*Proc. Natl. Acad. Sci. USA*
Vol. 80, pp. 4403–4407, July 1983
Genetics

# Expression of a β-galactosidase gene containing the ribosomal protein 51 intron is sensitive to the *rna2* mutation of yeast

(yeast ribosomal protein genes/mRNA splicing/*lacZ* gene fusions)

JOHN L. TEEM AND MICHAEL ROSBASH

Department of Biology, Brandeis University, Waltham, Massachusetts 02254

*Communicated by Gerald R. Fink, March 14, 1983*

ABSTRACT **The temperature-sensitive mutation *rna2* causes the accumulation of higher molecular weight transcripts from the ribosomal protein 51 (rp51) gene of yeast and many other yeast ribosomal protein genes. We have determined the DNA sequence of the rp51 gene, confirming that it contains an intron and that the higher molecular weight transcript is an intron-containing precursor RNA. These data and other experiments suggest that the *rna2* mutation affects mRNA processing (splicing) and that the presence of an intron is sufficient to render expression of a gene sensitive to the *rna2* mutation. To test these hypotheses, we have inserted the rp51 intron into the coding region of a hybrid *Escherichia coli* β-galactosidase gene, thereby interrupting the open reading frame subsequent to the initiating methionine codon. Despite the presence of the intron, the β-galactosidase gene is expressed in yeast. Thus, the rp51 intron is properly excised from the normally intronless gene. The presence of the rp51 intron causes the β-galactosidase activity to be sensitive to the *rna2* mutation, consistent with the notion that this mutation affects gene expression at the level of splicing. The experiments suggest that an intron-containing β-galactosidase gene can be used in a general way to study mRNA splicing.**

The temperature-sensitive *rna2* mutation of *Saccharomyces cerevisiae* causes a dramatic decline in the synthesis of most ribosomal proteins when cells carrying the *rna2* mutation are shifted to the restrictive temperature. This effect appears to be quite specific for ribosomal protein synthesis—i.e., the rate of synthesis of total protein and of most nonribosomal proteins is relatively unaffected (1). After a shift to the restrictive temperature, mRNA isolated from cells that carry the *rna2* mutation is specifically depleted of most ribosomal protein mRNAs, suggesting that the decline in ribosomal protein synthesis is due to a decline in ribosomal protein mRNA levels (2). For many ribosomal proteins, the decrease in mature mRNA levels in cells carrying the *rna2* mutation is associated with increased levels of higher molecular weight transcripts (3–5). In the case of ribosomal protein 51 (rp51), the mature mRNA and the higher molecular weight transcript were shown by nuclease S1 mapping to be spliced and unspliced transcripts, respectively (3).

These data suggest that many yeast ribosomal protein genes contain introns and that the *rna2* mutation interferes with ribosomal protein mRNA processing. Here we present the rp51 DNA sequence and identify the rp51 intron. We also demonstrate that the rp51 intron, when inserted into a yeast *CYC1–lacZ* fusion gene (6), is correctly spliced from the fusion transcript, thereby allowing β-galactosidase expression. Splicing of the rp51 intron from the *CYC1–lacZ* fusion transcript is further shown to be defective in cells carrying the *rna2* mutation when shifted to the nonpermissive temperature.

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

## MATERIALS AND METHODS

**Strains.** The haploid *S. cerevisiae* strain RY26 (*α, rna2, ura3-1, ura3-2, ade1, ade2, tyr1, his7, canR, gal*) was a generous gift of Richard Young, and DB745 (*α, ade-100, leu2-3, leu2-112, ura3-52*) was constructed in the laboratory of D. Botstein. RY26 was mated to the haploid strains PT216-2D (*a, rna2, his4, ura3, cry1, mal2*) and PT216-6B (*a, ade1, ura3, cry1, mal2*), obtained from J. Haber, to generate the diploid strains JT2D (*rna2/rna2,ura3/ura3*) and JT6B (*rna2/RNA2, ura3/ura3*), respectively. The plasmid pLGSD5 (6) was provided by L. Guarente. Plasmid DNA preparations, restriction enzyme digestions, ligation reactions, and *Escherichia coli* transformations were performed as described (7). Yeast transformations were also performed as described (8).

**Media.** For temperature-shift experiments, cells were grown on supplemented minimal media (EMM): 6.7 g of yeast nitrogen base (without amino acids) per liter supplemented with amino acids, purines, and pyrimidines, except uracil, and 2% ethanol as a carbon source.

**β-Galactosidase Assay.** Assays were performed as described (9), and the units of activity were calculated (10). Supplemented minimal media plates without uracil and containing X-Gal indicator were prepared as described (9).

**DNA Sequence Analysis.** The method of Maxam and Gilbert (11) was used for sequence analysis of restriction fragments labeled at 5′ or 3′ termini.

