The EMBO Journal vol.13 no.2 pp.446–452, 1994

Poly(A) site selection in the yeast Ty retroelement requires an upstream region and sequence-specific titratable factor(s) *in vitro*

Wei Hou[2], Roland Russnak[1] and Terry Platt[1,2,3]

[1]Department of Biochemistry, University of Rochester Medical Center, Rochester, NY 14642 and [2]Department of Biology, University of Rochester River Campus, Rochester, NY 14627, USA
[3]Corresponding author

Communicated by N.J.Proudfoot

In the Ty retrotransposon of *Saccharomyces cerevisiae*, as in most retroelements, the polyadenylation site of the 5′ long terminal repeat (LTR) is ignored and the one in the 3′ LTR is efficiently used. We examine here the contribution to this poly(A) site selection of the region termed 'U3', corresponding to the upstream non-transcribed portion of the 5′ LTR. Using an established assay *in vitro*, we find that 3′ processing is accurate and efficient with an RNA substrate corresponding to most of the LTR, whereas none is detectable with a shorter transcript lacking the U3 region, thus explaining why the 5′ poly(A) site is ignored in genomic Ty mRNA. When *HIS4* coding RNA, representing 'non-specific' sequence, replaces the U3 region, the Ty polyadenylation site is activated to 50% of the wild-type level. Within one specific region (TS1) in U3, 90–95 nt upstream of the poly(A) site, the change of UAG*U*AU to UCGCAU reduces processing efficiency by half, to the non-specific level provided by other sequences or by a deletion of the TS1 region. Another region (TS2) near the poly(A) site appears to be independently responsible for the remaining half of the processing activity. Alteration of both TS1 and TS2 eliminates processing entirely. In competition assays, excess unlabeled U3, but not its mutated counterparts, reduces the processing of radiolabeled Ty mRNA, suggesting the involvement of some sequence-specific titratable factor(s) in the whole cell extract for U3-specific activation. Since we obtain similar reductions in competing *CYC1* mRNA processing with excess unlabeled Ty-derived RNAs, we conclude that polyadenylation of Ty element RNA, while possessing some unusual properties, also shares common components of the processing machinery used with other yeast mRNA transcripts.

Key words: poly(A) sites/3′ RNA processing/RNA–protein interactions/Ty retroelement/upstream activation

Introduction

The 3′ ends of most eukaryotic messenger RNAs are generated by a processing event consisting of coupled endonucleolytic cleavage and polyadenylation (Wahle and Keller, 1992 and references therein). The *cis*-acting signals for this RNA processing are well defined in mammalian cells and contain two main elements: a highly conserved AAUAAA sequence 10–30 nucleotides (nt) upstream of the polyadenylation site and a relatively diffuse GU- or U-rich downstream sequence. It has yet to be demonstrated that these 'core elements' are regulated at the 3′ ends of cellular genes. The situation found in retroid elements, however, is quite distinct. These elements, which replicate by reverse transcription, must program the host machinery to synthesize terminally redundant RNAs obligatory for strand transfer reactions during DNA synthesis. The end result is that the identical cleavage/polyadenylation site and GU-rich element are transcribed at both the 5′ and 3′ ends of the message. Efficient replication requires that these poly(A) sites be acted upon by mechanisms which suppress use of the 5′ copy while ensuring maximal use of the 3′ copy (for reviews see Imperiale and DeZazzo, 1991; Proudfoot, 1991). Two retroviruses, namely Rous sarcoma virus (Ju and Cullen, 1985) and type I human T-cell leukemia virus (Ahmed *et al.*, 1991) circumvent this problem by inclusion of the AAUAAA hexanucleotide signal only at the 3′ end of the transcript. In all other retroid elements studied, core elements (including AAUAAA) capable of programming some degree of cleavage and polyadenylation are transcribed at both ends of the genomic RNA. Total occlusion of the 5′ poly(A) site *in vivo* has been observed in cauliflower mosaic virus (CaMV; Sanfaçon and Hohn, 1990), spleen necrosis virus (SNV; Iwasaki and Temin, 1990), human immunodeficiency virus (HIV; Weichs an der Glon *et al.*, 1991; Cherrington and Ganem, 1992) and ground squirrel hepatitis virus (GSHV; Cherrington *et al.*, 1992) and is determined by the relatively short distance of the core elements from the promoter. Purified RNA substrates designed to represent the 5′ end of HIV genomic RNA and which harbor the core elements are partially competent for processing *in vitro* (Gilmartin *et al.*, 1992; Valsamakis *et al.*, 1992). Thus promoter-proximal occlusion within the cell is not simply due to a lack of RNA between the cap and the 3′ end processing signals, but instead may reflect some property of the early elongating RNA polymerase II (Weichs an der Glon *et al.*, 1993).

Part of the retroid solution to poly(A) site selection is provided by a third class of RNA signals (in addition to AAUAAA and the GU block) that influence 3′ end processing. These are upstream elements (USEs) which are transcribed only once, at the 3′ end of genomic retroviral RNA, and which further activate the core signals. Examples include U3 sequences within the retroviruses HIV (Brown *et al.*, 1991; DeZazzo *et al.*, 1991, 1992; Valsamakis *et al.*, 1991, 1992; Cherrington and Ganem, 1992; Gilmartin *et al.*, 1992) and SNV (Russnak and Ganem, 1990), as well as the PS1 element of the hepadnavirus GSHV (Russnak and Ganem, 1990; Russnak, 1991). USEs are not unique to the U3 regions of retroid elements, but include the GSHV PS2 element located within the R subregion (Russnak and Ganem, 1990; Cherrington *et al.*, 1992), and were in fact first identified at two non-retroid viral poly(A) sites, SV40 late

