The EMBO Journal vol.10 no.11 pp.3503-3511, 1991

Homologous mRNA 3′ end formation in fission and budding yeast

Tim Humphrey, Parag Sadhale[1], Terry Platt[1] and Nick Proudfoot

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK and [1]Department of Biochemistry, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, NY 14642, USA

Communicated by N.Proudfoot

Sequences resembling polyadenylation signals of higher eukaryotes are present downstream of the *Schizosaccharomyces pombe ura4*$^+$ and *cdc10*$^+$ coding regions and function in HeLa cells. However, these and other mammalian polyadenylation signals are inactive in *S.pombe*. Instead, we find that polyadenylation signals of the *CYC1* gene of budding yeast *Saccharomyces cerevisiae* function accurately and efficiently in fission yeast. Furthermore, a 38 bp deletion which renders this RNA processing signal non-functional in *S.cerevisiae* has the equivalent effect in *S.pombe*. We demonstrate that synthetic pre-mRNAs encoding polyadenylation sites of *S.pombe* genes are accurately cleaved and polyadenylated in whole cell extracts of *S.cerevisiae*. Finally, as is the case in *S.cerevisiae*, DNA sequences encoding regions proximal to the *S.pombe* mRNA 3′ ends are found to be extremely AT rich; however, no general sequence motif can be found. We conclude that although fission yeast has many genetic features in common with higher eukaryotes, mRNA 3′ end formation is significantly different and appears to be formed by an RNA processing mechanism homologous to that of budding yeast. Since fission and budding yeast are evolutionarily divergent, this lower eukaryotic mechanism of mRNA 3′ end formation may be generally conserved.

Key words: mRNA 3′ end formation/*S.cerevisiae*/*S.pombe*/*yeast.*

Introduction

The 3′ ends of eukaryotic messenger RNA are generally formed by endonucleolytic cleavage and polyadenylation of the nascent mRNA transcript by nuclear factors interacting with the highly conserved AAUAAA sequence (for review see Wickens, 1990). The budding yeast, *Saccharomyces cerevisiae*, provides a notable exception to this mechanism, where although mRNA 3′ ends are polyadenylated, they usually lack the canonical AAUAAA motif. The first indication of which sequences control mRNA 3′ end formation in *S.cerevisiae* came from the analysis of an iso-1-cytochrome C (*CYC1*) deletion mutation (*cyc1-512*). The absence of 38 bp from the 3′ end of this gene results in reduced levels of mRNA transcripts which exhibit extended 3′ termini (Zaret and Sherman, 1982). Analysis of the sequences in these extended transcripts revealed a base sequence homology also present in the original *CYC1* 38 bp deletion TAG...TAGT...TTT (Zaret and Sherman, 1984). These and further studies (Osborne and Guerente, 1989; Russo and Sherman, 1989) on the mRNA 3′ end formation signal of *CYC1* led to the conclusion that this signal promotes termination of transcription in *S.cerevisiae*. Indeed it seemed plausible that polyadenylation might be directly coupled to transcriptional termination. In spite of these predictions, Butler and Platt (1988) have demonstrated that *CYC1* transcripts synthesized *in vitro* are efficiently 3′ end cleaved and polyadenylated in *S.cerevisiae* whole cell extracts and that the 38 bp deletion near the *CYC1* 3′ end blocks this activity. A number of different *S.cerevisiae* mRNAs have similarly been shown to 3′ process and polyadenylate *in vitro* although with varying efficiency (Butler *et al.*, 1990). It therefore seems likely that *S.cerevisiae*, like mammals, forms the 3′ ends of its mRNA in a two step process involving both termination of transcription and cleavage/polyadenylation. However, these two processes may be more tightly linked in *S.cerevisiae* than they are in mammals.

We have analysed the signals involved in mRNA 3′ end formation for the fission yeast *Schizosaccharomyces pombe*, an organism in which the processes of mRNA transcription (Russell, 1983; Jones *et al.*, 1988) and splicing (Käufer *et al.*, 1985) appear to be more closely related to higher eukaryotes than to *S.cerevisiae*. We demonstrate that, despite the presence of higher eukaryotic polyadenylation signals, *S.pombe* instead utilizes different signals that are functionally equivalent to those of *S.cerevisiae*.

Results

Fission yeast possesses functional mammalian polyadenylation sites

A survey of the downstream non-coding region of *S.pombe* genes reveals the presence of AATAAA sequences (or close derivatives) near the position mapped as the mRNA 3′ terminus as shown in Figure 1. This raises the possibility that the mechanism of polyadenylation in *S.pombe* may resemble that of higher eukaryotes. We therefore began these studies by testing whether these mammalian-like poly(A) signals in *S.pombe* genes could function in a mammalian cell. Two genes were chosen for these experiments, the *ura4*$^+$ gene (Grimm *et al.*, 1988) encoding the enzyme orotidine 5′ phosphate decarboxylase, and the cell cycle control gene *cdc10*$^+$ (Aves *et al.*, 1985). Both genes possess AATAAA-like and GT-rich sequences, the second part of the mammalian poly(A) signal (McDevitt *et al.*, 1984; Gil and Proudfoot, 1984), upstream and downstream respectively of the provisionally mapped mRNA 3′ ends. Fragments (~200 bp) containing these sequences were inserted into the *Pvu*II site downstream of the polyadenylation signal of both a functional and non-functional, thalassaemic (AATAAG) human α-globin gene (Higgs *et al.*, 1983; Whitelaw and Proudfoot, 1986)

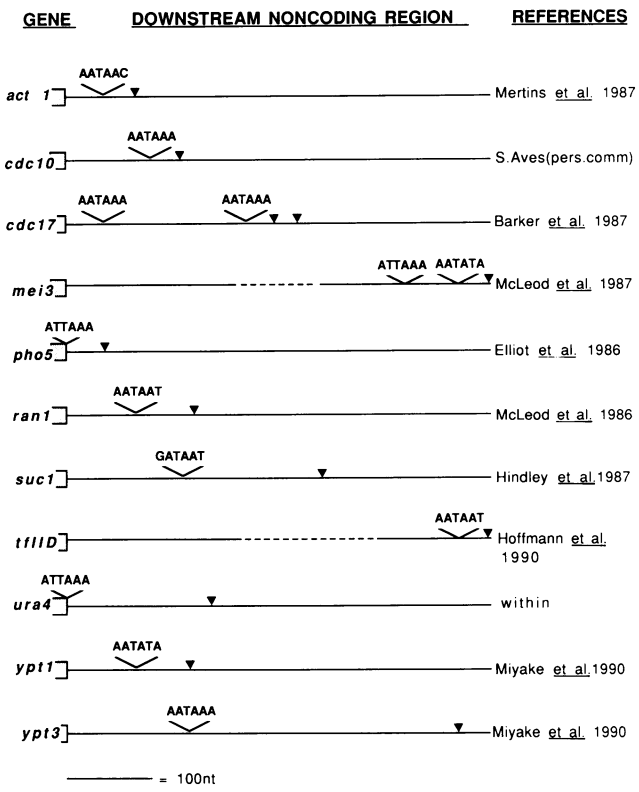

Fig. 1. Schematic diagram indicating the distance relationship between sequences most resembling the canonical higher eukaryotic polyadenylation signal (AATAAA) and the mRNA 3′ end (filled arrow) of all *S.pombe* genes known by authors to be mapped at time of publication. The coding region is represented as a grey block.

