The EMBO Journal Vol. 21 No. 15 pp. 4125–4135, 2002

Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination

Bernhard Dichtl[1], Diana Blank, Martin Sadowski, Wolfgang Hübner, Stefan Weiser and Walter Keller[1]

Department of Cell Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

[1]Corresponding authors
e-mail: bernhard.dichtl@unibas.ch or walter.keller@unibas.ch

RNA polymerase II (pol II) transcription termination requires co-transcriptional recognition of a functional polyadenylation signal, but the molecular mechanisms that transduce this signal to pol II remain unclear. We show that Yhh1p/Cft1p, the yeast homologue of the mammalian AAUAAA interacting protein CPSF 160, is an RNA-binding protein and provide evidence that it participates in poly(A) site recognition. Interestingly, RNA binding is mediated by a central domain composed of predicted β-propeller-forming repeats, which occurs in proteins of diverse cellular functions. We also found that Yhh1p/Cft1p bound specifically to the phosphorylated C-terminal domain (CTD) of pol II *in vitro* and in a two-hybrid test *in vivo*. Furthermore, transcriptional run-on analysis demonstrated that *yhh1* mutants were defective in transcription termination, suggesting that Yhh1p/Cft1p functions in the coupling of transcription and 3′-end formation. We propose that direct interactions of Yhh1p/Cft1p with both the RNA transcript and the CTD are required to communicate poly(A) site recognition to elongating pol II to initiate transcription termination.

Keywords: β-propeller repeats/pre-mRNA 3′-end formation/RNA polymerase II C-terminal domain/RNA–protein interactions/transcription termination

Introduction

Most eukaryotic mRNA precursors (pre-mRNAs) are matured by site-specific cleavage in the 3′-untranslated region followed by polyadenylation of the upstream cleavage product. A complex protein machinery catalyses these seemingly simple reactions (reviewed in Zhao *et al*., 1999). Initially, biochemical fractionation of yeast extracts defined cleavage factors IA and IB (CF IA and CF IB), cleavage factor II (CF II), poly(A) polymerase and polyadenylation factor I (PF I) as sufficient to catalyse cleavage and polyadenylation *in vitro* (Chen and Moore, 1992). More recent affinity purification schemes indicate, however, that CF II and PF I subunits and poly(A) polymerase are stably associated *in vivo* and form the cleavage and polyadenylation factor (CPF) (Ohnacker *et al*., 2000; Gavin *et al*., 2002).

Many mammalian pre-mRNA 3′-end processing factors have recognizable yeast homologues (reviewed in Shatkin and Manley, 2000). Cleavage and polyadenylation specificity factor (CPSF) is involved in the recognition of the highly conserved AAUAAA hexamer sequence, which occurs upstream of most higher eukaryotic polyadenylation sites (Keller *et al*., 1991). A major role in this process has been attributed to the CPSF 160 protein that might, however, require additional CPSF subunits to achieve specificity (Murthy and Manley, 1995). All yeast homologues of CPSF subunits are associated with CPF, suggesting that the two factors are at least partially functionally equivalent. Recently, we showed that CPF plays an important role in the recognition of the poorly conserved yeast poly(A) site sequences (Dichtl and Keller, 2001). The yeast homologue of CPSF 160, Yhh1p/Cft1p (which will be referred to as Yhh1p in the remainder of this paper), was implied in cleavage and polyadenylation of pre-mRNAs based on the loss of both activities in extracts that were depleted with a specific antiserum (Stumpf and Domdey, 1996).

Increasing evidence suggests that polymerase II (pol II) transcription and pre-mRNA processing are tightly coupled (for recent reviews, see Hirose and Manley, 2000; Proudfoot *et al*., 2002). Splicing (Yuryev *et al*., 1996), 5′-capping (Cho *et al*., 1997; McCracken *et al*., 1997a) and 3′-end formation (Dantonel *et al*., 1997; McCracken *et al*., 1997b; Barilla *et al*., 2001) have been linked to transcription, and in all cases interactions of processing factors with the C-terminal domain (CTD) of the largest subunit of pol II have been identified. Current models suggest that upon assembly of a transcription-competent complex at a gene promotor, the CTD is hyperphosphorylated. A change in the surface charge of the CTD is thought to promote the association of RNA processing factors. Escape into elongation occurs concomitant to CTD phosphorylation and RNA processing factors are believed to act on nascent substrates as these emerge from the transcription machinery.

Pre-mRNA 3′-end formation and transcription also appear to be interdependent events. On the one hand, a direct involvement of pol II in 3′-end formation was suggested in the mammalian system (Hirose and Manley, 1998). On the other hand, 3′-end formation activities are required for correct transcription termination (Birse *et al*., 1998). A functional polyadenylation signal is necessary for termination (reviewed in Proudfoot, 1989) and the mechanism by which the RNA signal is transduced to RNA pol II is commonly summarized by two major mechanistic models (Proudfoot *et al*., 2002). The observation that mutants ablating 3′-end cleavage were defective in termination supported a cleavage-dependent mechanism (Birse *et al*., 1998). Substrate cleavage was suggested to produce an entry site for 5′–3′ exonucleases, which degrade the downstream cleavage product, 'catch-up' with the elongating polymerase and force it to

terminate. Alternatively, it was suggested that the dissociation of factors from pol II, due to association with the emerging RNA substrate, change the properties of pol II such that termination is initiated. The transcriptional coactivator PC4 (Sub1p in yeast) was proposed to act as an anti-terminator of pol II transcription through interaction with the 3′-end formation factor CstF64 in mammals and Rna15p in yeast (Calvo and Manley, 2001). A genetic screen for pol II termination factors in *Schizosaccharomyces pombe* suggested that links between termination and 3′-end formation might be integrated with cell cycle events (Aranda and Proudfoot, 2001). These authors showed that a C-terminal termination domain in Rna15p interacts with the cell-cycle specific transcription factor Res2p and that Res2p by itself is required for termination. The cleavage-independent model was supported further by electron microscopy studies of nascent transcripts in *Xenopus* oocytes that visualized correctly terminated but uncleaved RNAs (Osheim *et al*., 1999). A mammalian *in vitro* system that reproduced the requirement for a poly(A) signal in termination was established (Tran *et al*., 2001). This analysis suggested that extrusion of the poly(A) signal from pol II, but not processing of the RNA substrate, is required for termination. Last, but not least, sequence elements downstream of poly(A) sites were proposed to act as pause elements for pol II transcription that are important for termination (Birse *et al*., 1997).

Here we show that the CPF subunit Yhh1p is an RNA-binding protein, which is involved in poly(A) site recognition. Interestingly, the RNA-binding domain of Yhh1p is composed of predicted β-propeller-forming repeats, which also occur in proteins involved in pre-mRNA splicing and UV-damage recognition. Furthermore, Yhh1p specifically interacted with the phosphorylated CTD of pol II and *yhh1* mutant cells were defective in pol II termination. We propose that Yhh1p has an important function in the coupling of 3′-end formation and transcription termination through direct interactions with the RNA substrate and the CTD.

Results

YHH1 is required for cleavage and polyadenylation of pre-mRNA in vitro

To investigate the function of Yhh1p in pre-mRNA 3′-end formation, we isolated temperature-sensitive (ts) alleles of *YHH1* (see Materials and methods). Sequence analysis revealed seven mutations within *yhh1-3* and six mutations within *yhh1-6* that underlie the ts phenotype (Figure 1A). As shown in Figure 1B, *yhh1-3*, *yhh1-6* and *yhh1-12* strains ceased growth ~200 min after a shift to 37°C and the mutant strains did not form colonies at this temperature (Figure 1B). Extracts from wild-type and mutant strains were tested for *in vitro* 3′-end processing. All extracts were active in cleavage of a CYC1 pre-mRNA at 30°C (Figure 1C, upper panel, lanes 2–5). However, *yhh1-3* extracts had less cleavage compared with wild type at 30°C and this activity was lost at 36°C (lane 8); *yhh1-6* and *yhh1-12* extracts retained activity at 36°C (lanes 9 and 10). Specific polyadenylation of pre-cleaved CYC1 substrate (CYC1-Pre) was reduced in mutant extracts at 30°C (Figure 1C, lower panel, lanes 2–5) and more strongly affected at 37°C (lanes 7–10). *yhh1-3* extracts showed the tightest phenotype under these conditions (lane 8). Identical results were obtained when the extracts were tested for cleavage of GAL7 pre-mRNA and polyadenylation of a pre-cleaved GAL7 RNA (results not shown). These results suggested a requirement for Yhh1p in pre-mRNA 3′-end formation.

