Proc. Natl. Acad. Sci. USA
Vol. 77, No. 5, pp. 2546–2550, May 1980
Biochemistry

Structure of a split yeast gene: Complete nucleotide sequence of the actin gene in *Saccharomyces cerevisiae*

(intron/splice points/amino acid sequence)

D. GALLWITZ* AND I. SURES†

*Physiologisch-Chemisches Institut I, Universität Marburg, Lahnberge, D-3550 Marburg/Lahn, West Germany; and †Département de Biologie Animale, Université de Genève, CH-1224 Chêne-Bougeries, Switzerland

Communicated by Wolfgang Beermann, February 12, 1980

ABSTRACT **The complete nucleotide sequence of the actin gene from *Saccharomyces cerevisiae* has been determined. The coding region is interrupted by a 304-base-pair intervening sequence that is located within the triplet coding for amino acid 4. DNA sequences of the intron–exon junctions are similar to those found in higher eukaryotes and can be aligned such that the intron starts with the dinucleotide 5′-G-T-3′ and ends with 5′-A-G-3′. Regions of homology within the sequences upstream from the initiation codon and those following the termination codon have been detected between the yeast iso-1-cytochrome *c* gene and the actin gene. As deduced from the nucleotide sequence, yeast actin has 374 amino acid residues. Its primary structure, especially the NH_2-terminal third of the protein, is highly conserved during evolution.**

Actin is a major protein in all eukaryotic cells. It serves a number of functions and seems to be involved in such vital cellular processes as mitosis and cytokinesis (for review see refs. 1 and 2). It is well documented that in higher eukaryotic organisms several actin genes are expressed; the different actins, however, are similar in primary structure (3–6). Sequence comparison has also shown that the structure of actin proteins from different species is highly conserved during evolution. Actin from the lower eukaryote *Physarum polycephalum* differs from mammalian cytoplasmic actins in only 17 out of 375 amino acids (6).

Recently we have isolated by molecular cloning the actin gene from yeast, *Saccharomyces cerevisiae*. We found that there is very likely only one actin gene in this unicellular organism (7). Here we report the complete nucleotide sequence of the yeast actin gene and show that a chromosomal yeast gene coding for a protein is split. The sequence data suggest that the splicing mechanism in yeast might be similar to that of higher eukaryotes.

MATERIALS AND METHODS

Plasmids. A 5.1-kilobase (kb) *Pst* I DNA fragment from *S. cerevisiae* strain A364A (8) containing the actin gene was cloned with pBR322 as described (7). This plasmid, pYA208, was identified by using nick-translated DNA from clone pYA102 which harbors part of the yeast actin gene on a 3.93-kb *Hin*dIII DNA fragment and which was cloned in *Escherichia coli*, using *Dictyostelium* actin cDNA as hybridization probe (7). Restriction endonuclease sites were mapped (7, 9) and DNA fragments were purified by agarose or polyacrylamide gel electrophoresis in the Tris/borate/EDTA buffer system (10).

Nucleotide Sequence Analysis. DNA fragments were either 5′-end-labeled with 0.75 μM [γ-^{32}P]ATP (Amersham, 2000 Ci/mmol; 1 Ci = 3.7×10^{10} becquerels) and phage T4 polynucleotide kinase (11) or 3′-end-labeled with either [α-^{32}P]dCTP or [α-^{32}P]dGTP (New England Nuclear, 300–500 Ci/mmol) in 35 mM potassium phosphate, pH 7.5/3 mM $MgCl_2$/1 mM dithiothreitol/40 units of DNA polymerase I (Klenow fragment; Boehringer Mannheim) per ml (12). Nucleotide sequences were determined according to Maxam and Gilbert (13) with separated strands or with end-labeled fragments after secondary restriction enzyme cleavage. Labeled single- or double-stranded fragments were recovered from gel slices by electrophoresis into dialysis bags in 10 mM Tris–borate/0.2 mM EDTA, pH 8.3. Some DNA sequences were also determined by using the enzymatic method of Sanger *et al.* (14) as described (15).