(Figure 2B). These constructs were contained within a transient expression vector pSVed, which possesses the SV40 replication origin and enhancer sequences (Proudfoot *et al.*, 1984). Such constructs were transfected into HeLa cells together with a plasmid containing the large T antigen and rabbit β-globin genes which has the dual function of activating pSVed replication and providing β-globin mRNA as a transfection control. Cytoplasmic RNA was analysed by S1 mapping, using a probe spanning both the α-globin and *S.pombe* polyadenylation sites, as well as a separate probe for the 3′ ends of β-globin mRNA. If the α-globin or *S.pombe* polyadenylation sites are used, fragments of 217 nt or ~450 nt respectively would be protected from S1 digestion. The β-globin cotransfection control gives an S1 protected band of 175 nt (Levitt *et al.*, 1989). The S1 protection experiment shows bands of the size expected if the *ura4* and *cdc10 S.pombe* sequences were functioning as polyadenylation signals in HeLa cells (Figure 2B, lanes 4 and 8). These sequences do not function in the reverse orientation (Figure 2B, lanes 3, 5, 7 and 9) and work inefficiently when cloned downstream of a functional α-globin polyadenylation signal (Figure 2B, lanes 2 and 6). From the size of the S1 protected bands (Figure 2B, lanes 4 and 8) the site of cleavage within the *S.pombe* gene fragments was estimated to be ~20 nt downstream of the AATAAA sequence in HeLa cells. This was confirmed in the case of the *ura4* insert by RNase protection analysis (Figure 3B). This mRNA was found to be polyadenylated (not shown). These data show that *S.pombe* contains functionally recognized higher eukaryotic polyadenylation signals.

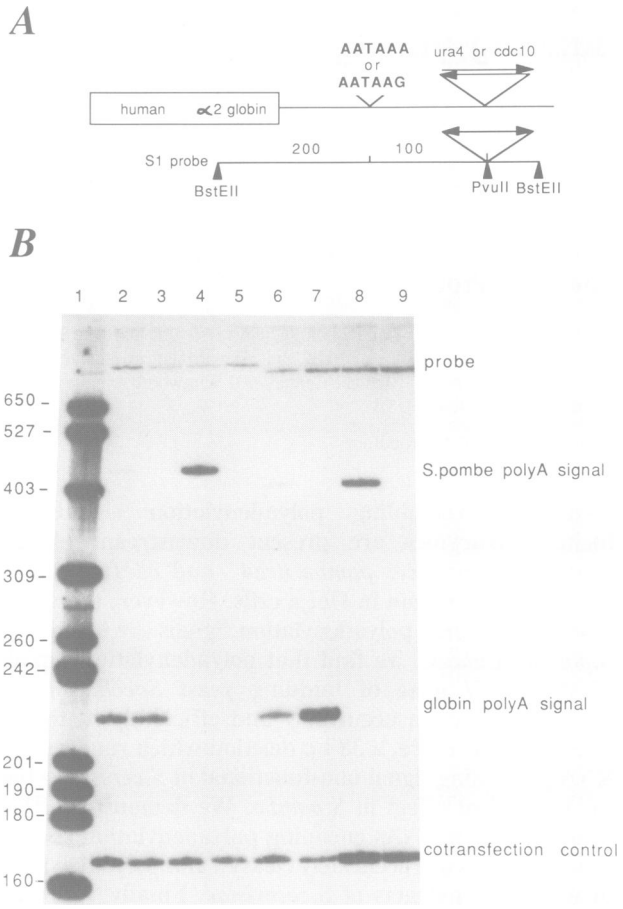

Fig. 2. Analysis of *S.pombe* sequences as higher eukaryotic polyadenylation signals. (**A**) Diagramatical representation of the human α-globin constructs transfected into HeLa cells and S1 probe. See text for further details. (**B**) S1 analysis of RNA from HeLa cells transfected with constructs containing *S.pombe* higher eukaryotic-like polyadenylation signals. This figure shows a 6% polyacrylamide denaturing gel from which the sizes of the S1 protected fragments were determined. Constructs transfected into HeLa cells are as follows; the *ura4* fragment inserted 5′−3′ into wild type human α-2 globin (α2W) *Pvu*II site (lane 2). The *ura4* fragment inserted 3′−5′ into α2W *Pvu*II site (lane 3). The *ura4* fragment inserted 5′−3′ into the thalassaemic α-2 globin (α2M) *Pvu*II site (lane 4). The *ura4* fragment inserted 3′−5′ into α2M *Pvu*II site (lane 5). Identical constructs except containing the *cdc10* fragment were transfected respectively (lanes 6−9). Lane 1, marker lane.

The 3′ ends of ura4 mRNA are different in HeLa cells and S.pombe

To map more precisely the 3′ end of the *ura4* mRNA in both HeLa cells and *S.pombe*, a riboprobe across the *Eco*RV−*Sca*I region of *ura4* was generated (Figure 3A); and the resulting RNase protection analysis is shown in Figure 3B. The hybrid human α-globin−*ura4* mRNA formed in HeLa cells with its 3′ end within the *ura4* sequence (Figure 2B, lane 4) is represented by an RNase protected band of 136 nt (Figure 3B, lane 1). The precise sequence position of this polyadenylation site is shown in Figure 3D (indicated by an open triangle). Surprisingly this band is not present in *S.pombe*, (Figure 3B, lane 2) even though a strong band is observed at 115 nt in both *S.pombe* and HeLa derived *ura4* 3′ transcripts (Figure 3B, lanes 1 and 2). However, this 115 nt band is also detected when an unprocessed synthetic RNA template spanning the whole *ura4* 3′ end