YHH1 is required for ACT1 poly(A) site recognition

To correlate the observed *in vitro* phenotypes with the lethality of *yhh1* cells *in vivo*, we performed northern blot analysis on total RNA extracted from strains grown at 23°C and after shift to 37°C. Probes specific for ADH1, CYH2 and CYC1 revealed mRNA levels in *yhh1* mutants grown at 37°C for up to 3 h, which were comparable to those in wild-type cells (Figure 2A, lanes 1–10). Some reduction in mRNA levels was observed in the mutants only after 4 h at 37°C. *rna15–1* cells, in contrast, showed a strong reduction of these mRNAs after 1 h at 37°C (lanes 11 and 12). Levels of the very stable PGK1 mRNA ($t_{\frac{1}{2}}$ = 45 min) were similar in all strains tested. STE3 mRNA, which is very unstable ($t_{\frac{1}{2}}$ = 4.2 min), was strongly reduced in both wild-type and mutant *YHH1* cells grown at 37°C; this RNA could not be detected in *rna15–1* cells due to their mating type. Levels of 18S rRNA and U14 snoRNA served as loading controls and were comparable in all lanes. Interestingly, we observed that ACT1 mRNA was strongly reduced in *yhh1-6* cells at early time points and also showed a smear of higher molecular weight species (lanes 7–10). The latter effect was more pronounced in *yhh1-3* cells (lanes 3–6); *rna15–1* cells showed a complete loss of ACT1 mRNA at 37°C (lane 12).

To resolve ACT1 mRNAs better, we treated total RNAs with RNase H targeted by an oligonucleotide complementary to the 3′ region of the message (ACT1-RNaseH). RNAs were analysed by polyacrylamide northern blot with a probe specific for the 3′-end of ACT1 mRNA. As shown in Figure 2B, at least four distinct ACT1 mRNA species (ACT1-I to -IV; lanes 1 and 2) can be distinguished, which originate from the use of alternative poly(A) sites (Mandart and Parker, 1995). Wild-type cells mainly use the most proximal poly(A) site (ACT1-I) and, to a lesser extent, upstream sites (ACT1-II to -IV). The same pattern of poly(A) site selection was observed in *yhh1* mutant cells at 23°C; however, after shift to 37°C, the mutant cells increasingly used minor upstream sites (sites III and IV) at the expense of the major downstream site (site I; lanes 3–8).

Analysis of the global poly(A) tail length distribution revealed that *yhh1* mutant strains did not show loss of poly(A) tails when grown at 37°C for up to 4 h (Figure 2C, lanes 3–10; data not shown for *yhh1-12*); in contrast, an *rna15–1* strain showed a marked loss of poly(A) tails after shift to 37°C for 1 h (lanes 11 and 12) (Minvielle-Sebastia *et al*., 1994). Since *yhh1* mutant strains ceased growth ~200 min after shift to 37°C, this indicated that lethality of *yhh1* mutants was not due to a general under-accumulation of mRNA. Instead, we propose that the formation of a subset of mRNAs with weak poly(A) sites requires Yhh1p for poly(A) site recognition (see Discussion). In agreement with the observed stability of mRNAs in *yhh1* mutant cells, we did not see a reduction of Yhh1p or other tested 3′-end formation factors (Rna15p, Rna14p, Pcf11p,

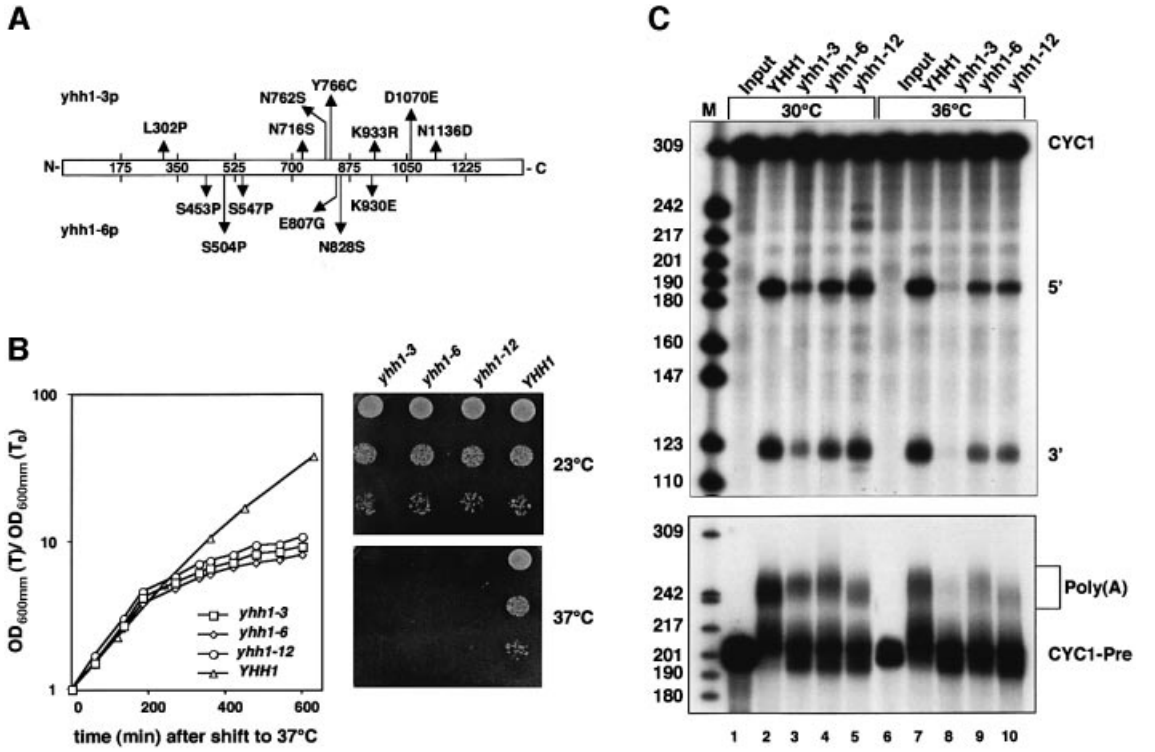

Fig. 1. Analysis of pre-mRNA 3′-end processing activity of *yhh1* mutant extracts *in vitro*. (**A**) Sequence analysis of *yhh1* mutant alleles revealed seven amino acid changes in yhh1-3p (top of the panel) and six amino acid changes in yhh1-6p (bottom of panel), which underlie the respective ts phenotypes. (**B**) Growth curves of wild-type and mutant *YHH1* strains after shift to 37°C. Ten-fold serial dilutions of cultures spotted on YPD plates followed by incubation at the indicated temperatures for 2 days. (**C**) *In vitro* cleavage (upper panel) and polyadenylation (lower panel) assays with extracts prepared from yeast strains as indicated. Input lanes represent mock-treated reactions. Length of marker bands is indicated on the left. For 3′-end cleavage, internally ^{32}P-labelled CYC1 RNA was used. The positions of full-length RNA (CYC1), 5′ and 3′ cleavage products are indicated. Cleavage was performed either at 30°C (lanes 1–5) or at 36°C (lanes 6–10). Specific polyadenylation was performed with internally ^{32}P-labelled CYC1-precleaved RNA that ends at the natural cleavage site. The positions of substrate (CYC1-Pre) and polyadenylation products [Poly(A)] are indicated. Polyadenylation assays were performed either at 30°C (lanes 1–5) or at 37°C, following a 5 min pre-incubation of extract and reaction mixture at this temperature (lanes 6–10).

Ysh1p, Fip1p and Pap1p) by western analysis of extracts prepared from strains following growth at 37°C for up to 4 h (results not shown).

Yhh1p is an RNA-binding protein

Aberrant selection of ACT1 poly(A) sites suggested that Yhh1p might interact directly with the pre-mRNA. To test for RNA-binding activity, we performed glutathione *S*-transferase (GST) pull-down experiments with Yhh1p that was expressed in *Escherichia coli* with a GST tag fused to the N-terminus and a His_6 tag at the C-terminus (GST-Yhh1p-H_6; Figure 3A, lane 1). The identity of the purified 183 kDa protein was confirmed by western blot with an anti-Yhh1p antibody (Figure 3A, lane 2) (Stumpf and Domdey, 1996). Figure 3B shows that CYC1 RNA was bound efficiently (lane 4). CYC1-Pre RNA, which lacked sequences downstream of the poly(A) site and CYC1-512 RNA, which has a 38 nucleotide deletion encompassing both the efficiency element and positioning element, were bound efficiently as well (lanes 7 and 10). Wild-type GAL7-1 and GAL7-3 RNA, which has a deletion in the efficiency element, were bound more weakly compared with CYC1 substrates (lanes 13 and 16). The observed interactions were specific for the substrate RNA since reactions contained excess unlabelled *E.coli* tRNA. No binding could be observed when GST or GST-Ysh1p-H_6 was used (lanes 3, 6, 9, 12 and 15; results not shown).

To evaluate the strength of the Yhh1p–RNA interactions, filter binding experiments were performed. Protein was incubated with labelled CYC1 or GAL7 RNA, and RNA–protein complexes retained on nitrocellulose filters were measured by scintillation counting (see Materials and methods). These analyses revealed apparent K_D values for the interaction of Yhh1p with CYC1 and GAL7 RNAs of 5 and 15 nM, respectively (data not shown).

