Functional analysis of mRNA 3' end formation signals in the convergent and overlapping transcription units of the *S.cerevisiae* genes *RHO1* and *MRP2*

Julie A.Peterson and Alan M.Myers*

Department of Biochemistry and Biophysics, Iowa State University, Ames, IA 50011, USA

Received June 9, 1993; Revised and Accepted October 12, 1993

ABSTRACT

The *Saccharomyces cerevisiae* genes *RHO1* and *MRP2* are convergently transcribed, with 281 base pairs separating their termination codons. Transcript mapping revealed at least 111 base pairs within the *RHO1-MRP2* intercoding region are transcribed in both directions. Transplacement experiments showed distinct sequences of 70 nt for *MRP2* and 179 nt for *RHO1* were sufficient for normal mRNA 3' end formation. The *MRP2* signal functioned in either orientation, although relatively inefficiently in the non-native orientation. This element contains a polyAT sequence essential for 3' end formation in both orientations. *RHO1* or *MRP2* 3' end formation was not affected by overproduction or elimination of the complementary, natural antisense transcript. In contrast, insertion of a strong promoter that extended antisense transcripts beyond their normal 3' ends inactivated either *MRP2* or *RHO1*. These data suggest that transcript termination in the compact yeast genome can be important to prevent inactivation of downstream genes as a result of antisense transcription.

INTRODUCTION

Nucleotide sequence analysis of chromosome III from the yeast *Saccharomyces cerevisiae* revealed very dense packaging of genetic coding information, with open reading frames located on average every 1.73 kilobase pairs (kb) (1). Such dense packaging is not likely to occur in multicellular eucaryotes, given the large discrepancy between gene number and DNA content. The unusually compact yeast genome may have evolved in part through elimination of non-coding DNA sequences between genes, which would explain the frequency of convergently transcribed open reading frames separated by 300 base pairs or less (1–6).

Genes closely opposed at the 3' end raise the question of whether convergent transcription influences gene expression. In animal cells transcription is known to extend a significant distance beyond each coding region. This situation in yeast would involve

production of long regions of antisense RNA as part of the normal transcription process, which might affect gene expression at a post-transcriptional level. Convergent transcription could also affect gene expression by promoter exclusion, i.e. active transcription of one gene might prevent initiation at the opposite promoter. Thus, the relatively short regions of intercoding DNA located between these convergently transcribed genes in yeast may provide efficient bidirectional transcription termination-and/or pre-mRNA cleavage and polyadenylation signals.

Correct mRNA 3' end formation is an essential step in gene expression (reviewed in Refs. 7–9). Mammalian transcript processing involves the sequence AAUAAA, known to specify cleavage and polyadenylation of transcripts extended hundreds to thousands of bases downstream from their mature termini. Mutations in this conserved sequence caused inefficient cleavage and polyadenylation, leading to decreased gene expression (10, 11). Most yeast genes lack the AAUAAA processing signal or any other single, conserved sequence element responsible for 3' end processing. Thus, *S. cerevisiae* may use multiple mechanisms for specification of mRNA 3' ends. Another potential difference between yeast and mammals regarding mRNA 3' end formation is the relationship between transcription termination and polyadenylation. All mRNAs in *S. cerevisiae* are thought to be polyadenylated, including those coding for histone proteins (12), suggesting transcription termination and polyadenylation are coupled events. Furthermore, transcription termination assays indicated a yeast 3' end formation signal sequence functions to stop processon of RNA polymerase (13, 14). Yeast precursor mRNAs, however, are cleaved *in vitro* in the presence of cell free extracts (15), suggesting transcription extends beyond the 3' end cleavage and polyadenylation site. Cleavage and polyadenylation are not necessarily coupled, because conditional mutants lacking poly(A) polymerase still cleave pre-mRNAs normally (16), and precleaved RNA is polyadenylated specifically *in vitro* (17, 18).

Several nucleotide sequence elements were identified in yeast that function to specify transcription termination and/or polyadenylation sites. The tripartite sequence TAG ... TA(T)GT ... TTT is essential for 3' end formation of *CYC1* mRNA (14,19). Related sequence elements were implicated in 3' end formation

