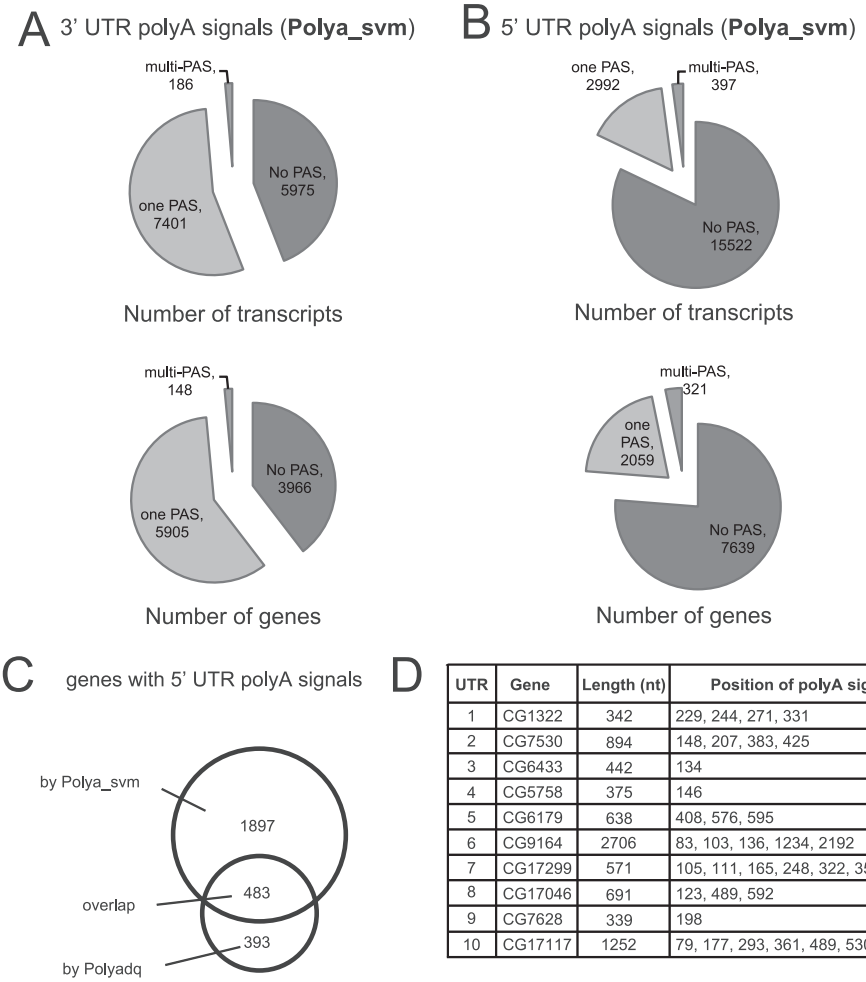
Circular RT-PCR (c-RT-PCR) was carried out as previously described (5). Purified PCR products were subcloned into a home-made plasmid T-vector (18) and sequenced.

The Adaptor RT-PCR assay has been previously described (23). For the reverse transcription, an adaptor-oligo(dT) was used: 5′-TAGAATTCAGCATTCGCTTCTTTTTTTTTTTTTTTT(C/G/A)-3′. In the next PCR, a gene-specific sense and adaptor-specific primers were used. Extension time for each PCR cycle was 1 min, which should amplify short transcripts terminated within the 5′-UTRs but limit the amplification of the full-length mRNAs. After two rounds of PCR using nested primers annealed to the 5′ ends of the 5′-UTRs, all visible bands on 2% agarose gels were purified, cloned into a plasmid T-vector, and sequenced.

RNAi. RNA interference (RNAi) knockdowns were achieved by incubating S2 cells with gene-specific double-strand RNA (dsRNA) as previously described (8, 12). The dsRNA probes were selected from the previously tested fragments listed in GenomeRNAi (http://rnai2.dkfz.de/GenomeRNAi/) (15). DNA fragments were PCR amplified from S2 cell genomic DNA. All PCR primers carried a T7 promoter sequence (5′-TAATACGACTCACTATAGGGA-3′) at the 5′ end, and the 3′ end corresponded to exonic regions of the gene of interest. For Rrp6 dsRNA, the 3′ ends were 5′-GCCTGCTGAACTTTTTCGAC-3′ (Rrp6_ds_FW) and 5′-AGCCGACACAAGAAGAGGAA-3′ (Rrp6_ds_RV). For CPSF-160 dsRNA, the 3′ ends were 5′-TCGGCTGGTTAACCGTAAAG-3′ (CPSF_ds_FW) and 5′-GTTCTGGAGCTAAGGCATCG-3′ (CPSF_ds_RV). For Pcf11 dsRNA, the 3′ ends were 5′-GCGAAGTGGCTTTCCTAGTG-3′ (Pcf11_ds_FW) and 5′-TCTCCCAAAAGGAATGATGC-3′ (Pcf11_ds_RV). For Cdk9 dsRNA, the 3′ ends were 5′-CATACTGTTGTCCTGGGGCT-3′ (Cdk9_ds_FW) and 5′-CAGCTATGCGGCTCCTTTAC-3′ (Cdk9_ds_RV). For CycT dsRNA, the 3′ ends were 5′-CATGGATGGTGGTACAGCAG-3′ (CycT_ds_FW) and 5′-AACTCCGATGACCAGTTTGG-3′ (CycT_ds_RV). For Fcp1 dsRNA, the 3′ ends were 5′-CCGAATCTTCGGAACGATAA-3′ (Fcp1_ds_FW) and 5′-CACCAGATGCTGAAAAAGCA-3′ (Fcp1_ds_RV). A LacZ dsRNA fragment was used as a control. The primer 3′ ends for LacZ dsRNA were 5′-CTGTCGTCGTCCCCTCAAAC-3′ (LacZ_ds_FW) and 5′-CGTTTCACCCTGCCATAAAG-3′ (LacZ_ds_RV), and the product was amplified from the pAc-LacZ plasmid. The dsRNAs were synthesized using the T7 RiboMAX Express RNAi system (Promega). A 7.5-μg aliquot of dsRNA was typically added to the cells (3-cm wheel; 1 ml of medium) just after transfection. Cells were incubated for 3 days before RNA purification.

RESULTS

Poly(A) signals are predicted in the 5′-UTR of many *Drosophila* transcripts. In a previous study, we accidentally discovered the presence of a functional poly(A) signal in the 5′-UTR of *D. melanogaster Ubx* mRNA (30). We sought to check whether poly(A) signals also exist in other 5′-UTRs. We ana-



D

UTR	Gene	Length (nt)	Position of polyA signals	E-value
1	CG1322	342	229, 244, 271, 331	0.070
2	CG7530	894	148, 207, 383, 425	0.017
3	CG6433	442	134	0.031
4	CG5758	375	146	0.440
5	CG6179	638	408, 576, 595	0.036
6	CG9164	2706	83, 103, 136, 1234, 2192	0.062
7	CG17299	571	105, 111, 165, 248, 322, 350, 384, 557	0.093
8	CG17046	691	123, 489, 592	0.036
9	CG7628	339	198	0.041
10	CG17117	1252	79, 177, 293, 361, 489, 530, 812	0.047

FIG. 1. Prediction of poly(A) signals in *Drosophila* 5′-UTRs. (A) Pie charts showing the proportion of 3′-UTRs with (light gray) or without (dark gray) poly(A) signals (PAS), as predicted by using the Polya_svm program; the smallest sector of the pie shows the portion of the 3′-UTR in which multiple poly(A) signals (multi-PAS) are predicted. The pie on top represents the number of transcripts with predicted poly(A) signals; the chart on the bottom shows the proportion of genes with at least one transcript with a predicted poly(A) signal. (B) Pie charts showing the proportions of transcripts (top) or genes (bottom) with at least one (light gray) or without and (dark gray) poly(A) signals in the 5′-UTR, as predicted with the Polya_svm program; the smallest sectors show transcripts/genes with predicted multi- poly(A) signals. (C) Venn diagram showing how many genes are predicted to have a transcript with at least a 5′-UTR poly(A) signal and how many of these are predicted by both the Polyadq and Polya_svm programs. (D) Information on the 10 5′-UTR sequences experimentally tested in this study.

lyzed all known *Drosophila* transcripts by using two programs previously developed for predicting mammalian poly(A) signals: Polya_svm and Polyadq (10, 34). Because the programs were developed for mammalian poly(A) signals, we first tested whether they could correctly identify poly(A) signals in the 3′-UTRs of known *Drosophila* transcripts. From the total collection of known *D. melanogaster* transcripts in Flybase, Polya_svm identified poly(A) sites in 7,587 of the 3′-UTRs (corresponding to 6,053 individual genes) of the total 13,562 annotated mRNAs (10,019 genes) (Fig. 1A); 186 of these 3′-UTRs (148 genes) have more than one poly(A) site predicted. The other program, Polyadq, identified fewer poly(A) signals within the known 3′-UTRs: 2,520 of the 3′-UTRs (2,036 genes) contain a single poly(A) signal and 242 (195 genes) contain multiple signals. Around 60% of the 3′-UTR signals found by Polyadq were also identified by Polya_svm. This initial analysis showed that both programs miss a large fraction of experimentally verified 3′-UTRs, yet both can effectively predict *Drosophila* poly(A) signals. Polya_svm is more sensitive than Polyadq, as previously shown with human sequences (10).

