MOLECULAR AND CELLULAR BIOLOGY, Mar. 1995, p. 1689–1697 Vol. 15, No. 3
0270-7306/95/$04.00+0

REF2 Encodes an RNA-Binding Protein Directly Involved in Yeast mRNA 3′-End Formation

ROLAND RUSSNAK, KEITH W. NEHRKE, AND TERRY PLATT*

Department of Biochemistry, University of Rochester, Rochester, New York 14642

Received 4 November 1994/Returned for modification 30 November 1994/Accepted 16 December 1994

The *Saccharomyces cerevisiae* mutant *ref2-1* (REF = RNA end formation) was originally identified by a genetic strategy predicted to detect decreases in the use of a *CYC1* poly(A) site interposed within the intron of an *ACT1-HIS4* fusion reporter gene. Direct RNA analysis now proves this effect and also demonstrates the *trans* action of the *REF2* gene product on cryptic poly(A) sites located within the coding region of a plasmid-borne *ACT1-lacZ* gene. Despite impaired growth of *ref2* strains, possibly because of a general defect in the efficiency of mRNA 3′-end processing, the steady-state characteristics of a variety of normal cellular mRNAs remain unaffected. Sequencing of the complementing gene predicts the Ref2p product to be a novel, basic protein of 429 amino acids (M_r, 48,000) with a high-level lysine/serine content and some unusual features. Analysis in vitro, with a number of defined RNA substrates, confirms that efficient use of weak poly(A) sites requires Ref2p: endonucleolytic cleavage is carried out accurately but at significantly lower rates in extracts prepared from Δ*ref2* cells. The addition of purified, epitope-tagged Ref2p (Ref2pF) reestablishes wild-type levels of activity in these extracts, demonstrating direct involvement of this protein in the cleavage step of 3′ mRNA processing. Together with the RNA-binding characteristics of Ref2pF in vitro, our results support an important contributing role for the *REF2* locus in 3′-end processing. As the first gene genetically identified to participate in mRNA 3′-end maturation prior to the final polyadenylation step, *REF2* provides an ideal starting point for identifying related genes in this event.

In all eucaryotic cells, the successful transfer of information from chromatin to the translational machinery depends on a series of mRNA processing events which include capping, splicing, and 3′-end formation. This latter reaction involves endonucleolytic cleavage followed by the addition of multiple adenylate residues (46) and is unique in that it liberates nascent mRNA to be transported from the nucleus and, in doing so, probably signals the elongating transcriptional apparatus to terminate. Thus, the selection and use of polyadenylation sites by 3′-end processing factors are of great biological interest and, in addition, provide a relatively simple system for analyzing a variety of RNA-protein and protein-protein interactions.

A desire to understand the molecular details of this event has provoked extensive biochemical fractionation of mammalian cellular extracts that are competent for cleavage and polyadenylation in vitro, and these efforts have resulted in the purification of a number of *trans*-acting factors. In the case of poly(A) polymerase (32, 47) and the 50-kDa (45) and 64-kDa (44) subunits of CstF (cleavage stimulation factor), which recognizes GU-rich elements downstream of the AAUAAA polyadenylation signal, the corresponding genes have also been isolated. Although significant progress has been made, complete understanding of the cellular mechanisms and the role played by 3′-end processing will be greatly facilitated by the identification of mutants defective in these reactions.

The budding yeast *Saccharomyces cerevisiae* is ideally suited to the analysis of a basic cellular event such as 3′-end processing, since it offers both a tractable genetic system as well as an established in vitro biochemical assay (4). This powerful complementary approach has attracted a number of investigators, but somewhat surprisingly, yeast mutants which influence directly those reactions involved in cleavage and polyadenylation remain elusive. The only mutation demonstrably involved directly in 3′-end maturation thus far is a conditional lethal allele of the *S. cerevisiae* poly(A) polymerase gene (*pap1-1* [29]), which was isolated by screening the Hartwell collection of *ts* yeast mutants for cleavage and polyadenylation in vitro. At nonpermissive temperatures, *pap1-1* cells produce mRNAs which have been properly cleaved at the 3′ end but are deficient in poly(A). In the described biochemical search of well over 100 *ts* strains that yielded *pap1-1*, no extracts defective in the cleavage reaction were identified. The gene encoding *PAP1* has now been isolated by complementation of this defect (29) as well as by classical biochemical means (24).

In this study we have characterized the yeast mutant *ref2-1*, which was originally obtained in a genetic selection designed in this laboratory to isolate genes involved in 3′-end processing (35). We describe the genetic attributes of the *REF2* gene and present evidence for its direct involvement in the endonucleolytic cleavage step of 3′-end processing. The finding that *REF2* is not essential for viability has enabled us to prepare whole-cell extracts depleted for Ref2p and to investigate the role of this protein in cleavage and polyadenylation reactions in vitro. A variant of Ref2p harboring a FLAG epitope at its N terminus (Ref2pF) was expressed in *Escherichia coli* and purified by affinity chromatography. Consistent with the genetic selection of *ref2-1*, this protein can stimulate 3′-end processing of select substrates in extracts devoid of the native protein that are otherwise defective. Ref2pF appears to bind RNA in the absence of other factors, although it contains no known RNA-binding motifs. We speculate that Ref2p may contribute to maximal poly(A) site use by providing an "anchor" to stabilize 3′-end processing complexes. The work described here corroborates the promise of our original selection and opens the door to a more complete genetic analysis of this event.

* Corresponding author. Mailing address: Department of Biochemistry, University of Rochester Medical Center, 601 Elmwood Ave., Rochester, NY 14642. Phone: (716) 275-8244. Fax: (716) 271-2683.

TABLE 1. *S. cerevisiae* strains used in this study

Strain	Genotype	Plasmid
DC35-15D[a]	*MAT***a** *REF2 ura3-52 trp1 leu2-3,112 HOL1 ade2 ade5 gal80 his4-401*::*ACT1-CYC1-HIS4*	
DR5-1D[a,b]	*MAT*α *ref2-1 ura3-52 trp1 leu2-3,112 HOL1 ade5 his4-401*::*ACT1-CYC1-HIS4*	
RR2[c]	*MAT***a** *REF2 ura3-52 trp1 leu2-3,112 HOL1 ade2 ade5 gal80 his4-401*::*ACT1-CYC1-HIS4* [*ACT1-CYC1-lacZ*; *TRP1*]	pSB17
RR3[d]	*MAT*α *ref2-1 ura3-52 trp1 leu2-3,112 HOL1 ade5 his4-401*::*ACT1-CYC1-HIS4* [*ACT1-CYC1-lacZ*; *TRP1*]	pSB17
RR4[c]	*MAT***a** *ref2*::*LEU2 ura3-52 trp1 leu2-3,112 HOL1 ade2 ade5 gal80 his4-401*::*ACT1-CYC1-HIS4*	
RR5[e]	*MAT***a** *ref2*::*LEU2 ura3-52 trp1 leu2-3,112 HOL1 ade2 ade5 gal80 his4-401*::*ACT1-CYC1-HIS4* [*ACT1-CYC1-lacZ*; *TRP1*]	pSB17
RR6-4A[f]	*MAT*α *REF2 ura3-52 leu2-3,112 ade2-101 lys2-801*	
RR6-4D[f]	*MAT***a** *ref2*::*LEU2 ura3-52 leu2-3,112 ade2-101 lys2-801 his3-*Δ*200*	
RR7[g]	*MAT*α *ref2-1 ura3-52 trp1 leu2-3,112 HOL1 ade5 his4-401*::*ACT1-CYC1-HIS4* [*ACT1-CYC1-lacZ*; *TRP1*] [*REF2*; *URA3*]	pSB17, YCpRF2
RR8[h]	*MAT***a** *ref2*::*LEU2 ura3-52 trp1 leu2-3,112 HOL1 ade2 ade5 gal80 his4-401*::*ACT1-CYC1-HIS4* [*ACT1-CYC1-lacZ*; *TRP1*] [*FLAG-RF2*; *URA3*]	pSB17, YCpGFRF2

[a] Derived from strain FC2-12B (27) integrated with pYAH83F (35) and created by Doug Campbell.
[b] Contains spontaneous mutation originally designated *tef2-1* (35).
[c] Transformant of DC35-15D.
[d] Transformant of DR5-1D.
[e] Transformant of RR4.
[f] Haploid segregants of diploid strain SC472 (33) containing one disrupted *REF2* allele.
[g] Transformant of RR3.
[h] Transformant of RR5.

MATERIALS AND METHODS

Strains and media. Genotypes of the *S. cerevisiae* haploid strains used in this study are given in Table 1. Both the wild-type (DC35-15D) and the *ref2-1* (DR5-1D) strains are deleted for the 5.5-kb portion of pYAH83F (35) which includes the *URA3* selective marker. We speculate that the deletion arose by intrachromosomal exchange between duplicate *Sph*I-*Bam*HI sequences (187 bp) flanking the YIp5 portion of the integrated DNA. The plasmid carrying the *ACT1-CYC1-lacZ* fusion gene (pSB17; a gift from Steve Baker) is a derivative of pYABI (22) in which the 83-bp *CYC1* poly(A) site was inserted into the *Xho*I site of the *ACT1* intron. It contains, also, a 1.45-kb *Eco*RI-*Bam*HI fragment harboring the CEN region of chromosome III. For the cloning of the wild-type REF2 gene in YCpRF2, see below.

