MOLECULAR AND CELLULAR BIOLOGY, Sept. 1992, p. 4215–4229
0270-7306/92/094215-15$02.00/0

Vol. 12, No. 9

Identification of Pre-mRNA Polyadenylation Sites in *Saccharomyces cerevisiae*

STEFAN HEIDMANN, BRIGITTE OBERMAIER, KARIN VOGEL, AND HORST DOMDEY*

Laboratorium für Molekulare Biologie-Genzentrum-der Ludwig-Maximilians-Universität München, Am Klopferspitz 18a, D-8033 Martinsried, Germany

Received 19 December 1991/Returned for modification 27 February 1992/Accepted 12 June 1992

In contrast to higher eukaryotes, little is known about the nature of the sequences which direct 3′-end formation of pre-mRNAs in the yeast *Saccharomyces cerevisiae*. The hexanucleotide AAUAAA, which is highly conserved and crucial in mammals, does not seem to have any functional importance for 3′-end formation in yeast cells. Instead, other elements have been proposed to serve as signal sequences. We performed a detailed investigation of the yeast *ACT1*, *ADH1*, *CYC1*, and *YPT1* cDNAs, which showed that the polyadenylation sites used in vivo can be scattered over a region spanning up to 200 nucleotides. It therefore seems very unlikely that a single signal sequence is responsible for the selection of all these polyadenylation sites. Our study also showed that in the large majority of mRNAs, polyadenylation starts directly before or after an adenosine residue and that 3′-end formation of *ADH1* transcripts occurs preferentially at the sequence PyAAA. Site-directed mutagenesis of these sites in the *ADH1* gene suggested that this PyAAA sequence is essential for polyadenylation site selection both in vitro and in vivo. Furthermore, the 3′-terminal regions of the yeast genes investigated here are characterized by their capacity to act as signals for 3′-end formation in vivo in either orientation.

Processing of mRNA precursors is an essential event in gene expression in eukaryotes. Generally speaking, the maturation of a primary transcript generated by RNA polymerase II can be divided into several distinct processing reactions: capping of the 5′ terminus of the primary transcript; modification of some adenosine residues; excision of intron sequences via splicing; and cleavage of the primary transcript at its 3′ end and then addition of a poly(A) tail.

The signals which govern cleavage of an mRNA precursor at its 3′ end and addition of the poly(A) tail have been well defined in mammals; in addition to some weakly conserved signals in the DNA region downstream of the polyadenylation site, there is a strongly conserved hexanucleotide sequence (AATAAA) located around 10 to 30 nucleotides upstream of the polyadenylation site (for a review, see reference 48). In contrast, the 3′ ends of the genes transcribed by RNA polymerase II in the yeast *Saccharomyces cerevisiae* do not contain such a strongly conserved nucleotide sequence. Only a subset of the yeast genes sequenced so far contain the hexanucleotide sequence AATAAA in their 3′ regions, and for *ADH2*, it was shown that mutation or deletion of that signal did not impair correct 3′-end formation in vivo (19). Instead, a number of other different nucleotide sequences have been proposed which might serve as signals for processing of pre-mRNA 3′ ends, such as the octanucleotide TTTTTATA and the tripartite sequence TAG . . .TA(T)GT. . TTT (15, 50). More recently, these two motifs have been slightly modified to TTTTTAT (20) and TAG. . .TATGTA (36), and there is also evidence for participation of the signal TATATA or TATCTA in the process of transcription termination and/or 3′-end processing (36). The presence or absence of these signals and the capacity of the 3′-end regions of some yeast genes to govern 3′-end formation in either orientation led to the postulation of different classes of polyadenylation sites (20). Most of the published work has dealt with the yeast iso-1-cytochrome *c* (*CYC1*) gene, in which a 38-bp deletion was reported to abolish transcription termination and 3′-end processing (6, 50). Closer investigation of the deleted sequence and of revertants that displayed normal levels of *CYC1* mRNA led to the formulation of the above-mentioned signal sequences (36, 50). The observation that these sequence motifs are necessary but not sufficient to direct 3′-end formation indicates the participation of additional regulatory elements which may be located at or near the actual polyadenylation site (19, 20, 36). In this study, we investigated in greater detail the actual polyadenylation sites in various yeast genes and found that the polyadenylation sites in a given yeast gene can be distributed over a wide sequence range. Site-directed mutagenesis and deletions of the *ADH1* polyadenylation sites suggested that, in combination with distinct 3′-end-processing signals, the sequence $Py(A)_n$ is a preferential site for polyadenylation.

MATERIALS AND METHODS

Bacterial and yeast strains and culture conditions. *Escherichia coli* XL1 Blue (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI*^{q} ZΔM15 Tn*10*(*tet*)] (5) was used for cloning procedures. Competent *E. coli* cells were produced by the method of Hanahan (14). *E. coli* was grown on Luria-Bertani medium supplemented with the appropriate antibiotics.

Yeast strain XJ 24-24a (*MAT***a** *adeb arg4-17 trp1-1 tyr7-1 aro7-1*) was the mRNA source of the cDNA library contained in the λZAPII vector (Stratagene). The mRNA for the rapid amplification of cDNA ends (RACE) employing the polymerase chain reaction (PCR) analysis was isolated from transformed DH484 cells (*MAT***a** *ade2-1 leu2-3 leu2-112 can1-100 trp5-48 ura4-11 lys1-1*). The yeast whole-cell extracts were prepared from strain EJ101 (*MATα trp1 pro1-126 prb1-112 pep4-3 prc1-126*). Untransformed yeast cells were grown on yeast extract-peptone-dextrose medium. *S. cere-*

* Corresponding author.

visiae DH484 was transformed by the lithium acetate procedure (21). Transformants were selected on minimal medium supplemented with all necessary constituents except for leucine.

Enzymes and chemicals. Restriction endonucleases, T7 RNA polymerase, T4 DNA ligase, avian myeloblastosis virus reverse transcriptase, exonuclease III, and T4 polynucleotide kinase were obtained from Boehringer Mannheim. SP6 RNA polymerase, human placental RNase inhibitor, [γ-^{32}P]ATP (185 TBq/mmol), [α-^{32}P]dCTP (111 TBq/mmol), [α-^{32}P]UTP (29.6 TBq/mmol), and α-^{35}S-dATP (24 TBq/mmol) were from Amersham. Exonuclease VII was obtained from Bethesda Research Laboratories. *Taq* polymerase was purchased from Cetus Corporation. S1 nuclease was from Pharmacia.

The following oligonucleotides were synthesized with an automated DNA synthesizer from Applied Biosystems:

O1: 5′-AATAACCAAAGCAGCAACCT-3′
O2: 5′-CGACGTAACATAGTTTTTCC-3′
O3: 5′-TGGGGCTCTGAATCTTTC-3′
O4: 5′-GACTCGAGTCTAGAGAGCTCC$(T)_{17}$-3′
O5: 5′-GACTCGAGTCTAGAGAGCTCC-3′
O6: 5′-CAAGAATTCACCAACTGGGACG-3′
O7: 5′-GATCTAAGCTTGCGAATTCTCTAGAGTCGACCTGCAGA-3′
O8: 5′-GATCTCTGCAGGTCGACTCTAGAGAATTCGCAAGCTTA-3′

Reverse transcription and amplification of specific polyadenylated mRNAs by RACE-PCR. RACE-PCR was done by the method of Frohman et al. (9) with some modifications. Poly(A)$^+$ RNA (2 μg) was incubated at 42°C for 90 min in a 20-μl reaction mixture containing 50 mM KCl, 20 mM Tris Cl (pH 8.4), 6 mM $MgCl_2$, 1 mM each deoxynucleoside triphosphate (dNTP), 1 mM dithiothreitol, 2.4 pmol of oligonucleotide O4, 7 U of RNAsin, and 25 U of avian myeloblastosis virus reverse transcriptase. After inactivation of the reverse transcriptase for 3 min at 95°C, the reaction mixture was brought to a volume of 1 ml with TE (10 mM Tris Cl [pH 7.5], 0.1 mM EDTA) to dilute oligonucleotide O4. A 5-μl sample of this cDNA preparation was used in a PCR amplification in 50 μl containing 50 mM KCl, 20 mM Tris Cl (pH 8.4), 6 mM $MgCl_2$, 100 μM each dNTP, 5 pmol of both O5 and O6, and 1 U of *Taq* DNA polymerase. The reaction mixtures were overlayered with 50 μl of mineral oil and incubated in a Perkin-Elmer thermal cycler with the following parameters. For the first cycle, denaturation lasted for 3 min at 94°C, annealing for 3 min at 50°C, and extension for 40 s at 72°C (since the gene-specific upstream primer O6 with the 5′-protruding *Eco*RI site did not exactly match the template, it was necessary to run the first amplification cycle at a lower annealing temperature). The subsequent 30 PCR cycles were performed with the following parameters: denaturation for 40 s at 94°C, annealing for 30 s at 60°C, and extension for 40 s at 72°C. A 15-μl sample of the reaction mixtures was analyzed on an agarose gel; to 20 μl of the samples, 4 μl of a solution containing 2.5% sodium dodecyl sulfate, 130 mM EDTA, and 2 mg of proteinase K per ml was added, and the mixture was incubated for 45 min at 37°C to digest the *Taq* DNA polymerase. After extraction twice with phenol and once with chloroform, the PCR products were precipitated with ethanol and then digested with *Eco*RI and *Xba*I (sites present in the primers) and ligated into an *Eco*RI-*Xba*I-digested pBluescriptKSII$^-$ vector (Stratagene).

Plasmid constructions and manipulations. The 3′-end region of the yeast *ADH1* gene was isolated as a 463-bp *Bam*HI-*Hin*dIII fragment from plasmid pAAH5 (2) and cloned into a *Bam*HI-*Hin*dIII-digested pGEM2 vector (Promega) to provide plasmid pgADH463, from which transcription with T7 RNA polymerase led to transcripts in sense orientation. The insert was shortened from its 3′ end by treatment with exonucleases III and VII after digestion with *Bam*HI and *Sac*I by the method of Yanisch-Perron et al. (49). The insert was also shortened from its 5′ end up to two nucleotides upstream of the *ADH1* translational stop codon by in vitro mutagenesis. This fragment had been shown earlier to direct 3′-end formation efficiently in vivo (45). The combination of deletions from both ends, each of them still working in vivo, resulted in a fragment of 91 bp. The corresponding complementary oligonucleotides with additional 5′-protruding GATC ends were synthesized, hybridized, and cloned in both orientations into the *Bam*HI-cleaved vector pBluescriptIIKS$^-$ to yield the plasmids pKSADH91s and pKSADH91a, where s (sense) and a (antisense) refer to the orientation of the insert relative to the T7 promoter.

The *ACT1* terminator region was obtained as a 354-bp *Sau*3AI fragment from the plasmid pYactI (32). This region was cloned into a *Bam*HI-cleaved pGEM2 vector, so that transcription with SP6 RNA polymerase from this plasmid (pgACT) yielded RNAs in the sense orientation.

The *YPT1* terminator region was obtained as a 494-bp *Sau*3AI fragment from the plasmid pYactI (32) and was cloned similarly into the *Bam*HI-digested vector pBluescript IIKS$^-$ in the sense orientation.