Next, we used these two programs to search for poly(A) signals in *Drosophila* 5′-UTRs. Of 18,911 annotated 5′-UTRs (including alternative transcripts), Polya_svm predicted poly(A) signals in 3,389 sequences (2,380 genes) (Fig. 1B). Of these 5′-UTRs, 397 (321 genes) had predicted multiple poly(A) signals. As expected, Polyadq identified poly(A) signals in fewer sequences: 1,101 of the 5′-UTRs (876 genes) with a single hit and 112 (94 genes) with multiple signals; 483 of these genes are predicted to contain 5′-UTR poly(A) signals by both programs (Fig. 1C) (lists reporting all the hits of both programs are shown in the supplemental material). In summary, this analysis clearly predicted that putative poly(A) signals are common in the 5′-UTRs of *Drosophila* transcripts. Furthermore, the number of putative poly(A) signals is likely

to be an underestimate, because both programs missed poly(A) signals in the 3′-UTRs of experimentally verified transcripts.

Experimental validation of poly(A) signals from *Drosophila* 5′-UTRs. Despite the strong bioinformatical prediction, a key issue is whether these 5′-UTR sequences can indeed function as poly(A) signals. To test for functionality, we selected 10 of the 5′-UTRs with putative strong signals: we called these UTR-1 to UTR-10 (Fig. 1D). These 5′-UTRs, ranging from 340 nt to 2,700 nt, were amplified from genomic DNA and inserted into the intergenic spacer of a dicistronic reporter. The dicistronic reporter was based on the *Adh-Adhr* dicistronic gene of *D. melanogaster* (5, 30); in this reporter we replaced *Adhr* with the coding region of firefly luciferase (*Luc*) (Fig. 2A). The expectation was that if the insert contained a functional poly(A) signal, then a monocistronic mRNA of the upstream reporter gene would be produced. If the putative poly(A) signal was nonfunctional or weakly functional, a longer dicistronic mRNA would be generated by use of the SV40 poly(A) signal downstream of the *Luc* gene. Two parallel sets of reporters were made, with either the genomic intron-containing *Adh* or the cDNA (Fig. 2A); similar reporters with the original *Adh* poly(A) signal serve as positive controls. The constructs were transfected into *Drosophila* S2 cells, and total RNA was isolated 24 to 48 h posttransfection and analyzed by Northern blotting with either *Adh*- or *Luc*-specific probes (Fig. 2B). In these experiments an enhanced green fluorescent protein (EGFP)-expressing plasmid was cotransfected, and the level of EGFP mRNA was used to normalize transfection variation.

The reporter with the endogenous *Adh* poly(A) signal produced, as expected, an abundant monocistronic mRNA and very little dicistronic mRNA (Fig. 2B, lanes 1 and 2 and 13 and 14): the dicistronic transcript only became clearly visible by probing the Northern blot with Luc-specific probes (Fig. 2B, top panels). In most assays, whether for the endogenous *Adh-Adhr* gene or for cDNA-derived reporters, the dicistronic transcript is not abundant (6, 30). Most of the reporters carrying the putative poly(A) signals from 5′-UTRs produced high levels of monocistronic *Adh* mRNA, confirming that these sequences function as poly(A) signals (Fig. 2B, lanes 3 to 12 and 15 to 24). The *Adh* mRNA abundance varied between reporters, indicating that the strength of the poly(A) signal varies between 5′-UTRs. In particular, sequence UTR-4 appeared to be stronger than the *Adh* poly(A) signal (Fig. 2B, lanes 9 and 10). As expected, high monocistronic transcript levels generally correlated with low dicistronic mRNA levels, and vice versa. For example, the reporters with UTR-1 and UTR-10 produced less *Adh* mRNA but more *Adh-Luc* mRNA (Fig. 2B, lanes 3 and 4 and lanes 23 and 24). In contrast, the UTR-4 reporter yielded more *Adh* mRNA but a just-visible *Adh-Luc* mRNA band (Fig. 2B, lanes 9 and 10).

The data also suggest that the presence of introns in *Adh* has relatively little influence on the usage of intergenic poly(A) signals. Sometimes the intron-containing reporter produced more *Adh* mRNA (UTR-3, UTR-6, and UTR-10) and sometimes it produced less *Adh* (UTR-2, UTR-5, and UTR-9). Overall, though, the intron-containing reporters tended to produce less dicistronic transcript, suggesting that the introns can enhance intergenic 3′-end processing.

As negative controls, we tested 5′-UTR sequences that are not predicted to contain poly(A) signals by using both Polyadq and Polya_svm (Neg-1, CG10192; Neg-2, CG2556, Neg-3, CG10808; Neg-4, CG7359; Neg-5, CG8171). We found that four of these sequences (Neg-2, Neg-3, Neg-4, and Neg-5) did not show poly(A) signal activity: they produced only trace amounts of *Adh* mRNA but relatively high levels of readthrough *Adh-Luc* mRNA compared to those produced by the reporters with the *Adh* poly(A) signal (Fig. 2C, lanes 3 to 6). In addition, deleting the intergenic spacer resulted in the loss of the *Adh* mRNA in both intron-containing and intronless reporters (Fig. 2D). Unexpectedly, the Neg-1 sequence produced a significant amount of *Adh* mRNA (Fig. 2C, lane 2); in this sequence there was no AAUAAA or AUUAAA hexamer, but there was a GAUAAA at position 200; this hexamer accounts for 1.75% of human and 1.16% of mouse poly(A) signals (36) and could explain the poly(A) activity we detected. Finding that one of the five negative-control sequences was also functionally active further suggested that the number of genes with poly(A) signals in the 5′-UTR might be higher than that predicted by the bioinformatics software.

To further characterize the 3′ end of mRNA produced by the reporters, we used a c-RT-PCR assay followed by cloning and sequencing of the PCR products to map the 3′ ends of the mRNAs (see Materials and Methods). We found that the monocistronic mRNA produced by the reporter with the *Adh* poly(A) signal is polyadenylated at the same position as the endogenous transcript, in both intron-containing and intronless reporters (Fig. 2E shows a map of observed polyadenylation sites). The size differences between spliced and nonspliced *Adh* mRNAs (Fig. 2B, compare odd and even lanes) are therefore solely due to the inclusion of the small 5′-UTR exon, which is only present in the genomic reporter (Fig. 2A). We also used c-RT-PCR to characterize the 3′ ends of the *Adh* mRNAs produced by the reporters with UTR-4 and UTR-6. We found that in both cases the 3′ ends were generated by cleavage just downstream of the AAUAAA sequence—20 to 30 nt after the hexamer within UTR-4 and 12 to 21 after that within UTR-6 (Fig. 2E). Sequencing indicated these mRNA have poly(A) tail lengths of up to 60 to 80 nt, similar to those of the endogenous *Adh* transcripts (5).

In summary, our experiments demonstrated that the putative poly(A) signals derived from the 5′-UTRs that we tested do serve as functional poly(A) signals in the manner predicted by the bioinformatical analyses: they drive 3′-end cleavage downstream of the AAUAAA hexamer and generate mRNAs with polyadenylated 3′ ends that are indistinguishable from those generated by a standard 3′-UTR-derived poly(A) signal.

