MOLECULAR AND CELLULAR BIOLOGY, Dec. 1993, p. 7836–7849
0270-7306/93/127836-14$02.00/0

Vol. 13, No. 12

Signals That Produce 3′ Termini in *CYC1* mRNA of the Yeast *Saccharomyces cerevisiae*

PATRICK RUSSO,[1]† WEN-ZHUO LI,[1] ZIJIAN GUO,[1] AND FRED SHERMAN[1,2]*

Departments of Biochemistry[1] and Biophysics,[2] University of Rochester School of Medicine and Dentistry, Rochester, New York 14642*

Received 11 June 1993/Returned for modification 15 July 1993/Accepted 16 September 1993

The *cyc1-512* mutant was previously shown to contain a 38-bp deletion, 8 nucleotides upstream from the major wild-type poly(A) site, in the *CYC1* gene, which encodes iso-1-cytochrome *c* of the yeast *Saccharomyces cerevisiae*. This 38-bp deletion caused a 90% reduction in the *CYC1* transcripts, which were heterogeneous in size, aberrantly long, and presumably labile (K. S. Zaret and F. Sherman, Cell 28:563–573, 1982). Site-directed mutagenesis in and adjacent to the 38-bp region was used to identify signals involved in the formation and positioning of *CYC1* mRNA 3′ ends. In addition, combinations of various putative 3′ end-forming signals were introduced by in vitro mutagenesis into the 3′ region of the *cyc1-512* mutant. The combined results from both studies suggest that 3′ end formation in yeast cells involves signals having the following three distinct but integrated elements acting in concert: (i) the upstream element, including sequences TATATA, TAG · · · TATGTA, and TTTTTATA, which function by enhancing the efficiency of downstream elements; (ii) downstream elements, such as TTAAGAAC and AAGAA, which position the poly(A) site; and (iii) the actual site of polyadenylation, which often occurs after cytidine residues that are 3′ to the so-called downstream element. While the upstream element is required for efficient 3′ end formation, alterations of the downstream element and poly(A) sites generally do not affect the efficiency of 3′ end formation but appear to alter the positions of poly(A) sites. In addition, we have better defined the upstream elements by examining various derivatives of TATATA and TAG · · · TATGTA, and we have examined the spatial requirements of the three elements by systematically introducing or deleting upstream and downstream elements and cytidine poly(A) sites.

Both the proper termination of mRNA transcription and the processing of the transcripts' 3′ ends are crucial steps in gene regulation, which, if aberrant, can lead to drastic reductions in gene expression. Such a case was found in the mutant allele *cyc1-512* of the *Saccharomyces cerevisiae* gene *CYC1*, which encodes iso-1-cytochrome *c* (19). *cyc1-512* strains displayed a 90% decrease in *CYC1*$^+$ mRNA and iso-1-cytochrome *c* that was caused by the deletion of a 38-bp region 8 nucleotides (nt) upstream from the normal poly(A) site. This *cyc1-512* mutation also caused aberrantly long transcripts with many discrete 3′ termini ranging from the wild-type poly(A) site to endpoints greater than 2,000 nt downstream. Butler and Platt (2) found that the same 38-bp region was necessary for the proper cleavage of extended *CYC1* transcripts in vitro. In addition, the same 38-bp region was found by Russo and Sherman (16) to be able to cause the termination of transcription in vivo. This conclusion concerning transcription termination was based on the loss of plasmid stability conferred by the *CEN3* element when transcription impinged upon this region, and restoration of plasmid stability by the introduction of a segment encompassing the 38-bp region, between the transcription promoter and the *CEN3* element. In another study, signals required for 3′ end formation were investigated by examining intragenic revertants and oligonucleotide-directed alterations that reverted the *cyc1-512* defect (15). The study of Russo et al. (15) revealed that the production of yeast mRNA 3′ termini may involve at least two functionally distinct sequence elements working in concert. One type of element determines the sites of preferred 3′ mRNA termini, as represented by the *cyc1-512* termini. The second type of element, which includes TAG · · · TATGTA and TATATA motifs, operates at a distance to enhance the use of the downstream 3′ preferred sites. Further support for the involvement of these sequences in 3′ endpoint formation comes from studies of Yu and Elder (18) and Irniger et al. (7, 8). Another study (5) indicated that TTTTTATA also may be involved in 3′ end formation. In contrast, Osborn and Guarente (11), using deletion analysis through the 38-bp region, suggested that no specific sequence signals are involved and that only A+T-rich regions are necessary for mRNA 3′ end formation in yeast cells.

To further our understanding of which sequences in the 38-bp region of the *CYC1* gene are involved in mRNA 3′ end formation, we have systematically made alterations throughout this region and examined the corresponding levels of *CYC1* mRNA and iso-1-cytochrome *c* as well as the efficiency and position of *CYC1* mRNA 3′ end formation. Similar analyses were performed on alterations made in the 3′ region of the *cyc1-512* allele that gave rise to putative 3′ end-forming signals and that resulted in various degrees of restitution of *CYC1* expression. Our results suggest that three distinct but integrated elements are involved in yeast mRNA 3′ end formation: the upstream element, represented by the sequences TATATA, TAG · · · TATGTA, and TTT TTATA, which enhance the efficiency of the downstream element; the downstream elements, represented by such sequences as TTAAGAAC and AAGAA, which position the poly(A) sites; and the actual site of polyadenylation that

* Corresponding author. Electronic mail address: fsrm@bphvax.biophysics.rochester.edu.

† Present address: Institute of Biotechnology, University of Helsinki, SF-00380 Helsinki, Finland.

often occurs after cytidine residues that are 3′ to the downstream element.

MATERIALS AND METHODS

Nomenclature and strains. The designation *CYC1*$^+$ denotes the wild-type allele encoding iso-1-cytochrome *c*. *CYC1* refers to any allele at this locus. The lowercase designations *cyc1-512*, *cyc1-965*, etc., denote altered *CYC1* alleles, irrespective of whether the allele causes deficiency.

Escherichia coli HB101 (*pro leu thr lacY hsdR endA recA rpsL20 ara-14 gal2 xyl-15 mtl-1 supE4-1*) was used for the propagation of plasmid vectors. *E. coli* BW313 (*dut ung thi-1 relA spoT1* F′ *lysA*) was used in the process of site-directed in vitro mutagenesis. *S. cerevisiae* B-7467 (*MAT***a** *cyc1-Δ1 CYC7*$^+$ *ura3-52*) served as the recipient of the YCp50 plasmid derivatives containing wild-type and altered *CYC1* genes (15).

Oligonucleotide site-directed mutagenesis in vitro. Oligonucleotides used for site-directed mutagenesis, for polymerase chain reaction (PCR), and for sequencing were synthesized with an Applied Biosystems 380A DNA synthesizer. Oligonucleotide-directed mutagenesis in vitro was performed by the method of Kunkel et al. (9). Plasmid pAB458 (3a) is a derivative of plasmid pEMBLY:32 (2b), which contains the wild-type *CYC1*$^+$ gene in a *Bam*HI-*Hin*dIII fragment. Plasmid pAB639 is also a derivative of pEMBLY:32, but it contains the fragment encompassing the *cyc1-512* allele. Site-directed deletions and alterations were made in the 3′ region of the *CYC1*$^+$ gene in pAB458, while putative 3′ end-forming sequences were inserted in the 3′ region of the *cyc1-512* allele in pAB639.

Plasmid constructions. Plasmid pAB640 was made by inserting a 4.7-kb *Bam*HI-*Eco*RI fragment, containing the entire *CYC1*$^+$ gene, into the yeast shuttle vector YCp50 (14). Plasmid pAB650, also a YCp50 derivative, contains a 4.7-kb *Bam*HI-*Eco*RI fragment that has the *cyc1-512* allele. Site-directed mutations, made in plasmid pAB458 or pAB639, were introduced into plasmids pAB640 and pAB650 as *Kpn*I-*Hin*dIII fragments.

Iso-1-cytochrome *c* determination. The relative amounts of iso-1-cytochrome *c* produced in *S. cerevisiae* B-7467 harboring different *CYC1* mutants were determined by low-temperature (−196°C) spectroscopic examination of intact cells (17) and by comparisons with yeast strains having known amounts of iso-1-cytochrome *c*.

Northern (RNA) blot analysis. Northern blot analysis of total RNA isolated from *S. cerevisiae* B-7467 and transformants was performed exactly as outlined by Russo et al. (15).

PCR mapping of mRNA 3′ termini. The 3′ termini of *CYC1* mRNAs were estimated from the sizes of fragments generated by reverse transcription PCR as described by Russo et al. (15), but with the following different primers and modifications:

```
Primer 1,                                          AA
5′-GACGAGCTCGGATCCTGCAGTTTTTTTTTTTTTTTTTCC-3′
                                                   GG
Primer 2,                                           T
5′-GACGAGCTCGGATCCTGCAGTTTTTT-3′
Primer 3,
5′-ATTCCTGGTACCAAGATGGCCTTTGG-3′
Primer 4,
5′-CACGCTTACATTCACGCCCTCCCCC-3′
```

Primer 1 was made degenerative at the last two nucleotides of the 3′ end to produce an equal mix of 12 primers. Instead of separating the samples on ethidium bromide-stained agarose gels, we radiolabeled primer 4 with ^{32}P, using the T4 polynucleotide kinase reaction of Maxam and Gilbert (10). Subsequently, PCR products were separated on 6 or 8% polyacrylamide sequencing gels (10), with a sequencing ladder to act as a size standard.

PCR sequencing of mRNA 3′ termini. The poly(A) sites of *CYC1* mRNAs were precisely determined by sequencing PCR fragments encompassing 3′ termini. The mRNA 3′ termini were amplified by PCR with primer 2 and primer 3 in a 100-μl reaction volume as described above. The amplified DNA fragments were separated from excess primers and deoxynucleoside triphosphates by using a Millipore Ultrafree-MC filter unit (Polysulfone type PTTK ultrafiltration membrane; 30,000-NMWL [nominal molecular weight limit] cutoff) as recommended by the manufacturer. The concentrate was resuspended in a final volume of 25 μl of double-distilled H_2O. Sequencing of the PCR products was performed by using Sequenase, version 2.0 (U.S. Biochemical). Five to 7 μl of the purified DNA fragments was mixed with 1.5 pmol of primer 4 and Sequenase reaction buffer. Standard Sequenase reaction conditions were used for the sequencing reaction.

