Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 4211–4214, May 1995
Biochemistry

Redundant 3′ end-forming signals for the yeast *CYC1* mRNA

ZIJIAN GUO*, PATRICK RUSSO*†, DING-FANG YUN*, J. SCOTT BUTLER‡, AND FRED SHERMAN*¶

Departments of *Biochemistry and ‡Microbiology and Immunology, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642; and †Institute of Biotechnology, University of Helsinki, Valimotie 7 SF-00380, Helsinki, Finland

Contributed by Fred Sherman, January 3, 1995

ABSTRACT **The *cyc1-512* mutation is a 38-bp deletion in the 3′ untranslated region of the *CYC1* gene, which encodes iso-1-cytochrome *c* in *Saccharomyces cerevisiae*. This deletion caused a 90% reduction in the levels of the *CYC1* mRNA and protein because of the absence of the normal 3′ end-forming signal. Although the 3′ end-forming signal was not defined by previous analyses, we report that concomitant alteration by base-pair substitution of three 3′ end-forming signals within and adjacent to the 38-bp region produced the same phenotype as the *cyc1-512* mutation. Furthermore, these signals appear to be related to the previously identified 3′ end-forming signal TATATA. A computer analysis revealed that TATATA and related sequences were present in the majority of 3′ untranslated regions of yeast genes. Although TATATA may be the strongest and most frequently used signal in yeast genes, the *CYC1*$^+$ gene concomitantly employed the weaker signals TTTATA, TATGTT, and TATTTA, resulting in a strong signal.**

The proper formation of mRNA 3′ ends and transcription termination are essential for the optimal expression of genes in eukaryotes. Point mutations in the mRNA 3′ end-forming signal, or a deletion of the signal, can lead to drastic decreases in gene expression (1–4). Such a case was found in the yeast *cyc1-512* mutant (5). In addition to displaying a 90% decrease in *CYC1*$^+$ mRNA and iso-1-cytochrome *c*, the *cyc1-512* mutation also caused the formation of aberrantly long transcripts with many discrete 3′ termini ranging from the wild-type poly(A) site to endpoints >2000 nt downstream (5–9). Butler and Platt (10) found that the same 38-bp region deleted in *cyc1-512* 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 (11) to be able to cause the termination of transcription *in vivo*. Although this *cyc1-512* mutation and its properties were uncovered >10 years ago (5), and although this *cyc1-512* mutation was instrumental for defining 3′ end-forming signals by examining sequences that restored the levels of the *CYC1* mRNAs (6), the exact 3′ end-forming signals of the wild-type *CYC1* gene were not previously determined. Because of the inability to inactivate the 3′ end-forming signal by nested deletions and multiple base-pair substitution within the 38-bp region, workers suggested that the 3′ end-forming signal was only an A+T-rich region (7) or a symmetrical RNA structure (8). Russo *et al.* (9) suggested that the *CYC1* 3′ end-forming signals may have functional redundancy, whereas Wu (12) suggested three TA pairs may be important.

Signals required for 3′ end formation were uncovered by examining intragenic revertants and oligonucleotide-directed alterations that restored the *cyc1-512* defect (6, 9). These results suggested that 3′ end formation in yeast involves signals having the following three distinct but integrated elements, which act in concert: (*i*) the upstream elements, 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 polyadenylylation, which often occurs after cytidine residues that are 3′ to the so-called downstream element. This paper presents evidence that redundant upstream signals, which are located within and adjacent to the 38-bp region, are responsible for the efficiency of *CYC1* mRNA 3′ end formation. Concomitant alteration of the three upstream signals abolished *CYC1* mRNA 3′ end formation *in vivo* and *in vitro* and caused at least a 90% reduction in the levels of *CYC1* mRNA and protein, a phenotype similar to the *cyc1-512* mutation.

MATERIALS AND METHODS

Genetic Nomenclature and Yeast 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*, *cyc1-1316*, etc., denote altered *CYC1* alleles, irrespective of whether or not the allele caused a deficiency.

