THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1984 by The American Society of Biological Chemists, Inc.
Vol. 259, No. 14, Issue of July 25, pp. 9218-9224, 1984
Printed in U.S.A.

A Yeast Ribosomal Protein Gene Whose Intron Is in the 5′ Leader*

(Received for publication, March 22, 1984)

Gopa Mitra‡ and Jonathan R. Warner

From the Departments of Cell Biology and Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461

The complete gene encoding yeast ribosomal protein 29, a component of the 60 S ribosomal subunit of *Saccharomyces cerevisiae* has been isolated and its nucleotide sequence determined. The coordinates of transcription initiation and termination of the rp29 gene have been mapped. Transcription appears to start at several sites, spanning nearly 30 nucleotides. The gene is transcribed into a precursor RNA molecule of 1110 to 1085 nucleotides from which an intron of 458 nucleotides is spliced out. The splice junction has been determined by comparing the nucleotide sequences of the rp29 gene with those of a homologous cDNA clone. Interestingly, the intron of the rp29 gene resides in the 5′ untranslated region of the rp29 mRNA, the first instance where this has been observed in yeast. The single open reading frame in the rp29 gene encodes a very basic protein of 155 amino acid residues.

The ribosome is of interest for two distinct reasons. On the one hand, it is a vital piece of the cell's machinery which has been highly conserved through evolution (1). Therefore, detailed knowledge of the structure of its components may lead to an understanding both of its function and of the evolutionary constraints upon it. On the other hand the synthesis of the ribosome is closely controlled by the cell, as demonstrated so clearly both for *Escherichia coli* (2) and for eukaryotes (3). Analysis of the genes which code for the components of the ribosome may lead to an understanding of the mechanisms involved in the coordinate regulation of the synthesis of these components.

We have, therefore, undertaken an extensive study of the ribosomal protein genes of the yeast, *Saccharomyces cerevisiae*. More than 20 of these genes have been cloned by several laboratories (4-7). A number of features of the structure and the regulation of these genes have been reported (reviewed in Ref. 3). Perhaps the most interesting finding to date is that most of the ribosomal protein genes have an intron, while very few other yeast genes do.

We now report a detailed analysis of the gene that codes for ribosomal protein 29 (rp29) (7). The entire gene has been cloned and sequenced. The location of the initiation and termination sites of both transcription and translation have been identified. An unusual feature of this gene is that an intron of 458 nucleotides lies within the 5′ untranslated region of its transcript. This has not previously been observed in a yeast gene. The predicted polypeptide is very basic, with a net positive charge of approximately +33.

MATERIALS AND METHODS

Yeast, Bacteria, Phage, and Plasmid—RNA was prepared from strain *ts368,MATa,ade1,ade2,ura1,tyr1,his7,lys2,gal1,rna2*^{ts} (8), a temperature-sensitive mutant of *S. cerevisiae* strain A364A. The yeast genomic DNA clone bank on the vector YEP13 (6) was propagated in *E. coli* C600. The yeast cDNA clone bank in the vector pMAC561 (9), a generous gift of G. McKnight (Zymos, Inc., Seattle), was propagated in *E. coli* RR1. The *E. coli* strain K12JM103 (10) was used for transformation with the phages M13 mp8 and mp9 containing subcloned DNA fragments of the yeast gene. Plasmid pRP29, containing the *Eco*RI/*Bam*HI fragment of the rp29 gene in pBR322, was described previously (11).

Enzymes and Chemicals—Restriction endonucleases and T4 DNA ligase were obtained from New England Biolabs. DNA polymerase I of *E. coli* and its Klenow fragment were obtained from Boehringer Mannheim. T4 polynucleotide kinase was obtained from P-L Biochemicals, avian myeloblastosis virus reverse transcriptase from Life Sciences Inc., and S1 nuclease from Sigma. [^{3}H]Lysine (94.8 Ci/mmol) was obtained from Amersham Corp. and [^{35}S]methionine (1058 Ci/mmol) from New England Nuclear.

Isolation of Recombinant Clones—Clone banks containing either yeast genomic DNA or cDNA were screened by the method of Grunstein and Hogness (12).

RNA Preparation—Total yeast RNA was isolated and poly(A$^+$) RNA purified from it as described previously (13).

DNA Preparation, Labeling, and Sequencing—Plasmid DNA preparations, restriction enzyme digestions, and ligation reactions were performed as described (14). DNA fragments were either 5′-end labeled with [γ-^{32}P]ATP (Amersham Corp., 3000 Ci/mmol) and T4 polynucleotide kinase or 3′-end labeled with deoxyribonucleoside[α-^{32}P]triphosphates (400 Ci/mmol) and reverse transcriptase (14). Fine structure mapping of restriction sites was carried out by the method of Smith and Birnstiel (15). DNA fragments were subcloned in phages M13 mp8 and mp9 and their sequence determined by the dideoxy chain termination method (16).

S1 Nuclease Mapping—S1 mapping was carried out according to Berk and Sharp (17) with end-labeled DNA. Hybridizations were performed at 50 °C in 80% formamide (deionized), 0.4 M NaCl, 0.4 M PIPES[1] (pH 6.5), and 1 mM EDTA for 4 h. 20 μg of poly(A$^+$) RNA were used for each experiment. Nuclease S1 digestion was done at 37 °C for 30 min with either 30 or 60 units of enzyme. The protected DNA fragments were precipitated with isopropanol using *E. coli* tRNA as carrier and were analyzed on 8% polyacrylamide-sequencing gels, containing 7 M urea.