RESULTS

Overall strategy. The nucleotide sequence requirement for 3′ end formation was investigated by making site-directed alterations in the 3′ untranslated regions of two plasmids, pAB640, which contains the wild-type *CYC1*$^+$ allele, and pAB650, which contains the *cyc1-512* allele. Deletions and replacements were introduced into the *CYC1*$^+$ plasmid for possible diminutions or repositioning of 3′ ends, whereas insertions and replacements were introduced in the *cyc1-512* plasmid for possible enhancement or formation of 3′ ends. The single-copy plasmids were introduced into strain B-7467, which lacks the chromosomal *CYC1* gene. The levels of the *CYC1* mRNAs were estimated by Northern analysis and from the amount of iso-1-cytochrome *c* as determined by low-temperature (−196°C) spectroscopic examination of intact cells. Initially, the positions of the 3′ ends were estimated from the sizes of PCR-generated fragments. More precise determination of certain transcript 3′ ends was accomplished by sequencing the PCR fragments. The following are summarized in Table 1: allele, strain, and plasmid designations; mutational alterations; and the results of the spectral, Northern, and PCR 3′ end mapping analyses. The alterations in the *CYC1*$^+$ and *cyc1-512* plasmids are shown in Fig. 1 and 2, respectively. The results from sequencing the PCR fragments are presented in Table 2. The direct results from Northern analysis, PCR 3′ end mapping, and the sequencing of PCR fragments are presented in Fig. 3, 4, and 5, respectively. The numbers designating the positions of the 3′ ends correspond to the assignment of the adenosine nucleotide of the ATG initiator codon as position 1. Each number in brackets in Tables 1 and 2 denotes the position of the preceding nucleotide, using the numbering convention stated above. The numerical values of the endpoints were adjusted for the number of nucleotides deleted or added.

Control values. As expected, the negative control strains, B-7467 and B-8034, lacked *CYC1*$^+$ mRNA and iso-1-cytochrome *c*. The extrachromosomal wild-type *CYC1*$^+$ gene of plasmid pAB640 in strain B-8035 produced the normal level of iso-1-cytochrome *c* and displayed the normal transcript size of approximately 630 nt. PCR 3′ end mapping revealed the major mRNA 3′ endpoint to be around position 502 and

TABLE 1. Properties of mutants with alterations in the 3′ region of *cyc1-512* and *CYC1*[a]

CYC1 allele	Strain	Plasmid	Description or alteration	% Normal iso-1[b]	*CYC1* mRNA (nt)	
					Size (Fig. 2)	Approx positions of 3′ endpoints (Fig. 3)
Controls						
cyc1$^-$	B-7467	None	Absence of *CYC1*$^+$	<5	None	None
cyc1$^-$	B-8034	YCp50	Absence of *CYC1*$^+$	<5	None	None
CYC1$^+$	B-8035	pAB640	*CYC1*$^+$ in YCp50 (Fig. 1A)	100	630	468, 475, 502, 525
cyc1-512	B-8036	pAB650	38-bp deletion (Fig. 2A)	10	630 and larger	429, 435, 502, 524, 560, 576, and more downstream
Deletions and replacements in the 38-bp region of *CYC1*$^+$ (Fig. 1B)						
cyc1-984	B-8056	pAB649	*Pst*I ΔGTTATGTTAGTATTAAGAAC[488]	100	630	(No 502) 522
cyc1-985	B-8057	pAB648	ΔTTAAGAACGTTATTTAT[497]	100	630	(No 502) 522
cyc1-983	B-8055	pAB647	ΔTTTTTTATAGTTATGTTAGTA[480]	80	630 and larger	502, 522, 557
cyc1-982	B-8054	pAB641	ΔGTCTAGGTCC[451]	100	630	502, 523
cyc1-978	B-8050	pAB642	ΔTTTTTTTATA[468]	100	630	502, 523
cyc1-977	B-8049	pAB643	ΔGTTATGTTAGT[479]	100	630	466, 473, 502, 522
cyc1-979	B-8051	pAB644	ΔATTAAGAACG[489]	100	630	(No 502) 522
cyc1-980	B-8052	pAB645	ΔTTATTTATAT[499]	100	630	511, 522
cyc1-981	B-8053	pAB646	ΔTTTCAAATTT[508]	100	630	511, 521
cyc1-1183	B-8501	pAB913	TTAAGAAC[488]→TTCTGCAG	100	ND	502, 511, 522
cyc1-1184	B-8502	pAB914	ΔC[502]	100	ND	502, 522
Creation of putative 3′ end-forming signals in *cyc1-512* (Fig. 2B)						
cyc1-965	B-8037	pAB651	TATATA[357]	100	630	381, 396, 402, 429, 435, 502, 522, 556
cyc1-966	B-8038	pAB733	CATATA[358]	40	630	502, 522, 556
cyc1-968	B-8040	pAB735	TATACA[358]	30	630 and larger	429, 435, 502, 524
cyc1-967	B-8039	pAB734	TACATA[358]	90	630	502, 522, 554
cyc1-969	B-8041	pAB736	CACACA[358]	10	630 and larger	502, 522, 554, and more downstream[c]
cyc1-970	B-8042	pAB737	TAGATATATGTAA[363]	100	630 and larger	502, 522
cyc1-971	B-8043	pAB738	TAGTTATGTT[359]	30	630 and larger	502, 522, 554
cyc1-972	B-8044	pAB739	TTTTTTTATA[355]	65	630 and larger	502, 522, 554
cyc1-1006	B-8116	pAB911	TTTTTTTATA 14 nt 5′ to 502	90	630	(No 502) 522, 556
cyc1-1007	B-8115	pAB912	TATATA 14 nt 5′ to 502	100	630	(No 502) 522, 556
cyc1-973	B-8045	pAB740	*Pst*I AATAAA 5 nt 5′ to 502	10	630 and larger	(No 502) 511, 523, and more downstream[c]
cyc1-974	B-8046	pAB654	*Pst*I TTTTTATA 3 nt 5′ to 502	10	630 and larger	457, 502, 523, 557, and more downstream[c]
cyc1-975	B-8047	pAB653	*Pst*I TATATA 3 nt 5′ to 502	40	630 and larger	457, 502, 523, 557, and more downstream[c]
cyc1-976	B-8048	pAB655	*Pst*I TAGTTATGTT 1 nt 5′ to 502	10	630 and larger	455, 501, 523, and more downstream[c]
cyc1-1193	B-8512	pAB652	*Pst*I 11 nt 5′ to 502	10	630 and larger	457, 502, 523, 557, and more downstream[c]
Combinations of putative 3′ end-forming signals created in the 3′ region of *cyc1-512* (Fig. 2B)						
cyc1-1185	B-8503	pAB915	TATATA-13 nt-TTAAGAAC[378]	100	ND	396, 402, 429, 435, 504, 524
cyc1-1186	B-8504	pAB916	TATATA-15 nt-AAGAA[377]	95	ND	396, 402, 429, 435, 502, 524
cyc1-1187	B-8505	pAB917	TATATA-15 nt-AATAA[377]	100	ND	381, 396, 402, 429, 435, 502, 524
cyc1-1188	B-8506	pAB918	No TATATA, just AAGAA[379]	10	ND	396, 403, 502, 524, 557, and more downstream
cyc1-1189	B-8507	pAB919	TATATA-25 nt-AAGAA[388]	100	ND	382, 402, 412, 429, 502, 524
Combination of putative 3′ end-forming signals created in the coding region of *cyc1-512* (Fig. 2B)						
cyc1-1190	B-8508	pAB920	TATATA-12 nt-AAGAA[321]	20	ND	502, 524, 557, and more downstream[c]
Combination of putative 3′ end-forming signals created 3′ to the 502 poly(A) site of *cyc1-512* (Fig. 5B)						
cyc1-1191	B-8509	pAB921	TATATA-15 nt-AAGAA[548]	100	ND	(No 502) 556, 564, 571, 581, 588, and more downstream
Combinations of putative 3′ end-forming signals created in *CYC1* (Fig. 2B)						
cyc1-1192	B-8510	pAB922	TATATA-13 nt-TTAAGAAC[378]	150	ND	396, 402, 502, 522
cyc1-1193	B-8511	pAB923	TATATA-15 nt-AAGAA[377]	95	ND	396, 402, 502, 522

[a] Δ denotes the deletion of the sequence that follows. Sequences of the alterations are presented in Fig. 1B and 2B. Each numbers in brackets denotes the position of the preceding nucleotide relative to the ATG translation initiator codon, where A is assigned +1; numbers for endpoint positions follow the same convention. The most prominent 3′ endpoint (the darkest band on the 3′ end mapping gel) is underlined; two are underlined when both are approximately equal. *Pst*I, position where a *Pst*I restriction site was created by site-directed mutagenesis; ND, not determined.

[b] Amount of the iso-1-cytochrome *c* protein relative to the wild-type level, as determined by spectral analysis.

[c] 3′ endpoint pattern similar to that of *cyc1-512*.