* To whom correspondence should be addressed

Table 1. Plasmids

Name	Insert	Nucleotides Included ¹	Insert Source	Recipient Vector	Indicator
pJP250	1404 bp <i>HincII</i>	–982–422	pG89/ST3 ²	pUC118	
pRP	396 bp <i>EcoRI</i> – <i>HincII</i>	27–422	pREL+27 ³	pUC118	
pMp	333 bp <i>BglIII</i> – <i>EcoRI</i>	1050–1382	pG89/ST3 ²	pUC119	
pJP252	397 bp <i>HpaI</i> – <i>BglIII</i>	654–1050	pG89/ST3 ²	pBS4	
pJP207	1570 bp <i>EcoRI</i>	–190–1380	pG89/ST3 ²	pUC118E ⁵	
pJP208	1767 bp <i>BamHI</i> ; (<i>HIS3</i>) ⁶		pUC18H ⁷ <i>HpaI</i> site	pJP207;	JP-8
pJP210	1767 bp <i>BamHI</i> (<i>HIS3</i>) ⁸		pUC18H ⁷ <i>NruI</i> site	pJP207;	JP-9
pJP211	1767 bp <i>BamHI</i> (<i>HIS3</i>) ⁶		pUC18H ⁷ <i>NruI</i> site	pJP207;	JP-10
pJP214	1168 bp <i>HindIII</i> (<i>URA3</i>) ⁹		pUC18U ⁷ <i>HpaI</i> site	pJP207;	JP-12
pJP215	1168 bp <i>HindIII</i> (<i>URA3</i>) ¹⁰		pUC18U ⁷ <i>HpaI</i> site	pJP207;	JP-13
pJP205	1570 bp <i>EcoRI</i>	–190–1380	pG89/ST3 ²	pBS4	
pJP212	pJP205 with artificial <i>NruI</i> site at nt 831				
pJP220	1168 bp <i>HindIII</i>		pUC18U ⁷	pUC 118	
pJP227	1168 bp <i>HindIII</i>		pUC18U ⁷	YEp351 ¹¹	
pJP224	245 bp <i>HpaI</i> – <i>NruI</i> ¹²	654–899	pG89/ST3 ²	pJP220	RHN
pJP223	245 bp <i>HpaI</i> – <i>NruI</i> ¹³	654–899	pG89/ST3 ²	pJP220	MNH
pJP221	180 bp <i>HpaI</i> – <i>NruI</i> ¹²	654–834	pJP212	pJP220	RHN*
pJP226	180 bp <i>HpaI</i> – <i>NruI</i> ¹³	654–834	pJP212	pJP220	MN*H
pJP222	70bp <i>NruI</i> – <i>NruI</i> ¹²	832–899	pJP212	pJP220	RN*N
pJP225	70bp <i>NruI</i> – <i>NruI</i> ¹³	832–899	pJP212	pJP220	MNN*
pJP230	1431 bp <i>HindIII</i>		pJP223	YEp351 ¹¹	MNH
pJP232	1431 bp <i>HindIII</i>		pJP224	YEp351 ¹¹	RHN
pJP229	1348 bp <i>HindIII</i>		pJP221	YEp351 ¹¹	RHN*
pJP233	1348 bp <i>HindIII</i>		pJP226	YEp351 ¹¹	MN*H
pJP228	1238 bp <i>HindIII</i>		pJP222	YEp351 ¹¹	RN*N
pJP231	1238 bp <i>HindIII</i>		pJP225	YEp351 ¹¹	MNN*
pJP236	pJP223 with nts 878–893 deleted		----	----	MNHΔ
pJP237	1415 bp <i>HindIII</i>		pJP236	YEp351 ¹¹	MNHΔ
pJP240	pJP224 with nts 878–893 deleted				RHNΔ
pJP241	1415 bp <i>HindIII</i>		pJP240	YEp351 ¹¹	RHNΔ
pJP238	pJP225 with nts 878–893 deleted		----	----	MNN*Δ
pJP239	1332 bp <i>HindIII</i>		pJP238	YEp351 ¹¹	MNN*Δ
pJP234	pJP222 with nts 878–893 deleted		----	----	RN*NΔ
pJP235	1222 bp <i>HindIII</i>		pJP234	YEp351 ¹¹	RN*NΔ
pJP242	pJP238 with AT rich sequence replacing nt 878–893		----	----	MNN*R
pJP243	1348 bp <i>HindIII</i>		pJP242	YEp351 ¹¹	MNN*R

1. Nt numbers are defined relative to the 5' nt of the *RHO1* initiation codon as +1, and are shown only for fragments from the *RHO1-MRP2* region.

2. Ref. 4.

3. J. Johnson and A. Myers (unpublished).

4. Stratagene Cloning Systems, La Jolla, CA.

5. pUC118 modified by replacement of the region between the two *PvuII* sites (containing the multiple cloning site) with a unique *EcoRI* site.

6. Oriented so that the *HIS3* and *RHO1* coding regions are on the same strand.

7. pUC18 with an inserted 1767 bp *BamHI* fragment containing *HIS3* (pUC18H), or an inserted 1168 bp *HindIII* fragment containing *URA3* (pUC18U).

8. Oriented so that the *HIS3* and *RHO1* coding regions are on opposite strands.

9. Oriented so that the *URA3* and *RHO1* coding regions are on the same strand.

10. Oriented so that the *URA3* and *RHO1* coding regions are on opposite strands.

11. Ref. 33.

12. *RHO1* coding strand is transcribed from the *URA3* promoter.

13. *RHO1* noncoding strand is transcribed from the *URA3* promoter.

in other genes including *ADH2*, *ARO4*, *TRP1*, and *TRP4* (20, 21) and in the polyadenylation signal of cauliflower mosaic virus functioning in yeast (8). A 26 bp, AT-rich sequence containing an alternating T-A element is required for 3' end formation in *GAL7* RNA (18,22), and an AT repeat element was also

implicated in termination of the *CYC1* mRNA (23). Finally, the sequence TTTTATA was essential for transcription termination of a *Drosophila melanogaster* gene expressed in yeast (24).

This study characterized transcription of the 281 nt region located between the convergently transcribed yeast genes *RHO1* and *MRP2* (3,4). The results show this intercoding sequence contains two adjacent, functionally independent 3' end formation signals, each of which contains a sequence motif implicated previously as a yeast polyadenylation signal. The 3' ends of the *RHO1* and *MRP2* mRNAs are natural antisense RNAs complementary to each other, however, neither transcript influences production of its counterpart. Insertion of a promoter that directed transcription beyond the normal polyadenylation sites inactivated the gene located downstream, suggesting the intercoding region functions not only to specify polyadenylation sites, but also to terminate transcription.

MATERIALS AND METHODS

DNA manipulations and plasmid construction

DNA manipulations were carried out by standard procedures (25, 26). The yeast cDNA library was supplied by Clontech Laboratories, Inc. (Palo Alto, CA). Plasmids are described in Table 1, with the exception of pREL+27. Plasmid pREL+27 contains a 1359 base pair *EcoRI* fragment in pUC118 (27), comprising *RHO1* nucleotides (nt) 27–1385 (nt numbers are defined using the A residue of the *RHO1* initiation codon as nt 1). This insert was constructed by exonucleolytic digestion of the 1.6 kb *EcoRI* fragment containing *RHO1* (3), and subsequent insertion of an *EcoRI* linker at nt 27.

Plasmid pJP212 was formed from pJP205 by site directed mutagenesis (28) using oligonucleotide JP28 (Table 2) to create an *NruI* site at nt 831. Site specific mutagenesis also was used to remove from several plasmids the AT repeat sequence TA16 (nts 878–893). Thus, mutagenesis of pJP223 and pJP225 with oligonucleotide JP32 formed pJP236 and pJP238, respectively, and mutagenesis of pJP222 and pJP224, with oligonucleotide JP33 formed pJP234 and pJP240, respectively. Plasmid pJP242 was formed from pJP238 using oligonucleotide JP34 to insert a TA rich, random sequence into the previous TA16 deletion. All insertions in *URA3* were confirmed by nucleotide sequence analysis (29) using oligonucleotide primer JP31 (data not shown).

