MOLECULAR AND CELLULAR BIOLOGY, July 1994, p. 4633–4642
0270-7306/94/$04.00+0

Vol. 14, No. 7

Flexibility and Interchangeability of Polyadenylation Signals in *Saccharomyces cerevisiae*

STEFAN HEIDMANN,† CATHERINE SCHINDEWOLF, GABI STUMPF, AND HORST DOMDEY*

Laboratorium für molekulare Biologie-Genzentrum der Ludwig-Maximilians-Universität München, D-82152 Martinsried, Germany

Received 7 February 1994/Returned for modification 18 March 1994/Accepted 20 April 1994

Various signal motifs have been reported to be essential for proper mRNA 3′-end formation in the yeast *Saccharomyces cerevisiae*. However, none of these motifs has been shown to be sufficient to direct 3′-end processing and/or transcription termination. Therefore, several structural motifs have to act in concert for efficient 3′-end formation. In the region upstream of the three polyadenylation sites of the yeast gene for alcohol dehydrogenase I (*ADH1*), we have identified a hitherto unknown signal sequence contained within the octamer AAAAAAAA. This motif, located 11 nucleotides upstream of the first *ADH1* polyadenylation site, is responsible for the utilization of this site in vitro and in vivo, since mutational alteration drastically reduced 3′-end formation at this position. Insertion of 38 *ADH1*-derived nucleotides encompassing the $(A)_8$ motif into the 3′-end formation-deficient *cyc1-512* deletion mutant restored full processing capacity in vitro. Insertion of the octamer alone did not restore 3′-end formation, although mutation of the $(A)_8$ motif in the functional construct had abolished 3′-end processing activity almost completely. This demonstrates that the sequence AAAAAAAA is a necessary, although not sufficient, signal for efficient mRNA 3′-end formation in *S. cerevisiae*.

mRNA 3′-end formation is an essential processing step during the maturation of a primary transcript towards a functional mRNA molecule (reviewed in reference 32). Abolishing proper 3′-end formation can lead to massive impairment of gene expression (25, 33). In higher eucaryotes, the signals on the precursor RNA governing this process have been well defined. Besides the highly conserved hexanucleotide AAUAAA 10 to 35 nucleotides upstream of the polyadenylation site (27), a less conserved GU- or U-rich downstream element is involved in the process of pre-mRNA cleavage and polyadenylation (7, 21, 22).

In contrast, for RNA polymerase II-transcribed genes of the yeast *Saccharomyces cerevisiae*, it has been proven to be difficult to define signals for mRNA 3′-end formation. The sequence AATAAA is present in only approximately 50% of yeast genes and seems to play no definitive role in this process (14). Alternatively, the sequence motifs TAG...TATGTA, TATATA, TACATA (29), TTTTTATA (10), and $(AT)_9$ (1, 6) have been proposed to play important roles in 3′-end formation. Although these motifs are necessary in the cases investigated, they are probably not of general significance and appear to be insufficient to specify 3′-end formation (1, 16). Several reports have presented evidence that it is not a single short sequence element which is essential for directing 3′-end processing in *S. cerevisiae*. For example, no single 8-bp linker substitution in the 3′-terminal region of the yeast gene coding for iso-1-cytochrome *c* (*CYC1*) could abolish polyadenylation in vivo (26). Investigation of the 3′-terminal regions of the *GAL1*, *GAL7*, and *GAL10* genes highlighted the probably degenerate and redundant nature of 3′-end processing determinants (31). Further, it was shown that the sequences at the polyadenylation sites themselves influence polyadenylation site selection in vitro and in vivo (9). In a recent study, Russo and coworkers (28) performed a detailed mutational analysis of the *CYC1* 3′-terminal region to determine the signals responsible for 3′-end formation in vivo. They proposed that three sequence elements work together: (i) a so-called upstream element (like TATATA, TAG...TATGTA, or TTTTTATA) which can enhance the activity of (ii) a so-called downstream element (like TTAAGAAC or AAGAA) to trigger 3′-end formation at (iii) the polyadenylation sites.

While investigating the 3′-end formation reaction in the yeast alcohol dehydrogenase I gene (*ADH1*), we have identified a new signal element, a stretch of eight consecutive adenosine residues, which we defined as a downstream element according to the classification mentioned above. This element has to be located at a proper distance from the polyadenylation site and is necessary for utilization of the first *ADH1* polyadenylation site. Insertion of *ADH1*-derived sequences at the deletion point of the *cyc1-512* mutation, which has been shown to be largely nonfunctional in 3′-end formation in vivo and in vitro (4, 33), restored processing activity in vitro. Furthermore, we show that this activity of the *ADH1* fragment is dependent on the $(A)_8$ stretch. However, this signal is not sufficient, since insertion of the adenosine octamer alone did not result in restoration of mRNA 3′-end formation in vitro.

MATERIALS AND METHODS

Bacterial and yeast strains and culture conditions. *Escherichia coli* XL 1 Blue {*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac*[F′ *proAB lacI*^{q}ZΔM15 Tn*10*(*tet*)]} (3) was used for cloning procedures. Competent *E. coli* cells were produced as described by Hanahan (8). *E. coli* CJ236 [*dut-1 ung-1 thi-1 relA1*/pCJ105(*cm*)] was the host for production of the uracil-containing single-stranded phagemid DNA required for high-efficiency mutagenesis. *E. coli* was grown on Luria-Bertani medium supplemented with the appropriate antibiotics.

The mRNA for the nuclease S1 and rapid amplification of cDNA ends (RACE)-PCR analyses was isolated from transformed yeast DH484 (*MAT***a** *ade2-1 leu2-3 leu2-112 can1-100*

* Corresponding author. Mailing address: Genzentrum München, Am Klopferspitz 18a, D-82152 Martinsried, Germany. Phone: (0)89-85783941. Fax: (0)89-85783795.

† Present address: Seattle Biomedical Research Institute, Seattle, Wash.

TABLE 1. Oligodeoxynucleotides used for in vitro site-specific mutagenesis

Mutant	Oligonucleotide sequence
HCmut1	5′-TCTTATGATTTATGAAAAAATATATTAAATAAGTTAT-3′
HCmut2	5′-TCTTATGATTTATGACCCGGGATATTAAATAAGTTAT-3′
A8mut1	5′-TAAATAAGTTATAAACAAAATAAGTGTATAC-3′
A8mut3	5′-ATTAAATAAGTTATATGCAAAATAAGTGTATAC-3′
A8mut5	5′-TATTAAATAAGTTATCTGCAGAATAAGTGTATACAA-3′
A8mut7	5′-TATTAAATAAGTTATCTGCAGCTTAAGTGTATACAAAT-3′
DIS+2	5′-AAAAAAATAAGTGTATGTACAAATTTTAAAGT-3′
DIS+4	5′-AAAAAAATAAGTGTATCTGTACAAATTTTAAAGT-3′
DIS+8	5′-AAAAAAATAAGTGTATCTCTCTGTACAAATTTTAAAGT-3′
DIS+16	5′-AAAAAAATAAGTGTATCTCTCTCTCTCTCTGTACAAATTTTAAAGT-3′
cycΔ	5′-AGAAAAATTTGAAATATAAATAGGGACCTAGACTTCAG-3′
cycΔ+8	5′-GAAAAATTTGAAATATAAATATTTTTTTTTGGGACCTAGACTTC-3′
cycΔ+18	5′-GAAAAATTTGAAATATAAATATTTTTTTTTATAACTTATTGGGACCTAGACTTC-3′
cycΔ+28	5′-GAAAAATTTGAAATATAAATATTTTTTTTTATAACTTATTTAATAATAAAGGGACCTAGACTTC-3′
cycΔ+38	5′-GAAAAATTTGAAATATAAATATTTTTTTTTATAACTTATTTAATAATAAAAATCATAAATGGGACCTAGACTTC-3′
cycΔ+38mutA8	5′-GAAAAATTTGAAATATAAATAAGCTGCAGATAACTTATTTAATAATAAAAATCATAAATGGGACCTAGACTTC-3′
cycΔ+38mutHC	5′-ATAACTTATTTAATAGCTGCAGTCATAAATGGGACCT-3′

