MOLECULAR AND CELLULAR BIOLOGY, Nov. 1995, p. 5983–5990 Vol. 15, No. 11
0270-7306/95/$04.00+0

3′-End-Forming Signals of Yeast mRNA

ZIJIAN GUO AND FRED SHERMAN*

Department of Biochemistry, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642

Received 12 June 1995/Returned for modification 12 July 1995/Accepted 25 August 1995

It was previously shown that three distinct but interdependent elements are required for 3′ end formation of mRNA in the yeast *Saccharomyces cerevisiae*: (i) the efficiency element TATATA and related sequences, which function by enhancing the efficiency of positioning elements; (ii) positioning elements, such as TTAAGAAC and AAGAA, which position the poly(A) site; and (iii) the actual site of polyadenylation. In this study, we have shown that several A-rich sequences, including the vertebrate poly(A) signal AATAAA, are also positioning elements. Saturated mutagenesis revealed that optimum sequences of the positioning element were AATAAA and AAAAAA and that this element can tolerate various extents of replacements. However, the GATAAA sequence was completely ineffective. The major cleavage sites determined in vitro corresponded to the major poly(A) sites observed in vivo. Our findings support the assumption that some components of the basic polyadenylation machinery could have been conserved among yeasts, plants, and mammals, although 3′ end formation in yeasts is clearly distinct from that of higher eukaryotes.

Unlike prokaryotic mRNA 3′ ends, which are formed primarily by termination events, higher eukaryotic mRNA 3′ ends are formed by endonucleolytic cleavage and subsequent polyadenylation, with transcription continuing beyond the poly(A) site. Extensive analysis has shown that the mammalian poly(A) signal contains three sequence elements (see reference 40 for a review). The most highly conserved is the hexanucleotide AAUAAA, which is found 10 to 30 nucleotides (nt) upstream of the actual cleavage site. This sequence has been shown to be essential for proper 3′ end formation in vivo and for both cleavage and polyadenylation in vitro. A less conserved U-rich or GU-rich element, usually situated immediately downstream of the poly(A) addition site, may serve to modulate the efficiency of the usage of poly(A) sites. The third element, the poly(A) addition site itself, is usually a dinucleotide CA, the second nucleotide of which is invariably an adenosine residue. Cleavage occurs 3′ to the adenosine residue, which is followed by the addition of the poly(A) tail.

In contrast to the higher eukaryotes, the yeast *Saccharomyces cerevisiae* appears to use different but still complex signals to direct mRNA 3′ end formation. The first poly(A) signal to be characterized in *S. cerevisiae* was destroyed by a 38-bp deletion of the *CYC1* gene near its 3′ end (42). This deletion mutation, *cyc1-512*, abolishes the normal 3′ end formation of the *CYC1* mRNA, resulting in low levels of *CYC1* mRNA with extended 3′ ends. Butler and Platt found that the same 38-bp region was necessary for the proper cleavage of extended *CYC1* transcripts in vitro (4). In addition, a study of the dependence of plasmid stability on transcription across a *CEN* element indicated that the same 38-bp region caused transcription termination in vivo (33). This *cyc1-512* mutation has been used to investigate the signals required for 3′ end formation. Several motifs—TATATA, TAG· · ·TATGTA, and TACA TA—were identified by examining intragenic revertants and oligonucleotide-directed alterations that reverted the *cyc1-512* defect (32). Further support for the involvement of these and related sequences in 3′ end formation came from studies of other altered yeast genes (1, 20, 21, 41) and from the insertion of heterologous fragments into the yeast genome (16, 17, 22).

Further analysis of the deletions of the *CYC1*$^+$ 38-bp region and the introduction of specific alterations in the *cyc1-512* mutant led Russo et al. (31) to suggest that 3′-end-forming signals are composed of three distinct but interacting elements. The first and second were previously denoted, respectively, as the upstream and downstream elements but are now referred to as, respectively, the efficiency and positioning elements. The third element is the sites of polyadenylation. The efficiency elements include such sequences as TATATA, TAG· · ·TAT GTA, TTTTTATA, and their derivatives, and they function to enhance the activity of the positioning elements. Russo et al. (31) determined that the positioning elements were composed of at least the TTAAGAAC and AAGAA sequences and that these elements directed the positions of poly(A) sites, which occurred 16 to 27 nt downstream, usually after pyrimidine residues followed by one or more adenosine residues.

In this study, we have systematically characterized the sequence of the positioning element and further investigated its properties. Like the efficiency element, the positioning element is also highly degenerate. In addition to TTAAGAAC and AAGAA, several other A-rich sequences qualified as positioning elements by their ability to position new poly(A) sites. Among them, the AATAAA sequence had the highest capacity. Saturated mutagenesis of the AATAAA sequence revealed that AAAAAA and AATAAA are almost equally efficient, while other mutations caused a reduction in efficiency which varied with the position in the sequence and the base that was introduced. In particular, an AATAAA-to-GATAAA mutation abolished its activity. Furthermore, it was found that AAUAAA, or its functionally related sequences, combined with an upstream element, UAUAUA, was sufficient to direct pre-mRNA cleavage in vitro. Our results suggest that TAT ATA and AATAAA, or their related sequences, determine 3′ end formation in *S. cerevisiae*, although the selection of the actual poly(A) site can be slightly affected by flanking sequences.

* Corresponding author. Mailing address: Department of Biochemistry, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642. Phone: (716) 275-2766. Fax: (716) 271-2683. Electronic mail address: fsrm@bphvax.biophysics.rochester.edu.

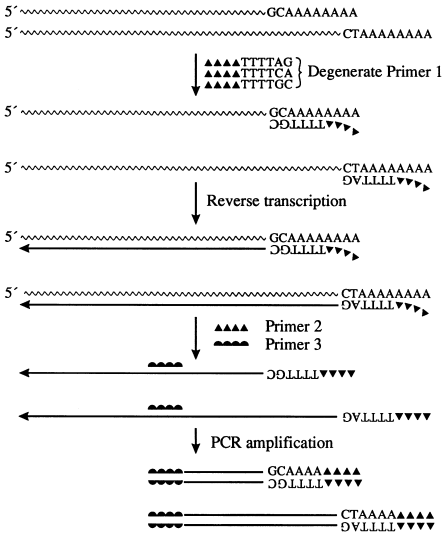


FIG. 1. Schematic representation of the method for the PCR mapping of the 3′ ends of the *CYC1* mRNAs. To help position the primer at the exact junction of the mRNA 3′ end and the poly(A) tail, primer 1 was made degenerate at the last two nucleotides of the 3′ end, resulting in a mixture of 12 primers (31). After degenerate primer 1 is hybridized to yeast total RNA, reverse transcription is performed to produce cDNA, which is subsequently amplified by PCR with primer 2 and ^{32}P-end-labeled primer 3.