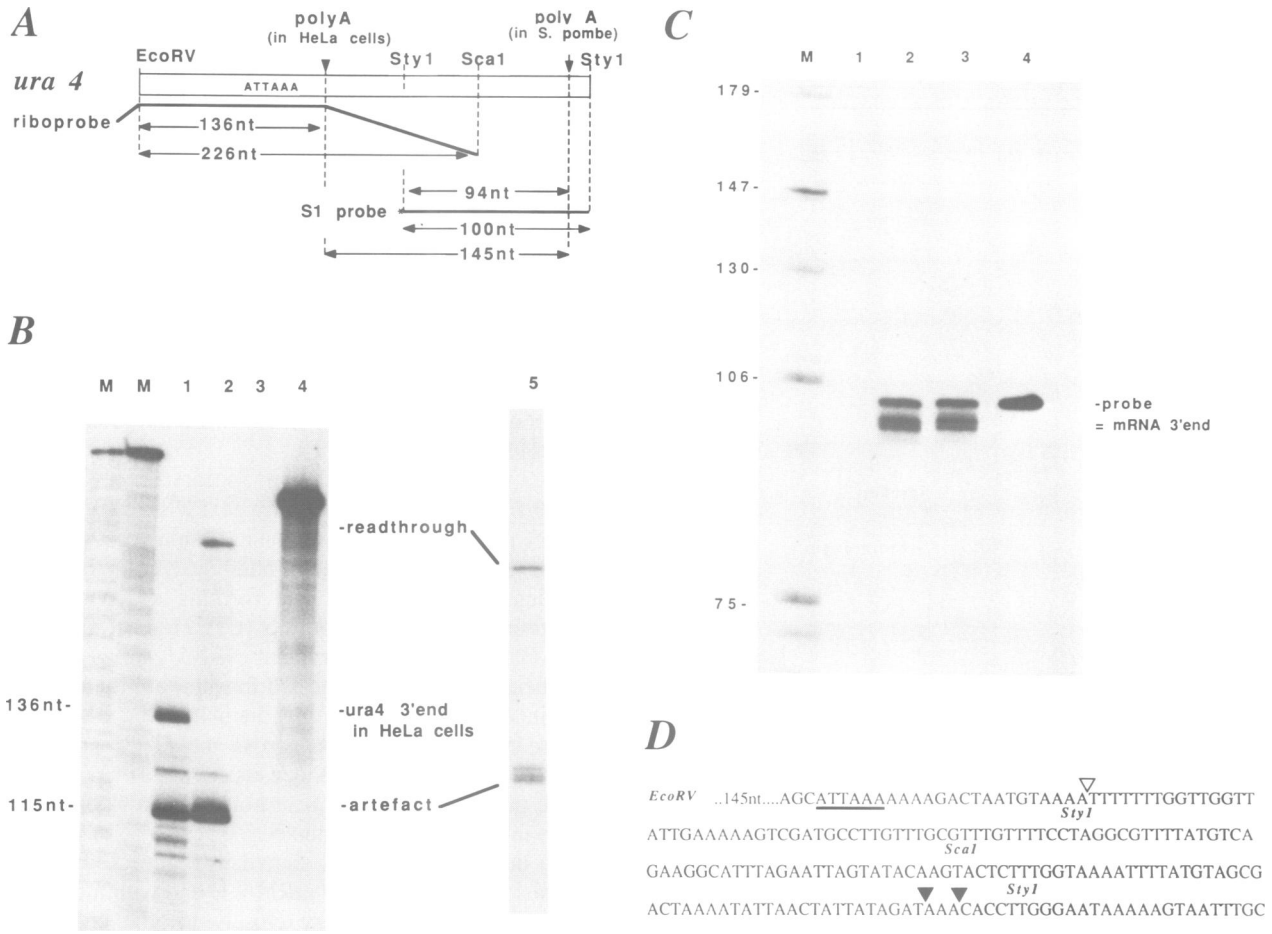

Fig. 3. Analysis of *ura4* mRNA 3′ end formation in HeLa cells and *S.pombe*. (**A**) Schematic diagram to indicate sites of *ura4* mRNA 3′ end formation in HeLa cells and in *S.pombe* as determined by RNase and S1 protection analysis. (**B**) RNase protection analysis across *ura4* higher eukaryotic-like polyadenylation signal (*Eco*RV−*Sca*I fragment) in HeLa cells and *S.pombe*. A 6% polyacrylamide denaturing gel is shown depicting the lengths of RNase protected bands of the *ura4* riboprobe hybridized to: cytoplasmic RNA from HeLa cells transfected with human α-globin−*ura4* hybrid (lane 1); total RNA from *S.pombe* (lane 2); tRNA (lane 3); SP6 *Eco*RV−*Sca*I sense transcript (lane 5), or the SP6 antisense probe alone (lane 4). Sizes were determined against a DNA sequence ladder (M). (**C**) Mapping the 3′ end of the *S.pombe ura4* mRNA. A 6% polyacrylamide denaturing gel is shown depicting the lengths of S1 protected bands of *ura4* S1 probes hybridized to: tRNA (lane 1); total RNA from wild type *S.pombe* (lane 2); total RNA from *leu*$^-$, *ade*$^-$, *ura*$^-$ *S.pombe* transfected with PUS construct containing a non-truncated *ura4* gene (see text for details) (lane 3). Probe alone (lane 4), Markers, (M). (**D**) Sequence of 3′ non-coding region of the *S.pombe ura4* gene. The mapped *S.pombe ura4* mRNA 3′ ends formed in HeLa cells (panel B lane 1) and in *S.pombe* (panel C lanes 2 and 3) are indicated with open and filled triangles respectively. The higher eukaryotic-like polyadenylation signal is underlined.

*Eco*RV−*Sca*I region is probed (Figure 3B, lane 5). This signal is therefore generated by RNA−RNA breathing and does not represent an authentic 3′ end. From S1 analysis, using the end labelled *Sty*I DNA probe (Figure 3A) we detect a doublet band (Figure 3C, lanes 2 and 3), 145 nt and 148 nt downstream from the *ura4* mammalian-like poly(A) signals. These bands define authentic 3′ ends of the *ura4* mRNA in *S.pombe* (the sequence positions denoted by filled triangles are indicated in Figure 3D). These results suggest that the 3′ end formation machinery of HeLa cells and *S.pombe* are not functionally equivalent.

Mammalian polyadenylation signals are non-functional in S.pombe

We next carried out a reciprocal experiment to test whether higher eukaryotic polyadenylation signals could function in *S.pombe*. The SV40 early promoter has been previously shown to be expressed well in *S.pombe* and initiates at the same site as in mammalian cells (Jones *et al.*, 1988). Furthermore, the intron from the SV40 early gene transcript has been shown to be correctly excised by *S.pombe* (Käufer *et al.*, 1985). To test whether the polyadenylation signal of the SV40 early transcript is recognized by *S.pombe* pSMCAT was transformed into a *leu*$^-$ strain of *S.pombe*. This plasmid contains the *S.cerevisiae LEU2* gene with the chloramphenicol acetyl transferase (CAT) coding sequences cloned between the SV40 early promoter and polyadenylation signal (Jones *et al.*, 1988). Total RNA from transformed *S.pombe* was isolated and hybridized to an SP6 generated riboprobe spanning the SV40 early polyadenylation site, as depicted in Figure 4A. Digestion of this RNA−RNA hybrid with RNase A and T1 gave rise to a protected band of 139 nt (Figure 4B, lane 4). The size of this band corresponds to an intact SV40 mRNA transcript in which the polyadenylation signal has been ignored. If the SV40 polyadenylation signal was recognized by a mechanism akin to that in higher eukaryotes, a riboprobe protection band of 83 nt would be obtained.

The human α-2 globin and adenovirus L3 polyadenylation signals were also tested in *S.pombe*. DNA fragments

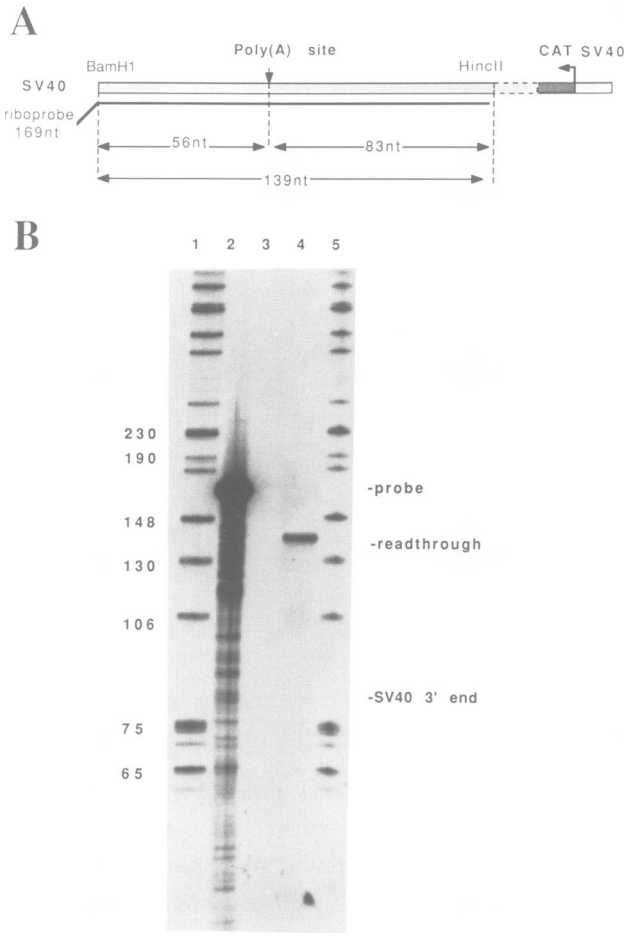

Fig. 4. Analysis of the SV40 polyadenylation signal in *S.pombe*. (**A**) Schematic diagram of construct pSMCAT transformed into *S.pombe* (*leu*$^-$ strain). Polyadenylation site and RNase protection sizes expected if polyadenylation signal is or is not recognized, are indicated. (**B**) RNase protection analysis across SV40 early polyadenylation signal. A 6% polyacrylamide denaturing gel is shown depicting the lengths of RNase protected bands of the SV40 riboprobe alone (lane 2), hybridized to tRNA (lane 3), and hybridized to total RNA from *S.pombe* (*leu*$^-$ strain) transformed with pSMCAT (lane 4). Sizes were determined against DNA markers (lanes 1 and 5).