**cDNA Synthesis.** Poly(A)$^+$ RNA was isolated as described (3) from RY26 grown on YM-1 media (12) with 2% glucose and from pHZ18-transformed DB745 grown in supplemented minimal media with 2% galactose. The synthetic oligonucleotide DNA primer was a gift of R. Schwartz and D. Engelke. The primer was labeled with $^{32}P$ to a specific activity of $\approx 10^8$ cpm/μg by using T4 polynucleotide kinase (11). After purification from DEAE-cellulose, the radioactive primer was isolated from a 15% acrylamide sequencing gel and was repurified from DEAE-cellulose. Labeled primer ($10^5$ cpm) was incubated at 42°C with 5 μg of poly(A)$^+$ RNA in 15 μl of 0.4 M Tris·HCl, pH 8.0/0.32 M NaCl/4 mM EDTA (1.6× RT buffer) for 60 min. The reaction was adjusted to 20 μM dATP, dGTP, dCTP, and dTTP, 6 mM $MgCl_2$, 12.5 μM dithiothreitol, 100 μg of actinomycin D per ml, and 19 units of reverse transcriptase added to obtain a final volume of 25 μl; it was then incubated at 25°C for 3 hr. For sequence analysis of primer extended cDNA, dideoxynucleosides were added to the reaction to the concentrations specified (13). The cDNA reaction products were analyzed on 15% acrylamide sequencing gels.

## RESULTS

The general features of the rp51 gene structure have been previously determined by restriction mapping, subcloning, RNA

Abbreviations: bp, base pair(s); kb, kilobase(s).

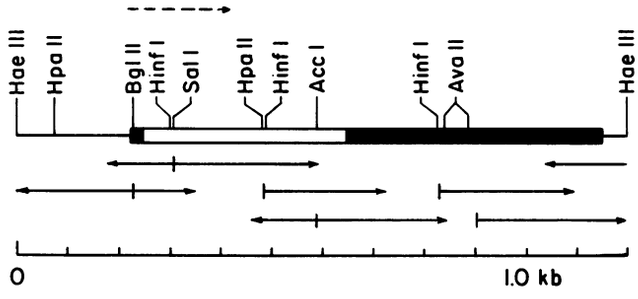



FIG. 1. Structure of the rp51 gene. The *Hae* III fragment of 1.2 kilobases (kb) which contains the rp51 gene is shown. Exons are represented by the two black boxes and the intron is in between. The direction of transcription is indicated by the dashed arrow. Solid arrows extend away from those restriction sites labeled for DNA sequence analysis, indicating the direction and extent of analysis. Both strands have been subjected to sequence analysis for the exon and intron regions, except for 100 bp of intron located between *Sal* I and *Hpa* II sequenced only on one strand. The rp51 DNA plasmid, pY11-138 (14), was used to determine the rp51 DNA sequence, including 60 bp upstream from the *Bgl* II site. Subcloned DNA from λ11-138 (15) was used to overlap and extend the DNA sequence upstream from the *Bgl* II site.

blotting, and nuclease S1 mapping (3, 14). These data indicate that a single *Bgl* II site lies within a small 5′ exon (exon 1) and that the major portion of the rp51 coding sequence lies within a second exon (exon 2), separated from exon 1 by an intron of about 400 base pairs (bp), as shown in Fig. 1.

By using the restriction sites and sequence analysis strategy shown in Fig. 1, the rp51 DNA sequence was determined by the method of Maxam and Gilbert (11) and is shown in Fig. 2. The sequence G-G-T-A-T-G occurs in the rp51 sequence 25 bp downstream from the *Bgl* II site (at position +4). This sequence conforms to the consensus 5′ donor splice sequence (R-G-T-X-X-G, where R = purine and X = any nucleoside) found in other eukaryotic mRNAs (16). Nuclease S1 mapping experiments have shown that splicing of rp51 mRNA takes place at approximately this position (data not shown). This and other experiments (see below) indicate that this sequence is the rp51 5′ donor site. The rp51 intron DNA sequence following the 5′ donor site is very A+T-rich and has stop codons in all three reading frames. At a position 398 bp downstream from the 5′ donor site there begins a single large open reading frame which terminates at the stop codon TAA. Translation of this open reading frame generates the sequence of a basic protein of 135 amino acids, a size sufficient to account for the rp51 protein. Directly preceding the open reading frame (at position +396), the sequence T-A-A-T-A-G occurs, which differs by one nucleotide from the consensus 3′ acceptor splice site sequence Y-Y-X-Y-A-G (Y = pyrimidine and X = any nucleoside) (16). Nuclease S1 mapping experiments indicate that this is the 3′ acceptor site and that no other splices occur within exon 2 (data not shown).