trp5-48 ura4-11 lys1-1) cells. Yeast whole-cell extracts were prepared from strain EJ101 (*MATα trp1 pro1-126 prb1-112 pep4-3 prc1-126*) in accordance with the protocol of Butler et al. (5). Untransformed yeast cells were grown on yeast extract-peptone-dextrose (YPD) medium. *S. cerevisiae* DH484 was transformed by the lithium acetate procedure (18). Transformants were selected on minimal medium supplemented with all of the necessary constituents except leucine.

Enzymes and chemicals. Restriction endonucleases, T4 DNA ligase, avian myeloblastosis virus reverse transcriptase, and T4 polynucleotide kinase were obtained from Boehringer Mannheim. T7 RNA polymerase was from Stratagene. Human placental RNase inhibitor, [α-^{32}P]dCTP (111 TBq/mmol), and [α-^{32}P]UTP (29.6 TBq/mmol) were from Amersham. *Taq* DNA polymerase was purchased from Cetus Corporation, and S1 nuclease was from Pharmacia.

Plasmid constructions and manipulations. The *ADH1* 3′-terminal region was obtained as a 316-bp *Bgl*II fragment from plasmid pSH101 (9) and cloned into *Bam*HI-cleaved vector pBluescriptIIKS$^-$ (Stratagene) for the generation of in vitro transcripts (see Fig. 1). For the in vivo analyses, wild-type sequences of *ADH1* and derivatives thereof were cloned and mutagenized in the vector pSH101. The *CYC1* terminator region was obtained as a 251-bp *Hin*dIII-*Sma*I fragment from a pGEM2 vector (Promega) containing the *CYC1* 3′-terminal region in its *Bam*HI site (9) and cloned into *Hin*dIII-*Sma*I-cleaved vector pBluescriptIISK$^-$. Transcription with T7 RNA polymerase from this plasmid (pSKCYC) yielded RNAs in the sense orientation.

All sequence alterations were done by site-directed mutagenesis by the method of Kunkel (20). The sequences of the oligonucleotides used are compiled in Table 1.

DNA sequencing. Plasmids containing the cloned RACE-PCR products were sequenced with an Applied Biosystems 373A DNA Sequencer using fluorescence-labeled primers and the thermal cycling protocol recommended by the manufacturer.

In vitro RNA synthesis and processing reactions. In vitro RNA synthesis was performed as described previously (23); in the reactions, 7mG(5′)ppp(5′)G was included in sixfold excess compared with the concentration of GTP so that most of the transcripts were capped at their 5′ ends. For T7 RNA polymerase runoff transcripts, the *ADH1*-derived templates were linearized with *Xho*I and the *CYC1*-derived templates were linearized with *Sma*I. 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 (4). To suppress polyadenylation, the reactions contained cordycepin triphosphate in equimolar amounts in addition to ATP (5).

RNA analysis. Yeast total RNA was prepared by the hot phenol method as described by Köhrer and Domdey (19). Poly(A)$^+$ RNA was isolated with an mRNA purification kit purchased from Pharmacia.

Reverse transcription and amplification of specific polyadenylated mRNAs by the RACE-PCR technique and nuclease S1 analyses were done as described previously (9).

RESULTS

The signal sequence TTTTTAT proposed by Henikoff and Cohen (10) has no functional importance for 3′-end processing of *ADH1* pre-mRNAs. In a previous study, we have shown that cleavage of yeast *ADH1* pre-mRNAs and subsequent polyadenylation occur preferentially at three sites in the Py AAA context (9; Fig. 1). Clearly, these sequences are not sufficient to direct the 3′-end processing reactions on their own, because with a triple mutation of the three *ADH1* polyadenylation sites, 3′-end formation still occurred in vitro and in vivo. Therefore, we searched in the upstream region for sequences with homology to known yeast polyadenylation signals. No perfect match was found with the dipartite sequence TAG... TATGTA nor with the motifs TATATA and TACATA (29). However, the sequence TTTTTAT located 23 nucleotides downstream of the translational stop codon (Fig. 2A) displays a high level of homology to the sequence TTTT TATA, which was shown to be crucial for the transcription termination activity of a *Drosophila* DNA fragment in *S. cerevisiae* (10). By using site-directed mutagenesis, we altered the corresponding nucleotides (i) to the complementary sequence AAAAATA and (ii) to CCCGGGA, thereby increasing the G+C content in this region (Fig. 2A). Synthetic transcripts containing these mutations were processed in vitro

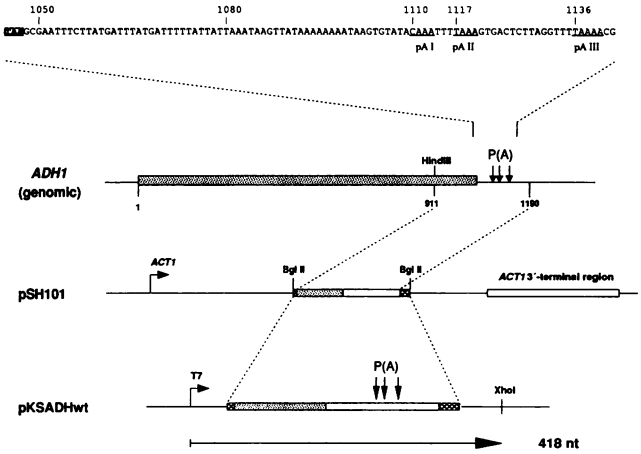


FIG. 1. Schematic presentation of the 3′-terminal region of the *ADH1* gene. The stippled box represents the protein-coding region. The positions of the three *ADH1* polyadenylation sites described [P(A)] are indicated by vertical arrows. The nucleotide sequence from the translational stop codon (boxed) to the three polyadenylation sites (pAI at position 1110, pAII at position 1117, and pAIII at position 1136) is shown on the top. Nucleotide numbering corresponds to that of Bennetzen and Hall (2). The *ADH1*-specific sequences between positions 911 and 1190 are contained in plasmid pSH101 (9), which allows production of mRNAs in yeast cells under the control of the *ACT1* promoter. pKSADHwt is a derivative of pBluescriptIIKS$^-$ containing the *ADH1* 3′-terminal region. The cross-hatched parts represent sequences derived from in vivo construct pSH101 which were retained during the cloning procedures. The in vitro runoff transcript with *Xho*I-cleaved plasmid pKSADHwt as the template is depicted at the bottom. nt, nucleotides.