The 243-bp *CYC1* terminator region was assembled with synthetic oligodeoxynucleotides. It corresponds to the 236-bp *Taq*I fragment used by Butler and Platt (6) with additional *Sau*3AI sites at both ends. This fragment was cloned into the *Bam*HI-cleaved vector pGEM2 in the sense orientation relative to the T7 promoter.

For in vivo analysis, all investigated fragments were inserted into a truncated *ACT1* transcription unit starting with the *Bam*HI site upstream of the promoter and ending at the internal *Hin*dIII site with the *ACT1* intron and the internal 321-bp *Bgl*II fragment deleted (32). This transcription unit was contained as a 933-bp *Bam*HI-*Hin*dIII fragment in the yeast-*E. coli* shuttle vector YEp351 (16), to yield the plasmid pSH50 (see Fig. 5). The 5′ overhangs of the 463-bp *Hin*dIII-*Bam*HI *ADH1* terminator fragment were filled in with Klenow polymerase, and *Hin*dIII linkers were ligated to the ends. The product was cleaved with *Hin*dIII and cloned into the *Hin*dIII-digested pSH50 to yield the plasmids pSH200 and pSH200a with the *ADH1* terminator in the sense and antisense orientations, respectively. The 354-bp *ACT1*, the 243-bp *CYC1*, and the 494-bp *YPT1* 3′-terminal regions were cloned as *Sau*3AI fragments in both orientations into the *Bgl*II site of pSH200.

The *ADH1* deletion fragments were tested in a different construct. First, the 3′-terminal region of the *ACT1* gene was cloned in the sense orientation, as a 386-bp *Sma*I-*Alu*I fragment isolated from plasmid pgACT, into the Klenow filled-in *Hin*dIII site of pSH50 (see Fig. 5) to yield plasmid pSH60. A subsequent in vitro mutagenesis was necessary to destroy the *Eco*RI site in the coding region of the *LEU2* gene contained in the vector. Therefore, the origin of replication of the single-stranded bacteriophage f1 was cloned as a 529-bp *Sac*I-*Bam*HI fragment obtained from the plasmid pUCf1 (Pharmacia) into the *Sac*I-*Bam*HI-cleaved plasmid pSH60. The mutagenesis was performed by using the Bio-Rad in vitro mutagenesis kit based on the method described by Kunkel (23), and the *Eco*RI site of the YEp351 polylinker was destroyed by Klenow fill-in and subsequent religation.

The two complementary oligonucleotides O7 and O8 were cloned into the *Bgl*II site of the truncated actin transcription unit to yield the plasmid pSH100. The deletion clones of the *ADH1* 3′-end region could then easily be cloned in the sense orientation into pSH100 as *Eco*RI-*Hin*dIII fragments obtained from the pGEM plasmids.

The DNA inserts were subcloned by standard procedures (25).

DNA sequencing. The λZAPII cDNA clones were sequenced by the chain termination method of Sanger et al. (39) as modified for supercoiled plasmids (8). The cloned RACE-PCR products were sequenced with the 373A DNA Sequencer from Applied Biosystems, using fluorescence-labeled primers and the thermal cycling protocol according to the manufacturer's protocols.

In vitro RNA synthesis and processing reactions. In vitro RNA synthesis was performed as described previously (28); in the reactions, 7mG(5′)ppp(5′)G was included in 10-fold excess compared with the concentration of GTP so that most of the transcripts were capped at their 5′ ends. The products were separated by electrophoresis on 6% polyacrylamide–8 M urea gels. The excised and eluted RNAs were used in in vitro processing reactions as described by Butler and Platt (6). The yeast extracts were prepared either as described by Lin et al. (24) or following a modified protocol by Butler et al. (7).

RNA analysis. Yeast total RNA was prepared by the hot-phenol method as described by Köhrer and Domdey (22). Poly(A)$^+$ RNA was isolated with an mRNA purification kit purchased from Pharmacia. For Northern (RNA) blotting, 1.5 μg of glyoxylated poly(A)$^+$ RNA was separated on a 1.5% agarose gel and transferred (46) to Hybond N nylon membranes (Amersham). The oligonucleotides O1, O2, and O3 served as hybridization probes, with O1 hybridizing immediately downstream of the actin intron (220 nucleotides upstream of the *Bgl*II site), O2 hybridizing 60 bp, and O3 hybridizing 179 bp downstream of the *Bgl*II site (see Fig. 5B).

Nuclease S1 analysis. Nuclease S1 mapping of the 3′ ends of the mRNAs was done by the method of Berk and Sharp (4) with some modifications. The DNA probes were isolated as *Bgl*II-*Pst*I fragments from plasmid pSH101 and its derivatives, and therefore their ends were not able to hybridize to genuine *ADH1* mRNA. The probes were end labeled with [α-^{32}P]dCTP and the Klenow fragment of DNA polymerase I at the *Bgl*II site. The labeled DNA fragments (50,000 cpm each) were denatured by boiling for 2 min and quenched on ice, and 2 μg of poly(A)$^+$ RNA and 3 μg of carrier tRNA were added. The mixture was ethanol precipitated and resuspended in 20 μl of hybridization solution containing 80% deionized formamide, 40 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)], 0.4 M NaCl, and 1 mM EDTA. The samples were heated for 10 min at 65°C and hybridized for 16 h at 30°C. A 300-μl portion of an ice-cold solution containing 0.28 M NaCl, 50 mM sodium acetate (pH 5.0), 4.5 mM $ZnSO_4$, 20 μg of single-stranded salmon sperm DNA per ml, and 600 U of nuclease S1 per ml was added. After 90 min of incubation at 16 or 23°C, reactions were stopped by adding 80 μl of stop solution containing 3.6 M ammonium acetate, 20 mM EDTA, and 40 μg of tRNA per ml. The low temperatures were chosen to avoid double-strand cleavage by nuclease S1 in the base-paired A+T-rich regions of the probes. The samples were extracted with phenol-chloroform and chloroform and precipitated with ethanol. Reaction products were separated on a 6% polyacrylamide–8 M urea denaturing gel.

RESULTS

Efficiency of 3′-end processing varies with different RNA substrates. As previously shown, in vitro transcripts spanning the 3′-end regions of a number of yeast genes can be accurately cleaved and polyadenylated in a yeast whole-cell extract (1, 6, 7, 38). To extend these analyses and to improve our understanding of the mechanism of 3′-end processing we subjected synthetic RNA transcripts representing the 3′ ends of the yeast *ACT1*, *ADH1*, and *YPT1* genes (Fig. 1A) to in vitro 3′-end-processing reactions. It seemed reasonable to examine in vitro 3′-end processing of the corresponding antisense transcripts, because for the *ADH1* and *YPT1* genes, antisense transcripts had been described (20, 47). Furthermore, as will be shown below, the corresponding DNA fragments directed efficient 3′-end formation in either orientation in vivo. All these RNAs were subjected to in vitro 3′-end-processing reactions under conditions previously described (6) in poly(A) polymerase-proficient and -deficient yeast extracts, respectively, and the products were analyzed by polyacrylamide gel electrophoresis (Fig. 2). The poly(A) polymerase-deficient extract was especially useful for the analysis of very inefficient processing reactions, since the small amounts of processed products were not spread out over a rather broad size range as a result of polyadenylation. The results are summarized in Table 1.

The 3′-end-processing efficiency of the *ACT1* sense transcript was extremely low (Fig. 2), and faint bands corresponding to processing products were observed only after incubation in a partially fractionated extract prepared by the method of Butler et al. (7). Only after several days of exposure were strong signals detected (data not shown), which corresponded roughly to the previously described polyadenylation sites (11). A distinctly higher processing efficiency was observed for the *YPT1*, *ADH1* (Fig. 2), and *CYC1* (data not shown) transcripts in the sense orientation. The derived cleavage positions are in good, although not perfect, accordance with the polyadenylation sites previously determined in vivo (3, 6, 10).

Analysis of the corresponding antisense transcripts showed a rather heterogeneous pattern. Whereas the *ACT1* and *ADH1* antisense transcripts yielded clearly visible cleavage and polyadenylation products, no such products could be observed when the antisense transcript of *CYC1* (data not shown) was subjected to a processing reaction. Possible processing products of *YPT1* did not emerge until after several days of exposure. These results were obtained with both polyadenylation-proficient and -deficient extracts.

Processing of the *ACT1* antisense transcript yielded four 5′-cleavage products in a narrow molecular weight range (Fig. 2, lower part). These products were very efficiently polyadenylated with an approximately 70-nucleotide poly(A) tail, which matches the previously determined lengths of poly(A) tails in yeast transcripts (13). Furthermore, the weak signals approximately 10 nucleotides above the 5′-cleavage products might indicate a biphasic polyadenylation process as described in higher eukaryotes (42), suggesting a similar mechanism in yeast cells. It is unlikely that these signals represent further 5′-cleavage products, because then they should also occur among the reaction products from tACT1a when incubated in the polyadenylation-deficient extract.

Sequences downstream from *ADH1* polyadenylation sites contribute to efficiency of the in vitro 3′-end-processing reaction. A set of experiments was designed to find out which part of the 3′-terminal region of the yeast *ADH1* gene was required to direct efficient 3′-end processing and polyadeny-

