Proc. Natl. Acad. Sci. USA
Vol. 77, No. 7, pp. 3912–3916, July 1980
Biochemistry

Isolation and sequence of the gene for actin in *Saccharomyces cerevisiae*

(DNA sequence/intervening sequence/cloning)

RAY NG AND JOHN ABELSON

Department of Chemistry, University of California at San Diego, La Jolla, California 92093

Communicated by Norman Davidson, April 11, 1980

ABSTRACT **The yeast *Saccharomyces cerevisiae* is known to contain the highly conserved and ubiquitous protein actin. We have used cloned actin sequences from *Dictyostelium discoideum* to identify and clone the actin gene in yeast. Hybridization to genomic fragments of yeast DNA suggest that there is a single actin gene in yeast. We have determined the nucleotide sequence of that gene and its flanking regions. The sequence of the gene reveals an intervening sequence of 309 base pairs in the coding sequences at the 5′ end of the gene. The existence and location of the intervening sequence was verified by using the dideoxy chain termination technique to determine the sequence at the 5′ terminus of the actin mRNA. The similarity of the splice junction sequences in this gene to those found in higher eukaryotes suggests that yeast must possess a similar splicing enzyme.**

Actin, a ubiquitous protein in higher organisms, plays an essential role in cell motility and structure. Although the functions of actin are diverse, ranging from muscle contraction to chromosome movement, the structure of the 42,000-dalton protein is highly conserved, both among the various types of actins found in higher organisms and in evolution (1, 2). The conservation of size and properties of actin has made it possible to identify actin in a number of lower eukaryotes. For example, the actin of *Dictyostelium discoideum* has been studied extensively (3, 4); the amino acid sequence of *Physarum polycephalum* actin is known (5). Recently, two groups have isolated actin in baker's yeast, *Saccharomyces cerevisiae* (ref. 6; R. Scheckman, personal communication).

Because yeast is nonmotile, actin most likely functions in cytoskeletal integrity, in chromosome condensation, or in the process of budding (2, 7, 8). The highly developed genetics and molecular biology of yeast make this a promising system for studying the genetics of actin and the control of its expression.

We have taken advantage of actin's highly conserved amino acid sequence to characterize and isolate the yeast actin gene. The *D. discoideum* actin gene hybridizes with a single *Eco*RI fragment in yeast, suggesting that yeast contains only one actin gene. We have cloned this gene and determined its nucleotide sequence. Comparative studies with the nucleotide sequence of yeast actin mRNA establish the existence of a 309-base-pair (bp) intervening sequence at the 5′ end of the gene—the first chromosomal gene in yeast found to have an intervening sequence (9). Yeast tRNA genes contain intervening sequences (10, 11), and recent studies indicate that intervening sequences also exist in the yeast mitochondrial genes (12, 13). However, several yeast chromosomal genes that do not have intervening sequences have been characterized (14–16).

The splice junction sequences found in the yeast actin gene bear a strong resemblance to those found in the genes of other eukaryotes (17, 18). This suggests that yeast contains an RNA splicing mechanism like that found in higher organisms. In comparison, there are approximately 17 actin genes in *D. discoideum* which apparently do not have intervening sequences (R. A. Firtel, personal communication). On the other hand, recent findings show that *Drosophila melanogaster* has six actin genes and that the two genes examined contain an intervening sequence at a similar position near the 5′ end of the gene (ref. 19; N. Davidson, personal communication).

