Nucleic Acids Research, 1998, Vol. 26, No. 21 **4965–4974**

A specific RNA–protein interaction at yeast polyadenylation efficiency elements

Shaoxiong Chen and Linda E. Hyman*

Department of Biochemistry, Tulane University School of Medicine, 1430 Tulane Avenue, New Orleans, LA 70112-2699, USA

Received June 12, 1998; Revised and Accepted September 18, 1998

ABSTRACT

The specific RNA–protein interactions responsible for the production of mature 3′ ends of eukaryotic mRNAs are not well understood. Sequence elements at the 3′ ends of yeast genes have been identified that specify the position of the poly(A) site and the efficiency of polyadenylation. To provide additional insights into the interaction between important sequences that direct 3′-end formation *in vivo* and nuclear proteins, we utilized gel mobility shift assays and UV-crosslinking studies. The data indicate that a protein, with an apparent molecular weight of 80 kDa, interacts specifically with pre-mRNA at the $(UA)_3$ efficiency element. Although the interaction is specific, it can be competed by RNA sequences that do not contain the same type of efficiency element; that is, a sequence lacking a $(UA)_3$ repeat. This result implies that the protein binding site is flexible. Using immunoprecipitation techniques, the protein has been identified as Hrp1, a heteronuclear RNA binding protein. The role of Hrp1p in 3′-end formation including RNA processing and transcription termination is addressed.

INTRODUCTION

The 3′ ends of eukaryotic mRNA are marked by a polyadenylate [poly(A)] tail. The poly(A) is added onto the pre-mRNA post-transcriptionally as a result of a two step process: endonucleolytic cleavage of the pre-mRNA followed by the polymerization of the adenylate residues (for reviews see 1,2). Significant progress has been made in understanding the machinery that directs these pre-mRNA processing events *in vitro*. In both mammalian cells and yeast, the cleavage and polyadenylation reactions can be performed using pre-mRNA substrates and cellular extracts (3,4).

In yeast, additional genetic approaches have been utilized to identify components of the processing machinery (5–9). Taken together, the current understanding of the cleavage reaction indicates that three factors (CF IA, CF IB and CF II) are required (10,11). The proteins that constitute each factor have been identified and many are homologous to proteins identified in mammalian cells. The polyadenylation reaction is also well understood. The poly(A) polymerase (PAP) has been identified from both mammalian cells and yeast (12,13). In yeast, in addition to PAP, a polyadenylation factor I (PF I), a poly(A) binding protein (PAB I), and a third factor, CF I, are required for poly(A) addition. The identification of the components of the processing reactions is a first step toward understanding the mechanism by which the processing machinery recognizes the poly(A) site and allows the poly(A) polymerase to add the tail onto the mature end of the mRNA.

In order to fully understand the mechanism of 3′-end formation it is also necessary to identify the *cis*-acting signals that determine the site of cleavage and polyadenylation, as well as the efficiency of this process. A comparison of sequences surrounding the poly(A) site of a large number of yeast genes does not reveal a highly conserved signal for cleavage and polyadenylation. However, mutational analysis revealed that a tri-partite signal exists at the ends of yeast genes (14,15). This signal consists of an element that influences the efficiency of the reaction (usually a UA-rich element), followed by a sequence that directs the position of the cleavage (usually an A-rich element), followed by the poly(A) site itself (usually PyA). Although not found uniformly in all yeast genes in this simple arrangement, a synthetic signal consisting of the efficiency, positioning and poly(A) site elements is necessary and sufficient to direct 3′-end formation *in vivo*.

Insight into the role of the efficiency element has recently been gained by the demonstration that a GST–Hrp1 (heterogeneous nuclear ribonucleoprotein, the sole component of CF IB) fusion protein binds to a pre-cleaved RNA *in vitro* at the UA element, which is essential for 3′-end formation *in vivo* (16). Additional proteins such as RNA 15p, and the 105 kDa protein in CF II, can also interact with the 3′-end RNA sequence (5,10,11). Although the RNA 15p binding site is unknown, the 105 kDa protein interacts with a UA-sequence in pre-mRNA only in the presence of ATP. It is not clear why both GST–Hrp1 and 105 kDa proteins binds the same UA element *in vitro*.

Although much has been learned regarding the mechanism of RNA processing at the 3′ end using *in vitro* assays, it is important to note that in these studies, the RNA processing events are uncoupled from transcription by RNA polymerase II. However, several lines of evidence suggest that transcription and 3′-end formation are coupled to each other *in vivo*. For example, in the absence of termination, downstream promoter elements may be adversely affected by readthrough from an upstream promoter

*To whom correspondence should be addressed. Tel: +1 504 584 2941; Fax: +1 504 584 2739; Email: lhyman@mailhost.tcs.tulane.edu

(17–20). Mutation of the *cis*-acting signal, critical for cleavage and polyadenylation in mammalian cells and yeast results in aberrant termination of transcription by RNA polymerase II (21–24). However the most direct evidence for coupling of the processing machinery with transcription termination comes from two recent observations. First, mutations in the known yeast processing factors Rna14, Rna15 and Pcf11 result in increased readthrough transcription (25). Secondly, processing factors, including those that constitute the mammalian CPSF (cleavage and polyadenylation specificity factor) are associated with the phosphorylated C-terminal domain of RNA polymerase II (26–28). This suggests a model whereby processing factors are loaded onto the polymerase early in the transcription process, and 'unloaded' following synthesis of the polyadenylation signal at the end of a gene.

In order to fully examine the process of 3′-end formation, we reasoned that an *in vitro* system that is competent in transcription and 3′-end formation may provide new insights into the mechanism of 3′-end formation. In contrast to the highly purified protein factors that process exogenously added pre-mRNA, nuclear extracts synthesize the pre-mRNA *de novo*, and may utilize factors that are specific to both polyadenylation and transcription termination. Using nuclear extracts, we demonstrated previously that RNA polymerase II recognizes termination sites, resulting in the production of transcripts that end ~100 nt downstream of the polyadenylation site (24). These transcripts are not polyadenylated and are suggestive of a pre-mRNA product of transcription that is neither cleaved nor polyadenylated. For this reason, we have chosen to examine the RNA–protein interactions important for 3′-end formation and transcription termination in a system that may more closely mimic the events that occur *in vivo*.

MATERIALS AND METHODS

Reagents and enzymes

Protease K, RNasin, rNTP, dNTP, and all restriction enzymes were purchased from Promega. Vent DNA polymerase, T4 DNA ligase and Klenow fragment were from New England Biolabs. T1 RNase was from Boehringer Mannheim. RNase A was from Sigma. [^{32}P]UTP was from Amersham. All chemicals were supplied by Fisher. Plasmid pT7T319U was from Pharmacia.

Plasmid construction

Plasmid pL101 and *ADH2* 3′-end sequence was described previously (23) and is shown in Figure 1A. The plasmid pBEVY-U (Fig. 1A) was kindly provided by Dr Charles Miller and described (29). It is a yeast 2 µm plasmid containing two promoters for bi-directional cloning and expression in yeast. One promoter, from the *ADH1* gene, is followed by the *ADH2* 3′ termination. The other promoter *GPD* is followed by the *ADH1* 3′ termination signal. This *ADH1* 3′-end sequence is sufficient for 3′-end formation and has been described previously (30). Plasmid p1GA (map not shown) was constructed on the basis of pBEVY-U by removing the *ADH2* 3′ end using *Sma*I and *Xho*I, filling in the ends with Klenow fragment and ligating the blunt ends with T4 DNA ligase. Plasmid pGA101 (map not shown) was made by inserting the *Bam*HI–*XhoI ADH2* 3′-end fragment (amplified by PCR with restriction enzyme-containing site primers PM1, 5′-GCGCTGCAGGATCCGACACTTCTAAATAAGCGG-3′ and PM2, 5′-CCGCTCGAGGGCATGCGAAGGAAAATGAG-3′) into *Bam*HI–*SalI* sites of p1GA. Therefore, ADH 2 3′ end was located between the GPD promoter and the *ADH1* 3′ end in plasmid pGA101. Plasmid p2GA was obtained by removing the *ADH1* 3′ end from pGA101 using *Pst*I and *Pvu*II, filling in the ends and ligating the blunt ends. Plasmid pL407 is the same as pL101 except that the *ADH2* 3′ end contains TATATA(74–79) element deletion. Plasmid p2GAdel(TA)$_3$ is the same as p2GA except that the *ADH2* 3′ end contains TATATA(74–79) element deletion.