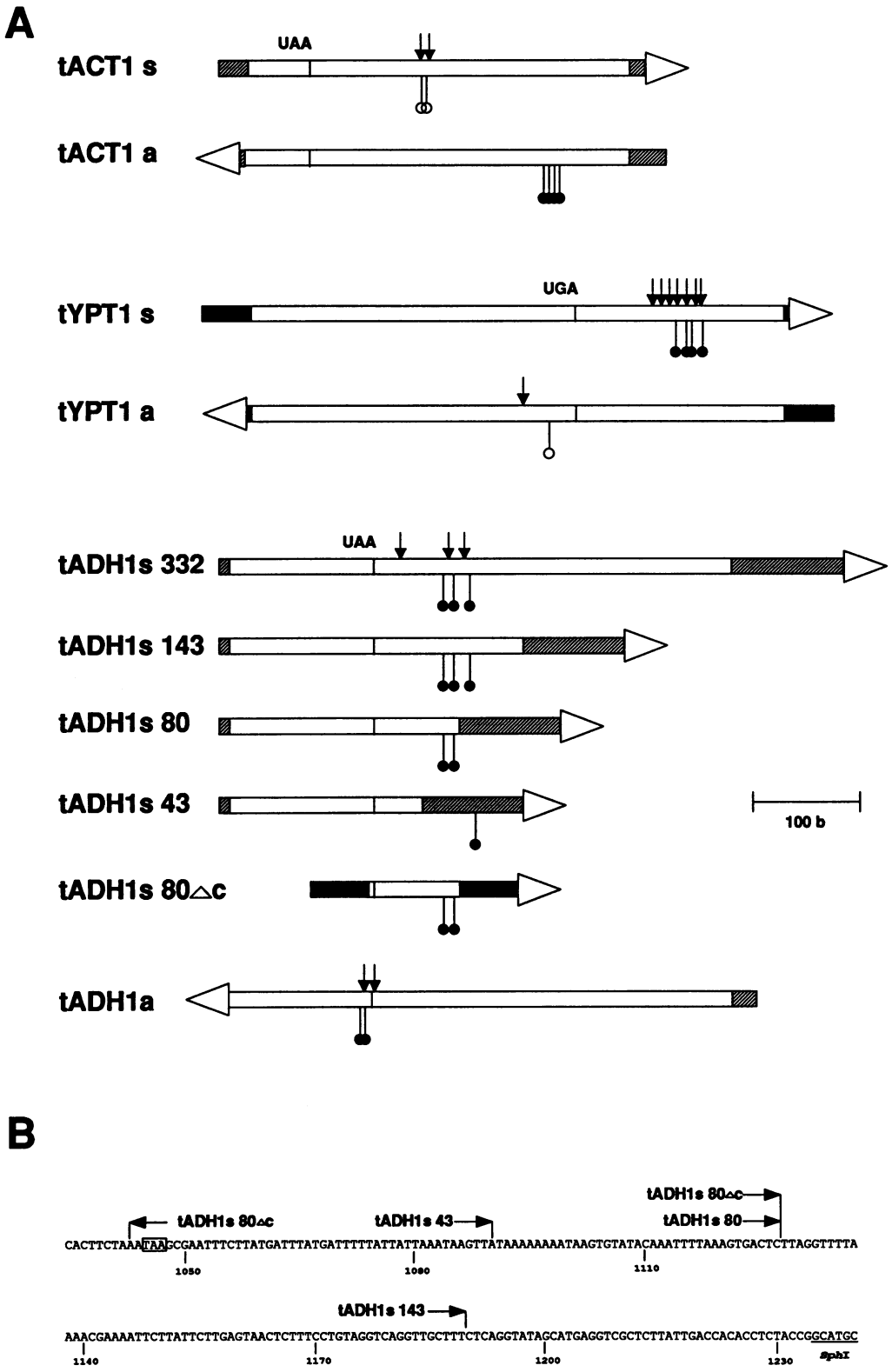


FIG. 1. (A) Schematic presentation of the SP6 and T7 transcripts tested for their capacity to become processed in vitro in yeast whole-cell extracts. Open bars represent the RNA sequences derived from yeast genomic sequences. Hatched bars represent sequences from the pGEM2 vector. Dark bars represent those from the pBluescript vector. The position of the translational stop codon is indicated. s and a denote the sense and antisense orientations of the transcripts, respectively. The vertical arrows point to the polyadenylation sites described previously in the literature. The pins indicate the cleavage positions observed in this study in the in vitro processing reactions. Closed pinheads indicate high processing efficiencies; open pinheads indicate very low processing efficiencies. Cleavage at the latter sites was only

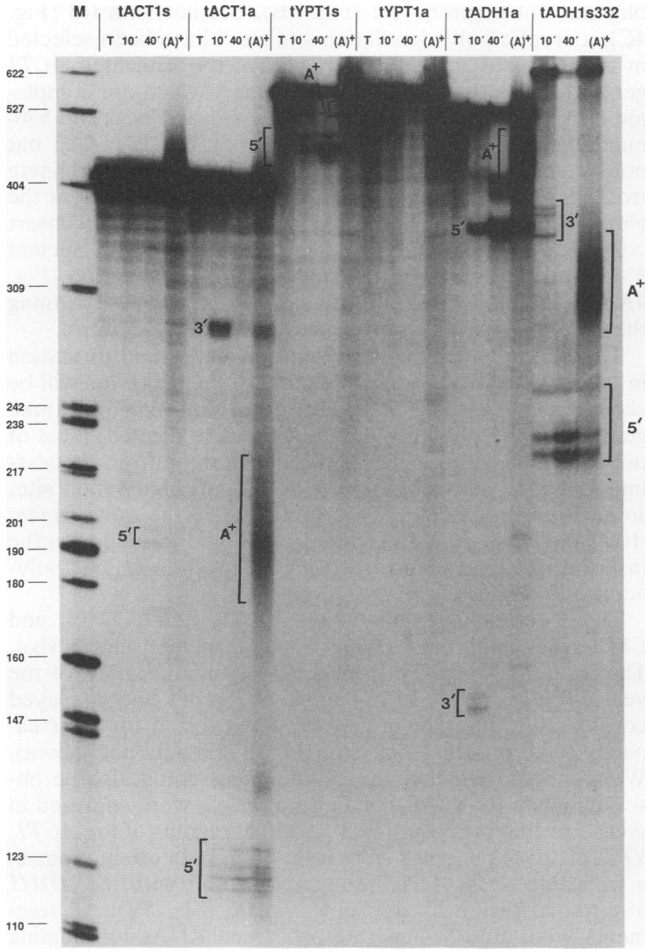


FIG. 2. Polyacrylamide gel analysis of the products formed during the in vitro 3'-end-processing reaction of transcripts shown in Fig. 1A. The radioactively labeled precursor RNAs (T) were incubated either in a polyadenylation-deficient extract for the indicated time or in a polyadenylation-proficient extract (A^+) for 40 min. The 5'-, 3'-, and polyadenylated 5'-cleavage products (A^+) are indicated. M, molecular weight marker (pBR322 DNA digested with *Hpa*II).

lation in vitro. A series of deletion clones was constructed and the corresponding RNA polymerase T7 transcripts (Fig. 1) were subjected to in vitro processing reactions. To improve the discrimination of unprocessed precursor molecules and processed products in the subsequent gel analyses, all transcripts contained, in addition to the *ADH1* sequences, vector-derived sequences at their 3' ends (as indicated in Fig. 1). Transcript tADH1s332 (Fig. 1), which contains 137 3'-terminal coding nucleotides of the *ADH1* gene plus the following 332 untranslated nucleotides, was processed very efficiently (Fig. 3A). Most of the transcript was converted after 7.5 min into three specific cleavage products (thick arrows in Fig. 3A) and polyadenylated material. By one-

TABLE 1. 3'-end-processing properties of several transcripts in vitro

3'-end region of gene	Orientation	Processing in extract A[a]	Processing in extract B[a]
ACT1	Sense	−	+
ACT1	Antisense	++	++
YPT1	Sense	+	++
YPT1	Antisense	−	+
ADH1	Sense	++	++
ADH1	Antisense	++	++
CYC1	Sense	++	ND[b]
CYC1	Antisense	−	−

[a] Extract A was prepared by the method of Lin et al. (24), and extract B [which lacked poly(A) polymerase activity] was prepared by the method of Butler and Platt (6).
[b] ND, not determined.

dimensional RNase T_1 fingerprint analysis, the RNAs contained in the three observed bands were characterized as cleavage products containing the 5' part of the primary transcript (data not shown). However, the cleavage sites did not coincide exactly with the three polyadenylation sites described by Bennetzen and Hall (3) and the single polyadenylation site found by Irniger et al. (20). These sites had been determined by nuclease S1 analysis. If the published polyadenylation sites had been used, RNA 5'-cleavage products of 171 (3), 214 (3), and 230 (3, 20) nucleotides would have been expected. Instead, the fragments obtained had lengths of approximately 210, 217, and 236 nucleotides. The observed as well as the previously published processing sites are indicated in Fig. 1A.

From transcript tADH1s143, the same 5'-cleavage products were obtained as with tADH1s332 (Fig. 3A). Furthermore, it was possible to detect the corresponding 3'-cleavage products (thin arrows in Fig. 3A) and to characterize them by RNase T_1 fingerprint analysis (data not shown). Transcript tADH1s80 (Fig. 1) was processed at the two remaining polyadenylation sites; however, the efficiency of the reaction was considerably reduced (especially cleavage of the more downstream site), and the polyadenylated products were only barely detectable (Fig. 3A). When the transcript tADH1s43, in which all three polyadenylation sites were deleted, was subjected to the processing reaction, no bands which could be attributed to cleavage in the *ADH1* sequence were detected (Fig. 3A). Assuming that the 240-nucleotide RNA which appeared after 40 min (Fig. 3A) is a capped 5'-cleavage product, then this RNA could only originate from a cryptic processing site in the attached 94 nucleotides of vector sequence. Specific processing products from tADH1s80Δc, the smallest of the investigated transcripts, could be detected only after a long exposure of the gel (Fig. 3B), and only when a polyadenylation-deficient extract (extract B in Table 1) was used.

All truncated *ADH1* 3'-end sequences containing natural polyadenylation sites work efficiently in vivo. All *ADH1* 3'-end fragments which had been analyzed in vitro were also tested by Northern blotting (Fig. 4A to C) to determine whether

detectable when the reaction was done in a polyadenylation-deficient extract and when the gels were overexposed. The naming of the *ADH1*s deletion clones refers to the number of *ADH1*-derived nucleotides downstream of the stop codon contained in the respective clones. In the construct tADH1s80Δc, all *ADH1*-specific coding sequences up to 2 nucleotides upstream of the stop codon were deleted. b, bases. (B) Partial nucleotide sequence of the *ADH1* 3'-terminal region. The translational stop codon is boxed. The boundaries of the various *ADH1* deletion clones are indicated. Nucleotide numbering corresponds to reference 3.

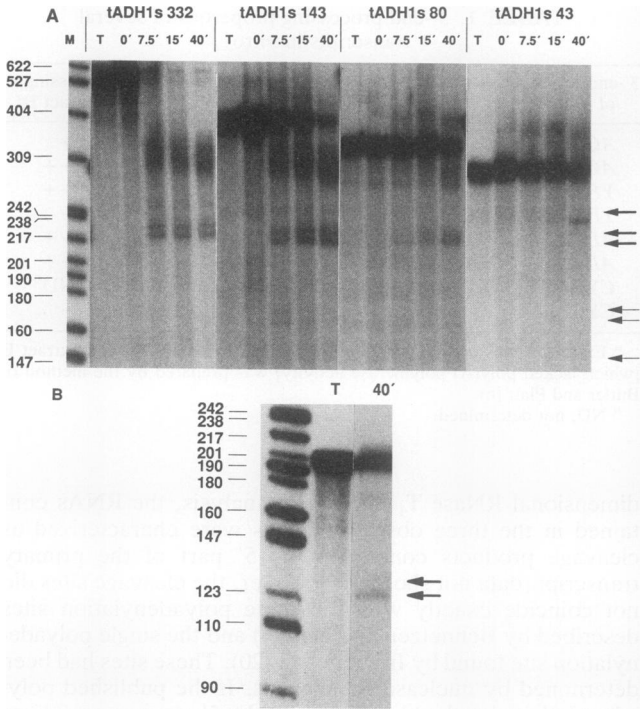


FIG. 3. Polyacrylamide gel analysis of the products formed during the in vitro 3'-end-processing reaction of the *ADH1* deletion transcripts shown in Fig. 1. (A) Processing of the transcripts derived from the deletion clones contained in the pGEM2 vector. The labeled precursor RNA (T) was incubated in a polyadenylation-proficient yeast whole-cell extract for the time indicated at top of each lane (', minutes). The thick arrows point to the three 5'-cleavage products common to all transcripts if they contain the corresponding polyadenylation sites. The thin arrows mark the corresponding 3'-cleavage products of tADH1s143. M, molecular weight marker (pBR322 DNA digested with *Hpa*II). (B) Processing products of the transcript derived from the clone tADH1s80Δc contained in the pBluescript vector when incubated in a polyadenylation-deficient extract. The arrows point to the two 5'-cleavage products. T, precursor RNA. Molecular weight marker, pBR322 DNA digested with *Hpa*II.

they were able to direct 3'-end formation in vivo when contained in a truncated *ACT1* transcription unit (pSH100, Fig. 5A). The oligodeoxynucleotides used for probing the blots were designed to help in distinguishing between RNA molecules ending in the inserted region (O1, Fig. 4A) and readthrough products (O2 and O3, Fig. 4B and C). The oligodeoxynucleotides also hybridized with the endogenous *ACT1* mRNA, providing an internal standard for the amount of RNA loaded in each lane.