© Oxford University Press

(Carswell and Alwine, 1989; Schek *et al.*, 1992) and adenovirus L1 (DeZazzo and Imperiale, 1989; Wilson-Gunn *et al.*, 1992). The possible involvement of USEs in 3′ end formation of yeast cellular genes is suggested by point mutations in the *ADH2* gene ~80 nt upstream of its poly(A) site (Hyman *et al.*, 1991), although whether these affect processing, stability or termination of transcripts has not been determined. Plant genes, such as pea *rbcS* (Mogen *et al.*, 1992) and CaMV (Sanfaçon *et al.*, 1991), also commonly have one or more USEs, possibly to compensate for the general lack of an AAUAAA consensus in plant 3′ signals. Interestingly, two upstream elements were also found to be important for 3′ end formation of CaMV RNA in yeast cells (Irniger *et al.*, 1992). Some of the viral sites appear functionally related, since the HIV USE can activate GSHV (Russnak and Ganem, 1990) and SV40 late (Valsamakis *et al.*, 1991) core poly(A) elements, and there is substantial sequence homology found between GSHV PS1 and the SV40 late USE (Schek *et al.*, 1992), as well as between the latter and the USEs of plant genes (Mogen *et al.*, 1992).

Non-viral retroid elements include the family of retrotransposons. In *Saccharomyces cerevisiae*, the multi-copy Ty element has a mode of transposition resembling retroviral replication (for reviews see Boeke, 1989; Boeke and Sandmeyer, 1991) in that new copies are generated via reverse transcription of terminally redundant mRNA. The organization of Ty conforms to the general anatomy of linear retroelements; these include other transposable elements (*Drosophila copia* and *gypsy*) as well as retroviruses. An internal segment of several kilobases is flanked by long terminal repeats (LTRs) of ~340 bp each. In Ty, the transcription start site for the 5.7 kb genomic RNA is within the 5′ LTR ~50 bp upstream of the site which, in the 3′ LTR, is used for polyadenylation. A progressive 5′ to 3′ deletion analysis of 3′ end forming signals *in vivo* has previously identified two regions, TS1 [−97 to −75; numbering is relative to the poly(A) site] and TS2 (−50 to −27), that contribute equally to 3′ end formation (Yu and Elder, 1989). Both regions contain sequences with almost perfect identity to the bipartite 3′ end processing motif proposed by Russo *et al.* (1991), which is located in many, but by no means all, yeast genes. TS1 is analogous to retroviral U3 USEs since it is transcribed only at the 3′ end of genomic RNA and together with TS2 ensures maximal levels of processing. Similarly, sequences harboring TS2 and the associated poly(A) site are analogous to core elements in retroviruses in that they are transcribed also at the 5′ terminal end of genomic RNA. Promoter-proximal poly(A) site occlusion is presumed to be operative in Ty since truncated, polyadenylated transcripts of 45 nt have not been detected (Yu and Elder, 1989). Thus one of the questions which remains is why the TS2 element in the 5′ poly(A) site is non-functional (Yu and Elder, 1989).

In the present work we have resolved this issue, as well as duplicated *in vitro* the other aspects of Ty poly(A) site regulation, using an assay for cleavage and polyadenylation developed in this laboratory (Butler and Platt, 1988). Our results reveal a simple but novel type of promoter-proximal occlusion in which TS2 and poly(A) site sequences are not recognized when part of 5′ truncated RNA molecules. Use of these core elements proceeds on extended substrates and is activated fully by the specific sequence provided by TS1. Competition experiments with RNAs containing only U3 sequences suggest that the interaction between some titratable factor(s) and the UAGUAUGUA motif within TS1, but not a mutated variant, is responsible for specific activation. Similar experiments show that U3 sequences can compete processing of *CYC1* mRNA, implying that recognition by some processing components is more general. Overall, our data provide the first direct evidence for sequence-specific RNA binding by a yeast factor involved in 3′ end formation.

Results

Efficient 3′ processing of Ty elements requires U3 sequences in vitro

To examine differential function of the 5′ and 3′ poly(A) sites of Ty, it was first necessary to see whether Ty mRNA would be processed *in vitro* using our established assay. Thus, SP6 transcripts, representing the desired LTR-derived substrates that were made from pGEM3Z templates *in vitro* (Figure 1B), were tested in yeast whole cell extracts, in which we have previously shown that pre-mRNAs corresponding to a variety of yeast genes are accurately processed (Butler *et al.*, 1990; Sadhale *et al.*, 1991). We found that our Ty pre-mRNA derivative was processed at a site consistent with the predominant 3′ end mapped *in vivo*, ~45 bp upstream of the end of the 3′ LTR (Yu and Elder, 1989; Figure 1C, lane 2). The identity of the cleavage and polyadenylation products was verified by examining a reaction containing CTP, which allows cleavage but not polyadenylation (data not shown). A 35 nt deletion in the U5 region, leaving no more than 10 nt of 3′ downstream sequences, had a minimal effect on the processing efficiency *in vitro* (compare Figure 1C, lane 2 with lane 4).

Since the primary difference between the 5′ and 3′ poly(A) site is the presence of U3 sequences upstream of the R region (see Figure 1A), we tested RNAs corresponding to the two sites. When the U3 region was missing, as it would be in the 5′ poly(A) site, the remaining RU5 sequences had no detectable activity; these were not reactivated by extension of the downstream region with a 100 bp non-specific segment derived from pGEM3Z vector (Figure 1C, lanes 6 and 8). U3 sequences themselves do not contain a cryptic cleavage site that can be recognized *in vitro* (Figure 1C, lane 10). Therefore, efficient 3′ processing of Ty element RNA *in vitro* requires the U3 region and its removal inactivates the poly(A) site, whereas sequences downstream of the poly(A) site are not essential for activation. This simple explanation is sufficient to account for poly(A) site selection in Ty elements, i.e. the poly(A) site in the 3′ LTR is activated by adjacent U3 sequences in the transcript; the 5′ LTR site is inactive because all upstream sequences are lacking.