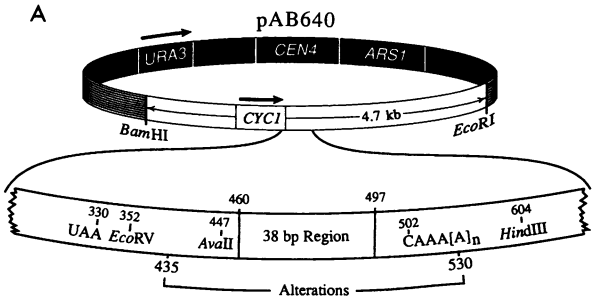
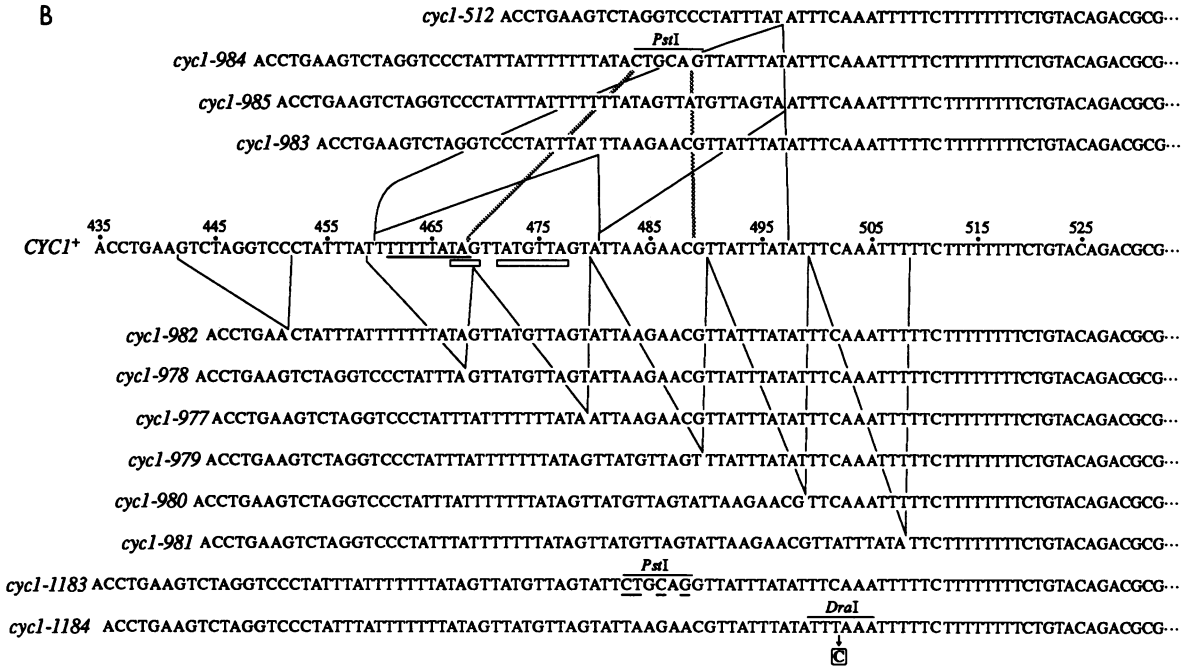


FIG. 1. (A) Schematic representation of the *CYC1* yeast expression plasmid pAB640. The approximately 4.7-kb *Bam*HI-*Eco*RI fragment encompasses the *CYC1*$^+$ gene and flanking regions. The 3′ region of the *CYC1* gene is expanded at the bottom and shows the positions of pertinent nucleotides relative to the A of the ATG initiator codon, which is assigned position 1; 330 corresponds to the last nucleotide of the *CYC1* termination codon UAA; 354, 447, and 604 denote the nucleotides directly preceding the cleavage points of the restriction enzymes *Eco*RV, *Ava*II, and *Hin*dIII, respectively; 460 and 497 are the flanking positions of the 38-bp region deleted in the *cyc1-512* mutant; 502 is the position of the cytidine preceding the wild-type poly(A) site; and 435 to 530 defines the range of the site-directed mutations, which are presented in panel B. (B) Details of site-directed mutations made in the *CYC1* gene. The allelic designations are shown at the left of each sequence. A portion of the wild-type *CYC1*$^+$ sequence is shown in the center, numbered by the convention described above. The deleted sequences are shown on the *CYC1*$^+$ wild-type sequence between two lines that converge at the breakpoints, revealing the new sequences generated by the deletions. The putative 3′ end-forming sequences TTTTTATA (5) and TAG · · · TATGTTA (15) are indicated in the *CYC1*$^+$ wild-type sequence by an underline and an open bar, respectively. The nucleotide substitutions in the allele *cyc1-1183*, in which the sequence AAGAA is changed to CTGCA and a *Pst*I site is generated, are denoted by underlines. The cytidine deleted at position 502 in the allele *cyc1-1184* is indicated.



minor endpoints to be around position 525. PCR sequencing showed that the major *CYC1*$^+$ polyadenylation site was after the cytidine at position 502, but the exact position was indeterminant because three adenosine nucleotides immediately follow the cytidine at 502. The *cyc1-512* allele of plasmid pAB630 in strain B-8036 showed similarities to the original *cyc1-512* mutant for iso-1-cytochrome *c* and *CYC1* mRNA levels and for the mRNA 3′ endpoint pattern. These results assured us that *CYC1* mRNA and iso-1-cytochrome *c* expression from these alleles could be used as references for the expression of *CYC1* mRNA by the following mutations.

Deletions and replacements in the 38-bp region of *CYC1*$^+$. The six deletions *cyc1-977* to *cyc1-982* lacked 10 or 11 contiguous nt in and around the 38-bp region of the wild-type *CYC1*$^+$ allele, with no substitutions. Although none of these deletions produced observable changes in iso-1-cytochrome *c* protein or *CYC1* mRNA amounts relative to wild-type *CYC1*$^+$ expression, some of the deletions affected *CYC1* mRNA 3′ endpoint position. For example, deleting the sequence TTTTTTTATA, resulting in allele *cyc1-978*, which includes the first 9 nt of the 38-bp region and the putative 3′ end-forming signal TTTTTATA proposed by Henikoff and Cohen (5), caused 3′ endpoints around 523 to increase in use relative to the 502 position. Furthermore, PCR sequencing results showed that polyadenylation was occurring at both the wild-type poly(A) site around position 502 and down-

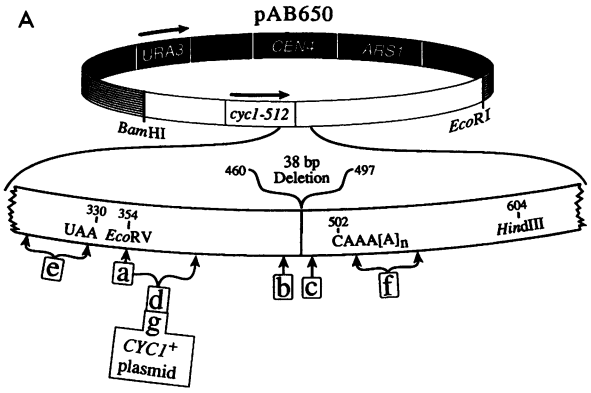


FIG. 2. (A) Schematic representation of the *cyc1-512* yeast expression vector. Plasmid pAB650 was derived from pAB640 by replacing the wild-type *CYC1*$^+$ *Kpn*I-*Hin*dIII fragment with the homologous fragment from the *cyc1-512* mutant that lacked the 38-bp region (nt 460 to 497). The numbering convention is the same as in Fig. 1. The letters under the expanded region present the approximate locations of the site-directed mutations, which are shown in detail in panel B. (B) Details of site-directed mutations made in *cyc1-512* (a to f) and *CYC1*$^+$ (g). The allele designations are displayed on the left of each sequence. The putative 3′ end-forming signals created by site-directed mutagenesis are underlined. The position of the 38-bp deletion is denoted by "DELETION."

B

(a)

```
                    350 EcoRV 360          455 DELETION          510       520
cyc1-512  TTTGTCGATATCATGTAA···TATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-965  TTTGTCTATATAATGTAA···TATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-966  TTTGTCATATATGTAA···TATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-967  TTTGTCTACATATGTAA···TATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-968  TTTGTCTATACATGTAA···TATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-969  TTTGTCACACATGTAA···TATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-970  TTTGTAGATATATGTAA···TATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-971  TTTGTAGTTATGTTGTAA···TATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-972  TTTTTTTATATCATGTAA···TATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
```

(b)

```
                    AvaII  455 DELETION          510       520
cyc1-512*  AAGTCTAGGTCCCTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-1006  AAGTCTTTTTTTATACCCTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-1007  AAGTCTATATACCCTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
```

(c)

```
                    PstI   455 DELETION          510       520
cyc1-1193  TAGGTCCCCTGCAGTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-973   TAGGTCCCCTGCAGAATAAATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-974   TAGGTCCCCTGCAGTTTTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-975   TAGGTCCCCTGCAGTATATATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
cyc1-976   TAGGTCCCCTGCAGTAGTTATGTTTCAAATTTTTCTTTTTTTTCTGTACAGACGCG···
```

(d)

```
                350 EcoRV 360       370       380       390       400
cyc1-512   TCGATATCATGTAATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCACATCC···
cyc1-1185  TCTATATAATGTAATTAGTTATTAAGAACTTACATTCACGCCCTCCCCCCACATCC···
cyc1-1186  TCTATATAATGTAATTAGTTATGAAGAACTTACATTCACGCCCTCCCCCCACATCC···
cyc1-1187  TCTATATAATGTAATTAGTTATGAATAACTTACATTCACGCCCTCCCCCCACATCC···
cyc1-1188  TCGATATGATGTAATTAGTTATGTCAAGAATACATTCACGCCCTCCCCCCACATCC···
cyc1-1189  TCTATATAATGTAATTAGTTATGTCACGCTTACAAGAAGCCCTCCCCCCACATCC···
```

(e)

```
              -300— TRANSLATED REGION — 320———|END 333
cyc1-512   TTAATTACCTACTTAAAAAAAGCCTGTGAGTAAACA···
cyc1-1190  TTATATACCTACTTAAAAAAAGAATGTGAGTAAACA···
```

(f)

```
                520       530       540       550       560       570
cyc1-512   CTGTACAGACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTT···
cyc1-1191  CTGTATATAAGCGTGTACGCATGTAAGAATATACTGAAAACCTTGCTTGAGAAGGTTT···
```

(g)

```
                350 EcoRV 360       370       380       390       400
CYC1+      TCGATATCATGTAATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCACATCC···
cyc1-1192  TCTATATAATGTAATTAGTTATTAAGAACTTACATTCACGCCCTCCCCCCACATCC···
cyc1-1193  TCTATATAATGTAATTAGTTATGAAGAACTTACATTCACGCCCTCCCCCCACATCC···
```

TABLE 2. Sequences of PCR-amplified regions encompassing 3′ endpoints

CYC1 allele	Sequence of 3′ endpoint (Fig. 5)[a]
CYC1$^+$	TTTATATTTC[502]**A$_{18}$CTGCAGG**
cyc1-965	ATTTC[502]**A$_{18}$CTGCAGGATCC**
cyc1-978	TATATTTC[502]**A$_n$** and TTTTTTTTC[520]**A$_n$**
cyc1-979	TTTCAAATTTTTCTTTTTTTTTC[520]**A$_n$** and TTTTTTTTCTGTAC[525]**A$_n$**
cyc1-981	TATTTATATTC[511]**A$_n$**
cyc1-983	TTTCAAATTTTTCTTTTTTTTC[520]**A$_n$** and TTTTTTTTCTGTAC[525]**A$_n$**
cyc1-984	TTTCAAATTTTTCTTTTTTTTC[520]**A$_n$** and TTTTTTTTCTGTAC[525]**A$_n$**
cyc1-985	TTTCAAATTTTTCTTTTTTTTC[520]**A$_{18}$**
cyc1-1007	TTTTTCTTTTTTTTC[520]**A$_n$** and TTTTTCTTTTTTTTCTGTAC[525]**A$_n$**
cyc1-1183	TTTCAAATTTTTCTTTTTTTTC[520]**A$_n$** and TTTTTCTTTTTTTTCTGTAC[525]**A$_n$**