Saccharomyces cerevisiae strain B-7467 (*MAT***a** *cyc1-Δ1 CYC7*$^+$ *ura3-52*) was used as the recipient of the YCp50 plasmid derivatives containing wild-type and altered *CYC1* gene. Yeast whole-cell extracts were prepared from yeast strain EJ101 (*MATα trp1 pro1-126 prb1-112 pep4-3 prc1-126*) (10).

Determination of Cytochrome *c* Content. Total amounts of cytochrome *c* were determined by spectroscopic examination of intact cells at −196°C (13) and by comparing the intensities of the c_α bands at 547 nm to the c_α bands of strains having known amounts of cytochrome *c*. A more accurate determination of cytochrome *c* content in intact cells was made by low temperature (−196°C) spectra recording with a modified Cary 14 spectrophotometer (14).

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.* (15). PCR mapping of *CYC1* mRNA 3′ termini were performed exactly as outlined by Russo *et al.* (9). Northern blot analysis of yeast total RNA was performed as described by Li and Sherman (16) with minor modifications. ^{32}P-end-labeled oligonucleotides OL90.151 and OL93.210 were used as the *URA3* and *CYC1* probes, respectively. Autoradiography was performed at −70°C with an intensifying screen for 5 hr for *CYC1* mRNA and 50 hr for *URA3* mRNA. An LKB 2222-010 UltraScan XL laser densitometer was used to determine the relative densities of the *CYC1* and *URA3* mRNA bands.

Plasmid Constructions. Plasmid pAB458, containing the wild-type *CYC1*$^+$ gene in a *Bam*HI/*Hin*dIII fragment, was used for site-directed mutagenesis in the 3′ untranslated region of the *CYC1*$^+$ gene. Plasmids pAB640 and pAB650 were constructed by inserting a 4.7-kb *Bam*HI/*Eco*RI fragment, containing the entire *CYC1*$^+$ or *cyc1-512* gene, respectively, into the yeast shuttle vector YCp50 (17). Site-directed muta-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

¶To whom reprint requests should be addressed.

tions, made in the plasmid pAB458, were introduced into plasmid pAB640 as *Kpn* I/*Hin*dIII fragments.

***In Vitro* Processing.** A T7 promoter sequence was fused to the *CYC1* 3′ untranslated region by PCR amplification of the DNA sequences (nt 343–599) of *CYC1*$^+$ or mutant alleles from the plasmids pAB640, pAB650, pAB1150, etc., with two primers, OL94.100 (ATTAAAGCCTTCGAGCGTCCC) and OL94.101 (GGATCCTAATACGACTCACTATAGGGAGGTCCTTTGTCGATATCATG), which contained the 29-nt T7 promoter sequence. The PCR product was used to synthesize capped RNA transcripts (263 nt) using an *in vitro* transcription kit (Ambion, Austin, TX) according to the manufacturer's recommendations. Whole cell extract preparations and *in vitro* processing reactions were conducted as described by Butler and Platt (10). To suppress polyadenylylation, the reaction mixtures contained 2 mM CTP instead of ATP as the energy source.

Computer Analysis. The University of Wisconsin Genetics Computer Group (GCG) software (version 7.0), including FINDPATTERNS and FETCH, was used to search for the hexanucleotide sequences TATATA, TATTTA, TATGTA, and TTTATA in the yeast gene sequences from the GenBank data base.