MATERIALS AND METHODS

Genetic nomenclature and yeast strains. The gene symbol *CYC1*$^+$ denotes the wild-type allele encoding iso-1-cytochrome *c* in the yeast *S. cerevisiae*, whereas *CYC1* refers to any allele at this locus. The lowercase designations *cyc1-512*, *cyc1-965*, etc., denote altered *CYC1* alleles, irrespective of whether or not the allele causes a deficiency.

S. cerevisiae B-7467 (*MAT***a** *cyc1-Δ1 CYC7*$^+$ *ura3-52*) was used as the recipient of the YCp50 plasmid derivatives containing altered forms of the *CYC1* gene (31). Yeast whole-cell extracts were prepared from yeast strain EJ101 (*MAT*α *trp1 pro1-126 prb1-112 pep4-3 prc1-126*) for the in vitro cleavage assay (4).

Plasmid constructions. Plasmid pAB1550, containing the *cyc1-512* gene in a *Bam*HI-*Hin*dIII fragment, was used for site-directed mutagenesis in the 3′ untranslated region of *cyc1-512*. Plasmid pAB650 was constructed by inserting a 4.7-kb *Bam*HI-*Eco*RI fragment containing the entire *cyc1-512* gene into the yeast shuttle vector YCp50 (31). Site-directed mutations made in plasmid pAB1550 were introduced into plasmid pAB650 as *Kpn*I-*Hin*dIII fragments.

Nucleic acid manipulations. Oligonucleotides used for site-directed mutagenesis, sequencing, and PCR were synthesized on an Applied Biosystems DNA synthesizer (model 380A). Oligonucleotide-directed mutagenesis was performed in vitro by the method of Kunkel et al. (23). PCR mapping of *CYC1* mRNA 3′ termini was performed exactly as outlined by Russo et al. (31), except that only one round of PCR amplification (25 cycles) was conducted with primer 2 and primer 3, as represented schematically in Fig. 1. The PCR products were separated on an 8% polyacrylamide-urea sequencing gel along with sequencing ladders. Autoradiography was performed at −70°C for 16 h with an intensifying screen. An LKB 2222-010 UltraScan XL laser densitometer was used to determine the densities of the bands. Northern (RNA) blot analysis of yeast total RNA was carried out as described by Li and Sherman (24), with minor modifications. ^{32}P-end-labeled oligonucleotide OL93.210 (GGGAGGGCGTGAATG TAAGCGTGACATAACTAAT) was used as the *CYC1* probe. After the filter was prehybridized for 16 h in the hybridization buffer, the probe was added to a final concentration of 4×10^5 cpm/ml. Hybridization was performed at room temperature for 16 h. The filter was washed with 1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate two times at room temperature and once at 42°C. Autoradiography was performed at −70°C for 24 h with two intensifying screens.

In vitro cleavage assay. A T7 promoter sequence was fused to the *CYC1* sequence by PCR amplification of the DNA sequence (bp 287 to 458) of the mutant *CYC1* alleles with two primers, OL95.66 (TAAATAGGGACCTAG ACTTCAGGTTG) and OL95.53 (GGATCCTAATACGACTCACTATAGGG ACAGAAACGACTTAATTACC), which contains a 29-bp T7 promoter sequence. The PCR product was used to synthesize capped ^{32}P-labeled RNA transcripts (175 nt) with an in vitro transcription kit (Ambion, Austin, Tex.) according to the manufacturer's recommendations. Whole-cell-extract preparations and in vitro processing reactions were conducted as described by Butler and Platt (4), with minor modifications. Briefly, components of the 10-μl reaction mixture were mixed at room temperature and incubated at 30°C for 1 h. To suppress polyadenylation, the reaction mixture contained 2 mM CTP instead of ATP as the energy source. Reactions were stopped by adding 2 μl of stop solution, and then 25 μl of ethanol was added to precipitate RNA. The cleavage products were then separated on an 8% polyacrylamide sequencing gel and visualized by autoradiography at −70°C for 24 h.

RESULTS

The control *CYC1* allele. We have characterized positioning elements with derivatives of *cyc1-965* in the plasmid pAB651 (Fig. 2A), which has the following traits. First, *cyc1-965* is a derivative of the *cyc1-512* allele (31), in which the efficiency and positioning elements of the *CYC1* poly(A) signal have been deleted (Fig. 2A), thus diminishing or eliminating the interference of the wild-type *CYC1* poly(A) signal. Second, this *cyc1-965* allele contains a strong efficiency element, TATATA, at the *Eco*RV site of *cyc1-512* (31) which restored the efficiency of 3′ end formation and increased the *CYC1* mRNA level from 10 to 100% (Fig. 2B), facilitating the detection of its mRNA by Northern blotting and PCR 3′ end mapping. Finally, we previously showed that when the *CYC1* positioning element TTAAGAAC (or AAGAA) was created downstream of the TATATA element of *cyc1-965*, new major poly(A) sites appeared approximately 20 nt downstream of this positioning element (31) (Fig. 2B). We reasoned that any sequences which had a function similar to that of TTAAGAAC could be identified and characterized if they were created at this same position.