The sequence of the rp51 mRNA was determined in the vicinity of the splice junction that was predicted from the DNA sequence analysis and from nuclease S1 mapping experiments. An oligonucleotide of 19 nucleotides, complementary to rp51 mRNA at a position 10 bp downstream from the 3′ acceptor site

```
-250       -240       -230       -220       -210       -200       -190       -180       -170       -160
GGCCTGTCCA GAGGACAATA GACCAATGGC AGAGACGAGG TTTGTTGGGT ACGTGCAACA GCGCCTGGGC TCAGCCGGCT TGCTGTGGGC AACGGAATGG

-150       -140       -130       -120       -110       -100       -90        -80        -70        -60
TGCTTCCTAA GCCAATCTTT CATGCTGTAG CTATCGTCGT GATCGCTTTT ACTGGCATAT TAGCGTAATA TGTATGAAAT TAGGTATTAA TCAAGCATTA

-50        -40        -30        -20        -10        MET        10         20         30         40
ATCGACTTAA TTCTAAGAAA AGTCAAGATC TCGAGACTAG CAATAACAAA ATG  GTATGTT AATATGGACT AAAGGAGGCT TTTAAGGACA

        50         60         70         80         90         100        110        120        130        140
CGTAATATTG AGTCGACATG CGCAATAAAG TCATACAATA AATAAGTTAA AGAAAACATA TGACGATATT TTCAGACGAT ATGGAACTGC GATACCGAAA

       150        160        170        180        190        200        210        220        230        240
AAGTGGTGTG ATTGAAATGA TACTCATGTT TTGGCTATAG TTGCAGTGGC TTTCCAAAAA ATTGGGAGTT AAAATGCCTA TAGCAGTATC ACCGGAGATT

       250        260        270        280        290        300        310        320        330        340
CAGTCTCTTG TGGATCTTCT TTTTGCATAG TTGAACAAGG GATAATATGG CGCCATGAAC TGCCTCTACA CATTATAATG TGTTTTTGAT ATCAGTATAC

                                                                                                        10
       350        360        370        380        390        400   GLY ARG VAL ARG THR LYS THR VAL LYS ARG
TAACAAGTTG AATTGCATTT ACAAACTTTT TATTTTGTAT TGCTTTTCGT CATTTTAATAG  GGT AGA GTT AGA ACC AAG ACC GTC AAG CGT

                                  20                                        30
ALA SER LYS ALA LEU ILE GLU ARG TYR TYR PRO LYS LEU THR LEU ASP PHE GLN THR ASN LYS ARG LEU CYS ASP GLU ILE
GCT TCT AAG GCT TTG ATT GAA CGT TAC TAT CCA AAG TTG ACT TTG GAT TTC CAA ACC AAC AAG AGA CTT TGT GAT GAA ATC

     40                                        50                                        60
ALA THR ILE GLN SER LYS ARG LEU ARG ASN LYS ILE ALA GLY TYR THR THR HIS LEU MET LYS ARG ILE GLN LYS GLY PRO
GCC ACT ATC CAA TCC AAG AGA TTG AGA AAC AAG ATT GCT GGT TAC ACC ACC CAT TTG ATG AAG AGA ATC CAA AAG GGT CCA

               70                                        80                                        90
VAL ARG GLY ILE SER PHE LYS LEU GLN GLU GLU GLU ARG GLU ARG LYS ASP GLN TYR VAL PRO GLU VAL SER ALA LEU ASP
GTT AGA GGT ATC TCT TTC AAA TTG CAA GAA GAA GAA AGA GAA AGA AAG GAC CAA TAC GTC CCA GAA GTC TCT GCT TTG GAC

                          100                                       110
LEU SER ARG SER ASN GLY VAL LEU ASN VAL ASP ASN GLN THR SER ASP LEU VAL LYS SER LEU GLY LEU LYS LEU PRO LEU
TTG TCT CGT TCT AAC GGT GTT TTG AAC GTT GAC AAC CAA ACT TCT GAC TTG GTT AAA TCT TTG GGT TTG AAG TTG CCA TTA

120                                       130
SER VAL ILE ASN VAL SER ALA GLN ARG ASP ARG ARG TYR ARG LYS ARG VAL             820        830        840
TCT GTT ATC AAC GTT TCT GCC CAA AGA GAC AGA CGT TAC AGA AAG AGA GTT TAA AATTAAATTAG AAAGCTATTT AAAATAATTT

       850        860        870        880        890        900        910        920        930        940
ACTATTCAAA ATATTTGCCT TTTCTTTTTT AATTTTTGTT TATTCTTTAA TGTATAATTA AATAAAAAAA TATTATTATA TTTACTAATT AAGCGAAGCG

       950        960
TTTTATGTAG CTCCTTGGCC
```

FIG. 2. DNA sequence of the rp51 gene. The DNA sequence of the 1.2-kb *Hae* III fragment (Fig. 1) that includes the rp51 gene is shown. The sequence complementary to the oligonucleotide primer (5′ C-G-C-T-T-G-A-C-G-G-T-C-T-T-G-G-T-T-C 3′) in exon 2 is underlined.