Synthesis of *ADH2* 3′ end antisense probes

The oligodeoxynuclotides PM3, 5′-GACACTTCTAAATAAGCGG-3′; PM4, 5′-GCGCTAATACGACTCACTATAGGGCAAACGCGGTGGGAGC-3′; PM5, 5′-GTGTTCGTTATGTACGGC-3′; and PM6, 5′-GCGCTAATACGACTCACTATAGGGCATGCGAAGGAAAATG-3′ were synthesized. The underlined DNA sequence is the T7 promoter. The template for synthesis of *ADH2* 3′-end antisense RNA probe A (Fig. 1B) was made by amplifying upstream *ADH2* 3′ end with primers PM3 and PM4. The template for synthesis of *ADH2* 3′ end antisense RNA probe B (Fig. 1B) was made by amplifying downstream *ADH2* 3′ end with primers PM5 and PM6. The PCR fragments were transcribed by T7 RNA polymerase in the presence of [^{32}P]UTP and RNAs were purified.

Yeast transformation

Plasmids pL101, pL407, p1GA, p2GA and pBEVY-U were transformed into yeast strain YPH 98 (*MATa ade2-101 lys2-801 ura3-52 trp1-1 leu2-1*). Yeast transformation was performed as described (31). Cells were streaked onto complete minimal media plates without uracil and with dextrose (CM-URA DEX plates).

Yeast total RNA preparation, northern blot analysis and RT–PCR mapping of mRNA poly(A) sites

Yeast cells containing p1GA, p2GA and pBEVY-U were grown in 5 ml of CM-URA Dex to an OD_{600} of ~0.8–1.0. Yeast cells containing pL101 and pL407 were grown in 5 ml of complete minimal media without uracil and with galactose to an OD_{600} ~0.5. Yeast total RNA was prepared according to the described protocol (32). RNAs were separated by electrophoresis on 1.5% agarose-formaldehyde gels and northern analyses were carried out as described (23). Mapping the 3′ ends of mRNAs was described previously (33).

Preparation of yeast nuclear extracts and *in vitro* transcription

Yeast nuclear extracts were prepared from strain BJ926 (*MATa/a trp1/+prc1-126/prc1-126 pep4-3/pep4-3 prb1-1122/prb1-1122 can1/can1*) as described (34). The resulting nuclear extracts were stored at –80°C. Protein concentration was determined using Bio-Rad protein assay. For each *in vitro* transcription reaction, 130 µg of extracts and 0.5 µg supercoiled DNA template were incubated in a total volume of 50 µl. The *in vitro* transcription reaction and RNA extraction were described previously (35). *In vitro* transcribed RNAs were analysed by northern blot.