E. coli DH5α was used exclusively for the maintenance of all plasmids as well as expression of Ref2pF (see below). Yeast strains were grown at 30°C in either YPD medium or SD medium supplemented with the required additives (42). Histidinol selective medium contained 642 mg of histidinol per liter in place of histidine. X-Gal indicator plates were made with 100 mM NaH_2PO_4 (added as a 1 M stock at pH 7.0) and 40 μg of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) per ml.

Yeast cells were transformed by a modified version of the lithium acetate procedure (20). Twenty milliliters of cells at an optical density at 600 nm of 0.5 was washed with water and incubated for 30 min at 30°C in 100 μl of 100 mM lithium acetate in TE (10 mM Tris [pH 7.5], 1 mM EDTA). One to 5 μg of transforming DNA and 40 μg of sheared calf thymus carrier DNA were added, and the incubation was continued for 1 h. The cell-DNA mixture was added to 700 μl of 40% polyethylene glycol 8000 in TE and, after a vortex mixing, was incubated for 30 min at 30°C. The cells were then heat shocked at 42°C for 5 min, washed three times in TE, and plated onto selective medium.

Isolation, cloning, and disruption of the *REF2* gene. RR3 was transformed with a yeast genomic DNA library (average insert size, ~15 kb) cloned into YCp50 (34). Ura^+ transformants (3,000 to 4,000) were replica plated onto both histidinol and X-Gal plates, and three white, Hol^- revertants were identified. Analysis of the recovered plasmids, which were retested for complementing ability, revealed two identical isolates which overlapped the third clone designated as YCp10.1. A 4-kb *Xba*I-*Pvu*II fragment from YCp10.1 was cloned into pGEM3Zf(+) (Promega) to create pRF2.3Z, the parental plasmid used for sequencing of the *REF2* open reading frame. A *Pvu*II (232)-*Hin*dIII (883) subclone, pRF2.P/H, was used to generate strand-specific RNA probes for Southern and Northern (RNA) analysis (see below). In the construction of YCpRF2, the 2.6-kb *Xba*I-*Bgl*II region was excised from pRF2.3Z as a *Pst*I (polylinker)-*Bgl*II fragment and cloned into YCplac33 (11) digested with the same enzymes. To make pRF2LEU2, a 3.0-kb *Bgl*II fragment harboring the entire *LEU2* gene was purified from CV13, end repaired with Klenow enzyme, and cloned into pRF2.3Z digested with *Hpa*I. To create *REF2*-inactivated lines, transformation of the appropriate yeast strain was carried out with *Pst*I-*Bgl*II-digested pRF2LEU2, and Leu^+ colonies were isolated. Proper integration was confirmed by Southern analysis of genomic DNA with a probe complementary to the remaining *REF2* sequences and transcribed from pRF2.P/H as described above. The transformed diploid strain SC472 was further sporulated and dissected to obtain strains RR6-4A and RR6-4D, which arose from the same ascospore.

Sequencing of the *ref2-1* allele. Genomic DNA was purified from strains DC35-15D and DR5-1D by the protocol of Philippsen et al. (30). Amplification by PCR was carried out with oligonucleotides ref2.4 (CACTGTGTCGACG GATCCATATGTCAGCACCCTGTTC), which straddles the putative initiation codon (underlined), and ref2.1, which is complementary to the sequence 1254-CGAGGATATGTCAATC-1269 situated 18 bp upstream of the wild-type TGA stop codon. The 1.3-kb reaction products were gel purified, repaired with T4 DNA polymerase, and cloned into *Sma*I-digested pGEM3Zf(+). Multiple clones generated from either strain were sequenced to control for *Taq* DNA polymerase-induced mutations. In addition to the frameshift mutation identified in the *ref2-1* allele (see Fig. 3), these strains contain a silent C-to-A change at position 840 which probably represents a polymorphism from the strain used to generate the genomic DNA library used in the complementation analysis.

RNA procedures. Total RNA was extracted from 50 ml of cells (optical density at 600 nm, 1) by the glass bead disruption method of Elder et al. (10). $Poly(A)^+$ RNA fractionation and analysis by Northern blot hybridization were carried out as described by Russnak and Ganem (36). Washing of membranes was carried out for 20 min at room temperature, for 20 min at 65°C in 1× SSPE (1× SSPE is 0.18 M NaCl, 10 mM $NaPO_4$, and 1 mM EDTA [pH 7.7])–0.1% sodium dodecyl sulfate (SDS), and then twice for 20 min (each) at 65°C in 0.1× SSPE–0.1% SDS. For quantification, exposed films were analyzed in an LKB Ultrascan XL scanning microdensitometer. All hybridization probes were ^{32}P-labeled antisense RNAs synthesized by SP6 or T7 RNA polymerase on linearized pGEM-derived plasmids. Plasmid SB7 (a gift from Steve Baker) contains a 644-bp *Bam*HI-*Sal*I sequence from pYAH83F which includes 489 bp of the *ACT1* promoter and initial coding sequence, 66 bp of the *ACT1* intron, and the 83-bp *CYC1* poly(A) site cloned into pGEM1 (Promega). Plasmids lacZ.3Z, H2B2.3Z, CBP1.3Z, and MCM1.3Z harbor the following sequences, respectively, in pGEM3Zf(+): a 624-bp *Hpa*I fragment (sequence positions 1725 to 2349) derived from the *lacZ* portion of pSB17 (described above), a 728-bp *Hin*dIII-*Bgl*II region containing the 3′ end of the *H2B2* gene (16) (DNA kindly provided by Mike Briggs), a 388-bp *CBP1 Hin*dIII-*Nsi*I segment (sequence positions 803 to 1191) purified from a derivative of clone p1-3.5 described by Mayer and Dieckmann (26), and a 900-bp *MCM1 Eco*RI (558)-*Bam*HI region taken from pXB3.4 (28), a gift from Randy Elbe.

Expression and purification of bacterial Ref2pF. Unique *Sal*I, *Bam*HI, and *Nde*I restriction sites were introduced immediately upstream of the *REF2* start codon by PCR amplification of the gene within pRF2.3Z with the oligonucleotides ref2.4 and ref2.1 described above. A *Sal*I-*Xba*I (261) subfragment was reintroduced into a *Sal*I (polylinker)-*Xba*I-digested derivative of pRF2.3Z in which the *Xba*I site within the polylinker was destroyed by partial digestion, Klenow repair, and religation. From the resulting plasmid, pRF2.3Z(Nde), an *Nde*I-*Bgl*II (1858) fragment, harboring the *REF2* coding region and 3′ untranslated region (UTR), was cloned into *Nde*I-*Bam*HI-digested pFLAG(S)-7 (7) to create pFLAG.RF2. For pTrcFRF2, an *Nco*I-*Eco*RI (1434) fragment from pFLAG.RF2 was cloned into pTrcHisC (Invitrogen Corp.) digested with the same enzymes.

Affinity purification of Ref2pF was carried out by the method of Chiang and Roeder (7). A 100-ml culture of DH5α cells harboring pTrcFRF2 was grown to an optical density at 600 nm of 1 in Luria-Bertani medium at 37°C and induced for 1 h with 500 μM isopropyl-β-D-thiogalactopyranoside (IPTG). Cells were