RESULTS AND DISCUSSION

***CYC1* mRNA 3′ End-Forming Signals.** Possible upstream elements in the 38-bp region were tested by sequentially altering the previously identified 3′ end-forming signals, including TAG···TATGTT and TTTTTATA, as well as TATTTA, which was recently demonstrated to enhance 3′ end formation of *cyc1-512* mRNA (Z.G. and F.S., unpublished data). The results, summarized in Fig. 1, include the amounts of both iso-1-cytochrome *c* and the 630-nt *CYC1* transcript (Fig. 2), as well as the efficiency of *in vitro* cleavage of pre-mRNA of the normal or mutant *CYC1* alleles (Fig. 3). In and adjacent to the 38-bp region, there are two copies of the sequence TATTTA; changing either one or both of them to TAGATC by site-directed mutagenesis had no or little effect on either the level of *CYC1* mRNA or iso-1-cytochrome *c* and on *in vitro* cleavage efficiency (*cyc1-1316, cyc1-1317*, and *cyc1-1319*). On the other hand, simultaneous mutation of the TTTTTATA and TAG···TATGTT sequences caused a 50% reduction in iso-1-cytochrome *c* level, a 95% reduction in the efficiency of *in vitro* cleavage, and an 87% reduction in the level of the 630-nt transcript (*cyc1-1318*). Although introducing four C residues changed the A+T content in the 38-bp region, the resulting sequence is still A+T-rich. Furthermore, we do not believe that the slight decrease in A+T content is the major cause of the defect in 3′ end formation, because our previous study showed that the creation of a *Pst* I restriction site, which similarly decreased A+T content in the 38-bp region, had no effect on the efficiency of 3′ end formation (9).

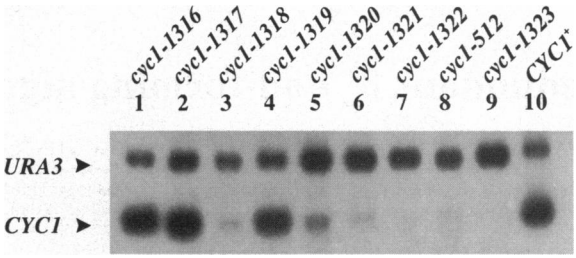

FIG. 2. Northern blot analysis of the 630-nt *CYC1* mRNA. Total yeast RNA, transferred to a nitrocellulose filter, was first hybridized to a *URA3* probe. The full-length *URA3* mRNA from the wild-type *URA3* gene on the plasmids is indicated. Some truncated *URA3* transcripts from the chromosomal *ura3-52* allele were also observed but are not shown. After autoradiographs were obtained, the *URA3* probe was removed by washing and a *CYC1* probe was hybridized to the same filter. The 630-nt *CYC1* mRNA was quantitated by densitometry using the full-length *URA3* mRNA as an internal standard.

Apparently, the two sequences destroyed in *cyc1-1318* were not the only 3′ end-forming signals in the 38-bp region. Compared with the *cyc1-1318* allele, the defect of *cyc1-512* is still more severe. To examine whether the residual 3′ end-forming capacity in *cyc1-1318* is due to the TATTTA sequences, either one or both of these sequences were additionally altered. The result revealed that a mutation of the upstream TATTTA sequence did not cause a further reduction of 3′ end formation efficiency. In contrast, alteration of the downstream sequence or both sequences resulted in deficiencies in iso-1-cytochrome *c*, the 630-nt transcript, and *in vitro* cleavage efficiency, similar to those in the *cyc1-512* strain. This result suggested that three distinct sequences (*i*) TTTTTATA, (*ii*) TAG···TATGTT, and (*iii*) TATTTA in the 38-bp region function redundantly in *CYC1* mRNA 3′ end formation. Although the downstream TATTTA sequence is 3′ to the AAGAA sequence, it is thought to be an upstream element based on the following observations. First, alteration of it had an effect on the levels of iso-1-cytochrome *c*, *CYC1* mRNA, and *in vitro* cleavage efficiency. Second, mutations in this element dramatically decreased the usage of the poly(A) site at nucleotide 522 but had little effect on the poly(A) site at nucleotide 502 (Fig. 4), indicating that it functions not by