containing these polyadenylation signals were inserted downstream of the *S.pombe ura4* gene (Grimm *et al.*, 1988) with its 3′ end deleted, within the plasmid PUS (Figure 5A and Materials and methods). The ability of these higher eukaryotic fragments to form mRNA 3′ ends was determined by transforming *ura4*$^-$ *Leu*$^-$ *S.pombe* with PUS constructs (2 μm, *LEU2*) and assaying for 3′ end formation within these regions by RNase protection analysis using total RNA from transformed *S.pombe*. Specific SP6 derived riboprobes of α-2 globin and L3 adenovirus poly(A) sites were used for the RNase protection analysis. The sizes of bands expected from RNase protection analysis if the human α-globin and adenovirus polyadenylation signals were functional in *S.pombe* are shown in Figure 5A. The RNase protection analysis of the *ura4*$^-$ *human* α-2 globin and *ura4*−L3 hybrid gene constructs are shown in Figure 5B, lanes 3 and 7 respectively. Only RNA products corresponding to readthrough transcripts of the higher eukaryotic polyadenylation signals are obtained. The level of message reading through the α-2 globin poly(A) site is comparable with that of a thalassaemia mutant (AATAAG) polyadenylation signal in an otherwise identical construct (Figure 5B, compare lanes 3 and 4). Similarly a mutated L3 poly(A) site (AAGAAA) shows RNA mismatch cleavage products (Figure 5B, lane 8) which have a combined band intensity approximately equal to that of the readthrough level across the wild type L3 polyadenylation signal (Figure 5B, lane 7). Furthermore, from Northern blot analysis these signals do not cause 3′ end formation outside of the riboprobed region (data not shown). These results as well as the previous RNase protection analysis of *ura4* mRNA (Figure 3B, lane 2) demonstrate that higher eukaryotic polyadenylation signals are inactive in *S.pombe*.

S.pombe and S.cerevisiae have interchangeable polyadenylation signals

Since *S.pombe* polyadenylation appears to be a process distinct from that in mammals, we next wished to determine whether mRNA 3′ end formation in fission yeast resembled that of the distantly related budding yeast. Fragments encoding the essential 3′ end forming sequence of the *S.cerevisiae* iso-1-cytochrome C (*CYC1*) gene, and a non-functional 38 bp deletion mutant *cyc1-512* (*cyc1*Δ) (Zaret and Sherman, 1982) were tested for their ability to form 3′ ends in *S.pombe* by inserting them into the PUS vector (Figure 5A). Transcription across the *CYC1* and *cyc1*Δ sequences from the above constructs was mapped using RNase protection analysis with homologous riboprobes as shown in Figure 5A. The nucleotide sequence of the 38 bp region deleted in *cyc1*Δ is shown in Figure 7. From this analysis, the *S.pombe ura4* mRNA 3′ end within the *CYC1* fragment maps to, or very closely to, the endogenous *S.cerevisiae CYC1* mRNA 3′ end (Figure 5C, lanes 3 and 4). The pattern, however, is displaced due to the presence of polylinker sequence (see Figure 5A). Virtually no mRNA 3′ end formation within the *cyc1*Δ fragment is detected (Figure 5C, lane 5).

The mRNA 3′ ends of a number of *S.cerevisiae* genes have been shown to be generated by RNA processing *in vitro* (Butler *et al.*, 1990). To determine whether this is also the case in *S.pombe*, a 192 nt SP6 pre-mRNA was synthesized *in vitro* which contained a 100 nt sequence encoding the 3′ end of the *S.pombe ura4* gene (see Figure 6A). This was incubated in *S.cerevisiae* whole cell extract, under cleavage and polyadenylation conditions (Butler and Platt, 1988). *In vitro* polyadenylation of the *ura4* pre-mRNA is clearly observed, as depicted by the characteristic heterogeneous increase in pre-mRNA length with increasing incubation time (Figure 6B, lanes 3−7). These conditions are also shown to be optimal for the specific polyadenylation of the *S.cerevisiae CYC1* pre-mRNA (Figure 6B, lanes 1 and 2). When incubated under cleavage conditions, in which polyadenylation is inhibited, specific cleavage products of the *ura4* pre-mRNA are also observed with increasing incubation time (Figure 6B, lanes 8−11). The 3′ half molecule from the cleaved pre-mRNA is not observed under these conditions. To determine whether the cleaved product and not the pre-mRNA was being specifically polyadenylated, a longer (336 nt) transcript was generated and incubated under polyadenylation and cleavage conditions (Figure 6C, lanes 1 and 2 respectively). Figure 6C demonstrates that it is the cleaved products that are polyadenylated. The site of cleavage of the *ura4* pre-mRNA

