Introns in histone genes alter the distribution of 3' ends

Niranjan B.Pandey, Nunta Chodchoy, Ta-Jen Liu and William F.Marzluff

Department of Chemistry and Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA

Received March 22, 1990; Revised and Accepted May 7, 1990

ABSTRACT

Chimeric genes were constructed which contained either a histone or globin promoter, a human α -globin coding region as a cDNA or containing one or both intervening sequences, and the 3' end of a mouse histone H2a gene. The genes were introduced into mouse L cells or Chinese Hamster Ovary cells. The genes containing at least one intervening sequence produced two mRNAs in about equal amounts, one which ended at a cryptic polyadenylation site 33 nucleotides 3' to the normal histone mRNA 3' end and one which ended at the normal histone 3' end. In contrast, the same construct containing a globin cDNA yielded mRNA ending only at the correct histone 3' end. Similar proportions of polyadenylated and non-polyadenylated mRNA were obtained when the cryptic polyadenylation signal was replaced with the globin polyadenylation signal. More than 90% of the transcripts were accurately spliced. All of the unspliced transcripts had histone 3' ends.

INTRODUCTION

The 3' end of a eucaryotic messenger RNA is formed by a cleavage reaction which releases the transcript from the chromatin template (reviewed in ref. 1). Most mRNA precursors end with a polyA sequence and contain intervening sequences which are subsequently removed. The sole exception are the replication-dependent histone genes in higher eucaryotes. These genes lack intervening sequences and code for mRNAs which end in a highly conserved stem-loop structure with a six-base stem and four base loop (2). The unique 3' ends on histone mRNAs have been implicated in regulation of histone mRNA stability (3–7), the determination of the level of expression of different histone genes (8) and in the cell-cycle regulation of histone mRNA synthesis (9,10).

The sequence requirements for histone 3' end formation have been defined in both sea urchin (11) and mammalian (9) histone genes and consist of two components: the stem-loop sequence and a purine-rich sequence 10–20 nucleotides 3' of the stem-loop. Endonucleolytic cleavage occurs between these two conserved sequences (12,13) resulting in formation of mature histone mRNA. The nucleolytic cleavage requires at least two components, one of which is the U7 snRNP. The U7 snRNP interacts with the purine-rich sequence (13–17). Similarly the formation of polyadenylated mRNA requires two sequences, a

highly conserved AAUAAA sequence 15–20 nucleotides from the 3' end of the mRNA and a downstream GU rich sequence which is less highly conserved (reviewed in ref. 1). Cleavage occurs between these two sequences and the polyA tail is then added.

There is extensive biochemical evidence that many of the intervening sequences are removed after the polyadenylated transcript is formed. However, both electron microscopic evidence (18) and studies of genes with alternative splicing and polyadenylation sites, indicate that there is a commitment to splicing as well as some splicing occurring prior to 3' end formation. The calcitonin gene is spliced by one pathway to yield calcitonin in the parathyroid gland and is spliced by a different pathway to yield the calcitonin gene related peptide (CGRP) mRNA in brain cells and the two mRNAs utilize different polyadenylation sites (19). The 3' end of the respective mRNAs is determined by the splicing pattern chosen (20). Here we report that introducing either (or both) intron(s) from the human α -globin gene into a gene containing the histone 3' end interferes with histone 3' end formation and favors usage of cryptic 3' polyadenylation sites. This effect is independent of the promoter used. In contrast genes containing globin sequence as a cDNA gave rise to mRNAs with only histone 3' ends. These results suggest that a commitment to splicing during transcription may help direct polyadenylation, hence interfering with histone 3' end formation.

MATERIALS AND METHODS

Construction of histone-globin chimeric genes

Histone globin chimeric genes were constructed in the plasmid pUC18. The complete sequence of the H2a and H3-164 histone genes used in this study has been reported (22–24). The intact human α 2-globin gene (21) was digested with BglII in the 3' untranslated region and the end blunted with T4 DNA polymerase and then digested with PstI. The 560 nt SstI to SstI fragment containing the histone H2a 3' end was cloned into pUC18. This plasmid was digested with HincII and PstI and the 1.4kb PstI-BglII fragment from the α -globin gene was inserted to give the GG(1,2)H gene. The 1.1 kb PstI to NcoI fragment containing the histone H3-614 promoter (22,23) was introduced into this gene after removing the globin promoter by digestion with PstI and NcoI to give the HG(1,2)H gene. The NcoI site spans the ATG initiation codon in both the histone and globin gene. The human α 2-globin cDNA clone was a generous gift of Dr. Bernard

Figure 1. Nucleotide and amino acid sequence of bovine *myf* cDNA.

Forget. Using the NcoI site in the first exon, the HindIII site in the second exon, or the BstEII site in the third exon, genes containing either the first, HG(1)H, or the second, HG(2)H, intron were constructed by replacing either the NcoI-HindIII fragment or the HindIII-BstEII fragment in the HG(1,2)H gene with the corresponding fragment from the cDNA clone. The HIH gene was constructed by replacing the 3' portion of the HG(1)H gene from the HindIII site at codon 89 of the α -globin gene with the 3' end of the H2a gene from the PstI site at codon 85. An AluI fragment containing the mouse globin polyadenylation site (4) was attached either at the HpaII site 35 nucleotides 3' to the end of the H2a-614 gene or at the XbaI site 320 nucleotides 3' to the end of the H2a-614 gene and the HIH gene. These genes are diagramed in Figure 1.