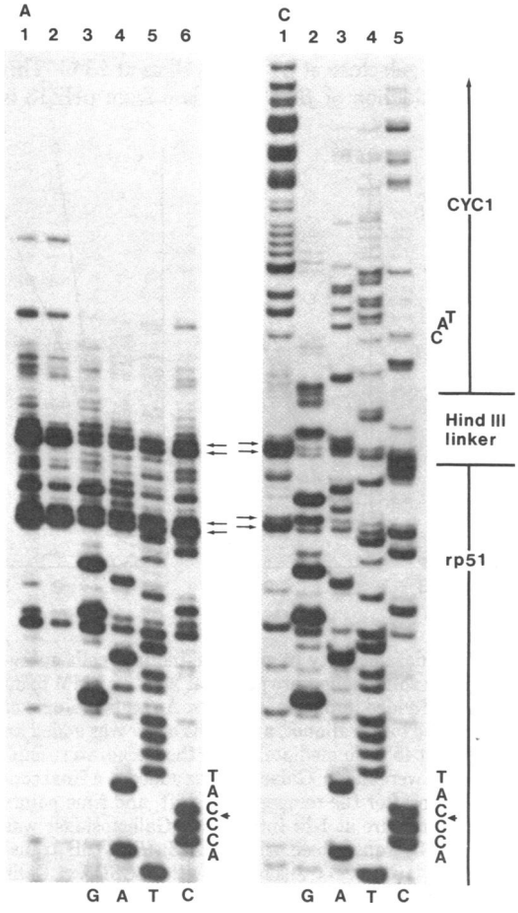
(see Fig. 2), was used as a primer with yeast poly(A)$^+$ RNA as a template for cDNA synthesis. The cDNA produced in the reaction is an extension of the oligonucleotide primer across the splice junction of the rp51 mRNA. The cDNA sequence ladder, generated by including dideoxynucleoside terminators in the synthesis reaction, is shown in Fig. 3*A*, lanes 3–6. The cDNA sequence generated is colinear with that of the 3′ exon up to (and including) the glycine codon (GGU), which is directly adjacent to the consensus 3′ acceptor splice site (Fig. 3*B*). Beyond the splice point, the sequence then corresponds to that of the 5′ exon, beginning with a methionine codon (AUG) that directly precedes the 5′ donor splice sequence. The rp51 mRNA sequence precisely defines the splice junction and confirms the predictions of the DNA sequence analysis.

There are two major cDNA products (marked by double arrows in Fig. 3*B*), separated by 6 bp. Nuclease S1 mapping experiments (data not shown) indicate that these correspond to the two major rp51 mRNA 5′ termini. The AUG that precedes the 5′ donor splice site must be the initiator methionine codon because there are no alternative AUG codons in any reading frame within the limits of exon 1. Thus, this 5′ exon contains only a single codon, and the mRNA sequence preceding this AUG corresponds to a short 5′ untranslated sequence.

The rp51 intron was introduced into the DNA sequence encoding the galactose-inducible *CYC1–lacZ* fusion gene in the plasmid pLGSD5. In yeast cells transformed with pLGSD5, β-galactosidase activity is observed only when cells are grown on the carbon source galactose (6). An outline of the construction of an intron-containing derivative of pLGSD5 (designated pHZ18) is shown in Fig. 4. In brief, a 615-bp *Bgl* II/*Ava* II restriction fragment from the rp51 gene (including the entire 5′ exon sequence, the intron, and 63 codons from the 3′ exon) was inserted into the coding region of the *CYC1–lacZ* fusion such that the open reading frame of the fusion gene was interrupted by the rp51 intron. The open reading frame sequence that begins with the *CYC1* start AUG continues into the rp51 exon 1 sequence (beginning at the *Bgl* II end of the fragment), including the rp51 AUG codon, which directly precedes the 5′ donor splice site. Translation beyond this point is terminated by stop codons within the rp51 intron. However, splicing of the rp51 intron from the transcript allows translation to continue in frame through the rp51 3′ exon towards the *Ava* II end of the inserted rp51 fragment. The *Ava* II end of the fragment has been joined in frame to the *lacZ* gene in pLGSD5 so that splicing should also allow *lacZ* expression. Thus, a RNA transcribed from the intron-containing fusion gene cannot be translated to produce β-galactosidase activity in yeast unless splicing occurs. Removal of the intron allows translation, initiated at the *CYC1* start codon, to proceed through the *lacZ* sequence.

The pHZ18 plasmid was transformed into yeast and *URA3* transformants were assayed on β-galactosidase indicator plates. As expected, no β-galactosidase activity was detected in cells grown on glucose, whereas cells grown on galactose were lacZ$^+$. This result suggests that the intron is correctly removed from the hybrid gene RNA (see *Discussion*).