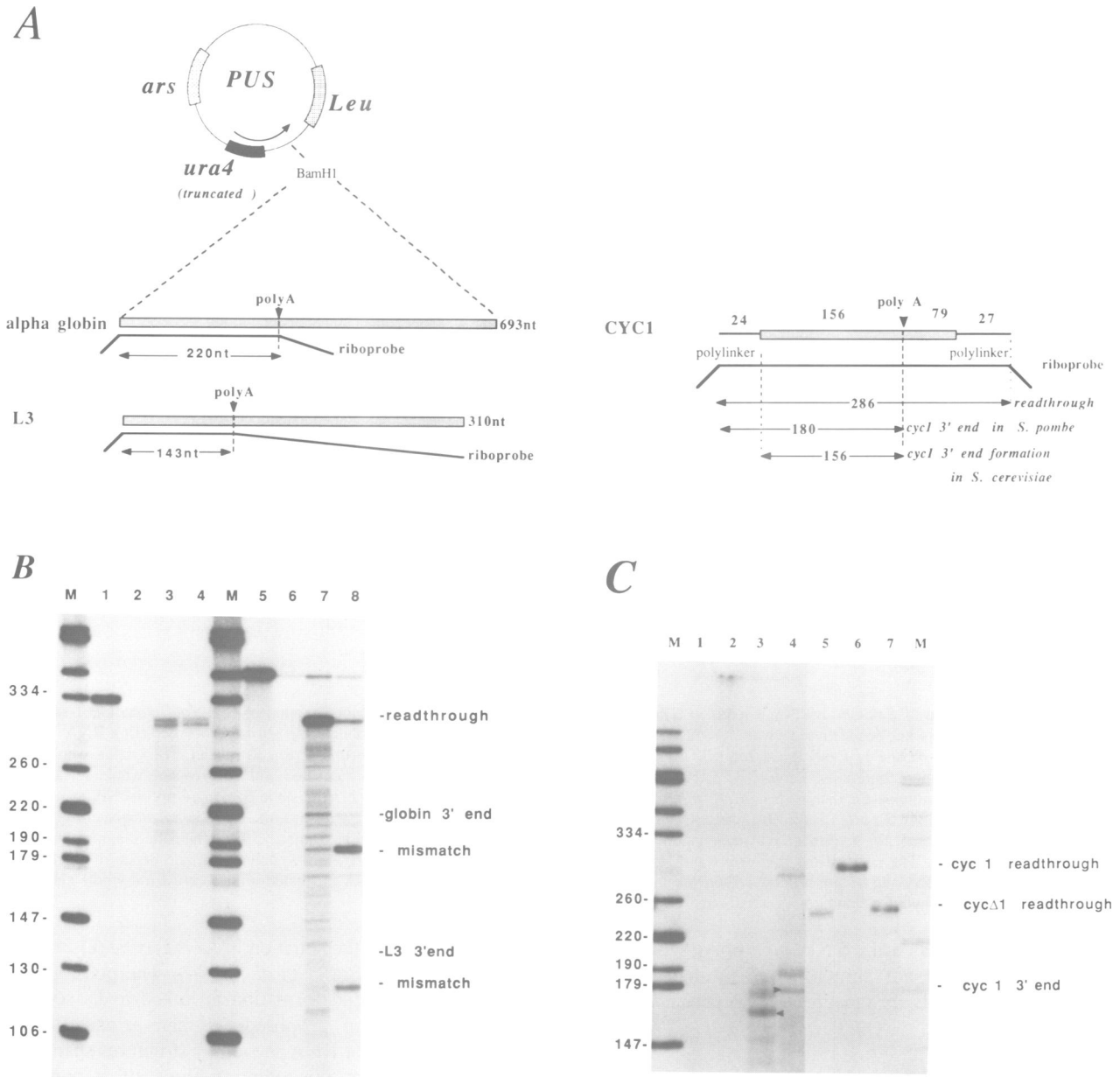

Fig. 5. Analysis of various polyadenylation signals in *S.pombe* by RNase protection. (**A**) Schematic diagram of constructs transformed into *S.pombe*. Polyadenylation sites and RNase protection sizes expected if polyadenylation signals are being recognized are indicated. (**B**) RNase protection analysis across higher eukaryotic polyadenylation signals and point mutations thereof in *S.pombe*. A 6% polyacrylamide denaturing gel is shown depicting the lengths of RNase protected bands of the globin riboprobe alone (lane 1) or hybridized to: tRNA (lane 2); total RNA from *S.pombe* transformed with PUS−α-globin (lane 3); total RNA from *S.pombe* transformed with PUS−α-globin with point mutation (AATAAG) (lane 4). The L3 riboprobe alone (lane 5) or hybridized to: tRNA (lane 6); total RNA from *S.pombe* transformed with PUS−L3 (lane 7); total RNA from *S.pombe* transformed with PUS−L3 with point mutation (AAGAAA) (lane 8). Markers (M). (**C**) RNase protection analysis of *S.cerevisiae CYC1* 3′ fragment in *S.pombe*. A 6% polyacrylamide denaturing gel is shown depicting the lengths of RNase protected bands of *CYC1* and *cyc1*Δ riboprobes hybridized to various RNAs. Lanes 1 and 2; *CYC1* and *cyc1*Δ riboprobes respectively hybridized to RNA. Lane 3; *CYC1* riboprobe hybridized to total RNA from wild type *S.cerevisiae*. Lane 4; *CYC1* riboprobe hybridized to total RNA from *S.pombe* PUC−*CYC1* transformant. Lane 5; *cyc1*Δ riboprobe hybridized to total RNA from *S.pombe* PUS−*cyc1*Δ transformant. Lanes 6 and 7 are *CYC1* and *cyc1*Δ probes alone, respectively.

in vitro using *S.cerevisiae* extracts (Figure 6B) matches the *S.pombe ura4* mRNA 3′ end *in vivo* (see Figure 3C). In both cases a doublet band pattern is observed. To show that cleavage and polyadenylation of *S.pombe* synthetic pre-mRNA by *S.cerevisiae* extract was a general phenomenon, these experiments were extended to a second pre-mRNA (Figure 6D and E). Downstream non-coding regions of the *S.pombe tfIID*, in which the mRNA 3′ end has been accurately mapped by comparison of cDNA and genomic sequences (Hoffmann *et al.*, 1990), was subcloned into an appropriate vector and synthetic pre-mRNAs were generated. The sizes of bands expected if the polyadenylation sites used *in vitro* matched those *in vivo* is shown in Figure 6D. Figure 6E shows the results of this *in vitro* polyadenylation experiment. The untreated pre-mRNA shown in lane 1 is converted into a product of ~460 nt (the predicted size) under cleavage conditions (lane 3) while both the cleaved and polyadenylated products are obtained under polyadenylation conditions (lane 2: this lane has lower levels of input pre-mRNA). This additional data on the *S.pombe tfIID*