Primer Extension Analysis—10 μg of poly(A$^+$) RNA were hybridized to 5′-end labeled DNA primer under conditions identical to those for S1 mapping. The hybrids were ethanol precipitated and exposed to reverse transcriptase (500 units/ml) as described (18). Extension reactions were carried out in 50 mM Tris-HCl (pH 8.3), 10 mM $MgCl_2$, 50 mM KCl, 10 mM dithiothreitol, 40 μg of actinomycin D per ml, and 1 mM each dATP, dGTP, dCTP, and dTTP at 41 °C for different time periods ranging from 30 min to 3 h. Following alkaline hydrolysis and phenol extraction the cDNA products were analyzed on an 8% polyacrylamide urea sequencing gel containing 7 M urea.

Partial N-terminal Sequence Determination—Messenger RNA specific for the rp29 gene was purified by hybridization of poly(A$^+$) RNA

* This research was supported by Grant GM25532 from the National Institutes of Health, Grant MV-203N from the American Cancer Society, and Grant CA13330 from the National Cancer Institute. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address, Department of Molecular Genetics, Hoffman-La Roche Inc., Nutley, New Jersey.

[1] The abbreviations used are: PIPES, 1,4-piperazinediethanesulfonic acid; kb, kilobase; bp, base pair(s).

to DNA of pRP29 immobilized on a nitrocellulose filter (14). The purified mRNA was released by heating for 45 s and was translated *in vitro* in a wheat germ extract in the presence of [^{3}H]lysine and [^{35}S]methionine as described (7). The translation products were subjected to electrophoresis on a 10–15% polyacrylamide gradient gel containing sodium dodecyl sulfate. After autoradiography, the desired band was excised, the protein was electroeluted, and was subjected to automated Edman degradation (19).

RESULTS

Isolation of Gene Coding for rp29

The gene coding for ribosomal protein 29 was initially identified in a recombinant λ phage clone A13 (7). Some sequences of the gene were found within a 1.5-kb *Eco*RI/*Bam*HI restriction fragment, subcloned into plasmid pRP29 (Fig. 1*A*) (11). pRP29 sequences were used to probe a Northern blot of RNA isolated from *ts*368, a temperature-sensitive strain carrying the *rna2* mutation, which blocks splicing at the restrictive temperature (20). The results suggested that the rp29 gene has an intron (7), since a transcript of 1050 nucleotides accumulates at the restrictive temperature, while the mature mRNA has only 620 nucleotides.

To ask whether pRP29 contained the complete gene, nuclease S1 mapping experiments were performed using two *Bgl*II sites that lie within the structural gene (Fig. 1*A*). A *Bgl*II/*Pst*I fragment (3.2 kb) of pRP29 was 3′-end labeled at the *Bgl*II site and hybridized to poly(A$^+$) RNA isolated from strain *ts*368 maintained at either 23 or 36 °C. After S1 digestion, a protected band of 300 nucleotides was observed for both RNA samples. A similar experiment with a 1.12-kb of *Bgl*II/*Pst*I fragment, 5′-end labeled at the *Bgl*II site, led to the protection of a band of 220 bp when poly(A$^+$) RNA isolated from the cells grown at 23 °C was used for hybridization and the protection of a band of 400 bp when poly(A$^+$) RNA isolated from the cells grown at 36 °C was used (Fig. 1*A*). Since the transcript of rp29 is approximately 1050 nucleotides, from which an intron of approximately 430 nucleotides is removed (7), it is clear that the origin of transcription must lie approximately 320 nucleotides upstream of the *Eco*RI site.

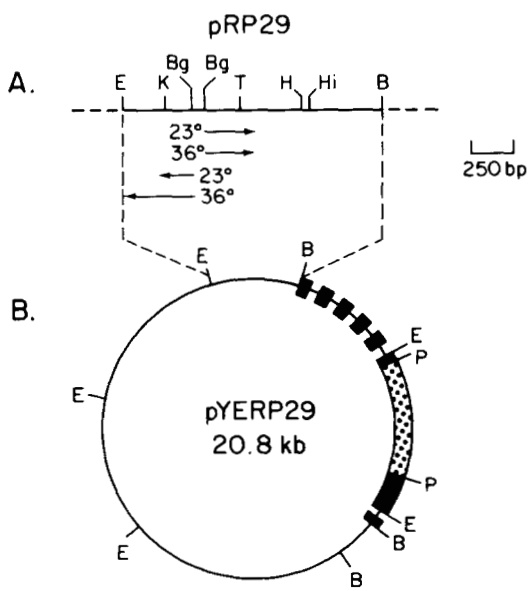

FIG. 1. **Isolation of the complete gene for ribosomal protein 29.** *A*, partial restriction map of the 1.5-kb *Eco*RI/*Bam*HI fragment of λA13 (6) subcloned into pBR322 to form the plasmid pRP29 (11). The *heavy broken line* shows pBR322 sequences. *Arrows* indicate direction and extent of protected fragments after S1 nuclease mapping (see text). The *Pst*I site of the *Bgl*II-*Pst*I fragments is that of the vector pBR322 (not shown). *B*, physical map of the plasmid pYERP29 showing restriction sites on the 10.5-kb *Bam*HI insert containing the rp29 gene (*thin line*) and the vector YEP13 (*heavy broken line*, pBR322 sequences; *heavy continuous line*, 2 μm DNA sequences; *dotted line*, LEU2 structural gene). *B*, *Bam*HI; *Bg*, *Bgl*II; *E*, *Eco*RI; *H*, *Hpa*II; *Hi*, *Hin*dIII; *K*, *Kpn*I; *T*, *Taq*I.