MATERIALS AND METHODS

Construction of Plasmids Carrying the Yeast Actin Gene. To detect colonies bearing the yeast actin gene, we made use of the fact that actin is a highly conserved protein. An actin gene from *D. discoideum* had already been cloned by R. A. Firtel and his colleagues and was made available to us (4). Radioactive cRNA from the actin fragment of the plasmid pcDd actin B was used to screen a collection of colonies bearing yeast *Hin*dIII fragments inserted into the plasmid pBR313 (20). One of the colonies was positive, and the plasmid isolated from this clone is designated "pYact II." This plasmid proved to contain only a part of the yeast actin gene. Consequently, a yeast clone bank, consisting of 5000 colonies, in which yeast *Eco*RI DNA fragments had been inserted into the plasmid pBR322 was constructed. This collection of colonies was screened by the method of Grunstein and Hogness (21) with a radioactive cRNA produced by *in vitro* transcription of the 1.6-kilobase (kb) *Hin*dIII fragment from pYact II. One of the colonies in this collection hybridized and harbored a plasmid carrying a 3.8-kb *Eco*RI insert which contained the entire yeast actin gene. This plasmid is designated "pYact I."

DNA Preparation. Plasmid DNA was isolated by using a modification of the Clewell and Helinski procedure (22).

Yeast DNA was isolated from *S. cerevisiae* (2180-1A). Cells were grown at 30°C to A_{660} of 2–3 in YPD medium, and the DNA was extracted and purified by the method of Sherman and Fink (23), with Hoescht dye 33258 in place of 4′, 6-diamidine-2-phenylindole.

RNA Preparation. Total yeast RNA was isolated by using a modification of the Hereford and Rosbach procedure (24). Poly(A)-containing RNA was purified by chromatography of total RNA on poly(U)-Sepharose (4).

Nick Translation. DNA was made radioactive by the nick translation procedure using DNA polymerase I and [α-^{32}P]-deoxynucleoside triphosphate as described (25).

DNA Sequence Determination. DNA was analyzed according to the method of Maxam and Gilbert (26). DNA copies of actin mRNA were analyzed by the method of Sures *et al.* (27) which was a modification of the Zimmern and Kaesberg procedure (28).

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.

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

RESULTS

Yeast Contains a Single Actin Gene. In order to identify and quantitate the actin gene(s) in yeast, total yeast DNA was digested with either *Hin*dIII or *Eco*RI. The fragments were separated by electrophoresis in a 1.2% agarose gel. The DNA was denatured *in situ* and the fragments were transferred to nitrocellulose filter paper by using the methods developed by Southern (29). Fragments containing the actin sequence were identified by hybridization at 60°C to a nick translated, actin-specific probe from *D. discoideum*. Two *Hin*dIII fragments (approximate sizes, 3.7 and 2.5 kb) and a single *Eco*RI fragment (3.8 kb) hybridized with the actin probe. Additional experiments by Norma Neff (personal communication) showed that yeast *Bam*HI and *Pst* I fragments hybridized to pYact I DNA, in agreement with our restriction map. As a result, it appears that yeast has a single actin gene, although the unlikely possibility of an identical repeat of the large DNA segment containing the actin gene cannot be ruled out.

In our first attempts to clone the yeast actin gene, a collection of colonies containing yeast *Hin*dIII fragments inserted into the plasmid pBR313 were screened with the *D. discoideum* probe. This collection contained the plasmid pYact II with the 2.5-kb *Hin*dIII fragment. Subsequently, an *Eco*RI library was constructed by ligation of yeast *Eco*RI fragments into the vector pBR322. We isolated a plasmid, pYact I, which contained the 3.8-kb *Eco*RI fragment. This fragment hybridized to the actin probe and was shown by DNA sequence analysis to contain the complete actin gene. Restriction endonuclease maps of the plasmids pYact I and pYact II are shown in Fig. 1.

Proof Through DNA Sequence Analysis that pYact I Contains the Actin Gene. The restriction endonuclease analysis of total yeast DNA suggested that there is a single actin gene which can be separated into two fragments by digestion with *Hin*dIII. DNA sequence analysis at the *Hin*dIII site in pYact I proved that this site is, in fact, located in the actin coding sequence. The DNA sequence in this region predicted an amino acid sequence closely homologous with the amino acid sequence of *Physarum* actin and with vertebrate actins as well (5, 31). The *Hin*dIII site was located at amino acid position 255. Further work led to a nearly complete sequence of the actin gene (Fig. 2).