Yu and Elder (1989) observed that RU5, compared with U3RU5, retained 50% activity *in vivo* when placed downstream of the *GAL1* gene. Since RU5 sequences alone were inactive *in vitro*, this suggested that there might be a general contribution of 5′ flanking sequences. We therefore inserted 140 nt of *HIS4* coding sequences directly upstream of the inactive RU5 substrate, and found that 50% of the wild-type processing was restored (Figure 2A, H4RU5 and Figure 2B, lanes 5−8). Although we cannot rule out contributions from specific sequences in *HIS4*, the 50% activity agrees with Yu and Elder's observations (1989) using *GAL1* in the same position.

To ask how the U3 region improves efficiency above this

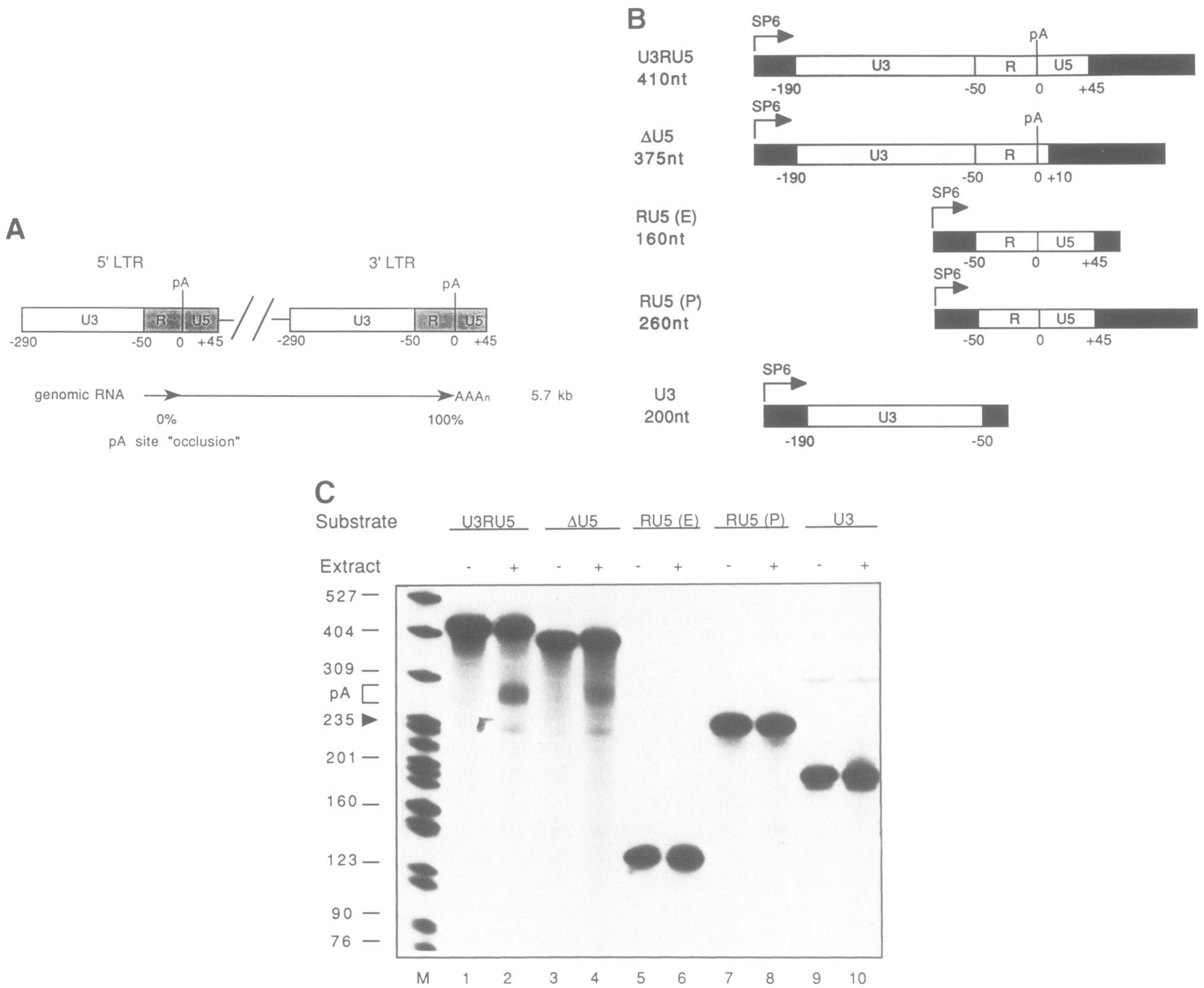

Fig. 1. U3 sequences are required for efficient processing *in vitro*. (**A**) Schematic diagram of Ty mRNA expression and poly(A) site utilization. 5′ and 3′ LTRs and the positions of the U3, R and U5 regions are shown. The Ty polyadenylation site is numbered as zero and the positions of the U3, R and U5 regions are marked. Ty genomic RNA is 5.7 kb long, as indicated (Boeke, 1989; Sandmeyer, 1992). The 3′ ends of Ty genomic RNA are formed exclusively within the 3′ LTR at the R−U5 junction, whereas none are formed in the 5′ LTR, though identical sequences (R-U5, shaded) are transcribed from each LTR. (**B**) Schematic diagram of the transcript templates tested. The different Ty LTR sequences (shown as open bars) were cloned in pGEM3Z and transcribed by phage SP6 RNA polymerase. Transcribed vector sequences are represented by solid bars. (**C**) Processing of pre-mRNAs. Uniformly labeled pre-mRNAs transcribed *in vitro* were incubated in yeast whole cell extracts in the presence of 2 mM ATP, 2% PEG, 1 mM Mg^{2+} and 75 mM potassium acetate at 30°C for 20 min. The RNAs were extracted and analyzed on a 4% denaturing polyacrylamide gel. DNA markers were generated from *Msp*I-digested pBR322 plasmid. Polyadenylated products are indicated as ‘pA’, and the 5′ cleavage product is 235 nt in length. 3′ cleavage products are not capped and are easily degraded in the reactions. The RNA alone controls are shown as ‘extract −’.