Because the HZ18 fusion mRNA also contains the rp51 exon sequences complementary to the primer, cDNA can be synthesized with the fusion mRNA as well as rp51 mRNA as a template. Thus, poly(A)$^+$ RNA, prepared from pHZ18-transformed DB745 cells grown on glucose or on galactose, was used as a template for cDNA primer extension experiments by using the same oligonucleotide primer employed above. Because little or no fusion mRNA is synthesized in the absence of galactose (6), RNA isolated from cells grown on glucose produce only the cDNA



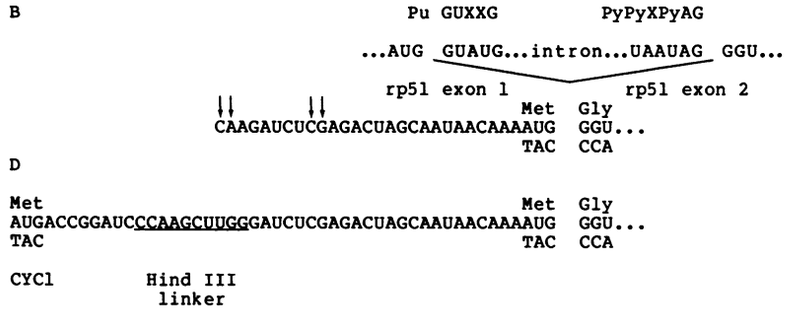




FIG. 3. cDNA synthesis from oligonucleotide-primed mRNA. Poly(A)$^+$ RNA was hybridized to the oligonucleotide primer complementary to the 3′ exon sequence of rp51 (see Fig. 2). cDNA extensions of the primer were synthesized and electrophoresed on 15% sequencing gels. (*A*) Lane 1, poly(A)$^+$ RNA obtained from RY26 cells grown at 23°C. Lane 2, poly(A)$^+$ RNA obtained from RY26 cells grown at 34°C. Lanes 3–6, the same reaction as lane 1 with dideoxynucleotide terminators included. The position of the splice is shown by an arrowhead. Below the arrowhead is the cDNA sequence C-C-A complementary to the glycine codon of exon 2 and adjacent to the 3′ acceptor site. Above the arrowhead is the cDNA sequence T-A-C complementary to the methionine codon that precedes the 5′ donor splice in exon 1. Double arrows indicate the two major cDNA doublets corresponding to the 5′ ends of rp51 mRNA. There exists in lane 2 a high molecular weight band, presumably corresponding to a primer extension product of rp51 precursor RNA (3). This region of the gel is not shown in the figure. (*B*) Sequence data from *A* represented as the RNA, indicating the splice point and 5′ ends of rp51 mRNA. The rp51 consensus splice sequences at the splice junction are shown. Pu, purine; Py, pyrimidine; and X, any nucleoside. (*C*) Lane 1, poly(A)$^+$ RNA obtained from pHZ18-transformed DB745 cells grown in galactose media. Lanes 2–5, the same reaction as lane 1 with dideoxynucleoside terminators included. Double arrows refer to the major cDNA products resulting from primer-extended rp51 mRNA. The position of the splice is indicated by the arrowhead as in *A*. Portions of the sequence ladder representing rp51, the *Hin*dIII linker, and *CYC1* are indicated. The T-A-C sequence shown in the *CYC1* region of the ladder is the cDNA complement of the *CYC1* start codon. (*D*) Sequence data from *C* represented as the RNA, indicating the splice point of HZ18 mRNA.