Cell culture and transfection

Mouse tk⁻ L cells were cotransfected with the Herpes virus thymidine kinase gene and the histone gene constructs using calcium phosphate precipitation as described previously (25). Chinese hamster ovary cells (CHO cells) were cotransfected with pSV2neo (26) and the histone gene constructs using the polybrene procedure (27) as described previously (8,28). RNA samples were prepared from pools of 30–100 colonies which were growing exponentially (<50% confluent).

Preparation and analysis of RNA

Total cell RNA was prepared from exponentially growing cells as described previously (29). Mouse L cells were fractionated into nuclear and cytoplasmic fractions as described previously (25). The RNAs were analyzed by a quantitative S1 nuclease mapping procedure (23) using appropriate end-labeled restriction

enzyme sites. The probes used for each assay are indicated in the figure legends. The S1-resistant DNA fragments were resolved by gel electrophoresis and detected by autoradiography. The films were scanned with a densitometer to quantitate the amount of the different RNA transcripts.

RESULTS

As part of our studies on the regulation of histone mRNA stability, we replaced the 3' end of the human α 2-globin gene (containing both intervening sequences) with the 3' end of a mouse histone H2a gene and found that about 50% of the mRNAs formed did not have histone 3' ends (3). We have investigated this phenomenon in detail by constructing a number of chimeric histone-globin genes, which are shown in Fig. 1. All of the constructions in Fig. 1 contain the mouse histone H2a-614 gene 3' end (8) starting at the SstI site 13 bases 5' of the stem-loop. This fragment is sufficient for efficient 3' end formation (8, 30). Either the human α 2-globin (21) or the mouse histone H3-614 promoter (24) were used. The NcoI site spanning the ATG initiation codon of both the α -globin gene and the histone H3-614 gene was used to construct these genes. Genes containing the histone H3-614 promoter and 5' untranslated region and the complete globin coding region with both introns [HG(1,2)H], either intron [HG(1)H or HG(2)H], or the globin cDNA (HcGH) all yield mRNAs which would code for authentic globin proteins. In addition a gene (HIH) was constructed which contained the first intervening sequence of the globin gene up to codon 89 and the last 199 nucleotides of histone H2a-614 sequence.

In Fig. 1 are also shown the constructs which have a globin polyadenylation signal attached varying distances 3' to genes

141
 Arg End Polylinker
 CGT TAA GCTGGAGCCTCGGTGGCCATGCTTCTTGCCC GACTCTAGAGGATCCCCGGGTACC
 * *
 GAGCTCCCAAAAAGGCTCTTTTCAGAGCCACCCACTGAATCAGATAAAGAGTTGTGTACGGTA
 *
 GCCGGTCTTGGTGGCCTTTCCTGGCCCTGCCCAAAAGTAAGAGTGTTTCGTGTTCTCTGCTTC
 *
 CTTGCCCATCTGCCGTGGTTTACCCTGGTTAGTTTA

Figure 2. Sequence of the 3' flanking region of the globin-histone chimeric genes. The sequence starts at the last codon of the globin gene. The polylinker sequence inserted during cloning is shown, followed by the 3' end of the H2a gene starting at the SstI site. The asterisks indicate the termini of the three transcripts found from this gene. The stem-loop and the purine-rich region required for both histone 3' end formation and which is also the cryptic polyadenylation signal are underlined.

Figure 3. The transcripts formed are independent of the promoter. A. Mouse L cells were transfected with the GG(1,2)H and HG(1,2)H genes and total cell RNA prepared and analyzed by S1 nuclease mapping, using the HcGH gene 3' end-labeled at the BstEII site in the third exon of the globin gene as a probe. The protected DNA fragments were resolved by gel electrophoresis and detected by autoradiography. The fragments map to the sites indicated in Fig. 3A. A schematic of the S1 nuclease assay is shown below the figure. The protected fragments are: H2a, RNAs ending at the normal histone mRNA end; H2a(A₁⁺), RNAs ending at the first cryptic polyadenylation site; H2a(A₂⁺), RNAs ending at the second cryptic polyadenylation site. Lane 1: 30 µg of RNA from cells transfected with the GG(1,2)H gene; Lane 2: RNA from untransfected tk⁻ mouse L cells; Lane 3: 5 µgm of RNA from cells transfected with the HG(1,2)H gene; lane 4: pUC18 digested with HpaII and 3' end labeled. B. Mouse L cells were transfected with the intact mouse H2a-614 gene and the transcripts from untransfected (lanes 1–3) and transfected cells (lanes 4–6) analyzed by S1 nuclease mapping using the H2a-614 gene labeled at the 3' end of the Aval site as a probe. A schematic of the S1 nuclease assay is shown below the figure. 20 µgms of total cell RNA (lanes 1 and 4) and equivalent fractions of polyA⁻ (lanes 2 and 5) and polyA⁺ RNA (lanes 3 and 6) were analyzed. RNAs terminating at the cryptic polyadenylation site are indicated as H2a(A₁⁺). Lane M is pUC18 digested with HpaII and 3' end labeled.

containing the histone H2a-614 3' end. The H2a-35 polyA gene is the intact H2a gene with a restriction fragment containing the mouse α -globin polyadenylation signal attached 35 nucleotides from the 3' end of the histone mRNA. The polyadenylation signal is 95 nucleotides from the start of this fragment putting the AATAAA signal 130 nucleotides from the end of the histone mRNA. Similarly H2a-320 is the H2a-614 gene with the same polyadenylation signal 421 nucleotides from the histone 3' end.