	1	2	Downstream	3	502 ↓	Percent normal: Iso-1	Percent normal: mRNA	Percent normal: Cleavage	*CYC1* mRNA major 3′ endpoints
CYC1$^+$	TATTTA	TTTTTTATAGTTATGTTAGTATT	AAGAA	CGTTATTTA	TATTTC	100	100	100	502, 522
cyc1-1316	TATTTA	TTTTTTATAGTTATGTTAGTATT	AAGAA	CGTTAGATC	TATTTC	100	98	82	502
cyc1-1317	TAGATC	TTTTTTATAGTTATGTTAGTATT	AAGAA	CGTTATTTA	TATTTC	100	90	100	502, 522
cyc1-1318	TATTTA	TTTTTTCTCGTCATGTCAGTATT	AAGAA	CGTTATTTA	TATTTC	50	13	7	502, 522
cyc1-1319	TAGATC	TTTTTTATAGTTATGTTAGTATT	AAGAA	CGTTAGATC	TATTTC	100	93	86	502
cyc1-1320	TAGATC	TTTTTTCTCGTCATGTCAGTATT	AAGAA	CGTTATTTA	TATTTC	45	7	8	502, 522
cyc1-1321	TATTTA	TTTTTTCTCGTCATGTCAGTATT	AAGAA	CGTTAGATC	TATTTC	10	3	<5	502
cyc1-1322	TAGATC	TTTTTTCTCGTCATGTCAGTATT	AAGAA	CGTTAGATC	TATTTC	5	1	<5	502
cyc1-512	TATTTA T	—			ATTTC	10	3	0	502, 522, 560
cyc1-1323	TAGATC T	—			ATTTC	5	1	0	502

FIG. 1. The 3′ end-forming signals of the *CYC1*$^+$ gene and the mutations created in the 3′ untranslated region of *CYC1*$^+$ and *cyc1-512*. Allele numbers are shown on the left of each sequence. The 3′ end-forming signals, including the upstream elements and a downstream element, are boxed, and the upstream elements are numbered 1–3. Element 2 encompassed two overlapping signals. The site-directed mutations within each of the signals are presented in solid boxes. The position of the 38-bp deletion in *cyc1-512* and *cyc1-1323* is denoted. The major 3′ endpoint of *CYC1* mRNA at nucleotide 502 is marked by an arrow. The amount of the iso-1-cytochrome *c* protein relative to the wild-type level was determined by spectral analysis. The percentage normal mRNA refers to only the 630-nt *CYC1* transcript (see Fig. 2). The major 3′ endpoints were estimated from the results shown in Fig. 4, where the A residue of *CYC1* ATG initiator codon is assigned position 1.

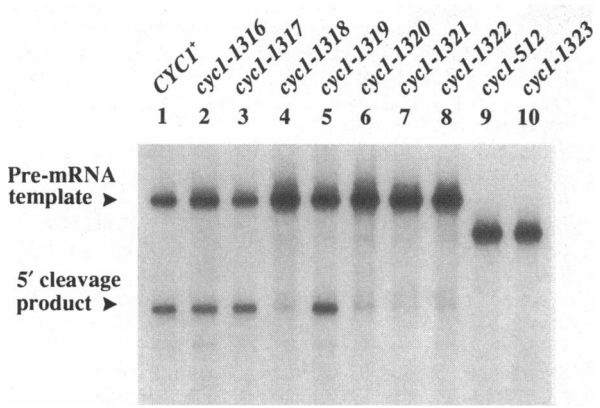

FIG. 3. *In vitro* cleavage assay. The capped pre-mRNA templates (*CYC1* 3′ untranslated region) were synthesized from plasmids pAB640, pAB650, and pAB1150–pAB1157, each of which contained one of the *CYC1* alleles listed in Fig. 1. Note that the pre-mRNA transcript from *cyc1-512* and *cyc1-1323* contained a 38-nt deletion. One major 5′ cleavage product was observed and is indicated, but no 3′ fragment accumulated after the cleavage reaction.

enhancing the efficiency of the element AAGAA, which positions the 502 poly(A) site (9). It is possible that this TATTTA sequence, along with an unidentified element located downstream of it, is mainly involved in the *CYC1* mRNA 3′ end formation at nucleotide 522, while the upstream signals TTTTTATA and TAG···TATGTT, along with the downstream AAGAA element, control the usage of the 502 poly(A) site. In the *cyc1-512* allele, there is still a TATTTA sequence, mutation of which caused a slight but observable further reduction in iso-1-cytochrome *c* and *CYC1* mRNA, suggesting it functions inefficiently in the *cyc1-512* background and may be cryptic in the wild-type condition.