In order to isolate the entire gene encoding rp29, a clone bank in the vector YEP13 (6) was probed with the *Eco*RI-*Hin*dIII fragment of pRP29. The physical map of the plasmid pYERP29 containing the complete gene for rp29 is shown in Fig. 1*B*.

Sequence of rp29 Gene

The structure of the rp29 gene and the map of its restriction sites is shown in Fig. 2. Its sequence was determined by applying the dideoxy chain termination method (16) to fragments subcloned in derivatives of phage M13.

The complete nucleotide sequence of the rp29 gene is shown in Fig. 3. The origin of the numbering is the initiation triplet AUG of the open reading frame (see below).

Analysis of rp29

Location of the Coding Region—The single open reading frame of significant length in the gene for rp29 is indicated in Fig. 3. That this is the frame used by the cell's ribosomes is supported by two observations. 1) A partial NH_2-terminal sequence of the rp29 protein was determined by analysis of rp29 polypeptides translated *in vitro* from hybrid-selected mRNA corresponding to the rp29 gene. As predicted from the nucleotide sequence, methionine and lysine were identified as the first two amino acids. 2) To confirm the location of the splice sites, a cDNA clone for rp29 was isolated (see below). The end points of the cDNA clone are indicated in Fig. 3. This clone was constructed in an expression vector whose transcription is driven by the yeast *ADC1* promoter but whose translation must initiate within the cDNA (9). This plasmid was introduced into yeast and, by analysis on two-dimensional polyacrylamide gels, the synthesis of rp29 in the transformant was compared to that of other ribosomal proteins and to that in other strains. It is clear from Table I that the cDNA clone is responsible for the increased synthesis of rp29. Therefore, the cDNA clone must have the AUG which initiates translation. Since AUG (+1 to +3) is the first such codon in the cDNA, this must be the initiator triplet.

Structure of the Polypeptide—The predicted amino acid sequence of the protein rp29 is presented in Fig. 3. It is a basic protein of 155 amino acid residues with a molecular weight of 17,600 daltons. The protein is very basic, with 30 lysine and 16 arginine residues. The sequence of the rp29 gene reveals a strong bias in codon usage, similar to that found in other yeast genes (21). For example, 7 out of 8 glycine residues

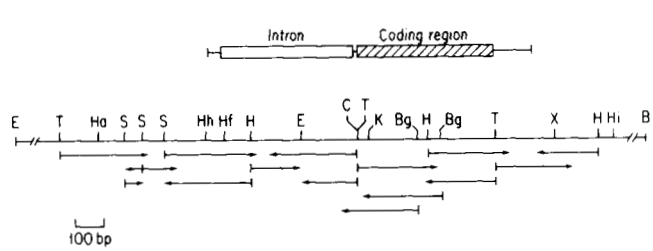

FIG. 2. **Restriction endonuclease map of DNA fragment carrying the rp29 gene and the strategy for DNA sequence determination.** *Arrows* indicate the direction and extent of sequence determination from the indicated sites. *B*, *Bam*HI; *Bg*, *Bgl*II; *C*, *Cla*I; *E*, *Eco*RI; *H*, *Hpa*II; *Hf*, *Hin*fI; *Hi*, *Hin*dIII; *Ha*, *Hae*III; *Hh*, *Hha*I; *K*, *Kpn*I; *S*, *Sau*3A; *T*, *Taq*I; *X*, *Xba*I.

