The EMBO Journal vol.13 no.10 pp.2432–2440, 1994

The site of 3′ end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP

Elizabeth C.Scharl and Joan A.Steitz

Department of Molecular Biophysics and Biochemistry, Howard Hughes Medical Institute, Boyer Center for Molecular Medicine, Yale University School of Medicine, 295 Congress Avenue, New Haven, CT 06536-0812, USA

Communicated by J.A.Steitz

Two conserved elements direct the 3′ end processing of histone messenger RNA: a stem–loop structure immediately upstream of the site of cleavage and the histone downstream element (HDE), located 12–19 nucleotides downstream of the stem–loop in the pre-messenger RNA. We studied the role of these two elements by systematically inserting up to 10 C residues between them in the mouse H2A-614 histone pre-mRNA. 3′ End mapping of RNAs processed *in vitro* demonstrated that as the HDE is moved downstream, the site of cleavage correspondingly moves 3′. In addition, the efficiency of processing declines. In the wild-type substrate, cleavage occurs 3′ of an A residue; modest increases in the efficiency of processing of the insertion mutants were observed when an A residue was placed at the new cleavage site. The results of psoralen cross-linking studies and immunoprecipitations using anti-trimethylguanosine antibodies indicated that the decreased processing efficiency of the insertion mutants is not due to impaired binding of the U7 small nuclear ribonucleoprotein (snRNP). We conclude that the mammalian U7 snRNP acts as a molecular ruler, targeting enzymatic components to cleave histone pre-mRNAs a fixed distance from its binding site, the HDE.
Key words: histone gene expression/RNA processing/small nuclear RNA/U7 snRNP

Introduction

Cell cycle-regulated histone messenger RNAs (mRNAs) are different from most RNA polymerase II transcribed messages in that they are not spliced and not polyadenylated. Instead, their 3′ ends are formed by an endonucleolytic cleavage event which requires both *cis*- and *trans*-acting factors (for reviews, see Birnstiel and Schaufele, 1988; Mowry and Steitz, 1988; Marzluff, 1992). Two conserved elements are found on histone pre-messenger RNAs (pre-mRNAs): a stem–loop immediately upstream of the site of cleavage and a sequence that includes a stretch of purine nucleotides located downstream of the cleavage site (histone downstream element, HDE). The stem–loop structure contains a six base pair stem and a four nucleotide loop that is highly conserved in sequence (for a review, see Marzluff, 1992).

Studies of histone mRNA 3′ end formation have focused on the sea urchin and mammalian systems. The U7 small nuclear ribonucleoprotein (snRNP) is required for the maturation of histone messages in both systems (Galli *et al.*, 1983; Strub and Birnstiel, 1986; Mowry and Steitz, 1987b; Soldati and Schümperli, 1988). Genetic suppression experiments have demonstrated that the 5′ end of the U7 snRNA forms base pairs with the HDE (Schaufele *et al.*, 1986; Bond *et al.*, 1991). Moreover, the latter studies revealed that the sequence of the snRNA–pre-mRNA duplex can be quite radically altered and still allow processing. The U7 snRNA contains a 2,2,7-trimethylguanosine (m_3G) 5′ cap; its snRNP particle is recognized by anti-Sm antibodies. In the mammalian U7 snRNP, two unique proteins of 14 and 50 kDa (Smith *et al.*, 1991; Mital *et al.*, 1993) have also been identified. In addition to the U7 snRNP, a cell cycle-regulated heat-labile factor is necessary for 3′ end processing in the mammalian system (Gick *et al.*, 1987; Lüscher and Schümperli, 1987). A mammalian hairpin binding factor (HBF) that is precipitable with Sm antibodies has been shown to bind to the stem–loop (Mowry and Steitz, 1987a; Vasserot *et al.*, 1989).

Although the stem–loop and the HDE elements are required in both sea urchin and mammalian histone pre-mRNAs, there are differences between the two systems. In sea urchin histone genes, a consensus ACCA sequence follows the stem–loop, the HDE consensus is CAAGAAAGA with the last five nucleotides being absolutely conserved, and the distance between the stem–loop and the HDE is always 10 nucleotides. Insertion of six nucleotides between these two elements has been observed to destroy all detectable processing in *Xenopus* oocytes (Georgiev and Birnstiel, 1985). In the majority of mouse and human histone genes whose expression has been characterized, a consensus of ACCCA follows the stem–loop, a consensus of G/AAAAGAGCTG has been derived for the HDE (Wells, 1986) with no sequence block being absolutely conserved, and the distance from the stem–loop to the HDE is more variable (12–19 nucleotides). Yet, the constraints on the distance between the stem–loop and the HDE that are apparent from sequence comparisons suggest that these two elements and the factors they bind may interact to facilitate processing of histone pre-mRNAs.

Here we have studied the effect of systematically moving the HDE further downstream of the stem–loop in a mouse histone H2A pre-mRNA. Most strikingly, the site of 3′ end formation *in vitro* moves in concert with the movement of the HDE. Decreases in the efficiency of processing as the HDE and the stem–loop are separated can be traced in part to a preference for an A residue at the cleavage site. However, data indicating that the U7 snRNP is able to recognize and bind the mutant substrates regardless of the placement of the HDE suggest that a particular geometry of the processing complex is necessary for efficient cleavage.

Results

Separating the HDE from the stem–loop alters both the site and efficiency of 3′ end formation

A series of mutations were made in the mouse H2A-614 pre-mRNA, which was previously well characterized with respect to HDE–U7 snRNA interactions (Bond *et al.*, 1991). The HDE was moved progressively downstream of the stem–loop by insertion of 1–10 C residues between these two structural elements (Figure 1). Five control constructs were used: 5Cswap contains both an insertion of five Cs and an exchange of the nucleotides ACCCA (nucleotides 30–34) which immediately follow the stem–loop with the next five nucleotides (nucleotides 35–38 plus one of the inserted Cs); 5Cswap therefore relocates the sequence normally upstream of the cleavage site to a position five nucleotides downstream. Δhp is missing the stem–loop of the histone pre-mRNA (nucleotides 14–29). 9C-A_9 and 9C-A_8 replace one of the inserted C nucleotides with an A residue at position +9 or +8 relative to the wild-type cleavage site, respectively. Therefore, 9C-A_8 reconstructs the ACCCA sequence found upstream of the wild-type cleavage site, while 9C-A_9 has the sequence ACCCCA upstream of the +9 position. HDEpu→py has the six purines (AAAGAG) of the histone downstream element changed to UUUCUC (nucleotides 46–51; Bond *et al.*, 1991).

These mutant histone transcripts were tested in an active *in vitro* processing system (Mowry and Steitz, 1987a). Internally [α-^{32}P]UTP-labeled substrates were incubated in HeLa nuclear extract for 1 h at 30°C. Figure 2 shows that as the distance between the HDE and the stem–loop increased by the insertion of three or more C residues, the size of the processed product changed. Specifically, as the HDE was moved further downstream, the length of the processed product increased by approximately a corresponding number of nucleotides (compare lanes 4–11 with lanes 2 and 3). Interestingly, the 3C substrate (lane 4) appeared to be processed both at the wild-type site and at a new site which generated a longer product. These results indicate that the HDE plays a major role in determining the position of the cleavage site on the histone pre-mRNA.