DNA fragments which contained at least the two upstream polyadenylation sites (pSH101 and pSH102) very efficiently triggered 3'-end formation, and no readthrough products were detected (Fig. 4A to C). In contrast, the insert which corresponds to tADH1s43 (construct pSH103 in Fig. 5A) generated a substantial proportion of expected readthrough product(s) terminating at the *ACT1* polyadenylation site(s) (Fig. 4A to C). The corresponding T7 transcript, tADH1s43, had not been processed in vitro, and, as will be shown later, this region does not contain any polyadenylation site generally used in vivo. Surprisingly, two additional poly(A)$^+$ RNA species of around 720 and 570 nucleotides were detected. Since the 720-nucleotide RNA hybridized with the oligonucleotide probe O2 (Fig. 4B), but not with O3 (Fig. 4C), a cryptic polyadenylation site must have been selected in the downstream (coding) region of the truncated *ACT1* gene just in between the two sequences which are complementary to the hybridization probes O2 and O3. The 570-nucleotide RNA detected only by probe O1 (Fig. 4A), but not by probe O2 or O3 (Fig. 4B and C), seems to originate from a cryptic polyadenylation site within the 3' end of the cloned insert. In pSH104 and pSH104a, the 91-bp insert corresponding to tADH1s80Δc (Fig. 5A) directed efficient 3'-end processing when inserted in either orientation (Fig. 4A to C), although the in vitro-synthesized RNAs spanning these regions had been processed only poorly in vitro.

These results demonstrate that efficient 3'-end formation in vivo still took place after deletion of the major (as will be shown below) in vivo polyadenylation site. When all three genuine *ADH1* polyadenylation sites were deleted, most of the corresponding poly(A)$^+$ mRNAs derived from processing at the downstream-located *ACT1* polyadenylation site, some from processing at a cryptic site in the downstream *ACT1* coding region, and some from 3'-end formation in the inserted *ADH1* sequence through the use of a cryptic polyadenylation site.

The 3'-terminal regions of the *ADH1*, *ACT1*, *YPT1*, and *CYC1* genes act as bidirectional 3'-end formation sites in vivo. The transcripts derived from the 3'-terminal regions of the yeast *ACT1*, *ADH1*, *YPT1*, and *CYC1* genes had displayed considerably different properties when tested for their capacity to be processed in vitro (Fig. 2 and data not shown). We wondered whether these differences could also be observed when the corresponding fragments were analyzed in vivo. For that purpose, the 3'-terminal regions of the *ACT1*, *YPT1*, and *CYC1* genes were inserted in both orientations in a truncated yeast *ACT1* transcription unit with the *ADH1* 3'-terminal region located downstream (Fig. 5B). All fragments were able to generate processed RNAs terminating within the inserted regions (Fig. 4D and E) irrespective of their orientation. As a result of the differing distances of the polyadenylation sites from the 5' ends of the inserted fragments, the RNAs display different sizes. Interestingly, the basic constructs pSH50, pSH200, and pSH200a yielded smaller amounts of RNA hybridizing to probe O1. For pSH50, this phenomenon might be explained by the lack of an inserted terminator region. For pSH200 and pSH200a, we can only speculate that the lack of an insertion upstream of the *ADH1* terminator has led to a lower transcriptional activity or lower stability of the corresponding RNAs.

Except for *CYC1* antisense, all inserts worked as efficient 3'-end-processing regions, i.e., no readthrough transcripts containing sequences beyond the inserted 3'-end regions could be detected. In the case of *CYC1* antisense, approximately 10% of the RNA molecules originating from the *ACT1* promoter on the plasmid were not processed within the inserted sequence (1.1-kb band in Fig. 4D and E). The short transcripts which were detected by hybridization with oligonucleotide O3 (Fig. 5B) in the RNA from the *YPT1* antisense and *CYC1* antisense transformants (Fig. 4E) could not originate from the actin promoter because (i) they are too short to span the region from the *ACT1* promoter to the region complementary to oligodeoxynucleotide O3 and (ii) they did not hybridize with oligodeoxynucleotide O1. Therefore cryptic promoters have to be postulated in these two 3'-end regions.

The nonabundant short RNAs which were detectable in the *ACT1* sense, *ACT1* antisense, *CYC1* sense, and *YPT1* sense RNA preparations (Fig. 4E) could either derive from

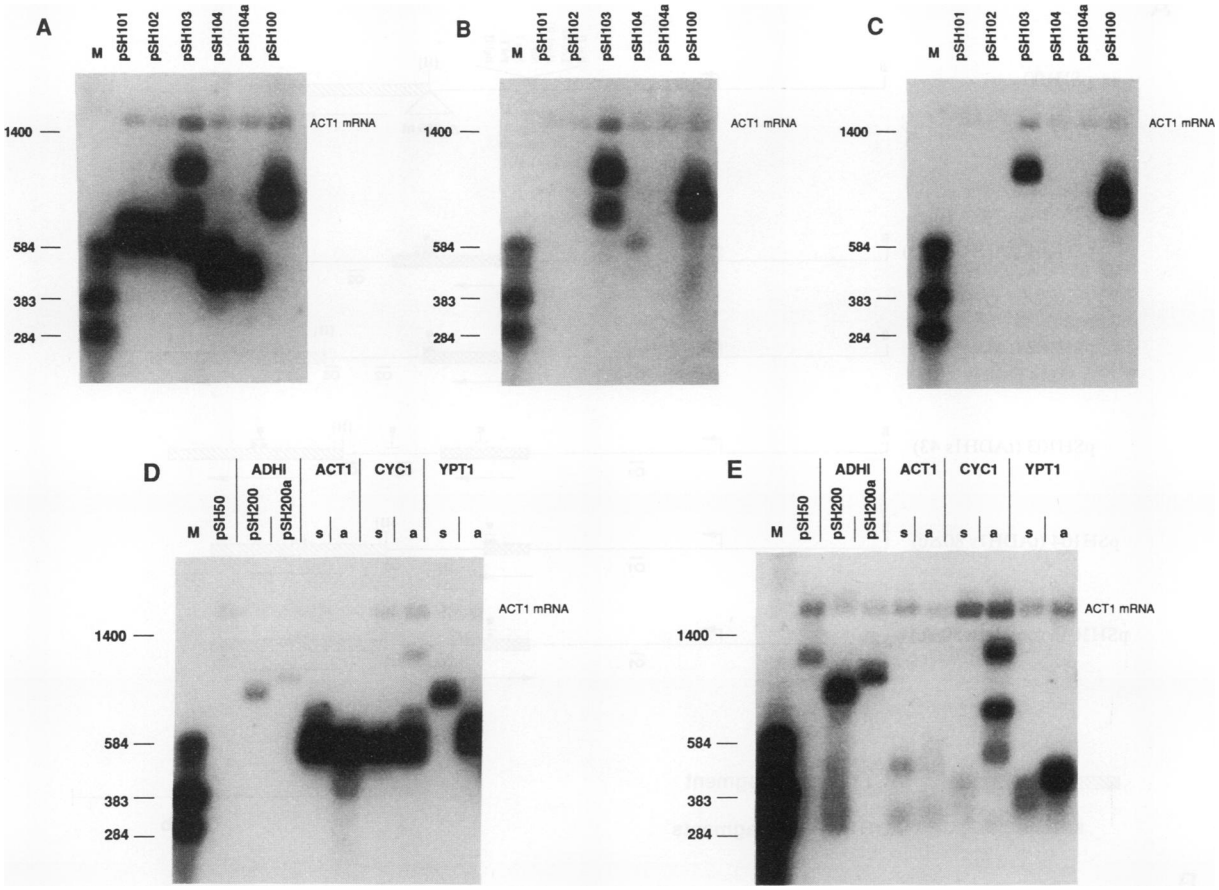


FIG. 4. Northern blot analysis of the constructs shown in Fig. 5. In each lane, 1.5 μg of poly(A)$^+$ RNA isolated from cells transformed with the constructs shown in Fig. 5A (A to C) and in Fig. 5B (D and E) were separated. The blots were hybridized with the oligonucleotides O1 (A and D), O2 (B), and O3 (C and E). M, molecular weight marker (radiolabeled in vitro transcripts with their lengths indicated at the left of the autoradiographs; the 1,400-nucleotide marker band derives from a nonradioactive RNA marker). The orientation of the inserted fragment is indicated by s (sense) and a (antisense). pSH200a corresponds to the construct pSH200 (Fig. 5B) with the *ADH1* 3′ region inserted in the antisense orientation. Numbers on left show size in nucleotides.

weak cryptic promoters or represent the 3′-cleavage products of the processing reactions.

Exact determination of in vivo polyadenylation sites by cDNA sequence analysis. The published 3′ ends of *ADH1* mRNA (3, 20) did not correspond exactly with those generated by the in vitro processing reaction (see above). To clarify this observed discrepancy and also to review and control the published polyadenylation sites of other yeast genes, we determined the exact polyadenylation sites used in vivo for a limited number of yeast genes. Most of the published data had been obtained by nuclease S1 mapping, a technique which does not permit nucleotide-precise assignments, because of the end-nibbling activity of nuclease S1 (12, 31, 43, 44) and its nucleolytic activity in A+T-rich base-paired regions (26).

Two different approaches were used to establish the polyadenylation sites. First, a yeast λZAPII cDNA library was screened with *CYC1*-, *ADH1*-, and *ACT1*-specific oligodeoxynucleotide probes, and the plasmids of hybridizing plaques were prepared after in vivo excision reactions and analyzed. Second, RACE-PCR, a less time-consuming approach, was employed for cDNA synthesis and amplification of poly(A)$^+$ RNA of yeast cells transformed with the constructs shown in Fig. 5B (9). As a sequence-specific upstream primer, oligonucleotide O6 was used which hybridized 10 nucleotides upstream of the *ACT1 Bgl*II site (Fig. 5B), so that any RNA molecule beginning at the *ACT1* promoter and ending in any of the inserted fragments, or in the downstream following terminator region, as well as the endogenous *ACT1* mRNA molecules, should be amplified. The combined results of the cDNA sequence analyses are shown in Fig. 6.

For *CYC1*, three different polyadenylation sites were found in 11 independent cDNA clones (Fig. 6A). In 9 of 11 clones, the polyadenylation site coincided exactly with the polyadenylation and in vitro processing site previously determined (6).