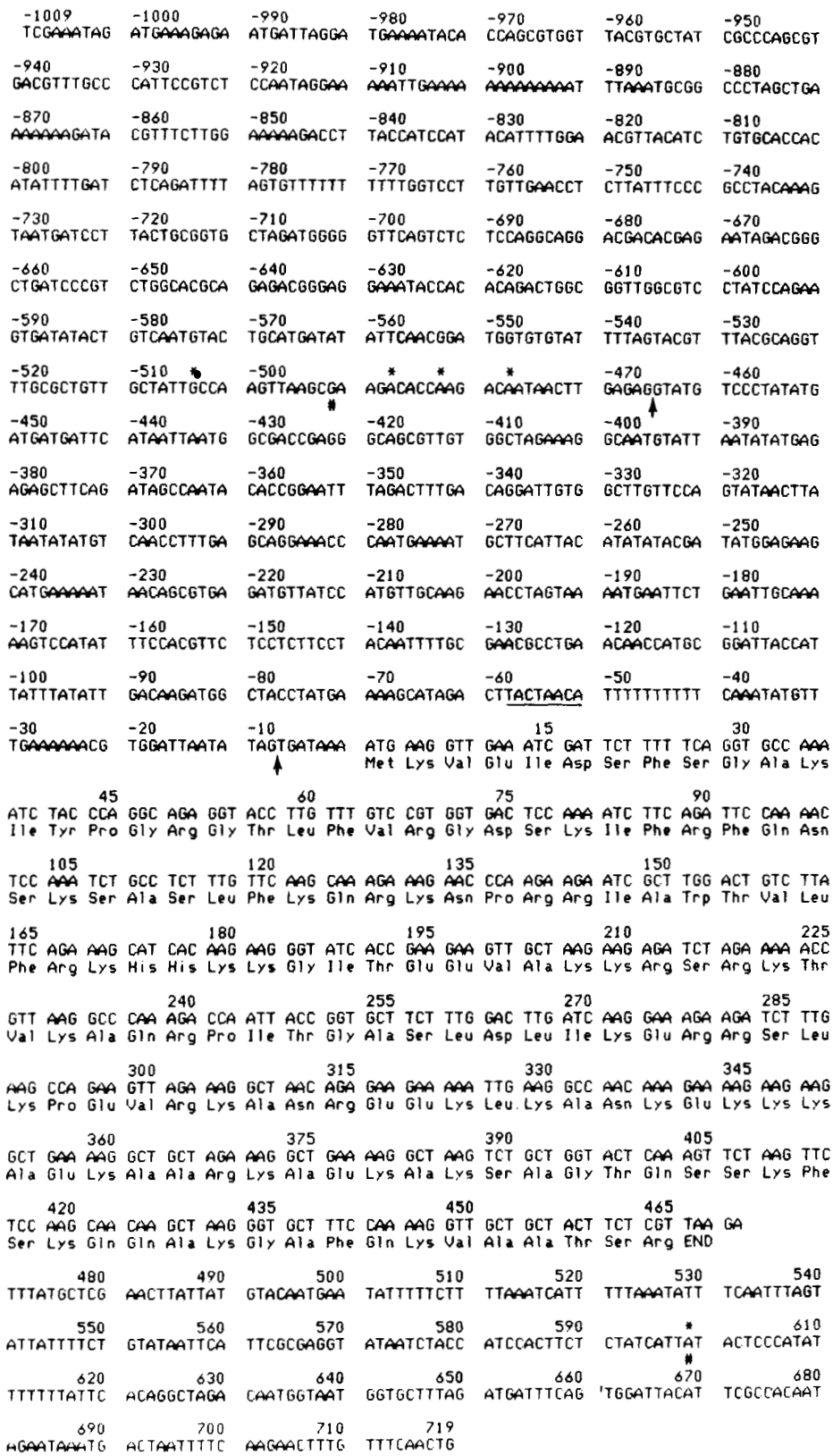

FIG. 3. **DNA sequence of the rp29 gene and the predicted amino acid sequence of the protein.** The 5′ donor and 3′ acceptor splice junctions of the intervening sequence are marked by *arrows*. The sequence 'TACTAACA' which specifies the 3′ splice site is *underlined*. The *asterisks* in the 5′ and 3′ flanking regions mark the sites for the ends of the mRNA. The # *symbol* marks the ends of the cDNA clone.

are coded by GGU, 14 out of 16 arginine codons by AGA, and all 10 glutamic acid codons by GAA.

The distribution of basic, acidic, and hydrophobic amino acid residues in rp29 is shown in Fig. 4. In the amino half of the protein only 5 acidic amino acid residues are present; they are also rare near the carboxyl terminus. A cluster of hydrophobic residues occurs at positions 49–55.

Analysis of the Intron

Northern gel analysis suggested that the intron of the rp29 gene has approximately 430 nucleotides (7). Splice sites in yeast genes appear to be defined by rigidly conserved sequences (22). The 5′ splice immediately precedes a GTATGT;

TABLE I

Excess synthesis of rp29 in cells bearing a cDNA clone

The synthesis of individual ribosomal proteins was measured as described previously (13). Three strains of yeast were grown to midlogarithmic phase and labeled for 3 min with [^{35}S]methionine. Total protein was prepared, mixed with protein from cells uniformly labeled with [^{3}H]methionine, and subjected to two-dimensional gel electrophoresis. Individual proteins were cut from the gel, solubilized, and the ^{35}S to ^{3}H ratio determined. This ratio is a measure of the rate of synthesis of the individual protein. Strains J451 and J501 are controls; strain J457 carries the cDNA clone of rp29 under the transcriptional control of the ADC1 promoter (9).

Ribosomal protein	^{35}S/^{3}H		
	Strain J451	Strain J501	Strain J457
22	1.58	1.91	1.81
23	1.54	2.17	1.63
27	1.08	1.19	1.14
28	1.76	1.93	1.79
29	1.21	1.40	5.76

the 3′ splice occurs after the first AG beyond a TACTAAC. The gene for rp29 contains each of these characteristic sequences only once. We first surmised, therefore, that the 5′ splice occurs after nucleotide (−466) (G)/GTATGT and the 3′ splice occurs after AG(T) (nucleotide −8), 43 nucleotides downstream from TACTAAC (−58 to −52). This is consistent with S1 nuclease mapping (Fig. 5), which indicates that the 3′ splice site of the intron is 217 bp upstream of the *Bgl*II site (+211).

In order to confirm the locations of the splice junction, we obtained a yeast cDNA bank from G. McKnight (9) and identified a clone carrying the rp29 cDNA insert by using the *Eco*RI/*Hin*dIII fragment (Fig. 2) as a probe. Fig. 6 shows a partial restriction map of the plasmid carrying the cDNA insert; the *arrows* indicate the extent of sequence determination. The sequence of the rp29 cDNA across the splice junction is shown in Fig. 7. The cDNA sequence is colinear with that of the four amino acid residues of the coding region, including ATG, and 7 nucleotides upstream, which are directly adjacent to the 3′ acceptor splice site. Beyond that the sequence corresponds to the 5′ region of the gene, starting with the sequence 5′GAGAG3′ that directly precedes the 5′ donor splice sequence. Hence the cDNA sequence confirms precisely the splice sites predicted from S1 mapping and DNA sequence analysis.