It is interesting to note that there is a difference between the *cyc1-983* allele (9), a deletion mutation of the TTTTTATA and TAG···TATGTT sequences, and the *cyc1-1318* allele. 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. Compared with *cyc1-1318*, the defect of *cyc1-983* seems to be alleviated. We interpret this observation to indicate that a proper space between the 3′ end-forming signals and the poly(A) site is required for optimal function. Possibly, the seemingly nonfunctional upstream TATTTA sequence could be activated in *cyc1-983* because it was brought closer to the poly(A) site by the deletion.

The *cyc1-1318* allele produces ≈50% of the normal amount of iso-1-cytochrome *c*. In contrast, the 630-nt transcript of this allele decreased to 13% of the normal level. This obvious discrepancy between levels of mRNA and protein was also observed in some of the other mutant alleles (Fig. 1) and probably resulted from the inefficiency of mRNA 3′ end formation. Some abnormally long transcripts of the mutant alleles could be detected after overexposure of Northern blots and PCR 3′ end mapping gels. These abnormally long transcripts are probably translated and contribute to the total level of iso-1-cytochrome *c*.

A Computer Analysis Revealed That TATATA and Related Sequences Were Present in the Majority of 3′ Untranslated Regions of Yeast Genes. Previous studies revealed several 3′ end-forming signals in the *CYC1*$^+$ gene and *cyc1-512* revertants (6, 9). Among them, TATATA seems to have the highest efficiency for forming 3′ ends. TACATA and TAG···TATGTA acted moderately, and TTTTTATA, TAG···TATGTT, and TATTTA acted relatively inefficiently. It should be noted that if TAG of TAG···TATGTA(T) and the first two T

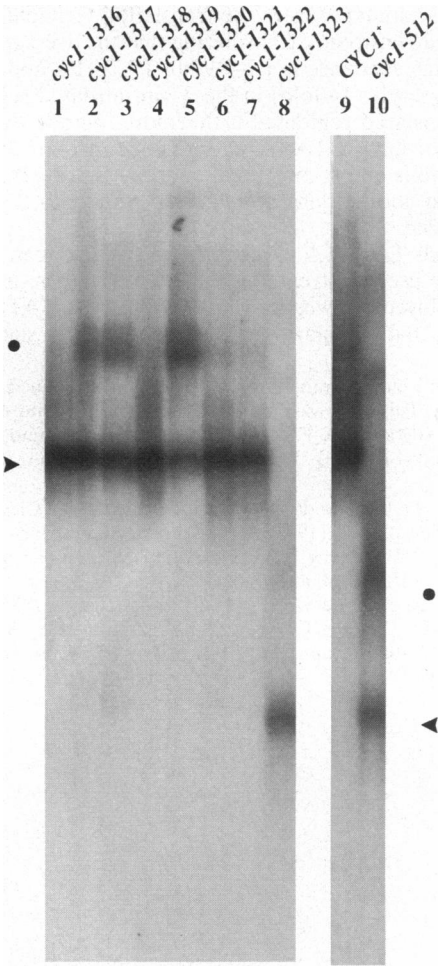

FIG. 4. PCR mapping of the *CYC1* mRNA 3′ ends. The ^{32}P-radiolabeled 3′ end mapping products were electrophoresed through a 6% polyacrylamide/urea sequencing gel along with a sequencing ladder (not shown), which was the product of a sequencing reaction using the control DNA M13mp18 and −40 primer included in the Sequenase version 2.0 kit (United States Biochemical). The allele number for each sample is shown above the lanes. The 502 and 522 positions are marked by an arrowhead and a solid circle, respectively.

residues of TTTTTATA are omitted, all the moderate and weak signals have only a 1- or 2-nt deviation from the strongest signal, TATATA. This relationship was also observed by Irniger and Braus (18, 19) in a saturation mutagenesis study of a 3′ end-forming signal of cauliflower mosaic virus, which functions in yeast. In their system, the sequences TATATA, TTTATA, TATTTA, and TATGTT were, respectively, 132%, 42%, 63%, and 26% efficient in 3′ end formation compared to the sequences TATGTA and TACATA, both of which were ≈100% efficient. For the sake of completion, it should be mentioned that the octamer AAAAAAAA can act as a highly efficient 3′ end-forming signal in conjunction with the other sequence(s) (20).