To characterize further the RNA-binding properties of Yhh1p, we performed RNase H protection experiments on a short CYC1 substrate (sCYC1; Figure 3C) (Dichtl and Keller, 2001). Body-labelled RNA was pre-incubated in the absence or presence of protein and subsequently supplemented with antisense DNA oligonucleotides and RNase H. Oligos 2–8 formed RNA–DNA hybrids efficiently, which targeted RNase H (Figure 3D, lanes 5–11) and formation of distinct cleavage products showed that RNase H cleavage was specifically directed by antisense oligonucleotides. Oligo 1 was not functional in this assay (lane 4). GST-Yhh1p-H_6 prevented DNA–RNA hybrid formation of oligos 6, 7 and 8 (lanes 18–20). Cleavage directed by oligo 2 was reduced in multiple experiments (lane 14). No protection could be observed when oligos 3,

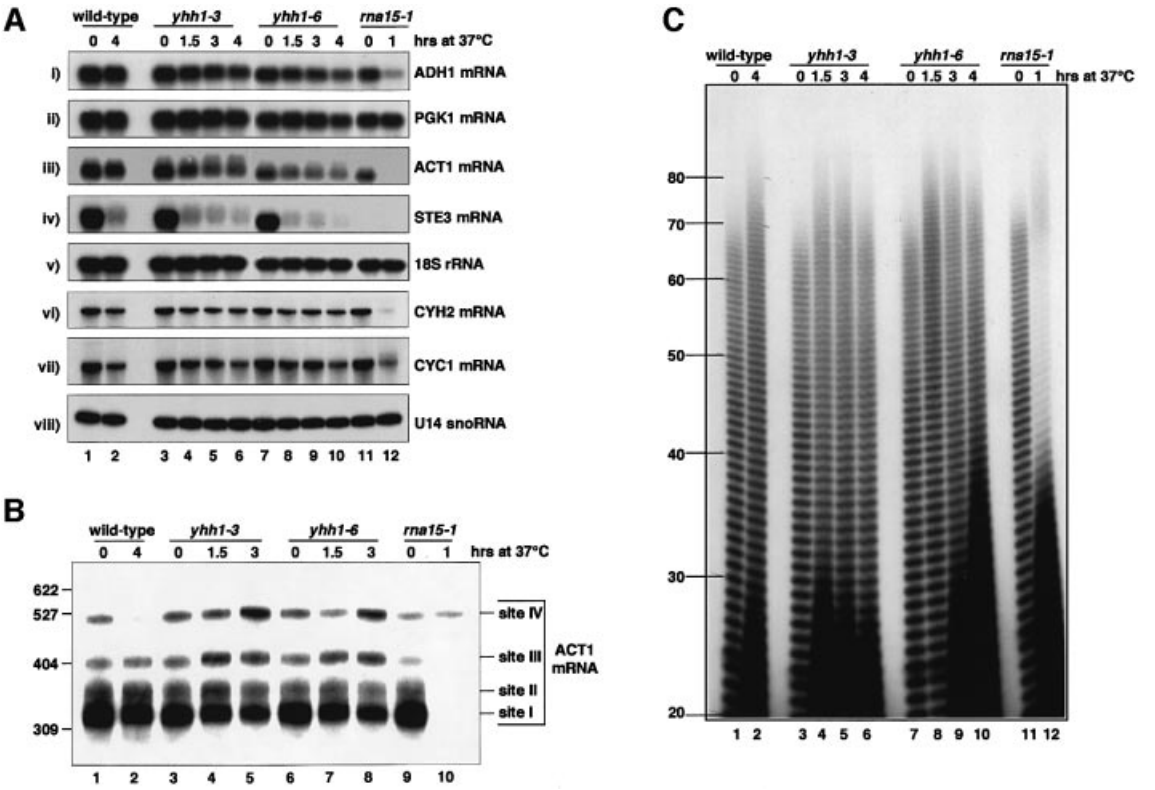

Fig. 2. *yhh1* mutant cells are defective in recognition of the ACT1 poly(A) site. (**A**) Northern analysis of total RNA extracted from wild-type and mutant *YHH1* cells grown at 23°C, or following a shift to 37°C for 1.5, 3 and 4 h. For control, an *rna15-1* strain grown at 23°C and 37°C was analysed in parallel. RNAs were separated on formaldehyde/1.2% agarose gels (panels i–v) or 8.3 M urea/8% polyacrylamide gels (panels vi–viii). Filters were developed with random-primed labelled probes or end-labelled oligonucleotides directed against the RNA species indicated on the right of each panel. (**B**) Analysis of ACT1 poly(A) site usage in wild-type and *yhh1* and *rna15-1* mutant cells as indicated. Total RNAs extracted from strains grown as described in (A) were treated with RNase H and oligonucleotide ACT1-RNaseH, and analysed by northern blotting. A random-primed probe complementary to the ACT1 3′-end region (nucleotides 930–1536) was used. The positions of RNAs polyadenylated at sites I to IV are indicated, as well as molecular weight markers. (**C**) Poly(A) tail analysis with total RNA extracted from indicated *YHH1* strains after growth at 23°C (0 h) or after shift to 37°C for 1.5, 3 and 4 h. Also analysed was an *rna15-1* strain grown at 23°C or after shift to 37°C for 1 h. Poly(A) tail length is indicated on the left. Strong signals of short RNA species in lanes 2, 4, 5, 6, 9, 10 and 12 were not reproducible and probably result from incomplete digestion of non-poly(A) sequences.

4 and 5 were used (lanes 15–17). Thus, the protein bound with high affinity to sequences encompassing the poly(A) site of sCYC1. A weaker binding site resided in the efficiency element of sCYC1.

These results were corroborated by S7 nuclease protection experiments. Figure 3E shows that pre-incubation of the RNA substrate with increasing amounts of GST-Yhh1p-H_6 resulted in protection of sequences encompassing the sCYC1 poly(A) site (position 47–75) from S7 digestion. In contrast to RNase H protection experiments, this assay did not resolve binding of Yhh1p at efficiency element positions (nucleotides 19–30), possibly due to a higher stringency of the S7 protection assay.

In summary, our experiments demonstrated that Yhh1p is an RNA binding protein, which bound to CYC1 elements encompassing the poly(A) site. Since interactions observed with sequences downstream of the poly(A) site in RNase H and S7 protection experiments were not essential for binding of CYC1 RNA, as found by GST pull-down, we conclude that multiple RNA elements encompassing the CYC1 poly(A) site and possibly the poly(A) site itself are recognized. Therefore, no single sequence element was bound specifically by Yhh1p. Nevertheless, the previous observation that poly(A) site recognition by CPF (Dichtl and Keller, 2001) involves sequence elements which are also bound by Yhh1p is compatible with the idea that Yhh1p participates in poly(A) site recognition (see Discussion).

Yhh1p carries a central RNA-binding domain

The sequence of Yhh1p does not contain a known RNA-binding motif that was detectable with comparative protein analysis programs (see Materials and methods). To identify sequences required for RNA binding, N- and C-terminally truncated Yhh1p proteins were employed in GST pull-down assays. Figure 4A shows that truncation of up to 608 C-terminal residues did not affect binding of CYC1 RNA compared with full-length protein (lanes 3–6). In contrast, proteins missing >691 C-terminal residues were unable to bind (lanes 7–10). Deletion of 1016 N-terminal amino acids (aa) completely abolished RNA binding (lane 16); removal of up to 500 N-terminal residues had no effect (lanes 11–13), whereas proteins missing 583 or 663 aa, respectively, showed decreasing RNA-binding ability (lanes 14 and 15). These results indicated that a central domain within aa 500 and 750 of Yhh1p harboured the RNA-binding domain.