These results demonstrate, surprisingly, that the intron of the rp29 gene is within the 5′ leader of the transcript. While the splice sites of most yeast genes are within a few codons of the translation initiation site, this is the first report of a splice site in the noncoding region of a transcript.

Analysis of the Transcription Unit

Initiation of Transcription—The results of cDNA sequence determination indicate that the 5′ end of the rp29 gene is at least 492 nucleotides upstream from the initiator ATG codon (Fig. 3). In order to determine more precisely the site for the initiation of transcription, primer extension analysis was carried out. A 37-bp *Kpn*I-*Cla*I fragment, 5′-end labeled at the *Kpn*I site, (+51; Fig. 3) was hybridized to poly(A^{+}) RNA isolated from strain *ts*368 grown at the permissive temperature. This hybrid was then used as a primer for elongation by reverse transcriptase. Fig. 8*A* shows four major cDNA products, of sizes approximately 76, 81, 86, and 102 nucleotides, respectively. These results suggest that the rp29 gene probably has heterogenous transcription initiation sites which map at positions −478(A), −483(A), −488(A), and −504(G) (Fig. 3). Although the −504 initiation site is the weakest of the four, it must be the site responsible for the RNA transcribed into cDNA in our clone, which starts at position −493 (Fig. 3). In the rp29 gene, TATA-like sequences are present 38, 58, and 81 bp upstream of the first transcriptional start site at positions −543, −563, and −586, respectively. These sequences resemble the TATA box sequence which is present in most eukaryotic genes and is implicated in the eucaryotic RNA polymerase recognition site (23).

Termination of the rp29 mRNA—S1 mapping indicated that the 3′ end of the rp29 mRNA is approximately 300 nucleotides downstream from the *Bgl*II site (Fig. 1). A more accurate S1 analysis, using a fragment 3′ labeled at the *Taq*I site (+478) revealed two potential 3′ ends, with a weak band corresponding to position +596 and a strong one at position +599. The sequence of the 3′ end of the rp29 cDNA clone (Fig. 8*B*; Fig. 3) shows that the poly(A) tract starts at +599, 131 nucleotides downstream from the TAA stop codon, consistent with results of the S1 experiment. The spliced mRNA, therefore, is 645 to 619 nucleotides in length and arises from a molecule of 1103 to 1077 nucleotides.

DISCUSSION

Ribosomal proteins are in general small and basic. rp29 is no exception. Its position on a two-dimensional gel suggested it would be one of the most basic of the ribosomal proteins (24), and the sequence bears this out. A crude estimate, based on subtracting the number of aspartic and glutamic residues from the number of lysine and arginine residues, yields a net positive charge of (46 − 13)/155 = 0.21/residue. The values for other yeast ribosomal proteins are 0.13/residue for L29(CYH) (25) and 0.10/residue for L3(TCM) (26). In comparison yeast histone H2B has a value of 0.12/residue (27). It is evident from Fig. 4 that the basic residues of rp29 are nonrandomly distributed within the protein, with the amino terminus being particularly basic. This is also true for L29 (25) and rp51 (28) but less so for some other ribosomal proteins, *e.g.* S10 (29).

The primer extension experiment shown in Fig. 8 suggests that there are four transcription initiation sites. The method is, however, subject to the uncertainty that secondary structure can cause premature termination by the reverse transcriptase. Indeed, only a small fraction of the extended primer includes the sequences at the 5′ end of the cDNA clone. Nevertheless, there is no obvious secondary structure that would cause termination of reverse transcriptase at the downstream sites. Furthermore, multiple transcription initiation sites are common in yeast (30). Therefore, we conclude ten-

FIG. 4. **Schematic representation of the amino acid sequence of rp29 protein.** The basic residues, arginine and lysine, are indicated above the *horizontal line*, the acidic residues, glutamic and aspartic acid, *below the line. Dots on the line* indicate the hydrophobic residues: tryptophan, phenylalanine, valine, leucine, isoleucine, and alanine.