An Intervening Sequence Separates the First Three Amino Acid Residues from the Rest of the Actin Coding Sequence. Several lines of evidence from the DNA sequence of the yeast actin gene suggested that an intervening sequence occurs in the gene between the third and fourth codons. The DNA sequence of the actin gene predicted an amino acid sequence similar to the *Physarum* sequence for most of the 374 residues (5). This similarity did not exist, however, at the NH_2 terminus of the coding sequence and, most important, a methonine initiator codon in phase with the rest of the coding sequence was lacking. The DNA sequence predicted an NH_2-terminal sequence of Ile-Cys-Leu, an amino acid sequence radically different from that of all known actins (31). Further analysis revealed a sequence 309 bp upstream from this region that contained a methonine initiator codon followed by the codons for Asp-Ser-Gly, an NH_2-terminal sequence compatible with other actins. Although the sequences surrounding this proposed intervening sequence were homologous with the sequences at the splice junctions in higher eukaryotes, the existence of a 309-bp intervening sequence could only be considered plausible from this evidence. To define the limits of the intervening sequence requires a comparison of the gene and mRNA sequences.

Nucleotide Sequence of Actin mRNA in the Region of the Splice Junction Confirms the Existence of the Intervening Sequence. To determine the sequence of the region near the putative splice junction, we used the dideoxy chain-terminating primer-extension technique described by several groups (27, 32). This procedure obviates the need to purify the actin mRNA or to clone its cDNA. A 75-bp *Hpa* II/*Hin*fI fragment (amino acid positions 22–50) close to the putative splice junction was specifically hybridized to the actin message in poly(A)RNA. Reverse transcriptase was then used to extend the primer toward the 5′ end of the mRNA. If an intervening sequence indeed existed, one would expect the mRNA sequence to suddenly deviate from the DNA sequence at the splice point. This is exactly what we found. Using the known DNA sequence for comparison, we were able to deduce a vital portion of the mRNA sequence, despite the appearance of some extraneous bands (Fig. 3). This sequence is colinear with the gene until codon 3. Then, after a hiatus of 309 nucleotides, the proposed