The H2H-polyA gene also has the globin polyadenylation signal attached 421 nucleotides 3' of the end of the histone mRNA. Since the polyA tail is added 20 nucleotides after the AATAAA signal, the polyA tail will be added either 150 (in the H2a-35 gene) or 441 nucleotides (in the H2a-320 gene) from the end of the histone mRNA. All of these genes were introduced into tk⁻ mouse L cells or into CHO cells and pools (30–100 colonies) of stable transfectants selected and analyzed. RNA was prepared

from exponentially growing cells to maximize the amount of mRNA expressed from these genes.

Figure 2 shows the sequence of the 3' untranslated region of the globin-histone chimeras. The ends of the different transcripts detected are indicated by an *; the first * is at the end of the normal histone mRNA and the other two *'s mark the major and minor attachment sites of the polyA tail. The stem-loop and the portion of the purine-rich signal which is also the cryptic polyadenylation signal are underlined. The cryptic polyadenylation signal (AGATAAA) is part of the normal purine-rich signal recognized by U7 snRNA during histone 3' end formation (15).

The presence of intervening sequences alters the distribution of 3' ends

The 3' ends of two of the histone-globin chimeric genes [HG(1,2)H and GG(1,2)H] are shown in figure 3. These genes are identical except for the promoters and 5' untranslated regions. To map the transcripts an S1 nuclease assay was used. The HcGH gene was 3' end labeled at the BstEII site at codon 107 in the third exon of the globin gene. This assay maps the 3' ends of all the transcripts (spliced and unspliced). Below we address the question of the splicing of different transcripts (see Fig. 6). The HG(1,2)H and GG(1,2)H genes yielded mRNAs with an identical pattern of 3' ends, with about 50% of the RNAs ending at the histone 3' end and the rest ending at two other sites, the prominent site 33 nucleotides 3' of the normal histone 3' end and the minor site 34 nucleotides farther downstream (Fig. 3A, lanes 1,3). Both of these RNAs are polyadenylated as judged by binding to oligo(dT) cellulose (not shown but see Fig. 4B). Since these genes differ only in the promoter region, the promoter did not affect the 3' end which was utilized. There was much less RNA produced from the cells transfected with the GG(1,2)H gene compared with the HG(1,2)H gene probably since the globin promoter is expressed much less efficiently than the histone promoter. The great majority of the transcripts from both the GG(1,2)H (not shown) and HG(1,2)H genes were spliced (see Fig. 6.)

We have previously shown that a small proportion (about 1%) of the transcripts from the endogenous H2a-614 genes end at these cryptic polyadenylation sites (Fig. 3B, lanes 1,3, also see ref. 4). Transfection of the intact H2a-614 gene into mouse L cells resulted in increased expression of H2a-614 mRNA (which is expressed in high amounts from the endogenous gene) but did not significantly alter the per cent of transcripts which ended at the cryptic polyadenylation site (Fig. 3B, lanes 4 vs. 1). These results suggested that the polyadenylated 3' ends were formed because there were intervening sequences present in these genes and were not a result of the transfection procedure.

Effect of the position of introns on 3' end formation

To determine whether there was an effect of the numbers of introns in the transcript, the distance of the intron from the 3' end, and whether either of the human α -globin introns could cause this effect, the mRNAs expressed from the HG(1)H, HG(2)H, and HcGH genes were studied. These genes were transfected into both CHO and mouse L cells. Each of these genes gives rise to the same mRNA after splicing. Figure 4A shows the transcripts formed from chimeric histone-globin genes containing either the first or second intron, or the HcGH construct which lacks introns but would encode the same mature mRNA. Probes mapping the 3' ends of all transcripts from the HindIII site at codon 89 of the globin coding region (Fig. 4A, lanes 1-4) and

Figure 4. Transcripts from the histone-globin chimeric genes. A. The HIH, HG(1)H, HG(2)H, and HcGH genes were introduced into CHO cells by stable transfection and the RNA products analyzed by S1 nuclease mapping. The probes and RNAs used are indicated above the lanes. The probes used were the HIH gene 3' end labeled at the HindIII site in the globin gene (lane 1) or the HcGH gene 3' end labeled at the HindIII site in the second exon of the globin gene (lanes 2-4). In lanes 5-7 the HcGH probe labeled at the NcoI site was used, which will detect only processed transcripts. 5 μ gms of RNA was analyzed in lane 1 and 30 μ gms of RNA was analyzed in lanes 2-7. Schematics of the S1 nuclease assays are shown below the figure. Lane 1: HIH; Lanes 2 and 7: HG(1)H; Lanes 3 and 6: HG(2)H; Lanes 4 and 5: HcGH. B. The polyA⁺ and polyA⁻ transcripts are cytoplasmic. Mouse L cells transfected with the HIH gene were fractionated into nuclear (N) and cytoplasmic fractions. The RNA from the cytoplasmic fraction was fractionated into A⁺ and A⁻ RNAs and each fraction analyzed by S1 nuclease mapping using the HIH gene labeled at the 3' end of the HindIII site as in A. The position of marker DNAs is shown at the right. Lane T was 5 μ gms of total cell RNA and equivalent proportions of polyA⁺ and polyA⁻ RNA were analyzed.