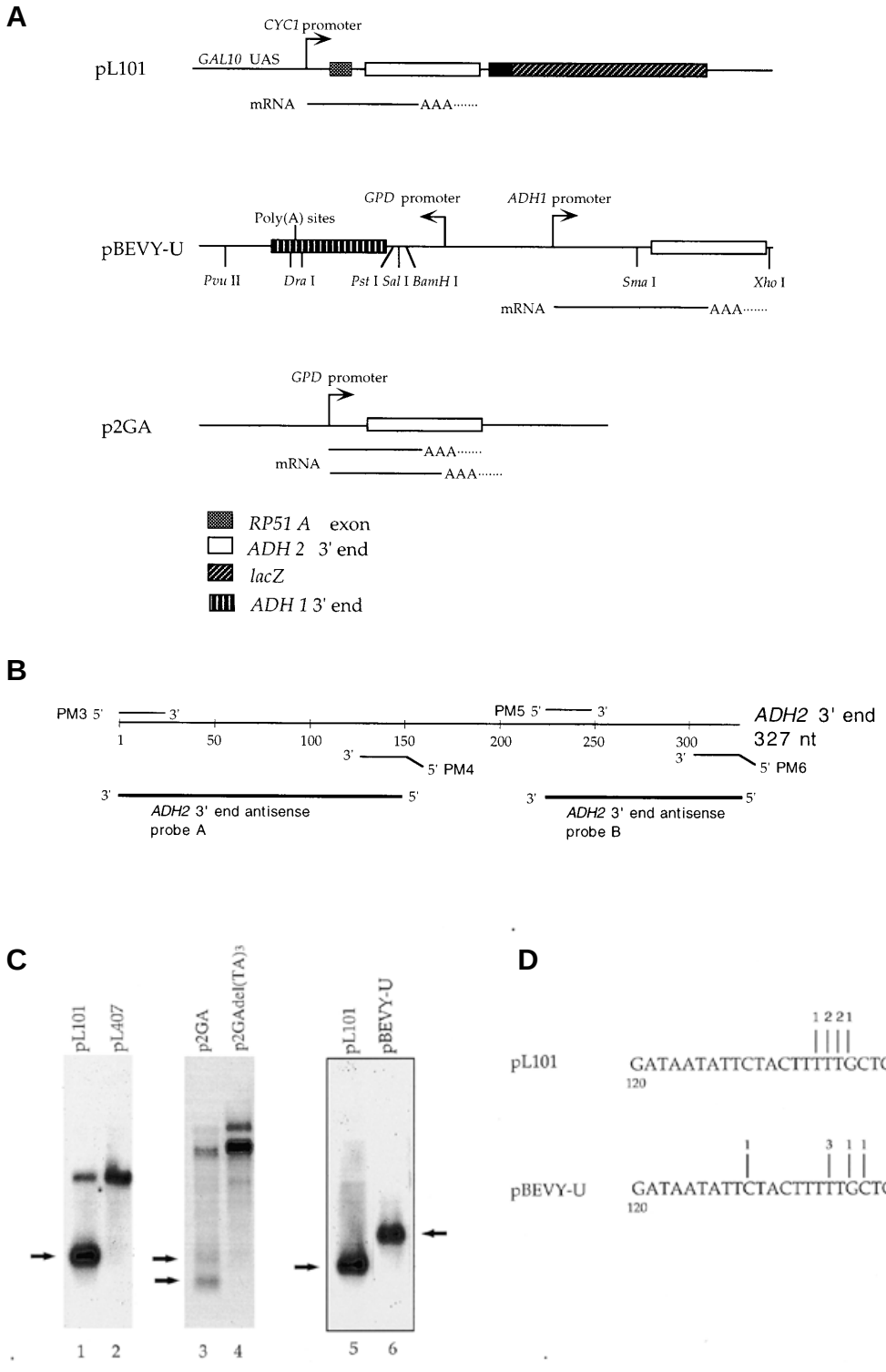

Figure 1. Plasmids used in this study and *in vivo* RNA analysis. (**A**) Schematic diagram of the three reporter plasmids (not to scale). Plasmid pL101 contains the *ADH2* 3′ end inserted into intron of the *RP51 A* gene under the control of *CYC1* promoter and the *GAL* UAS. Plasmids pBEVY-U and p2GA contains the *ADH2* 3′ end under the control of the *ADH1* and *GPD* promoters, respectively. (**B**) *ADH2* 3′-end antisense probes A and B indicated by thick lines. (**C**) *In vivo* RNAs were analyzed by northern blot using *ADH2* 3′-end antisense RNA probe A. Arrows indicate the *in vivo* RNA products. Lanes 1 and 2, 10 µg RNA extracted from yeast cells containing plasmids pL101 and pL407, respectively; lanes 3 and 4, 2.5 µg *in vivo* RNA extracted from yeast cells containing plasmids p2GA and p2GAdel$(TA)_3$, respectively; lanes 5 and 6, 3 µg RNA extracted from yeast cells containing plasmids pL101 and pBEVY-U, respectively. (**D**) Mapping the poly(A) sites. The cDNAs were obtained by RT–PCR from RNAs extracted from yeast cells containing plasmids pL101 and pBEVY-U cloned into the pT7T3. Six clones from each sample were sequenced. The lines above the letters indicate last nucleotides immediately prior to the poly(A) tail. The numbers above the vertical lines indicate the number of clones in which that site is used for adding poly(A) tail.

Synthesis of different RNA fragments

DNA templates for synthesizing each *ADH2* 3′ end RNA fragment were made using PCR with a 5′ sense primer containing the T7 promoter sequence and a 3′ antisense primer. RNA probes were labeled with [^{32}P]UTP during *in vitro* transcription, and RNA competitors were transcribed in the absence of [^{32}P]UTP. For making sense *ADH1* RNA fragment A, the *Pst*I–*Dra*I *ADH1* 3′-end fragment from pBEVY-U (Fig. 1A) was cloned into *Pst*I–*Sma* I cleaved vector pT7T319U. The new plasmid pT7T319U*ADH1* 3′ end was cut by *Eco*RI and transcribed by T7 RNA polymerase to synthesize sense *ADH1* RNA fragment A. In order to make the *ADH1* RNA fragment B for UV-crosslinking study, primers 324T7 (5′-GCGCTAATACGACTCACTATAGG-GTTGACACTTCTAAATAAG-3′) and 325 (5′-AAATTTGTA-TACACTTA-3′) were used to amplify template *ADH1* 3′ end in plasmid pBEVY-U. The PCR fragments containing 87 bp just upstream of poly(A) sites (30) were transcribed by T7 RNA polymerase. Both *ADH1* RNA fragments A and B contain the *cis*-acting sequences upstream of poly(A) site (30), but *ADH1* RNA fragment A is longer. All RNA probes and competitors were purified by electrophoresis in 5% acrylamide–7 M urea gels, overnight elution at 24°C in 0.5 M ammonium acetate, 1 mM EDTA, and 0.1% SDS, followed by phenol extraction and ethanol precipitation.

Gel mobility shift assay

RNAs were incubated for 15 min at 24°C with 90 µg of nuclear extracts in a 20 µl mixture using the *in vitro* transcription conditions (35). Optimization of the reaction included: varying the pH (range 6.9–8.1); the K acetate (0–70 mM); and the $MgSO_4$, $MgCl_2$, NaCl and KCl (0–100 mM). The absence of Mg acetate had no effect on specific RNA–protein binding. The final reaction mixture contained 10 mM HEPES (pH adjusted by KOH to 7.8), 0.025 mM of each rNTP, 3 mM $MgSO_4$, 3 mM $MgCl_2$, 5 mM EDTA, 2.5 mM DTT, 20 U RNasin, 5 µg tRNA, 10% glycerol and 90 µg yeast nuclear extracts. Competition assays were performed by mixing ^{32}P-labeled RNA with 20-, 40-, 80-, 160-, 320- and 640-fold molar excess of unlabeled competitor RNAs prior to the addition of nuclear extracts. The reaction mixtures were run on a 5% polyacrylamide (60:1)-1× Tris–glycine gel at 4°C. Gels were dried and visualized by a conventional autoradiography or by a PhosphorImager.

UV-crosslinking

RNAs were incubated with 90 µg as described above for the gel mobility shift assay. After binding, each mixture was digested with 60 U RNase T1 at 24°C for 15 min and subsequently exposed to short wavelength UV light (254 nm) on ice at a 6 cm distance for 40 min. Samples were digested with 25 µg RNase A only or with 25 µg RNase A and 3 µg proteinase K for 15 min at 37°C and analyzed by 10% SDS–PAGE. The gel was dried and subjected to autoradiography or visualized by scanning the gel with a PhosphorImager.

Immunoblotting

18 µg yeast nuclear extracts and 0.14 µg H6Hrp1 were separated on a 10% SDS–PAGE, blotted and immunostained with mouse polyclonal antibodies against H_6Hrp1. Immunoblots were carried out as described (36), using a 1:2000 dilution of mouse polyclonal serum against H_6Hrp1. Both mouse polyclonal serum and H_6Hrp1 protein were generous gifts from Marco Kessler and Claire Moore.

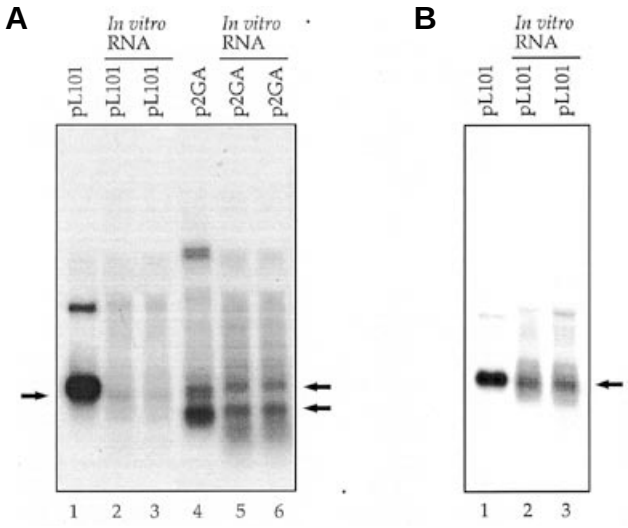

Figure 2. *In vitro* transcription using templates pL101 and p2GA. Both *in vivo* and *in vitro* RNAs were analyzed by northern blot using *ADH2* 3′-end antisense RNA probe A. (**A**) Lanes 1 and 4, 5 µg of *in vivo* RNA extracted from yeast cells containing plasmids pL101 and p2GA, respectively; lane 2, RNA transcribed *in vitro* using template pL101; lane 5, RNA transcribed *in vitro* using template p2GA; Lanes 3 and 6 are duplicates of lanes 2 and 5, respectively. The left arrow indicates the *in vitro* pL101 RNA. The right arrows indicate the *in vitro* p2GA RNA. (**B**) Longer exposure of *in vitro* transcribed RNA. Lane 1, 1 µg of *in vivo* pL101 RNA; lanes 2 and 3, *in vitro* transcribed pL101 RNA, as indicated by the right arrow.

Immunoprecipitation

Immunoprecipitation (IP) was performed following the protocol described by Santa Cruz Biotechnology, Inc. Briefly, after UV-crosslinking and RNase A digestion, Buffer IP (1× PBS, 0.01% NP40, 0.1% SDS) was added to each tube and mixed. Subsequently, 1 µl anti-Hrp1 mouse serum or 1 µl pre-immune mouse serum was added and incubated at 4°C for 1 h; 20 µl protein A agarose (Santa Cruz) was incubated with 180 µg yeast nuclear extracts in 200 µl Buffer IP at 4°C for 30 min. The beads were collected by centrifugation and mixed with each sample. The reactions were incubated at 4°C for 1 h. Immunoprecipitates were collected by centrifugation, the pellet washed twice with buffer IP, repeating centrifugation steps as above. After final wash, the pellet was resuspended in 40 µl 1× electrophoresis sample buffer.