The *GAL7* core signal also contains a positioning element. Although TTAAGAAC and AAGAA are involved in the positioning of the poly(A) site in the *CYC1* gene (31), we cannot rule out the possibility that other sequences also function as a positioning element. In fact, many yeast mRNAs do not possess this sequence in their 3′ untranslated region, suggesting that sequences other than AAGAA may be involved. To approach this problem, we considered other well-defined yeast poly(A) signals, e.g., the poly(A) signal of the *GAL7* gene. Abe et al. (1) demonstrated that a 26-bp sequence, $(AT)_6$AAT AATGACATCAT, located 7 nt upstream from the poly(A) site of *GAL7* mRNA (34), appeared to be necessary and sufficient to specify mRNA 3′ end formation both in vivo and in vitro. However, because the $(AT)_6$ sequence, which seems to function as an efficiency element, is incapable of signaling 3′ end formation, we reasoned that the downstream 14-bp sequence contained a necessary signal, probably a positioning element functionally similar to AAGAA. Considering that the major poly(A) sites are often 10 to 20 nt 3′ to the positioning element (31), we introduced the sequence AATAATGA, but not CATCAT, which was believed to be too close to the poly(A) site. The *cyc1-1349* mutation was made by creating the AATAATGA sequence downstream of the TATATA element of *cyc1-965*. PCR mapping analysis showed that this mutation caused a shift of the major 3′ ends of *CYC1* mRNA from positions 502 and 520 to 387, 396, and 402 (Fig. 3, lanes 6 and 12). Accordingly, the *cyc1-1349* mRNA was shorter than that of *cyc1-965* (Fig. 4, lane 6), which could be accounted for by the shift of poly(A) sites.

In a hypothesis that was in contrast to our result, Abe et al. (1) suggested that a sequence similar to CATCAT may be involved in 3′ end formation, although no supporting experimental evidence was presented.

AATAAA has the highest capacity in positioning new poly(A) sites. Because both TTAAGAAC and AATAATGA are A-rich sequences, several other A-rich sequences were tested for their ability to direct 3′ end formation around position 400 (Fig. 3A and B). The results are summarized in Fig. 5.

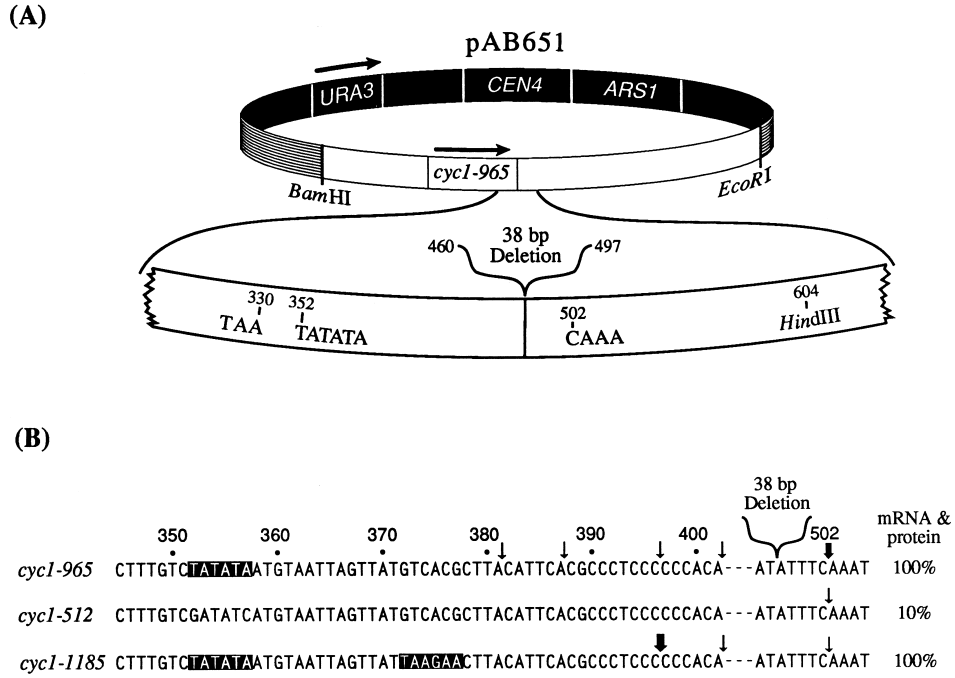


FIG. 2. (A) Schematic representation of the *cyc1-965* yeast expression plasmid pAB651. The 4.7-kb *Bam*HI-*Eco*RI fragment encompasses the *cyc1-965* gene and flanking regions. The 3′ region of the *cyc1-965* gene is expanded at the bottom, and the positions of pertinent nucleotides are also shown. The A of the ATG initiator codon is assigned position 1, 330 corresponds to the last nucleotide of *CYC1* termination codon TAA, 352 denotes the first nucleotide of efficiency element TATATA, 460 and 497 are the flanking positions of the deleted 38-bp region, and 502 is the position of the cytidine preceding the major poly(A) site. (B) Comparison of the *cyc1-965*, *cyc1-512*, and *cyc1-1185* alleles. The allele designations are displayed on the left of each sequence. The efficiency and positioning elements created in *cyc1-965* and *cyc1-1185* are shown in reverse type. The thick and thin arrows, respectively, denote the strong and weak poly(A) sites. The position of the 38-bp deletion is indicated at junction 497.

First, TTAAAAAC created in *cyc1-1345* was found to shift the major poly(A) sites to positions 396 and 402, producing a poly(A) site pattern similar to that of TTAAGAAC (*cyc1-1185*) (31). A related sequence, TTAACAAC (*cyc1-1348*), also positioned poly(A) sites to 396 and 402 (Fig. 3, lanes 5 and 11). It was noted that these A-rich sequences had the following two effects on the positioning of poly(A) sites: they enhanced the use of the poly(A) sites around position 400 and decreased polyadenylation at the 502 site. Therefore, the increased amount of polyadenylation in the position 400 region, compared with the level at the 502 site, reflected the effectiveness of the positioning element. Normalized to the total intensities of the bands around position 400 (sites 396 and 402 in *cyc1-1345* and *cyc1-1348* and 387, 396, and 402 in *cyc1-1349*), the intensity of the band at the 502 site in *cyc1-1348* (TTAACAAC) was approximately twice as much as those in *cyc1-1345* (TTAAAA AC) and *cyc1-1349* (AATAATGA), both of which were similar, indicating that TTAACAAC was less efficient in positioning poly(A) sites. Another A-rich sequence, the mammalian poly(A) signal AATAAA, was also tested in this system (Fig. 3, lanes 4 and 10). In this allele (*cyc1-1347*), surprisingly, the normalized intensity of the band at the 502 site was 50% less than that in *cyc1-1345* (TTAAAAAC), indicating that AATA AA had the highest capacity among these sequences. As a control, a GA-rich sequence, GGAGA (or TTGGAGAC) (*cyc1-1346*), was completely ineffective (Fig. 3, lanes 3 and 9). Although all the A-rich sequences tested here qualified as a downstream element, their capacities varied from one to another, suggesting that a specific sequence, rather than just an A-rich one, is determining 3′ end formation.