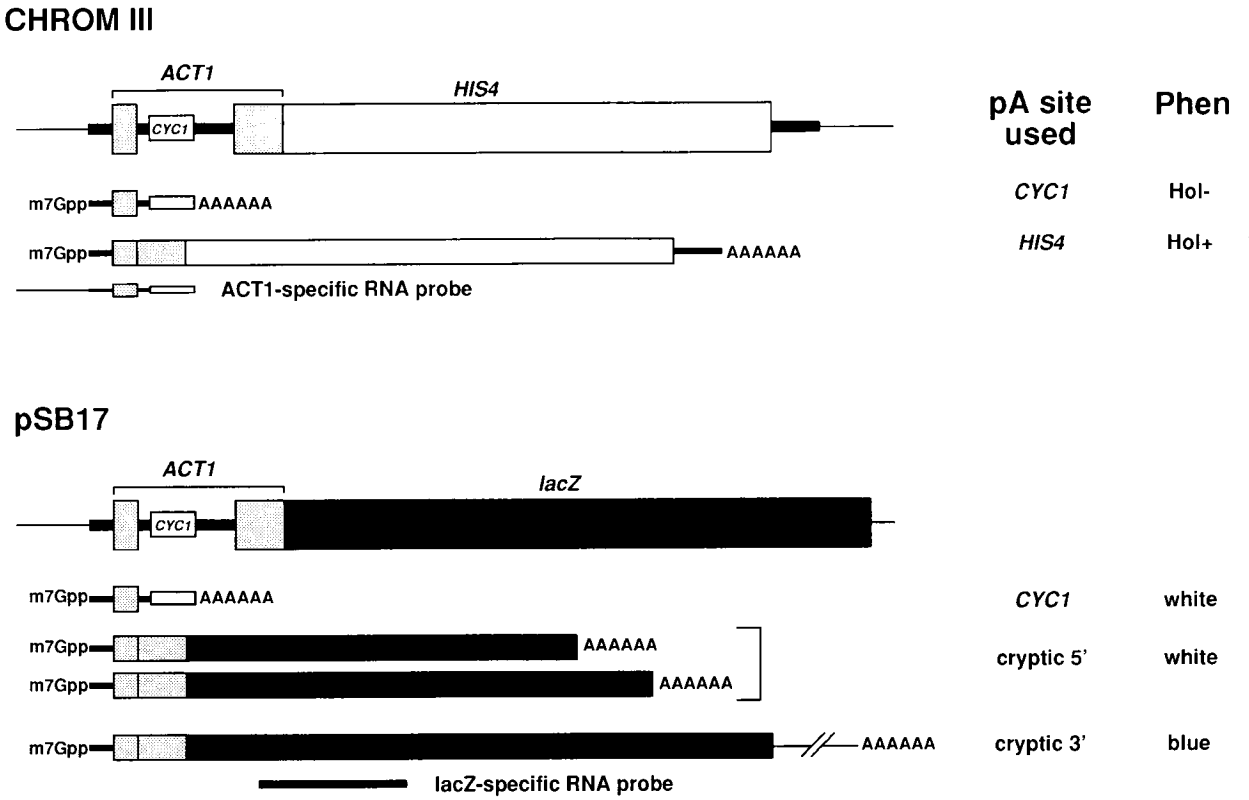

FIG. 1. Diagram of polyadenylated RNAs generated from the *ACT1-CYC1-HIS4* and *ACT1-CYC1-lacZ* tandem poly(A) site reporter genes. Strains harboring the *ACT1-CYC1-HIS4* construct resulted from integration of pYAH83F into the *his4-401* allele on chromosome (CHROM) III (35). The *ACT1-CYC1-lacZ* fusion was carried on a CEN-based vector and maintained at a low copy number. The indicated antisense RNA probes used in the Northern analysis shown in Fig. 2A were generated from pSB7 and placZ.3Z (see Materials and Methods). Listed to the right are the poly(A) (pA) sites used to generate the termini as well as the expected phenotypes (Phen) of cells accumulating that RNA type. The Hol designation refers to growth on histidinol medium, while color development at the bottom right refers to that observed on X-Gal plates (see text).

harvested and lysed at 4°C by sonication in 3 ml of a solution of 20 mM Tris (pH 7.9), 0.2 mM EDTA, 0.5 M NaCl, 10 mM β-mercaptoethanol, 0.2 mM phenylmethylsulfonyl fluoride, 1 mg of pepstatin per ml, 1 mg of leupeptin per ml, and 20% glycerol. One half of the lysate was incubated for 4 h at 4°C with 50 μl of anti-FLAG M2 agarose (IBI). The resin was washed 10 times in 10 ml of a solution of 20 mM Tris (pH 7.9), 0.2 mM EDTA, 300 mM KCl, 1 mM dithiothreitol, 0.5 mM phenymethylsulfonyl fluoride, and 20% glycerol and transferred to a microcentrifuge tube. Bound fusion protein was eluted two times with 200 μl of the FLAG peptide (0.2 mg/ml; IBI) in the same buffer at 4°C for 20 min, and supernatants were pooled and dialyzed against a solution of 10 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES; pH 7.0), 0.2 mM EDTA, 50 mM potassium acetate and 50% glycerol. We estimate from SDS-polyacrylamide gel electrophoresis (PAGE) and silver staining that yields were typically 50 μg and that the FLAG-Ref2p preparations were approximately 90% pure.

3′-end processing in vitro and RNA binding. Yeast processing extracts were prepared from strains RR6-4A and RR6-4D as described by Butler et al. (5), except that the yeast cells were grown to an optical density at 600 nm of 7.0 and the spheroplasts were equilibrated in 0.2 M KCl buffer prior to lysis. Processing reactions were carried out at 30°C for the indicated times in a total volume of 10 μl which included 1 nM ^{32}P-labeled RNA, 10 mM HEPES (pH 7.0), 1 mM magnesium acetate, 75 mM potassium acetate, 10 mM dithiothreitol, 2% polyethylene glycol, 2 mM ATP, and 2 μl of extract. When indicated, 2 mM CTP was substituted for ATP or reaction mixtures were supplemented with 1 μl (~10 ng or 20 nM) of purified Ref2pF. Reactions were stopped by the addition of 1 μl of a solution of 0.14 M EDTA, 2.5% SDS, and 2 mg of proteinase K per ml and then incubation at 37°C for 15 min. The reaction products were extracted, precipitated, and separated on 4% denaturing acrylamide gels.

For the bandshift experiments, various amounts of Ref2pF were incubated with 1 nM radioactive RNA in 10 μl of a solution of 10 mM HEPES (pH 7.0), 0.2 mM EDTA, 50 mM potassium acetate, and 2 mg of heparin per ml for 5 min on ice. Reaction mixtures were electrophoresed directly on a 2% agarose gel in 0.5× Tris-borate-EDTA, dried onto DE81 paper, and visualized by autoradiography. For the competition experiments, nucleic acids were mixed on ice and added to a reaction mixture containing 2 μl of Ref2pF. RNA homopolymers (average length, 500 to 1,000 nucleotides [nt]; Boehringer Mannheim) were used at a final nucleotide concentration of 5 or 20 μM as indicated.

Construction of the templates used to synthesize the Ty RNA substrates U3RU5, H4RU5, and ΔTS1/2 has been described elsewhere in detail (17). Wild-type *CYC1* and *cyc1-512* RNAs were transcribed from pGYC1 and pG200R, respectively, which are described by Butler and Platt (4); T/H RNA was transcribed from pG2T/H (43), while the *GAL7* RNA was transcribed from a derivative of pSB28 (5) and included sequences 70 nt upstream and 120 nt downstream of the poly(A) site.

RESULTS

Identification of *trans*-acting mutants defective in poly(A) site use. Ruohola et al. (35) described a genetic selection designed to identify spontaneous, nonlethal mutations in genes involved in mRNA 3′-end formation. In their scheme, normal yeast cells carrying an integrated *ACT1-HIS4* fusion gene with a *CYC1* cleavage and polyadenylation site inserted within the intron of the *ACT1* portion failed to grow on the histidine precursor, histidinol, i.e., these cells were phenotypically Hol$^-$. This was attributed to efficient use of the integrated *CYC1* poly(A) site and resulted in a high proportion of truncated RNAs [260 nt excluding the poly(A) tail]. This is schematically depicted in Fig. 1, and an example is shown in Fig. 2A (lane 1). A Hol$^+$ revertant, *tef2-1* (renamed here as *ref2-1*), was isolated and shown to contain a recessive, and therefore likely *trans*-acting, mutation. The nature of the defect, given that a Hol$^+$ phenotype could arise from several types of mutations, was not determined at that time.

We have now demonstrated in Fig. 2A (lane 2) that in *ref2-1*

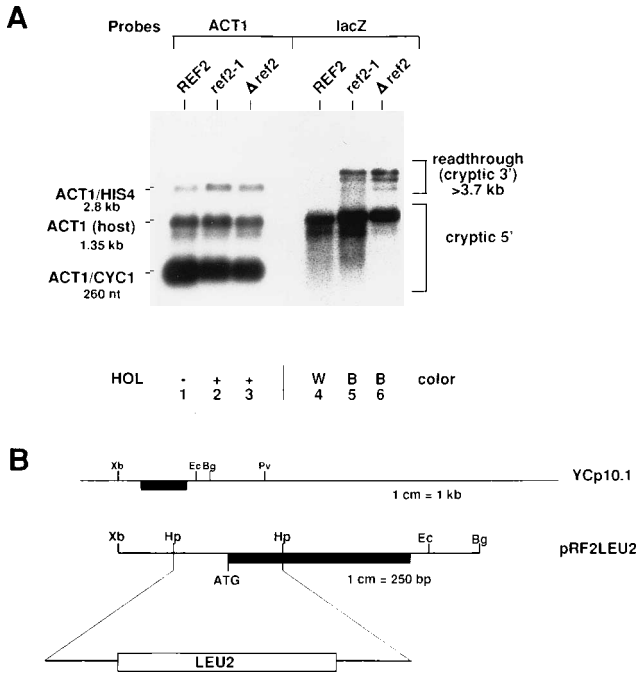

FIG. 2. Decreased use of tester poly(A) sites in *ref2-1* and Δ*ref2* cells. (A) Approximately 1 μg of poly(A)$^+$ RNA isolated from the indicated strains was analyzed by Northern blot hybridization. All strains contain the integrated *ACT1-CYC1-HIS4* gene. RNAs arising from this locus were detected with an *ACT1*-specific probe as shown in lanes 1, 2, and 3. This probe did not cross-hybridize to endogenous *CYC1* mRNA, which comigrates with the truncated RNA species in these gels (data not shown). The strains shown in lanes 4, 5, and 6 harbor, in addition, pSB17 from which the observed *lacZ*-specific RNAs were generated. The various RNA species and the probes used for their detection are schematized in Figure 1. (B) Strategy for the creation of the Δ*ref2* strains used in this study. Cells were transformed with the shown segment of pRF2LEU2 (see Materials and Methods for construction) in which a 0.7-kb *Hpa*I region was replaced with a functional *LEU2* gene. Also shown is a partial restriction map of the yeast DNA insert within the original complementing plasmid YCp10.1. The black bar denotes the *REF2* coding region.