RESULTS

The effect of the $(UA)_3$ deletion on 3′-end formation *in vivo*

Previous studies have demonstrated that the sequences located at the end of the *ADH2* gene constitute a strong termination signal (23,24). By examining a 327 nt fragment inserted within an intron in a chimeric gene linked to *LacZ* (Fig. 1A), we determined that transcription termination occurs with high efficiency (95% termination) at the *ADH2* 3′ end. Four point mutations were identified within the terminator sequence that resulted in

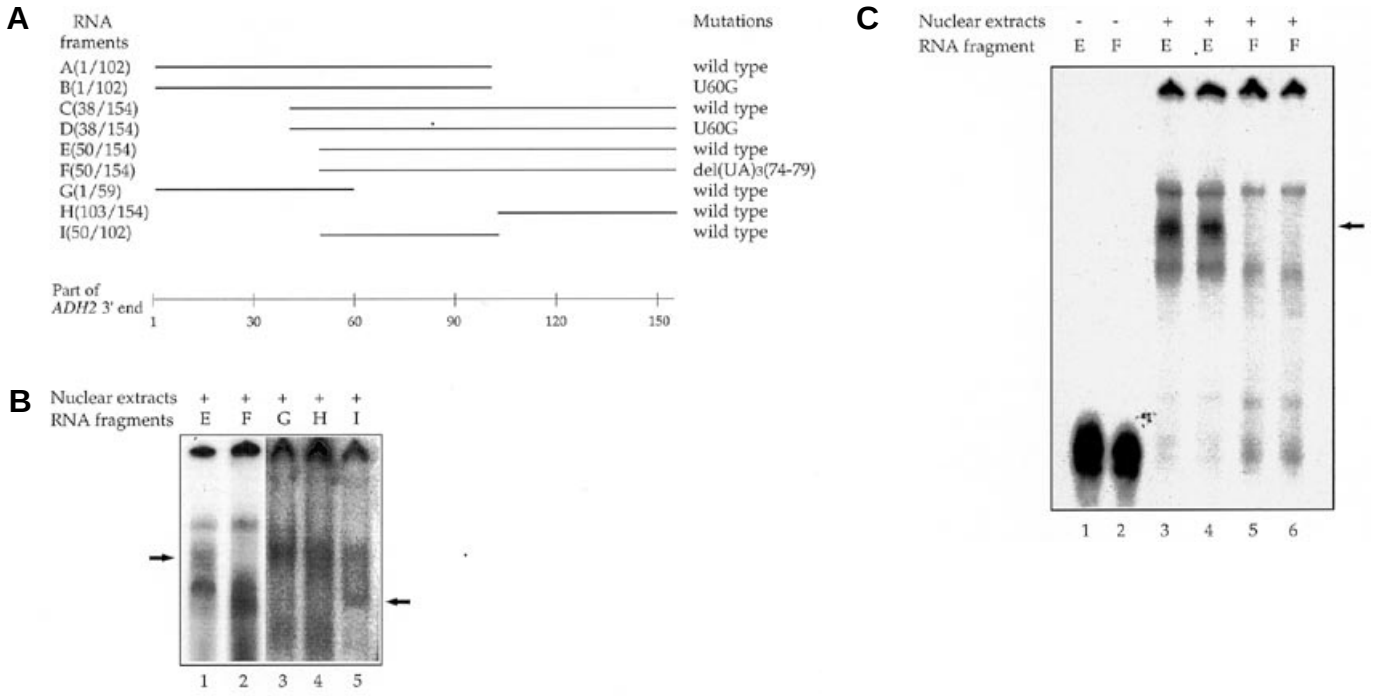

Figure 3. RNA fragments and RNA–protein binding. (**A**) The RNA fragments tested by the gel mobility shift assay. The nucleotide numbering corresponds to that described (23). (**B**) ^{32}P-labeled RNA fragments E–I were incubated with yeast nuclear extracts and analysed on a native polyacrylamide gel. The left arrow indicates the unique band in lane 1. The right arrow indicates the unique band in lane 5. (**C**) Gel mobility shift assay using the optimum binding conditions for RNA–protein interaction. The specific band is indicated by the right arrow. Lanes 1 and 2, 11 pmol wild-type RNA (E) and 11 pmol mutant RNA (F) probes, respectively; lanes 3 and 4 (identical reactions), 11 pmol wild-type RNA probes (E) incubated with nuclear extracts; Lanes 5 and 6 (identical reactions), 11 pmol mutant RNA (F) incubated with nuclear extracts.

decreased efficiency of termination of transcription. However, these mutations are located upstream of the presumptive efficiency and positioning elements described by Guo and Sherman (15). To determine if the *ADH2* termination site utilizes an unusual signal or contains a consensus-like efficiency element, we examined the effect of a deletion of a $(UA)_3$ sequence, a good candidate for the efficiency element in the *ADH2* 3′ end. Deletion of this efficiency element (pL407) results in increased readthrough *in vivo* as indicated by the production high levels of β-galactosidase activity compared with cells containing the wild-type plasmid (data not shown). RNA isolated from cells containing the wild-type plasmid (pL101) or the plasmid containing the efficiency element deletion (pL407) was analyzed by northern blotting using an *ADH2* 3′-end antisense probe A, as shown in Figure 1B, and the predicted terminated transcript was seen only in wild-type cells (Fig. 1C, lanes 1 and 2). Thus, the $(UA)_3$ sequence is required for 3′-end formation *in vivo*.

The effect of the *ADH2* 3′ end context on the choice of its poly(A) sites

A major feature of the chimeric gene used to access termination efficiency is the presence of the 3′-end formation signal within an intron. To test if the terminator context might have effects on mRNA 3′-end formation, we utilized two new templates in which the intron is removed and transcription is driven either by the *GPD* or the *ADH1* promoter (see Fig. 1A for a comparison of these plasmids). In addition, as the promoters are different in all three plasmids, we can determine if selection of the poly(A) site is influenced by the promoter. RNA was isolated from cells containing the plasmids pL101, pBEVY-U and p2GA, and subjected to northern blot analysis as shown in Figure 1C. Note that within the p2GA context, a deletion of the same $(UA)_3$ efficiency element also results in complete readthrough (lanes 3 and 4).

To establish if the polyadenylation sites are the same in both the pL101 context and the pBEVY context, RT–PCR was used to generate the appropriate DNA fragments that were then cloned and sequenced. The results of this analysis are shown in Figure 1D, and demonstrate that the poly(A) tail is added onto the pre-mRNA at the same region within the *ADH2* 3′-end sequence regardless of the intron context of the terminator. In contrast, RNA from cells containing p2GA contain a major and a minor species (not mapped). The difference between the poly(A) site selection may be promoter specific or due to contextual effects.