Since AATAAA is the most effective signal, its function was further characterized in detail. Two mutants, *cyc1-1350* and *cyc1-1351*, were created solely to move the AATAAA sequence 6 bp upstream or downstream of that in *cyc1-1347*. As expected, while the major 3′ end of *cyc1-1347* mRNA is at position 396, the major poly(A) sites of *cyc1-1350* and *cyc1-1351* mRNAs were shifted to positions 387 and 402, respectively (Fig. 3C, lanes 13 and 14), indicating that AATAAA is positioning the poly(A) sites without regard to sequence contexts. We have reported that the use of the positioning element may be sensitive to the spacing between the efficiency and positioning elements (31). This result indicates that the positioning element functions efficiently within a 12-bp region, from 9 bp to 21 bp downstream of TATATA. Whether greater or smaller spacing can be tolerated has yet to be determined.

Saturation mutagenesis reveals that the positioning element is highly degenerate. The effects of single-base changes in each position of AATAAA were investigated by saturation mutagenesis of this sequence in *cyc1-1347*. The result of PCR 3′ end mapping is shown in Fig. 3C and summarized in Fig. 6. Remarkably, while nucleotide T of the AATAAA sequence is absolutely required for its function in metazoans (37), variations at this position in *cyc1-1358*, *cyc1-1359*, or *cyc1-1360* caused little reduction in the efficiency of 3′ end formation in vivo at positions 396 and 402. On the other hand, mutations at the other positions reduced the efficiency. The extent of the reduction varied with the position and the base used for replacement. In particular, an AATAAA-to-GATAAA mutation (*cyc1-1354*) totally abolished its activity (Fig. 3C, lane 17). This observation was supported by Northern blot analysis of the mutant *CYC1* mRNAs (Fig. 4). While all mutants containing a strong positioning signal (*cyc1-1350*, *cyc1-1351*, *cyc1-1358*, etc.) produced a shorter transcript (530 nt), the *cyc1-1354* mRNA was approximately 600 nt long (Fig. 4, lane 11). In addition to

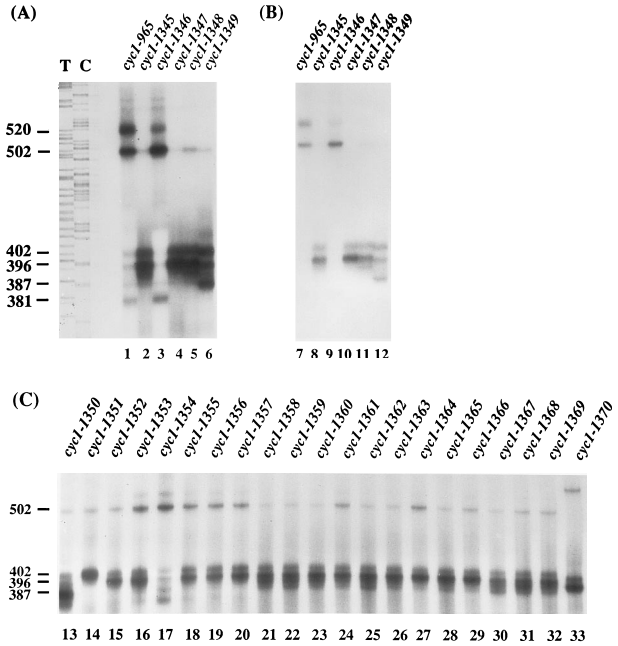


FIG. 3. The AATAAA sequence is the efficiency element of the yeast 3′-end-forming signal. Mapping of the 3′ ends of the transcripts from the *cyc1* mutants in two experiments is shown. The ^{32}P-radiolabeled 3′ end mapping products were electrophoresed through 8% polyacrylamide-urea sequencing gels along with sequencing ladders T and C (A), which were products of sequencing reactions with the control DNA M13mp18 and −40 primers included in the Sequenase version 2.0 kit (United States Biochemicals). The allele numbers for the *CYC1* mutants are displayed above each lane. With the sequencing ladders being used as size standards, the lengths of the PCR products were estimated at an accuracy of ±2 bp. Subsequently, the approximate positions of the poly(A) sites were arrived at, and they are shown on the left of the panels (the A of the ATG initiator codon was assigned position 1). (A and B) Two exposures from the first experiment. (C) An exposure from the second experiment.

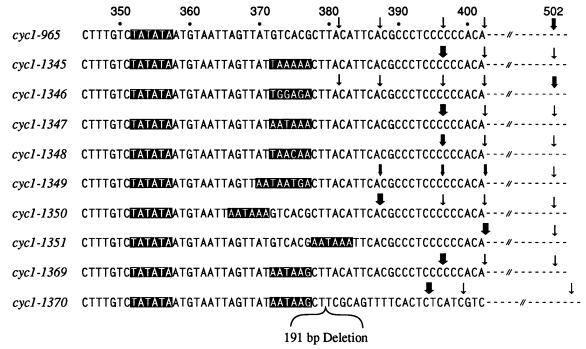


FIG. 5. Details of site-directed mutations derived from *cyc1-512* and the corresponding poly(A) sites, as determined by PCR 3′ end mapping. Allele numbers are shown on the left of each sequence. The alterations are presented in reverse type. Major and minor 3′ termini are denoted, respectively, by thick and thin arrows. The deletion mutation within *cyc1-1370* is also indicated.

the 530-nt transcript, a 600-nt transcript appeared in the mutants containing relatively weak signals, i.e., *cyc1-1353*, *cyc1-1355*, *cyc1-1356*, *cyc1-1357*, *cyc1-1361*, and *cyc1-1364*. Apparently, like the efficiency element, the positioning element is also highly degenerate. This raises the possibility that many yeast mRNAs that do not possess the canonical AAUAAA motif in their 3′ untranslated regions may use variations of this sequence for 3′ end formation. For example, in addition to the TTAAGAAC sequence of *CYC1* and AATAATGA of *GAL7*, an octamer sequence, AAAAAAAA, in the *ADH1* gene was found to be a necessary signal for mRNA 3′ end formation (15). In agreement with this conclusion, it has been shown that deletion or mutation of the AATAAA sequence in the *ADH2* gene (19) and a cauliflower mosaic virus sequence (22) caused a reduction in efficiency but did not abolish mRNA 3′ end formation, suggesting that there may be cryptic but less efficient positioning elements which manifest themselves when AATAAA is deleted.