[a] Each number in brackets refers to the position of the preceding nucleotide, which is the final transcribed residue. (The numbering system is described in Table 1, footnote *a*). Nucleotides in boldface indicate sequences encoded by the poly(A) tail and the oligonucleotide primers (see Materials and Methods, PCR mapping of 3′ termini.) **A$_{18}$** denotes 18 adenosine residues, which corresponds to the number of thymidine residues in the oligonucleotide primers used to synthesize the cDNA; **A$_n$** denotes an indeterminant number of repeated adenosines. These represent the most prominent signals seen on the sequencing gels, although other, weaker sequences could be observed for some alleles (Fig. 5).

stream immediately after the cytidine at position 520. This finding suggests that this deletion partially diminished the efficiency of 3′ end formation at the major wild-type position, which in turn caused downstream minor endpoints of the original wild-type allele to be more efficiently utilized.

A more dramatic effect on mRNA 3′ end position was seen in *cyc1-979*, which has a deletion of ATTAAGAACG. This deletion caused the complete disuse of the major wild-type 3′ endpoint around 502. Instead, 3′ endpoints around 522 became the major mRNA 3′ termini according to PCR mapping. PCR sequencing confirmed that the wild-type poly(A) was no longer used and showed that polyadenylation occurred immediately after the cytidine at 520 and probably immediately after a cytidine at 525. Additional mutations which corroborate the importance of signals within this deletion responsible for positioning the site of polyadenylation will be discussed below. Two other mutations which changed the position of the major polyadenylation site deleted the sequences TTATTTATAT and TTTCAAATTT, resulting in alleles *cyc1-980* and *cyc1-981*, respectively. Both deletions caused the major poly(A) site to occur after the cytidine at position 511. Interestingly, though each deletion removed completely separate sequences of the same size, 10 bp, between the sequence TTAAGAACG and the new poly(A) site at 511, both mutants conserved a distance of 16 bp between the center of TTAAGAACG and the new poly(A) site at 511. Furthermore, the actual site of polyadenylation utilized by both mutant alleles again appears to have a bias for cytidine residues.

Because none of the short deletions produced the same drastic effects as the original 38-bp deletion did, larger deletions in the 38-bp region of *CYC1* were analyzed. For instances, the first 21 bp, TTTTTTATAGTTATGTTAGTA, of the 38-bp region were deleted in allele *cyc1-983*. Spectral analysis of the *cyc1-983* mutant revealed a 20% diminution in the level of iso-1-cytochrome *c* compared with the *CYC1*$^+$ wild-type strain. Northern analysis showed no obvious reduction in the amount of *CYC1* transcripts, but there was the appearance of aberrantly longer transcripts above the normal-size band. 3′ end mapping revealed that termini centered around 502 were less intense than the endpoints centered around 522, the inverse of the situation in the *CYC1*$^+$ wild-type gene, and additional termini around 557 appeared. Furthermore, 3′ end sequencing also suggested that the major poly(A) site was no longer after the cytidine at 502 but instead after cytidine nucleotides at 520 and 525. This decrease in use of the endpoint at 502 and increase in endpoints at 520 and 525 is consistent with but greater than that seen when the first 10 bp of this deletion were removed in allele *cyc1-978*. This may be explained by the simultaneous deletion of the putative mRNA 3′ end-forming signals TTTTTATA (5) and TAG · · · TATGTA (15) by this 21-bp deletion, both of which may affect the efficiency of 3′ end formation at the wild-type poly(A) site. In addition, the increase in use of downstream poly(A) sites suggests the existence of other 3′ end-forming signals in the area, and the preference for cytidine residues at the polyadenylation site is reinforced.

The next larger deletion analyzed, in allele *cyc1-985*, lacked the last 17 bp, TTAAGAACGTTATTTAT, of the

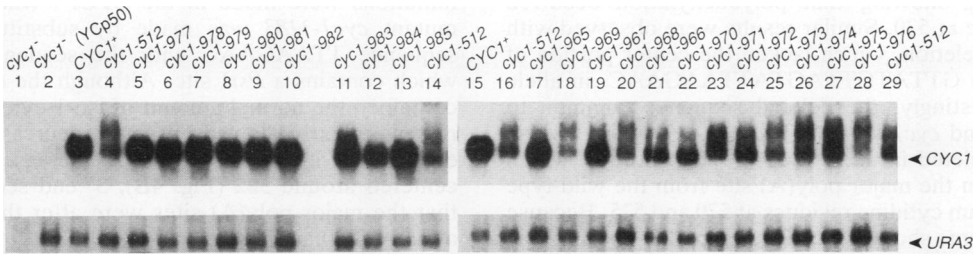


FIG. 3. mRNA blot analysis of site-directed mutants of *CYC1*$^+$ and *cyc1-512*. Northern analysis was used to detect gross changes in amount or size of *CYC1*-specific transcripts. Blots were first probed with a *CYC1*-specific probe, autoradiographed, washed to completely remove the *CYC1* signal, and then probed with a *URA3*-specific probe, which constituted the internal control. This hybridization regimen was carried out because a truncated transcript from the genomic *ura3-52* gene appears on Northern blots in the area of the *CYC1* transcript. Lanes 1 and 2 represent yeast strains B-7467 and B-8034, respectively. The alleles designated above the lanes are described in Table 1.

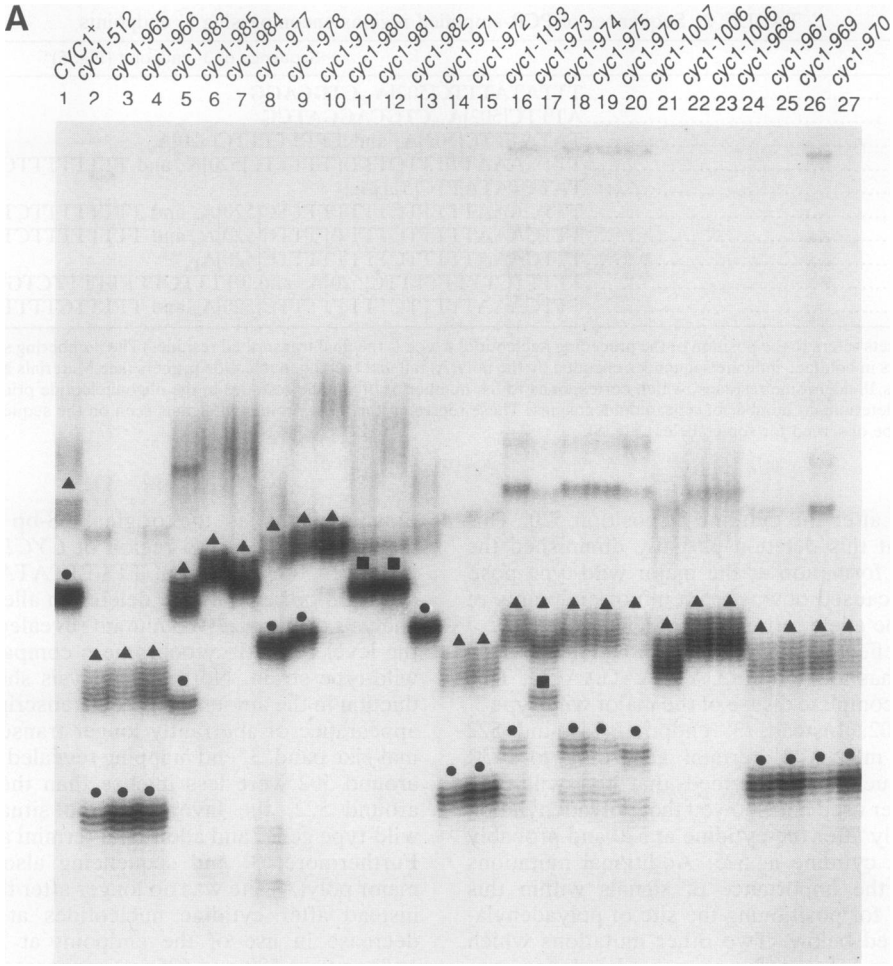


FIG. 4. 3′ end mapping of transcripts from site-directed mutants of *CYC1*$^+$ and *cyc1-512*. Instead of agarose gel electrophoresis, DNA sequencing gels were used to separate the ^{32}P-radiolabeled 3′ end mapping products; also, sequencing ladders were used as size standards. The allele number for each sample is displayed above the lanes and described in Table 1. The following symbols displayed above certain bands correspond to the following approximate endpoints, in which the A of the ATG initiator codon is assigned position 1: ●, 502; ▲, 522; ■, 511; ○, 396; □, 402; and △, 571. Multiple bands are seen around many of the endpoints and may be caused by either an artifact of the PCR amplification or by actual heterogeneity of 3′ ends produced around a certain nucleotide. In either case, the values specified above were arrived at by taking an average of the band sizes. The 502 position determined for the major *CYC1*$^+$ wild-type endpoint matches the previously reported value (1). Furthermore, the mutant endpoints showed internal consistencies, which also corresponded to the results of the 3′ end sequencing (Fig. 5; Table 2).

38-bp region; although this mutant expressed normal levels of iso-1-cytochrome *c* and *CYC1* mRNA, it showed an almost complete lack of the wild-type poly(A) site. Instead, endpoints around 522 became the major poly(A) site, with PCR sequencing showing that polyadenylation occurred after the cytidine at 520. Similar results were observed with the last larger deletion, which lacked the center portion of the 38-bp region GTTATGTTAGTATTAAGAAC, in allele *cyc1-984*. Interestingly, the deleted sequence common to both *cyc1-985* and *cyc1-984* is TTAAGAAC, which is also the sequence deleted in allele *cyc1-979*, all three of which show a change in the major poly(A) site from the wild-type 502 to downstream cytidine residues at 520 and 525. Because none of the mutants showed a change in total 3′ end-forming efficiency, the signals within the sequence TTAAGAAC appear to be involved in determining the position of poly(A) sites. In addition, the polyadenylation sites utilized by all three mutants occurred after the same cytidine nucleotides, further suggesting a commonality in the lesion, redundancy in 3′ end-forming signals, and preference for cytidines at poly(A) sites.