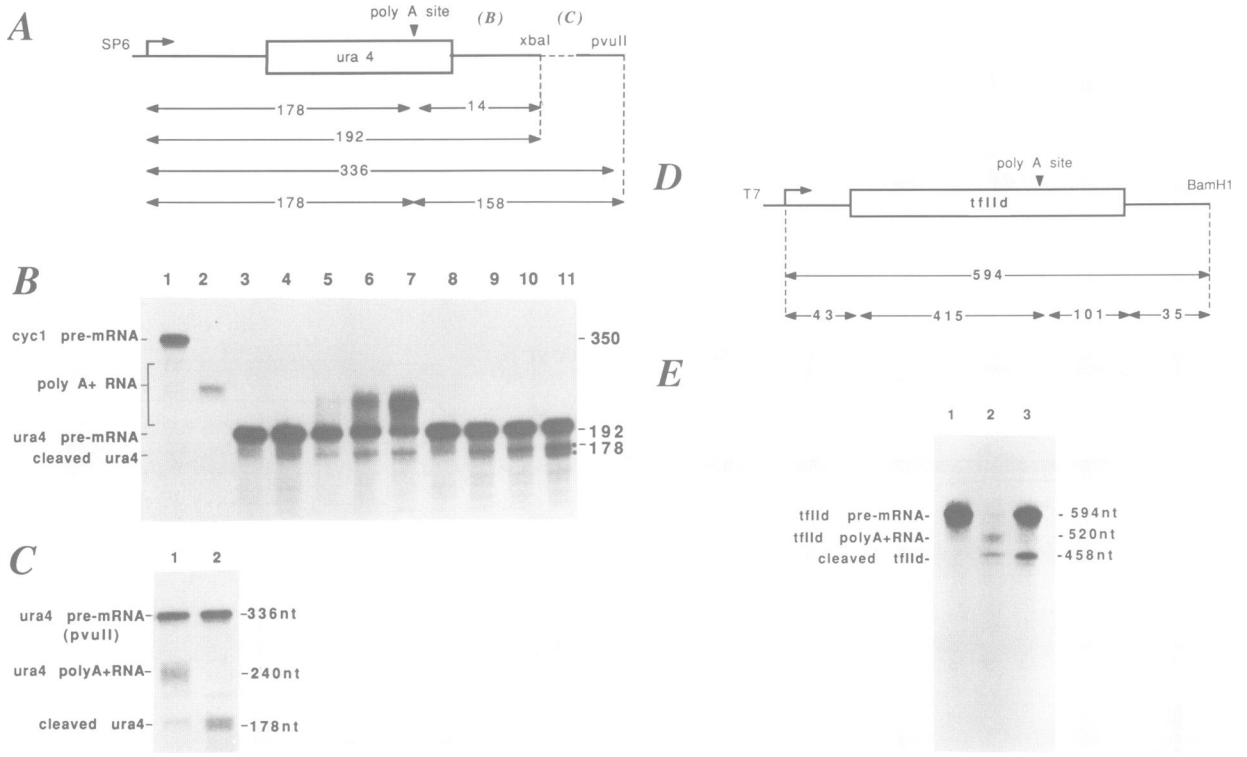

Fig. 6. *In vitro* processing of *S.pombe* mRNA 3′ ends in *S.cerevisiae* whole cell extract. (**A**) Schematic diagram of SP6 *ura4* clone and transcripts. The sizes of the transcripts and respective cleavage products are shown. (**B**) A 5% polyacrylamide denaturing gel is shown depicting RNA processing of pre-mRNAs. Lane 1, *Pvu*II linearized pGYC1 transcript. Lane 2, pGYC1 transcript incubated for 30 min in *S.cerevisiae* whole cell extract under polyadenylation conditions. Lanes 3–7; *Xba*I linearized SP6 *ura4* pre-mRNA (192 nt) incubated with *S.cerevisiae* whole cell extract under polyadenylation conditions for 0 min, 5 min, 15 min, 30 min and 1 h respectively. Lanes 8–11: the same SP6 *ura4* pre-mRNA incubated under cleavage conditions for 5 min, 15 min, 30 min and 1 h respectively. (**C**) *Pvu*II linearized SP6 *ura4* pre-mRNA (336 nt) incubated with *S.cerevisiae* whole cell extract for 1 h under polyadenylation conditions (lane 1) and cleavage conditions (lane 2). (**D**) Schematic diagram of pGEM *tfIId* clone and transcripts. The sizes of the transcripts and respective cleavage products are shown. (**E**) A 5% polyacrylamide denaturing gel is shown depicting RNA processing of pre-mRNAs. *TfIId* pre-mRNA (594 nt) (lane 1). *TfIId* pre-mRNA incubated with *S.cerevisiae* whole cell extract for 30 min under polyadenylation conditions (lane 2) and cleavage conditions (lane 3).

poly(A) site as well as a third *S.pombe* poly(A) site from the *ypt1* gene (data not shown) confirm that *S.pombe* polyadenylation signals are recognized by *S.cerevisiae* RNA processing activities.

S.pombe polyadenylation signals like those of S.cerevisiae are AT rich but possess no obvious sequence motif

In an attempt to identify a consensus signal for mRNA 3′ end formation in *S.pombe*, a number of novel *S.pombe* cDNA 3′ ends were isolated and sequenced. These sequences were compared with those sequences published in which the 3′ end of the *S.pombe* genes have been accurately mapped. A number of sequence homologies can be found in common with both higher eukaryotic polyadenylation signals and those of *S.cerevisiae* (Figure 7). In particular the sequence TTTTTAT or close derivatives (underlined) is present in many although not all of the cDNA sequences, and is similar to the *S.cerevisiae* termination signal proposed by Henikoff *et al.* (1983). The AT-rich sequence (dotted underlined) motif defined as a polyadenylation signal in the *S.cerevisiae GAL7* gene (Abe *et al.*, 1990) is present in two of the cDNAs. Finally the termination signal (bracketed) based on the *S.cerevisiae CYC1* gene (Zaret and Sherman, 1982) is only present in one of the *S.pombe* cDNAs. However, in general the 3′ ends of *S.pombe* appear not to share any common motif. This apparent absence of sequence motif clearly resembles the situation in *S.cerevisiae* (Osborne and Guerente, 1989; Butler *et al.*, 1990). The possibility that yeast mRNA 3′ end formation is linked to its secondary structure has been proposed (Hyman *et al.*, 1991). A computer search for RNA secondary structures spanning the entire downstream noncoding regions of a number of *S.pombe* genes was performed. However, no common structures were found (not shown). Interestingly, a property common to all DNA sequences encoding mRNA 3′ ends in both *S.pombe* and *S.cerevisiae* is the extreme AT richness of these regions (see Figure 7). Such AT richness is shared with the sequences spanning the mRNA 3′ ends in *S.cerevisiae*, and is also a property shared with sequences involved in rho-dependent termination in prokaryotes (Osborne and Guerente, 1989). Clearly, the apparently redundant nature of the sequences involved in 3′ end formation of yeast obscures their precise definition.

Discussion

The results presented in these studies demonstrate that the mechanism of mRNA 3′ end formation in *S.pombe* closely resembles that of *S.cerevisiae*. First, sequences encoded within the *S.cerevisiae CYC1* fragment, which bear no apparent resemblance to those of higher eukaryotes, are capable of mRNA 3′ end formation in *S.pombe*. Secondly, the site of cleavage within this fragment in both *S.cerevisiae*

NAME	SEQUENCES UPSTREAM OF POLY (A) SITE	AT%	REFERENCE
	AGATCTTGTTTGAACTATAAATACTATAAGTAAATTTTGAGTA$_n$	80	
	CGTTGTTTGAACTATAATCTATAAATAAATTTTATGATTTTGATA$_n$	82	
	ATAGTTATGAGTATGACAGTTGTAAATAAATTTTGTTCTTGTGATA$_n$	76	
	AATTGGCGTGTTATGTAGTTAATCAATGACGCGTTGACATAAACA$_n$	65	
	ATTCTTAGTGCTTTTAAATACAAAAAGTGTGTAGGCTTGGTAATTA$_n$	72	
	GTACATATAAAGGCTTAGAGTAGCTAATAAAAAGTCTACATTCTA$_n$	71	
	TTTATAGAATGCAACGAAATTTTTTATCCTTTTTGCGTTTTGGATTA$_n$	74	
	ATGTGAGCAAATATGATCGATAATATATATATTTTGTATATATGA$_n$	80	
	TGTTCTCGAAAAATTTAATTAATAACTATTTTATTTTGTCGCTAGA$_n$	78	
L40c	ATTACGCTTTCATAAAAATAAATCGCCGTGTTGGTTTTATTTTACCA$_n$	70	Beltrame et al. 1987
ran1	CGGGATTTGAATTTTAACATTGTGTAAATAGCTAACATTTTATGCA	72	McLeod et al. 1986
tfIID	TTTTTTCATGATCGTGTAATAATAATAATAATCTTTTAAAAATTGA$_n$	85	Hoffmann et al. 1990
ura4	CTTTGGTAAAATTTTATGTAGCGACTAAAATATTAACTATTATAGA	78	within
ypt1	CTTTTAATATATTTTTATATACATACGTTTTACGTACTTCGCTGTCA$_n$	74	Mikaye et al. 1990
ypt3	ATTTGTAGGGTTACTTTAATAGAGAATTTGTCCAGATATTTTCAATA$_n$	74	Mikaye et al. 1990
ADH2*	GATGACAGTGTTCGAAGTTTCACGAATAAAAGATAATATTCTACTT	70	Russell et al. 1983
CYC1*	TTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCA	85	Butler and Platt, 1988
GAL7*	CGGAGTGACAATATATATATATATATATAATAATGACATCATTATCA	79	Abe et al. 1990
	SEQUENCES DOWNSTREAM OF POLY(A) SITE		
ura4	AACACCTTGGGAATAAAAAGTAATTTGCTATAGTAATTTATTAAACA	77	within
tfIID	ATAATTTGGTTTTAGTCTGAAGAGCGTATATCGACAGTTCTTAAAAA	71	Hoffmann et al. 1990
ypt1	ATTCCTTTTGTGAGTTTCATTCGCTGACGGTGAACTGCTAATATATA	63	Mikaye et al. 1990
ypt3	ATTCATACAGCTCATGTTTAATTTCCATGATATATCCTTTATAAGAG	72	Mikaye et al. 1990