The poly(A) site is mainly determined by the efficiency and positioning elements. It was reported that the 3′ end formation of the *ADH1* transcripts occurs preferentially at the sequence $Py(A)_N$ (14). We also found that the 3′ ends of the *CYC1* transcript were formed preferentially at cytidine nucleotides (31), suggesting that the selection of the actual poly(A) site may be affected by adjacent sequences. In this regard, it should be noted that approximately 20 bp downstream of the positioning elements, at a position where the poly(A) sites are formed, an unusual CCCCCC sequence is found (Fig. 5). To test whether this region plays a role in determining mRNA 3′ ends, we deleted a 191-bp fragment immediately downstream of the AATAAG sequence in *cyc1-1369*. This mutation, *cyc1-1370*, produced a transcript with a major poly(A) site approximately 17 nt downstream from the AAUAAG element (Fig. 3C, lane 33), indicating that the poly(A) site is mainly determined by the efficiency and positioning elements, though the flanking sequences may slightly affect the precise selection of poly(A) sites.

The positioning element AAUAAA, combined with UAUAUA, directs in vitro processing of pre-mRNAs. It has been demonstrated that many but not all yeast pre-mRNAs can be specifically cleaved and polyadenylated in vitro by a yeast cell extract (1, 4, 5). To test whether the AAUAAA sequence, combined with a strong efficiency element, UAUAUA, is sufficient to direct pre-mRNA cleavage in vitro, we synthesized RNA templates extending beyond the created TATATA···AATAAA sequence. The result showed that several in vitro cleavage prod-

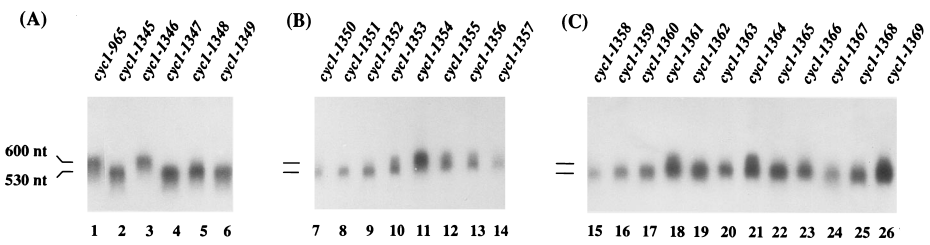


FIG. 4. AATAAA and its related sequences direct the appearance of a shorter *CYC1* transcript. *CYC1* mRNA from the mutant strains was detected by Northern blot analysis in three separate experiments (A, B, and C). Allele numbers for each mutant are displayed above the lanes. The positions of 600- and 530-nt *CYC1* transcripts are indicated.

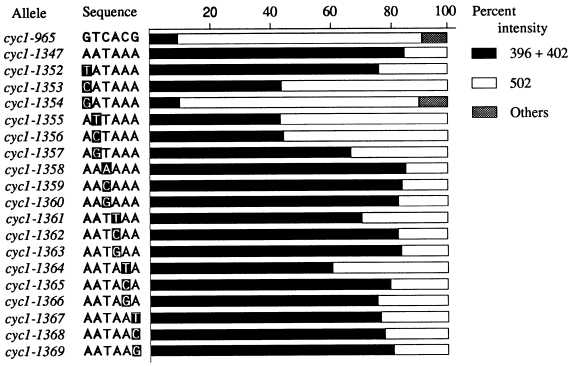


FIG. 6. Summary of the effects of single base changes in the AATAAA sequence. Allele numbers and the corresponding *CYC1* sequences from positions 372 to 377 are shown on the left. Single base changes in the AATAAA sequence are presented in reverse type. Because the efficiency of PCR amplification is inversely proportionate to the distance separating the two primers, the relative amounts of shorter PCR products are higher than those of the corresponding mRNAs. These values were corrected with the *R* values (amplification efficiencies) for the PCR products around position 400 (160 bp) and at position 502 (220 bp). These values corresponded to approximately 0.72 and 0.62, respectively (7). By using the formula $Y = A\ (1 + R)^n$, where *Y* is the yield of amplification, *A* is the initial amount of cDNA, *R* is the amplification efficiency, and *n* is the number of cycles (7), we estimated that the quantity of the PCR bands around position 400 (Fig. 3) was 4.5-fold more than the actual amount of the transcripts. Therefore, the percentage of PCR band intensities was accordingly corrected. The data in this figure are the averages of two experiments, one of which is shown in Fig. 3C. The solid, empty, and stippled horizontal bars denote the corrected percentages of the intensities of the bands at the indicated positions.

ucts accumulated after 60 min of incubation (Fig. 7). The major cleavage site appeared to be at or near the position corresponding to the major in vivo poly(A) site, as estimated by comparison with a 93-nt transcript. More importantly, the in vitro cleavage at this site was dependent on a functional positioning element. A nonfunctional GAUAAA sequence did not cause the appearance of the major band (Fig. 7, lane 5). On the other hand, the minor in vitro cleavage sites appeared to be not directed by the created positioning elements. Furthermore, transcripts corresponding to these minor sites were not detected in vivo by the PCR mapping assay (Fig. 3). The discrepancy between the in vivo poly(A) sites and in vitro cleavage sites has also been observed in other yeast genes (1, 10).

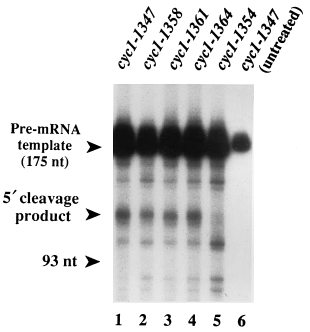
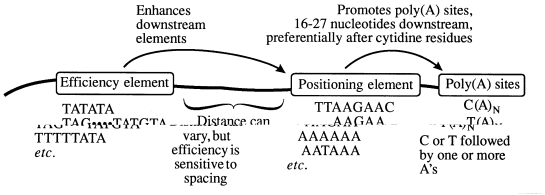


FIG. 7. UAUAUA, AAUAAA, and related sequences direct pre-mRNA cleavage in vitro. The capped ^{32}P-radiolabeled pre-mRNA templates were synthesized from plasmids pAB1414, pAB1425, pAB1423, pAB1428, and pAB1421, each of which contained one of the *CYC1* alleles (which are denoted at the top of the figure). One major and several minor 5′ cleavage products were observed in all the alleles, except for in *cyc1-1354*. No corresponding 3′ fragments were found to accumulate after the cleavage reaction. The pre-mRNA templates, the major 5′ cleavage products, and the position of a 93-nt transcript are indicated on the left.