The 5′-UTR poly(A) signals are silent in the endogenous genes. Given that the 5′-UTR poly(A) signals are functional when placed into reporter genes, the question was whether these signals are also used in flies during transcription of the endogenous genes. For the 10 genes we tested, we found no evidence of truncated transcripts with 3′ ends downstream of the predicted poly(A) signals in the extensive collection of EST databases available in Flybase. For CG17046 (UTR-8), we found a 3′ EST that ends in the 5′-UTR (GenBank accession number EC267859), but the 3′ end sequence did not coincide with any predicted poly(A) signal: the closest hexamer to the 3′ end is 107 nt upstream. In addition, the 10 5′-UTRs we tested

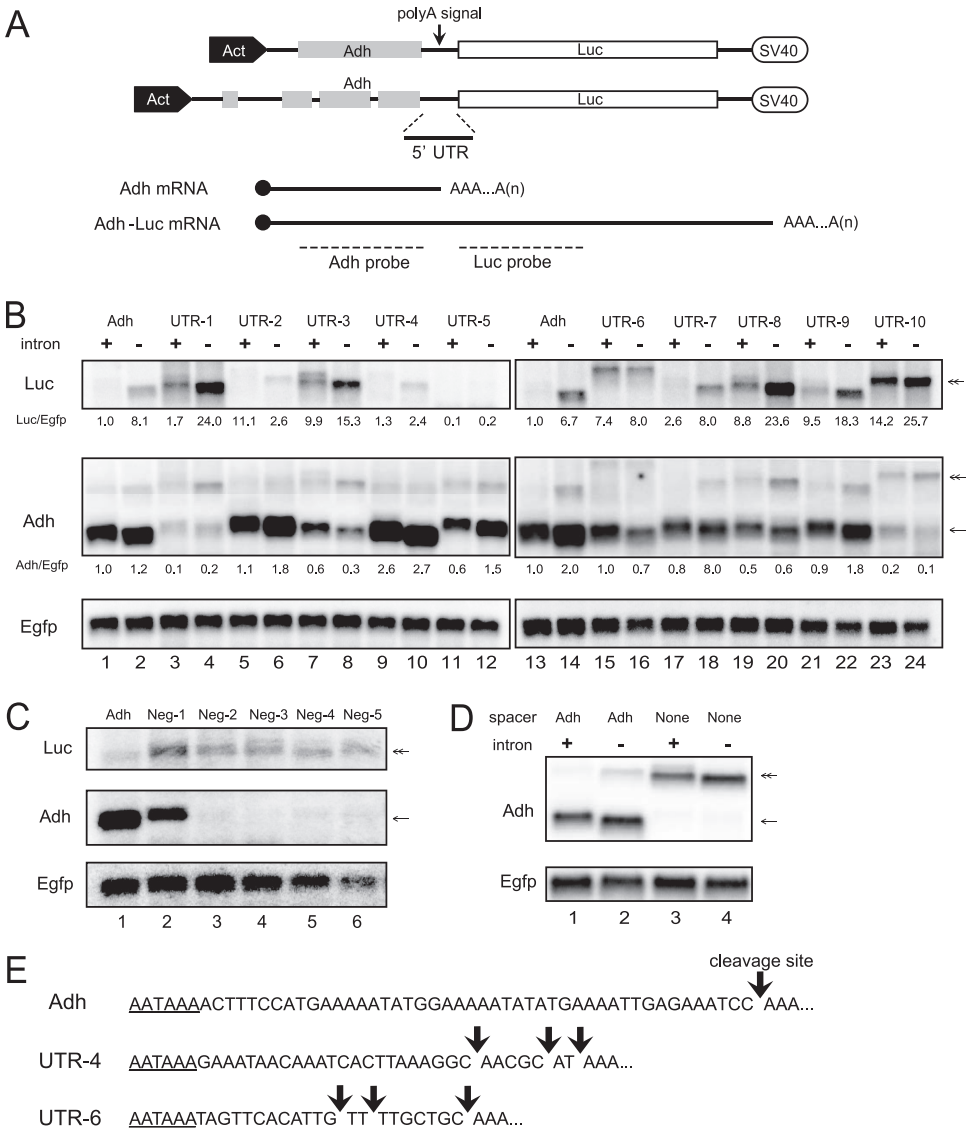


FIG. 2. 5′-UTR sequences contain functional poly(A) signals. (A) Schematics for the *Adh-Luc* dicistronic reporters. Boxes represent exons, and lines show UTRs and introns. Act stands for the *Drosophila* actin-5C promoter, and SV40 stands for the SV40 late poly(A) signal present in pAc5.1/V5-His (Invitrogen). The cDNA version of *Adh* (upper diagram) encodes the full *Drosophila Adh* open reading frame. The genomic version of *Adh* starts from the adult transcription start site and includes the 5′-UTR exon (6). The arrow is pointing at the *Adh* poly(A) signal in the intergenic spacer. Twenty-two reporters were constructed with the 10 5′-UTRs listed in Fig. 1D and with the *Adh-Adhr* spacer; half of the reporters carried the genomic *Adh* sequence, and the others carried the cDNA derivative. A schematic of the expected monocistronic and dicistronic transcripts is drawn below, with Northern blot probes indicated by dotted lines below the schematic. (B) Northern blots from analysis of total RNA from transfected S2 cells. Cells were cotransfected with a plasmid expressing EGFP to normalize for transfection variations. The *Adh-Luc* mRNAs were first detected with the *Luc*-specific probe (top panel), and then the filter was stripped and reprobed for *Adh* (middle panel). The single-arrowed line points to the *Adh* monocistronic mRNA, and doubled-arrowed lines indicate the readthrough dicistronic *Adh-Luc* transcript. Relative quantification of the signal intensities is reported below; bands intensities were normalized by dividing the relative intensity of the Egfp band (visualized with a phosphorimager and quantified with Quantity One program [Rio-Rad]). Intensities of the bands are relative to that of *Adh* (middle panel) or *Luc* (top) in lanes 1 or 13, respectively. (C) Results of experiments similar to that in panel B, with 5′-UTRs predicted not to carry poly(A) signals. (D) Results of experments similar to those in panels A and B, with no intergenic spacer between *Adh* and *Luc*. (E) Sequence of the 3′ end of the transcripts produced by the reporters with the poly(A) signals, as indicated (*Adh*, UTR-4, and UTR-6). Observed poly(A) sites are indicated by arrows and are based on sequencing of several clones (for *Adh*, $n = 3$; for UTR-4, $n = 4$; for UTR-6, $n = 4$).

were all from genes well expressed during the *Drosophila* life cycle, indicating that the presence of the 5′-UTR poly(A) signals does not necessarily inhibit transcription.

To further check whether these 5′-UTR signals are used at some stage of the life cycle, we performed an adaptor oligo(dT)–RT-PCR assay and searched for short transcripts with premature 3′ ends (see Fig. S1 in the supplemental material). We analyzed total RNA from S2 cells, embryos, first-instar larvae, second-instar larvae, and adult flies. After two rounds of PCR (totalling 50 to 60 cycles) with nested primers, all visible bands were purified, subcloned, and sequenced. We found that almost all PCR products were specific; however,

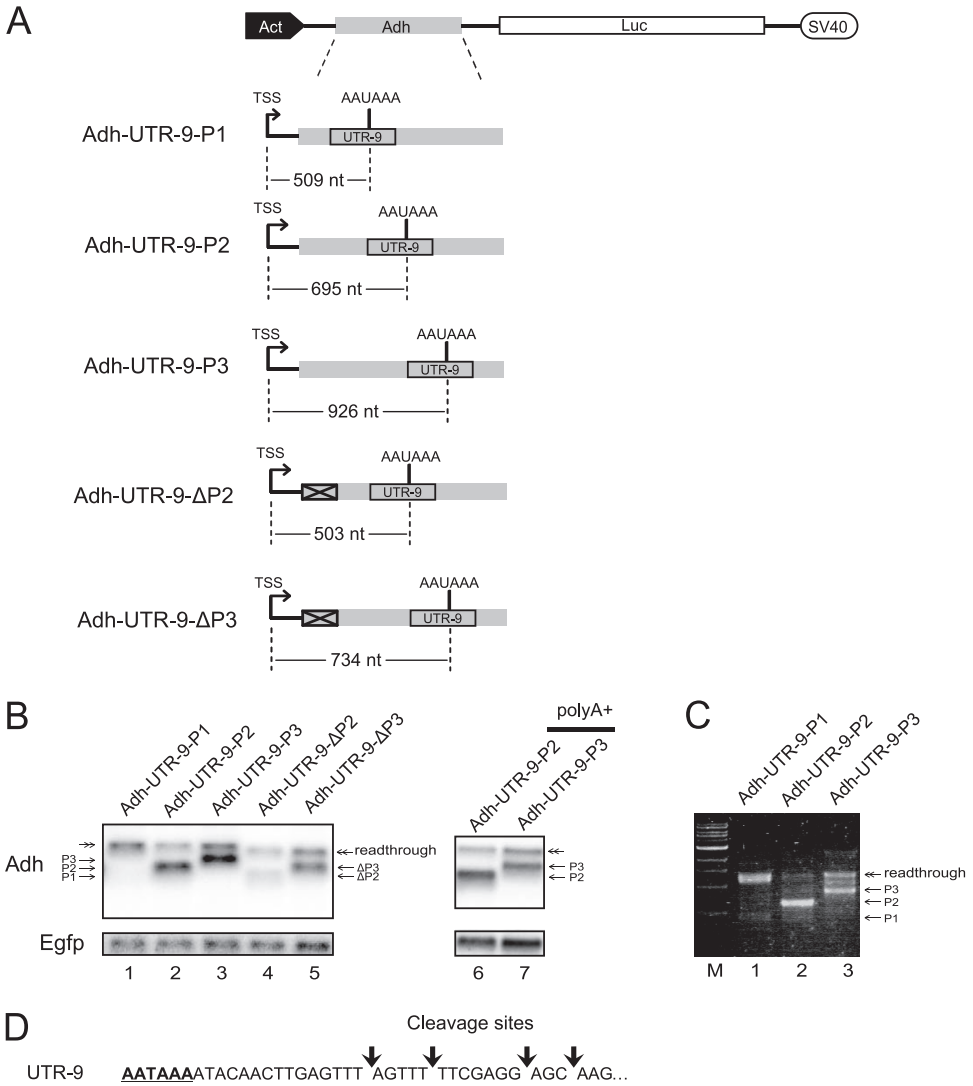


FIG. 3. Poly(A) signals become silent when close to the 5′ end. (A) Schematics of reporters with UTR-9 inserted at different positions in the *Adh* coding region. The distance from TSS to AAUAAA is indicated below each schematic. (B) Northern blots of total RNA and poly(A)$^+$ RNA of S2 cells transfected with the reporters shown in panel A; the probes used were those described for Fig. 2. (Left panel) The top band (with double-arrowed line) is the readthrough mRNA processed at the intergenic *Adh* poly(A) signal. Truncated transcripts processed at early poly(A) sites are indicated (P1, P2, and P3); ΔP2 and ΔP3 indicate mRNA derived by the deletion derivative lacking the initial region of *Adh*. (C) Agarose gel showing the DNA fragments produced by the adaptor RT-PCR assay of total RNA extracted from cells transfected with the indicated reporters. (D) Location of the poly(A) sites in the P1, P2, and P3 transcripts shown in panel C (based on sequencing of several clones of the P1, P2, and P3 RT-PCR fragments).