To further investigate the importance of signals within the sequence TTAAGAAC and the cytidine at 502, two more mutations were made in the *CYC1*$^+$ wild-type gene. The mutant *cyc1-1183* was made by substituting the normal sequence TTAAGAAC with the sequence TTCTGCAG, which contains a *Pst*I site. Although the *cyc1-1183* mutant contained the normal amount of iso-1-cytochrome *c*, 3′ end mapping clearly showed a dramatic decrease in the use of the endpoint at 502 and a substantial increase in the endpoints centered around 522 (Fig. 4B). 3′ end sequencing showed that the major poly(A) sites were after the cytidine nucleotides at 520 and 525.

The effects of deleting the cytidine at 502 was investigated with allele *cyc1-1184*. Spectral analysis of the *cyc1-1184* strain revealed no obvious change in the amount of iso-1-cytochrome *c* compared with the wild-type *CYC1*$^+$ strain, but 3′ end mapping showed that poly(A) sites at 502 and 522

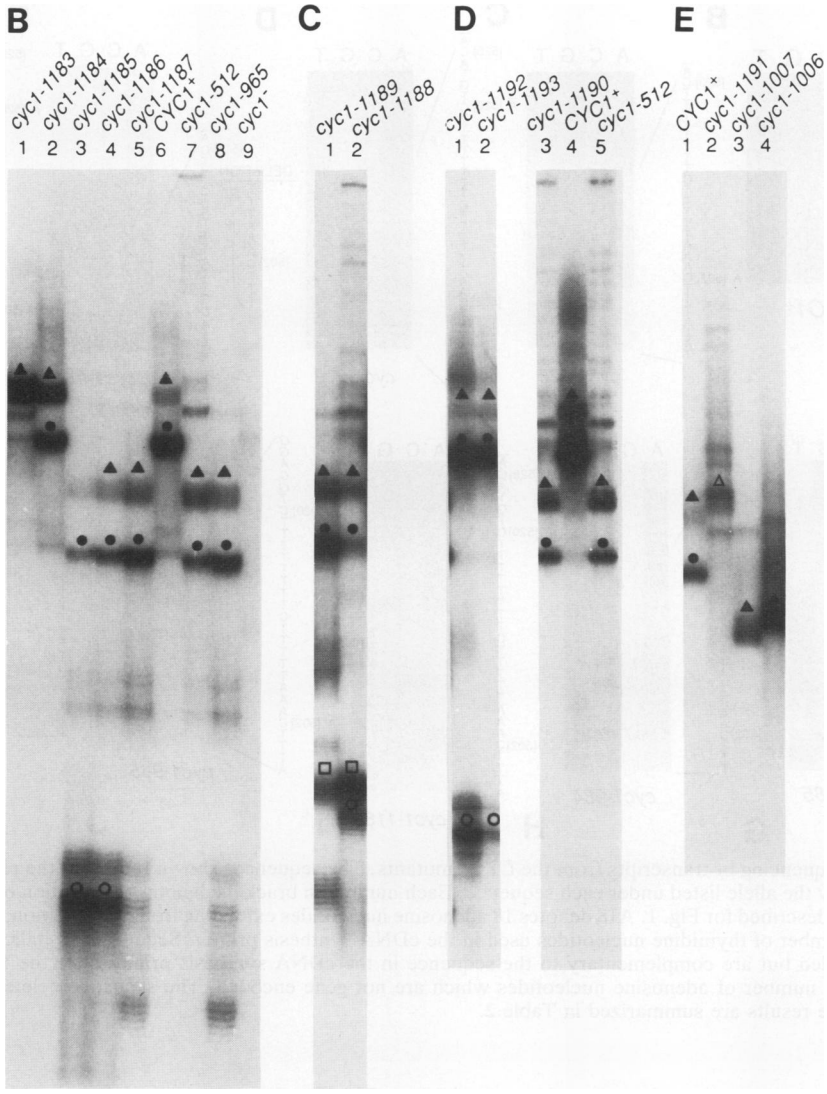


FIG. 4—*Continued.*

were more equally used compared with the wild type. This finding suggests that although the cytidine at 502 is not absolutely necessary for 3′ end formation, it does appear to affect the efficiency of 3′ end formation around 502.

Formation of 3′ end signals in *cyc1-512*. To further characterize certain putative 3′ end-forming signals in *S. cerevisiae*, signals were recreated at different positions in the 3′ region of the *cyc1-512* allele, and their effects on *CYC1* expression were determined by the methods used for the analyses described above. The results are summarized in Tables 1 and 2; the original Northern, 3′ end mapping, and 3′ end sequencing results are presented in Fig. 3, 4, and 5, respectively. All mutations constructed in the *cyc1-512* mutant are represented in Fig. 2B.

Initially, eight different sequences were created by site-directed mutagenesis around position 355, approximately 100 bp 5′ to the 38 bp deletion of the *cyc1-512* mutation, resulting in the destruction of an *Eco*RV site. Of the eight, three sequences, TATATA, TAGATATATGTAA, and TAC ATA in alleles *cyc1-965*, *cyc1-970*, and *cyc1-967*, respectively, restored *CYC1* expression to normal or near-normal levels. These sequences were tested because similar sequences at different locations in the *cyc1-512* mutation were found to restore *CYC1* expression (15). Furthermore, all three sequences appeared to enhance the use of the wild-type poly(A) site around 502. This was confirmed for allele *cyc1-965* by PCR sequencing, which showed the major poly(A) site to be after the cytidine at 502. [In Fig. 5I, the weaker sequence seen extending past 502 alongside the adenosines of the poly(A) tail probably represent less utilized endpoints downstream.] In contrast, the created similar sequences, TTTTTTTATA, CATATA, TATACA, and TAG TTATGTT, in the same location as the above sequences in alleles *cyc1-972*, *cyc1-966*, *cyc1-968*, and *cyc1-971*, respectively, had lesser and varying effects on restoring *CYC1* expression and enhancing polyadenylation at the 502 site, whereas the sequence CACACA in allele *cyc1-969* had no effect at all. These results suggest that sequences such as TATATA and TAGATATATGTAA and related sequences contain signals which enhance the use of the most proximal weak downstream poly(A) sites originally seen in *cyc1-512* and not termini further downstream. These sequences also

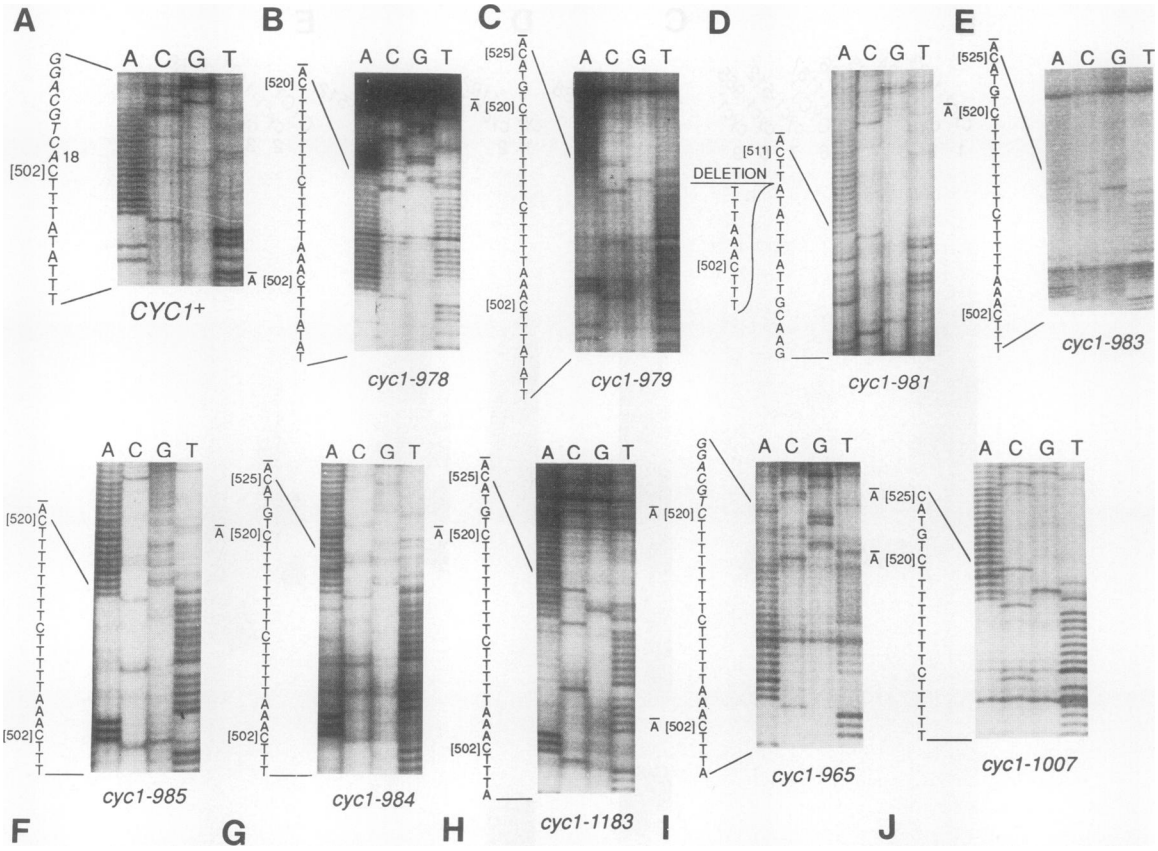


FIG. 5. PCR 3′ end sequencing of transcripts from the *CYC1* mutants. The sequences shown represent the regions near the poly(A) sites of transcripts produced by the allele listed under each sequence. Each number in brackets denotes the position of the nucleotide to the right, numbered by convention described for Fig. 1. A18 denotes 18 adenosine nucleotides extending from that position, which are not gene encoded and correspond to the number of thymidine nucleotides used in the cDNA synthesis primer. Sequences in italics also represent nucleotides that were not gene encoded but are complementary to the sequence in the cDNA synthesis primer after the 18 thymidine nucleotides. $\overline{A}$ denotes an indeterminate number of adenosine nucleotides which are not gene encoded. The sequence deleted in the *cyc1-981* mutant is indicated (Fig. 1B). These results are summarized in Table 2.

show some specificity because they functioned at varying efficiencies.