DISCUSSION

The redundancy and degeneracy of yeast polyadenylation signals make predictions of the elements specifying 3′ end formation difficult. The efforts to inactivate yeast poly(A) signals by deletion or base pair substitutions have led to complicated results (10, 31). A possible solution to this problem is to use sequences from organisms other than yeasts or yeast sequences in which the poly(A) signal has been deleted, e.g., *cyc1-512*. By selection for the sequences restoring 3′ end formation in the poly(A) signal-free sequences, some elements of the yeast poly(A) signal have been successfully identified (32). In this study, the positioning element has been characterized by use of the same strategy. A general summary of the 3′-end-forming signals of yeast mRNA is presented in Fig. 8. Although the positioning element AAGAA positions poly(A) sites of *CYC1* mRNA, this sequence is not conserved in all yeast genes. We demonstrated here that part of the *GAL7* core signal, AATAATGA, functioned as a positioning element in a *cyc1-965* context. It is interesting to note that AATAATGA caused a pattern of poly(A) sites different from that of AA GAA. In *cyc1-1349*, in addition to positions 402 and 396, the use of the 387 site was also enhanced (Fig. 5), which was not observed in *cyc1-1186* (31). The most likely explanation could be that the 5′ boundary of the AATAATGA sequence is 3 bp upstream of that of AAGAA in *cyc1-1186*. The AATAA element of the AATAATGA sequence is probably the element responsible for 3′ end formation at position 387. We also demonstrated that several A-rich sequences, including the vertebrate poly(A) signal AATAAA, can function as positioning elements. It has been proposed that because AAUAAA and similar sequences are often found 20 to 30 nt upstream of the poly(A) sites of yeast mRNAs, this sequence has a possible role in mRNA 3′ end formation (2). However, no supporting evidence was reported. Deletion or base pair substitution analysis of AAUAAA showed that it was not absolutely required in the *ADE8* and *ADH2* genes (17, 19). These observations, along with our results in this study, are reminiscent of the study of the *CYC1* efficiency elements in which we found that none of the 10-bp deletions within the 38-bp region can completely abolish *CYC1* mRNA 3′ end formation because of the redundancy of the signal (13, 31).

As expected, saturation mutagenesis revealed that the positioning element is highly degenerate. Only one mutation, AAT AAA to GATAAA, abolished its activity. A guanosine residue at the first position could have an inhibitory effect on the immediately downstream sequence. We have observed that a TTAAGAAC sequence worked better than TGAAGAAC (31). While AATAAG acted strongly in *cyc1-1369* (Fig. 3C, lane 32), a GAATAA sequence (*cyc1-1187*) had little activity in the positioning of poly(A) sites (31). Therefore, the activity of 5-bp signals, such as AAAAA, AATAA, AAGAA, etc., may be greatly affected by the sequence context. On the other hand, the function of AATAAA seems to be not inhibited by a

TABLE 1. Putative efficiency and positioning elements for 3′ end formation of some yeast mRNAs[a]

Gene	Sequence
ACT1	*TAA*· · ·20· · ·TGTTTATGTATGTACCTCTCTCTCTATTTCTATTTTTAAACCACCCTCTCAATAAAATAAAAATAATAAAGATTT ↓ TTAAGG
ADH1	*TAA*· · ·7· · ·TCTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTT ↓ AGGTTTTAAA
CYC1	*TAA*· · ·120· · ·CCTATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAAATT ↓
GAL7	*TAA*· · ·100· · ·AGTGACAATATATATATATATATATATAATAATGACATCATTATCTGTAAAT ↓
GCN4	*TGA*· · ·12· · ·TTTATTTTATATTTTTTATTTCATTCTCGTGTATAACGAAATAGATACATTCACTTAGATAAGAATTTAATCTTT ↓ TTATGCCAATTT
MRP2	*TAA*· · ·14· · ·AAAGTATATATATATATATATAGGATACGCACTTGTAAATAAATACATGATAATTTTTATTCTTCTTTAAT ↓
RHO1	*TAG*· · ·154· · ·TCTTCTCTACATATAAGTGTATTGCTCAGTAAGTATTATCATCTATTAAAG

[a] The 3′ regions of seven yeast genes in which major poly(A) sites have been precisely mapped by sequencing the corresponding cDNAs (10, 14, 27, 31, 34) are shown. The translation termination codons for each of the genes are in italics and are followed by numbers designating the number of base pairs between the termination codon and the remaining sequence shown. The arrows denote the major poly(A) sites. The putative efficiency and positioning elements directing 3′ end formation at the major poly(A) site are, respectively, double underlined and single underlined. In some cases, there could be more than one element for one poly(A) site.

guanosine nucleotide immediately 5′ to this sequence (*cyc1-1351*) (Fig. 3C, lane 14).

Our saturation mutagenesis result was consistent with a result reported by Henikoff et al. (17), who found that the AATAAA sequence was not absolutely required for mRNA 3′ end formation in yeast cells. Henikoff et al. (17) studied a transcription unit from the *Drosophila melanogaster ADE8* segment in yeast cells. The *ADE8* gene produces a 1,000-nt-long transcript with three major poly(A) sites within a 30-nt region. An AATAAA sequence can be found approximately 10 to 30 bp upstream of the major poly(A) sites. It was demonstrated that the gene still gave a normal 1,000-nt transcript when the sequence AATAAA was deleted. Therefore, it was proposed that AATAAA did not play a role in mRNA 3′ end formation. However, it should be noted that deletion of sequence AAT AAA eliminated the most downstream one of the three major poly(A) sites, which was situated approximately 30 bp 3′ to the AATAAA sequence. From 10 to 20 bp upstream of the AAT AAA sequence are two copies of the AAAAA sequence, which we suggest directs the other two major poly(A) sites. Probably because of the presence of the AAAAA sequences, deletion of AATAAA did not cause a dramatic effect on the length and steady-state level of *ADE8* mRNA. In another report, the function of an AAUAAA sequence located 15 nt upstream of the *ADH2* mRNA poly(A) site was studied by Hyman et al. (19). The AATAAA sequence was first changed to AATTCA by site-directed mutagenesis. The 2-bp alteration did not result in a significant change in either the *ADH2* mRNA level or the position of the mRNA 3′ end. Our interpretation of this is that even though the two adenosine residues were changed to a thymidine and a cytidine, this region was still A rich. The resulting TCAAGATAA sequence, instead of AATAAA, could function as a positioning element. In another plasmid, the AA TAAA sequence was deleted. This mutant produced a lower level of normal *ADH2* mRNA, indicating the importance of this sequence.