Figure 5. Usage of distal polyadenylation sites. The H2a-35 polyA, H2a-320 polyA and H2a-polyA genes were introduced into CHO cells by transfection. The 3' ends of the transcripts were mapped using S1 nuclease. Schematics of the S1 nuclease assays are shown below the figure. A. 5 μ gms of RNA from cells transfected with the H2a-35 polyA gene was analyzed. The probe was the intact mouse H2a-614 gene 3' end labeled at the AvaI site at codon 20. The protected fragments are H2a, protection of the hamster H2a mRNAs; H2a-614, transcripts from the transfected gene terminating at the normal histone 3' end; A⁺(G), protection of longer transcripts which map to the point of attachment of the globin polyadenylation site to the H2a-614 gene. Lane T: Total cell RNA; Lanes A⁺ and A⁻, polyA⁺ and polyA⁻ mRNA respectively. RNA from five times as many cells was analyzed in the A⁺ lane compared with the total and the A⁻ lanes. The bands labeled P are derived from the probe. B. 10 μ gm of RNA from cells transfected with the H2a gene (lane 1), 30 μ gm of RNA from cells transfected with the H2a-polyA gene (lanes 2, 3 and 5) or 30 μ gm of RNA from cells transfected with the H2a-320 polyA gene (lane 4) were analyzed by S1 nuclease mapping. In lanes 1 and 2, the probe was the H2a gene labeled at the HindIII site in the last exon. In lane 3 the same RNA was probed with the H2a-polyA gene labeled at the XbaI site which is 121 nts from the globin polyadenylation site. In lane 4 the probe was the H2a-320polyA gene labeled at the AvaI site at codon 20. An equal amount of RNA from the cells transfected with the H2a gene was analyzed for comparison in lane 5. Schematics of the S1 nuclease assays are shown below the figure. The protected fragments are H2a, protection of the hamster H2a RNAs; H2a-614, protection to the normal 3' end of the mouse H2a-614 gene; H2a(A⁺), protection to the first cryptic polyadenylation site; A⁺(G), protection to the globin polyadenylation site. The same samples were analyzed in lanes 2 and 5 except that electrophoresis was for a longer time in lane 5.

the complete transcript from the NcoI site at the ATG codon (Fig. 4A, lanes 5–7) were used. The genes containing either of the two globin introns gave the same distributions of 3' ends (Fig. 4A, lanes 2,3, 6,7) as the HG(1,2)H gene (Fig. 3A, lane 3) containing both introns, with about 50% of the transcripts ending at the histone 3' end and 50% ending at the cryptic polyadenylated site. Note that the probe labeled at the NcoI site could protect both spliced and unspliced transcripts and hence does not explicitly address whether the transcripts are spliced (see Fig. 6).

However, more than 95% of the transcripts from the HcGH gene which lacked introns had the histone 3' end (Fig. 4A, lane 4,5). This conclusively shows that the alteration in the distribution of 3' ends was due to the presence of an intervening sequence in the gene. Placing the intron different distances from the histone 3' end, 500 nts in the HG(1)H gene (Fig. 4A, lane 2) or 150 nts in the HG(2)H gene (Fig. 4A, lane 3), did not alter the relative proportion of the 3' ends formed.

To see whether including additional histone sequences prior to the 3' processing signal affected the processing of the histone-globin chimeric genes, the H2H gene, which contains a globin intron and 199 nucleotides of histone sequence was constructed (Fig. 1). This gene gave transcripts with the same distribution of 3' ends, ending at both the histone end and at the cryptic polyadenylation site (Fig. 4A, lane 1; 4B, lane T). The cells transfected with the H2H gene were fractionated into nucleus and cytoplasm and the RNAs fractionated into polyA⁺ and polyA⁻ RNAs. Both the polyA⁺ and polyA⁻ transcripts are

predominantly cytoplasmic (Fig. 4B). Similar results were found for all of these genes in both mouse L cells and CHO cells (data not shown). The amount of expression in mouse L cells varied in different transfections due to variations in copy number, while the amount of expression in CHO cells was similar (and much less, due to the reduced copy number) for all the constructs (25, 28).

The presence of introns allows usage of distal 3' polyadenylation signals

Since the cryptic polyadenylation signal is normally an inefficient signal for polyadenylation, we placed a globin polyadenylation downstream of the histone 3' end. To test the effect of introns on utilization of the globin polyadenylation signal, the same signal was also attached 3' of the H2H gene. The mouse α -globin polyadenylation signal was placed either 130 or 421 nucleotides 3' to the intact H2a-614 gene or 421 nucleotides 3' of the H2H gene. When the globin polyadenylation signal was placed close to the histone 3' end in the intact gene, less than 5% of the transcripts were polyadenylated (Fig. 5A) and we have shown previously that these all end at the globin 3' end (8). The domination of the histone 3' signal is due to the fact that the H2a-614 3' processing signal was highly efficient (8).

The globin polyadenylation signal was then placed 421 nucleotides from the histone 3' end of the intact H2a gene and the H2H gene. About 40% of the transcripts from the H2H-polyA