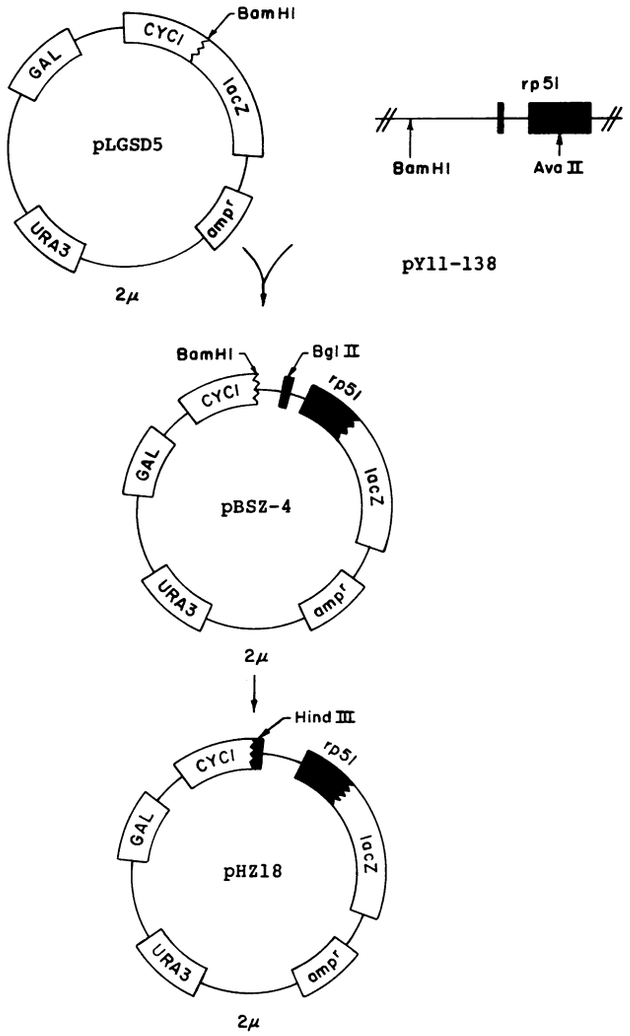



FIG. 4. Construction of pHZ18. pLGSD5 plasmid DNA was linearized with *Bam*HI at the junction where the *CYC1* start codon has been fused to the *lacZ* gene. The plasmid pY11-138 containing the rp51 gene (14) was cut with *Bam*HI and *Ava* II, and a 1.2-kb fragment containing exon 1, the intron, and 63 codons of exon 2 was isolated. rp51 exon sequences are indicated by the two black segments, the rp51 intron lying in between. The *Bam*HI end of the fragment was ligated to *Bam*HI-digested pLGSD5 DNA. Klenow fragment was then used to fill in both the *Ava* II end of the fragment as well as the remaining *Bam*HI end of the vector. The filled-in *Ava* II site was then ligated to the filled-in *Bam*HI site, joining the rp51 exon 2 sequence in frame to the *lacZ* coding sequence and circularizing the plasmid DNA to produce pBSZ-4. pBSZ-4 DNA was cut with *Bgl* II and *Bam*HI and the vector fragment was isolated. The *Bam*HI end and the *Bgl* II end of the fragment were filled in with Klenow fragment. *Hin*dIII linkers were added, and the plasmid DNA was circularized by ligation. Inclusion of a single *Hin*dIII linker between the filled-in *Bam*HI and *Bgl* II ends of the vector correctly joins the reading frame of rp51 exon 1 to the reading frame initiated at the *CYC1* start codon, resulting in pHZ18. A plasmid that was ligated without including the *Hin*dIII linker (pHZ18-1) was also isolated. amp$^r$, ampicillin resistance.

products corresponding to primer-extended rp51 mRNA identical to those seen in untransformed cells (as in Fig. 3*A*, lane 1). However, cDNA generated with mRNA from cells grown on galactose contained an additional set of larger molecules (Fig. 3*C*, lane 1). Dideoxynucleoside terminators were included in the cDNA synthesis reaction to generate a sequence ladder for these primer-extended products (Fig. 3*C*, lanes 2–5). This sequence ladder (the sum of the cDNA sequence from rp51 mRNA as well as HZ18 mRNA) is identical to the rp51 cDNA sequence up to the position of the two intense doublet bands that represent the 5′ ends of rp51 mRNA, because both rp51 and HZ18 mRNA share the same sequence up to this point. The lack of any significant sequence heterogeneity in the vicinity of the splice joint suggests that the rp51 intron is correctly removed from the HZ18 primary transcript, because only cDNA synthesized from properly spliced HZ18 mRNA could generate the same rp51 sequence as cDNA synthesized from mature rp51 mRNA (see *Discussion*). Extending beyond the position of the rp51 5′ ends, the sequence corresponds to that of the HZ18 fusion gene, which includes the sequence of a *Hin*dIII linker (inserted during the HZ18 construction; see Fig. 4), followed by the sequence of *CYC1* (Fig. 3*D*).

The cDNAs synthesized from HZ18 mRNA are heterogeneous in length (Fig. 3*C*, lane 1), suggesting that the HZ18 fusion mRNA has the same heterogeneous 5′ ends as *CYC1* mRNA (17). This result is not unexpected, because the 5′ ends of pLGSD5 fusion mRNA are the same as the 5′ ends of *CYC1* mRNA (6).