Strains and media

Escherichia coli strain TG-1 (Amersham Corp., Arlington Hts., IL) was used for all plasmid amplifications. Except where noted, yeast strains used in this study were derived from wild type strain W303-1B (*MATα ade2, his3, leu2, trp1, ura3*) or the isogenic diploid strain W303 (30). Yeast strains JP-8, JP-9, JP-10, JP-12, and JP-13 were constructed by gene replacement using recombinant *EcoRI* fragments excised from plasmids pJP208, pJP210, pJP211, pJP214, and pJP215, respectively (Table 1). The locations and orientations of all chromosomal insertions were confirmed by Southern hybridization analysis (data not shown).

The following media were used: YPD (1% yeast extract, 2% peptone, 2% glucose); TWO (1.2 M sorbitol, 2% glucose, 0.7% yeast nitrogen base without amino acids); WO (2% glucose, 0.7% yeast nitrogen base without amino acids); EG (1% yeast extract, 2% peptone, 2% glycerol, 2% ethanol); KAc (1% potassium acetate, 0.05% glucose, 0.1% yeast extract). Minimal media were supplemented as needed with uracil, leucine, histidine, tryptophan and adenine at 20 μg/ml.

Table 2. Oligonucleotides

Name	Sequence (5' to 3')	Position ¹	Strand ²
JP26	CCAGGGTTTATCAATGCTCG	1220–1201	B
JP27	GACCATGTGCATACCCAGAC	224–243	A
JP28	ATCATCTATTCGCGAAAGAAGAAT	822–841	A
JP29	CTTCTTAGCTTTACCATTCG	588–569	B
JP30	CATTCGTAGTATCCGGTTGC	491–472	B
JP31	GATGTTTCGTACCACCAAGGA	335–354	C
JP32	CCATGCGAAAGGGATACGCAC ³	900–868	B
JP33	GTGCGTATCCCTTTTCGCATGG ³	868–900	A
JP34	CATGCGAAAGAATTTTAAACA- ACTTAGGATACGCAC ³	900–868	B
SP103	CAGGCACCATCACC	62–49	B
p323	GCGTCGACTCTAGAGAATTC(T) ¹⁷	----	
p324	GCGTCGACTCTAGAGAATTC	----	

1. Nucleotide numbers are defined relative to the 5' nucleotide of the *RHO1* initiation codon as +1

2. 'A' indicates oligonucleotides from the *RHO1* coding strand, and 'B' marks those from the *RHO1* non-coding strand. 'C' indicates an oligonucleotide from the *URA3* coding strand (31).

3. The first four bases of JP32, the last four bases of JP33, and the first three bases of JP bases in the *URA3* gene.

Nucleic acid hybridization

Double stranded DNA probes were synthesized by random primer labeling (31). *MRP2* probe BE333 was excised from plasmid pMP by digestion with *EcoRI* and *HindIII*. *RHO1* probe EC396 was excised from plasmid pRP by digestion with *EcoRI* and *BamHI*. *RHO1* probe EE150 was excised from plasmid pG89/ST3 by digestion with *EcoRI*. *URA3* transcripts were detected with probe D1168 excised from plasmid pUC18U by digestion with *HindIII*. To form single stranded RNA probe ssBH, plasmid pJP252 was linearized with *EcoRI* and transcribed in vitro using bacteriophage T3 RNA polymerase.

Total yeast RNA used for nonhem hybridization analysis was extracted from cells grown to early log phase as described (25). Polyadenylated RNA was isolated by affinity chromatography using a poly-U Sepharose column (Pharmacia LKB Biotechnology, Piscataway, NJ). Northern hybridization analyses using double stranded DNA probes used standard procedures for formaldehyde gel electrophoresis (25). Northern hybridizations using single stranded RNA probes were as follows. Blots were soaked for six hours at 65°C in prehybridization solution (10% dextran sulfate, 1M NaCl, 50% formamide, 1% SDS, 0.1 mg/ml salmon sperm DNA). After adding probe the filters were hybridized overnight at 65°C. They were washed twice for 20 min in 2× SSC, 0.1% SDS at room temperature, and twice for 20 min at 65°C in 0.1% SSC, 0.1% SDS.

SI nuclease protection assays

SI nuclease protection assays followed standard procedures (25,32). The 5' end-labeled probe complementary to *RHO1* mRNA was synthesized from oligonucleotide SP103 (nts 62–49 of the *RHO1* noncoding strand). SP103 was radioactively labeled using γ -³²P-ATP (NEN Research Products) and polynucleotide kinase, then annealed to single stranded DNA coding for *RHO1* in single stranded phagemid pJP250. The oligonucleotide was extended through the 5' noncoding region of *RHO1* using DNA polymerase I Klenow fragment. The resulting double stranded DNA was cleaved with *BamHI* at the unique site restriction upstream of *RHO1* in the multiple cloning sequence of plasmid pJP250. The 1049 nucleotide single stranded probe fragment was purified by alkaline agarose gel electrophoresis and annealed to total cellular RNA. The remaining reactions were carried out

as recommended (25). The region of the probe protected from SI nuclease digestion was determined using a sequencing ladder synthesized using oligonucleotide SP103 as a primer and single stranded pJP250 as the template.

Amplification of cDNA ends by PCR

Rapid amplification of cDNA ends (RACE) used the polymerase chain reaction (PCR) procedure of Frohman *et al.* (33), with minor modifications. Oligonucleotide JP29 was used as a primer for the reverse transcriptase reaction to synthesize the 5' end of *RHO1* cDNA. Polydeoxyadenosine tails were added to the 5' end of the *RHO1* cDNA using terminal deoxynucleotidyl transferase (Gibco BRL). The 5' end of *RHO1* cDNA was amplified by PCR using oligonucleotides JP30 (25 pmoles), p323, (10 pmoles), and p324 (25 pmoles). Oligonucleotide p323 anneals to the polydeoxyadenosine tail of the cDNA during the first amplification cycle. Oligonucleotide p324 anneals to the 5' end sequence provided by p323 during subsequent rounds of PCR amplification.

The 5' end of the *RHO1* cDNA was excised from the amplified fragments by digestion with *HaeIII* and *XbaI*. A *HaeIII* restriction site is located at nt 97 in *RHO1*, and a *XbaI* site is located in PCR amplification primer p324, upstream of the polyadenylation site in the amplified cDNA. These fragments were cloned in phagemid pUC118 and their nucleotide sequences determined by the chain termination method (29).