Figure 6. Processing of the chimeric mRNAs. A. 5 μ gm of total cell RNA was prepared from mouse L cells transfected with the HIH, HG(1,2)H, HG(1)H, HG(2)H and HcGH genes and analyzed by S1 nuclease mapping. The probes and RNAs used are indicated above the lanes. The probes used were: the HIH gene 3' end labeled at the NcoI site (lanes 1–4); the HG(2)H gene 3' end labeled at the NcoI site (lanes 5,6); and the HG(1,2)H gene labeled at the NcoI site (lanes 7). Schematics of the S1 nuclease assays are shown below the figure. In lanes 1–4 a probe to the H3-614 gene was included as an internal control. This probe protects a 174 nucleotide fragment labeled H3. The protected fragments are: a 96 nucleotide fragment from the HIH probe extending to the boundary of the first intervening sequence (lanes 1–4) labeled G(S) or a 297 nucleotide fragment from the HG(2)H probe extending to the boundary of the second intervening sequence (lanes 5,6); US, unspliced RNAs extending to the end of the HIH gene (603 nts, lane 1), the point of divergence of the HIH gene and the HG(1,2)H or HG(1)H genes (388 nts, lanes 2,3) or the end of the HG(2)H gene (622 nts, lane 5). The fragment labeled I2 represents transcripts with the first but not the second intron removed from the HG(1,2)H gene. Lane 1: HIH; Lane 2: HG(1,2)H; Lane 3: HG(1)H; Lane 4: HcGH; Lane 5: HG(2)H; Lane 6: HcGH. Lane 7: HG(1,2)H gene. Lane M: pUC 18 digested with HpaII. The position of the marker fragments are indicated. B. RNA from the cells transfected with the HG(1,2)H gene was fractionated into polyA⁺ and polyA⁻ RNA and equal proportions of each fraction assayed by S1 nuclease mapping using the HIH probe labeled at the NcoI site to map the spliced transcripts (left panel) and the unspliced transcripts (center panel). Lane t is a tRNA control. Because of the small amounts of unspliced RNA, 10 times as much RNA was analyzed in the center panel than the left panel. The exact position of the 3' end was mapped using 30 μ gms of RNA mapped with the HG(1,2)H gene labeled at the HindIII site as a probe (right panel). This probe will detect only unspliced transcripts. C. Total cell RNA from mouse L cells transfected with the HIH-polyA gene was fractionated into polyA⁺ and polyA⁻ fractions. 15 μ gms of total cell RNA (lane T) and equivalent proportions of the polyA⁺ and polyA⁻ RNA (lanes A⁺ and A⁻) were analyzed by S1 nuclease mapping using the HIH-polyA gene labeled at the NcoI site as a probe. We included a probe for the 5' portion of the histone H3-614 gene (174 nts, NcoI to SalI) labeled at the NcoI site which protects the hamster histone H3 mRNAs (H3_H), which are poly A⁻. Band P is derived from the probe. Lane M is marker pUC18 digested with HpaII.

gene, which contains an intron, were polyadenylated (Fig. 5B, lane 2). Most of the transcripts utilized the globin polyadenylation signal (Fig. 5B, lanes 2 and 3), although this was 410 nucleotides 3' to the cryptic polyadenylation signal which was used efficiently in the absence of the globin polyadenylation signal (Fig. 5B, lane 1). Thus the addition of the globin polyadenylation signal did not change the relative amounts of polyadenylated versus non-polyadenylated mRNAs formed, but the efficient distal globin

polyadenylation site was dominant to the proximal cryptic site. In contrast, when the polyadenylation signal was 421 nucleotides 3' of the intact histone gene, there was no polyadenylated RNA ending at the globin 3' end detected (Fig. 5B, lane 4). Thus, an efficient polyadenylation site 3' of a histone gene was only utilized if there was an intron in the gene, and the proportion of transcripts polyadenylated (about 40%) was the same as when there was an inefficient polyadenylation site.

Efficiency of processing of the chimeric transcripts

The experiments described in Figs. 3 and 4 have mapped the 3' ends of all the transcripts, without regard as to whether they were spliced. To determine whether the transcripts were spliced, an S1 nuclease assay was performed using the H1H or HG(2)H gene (containing intervening sequences) 3' end labeled at the NcoI site spanning the ATG codon as a probe. These probes give a short (96 nucleotide) protected fragment extending to the 5' boundary of the first intervening sequence for the spliced transcripts and a long fragment extending to the 3' end of the transcript for the unspliced transcripts. In Fig. 6A (lanes 1–4) the H1H gene (which contains the first intervening sequence of the globin gene) was used as a probe to map the RNAs from the H1H, HG(1,2)H, HG(1)H and HcGH genes. A histone H3 probe labeled at the NcoI site of the H3 gene and extending to the SalI site (174 nucleotides) was also included in the assay as an internal control (Fig. 6A, lanes 1–4). The spliced RNAs all protect a 96 nucleotide fragment. In addition, there are longer fragments (designated US) due to protection by unspliced transcripts (388 nucleotides from the HG(1,2)H and HG(1)H RNAs, mapping to where they diverge from the H1H gene, and 603 nucleotides mapping to the end of the H1H RNA). As expected, the HcGH gene maps only a single band due to protection to the end of the first intron in the probe (Fig. 6A, lanes 4 and 6). The HG(2)H gene was used to map the transcripts from this gene and there was a great majority of spliced transcripts and a small amount of unspliced transcripts found in steady-state RNA (Fig. 6A, lane 5). Thus the great majority of steady-state transcripts which accumulate from all of these genes are spliced.

The H1H probe mapped all of the unspliced transcripts from the HG(1)H and HG(1,2)H as a single protected fragment of 388 nts (Fig. 6A lanes 2,3). The relative amount of these fragments was similar to the relative amount of unspliced RNAs from the H1H gene and the HG(2)H gene (Fig. 6A, lanes 1 and 5) suggesting that there were not a large number of unspliced transcripts with heterogeneous ends which extended past the histone 3' end. To definitively map the unspliced transcripts from the HG(1,2)H gene we used the HG(1,2)H gene labeled at the NcoI site as a probe. We detected unspliced transcripts which had the first intron but not the second intron removed (12, 405 nucleotides) as well as completely unspliced transcripts (Fig. 6A, lane 7). The relative intensities of the unspliced transcripts compared to the spliced transcripts were similar to those seen using the H1H gene (Fig. 6A, lane 2) as a probe. Thus we conclude that the majority of the unspliced RNAs had defined 3' ends. The transcripts from both the GGH (data not shown) and HG(1)H genes were also spliced efficiently, with only small amounts of unspliced transcripts present in steady-state RNA.