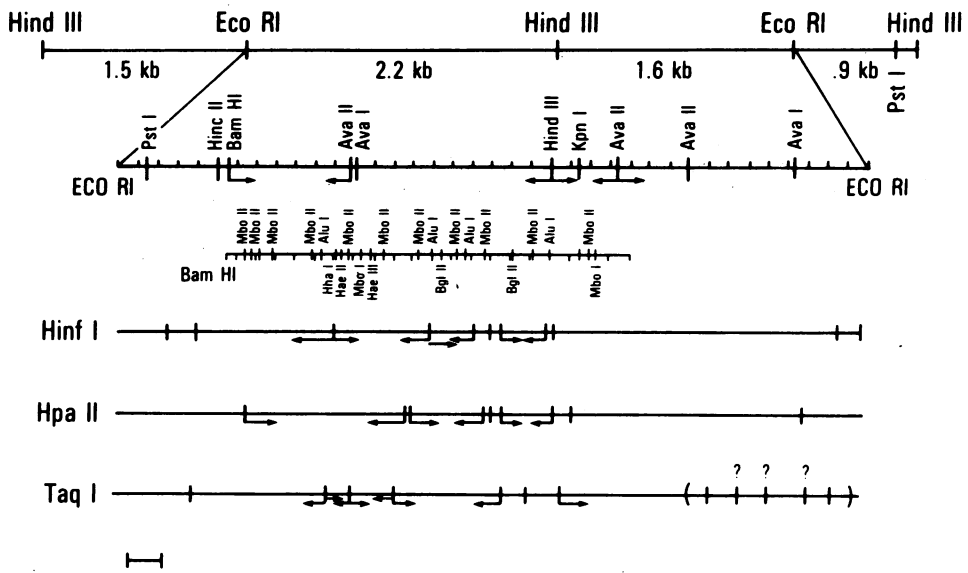

FIG. 1. Genomic organization of the *Eco*RI and *Hin*dIII sites based on hybridization experiments is shown on the top line. The 3.8-kb *Eco*RI fragment was mapped by the method of Smith and Birnstiel (30). The locations of some of the less-common restriction endonuclease sites appear on the second line; a map of sites derived by DNA sequence determination is given below it. A detailed restriction map for the enzymes *Hpa* I, *Hin*fI, and *Taq* I is shown. Arrows indicate the direction and extent of sequence analysis at each site. Scale represents 0.1 kb.