In this study, several cleavage products were detected when pre-mRNA templates containing UAUAUA and AAUAAA (or related sequences) were tested with a yeast cell extract. The major one appeared to be directed by the AAUAAA sequence. A GAUAAA sequence or the wild-type GTCACG sequence did not cause the accumulation of this cleavage product (Fig. 7, lane 5) (data not shown). On the other hand, other cleavage sites did not appear to be specified by AAUAAA and related sequences. Interestingly, the use of these minor sites was increased in *cyc1-1354* (Fig. 7, lane 5), indicating that the total cleavage capacity was not greatly affected by the inactivation of the positioning element. This observation is consistent with our previous in vivo results, which showed that without a change of the efficiency element, the introduction of new poly(A) sites by creation of a functional positioning element usually decreased the use of other preexisting sites and vice versa (31).

In addition to the AAUAAA motif and the downstream GU-rich or U-rich element of mammalian systems, upstream sequences that contribute to the efficiency of mRNA 3′ end formation were also identified in the mammalian viruses simian virus 40 (6, 36), adenovirus (8, 9), hepatitis B virus (29), and human immunodeficiency virus type 1 (3, 11, 12, 38, 39). These elements function in an orientation- and position-dependent manner, and several appear to be functionally analogous. After a comparison of these upstream sequences, Schek et al. (36) proposed a consensus sequence, TATTTGTR, which was found in most of the upstream elements. In plants, a study of the polyadenylation signal from the cauliflower mosaic virus revealed that in addition to the essential element AATAAA, upstream sequences also have a great effect on processing efficiency, suggesting that one or several positively acting upstream elements are present (35). An oligonucleotide consisting of the cauliflower mosaic virus upstream sequences could activate a normally silent exogenous poly(A) signal if it were inserted upstream of its AATAAA motif. Strikingly, the cauliflower mosaic virus upstream element also contains two copies of UAUUUGUA, the conserved motif found in the mammalian virus upstream element. We suggest that the UAUUU GUA motif could be evolutionarily related to yeast efficiency element UAUAUA. In this study, we have shown evidence that highly conserved vertebrate poly(A) signal AAUAAA and related sequences are essential in the positioning of poly(A) sites in yeast cells. Furthermore, the bias for CA or $Py(A)_N$ as the actual poly(A) addition site has been observed not only in

mammalian systems but also in *S. cerevisiae*. Therefore, we believe that it is possible that some components of the basic polyadenylation machinery have been conserved among yeasts, plants, and mammals, although 3′ end formation in yeasts is clearly distinct from that in higher eukaryotes.

Although the efficiency and positioning elements have been characterized in detail in the *CYC1* gene, it remains to be seen if all other yeast genes utilize these elements for mRNA 3′ end formation (Table 1). As we described above, the *GAL7* gene appears to possess recognizable efficiency and positioning elements. The *ADH1* gene contains a strong positioning element, AAAAAAAA, while its efficiency element needs to be defined (15). Recently, Egli et al. (10) reported an extensive analysis of the *GCN4* mRNA 3′-end-forming signal. The major poly(A) site of *GCN4* mRNA is at a CAA sequence 95 bp downstream of the translation termination codon. Deletion of a 15-bp sequence located 30 bp upstream of the major poly(A) site resulted in the appearance of two additional poly(A) sites farther downstream. This 15-bp region contains an A-rich sequence, AAATAGA, which is worthy of further investigation to see if it is the element responsible for poly(A) site selection. Further deletion of another A-rich sequence, AAGAA, located 20 bp upstream of the wild-type poly(A) site appeared to decrease the usage of the wild-type poly(A) site, while the additional downstream poly(A) sites were not affected. The sequences determining the efficiency of *GCN4* mRNA 3′ end formation appeared to be dispersed in a relatively long region and remain to be elucidated. However, it was suggested that the TTTT TAT and TGTATA motifs could be involved (10).

It has been noted that DNA segments from several yeast genes, including *ADH1*, *CYC1*, *ARO4*, etc., can direct mRNA 3′ end formation in both orientations, while the 3′-end-forming signals of *GAL7*, *GCN4*, *PHO5*, etc., are functional in a strict orientation-dependent manner. On the basis of this observation, it was suggested that there were two different processing mechanisms for the two classes of 3′-end-forming signal (21). However, our study of the 3′-end-forming signals of *CYC1* and *GAL7*, which belong to two different classes, showed that they consisted of functionally similar efficiency and positioning elements and that the actual poly(A) site of both genes was at a PyAAA sequence (34, 42), suggesting that the 3′-end-forming signals of the two genes are not fundamentally different. Actually, the fact that a variety of pre-mRNAs, including *GAL7* and *CYC1*, are processed in the same cell-free system to yield mature mRNAs suggests that a common mechanism underlies this processing. This notion is consistent with a recent report by Preker et al. (28) that the *FIP1* gene, which encodes a yeast polyadenylation factor, was required for both *GAL7* and *CYC1* pre-mRNA processing. If the yeast mRNA 3′ ends are generated by a common processing mechanism, it would be tempting to further suggest that all the yeast genes use essentially the same signals for mRNA 3′ end formation. The apparent lack of consensus sequences may just be an inevitable result of signal redundancy and sequence variations.

ACKNOWLEDGMENTS

We thank J. S. Butler for providing the yeast cell extract and L. L. Comfort for the synthesis of the oligonucleotides used in this study.