To test whether the above-mentioned sequences could function at different locations, the putative signals TTTT TATA, TATATA, TAGTTATGTT, and AATAAA (a higher eukaryotic mRNA 3′ end-processing signal) were created 1 to 5 nt upstream from the 502 position of *cyc1-512*, in alleles *cyc1-974*, *cyc1-975*, *cyc1-976*, and *cyc1-973*, respectively (Fig. 2B). Because a *Pst*I site was also created immediately 5′ to each of the four putative signals, a control containing only a *Pst*I site in the same location was made and designated allele *cyc1-1193*. Creating the sequences TTTTTATA, TAGTTATGTT, and AATAAA at this location appeared to neither increase the iso-1-cytochrome *c* level nor change *CYC1* mRNA banding patterns on Northern blots compared with the *cyc1-512* mutant. In contrast, the creation of TATATA 3 nt upstream from the 502 site in allele *cyc1-975* increased iso-1-cytochrome *c* to 40% of the normal level, though this is significantly less than the 100% wild-type level observed when TATATA was positioned 100 nt upstream from 502 in the allele *cyc1-965*. Northern analysis also displayed a slight increase in both the wild-type and aberrant transcripts of *cyc1-975*. The addition of only the *Pst*I site in the *cyc1-1193* mutant neither increased nor decreased iso-1-cytochrome *c* and *CYC1* transcript levels relative to *cyc1-512*. Furthermore, all of these mutants except one produced 3′ endpoint patterns similar to each other and to that of the original *cyc1-512* mutant, which suggests some type of inhibition in their ability to enhance the use of downstream poly(A) sites. The exception was *cyc1-973*, which had the higher eukaryotic mRNA 3′ end-processing signal AATAAA created 5 nt upstream from 502; although *cyc1-512* was not restored by this mutation, it did appear to change the poly(A) site from 502 to a new site around 511. In contrast to these dysfunctional mutants, repositioning the sequences TTTTT TATA and TATATA approximately 14 nt upstream from the 502 position of *cyc1-512* resulting in alleles *cyc1-1006* and *cyc1-1007*, respectively, restored iso-1-cytochrome *c* and *CYC1* mRNA levels to normal or near-normal amounts. Interestingly, neither sequence enhanced the use of the 502 poly(A) site. Instead, both utilized endpoints around 522, with the sequencing of *cyc1-1007* transcripts confirming that the major poly(A) sites were directly after the cytidine nucleotides at 520 and 525. In toto, these results suggest that the spacing between the upstream enhancing sequences, such as TATATA, and the downstream poly(A) sites is a factor that can affect the efficiency of 3′ end formation and poly(A) site recognition. In addition, it appears that other, as yet undefined signals in the *cyc1-512* 3′ region are able to

determine weak poly(A) sites at 502, 520, 525, and other downstream sites.

Combinations of putative 3' end-forming signals. Previously in this study, the sequence TTAAGAAC from the 38-bp region was implicated in determining the position of the *CYC1* wild-type poly(A) site at 502. To test whether this sequence and the related sequence AAGAA could generate mRNA 3' ends in combination with so-called upstream sequences such as TATATA, these two putative mRNA 3' end-forming signals were created together at different locations in the *cyc1-512* and *CYC1*$^+$ loci (Fig. 2B). The first such mutant allele tested, *cyc1-1185*, had the sequence TATATA-13 nt-TTAAGAAC created approximately 80 nt upstream from the 502 position of *cyc1-512*. Originally in *cyc1-965*, the creation of the sequence TATATA at the same location restored *cyc1-512* to normal expression levels and increased the use of the 502 poly(A) site. In comparison, creating the sequence TTAAGAAC 13 nt downstream from the TATATA sequence also restored *CYC1* expression to normal amounts, but instead of the use of the 502 poly(A) site being enhanced, new endpoints around 396 and 402 became the major poly(A) sites. Similar results were observed for allele *cyc1-1186*, which combined the sequences TATATA-15 nt-AAGAA at the same location as in *cyc1-1185*. However, in *cyc1-1186*, the major poly(A) sites around 396 and 402 appeared less intense, while termini around 502 appeared more intense (Fig. 4B), relative to *cyc1-1185*. This finding suggests that the sequence TTAAGAAC is more efficient than just AAGAA in forming poly(A) sites. In contrast, the sequence TATATA-15 nt-AATAA, created in allele *cyc1-1187* at the same location as in the above-mentioned combinations, did not generate the major poly(A) sites at 396 and 402. Instead, endpoints around 502 were more efficiently utilized and became the major poly(A) site as was the case when only the upstream sequence TATATA was created in *cyc1-965*. Thus, the sequence AATAA had no significant effect on determining the position of the poly(A) site compared with the sequences TTAAGAAC and AA GAA. Relatedly, creating only the sequence AAGAA in allele *cyc1-1188* at the same location as above, but without the upstream TATATA, did not restore *cyc1-512* but did appear to generate weak endpoints around 396 and 402, although the *cyc1-512* endpoint pattern was still observable. Apparently, sequences like TTAAGAAC and AAGAA may be involved in determining poly(A) sites, but upstream sequences such as TATATA are necessary for the efficient use of those sites.

Context and spacing effects were investigated with the *cyc1-1189* mutant, which contained the sequence TATATA-25 nt-AAGAA approximately 70 nt upstream from the 502 position of *cyc1-512* (Fig. 2B). This mutation did restore iso-1-cytochrome *c* expression to normal levels, but it was the 502 endpoint that was utilized the most, although new but weak poly(A) sites around 402 and 412 were generated. This apparent dysfunction of the AAGAA element might be explained by the inhibitory effects of the surrounding sequences and/or nonoptimal spacing between the upstream sequence TATATA and AAGAA.

The last two mutations in this series were used to test whether the combination TATATA-spacer-AAGAA could function at different locations in the *cyc1-512* 3' region. For instance, allele *cyc1-1190* contained the sequence TATATA-12 nt-AAGAA within the iso-1-cytochrome *c* coding region (Fig. 2B), and although the substitutions Ile-100→Tyr and Ala-106→Glu were made to create this sequence, neither the wild-type reading frame nor the position of the termination codon was altered. Surprisingly, neither TATATA nor AAGAA appeared to function in this location, because no new poly(A) sites were detected downstream from AAGAA, and the 502 endpoint was not enhanced, leading to no restorations of *cyc1-512* at all. In contrast, when the sequence TATATA-15 nt-AAGAA was created approximately 20 nt downstream from the 502 position of *cyc1-512* in allele *cyc1-1191*, both appeared to function efficiently, as suggested by the production of a strong poly(A) site around 571 and the complete restoration of *cyc1-512*. Apparently, these putative 3' end-forming signals cannot function properly when positioned inside the coding region of the *CYC1* gene. Furthermore, these same signals appear to function in a 5'-to-3' direction because when they are placed downstream of the 502 site, polyadenylation at this site is defunct, though new sites downstream from the created signals are efficiently utilized.

3' end-forming signal combinations in *CYC1*. The last question addressed by our experiments was whether the above-mentioned combination of putative mRNA 3' end-forming signals could function in the 3' region of the wild-type *CYC1*$^+$ gene. For example, the *cyc1-1192* mutant contained the sequence TATATA-13 nt-TTAAGAAC at the same location as in *cyc1-1185* but in the *CYC1*$^+$ wild-type allele (Fig. 2B). Unexpectedly, spectral analysis of the *cyc1-1192* mutant showed an increase in iso-1-cytochrome *c* level to approximately 150% of the wild-type level. Furthermore, 3' end mapping revealed the appearance of new endpoints around 396 and 402, but these were less intense than the termini around 502, suggesting that the wild-type 3' end-forming signals were more efficient. Finally, our last mutant allele, *cyc1-1193*, contained the sequence TATATA-15 nt-AAGAA in approximately the same context as in the previous combination, *cyc1-1192* (Fig. 2B). Spectral analysis of the *cyc1-1193* strain showed a slight diminution from the wild-type level of iso-1-cytochrome *c*. 3' end mapping revealed that the 502 endpoint remained the most intense band, but new termini were observed around 396 and 402, though less intense than similar endpoints in *cyc1-1192* mutant. Again, it appears that the sequence TTAAGAAC is more efficient than AAGAA, but the *CYC1*$^+$ wild-type mRNA 3' end-forming signals are still the most efficient.

DISCUSSION

3' end-forming signals in the *CYC1*$^+$ 38-bp region. In this study, mRNA 3' end-forming signals have been characterized by altering the 3' regions in the *CYC1*$^+$ and *cyc1-512* genes. Initially, deletions were made in the wild-type *CYC1*$^+$ 38-bp region, which is required for normal 3' end formation (19). Although only one of the partial deletions diminished overall *CYC1* mRNA production, some deletions had obvious effects on the position of poly(A) sites. For instance, deletion of TTTTTTTATA from the 38-bp region in *cyc1-978* caused 3' endpoint production around the 520 cytidine to increase and become equal to endpoints around 502 (in wild-type *CYC1*$^+$ mRNA, the 3' endpoints around 502 are utilized much more than those at 520); however, total mRNA 3' end formation in *cyc1-978* was not affected (Fig. 6A). This deletion was designed primarily to remove the putative 3' end-forming signal, TTTTTATA (5), and reposition an upstream TA to maintain the putative signal TAG · · · TAT GTA (15) (Fig. 1B). Although deletions of portions of this putative signal in *cyc1-978* had no obvious effects, deletions of both putative signals simultaneously in *cyc1-983* caused a decrease in levels of iso-1-cytochrome *c*, produced aber-