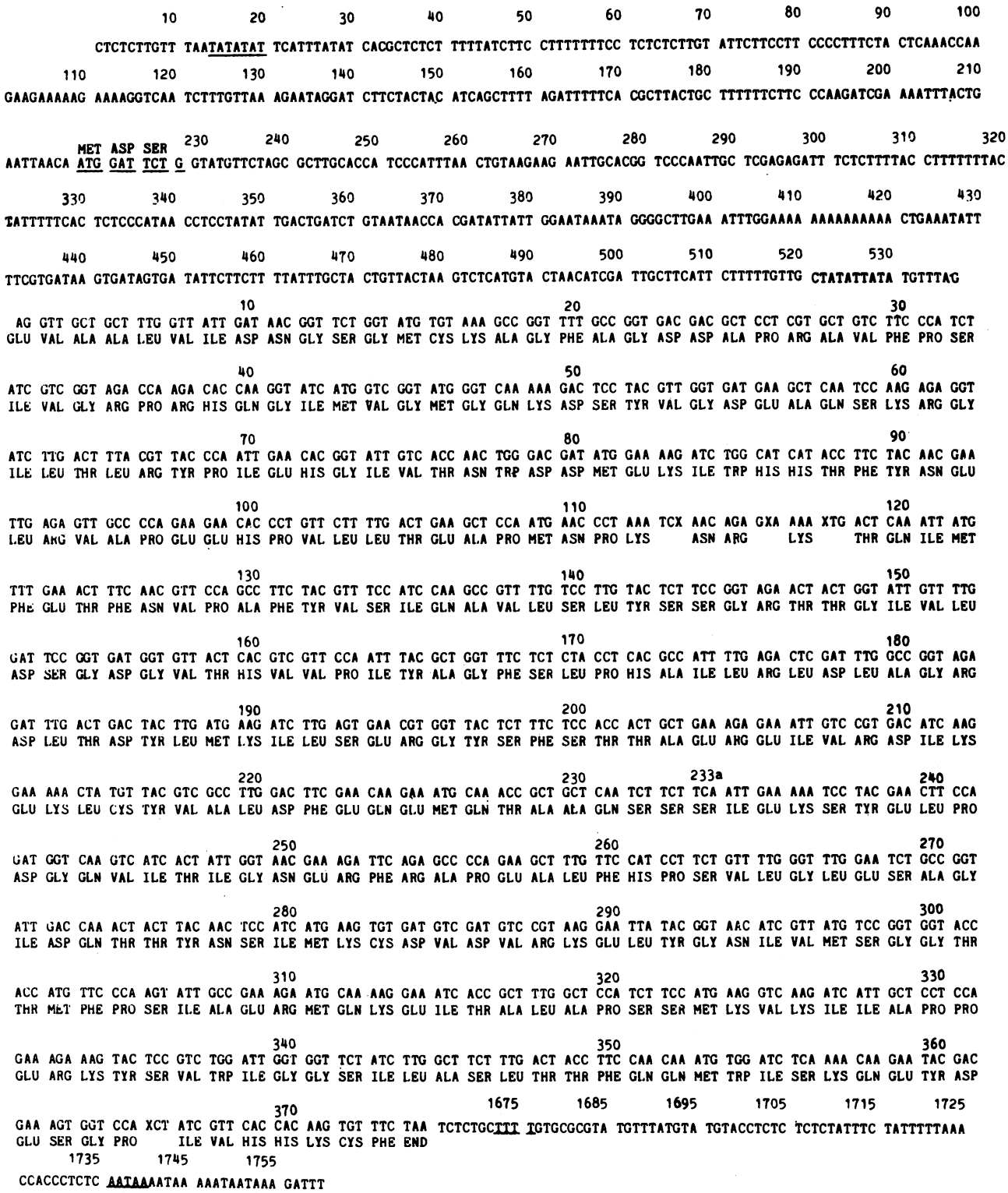

FIG. 2. DNA sequence of yeast actin gene and amino acid sequence of the actin. The first two codons (for Asp and Ser) are separated from the main body of the gene by a 309-bp intervening sequence. Uncertainties exist at codons 113, 116, 118, and 366 (indicated by "X"). Approximately 200 bp upstream from the initiation codon is the putative promoter site, T-A-T-A-T-A-T.

intron region, the sequence again becomes colinear. The sequence of the mRNA definitely establishes the existence of a 309-bp intervening sequence in the yeast actin gene. The interruption occurs between codons 2 and 3, giving a predicted NH_2-terminal sequence for yeast actin of Asp-Ser-Glu-Val-Ala.

Possible Transcription and Translation Signals in the Actin Gene. Comparison of the yeast actin gene sequences to the sequences of other eukaryotic genes has allowed the identification of several transcriptional and translational control sequences. If it is assumed that the longest DNA copies in the dideoxy primer-extension experiment are full length, then transcription starts approximately 140 ± 30 bp away from the iniatiation codon (unpublished data). A potential promoter recognition sequence, T-A-T-A-T-A-T, similar to that seen in other eukaryotic genes, is found in a region 30–50 bp upstream from