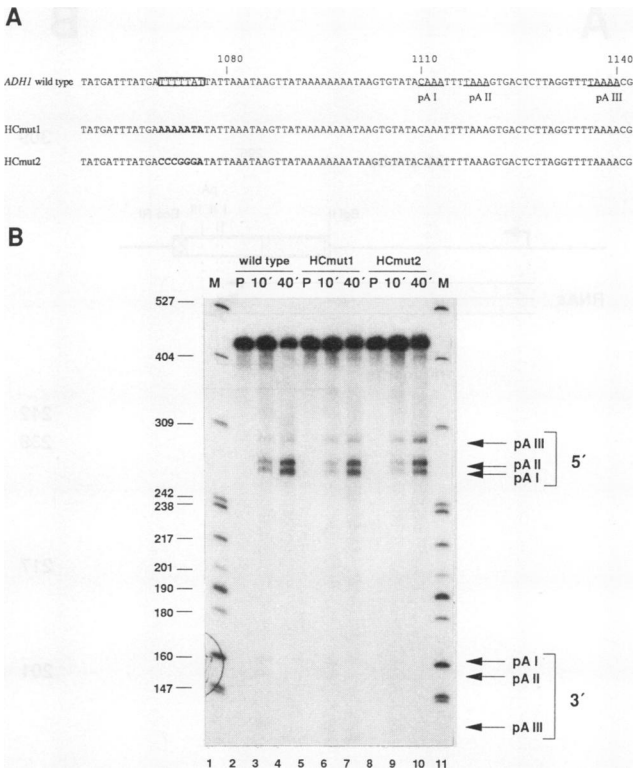
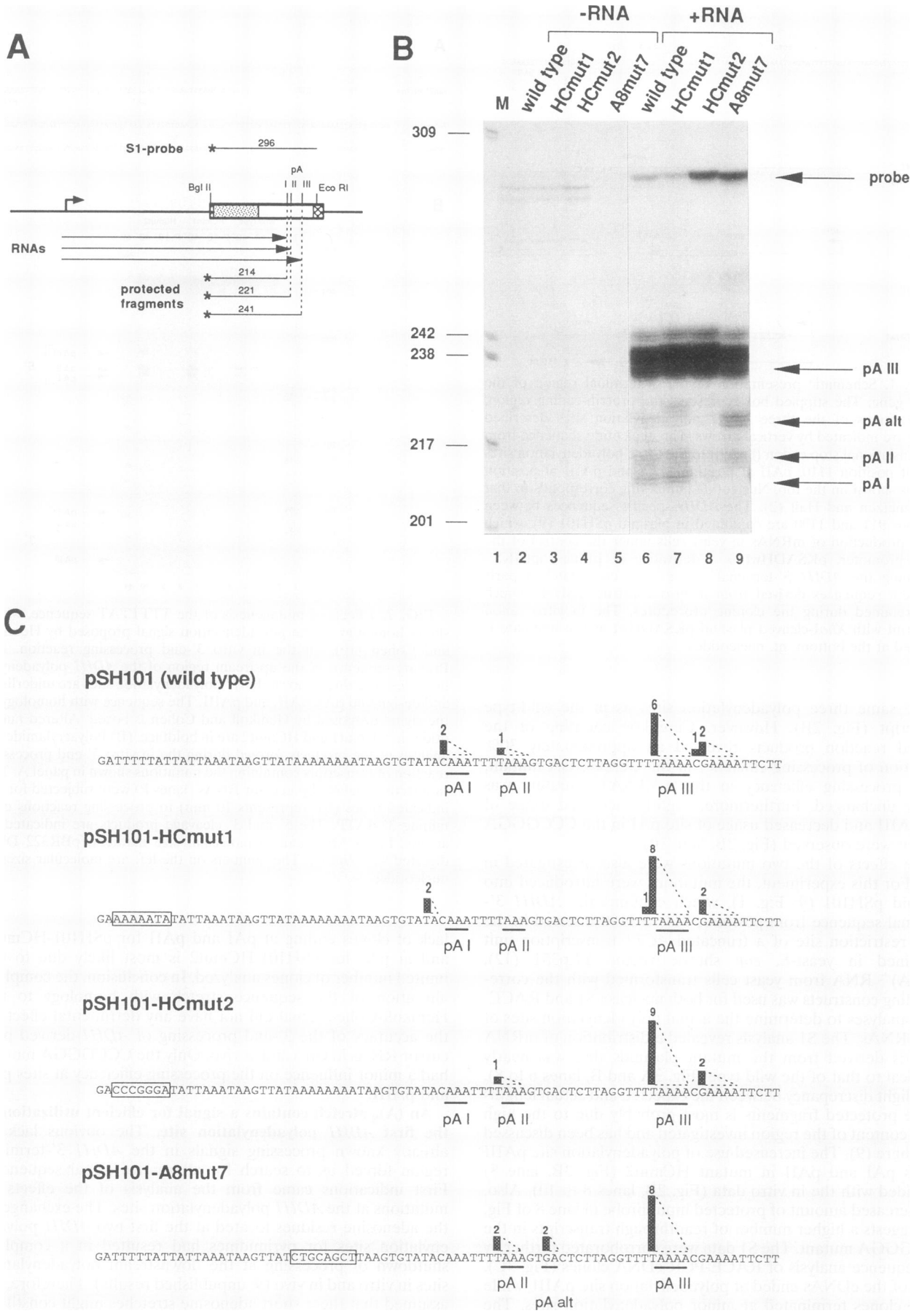


FIG. 2. Effects of mutagenesis of the TTTTTAT sequence, which shows homology to the polyadenylation signal proposed by Henikoff and Cohen (10), on the in vitro 3′-end processing reaction. (A) Primary sequence of the upstream region of the *ADH1* polyadenylation sites. The three known *ADH1* polyadenylation sites are underlined and designated pAI, pAII, and pAIII. The sequence with homology to the signal proposed by Henikoff and Cohen is boxed. Altered nucleotides in HCmut1 and HCmut2 are in boldface. (B) Polyacrylamide gel analysis of the products formed during the in vitro 3′-end processing reaction of transcripts containing the mutations shown in panel A. The radioactively labeled precursor RNAs (lanes P) were subjected for the indicated times (10′ represents 10 min) to processing reactions containing 3′ dATP. The 5′ and 3′ cleavage products are indicated by arrows. Lanes M contained molecular size markers (pBR322 DNA digested with *Hpa*II). The numbers on the left are molecular sizes in nucleotides.

at the same three polyadenylation sites as in the wild-type transcript (Fig. 2B). However, Cerenkov counting of the excised reaction products revealed an approximately 30% reduction of processing efficiency in the CCCGGGA mutant, while processing efficiency in the AAAAATA mutant was nearly unchanged. Furthermore, slightly increased usage of site pAIII and decreased usage of site pAI in the CCCGGGA mutant were observed (Fig. 2B, lane 10).

The effects of the two mutations were also investigated in vivo. For this experiment, the mutations were introduced into plasmid pSH101 (9; Fig. 1), which contains the *ADH1* 3′-terminal sequence from positions 911 to 1190 inserted in the *Bgl*II restriction site of a truncated *ACT1* transcription unit contained in yeast-*E. coli* shuttle vector YEp351 (12). Poly(A)$^+$ RNA from yeast cells transformed with the corresponding constructs was used for both nuclease S1 and RACE-PCR analyses to determine the actual polyadenylation sites of the mRNAs. The S1 analysis revealed a distribution of mRNA 3′ ends derived from the mutant plasmids that was nearly identical to that of the wild type (Fig. 3A and B, lanes 6 to 8). The slight discrepancy between the expected and actual lengths of the protected fragments is most probably due to the high A+T content of the region investigated and has been discussed elsewhere (9). The increased use of polyadenylation site pAIII versus pAI and pAII in mutant HCmut2 (Fig. 3B, lane 8) coincided with the in vitro data (Fig. 2B, lanes 8 to 10). Also, the increased amount of protected input probe in lane 8 of Fig. 3B suggests a higher number of readthrough transcripts in the CCCGGGA mutant. The S1 data were corroborated further by the sequence analysis of RACE-PCR cDNA clones (Fig. 3C). Most of the cDNAs ended at polyadenylation site pAIII, while a few clones terminated at minor polyadenylation sites. The lack of clones ending at pAI and pAII for pSH101-HCmut1 and at pAI for pSH101-HCmut2 is most likely due to the limited number of clones analyzed. In conclusion, the complete alteration of the sequence stretch with homology to the Henikoff-Cohen signal did not have any detrimental effect on the accuracy of the 3′-end processing of *ADH1*-derived precursor RNAs in vitro and in vivo. Only the CCCGGGA mutant had a minor influence on the processing efficiency at sites pAI and pAIII.