most products were generated by misannealing of the oligo(dT) at stretches of ~8 As in the 5′-UTRs. Of the 37 predicted 5′-UTR poly(A) signals, we found evidence for only three polyadenylated transcripts ending in the 5′-UTR: UTR-2 (of CG7530), UTR-5 (of CG6179), and UTR-8 (of CG17046) (see Fig. S1). One of these, UTR-8, could only be detected in adult flies. In summary: for most of the transcripts we analyzed, there was no evidence of truncated transcripts within the 5′-UTRs; for those that we found some evidence, they must be present at very low levels, and we could not rule out the possibility that these 3′ ends were from readthrough transcripts initiated at promoters of upstream genes.

Close proximity to the transcription start site silences poly(A) signals. The finding that 5′-UTR-derived putative poly(A) signals were very active when they are placed at the 3′ end but not in their natural location at the 5′ end raised the possibility that the position where a poly(A) signal is within a gene may determine whether it is recognized as functional. We tested this putative positional effect by putting putative poly(A) signals derived from the 5′-UTRs into different locations in reporter genes. Initially, we put UTR-9 into three different positions in the *Adh* gene: in the resulting three constructs, the distances between the AAUAAA and TSS were 509 nt (construct P1), 695 nt (P2), and 926 nt (P3) (Fig. 3A). Northern blot analysis of total RNA from cells transfected with these constructs showed that at P2 and P3 the UTR-9 poly(A) signal is very active, producing high levels of the expected truncated *Adh* transcripts (Fig. 3B, Adh panel, lanes 2 and 3).

In contrast, the reporter with the poly(A) signal at P1 generated only trace amounts of the expected truncated transcript (Fig. 3B, lane 1). Construct P1 appeared to be transcribed well, because the readthrough *Adh-Luc* transcript processed at the intergenic *Adh* poly(A) site was produced at a relatively high level (Fig. 3B). As expected, the truncated transcripts were polyadenylated, since they could be detected by oligo-d(T)-primed RT-PCR (Fig. 3C) and were retained in oligo(dT)-selected RNA (Fig. 3B, lanes 6 and 7). Sequencing of the RT-PCR products showed that polyadenylation takes place 11 to 26 nt downstream of the AAUAAA in UTR-9 (Fig. 3D) and that their poly(A)s are of similar length. These results further indicate that the 3′ ends of these transcripts are generated by the standard cleavage and polyadenylation reaction.

Next, we tested whether the low activity of the poly(A) signal in construct P1 might be due to the proximity of inhibitory sequences located at the beginning of the *Adh* coding region. To do this we made a construct in which we deleted the first 192 nt of the *Adh* coding region, moving P2 closer to the TSS. In the resulting Adh-UTR-9-ΔP2 construct, the AAUAAA was 503 nt from the TSS, a similar distance as for construct P1 (Fig. 3A). The transcript level was low, as with the initial Adh-UTR-9-P1 reporter (Fig. 3B, lane 4 versus 2). However, deletion of the same 192-nt sequence from the reporter Adh-UTR-9-P3, to yield Adh-UTR-9-ΔP3, caused only a moderate reduction (Fig. 3B, lane 5 versus 3). These experiments indicate that the UTR-9 poly(A) signal becomes silent when it is close to the TSS, regardless of the upstream flanking sequence. We observed a similar position dependence with a shorter derivative of UTR-9 (Fig. 4A, S-UTR-9); this sequence had a slightly weaker poly(A) signal activity (Fig. 4B versus 2B). As for the full-length UTR-9, the poly(A) signal in S-UTR-9 appeared to be silent when at position P1 (Fig. 4C, lane 1) but very active at P2 and P3 (Fig. 4C, lanes 2 and 3).

The production of truncated *Adh* mRNAs requires the presence of the AAUAAA hexamer in S-UTR-9: deletion of the hexamer prevented production of the truncated *Adh* transcript regardless of where it was inserted (Fig. 4C, lanes 4 to 6). Surprisingly, deletion of the AAUAAA hexamer did not greatly increase the level of the readthrough transcript. Although it has been reported that the hexamer alone can induce Pol II pausing and termination (24, 26), our data suggest that sequences flanking the hexamer can also negatively affect Pol II elongation through the intergenic region.

Next, to test if the position-dependent effect is applicable to other poly(A) signals, we inserted a strong SV40 poly(A) signal at positions P1, P2, P3, ΔP2, and ΔP3, as described above. The results showed that the SV40 poly(A) signal at P1 was also silenced, producing much less transcript than at P2 or P3 (Fig. 5C, lane 1 versus lanes 2 and 3; lanes 6 to 8 show the same transcripts in poly(A)-selected RNA). Moving the SV40 signal closer to the TSS by deletion of the first 192 nt of *Adh* also reduced its activity, as in the UTR-9-based reporters (Fig. 5C, lane 4 versus 2). However, the SV40 poly(A) signal produced little readthrough transcript, unlike the reporters with UTR-9; the SV40 poly(A) signal appeared to be stronger, yielding, relative to the other signals, more of the truncated transcript at all three positions. It is also possible that the levels of the readthrough transcripts were affected by changes in mRNA stability; in particular, both the UTR-9 and the SV40 sequences have premature termination codons (PTCs), making the corresponding readthrough transcripts potential NMD targets. We investigated this possibility with the short derivative of UTR-9 (S-UTR9), because this sequence, unlike the SV40 sequence, retains only one in-frame PTC, which can be removed by site-specific mutagenesis. We found that the transcript levels were not affected by the PTC presence; therefore, excluding that NMD could target these transcripts (Fig. 4C, lanes 4 to 6). Furthermore, RNAi knockdown of UPF1 (an essential protein for NMD) does not increase the amount of readthrough transcript with the Adh-UTR-9-P1 reporter (data not shown).

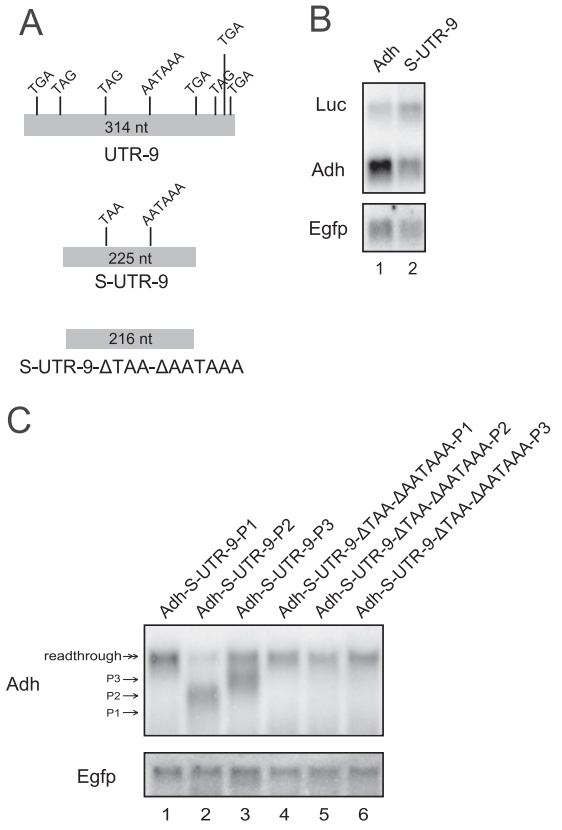


FIG. 4. 3′-end processing requires the AAUAAA hexamer. (A) Schematics of the UTR-9 derivatives with or without AAUAAA and in-frame stop codons. (B) Northern blot analysis of reporters with the shorter UTR-9 derivative (shown in panel A) inserted between *Adh* and *Luc* as in Fig. 2; probes and labeling are as described for Fig. 2. (C) Northern blot analysis of cells transfected with reporters containing the S-UTR-9 derivative at positions P1, P2, and P3 in *Adh*, as in Fig. 3; the S-UTR-9-ΔTAAΔAATAAA constructs lack all in-frame stop codons and AAUAAA.