A second effect of moving the HDE further downstream is that, in general, the efficiency of cleavage progressively declines (Table I, column 2). The wild-type substrate (Figure 2, lane 2) was processed to ~50%, while a single C insertion (1C) reproducibly increased the levels of product (70% processed, lane 3). Insertion of additional Cs (lanes 4, 5, 7, 8 and 11) decreased processing efficiency to ~5% for the 10C mutant. No detectable processing (<1%) was seen with the HDEpu→py mutant (lane 1), consistent with previous observations with this substrate (Bond *et al.*, 1991). The Δhp mutant also showed severely decreased processing (<1%, lane 12), as reported for similar substrates (Mowry *et al.*, 1989; Vasserot *et al.*, 1989). The processing efficiency of the remaining mutants is considered below.

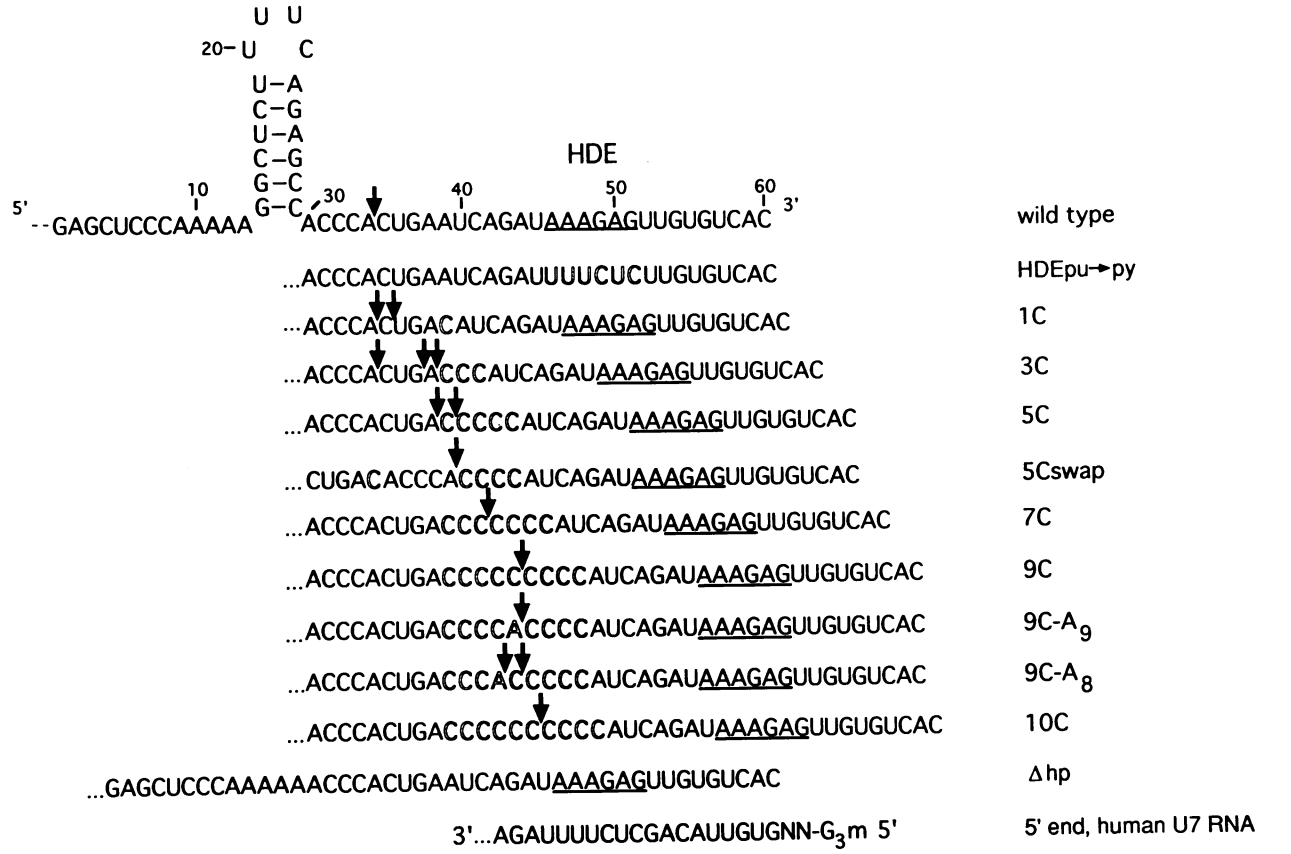

Fig. 1. Substrates used for *in vitro* processing reactions. Shown is the sequence of the mouse H2A-614 pre-mRNA (Hurt *et al.*, 1989) produced as a T3 RNA polymerase run-off transcript. At the 5′ end are 14 additional nucleotides which derive from vector sequences. Nucleotides are numbered starting with the mouse H2A-614 sequence. The arrows denote observed cleavage sites. The purine stretch within the HDE is underlined. Mutant sequences are shown below that of wild type, with inserted or altered nucleotides indicated by block letters. The 5′ portion of the human U7 snRNA sequence is also shown (Mowry and Steitz, 1987b).

3′ End mapping of processed products

Each substrate was transcribed in the presence of [γ-^{32}P]GTP and subjected to both alkali hydrolysis and partial digestion with RNases U_2 and T_1 to produce sequencing standards. Uncapped [α-^{32}P]UTP-labeled substrates were processed *in vitro*; the products were gel purified and fractionated in lanes adjacent to the standards (Figure 3). Whereas the products of hydrolysis and RNase digestion possess 3′ phosphate groups, cleavage of a histone substrate generates a 3′ hydroxyl group (Streit *et al.*, 1993). Thus, the mobilities of the processed products were expected to be slightly retarded relative to the sequencing standards of the same length. Conclusions based on the results shown in Figure 3 were verified by repeating these analyses using processed substrates and standards labeled by kinasing their 5′ ends instead (data not shown); identical 3′ end assignments were obtained.

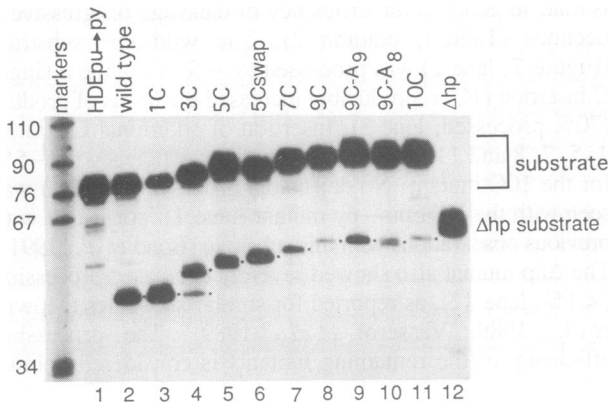

Fig. 2. *In vitro* processing of mouse H2A-614 mutants. One fmol of internally labeled substrate was incubated in HeLa cell nuclear extract for 1 h at 30°C; the RNAs were then resolved on an 8% polyacrylamide/8.3 M urea gel. Input substrates are indicated and dots mark processed products. Lanes 1 and 10 contained HDEpu→py and Δhp, respectively; both lack detectable amounts of processed products (<1%). Lanes 2–9 show wild-type substrate and the insertion mutants. Wild-type substrate is 74 nucleotides, while its processed product is 48 nucleotides.

In most cases, as the HDE moved further downstream of the stem–loop, the observed 3′ end moved a corresponding number of nucleotides downstream (see Figure 1). Figure 3A shows the wild-type processed product ends with the A at position 34. The 3′ end of the 1C substrate mapped to the same site, with an additional minor band whose 3′ end mapped one nucleotide downstream (Figure 3B). The processing of the 3C insertion mutant produced two different bands, both of which were analyzed (Figure 3C and D). The 3′ end of the faster migrating product corresponded to the wild-type A residue (nucleotide 34); the slower migrating band resolved into two different processed products on the 3′ end mapping gel, corresponding to a mixture of nucleotides 37 and 38 (Figure 2, lane 4). The 5C insertion mutant also yielded two different processed products whose 3′ ends mapped to either nucleotide 38 or the first of the inserted Cs (Figure 3E), while the processed product of 5Cswap ended with an A (Figure 3F). Substrates 7C, 9C and 10C each produced a single processed product with ends after positions +7, +9 and +10 relative to the wild-type site, respectively (Figure 3G, H and K). The A substitution in 9C-A_9 yielded the same product terminus as 9C (after position +9) (Figure 3I), but 9C-A_8 yielded products with ends at both positions +8 and +9 (Figure 3J).