An $(A)_8$ stretch contains a signal for efficient utilization of the first *ADH1* polyadenylation site. The obvious lack of already known processing signals in the *ADH1* 3′-terminal region forced us to search for alternative signal sequences. First indications came from the analysis of the effects of mutations at the *ADH1* polyadenylation sites. The exchange of the adenosine residues located at the first two *ADH1* polyadenylation sites for pyrimidines had resulted in a complete shutdown of processing at the downstream polyadenylation sites in vitro and in vivo (9; unpublished results). Therefore, we assumed that these short adenosine stretches might constitute


A
S1-probe
296
pA
I II III
Bgl II
Eco RI
RNAs
protected fragments
214
221
241
B
-RNA
+RNA
M
wild type
HCmut1
HCmut2
A8mut7
wild type
HCmut1
HCmut2
A8mut7
309
242
238
217
201
probe
pA III
pA alt
pA II
pA I
1 2 3 4 5 6 7 8 9
C
pSH101 (wild type)
GATTTTTATTATTAAATAAGTTATAAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTCTT
pA I
pA II
pA III
pSH101-HCmut1
GAAAAAAATATTAAATAAGTTATAAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTCTT
pA I
pA II
pA III
pSH101-HCmut2
GACCCGGGATATTAAATAAGTTATAAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTCTT
pA I
pA II
pA III
pSH101-A8mut7
GATTTTTATTATTAAATAAGTTATCTGCAGCTTAAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTCTT
pA II
pA III
pA alt

some sort of processing signal. We suspected that, consistent with this assumption, the consecutive row of eight adenosine residues located 10 nucleotides upstream of the first *ADH1* polyadenylation site might be a processing signal as well. To test this hypothesis, we changed one, three, five, and seven nucleotides of this motif to obtain the sequences AAAC AAAA (A8mut1), ATGCAAAA (A8mut3), CTGCAGAA (A8mut5), and CTGCAGCT (A8mut7), respectively. Synthetic transcripts containing these mutations were subjected to in vitro processing reactions (Fig. 4). The single A-to-C mutation in A8mut1 had no detectable effect (Fig. 4B, lane 8). The three-nucleotide exchange in A8mut3 led to a clear-cut decrease of cleavage at the first polyadenylation site and a concomitant increase of cleavages at alternative site pA_{alt} and at pAIII compared with the wild-type situation (Fig. 4B, lane 9). These effects became even more predominant when five or seven nucleotides were exchanged within the $(A)_8$ motif (Fig. 4B, lanes 10 and 11). In A8mut7, cleavage at pAI was almost completely abolished, the products deriving from cleavage at sites pAII and pA_{alt} were formed in equal amounts, and processing at site pAIII was strongly favored (Fig. 4B, lane 11).

The effect of mutation A8mut7 was also assayed in vivo. Both nuclease S1 analysis and sequencing of cDNA clones confirmed that pAI was no longer used in this mutant, but polyadenylation at alternative site pA_{alt} occurred (Fig. 3B and C). Therefore, the in vivo and in vitro results strongly suggest that the $(A)_8$ stretch contains a hitherto unknown signal for 3′-end formation at the first polyadenylation site of the *ADH1* gene.

The distance between the 3′-end processing signal and the polyadenylation sites can vary in a certain range. To test if the distance between the newly identified signal motif within the $(A)_8$ stretch and the polyadenylation sites also has to be defined, we inserted 2, 4, 8, and 16 nucleotides by site-directed mutagenesis. The insertions created an *Rsa*I restriction site and additional alternating pyrimidine stretches which had been shown not to serve as preferential polyadenylation sites (9). The results of the corresponding in vitro 3′-end processing reactions are shown in Fig. 5B. With an insertion of up to eight nucleotides, cleavage occurred at the three genuine wild-type sites (Fig. 5B, lanes 8 to 10). The sizes of the 5′ cleavage products were shifted in accordance with the additional nucleotides in the upstream regions, and the 3′ cleavage products migrated at the same positions as in the wild type. Interestingly, the relative processing efficiencies had changed in the mutants. Cleavage efficiency decreased at polyadenylation site pAII compared with the wild type, in which the products corresponding to cleavage at sites pAI and pAII were formed in equal amounts. In mutant DIS+16, a substantial proportion of the processing products arose from cleavage in the alternating TC sequence of the inserted spacer fragment (cleavage site

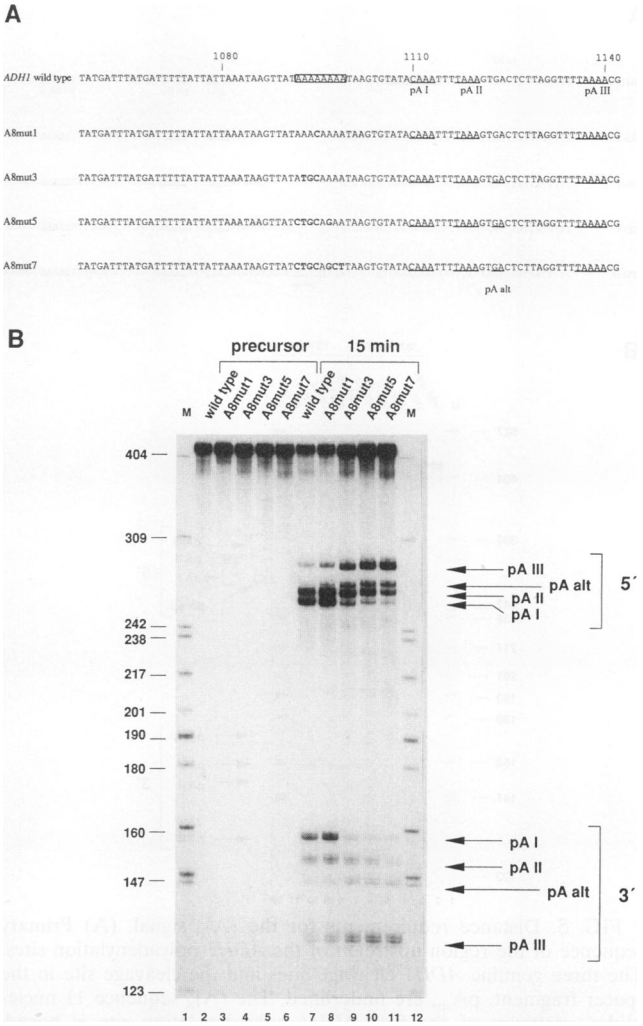


FIG. 4. Influence of the $(A)_8$ stretch on polyadenylation site selection in vitro. (A) Primary sequence of the upstream region of the *ADH1* polyadenylation sites. The three genuine *ADH1* polyadenylation sites and alternative polyadenylation site pA_{alt} are underlined. The $(A)_8$ sequence 11 nucleotides upstream of the first *ADH1* polyadenylation site is boxed. Altered nucleotides are in boldface. (B) Polyacrylamide gel analysis of the products formed during the in vitro 3′-end processing reaction of transcripts containing the mutations shown in panel A. The arrows indicate the 5′and 3′ cleavage products. Lanes M contained molecular size markers (pBR322 DNA digested with *Hpa*II). The numbers on the left are molecular sizes in nucleotides.