cells, increased histidinol dehydrogenase activity correlated with a twofold elevation in steady-state levels of full-length, properly spliced and polyadenylated *ACT1-HIS4* RNA and that levels of the truncated RNA were concomitantly reduced by 50%. The probe used in this experiment cross-hybridized to host cell *ACT1* RNA, which provided a convenient internal control to standardize the amount of poly(A)$^+$ RNA applied to the gel. The observed shift in the transcriptional pattern was consistent with a partial defect in 3′-end formation at the *CYC1* poly(A) site embedded within the *ACT1* intron but not at the downstream *HIS4* or the endogenous *ACT1* poly(A) sites. RNase mapping of the truncated transcript verified that the same site of polyadenylation was used in *ref2-1* as in the parental strain, indicating that the specificity of the endonucleolytic cleavage reaction was not altered (data not shown). From our analysis of many *trans*-acting Hol$^+$ mutants, including *tef1-3* from Ruohola et al. (35), we also observe in the majority of cases a distinct molecular phenotype whereby levels of both the truncated and the extended RNAs are elevated relative to the *ACT1* control message. We have not characterized these potentially interesting mutants further, since they are incompatible with a defect in 3′-end processing.

As a further test of the *trans*-acting nature of the *ref2-1* mutation, an *ACT1-lacZ* fusion gene containing the same inserted 83-bp *CYC1* poly(A) site and maintained on a low-copy-number plasmid (pSB17 [Fig. 1]) was introduced into both wild-type and mutant cells. Poly(A)$^+$ RNA was fractionated from the resulting transformants, and equal amounts, as determined by ethidium bromide staining, were hybridized to a probe complementary to *lacZ* (Fig. 2A). This analysis revealed a variety of transcripts which, on the basis of length and on the assumption that they are 5′ coterminal, must be generated by cleavage and polyadenylation at cryptic sites within the *lacZ* coding region (designated cryptic 5′ sites [Fig. 1 and 2]). Small RNAs processed at the *CYC1* insert cannot be detected with this downstream probe. As shown in Fig. 2A (lane 5), longer, extended RNAs became more apparent in *ref2-1* cells and were interpreted to represent the cumulative readthrough of both the cryptic poly(A) sites within *lacZ* and the upstream *CYC1* poly(A) site. These extended RNAs were presumably processed at cryptic sites further downstream and are competent for translation of functional β-galactosidase since *ref2-1* cells harboring pSB17 became dark blue, whereas *REF2* cells were pale blue on X-Gal indicator plates. The defect in *ref2-1* apparently influences directly the use of the cryptic poly(A) sites found within *lacZ*, since differential color development and accumulation of extended RNAs was observed also between wild-type and *ref2-1* cells transformed with a reporter gene without an inserted *CYC1* poly(A) site (data not shown). This also ruled out the formal possibility that *ref2-1* was an up-regulatory mutation whereby splicing of the *ACT1* intron predominated over the use of the competing poly(A) site within it.

Isolation of the *REF2* gene. The fortuitous observation that *ref2-1* and wild-type cells transformed with *lacZ* indicator plasmids can be distinguished colorimetrically provided us with a double screen for the complementation analysis. Thus, a *ref2-1*/pSB17 strain (Hol$^+$, blue) was transformed with a low-copy-number yeast genomic DNA library (34), and from among the Ura$^+$ transformants we searched for colonies which remained relatively white on X-Gal and could no longer grow on histidinol. Three were identified, and the DNA inserts from the recovered complementing plasmids were found to be identical in two cases and to overlap the third, YCp10.1, which is shown in Fig. 2B. A complementing 2.6-kb *Xba*I-*Bgl*II genomic DNA fragment was sequenced, revealing the open reading frame shown in Fig. 3. To affirm that this region is actively transcribed, Northern hybridization analysis with an antisense RNA derived from the *Pvu*II site at 232 to the *Hin*dIII site at 883 of the putative *REF2* coding sequence identified a single polyadenylated mRNA species of approximately 2 kb (data not shown).

Evidence that we had cloned the authentic *REF2* gene and not an extragenic suppressor was provided by the identification of a frameshift mutation in the original *ref2-1* allele (see Materials and Methods) (Fig. 3). We note also that disruption of the *REF2* gene (see below) in a strain harboring the *ACT1-CYC1-HIS4* or *ACT1-CYC1-lacZ* reporter genes resulted in altered RNA profiles almost identical to those observed in the *ref2-1* isolate (Fig. 2A, lanes 3 and 6 versus lanes 2 and 5) and that an extensive search for high-copy-number suppressors of *ref2-1* resulted in the isolation only of clones which overlap those described above.

REF2 codes for a putative protein of 429 amino acids (aa) with a predicted molecular weight of 47,792. If unmodified, it is an extremely basic protein (pI, approximately 10) with high content levels of lysine (14.4%) and serine (12.6%). Ref2p does not show any significant homology to proteins in the databases, nor does it share any distinguishable motifs as defined elsewhere (1a). This protein does, however, possess some unusual features. One is a hyper alternating charge run (KEK sequence) from residues 208 to 220 composed almost entirely of lysine and glutamic acid conspicuously arranged into alter-

```
-119  gagtaagtgaggttaaagctcaaaaacctacaagcaaaggcacaaggaaatcgaaaggac
 -59  caagaaaggaaaagaaaagaaacgatacgataagtaaagacactgtggaagaattgaaca

   1  MetSerAlaProValProGlnLeuValAsnIleSerHisAlaLeuGlnAlaSerThrIle
   1  ATGTCAGCACCTGTTCCTCAATTAGTAAATATTTCTCATGCCTTGCAAGCTTCCACTATC

  21  GlnGlnIleArgLeuAspMetValAspPheAsnLysAspCysLysLeuSerSerIleGln
  61  CAGCAAATTCGTTTGGATATGGTAGATTTTAATAAAGATTGCAAATTATCTTCCATTCAA

  41  LeuAlaArgIleAspLysTyrIleAspSerLeuGlnAlaAlaLeuAsnGlnPheThrLys
 121  CTGGCAAGAATTGACAAGTATATCGATTCTTTGCAAGCGGCTTTGAATCAGTTTACCAAA

  61  AspAsnLeuHisIleGluArgLysGluLysAsnValThrGluAlaAspIleGlnLeuTyr
 181  GATAACTTGCACATAGAACGGAAGGAAAAGAACGTGACTGAAGCAGATATACAGCTGTAT

  81  SerGlyLeuLysSerMetTyrLeuAspTyrLeuAsnGlnLeuIleLysLeuLysHisGlu
 241  TCTGGCTTAAAATCAATGTATCTAGATTATTTGAACCAGCTAATTAAACTAAAGCATGAG

 101  LysGlnHisHisSerThrProProIleAlaAsnAspValSerLeuAspPhePheValAsn
 301  AAACAACACCACTCTACACCGCCCATCGCTAATGATGTCTCCCTGGACTTCTTCGTTAAC

 121  GlnLeuProLysPheSerProGluGluArgLysAsnTyrIleAspAsnLeuIleLeuAsn
 361  CAATTGCCTAAGTTTTCTCCCGAGGAAAGGAAAAATTACATTGACAATTTAATCCTGAAT

 141  LysAsnSerHisAsnArgLeuSerLysMetAspGlyLeuValAspAlaValIleAsnLeu
 421  AAAAATAGCCATAATCGATTATCTAAAATGGATGGTCTGGTAGATGCAGTCATCAATTTA

 161  CysValLeuAspThrSerValAlaGluAsnValArgSerTyrMetLysLeuLeuAspThr
 481  TGCGTCTTGGATACTTCCGTTGCTGAAAATGTACGATCCTATATGAAATTATTAGATACA

 181  LeuGlyPheGlnLysGlySerAsnSerThrGlyThrLysAlaAsnLeuLysLysLysLeu
 541  TTGGGTTTCCAAAAGGGTTCGAATAGTACTGGTACAAAGGCCAACCTCAAAAAGAAATTG

 201  AlaSerSerLysAlaLysIleLysAspSerGluLysGluLysGluLysGluLysAspLys
 601  GCTAGTTCGAAAGCAAAGATAAAGGATTCAGAAAAAGAAAAGGAGAAGGAGAAGGATAAA

 221  SerLysValLysMetLysThrLysLeuLysProSerProLeuLeuAsnAsnAspAspLys
 661  TCAAAAGTCAAGATGAAAACTAAATTAAAACCTTCTCCTTTGCTCAATAACGATGACAAA

 241  AsnSerSerProSerProThrAlaSerThrSerSerMetLysLysLeuLysSerGlyLeu
 721  AATTCTTCTCCATCGCCTACTGCATCTACCTCTTCAATGAAGAAATTGAAGTCGGGTTTA

 261  PheAsnLysAsnGluAlaLysSerThrGluSerLeuProThrSerSerLysLysLysLeu
 781  TTCAATAAAAATGAAGCTAAGTCTACAGAATCTCTACCTACTTCTTCCAAGAAAAAACTA

 281  SerPheSerLysTyrLeuAsnLysAspAspAlaAspMetThrLysLeuGlyThrLysArg
 841  TCATTTTCTAAATATCTGAACAAGGATGACGCAGATATGACCAAGCTTGGGACTAAACGG

 301  SerIleAspValAspPheLysValAsnProGluAlaSerThrValAlaSerAsnIleIle
 901  TCAATAGATGTGGATTTCAAAGTCAACCCCGAAGCATCCACGGTGGCTTCTAATATCATA

 321  SerSerSerThrSerGlySerSerThrThrThrValAlaThrProAlaSerSerGluGlu
 961  TCTTCGTCAACGTCAGGATCGTCAACCACAACGGTAGCGACTCCTGCTTCTTCAGAAGAG

 341  ProLeuLysLysLysThrLysIleSerValGlnAspSerAsnValGlnSerIleLeuArg
1021  CCCTTAAAAAAAAAAACCAAAATATCCGTGCAAGACTCTAATGTACAATCGATTTTGAGA
                    ProLysTyrProCysLysThrLeuMetTyrAsnArgPhe *

 361  AsnGlyLysProLysLysAlaArgIleSerSerIleLysPheLeuAspAspSerGlnLeu
1081  AATGGTAAACCGAAAAAAGCACGCATAAGTAGCATCAAATTTTTGGATGATTCCCAACTA

 381  IleLysValTyrGlyAspAspLeuProAsnGlnGlyLeuGlnValSerProThrGlnLeu
1141  ATAAAAGTTTACGGTGACGATCTACCGAACCAAGGGCTACAAGTTTCTCCTACTCAATTG

 401  LysLysIleLeuLysProPheLysGluGlyGluProLysGluIleIleLeuPheGluAsp
1201  AAAAAAATTCTGAAACCATTCAAGGAGGGGGAACCGAAGGAAATTATATTGTTCGAGGAT

 421  MetSerIleLysLeuAsnLeuLeuIle *
1261  ATGTCAATCAAATTAAACCTCTTGATTTGAtgtttctgaagaacacaaacagtgatgact

1321  atatggatatatccgagactaaaggtggcccaatacattgtgaaacaagaccccgttgat
1381  ctatagaaaaaatttcaatcatttcaacccggacttgaataaaaggccgccaagaattct
1441  taaatggaaatacgaactcaaccccgactatagcaaaggctttcggtaaaaatagtttat
1501  tactaaggaaggacagaggtggtttgccatacaagcatgtccccatagtaaaaagaaata
1561  aatatcctccaagaccagtacactaagctatgacatgctgatacataaacatgtagtata
```