The unspliced RNAs have predominantly histone 3' ends

All the unspliced RNAs from the H1H and HG(2)H genes had a single 3' end (Fig. 6A, lanes 1 and 5) the size expected for the histone 3' end. We show in Fig. 6B and 6C that all the unspliced RNAs from the HG(1,2)H and H1H genes have histone 3' ends. To further define the unspliced transcripts, RNA was prepared from mouse L cells transfected with the HG(1,2)H gene and fractionated into polyA⁺ and polyA[−] RNAs by chromatography on oligo (dT) cellulose. Separate gels showing the assays of the fractionated RNA for spliced RNAs (Fig. 6B, left panel) and unspliced RNAs are shown (Fig. 6B, center panel). About 50% of the spliced RNAs were polyadenylated as expected from the results in Fig. 3 and Fig. 4. None of the unspliced RNAs

were retained on dT-cellulose, suggesting that they have histone 3' ends. To definitively map the 3' ends the HG(1,2)H gene was labeled at the HindIII site in the second exon and used as a probe. This probe protects only unspliced RNAs since the HindIII site is only 30 nts from the start of the second intron. A 396 nt fragment will be protected by transcripts which extend to the histone mRNA 3' end and a 430 nt fragment by transcripts extending to the polyadenylation site. As shown in Fig. 6B (right panel, lane T), only one protected fragment of 396 nts is seen, showing that all the unspliced RNAs end at the histone 3' end.

Similar results were seen with transcripts from the H1H-polyA gene (Fig. 6C). Again most of the transcripts were spliced and the unspliced transcripts had predominantly histone 3' ends (Fig. 6C, lane T). The spliced RNAs were found in both the polyA⁺ and polyA[−] fractions indicating that both types of transcripts were spliced (Fig. 6C, lanes A⁺, A[−]). The unspliced RNAs had histone 3' ends, as determined by their size and the fact that they are not retained on dT cellulose (Fig. 6C, right panel). These results strongly suggest that the transcripts with histone 3' ends are spliced more slowly (or inefficiently) than the polyadenylated transcripts resulting in significant amounts of unspliced non-polyadenylated RNA in steady-state RNAs.

DISCUSSION

The replication-dependent histone genes are the only class of genes coding for proteins in higher eucaryotes which lack intervening sequences and which do not code for polyadenylated mRNAs. The results reported here suggest a link between these two unique properties of histone genes, and suggests that efficient formation of histone 3' ends is not compatible with the presence of intervening sequences. In the histone H2a-614 gene the cryptic polyadenylation signal utilized (AGATAAA) is part of the purine-rich signal normally recognized for formation of the histone 3' end (Fig. 2, ref. 15). We observed that introducing an intron into this histone gene results in alteration of the 3' ends which are formed on the transcripts, favoring the formation of a polyadenylated transcript rather than transcripts with a normal histone 3' end. This is due to the presence of a functional intron in the gene. Either α -globin intron, which do not share any common sequences other than the splicing signals (21), alters the formation of 3' ends. This result is not due to the longer histone transcript since simply moving the histone 3' end further 3' to the gene does not alter the pattern of 3' end formation (30, and our unpublished results).

Our results suggest that the nascent transcript containing introns is probably directed to a polyadenylation pathway since the proportion of polyadenylated transcripts formed is the same whether there is an efficient or inefficient polyadenylation site and the formation of the polyadenylated transcripts is not dependent on the distance between the histone 3' end and the polyadenylation signal. We cannot rule out the possibility that a portion of the transcripts with histone 3' ends are formed subsequent to an initial polyadenylation event, although we feel that this is unlikely. Since equal amounts of total transcripts were formed from the HGH genes containing introns and the HcGH gene which lacked introns, there was no effect of introns on the overall expression of the gene. This suggests that the effect of the introns is to alter the distribution of 3' ends, and not to 'rescue' transcripts which would otherwise be degraded. The exact mechanism by which the distribution of 3' ends is altered is not known, although the lack of dependence on the distance of the

3' end from the intron argues against a simple physical interference by the spliceosome.

For some genes, e.g. β -globin (31), introns are required for the appearance of mature mRNA in the cytoplasm although the intronless genes are transcribed at the same rate as the wild-type gene. However, histone genes which have the 3' end replaced by a polyadenylation signal are efficiently expressed even though they lack introns (4). Likewise there are histone genes which lack introns and produce polyadenylated histone mRNAs (32–34). Thus histone genes differ from the β -globin genes in the requirement for an intron for production of polyadenylated mRNAs.

Our results are in contrast to the results of Proudfoot and coworkers who found that a globin polyadenylation site is dominant to the 3' end of a mouse histone H4 gene (31). When the two ends were present in tandem 3' to a histone H4 gene, the globin polyadenylation site was preferentially used. We have recently shown that one of the factors determining the amount of expression of individual mouse histone genes is the efficiency of 3' end formation (8,28). The histone H2A-614 gene is expressed in very high amounts (23), partly due to the high efficiency of the 3' end formation. The mouse H4 gene used in the experiments of Proudfoot and coworkers (35) is expressed in low amounts (36), possibly due in part to inefficient 3' end formation. Recently several examples of vertebrate histone genes have been reported which give rise to both normal polyA⁺-histone transcripts and polyadenylated transcripts in a tissue-specific (32) or cell-cycle (33,34) dependent manner.