RACE amplification of cDNA 3' ends was performed basically in the same way as 5' end amplification. The reverse transcriptase primer was p323, expected to anneal to poly A tails of mRNA. Amplification of *RHO1* mRNA 3' ends used oligonucleotides p324 and JP27 (25 pmoles each). The 3' ends were excised from the amplified fragments by digestion with *EcoRI* (site located in p324) and *HincII* (site at nt 654). Amplification of *MRP2* mRNA 3' ends used oligonucleotides p324 and JP26 (25 pmoles each). The 3' ends were excised from the amplified fragments by digestion with *EcoRI* (site located in p324) and *BglII* (site at nt 1047). The 3' ends of cDNAs produced from *URA3-RHO1* or *URA3-MRP2* gene fusions were amplified using oligonucleotides JP31 and p324. The 3' ends were excised from the amplified fragments by digestion with *XbaI* (site in p324) and *EcoRV* (site at *URA3* nt 413 (34), and cloned in phagemid pUC119.

RESULTS

Physical mapping of the *RHO1* and *MRP2* transcripts

Northern hybridization analysis determined the approximate size and corresponding DNA location of the *RHO1* and *MRP2* transcripts. *RHO1* probe EC396 produced a broad band centered at approximately 1050 nt, whereas *MRP2* probe BE333 produced a sharper band of approximately 500 nt (Figure 1). The *RHO1* coding region comprises 627 nt (3), indicating the transcript contains more than 400 nt of non-coding sequence. The *MRP2* transcript contains approximately 150 nt in excess of its 345 nt coding region (4).

RHO1 mRNA contains less than 200 nt upstream of the initiation codon, because probe EE150 did not detect the transcript (Figure 1). This was confirmed by S1 nuclease mapping, which detected one cluster of 5' ends centered at nt -197 and a fainter cluster at nt -74 (data not shown). Further confirmation was obtained by sequence analysis of cDNAs corresponding to the 5' end of the *RHO1* transcript, amplified from total polyadenylated RNA. Two different 5' end clones were obtained, beginning at nt -196 or nt -122. The latter site was not observed by nuclease protection and thus may have derived from premature termination during reverse transcription.

The overall size and 5' end determination of *RHO1* mRNA indicated the transcript must extend approximately 200 nt beyond the translation termination codon, to within approximately 80 nt of the *MRP2* coding region. S1 nuclease protection assays previously located the *MRP2* transcription start sites within 50 nt of the initiation codon (4). Together with the size of the *MRP2* transcript this information indicated polyadenylation sites are located approximately 100 nt downstream of the translation

termination codon. Taken together, the physical mapping data suggested *MRP2* and *RHO1* transcripts overlap.

Sequence analysis of *RHO1* and *MRP2* cDNA clones determined specific polyadenylation sites and thereby defined the minimum extent of transcriptional overlap (Figure 2). Three independent *RHO1* cDNA clones isolated from a λ gt10 genomic library contained poly(A) tails adjacent to nucleotides 814, 831 and 859. Additional 3' end cDNA clones were obtained by PCR amplification. Polyadenylation sites at nts 831 and 859 were again detected in the PCR products, which also revealed seven additional sites. Of the ten total polyadenylation sites detected, six were identified in multiple, independent clones. The *RHO1* polyadenylation sites fell into three clusters, ranging from nts 761–768, nts 814–859, and nts 899–919. These sites are located from 131 to 289 nt beyond the termination codon, in agreement with the physical mapping data. The most distal polyadenylation site detected corresponds to a nucleotide within the *MRP2* coding region (Figure 2).

Sequence analysis of *MRP2* cDNAs revealed poly(A) tails adjacent to five nucleotides scattered from nts 860–808 (Figure 2). Three of the five sites were detected in multiple, independent cDNA clones. This cluster of nucleotides is located 52 to 104 nt distal to the termination codon, in agreement with the previous physical mapping data. Taken together the transcript mapping data indicated at least 111 base pairs (nt 808–919) in the *RHO1*-*MRP2* intercoding region are transcribed in both directions (Figure 2).

Functional independence of the *RHO1* and *MRP2* transcript termination mechanisms

The close approximation of polyadenylation sites in the convergent *MRP2* and *RHO1* transcripts (Figure 2) raised the possibility that transcriptional overlap influences transcript termination. To test this hypothesis, different segments of yeast DNA were inserted into the chromosome between the two genes such that one of the overlapping transcripts would be eliminated. Insertions were at the *HpaI* site 22 nt downstream of the *RHO1* termination codon, or the *NruI* site 10 nt downstream of the *MRP2* stop codon (Figure 3). *RHO1* transcription in strain JP-12 was expected to end within the inserted fragment at the *URA3* termination site, eliminating the natural antisense RNA for *MRP2*. This was verified by northern hybridization analysis using single stranded RNA probe ssBH (nts 654–1050 of the *MRP2* coding strand) (Figure 1), which detects antisense RNA complementary to *MRP2* mRNA. This natural antisense RNA was undetectable by northern hybridization in strain JP-12 (Figure 4). Strain JP-13, identical to strain JP-12 except for the orientation of the *URA3* insert (Figure 3), also lacked antisense RNA complementary to *MRP2* mRNA (Figure 4).

Northern hybridization and nucleotide sequence analysis of cDNA clones were used to determine the effects of each insertion on 3' end formation of *MRP2* transcripts. Probe BE333 (Figure 1) detected *MRP2* transcripts of normal size and abundance in JP-12 and JP-13 (data not shown). cDNA clones from these strains revealed polyadenylation sites at nts 808, 815, 832 and 849, all detected previously in wild type mRNA (Figure 2). Other polyadenylation were found in the same general location, at nts 799, 801, 810, 811, 822, 829, 831, and 842. Thus, absence of the natural antisense mRNA did not affect the general location of polyadenylation in *MRP2* transcripts.

Strain JP-8 allowed examination of *MRP2* mRNA 3' end formation in the presence of excess natural antisense mRNA.

Figure 1. Size determination of *RHO1* and *MRP2* transcripts. (A) The restriction map of the *RHO1*-*MRP2* locus represents 1720 nt and is drawn to scale. Recognition sites are indicated for *EcoRI* (E), *HincII* (C), *HpaI* (H), *NruI* (N), and *BglII* (B). The *EcoRI* site denoted 'E*' is derived from an oligonucleotide linker ligated to an exonucleolytic digestion product; it is not present in the genome. Restriction fragments used as double stranded hybridization probes are indicated. The dotted arrow indicates the location of single-stranded RNA probe ssBH, and points in the 5'–3' direction. (B) *RHO1* and *MRP2* transcripts were detected in northern blots using probes EC396 and BE333, respectively. An RNA size ladder was run on a portion of the same gel. Lanes 1–3 contain 10, 20 and 30 μ g, respectively, of total yeast RNA.