In 17 independent clones representing *ADH1* cDNAs, three different sites were detected (Fig. 6B). These three sites correspond perfectly to the observed in vitro processing sites. The two upstream sites surround closely the second putative polyadenylation site located by nuclease S1 analysis (3). The most downstream site coincides well with the most downstream site previously described (3, 20). None of the 17 cDNA clones contained an insert that would end at or near the proposed most upstream polyadenylation site, which is located around 24 nucleotides downstream of the *ADH1* translation termination codon (3).

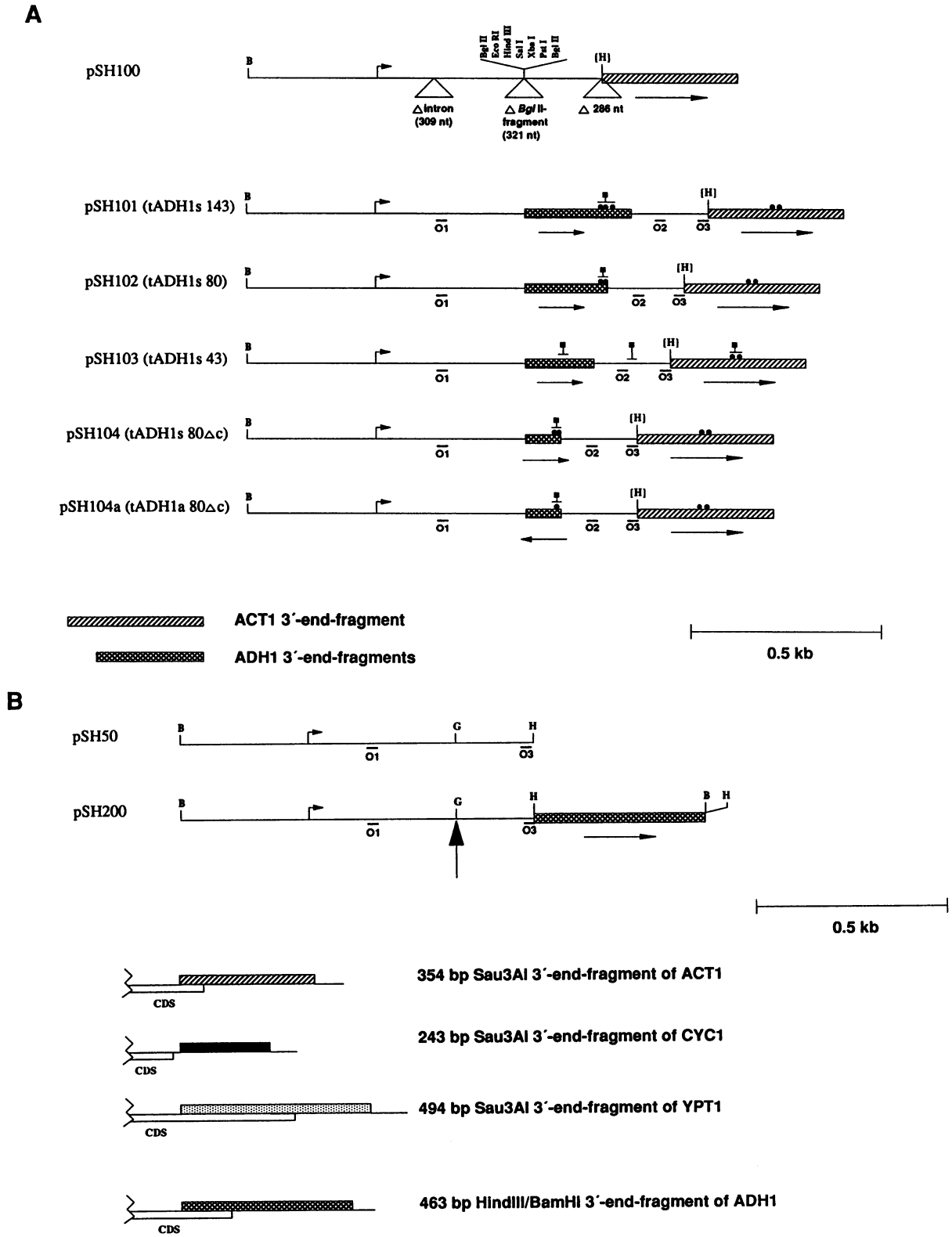


FIG. 5. Schematic presentation of the constructs which were tested for their ability to direct 3'-end formation in vivo. All constructs are based on a truncated *ACT1* transcription unit and are contained in the yeast-*E. coli* shuttle vector YEp351. (A) Constructs for the in vivo analyses of *ADH1* deletion fragments corresponding to the transcripts shown in Fig. 1. On top of the figure the basic construct is shown; deleted *ACT1* sequences are indicated by triangles. For convenient cloning of the *ADH1* deletion fragments (cross-hatched bars), a synthetic polylinker composed of O7 and O8 was first inserted in the single *Bgl*II site. The small arrow indicates the transcriptional start site. Arrows below the bars indicate the sense or antisense orientation of the inserted fragments. O1, O2, and O3 show the positions to which the

Seven different polyadenylation sites were determined in 12 independent *YPT1* cDNA clones (Fig. 6C). They show a good agreement with the published data (10).

Strikingly, for *ACT1*, 12 different polyadenylation sites were found in 46 independent clones (Fig. 6D). Only two sites had been described previously (11); both of them coincide with the most abundant cDNA clones (14 of 46 and 11 of 46, respectively). The most downstream site is located 182 nucleotides away from the second polyadenylation site described previously (11).

Figure 6E to H illustrates the polyadenylation sites which were used when the *CYC1*, *ADH1*, *YPT1*, and *ACT1* 3'-terminal regions had been inserted in their antisense orientation. These results again demonstrate that the polyadenylation sites can be distributed over a wide sequence range. Note that in the case of *CYC1* antisense, the most downstream polyadenylation site identified is located 5 to 6 nucleotides outside the *CYC1* insert (Fig. 6E).

Since in most cases (119 of 127 independent clones, i.e., 93.7%) polyadenylation occurred at sites with one or more adenosine residues located at these positions, it was not possible to determine the exact 3'-end-processing sites. As a result of this ambiguity, it was also not possible to generate from the compilation of determined polyadenylation sites a unique common sequence motif which might be connected with polyadenylation site selection. On the other hand, one might speculate that exactly this small stretch of adenosine residues is a contributing factor for polyadenylation site selection. This assumption is further supported by the fact that all major polyadenylation sites found (with the exception of *ACT1* antisense) comprise two or more adenosine residues.

Frohman et al. (9) had measured 34 nucleotides as the maximum length of the poly(A) tail in the cDNA clones derived from RACE-PCR, although no limiting precautions had been taken. Therefore, it was astonishing that the poly(A) tails of the RACE-PCR cDNA clones studied here reached lengths of up to 80 nucleotides with a mean value of 27 nucleotides (data not shown). Furthermore, in the case of *CYC1* antisense in a sampling of 24 investigated recombinants, we did not obtain any clone representing a read-through event. One might presume that the *CYC1*-derived sequences could have folded in a secondary structure which blocked reverse transcription by avian myeloblastosis virus reverse transcriptase.

Mutational analysis of *ADH1* polyadenylation sites. The experiments described so far have shown that the chosen transcripts displayed an unexpected heterogeneity in polyadenylation site selection, with the consequence that it was not possible to formulate a consensus motif for the 3'-end-processing site.

On the other hand, it was obvious that 3'-end processing of the *ADH1* transcript preferentially occurs in the sequence context PyAAA. To determine whether the adenosine residues in this motif are of any functional importance, we did a series of mutagenesis experiments. The *ADH1* transcript was chosen because (i) the *ADH1* transcripts were processed very efficiently in vitro, and (ii) the in vivo and in vitro polyadenylation sites correlated well for all three established sites.

In Fig. 7A, the different mutations are shown (boxed sequences) which either replaced the adenosine stretches by pyrimidines or which generated a prolonged adenosine stretch in the *ADH1* polyadenylation region. In vitro transcripts (derivatives of tADH1s143, Fig. 1) containing these mutations were subjected to in vitro processing reactions with a polyadenylation-deficient extract, with the results shown in Fig. 7B. The reaction products obtained from the wild-type construct accounting for cleavages at sites 2, 4, and 5 have already been shown in Fig. 3. Changing any of the three adenosine stretches into pyrimidine stretches almost completely abolished the use of the respective processing site in vitro (Fig. 7B). In the transcript tADH1s143-Mut2, which contains an alteration in the adenosine stretch of site 4, the mutation additionally abolished processing at site 2. The mutation of the third adenosine stretch (tADH1s-Mut3) resulted in a downstream shift of the cleavage site to the sequence CGAAAA (site 1).

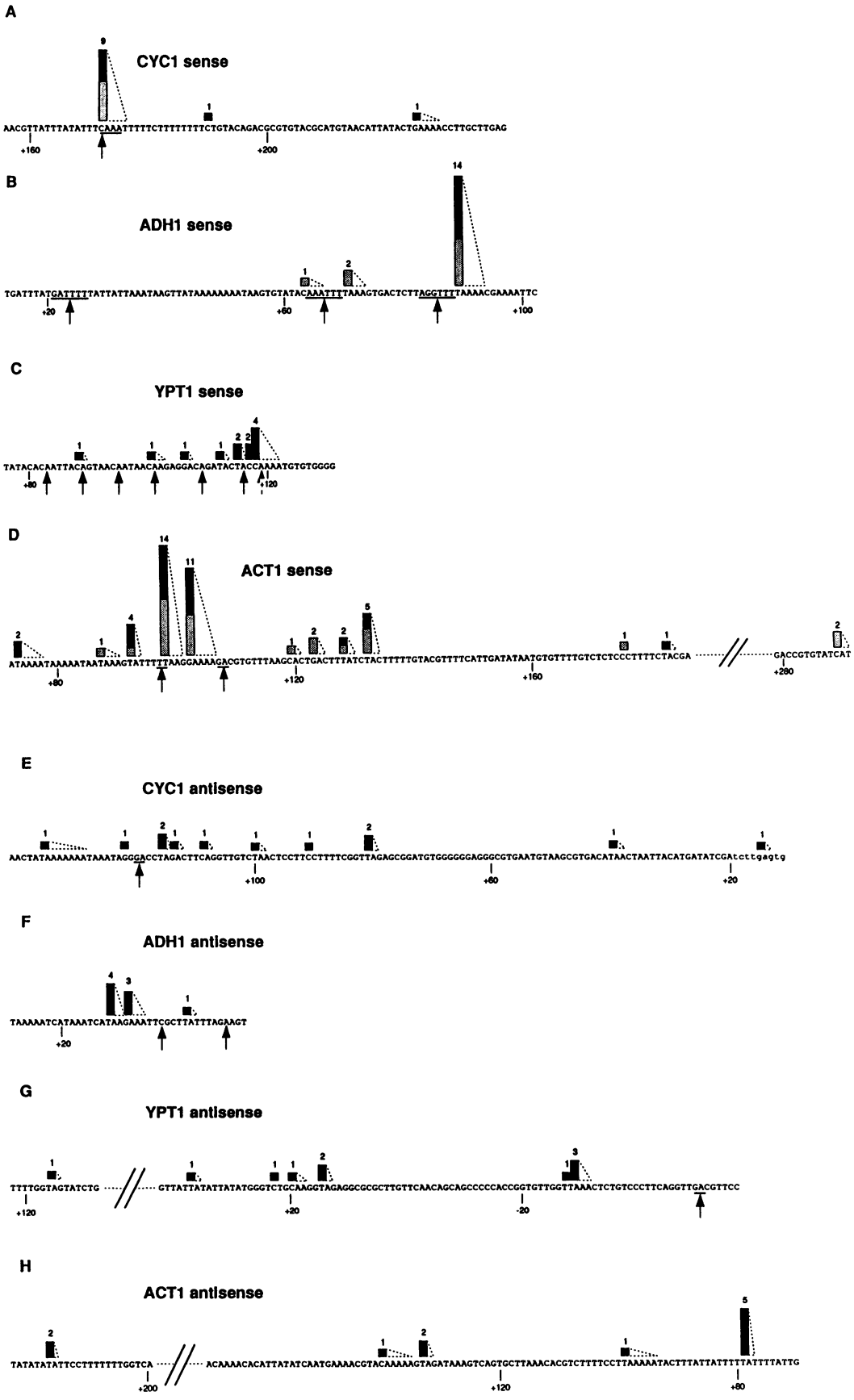
In the mutant construct tADH1s143-Mut4, the pyrimidine residues separating rows 1 and 2 adenosine were changed into adenosines, so that a contiguous stretch of 10 adenosine residues was created. This mutation resulted in a new cleavage pattern. Processing at the first site still occurred, but the observed cleavage products were scattered over a rather broad range, which suggests that cleavage took place at several positions within this adenosine stretch. Furthermore, this mutation, in contrast to tADH1s143-Mut2, did not inhibit the use of the third in vivo-established polyadenylation site.