FIG. 3. In vivo analysis of various mutations in the *ADH1* 3′-terminal region. (A) Schematic presentation of the input probe used for nuclease S1 analysis and of the expected protected fragments for the wild-type situation (construct pSH101). Arrows indicate mRNAs originating at the *ACT1* promoter and ending at the *ADH1* polyadenylation sites. The S1 DNA probe (a 296-bp *Bgl*II-*Eco*RI fragment labeled at the *Bgl*II site) is shown at the top. Expected DNA fragments protected by mRNAs ending at one of the three *ADH1* polyadenylation sites are depicted together with their respective lengths at the bottom. (B) Nuclease S1 mapping of the mRNA 3′ ends generated in yeast cells transformed with *ADH1* constructs containing the various mutations. Nuclease S1 digestion was performed either without (−RNA) or with (+RNA) RNA at 16°C for 90 min. The arrows indicate fragments protected by polyadenylation at the corresponding sites. Lane M contained molecular size markers (pBR322 DNA digested with *Hpa*II). The numbers on the left are molecular sizes in nucleotides. (C) DNA sequence context of the *ADH1* polyadenylation region and the three corresponding sequences in which upstream nucleotide stretches were mutagenized. The mutated sequences are boxed. In vitro cleavage sites are underlined. The bars above the DNA sequences indicate the polyadenylation sites as determined by sequencing of RACE-PCR cDNA clones. The height of each bar represents the abundance of the corresponding clone, and the actual number of clones is given above each bar. The dotted triangles attached to the bars reflect uncertainty about their exact positions due to the presence of adenosine residues at these sites.

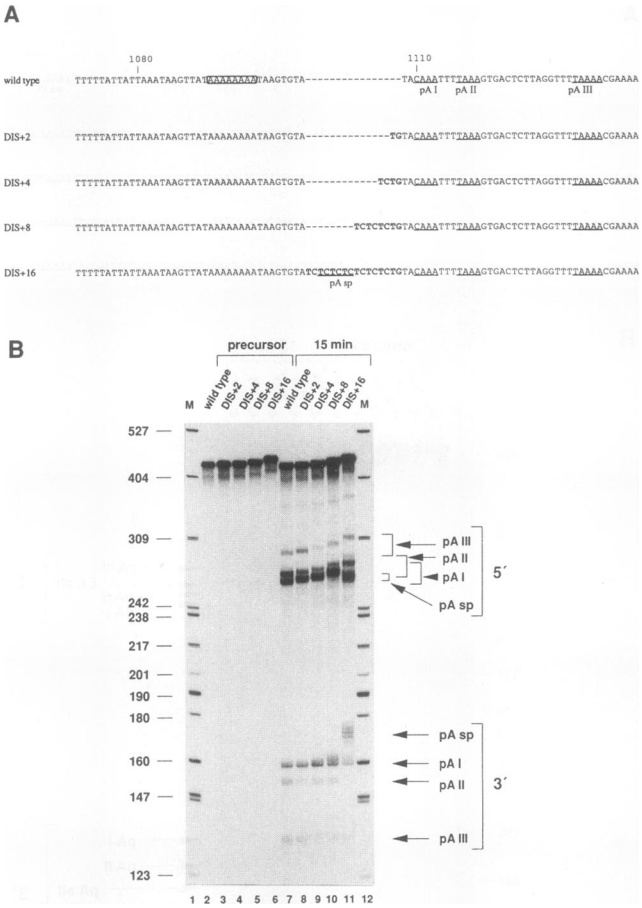


FIG. 5. Distance requirements for the $(A)_8$ signal. (A) Primary sequence of the region upstream of the *ADH1* polyadenylation sites. The three genuine *ADH1* cleavage sites and the cleavage site in the spacer fragment, pA_{sp}, are underlined. The $(A)_8$ sequence 11 nucleotides upstream of the first *ADH1* polyadenylation site is boxed. Inserted nucleotides are in boldface. Constructs DIS+2, DIS+4, DIS+8, and DIS+16 contain an additional 2, 4, 8, and 16 nucleotides, respectively. (B) Polyacrylamide gel analysis of the products formed during the in vitro 3′-end processing reaction of transcripts containing the mutations shown in panel A. The arrows indicate the 5′ and 3′ cleavage products. Lanes M contained molecular size markers (pBR322 DNA digested with *Hpa*II). The numbers on the left are molecular sizes in nucleotides.

pA_{sp}; Fig. 5B, lane 11). This implies that the system was unable to bridge the distance of an additional 16 nucleotides to reach the otherwise preferentially used CAAA sequence of polyadenylation site pAI. Instead, processing occurred within a region located an appropriate distance (around 14 nucleotides) downstream of the signal, although the processing sites themselves did not display an optimal sequence context. As can be judged from the pattern of the 3′ cleavage products, cleavage occurred every second nucleotide. Most likely, the precursors were cleaved after the cytidine residues in the alternating TC stretch, since it was shown that 3′-end processing in the *CYC1* 3′-terminal region preferentially takes place after cytidine residues, which are in some cases flanked by thymidine nucleotides (28).

The $(A)_8$ sequence works as a processing signal when introduced in a heterologous context. To find out if the newly defined signal is also effective in a heterologous sequence context, *ADH1*-derived fragments of various size were inserted into the *cyc1-512* mutant. This mutant is characterized by a deletion of 38 bp upstream of the *CYC1* polyadenylation site, which results in the destruction of proper *CYC1* mRNA 3′-end formation in vivo and in vitro (4, 33). In the literature, the deletion point is generally fixed 5 nucleotides upstream of the *CYC1* polyadenylation site (the sequence CAAA with the C located at position 502) (28–30, 33). However, since the 38-bp deletion occurred between two direct 7-bp repeats, one of which is retained in the *cyc1-512* mutant, the deletion point can be located at any position between 5 and 12 bp upstream of the *CYC1* polyadenylation site. To obtain comparable spacing of the newly identified *ADH1* upstream signal and the *CYC1* polyadenylation site, we regarded the deletion point to be 12 bp upstream of the C at position 502. Therefore, *ADH1*-specific sequences located in the corresponding region relative to the first *ADH1* polyadenylation site were inserted at position 452 of *cyc1-512*. The various insertions had the $(A)_8$ stretch at the 3′ end in common and differed in the number of preceding nucleotides. Parts of the wild-type sequences of the *CYC1* and *ADH1* genes, as well as the sequences of the insertions, are shown in Fig. 6A. In the first mutant of this series (cycΔ+38), we replaced the 38 *CYC1*-derived nucleotides, which are absent in the *cyc1-512* mutant (cycΔ), with an *ADH1*-derived sequence identical in length [the $(A)_8$ sequence plus the 30 preceding nucleotides]. The in vitro processing pattern of a precursor RNA carrying this mutation was indistinguishable from the pattern obtained with the wild-type *CYC1* construct (compare lanes 9 and 11 of Fig. 6B). This result shows that the *ADH1*-specific sequence is functionally equivalent to the corresponding *CYC1*-specific sequence. When shorter fragments were inserted, cleavage efficiency decreased, albeit the precursor molecules were all processed at the genuine *CYC1* polyadenylation site as judged by the migration of the 3′-end cleavage products (Fig. 6B, lanes 12 to 14). The shortest insertion, which represents exclusively the $(A)_8$ stretch, triggered only a very weak processing reaction (Fig. 6B, lane 14). Therefore, the $(A)_8$ stretch by itself is not sufficient to restore the processing capacity of the *cyc1-512* precursor. To test whether the $(A)_8$ stretch, while insufficient to restore processing, is necessary for in vitro processing in the *CYC1* context, we introduced the same 7-bp alteration within this motif which nearly abolished 3′-end formation at the first polyadenylation site of the *ADH1* gene. As can be seen in Fig. 6B, lane 15, this mutation led to a nearly complete shutdown of the processing reaction at the *CYC1* polyadenylation site, thereby demonstrating the functional importance of this sequence, even in the *CYC1* environment.