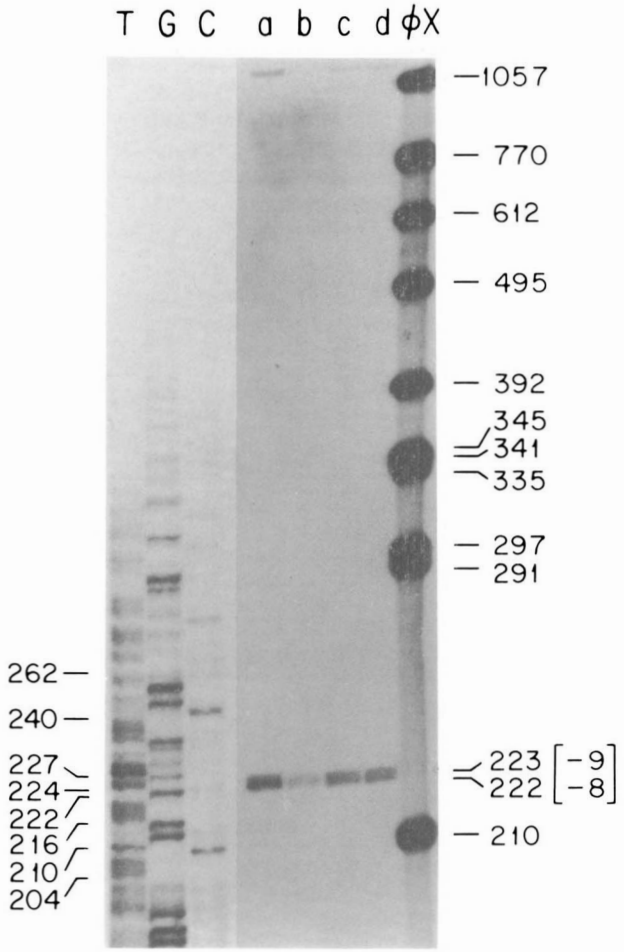

FIG. 5. **S1 nuclease mapping of the 3′ acceptor splice site of the intron of rp29 gene.** The 1.12-kb *Bgl*II/*Pst*I fragment of plasmid pRP29 containing the *Kpn*I site of the gene (Fig. 1*A*, *Pst*I site being within the pBR322 sequences) was 5′-end labeled at the *Bgl*II site (+211), hybridized to poly(A$^+$) RNA from yeast strain *ts*368 grown at 23 °C, and treated with S1 nuclease for 30 min at 37 °C. *Lanes a–d* show the protected fragments under various conditions: *a*, 40 μg of RNA, 30 units of S1; *b*, 20 μg of RNA, 60 units of S1; *c*, 20 μg of RNA, 30 units of S1; *d*, 40 μg of RNA, 60 units of S1. The C, G, and T ladder of a known sequence and radiolabeled *Hinc*II fragments of ϕX174 RF DNA were used to measure the length of the protected fragments. The size of the protected piece is indicated, together with its coordinate (−8).

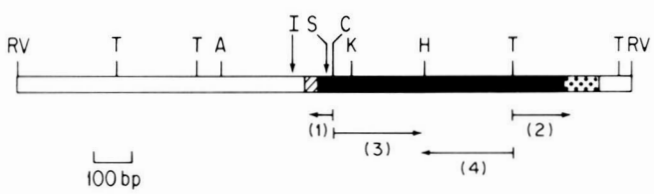

FIG. 6. **Partial restriction map of the 1.5-kb *Eco*RV fragment of the expression vector pMAC561 (9), carrying the cDNA insert of rp29 gene.** *Open bars*, vector sequences; *solid bar*, rp29 cDNA; *hatched bar*, dC tailing; *dotted bar*, dT tailing. *Vertical arrows* marked *I* and *S* indicate the transcription initiation sites of the *ADC1* promoter and the virtual splice junction of the intron of rp29 cDNA, respectively. *Horizontal arrows* indicate the extent and directions of sequence determination from the indicated sites.

tatively that initiation occurs as indicated in Fig. 3. It should be noted that because of the very short sequence 5′ to the splice site, conventional S1 analysis of these transcripts is not feasible.

The sequences upstream have some interesting features. In

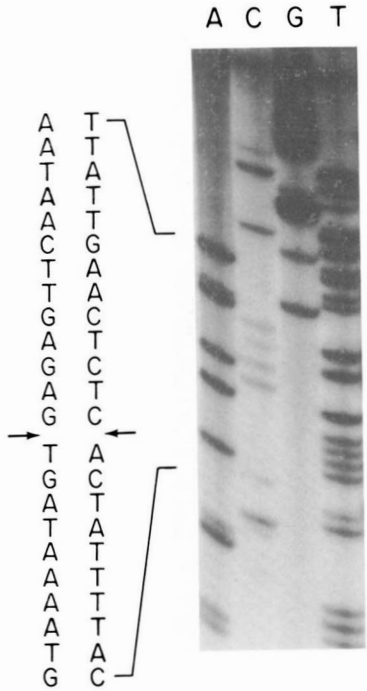

FIG. 7. **Sequencing gel showing the splice junction of the rp29 cDNA (Fig. 6 (*1*)).** Sequence is that of the antisense strand. Corresponding nucleotides of the sense strand are shown at *left*. The position of the splice is indicated by the *arrow*.

the rp29 gene TATA-like sequences are present 39, 59, and 82 bp upstream of the first transcriptional state site at positions −543, −563, and −586, respectively. These sequences resemble the "Hogness box" sequence that is usually present approximately 30 bp upstream of the capping site in most eucaryotic genes and is implicated in the eukaryotic RNA polymerase recognition site (23). Another sequence thought to be important for transcription initiation in eukaryotes is the CAAT sequence that has been found 80 bp upstream from the initiation of RNA synthesis (31). In the rp29 gene, TCAAT appears at position −579, 75 bp upstream from the first transcription start site.

In the 3′ flanking region of rp29 gene, the sequence TATGTA followed by TTT is present 20 nucleotides downstream from the TAA stop codon. This sequence conforms to a proposed consensus sequence TAG- - -TATGTA- - -TTT found in the 3′ nontranslated region of yeast genes, which has been implicated in efficient polyadenylation of mRNAs (32). It is not yet clear whether termination of transcription takes place at this site.