A diploid strain homozygous for the *rna2* mutation was transformed with pHZ18 to determine if β-galactosidase expression from the HZ18 fusion gene is affected by the *rna2* mutation at the nonpermissive temperature. Cells that have been maintained at 23°C rapidly synthesize β-galactosidase activity upon addition of galactose, reaching about 5 units of activity after 3 hr. In contrast, cells that were shifted to 34°C prior to addition of galactose do not accumulate detectable β-galactosidase activity (Fig. 5*A*). The data suggest that the expression of the HZ18 fusion gene is inhibited at the nonpermissive temperature in a *rna2* background. Similar results were obtained with a pHZ18-transformed *rna2* haploid strain (data not shown). To determine whether the failure to accumulate β-galactosidase activity at the nonpermissive temperature is specific to the *rna2* mutation, the experiment was repeated with a diploid strain heterozygous for the *rna2* mutation (Fig. 5*B*). In this strain β-galactosidase can be induced with galactose at 34°C as well as at 23°C. This indicates that the induction of β-galactosidase from pHZ18 is

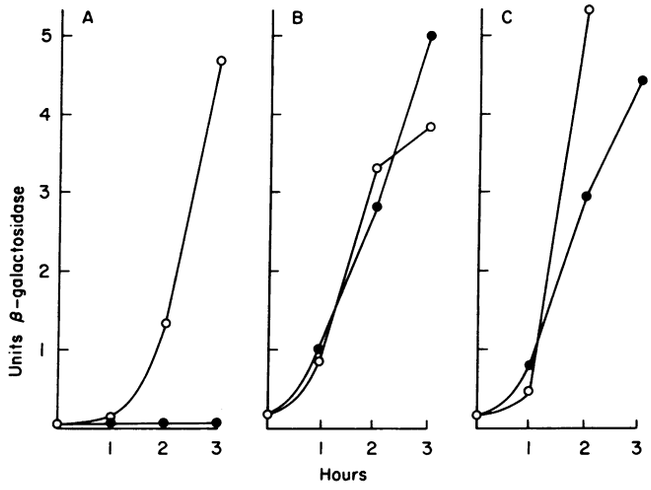



FIG. 5. β-Galactosidase activity in galactose-induced cells subsequent to a temperature shift. Cells were grown at 23°C on EMM to an $OD_{610}$ of 0.3 and were divided into equal aliquots. An equal volume of EMM at 23°C was added to one aliquot, and to the other was added an equal volume of EMM at 45°C. Immediately after the temperature shift samples of each culture were taken. Galactose was added to a final concentration of 2%, 3 min after the temperature shift, and time points were taken from each culture at 1-hr intervals. β-Galactosidase was then assayed. (*A*) JT2D transformed with pHZ18; (*B*) JT6B transformed with pHZ18; and (*C*) JT2D transformed with pLGSD5. ○, Cells maintained at 23°C; and ●, cells maintained at 34°C.

not temperature sensitive in the presence of a wild-type *RNA2* allele. Moreover, diploid cells homozygous for *rna2* and transformed with pLGSD5 (in which the *CYC1–lacZ* fusion has no intron) can express β-galactosidase upon galactose induction at the nonpermissive temperature (Fig. 5C). These data suggest that processing of HZ18 mRNA is affected at the nonpermissive temperature by the *rna2* mutation.

To demonstrate that translation of HZ18 mRNA is initiated before the 5′ donor site, a different plasmid was constructed (pHZ18-1) that is identical to pHZ18 except that the *Hin*dIII linker was omitted during the insertion of the restriction fragment containing the rp51 intron (see Fig. 3). The omission of the 10-bp *Hin*dIII linker shifts the reading frame of a RNA encoding a protein initiated before the rp51 sequence, resulting in premature chain termination of both spliced and unspliced RNA. The reading frame of a protein initiated after the beginning of the rp51 sequence should be unaffected by the omission of the *Hin*dIII linker. Consistent with the notion that translation initiates at the *CYC1* methionine, no β-galactosidase activity was observed in pHZ18-1-transformed DB745 cells grown in the presence of galactose (data not shown).

## DISCUSSION

Previous nuclease S1 mapping experiments indicated that rp51 mRNA is spliced (3). Furthermore, a RNA corresponding to an unspliced rp51 RNA was shown to accumulate in *rna2* cells at the nonpermissive temperature. The sequence analysis of the rp51 gene presented here confirms these data and shows that the gene contains a single 398-bp intron that separates a small 5′ exon (containing a single methionine codon) from a 3′ exon containing the remainder of the rp51 coding sequence.

A detailed comparison of the rp51 intron sequence with other intron-containing yeast genes reveals a number of interesting features. These comparisons, and a comparison of the entire rp51 gene with other yeast ribosomal protein genes that have been subjected to sequence analysis, will be presented elsewhere. However, we note that within all nonmitochondrial yeast mRNA introns that have been subjected to sequence analysis to date, the conserved sequence T-A-C-T-A-A-C occurs between 20 and 50 nucleotides from the 3′ acceptor splice site. The conservation of this sequence and its position within these introns suggest that it provides some function required for splicing. This intron-containing β-galactosidase gene provides a means to test this hypothesis and to define more generally the intron sequences important for mRNA splicing.