We also examined in the *CYC1* context the importance of the sequence stretch which is homologous to the polyadenylation signal sequence proposed by Henikoff and Cohen (10). In mutant cycΔ+38mutHC, the sequence TTTTTAT was changed to CTGCAGC, analogous to the sequence alteration in mutant cycΔ+38mutA8 (Fig. 6A). The in vitro processing pattern of transcripts containing this mutation clearly showed that this alteration had little or no effect on both cleavage accuracy and efficiency (Fig. 6C).

In conclusion, these results demonstrated unequivocally that the sequence AAAAAAAA is necessary, but not sufficient, for the 38-bp fragment derived from the *ADH1* 3′-terminal region to work as an efficient processing signal when inserted in the *cyc1-512* deletion mutant. In contrast, the sequence stretch with homology to the Henikoff-Cohen signal is not essential for this 38-bp fragment to direct 3′-end processing.

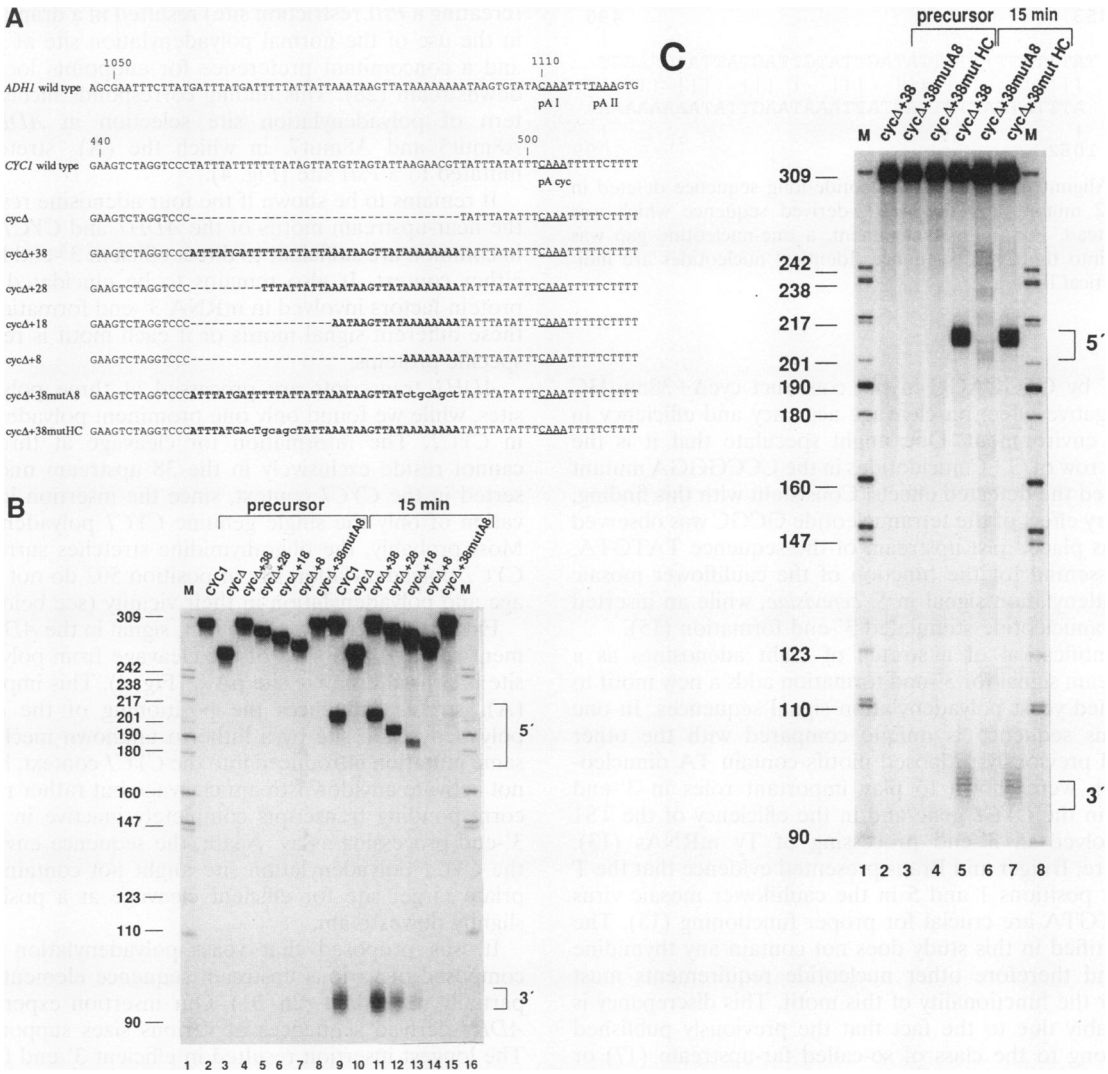


FIG. 6. Effects of inserting *ADH1*-specific sequences in the deletion point of the *cyc1-512* mutant. (A) Primary sequences of the upstream regions of the *ADH1* and *CYC1* polyadenylation sites. The cleavage and polyadenylation sites are underlined. The inserted *ADH1*-derived sequences are in boldface, and altered nucleotides are in lowercase letters. (B and C) Polyacrylamide gel analysis of the products formed during the in vitro 3′-end processing reaction of transcripts containing the mutations shown in panel A. The 5′ and 3′ cleavage products are indicated. Lanes M contained molecular size markers (pBR322 DNA digested with *Hpa*II). The numbers to the left are molecular sizes in nucleotides. The effects of nucleotide alterations in the polyadenylation signal sequence proposed by Henikoff and Cohen (10) and the $(A)_8$ stretch are compared in panel C.