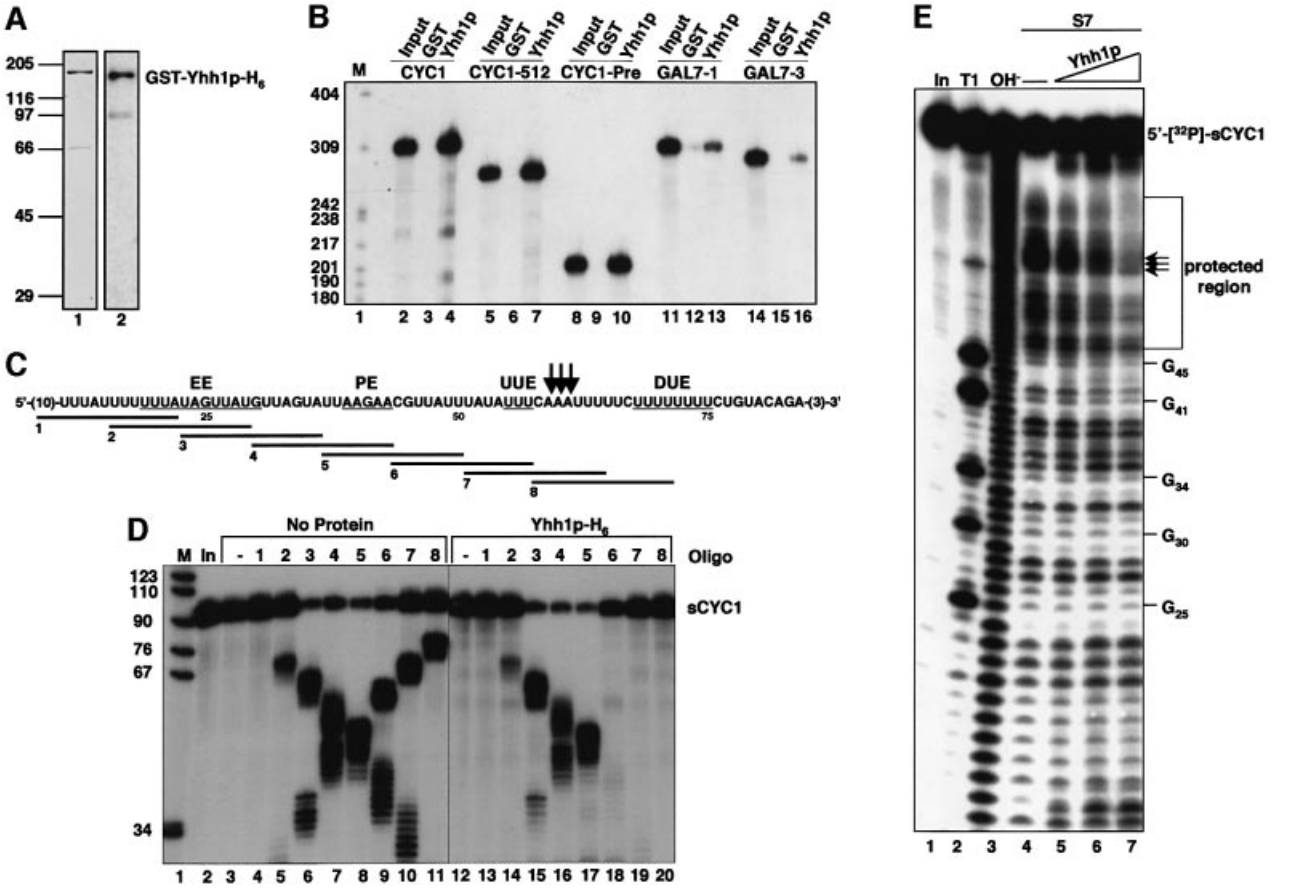

Fig. 3. Yhh1p is an RNA-binding protein. (**A**) Purified GST-Yhh1p-H_6 expressed in *E.coli* was subjected to SDS–PAGE and visualized by Coomassie Brilliant Blue staining (lane 1) or detected after western transfer with an anti-Yhh1p serum (lane 2). Positions of recombinant protein and marker bands are indicated. (**B**) GST pull-down experiments with 0.5 µg GST (lanes 3, 6, 9, 12 and 15) or GST-Yhh1p-H_6 (Yhh1p; lanes 4, 7, 10, 13 and 16) and *in vitro* transcribed, ^{32}P-labelled RNAs. Lanes 2, 5, 8, 11 and 14 show 10% of RNA included in binding reactions (Input). Length of marker bands (M; lane 1) is indicated on the left. (**C**) Sequence of sCYC1 RNA. Only CYC1-derived nucleotides are shown; vector encoded nucleotides are indicated by numbers in parentheses. Efficiency (EE) and positioning elements (PE), and upstream (UUE) and downstream (DUE) U-rich elements are underlined. The arrows mark poly(A)-site positions. Sequences complementary to 14-mer DNA oligonucleotides (1–8) are indicated. (**D**) RNase H protection analysis with Yhh1p and sCYC1 RNA. Internally ^{32}P-labelled sCYC1 RNA was incubated with binding buffer (No protein; lanes 2–11) or GST-Yhh1p-H_6. The GST moiety was removed from recombinant protein by cleavage with TEV protease included in the binding reaction. Protein containing assays are therefore indicated by Yhh1p-H_6 (lanes 12–20). After pre-incubation, the substrate was cleaved by addition of RNase H and oligonucleotides as indicated. For control, oligonucleotides were omitted in lanes 3 and 12. The RNA input (In) is shown in lane 2. *Hpa*II-digested pBR322 fragments were 5′-end labelled and served as marker bands (M; lane 1). (**E**) S7 nuclease protection analysis. 5′-[^{32}P]-labelled sCYC1 RNA was pre-incubated in the absence (lane 4) and presence of increasing amounts of GST-Yhh1p-His_6 (100, 200 and 400 ng; lanes 5–7), followed by the addition of S7 nuclease. Reaction products were separated on 12% polyacrylamide/8.3 M urea gels. Also analysed were the RNA input (In; lane 1), a partial T1 nuclease digest (T1; T1 cuts RNA after guanosines) and a partial alkaline hydrolysis of the input RNA (OH^-; lane 3). To the right of the panel, the migration of full-length RNA and the position of guanosines in the RNA substrate are indicated. Arrows mark the *in vitro* cleavage sites and the region protected by GST-Yhh1p-His_6 is indicated.

To test this, we produced fragments that either partially or completely spanned the proposed domain (Figure 4B). In GST pull-down experiments, we found that fragments containing aa 501–666, 501–749 and 584–749 showed efficient binding of CYC1 RNA, which was comparable to full-length protein (Figure 4C, lanes 3, 4, 5 and 7). Proteins containing aa 584–666 or 664–749 (lanes 6 and 8) showed strongly reduced but detectable binding activity compared with GST (lane 2). These results were corroborated by gel mobility shift assays with the sCYC1 RNA. As shown in Figure 4D, fragments containing residues 501–666 (lanes 2 and 3), 501–749 (lanes 4 and 5) and 584–749 (lanes 8 and 9) were most efficient in formation of a slower migrating sCYC1 species. The fragment containing aa 584–666 was less efficient (lanes 6 and 7) and the fragment containing aa 664–749 was not active in this assay (lanes 10 and 11). To test whether the RNA-binding properties of the isolated RNA-binding domain were the same as with the full-length protein, we performed RNase H protection experiments with the fragment containing aa 501–749 as described in Figure 3D. The observed protection pattern paralleled that found with full-length protein (data not shown), demonstrating that the central RNA-binding domain in Yhh1p was sufficient for interaction with sCYC1.

The Yhh1p RNA-binding domain is composed of predicted β-propeller repeats

As observed by Caspary *et al*. (1999), Yhh1p and its homologues (CPSF 160 proteins) display sequence similarity to proteins associated with the spliceosomal U2 snRNP (SAP130) (Das et al., 1999) and to *Xeroderma*

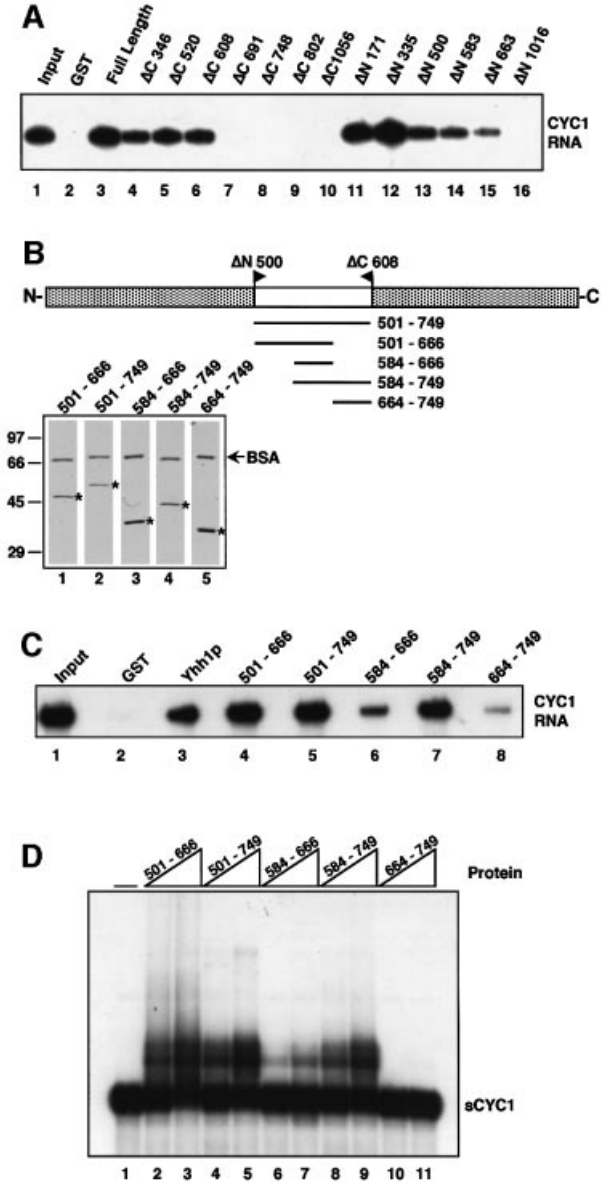

Fig. 4. Delineation of the Yhh1p RNA-binding domain. (**A**) GST pull-down with *in vitro* transcribed, ^{32}P-labelled CYC1 RNA and 0.5 μg GST (lane 2), GST-Yhh1p-H_6 (Yhh1p; lane 3), C-terminal (lanes 4–10) and N-terminal truncations of Yhh1p (lanes 11–16) as indicated on top of the panel. Lane 1 shows 10% of RNA in binding reactions (Input). (**B**) Representation of internal Yhh1p domains that were expressed as N-terminal GST fusions and with a C-terminal His_6 tag. For quality control, recombinant proteins (indicated by an asterisk) were resolved by SDS–PAGE and stained with Coomassie Brilliant Blue. Migration of BSA that was included in the final elution buffer is indicated. (**C**) ^{32}P-labelled CYC1 RNA was tested for interaction with 0.5 μg GST (lane 2), GST-Yhh1p-H_6 (Yhh1p; lane 3), and Yhh1p domains as indicated in (B) (lanes 4–8). Lane 1 shows 10% of RNA in binding reactions (Input). (**D**) Mobility shift experiments with Yhh1p domains as indicated in (B). ^{32}P-labelled sCYC1 RNA was incubated with increasing amounts (200 and 400 ng) of protein and analysed by native polyacrylamide gel electrophoresis. Protein was omitted in lane 1. The position of sCYC1 is indicated.

pigmentosum group E proteins (XPE) that are implicated in UV-damage recognition (Takao *et al.*, 1993). A computational analysis suggested that degenerate repeat sequences within these proteins form β-propeller structures (Neuwald and Poleksic, 2000). We searched for sequences displaying homology to the Yhh1p RNA-binding domain using the NCBI BLAST server, and identified CPSF 160, SAP130 and XPE proteins. A sequence element in Yhh1p that is sufficient for RNA binding (aa 503–684) showed the most significant similarities. Figure 5 shows that the active Yhh1p RNA-binding domains, including aa 501–666 and 584–749, aligned with predicted β-propeller repeats IV–V and V–VIII, respectively (Figure 5) (Neuwald and Poleksic, 2000). These observations suggest the testable hypothesis that members of these protein families contain a nucleic acid binding motif that is related to the RNA-binding domain of Yhh1p.