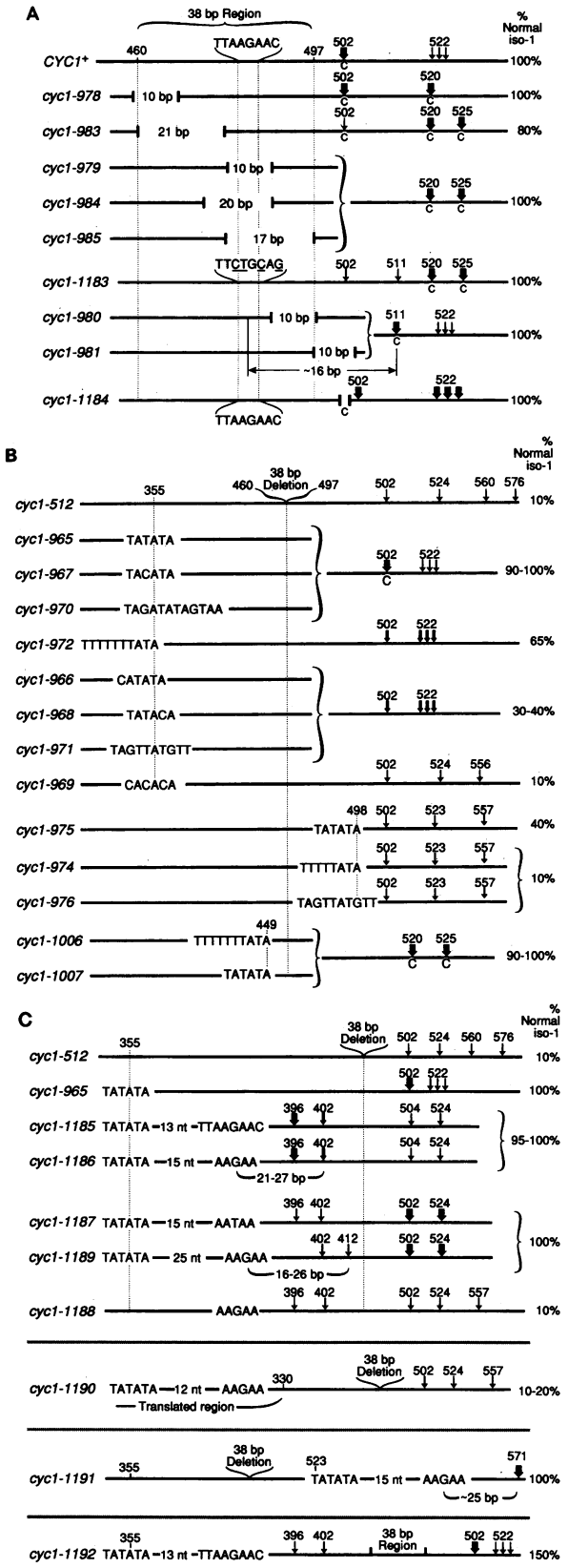


FIG. 6. Schematic representation and summary of the data presented in Table 1.

rantly long transcripts (Fig. 3), and shifted the major *CYC1* poly(A) site from 502 to 520 (Fig. 6A). Apparently, deletions of either putative signal individually in *cyc1-978* or *cyc1-977* caused little if any effect on total *CYC1* mRNA 3′ end production, though deleting TTTTTTTATA did cause a change in the pattern of poly(A) utilization. In contrast, concomitant deletions of both putative signals in *cyc1-983* caused a partial decrease in overall *CYC1* mRNA production and a dramatic change in 3′ end positioning. This finding suggests that both sequences are part of a larger integrated and probably redundant 3′ end-forming signal, which functions by enhancing the use of certain downstream poly(A) sites, e.g., 502, 520, and 525. Surprisingly, deletion of the sequence TTAAGAAC in the three mutants *cyc1-979*, *cyc1-985*, and *cyc1-984* and the substitution of AAGAAC with CTGCAG in *cyc1-1183* repositioned the major 3′ endpoint from 502 to a cytidine at 520, with no change in total 3′ end production (Fig. 6A). Furthermore, two different 10-bp deletions, between the sequence TTAAGAAC and the wild-type poly(A) sites in alleles *cyc1-980* and *cyc1-981*, both showed a shift in the major 3′ endpoint from 502 to a cytidine at position 511 (Fig. 6A). Although one of these 10-bp deletions did remove the 502 wild-type poly(A) site, results from both deletions imply that a certain distance, approximately 16 nt, must be maintained between some signal in the sequence TTAAGAAC and the actual poly(A) site. Similarly, when Fitzgerald and Shenk (3) made deletions between the higher eukaryotic mRNA 3′ end-processing signal AAUAAA and the normal poly(A) site of the simian virus 40 late genes, they observed that a distance of approximately 11 to 19 nt was maintained between AAUAAA and newly created poly(A) sites downstream. Therefore, it is possible that signals within the sequence TTAAGAAC function in a similar manner to AAUAAA, by playing a part in defining mRNA 3′ end cleavage and polyadenylation sites in *S. cerevisiae*. It may also be important that the deletion of sequences upstream from both TTAAGAAC (*cyc1-983* [this report]) and AAUAAA (3) caused additional poly(A) sites downstream from the normal site to be utilized, which may further imply functional similarities between yeast and higher eukaryotic mRNA 3′ end-forming mechanisms. Finally, the repeated appearance of cytidine nucleotides at or near the poly(A) sites (502, 511, 520, and 525) suggests that 3′ ends are formed preferentially at cytidine nucleotides. This may be related to the findings of Heidmann et al. (4), who observed that approximately 94% of poly(A) sites occur at Py(A)$_n$ sequences, i.e., cytidine or thymidine residues, followed by one or more adenosine residues in normal and mutant forms of four yeast genes. Interestingly, the deletion of the cytidine at 502 in the *cyc1-1184* allele caused the 3′ end-forming pattern to appear more equally distributed between termini around 502 and 522 compared with the wild-type pattern (Fig. 6A). This finding further suggests that although cytidine residues may be preferred at poly(A) sites, they are not absolutely essential.

In summary, the results obtained from the mutations made in and around the *CYC1*$^+$ 38-bp region suggest the following hypotheses. (i) Sequences in and around the 38-bp region (e.g., TTTTTATA, TAG · · · TATGTA, TTAAGAAC, and others yet undefined) function in a cooperative and partially redundant manner to generate mRNA 3′ ends. (ii) Signals within such sequences as TTTTTATA and TAG · · · TA TGTA affect the efficiency of 3′ end formation at downstream preferred poly(A) sites. (iii) Signals within the sequence TTAAGAAC aid in defining preferred sites of polyadenylation approximately 16 nt downstream. (iv) mRNA 3′

end formation and polyadenylation preferentially occur after cytidine nucleotides.

Some of these hypotheses were tested by the creation of the above-mentioned putative 3′ end-forming signals and related sequences, individually and in combinations in the *cyc1-512* mutant, and will be discussed in further detail below.

Different putative 3′ end-forming signals restored the *cyc1-512* deficiency to different efficiencies. The abilities of certain putative 3′ end-forming signals to enhance the use of weak downstream poly(A) sites were first tested by creating those putative signals approximately 100 bp upstream from the 38-bp deletion in *cyc1-512*. Some of these signals were previously suggested by Russo et al. (15) and others (5). The sequences (and *cyc1* mutants) which restored *cyc1-512* to wild-type or near-wild-type *CYC1*$^+$ mRNA levels were as follows: TATATA (*cyc1-965*), 100%; TAGATATATGTAA (*cyc1-970*), 100%; and TACATA (*cyc1-967*), 90%. In contrast, other, similar sequences were not as efficient or were ineffective. These included the following: TTTTTATA (*cyc1-972*), 65%; TAGTTATGTT (*cyc1-971*), 30%; CAT ATA (*cyc1-966*), 40%; TATACA (*cyc1-968*), 30%; and CAC ACA (*cyc1-964*), 10%, which had no effect. Although the above-mentioned sequences were created approximately 100 bp upstream from the 502 poly(A) site of *cyc1-512*, those that increased mRNA levels the most enhanced the use of the 502 endpoint (Fig. 6B). It could be argued that instead of enhancing 3′ endpoint formation, these sequences increased the stability of the *CYC1* mRNAs. However, if this is the case, all of the weak poly(A) sites utilized by *cyc1-512* should be equally enhanced, not just endpoints around 502. Alternatively, these sequences may represent the so-called upstream signals that were suggested by Russo et al. (15) and which appeared to function by enhancing the use of the nearest downstream preferred poly(A) sites, one such preferred site being around 502. The notion of preferred poly(A) sites is proposed because even in *cyc1-512*, in which the 38-bp deletion caused a 90% reduction in polyadenylation at the 502 site, the 502 site is still utilized, albeit at a much lower efficiency. Furthermore, it appears that these upstream signals can function from a distance, do not produce new major poly(A) sites, and show sequence specificity. This specificity is demonstrated by the three sequences TACATA, CATATA, and TATACA, which all have the same A+T composition but function at different efficiencies. This is important to note, because Osborn and Guarente (11) suggested that mRNA 3′ end formation in *S. cerevisiae* was controlled by the A+T richness of a sequence, not by specific signals. Sequence specificity is further supported by the fact that although alternating pyrimidine and purine sequences (e.g., TATATA, TACATA, TATACA, and CAT ATA) may appear to be sufficient for 3′ end formation, all restored *cyc1-512* to some degree; the sequence CACACA had no observable 3′ end-forming ability.

In contrast to the results presented above, when the sequences TTTTTATA (*cyc1-974*), TATATA (*cyc1-975*), and TAGTTATGTT (*cyc1-976*) were created 1 to 3 bp upstream from 502 in *cyc1-512*, only TATATA produced an observable increase in *CYC1* expression, though significantly less than when created 100 bp upstream from 502, as in the *cyc1-965* mutant (Fig. 6B). Interestingly, the sequences TTTTTATA and TATATA, repositioned 14 bp upstream from the 502 endpoint in alleles *cyc1-1006* and *cyc1-1007*, respectively, increased *CYC1* mRNA levels to normal levels but did so by enhancing poly(A) sites after cytidine residues at 520 and 525, not at 502 (Fig. 6B). These findings suggest that when the so-called upstream signals TTTTTATA and TATATA are only 1 to 3 nt 5′ from the 502 poly(A) site, some type of spatial inhibition occurs between the upstream signal and whatever signal is involved in defining the preferred poly(A) site at 502. In contrast, when the spacing between the upstream signal and 502 was increased to about 14 nt, the upstream signals TTTTTATA and TATATA were no longer inhibited but were still too close to 502 to recognize it as a poly(A) site, resulting in the enhancement of preferred poly(A) sites further downstream. These so-called signals may represent protein and/or small nuclear ribonucleoprotein recognition sites that need a minimum separation to function properly. Whatever the mechanisms, it appears that the functional efficiency of the putative upstream sequences is affected by its position in the 3′ region of *cyc1-512*; more evidence for this context effect is discussed below. Lastly, when the higher eukaryotic mRNA 3′ end-processing signal, AATAAA, was placed 5 nt upstream from the 502 endpoint (*cyc1-973*), it appeared to shift the 3′ endpoint from 502 to 511 and 520 but did not increase 3′ end formation at these sites. While it is now generally agreed that the higher eukaryotic signal AAUAAA does not function or at least does not function efficiently in yeast cells (2, 6, 8), our results suggest that it may be able to define a preferred poly(A) site, albeit at a low efficiency. If this is true, then upstream signals might be able to enhance new poly(A) sites defined by AAUAA, a suggestion worth testing.