The 3' end of histone mRNA plays an important role in regulation of histone mRNA levels during the cell cycle, both due to a regulation of 3' end formation (10, 37) and to regulation of histone mRNA degradation (7). Seiler-Tuyns and Paterson (38) have reported that a histone H4 gene containing an intron does not show normal cell cycle regulation, but is instead expressed constitutively. Since the authors did not comment on the reading frame of their transcripts it is not possible to ascertain whether the transcripts with histone ends would show normal regulation of RNA stability (5). There were also longer transcripts formed from this gene which they did not characterize but which could well represent polyadenylated transcripts and could have resulted from the presence of an intron in the gene.

A nascent spliceosome may direct 3' end formation

The finding that an intervening sequence in a gene affects the formation of the 3' end implies that the intervening sequence is recognized prior to 3' end formation. This is consistent with the results of Leff and coworkers (20) who showed that the splicing pattern chosen in the calcitonin CGRP gene determines which polyadenylation site is used. It is also consistent with the electron microscopic results of Osheim and coworkers (18) that RNP particles are assembled at splice junctions during transcription. Recently Robberson and coworkers (39) have presented evidence that exons (rather than introns) may be recognized to define the splicing pathway of pre-mRNA, resulting in the small size of the great majority of internal exons. They further suggest that the 3' donor site of the last intron plays a role in recognition of the polyadenylation signal (39, S. Berget, personal communication). This proposal is entirely consistent with our data which demonstrates that a polyadenylation site, rather than the histone 3' processing site is recognized preferentially when a gene contains an intron.

It is most likely that this recognition occurs at the level of the RNA rather than the DNA. A nascent spliceosome may rapidly

Fig. 7. Model of the effect of an intervening sequence on 3' end formation. The DNA is shown as straight lines, the exons of the RNA by a jagged line, an intron by a smooth line, the normal 3' end of the message by the stem-loop structure, the spliceosome by a slanted rectangle, a generic polyA signal by AAUAAA, and the putative linker between the spliceosome and the polyadenylation signal by a shaded triangle. The model illustrates the link between commitment to splicing and polyadenylation resulting when an intron was inserted in a histone gene.

form on the RNA during transcription. This nascent spliceosome might interfere with the binding of the factors involved in histone 3' end formation or it might act positively to help direct polyadenylation. Factors on the spliceosome might interact with the polyadenylation machinery and promote polyadenylation (Fig. 7). The result is that cryptic polyadenylation sites are used relatively efficiently if an intervening sequence is inserted into the histone gene. Since the same results were found regardless of the distance the intervening sequence was located relative to the 3' end (Fig. 4A), we favor a positive effect of the nascent spliceosome on polyadenylation.

Several aspects of this phenomenon are worth noting. The effect is seen even when the intervening sequence is very close to the 3' end (150 nucleotides from the 3' border of the intervening sequence to the 3' end) suggesting that the factors responsible must act within less than five seconds (assuming a transcription rate of 30 nucleotides per second) as the RNA polymerase traverses the DNA between the end of the intron and the 3' end. Even if the key recognition point is at the 5' end of the intron, the small size of these introns (about 100 nucleotides) means that the transcription of the intervening sequence and the commitment to a different pathway of 3' end formation are rapidly coupled.

The effect extends for a relatively long distance past the histone 3' end. We have recently shown that transcription extends at least 500 nucleotides 3' of the H2a-614 gene (our unpublished results). If an efficient polyadenylation signal is placed 421 nucleotides 3' of the end of a histone gene containing introns then this site is used (and is dominant to the cryptic polyadenylation site otherwise used). This result suggests that once the nascent transcript is committed to polyadenylation then the most efficient site available may be selected. Efficient polyadenylation sites located 3' to a 'normal' histone gene are not utilized significantly, even when they are located very close to the normal histone 3' end (Fig. 5A). These results again suggest that something 'marks' the nascent transcript to influence the formation of the 3' end.

A second conclusion that can be drawn from these results is that the transcripts with histone 3' ends are not spliced as efficiently as transcripts containing polyadenylated ends.

However, over 90% of the steady-state RNAs were correctly processed. All of the unspliced transcripts which we detected in steady-state RNA had histone 3' ends. This result suggests that the formation of the histone 3' end may reduce the rate of splicing, although it should be emphasized that the great majority of transcripts with histone 3' ends were correctly spliced as was also found by Seiler-Tuyns and Paterson (38).

There was not a detectable effect of the coding region or promoter sequences on the formation of the 3' ends. This is in contrast to the results with U1 snRNA, another RNA polymerase II transcript, where it was found that formation of the 3' end of these RNAs is dependent upon having an snRNA promoter (40,41). These results suggest a tight coupling of all phases of nuclear RNA metabolism: transcription, splicing, 3' end formation and transport. The histone mRNAs may be handled by a different nuclear system from spliced, polyadenylated mRNAs in all phases of their intranuclear metabolism because of their unique structure.

ACKNOWLEDGMENTS

This work was supported by grant GM29832 from the NIH to W.F.M. We thank Dr. Bernard Forget (Yale University) for a gift of the human $\alpha 2$ -globin cDNA clone.