Role of sequence at the cleavage site

Although processing occurs 3′ to A and 5′ to C in the wild-type substrate, other nucleotides flank the cleavage site in the insertion mutants. Specifically, both Gs and Cs can appear at the 3′ ends of processed mutant substrates, resulting from cleavage upstream of C, U or A (Figure 1). Since the 3′ ends differed for the various mutants, it is unlikely that the products were first cleaved and then underwent subsequent trimming in our extracts. We conclude that the

Table I. Histone pre-mRNA processing and α-TMG immunoprecipitations

Substrate	Amount of product formed (%)	Amount of substrate precipitated (%)		
		No oligonucleotide present	2′-*O*-Me anti-U1 present	2′-*O*-Me anti-U7 present
HDEpu→py	<1	3 ± 1.1	1 ± <1	2 ± <1
wt	51 ± 8	5 ± 3.2	8 ± <1	2 ± <1
1C	70 ± 8	5 ± 2.7	5 ± 1	1 ± <1
3C upper band	26 ± 4	10 ± 5.4	8 ± 1	1 ± <1
3C lower band	10 ± 3	10 ± 5.4	8 ± 1	1 ± <1
5C	25 ± 4	11 ± 3.0	14 ± 6	3 ± 1
5Cswap	39 ± 5	7 ± 1.9	5 ± 1	1 ± <1
7C	12 ± 2	11 ± 1.3	16 ± 5	2 ± <1
9C	5 ± <1	13 ± 5.2	9 ± 2	2 ± <1
9C-A_9	21 ± 3	–	7 ± <1	4 ± 2
9C-A_8	8 ± 1	–	11 ± 4	3 ± 2
10C	5 ± <1	13 ± 4.7	7 ± 1	3 ± <1
Δhp	<1	8 ± 3.7	6 ± <1	5 ± 1

Column 2: bands were quantitated using a Molecular Dynamics PhosphorImager and percent processing determined as follows: percent processing = [(processed × 2)/(processed × 2 + remaining substrate)] × 100; seven of the 14 labeled U residues are present in the processed products. Data from three experiments were averaged. When processed products differing by one nucleotide in length (1C, 3C upper band, 5C and 9C-A_8) were obtained, the value shown represents the sum of both bands. 3C lower band corresponds to processed product with a wild-type 3′ end; 3C upper band represents products terminating at positions +3 and +4 relative to the wild-type cleavage site. Column 3: the amount of precipitation was calculated as follows: percent precipitation = pellet/(pellet + supernatant) × 100. Five different experiments, involving two different preparations of HeLa cell nuclear extract, were used to generate the data. Columns 4 and 5: data were averaged from two experiments.

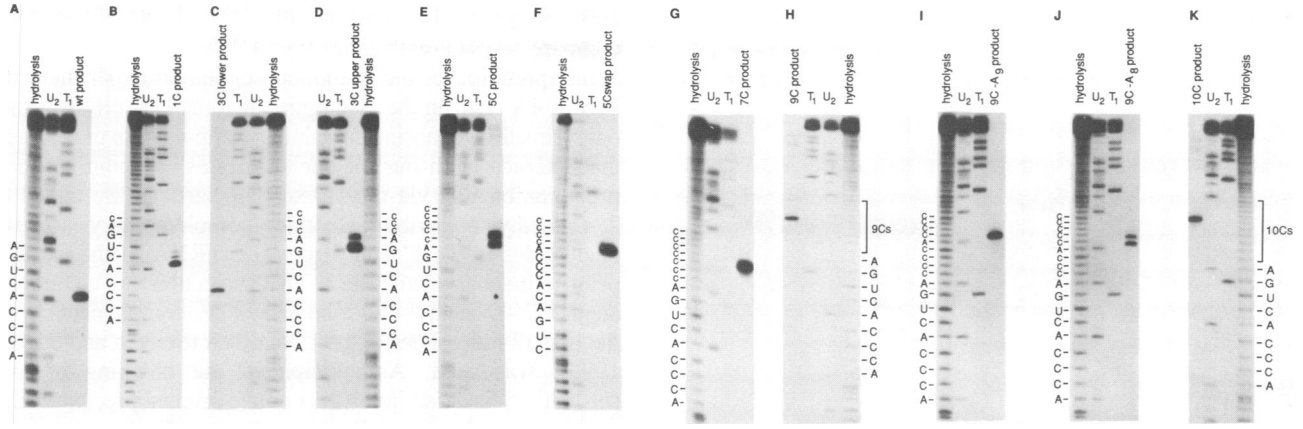

Fig. 3. 3′ End mapping of processed products. [γ-^{32}P]GTP-labeled substrates were subjected to partial alkali hydrolysis (hydrolysis), partial RNase U_2 digestion (U_2) or partial RNase T_1 digestion (T_1). Product lanes show uncapped [α-^{32}P]UTP-labeled processed products from the substrates indicated. (**A**) is wild type, (**B**) is 1C, (**C**) and (**D**) are 3C, (**E**) is 5C, (**F**) is 5Cswap, (**G**) is 7C, (**H**) is 9C, (**I**) is 9C-A_9, (**J**) is 9C-A_8 and (**K**) is 10C. Substrates were processed for 1 h in HeLa cell nuclear extract and processed products gel purified on a 5% acrylamide/8.3 M urea gel before analysis. The sequences shown start with nucleotides immediately following the stem−loop on the histone message and include residues through the inserted Cs.

histone 3′ cleavage activity does not have an absolute sequence specificity at its cut site.

On the other hand, it was possible that the decreased processing efficiency observed for most of the insertion mutants (Table I) resulted from altered sequences within a preferred cleavage region. 5Cswap, 9C-A_9 and 9C-A_8 had been constructed to investigate this possibility. 5Cswap places an A residue preceded by ACCC at position +5 relative to the original cleavage site. As a result, processing of 5Cswap increased by 50% relative to 5C (see Figure 2, lanes 5 and 6). Likewise, 9C-A_9 showed increased processing compared with 9C (lanes 8 and 9), although again, wild-type levels of cleavage were not obtained. 9C-A_8 provides an A at +8 relative to the wild-type cleavage site instead of at +9 and reconstructs the ACCCA sequence found at the wild-type cleavage site. Although processing was increased in 9C-A_8 relative to 9C (compare lanes 10 and 8), levels were significantly lower with 9C-A_8 than with 9C-A_9 (lane 9). Therefore, it seems that the presence of an A residue alone at a defined distance from the HDE of the histone H2A-614 substrate is more important for the efficiency of processing than the presence of the consensus ACCCA sequence.

Further support for this conclusion comes from an examination of the processing of the other insertion mutants (see Table I). For example, the +4 position (an A) is cleaved in addition to the predicted cut in the 3C and 5C insertion mutants. Moreover, in the 5Cswap mutant, where the +4 position changes from an A to a C, no cleavage occurs at position +4. We conclude that the cleavage component of the histone processing machinery prefers, but is not obligated, to cut 3′ of A residues.