Biohazard Considerations. Cloning experiments and growth of recombinant DNA plasmids were carried out under L2–B1 conditions following the rules of the Bundesminister für Forschung und Technologie of the Federal Republic of Germany. (L2–B1 equals P2–EK1 conditions according to the National Institutes of Health.)

RESULTS AND DISCUSSION

The Yeast Actin Gene is Split. As previously reported, we have cloned a 5.1-kb yeast DNA fragment (clone pYA208) containing the actin gene, using cloned *Dictyostelium* actin cDNA as hybridization probe (4, 7). Only part of this heterologous probe hybridized to yeast actin DNA. Nucleotide sequence analysis identified part of a 0.54-kb *Alu* I fragment from pYA208 DNA as the segment coding for amino acids 5–57 of actin (7). A restriction endonuclease map of the 5.1-kb *Pst* I fragment was established (7, 9), and the complete nucleotide sequence of the actin gene was determined as outlined in Figs. 1 and 2. As can be seen in Fig. 2, the coding region is interrupted by an intervening sequence. In front of amino acid residue 4 no initiation codon (ATG or GTG) was found in reading frame. This is illustrated in a sequencing gel shown in Fig. 3. Because actin from yeast, as isolated by affinity chromatography on DNase I/agarose (unpublished data) or by DEAE-cellulose chromatography (17) shows nearly identical behavior as mammalian actins in sodium dodecyl sulfate/polyacrylamide gels, its chain length is not expected to be significantly different from that of other actins. Therefore the actin coding region must be interrupted by an intron. Besides an ATG triplet that is out of reading frame and ends five nucleotides upstream from the GAG triplet coding for glutamic acid in position 4 (Figs. 2 and 3) there are two other ATG codons ending 49 and 299 base pairs upstream from the GAG codon. Both, however, are followed by termination codons. Twelve base pairs further upstream another ATG codon starts and is followed by a GAT triplet coding for aspartic acid. We believe that this ATG is the initiation codon of the actin gene for the following reasons: (*i*) All actin proteins whose amino acid sequences have been determined so far start with an acidic amino acid (5, 6, 18). (*ii*) The

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.

Abbreviation: kb, kilobase(s).

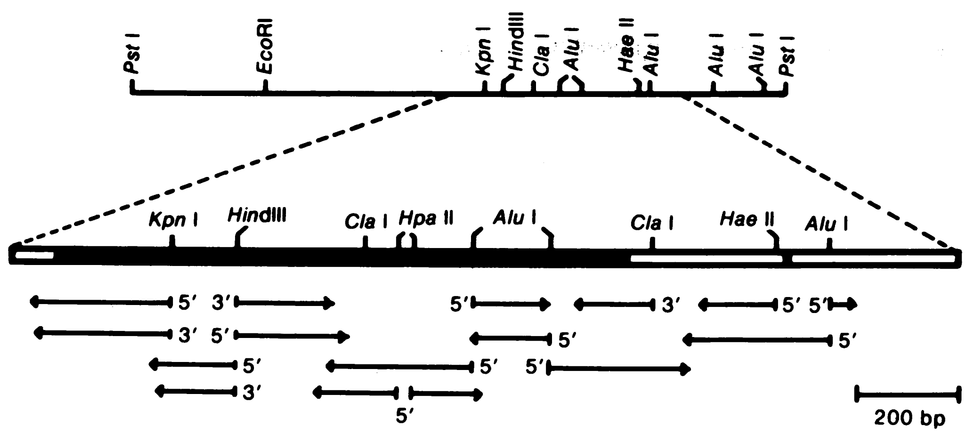


FIG. 1. Sequence determination strategy of the actin gene from *S. cerevisiae*. Cutting sites of restriction endonucleases used to fragment the 5.1-kb *Pst* I yeast DNA segment of pYA208 for sequence determination are indicated. The nucleotide sequences of 5′- or 3′-end-labeled restriction fragments were determined as shown by arrows. Thick line: coding region, which is interrupted by intron. Direction of transcription is from right to left. bp, Base pairs.

nucleotide sequences at the presumptive intron–exon junctions are similar to the "prototype sequence" found at intron–exon borders of other eukaryotic genes (19). The sequence can be aligned such that the intron of the actin gene starts with 5′-G-T-3′ and ends with 5′-A-G-3′. (*iii*) There is a striking sequence homology of 10 out of 12 nucleotides immediately