FIG. 3. Nucleotide and predicted protein sequences of *REF2*. Sequencing of the *ref2-1* allele (see Materials and Methods for details) revealed a missing A residue within the A stretch beginning at nt 1026. This frameshift results in the shown truncated protein of 358 aa compared with a full-length of 429 aa. Asterisks denote nonsense codons. Underlined is the 27-aa stretch from 204 to 230 in which lysine occupies every second position (a serine at position 210 is the lone exception). Embedded within this segment (double underline) lies the 13-aa hyperalternating charge run or KEK sequence (see text for discussion).

nating positive and negative charges. Also noteworthy is a stretch from aa 321 to 331 in which 10 of 11 residues are serine or threonine, and thus potential targets for phosphorylation. No statement regarding the function of these regions can be made yet, since both are found within the truncated protein encoded by the *ref2-1* loss-of-function allele.

mRNA metabolism in *REF2*-inactivated cells. To determine whether *REF2* was an essential gene, a wild-type homozygous diploid was transformed with the knockout construct pRF2LEU2, which effectively replaces the promoter and N-terminal portion of the coding region with a functional *LEU2* gene, as shown in Fig.

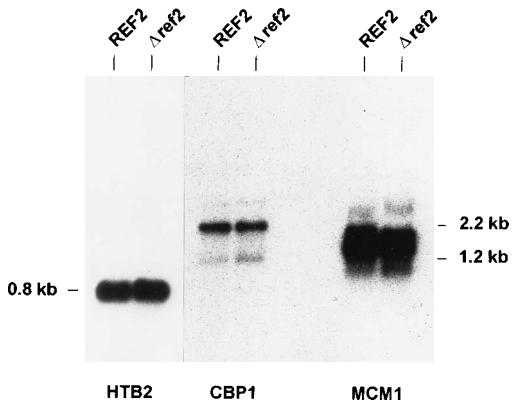

FIG. 4. Analysis of steady-state levels of normal cellular mRNAs in wild-type and Δ*ref2* yeast cells. Equal amounts of poly(A)$^+$ RNA, as determined by ethidium bromide staining after electrophoresis, were hybridized with the specific RNA probes indicated (see Materials and Methods for exact sequences). Strains used in this analysis were RR6-4A and RR6-4D, which arose from the same ascospore after sporulation of a *REF2*/*ref2*::*LEU2* diploid.

2B. After sporulation and tetrad dissection, a 2:2 segregation of large, Leu$^-$ and small, Leu$^+$ colonies was observed. This small colony size was attributed to an approximate 1.5-fold reduction in the vegetative growth rate of haploid strains carrying the disrupted *ref2*::*LEU2* allele, as measured in liquid culture, and not from a defect in spore development. This impaired growth, which had also been observed for the *ref2-1* strain, was not further exacerbated at either low or high temperatures. Thus, *REF2* is essential for normal growth but not for viability. In addition to an extended doubling time, Δ*ref2* cells display abnormal morphologies and multiple budding.

The artificial construct used in the genetic selection contained *CYC1* and *HIS4* poly(A) sites situated in tandem (Fig. 1A). As stated above, our ability to detect and accurately quantify reductions in upstream *CYC1* poly(A) site function (Fig. 2A, lanes 1, 2, and 3) relied on the assumption that levels of host *ACT1* mRNA remain fixed in mutant cells. Similarly, the stability of extended RNAs, processed at the downstream *HIS4* site, had to remain constant in order to measure increased readthrough. Both situations demanded that 3′-end processing not be compromised at either poly(A) site in Δ*ref2* cells. We decided, therefore, to examine the steady-state levels of other cellular mRNAs in wild-type versus Δ*ref2* cells and found, as shown in Fig. 4, that there were no quantitative or qualitative differences in the mRNAs transcribed from the *HTB2*, *CBP1*, and *MCM1* genes.

Ref2p participates in the 3′-end cleavage reaction. In an attempt to determine whether Ref2p was directly involved in the biochemistry of poly(A) site use and not some other cellular process influencing 3′-end formation, we carried out cleavage and polyadenylation reactions in vitro with 40% ammonium sulfate fractions (see Materials and Methods) prepared from strains RR6-4A (wild type) and RR6-4D (*ref2*::*LEU2*). In Fig. 5, we tested three distinct pre-mRNAs, all of which contain the poly(A) site of the yeast retrotransposable element Ty and are of equal length (250 nt upstream and 150 nt downstream of the pA site). U3RU5 contains wild-type long terminal repeat sequences, while substrates H4RU5 and ΔTS1/2 are deficient in one or both, respectively, of the two specific signals (designated TS1 and TS2) which are required for maximal Ty poly(A) site use. Our choice was based on previous studies (17) which demonstrated that these three substrates are processed with differing kinetics in vitro.

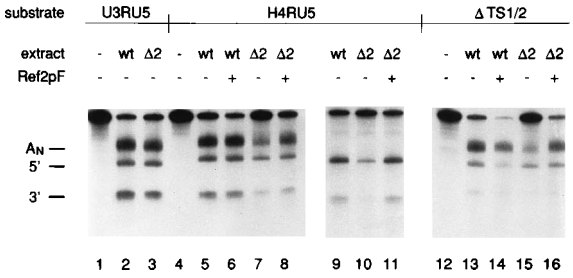

FIG. 5. Processing in vitro of yeast Ty-derived poly(A) sites with extracts prepared from wild-type or Δ*ref2* cells and complementation with recombinant Ref2pF. The pre-mRNAs U3RU5 (lanes 1 to 3), H4RU5 (lanes 4 to 11), and ΔTS1/TS2 (lanes 12 to 16) were incubated as described in Materials and Methods. Reaction mixtures contained either ATP (lanes 1 to 8 and 12 to 16) or CTP (lanes 9 to 11) and were run with yeast whole-cell extracts prepared from strains RR6-4A (lanes 2, 5, 6, 9, 13, and 14) and RR6-4D (lanes 3, 7, 8, 10, 11, 15, and 16) or with extract buffer alone (lanes 1, 4, and 12) and then electrophoresed on 4% denaturing acrylamide gels. The reaction mixtures shown in lanes 6, 8, 11, 14, and 16 also contain 1 μl of Ref2pF purified from *E. coli*. The reaction mixtures in lanes 1 to 11 were run for 5 min, while those in lanes 12 to 16 proceeded for 30 min.