Lowered processing efficiency is not correlated with lack of binding of the U7 snRNP to the insertion substrates

The severely decreased processing activity of the longer insertion mutants could arise in several ways. The U7 snRNP might be unable to recognize the substrates when its known binding site, the HDE (Schaufele *et al.*, 1986; Bond *et al.*, 1991), is distanced from the stem−loop. Alternatively, a precise juxtaposition of factors binding to the stem−loop and the U7 snRNP base pairing with the HDE may be essential for productive interactions and efficient cleavage.

To distinguish between these alternatives, binding of the U7 snRNP to the various substrates was assessed by immunoprecipitation with anti-trimethylguanosine (α-TMG) antibodies (Figure 4). Internally ^{32}P-labeled histone mRNA substrates (uncapped) were incubated under processing conditions for 15 min at 30°C to allow complex formation. After exposure to α-TMG antibodies, which recognize the 2,2,7-trimethylguanosine cap structure on U7 snRNA, RNAs were isolated and gel fractionated. Lanes 1−12 of Figure 4A show pellets, while lanes 13−24 contain supernatants. The supernatants both provide a control for degradation of the substrates and monitor the extent of processing during the 15 min incubation. Background precipitation was assessed in two ways: by a complete reaction minus the α-TMG antibody (lanes 12 and 24) and by using two substrates defective in the HDE. One of these was wild-type processed product (lanes 1 and 13), which lacks the entire downstream region (including the HDE); the second was the HDEpu→py mutant (lanes 3 and 15), which has the purines of the HDE altered and shows no detectable cleavage *in vitro* (see Figure 2, lane 1). The processed product was not detectable in the α-TMG precipitate (Figure 4A, lane 1), as expected. In contrast, the HDEpu→py was precipitated to a low level, which is considered below (lane 3).

The wild-type and insertion substrates showed an unexpected but interesting pattern of immunoprecipitation by the α-TMG antibody (Figure 4A, lanes 2−10). The measured binding of each substrate appeared to be inversely proportional to its ability to be processed (Table I). Wild-type and 1C substrates processed with the greatest efficiency, and exhibited the weakest precipitability with α-TMG antibodies (Figure 4A, lanes 2 and 4). Insertion substrates 3C−10C, which process progressively less efficiently, precipitated to greater extents (lanes 5, 6 and 8−10). These results suggested that good substrates may be recognized rapidly and quickly cleaved, leading to dissociation of the processing complex and loss of α-TMG precipitability. This interpretation was substantiated by comparing the

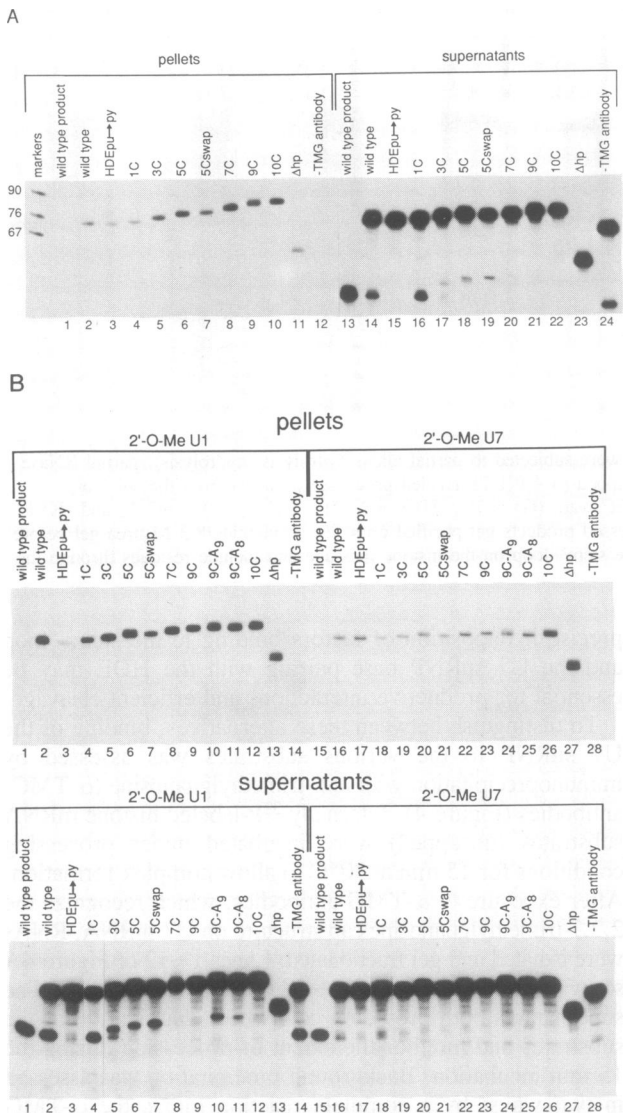

Fig. 4. Anti-TMG immunoprecipitation of labeled histone transcripts. **(A)** Uncapped histone substrates were incubated in HeLa nuclear extract under processing conditions for 15 min at 30°C to allow assembly of complexes, which were then selected with α-TMG antibodies and fractionated on an 8% polyacrylamide/8.3 M urea gel. Lanes 1–12 are pellets, lanes 13–24 are the corresponding supernatants. Lanes 12 and 24 contain reactions which omitted the α-TMG antibody and serve as a measure of background; the reactions in lanes 1 and 13 contained gel-purified wild-type processed product, which is cleaved after nucleotide 34 and does not contain the HDE. **(B)** The reactions were the same as described above except that a 2′-*O*-methyl-oligoribonucleotide was added. Lanes 1–14 show pellets and supernatants from reactions containing an anti-U1 2′-*O*-methyl-oligoribonucleotide; an anti-U7 2′-*O*-methyl-oligoribonucleotide was present in reactions in pellet and supernatant lanes 15–28. Note that supernatant lane 8 is slightly underloaded.

immunoprecipitation of the 5C with 5Cswap and both 9C-A_9 and 9C-A_8 with the 9C substrate. 5Cswap, which is processed to a greater extent than 5C (see Figure 2, compare lanes 5 and 6), was precipitated less efficiently (Figure 4A, lanes 6 and 7; Table I). 9C-A_8, which was processed to a slightly greater extent than 9C, precipitated to about the same extent; 9C-A_9 was processed more efficiently than 9C and precipitated somewhat less efficiently (see Table I). Δhp was also precipitated, although to a slightly lesser extent than very poor processing substrates like 9C and 10C (Figure 4A, compare lane 11 with lanes 9 and 10).

Interpretation of the immunoprecipitation experiment in Figure 4A rests on the assumption that the U7 snRNP is the only trimethylguanosine-containing component in the extract that interacts with our histone substrates. To confirm that precipitation was via the U7 snRNP particle, we added a 2′-*O*-methyl-oligoribonucleotide complementary to 18 nucleotides at the 5′ end of the U7 snRNA to the extract prior to incubation and the α-TMG selection. As a control, another 2′-*O*-methyl-oligoribonucleotide, complementary to the first 12 nucleotides of the U1 snRNA (Seiwert and Steitz, 1993), was used. As anticipated, the blocking of U1 snRNA's 5′ end had little effect on the immunoprecipitability of the various histone substrates (compare Figure 4A, lanes 1–12, with Figure 4B, pellet lanes 1–14) and on the ability of the substrates to be processed (compare Figure 4A, lanes 13–24, with Figure 4B, supernatant lanes 1–14). In contrast, blocking the 5′ end of the U7 snRNA reduced precipitation of the histone substrates to background levels (<4%, Figure 4B, pellet lanes 15–28; Table I) and also interfered with processing (Figure 4B, supernatant lanes 15–28). The only anomalous data were with Δhp, which we expected to precipitate negligibly when the 5′ end of the U7 snRNP was blocked. Its intermediate level of precipitation, which was not lowered by the oligoribonucleotide complementary to either U1 or U7 (see Table I), suggests that perhaps another TMG-containing extract component may bind this severely mutated substrate. The results for all the other mutants argue that immunoprecipitation of the histone pre-mRNA substrates with α-TMG is via the U7 snRNP; they also confirm previous data showing that base pairing of the 5′ end of U7 snRNA with the HDE is essential for substrate binding and processing (Schaufele *et al.*, 1986; Bond *et al.*, 1991; Grimm *et al.*, 1993).