This work was supported by National Institutes of Health grant RO1 GM 12702.

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. **Brown, P. H., L. S. Tiley, and B. R. Cullen.** 1991. Efficient polyadenylation within the human immunodeficiency virus type 1 long terminal repeat requires flanking U3-specific sequences. J. Virol. **65:**3340–3343.
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. **Carswell, S., and J. C. Alwine.** 1989. Efficiency of utilization of the simian virus 40 late polyadenylation site: effects of upstream sequences. Mol. Cell. Biol. **9:**4248–4258.
7. **Chelly, J., D. Montarras, C. Pniset, Y. Berwald-Netter, and J. C. Kaplan.** 1990. Quantitative estimation of minor mRNAs by cDNA-polymerase chain reaction. Eur. J. Biochem. **187:**691–698.
8. **DeZazzo, J. D., E. Falck-Pedersen, and M. J. Imperiale.** 1991. Sequences regulating temporal poly(A) site switching in the adenovirus major late transcription unit. Mol. Cell. Biol. **11:**5977–5984.
9. **DeZazzo, J. D., and M. J. Imperiale.** 1989. Sequences upstream of AAUAAA influence poly(A) site selection in a complex transcription unit. Mol. Cell. Biol. **9:**4951–4961.
10. **Egli, C. M., C. Springer, and G. H. Braus.** 1995. A complex unidirectional signal element mediates *GCN4* mRNA 3′ end formation in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **15:**2466–2473.
11. **Gilmartin, G. M., E. S. Fleming, and J. Oetjen.** 1992. Activation of HIV-1 3′ processing in vitro requires both an upstream element and TAR. EMBO J. **11:**4419–4428.
12. **Gilmartin, G. M., E. S. Fleming, J. Oetjen, and B. R. Graveley.** 1995. CPSF recognition of an HIV-1 mRNA 3′-processing enhancer: multiple sequence contacts involved in poly(A) site definition. Genes Dev. **9:**72–83.
13. **Guo, Z., P. Russo, D.-F. Yun, J. S. Butler, and F. Sherman.** 1995. Redundant 3′ end-forming signals for the yeast *CYC1* mRNA. Proc. Natl. Acad. Sci. USA **92:**4211–4214.
14. **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.
15. **Heidmann, S., C. Schindewolf, G. Stumpf, and H. Domdey.** 1994. Flexibility and interchangeability of polyadenylation signals in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **14:**4633–4642.
16. **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.
17. **Henikoff, S., J. D. Kelly, and E. H. Cohen.** 1983. Transcription terminates in yeast distal to a control sequence. Cell **33:**607–614.
18. **Hyman, L. E., and C. L. Moore.** 1993. Termination and pausing of RNA polymerase II downstream of yeast polyadenylation sites. Mol. Cell. Biol. **13:**5159–5167.
19. **Hyman, L. E., S. H. Seiler, J. Whoiskey, 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.
20. **Irniger, S., and G. H. Braus.** 1994. Saturation mutagenesis of a polyadenylation signal reveals a hexanucleotide element essential for mRNA 3′ end formation in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **91:**257–261.
21. **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.
22. **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.
23. **Kunkel, T. A., J. D. Roberts, and R. A. Zakoar.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154:**367–382.
24. **Li, W.-Z., and F. Sherman.** 1991. Two types of TATA elements for the *CYC1* gene of the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. **11:**666–676.
25. **Osborne, B. I., and L. Guarente.** 1988. Transcription by RNA polymerase II induces changes of DNA topology in yeast. Genes Dev. **2:**766–772.
26. **Osborne, B. I., and L. Guarente.** 1989. Mutational analysis of a yeast transcriptional terminator. Proc. Natl. Acad. Sci. USA **86:**4097–4101.
27. **Peterson, J. A., and A. M. Myers.** 1993. Functional analysis of mRNA 3′ end formation signals in the convergent and overlapping transcription units of the *S. cerevisiae* genes *RHO1* and *MRP2*. Nucleic Acids Res. **21:**5500–5508.
28. **Preker, P. J., J. Lingner, L. Minvielle-Sebastia, and W. Keller.** 1995. The *FIP1* gene encodes a component of the yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase. Cell **81:**379–389.
29. **Russnak, R., and D. Ganem.** 1990. Sequences 5′ to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses. Genes Dev. **4:**764–776.
30. **Russo, P.** 1995. *Saccharomyces cerevisiae* mRNA 3′ end forming signals are also involved in transcription termination. Yeast **11:**447–453.

31. **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.
32. **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.
33. **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.
34. **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.
35. **Sanfacon, H., P. Brodmann, and T. Hohn.** 1991. A dissection of the cauliflower mosaic virus polyadenylation signal. Genes Dev. **5:**141–149.
36. **Schek, N., C. Cooke, and J. C. Alwine.** 1992. Definition of the upstream efficiency element of the simian virus 40 late polyadenylation signal by using in vitro analyses. Mol. Cell. Biol. **12:**5386–5393.
37. **Sheets, M. D., S. C. Ogg, and M. P. Wickens.** 1990. Point mutations in AAU AAA and the poly(A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation *in vitro*. Nucleic Acids Res. **18:**5799–5805.
38. **Valsamakis, A., N. Schek, and J. C. Alwine.** 1992. Elements upstream of the AAUAAA within the human immunodeficiency virus polyadenylation signal are required for efficient polyadenylation in vitro. Mol. Cell. Biol. **12:**3699–3705.
39. **Valsamakis, A., S. Zeichner, S. Carswell, and J. C. Alwine.** 1991. The human immunodeficiency virus type 1 polyadenylation signal: a 3′ long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylation. Proc. Natl. Acad. Sci. USA **88:**2108–2112.
40. **Wahle, E., and W. Keller.** 1992. The biochemistry of 3′-end cleavage and polyadenylation of messenger RNA precursors. Annu. Rev. Biochem. **61:**419–440.
41. **Yu, K., and R. T. Elder.** 1989. Some of the signals for 3′-end formation in transcription of the *Saccharomyces cerevisiae* Ty-D15 element are immediately downstream of the initiation site. Mol. Cell. Biol. **9:**2431–2444.
42. **Zaret, K. S., and F. Sherman.** 1982. DNA sequences required for efficient transcription termination in yeast. Cell **28:**563–573.