This strain has the 1.7 kb *Bam*HI fragment containing *HIS3* (35) inserted downstream of *RHO1* (Figure 3). Hybridization with probe ssBH showed the level of antisense RNA complementary to *MRP2* mRNA was greatly increased in JP-8 compared to its normal level on the *RHO1* transcript in wild type strains (Figure 4). This antisense transcript presumably arose from the *DED1* promoter downstream of *HIS3* (35) *MRP2* transcripts in JP-8 again were of normal size and abundance (data not shown). cDNA clones from JP-8 indicated polyadenylation sites at nts 832 and 849, also detected in wild type cells (Figure 2). Additional sites were detected in the same general location, at nts 809, 827, 829, 838, and 842. Thus, neither elimination nor excessive production of the overlapping transcript had noticeable effects on *MRP2* expression.

Strain JP-9 tested whether convergent transcription affected 3' end formation of *RHO1* mRNA. This strain contains the 1.7 kb *HIS3* fragment inserted at the *Nru*I site in the intergenic region (Figure 3). This insertion is expected to produce high levels of natural antisense RNA for *RHO1*, as predicted from the analysis of strain JP-8. Neither the size nor the abundance of *RHO1* transcripts detected by probe EC396 was altered in JP-9 (data

not shown). Polyadenylation sites were detected in cDNA clones from JP-9 at nts 761, 763, 768, 831 and 859, all detected in wild type cells (Figure 2). An additional site was detected at nt 863. Thus, like *MRP2*, excessive antisense transcription of the 3' flanking region does not affect *RHO1* expression.

Dissection of the *RHO1* transcript termination signals

To functionally dissect the 3' end formation signals of *RHO1* or *MRP2*, various portions of the intercoding region were inserted in *URA3* (34) at the internal *Nco*I site (Figure 5). Each construct was introduced into wild type strain W303-1B as part of a multicopy plasmid comprising the vector YEp351 (36). Presence of a functional 3' end processing signal in the inserted sequence was detected by observation of truncated *URA3* transcripts in northern hybridization analysis.

Plasmid pJP232 contains nts 654–899 of the intergenic region inserted in *URA3* (Figure 5). This insert, termed RHN (*RHO1* orientation, from *Hpa*I to *Nru*I) contains all but 36 nt of the intergenic region and includes all polyadenylation sites detected in wild type *RHO1* and *MRP2* transcripts (Figure 2). The control strain for this experiment, JP-14, contains an unmodified *URA3*

Figure 2. *RHO1* and *MRP2* polyadenylation sites. (A) Representative sequence data shows the antisense strand of 3' end cDNA clones. The nucleotide number of the polyadenylation site is shown below each sequence. (B) The sequence between the *RHO1* and *MRP2* termination codons (nts 625–924) is shown. Arrows show the direction of transcription. Termination codons are marked by triple asterisks. Polyadenylation sites determined using PCR are marked by 'X' or 'T' depending on which of two independent cDNA pools was used to obtain the clone. 'O' marks polyadenylation sites determined using the λgt10 yeast cDNA library. Each mark represents one cDNA clone.

Figure 3. Chromosomal insertions within the *RHO1*-*MRP2* intercoding region. The position of an inserted fragment bearing either *URA3* or *HIS3* is indicated for each strain. Recognition sites for *EcoRI* (E), *HpaI* (H), and *NruI* (N) are shown. The dotted arrow indicates the transcription initiation site of the *DED1* gene, and points in the 5'–3' direction.

Figure 4. Transcription of *MRP2* antisense RNA. RNA from the indicated strains was probed with single stranded RNA probe ssBH complementary to 3' sequences downstream of *RHO1* (Figure 1). Total RNA was analyzed except for the wild type sample denoted (A+), which was a polyadenylated fraction. Probing the JP-12 and JP-13 samples with fragment EC396 (Figure 1) revealed a *RHO1* transcript signal of approximately the same intensity as that observed in wild type RNA (data not shown). Molecular sizes were estimated using an RNA size ladder run on a portion of the same gel.

gene as part of the transforming plasmid. Northern hybridization showed JP-14 contained elevated levels of normal *URA3* transcripts compared to the parent W303-1B, as expected from the additional copies of *URA3* (Figure 6A,B). In contrast, transcripts from the gene containing insert RHN were shorter than native *URA3* transcripts (Figure 6A). Their length of approximately 650 nt corresponds to the size expected if insert RHN caused cleavage and polyadenylation at native *RHO1* 3' ends.

To localize further the sequences responsible for truncation of *URA3* transcripts, RHN was divided into two smaller

Figure 5. Fragments tested for 3' end formation signals. The map at the top of the figure represents *RHO1*-*MRP2* nts 580–980. The figure is drawn to scale. Recognition sites are indicated for *HpaI* (H) and *NruI* (N). 'N*' represents a *NruI* site constructed by site-specific mutagenesis. Each line represents a fragment of the *RHO1*-*MRP2* intercoding region inserted into the *URA3* coding sequence at the unique *NcoI* site (34). The solid arrows show the direction of *URA3* transcription. The open triangle marked 'ΔTA16' indicates deletion of the TA16 sequence element. The open triangle marked '+16' indicates TA16 was replaced by a 16 nt scrambled, AT-rich sequence.

fragments using an artificial restriction site, *NruI**, positioned at nt 831 by site directed mutagenesis (Figure 5). Fragment RHN* comprises nts 654–834, and fragment RN*N comprises nts 834–899. Insertion of RHN* caused truncation of *URA3* transcripts (Figure 6A). In contrast, the majority of *URA3* transcripts were full length when fragment RN*N was inserted, although a relatively small amount of truncated transcript was present (Figure 6A). Thus, a functional *RHO1* 3' end formation signal is located within fragment RHN*. An additional 3' end formation signal is located within fragment RN*N (nt 834 to nt 899), although this sequence is far less efficient than the upstream signal.