Psoralen cross-linking confirms U7 snRNP association with mutant substrates

To provide additional direct evidence for U7 snRNP interaction with the mutant substrates, psoralen cross-linking experiments were performed. Since psoralen intercalates into double-stranded regions in nucleic acids and covalently cross-links the two strands, close associations between two RNAs can be identified (Cimino *et al.*, 1985). Internally labeled histone pre-mRNAs were incubated for 15 min under processing conditions to allow complex formation and then irradiated on ice in the presence of psoralen. Total RNA was isolated and gel fractionated, revealing two slowly migrating cross-linked bands for the wild-type substrate, as well as for the C-insertion substrates (Figure 5A, odd-numbered lanes).

The identity of the cross-linked bands indicated by dots in Figure 5A was determined by specific RNase H cleavage (even-numbered lanes). Both the slower and faster migrating bands showed altered mobility after treatment in the presence of a deoxyoligonucleotide complementary to U7 RNA, while a 5S rRNA oligonucleotide had no effect (data not shown). Note that the faint bands in lanes 1 and 2 (containing HDEpu→py, which should not bind the U7 snRNP) and the bands not marked by dots in lanes 11–24 are not shifted, indicating that U7 RNA is not part of these species. A

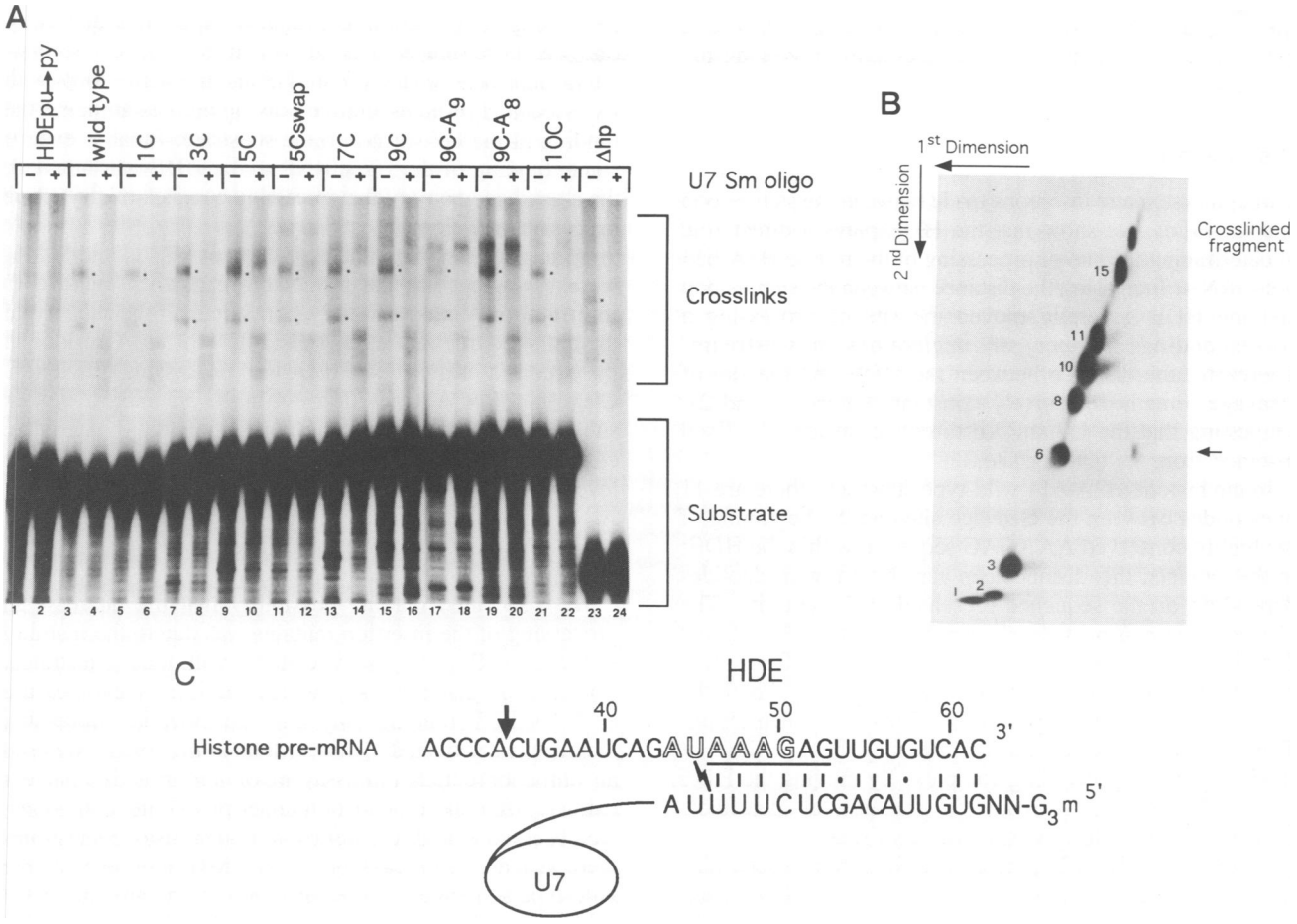

Fig. 5. Identification and mapping of histone pre-mRNA–U7 RNA psoralen cross-links. (**A**) Various histone substrates were psoralen cross-linked and then RNase H digested in the absence (odd-numbered lanes) or presence (even-numbered lanes) of a deoxyoligonucleotide complementary to the Sm site of U7 snRNA (U7 Sm oligo). Dots mark U7-specific cross-links which show altered mobility after RNase H digestion. (**B**) The slower migrating U7-specific cross-linked species of the 9C-insertion substrate was gel purified, RNase T_1 digested, and fractionated on a two-dimensional gel containing 20% polyacrylamide/8.3 M urea as described in Materials and methods. The substrate was labeled with [α-^{32}P]UTP and [α-^{32}P]ATP. The sizes of the RNase T_1 fragments, including the RNase T_1 fragment containing the covalently attached U7 fragment, are marked. Identification of the fragments was based on fractionation of T_1 products in one dimension and on the expected relative intensity of single-labeled transcripts using α-^{32}P-labeled ATP, CTP or UTP; this showed that the terminal 4mer (not designated) migrated just below the 8mer. The arrow marks the RNase T_1 fragment generated by photoreversal. (**C**) Diagram of the deduced psoralen cross-linking site. Partial sequences of the mouse histone H2A-614 and the human U7 snRNP are shown. The purine stretch of the HDE is underlined. The histone substrate RNase T_1 fragment which contains the cross-link to U7 RNA is shown in bold, with a lightning blot representing the bases proposed to be involved in the cross-link.

deoxyoligonucleotide complementary to the 5′ end of the histone pre-mRNA targeted both U7-cross-linked species for RNase H cleavage (not shown), demonstrating that pre-mRNA and not processed histone message is involved. These results confirm our conclusion that the C-insertion substrates and the Δhp mutant interact with the U7 snRNP, despite their impaired ability to undergo processing.