In vitro transcription assay using different promoters

In addition to the ability of the *ADH2* 3′-end signal to direct efficient termination *in vivo*, we have also shown that RNA polymerase II recognizes the sequence during *in vitro* transcription and generates RNA that is ~100 nt longer than the same RNA synthesized *in vivo*. The template used for that study was the supercoiled plasmid, pL10l. As the transcription extracts are prepared from cells grown under repressing conditions (i.e. in dextrose), the resulting transcription presumably reflects a basal level of activity. Given that the promoter in the pL101 plasmid is regulated by galactose induction, we reasoned that we might improve upon the yields in this reaction if we utilized a new

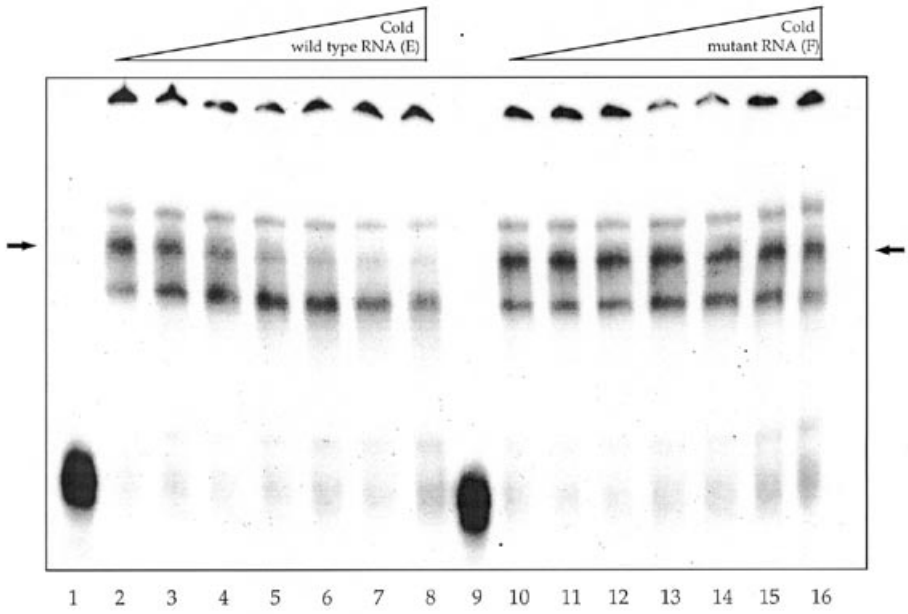

Figure 4. Competition assay using cold wild-type RNA (E) and mutant RNA (F). Increasing molar amounts of cold wild-type RNA and mutant RNA were mixed with 11 pmol ^{32}P-labeled wild-type RNA (E) before adding nuclear extracts. Lanes 1 and 9, wild-type RNA probes (E) alone; lanes 2–8, increasing molar amount of cold wild-type RNA (0, 20, 40, 80, 160, 320 and 640×); lanes 10–16, increasing molar amount of cold mutant RNA (0, 20, 40, 80, 160, 320 and 640×). The left arrow indicates the specific band can be competed gradually by the wild-type RNA. The right arrow indicates that the specific band cannot be competed by cold mutant RNA.

template that contains the same termination sequence driven by the constitutive, highly expressed *GPD* promoter. To test the promoter effects on the efficiency of the transcription reaction, we compared RNA synthesized *in vitro* off both types of templates. Indeed, the p2GA template directs the transcription of the *ADH2* terminated transcript more efficiently compared with the pL101 template, as shown in Figure 2A (compare lanes 2 and 3 with lanes 5 and 6). Quantitation of the signals seen on northern blots demonstrate an increase in the *ADH2* signal of ~3-fold.

A specific RNA–protein interaction in the nuclear extracts

Having established that the *in vitro* transcription reaction mimics the events that occur *in vivo*, we reasoned that we might utilize these extracts to provide additional insights into the relationship of a transcriptional termination sequence with specific nuclear proteins. In order to investigate specific RNA–protein interactions we utilized an RNA gel mobility shift assay. We focused on that part of the 327 nt RNA fragment that contained the most likely candidates for valid efficiency and positioning elements within the *ADH2* termination sequence. We optimized the RNA substrate for binding by testing several overlapping fragments of the *ADH2* 3′-end sequence. ^{32}P-labeled RNA was prepared corresponding to each of the fragments shown in Figure 3A. Primers containing the T7 promoter sequence were combined with gene-specific sequences to generate PCR fragments that would serve as a template for T7 RNA polymerase. To assess the specificity of the RNA–protein interaction, both wild-type and mutant RNAs were synthesized. The mutant RNAs contained a previously described U60G mutation (23) or the deletion of the $(UA)_3$ sequence, either of which had a clear negative effect on 3′-end formation *in vivo*. Each fragment was then tested for the ability to bind to protein from the nuclear extract under the same conditions used for the transcription reactions. Of the nine different RNAs tested in the gel shift assay, no differences between the wild-type and the mutants were observed between fragments A and B or C and D (data not shown). However, fragment E clearly showed a specific gel-shifting pattern that was eliminated in the comparable fragment containing the $(UA)_3$ deletion (fragment F, Fig. 3B). Fragment E contains *ADH2* 3′-end sequences from position 50 to 154, including the efficiency element. To localize the protein binding site(s) further, truncated RNAs were synthesized (Fig. 3A, fragments G–I). Only fragment I, containing the $(UA)_3$ repeat, showed a unique gel-shifting pattern establishing that the protein binding site is present within this sequence (Fig. 3B).

To confirm these observations we optimized conditions for binding of protein to the E RNA fragment. Specifically, the parameters for pH, salt concentrations ($MgCl_2$, KCl, NaCl, $MgSO_4$, and K- and Mg-acetate), were determined as described in Materials and Methods (data not shown). Upon optimization, the RNA–protein interaction was very clear, as shown in Figure 3C. These conditions were utilized for all further experiments.

We reasoned that if the RNA–protein interaction we observed was indeed specific, then the gel-shift pattern should be competed by wild-type RNA (RNA corresponding to fragment E), but not by RNA containing a mutation that eliminates 3′-end formation *in vivo* (RNA corresponding to fragment F). Thus, competition experiments were performed whereby unlabeled RNA was added in increasing amounts to the reaction mix. As shown in Figure 4, the unlabeled wild-type RNA competes with the labeled probe for protein binding; in contrast, unlabeled RNA containing the $(UA)_3$ deletion does not compete, even when present in 640-fold molar excess. Taken together with the observation that a RNA binding