As shown in Fig. 5, lanes 2 and 3, the wild-type U3RU5 precursor is processed efficiently and accurately in both normal and Δ*ref2* extracts. Similar results were observed with a wild-type *CYC1* substrate (data not shown). In contrast, processing of the mutated counterparts H4RU5 and ΔTS1/2 was dramatically reduced in extracts from Δ*ref2* cells compared with that in the wild type (Fig. 5, lanes 5 and 7 and lanes 13 and 15). We point out that in the reaction results shown in Fig. 5, lanes 2 and 5, the H4RU5 substrate appears to have been processed as efficiently as U3RU5 in wild-type extracts. In fact, the more-detailed kinetic analysis shown in Fig. 6A indicated that by 5 min the accumulation of processed H4RU5 RNA was approximately 80% that of U3RU5. Still, this does not agree with our earlier reports suggesting that H4RU5 processing was decreased twofold (17), and we attribute the discrepancy to strain variability or to the use of extracts which were, in this case, more robust.

The observed decrease in the processing in Ref2p-depleted extracts was attributed to a defect at the endonucleolytic cleavage step, as judged by reactions in which CTP had been substituted for ATP as an energy source (Fig. 5, compare lanes 9 and 10). Also, quantification of total processing reactions such as those in Fig. 5, lanes 5 and 7, showed that the proportion of cleaved RNA which further became polyadenylated remained constant in Δ*ref2* extracts, an indication that the coupling of cleavage and polyadenylation was not affected.

Differential cleavage in vitro was due not to a general decrease in the concentration of processing factors in slowly growing Δ*ref2* cells but rather to the absence of Ref2p, since in all cases (Fig. 5, lanes 8, 11, and 16) processing was reconstituted fully by the addition of ~10 ng of recombinant epitope-tagged protein, purified from bacteria by anti-FLAG chromatography (see Materials and Methods for details). This stimulation was not due to some indirect mechanism, since Ref2pF had little effect on H4RU5 and ΔTS1/2 processing in wild-type extracts (Fig. 5, lanes 6 and 14). For the reaction results shown in lane 14, although less total RNA was applied to the gel, the fraction which was processed remained equivalent to that shown in lane 13 (Fig. 5). We note also that Ref2pF functions in vivo since the altered gene, driven by a *GAL10* promoter and introduced on a low-copy-number plasmid (YCpGFRF2; see strain RR8 in Table 1), resulted in reversion of the Δ*ref2* strain phenotype from Hol^+ and blue on X-Gal to Hol^- and white (data not shown).

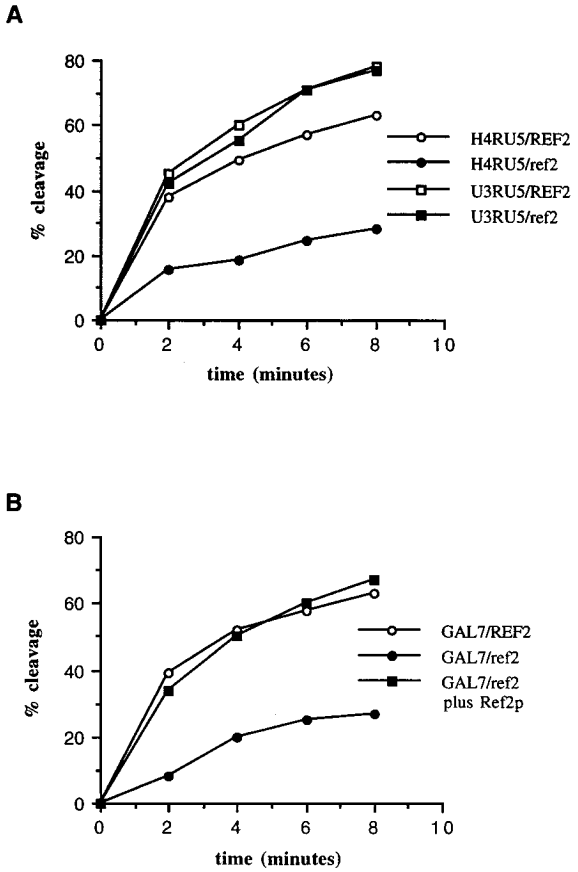

FIG. 6. Kinetic analysis of cleavage in vitro. Reaction mixtures containing CTP were run at 30°C in a total volume of 50 μl. At the indicated times, 10-μl aliquots were removed, and the reaction was terminated as described in Materials and Methods. The percentage of RNA cleaved was determined by PhosphorImager quantitation of the radiolabeled bands on gel electrophoresis and was normalized for specific activity. (A) Cleavage of the U3RU5 and H4RU5 Ty poly(A)(substrates shown in Fig. 5 and carried out in extracts prepared from either wild-type (REF2) or Δ*ref2* (ref2) strains. (B) As above but with a pre-mRNA containing the *GAL7* poly(A) site. Also shown is a time course of cleavage in which purified recombinant Ref2pF was added to the Δ*ref2* extract (solid squares).

Figure 6A shows a time course of RNA cleavage for the U3RU5 and H4RU5 substrates discussed above; Fig. 6B shows a time course for a *GAL7* poly(A) precursor which contains only 70 nt of specific sequences upstream of the site of cleavage and is, therefore, processed less efficiently than substrates which are 5′ extended. This analysis revealed that the observed decreases in accumulated levels of cleavage product in extracts prepared from Δ*ref2* cells correlated with a reduction in the reaction rate and, as shown for the *GAL7* RNA, that this rate was stimulated with the addition of purified Ref2pF.

RNA-binding properties of Ref2p. The finding that Ref2p was involved in an RNA-processing event prompted us to examine whether it associated with RNA in the absence of other processing factors. To this end we carried out the set of bandshift experiments whose results are shown in Fig. 7. In Fig. 7A (lanes 1 to 6), a more slowly migrating complex was observed to form on a 320-nt processing-competent *CYC1* substrate with the addition of increasing amounts of Ref2pF. As shown in Fig. 7B, Ref2pF also bound readily the following substrates: (i) the 400-nt Ty U3RU5, H4RU5, and ΔTS1/2 RNAs, which vary approximately 10-fold in their abilities to be

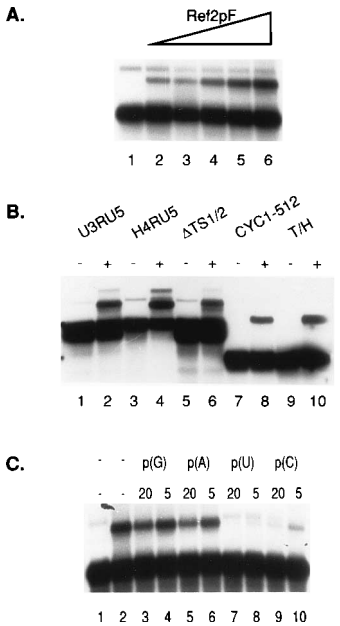

FIG. 7. Analysis of the RNA-binding properties of Ref2pF by gel shift. Uniformly labeled RNAs (at a 1 nM final concentration) were incubated with purified Ref2pF for 5 min at 4°C in buffer containing 2 mg of heparin per ml, and the reaction mixtures were electrophoresed on a 2% agarose gel in 0.5× TBE at 4°C. (A) Lanes 1 to 6, binding of a *CYC1* pre-mRNA with 0, 1, 2, 3, 4, or 5 μl, respectively, of Ref2pF. (B) The indicated RNAs were incubated with either Ref2pF storage buffer alone (lanes 1, 3, 5, 7, and 9) or 2 μl of Ref2pF (lanes 2, 4, 6, 8, and 10). (C) The indicated ribonucleotide homopolymers at nucleotide concentrations of 5 or 20 μM were mixed with labeled *CYC1* RNA prior to the addition of 2 μl of Ref2pF. Lane 1, migration of RNA in the absence of protein; lane 2, complex formation with no competitor.

cleaved and polyadenylated, (ii) a 280-nt *cyc1-512* mutant RNA harboring a 38-bp deletion which totally abolishes 3′-end processing in vitro (4), and (iii) a T/H RNA composed of bacterial and vector sequences (43). We conclude from these experiments that although Ref2pF participates in the processing reaction, it cannot by itself recognize the required specific signals. The observed binding was not simply due to the electrostatic attraction between a basic protein and the phosphate backbone of nucleic acid, since (i) RNA-protein complexes were stable in the presence of 2 mg of heparin per ml and (ii) the binding of Ref2pF to RNA was not inhibited with a 50-fold molar excess of either single-stranded or double-stranded DNA (data not shown). In fact, Ref2pF displayed a preferential affinity for pyrimidine bases, since binding to *CYC1* RNA was easily inhibited by U and C ribonucleotide homopolymers but not G or A (Fig. 7C).