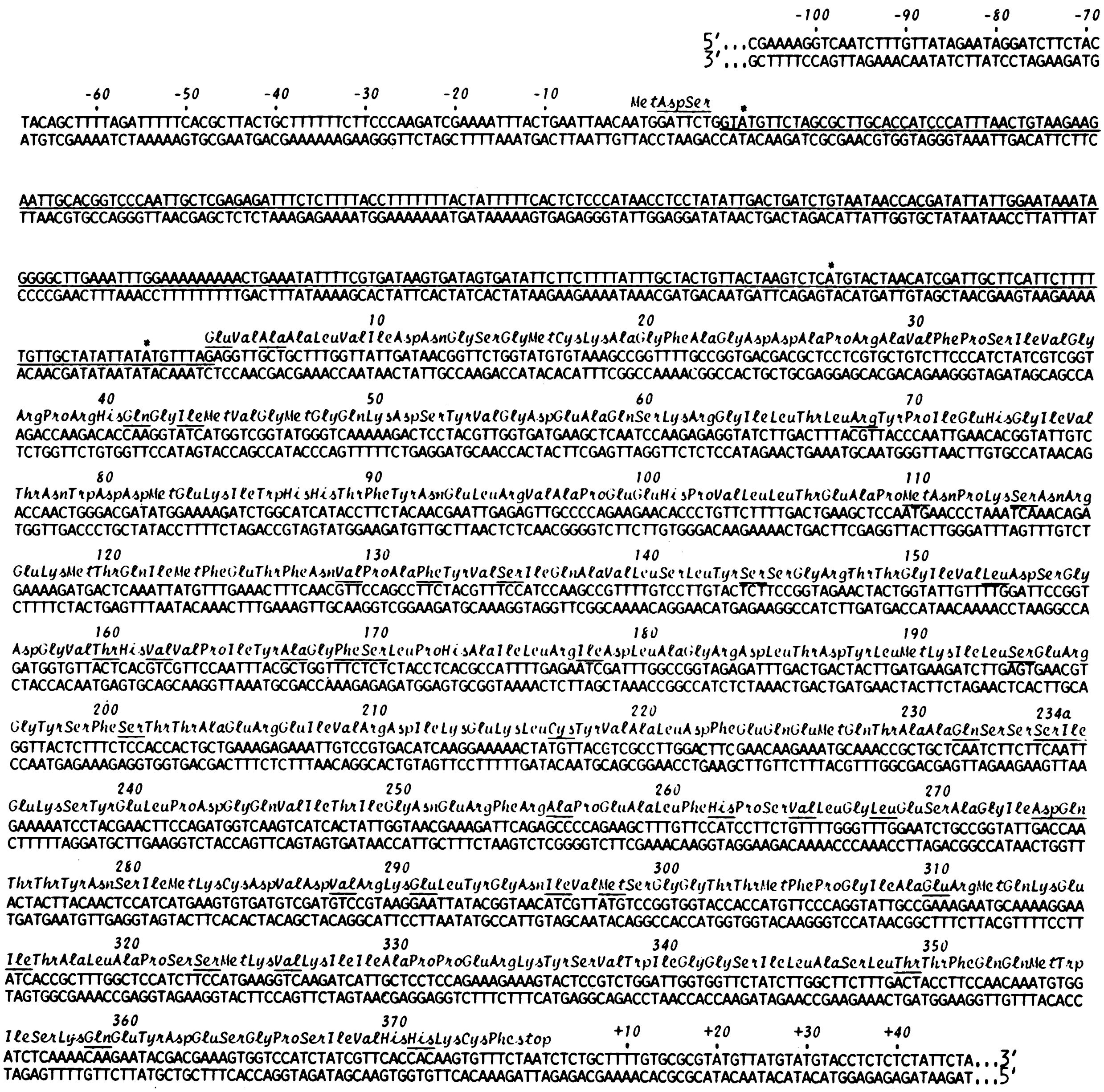


FIG. 2. Nucleotide sequence of the yeast actin gene. Amino acids of actin as deduced from the DNA sequence are numbered as proposed (6, 16). Amino acids underlined are different from the corresponding residues in *Physarum* actin. ATG codons within the 304-base-pair intron are marked by asterisks.