Yhh1p functions in coupling of 3′-end formation and transcription termination

We provided evidence that Yhh1p is involved in poly(A) site recognition. Since a functional poly(A) signal is required for correct termination of pol II transcription (reviewed in Proudfoot, 1989) we addressed the possible involvement of Yhh1p in termination by transcriptional run-on analysis (TRO) with the well studied CYC1 gene (Birse *et al.*, 1998). The distribution of run-on transcripts over single-stranded DNA probes (P1–P6; see Figure 6A) was analysed in wild-type and mutant *yhh1-3* and *yhh1-6* cells following growth at 25°C and after shift to 37°C for 1 h. Correct termination results in run-on signals over probes P1–P3 but not P4–P6 (Birse *et al.*, 1998). Slot hybridizations in Figure 6B showed that transcription terminated efficiently in wild-type cells that were grown at 25 or 37°C. In contrast, at both permissive and restrictive temperature, *yhh1* mutant cells displayed increased run-on signals, also over probes P4 and P6, reflecting an increased pol II density past the correct termination site. In *yhh1-3* and *yhh1-6* cells, read-through was increased ~10-fold and up to 20-fold, respectively (Figure 6C). Interestingly, run-on signals over probe 2 (which lies inside the coding region of CYC1) were increased in *yhh1-3* cells after shift to 37°C, indicating that these cells were also somewhat impaired in pol II elongation.

The observed termination defect in *yhh1* mutants prompted us to test whether the protein interacted with the CTD of pol II. Figure 7A shows a GST pull-down experiment in which *in vitro* translated Yhh1p bound efficiently to GST–CTD that was phosphorylated with HeLa nuclear extract (see Materials and methods; Figure 7A, lane 4). No interactions were observed with GST alone or non-phosphorylated GST–CTD (lanes 2 and 3). Ysh1p/Brr5p did not interact with either form of the CTD (lanes 3 and 4) and the CF IA subunit Pcf11p showed increased binding to phosphorylated compared with non-phosphorylated GST–CTD (lanes 3 and 4) (Barilla *et al.*, 2001; M.Sadowski, W.Hübner and W.Keller, manuscript in preparation). Since these experiments employed HeLa nuclear extract for CTD phosphorylation and rabbit reticulocyte lysate to obtain *in vitro* translated Yhh1p, the possibility remains that the observed interactions were mediated by a contaminating mammalian phospho-CTD binding protein. To test whether the interaction between Yhh1p and the CTD also occurred *in vivo*, we used a *GAL4*-based two-hybrid system and assayed for activation of *HIS3* transcription. A plasmid carrying the *GAL4* DNA binding domain fused to the CTD (pBD–CTD) and

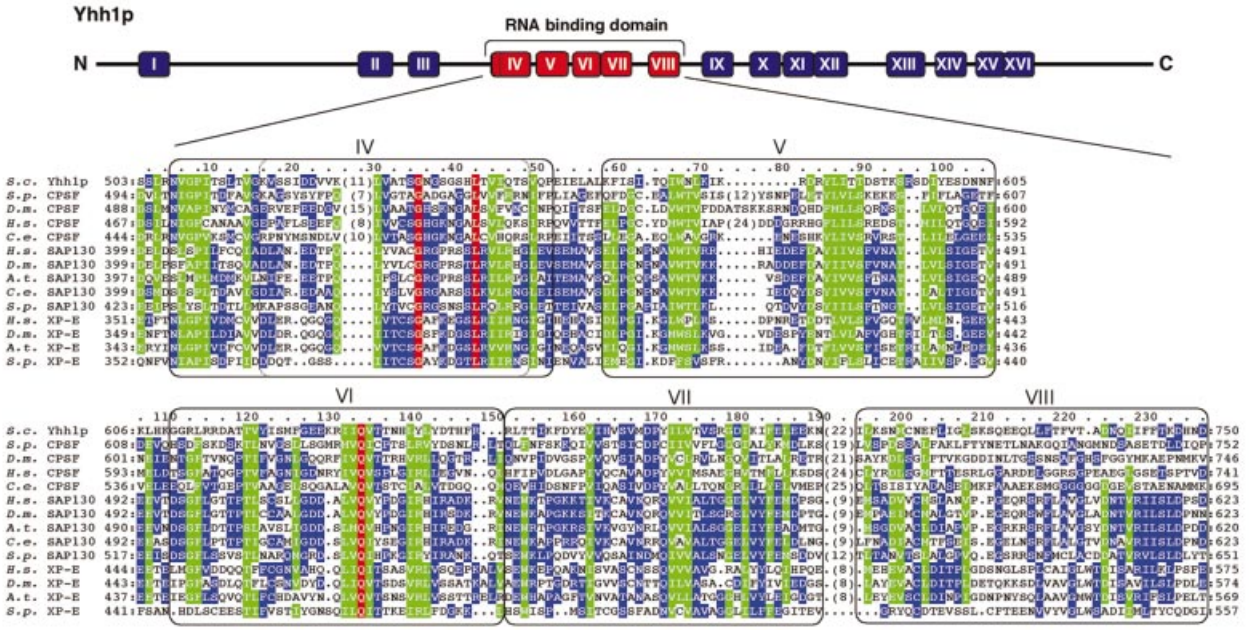

Fig. 5. Alignment of sequences related to the Yhh1p RNA-binding domain. Sequences similar to the RNA-binding domain of Yhh1p were aligned as described in Materials and methods. Invariant residues are shown in red; green indicates conserved and similar amino acid positions which are present in 70–100% of the aligned sequences, and conserved and similar residues occurring in 35–70% of the aligned sequences are shown in blue. Also indicated are sequence repeats within Yhh1p (boxes numbered I–XVI) that were predicted to form β-propeller-like structures (Neuwald and Poleksic, 2000). The sequence repeats that align with the active RNA-binding fragments of Yhh1p (aa 501–666, repeats IV–VII, and aa 584–749, repeats V–VIII) are shown in red. DDBJ/EMBL/Genbank accession numbers are: *Homo Sapiens* (*H.s.*) SAP130: CAB56791; *Drosophila melanogaster* (*D.m.*) SAP130: AAF47416; *Arabidopsis thaliana* (*A.t.*) SAP130: CAB75754; *Caenorhabditis elegans* (*C.e.*) SAP130: T32916; *S.pombe* (*S.p.*) SAP130: BAA86918; *H.s.* XP-E: CAA05770; *A.t.* XP-E: CAB81084; *D.m.* XP-E: AAF54901; *S.p.* XP-E: T37876; *D.m.* CPSF: AAF58240; *H.s.* CPSF: Q10570; *C.e.* CPSF: AAF36067; *S.p.* CPSF: T39636; *S.cerevisiae* (*S.c.*) Yhh1p: AAB64737 (YDR301w). Several more weakly related protein sequences exist in the database that were not included in this alignment.

plasmids containing the *GAL4* activation domain fused to test genes (plasmids pACT-YHH1, pACT-YSH1 and pACT-PCF11) were co-transformed into strain Y190 and growth was tested on synthetic medium lacking histidine. Figure 7B shows that *YHH1* and *PCF11*, but not *YSH1* or empty vector, promoted a His^+ growth phenotype under stringent conditions (35 mM 3′-aminotriazol). To address whether the termination defect in *yhh1* mutant cells resulted from impaired interaction with the CTD, we tested the *yhh1-6* gene for interaction with the CTD in the two-hybrid assay. We could not detect a reduced His^+ growth phenotype with the mutant allele, indicating that the termination defect did not result from a reduced interaction with the CTD (data not shown). Irrespective of this observation the interactions of Yhh1p with both the RNA substrate and the CTD in combination with the read-through defect of *yhh1* mutant cells support the idea that Yhh1p functions to communicate poly(A) site recognition to pol II.