REFERENCES

1. Birnstiel, M.L., Busslinger, M., and Strub, K. (1985) *Cell* 41, 349–359.
2. Hentschel, C.C., and Birnstiel, M.L. (1981) *Cell* 25, 301–313.
3. Pandey, N.B., and Marzluff, W.F. (1987) *Mol. Cell. Biol.* 7, 4557–4559.
4. Levine, B.J., Chodchoy, N., Marzluff, W.F., and Skoultschi, A.S. (1987) *Proc. Natl. Acad. Sci. USA* 84, 6189–6193.
5. Graves, R.A., Pandey, N.B., Chodchoy, N., and Marzluff, W.F. (1987) *Cell* 48, 615–626.
6. Capasso, O., Bleecker, G.C., and Heintz, N. (1987) *EMBO J.* 6, 1825–1832.
7. Marzluff, W.F. and Pandey, N.B. (1988) *Trends in Biochemical Sciences* 13, 49–52.
8. Liu, T.-J., Levine, B.J., Skoultschi, A.I. and Marzluff, W.F. (1989) *Mol. Cell Biol.* 9, 3499–3508.
9. Stauber, C., Luscher, B., Eckner, R., Lotscher, E., and Schumperli, D. (1986) *EMBO J.* 5, 3297–3303.
10. Luscher, B., and Schumperli, D. (1987) *EMBO J.* 6, 1721–1726.
11. Birchmeier, C., Folk, W. and Birnstiel, M.L. (1983) *Cell* 28, 739–745.
12. Gick, O., Kramer, A., Keller, W., and Birnstiel, M.L. (1986) *EMBO J.* 5, 1319–1326.
13. Mowry, K.L., and Steitz, J.A. (1987) *Mol. Cell. Biol.* 7, 1663–1672.
14. Mowry, K.L., and Steitz, J.A. (1987) *Science* 238, 1682–1687.
15. Mowry, K.L., and Steitz, J.A. (1988) *Trends in Biochemical Sciences* 13, 447–451.
16. Cotten, M., Gick, O., Vasserot, A., Schaffner, G. and Birnstiel, M.L. (1988) *EMBO J.* 7, 801–808.
17. Soldati, D., and Schumperli, D. (1988) *Mol. Cell. Biol.* 8, 1518–1524.
18. Osheim, Y.N., Miller, Jr., O.L., and Beyer, A.L. (1985) *Cell* 43, 143–151.
19. Amara, S.G., Evans, R.M. and Rosenfeld, M.G. (1984) *Mol. Cell Biol.* 4, 2151–2160.
20. Leff, S.E., Evans, R.M. and Rosenfeld, M.G. (1987) *Cell* 48, 517–524.
21. Proudfoot, N.J., and Maniatis, T. (1980) *Cell* 21, 537–544.
22. Graves, R. A., Wellman, S.E., Chiu, I.-M., and Marzluff, W.F. (1985) *J. Mol. Biol.* 183, 179–194.
23. Hurt, M.M., Chodchoy, N. and Marzluff, W.F. (1989) *Nuc. Acids Res.* 17, 8876.
24. Taylor, J.D., Wellman, S.E., and Marzluff, W.F. (1986) *J. Mol. Evol.* 223, 242–249.
25. Alterman, R.B., Sprecher, C., Graves, R.A., Marzluff, W.F., and Skoultschi, A.S. (1985) *Mol. Cell. Biol.* 5, 2316–2324.
26. Southern, E.M., and Berg, P. (1982) *J. Mol. and Appl. Genet.* 1, 327–341.
27. Chaney, W.G., Howard, D.R., Follard, J.W., Sallustio, S., and Stanley, P. (1986) *Somatic Cell and Mol. Genet.* 12, 237–244.
28. Levine, B.J., Liu, T.-J., Marzluff, W.F. and Skoultschi, A.I. (1988) *Mol. Cell. Biol.* 8, 1887–1895.
29. Graves, R.A., and Marzluff, W.F. (1984) *Mol. Cell. Biol.* 4, 351–357.
30. Chodchoy, N., Levine, B.J., Sprecher, C.S., Skoultschi, A.S., and Marzluff, W.F. (1987) *Mol. Cell. Biol.* 7, 1039–1047.
31. Buchman, A.R. and Berg, P. (1988) *Mol. Cell Biol.* 8, 4395–4405.
32. Kirsh, A.L., Groudine, M. and Challoner, P.B. (1989) *Genes Dev.* 3, 2172–2179.
33. Mannironi, C., Bonner, W.M. and Hatch, C.L. (1989) *Nuc. Acids Res.* 17, 9113–9126.
34. Cheng, G., Nandi, A., Clerk, S. and Skoultschi, A.I. (1989) *Proc. Nat. Acad. Sci. USA* 86, 7002–7006.
35. Whitelaw, E., Coates, A. and Proudfoot, N.J. (1986) *Nuc. Acids Res.* 17, 7059–7070.
36. Seiler-Tuyns, A., and Birnstiel, M.L. (1981) *J. Mol. Biol.* 151, 607–625.
37. Stauber, C. and Schumperli, D. (1988) *Nuc. Acids Res.* 16, 9399–9413.
38. Seiler-Tuyns, A. and Paterson, B.M. (1986) *Nuc. Acids. Res.* 14, 8845–8862.
39. Robberson, B.L., Cote, G.J. and Berget, S.M. (1990) *Mol. Cell Biol.* 10, 84–94.
40. Hernandez, N., and Weiner, A.M. (1986) *Cell* 47, 249–258.
41. Neuman de Vegvar, H.V., Lund, E., and Dahlberg, J.E. (1986) *Cell* 47, 259–266.