‘non-specific’ level of 50%, we made a 22 nt deletion (designated as ΔTS1, see Figure 2A) in a region of U3 implicated in 3′ end formation *in vivo* (Yu and Elder, 1989). *In vitro*, this deletion decreased processing efficiency to the same level as the H4RU5 substrate (Figure 2B, lanes 13−16). Sequences within TS1 conform to the Russo *et al.* (1991) bipartite motif UA**G**/**U**AUGUA (see Figure 2A, ΔTS1). Analysis of a *cyc1-512* revertant indicated that the bold A and U within this motif restore 3′ end formation, while Cs at these same positions do not (Russo *et al.*, 1991). To test the importance of these two nucleotides in the U3 motif, we substituted both the A at position −93 and the T at −91 by C, and found that U3-specific activation was abolished (Figure 2A, ‘TS1-93/91C’; Figure 2B, lanes 9−12). To summarize, (i) the RU5 substrate alone is inactive for processing *in vitro*; (ii) 50% of the activity can be restored by adding any of several sequences upstream, including *GAL1* (*in vivo*), *HIS4*, U3(ΔTS1) or U3(TS1-93/91C) regions (*in vitro*) and (iii) U3 with an intact TS1 sequence is maximally active.

TS1 and TS2 appear to contribute independently to efficient 3′ processing

Since the deletion of TS1 in the U3 region reduced the processing activity to a non-specific level, i.e. to ~50%, other signals for processing must account for the remaining activity. Indeed, a region designated as TS2 in R by Yu and Elder (1989) was also found to be important for 3′ end formation *in vivo*. In their analysis, progressive 5′ deletions eliminating both TS1 and TS2 abolished 3′ end formation completely, whereas deletions only affecting TS1 retained partial activity. This observation indicated that TS2 might

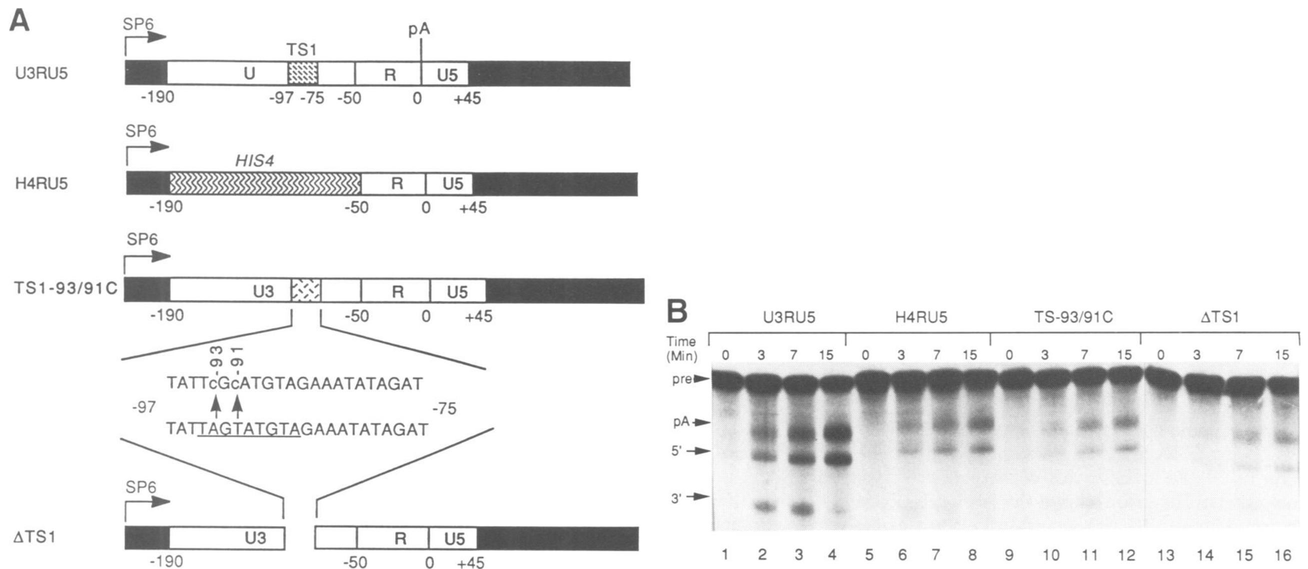

Fig. 2. U3 activation depends on both specific and non-specific sequences. **(A)** Schematic diagram of the substrates. In H4RU5, 140 nt of *HIS4* coding sequence were inserted upstream of the RU5 region, replacing the U3 sequence. A 22 nt deletion was made within U3 from positions −97 to −75, designated as ΔTS1. The perfect match to the bipartite motif (Russo *et al.*, 1991) is underlined. A 2 nt substitution within TS1, at positions −93 and −91, was carried out to create TS1-93/91C. pGEM3Z sequences are shown as solid bars. **(B)** A time course of processing. SP6-transcribed, uniformly ^{32}P-labeled pre-RNAs were incubated in yeast whole cell extract for the indicated time, before stopping the reactions with 0.5% SDS and 2 mg/ml proteinase K. RNAs were analyzed as described in Figure 1B. Polyadenylated products, 5′ and 3′ cleavage products are indicated as pA, 5′ and 3′ respectively.