Discussion

Yhh1p is an RNA-binding protein and several lines of evidence support the proposal that it contributes to poly(A) site recognition. Yeast poly(A) site sequences can be redundant and are highly degenerate (Graber *et al*., 1999; van Helden *et al*., 2000). Recognition of yeast poly(A) sites is likely to rely on the simultaneous recognition of several sequence elements that are bound by a number of different *trans*-acting factors (Valentini *et al*., 1999; Dichtl and Keller, 2001; Gross and Moore, 2001). Thus, it is not to be expected that a single protein would be sufficient for poly(A) site recognition or that RNA-binding proteins involved in this process would show specificity for a single RNA element. We demonstrated previously that interactions of CPF with U-rich elements encompassing the CYC1 poly(A) site are essential for cleavage *in vitro* and polyadenylation *in vivo* (Dichtl and Keller, 2001). U-rich elements occur immediately upstream and downstream of many yeast poly(A) sites (Graber *et al*., 1999) and previous analyses revealed that RNA-binding components of CPF preferentially interact with these sequences (Yth1p and Ydh1p/Cft2p; Barabino *et al*., 2000; Dichtl and Keller, 2001). Our findings reported here, i.e. that the same U-rich elements are also recognized by Yhh1p, is in agreement with the idea that the protein contributes to the RNA-binding specificity of CPF. We predict that CPF forms an RNA-binding surface that makes multiple contacts at the poly(A) site region. A future challenge will be to determine how the RNA-binding proteins of CPF are topologically arranged on an RNA substrate to achieve specific recognition of a poly(A) site.

The deficiency of *yhh1-3* mutant extracts in cleavage and polyadenylation at the non-permissive temperature *in vitro* suggested a direct requirement for Yhh1p in 3′-end formation. However, global poly(A)-tail length distribution and steady-state levels of most mRNAs tested were only weakly affected in mutant strains. Therefore, we propose that reduced cleavage and polyadenylation activities in *yhh1* mutant extracts *in vitro* do not reflect a

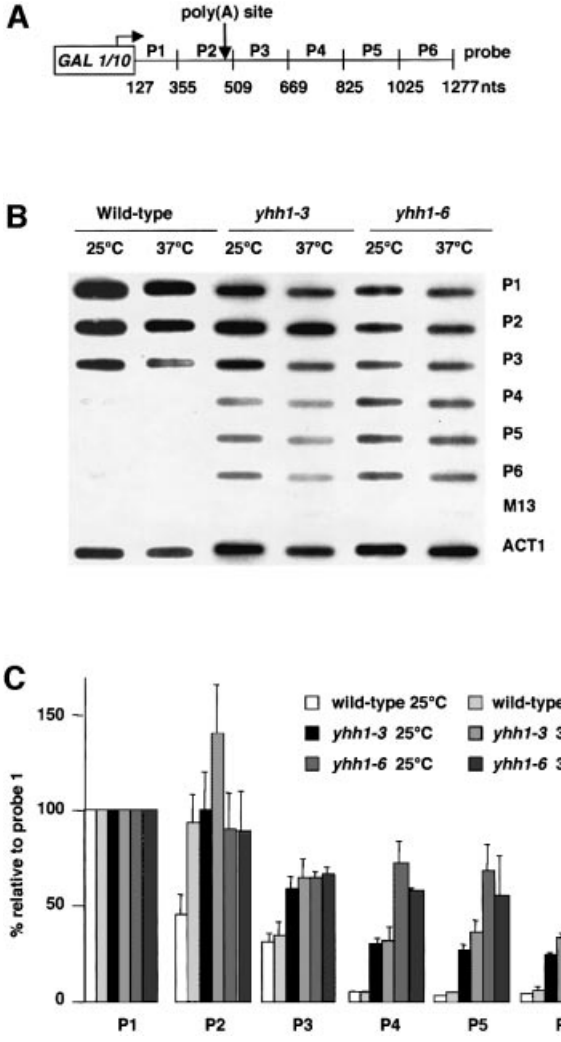

Fig. 6. Yhh1p is required for pol II transcription termination. (**A**) Representation of the pGAL-CYC1 gene that was analysed by transcriptional run-on (TRO). M13 probes (P1–P6) complementary to regions of the CYC1 gene (indicated in nucleotides relative to the transcription start site) and the poly(A) site are indicated. (**B**) Slot hybridizations obtained following TRO (Birse *et al*., 1998) of wild-type and mutant *yhh1* cells grown at 25°C or following shift to 37°C for 1 h. P1–P6 represent CYC1 probes as described in (A), M13 and ACT1 probes served as controls. (**C**) Quantification of run-on data. Each indicated strain was tested at both 25 and 37°C in three separate experiments. All values obtained were corrected by subtraction of the M13 background signal and normalized to the value of probe 1, which was fixed at 100%.

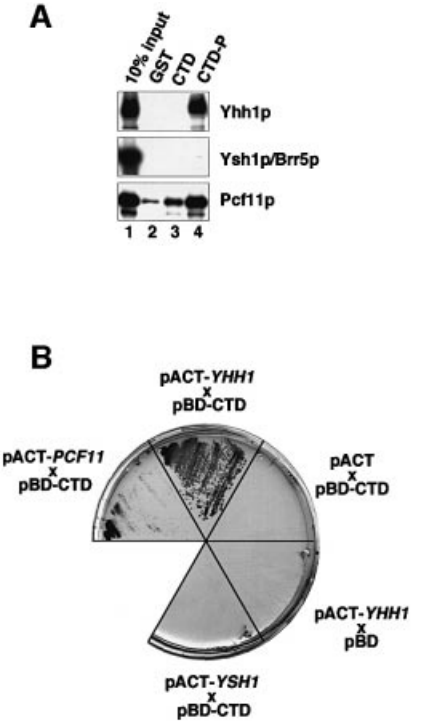

Fig. 7. Yhh1p interacts with the phosphorylated CTD. (**A**) Pull-down experiments with 1 μg GST (lanes 2), GST–CTD (lane 3) and phosphorylated GST–CTD (lane 4) and *in vitro* translated, [^{35}S]methionine-labelled proteins as indicated to the right of each panel. Bound proteins were separated by SDS–PAGE and visualized by autoradiography. Input (lane 1) shows 10% of *in vitro* translation reactions in the binding assay. (**B**) Yeast two-hybrid analysis. A plasmid carrying the CTD fused to the GAL4 DNA binding domain (pBD–CTD) and plasmids carrying test genes fused to the *GAL4* activation domain (pACT–YHH1, pACT–YSH1 and pACT–PCF11) were co-transformed into strain Y190. Activation of *HIS3* expression was tested by plating the indicated yeast strains on synthetic medium lacking histidine and containing 35 mM 3-amino-1,2,4-triazole. For control, the pBD–CTD and pACT–YHH1 vectors were tested in the presence of empty pACT and pBD vectors, respectively.

general defect in the catalytic steps of pre-mRNA 3′-end formation. Instead, Yhh1p might be essential for the formation of a subset of mRNAs with weak poly(A) sites that require its contribution in poly(A) site definition. In accordance with this hypothesis, we observed that recognition of the weak *ACT1* poly(A) site is deficient in *yhh1* mutant cells.

We delineated a novel RNA-binding domain in Yhh1p. Two minimal protein fragments containing residues 501–666 and 584–749, respectively, displayed RNA-binding activity that was comparable to full-length protein. The observation that only weak binding activity could be detected with the fragment common to both of these domains suggested that the RNA-binding domain might be constituted by an overlapping bipartite motif. Possibly, folding constraints necessitate the presence of sequences in addition to the sequence that is shared by the two minimal fragments to produce a polypeptide that is active in RNA binding, and structural analysis will be required to define this domain further. The RNA-binding domains show significant sequence similarity to higher eukaryotic homologues of Yhh1p, SAP 130 and XPE proteins. This similarity results from degenerate sequence repeats, which have been predicted to form multiple β-propeller-like structures (Neuwald and Poleksic, 2000). These structures are formed by the large WD repeat protein family (Smith *et al*., 1999), but have not previously been implicated in RNA binding. Our observations that the Yhh1p RNA-binding domain aligns with β-propeller repeat sequences suggest that this structure can function in RNA binding. A common functional role for the CPSF 160, SAP130 and XPE protein families in nucleic acid binding was suggested by Caspary *et al*. (1999). CPSF 160 is the only protein besides Yhh1p that has been shown to interact with RNA (Murthy and Manley, 1995). It was proposed that CPSF 160 carries a highly degenerate RNP-type RNA-binding motif (Murthy and Manley, 1995), which lies N-terminal to the sequence similarity found with the

Yhh1p RNA-binding motif. The identification of the RNA-binding domain in Yhh1p now offers the possibility to test whether a similar domain is functional in CPSF 160, SAP 130 and XPE proteins.

We found that Yhh1p specifically interacted with the phosphorylated CTD. The CF IA component Pcf11p was shown to interact with the CTD (Barilla *et al*., 2001), but no interactions were identified for CPF subunits. Experiments with CF II (a subfactor of CPF) suggested that at least one of its components binds the CTD (Rodriguez *et al*., 2000). Therefore, we propose that Yhh1p mediates or contributes to CTD binding of CPF. The termination defect that was observed in *yhh1* mutants highlights the importance of Yhh1p in coupling 3′-end formation and termination. Notably, several other mutant alleles of CPF components (*pap1-1*, *fip1-1*, *yth1-1*) were previously found not to affect termination (Birse *et al*., 1998). Components of CF IA that were previously shown to be required for termination (Birse *et al*., 1998; Aranda and Proudfoot, 2001) interact with either the CTD (Pcf11p) or the RNA substrate (Rna15p). In contrast, Yhh1p interacts with both the RNA and the CTD, and would thus fulfil a more stringent requirement attributable to a factor that is directly involved in signalling poly(A) site recognition to pol II. According to current models, we suggest that Yhh1p mediates the association of CPF with the phosphorylated CTD and that CPF travels with elongating pol II. Once the poly(A) site of the transcript emerges, CPF binds to the RNA, and dissociation of Yhh1p from the CTD might provide a vital signal for pol II to terminate transcription. Mammalian CPSF has been shown to interact with the CTD (McCracken *et al*., 1997b). Considering the evolutionary conservation of yeast and mammalian 3′-end formation factors, we suggest that CPSF 160 might also serve a function in termination that involves its properties in recognizing the AAUAAA signal.