To further investigate the generality of the observation that poly(A) signals do not function when located close to the 5′ end, we analyzed different reporter genes and other poly(A) signals. We inserted the bovine growth hormone gene (BGH) poly(A) signal at three positions in the *Escherichia coli* β-galactosidase gene (*lacZ*); in these constructs the AAUAAA hexamer was at 204, 404, and 704 nt from the TSS (Fig. 5A). In another set of reporters we inserted the UTR-4 poly(A) signal at three positions in the firefly luciferase gene (*Luc*), placing the AAUAAA sequence at 253,

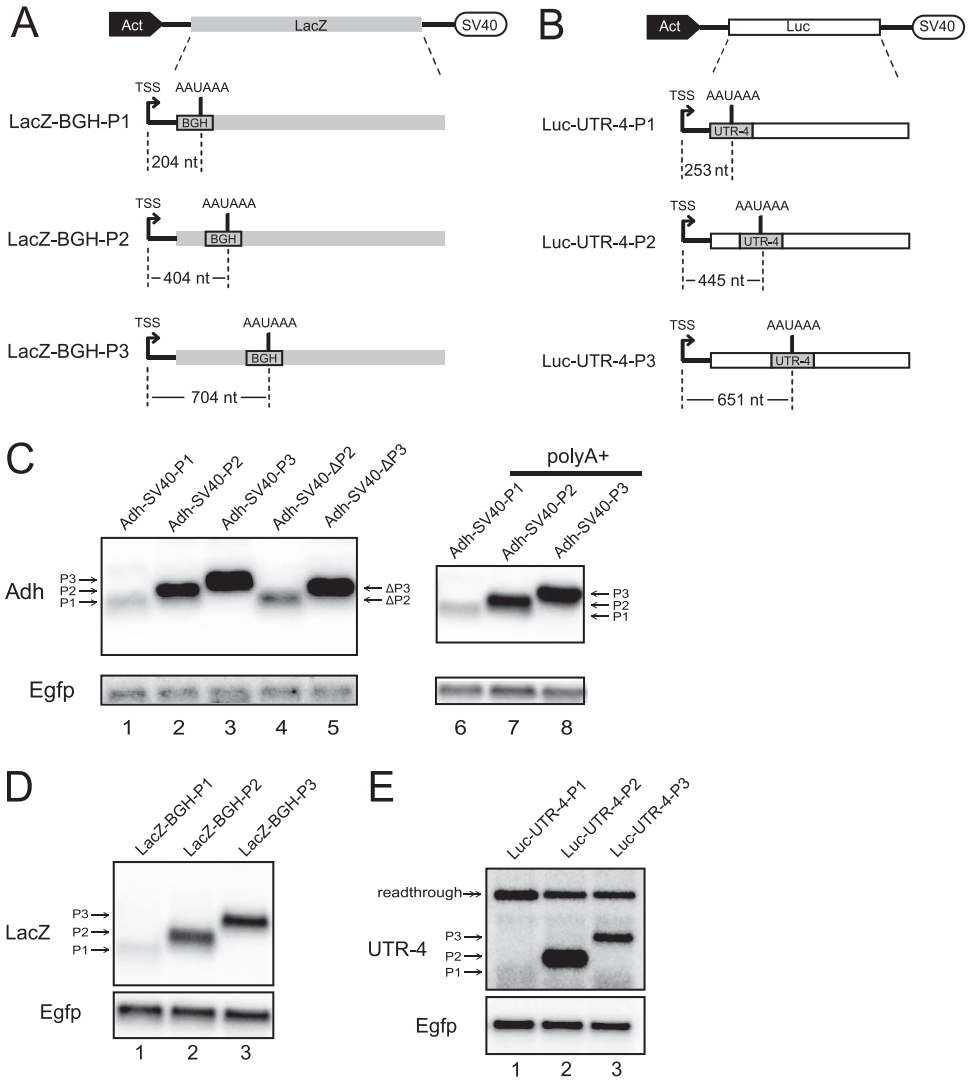


FIG. 5. Proximity to the TSS silences poly(A) signals in different reporter genes. (A) Schematics of *lacZ* reporters with the BGH poly(A) signal inserted at different positions. Distances from the TSS to AAUAAA are indicated below each schematic. (B) Schematics of *Luc* reporters with UTR-4 inserted at different positions. Distances from the TSS to AAUAAA are indicated below the schematics. (C) Northern blots of total RNA and poly(A)$^+$ RNA of S2 cells transfected with reporters with the SV40 late poly(A) signal at positions P1, P2, P3, ΔP2, and ΔP3 in *Adh*. (D) Northern blots of total RNA from S2 cells transfected with the reporters in shown panel A. A BGH-specific probe was used. The mRNAs polyadenylated at the three positions are indicated by arrows. (E) Northern blots of total RNA from S2 cells transfected with the reporters shown in panel B. A UTR-4-specific probe was used. The mRNAs polyadenylated at the three positions are indicated by arrows. Readthrough sequences are polyadenylated at the SV40 poly(A) signal in the plasmid.

445, and 862 nt from the 5′ end (Fig. 5B). With the *lacZ*-based reporters, we found that BGH poly(A) signal produced very little mRNA at P1 compared to P2 and P3 (Fig. 5D). Similarly, with the Luc reporters we found that when located at P1, the UTR-4 poly(A) signal was silent, while it was highly active when further downstream at P2 and P3 (Fig. 5E). In these latter reporters, for unknown reasons the signal seemed weaker at P3 than at P2 (Fig. 5E).

To assess whether the position on the pre-mRNA affects poly(A) signal recognition in other organisms, we made similar constructs driven by the CMV promoter and tested them in human HEK 293T cells (Fig. 6A). In these reporters, the SV40 poly(A) signal was placed at three alternative positions, P1, P2, and P3 in *Adh*, as in the *Drosophila* reporters described above. The reporters were transiently transfected in cells, and the RNA was assayed by Northern blotting of total RNA. We found the same positional effect as in *Drosophila*: when placed at P1, the SV40 signal was silent, while it was very active further downstream at positions P2 and P3 (Fig. 6B).

In summary, the results indicate that it is common for poly(A) signals to become silent when placed close to the TSS in both *Drosophila* and human cells. The minimum distance at which the silencing occurs varies among different transcripts (~200 to 500 nt); it seems that the distance over which the poly(A) signal can be efficiently recognized also depends on the strength of the poly(A) signal and the gene context.

Transcripts generated at early poly(A) signals are not unstable. Next we asked whether the reason for the low steady-

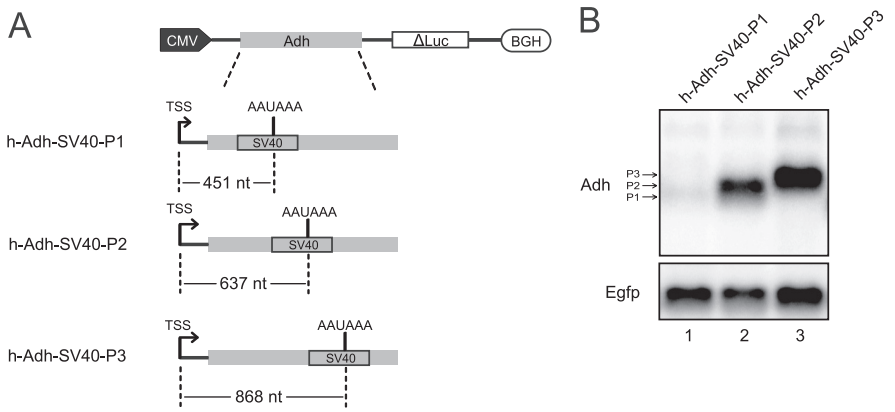


FIG. 6. Proximity to TSS silences the poly(A) signal in human cells. (A) Schematics of *Adh-ΔLuc* (truncated Luc) reporters in pCDNA 3.1 plasmid (Invitrogen). The constructs are flanked by the CMV promoter and a BGH poly(A) signal. The SV40 poly(A) signal was inserted at P1, P2, or P3 in *Adh*. Distances from the TSS to AAUAAA are indicated below each diagram. (B) Northern blots of total RNA from 293T cells transfected with the reporters shown in panel A. Probes are those described for Fig. 2. The mRNAs generated at each position are indicated.