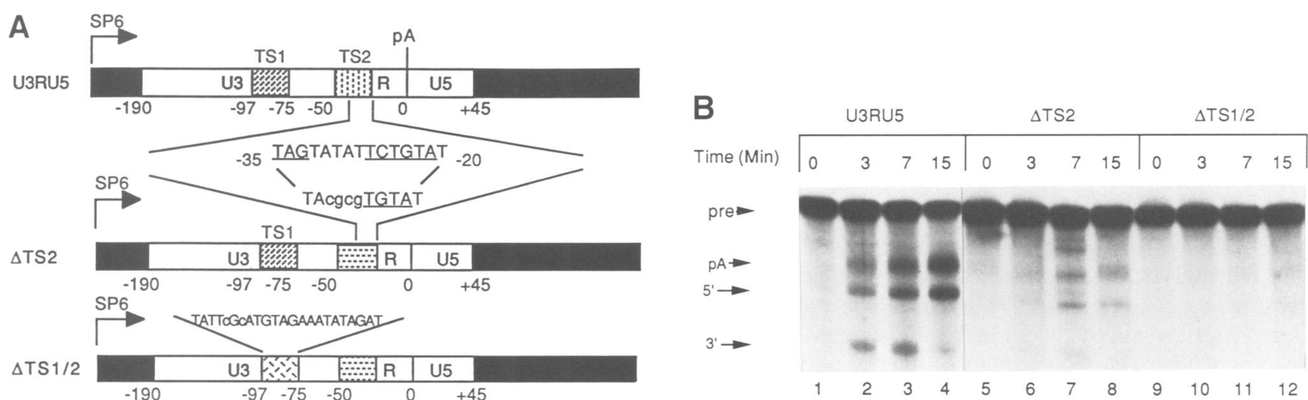

Fig. 3. TS1 and TS2 both contribute to the processing of Ty elements. **(A)** Schematic diagram of the pre-RNAs analyzed in panel B. In TS2 (as defined by Yu and Elder, 1989) we altered the 3′ portion as shown. The sequence similar to the Russo motif is underlined. The deletion and substitution are indicated in lower case. **(B)** Time course of processing reactions. The reactions were conducted as described in Figure 2B.

also encompass a processing signal for Ty elements. We therefore made a deletion/substitution in the 3′ portion of the TS2 region, replacing GUAUAUUC with CGCG, and a time course of 3′ processing was performed *in vitro* (Figure 3A, ΔTS2 and Figure 3B, lanes 5−8). As shown in Figure 3B, processing of ΔTS2 was at about 50% of the wild-type, quite similar to the level with the ΔTS1 substrate. When both TS2 and TS1 were altered, the processing activity was almost completely eliminated (Figure 3A, ΔTS1/ΔTS2; Figure 3B, lanes 9−12). The results demonstrate that TS1 in U3 and TS2 in R contribute about equally and independently to the maximal 3′ processing of Ty elements *in vitro*. Interestingly, TS2 also displays similarities to the bipartite motif, indicating that similar signals for 3′ processing might work additively at the single Ty poly(A) site.

U3 activation requires some titratable factor(s)

We have demonstrated above that U3 contributes specifically to activation of Ty 3′ processing *in vitro* and we have identified one region that is involved. How does this activating element exert its influence from nearly 90 nt upstream of the poly(A) site? To investigate whether specific interactions between U3 and some protein factor(s) might be responsible for the activation, we performed competition assays *in vitro*. In these experiments, processing of radiolabeled U3RU5 wild-type RNA substrate was examined in the presence of a 10- or 30-fold molar excess of unlabeled competitor RNA. Figure 4 demonstrates that processing of labeled U3RU5 was competed by a 30-fold excess of its unlabeled counterpart (lane 4) and to a lesser degree by the shorter U3 fragment (lane 6). At the same time, neither U3 sequences with the TS1 deletion or with the 2 nt substitution nor the *HIS4* sequence compete (lanes 7−12). This competition experiment suggests that some titratable factor must interact with one or both of the mutated nucleotides within the TS1 motif to cause U3 activation of 3′ processing of Ty RNA *in vitro*. It supports the notion that the *HIS4* sequence does not have a specific binding site for this factor.

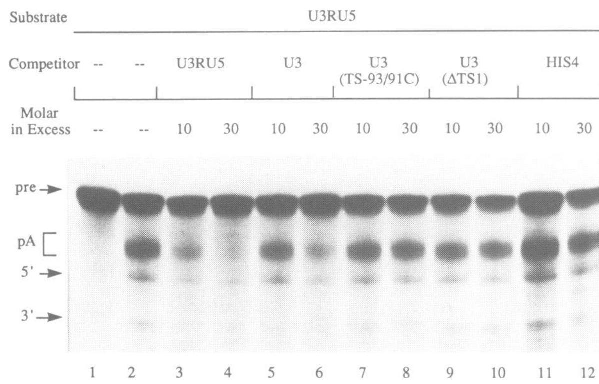

Fig. 4. Specific competition of Ty processing with U3 sequences. The ^{32}P-labeled U3RU5 substrate was processed *in vitro* in the presence of a 10- or 30-fold molar excess of the indicated unlabeled RNA competitors. The processing of the labeled U3RU5 was analyzed as described above. The U3 sequence itself and its derivatives containing TS1 deletion and 2 nt substitution, as well as the *HIS4* coding sequence, were used as competitors.

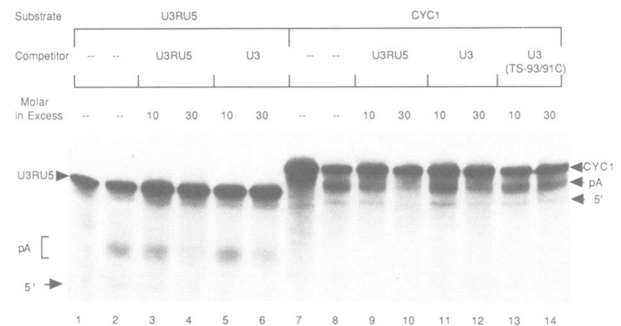

Fig. 5. Competition of *CYC1* processing with Ty elements. Labeled *CYC1* substrates were processed *in vitro* in the presence of similar unlabeled competitors as in Figure 4. Competitions of U3RU5 substrates with U3RU5 and U3 are also shown. The processed products of both U3RU5 and *CYC1* pre-mRNA are indicated on the left and right respectively.