3′ end-forming signals function in concert. The results presented above suggest the existence of three classes of sequence elements which function in concert to form mRNA 3′ ends in *S. cerevisiae*. The first class are the so-called upstream elements, including such sequences as TATATA, TAG · · · TATGTA, TACATA, and TTTTTATA, which enhance the 3′ end-forming function of downstream elements. These so-called downstream elements represent the second class of signals, which include the sequences TTAAGAAC and AAGAA, and appear to delineate the approximate site of mRNA 3′ end formation. The third class of elements are 3′ to the downstream elements and represent preferred sites of polyadenylation, which often occurred after cytidine nucleotides in *CYC1* wild-type and mutant mRNAs. Evidence to further support our model comes from the combination of putative signals constructed in the *cyc1-512* mutant. The first such combination placed the so-called upstream signal TATATA 13 nt 5′ to the so-called downstream signal TTAAGAAG, in allele *cyc1-1185*. When only the TATATA signal was created in the same place, in allele *cyc1-965*, it enhanced the 502 endpoint approximately 100 bp downstream (Fig. 6C). In contrast, the most abundant termini in the *cyc1-1185* mutant were around a group of cytidine nucleotides at position 396. Thus, the addition of the sequence TTAAGAAC, downstream from TATATA, enhanced or defined new endpoints, 21 to 27 nt downstream from its position (Fig. 6C). When TTAAGAAC was shortened to AAGAA, but still in combination with an upstream TATATA in allele *cyc1-1186*, the same new endpoints were generated, though less efficiently (Fig. 6C). As might be predicted by our model, the sequence AAGAA without an upstream TATATA, as in allele *cyc1-1188*, was unable to enhance 3′ endpoint production in *cyc1-512* but was able to generate the same endpoints seen in *cyc1-1185* and *cyc1-1186* around 396 and 402, though at a much lower efficiency (Fig. 6C). Furthermore, the specificity of these signals is illustrated by allele *cyc1-1187*, in which the sequence AATAA replaced AAGAA downstream from TATATA. In this case, endpoints around 502, not poly(A) sites around 396 and 402,

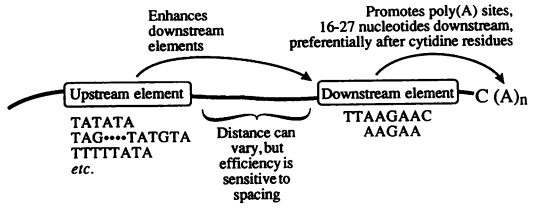


FIG. 7. Schematic representation of the major conclusions concerning the sequences and properties of the upstream and downstream elements.

were enhanced the most (Fig. 6C); these results are similar to those with *cyc1-965*, in which only the upstream signal TATATA was created at the same location.

Although the putative signal combination of TATATA-15 nt-AAGAA functioned in the above-mentioned situation, it was less efficient in other contexts. For instances, changing the distance between TATATA and AAGAA from 15 nt, as in *cyc1-1186*, to 25 nt, in *cyc1-1189*, decreased the use of the new termini generated proximally downstream from the AAGAA element (Fig. 6C). However, a similar spacing appeared to be maintained between the AAGAA element and newly generated 3′ ends, as summarized below:

cyc1-1186, TATATA-15 nt-AAGAA-21 to 27 nt-3′ ends
cyc1-1189, TATATA-25 nt-AAGAA-16 to 26 nt-3′ ends

This range of distance between the so-called downstream signals and the actual poly(A) sites, approximately 16 to 27 nt, is similar to the distance maintained by the two deletions made between the wild-type sequence TTAAGAAC and the new endpoint at 511 in *cyc1-980* and *cyc1-981* (Fig. 6A). Furthermore, this distance range is within the range, 5 to 30 nt, observed between the mRNA 3′ end-processing signal AAUAAA and poly(A) sites in higher eukaryotes (13). Although mRNA 3′ end processing in yeasts and higher eukaryotes is not identical, it appears that similarities in mechanisms exist.

Another context effect was observed in *cyc1-1190*, in which the combined sequence TATATA-12 nt-AAGAA was positioned within the coding region of the *cyc1-512* gene. At this location, neither the upstream nor downstream signals appeared to function because not only were no new poly(A) sites seen, but also the 502 endpoint was not enhanced (Fig. 6C). [Although AATAA was not functional in *cyc1-1187*, the upstream TATATA enhanced the 502 poly(A) site.] This finding suggests that other sequences, either specific or general, in the 3′ untranslated region may be necessary for the function of these putative signals. The dysfunction of these sequences in the coding region context may not be surprising because of their possible deleterious effects on normal gene expression when positioned within or too close to a coding region; however, the mechanism for their inactivation is unclear.

In contrast, 3′ end formation occurred efficiently in the *cyc1-1191* mutant that contained the TATATA-15 nt-AAGAA sequence approximately 20 nt downstream from the 502 site of the *cyc1-512* mutant. Generating this combination 3′ to the 502 position restored iso-1-cytochrome *c* to the normal level and produced a prominent endpoint around 571 (Fig. 6C). Not only does this result show that these signals can function at other locations, but it also suggests that these 3′ end forming signals function in a 5′ to 3′ direction, because placing TATATA-15 nt-AAGAA downstream of 502 did not enhance the upstream 502 endpoint.

Lastly, the sequence specificity and greater efficiency of the wild-type *CYC1* 3′ end-forming signals were demonstrated in *cyc1-1192*, in which the sequence TATATA-13 nt-TTAAGAAC was created approximately 100 bp upstream from the *CYC1*$^+$ 38-bp region. In this case, the wild-type poly(A) site at 502 remained the major endpoint, though new much weaker poly(A) sites appeared to be generated by the putative signals at positions 396 and 402 (Fig. 6C). This may suggest that whatever factors (protein and/or RNA) recognize these signals, the wild-type sequences in the 38-bp region have a greater affinity than the putative signals that we have created upstream.

Proposed model for mRNA 3′ end formation in *S. cerevisiae*. Figure 7 summarizes schematically the sequence elements that we believe work in concert to generate mRNA 3′ endpoints in *S. cerevisiae* and how they might interact. For example, the so-called downstream elements, represented by the related sequences TTAAGAAC and AAGAA, appear to define the positions of poly(A) sites within a range 16 to 27 nt downstream from their centers. In addition, cytidine residues appear to be preferred at the actual poly(A) site, although they are not absolutely necessary. Although the downstream signals may define preferred poly(A) sites, the so-called upstream elements, represented by the sequences TATATA, TAG · · · TATGTA, and TTTTTATA and related sequences, are necessary to enhance the utilization of these preferred poly(A) sites. We wish to emphasize that it is not our intention to suggest that these are the only signals in *S. cerevisiae* involved in mRNA 3′ end formation; almost certainly they are not. Furthermore, we do not wish to imply that these elements are completely discrete signals, because it appears that they must function together for efficient poly(A) site production. A similar proposal has been made for normal mRNA 3′ end formation in eukaryotes, which appears to be dependent on both a functional AAUAAA and a transcription terminator (12). How the upstream and downstream elements perform their functions is a question outside the scope of our experiments, although it is conceivable that a signal within the sequence TTAAGAAC is acting as a recognition site for a factor which produces mRNA 3′ endpoints by cleavage and subsequent polyadenylation. The upstream element may enhance this cleavage by acting as a binding site for this putative cleavage factor. Furthermore, this combined effort may result in the termination of transcription soon after the cleavage site. Although our speculations remain to be elucidated, it is clear that the production of mRNA 3′ endpoints in *S. cerevisiae* is a complex process utilizing specific sequences that work in concert and may have functional similarities to counterparts in higher eukaryotes.

ACKNOWLEDGMENT

This investigation was supported by Public Health Service research grant R01 GM12702 from the National Institutes of Health.

REFERENCES

1. **Boss, J. M., S. Gillam, R. S. Zitomer, and M. Smith.** 1981. Sequence of the yeast iso-1-cytochrome *c* mRNA. J. Biol. Chem. **256:**12958–12961.
2. **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.
2a. **Cardillo, T.** Unpublished data.
2b. **Ceseareni, G.** Unpublished data.

3. **Fitzgerald, M., and T. Shenk.** 1981. The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell **24**:251–260.
4. **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.
5. **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.
6. **Hyman, L. A., 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.
7. **Irniger, S., C. M. Egli, and G. Braus.** 1991. Different classes of polyadenylation sites in the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. **11**:3060–3069.
8. **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 to plants. Mol. Cell. Biol. **12**:2322–2330.
9. **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.
10. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–559.
11. **Osborn, B. I., and L. Guarente.** 1989. Mutational analysis of a yeast transcriptional terminator. Proc. Natl. Acad. Sci. USA **86**:4097–4101.
12. **Proudfoot, N. J.** 1991. Poly(A) signals. Cell **64**:671–671.
13. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′ non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
14. **Rose, M. D., P. Novick, J. H. Thomas, D. Botstein, and G. R. Fink.** 1987. A *Saccharomyces cerevisiae* genomic bank based on a centromere-containing shuttle vector. Gene **60**:237–243.
15. **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.
16. **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.
17. **Sherman, F., and P. P. Slonimski.** 1964. Respiration-deficient mutants of yeast. II. Biochemistry. Biochim. Biophys. Acta **90**:1–15.
18. **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.
19. **Zaret, K. S., and F. Sherman.** 1982. DNA sequence required for efficient transcription termination in yeast. Cell **28**:563–573.