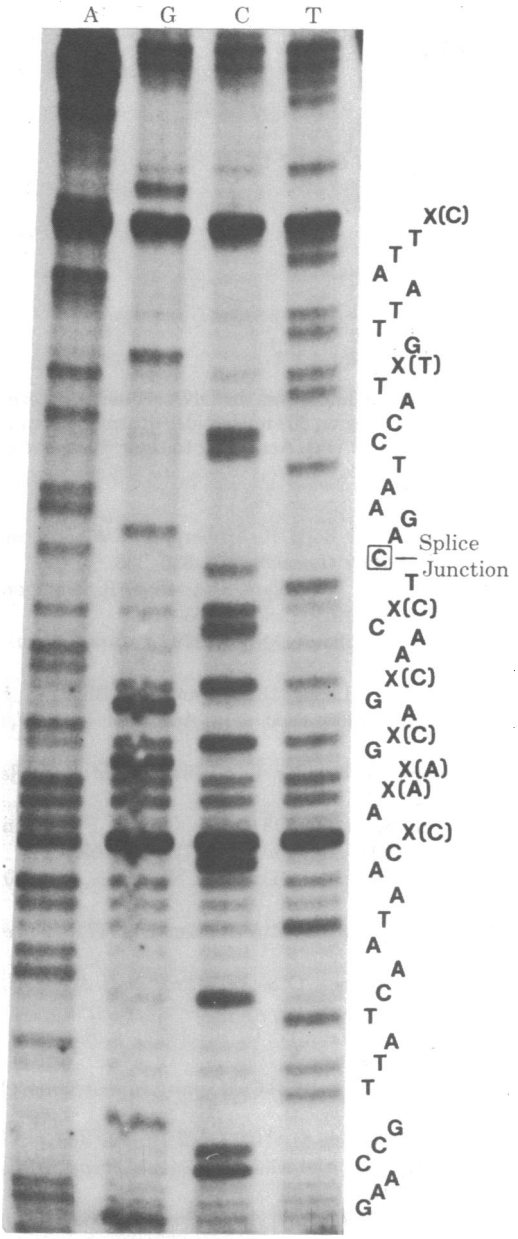

FIG. 3. The sequence of the mRNA was determined using the dideoxy primer-extension technique as reported by Sures *et al.* (27). Poly(A) RNA (125 μg) was hybridized to 5 pmol of a 75-bp strand-separated *Hpa* II/*Hin*fI fragment (codons 22–25). Reverse transcriptase was then used to extend the primer, and the DNA products were electrophoresed on a thin sequencing gel (33). Using the known DNA sequence as a guide, we were able to determine that the mRNA sequence was colinear with the gene, except that it lacked the sequence of 309 nucleotides immediately following codon 3. The location of the splice junction is indicated. Xs indicate regions of uncertainties. The letter inside the parentheses is the expected nucleotide.

this site (34) (Fig. 2). The actin gene also uses an A-U-G-G translation start sequence that is found in most higher eukaryotic initiation sites (15, 35). A pentanucleotide—A-A-T-A-A—found near the end of the 3′ untranslated region of all eukaryotic mRNAs (36, 37) also appears 70 nucleotides after the termination codon. Poly(A) addition occurs 18 bases after this site in rabbit, human, and mouse β-globin mRNA (36). Further experimentation is needed to prove that the regions discussed have real roles in transcriptional or translational control.

DISCUSSION

In this paper we report the isolation, identification, and nucleotide sequence of the actin gene from yeast. The nucleotide sequence establishes (with a few remaining uncertainties) the amino acid sequence of yeast actin. The NH_2-terminal sequence, Asp-Ser-Glu-Val-Ala, is one amino acid shorter than that in *Physarum*, *Dictyostelium*, and rabbit skeletal muscle (α) actins but is the same length as in bovine brain (β and γ) and chicken gizzard (γ) actins (5, 31). The amino acid sequence differs from that of *Physarum* in 43 positions, for an 11.6% variation. By contrast, *Physarum* actin differs from mammalian cytoplasmic actin in only 12 residues, a variation of 3.2% (5). One of the obvious contributions of the new DNA sequence-determination technologies will be the detailed study of the changes of gene sequence in evolution. The easy access to actin sequences due to the conserved homologies should make the actin genes among the most studied in this regard.

The sequence of the yeast actin gene reveals an extreme bias in codon usage similar to that found in other yeast genes whose sequences have been determined (14–16). We note, for example, that the 28 glycine residues are all coded for by GGU. There is a similar unique selection of codons for cysteine (UGU), asparagine (AAC), glutamine (CAA), and tyrosine (UAC). Similar cases of extreme preference in the use of codons is seen for several other amino acids. The significance of this bias is not clear, but it could have practical importance in the design of synthetic DNA sequences for use in the isolation of specific yeast genes (14).

The yeast actin intron resembles other intervening sequences found in eukaryotic organisms; it is extremely A+T rich (68% A+T), and a pyrimidine tract exists near the 3′ end. The splice junction sequences are similar to those found at other exon–intron regions and conform to rules postulated by several groups (17, 18). Seif *et al.* (17) found that the 5′ junction of most eukaryotic and viral genes are Pu↓G-T-X-X-G, and that the 3′ side has a Py-Py-X-Py-A-G↓ sequence. They also reported that no dinucleotide A-G appears within 13 nucleotides of the terminal A-G of the 3′ end of the intervening sequence. The yeast actin junction sequences are as follows: 5′ G↓G-T-A-T-G ··· G-T-T-T-A-G↓A-G 3′. Because of a terminal repeat at the splice junction, we cannot unambiguously assign the origin of the G in the mRNA (Fig. 4). However, according to the above rules, the G probably comes from the 5′ side of the intron.