Fig. 7. Sequence analysis of *S.pombe* and *S.cerevisiae* mRNA 3′ ends. The polyadenylation sites of sequenced cDNAs containing a minimum of 11 adenosine residues (A_n) at the 3′ end are aligned together with known *S.pombe* cDNA 3′ ends, *S.pombe* mRNA 3′ ends accurately mapped by S1 analysis (A), and DNA sequences encoding well studied *S.cerevisiae* mRNA 3′ ends (*). The polyadenylation sites lie at the 3′ end of the sequence where the upstream signals are compared, and at the 5′ end where the downstream signals are compared. Homologies or partial homologies to the polyadenylation consensus signals in higher eukaryotes and *S.cerevisiae* are indicated in bold, or underlined and bracketed respectively. The 38 bp deleted from *CYC1* in *cyc1Δ* is underlined in bold. The AT richness of each of these sequences is indicated.

and *S.pombe* are identical or very close, implying sequence recognition by an equivalent factor(s) in *S.pombe*. Thirdly the deletion mutant, *cyc1-512* fragment, which is very inefficient in *S.cerevisiae in vivo* (Zaret and Sherman, 1982) and *in vitro* (Butler and Platt, 1988) is also very inefficient in *S.pombe*, implying a common sequence requirement. Fourthly, unlike higher eukaryotic polyadenylation signals (Butler *et al.*, 1990), *S.pombe* pre-mRNAs can be accurately and efficiently cleaved in an *S.cerevisiae* whole cell extract, demonstrating *S.pombe* mRNA sequence recognition by soluble cleavage and polyadenylation factors from *S.cerevisiae*. Furthermore this strongly suggests that mRNA 3′ ends in *S.pombe* are formed by cleavage and polyadenylation as opposed to transcriptional termination. Finally, like *S.cerevisiae*, sequences surrounding the polyadenylation sites have no apparent consensus signal but are extremely AT rich. From these results it would appear that the sequences and factors required for mRNA 3′ end formation in fission and budding yeast are functionally homologous.

These results are surprising, since in other aspects of gene expression *S.pombe* appears to have features in common with

higher eukaryotes. For example, the SV40 early promoter (Jones *et al.*, 1988) and intron (Käufer *et al.*, 1985) are accurately and efficiently recognized by *S.pombe*. The SV40 early gene polyadenylation site, however, is clearly not recognized. It is therefore even more intriguing that sequences which are functional as mammalian polyadenylation signals are present but apparently inactive at the 3′ ends of *S.pombe* genes.

From phylogenetic studies of ribosomal RNA, fission and budding yeast are evolutionarily divergent (1200 Myr) (Huysmans *et al.*, 1983). Since mRNA 3′ end formation in fission and budding yeast would appear to be functionally homologous, this lower eukaryotic mechanism of mRNA 3′ end formation may be generally conserved. It is interesting to note that analysis of 3′ untranslated sequences of plant genes also revealed the presence of an AATAAA-like sequence (Joshi, 1987). However, the majority of these sequences are slightly degenerate, as is the case in *S.pombe*. Although recent work shows that this sequence is required (Ingelbrecht *et al.*, 1989; Sanfacon *et al.*, 1991) plant cells do not properly recognize higher eukaryotic polyadenylation signals (Hunt *et al.*, 1987). Therefore, plants may possess features of both the lower and higher eukaryotic mRNA 3′ end formation systems.

We have so far only imprecisely defined the signals required for *S.pombe* mRNA 3′ end formation. The termination signal proposed by Zaret and Sherman (1984), TAG...TAGT...TTT, is not obviously present in 3′ end sequences of the *S.pombe* genes, although derivations of this sequence and others can be identified. This apparent absence of sequence motif clearly resembles the situation in *S.cerevisiae* (Osborne and Guerente, 1989; Butler *et al.*, 1990). It seems possible that both *S.pombe* and *S.cerevisiae* will not have precise RNA 3′ end formation/polyadenylation signals. When the sequences both upstream and downstream of the *S.pombe tfIID*$^+$, *ura4*$^+$ and *yptI*$^+$ polyadenylation sites, which are presumably cleaved by a common RNA endonuclease in *S.cerevisiae* extract, are compared, no obvious sequence homology is found (see Figure 7). This is also the case when different *S.cerevisiae* RNA templates cleaved *in vitro* are compared (Butler *et al.*, 1990). Furthermore, extensive mutagenesis through the 3′ end sequence of the *S.cerevisiae CYC1* gene has failed to identify any one essential sequence element; rather, the required sequence appears to be rather diffuse and redundant (Osborne and Guerente, 1989; Russo and Sherman, 1989). Even so, we are currently attempting to define better the 3′ end formation signals of *S.pombe ura4*$^+$ mRNA. The functional interchangeability of *S.cerevisiae* and *S.pombe* sequence elements for 3′ end formation suggests that a further comparative analysis together with yeast genetic approaches (Ruohola *et al.*, 1988) will facilitate their identification.

The practical implications of these results are that 3′ processing signals need not be a major consideration with respect to gene expression when wishing to complement *S.cerevisiae* mutants functionally with genes from *S.pombe* and vice versa. However, it is possible that expression of higher eukaryotic genes in yeast may be complicated by inappropriate mRNA 3′ end formation.

Materials and methods

Human α-2 globin S pombe 3′ end construction

The *Eco*RV−*Sca*I fragment of *ura4* (226 bp) or the *Ssp*I−*Hin*dIII fragment of *cdc10*$^+$ (Aves *et al.*, 1985) (155 bp) was filled in with DNA polymerase (Klenow) and cloned into the *Pvu*II site in both orientations of both the wild type and thalassaemic human α-2 globin gene in pSVed (Whitelaw and Proudfoot, 1986).

Construction of PUS plasmids

The *Hin*dIII−*Eco*RV fragment of the *S.pombe ura4* gene was filled in and inserted into the *Sma*I site of Pirt 2 (2μ, *LEU2*) (EMBO Practical Course Manual, 1988) in the orientation indicated by the arrows representing *ura4* transcription (Figure 3A). The fragments indicated were all blunt-ended and cloned into a blunt-ended *Bam*HI site 5′−3′ with respect to the *ura4* gene. Fragments were taken from pON140:pGEM4 (Promega Biotec) vector containing a 1.76 kb *ura4* fragment as constructed by Olaf Nielsen. PUSα-globin contains the *Bst*EII fragment from *a*2W3′PSpSVED or the thalassaemic point mutant (Whitelaw and Proudfoot, 1986). Construct PUS L3 contains the *Eco*RI fragment from SP6 L3 or the non-functional point mutant (Humphrey *et al.*, 1987). PUS CYC1 and CYC1Δ contain the *Hin*dIII−*Eco*RV fragments of pGYC1 and pG200 R respectively (Butler and Platt, 1988).

Transfection, transformation and RNA preparation

Subconfluent HeLa cells were transfected with DNA constructs and the transiently expressed cytoplasmic RNA transcripts were purified as previously described (Levitt *et al.*, 1989). *S.pombe* (strain: *leu1-32 ade6-704 ura4*$^-$ *D18 h*$^+$) was transformed with the PUS constructs using the LiCl method (Bröker, 1987) and *leu*$^+$ prototrophic transformants were selected for with minimal medium containing 75 mg/l adenine and uracil. Total RNA from *S.pombe* was isolated as described in the EMBO Practical Course Manual (1988).