The mutagenesis of all three genuine polyadenylation sites resulted in a completely new 3'-end-processing pattern. Four 5'-cleavage products were generated, with the most prominent product ending at the GA sequence following the second adenosine stretch in the wild-type sequence (site 3 in Fig. 7). A faint signal indicating a cleavage reaction at this site had also been detected in the reaction products of all other transcripts except for tADH1s143-Mut4. Strikingly, sites 5 and 4 were still used, although the corresponding sequences had been mutated and were rarely used at all in the constructs in which they had been the sole mutation. In addition, processing at a new site (site 6) was observed.

None of the mutations impaired the in vitro polyadenylation of the 5'-cleavage products as assayed by doing the processing reactions in a polyadenylation-proficient extract (data not shown).

The effects of the mutations were also assayed in vivo. The mutations were introduced in construct pSH101 (Fig. 5A), and poly(A)$^+$ RNA from the respective transformants (pSH101-Mut1 to pSH101-Mut5) was analyzed. Both RACE-PCR followed by cDNA sequence analysis and nuclease S1 mapping were used for the determination of the 3' ends formed in vivo.

corresponding oligodeoxynucleotides hybridize. Dots indicate the in vitro-determined cleavage sites; square-headed pins indicate the in vivo-generated mRNA 3' ends determined by Northern blot analysis (Fig. 4A to C). B, *Bam*HI; [H], *Hin*dIII site which was destroyed during the cloning procedures. (B) Constructs for the in vivo analyses of the 3'-terminal regions of the yeast *ACT1*, *CYC1*, and *YPT1* genes. The respective 3'-terminal regions, which are shown as shaded bars with their chromosomal locations at the bottom of the figure, were inserted in both orientations in the *Bgl*II site of pSH200. The small arrow indicates the transcriptional start site. O1 and O3 show the positions to which the corresponding oligodeoxynucleotides hybridize. B, *Bam*HI; G, *Bgl*II; H, *Hin*dIII; CDS, coding sequence. The thick arrow indicates the insertion site of the various 3'-terminal fragments.


A
CYC1 sense
AACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAG
+160
+200
B
ADH1 sense
TGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTC
+20
+60
+100
C
YPT1 sense
TATACACAATTACAGTAACAATAACAAGAGGACAGATACTACCAAAATGTGTGGGG
+80
+120
D
ACT1 sense
ATAAAATAAAAATAATAAAGTATTTTTAAGGAAAAGACGTGTTTAAGCACTGACTTTATCTACTTTTTGTACGTTTTCATTGATATAATGTGTTTTGTCTCTCCCTTTTCTACGA
GACCGTGTATCAT
+80
+120
+160
+280
E
CYC1 antisense
AACTATAAAAAAATAAATAGGGACCTAGACTTCAGGTTGTCTAACTCCTTCCTTTTCGGTTAGAGCGGATGTGGGGGGAGGGCGTGAATGTAAGCGTGACATAACTAATTACATGATATCGAtcttgagtg
+100
+60
+20
F
ADH1 antisense
TAAAAATCATAAATCATAAGAAATTCGCTTATTTAGAAGT
+20
G
YPT1 antisense
TTTTGGTAGTATCTG
GTTATTATATTATATGGGTCTGCAAGGTAGAGGCGCGCTTGTTCAACAGCAGCCCCCACCGGTGTTGGTTAAACTCTGTCCCTTCAGGTTGACGTTCC
+120
+20
-20
H
ACT1 antisense
TATATATATTCCTTTTTTTGGTCA
ACAAAACACATTATATCAATGAAAACGTACAAAAAGTAGATAAAGTCAGTGCTTAAACACGTCTTTTCCTTAAAAATACTTTATTATTTTTTATTTTATTG
+200
+120
+80

The results of the cDNA sequence analysis (numbered bars in Fig. 7A) revealed, besides the occurrence of additional downstream polyadenylation sites in the wild-type construct pSH101, that all mutations had an effect on polyadenylation site selection in vivo. The mutations Mut1, Mut2, and Mut3 (boxed sequences in Fig. 7A) strongly or even completely (Mut1 and Mut2) inhibited processing at the respective sites. On the other hand, the mutation Mut2 did not have such a drastic effect on the use of the downstream-located major polyadenylation site as had been observed in vitro (Fig. 7B). The three mutations in construct Mut5 caused a similarly complete change of the polyadenylation pattern as had been observed in vitro.

Although the cDNA sequence data provided a very precise picture of the polyadenylation sites used, the limited number of analyzed cDNAs could not be used to reach a statistically relevant conclusion about the frequency with which a specific polyadenylation site was chosen, so that in addition S1 nuclease mapping was performed (Fig. 7C). When the reaction was done at 23°C, the data obtained for the wild-type construct (lane 15 in Fig. 7C) were in good agreement with the previous S1 mapping data for *ADH1* mRNA (3, 20). However, at first it was difficult to correlate the S1 results with our cDNA sequence data. The strong signal in lane 15 corresponds to a protected DNA fragment terminating at the two G residues designated by an arrow in Fig. 7A, wild type. This fragment is 8 or 9 nucleotides shorter than an expected fragment, if it were protected by an RNA polyadenylated at site 2. In that case, the region of complementarity would reach up to the T residue complementary to the last adenosine residue in the sequence GGUUUUAAAA. The protected fragment due to processing at site 5 is also slightly too short, and products resulting from polyadenylation at site 4 could not be detected at all.

These discrepancies most likely derive from the well known tendency of nuclease S1 to end nibble into legitimate duplexes (12, 31, 43, 44). This is especially true for A+T-rich sequences, which are characteristic for the 3′-end regions of yeast genes. This fact led us to hypothesize that nuclease S1 under standard conditions degrades the hybrids in the terminal A · T and U · A base-paired regions up to the next G · C base pair(s). This would also explain why we could not detect a fragment protected by RNAs polyadenylated at site 4: the corresponding hybrids were apparently degraded down to the next G · C base pair, which is located at site 5. In Mut1, a cytosine residue had been introduced 2 nucleotides downstream of this base pair, and consequently in the S1 analysis, a protected fragment appeared which is 2 nucleotides longer (Fig. 7C, lane 16). Following this argumentation, the S1 data do corroborate the results obtained by cDNA sequence analysis: in Mut3 and Mut5, the use of the major polyadenylation site 2 is abolished, and in Mut2, it is slightly reduced (Fig. 7C, lanes 18, 20, and 17, respectively). Furthermore, the downstream shift from site 2 to site 1 in Mut3 (Fig. 7C, lane 18) and the use of sites between sites 4 and 2 in Mut2 and Mut5 (Fig. 7C, lanes 17 and 20, respectively) could be observed.

To support our interpretation, we did the S1 mapping experiments under less stringent conditions, i.e., with a reaction temperature of 16°C (Fig. 7C, lanes 9 through 14), at which nuclease S1 is still active (38). First, the intensity of the signals in lanes 9 to 14 increased when compared with lanes 15 to 20, presumably because the label at the base-paired 3′ end of the probe was not removed by nuclease S1 at 16°C to the same extent as at 23°C. Second, at the lower temperature, the lengths of the protected fragments increased in most cases by some nucleotides. This is probably not due only to incomplete digestion of noncomplementary parts of the probe, but rather reflects the presence of the postulated A · T and U · A base-paired regions at the ends of the RNAs. This becomes evident when comparing, for example, the signals of DNA molecules protected by the wild type- and Mut2-derived RNAs, migrating at position 5 (Fig. 7C, lanes 9 and 15 and 11 and 17, respectively). Extended protected fragments could be detected in lane 9, indicating polyadenylation at site 4, while such fragments are missing in lane 11 (Mut2). Since in the latter case site 4 had been mutagenized, we did not expect these fragments, in accordance with our interpretation. But if these extended fragments generated at the lower temperature were the result of incomplete digestion, they should have been present also in lane 11, which is not the case.

DISCUSSION

The mechanism by which mRNA 3′ ends are formed in the yeast *S. cerevisiae* is still elusive. There have been findings which support a mechanism similar to that in prokaryotes, in which the 3′ ends are generated by transcription termination (34). On the other hand, it was shown that synthetic mRNA precursors can be processed in a cell-free yeast in vitro system, providing evidence that the mechanism of mRNA 3′-end formation is similar in higher eukaryotes and yeasts (6). Still, there are some clear differences. Several studies have shown that in *S. cerevisiae*, transcription does not proceed far beyond the polyadenylation site (33, 34, 37) and that yeast genes do not possess strictly conserved consensus signals such as AATAAA in their 3′ regions which determine the site of cleavage and polyadenylation in higher eukaryotes. Extensive analysis of the *CYC1* 3′-end region has stressed the importance of a 38-bp fragment in the process of transcription termination and/or pre-mRNA processing (6, 35, 36, 50). However, the exact definition of specific sequence signals responsible for the in vitro processing reaction had not been achieved.