In comparing the actin genes between species, it is interesting to observe that *Dictyostelium* contains many more actin genes (17) than does yeast, but the genes most likely do not contain intervening sequences. *Drosophila* has six actin genes, and the two genes examined have an intervening sequence (19). The *Drosophila* intervening sequence is larger than that of yeast but has a similar location near the 5′ end of the gene (19). The highest variation among actin amino acid sequences occurs in two regions (31): between residues 2 and 18 and between residues 259 and 298. The location of the intervening sequence between codons 2 and 3 suggests that the intervening sequence could have a role in promoting functional variations in the actin sequence. This could have been through DNA recombination, a mechanism likely to have been operative in the evolutionary past, or through alternative modes of splicing, as in the case of the adenovirus late mRNAs (38) in the present. The latter mechanism cannot act at a high efficiency in vegetative cells because there appears to be one predominant yeast actin mRNA (Fig. 3).

In our laboratory we have been interested in the mechanism of RNA splicing. We have used the yeast tRNA system to study this reaction (39). Comparison of the splice junctions in tRNA genes with those in nuclear genes coding for proteins has dis-

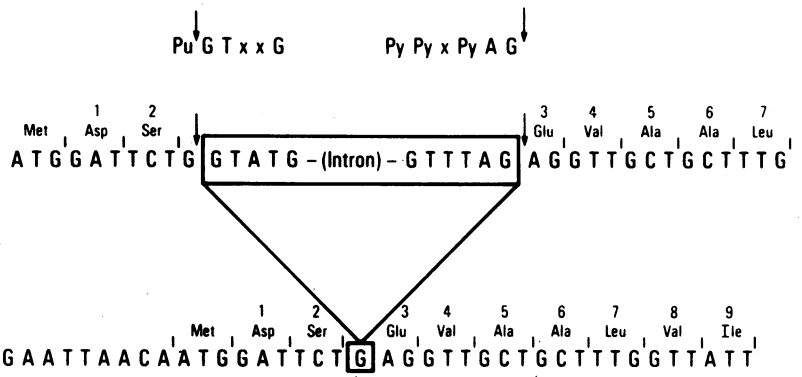

FIG. 4. The 5′ end of the actin gene with its intervening sequence is outlined above. The bottom line shows the DNA sequence determined from the mRNA. The mRNA sequence shows that the 5′ end of the actin message originates from a region 309 nucleotides upstream from codon 3. The yeast actin splice junction conforms to the canonical splice sequence postulated by Seif *et al.* (17). According to their rules, one would expect the terminally redundant G to come from the 5′ side of the intervening sequence.

couraged the possibility that the two different splicing reactions are carried out by the same enzyme system. The existence of an intervening sequence in the actin gene dictates that yeast must have a mRNA splicing system similar to that found in higher organisms. The yeast system may prove useful in investigation of the mechanism of mRNA splicing.

The results reported here open the way to two avenues of research. The actin gene can be used to study the regulation and expression of a split gene in yeast. The existence of a single actin gene offers the possibility for obtaining conditional lethal mutants of actin. Such mutants would provide us with an opportunity to study the function of actin in yeast.

The data presented in this paper establish the yeast actin gene as a nearly "classic" eukaryotic gene. In evidence are a typical promoter recognition site, an intervening sequence with canonical splice sequences, and a poly(A) addition site near the 3′ end of the gene. This evidence adds to a long list of features that establish yeast as an excellent model eukaryote.

We thank Dr. J. Beckmann, Dr. S. Kang, P. Johnson, J. Romeo, J. Rose, and L. Velten for their technical assistance. We are grateful to Dr. R. Firtel for providing us with an actin probe from *Dictyostelium* and to Dr. N. Davidson for unpublished data. We appreciate the efforts of M. Romashko in helping prepare this manuscript. Last, we appreciate the encouragement of the following people: Dr. H. Sakano, E. Smith, and especially, J. Velten and Dr. K. Janssen. This research was supported by Grant CA-10984 from the National Institutes of Health.