state accumulation of mRNA polyadenylated at early sites might simply be that such transcripts are unstable; they might be efficiently produced but rapidly degraded by some mRNA quality control pathway. Many recent studies have indicated that aberrant transcripts are rapidly degraded by the nuclear exosome (19). Therefore, we were interested in assessing whether the exosome also targets prematurely polyadenylated transcripts. We depleted nine exosome subunits (Rrp6, Dis3/Rrp44, Rrp41/Ski6, Mtr3, Rrp40, Rrp46, Rrp42, Csl4, and Rrp4) in S2 cells by RNAi and transfected them with some of the *Adh-Luc* reporters described above. Initially, we carried out an RNAi screen using the Adh-SV40-P1 reporter, and we found that none of the RNAi depletions appreciably increased the level of the small transcripts processed at the 5′-proximal P1 site (see Fig. S2 in the supplemental material). In particular, we assessed the effect of depleting Rrp6, the exonucleolytic active nuclear-specific subunit of the exosome (2); we observed no recovery of the P1 truncated transcript, and instead the levels of the transcripts were reduced (Fig. 7A). Depletion of Rrp6 also appeared to reduce the level of the full-length *Adh* mRNA of the original *Adh-Luc* reporter (Fig. 7C, lane 1 versus 6). It appears that the nuclear exosome might stimulate mRNA biogenesis rather than hampering it. Depletion of the poly(A) factors CPSF-160 or Pcf11 reduced mRNA levels similar to Rrp6 knockdown (Fig. 7C). Furthermore, double depletion of CPSF-160 plus Rrp6 or of Pcf11 plus Rrp6 had a cumulative effect that further reduced mRNA levels (Fig. 7C). In these experiments, RNAi also affected the EGFP transcript; however, similar results were observed in many experimental repeats, and levels of the 18S rRNA indicated no significant variability in the assay. In summary, these results indicate that the truncated transcripts produced by early 3′-end processing are probably not subjected to exosome-mediated mRNA surveillance.

It can be argued that something else other than the exosome is responsible for the degradation. To address the issue further, we cloned the Adh-UTR9-P1 and -P2 constructs in front of an ecdysone-inducible promoter and assessed them in S2 cells as described above (see Fig. S3A in the supplemental material). The expectation was that if the P1 transcript is made but

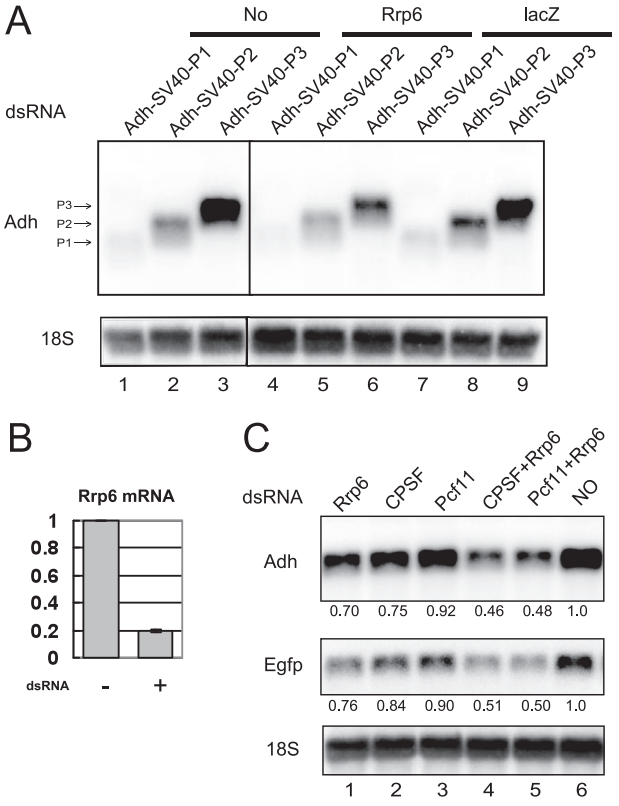


FIG. 7. Rrp6 depletion does not recover truncated mRNA levels. (A) Northern blot analysis of transcripts in Rrp6-depleted S2 cells transfected with Adh-SV40-P1, Adh-SV40-P2, or Adh-SV40-P3. Mock experiments without dsRNA (lanes 1 to 3) or with off-targeting LacZ dsRNA (lanes 7 to 9) are also shown. Adh probe was that shown in Fig. 2, and the 18S rRNA was previously described (9) (B) Real-time RT-PCR measurements of Rrp6 mRNA depletion relative to control cells not treated with dsRNA. The level of Rrp6 mRNA was normalized to that of Rpl32. (C) Northern blots of total RNA from S2 cells transfected with the *Adh-Luc* reporter (as for Fig. 2, with the *Adh* poly(A) signal in the intergenic region); the cells were treated with dsRNA against Rrp6, CPSF-160, Pcf11, CPSF plus Rrp6, or Pcf11 plus Rrp6. Quantifications (below the blots) are band intensities relative to that in the control not treated with dsRNA; values were standardized to the relative intensity of the 18S band.

quickly degraded, then we should detect relatively more of it at early time points than later in the induction. In these experiments the readthrough could be detected 20 min after induction and became apparent after 40 min, yet only very low levels of the short transcript could be detected with the P1 reporter (see Fig. S3B), similar to what was observed with steady-state RNA. Furthermore, the level of P1 seemed to increase linearly with that of the readthrough. The P2 truncated transcript, which as expected was more abundant, also accumulated at the same rate as the readthrough (see Fig. S3C). With the inducible reporters, it was also more apparent that there was more readthrough transcript when the signal was at P1 than at P2. Skipping the upstream signal would be expected to increase the production of readthrough. This was not always apparent in our experiments with steady-state RNA (e.g., Fig. 3B, lanes 4 and 5); perhaps this was due to different transcripts having slightly different stabilities.

In summary, the data indicate that truncated transcripts are not intrinsically unstable. Thus, their low levels might be primarily caused by a failure in 3′-end processing at the early poly(A) sites.

Early poly(A) signals are more sensitive to Pcf11 depletion. Early poly(A) signals might be skipped because key processing factors are inefficiently recruited to short nascent transcripts. We reasoned that depletion of some 3′-end processing proteins might affect earlier sites more than later ones. We tested this possibility in S2 cells that had been depleted of CPSF-160, CstF-64, or Pcf11 by RNAi. Cells were transfected with the *Adh-Luc* reporters with poly(A) signals at different distances from the TSS (as in Fig. 3A), and transcript levels were analyzed by Northern blotting. Depletion of CPSF-160 and CstF-64 caused a general reduction in the transcript levels regardless of the relative position of the poly(A) signals (Fig. 7C and data not shown). However, depletion of Pcf11 appeared to comparably affect the early poly(A) signals more: in cells partially depleted of Pcf11 the ratio of truncated transcript to readthrough was clearly decreased for both P2 and P3 mRNAs (Fig. 8A and B, lanes 2 and 3 versus 5 and 6).

Pcf11 recruitment may be affected by Pol II CTD phosphorylation. In an attempt to change the Ser2P level, we knocked down Cdk9 and CycT components of p-TEFb (27) and assessed the effects on the same reporters (Fig. 8D). Depletion of either protein reduced the level of all transcripts, and the decrease was most apparent in cells depleted of both (Fig. 8D). However, the P2, P3, and readthrough transcripts were equally reduced. We also assessed the effect of depleting the CTD phosphatase Fcp1: Fcp1 has the opposite effect of Cdk9, reducing CTD Ser2 phosphorylation (11). In Fcp1-depleted cells there was slighter more of the P2 and P3 transcripts than the longer readthrough transcript (Fig. 8F, lanes 2 and 3 versus 5 and 6); Fig. 8G shows quantification of truncated transcript/readthrough ratios. However, contrary to expectation, depletion of Fcp1 did not recover the level of transcripts at P1-we found only one reporter (Luc-UTR-4-P1) in which depletion of Fcp1 produced a slight increase in the amount of P1 (data not shown). However, it is possible that the failure to activate the 5′ poly(A) site may be due to an insufficient reduction of Fcp1 protein (this could not be checked directly, due to the lack of an Fcp1 antibody). In agreement with this view, we found no obvious change in the global level of Ser2P in Fcp1-depleted S2 cells (see Discussion) (see also Fig. S4 in the supplemental material).

In summary, the results indicate that poly(A) signals that are relatively close to the 5′ end of the pre-mRNA are more sensitive to changes in Pcf11 levels, yet depletion of Pcf11 seems to inhibit 3′-end processing over a region greater than that where poly(A) signals fail to function. The mechanism which silences 5′-proximal poly(A) signals might be linked to low Ser2 phosphorylation of the CTD; Fcp1 is probably contributing to this mechanism, but based on our data it is not clear whether it is sufficient.

DISCUSSION

The composite sequence that makes the poly(A) signal is the key determinant in the specification of the poly(A) site (13, 39), and it is generally expected that such sequences should be only found at the 3′ ends of genes. Contrary to this view, here we have reported that bioinformatic programs also predict the presence of poly(A) signals in the 5′-UTR of 24% of *Drosophila* genes. We have experimentally verified the function of a subset of these sequences when they are placed at the 3′ end of reporter genes: one was even more efficient than the endogenous poly(A) signal of *Adh*, one of the most highly expressed genes in *Drosophila*. The number of transcripts with putative poly(A) signals in their 5′-UTR is probably an underestimate, as the programs we used only predicted about half of the known poly(A) signals in 3′-UTRs.