The primary aim of this study was to provide a clearer definition and characterization of the site at which the actual 3′-end-processing reaction occurs. Our results, especially those obtained through the deletion experiments on the *ADH1* transcript, demonstrate that the sequences responsi-

FIG. 6. Schematic presentation of the distribution of the various 3′ ends determined by sequencing cDNA clones derived either from a yeast cDNA library or from cloned RACE-PCR fragments from the RNAs shown in Fig. 4D and E. The bars above the DNA sequence indicate the polyadenylation sites; their heights represent the abundance of the corresponding clone with the actual number of clones given on top of each bar. The dotted triangles attached to the bars reflect the uncertainty in the exact position because of the presence of adenosine residues at these sites. The lightly shaded portions of some bars represent the clones found by screening the λZAPII cDNA library, while the black parts represent cDNA ends determined from RACE-PCR clones. Nucleotide numbering starts with the base following the translational stop codon. Therefore, numbering in the antisense constructs proceeds from the right to the left. Black arrows indicate 3′ ends described in the literature, with the underlined sequences representing uncertainties mentioned in the corresponding publications. Lowercase letters indicate nucleotides flanking the subcloned insert.

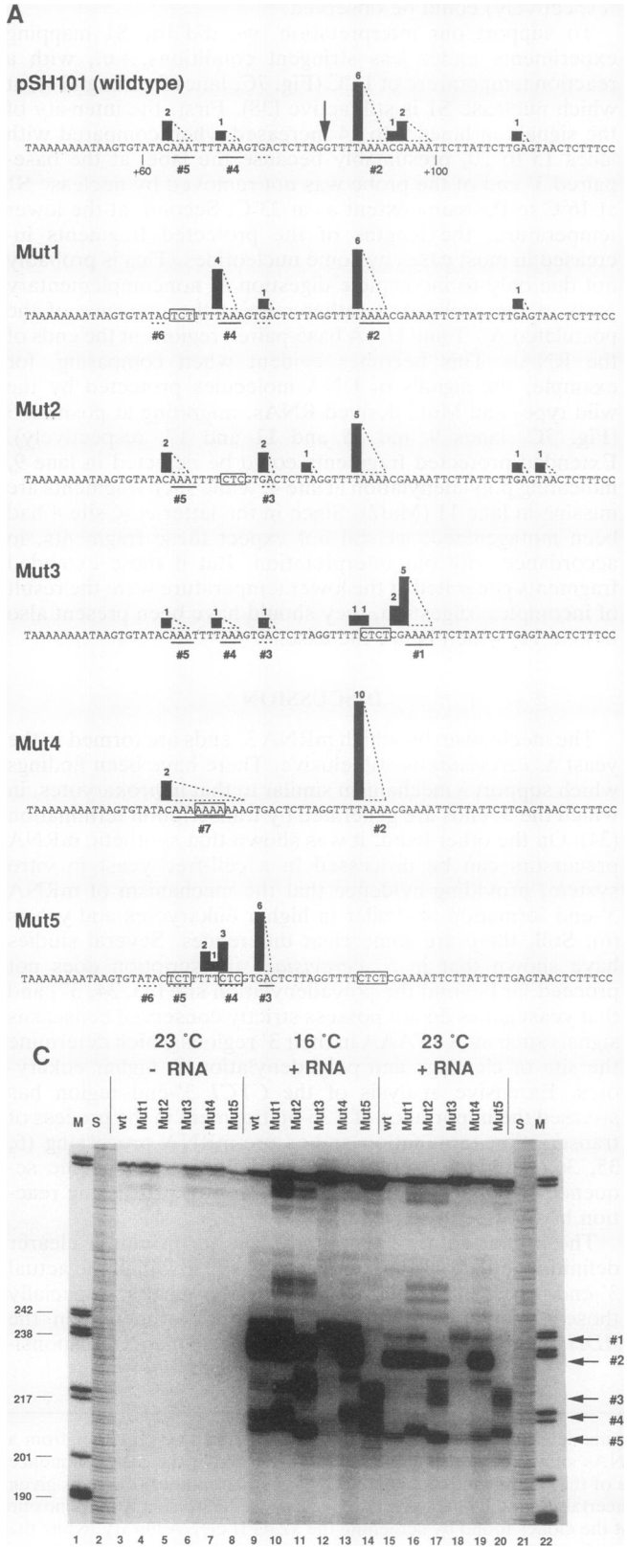


FIG. 7. Mutational analysis of the *ADH1* polyadenylation sites. (A) DNA sequence context of the *ADH1* polyadenylation region and the five corresponding sequences in which the in vitro- and in vivo-determined polyadenylation sites had been mutagenized. The mutated sequences are boxed. Major in vitro cleavage sites are underlined, and minor cleavage sites are indicated by dotted lines. The numbers beneath the lines refer to the products indicated in panel B. The bars above the sequences indicate the distribution and frequency of the polyadenylation sites used in vivo as determined by cDNA sequence analysis. The arrow denotes the position of the major in vivo polyadenylation site as determined by nuclease S1 analysis. (B) Gel electrophoretic analysis of the in vitro processing products of transcripts derived from tADH1s143 (Fig. 1) containing the mutations shown in panel A. The in vitro transcripts were incubated for the indicated time in a polyadenylation-deficient extract, and the products were separated on a 6% polyacrylamide–8 M urea denaturing gel. The different processing products are indicated by the numbers on the right. The numbers correspond to those in panel A. Corresponding 5′- and 3′-cleavage products share the same number. M, molecular weight marker (pBR322 DNA cleaved with *Hpa*II). (C) Nuclease S1 mapping of the mRNA 3′ ends generated in yeast cells transformed with *ADH1* constructs bearing the various mutations. Nuclease S1 digestion was performed either without RNA at 23°C (lanes 3 to 8) or with RNA at 16°C (lanes 9 to 14) or 23°C (lanes 15 to 20). The reaction products were separated on a 6% polyacrylamide–8 M urea gel. Numbered arrows indicate protected fragments due to polyadenylation at the corresponding sites. S, sequencing ladder; M, molecular weight marker (pBR322 DNA cleaved with *Hpa*II).

ble for 3'-end formation are not confined to a very short segment and that sequences both upstream and downstream of the polyadenylation site are involved in this recognition process. Apparently, these regions seem to confer a general signal for 3'-end processing, while sequence motifs closer to the polyadenylation site act as targeting signals for performing the cleavage reaction. We assume that these motifs are composed of two types of signals which are necessary for the efficient processing of a yeast mRNA precursor. The first signal is an upstream element such as the previously proposed sequence motifs UUUUUAUA (15), UAG. . .UA(U) GU. . .UUU (50), UAUAUA (36), UAUCUA (36), or minor deviations thereof, in a generally A+U-rich environment, which announce a processing event soon downstream. However, these signals are not sufficient to direct the choice of all the actual polyadenylation sites used. A second signal is encoded in the cleavage sites themselves. Our studies suggest that sequence motifs comprising a pyrimidine followed by one or more adenosines are used preferentially. The in vivo results having led to this conclusion cannot be dismissed simply as PCR or reverse transcription artifacts in our experiments arising from incorrect priming. In this case, no adenosine stretches significantly longer than the oligo(dT) moiety of the primer used for first-strand cDNA synthesis could have been found. There is also no obvious reason why some adenosine stretches should facilitate incorrect priming, while other adenosine stretches located in the close vicinity should not. For example, in the primary transcript of the *ADH1* gene, there is a contiguous stretch of eight adenosines 10 nucleotides upstream of the first determined polyadenylation site (Fig. 6), but no cDNA clones ending at this site were found.

The importance of these adenosine stretches for polyadenylation site selection has been substantiated by mutagenesis of these sequence elements in the 3'-end region of the *ADH1* gene. In general, the replacement of these A stretches by pyrimidines led to drastic reductions of the processing efficiency at these specific sites both in vitro and in vivo. In the triple mutant Mut5, the 3' ends of the RNAs clustered around the mutated sites 4 and 5, and the formerly inefficient cleavage site 3 was raised in the hierachy and used preferentially. Furthermore, the downstream-located sites 2 and 1 were completely inactivated, in vivo as well as in vitro. This suggests that some sort of signal necessary for the use of these latter two sites had been destroyed. The in vitro data of Mut2 also suggested an inactivating effect on site 2, which, however, was not confirmed by the in vivo data. This discrepancy could be explained by the different efficiencies with which the various sites were used in vitro and in vivo, although the reasons for the different processing efficiencies remain obscure. In vitro processing at a functional site 2 (the major site determined in vivo) always yielded a rather weak signal. If an upstream mutation was able to act inhibitory on the cleavage at site 2, the complete disappearance of the corresponding signal in vitro could be anticipated.

In general, these results resemble those which were obtained when the simian virus 40 late mRNA polyadenylation site, which is also characterized by a pair of adenosines, was mutagenized. Here, it was also not the efficiency but the accuracy of the in vitro processing reaction which was affected (40).

There is evidence that the Py(A)$_n$ motif might be necessary but not sufficient for 3'-end processing, since in the *cyc1-512* mutant in which the upstream signal had been deleted, but the genuine polyadenylation site had been retained, no in vitro processing of a precursor RNA spanning this region was observed (6).

Published data of yeast cDNA sequences provide further support for this Py(A)$_n$ preference as a polyadenylation site; for example, eight different 3' ends were described for the *CBP1* mRNAs, six of them ending with Py(A)$_n$ (27). A detailed investigation of the mRNA 3' ends of the *GAL7* locus both in vitro and in vivo revealed that 3'-end processing occurred in the sequence context Py(A)$_n$ (38). Similar results were described for higher eukaryotes: the poly(A) tail of two viral mRNAs was added to a genome-encoded adenosine residue (30, 41); and computer analysis of 63 transcripts revealed that polyadenylation started at an adenosine residue in 70% of the cases (40).

However, several recent assignments of in vivo-formed yeast mRNA 3' ends seem not to confirm this rule (20). We showed, as an example, that for the *ADH1* transcripts this discrepancy is probably due to double-strand degradation by nuclease S1, which was used for 3'-end mapping in the mentioned study. For *GAL7*, a similar contradiction had been observed between an RNase protection and a cDNA sequence analysis experiment (1, 38). Therefore, nuclease mapping studies should be considered as useful for conclusions about frequencies of specific RNA species and for rough estimates of the ends of these molecules, but the precise mapping of polyadenylation sites may require cDNA sequence analysis.

All the DNA fragments investigated in this study directed efficient 3'-end processing in vivo (except *CYC1* antisense). The corresponding transcripts, however, were processed with considerably different efficiencies in vitro. Butler et al. (7) speculated that the presence of an unknown signal in the coding region of a given gene might have an influence on the in vitro processing efficiency. However, although the *ACT1*s transcript contained 58 nucleotides of coding sequence, it was barely processed at all in vitro, so that additional signals outside of the investigated regions would have to be postulated. Alternative explanations could be the lack or low abundance of a specificity factor(s) required for efficient processing of these transcripts in the extracts, or secondary structure of the RNAs which blocked access of the processing machinery. Determination of the mRNA 3' ends formed in vivo furthermore revealed that the polyadenylation sites can be distributed more or less equally over a broad sequence range, and it is likely that even more sites would have been found if more cDNA clones had been analyzed. In vitro, this broad distribution would have led to a complex pattern of 5'- and 3'-cleavage products, and if all sites were processed with equal and low efficiencies, the detection of these products would have been nearly impossible. Similar discrepancies between in vivo and in vitro data concerning 3'-end formation of yeast mRNAs had already been described previously; e.g., fragments of the *ADH2* and *ACP2* genes directed 3'-end formation very efficiently in vivo, but transcripts spanning these regions were processed only poorly in vitro (19).