1. Pollard, T. D. & Weihing, R. R. (1974) in *Critical Reviews in Biochemistry*, ed. Fasman, G. (CRC, Cleveland, OH), Vol. 2, pp. 1–65.
2. Goldman, R., Polland, T. & Rosenbaum, J., eds. (1976) *Cell Motility* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Vols. 1–3.
3. Spudich, J. A. (1974) *J. Biol. Chem.* **249**, 6013–6020.
4. Kindle, K. L. & Firtel, R. A. (1978) *Cell* **15**, 763–778.
5. Vandekerckhove, J. & Weber, K. (1978) *Nature (London)* **276**, 720–721.
6. Koteliansky, V. E., Glukhova, M. A., Benjaman, M. V., Surguchov, A. P. & Smirnov, V. N. (1979) *FEBS Lett.* **102**, 55–58.
7. Rungger, D., Runggar-Brändle, E., Chaponnier, C. & Gabbiani, G. (1979) *Nature (London)* **282**, 320–321.
8. Lazarides, E. (1976) *J. Cell Biol.* **68**, 202–219.
9. Crick, F. (1979) *Science* **204**, 264–271.
10. Goodman, H. M., Olson, M. V. & Hall, B. D. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5453–5457.
11. Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R. & Rutter, W. J. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 190–194.
12. Church, G. M., Slonimski, P. P. & Gilbert, W. (1979) *Cell* **18**, 1209–1215.
13. Bos, J. L., Ting, C. H., Borst, P., Arnberg, A. C. & van Bruggen, E. F. J. (1978) *Nature (London)* **275**, 336–338.
14. Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. L. & Hall, B. D. (1979) *Cell* **16**, 753–761.
15. Holland, J. P. & Holland, M. J. (1979) *J. Biol. Chem.* **254**, 9839–9854.
16. Montgomery, D. L., Leung, D. W., Smith, M., Bhalit, P., Faye, G. & Hall, B. D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 541–545.
17. Seif, I., Khoury, G. & Dhar, R. (1979) *Nucleic Acids Res.* **6**, 3387–3398.
18. Breathnach, R., Benoist, C., O'Hara, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4853–4857.
19. Fryberg, E. A., Kindle, K. L., Davidson, N. & Sodja, A. (1980) *Cell* **19**, 365–378.
20. Beckman, J. S., Johnson, P. F. & Abelson, J. (1977) *Science* **196**, 205–208.
21. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
22. Clewell, D. B. & Helinski, D. R. (1972) *J. Bacteriol.* **110**, 1135–1146.
23. Sherman, F. & Fink, G. (1974) *Cold Spring Harbor Symp. Quant. Biol.* **20**, 77–80.
24. Herford, L. & Rosbach, M. (1977) *Cell* **10**, 453–462.
25. Maniatis, R., Jeffrey, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1184–1188.
26. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
27. Sures, I., Levy, S. & Kedes, L. H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1265–1269.
28. Zimmern, D. & Kaesberg, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4257–4261.
29. Southern, E. (1975) *J. Mol. Biol.* **98**, 503–508.
30. Smith, H. O. & Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3**, 2387–2398.
31. Vanderkerckhove, J. & Weber, K. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1106–1110.
32. Brownlee, G. G. & Cartwright, E. M. (1977) *J. Mol. Biol.* **114**, 93–117.
33. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107–110.
34. Baker, C. C., Herisse, J., Courtois, G., Galibert, F. & Ziff, E. (1979) *Cell* **18**, 569–580.
35. Kozak, M. & Shatkin, A. J. (1978) *Cell* **13**, 201–212.
36. Kondel, D. A., Tilghman, S. M. & Leder, P. (1978) *Cell* **15**, 1125–1132.
37. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
38. Berk, A. J. & Sharp, P. A. (1978) *Cell* **14**, 695–711.
39. Knapp, G., Beckmann, J. S., Johnson, P. F., Fuhrman, S. A. & Abelson, J. (1978) *Cell* **14**, 221–236.