Polyadenylation sites in the truncated *URA3* transcripts were determined by sequence analysis of PCR amplified cDNA clones. Fragment RHN directed polyadenylation at nts 831, 841, 851 and 859 within the insertion, and fragment RHN* caused polyadenylation at nts 833 and 834. Each of these locations is at or near polyadenylation sites of wild type *RHO1* transcripts (Figure 2). Truncated *URA3* transcripts resulting from insertion of fragment RN*N had 3' ends 38 to 52 nucleotides beyond the insertion, within the *URA3* coding region. These sequences are not related to polyadenylation sites of wild type *RHO1* transcripts. Detection of these processed transcripts confirmed that fragment RN*N acts as a weak 3' end formation signal.

Dissection of the *MRP2* transcript termination signals

Inserting portions of the *RHO1*-*MRP2* intergenic region into *URA3* in the *MRP2* orientation allowed detection of *MRP2* 3' end formation signals. Fragment MNH (*MRP2* orientation, from *NruI* to *HpaI*) is identical to fragment RHN except for its opposite orientation in *URA3* (Figure 5). This insertion efficiently caused truncation of *URA3* transcripts (Figure 6B). Fragment MNH was divided into two smaller fragments, MNN* (nts 834–899) and MN*H, (nts 654–834) (Figure 5). Fragment MNN* efficiently caused truncation of *URA3* transcripts, although a relatively small amount of run through transcript was present (Figure 6).

Figure 6. Termination of *URA3* transcripts by the *MRP2* or *RHO1* 3' end formation signals. Total RNA from the indicated strains was hybridized to *URA3* probe D1168. W303-1B is the wild type parent strain. The other strains contain a multicopy plasmid bearing *URA3* or recombinant genes with the indicated insertion in *URA3*. An RNA size ladder was run on a portion of the same gel. (A) *RHO1* 3' end formation signals. The right panel shows two lanes from the same gel as the left panel that were exposed to film for four days; all other exposures were overnight. (B) *MRP2* 3' end formation signals.

Fragment MN*H did not cause detectable truncation of *URA3* transcripts (Figure 6B). Polyadenylation sites in these truncated transcripts, determined from cDNA clones, were those detected in wild type *MRP2* transcript (nts 830 and 831 for MNH, nt 832 for MNN*). Thus a functional *MRP2* 3' end formation signal is located within the 70 nucleotides of fragment MNN*.

Mutagenesis of the *MRP2* and *RHO1* terminator elements

The *MRP2* terminator fragment MNN* contains a 16 nt sequence consisting of alternating T and A residues termed TA16 (nts 893–878). Because a similar sequence element was implicated previously as a transcription termination signal in *S. cerevisiae* (22), TA16 was examined for this function in *MRP2*. Deletion of TA16 from fragment MNH produced fragment MNHΔ (*MRP2* orientation, from *NruI* to *HpaI* with TA16 deleted [Δ]), which was tested for the ability to direct truncation of *URA3* transcription. The majority of *URA3* transcription proceeded through the MNHΔ insertion (Figure 6B). A minor amount of truncated *URA3* transcript was also detected, however, this material was evident as a broad band not matching the size of the precise termination product seen for the intact MNH terminator. Deletion of TA16 from fragment MNN*, forming fragment MNN*Δ, again eliminated the 3' end formation signal (Figure 6B).

The essential function of TA16 in 3' end formation could derive from sequence specificity, general adenine and thymine content, or spacing of other sequence elements in the signal. To discriminate between these possibilities TA16 was replaced by the 16 nt AT rich, scrambled sequence AATTTTAAACAACTTA. The scrambled sequence was inserted into fragment MNN* in place of TA16, producing fragment MNN*R (*MRP2* orientation, from *NruI* to **NruI*, with TA16 replaced by a scrambled AT rich sequence) (Figure 5). Insertion of this fragment in *URA3* did not cause truncation of the transcript

(Figure 6B). Thus, the essential function of TA16 could not be restored by an AT rich element of the same length.

TA16 is also present in fragment RN*N, which contains a relatively inefficient 3' end formation signal (Figure 6A). To determine if TA16 contributes to *RHO1* 3' end formation it was deleted from fragment RHN to form fragment RHNΔ. Fragment RHNΔ directed truncation of the majority of *URA3* transcripts, however, a significant amount of run-through transcript was present compared to insert RHN (Figure 6A). TA16 was deleted also from fragment RN*N, forming fragment RN*NΔ. In this instance only full length transcripts were detected (Figure 6A); the rare truncated products observed for the parent fragment RN*N were not present.

Effects of antisense transcription of the *MRP2* or *RHO1* coding region

Analysis of strain JP-8 had confirmed the *HIS3* genomic fragment contains a strong promoter downstream of *HIS3*, presumably that of *DED1* (35). In JP-8 this promoter directs transcription away from *RHO1*, through the intercoding region, towards the 3' end of *MRP2* (Figure 3). Attempts to construct a haploid strain identical to JP-8 except for orientation of the *HIS3* fragment were unsuccessful, suggesting antisense transcription of the *RHO1* coding region from the *DED1* promoter was lethal. To test this hypothesis the *HIS3* fragment was inserted into the *HpaI* site downstream of *RHO1* in the *his3/his3* diploid strain W303. Southern hybridization analysis confirmed the orientation of the insert was such that *DED1* would direct transcription of the *RHO1* antisense strand (data not shown). Analysis of 50 spore clones derived meiotically from this diploid showed all viable haploid progeny required histidine. Thus, meiotic progeny containing the *HIS3* insertion downstream of *RHO1* were inviable. *RHO1* is known to be essential for viability in yeast (3), suggesting the *HIS3* insertion inactivates *RHO1*.

MRP2 was also inactivated by a downstream insertion of the *HIS3*-containing fragment such that the *DED1* promoter directed transcription of the *MRP2* antisense strand. Strain JP-10 contains *HIS3* inserted in this orientation at the *NruI* site downstream of *MRP2* (Figure 3). JP-10 is unable to grow on non-fermentable medium (EG), indicating the insertion causes respiratory deficiency. The mutant strain complements the respiratory deficiency of a ρ^0 tester strain lacking mitochondrial DNA, indicating a nuclear mutation is responsible for the inability to grow on the nonfermentable substrate. This type of respiratory deficiency is the phenotype caused by *mrp2* mutations (4). Respiratory deficiency was not observed when *HIS3* was inserted in the opposite orientation, in strain JP-9.