Materials and methods

Yeast and E.coli procedures

Manipulations and growth of *E.coli* and *Saccharomyces cerevisiae* were performed by established procedures. For temperature shift experiments, yeast cultures were grown to mid-exponential phase at 23°C in YPD, diluted with pre-warmed medium and transferred to a water bath at 37°C. Isolation of temperature-sensitive *yhh1* alleles is described in the Supplementary data available at *The EMBO Journal* Online. Genotypes of yeast strains used are: W303 (*ade2; leu2; ura3; trp1-1; his3*; Thomas and Rothstein, 1989); MO2 (*ura3-1/ura3-1; trp1Δ/trp1Δ; ade2-1/ade2-1*; *leu2-3,112/leu2-3,112; his3-11,15/his3-11,15; YHH1/yhh1::TRP1*); YBD40 [*ura3; ade2-1; leu2,3-112; his3-11,15; trp1Δ; yhh1::TRP1* (pBD57; *YHH1-URA3-CEN*)]; YBD57 [*ura3; ade2-1; leu2,3-112; his3-11,15; trp1Δ; yhh1::TRP1* (*yhh1-3-URA3-CEN*)]; YBD58 [*ura3; ade2-1; leu2,3-112; his3-11,15; trp1Δ; yhh1::TRP1* (*yhh1-6-URA3-CEN*)]; YBD63 [*ura3; ade2-1; leu2,3-112; his3-11,15; trp1Δ; yhh1::TRP1* (*yhh1-12-URA3-CEN*)]; Y190 (*ura3-52 trp1-901 ade2-101 leu2-3, 112 his3-200r gal4D gal80D URA3::GAL1-lacZ, LYS2::GAL1-HIS3 cyhr*; Clontech); *rna15–1* (*ura3-1, trp1-1, ade2-1, leu2-3,112, his3-11,15, rna15–1*; Minvielle-Sebastia *et al*., 1994).

Plasmids

Construction of plasmids used in this study is described in detail in the Supplementary data: pBD57 (*YHH1-URA3-CEN*), pBD64 (*YHH1-LEU2-CEN*), pBD75 (GST-Yhh1p-H_6), pACT–YHH1 and plasmids used for expression of N- and C-terminally truncated Yhh1p proteins and internal domains. Fusions of the CTD with GST (GST–CTD) and the GAL4 DNA binding domain (pBD–CTD) included the C-terminal 229 amino acids of Rpb1p (nucleotides 4514–5201).

Extract preparation and in vitro cleavage and polyadenylation assays

3′-end processing extracts were prepared as described previously (Ohnacker *et al*., 2001) Cleavage and polyadenylation assays were also carried out as described previously (Minvielle-Sebastia *et al*., 1994). When cleavage only was assayed, EDTA replaced MgAc and CTP replaced ATP. In a single reaction, 30–40 μg of total protein were incubated with pre-mix. For assays at elevated temperatures, pre-mix and extracts were first pre-incubated separately at 37°C for 5 min, combined and assayed at 37°C for 45 min. *In vitro* substrates were obtained by run-off transcription as described previously (Dichtl and Keller, 2001).

RNA analyses

Total RNAs were extracted from yeast using a hot phenol method, separated on 1.2% formaldehyde/agarose or 8.3 M urea/8% polyacrylamide gels and transferred to Hybond N^+ (Amersham Pharmacia Biotech, Buckinghamshire, UK) membranes. Randomly primed probes were generated with [α-^{32}P]dATP using the Boehringer Mannheim kit. Templates were produced by polymerase chain reaction (PCR) on yeast genomic DNA covering the following sequences (numbers relative to start codon): PGK1 (1–1251), CYH2 (1–450), ADH1 (1–1049), STE3 (1–600), CYC1 (1–629), ACT1 (1–677) and ACT1-3′ (930–1536). Oligonucleotides anti-18S rRNA (CAGACAAATCACTCCA) and anti-U14 snoRNA (TCACTCAGACATCCTAGG) were labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase. RNase H reactions on total RNA were carried out with oligo ACT1-RNaseH (CTGGGAACA-TGGTGGTACCACC) and RNAs were separated on 8.3 M urea/6% polyacrylamide gels. Poly(A)-tail analysis was performed as described previously (Minvielle-Sebastia *et al*., 1998). Transcriptional run-on analysis was performed according to Birse *et al*. (1998).

Recombinant protein expression and purification

Escherichia coli BL21 cells transformed with the respective plasmids were grown in 2× YT at 25°C to an OD_{600} of 2.0. After induction with 0.5 mM IPTG, incubation was continued for 6 h. Proteins carrying a C-terminal His_6 tag were first purified with Ni^{2+}-NTA (Qiagen) according to the manufacturer's instructions. GST fusion proteins and double-tagged constructs were purified on glutathione–Sepharose 4B as recommended (Amersham Pharmacia Biotech, Piscataway, NJ); however, all steps were performed at 4°C or on ice. Proteins were eluted with buffer G [10 mM reduced glutathione, 50 mM Tris–HCl pH 8.0, 0.01% NP-40, 0.01 mg/ml bovine serum albumin (BSA), 10% glycerol].

Protein–protein interactions

Phosphorylation of GST–CTD was performed as described previously (Hirose and Manley, 1998). GST, GST–CTD and phosphorylated GST–CTD proteins were bound to 100 μl glutathione–Sepharose 4B (Pharmacia) at 0.02 mg/ml in buffer P [phosphate-buffered saline containing 0.02% NP-40, 1 mM dithiothreitol (DTT), 0.4 μg/ml leupeptin, 0.7 μg/ml pepstatin] for 1 h at 4°C. Beads were washed five times with buffer P containing 1 M NaCl, equilibrated twice with buffer P, and resuspended in 1 ml buffer P supplemented with 0.1 mg/ml BSA. Phosphorylation of the CTD was controlled by immunodetection with an antibody directed against phosphoserine 5 (H14, BAbCO, Richmond, CA) or the unphosphorylated CTD of pol II (8WG16, BAbCO). *In vitro* translations were performed with the TNT-coupled transcription–translation system (Promega). For GST pull-down, 45 μl GST or GST-fusion protein were incubated with *in vitro* translated ^{35}S-labelled proteins for 1.5 h at 4°C. The resin was washed four times with 1 ml buffer P for 20 min at 4°C. Proteins were eluted in sample buffer and resolved by SDS–PAGE.

RNA–protein interactions

Pull-down experiments with GST fusion proteins and CYC1 and GAL7 wild-type and mutant RNAs, and RNase H protection experiments were performed as described previously (Dichtl and Keller, 2001). Filter-binding experiments were performed in a 50-μl volume in RNA binding buffer (RBB) [13 mM HEPES–KOH pH 7.9, 28 mM $(NH_4)_2SO_4$, 33 mM KCl, 1 mM $MgCl_2$, 1 mM EDTA, 0.2 mM DTT, 0.01% NP-40]. Internally labelled CYC1 or GAL7 RNA (20 fmol) were incubated with 0, 5, 10, 100, 400, 800 and 1500 fmol GST-Yhh1p-H_6 for 15 min at room temperature. The reactions were applied to NC20 nitrocellulose filters (Schleicher and Schuell, Dassel, Germany) that were pre-treated with RBB containing 10 μg *E.coli* tRNA. Following a wash step with 1.5 ml

RBB, the radioactivity retained on the filters was determined by scintillation counting with Emulsifier Safe liquid (Packard, Groningen, Netherlands). Apparent K_D values were determined as described by Wahle *et al*. (1993). S7 nuclease protection experiments, partial alkaline hydrolysis and nuclease T1 digest were performed as described previously (Knapp, 1989) with modifications. One hundred femtomoles of 5′-[^{32}P]-labelled sCYC1 RNA was pre-incubated in the absence and presence of 100, 200 and 400 ng of GST-Yhh1p-His_6 in buffer (25 mM Tris–HCl pH 8.0, 30 mM KAc, 5 mM MgAc, 1 mM DTT) for 10 min at 37°C. Then, 0.1 U of S7 nuclease was added and reactions were stopped after 6 min by the addition of loading buffer containing 100 mM EDTA. Reaction products were separated on 12% polyacrylamide/8.3 M urea gels. Electromobility shift assays were performed with proteins as indicated and 20 fmol sCYC1 RNA in 10 µl containing 100 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 0.5 U RNA Guard and 100 ng *E.coli* tRNA. Following pre-incubation at 30°C for 15 min, reactions were separated on native 8% polyacrylamide gels (acrylamide/bisacrylamide 80:1 in 25 mM Tris base, 25 mM boric acid, 1 mM EDTA) at room temperature for 4 h at 200 V.