CYC1 poly(A) site use depends on factors that can be titrated by Ty RNA derivatives

The *CYC1* gene, encoding *iso*-1-cytochrome *c*, provided the initial model system for analyzing yeast mRNA cleavage and polyadenylation *in vitro*. Deletion of 38 bp just upstream of its poly(A) site eliminated *CYC1* mRNA processing (Zaret and Sherman, 1982; Butler and Platt, 1988), suggestive of a single essential element. The LTRs of Ty elements conceivably represent a different class of polyadenylation sites, since we have shown here that two related elements (TS1 and TS2) separated by >50 nt, each contribute to poly(A) site function. To explore whether *CYC1* and Ty LTRs share particular recognition factors, an assay of *CYC1* 3′ processing *in vitro* was carried out with U3RU5, U3 or U3 mutated derivatives as unlabeled competitor. This substrate contains 250 nt of *CYC1* gene 3′ untranslated region (see Materials and methods) and an upstream 220 nt non-specific *Escherichia coli* sequence. Interestingly, *CYC1* 3′ processing activity was almost fully competed away by U3RU5 (Figure 5, lanes 9 and 10). U3 wild-type sequences alone, by contrast, only competed the processing of *CYC1* to a lower level, behaving similarly to the ability of U3 to compete U3RU5 processing (lanes 5 and 6 and 11 and 12). In this assay, a U3 sequence containing the 2 nt substitution in TS1 was completely unable to compete *CYC1* processing (lanes 13 and 14). This observation suggested that polyadenylation of Ty element RNA shares common components of the processing machinery used with other yeast mRNA transcripts.

Discussion

Using yeast whole cell extracts, we have investigated some of the components involved in 3′ processing of Ty element transcripts *in vitro*. We found that Ty mRNA was processed efficiently at a site that agrees with the location of the predominant 3′ end mapped *in vivo* (Yu and Elder, 1989). The alterations of putative *cis*-acting signals, by both deletions and substitutions, showed that the TS1 motif in the U3 region and the TS2 motif in the R region are critical, respectively, for U3-specific activation and the 'basal' processing activity in our *in vitro* assay. The latter could be observed when non-specific sequences were added 5′ to the inactive RU5 substrate, yielding 50% of the processing levels obtained with the U3RU5 substrate. The substitution of two specific nt within TS1, >90 nt upstream of the poly(A) site, reduced the processing efficiency of a U3RU5 substrate by half, to a non-specific level. Therefore we infer that full U3 activation of Ty processing *in vitro* depends on specific sequences within TS1 to enhance a general contribution from the presence of upstream RNA. The TS1 and TS2 regions appear to contribute about equally and independently to 3′ processing efficiency.

A solution to the problem of duplicated poly(A) sites is critical to the function of Ty as well as the LTR retroelements of higher eukaryotes (for reviews see Imperiale and DeZazzo, 1991 and Proudfoot, 1991). Our results lead directly to a simple explanation of the 5′ poly(A) site 'occlusion' in Ty: the transcribed 5′ LTR of Ty lacks U3, and we have shown that RU5 alone is inactive for 3′ processing. By contrast, the transcript into the polyadenylation site of the 3′ LTR contains the U3 region and is efficiently processed. Occlusion of the poly(A) site in the 5′ LTR is not due to inhibition by Ty genomic RNA immediately downstream of U5, since replacement by non-specific vector sequence 3′ to the poly(A) site did not restore its processing activity (see construct RU5(P); Figure 1). This finding, that deletion of U3 in Ty elements abolishes the 3′ processing activity *in vitro*, contrasts with HIV-1 pre-mRNA, in which RU5-like transcripts are generally good processing substrates *in vitro* (Gilmartin *et al.*, 1992; Valsamakis *et al.*, 1992). With Ty we have shown that replacing U3 by non-specific sequences restores partial processing activity *in vitro*; this explains why replacement of U3 by *GAL1* sequences *in vivo* retained 50% activity (Yu and Elder, 1989). We have further shown that enhancement of 3′ processing efficiency by specific U3 sequences plays an indispensible role for full utilization of the polyadenylation site. In higher cells, the stability of the processing complex correlates with the efficiency of forming 3′ ends (Ahmed *et al.*, 1991; Weiss *et al.*, 1991; Gilmartin *et al.*, 1992; Mann *et al.*, 1993). It is possible that non-specific RNA is needed to provide a 'loading' site for the processing machinery, forming a less stable than normal but partially functional complex, whereas a specific element is necessary for stabilizing the assembled processing components.

We have identified short motifs in each region related to Ty poly(A) site activation and utilization. TS1 contains

UAGUAUGUA and TS2 has UAG/UCUGUA in its sequence; both have similarities to the UAG/UAUGUA motif proposed by Russo *et al.* (1991). In one of their *cyc1-512* revertants which restores partial function, alterations of UCG/CAUGUA to UAG/UAUGUA were observed ~150 nt upstream of the *CYC1* poly(A) site, suggestive of TS1-like function (Russo *et al.*, 1991). The possibility that the 5′ second A and the fourth U within the UAG/UAUGUA motif may be essential for 3′ end formation *in vivo* and *in vitro* is consistent with an analysis of *CYC1* TS2-like signals [within 30 bp of the poly(A) site], which indicates that 'UA' dinucleotide pairs within a UAUAGUUA motif play a critical role in RNA 3′ processing (S.-Y.Wu and T.Platt, unpublished data).

Mutations within TS1 and TS2 motifs decrease the 3′ processing activity of the substrate, indicating that recognition of these elements by specific protein factors may be involved. Since it has been hard to interpret the results of our band-shift and UV cross-linking attempts in yeast crude extracts, we decided to carry out competition assays with Ty-derived RNAs. In particular, our competition experiments with U3 alone, carrying TS1 sequences, provide the first proof in yeast of some titratable factor(s) causing USE-like activation *in vitro* (Figure 4). U3 with the 2 nt substitution in TS1 failed to compete, suggesting that some processing component(s) may make specific contacts with short distinct signals. Whether one of the activities shown to be involved in specific cleavage of yeast precursors as fractionated by Chen and Moore (1992) might be such a factor, remains to be determined.