DISCUSSION

Identification of the termini of the *RHO1* and *MRP2* transcripts revealed that the 3' ends of these two mRNAs are complementary, and that the two transcription units overlap by at least 111 base pairs. These observations raised the question of how the cis acting signals specifying mRNA 3' end formation of both *RHO1* and *MRP2* are arranged relative to each other in the 281 base pair region separating the two coding sequences., and furthermore, whether the overlapping transcripts are significant for regulation of either *MRP2* or *RHO1* expression.

The 3' end formation signals of both *RHO1* and *MRP2* are located entirely between the *HpaI* and *NruI* sites, because transplacement of this region into *URA3* in either orientation

caused polyadenylation at the same sites observed in wild type *RHO1* or *MRP2* transcripts. Two functionally independent 3' end formation signals were identified in the intercoding region, one operating for *RHO1* and one for *MRP2*. The signals are relatively compact, because fully functional elements were located in regions of 179 bp for *RHO1* and 70 bp for *MRP2*. These data are in agreement with the results of Sadhale and Platt (17), who showed an 80 nt region of the *GAL7* 3' untranslated region is sufficient for cleavage and polyadenylation *in vitro* at native sites. The data indicate that two compact and independently functional mRNA 3' end formation signals are adjacent to each other in the intercoding region, with minimal amounts, if any, of spacer DNA separating the two signal sequences.

The *RHO1* element caused 3' end processing only in its native orientation. In contrast, the *MRP2* element was very efficient in its native orientation, but also caused 3'-end formation at low efficiency when inserted in *URA3* in the opposite orientation. In this orientation the element caused polyadenylation at sites within *URA3* sequences downstream of the *MRP2* insertion. The distance between the downstream end of the *MRP2* sequence and the chosen polyadenylation sites, however, is approximately the same in the mutant *URA3* gene and native *RHO1* (sites at nt 899 and 919; Figure 2). This result might be explained by a mechanism in which polyadenylation sites are located a fixed number of nucleotides downstream of a specific sequence element within the *MRP2* 3' end formation signal. Thus, in the chromosome, the same 70 bp region of DNA that provides a 3' end formation signal for *MRP2* transcripts may also signal 3' end formation of convergently transcribed *RHO1* transcripts extending beyond the major *RHO1* 3' end signal. According to this hypothesis, the distal *RHO1* polyadenylation sites at nt 899 and 919 (Figure 2) would be the result of a bidirectional 3' end formation signal in *MRP2*.

Both the *RHO1* and *MRP2* 3' end formation signals contain sequence motifs implicated previously in transcript processing in yeast. The sequence TAG-19 nt-TATG-9 nt-TT is located from nt 746 to nt 782 in the *RHO1* signal. This sequence resembles the tripartite motif found in *CYC1*, *TRP1*, *TRP4*, and *ARO4* (19,21). The *TRP1*, *TRP4*, and *ARO4* elements all functioned with relatively low efficiency as a bi-directional 3' end formation signal in transplacement experiments, in contrast to the *RHO1* element which is unidirectional. The tripartite element is necessary but not sufficient to direct 3' end formation (21). Thus, sequence elements in addition to the tripartite motif may be responsible for bidirectional function of the *TRP1*, *TRP4* and *ARO4* 3' end formation signals.

The *MRP2* 3' end formation signal contained a different sequence motif identified previously in the *GAL7* gene (22), specifically the TA16 sequence element (dAdT repeated eight times). Deletion analysis showed TA16 is essential for 3' end formation. The requirement of TA16 is likely to be sequence specific, because replacement of the element with an AT rich sequence of the same length did not restore 3' end formation function. The replacement sequence, however, did contain two G:C base pairs, leaving open the possibility that secondary structural features in this region provide the essential function. The obvious structural symmetry of TA16 suggests the bidirectional function of the *MRP2* 3' end formation signal stems from recognition of this sequence element. Sequence elements in addition to TA16 must function in the *MRP2* signal sequence, because of the clear discrepancy in 3' end formation efficiency between the native and inverted orientation.

Overlapping transcription units in yeast chromosomes have been observed repeatedly, indicating that antisense transcription occurs without preventing expression of the gene comprising the corresponding sense strand (5, 6). The current study provides another example of this situation. Furthermore, neither increasing the abundance of the antisense transcript nor eliminating antisense transcription had any noticeable effect on *MRP2* or *RHO1* expression, at the level of mRNA abundance, polyadenylation site selection, or genetic function, as long as the extent of antisense transcription was the same as that of wild type cells. Similarly, Malavasic and Elder (5) found that preventing overlapping transcription between *SP012* and *SP016* did not block expression of either gene.

We examined further whether increasing the length of antisense transcription beyond that which occurs in wild type cells can affect expression of the corresponding sense gene. To do so, a strong promoter was inserted downstream of the mRNA 3' end formation signal of *RHO1*, oriented so that antisense transcription would extend into the *MRP2* coding region. The results of this experiment suggest that high level antisense transcription can inactivate a gene if the convergent transcript extends into that gene's coding sequence. These conclusions are illustrated by comparing the effects of *HIS3* insertions in strains JP-8 and JP-10 (Figure 3). In JP-10 the *DED1* promoter directs transcription of the *MRP2* antisense strand, and resulted in inactivation of the gene. In JP-10 the *DED1* promoter is moved 240 bases downstream of *MRP2*, beyond the *RHO1* 3' end formation signal. In this instance transcription from *DED1* had no apparent effect on *MRP2* expression. The same situation could be inferred for *RHO1*. In strain JP-9 the presumed high level transcription of the 3' flanking region from the *DED1* promoter did not affect *RHO1* expression. In contrast, moving the promoter 240 bases, upstream of the *MRP2* 3' end formation signal, seemingly caused gene inactivation. Thus, we propose that efficient transcription termination functions are required in some instances to prevent interference with gene expression.