Computing

Protein and nucleotide sequence searches were performed with the BLASTP and TBLASTX programs on the NCBI server (Altschul *et al*., 1997) (http://www.ncbi.nlm.nih.gov). The Yhh1p RNA-binding domain and corresponding similar sequences were aligned with the CLUSTAL_X program (Thompson *et al*., 1997) and subsequent manual refinement.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online.

Acknowledgements

We thank G.Stumpf and H.Domdey for α-Yhh1p serum, N.Proudfoot for M13 phages and the pGAL-CYC1 plasmid, and I.Kaufmann and J.Wolf for comments on the manuscript. This work was supported by the University of Basel, the Swiss National Science Foundation, the European Community (via the Bundesamt für Bildung und Wissenschaft, Bern) and the Louis-Jeantet-Foundation for Medicine. B.D. was the recipient of an EMBO long-term fellowship.

References

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Aranda,A. and Proudfoot,N. (2001) Transcriptional termination factors for RNA polymerase II in yeast. *Mol. Cell.*, **7**, 1003–1011.

Barabino,S.L.M., Ohnacker,M. and Keller,W. (2000) Distinct roles of two Yth1p domains in 3′-end cleavage and polyadenylation of yeast pre-mRNAs. *EMBO J.*, **19**, 3778–3787.

Barilla,D., Lee,B.A. and Proudfoot,N.J. (2001) Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **98**, 445–450.

Birse,C.E., Lee,B.A., Hansen,K. and Proudfoot,N.J. (1997) Transcriptional termination signals for RNA polymerase II in fission yeast. *EMBO J.*, **16**, 3633–3643.

Birse,C.E., Minvielle-Sebastia,L., Lee,B.A., Keller,W. and Proudfoot,N.J. (1998) Cleavage of the primary transcript couples 3′-end RNA-processing with termination of pol II transcription. *Science*, **280**, 298–301.

Calvo,O. and Manley,J.L. (2001) Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation and termination. *Mol. Cell*, **7**, 1013–1023.

Caspary,F., Shevchenko,A., Wilm,M. and Seraphin,B. (1999) Partial purification of the yeast U2 snRNP reveals a novel yeast pre-mRNA splicing factor required for pre-spliceosome assembly. *EMBO J.*, **18**, 3463–3474.

Chen,J. and Moore,C. (1992) Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA. *Mol. Cell. Biol.*, **12**, 3470–3481.

Cho,E.J., Takagi,T., Moore,C.R. and Buratowski,S. (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev.*, **11**, 3319–3326.

Dantonel,J.-C., Murthy,K.G.K., Manley,J.L. and Tora,L. (1997) Transcription factor TFIID recruits factor CPSF for formation of 3′ end of mRNA. *Nature*, **389**, 399–402.

Das,B.K., Xia,L., Palandjian,L., Gozani,O., Chyung,Y. and Reed,R. (1999) Characterization of a protein complex containing spliceosomal proteins SAPs 49, 130, 145, 155. *Mol. Cell. Biol.*, **19**, 6796–6802.

Dichtl,B. and Keller,W. (2001) Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. *EMBO J.*, **20**, 3197–3209.

Gavin,A.C. *et al*. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141–147.

Graber,J.H., Cantor,C.R., Mohr,S.C. and Smith,T.F. (1999) Genomic detection of new yeast pre-mRNA 3′-end-processing signals. *Nucleic Acids Res.*, **27**, 888–894.

Gross,S. and Moore,C.L. (2001) Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3′-end formation. *Mol. Cell. Biol.*, **21**, 8045–8055.

Hirose,Y. and Manley,J.L. (1998) RNA polymerase II is an essential mRNA polyadenylation factor. *Nature*, **395**, 93–96.

Hirose,Y. and Manley,J.L. (2000) RNA polymerase II and the integration of nuclear events. *Genes Dev.*, **14**, 1415–1429.

Keller,W., Bienroth,S., Lang,K.M. and Christofori,G. (1991) Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3′ processing signal AAUAAA. *EMBO J.*, **10**, 4241–4249.

Knapp,G. (1989) Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol.*, **180**, 192–212.

Mandart,E. and Parker,R. (1995) Effects of mutations in the *Saccharomyces cerevisiae RNA14*, *RNA15* and *PAP1* genes on polyadenylation *in vivo*. *Mol. Cell. Biol.*, **15**, 6979–6986.

McCracken,S. *et al*. (1997a) 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.*, **11**, 3306–3318.

McCracken,S., Fong,N., Yankulov,K., Ballantyne,S., Pan,G., Greenblatt,J., Patterson,S.D., Wickens,M. and Bentley,D.L. (1997b) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, **385**, 357–361.

Minvielle-Sebastia,L., Preker,P.J. and Keller,W. (1994) *RNA14* and *RNA15* proteins as components of a yeast pre-mRNA 3′-end processing factor. *Science*, **266**, 1702–1705.

Minvielle-Sebastia,L., Beyer,K., Krecic,A.M., Hector,R.E., Swanson, M.S. and Keller,W. (1998) Control of cleavage site selection during mRNA 3′-end formation by a yeast hnRNP. *EMBO J.*, **17**, 7454–7468.

Murthy,K.G.K. and Manley,J.L. (1995) The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3′-end formation. *Genes Dev.*, **9**, 2672–2683.

Neuwald,A.F. and Poleksic,A. (2000) PSI-BLAST searches using hidden markov models of structural repeats: prediction of an unusual sliding DNA clamp and of beta-propellers in UV-damaged DNA-binding protein. *Nucleic Acids Res.*, **28**, 3570–3580.

Ohnacker,M., Barabino,S.M., Preker,P.J. and Keller,W. (2000) The WD-repeat protein Pfs2p bridges two essential factors within the yeast pre-mRNA 3′-end-processing complex. *EMBO J.*, **19**, 37–47.

Osheim,Y.N., Proudfoot,N.J. and Beyer,A.L. (1999) EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Mol. Cell*, **3**, 379–387.

Proudfoot,N.J. (1989) How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem. Sci.*, **14**, 105–110.

Proudfoot,N.J., Furger,A. and Dye,M.J. (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 502–512.

Rodriguez,C.R., Cho,E.J., Keogh,M.C., Moore,C.L., Greenleaf,A.L. and Buratowski,S. (2000) Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Mol. Cell. Biol.*, **20**, 104–112.

Shatkin,A.J. and Manley,J.L. (2000) The ends of the affair: capping and polyadenylation. *Nat. Struct. Biol.*, **7**, 838–842.

Smith,T.F., Gaitatzes,C., Saxena,K. and Neer,E.J. (1999) The WD repeat: a common architecture for diverse functions. *Trends Biochem. Sci.*, **24**, 181–185.

Stumpf,G. and Domdey,H. (1996) Dependence of yeast pre-mRNA 3′-end processing on CFT1: a sequence homolog of the mammalian AAUAAA binding factor. *Science*, **274**, 1517–1519.

Takao,M., Abramic,M., Moos,M.,Jr, Otrin,V.R., Wootton,J.C., McLenigan,M., Levine,A.S. and Protic,M. (1993) A 127 kDa component of a UV-damaged DNA-binding complex, which is defective in some xeroderma pigmentosum group E patients, is

homologous to a slime mold protein. *Nucleic Acids Res.*, **21**, 4111–4118.

Thomas,B.J. and Rothstein,R.J. (1989) Elevated recombination rates in transcriptionally active DNA. *Cell*, **56**, 619–630.

Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**, 4876–4882.

Tran,D.P., Kim,S.J., Park,N.J., Jew,T.M. and Martinson,H.G. (2001) Mechanism of poly(A) signal transduction to RNA polymerase II *in vitro*. *Mol. Cell. Biol.*, **21**, 7495–7508.

Valentini,S.R., Weiss,V.H. and Silver,P.A. (1999) Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3′-end formation. *RNA*, **5**, 1–9.

van Helden,J., del Olmo,M. and Perez-Ortin,J.E. (2000) Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. *Nucleic Acids Res.*, **28**, 1000–1010.

Wahle,E., Lustig,A., Jenö,P. and Maurer,P. (1993) Mammalian poly(A)-binding protein II. *J. Biol. Chem.*, **268**, 2937–2945.

Yuryev,A., Patturajan,M., Litingtung,Y., Joshi,R.V., Gentile,C., Gebara,M. and Corden,J.L. (1996) The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc. Natl Acad. Sci. USA*, **93**, 6975–6980.

Zhao,J., Hyman L. and Moore C. (1999). Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.*, **63**, 405–445.

Received March 6, 2002; revised May 28, 2002; accepted June 3, 2002

Note added in proof

Consistent with a function of Yhh1p in coupling of 3′-end formation and transcription, Licatalosi *et al*. (2002) showed that CPF subunits functionally interact with RNA polymerase II during transcription.

Licatalosi, D.D., Geiger,G., Minet,M., Schroeder,S., Cilli,K., McNeil,J.B. and Bentley,D.L. (2002) Functional interaction of yeast pre-mRNA 3′-end processing factors with RNA polymerase II. *Mol. Cell*, **9**, 1101–1111.