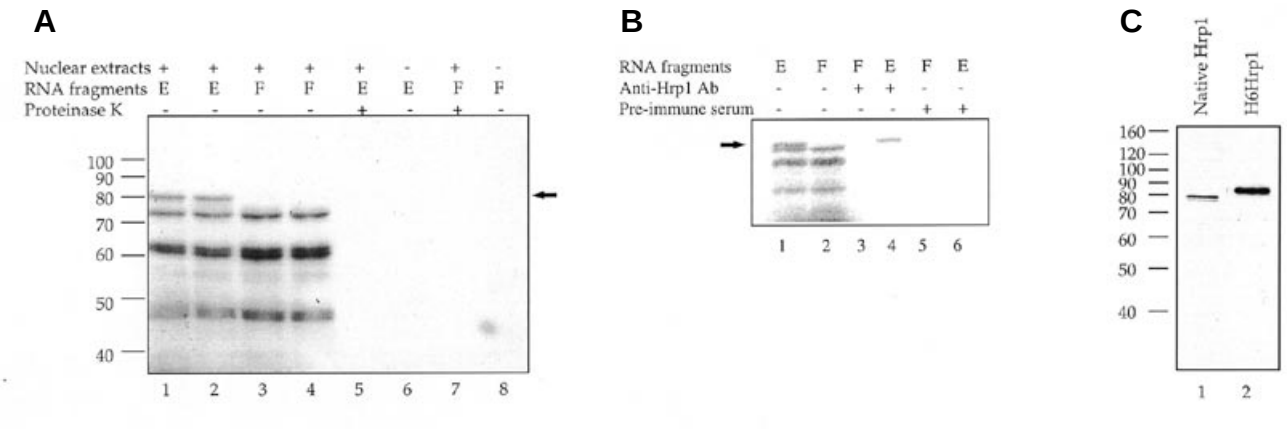

Figure 5. UV-crosslinking of proteins to RNAs, immunoprecipitation and immunoblotting analysis of Hrp1 protein. (**A**) 29 pmol ^{32}P-labeled wild-type (E) and 29 nM mutant (F) RNA were UV-crosslinked to proteins in yeast nuclear extracts. Lanes 1 and 2 (identical reactions), proteins crosslinked to wild-type RNA (E); lanes 3 and 4 (identical reactions), proteins crosslinked to mutant RNA (F); lane 5, proteinase K treatment of proteins crosslinked to wild-type RNA (E); lane 6, wild-type RNA (E) alone without nuclear extracts; lane 7, proteinase K treatment of proteins crosslinked to mutant RNA (F); lane 8, mutant RNA (F) alone without nuclear extracts. The arrow indicates the specific crosslinked protein species. (**B**) Proteins in yeast nuclear extracts were UV-crosslinked to 8 pmol RNA E or F and pre-immune serum or anti-Hrp1 antibodies were added for immunoprecipitation. Lane 1, nuclear extracts (NE) + RNA E; lane 2, NE + RNA F; lane 3, NE + RNA F + anti-Hrp1 Ab; lane 4, NE + RNA E + anti-Hrp1 Ab; lane 5, NE + RNA F + pre-immune serum; lane 6, NE + RNA E + pre-immune serum. (**C**) Yeast nuclear extracts (18 µg) and H_6Hrp1 protein (0.14 µg) were separated on a 10% SDS–PAGE, blotted and immunostained with mouse polyclonal antibodies against H_6Hrp1. Lane 1, native Hrp1 protein; lane 2, H_6Hrp1 protein.

substrate containing the $(UA)_3$ deletion does not allow formation of a gel shifted species, we conclude that the $(UA)_3$ sequence is necessary for pre-mRNA to interact with a protein present in the nuclear extract that may be required for 3′-end formation *in vivo*.

An 80 kDa protein detected by UV-crosslinking

To establish what protein is responsible (at least in part) for the gel mobility shift, a series of UV-crosslinking and immunoprecipitation experiments were conducted. In this set of experiments, the radiolabeled RNA was added to nuclear extracts under the optimal binding conditions determined for the gel mobility shift assay. After binding, samples were treated with RNAse T1, which digested the sequence around the UA element, and exposed to UV light, as described in Materials and Methods. After irradiation, samples were treated with RNase A, and examined by SDS–PAGE. The results are shown in Figure 5A. If the starting RNA substrate contained the UA sequence, an 80 kDa protein was detected, along with other non-specific proteins (lanes 1 and 2). However, when the $(UA)_3$ sequence was deleted (lanes 3 and 4), this 80 kDa protein did not bind to the mutant RNA, suggesting that the 80 kDa polypeptide binds at the UA sequence. As a GST–Hrp1p fusion protein has been shown to interact with a UA repeat present in the 3′ end of the *GAL7* mRNA (16), we next examined the possibility that this is the same protein that we are detecting in our UV-crosslinking assay.

To establish if the 80kDa protein is Hrp1p , we utilized a polyclonal antibody raised against purified Hrp1p to immunoprecipitate extracts that have been subjected to the UV-crosslinking. As shown in Figure 5B, the specific crosslinked protein is precipitated by anti-Hrp1p antibodies. Thus, we conclude that the protein responsible for the mobility shift and the specific interaction with the (UA) efficiency element is indeed Hrp1p.

Finally, to further extend our analysis, we used this antibody for western blot analysis on nuclear extracts. Interestingly, we detected two Hrp1p species (a major species of 79.5 kDa and a minor species of 77.3 kDa; Fig. 5C, lane 1). The recombinant H_6Hrp1 was used as a positive control (84 kDa; Fig. 5C, lane 2). Although the molecular mass of Hrp1 is slightly higher than that reported for Hrp1p (73 kDa), in our hands, recombinant Hrp1p (calculated molecular mass, 76 kDa) also migrated appropriately higher.

Hrp1p binding specificity

To determine if the protein interaction with the *ADH2* 3′-end sequence is specific to only this terminator, or is a more universal factor interacting with other efficiency elements, we asked if a different 3′-end signal could compete for binding with the *ADH2* 3′ end. To address this point, we examined the *ADH1* sequence as the *ADH1* and *2* genes are diverged within this region of the non-coding regions. We examined the *ADH1* 3′-end sequence for two reasons. First, it is as efficient a termination signal as the *ADH2* 3′-end sequence *in vivo*. Secondly, examination of the sequences known to be important for directing 3′-end formation in the *ADH1* gene do not reveal a UA repeat (or any of the accepted variations of this efficiency element) sequence. Thus, this gene does not seem to conform to the generalized processing signals found in many other yeast 3′ ends. If the *ADH1* 3′-end signal operates via a different mechanism for 3′-end formation, it should not effectively compete for binding with the *ADH2* 3′-end sequence. As shown in Figure 6A, the *ADH1* 3′ end competes effectively with the *ADH2* 3′-end sequence in a gel mobility shift assay.

To demonstrate directly that the competition with *ADH2* is due to Hrp1p binding, we performed a crosslinking experiment utilizing radiolabeled *ADH1* RNA in the presence of non-specific RNA. As shown in Figure 6B, the same 80 kDa protein that binds to *ADH2* RNA also binds to *ADH1* sequences. Finally, purified H_6Hrp1p can also be crosslinked to the *ADH1* sequence in the presence of non-specific RNA in a manner analogous to *ADH2* binding, albeit with reduced efficiency (Fig. 6C). Thus, from these experiments we conclude that although the two signals