DISCUSSION

The purpose of this study was to define signal sequences upstream of the *ADH1* polyadenylation sites which are responsible for proper mRNA 3′-end formation. During the last decade, a variety of sequences have been proposed which might be actively involved in 3′-end processing and/or termination in several yeast genes. In the 3′-terminal region of the yeast *ADH1* gene, no perfect match with any of the signals proposed earlier was detected. Only the sequence TTTTTAT, located 23 nucleotides downstream of the *ADH1* termination codon, displays significant homology to the signal TTTTTATA proposed by Henikoff and Cohen (10). In the original investigations, deletion of the last adenosine of this signal led to the formation of 50% read-through transcripts, stressing the importance of this single nucleotide (11). Deletion of the entire signal abolished proper 3′-end formation completely (10). Therefore, one might speculate that the sequence TTTTTAT contributes, at least in part, to 3′-end processing of *ADH1* transcripts. In the mutagenesis experiment in which the TTTT TAT sequence was replaced by AAAAATA, no significant influence on 3′-end formation, either in vitro or in vivo, was detected. This result is in agreement with the findings of Irniger et al. (16), who mutated the sequence TTTTTAT located in the 3′-terminal region of the *PHO5* gene to GTCT TGT. This mutation also caused no detectable reduction of 3′-end formation efficiency compared with the wild-type construct. In the CCCGGGA mutant of the *ADH1* gene, however, a decrease in efficiency and a slight alteration in the distribution of the mRNA 3′ ends were detected. Most likely, these effects are due to an inhibitory action of the G+C-rich sequence introduced, since they were not observed in the AAAAATA mutant. On the other hand, replacement of

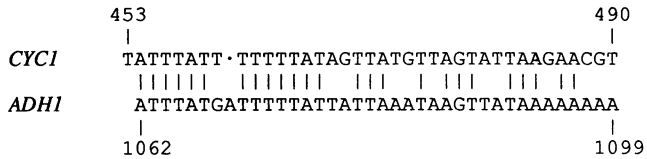


FIG. 7. Alignment of the 38-nucleotide-long sequence deleted in the *cyc1-512* mutant and the *ADH1*-derived sequence which was inserted instead. For optimal alignment, a one-nucleotide gap was introduced into the *CYC1* sequence. Identical nucleotides are indicated by vertical lines.

TTTTTAT by CTGCAGC in the construct cycΔ+38mutHC had no negative effect on cleavage accuracy and efficiency in the *CYC1* environment. One might speculate that it is the contiguous row of G+C nucleotides in the CCCGGGA mutant which caused the detected effects. Consistent with this finding, an inhibitory effect of the tetranucleotide GCGC was observed when it was placed just upstream of the sequence TATGTA, which is essential for the function of the cauliflower mosaic virus polyadenylation signal in *S. cerevisiae*, while an inserted ATAT tetranucleotide stimulated 3′-end formation (15).

The identification of a stretch of eight adenosines as a near-upstream signal for 3′-end formation adds a new motif to the published yeast polyadenylation signal sequences. In one respect, this sequence is unique compared with the other signals. All previously proposed motifs contain TA dinucleotides, which were shown to play important roles in 3′-end formation in the *CYC1* gene and in the efficiency of the TS1 region involved in 3′-end processing of Ty mRNAs (13). Furthermore, Irniger and Braus presented evidence that the T residues at positions 1 and 5 in the cauliflower mosaic virus signal TATGTA are crucial for proper functioning (15). The signal identified in this study does not contain any thymidine residue, and therefore other nucleotide requirements must account for the functionality of this motif. This discrepancy is most probably due to the fact that the previously published signals belong to the class of so-called far-upstream (17) or upstream (28) elements, while because of its proximity to the polyadenylation site, the $(A)_8$ signal belongs to the class of so-called near-upstream (17) or downstream (28) elements. We prefer the terms far- and near-upstream elements for describing the split nature of the yeast polyadenylation signals because the term downstream element may cause confusion when comparing the yeast signals with the situation in higher eucaryotes, in which the term downstream element denotes sequences downstream of the polyadenylation sites (7, 21, 22).

The fact that the 38 *ADH1*-derived nucleotides can substitute for the 38 nucleotides deleted in the *cyc1-512* mutant demonstrates the possible interchangeability of yeast polyadenylation signals. Comparison of the 38-bp stretch deleted in the *cyc1-512* mutant to the inserted *ADH1*-derived sequence revealed considerable overall homology (Fig. 7). However, we have demonstrated that the region with the best match (seven identical contiguous nucleotides), which resembles the signal sequence proposed by Henikoff and Cohen (10), is dispensable for proper 3′-end processing (Fig. 2, 3, and 6C). On the other hand, the $(A)_8$ stretch, which has been shown to be important for 3′-end processing in both the *ADH1* and *CYC1* contexts, has its counterpart in the distantly related sequence AA GAACGT of the *CYC1* 38-bp region. The latter motif overlaps the sequence TTAAGAAC identified by Russo et al. (28) as the near-upstream element for 3′-end formation in the *CYC1* gene. Interestingly, mutation of this motif to TTCTGCAG (creating a *Pst*I restriction site) resulted in a dramatic decrease in the use of the normal polyadenylation site at position 502 and a concomitant preference for endpoints located further downstream (28). This finding corresponds nicely to the pattern of polyadenylation site selection in *ADH1* mutants A8mut5 and A8mut7, in which the $(A)_8$ stretch was also mutated to a *Pst*I site (Fig. 4).

It remains to be shown if the four adenosine residues which the near-upstream motifs of the *ADH1* and *CYC1* genes have in common are sufficient to direct efficient 3′-end formation in either context. It also remains to be elucidated if common protein factors involved in mRNA 3′-end formation recognize these different signal motifs or if each motif is recognized by specific proteins.

ADH1 transcripts are processed at three polyadenylation sites, while we found only one prominent polyadenylation site in *CYC1*. The information for cleavage at three positions cannot reside exclusively in the 38 upstream nucleotides inserted in the *CYC1* context, since the insertion led to reactivation of only the single genuine *CYC1* polyadenylation site. Most probably, the oligothymidine stretches surrounding the *CYC1* polyadenylation site at position 502 do not favor cleavage and polyadenylation in their vicinity (see below).

Extensive alteration of the $(A)_8$ signal in the *ADH1* environment resulted in a shift of the cleavage from polyadenylation site pAI to alternative site pA_{alt} (Fig. 4). This implies that the $(A)_8$ stretch influences the positioning of the downstream polyadenylation site by a hitherto unknown mechanism. The same mutation introduced into the *CYC1* context, however, did not activate any downstream cleavage but rather rendered the corresponding transcripts completely inactive in the in vitro 3′-end processing assay. Again, the sequence environment of the *CYC1* polyadenylation site might not contain any appropriate target site for efficient cleavage at a position shifted slightly downstream.

It was proposed that yeast polyadenylation signals are composed of various upstream sequence elements which are partially redundant (26, 31). Our insertion experiments with *ADH1*-derived sequences of various sizes support this view. The longest insertion resulted in efficient 3′-end formation in vitro, while shorter fragments functioned less efficiently. The decrease in processing efficiency was paralleled by decreasing lengths of the insertions, implying a partial contribution of all of the 10-bp increments of the various inserted fragments. This observation cannot be dismissed as a simple restoration of proper spacing of some signal upstream of the 38-bp deletion in *cyc1-512* and the polyadenylation site itself. If that had been the case, both cycΔ+38 and cycΔ+38mutA8 transcripts should have acted as efficient substrates for 3′-end processing reactions.