To map the U7-specific cross-links on the histone pre-mRNA, two-dimensional gel analysis was performed. The two cross-linked bands seen with the wild-type and 9C insertion substrates (see Figure 5A, lanes 3 and 15) were isolated and subjected to complete RNase T_1 digestion. After fractionation in the first dimension, the psoralen cross-links were reversed by irradiation with 254 nm light and the resulting products were gel fractionated in the second dimension. A cross-linked RNase T_1 fragment migrates slower than expected in the first dimension, but after photoreversal runs true to size in the second dimension.

When the slower migrating U7-specific cross-linked species were analyzed, a spot corresponding to a six-nucleotide T_1 fragment migrated off the diagonal for both the wild-type (not shown) and 9C insertion mutant substrate (Figure 5B). Whereas two six-nucleotide T_1 fragments occur within the wild-type substrate, the 9C insertion substrate contains only one 6-mer, with the other converted to a 15-nucleotide fragment because of the 9C insertion (see Figure 1). These analyses therefore allowed localization of the cross-link to a six-nucleotide fragment which overlaps the HDE, the known site of interaction with U7 RNA; the only pyrimidine (a U) within this 6mer is in an ideal configuration for cross-linking (Figure 5C). Unfortunately, we were unable to map the faster migrating U7-specific cross-link. However, RNase H digestion targeted by three different deoxyoligonucleotides complementary to the histone pre-mRNA (see Materials and methods) suggested that this cross-link occurs close to the 3′ end of the substrate. A 2′-*O*-methyl-oligonucleotide complementary to the 5′ end of U7 RNA blocked the formation of both U7-specific cross-linked

species (data not shown), providing further evidence that cross-linking is dependent on an interaction involving the 5′ end of U7 RNA.

Discussion

Through systematic insertions made between the stem–loop and the HDE, we show that the HDE plays a direct role in determining the site of processing of the mouse H2A-614 pre-mRNA. Increasing the distance between the stem–loop and the HDE generally moved the site of processing a corresponding number of nucleotides downstream. Therefore, the distance between the HDE and the site of cleavage remained relatively constant (Figures 1 and 2), suggesting that the U7 snRNP directs cleavage at a fixed distance from its binding site.

In the histone H2A-614 wild-type substrate, there are 11 nucleotides between the cleavage site and the beginning of the highly conserved AAAGAG sequence within the HDE; in the mutants, this distance can vary by a few nucleotides depending on the sequence presented at the cut site. The cleavage component is clearly capable of cutting 3′ to C and G residues, even though an A is preferred. The 3C insertion mutant was interesting in that cleavage occurred at both the wild-type cut site and at positions +3 and +4 downstream. Thus, a span of three nucleotides may be the limit of flexibility of the processing complex. By moving the HDE five nucleotides downstream (5C mutant), all detectable cleavage at the wild-type site was abolished.

By binding to the HDE, the U7 snRNP serves to determine the cut site in the pre-mRNA; however, the U7 RNA itself is probably not directly involved in cleavage. The only sequences in U7 RNA that are highly conserved across species are the Sm binding site and the 5′ end (Strub *et al.*, 1984; De Lorenzi *et al.*, 1986; Mowry and Steitz, 1987b; Cotten *et al.*, 1988; Soldati and Schümperli, 1988; Gruber *et al.*, 1991; Phillips and Turner, 1991; Phillips and Birnstiel, 1992; Watkins *et al.*, 1992), which has been shown to base pair with the HDE. The conserved nucleotides at the 5′ end of U7 snRNA can be replaced as long as compensatory changes are introduced into the HDE of the substrate (Schaufele *et al.*, 1986; Bond *et al.*, 1991). The Sm site is likely to be bound by the core Sm proteins (Smith *et al.*, 1991), even though its sequence differs slightly (Mowry and Steitz, 1987b; Cotten *et al.*, 1988; Soldati and Schümperli, 1988; Gruber *et al.*, 1991; Phillips and Turner, 1991) from the Sm consensus sequence found in the spliceosomal snRNAs. In fact, this atypical Sm site appears to be important for activity of the U7 snRNP because when it is replaced by the consensus Sm site of the major snRNAs, no detectable levels of histone mRNA 3′ end processing were observed in the *Xenopus* oocyte (Grimm *et al.*, 1993). Finally, the 3′ terminal stem of U7 snRNA is conserved only in secondary structure, and can be mutated if some base pairing is maintained (Gilmartin *et al.*, 1988). Thus, it appears likely that a snRNP-associated protein(s), rather than a conserved and available sequence within the U7 snRNA itself, comprises the catalytic component of the 3′ end processing machinery.

The decrease in the processing efficiency that occurred as the HDE was moved away from the stem–loop (Table I, column 2) could have arisen in several ways. First, the processing component might require a specific sequence of one or more nucleotides at which to cleave. Second, interactions between the U7 snRNP and the stem–loop (with its associated proteins) may be disrupted, destabilizing the binding of the U7 snRNP to the substrate. Finally, even if stable binding of the U7 snRNP to the HDE was attained, the altered geometry of the processing complex might reduce processing efficiency.

To test the first hypothesis, 5Cswap, 9C-A_9 and 9C-A_8 were assessed for their processing efficiency relative to insertion mutations with only C residues in the vicinity of the cut site. In every case, substitution of an A residue at the expected cleavage site increased processing efficiency. Cleavage 3′ of A at position +4 in the 3C and 5C mutant substrates further supports our conclusion that the cleavage component prefers cutting downstream of A. Yet, wild-type levels of processing were not obtained with 5Cswap, 9C-A_9 or 9C-A_8, suggesting that neither an A residue alone, nor the ACCCA sequence, is sufficient to achieve maximal efficiency.

We also ruled out the possibility that the decrease in processing of the insertion mutants was due to the inability of the U7 snRNP to associate stably with these substrates. Rather than a gel-shift assay, which provides a clear measure of U7 snRNP binding only after an essential processing factor(s) is inactivated by heat (Melin *et al.*, 1992), we used an immunoprecipitation assay involving an active nuclear extract. After short incubation under processing conditions, we observed that all the insertion mutant histone substrates were at least as precipitable by α-TMG antibodies as the wild-type substrate (Figure 4A). In fact, the amount of U7 snRNP bound to the substrate and the processing efficiency of the mutant seemed to be inversely proportional. It appeared that with good substrates for processing, cleavage occurred and substrates were quickly released from the U7 snRNP, therefore losing their immunoprecipitability. With the poorer substrates, the U7 snRNP remained associated longer, even though effective contacts were not made. This explanation agrees with previous observations of Mowry and Steitz (1987a); in RNase protection experiments, the pre-mRNA substrate that processed less efficiently produced higher yields of snRNP-protected fragment.

In addition to the immunoprecipitation experiments, psoralen cross-linking studies provided direct evidence for the association of U7 RNA and the histone insertion substrates. RNase T1 digestions and two-dimensional gel analysis showed that one histone pre-mRNA–U7 RNA cross-link occurred within a six-nucleotide fragment which includes part of the HDE. This result confirms earlier genetic analyses of Schaufele *et al.* (1986) and Bond *et al.* (1991), which localized the base-pairing interaction on the U7 snRNA and histone pre-mRNA substrate.

Multiple elements, both in the pre-mRNA and *trans*-acting factors, contribute to the processing of histone substrates. As previously shown, the stem–loop, the HDE in the pre-mRNA, the U7 snRNP and a heat-labile activity are all required for efficient processing (for reviews, see Birnstiel and Schaufele, 1988; Mowry and Steitz, 1988; Marzluff, 1992). Our results indicate that two additional features of the substrate are important: the distance between the stem–loop and the HDE, and the nucleotide at the cleavage site. Clearly, the proper juxtaposition of all these elements

and factors they bind is critical for achieving processing. Fine tuning of these interactions is likely to play a role in determining the levels of one type of histone message with respect to others present in the cell.