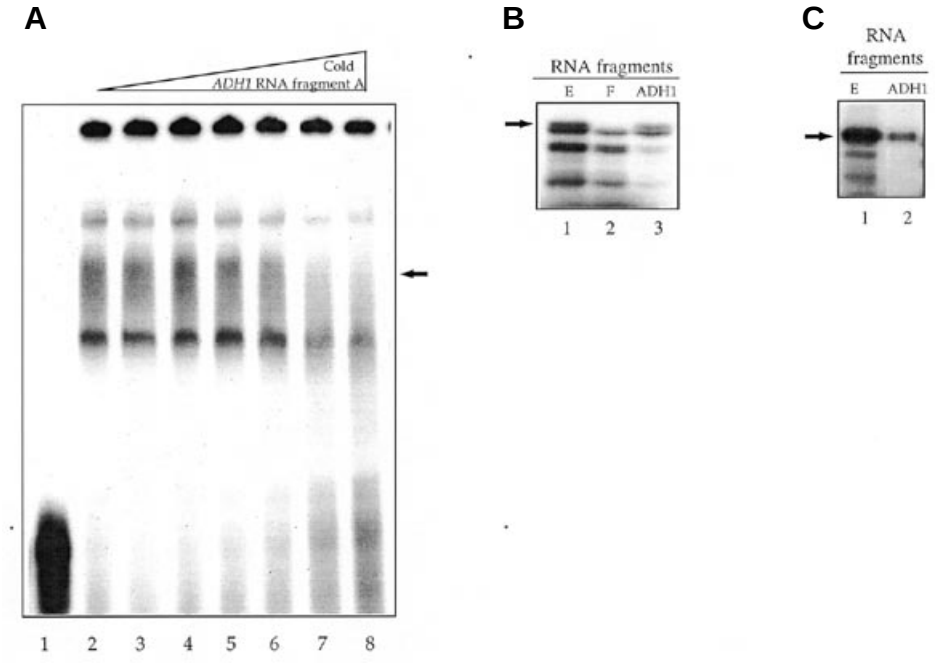

Figure 6. Competition assay and UV-crosslinking. (**A**) Competition assay using unlabeled the *ADH1* RNA fragment A. Lane 1, 11 pmol ^{32}P-labeled wild-type RNA (E) alone; lanes 2–8, increasing molar amounts of unlabeled *ADH 1* RNA fragments A (0, 20, 40, 80, 160, 320 and 640×). The right arrow indicates the specific band that can be competed. (**B** and **C**) UV-crosslinking of a protein in yeast nuclear extracts and the recombinant H_6Hrp1 to the *ADH1* RNA fragment B. The proteins were analyzed by 8% SDS–PAGE. (B) Proteins in yeast nuclear extracts were crosslinked to RNA fragments E, F and *ADH1* RNA fragment B. Lane 1, NE + RNA E; lane 2, NE + RNA F; lane 3, NE + *ADH1* RNA fragment B. (C) The recombinant H_6Hrp1 was crosslinked to the RNA fragment E and *ADH1* RNA fragment B, respectively. Lane 1, H_6Hrp1 + RNA E; lane 2, H_6Hrp1 + *ADH1* RNA fragment.

contain different efficiency element sequences, they interact with the same protein factor, Hrp1p.

The effect of the $(UA)_3$ deletion on transcription termination *in vitro*

We have demonstrated a specific RNA–protein interaction within the 3′ end of yeast mRNAs; however, it is not clear if the 80 kDa protein we are detecting is important for 3′-end processing, actual transcription termination or both. As noted, the two processes of polymerase dissociation (transcription termination) and RNA processing (cleavage and polyadenylation) are tightly linked in yeast and in mammalian cells *in vivo* making it difficult to study the individual reactions separately (21,22). However, we can study both processes separately, *in vitro*. The UA element is required for 3′-end processing using cellular extracts (37). As we have now identified a protein that interacts with pre-mRNA at the efficiency element, we wanted to determine if a mutation in the sequence directing this interaction would effect transcription termination. To address this point we utilized an *in vitro* transcription assay. Two sets of templates were examined for transcription *in vitro*. The first contained the wild-type *ADH2* sequence (Fig. 7A, lanes 2 and 3); the second contained the $(UA)_3$ deletion (Fig. 7A, lanes 4 and 5). After *in vitro* transcription, RNA was isolated and examined by northern blot analysis using an *ADH2* 3′-end antisense probe A (Fig. 1B) that detects RNA product. From the data shown in Fig 7A, it is clear that there is little difference in the transcription pattern for the two templates. Confirmation of this result was obtained when the p2GA and its mutant derivative plasmid were used as a template for *in vitro* transcription (Fig. 7B). In Figure 7B, *ADH2* 3′-end antisense probe B (Fig. 1B) was used. This probe is designed to detect the high molecular weight species of *in vivo* and *in vitro* RNAs as shown in Figure 2, lanes 4 and 5. However, *in vivo*, the $(UA)_3$ deletion results in complete readthrough (Fig. 1C, lanes 2 and 4). The difference between the *in vivo* and *in vitro* results allows us to conclude that the efficiency element is important for 3′-end formation because it directs the proper cleavage of the pre-mRNA transcript but it does not have an effect on the transcripts that we detect *in vitro*. This may mean that the *in vitro* synthesized RNAs are not the result of truly terminated transcript, but rather may reflect a pausing of the polymerase that, *in vivo*, can resume transcription.

DISCUSSION

In this report we demonstrate that an 80 kDa protein interacts with a yeast 3′ pre-mRNA at the efficiency elements of yeast polyadenylation signals. This interaction was revealed utilizing gel mobility shift assays and UV-crosslinking in crude nuclear extracts, methodologies that had not previously been successful in identifying specific interactions involved in mRNA 3′-end formation in yeast. Due to the complexities of the transcription and processing machinery, we reasoned that we might detect interactions that more closely resemble activities present in the nucleus if we used nuclear extracts, as opposed to purified components. Earlier work identified two purified proteins that specifically interacted with the (UA) efficiency element: GST–Hrp1 (99 kDa) and Cft2 (105 kDa). In utilizing yeast nuclear extracts that more closely resemble the *in vivo* situation, we examine

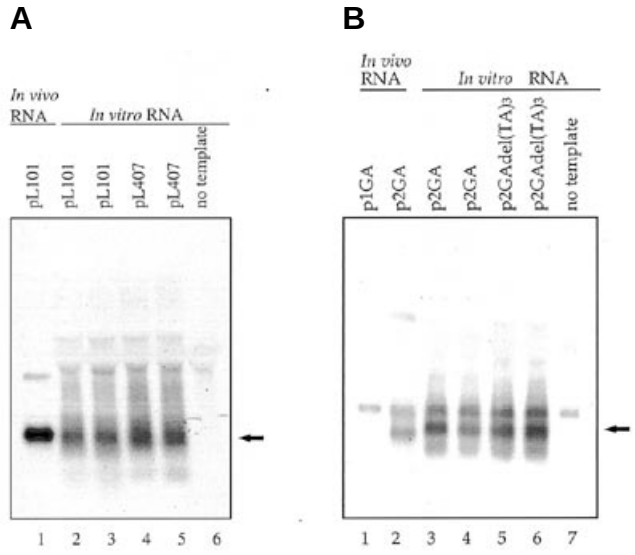

Figure 7. Northern blot analysis of *in vitro* transcribed RNA. The arrow indicates the *in vitro* transcribed RNA product. (**A**) *In vitro* transcription using templates pL101 and pL407. Lane 1, 1 µg *in vivo* RNA extracted from cells containing pL101; lanes 2 and 3 (identical reactions), *in vitro* RNA using template pL101; lanes 4 and 5 (identical reactions), *in vitro* RNA using template pL407; lane 6, no template. The *ADH2* 3′ end antisense probe A was used. (**B**) *In vitro* transcription using templates p2GA and p2GAdel(TA)$_3$. Lanes 1 and 2, 5 µg RNAs extracted from yeast cells containing plasmids p1GA and p2GA, respectively. The p1GA does not contain *ADH2* 3′-end fragment and was used as negative control; lanes 3 and 4 (identical reactions), RNAs extracted from *in vitro* transcription using template p2GA; lanes 5 and 6 (identical reactions), RNAs extracted from *in vitro* transcription using template p2GAdel(TA)$_3$; lane 7, RNAs extracted from *in vitro* transcription without template and used as a negative control for northern analysis. The *ADH2* 3′-end antisense probe B was used to detect the high MW species of RNA shown in Figure 2, lanes 4–6.

RNA–protein interaction at the UA element. The protein that we have identified is Hrp1p. Interestingly, although the *ADH1* 3′-end signal does not contain any of the UA-like efficiency elements determined by mutational analysis, it can compete for binding with the *ADH2* 3′-end signal, which does contain this consensus-like sequence. This suggests that the Hrp1p binding site is not strictly sequence specific, as it appears to bind to non-consensus efficiency elements. Thus, the same factor may recognize different sequences. This may account for the observation that *cis*-acting signals for 3′-end formation are not highly conserved in yeast.