Proper spacing between the newly identified signal and the corresponding polyadenylation sites seems to be important. In the series of insertion mutants including DIS+2 to DIS+16, a distance of up to 8 additional nucleotides can be bridged to cleave at preferential polyadenylation site pAI, demonstrating the flexibility of the system. The decreased use of polyadenylation site pAII in these mutants is most probably due to the enlarged distance between the signal and the site itself. The insertions may position pAI at a proper distance from the signal of site pAII. Virtually no cleavage in the inserted spacer fragments was observed in these cases. This finding argues against the recent report of preferential polyadenylation after cytidine residues (28) which do occur in the insertions. This discrepancy might be explained by assuming that potential polyadenylation sites can be ordered in a hierarchy. We have shown previously that sites containing adenosine residues are a

preferred target for cleavage and polyadenylation (9). In the *CYC1* system investigated by Russo et al. (28), no sequences with homology to sites we would expect to be used preferentially can be found in close proximity to many determined alternative polyadenylation sites. For example, the polyadenylation site located at cytidine 511 of the *CYC1* gene is surrounded by oligothymidine stretches, which obviously rank lower in the hierarchy of positions used for cleavage and polyadenylation. Therefore, preferential polyadenylation after cytidine residues may occur if there are no sequences with higher preference in the vicinity. Since preferential site pAI (CAAA) in insertion mutants DIS+2, DIS+4, and DIS+8 is still within range of the system, no 3′-end processing at the cytidines in the spacer fragments was observed. In mutant DIS+16, however, the system is no longer able to bridge the additionally inserted nucleotides efficiently. Thus, cleavage in the spacer fragment, which most probably occurred after the cytidine residues, was detected.

The organization of polyadenylation signals in *S. cerevisiae* is reminiscent of the situation described for the *rbcS* gene of higher plants, in which a far-upstream element interacts with multiple near-upstream elements, each of which is specific for a separate polyadenylation site (24). Applying this model, we wondered which sequences might constitute the near-upstream signals for *ADH1* polyadenylation sites pAII and pAIII. For pAIII, the near-upstream element is most probably located at the adenosine stretches of pAI and pAII, since the exchange of these purine stretches by pyrimidines led to complete shutdown of 3′-end processing at site pAIII (9; our unpublished results). Until now, we have not been able to define the near-upstream signal for pAII. None of the alterations of the $(A)_8$ stretch had a negative influence on 3′-end processing at this site. Therefore, we initially proposed that some sequence downstream of $(A)_8$ might be responsible for performing this task. However, changing the sequence from AAAAAAAATA AGTG to AAAAAACCCGGGTG had little or no effect on 3′-end processing in vitro (data not shown). Most likely, the remaining six adenosine residues serve as a signal for cleavage at both pAI and pAII. Further mutagenesis experiments will help to clarify the nature of the near-upstream element for pAII.

Despite the efforts of numerous laboratories, it has not been possible to define a universal consensus motif for pre-mRNA polyadenylation signals in *S. cerevisiae*. Although the proposed signals have been shown to be important in some genes, they are not essential in other genes. This study adds a new motif, AAAAAAAA, to the previously described signals TAG...TAT GTA, TATATA, TACATA (29), TTTTTATA (10), $(AT)_9$ (1, 6), and TTAAGAAC (28). Evidence emerges from this work and a number of previous publications that yeast polyadenylation signals are of a composite nature (1, 9, 14, 26, 31, 33). Several sequence elements and features, such as base composition and spacing between signals, might have to interact for efficient 3′-end formation. It remains to be elucidated how these different elements interact with each other and how involved protein factors recognize and utilize these redundant and degenerate elements for efficient pre-mRNA cleavage and polyadenylation.

ACKNOWLEDGMENTS

We thank J. Hani for helpful discussions and critical reading of the manuscript, C. Rommel for technical assistance, and B. Obermaier for DNA sequence analysis. Furthermore, we are indebted to T. Platt for sharing data prior to publication.

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. Hiraoka, and T. Fukasawa.** 1990. Signal sequence for generation of mRNA 3′ end in the *Saccharomyces cerevisiae GAL7* gene. EMBO J. **9**:3691–3697.
2. **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.
3. **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.
4. **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.
5. **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.
6. **Chen, J., and C. Moore.** 1992. Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA. Mol. Cell. Biol. **12**:3470–3481.
7. **Gil, A., and N. J. Proudfoot.** 1984. A sequence downstream of AAUAAA is required for rabbit β-globin mRNA 3′-end formation. Nature (London) **312**:473–474.
8. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
9. **Heidmann, S., B. Obermaier, K. Vogel, and H. Domdey.** 1992. Identification of pre-mRNA polyadenylation sites in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **12**:4215–4229.
10. **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.
11. **Henikoff, S., J. D. Kelly, and E. H. Cohen.** 1983. Transcription terminates in yeast distal to a control sequence. Cell **33**:607–614.
12. **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.
13. **Hou, W., R. Russnak, and T. Platt.** 1994. Poly(A) site selection in the yeast Ty retroelement requires an upstream region and sequence-specific titratable factor(s) in vitro. EMBO J. **13**:446–452.
14. **Hyman, L. E., S. H. Seiler, J. Whoriskey, and C. L. 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.
15. **Irniger, S., and G. H. Braus.** 1994. Saturation mutagenesis of a polyadenylylation signal reveals a hexanucleotide element essential for mRNA 3′ end formation in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **91**:257–261.
16. **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.
17. **Irniger, S., H. Sanfaçon, C. M. Egli, and G. H. Braus.** 1992. Different sequence elements are required for function of the cauliflower mosaic virus polyadenylation site in *Saccharomyces cerevisiae* compared with in plants. Mol. Cell. Biol. **12**:2322–2330.
18. **Ito, H., Y. Fukuda, K. Murata, and A. Kimura.** 1983. Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153**:163–168.
19. **Köhrer, K., and H. Domdey.** 1991. Preparation of high molecular weight RNA. Methods Enzymol. **194**:398–405.
20. **Kunkel, T. A.** 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA **82**:488–492.
21. **McDevitt, M. A., M. J. Imperiale, H. Ali, and J. R. Nevins.** 1984. Requirement of a downstream sequence for generation of a poly(A) addition site. Cell **37**:993–999.
22. **McLauchlan, J., D. Gaffney, J. L. Whitton, and J. B. Clements.** 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini. Nucleic Acids Res. **13**:1347–1368.
23. **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.

24. **Mogen, B. D., M. H. MacDonald, G. Leggewie, and A. G. Hunt.** 1992. Several distinct types of sequence elements are required for efficient mRNA 3′ end formation in a pea *rbcS* gene. Mol. Cell. Biol. **12:**5406–5414.
25. **Orkin, S. H., T.-C. Cheng, S. E. Antonarakis, and H. H. Kazazian.** 1985. Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human β-globin gene. EMBO J. **4:**453–456.
26. **Osborne, B. I., and L. Guarente.** 1989. Mutational analysis of a yeast transcriptional terminator. Proc. Natl. Acad. Sci. USA **86:**4097–4101.
27. **Proudfoot, N. J., and G. G. Brownlee.** 1974. Sequence at the 3′ end of globin mRNA shows homology with immunoglobulin light chain mRNA. Nature (London) **252:**359–362.
28. **Russo, P., W.-Z. Li, Z. Guo, and F. Sherman.** 1993. Signals that produce 3′ termini in *CYC1* mRNA of the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. **13:**7836–7849.
29. **Russo, P., W.-Z. Li, D. M. Hampsey, K. S. Zaret, and F. Sherman.** 1991. Distinct cis-acting signals enhance 3′ endpoint formation of *CYC1* mRNA in the yeast *Saccharomyces cerevisiae*. EMBO J. **10:**563–571.
30. **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.
31. **Sadhale, P. P., and T. Platt.** 1992. Unusual aspects of in vitro RNA processing in the 3′ regions of the *GAL1*, *GAL7*, and *GAL10* genes in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **12:**4262–4270.
32. **Wahle, E., and W. Keller.** 1992. The biochemistry of 3′-end cleavage and polyadenylation of messenger RNA precursors. Annu. Rev. Biochem. **61:**419–440.
33. **Zaret, K. S., and F. Sherman.** 1982. DNA sequence required for efficient transcription termination in yeast. Cell **28:**563–573.