As mentioned in the introduction, recognition of poly(A) signal is linked to Pol II termination. Our findings therefore raise the question of why such sequences are allowed to evolve at the beginning of genes, where they could potentially interfere with transcription. One obvious possibility is that 5′ signals are skipped because Pol II is not yet loaded with 3′-end processing factors during the first section of the elongation phase (20, 44). Probably, recruitment of processing factors remains inefficient until Ser2 in the CTD of Pol II becomes hyperphosphorylated (7). Our observation that depletion of the processing factor Pcf11 affects early poly(A) signals more than later ones supports this model. Furthermore, depletion of the CTD phosphatase Fcp1 increased the level of short transcripts more than longer ones, indicating that the skipping mechanism depends on low Ser2 phosphorylation. However, except for one case, depletion of Fcp1 did not activate the 5′ poly(A) signals that are skipped. While the lack of major effects might simply be due to incomplete depletion of the phosphatase, it is also possible that the signals are not affected by Fcp1 depletion, because Ser2 is not yet hyperphosphorylated when Pol II transcribes these early sites (7). For *S. cerevisiae* it has been reported that the phosphorylation state of the CTD at early stages of transcription prevents early poly(A) complex-dependent termination and instead favors Nrd1-dependent termination and generation of cryptic unstable transcripts (CUTs) (16, 38). CUTs are degraded by the nuclear exosome, but in our Northern blot assays, we did not find evidence of truncated transcripts that were enriched upon depletion of exosome components. All together, our data suggest that the transcript levels are low because of failure in 3′-end processing at the early sites, rather than selective degradation by mRNA surveillance mechanisms. Furthermore, since the truncated transcripts are

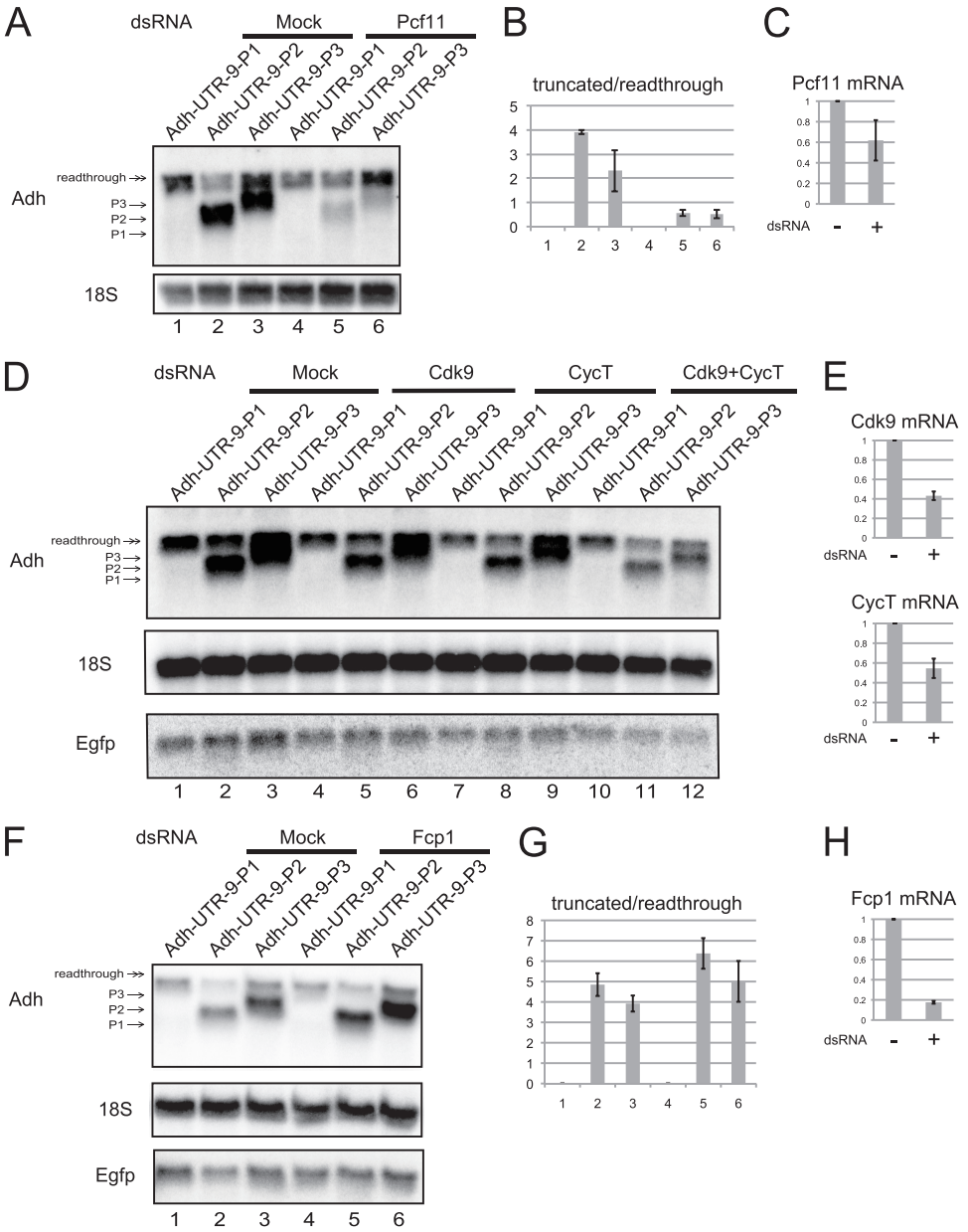


FIG. 8. Early poly(A) signals are more sensitive to Pcf11 depletion. (A) Northern blot analysis of total RNA extracted from Pcf11-depleted S2 cells transfected with Adh-UTR-9-P1, Adh-UTR-9-P2, or Adh-UTR-9-P3 as described for Fig. 3. Mock experiments involved cells not treated with Pcf11 dsRNA (lanes 1 to 3); the Adh probe was that shown in Fig. 2, and the 18S rRNA was that described for Fig. 7. (B) Band intensity quantification of the P2/readthrough and P3/readthrough transcript ratios detected by Northern blotting, with error bars based on two independent experiments. (C) Real-time RT-PCR quantification of Pcf11 mRNA in cells treated with dsRNA relative to control; Pcf11 mRNA levels were normalized to that of Rpl32. (D) Northern blot analysis of total RNA from S2 cells depleted of Cdk9, CycT, or Cdk9 and CycT and transfected with the same constructs as in panel A. (E) Real-time RT-PCR quantification of Cdk9 and CycT mRNA depletion, as described for panel C. (F) Northern blot analysis of total RNA from Fcp1-depleted S2 cells transfected with the same constructs as shown above. (G) P2/readthrough and P3/readthrough band intensity ratios detected by Northern blotting; error bars are based on two experiments. (H) Real-time RT-PCR quantification of the levels of Fcp1 mRNA in cells depleted via RNAi.

polyadenylated, it seems that it is primarily cleavage that is failing at these 5′ sites.

For most of the 5′-UTR poly(A) signals we have assayed, there is no evidence that they are used in flies; a few 5′-UTR signals are used but can be detected only by nested PCR, indicating that they are rarely recognized. The proposal is that 5′-UTR poly(A) signals are probably silent in the endogenous genes because they are too close to the 5′ end. In agreement with this view, we found, by using different reporter genes in both *Drosophila* and human cells, that the 5′-UTR sequences, as well as standard poly(A) signals, become silent when located near the 5′ end of reporter genes. The distance over which the signals are silenced varies between reporter genes (~500 nt from the TSS in *Adh*-based reporters, and ~200 to 250 nt in

LacZ- and *Luc*-based reporters). We propose that skipping of early signals is a general feature of the way poly(A) signals are recognized on nascent pre-mRNA. At this stage our conclusion is that most promoter-proximal poly(A) signals are intrinsically silent, and that is the explanation for why 5′-UTR poly(A) signals do not affect transcription of the endogenous genes. Future studies should address whether there is a link with Pol II pausing (24). In addition, it is feasible that for some of these genes 5′-UTR signals are means of regulation. In agreement with this view, the list of genes with 5′ signals (see the tables in the supplemental material) is significantly enriched ($P < 10^{-7}$) with genes regulated during postembryonic development (data analysis not shown).

The possibility that promoter-proximal poly(A) signals might be silent has not been systematically assessed for cellular genes, yet many studies have previously reported that proximity to the promoter can silence poly(A) signals in retroviral pre-mRNAs (39). Studies with the HIV-1 provirus have shown that the U1 snRNP binds a 5′ splice site immediately downstream of the 5′ long terminal repeat (LTR) poly(A) signal and prevents its usage; these studies concluded that it is the presence of the 5′ splice site rather than the physical proximity to the promoter that negatively regulates the early poly(A) signal (3, 4). In HIV-1 the 5′ LTR poly(A) signal is 254 nt downstream of the TSS; our results would predict that at this distance the poly(A) signals are intrinsically silent. In addition, the observation that nonretroviral poly(A) signals are active when the original 5′ LTR poly(A) signal is replaced in HIV-1 also seems to contrast with our prediction (40). But HIV-1 transcription requires the viral protein Tat (45); Tat directly interacts with P-TEFb and stimulates its CTD kinase activity (35, 42). We speculate that the poly(A) signal within the HIV 5′ LTR is not intrinsically silent, because the rapid Tat-dependent Ser-2 CTD hyperphosphorylation moves forward the recruitment of 3′-end processing factors.