Other explanations can be proposed to explain the observed inactivation of *RHO1* and *MRP2* by proximal insertion of the *DED1* promoter. Insertion of the *HIS3* sequence may eliminate any mRNA 3' end formation signal, so that polyadenylation of the mRNAs fails to occur. Another possibility is that a region of the *HIS3* insert transcribed in the 3' flanking region causes the transcripts to be rapidly degraded. Although neither of these possibilities can be ruled out by the available data, it is clear that the native 3' flanking region or mRNA 3' end formation signals are not required for functional expression of either *RHO1* or *MRP2*. This conclusion is drawn from the observation that plasmids bearing truncated versions of *RHO1* or *MRP2* ending at the *HpaI* or *NruI* sites, respectively, complement deletion mutations of the corresponding chromosomal gene; in these instances multiple transcripts extending into the plasmid sequence were detected by Northern hybridization analysis (JAP, unpublished results). Different sequences were located downstream of the two genes in these plasmids. Thus, in at least two instances non-native sequences immediately downstream of the *MRP2* or *RHO1* coding region support expression of the gene. Thus, we favor the hypothesis that the observed inactivation of *MRP2* and *RHO1* is the result of high level antisense transcription of the coding region.

Relatively high levels of convergent transcription seemingly were required to inactivate either *RHO1* or *MRP2*. This conclusion stems from the knowledge that both orientations of

the *HIS3*-containing fragment would insert an antisense promoter, that of either *DED1* or *PET56* (35). *DED1* transcripts, however, accumulate to a significantly higher level than *PET56* transcripts (35). The steady state level of the transcript produced from *DED1* in strain JP-8 confirmed the strength of this promoter (Figure 3).

Gene inactivation by convergent transcription could be either a *cis* or *trans* acting effect. A *cis* acting effect would indicate direct interference with transcription initiation or elongation, whereas a *trans* acting effect would suggest antisense RNA interfered with translation or another post-transcriptional level of gene expression. *RHO1* inactivation by convergent transcription from the *DED1* promoter is recessive in a diploid containing a wild type *RHO1* gene. In this instance, therefore, gene inactivation by convergent transcription is likely to be a *cis* acting effect. A *trans* acting effect can not be ruled out, however, because the gene dosage of *RHO1* relative to *DED1* varies in the heterozygous diploid and its haploid derivatives.

ACKNOWLEDGEMENTS

We thank C.Dieckmann and D.Larson for critical reading of the manuscript prior to submission for publication. This research was supported by award GM39254 from the National Institutes of Health.

REFERENCES

- Ohver, S.G. *et al.* (1992) *Nature*, 357, 38–46.
- Liu, Y. and Dieckmann, C.L. (1989) *Mol. Cell. Biol.*, 9, 3323–3331.
- Madaule, P., Axel, R. and Myers, A.M. (1987) *Proc. Natl. Acad. Sci. USA*, 84, 779–783.
- Myers, A.M., Crivellone, M.D. and Tzagoloff, A. (1987) *J. Biol. Chem.*, 262, 3388–3397.
- Malavasic, M.J. and Elder, R.T. (1990) *Mol. Cell. Biol.*, 10, 2809–2819.
- McGrath, J.P. and Varshavsky, A. (1989) *Nature*, 340, 400–404.
- Proudfoot, N. (1991) *Cell*, 64, 671–674.
- Wahle, E. and Keller, W. (1992) *Ann. Rev. Biochem.*, 61, 419–440.
- Wickens, M. (1991) *Trends Biochem. Sci.*, 15, 277–281.
- Higgs, D., Goodboun, S., Lamb, J., Clegg, J., Weatherall, D. and Proudfoot, N. (1983) *Nature*, 306, 398–400.
- Orkin, S., Cheng, T., Antonarkin, S. and Kazazian, H. (1985) *EMBO J.*, 4, 453–456.
- Farhmer, K., Yarger, J. and Hereford, L. (1980) *Nucleic Acids Res.*, 8, 5725–5737.
- Osborne, B. and Guarente, L. (1989) *Proc. Natl. Acad. Sci. USA*, 86, 4097–4101.
- Russo, P. and Sherman, F. (1989) *Proc. Natl. Acad. Sci. USA*, 86, 8343–8352.
- Butler, J. and Platt, T. (1988) *Science*, 242, 1270–1274.
- Patel, D. and Butler, S. (1992) *Mol. Cell. Biol.*, 12, 3297–3304.
- Sadhale, P.P. and Platt, T. (1992) *Mol. Cell. Biol.*, 12, 4262–4270.
- Chen, J. and Moore, C. (1992) *Mol. Cell. Biol.*, 12, 3471–3481.
- Zaret, K.S. and Sherman, F. (1982) *Cell*, 28, 563–573.
- Hyman, L.E., Seiler, S.H., Whoriskey, J. and Moore, C.L. (1991) *Mol. Cell. Biol.*, 11, 2004–2012.
- Irniiger, S., Egh, C.M. and Braus, G.H. (1991) *Mol. Cell. Biol.*, 11, 3060–3069.
- Abe, A., Hiraoka, Y. and Fukasawa, T. (1990) *EMBO J.*, 9, 3691–3697.
- Russo, P., Li, W.-Z., Hampsey, D.M., Zaret, K.S. and Sherman, F. (1991) *EMBO J.*, 10, 563–571.
- Henikoff, S. and Cohen, E.H. (1984) *Mol. Cell. Biol.*, 4, 1515–1520.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A., Seidman, J.G. and Struhl, K. (1989) *Current protocols in molecular biology*. John Wiley and Sons, New York.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning. A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Vieira, J. and Messing, J. (1987) *Methods Enzymol.*, 153, 3–11.
- Sayers, J., Schmidt, W. and Eckstein, F. (1988) *Nucleic Acids Res.*, 16, 791–802.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, 74, 5463–5467.
- Wallis, J.W., Chrebet, G., Brodsky, G., Rolfe, M. and Rothstein, R. (1989) *Cell*, 58, 409–419.
- Feinberg, A. and Vogelstein, B. (1983) *Analyt. Biochem.*, 132, 6–13.
- Berk, A. and Sharp, P. (1977) *Cell*, 12, 721–732.
- Frohman, M., Dush, M. and Marlin, G. (1988) *Proc. Natl. Acad. Sci. USA*, 998–9002.
- Rose, M.D., Grisafi, P. and Botstein, D. (1984) *Gene*, 29, 113–124.
- Struhl, K. (1985) *Nucleic Acids Res.*, 13, 8587–8601.
- Hill J.E., Myers, A.M., Koerner, T.J. and Tzagoloff, A. (1986) *Yeast*, 2, 163–167.