For the majority of the cloned yeast ribosomal protein genes analyzed thus far, a higher molecular weight RNA—corresponding presumably to an unspliced precursor RNA—accumulates in *rna2* cells shifted to the nonpermissive temperature (3–5). Indeed, for the gene that codes for S10 (5), DNA sequence analysis has confirmed the presence of an intron. The accumulation of unspliced RNAs at the nonpermissive temperature suggests strongly that mRNA splicing is adversely affected in the presence of *rna2* at the nonpermissive temperature. We have found that unspliced actin RNA (18, 19) also rapidly accumulates in *rna2* cells at the nonpermissive temperature (unpublished data). These observations make the prediction that an intron is sufficient to render a gene sensitive to the *rna2* mutation. The absence of β-galactosidase activity at the nonpermissive temperature in pHZ18-transformed cells mutant at the *rna2* locus is consistent with this hypothesis.

In constructing pHZ18, the rp51 intron was inserted between the initiating methionine codon from *CYC1* and the *lacZ* sequence. Two independent experiments suggest that the hybrid β-galactosidase RNA, transcribed from HZ18, is correctly spliced. First, the absence of any significant sequence heterogeneity in the rp51 portion of the cDNA sequence of the HZ18 mRNA (Fig. 3C) indicates that the hybrid β-galactosidase mRNA is spliced in the same position as rp51 mRNA. We would have expected to detect the intron sequence or some other sequence if a significant fraction of the HZ18 RNA was inefficiently or imprecisely spliced. Second, the mere presence of β-galactoside activity in response to galactose induction suggests that the HZ18 mRNA is spliced.

The results imply that the rp51 intron can be efficiently removed from other sequences into which it has been inserted. Recent experiments indicate that the rp51 intron can be efficiently removed from the yeast actin gene (20). All of these results imply that the specificity for mRNA splicing resides within or near the intron sequence itself.

Introns in nonribosomal protein yeast genes appear to be quite rare—even when the corresponding genes in metazoans contain introns (21–23). In contrast, introns in ribosomal protein genes appear to be relatively common (3–5). The significance of this unusual apportionment of introns among yeast genes is unclear. The effect of *rna2* on mRNA processing together with this unusual distribution of intron-containing genes can account for the apparent specificity of the *rna2* mutation for ribosomal protein gene expression.

We are grateful to L. Guarente, R. Schwartz, and D. Engelke for supplying reagents, R. Young and J. Haber for providing strains, and M. Gray, M. Osley, C. Pikielny, and L. Hereford for reviewing the manuscript. J.L.T. was supported by a National Institutes of Health Training Grant (GM 07122). This work was supported by a grant from the National Institutes of Health (GM 23549) to M.R.

1. Gorenstein, C. & Warner, J. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1547–1551.
2. Warner, J. R. & Gorenstein, C. (1977) *Cell* **11**, 201–212.
3. Rosbash, M., Harris, P. K. W., Woolford, J. L. & Teem, J. L. (1981) *Cell* **24**, 679–686.
4. Fried, H. M., Pearson, N. J., Kim, C. H. & Warner, J. R. (1981) *J. Biol. Chem.* **256**, 10176–10183.
5. Leer, R. J., van Raamsdonk-Duin, M. M. C., Molenaar, C. M. Th., Cohen, L. H., Mager, W. H. & Planta, R. (1982) *Nucleic Acids Res.* **10**, 5869–5878.
6. Guarente, L., Yocum, R. & Gifford, P. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7410–7414.
7. Gray, M. R., Colot, H. V., Guarente, L. & Rosbash, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6598–6602.
8. Hinnen, A., Hicks, J. & Fink, G. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1929–1933.
9. Osley, M. & Hereford, L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7689–7693.
10. Miller, J. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
11. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
12. Hartwell, L. H. (1967) *J. Bacteriol.* **93**, 1662–1670.
13. Klemenz, R. & Geiduschek, P. E. (1980) *Nucleic Acids Res.* **8**, 2679–2690.
14. Woolford, J. L., Hereford, L. M. & Rosbash, M. (1979) *Cell* **18**, 1247–1259.
15. Woolford, J. L., Jr., & Rosbash, M. (1981) *Nucleic Acids Res.* **9**, 5021–5036.
16. Seif, I., Khoury, G. & Dhar, R. (1979) *Nucleic Acids Res.* **6**, 3387–3398.
17. Faye, G., Leung, D. W., Tatchell, K., Hall, B. D. & Smith, M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2258–2262.
18. Ng, R. & Abelson, J. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3912–3916.
19. Gallwitz, D. & Sures, I. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2546–2550.
20. Langford, C., Nellen, W., Niessing, J. & Gallwitz, D. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1496–1500.
21. Scarpulla, R. C., Agne, K. M. & Wu, R. (1981) *J. Biol. Chem.* **256**, 6480–6486.
22. Smith, M., Leung, D. W., Gillam, S. & Astell, C. R. (1979) *Cell* **16**, 753–761.
23. Montgomery, D. L., Leung, D. W., Smith, M., Shalit, P., Faye, G. & Hall, B. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 541–545.