If the TATATA sequence is considered to be the strongest 3′ end-forming signal in yeast genes, this sequence should appear in 3′ untranslated regions more frequently than translated regions, either to form or to avoid forming 3′ ends, respectively. To test this, the GenBank data base was searched by using the University of Wisconsin GCG programs (version 7.0), including FINDPATTERNS and FETCH; the results revealed that among 1017 yeast nuclear genes whose translated regions or putative translated regions were known, only 22.5% of the genes contain one or more copies of TATATA sequences in their translated regions, while 52.1% of the genes contained the TATATA sequence(s) in their putative 3′ untranslated

regions, as estimated from the translation termination codon to 300 bp downstream. Considering that the average length of yeast coding sequence is 1.7–1.9 kb, TATATA appears much more frequently (14-fold) in the 3′ untranslated regions than in the translated regions. Furthermore, among those genes that do not have a TATATA sequence in their 3′ untranslated regions, most of them contain the weaker signals mentioned above (data not shown), similar to those in the *CYC1*$^+$ gene.

Although TATATA appears to be the strongest and most frequently used upstream signal in yeast genes, the *CYC1*$^+$ gene employed the weaker signals TTTATA, TATGTT, and TATTTA that cumulatively constitute a strong signal.

We thank Linda Comfort for synthesis of the oligonucleotides used in this study. This work was supported by National Institutes of Health Grant R01 GM12702 to F.S. and National Science Foundation Grant MCB 93116644 to J.S.B.

1. Higgs, D. R., Goodbourn, S. E. Y., Lamb, J., Clegg, J. B. & Weatherall, D. J. (1993) *Nature (London)* **306,** 398–400.
2. Orkin, S. H., Cheng, T. C., Antnarakis, S. E. & Kazazian, H. H. (1985) *EMBO J.* **4,** 453–456.
3. Pulak, R. & Anderson, P. (1993) *Genes Dev.* **7,** 1885–1897.
4. Rund, D., Dowling, C., Najjar, K., Rachmilewitz, E. A., Kazazian, H. H. & Oppenheim, A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 4324–4328.
5. Zaret, K. S. & Sherman, F. (1982) *Cell* **28,** 563–573.
6. Russo, P., Li, W.-Z., Hampsey, D. M., Zaret, K. S. & Sherman, F. (1991) *EMBO J.* **10,** 563–571.
7. Osborn, B. I. & Guarente, L. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4097–4101.
8. Sadhale, P. P. & Platt, T. (1992) *Mol. Cell. Biol.* **12,** 4262–4270.
9. Russo, P., Li, W.-Z., Guo, Z. & Sherman, F. (1993) *Mol. Cell. Biol.* **13,** 7836–7849.
10. Butler, J. S. & Platt, T. (1988) *Science* **242,** 1270–1274.
11. Russo, P. & Sherman, F. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8348–8352.
12. Wu, S.-Y. (1993) Ph.D. thesis (Univ. of Rochester, New York).
13. Sherman, F. & Slonimski, P. P. (1964) *Biochim. Biophys. Acta* **90,** 1–15.
14. Sherman, F., Stewart, J. M., Parker, J. H., Inhaber, E., Shipman, N. A., Putterman, G. J., Gardisky, R. L. & Margoliash, E. (1968) *J. Biol. Chem.* **243,** 5446–5456.
15. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987) *Methods Enzymol.* **154,** 367–382.
16. Li, W.-Z. & Sherman, F. (1991) *Mol. Cell. Biol.* **11,** 666–676.
17. Rose, M. D., Novick, P., Thomas, J. H., Botstein, D. & Fink, G. R. (1987) *Gene* **60,** 237–243.
18. Irniger, S., Sanfacon, H., Egli, C. M. & Braus, G. H. (1992) *Mol. Cell. Biol.* **12,** 2322–2330.
19. Irniger, S. & Braus, G. H. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 257–261.
20. Heidmann, S., Schindewolf, C., Stumpf, G. & Domdey, H. (1994) *Mol. Cell. Biol.* **14,** 4633–4642.