There are other RNA processing events in which cleavage occurs at a fixed distance from a recognition element in the substrate. For example, in *Xenopus laevis* and *Saccharomyces cerevisiae* tRNA splicing, the choice of splice sites is determined in part by the length of the anticodon stem (Mattoccia *et al.*, 1988; Reyes and Abelson, 1988; Baldi *et al.*, 1992). In *Escherichia coli*, RNase III also seems to apply a measuring system to determine its site of cleavage (Krinke and Wulff, 1990), which occurs 10–14 nucleotides from the 3′ end of the double-stranded stem recognized by the enzyme. In contrast, in pre-mRNA splicing, reactive nucleotides lie within the sequences bound by snRNPs (for reviews, see Green, 1991; Guthrie, 1991; Moore *et al.*, 1993).

The mechanism of polyadenylation of most RNA polymerase II transcripts exhibits interesting parallels with histone mRNA 3′ end processing. Polyadenylated and histone pre-mRNAs possess recognition elements both upstream and downstream of the cleavage site. In polyadenylation substrates, a highly conserved upstream AAUAAA sequence, and less rigidly conserved downstream elements rich in GU or U, are important (for a review, see Wahle and Keller, 1992). Cleavage is executed by a protein factor(s). Moreover, an A is found at the cleavage site, with cleavage occurring after CA in 59% of all polyadenylated transcripts (Sheets *et al.*, 1990). This situation is similar to mammalian histone genes, where in 45% the sequence ACCCA is found downstream of the stem–loop (for a compilation of sequences, see Bond *et al.*, 1991); however, in most cases the cleavage site has not been precisely mapped. It may be that common factors are involved in 3′ end cleavage of both histone and polyadenylated mRNAs.

Materials and methods

Construction of mouse H2A-614 mutants

A fragment of the mouse H2A-614 gene (Hurt *et al.*, 1989) was the wild-type substrate [HDEwt in Bond *et al.* (1991)]. Clones containing mutations were constructed (Maniatis *et al.*, 1982) using overlapping deoxyoligonucleotides, which were synthesized on an Applied Biosystems DNA synthesizer (John Flory, Yale University). Template deoxyoligonucleotides had the sequences 5′ GCGCGTCTAGAACACTCTTACTTTTGGGGCAGGGCCAGGAAAGGCCACCAAGACCGGCTACCGTGACACAACTCTTTATCTGATXTCAGTGGGTGGCTCTGAAAAGAGCCTTTTTGGGAGCT 3′ (where X = 1, 3, 5, 7 or 9 C residues) or 5′ GCGCGTCTAGAACACTCTTACTTTTGGGGCAGGGCCAGGAAAAGCCACCAAGACCGGCTACCGTGACACAACTCTTTATCTGATGGGGTGGGTGTCAGGGCTCTGAAAAGAGCCTTTTTGGGAGCT 3′ for 5Cswap. The primer used was 5′ CCCAAAAAGGCTCTTTTCAG 3′. The deoxyoligonucleotide used to generate Δhp was 5′ GCGCGTCTAGAACACTCTTACTTTTGGGGCAGGGCCAGGAAAGGCCACCAAGACCGGCTACCGTGACACAACTCTTTATCTGATTCAGTGGGTTTTTTGGGAGCT 3′, with a primer sequence of 5′ CCCAAAAAACCCACTGAATC 3′. The primers (100 ng) bound to their respective first-strand deoxyoligonucleotides (600 ng) were extended under polymerase chain reaction (PCR) conditions (Maniatis *et al.*, 1982) with an annealing temperature of 55°C. The PCR products were gel purified on a 5% polyacrylamide/8.3 M urea gel, digested with *Xba*I, and ligated into Bluescript SK+ (Stratagene), which had been digested with *Sac*I and *Xba*I. Sequences were confirmed by dideoxy sequencing. The 10C mutant was isolated while screening clones for 9C. 9C-A_9 and 9C-A_8 were generated by Kunkel mutagenesis using the *E.coli* 236 strain (*dut*$^-$*ung*$^-$) (Kunkel *et al.*, 1987). The deoxyoligonucleotides used to generate 9C-A_9 and 9C-A_8 were 5′ CTGATGGGGTGGGGTCAGTGGG 3′ and 5′ CTGATGGGGTGGGTCAGTGGG 3′, respectively. HDEpu→py was described by Bond *et al.* (1991).

Transcription templates were generated by digestion with *Mae*III and T3 run-off transcripts produced as described in Melton *et al.* (1984). All transcripts contain 14 nucleotides of vector sequence at their 5′ ends: 5′ GGGAAGAAAAGCUG 3′.

Nuclear extracts and processing reactions

Nuclear extracts were prepared from HeLa cells as described by Dignam *et al.* (1983), and including modifications of Heintz and Roeder (1984). HeLa cells were obtained from the National Cell Culture Center (Coon Rapids, MN). Processing reactions contained 60% (v/v) nuclear extract, 80 mM KCl (final concentration), 20 mM EDTA (final concentration), 0.5 μg yeast total RNA (Worthington Biochemical Corporation) and 1×10^{-15} mol (2×10^9 c.p.m./nmol) of substrate RNA. Reactions were performed in a final volume of 10 μl for 1 h at 30°C. Tris–HCl (pH 7.5), EDTA, SDS and proteinase K were added to a final concentration of 10 mM, 25 mM, 0.5% and 1 μg/μl, respectively, and incubated at 65°C for 15 min. Isolated RNA was then analyzed on an 8% polyacrylamide/8.3 M urea gel.

3′ end mapping of cleavage sites

[γ-^{32}P]GTP-labeled transcripts were subjected to alkaline hydrolysis and partial digestion with RNase U_2 and T_1 as described by Pharmacia. These digested substrates, along with internally [α-^{32}P]UTP-labeled gel-purified processed products, were fractionated on a 20% polyacrylamide/8.3 M urea gel.

Anti-trimethylguanosine immunoprecipitations

Anti-trimethylguanosine antibody (0.1 μg/10 μl processing reaction) and rabbit anti-mouse IgG (Oncogene Science) (5 μl/10 μl processing reaction) were incubated with protein A–Sepharose (Pharmacia) (4 mg/10 μl processing reaction) for 1 h at room temperature with mixing. The protein A–Sepharose mixture was then washed three times with 500 μl NET-2 buffer [50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 0.05% (v/v) Nonidet P-40] and then 200 μl NET-2 were added. Internally [α-^{32}P]UTP-labeled histone substrates (4×10^{-15} mol) were incubated under the processing conditions described above (final volume of 40 μl) for 15 min at 30°C and then added to antibody-bound protein A–Sepharose. Samples were incubated for 1 h at 4°C with mixing, washed four times with 500 μl cold NET-2, extracted with phenol–chloroform–isoamyl alcohol (50:50:1), ethanol precipitated and fractionated on an 8% polyacrylamide/8.3 M urea gel.

2′-*O*-Methyl RNA oligonucleotides (35 ng) were incubated under processing conditions (minus the histone substrate) in a 40 μl reaction for 10 min. Then the histone substrate was added and the reaction incubated for an additional 15 min at 30°C. The 2′-*O*-methyl RNA oligonucleotides were purchased from Oligos etc.: anti-U1, 5′ UGCCAGGUAAGUAUYYYYA 3′; anti-U7, 5′ CUAAAAGAGCUGUAACACBBBB 3′ (where Y indicates modified dC residues and B indicates modified 2′-dC residues).