The sequence similarity between TS1 and TS2 might reflect a similar role for these elements, such as recruitment of a common factor to these two sites. We hoped to test this possibility by asking whether U3 sequences could compete the processing of H4RU5 substrates containing only TS2. Although preliminary observations indicate that U3 does compete (data not shown), we realized that this experiment cannot distinguish between excess U3 sequestering the components which would bind it more tightly than H4 sequences and competing away a TS2-specific factor. In any event, when both TS1 and TS2 are present, the processing complex assembly appears to be additive and 3′ processing is efficient. If either one of the recognition motifs is absent or mutated, the complex is less stable. Since TS1 and TS2 are separated by 50 nt, RNA conformational changes may also be involved. It will be interesting to determine whether the spacing between the two motifs is important and how the RNA structure is involved in the 3′ end formation. If the resemblance between TS1 and TS2 reflects a functional duplication, activation of hepatitis B virus poly(A) site by multiple copies of its own essential PS1 element would provide a precedent for this mechanism (Russnak, 1991).

Whatever the recognition and activation mechanisms for Ty RNA 3′ processing, they are of general interest: the competition of *CYC1* 3′ processing by Ty elements and U3 sequences (Figure 5) indicates that yeast cellular transcripts and retroelement mRNAs share common components of the processing machinery. In higher eukaryotes, the core polyadenylation signals of retroelements and cellular genes are the same (AAUAAA plus GU box), and the retroelements usually have additional upstream stimulation sequences (lacking identifiable consensus). The downstream GU box is required and behaves in a distance-dependent manner, as does the upstream element of HIV-1 virus (Gilmartin *et al.*, 1992). In the case of Ty, however, no particular downstream sequence is required and the far upstream sequence (i.e. TS1) seems to share characteristics with the near upstream one (i.e. TS2) in terms of primary structure. This is reminiscent of the upstream revertants of *cyc1-512* which restore partial function to the *CYC1* poly(A) site (Russo *et al.*, 1991). Upstream effects on 3′ end formation in yeast have also been reported for CaMV (Irniger *et al.*, 1992) and *ADH2* (Hyman *et al.*, 1991), although whether these affect processing *per se* has not been determined. Our results, however, demonstrate an effect on processing directly, and our previous work on *GAL1* and *GAL10* mRNAs suggests that upstream regions may be involved in those genes as well, though no specific sequences have yet been identified (Sadhale and Platt, 1992). In these instances, as in Ty, it is as if the upstream motif can contribute independently to function at varying distances from the poly(A) site. Such additive action in 3′ processing would differ from the situation found in higher eukaryotic cells, in which one signal (AAUAAA) is indispensible.

In the context of our current results, the parallels between promotors and 3′ processing sites are intriguing. The cleavage site is analogous to a transcription start point, being precise but determined by a spatial relation to the upstream elements more than by a particular local sequence. The promotor TATA box and AAUAAA (higher cells) or bipartite-like motif (yeast) consensus elements are essential for the 'basal' activity. Combinatorial regulation can be conferred by UAS elements in promotors and by upstream elements of 3′ processing sites, using multiple factor binding sites to obtain additive effects where appropriate. We think it likely that a core motif binding factor, analogous to the DNA binding domains of transcription factors, may be required for the context-dependent recruitment of additional activation factors specific for each region. To the extent that these ideas are valid, identification of the titratable factors that recognize the TS1 and TS2 motifs of Ty will provide powerful tools for elucidating the regulation of 3′ mRNA processing in yeast.

Materials and methods

Strains and media

For preparation of yeast whole cell extracts the strain EJ101 (*MATa his⁻ pep4-3 prb1 prc1*) was used (Butler *et al.*, 1990). YEPD medium (1% yeast extract, 2% peptone and 2% dextrose) was used for routine cultivation of yeast strains.

Plasmid constructs

For the construction of U3RU5, a *Cla*I(−190)−*Hin*dIII(+45) fragment, encompassing 235 bp of Ty-D15 LTR, was excised from PK103 (gift from R.T.Elder), end-repaired with Klenow fragment and inserted into the *Sma*I site of pGEM3Z (Promega). In ΔU5, a *Cla*I(−190)−*Bam*HI(+10) fragment representing a 3′ truncated LTR of 197 bp was similarly cloned into pGEM3Z. Removal of U3-specific sequences to create RU5 was carried out by digesting U3RU5 with *Xho*I(−50) and *Bam*HI (polylinker), blunt-ending with Klenow and religating. This reconstituted a *Xho*I site at the DNA junction. Insertion of an end-repaired 150 bp *Sal*I−*Hin*dIII fragment derived from the *HIS4* coding region into the unique *Xho*I site of RU5 in the forward orientation yielded the H4RU5 construct. Mutagenesis of U3RU5 to TS1-93/91C, ΔTS1 and ΔTS2 was carried out using the uracil-substituted DNA method of Kunkel *et al.* (1987). To produce ssDNA, the *Bam*HI−*Eco*RI insert of U3RU5 was transferred to the phagemid pGEM3Zf(+) (Promega) digested with the same enzymes. To create ΔTS1/2, site-directed mutagenesis was done on the TS1-93/91C mutant. A chimeric fragment of non-specific *E.coli* sequence (220 nt) followed by

a 250 bp *Eco*RV–*Hin*dIII fragment of the *CYC1* gene 3′ untranslated region (150 bp are prior to the poly(A) site), was inserted into the pGEM3Z *Sma*I site to create the *CYC1* substrate used in the competition assay.

Yeast whole cell extracts and 3′ processing in vitro

Yeast processing extracts used in this work were 40% ammonium sulfate fractions as described by Butler *et al.* (1990); the RNA substrates tested were also made as described therein. Processing reactions were carried out essentially as described by Butler *et al.* (1990). Five femtomoles of ^{32}P-labeled pre-mRNA were incubated in a total reaction volume of 10 μl containing 20% yeast whole cell extracts, 2 mM ATP, 2% PEG, 1 mM Mg^{2+} and 75 mM potassium acetate for 20 min at 30°C. For the competition assays, unlabeled competitor RNA in a 10- or 30-fold excess was mixed with ^{32}P-labeled pre-mRNA. The reactions were stopped by the addition of SDS to 0.5% and treated with 2 mg/ml proteinase K. The RNAs were then extracted and analyzed on 4% denaturing polyacrylamide gels. Quantitation of the signals was done by laser densitometry scanning (LKB UltraScan XL). Efficiency of 3′ processing was measured by dividing the total signals of cleavage product and poly(A) product by that of the input RNA.