In summary, against the common view that the poly(A) signal sequence alone is sufficient to define poly(A) sites, our data clearly indicate that *in vivo* an important second determinant is the stage at which the sequence emerges from Pol II. The ability of skipping early poly(A) signals might have evolved to prevent premature transcription termination. The mechanism is probably conserved in eukaryotes: standard poly(A) signals placed near the 5′ end also become silent in *S. cerevisiae* (D. Libri, personal communication).

ACKNOWLEDGMENTS

We thank Matthias Soller, Alicia Hidalgo, and Steve Dove and their lab members for useful discussions and for sharing equipment, Domenico Libri and Steve Buratowski for valuable discussions and experimental suggestions, and Yi Jin Liew for bioinformatics help. We also thank Domenico Libri, Bob Michell, and Nick Proudfoot for critically reading the manuscript.

This study was supported by a Royal Society University Research Fellowship and a Wellcome Trust project grant to S.B. and by a Dorothy Hodgkin Postgraduate Award to J.G.

REFERENCES

1. **Ahn, S. H., M. Kim, and S. Buratowski.** 2004. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. Mol. Cell **13:**67–76.
2. **Allmang, C., et al.** 1999. The yeast exosome and human PM-Scl are related complexes of 3′-→ 5′ exonucleases. Genes Dev. **13:**2148–2158.
3. **Ashe, M. P., P. Griffin, W. James, and N. J. Proudfoot.** 1995. Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. Genes Dev. **9:**3008–3025.
4. **Ashe, M. P., L. H. Pearson, and N. J. Proudfoot.** 1997. The HIV-1 5′ LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. EMBO J. **16:**5752–5763.
5. **Brogna, S.** 1999. Nonsense mutations in the alcohol dehydrogenase gene of Drosophila melanogaster correlate with an abnormal 3′ end processing of the corresponding pre-mRNA. RNA **5:**562–573.
6. **Brogna, S., and M. Ashburner.** 1997. The Adh-related gene of Drosophila melanogaster is expressed as a functional dicistronic messenger RNA: multigenic transcription in higher organisms. EMBO J. **16:**2023–2031.
7. **Buratowski, S.** 2009. Progression through the RNA polymerase II CTD cycle. Mol. Cell **36:**541–546.
8. **Caplen, N. J., J. Fleenor, A. Fire, and R. A. Morgan.** 2000. dsRNA-mediated gene silencing in cultured Drosophila cells: a tissue culture model for the analysis of RNA interference. Gene **252:**95–105.
9. **Chan, H. Y., S. Brogna, and C. J. O'Kane.** 2001. Dribble, the Drosophila KRR1p homologue, is involved in rRNA processing. Mol. Biol. Cell **12:**1409–1419.
10. **Cheng, Y., R. M. Miura, and B. Tian.** 2006. Prediction of mRNA polyadenylation sites by support vector machine. Bioinformatics **22:**2320–2325.
11. **Cho, E. J., M. S. Kobor, M. Kim, J. Greenblatt, and S. Buratowski.** 2001. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes Dev. **15:**3319–3329.
12. **Clemens, J. C., et al.** 2000. Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways. Proc. Natl. Acad. Sci. U. S. A. **97:**6499–6503.
13. **Colgan, D. F., and J. L. Manley.** 1997. Mechanism and regulation of mRNA polyadenylation. Genes Dev. **11:**2755–2766.
14. **de Vries, H., et al.** 2000. Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. EMBO J. **19:**5895–5904.
15. **Gilsdorf, M., et al.** 2010. GenomeRNAi: a database for cell-based RNAi phenotypes. 2009 update. Nucleic Acids Res. **38:**D448–D452.
16. **Gudipati, R. K., T. Villa, J. Boulay, and D. Libri.** 2008. Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. Nat. Struct. Mol. Biol. **15:**786–794.
17. **Hirose, Y., and J. L. Manley.** 1998. RNA polymerase II is an essential mRNA polyadenylation factor. Nature **395:**93–96.
18. **Holton, T. A., and M. W. Graham.** 1991. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res. **19:**1156.
19. **Houseley, J., and D. Tollervey.** 2009. The many pathways of RNA degradation. Cell **136:**763–776.
20. **Licatalosi, D. D., et al.** 2002. Functional interaction of yeast pre-mRNA 3′ end processing factors with RNA polymerase II. Mol. Cell **9:**1101–1111.
21. **Lopez, F., S. Granjeaud, T. Ara, B. Ghattas, and D. Gautheret.** 2006. The disparate nature of "intergenic" polyadenylation sites. RNA **12:**1794–1801.
22. **McCracken, S., et al.** 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature **385:**357–361.
23. **Moucadel, V., F. Lopez, T. Ara, P. Benech, and D. Gautheret.** 2007. Beyond the 3′ end: experimental validation of extended transcript isoforms. Nucleic Acids Res. **35:**1947–1957.
24. **Nag, A., K. Narsinh, A. Kazerouninia, and H. G. Martinson.** 2006. The conserved AAUAAA hexamer of the poly(A) signal can act alone to trigger a stable decrease in RNA polymerase II transcription velocity. RNA **12:**1534–1544.
25. **Nunes, N. M., W. Li, B. Tian, and A. Furger.** 2010. A functional human poly(A) site requires only a potent DSE and an A-rich upstream sequence. EMBO J. **29:**1523–1536.
26. **Orozco, I. J., S. J. Kim, and H. G. Martinson.** 2002. The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template. J. Biol. Chem. **277:**42899–42911.
27. **Peterlin, B. M., and D. H. Price.** 2006. Controlling the elongation phase of transcription with P-TEFb. Mol. Cell **23:**297–305.
28. **Proudfoot, N.** 2004. New perspectives on connecting messenger RNA 3′ end formation to transcription. Curr. Opin. Cell Biol. **16:**272–278.
29. **Proudfoot, N. J., A. Furger, and M. J. Dye.** 2002. Integrating mRNA processing with transcription. Cell **108:**501–512.
30. **Ramanathan, P., J. Guo, R. N. Whitehead, and S. Brogna.** 2008. The intergenic spacer of the Drosophila Adh-Adhr dicistronic mRNA stimulates internal translation initiation. RNA Biol. **5:**149–156.
31. **Richard, P., and J. L. Manley.** 2009. Transcription termination by nuclear RNA polymerases. Genes Dev. **23:**1247–1269.
32. **Sadowski, M., B. Dichtl, W. Hubner, and W. Keller.** 2003. Independent functions of yeast Pcf11p in pre-mRNA 3′ end processing and in transcription termination. EMBO J. **22:**2167–2177.
33. **Shi, Y., et al.** 2009. Molecular architecture of the human pre-mRNA 3′ processing complex. Mol. Cell **33:**365–376.

34. **Tabaska, J. E., and M. Q. Zhang.** 1999. Detection of polyadenylation signals in human DNA sequences. Gene **231:**77–86.
35. **Tahirov, T. H., et al.** 2010. Crystal structure of HIV-1 Tat complexed with human P-TEFb. Nature **465:**747–751.
36. **Tian, B., J. Hu, H. Zhang, and C. S. Lutz.** 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res. **33:**201–212.
37. **Tian, B., Z. Pan, and J. Y. Lee.** 2007. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome Res. **17:**156–165.
38. **Vasiljeva, L., M. Kim, H. Mutschler, S. Buratowski, and A. Meinhart.** 2008. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat. Struct. Mol. Biol. **15:** 795–804.
39. **Wahle, E.** 1995. 3′-end cleavage and polyadenylation of mRNA precursors. Biochim. Biophys. Acta **1261:**183–194.
40. **Weichs an der Glon, C., J. Monks, and N. J. Proudfoot.** 1991. Occlusion of the HIV poly(A) site. Genes Dev. **5:**244–253.
41. **West, S., and N. J. Proudfoot.** 2008. Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination. Nucleic Acids Res. **36:**905–914.
42. **Yang, Z., et al.** 2005. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Mol. Cell **19:**535–545.
43. **Zhang, Z., J. Fu, and D. S. Gilmour.** 2005. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3′-end processing factor, Pcf11. Genes Dev. **19:**1572–1580.
44. **Zhang, Z., and D. S. Gilmour.** 2006. Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. Mol. Cell **21:**65–74.
45. **Zhu, Y., et al.** 1997. Transcription elongation factor P-TEFb is required for HIV-1 Tat transactivation in vitro. Genes Dev. **11:**2622–2632.