Hence, there does not seem to be a strict correlation of the in vitro and in vivo efficiencies of the polyadenylation sites, although it seems obvious, at least for the transcripts in sense orientation, that the major sites used in vivo are also targets for the in vitro processing reactions. However, the cleavage sites observed in vitro do not necessarily represent major sites used in vivo (e.g., *ADH1* sense, *ACT1* antisense). Further discrepancies of the in vivo and in vitro results became evident through the investigation of the *ADH1* deletion clones. Whereas the in vitro processing efficiencies

decreased with decreasing lengths of the precursors, the same fragments led to efficient 3′-end formation in vivo as long as genuine polyadenylation sites were contained in these fragments. The deletion of all three mapped genuine polyadenylation sites resulted in the synthesis of readthrough transcripts, but the inserted fragment still had some residual 3′-end formation capacity as indicated by the observed use of cryptic polyadenylation sites. This 3′-end formation capacity might result from the sequence motif TTTTTAT 23 nucleotides downstream of the *ADH1* gene translational stop codon, which is still present in the respective deletion clone. The corresponding in vitro transcript tADH1s43 was not cleaved in the *ADH1*-derived sequences. Still, there might have been a very low processing efficiency, generating products which were not distinguishable from degradation products. The observed 240-nucleotide processing product in Fig. 3A which is generated by cleavage in vector-derived sequences may correlate to the in vivo readthrough products ending in the following *ACT1* coding region. This suggests the existence of a signal in the truncated fragment which is responsible for 3′-end formation some 50 to 70 nucleotides downstream of the 3′ border of the insert and which exerts this effect when the genuine polyadenylation sites are deleted.

It has recently been shown that precursor RNAs spanning the 3′-end region of the *Schizosaccharomyces pombe ura4*$^+$ gene can be processed efficiently in an *S. cerevisiae* whole-cell extract (18). The mRNA 3′ ends were mapped to the sequence . . .GAUAAACA. . ., with the last template-derived nucleotide indicated. This site, as well as the polyadenylation sites of the *ypt1*$^+$, *ypt3*$^+$, and *tfIID*$^+$ genes (17, 29), contains a stretch of adenosines, suggesting that this characteristic is also important for 3′-end formation in *Schizosaccharomyces pombe* pre-mRNAs.

ACKNOWLEDGMENTS

We thank C. Schindewolf, J. Hani, and J. S. Elce for helpful discussions and critical reading of the manuscript and E. Lindner for providing the yeast cDNA library.

This work was supported by grant BCT0372-8 from the Bundesministerium für Forschung und Technologie and by donations from Hoechst AG, Wacker GmbH, and Boehringer Mannheim GmbH.

REFERENCES

1. **Abe, A., Y. Hirakoa, and T. Fukasawa.** 1990. Signal sequence for generation of mRNA 3′ end in the *Saccharomyces cerevisiae GAL7* gene. EMBO J. **9**:3691–3697.
2. **Ammerer, G.** 1983. Expression of genes in yeast using the ADCI promoter. Methods Enzymol. **101**:192–201.
3. **Bennetzen, J. L., and B. D. Hall.** 1982. The primary structure of the *Saccharomyces cerevisiae* gene for alcohol dehydrogenase I. J. Biol. Chem. **257**:3018–3025.
4. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
5. **Bullock, W. O., J. M. Fernandez, and J. M. Short.** 1987. XL1-Blue: a high efficiency plasmid transforming recA *Escherichia coli* strain with beta-galactosidase selection. BioTechniques **5**:376–378.
6. **Butler, J. S., and T. Platt.** 1988. RNA processing generates the mature 3′ end of yeast *CYC1* messenger RNA *in vitro*. Science **242**:1270–1274.
7. **Butler, J. S., P. P. Sadhale, and T. Platt.** 1990. RNA processing in vitro produces mature 3′ ends of a variety of *Saccharomyces cerevisiae* mRNAs. Mol. Cell. Biol. **10**:2599–2605.
8. **Chen, E. Y., and P. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**:165–170.
9. **Frohman, M. A., M. K. Dush, and G. R. Martin.** 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85**:8998–9002.
10. **Gallwitz, D., C. Donath, and C. Sander.** 1983. A yeast gene encoding a protein homologous to the human c-has/bas proto-oncogene product. Nature (London) **306**:704–707.
11. **Gallwitz, D., F. Perrin, and R. Seidel.** 1981. The actin gene in yeast *Saccharomyces cerevisiae*: 5′ and 3′ end mapping, flanking and putative regulatory sequences. Nucleic Acids Res. **9**:6339–6350.
12. **Green, M. R., and R. G. Roeder.** 1980. Definition of a novel promoter for the major adenovirus-associated virus mRNA. Cell **22**:231–242.
13. **Groner, B., and S. L. Philips.** 1975. Polyadenylate metabolism in the nuclei and cytoplasm of *Saccharomyces cerevisiae*. J. Biol. Chem. **250**:5640–5646.
14. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
15. **Henikoff, S., and E. H. Cohen.** 1984. Sequences responsible for transcription termination on a gene segment in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **4**:1515–1520.
16. **Hill, J. E., A. M. Myers, T. J. Koerner, and A. Tzagoloff.** 1986. Yeast/E. coli shuttle vectors with multiple unique restriction sites. Yeast **2**:163–167.
17. **Hoffmann, A., M. Horikoshi, C. K. Wang, S. Schroeder, P. A. Weil, and R. G. Roeder.** 1990. Cloning of the *Schizosaccharomyces pombe* TFIID gene reveals a strong conservation of functional domains present in *Saccharomyces cerevisiae* TFIID. Genes Dev. **4**:1141–1148.
18. **Humphrey, T., P. Sadhale, T. Platt, and N. Proudfoot.** 1991. Homologous mRNA 3′ end formation in fission and budding yeast. EMBO J. **10**:3503–3511.
19. **Hyman, L. E., H. S. Seiler, J. Whoriskey, and C. Moore.** 1991. Point mutations upstream of the yeast *ADH2* poly(A) site significantly reduce the efficiency of 3′-end formation. Mol. Cell. Biol. **11**:2004–2012.
20. **Irniger, S., C. M. Egli, and G. H. Braus.** 1991. Different classes of polyadenylation sites in the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. **11**:3060–3069.
21. **Ito, H., Y. Fukuda, K. Murata, and A. Kimura.** 1983. Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153**:163–168.
22. **Köhrer, K., and H. Domdey.** 1991. Preparation of high molecular weight RNA. Methods Enzymol. **194**:398–405.
23. **Kunkel, T. A.** 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA **82**:488–492.
24. **Lin, R.-J., A. J. Newman, S.-C. Cheng, and J. Abelson.** 1985. Yeast mRNA splicing *in vitro*. J. Biol. Chem. **260**:14780–14792.
25. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Maquat, L. E., A. J. Kinniburgh, E. A. Rachmilewitz, and J. Ross.** 1981. Unstable β-globin mRNA in mRNA-deficient β^0 thalassemia. Cell **27**:543–553.
27. **Mayer, S. A., and C. Dieckmann.** 1991. Yeast *CBP1* mRNA 3′ end formation is regulated during the induction of mitochondrial function. Mol. Cell. Biol. **11**:813–821.
28. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
29. **Miyake, S., and M. Yamamoto.** 1990. Identification of *ras*-related, *YPT* family genes in *Schizosaccharomyces pombe*. EMBO J. **9**:1417–1422.
30. **Moore, C. L., H. Skolnik-David, and P. A. Sharp.** 1986. Analysis of RNA cleavage at the adenovirus-2 L3 polyadenylation site. EMBO J. **5**:1929–1938.
31. **Murray, M. G.** 1986. Use of sodium trichloroacetate and mung

bean nuclease to increase sensitivity and precision during transcript mapping. Anal. Biochem. **158**:165–170.
32. **Ng, R., and J. Abelson.** 1980. Isolation and sequence of the gene for actin in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **77**:3912–3916.
33. **Osborne, B. I., and L. Guarente.** 1988. Transcription by RNA polymerase II induces changes of DNA topology in yeast. Genes Dev. **2**:766–772.
34. **Osborne, B. I., and L. Guarente.** 1989. Mutational analysis of a yeast transcriptional terminator. Proc. Natl. Acad. Sci. USA **86**:4097–4101.
35. **Ruohola, H., S. M. Baker, R. Parker, and T. Platt.** 1988. Orientation-dependent function of a short *CYC1* DNA fragment in directing mRNA 3′ end formation in yeast. Proc. Natl. Acad. Sci. USA **85**:5041–5045.
36. **Russo, P., W.-Z. Li, D. M. Hampsey, K. S. Zaret, and R. Sherman.** 1991. Distinct *cis*-acting signals enhance 3′ endpoint formation of *CYC1* mRNA in the yeast *Saccharomyces cerevisiae*. EMBO J. **10**:563–571.
37. **Russo, P., and F. Sherman.** 1989. Transcription terminates near the poly(A) site in the *CYC1* gene of the yeast *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **86**:8348–8352.
38. **Sadhale, P. P., R. Sapolsky, R. W. Davis, J. S. Butler, and T. Platt.** 1991. Polymerase chain reaction mapping of yeast *GAL7* mRNA polyadenylation sites demonstrates that 3′ end processing *in vitro* faithfully reproduces the 3′ ends observed *in vivo*. Nucleic Acids Res. **19**:3683–3688.
39. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
40. **Sheets, M. D., S. C. Ogg, and M. P. Wickens.** 1990. Point mutations in AAUAAA and the poly(A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation *in vitro*. Nucleic Acids Res. **18**:5799–5805.
41. **Sheets, M. D., P. Stephenson, and M. P. Wickens.** 1987. Products of in vitro cleavage and polyadenylation of simian virus 40 late pre-mRNAs. Mol. Cell. Biol. **7**:1518–1529.
42. **Sheets, M. D., and M. Wickens.** 1989. Two phases in the addition of a poly(A) tail. Genes Dev. **3**:1401–1412.
43. **Shenk, T. E., C. Rhodes, P. W. J. Rigby, and P. Berg.** 1975. Biochemical method for mapping mutational alterations in DNA with S1 nuclease: the location of deletions and temperature-sensitive mutations in simian virus 40. Proc. Natl. Acad. Sci. USA **72**:989–993.
44. **Sollner-Webb, B., and R. H. Reeder.** 1979. The nucleotide sequence of the initiation and termination sites for ribosomal RNA transcription in *X. laevis*. Cell **18**:485–499.
45. **Thierack, J. (University of Munich).** 1990. Personal communication.
46. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
47. **Thompson-Jäger, S., and H. Domdey.** 1990. The intron of the yeast actin gene contains the promoter for an antisense RNA. Curr. Genet. **17**:269–273.
48. **Wickens, M.** 1990. How the messenger got its tail: addition of poly(A) in the nucleus. Trends Biochem. Sci. **15**:277–281.
49. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
50. **Zaret, K. S., and F. Sherman.** 1982. DNA sequence required for efficient transcription termination in yeast. Cell **28**:563–573.