DISCUSSION

Yeast genes that are involved in the synthesis and function of poly(A) tracts found at the 3′ end of mRNA, including those coding for poly(A) polymerase (*PAP1* [24, 29]), poly(A)-binding protein (*PAB1* [40]), and poly(A) nuclease (*PAN1* [41]), have been described. *REF2*, presented here, is the first yeast gene identified that directly influences the use of poly(A) sites prior to the addition of a poly(A) tail. This assertion is supported by several observations. (i) In the yeast mutant *ref2-1* or a strain carrying a disrupted *REF2* gene, 3′-end formation is diminished at inefficient, or otherwise compromised, poly(A) sites such as that of *CYC1* embedded within an intron and cryptic bacterial *lacZ* sequences. (ii) Decreased endonucleolytic cleavage is observed in vitro with mutated Ty pre-mRNA substrates in extracts devoid of Ref2p. (iii) An epitope-tagged variant of Ref2p, overexpressed and purified from *E. coli*, stimulates this reaction to wild-type levels. Together, these results strongly suggest that Ref2p is required for optimal utilization of poly(A) sites.

As a mechanism of action, we favor the idea that Ref2p participates in the formation of a precleavage complex as part of the ribonucleoprotein structure itself. In mammals, poly(A) signals program the ordered cooperative assembly of at least 10 polypeptides, whose stability determines the efficiency of poly(A) site use and is influenced by the underlying RNA sequences (1, 12, 14, 48). In *S. cerevisiae*, small, discrete AU-rich sequences, perhaps related to AAUAAA in higher eucaryotes, and often working together, determine the efficiency of poly(A) site use (17–19, 37, 38). With strong organizing signals, such as TS1 and TS2 in Ty, the Ref2p protein is dispensible for efficient precleavage complex formation in vitro. On the other hand, for substrates in which TS1 has been replaced by nonspecific sequences (H4RU5) or both TS1 and TS2 have been disrupted (ΔTS1/2), Ref2p plays an increasingly important role. Previous studies in this laboratory have already established that the processing machinery utilizes nonspecific RNA (17). Ref2p, whose RNA-binding specificity has not been well characterized, is a candidate for contributing to these interactions, acting as an anchor to help stabilize the precleavage complex.

A plausible alternative view is that Ref2p may be a cofactor required for the function of the processing endonuclease. Thus, in mutant cells, the proper assembly of proteins would result in the establishment of a productive complex for processing, but cleavage would subsequently proceed at a slower rate. If this were the case, decreased processing would be expected to occur on all substrates. For wild-type *CYC1* RNAs (data not shown) as well as wild-type Ty U3RU5 (Fig. 5, compare lanes 2 and 3), this was not observed. Furthermore, this model does not necessarily demand that Ref2p bind RNA.

Does Ref2p have a stimulatory effect on other nuclear RNA processing events such as splicing? It would not be unreasonable to surmise that a nonspecific RNA-binding protein may have a more general function, for example, participating in the proper presentation of nascent RNAs to both splicing and 3′-end processing factors. We feel that such an interpretation is unlikely, because of the nature of the genetic selection in which *ref2-1* was identified. If this were true, any delay in *ACT1* intron removal, cotranscriptionally, from readthrough RNAs which arise from similarly decreased kinetics of *CYC1* poly(A) use would offset each other. In other words, readthrough transcripts that are not spliced as efficiently in this tightly coupled system will lead to reduced *HIS4* expression and, by exposing the intronic poly(A) site longer, should restore higher levels of truncated RNA, probably rendering the cells Hol$^-$.

Chen and Moore (6) have described two factors of unknown complexity, CFI and CFII, which in combination are sufficient for accurate and efficient cleavage of *GAL7* and *CYC1* pre-mRNAs. CF1 is required also for polyadenylation on precleaved substrates and thus is functionally analogous to the large heteromer CPSF which specifically binds AAUAAA in metazoan cells (2, 13, 21). It will be of interest to determine whether Ref2p is an integral part of one of these factors or, alternatively, an RNA-binding protein which interacts with them. Accessory RNA-binding proteins which stimulate poly(A) site use in mammals have been identified. They include the U1A protein as part of the U1 small nuclear ribonucleoprotein (25) and DSEF-1 (31), which have been shown

to stimulate use of the simian virus 40 late polyadenylation site from upstream and downstream positions, respectively.

Interestingly, human U1A protein, in its free form, can also behave as a negative regulator. In an example of an autoregulatory loop mechanism, U1A binds its own pre-mRNA at sites located 19 and 45 nt upstream of the AUUAAA pA signal (3) and inhibits polyadenylation but not cleavage via a direct interaction with poly(A) polymerase (15). In an attempt to determine whether Ref2p associates with yeast Pap1p, we have failed to suppress *ref2-1* with *PAP1* on a high-copy-number plasmid (a gift from Scott Butler) or to demonstrate a direct interaction with Pap1p in vivo by a two-hybrid test (8) (data not shown).

Our observations with *REF2* are reminiscent of those described for the *MUD1* gene, which codes for the yeast version of the U1A protein (23). Like Ref2p, the absence of U1A protein has little effect on splicing both in vivo and in vitro unless a mutated RNA substrate is presented. The authors have suggested that the U1A protein aids in the folding or maintenance of U1 RNA in an active configuration. Similarly, we propose that Ref2p contributes to the stability, and therefore the efficiency, of 3′-end processing complexes formed at yeast poly(A) sites.

We have further shown that Ref2pF purified from *E. coli* is competent for RNA binding at a relatively high heparin concentration (2 mg/ml). The ability to bind RNA in vitro is certainly congruent with a role in complex formation on pA sites, but it remains to be established whether this binding mediates Ref2p function in vivo. Proof will require the identification of mutations which simultaneously disrupt RNA binding, cleavage in vitro, and the ability to complement a Δ*ref2* strain. These experiments are now under way and should, in addition, help to define a novel type of RNA-binding domain, given that Ref2p shows little resemblance to other known RNA-binding proteins.

Another predicted feature of Ref2p, given that it functions as part of the 3′-end cleavage machinery, would be to form contacts with other members within the complex. Using a two-hybrid screen, we have recently identified another novel yeast protein which associates with Ref2p in vivo. Because of the nature of the assay, this specific interaction is presumed to occur in the absence of RNA as a cofactor. Efforts to immunopurify Ref2pF-containing complexes which assemble on poly(A) substrates are also currently under way.

Finally, an intriguing question arises from this work. It appears that Ref2p is not required for efficient processing of wild-type substrates in vitro, and in yeast cells totally depleted of Ref2p, cytoplasmic mRNA pools for a number of tested genes are unaffected. Why, then, do Δ*ref2* cells have a growth defect? One possibility is that there exists a class of poly(A) sites, yet to be discovered, which is highly responsive to Ref2p levels. A more likely alternative is that a slight decrease in the kinetics of cleavage at all poly(A) sites disturbs a tightly coupled cellular event. It is known that both in *S. cerevisiae* (39) and mammalian (9) cells, 3′-end processing is critical for termination of RNA polymerase II elongation. Therefore, a subsequent delay in transcription termination, which may occur in *ref2* cells, could allow polymerase to infringe upon and interfere with the expression of closely spaced adjacent genes. Whatever the answer, the study of a mutant partially defective in 3′-end cleavage will provide valuable insights into the role of this process in the expression of eucaryotic genomes and underscores the importance of identifying further mutations in this RNA-processing event.

ACKNOWLEDGMENTS

We acknowledge the members of the Eric Phizicky laboratory for helpful advice, for plasmids pCV13 and YCplac33, and for strain SC472. Special thanks go to Scott Butler for the yeast library and to Sandy Consaul for carrying out the microdissection. We are also grateful to Scott Butler and Shalini Pereira for critical reviews of the manuscript.

This work was supported in part by Public Health Service grant 5-R01-GM35658 to T.P. and a National Cancer Institute of Canada Senior Fellowship to R.R.

REFERENCES

1. **Ahmed, Y. F., G. M. Gilmartin, S. M. Hanly, J. R. Nevins, and W. C. Greene.** 1991. The HTLV-I Rex response element mediates a novel form of mRNA polyadenylation. Cell **64:**727–737.

1a. **Bairoch, A.** PROSITE dictionary of protein sites and patterns. Genetics Computer Group, Inc., Madison, Wis.