Psoralen cross-linking assays

[^{32}P]UTP-labeled histone pre-mRNAs were incubated under processing conditions as described above in a total of 20 μl for 15 min at 30°C. Then 0.5 μg of psoralen (4′-aminomethyl-4,5′,8-trimethylpsoralen; HRI Associates) was added to a final volume of 50 μl and irradiated with 365 nm light for 15 min on ice (Hausner *et al.*, 1990). RNA was isolated as described above and fractionated on an 8% polyacrylamide/8.3 M urea gel. Two-dimensional gels and photoreversal were performed as described in Wassarman and Steitz (1992).

RNase H digestion

RNase H digestions of gel-purified cross-linked RNAs were performed in 40 mM Tris–HCl (pH 7.5), 50 mM KCl, 5 mM $MgCl_2$, 0.5 μg of deoxynucleotide U7 Sm (CCTACTAGACAAATT) and 2 U RNase H (Boehringer Mannheim) for 1 h at 37°C. The deoxyoligonucleotides used to target the histone pre-mRNA were as follows: anti-hist 1 (GGAGCTCCAGCTTTTGTTCCCBBBB, where B indicates an abasic biotin) complementary to nucleotides 1–7 (Figure 1), anti-hist 2 (GATTCAGTGGGT) complementary to nucleotides 30–41 and anti-hist 3 (GTGACACAACTCTTTA) complementary to nucleotides 45–60.

RNase T_1 digestions of gel-purified cross-linked RNA were performed as described in Wassarman and Steitz (1993), except that incubations were for 1 h.

Acknowledgements

We thank Vic Myer, Brenda Peculis, Karen Montzka-Wassarman, David Wassarman, Erik Sontheimer, David Toczyski, Kazio Tycowski and the

rest of the Steitz lab for advice and discussions. We also thank Brenda Peculis and Vic Myer for critical reading of the manuscript. This research was supported by grant GM26514 from the National Institutes of Health.

References

Baldi,M.I., Mattoccia,E., Bufardeci,E., Fabbri,S. and Tocchini-Valentini,G.P. *Science*, **255**, 1404–1408.

Birnstiel,M.L. and Schaufele,F.J. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer-Verlag, New York, pp. 155–182.

Bond,U.M., Yario,T.A. and Steitz,J.A. (1991) *Genes Dev.*, **5**, 1709–1722.

Cimino,G.D., Gamper,H.B., Isaacs,S.T. and Hearst,J.E. (1985) *Annu. Rev. Biochem.*, **54**, 1151–1193.

Cotten,M., Octavian,G., Vasserot,A., Schaffner,G. and Birnstiel,M.L. (1988) *EMBO J.*, **7**, 801–808.

De Lorenzi,M., Rohrer,U. and Birnstiel,M.L. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 3243–3247.

Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.

Galli,G., Hofstetter,H., Stunnenberg,H.G. and Birnstiel,M.L. (1983) *Cell*, **34**, 823–828.

Georgiev,O. and Birnstiel,M.L. (1985) *EMBO J.*, **4**, 481–489.

Gick,O., Kramer,A., Vasserot,A. and Birnstiel,M.L. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 8937–8940.

Gilmartin,G.M., Schaufele,F., Schaffner,G. and Birnstiel,M.L. (1988) *Mol. Cell. Biol.*, **8**, 1076–1084.

Green,M.R. (1991) *Annu. Rev. Cell Biol.*, **7**, 559–599.

Grimm,G., Stefanovic,B. and Schümperli,D. (1993) *EMBO J.*, **12**, 1229–1238.

Gruber,A., Soldati,D., Burri,M. and Schümperli,D. (1991) *Biochim. Biophys. Acta*, **1088**, 151–154.

Guthrie,C. (1991) *Science*, **253**, 157–163.

Hausner,T.-P., Giglio,L.M. and Weiner,A.M. (1990) *Genes Dev.*, **4**, 2146–2156.

Heintz,N. and Roeder,R.G. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 2713–2717.

Hurt,M.M., Chodchoy,N. and Marzluff,W.F. (1989) *Nucleic Acids Res.*, **17**, 8876.

Krinke,L. and Wulff,D.L. (1990) *Nucleic Acids Res.*, **18**, 4809–4815.

Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.

Lüscher,B. and Schümperli,D. (1987) *EMBO J.*, **6**, 1721–1726.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Marzluff,W.F. (1992) *Gene Expr.*, **2**, 93–97.

Mattoccia,E., Baldi,I.M., Gandini-Attardi,D., Ciafre,S. and Tocchini-Valentini,G.P. *Cell*, **55**, 731–738.

Melin,L., Soldati,D., Mital,R., Streit,A. and Schümperli,D. (1992) *EMBO J.*, **11**, 691–697.

Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. *Nucleic Acids Res.*, **12**, 7035–7056.

Mital,R., Albrecht,U. and Schümperli,D. (1993) *Nucleic Acids Res.*, **21**, 1049–1050.

Moore,M.J., Query,C.C. and Sharp,P.A. (1993) In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.

Mowry,K.L. and Steitz,J.A. (1987a) *Mol. Cell. Biol.*, **7**, 1663–1672.

Mowry,K.L. and Steitz,J.A. (1987b) *Science*, **238**, 1682–1687.

Mowry,K.L. and Steitz,J.A. (1988) *Trends Biochem. Sci.*, **13**, 447–451.

Mowry,K.L., Oh,R. and Steitz,J.A. (1989) *Mol. Cell. Biol.*, **9**, 3105–3108.

Phillips,S.C. and Birnstiel,M.L. (1992) *Biochim. Biophys. Acta*, **1131**, 95–98.

Phillips,S.C. and Turner,P.C. (1991) *Nucleic Acids Res.*, **19**, 1344.

Reyes,V.M. and Abelson,J. (1988) *Cell*, **55**, 719–730.

Schaufele,F., Gilmartin,G.M., Bannworth,W. and Birnstiel,M.L. (1986) *Nature*, **323**, 777–781.

Seiwert,S.D. and Steitz,J.A. (1993) *Mol. Cell. Biol.*, **13**, 3135–3145.

Sheets,M.D., Ogg,S.C. and Wickens,M.P. (1990) *Nucleic Acids Res.*, **18**, 5799–5805.

Smith,H.O., Tabiti,K., Schaffner,G., Soldati,D., Albrecht,U. and Birnstiel,M.L. *Proc. Natl Acad. Sci. USA*, **88**, 9784–9788.

Soldati,D. and Schümperli,D. (1988) *Mol. Cell. Biol.*, **8**, 1518–1524.

Streit,A., Konig,T.W., Soldati,D., Melin,L. and Schümperli,D. (1993) *Nucleic Acids Res.*, **21**, 1569–1575.

Strub,K. and Birnstiel,M.L. (1986) *EMBO J.*, **5**, 1675–1682.

Strub,K., Galli,G., Busslinger,M. and Birnstiel,M.L. (1984) *EMBO J.*, **3**, 2801–2807.

Vasserot,A.P., Schaufele,F.J. and Birnstiel,M.L. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 4345–4349.

Wahle,E. and Keller,W. (1992) *Annu. Rev. Biochem.*, **61**, 419–440.

Wassarman,D.A. and Steitz,J.A. (1992) *Science*, **257**, 1918–1925.

Wassarman,K.M. and Steitz,J.A. (1993) *Genes Dev.*, **7**, 647–659.

Watkins,N.J., Phillips,S.C. and Turner,P.C. (1992) *Gene*, **120**, 271–276.

Wells,D.E. (1986) *Nucleic Acids Res.*, **14**, r119–r148.

Received on September 27, 1993; revised on February 25, 1994