Acknowledgements

We thank Dr R.T.Elder for kindly providing the PK103 plasmid. We thank members of our laboratory for discussions and encouragement, especially Shwu-Yuan Wu who made the *CYC1* substrate. We also thank Drs Scott Butler, Eric Phizicky and Elizabeth Grayhack for their critical comments and suggestions on this manuscript. This work was supported in part by Public Health Service grant 2-R01-GM35658 to T.P. and a National Cancer Institute of Canada Senior Fellowship to R.R.

References

Ahmed,Y.F., Gilmartin,G.M., Hanly,S.M., Nevins,J.R. and Greene,W.C. (1991) *Cell*, **64**, 727–737.
Boeke,D.J. (1989) In Berg,D. and Howe,M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 335–374.
Boeke,D.J. and Sandmeyer,S. (1991) In Broach,J.R., Jones,E.W. and Pringle,J. (eds), *The Molecular and Cellular Biology of the Yeast Saccharomyces cerevisiae*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 193–261.
Brown,P.H., Tiley,L.S. and Cullen,B.R. (1991) *J. Virol.*, **65**, 3340–3343.
Butler,J.S. and Platt,T. (1988) *Science*, **242**, 1270–1274.
Butler,J.S., Sadhale,P.P. and Platt,T. (1990) *Mol. Cell. Biol.*, **10**, 2599–2605.
Carswell,S. and Alwine,J.C. (1989) *Mol. Cell. Biol.*, **9**, 4248–4258.
Chen,J. and Moore,C. (1992) *Mol. Cell. Biol.*, **12**, 3470–3481.
Cherrington,J.M. and Ganem,D. (1992) *EMBO J.*, **11**, 1513–1524.
Cherrington,J.M., Russnak,R. and Ganem,D. (1992) *J. Virol.*, **66**, 7589–7596.
DeZazzo,J.D. and Imperiale,M.J. (1989) *Mol. Cell. Biol.*, **9**, 4951–4961.
DeZazzo,J.D., Kilpatrick,J.E. and Imperiale,M.J. (1991) *Mol. Cell. Biol.*, **11**, 1624–1630.
DeZazzo,J.D., Scott,J.M. and Imperiale,M.J. (1992) *Mol. Cell. Biol.*, **12**, 3699–3705.
Gilmartin,G.M., Fleming,E.S. and Oetjen,J. (1992) *EMBO J.*, **11**, 4419–4428.
Hyman,L.E., Seiler,S.H., Whoriskey,J. and Moore,C.L. (1991) *Mol. Cell. Biol.*, **11**, 2004–2012.
Imperiale,M.J. and DeZazzo,J.D. (1991) *New Biol.*, **3**, 531–537.
Irniger,S., Sanfaçon,H., Egli,C.M. and Braus,G.H. (1992) *Mol. Cell. Biol.*, **12**, 2322–2330.
Iwasaki,K. and Temin,H.M. (1990) *Genes Dev.*, **4**, 2299–2307.
Ju,G. and Cullen,B.R. (1985) *Adv. Virus Res.*, **30**, 179–223.
Kunkle,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
Mann,K.P., Weiss,E.A. and Nevins,J.R. (1993) *Mol. Cell. Biol.*, **13**, 2411–2419.
Mogen,B.D., MacDonald,M.H., Leggewie,G. and Hunt,A.G. (1992) *Mol. Cell. Biol.*, **12**, 5406–5414.
Proudfoot,N.J. (1991) *Cell*, **64**, 671–674.
Russnak,R. (1991) *Nucleic Acids Res.*, **19**, 6449–6456.
Russnak,R. and Ganem,D. (1990) *Genes Dev.*, **4**, 764–776.
Russo,P., Li,W.-Z., Hampsey,D.M., Zaret,K.S. and Sherman,F. (1991) *EMBO J.*, **10**, 563–571.
Sadhale,P.P. and Platt,T. (1992) *Mol. Cell. Biol.*, **12**, 4215–4229.
Sadhale,P.P., Sapolsky,R., Davis,R.W., Butler,J.S. and Platt,T. (1991) *Nucleic Acids Res.*, **19**, 3683–3688.
Sanfaçon,H. and Hohn,T. (1990) *Nature*, **346**, 81–84.
Sanfaçon,H., Brodmann,P. and Hohn,T. (1991) *Genes Dev.*, **5**, 141–149.
Schek,N., Cooke,C. and Alwine,J.C. (1992) *Mol. Cell. Biol.*, **12**, 5386–5393.
Valsamakis,A., Zeichner,S., Carswell,S. and Alwine,J.C. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 2108–2112.
Valsamakis,A., Schek,N. and Alwine,J.C. (1992) *Mol. Cell. Biol.*, **12**, 3699–3705.
Wahle,E. and Keller,W. (1992) *Annu. Rev. Biochem.*, **61**, 419–440.
Weichs an der Glon,C., Monks,J. and Proudfoot,N.J. (1991) *Genes Dev.*, **5**, 244–253.
Weichs an der Glon,C., Ashe,M., Eggermont,J. and Proudfoot,N.J. (1993) *EMBO J.*, **12**, 2119–2128.
Weiss,E., Gilmartin,G.M. and Nevins,J.R. (1991) *EMBO J.*, **10**, 215–219.
Wilson-Gunn,S.I., Kilpatrick,J.E. and Imperiale,M.J. (1992) *J. Virol.*, **66**, 5418–5424.
Yu,K. and Elder,R.T. (1989) *Mol. Cell. Biol.*, **9**, 2431–2444.
Zaret,K.S. and Sherman,F. (1982) *Cell*, **28**, 563–573.

Received on September 6, 1993; revised on October 19, 1993