2. **Bardwell, V. J., M. Wickens, S. Bienroth, W. Keller, B. S. Sproat, and A. I. Lamond.** 1991. Site-directed ribose methylation identifies 2′-OH groups in polyadenylation substrates critical for AAUAAA recognition and poly(A) addition. Cell **65:**125–133.

3. **Boelens, W. C., E. J. R. Jansen, W. J. van Venrooij, R. Stripecke, I. W. Mattaj, and S. I. Gunderson.** 1993. The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. Cell **72:**881–892.

4. **Butler, J. S., and T. Platt.** 1988. RNA processing generates the mature 3′ end of yeast *CYC1* mRNA *in vitro*. Science **242:**1270–1274.

5. **Butler, J. S., P. P. Sadhale, and T. Platt.** 1990. RNA processing in vitro produces mature 3′ ends of a variety of *Saccharomyces cerevisiae* mRNAs. Mol. Cell. Biol. **10:**2599–2605.

6. **Chen, J., and C. Moore.** 1992. Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA. Mol. Cell. Biol. **12:**3470–3481.

7. **Chiang, C.-M., and R. G. Roeder.** 1993. Expression and purification of general transcription factors by FLAG epitope-tagging and peptide elution. Pept. Res. **6:**62–64.

8. **Chien, C.-T., P. L. Bartel, R. Sternglanz, and S. Fields.** 1991. The two-hybrid system: a method to identify and clones genes for proteins that interact with a protein of interest. Proc. Natl. Acad. Sci. USA **88:**9578–9582.

9. **Edwalds-Gilbert, G., J. Prescott, and E. Falck-Pedersen.** 1993. 3′ RNA processing efficiency plays a primary role in generating termination-competent RNA polymerase II elongation complexes. Mol. Cell. Biol. **13:**3472–3480.

10. **Elder, R. T., E. Y. Loh, and R. W. Davis.** 1983. RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. Proc. Natl. Acad. Sci. USA **80:**2432–2436.

11. **Gietz, R. D., and A. Sugino.** 1988. New yeast-*Escherichia coli* shuttle vectors constructed with *in vitro* mutagenized yeast genes lacking six-base pair restriction sites. Gene **74:**527–534.

12. **Gilmartin, G. M., E. S. Fleming, and J. Oetjen.** 1992. Activation of HIV-1 pre-mRNA 3′ processing *in vitro* requires both an upstream element and TAR. EMBO J. **11:**4419–4428.

13. **Gilmartin, G. M., and J. R. Nevins.** 1989. An ordered pathway of assembly of components required for polyadenylation site recognition and processing. Genes Dev. **2:**578–587.

14. **Gilmartin, G. M., and J. R. Nevins.** 1991. Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-mRNA. Mol. Cell. Biol. **11:**2432–2438.

15. **Gunderson, S. I., K. Beyer, G. Martin, W. Keller, W. C. Boelens, and I. W. Mattaj.** 1994. The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. Cell **76:**531–541.

16. **Hereford, L., M. A. Osley, J. R. Ludwig II, and C. S. McLaughlin.** 1981. Cell-cycle control of yeast histone mRNA. Cell **24:**367–375.

17. **Hou, W., R. Russnak, and T. Platt.** 1994. Poly(A) site selection in the yeast Ty retroelement requires an upstream region and sequence-specific titratable factor(s) *in vitro*. EMBO J. **13:**446–452.

18. **Irniger, S., and G. H. Braus.** 1994. Saturation mutagenesis of a polyadenylation signal reveals a hexanucleotide element essential for mRNA 3′ end formation in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **91:**257–261.

19. **Irniger, S., C. M. Egli, and G. H. Braus.** 1991. Different classes of polyadenylation sites in the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. **11:**3060–3069.

20. **Ito, H., Y. Jukuda, K. Murata, and A. Kimura.** 1983. Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153:**163–168.

21. **Keller, W., S. Bienroth, K. M. Lang, and G. Christofori.** 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3′ processing signal AAUAAA. EMBO J. **10:**4241–4249.

22. **Larson, G. P., K. Itakura, H. Ito, and J. Rossi.** 1983. *Saccharomyces cerevisiae* actin-*Escherichia coli lacZ* gene fusions: synthetic-oligonucleotide-mediated deletion of the 309 base pair intervening sequences in the actin gene. Gene **22:**31–99.

23. **Liao, X. C., J. Tang, and M. Rosbash.** 1993. An enhancer screen identifies a gene that encodes the yeast U1 snRNP A protein: implications for snRNP protein function in pre-mRNA splicing. Genes Dev. **7:**419–428.
24. **Lingner, J., J. Kellerman, and W. Keller.** 1991. Cloning and expression of the essential gene for poly(A) polymerase from *S. cerevisiae*. Nature (London) **354:**496–498.
25. **Lutz, C. S., and J. C. Alwine.** 1994. Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal. Genes Dev. **8:**576–586.
26. **Mayer, S. A., and C. L. Dieckmann.** 1989. The yeast *CBP1* gene produces two differentially regulated transcripts by alternative 3′ end formation. Mol. Cell. Biol. **9:**4161–4169.
27. **Parker, R., and C. Guthrie.** 1985. A point mutation in the conserved hexanucleotide at a yeast 5′ splice junction uncouples recognition, cleavage, and ligation. Cell **41:**107–118.
28. **Passmore, S., G. T. Maine, R. Elbe, C. Christ, and B.-K. Tye.** 1988. *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of *MAT*α cells. J. Mol. Biol. **204:**593–606.
29. **Patel, D., and J. S. Butler.** 1992. Conditional defect in mRNA 3′ end processing caused by a mutation in the gene for poly(A) polymerase. Mol. Cell. Biol. **12:**3297–3304.
30. **Philippsen, P., A. Stotz, and C. Scherf.** 1991. DNA of *Saccharomyces cerevisiae*. Methods Enzymol. **194:**169–182.
31. **Qian, Z., and J. Wilusz.** 1991. An RNA-binding protein specifically interacts with a functionally important domain of the downstream element of the simian virus 40 late polyadenylation signal. Mol. Cell. Biol. **11:**5312–5320.
32. **Raabe, T., F. J. Bollum, and J. L. Manley.** 1991. Primary structure and expression of bovine poly(A) polymerase. Nature (London) **353:**229–234.
33. **Robinson, M. K., W. H. van Zyl, E. M. Phizicky, and J. R. Broach.** 1994. *TPD1* of *Saccharomyces cerevisiae* encodes a protein phosphatase 2C-like activity implicated in tRNA splicing and cell separation. Mol. Cell. Biol. **14:**3634–3645.
34. **Rose, M. D., P. Novick, J. H. Thomas, D. Botstein, and G. R. Fink.** 1987. A *Saccharomyces cerevisiae* genomic plasmid bank based on a centromere-containing shuttle vector. Gene **60:**237–243.
35. **Ruohola, H., S. M. Baker, R. Parker, and T. Platt.** 1988. Orientation-dependent function of a short *CYC1* DNA fragment in directing mRNA 3′ end formation in yeast. Proc. Natl. Acad. Sci. USA **85:**5041–5045.
36. **Russnak, R., and D. Ganem.** 1990. Sequences 5′ to the polyadenylation signal mediate differential poly(A) site use in hepatitis B virus. Genes Dev. **4:**764–776.
37. **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.
38. **Russo, P., W.-Z. Li, Z. Guo, and F. Sherman.** 1993. Signals that produce 3′ end termini in *CYC1* mRNA of the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. **13:**7836–7849.
39. **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.
40. **Sachs, A. B., and R. W. Davis.** 1989. The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. Cell **58:**857–867.
41. **Sachs, A. B., and J. A. Deardorff.** 1992. Translation initiation requires the PAB-dependent poly(A) ribonuclease in yeast. Cell **70:**961–973.
42. **Sherman, F., G. R. Fink, and C. Lawrence.** 1986. Methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
43. **Steinmetz, E. J., C. A. Brennan, and T. Platt.** 1990. A short intervening structure can block rho factor helicase action at a distance. J. Biol. Chem. **265:**18408–18413.
44. **Takagaki, Y., C. C. MacDonald, T. Shenk, and J. L. Manley.** 1992. The human 64-kd polyadenylation factor contains a ribonucleoprotein-type RNA binding domain and unusual auxiliary motifs. Proc. Natl. Acad. Sci. USA **89:**1403–1407.
45. **Takagaki, Y., and J. L. Manley.** 1992. A human polyadenylation factor is a G protein β-subunit homologue. J. Biol. Chem. **267:**23471–23474.
46. **Wahle, E., and W. Keller.** 1992. The biochemistry of 3′-end cleavage and polyadenylation of messenger RNA precursors. Annu. Rev. Biochem. **61:** 419–440.
47. **Wahle, E., G. Martin, E. Schiltz, and W. Keller.** 1991. Isolation and expression of cDNA clones encoding mammalian poly(A) polymerase. EMBO J. **10:**4251–4257.
48. **Weiss, E. A., G. M. Gilmartin, and J. R. Nevins.** 1991. Poly(A) site efficiency reflects the stability of complex formation involving the downstream element. EMBO J. **10:**215–219.