Previous experiments with the cloned rp29 gene (7) indicated that a transcript several hundred nucleotides larger than the homologous mRNA accumulated in cells unable to splice precursor mRNA due to a temperature-sensitive mutation in *rna*2 gene, when they were maintained at the restrictive temperature. These results suggested the presence of an intervening sequence in the rp29 gene. Sequence analysis of the rp29 gene and S1 nuclease mapping confirm these predictions and show that the gene contains a single intron of 458 bp, of which 64% are A-T. The sequences in the vicinity of the splice junctions were confirmed by sequencing the cDNA clone derived from rp29 mRNA. The hexanucleotide sequence GTATGT starting at −465 of the rp29 gene conforms to the 5′ donor splice sequence found in other sequenced yeast genes (22). The conserved septanucleotide sequence 'TACTAAC' is present within the introns of all nuclear protein-coding

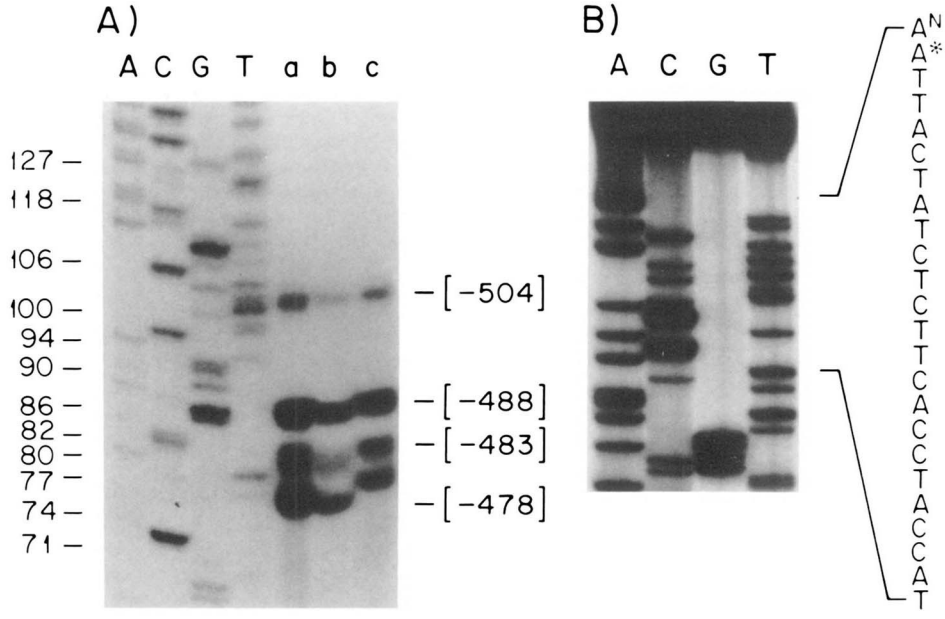

FIG. 8. **Mapping of the transcription units of rp29 mRNA.** *A*, primer extension analysis for mapping the 5′ ends of rp29 mRNA. A 37-bp *Kpn*I-*Cla*I fragment (Fig. 2), 5′-end labeled at the *Kpn*I site, was hybridized to poly(A$^+$) RNA from yeast strain *ts*368 grown at 23 °C. The resulting hybrid was extended using 500 units/ml of reverse transcriptase at 41 °C for various time periods. The radioactively labeled cDNA products of the primer extension experiment are shown in *lanes a*, *b*, and *c*. *a*, 30 min; *b*, 1 h; *c*, 3 h. The A, C, G, and T ladder of a known sequence was used to measure the length of the extension products. *B*, sequence of the 3′ end of rp29 cDNA. A *Taq*I/*Taq*I fragment of the rp29 cDNA (Fig. 6 (2)) was cloned into M13 mp8, and the sequence was determined by dideoxy chain termination method. The end of the mRNA site is marked by an *asterisk*, beyond which is a stretch of A residues corresponding to the poly(A) tail of the mRNA.

yeast genes between 20 to 55 bp upstream from 3′ acceptor splice site and has been suggested to be an essential element for splicing of yeast genes (22). In the rp29 gene, this sequence appears 43 nucleotides upstream from the 3′ splice site. One of the interesting features of the yeast split coding genes is the location of the splice site near the AUG initiation codon. For the ribosomal protein genes and also for yeast actin, the intron separates a small 5′ exon from a 3′ exon, containing the remainder of the coding sequence. On the other hand, in the rp29 gene, the intron is within the 5′ untranslated region. This is the first report of an intron in the leader region of a yeast transcript.

The occurrence of introns in the leader regions has been observed for some other eukaryotic genes, *e.g.* two rat preproinsulin genes, chicken ovalbumin, silk fibroin, etc. (reviewed in Ref. 33). The 5′ noncoding regions of many eukaryotic mRNAs have sequences complementary to that of the 3′ end of 18 S ribosomal RNA (34). Such interactions might contribute to ribosome binding and to the efficiency of translation. Hence it has been suggested that in some cases a splice in the leader is necessary to move such sequences near the initiation codon AUG (35). In the 5′ noncoding region of rp29 mRNA the nucleotides at positions −484 to −487 can form a stable complex with the 3′ end of 18 S rRNA with a free energy of −9.0 kcal/mol.