RNA analysis

The S1 probes were generated by end-labelling the homologous *Bst*EII fragments from α-2 pSVED, the *Eco*RI fragment of the rabbit β-globin gene and the *Sty*I fragment of the *S.pombe ura4* gene. The method of S1 analysis has been previously described (Levitt *et al.*, 1989). RNase protection analysis was performed as described by Melton *et al.* (1984). The riboprobes used in this paper were generated as follows: SP6 *ura4* was constructed by inserting the blunt-ended *Sty*I fragments of the *S.pombe ura4*$^+$ gene containing the polyadenylation site into the blunt-ended *Eco*RI site of pGEM-7Zf(+) (Promega). This construct was linearized with *Xba*I or *Pvu*II. The construct used to generate the SV40 early riboprobe was generated by cloning the *Bam*HI−*Hin*dIII fragment from SV40 into the equivalent sites in pSP64, to make pSP64SVBH. This construct was linearized with *Hin*cII. The L3 riboprobe was generated by cloning the *Eco*RI−*Eco*RI fragment of L3 (Humphrey *et al.*, 1987) into the *Eco*RI site of pGEM-7Zf(+) plasmid 5′−3′ (Promega). This resulting construct was then linearized with *Xba*I. The human α-2 globin riboprobe was constructed by cloning a filled in *Bst*EII−*Pvu*II fragment from α2W3′PSpSVED (Whitelaw and Proudfoot, 1986) into the *Sma*I site of pSP64 plasmid. This resulting construct was then linearized with *Xba*I. The *S.pombe tfIId Hin*dIII−*Dde*I fragment containing the polyadenylation signal was blunt-ended and cloned into a filled in *Eco*RI site of pGEM-7Zf(+). This construct was then linearized with *Bam*HI.

Pre-mRNA was generated from the above constructs by SP6 or T7 transcription using the cap-priming protocol (Contreras *et al.*, 1982; Melton *et al.*, 1984) in the presence of [^{32}P]GTP. *CYC1* and *cyc1*Δ riboprobes were generated by T7 transcription of *Hin*dIII linearized pGYC1 (Butler and Platt, 1988) or SP6 transcription of *Eco*RI linearized pG200R (Butler *et al.*, 1990) in the presence of [^{32}P]UTP.

In vitro RNA processing

Conditions for *in vitro* cleavage and polyadenylation are as previously described (Butler *et al.*, 1990). The *S.cerevisiae* whole cell extract was prepared as described previously (Butler *et al.*, 1990).

Sequencing of cDNAs

The *S.pombe* cDNA library was kindly supplied by Dallan Young. Sequencing was performed by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977).

Acknowledgements

We would like thank Dallan Young for the use of his cDNA library, Steve Aves, Martin Willer and Joanne Brindley for their contributions to this project, and also Shona Murphy and David Frendewey for their help and advice. We would like to thank Dr S.Moreno for the pSMCAT construct and Alex Hoffmann for the *S.pombe tfIID* gene. This project was supported in part by a Wellcome project grant to N.J.P., a NATO collaboration grant

to N.J.P. and T.P., and US PHS Grant no. 5-R01-35658 to T.P. T.H. is supported by an MRC studentship.

References

Abe,A., Hiraoka,Y. and Fukasawa,T. (1990) *EMBO J.*, **9**, 3691–3697.
Aves,S.J., Durkacz,B.W., Carr,A. and Nurse,P. (1985) *EMBO J.*, **4**, 457–463.
Barker,D.G., White,J.H.M. and Johnston,L.H. (1987) *Eur. J. Biochem.*, **162**, 659–667.
Beltrame,M. and Bianchi,M.E. (1987) *Nucleic Acids Res.*, **15**, 9089–9090.
Bröker,M. (1987) *Biotechniques*, **5**, 516–518.
Butler,J.S. and Platt,T. (1988) *Science*, **242**, 1270–1274.
Butler,J.S., Sadhale,P. and Platt,T. (1990) *Mol. Cell. Biol.*, **10**, 2599–2605.
Contreras,R., Cheroutre,H., Degrave,W. and Fiers,W. (1982) *Nucleic Acids Res.*, **10**, 6353–6362.
Eilliot,S., Chang,C., Schweingruber,M.E., Schaller,J., Rickli,E.E. and Carbon,J. (1986) *J. Biol. Chem.*, **261**, 2936–2941.
EMBO Practical Course Manual (1988) *Genetic Manipulation of Yeasts*. University of Rome, La Sapienza, Department of Cellular and Developmental Biology, September 1988.
Gil,A. and Proudfoot,N.J. (1984) *Nature*, **312**, 473–475.
Grimm,C., Kohli,J., Murray,J. and Maundrell,K. (1988) *Mol. Gen. Genet.*, **215**, 81–86.
Henikoff,S., Kelly,J.D. and Cohen,E.H. (1983) *Cell*, **33**, 607–614.
Higgs,D.R., Goodbourn,S.E.Y., Lamb,J., Clegg,J.B., Weatherall,D.J. and Proudfoot,N.J. (1983) *Nature*, **306**, 398–400.
Hindley,J., Phear,G., Stein,M. and Beach,D. (1987) *Mol. Cell. Biol.*, **7**, 504–511.
Hoffmann,A., Horikoshi,M., Wang,C.K., Schroeder,S., Weil,P.A. and Roeder,R.G. (1990) *Genes Dev.*, **4**, 1141–1148.
Humphrey,T., Christofori,G., Lucijanio,V. and Keller,W. (1987) *EMBO J.*, **6**, 4159–4168.
Hunt,A.G., Chu,N.M., Odell,J.T., Nagy,F. and Chua,N.H. (1987) *Plant Mol. Biol.*, **8**, 23–35.
Huysmans,E., Daws,E., Van den Berghe,A. and De Wachter,R. (1983) *Nucleic Acids Res.*, **11**, 2871–2880.
Hyman,L.E., Seiler,S.H., Whoriskey,J. and Moore,C.L. (1991) *Mol. Cell. Biol.*, **11**, 2004–2012.
Jones,R.H., Moreno,S., Nurse,P. and Jones,N.C. (1988) *Cell*, **53**, 659–667.
Joshi,C.P. (1987) *Nucleic Acids Res.*, **15**, 9297–9640.
Ingelbrecht,I.L.W., Herman,L.M.F., Dekeyser,R.A., Van Montagu,M.C. and Depicker,A.G. (1989) *Plant Cell*, **1**, 671–680.
Käufer,N.F., Simanis,V. and Nurse,P. (1985) *Nature*, **318**, 78–80.
Levitt,N., Briggs,D., Gil,A. and Proudfoot,N.J. (1989) *Genes Dev.*, **3**, 1019–1025.
McDevitt,M.A., Imperiale,M., Ali,H. and Nevins,J.R. (1984) *Cell*, **37**, 993–999.
McLeod,M. and Beach,D. (1986) *EMBO J.*, **5**, 3665–3671.
McLeod,M., Stein,M. and Beach,D. (1987) *EMBO J.*, **6**, 729–736.
Melton,D.A., Krieg,P.A., Rebagliahi,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
Mertins,P. and Gallwitz,D. (1987) *Nucleic Acids Res.*, **18**, 7369–7379.
Miyake,S. and Yamomoto,M. (1990) *EMBO J.*, **9**, 1417–1422.
Osborne,B.I. and Guerente,L. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 4097–4101.
Proudfoot,N.J., Rutherford,T.R. and Partington,G.A. (1984) *EMBO J.*, **3**, 1533–1540.
Ruohola,H., Baker,S.M., Parker,R. and Platt,T. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5041–5045.
Russell,P.R. (1983) *Nature*, **301**, 167–169.
Russo,P. and Sherman,F. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 8348–8352.
Sanfacon,H., Brodmann,P. and Hohn,T. (1991) *Genes Dev.*, **5**, 141–149.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Whitelaw,E. and Proudfoot,N.J. (1986) *EMBO J.*, **5**, 2915–2922.
Wickens,M. (1990) *Trends Biochem. Sci.*, **15**, 277–281.
Zaret,K.S. and Sherman,F. (1982) *Cell*, **28**, 563–573.
Zaret,K.S. and Sherman,F. (1984) *J. Mol. Biol.*, **176**, 107–135.

Received on June 3, 1991; revised on July 22, 1991