The mechanism of transcription termination by RNA polymerase II is complicated by the fact the mRNA does not end at the poly(A) site, but rather is generated by processing of a pre-mRNA. It is likely that once RNA polymerase transcribes past the site that will become the poly(A) tail, three interrelated events must occur: release of the enzyme from the template, cleavage of the nascent RNA, and polyadenylation. Several lines of evidence suggest that the release of the RNA polymerase II enzyme from the template is dependent on the cleavage event. For example, mutations that result in loss of cleavage activity also result in increased readthrough transcription (25) and mutations in the mammalian poly(A) signal result in aberrant transcription termination *in vivo*. Taken together, these results suggest that if the poly(A) site is not recognized, the transcription reaction will not terminate efficiently (21,22). However, recent work suggests that although polyadenylation signals may be required for termination, they are not sufficient to direct the release of RNA polymerase. The polyadenylation signals may work in concert with downstream elements (DSE) to pause RNA polymerase prior to the cleavage event (38,39). The data presented here support this hypothesis in that, *in vivo*, the (UA) efficiency element is clearly required for functional 3′-end formation, yet the *in vitro* results suggest that terminated transcripts are generated even in the absense of cleavage. It is possible that in the *in vitro* system, RNA polymerase pauses at a DSE, generating the pre-cleaved transcripts that are observed, but *in vivo* these are ‘chased’ into readthrough transcripts in the absence of cleavage. Alternative explanations include the possibility that the *in vitro* system is lacking important factors that may function in coupling the two reactions.

It is also interesting that different yeast genes appear to have different sequence requirements for transcription termination. For example, the 3′-end formation signals for the *FBP1* gene resemble the *ADH2* signal; in both genes, deletion of the efficiency element does not impair transcription termination or polymerase pausing. The assays used to examine the *FBP1* gene, transcription run-on analysis and plasmid instability assays, suggest that actual termination of transcription does occur in the absence of an efficiency element (40). Yet, other well-studied sequences, including the *CYC1* signal, appear to be dependent on the (UA)-like efficiency element for their termination activity *in vivo* (25,41,42). This raises the possibility that there are two types of signals directing 3′-end formation in yeast; the complex signal(s) such as those present in *ADH2* and *FBP1* and the more compact signal found in *GAL7* and *CYC1*. Whether these differences are inconsequential or reflect a regulatory function remains to be determined.

ACKNOWLEDGEMENTS

We would like to thank M.Kessler and C.L.Moore for providing Hrp1 antibodies and H_6Hrp1 purified protein and J.Nolan and X.Y.Zhang for useful suggestions. We also thank W.H.Baricos, T.Jackson, C.Magrath and R.Reger for critical reading of the manuscript and helpful discussions. These studies were supported by grants from the National Science Foundation (MCB9604295) and the Tulane University Cancer Center.

REFERENCES

1 Wahle,E. and Keller,W. (1996) *Trends Biochem. Sci.*, **21**, 247–250.
2 Colgan,D.F. and Manley,J.L. (1997) *Genes Dev.*, **11**, 2734–2766.
3 Moore,C. and Sharp,P. (1985) *Cell*, **41**, 845–855.
4 Butler,J.S. and Platt,T. (1988) *Science*, **242**, 1270–1274.
5 Minvielle-Sabastia,L., Preker,P.J. and Keller,W. (1994) *Science*, **266**, 1702–1705.
6 Chanfreau,G., Noble,S.M. and Guthrie,C. (1996) *Science*, **274**, 1511–1514.
7 Jenny,A., Minvielle-Sebastia,L., Preker,P.J. and Keller,W. (1996) *Science*, **274**, 1514–1517.
8 Stumpf,G. and Domdey,H. (1996) *Science*, **274**, 1517–1520.
9 Barabino,S., Hubner,W., Jenny,A., Minvielle-Sebastia,L. and Keller,W. (1997) *Genes Dev.*, **11**, 1703–1716.
10 Kessler,M., Zhao,J. and Moore,C.L. (1996) *J. Biol. Chem.*, **271**, 27167–27175.
11 Zhao,J., Kessler,M. and Moore,C.L. (1997) *J. Biol. Chem.*, **272**, 10831–10838.
12 Takagaki,Y., Ryner,L.C. and Manley,J.L. (1988) *Cell*, **52**, 731–742.
13 Patel,D. and Butler,J.S. (1992) *Mol. Cell. Biol.*, **12**, 3297–3304.
14 Guo,Z. and Sherman,F. (1996) *Mol. Cell. Biol.*, **16**, 2772–2776.

15 Guo,Z. and Sherman,F. (1996) *Trends Biochem. Sci*., **21**, 477–481.
16 Kessler,M.M., Henry,M.F., Shen,E., Zhao,J., Gross,S., Silver,P.M. and Moore,L.C. (1997) *Genes Dev.*, **11**, 2545–2556.
17 Cullen,B.R., Lomedico,P.T. and Ju,G. (1984) *Nature*, **307**, 241–245.
18 Proudfoot,N.J. (1986) *Nature*, **322**, 562–565.
19 Irniger,S., Egli,C.M., Kuenzler,M. and Braus,G. (1992) *Nucleic Acids Res*., **20**, 4733–4739.
20 Eggermont,J. and Proudfoot,N. (1993) *EMBO J*., **12**, 2539–2548.
21 Connelly,S. and Manley,J.L. (1988) *Genes Dev.*, **2**, 440–452.
22 Connelly,S. and Manley,J. (1989) *Mol. Cell. Biol*., **9**, 5254–5259.
23 Hyman,L.E., Seiler,S.H., Whoriskey,J. and Moore,C.L. (1991) *Mol. Cell. Biol*., **11**, 2004–2012.
24 Hyman,L.E. and Moore,C.L. (1993) *Mol. Cell. Biol*., **13**, 5159–5167.
25 Birse,C.E., Minvielle-Sebastia,L., Lee,B.A., Keller,W. and Proudfoot,N.J. (1998) *Science*, **280**, 298–301.
26 Dantonel,J.-C., Murthy,K.G.K., Manley,J.L. and Tora,L. (1997) *Nature*, **389**, 399–402.
27 McCracken,S., Fong,N., Yankulov,K., Ballantyne,S., Pan,G., Greenblatt,J., Patterson,S.D., Wickens,M. and Bentley,D.L. (1997) *Nature*, **385**, 357–361.
28 Steinmetz,E. (1997) *Cell*, **89**, 491–494.
29 Miller,C., Martinat,M.A. and Hyman,L.E. (1998) *Nucleic Acids Res*., **26**, 3577–3583.
30 Heidmann,S., Schindewolf,C., Stumpf,G. and Domdey,H. (1994) *Mol. Cell. Biol*., **14**, 4633–4642.
31 Firmenich,A. and Redding,K. (1993) *Biotechniques*, **14**, 712–718.
32 Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1994) *Current Protocols in Molecular Biology*. John Wiley and Sons, Inc., New York.
33 Chen,S., Reger,R., Miller,C. and Hyman,L.E. (1996) *Nucleic Acids Res*., **15**, 2885–2893.
34 Guthrie,C. and Fink,G.R. (1991) *Guide to Yeast Genetics and Molecular Biology*. Academic Press, Inc., San Diego, CA.
35 Lue,N.F. and Kornberg,R.D. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 8839–8843.
36 Kessler,M., Zhelkovsky,A., Skvorak,A. and Moore,C.L. (1995) . *Biochemistry*, **34**, 1750–1759.
37 Chen,J. and Moore,C.J. (1992) *Mol. Cell. Biol*., **12**, 3470–3481.
38 Eggermont,J. and Proudfoot,N.J. (1993) *EMBO J*., **12**, 2539–2548.
39 Birse,C.E., Lee,B.A., Hansen,K. and Proudfoot,N.J. (1997) *EMBO J*., **16**, 3633–3643.
40 Aranda,A., Perez-Ortin,J.E., Moore,C.L. and del Olmo,M. (1998) *RNA*, **4**, 303–318.
41 Russo,P. and Sherman,F. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 8348–8352.
42 Russo,P. (1995) *Yeast*, **11**, 447–453.