Sequence determination has been performed for a large number of nuclear genes isolated from the yeast *S. cerevisiae*. Of the nonribosomal protein genes only actin (36, 37) and MATa1 (38) have intervening sequences. On the other hand, 12 out of 16 ribosomal protein genes show accumulation of larger transcripts in an *rna2* mutant at the restrictive temperature, suggesting the presence of introns within the genes. For S10, L29, rp51, rp59, and now rp29, DNA sequence analysis has confirmed the existence of introns. Electron microscopic analysis of R-loops has suggested the presence of introns in several additional ribosomal protein genes of *Saccharomyces carlsbergensis* (39). Why, then, do the majority of the yeast ribosomal protein genes contain introns, which are generally so rare in the yeast genome? Why are the introns so near the 5′ end of the transcript? In the case of rp29 the 5′ exon is only 13 to 39 nucleotides long. It is tempting to speculate that these short exons are involved in the coordinate regulation of the synthesis of ribosomal proteins (3).

Acknowledgments—We are grateful to Dr. G. McKnight for the cDNA clone bank, to Dr. D. Shields for his help in determining the N-terminal amino acid sequence, to W. F. Schwindinger for use of his computer programs for DNA sequence analysis, and to Dr. J. Marmur for fruitful discussions.

REFERENCES

1. Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L., and Nomura, M. (eds) (1979) *Ribosomes: Structure Function, and Genetics*, University Park Press, Baltimore, MD
2. Lindahl, L., and Zengel, J. M. (1982) *Adv. Genet.* **21,** 53–121
3. Fried, H. M., and Warner, J. R. (1984) in *Recombinant DNA and Cell Proliferation* (Stein, J., and Stein, G., eds) pp. 169–192, Academic Press, New York
4. Woolford, J. L., Jr., Hereford, L. M., and Rosbash, M. (1979) *Cell* **18,** 1247–1259
5. Bollen, G. H. P. M., Cohen, L. H., Mager, W. H., Klaassen, A. W., and Planta, R. J. (1981) *Gene* **14,** 279–287
6. Fried, H. M., and Warner, J. R. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 238–242
7. Fried, H. M., Pearson, N. J., Kim, C. H., and Warner, J. R. (1981) *J. Biol. Chem.* **256,** 10176–10183
8. Hartwell, L. H., Mclaughlin, C. S., and Warner, J. R. (1970) *Mol. Gen. Genet.* **109,** 42–56
9. McKnight, G. L., and McConaughy, B. L. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 4412–4416
10. Messing, J. (1982) in *Genetic Engineering/Principles and Methods*

(Setlow, J. K., and Hollaender, A., eds) Vol. 4, pp. 19–34, Plenum Press, New York
11. Kim, C. H., and Warner, J. R. (1983) *Mol. Cell. Biol.* **3,** 457–465
12. Grunstein, M., and Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U. S. A.* **72,** 3961–3965
13. Warner, J. R., and Gorenstein, C. (1977) *Cell* **11,** 201–212
14. Maniatis, T., Fritsch, E. F., and Sambrook, J. (eds) (1982) *Molecular Cloning A Laboratory Manual* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
15. Smith, H. O., and Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3,** 2387–2398
16. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
17. Berk, A. J., and Sharp, P. A. (1977) *Cell* **12,** 721–732
18. Treisman, R., Proudfoot, N. J., Shander, M., and Maniatis, T. (1982) *Cell* **29,** 903–911
19. Shields, D., and Blobel, G. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 2059–2063
20. Rosbash, M., Harris, P. K. W., Woolford, J., and Teem, J. L. (1981) *Cell* **24,** 679–686
21. Bennetzen, J. L., and Hall, B. D. (1982) *J. Biol. Chem.* **257,** 3026–3031
22. Langford, C. J., and Gallwitz, D. (1983) *Cell* **33,** 519–527
23. Breathnach, R., and Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383
24. Warner, J. R. (1982) in *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression* (Strathern, J., Jones, E., and Broach, J. R., eds) pp. 525–560, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
25. Kaufer, N. K., Fried, H. M., Schwindinger, W. F., Jasin, M., and Warner, J. R. (1983) *Nucleic Acids Res.* **11,** 3123–3135
26. Schultz, L. D., and Friesen, J. D. (1983) *J. Bacteriol.* **156,** 8–14
27. Wallis, J. W., Hereford, L., and Grunstein, M. (1980) *Cell* **22,** 799–805
28. Teem, J. L., and Rosbash, M. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 4403–4407
29. Leer, R. J., van Raamsdonk-Duin, M. M. C., Molenaar, C. M. T., Cohen, L. H., Mager, W. H., and Planta, R. J. (1982) *Nucleic Acids Res.* **10,** 5869–5878
30. Faye, G., Leung, D. W., Tatchell, K., Hall, B. D., and Smith, M. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 2258–2262
31. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) *Nucleic Acids Res.* **8,** 127–142
32. Zaret, K. S., and Sherman, F. (1982) *Cell* **28,** 563–573
33. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872
34. Hagenbüchle, O., Santer, M., Steitz, J. A. and Mans, R. J. (1978) *Cell* **13,** 551–563
35. Gannon, F., O'Hare, K., Perrin, F., LePennec, J. P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Gaparin, A., Cami, B., and Chambon, P. (1979) *Nature (Lond.)* **278,** 428–434
36. Ng, R., and Abelson, J. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 3912–3916
37. Gallwitz, D., and Sures, I. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 2546–2550
38. Miller, A. M. (1984) *EMBO J.* **3,** 1061–1065
39. Bollen, G. H. P. M., Molenaar, C. M. T., Cohen, L. H., van Raamsdonk-Duin, M. M. C., Mager, W. H., and Planta, R. J. (1982) *Gene* **18,** 29–37