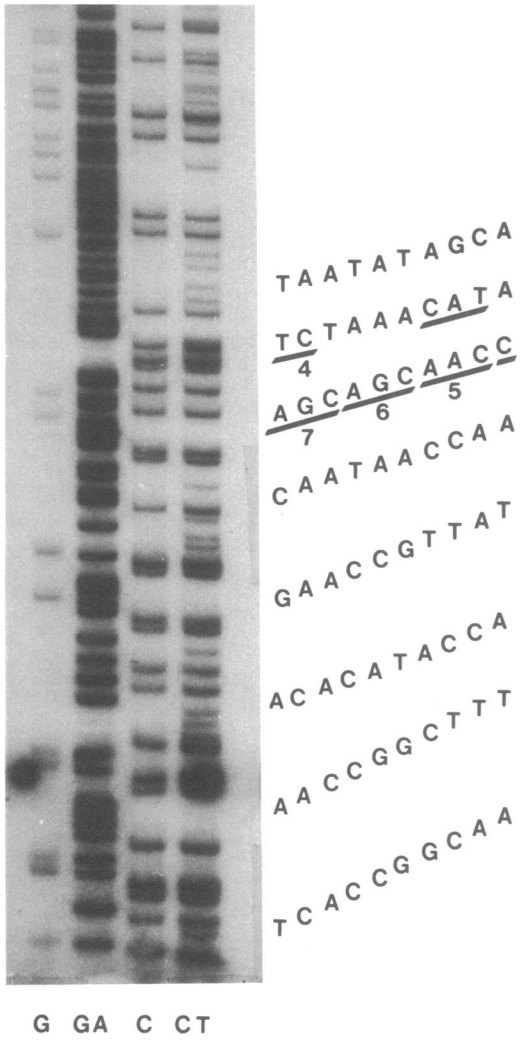


FIG. 3. Autoradiogram of an 8% polyacrylamide sequencing gel showing the intron–exon junction of the coding strand upstream from the amino acid in position 4. Base triplets complementary to those coding for amino acids 4–7 and to ATG codon out of reading frame are underlined.

upstream from the initiation codon between the yeast iso-1-cytochrome *c* gene and the actin gene (Fig. 5). (*iv*) A potential binding region for the 3′ end of 18S rRNA (20, 21) to actin mRNA exists at position −34 to −28. The sequence 5′-T-A-T-A-G-A-A-T-A-3′ (positions −88 to −80) resembles a "Pribnow box" structure (22), which was noted in the *Drosophila* histone genes (23) and is found in the 5′-flanking region of many eukaryotic genes, including the yeast iso-1-cytochrome *c* gene (15, 21, 24–26).

The proposed splicing points at the intron–exon junctions of the yeast actin gene are compared with the "prototype sequence" of Breathnach *et al.* (19) in Fig. 4. Although there is no sequence redundancy at the intron boundaries of the yeast actin gene, sequences surrounding the presumptive cutting sites

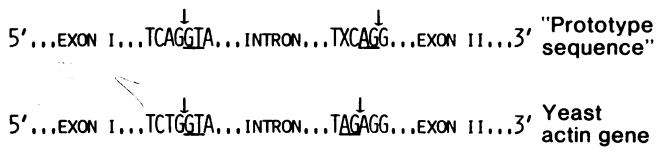
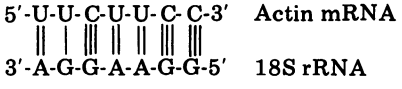


FIG. 4. Nucleotide sequence surrounding intron–exon junctions of yeast actin gene. For comparison the "prototype sequence" of genes from higher eukaryotes (19) is shown. Possible cutting sites are indicated by arrows.

are similar to those found in genes of higher eukaryotes. This suggests that a splicing mechanism similar to that of higher eukaryotes also exists in lower unicellular organisms such as yeast. A common enzymatic mechanism for excision and ligation therefore might have arisen early in eukaryotic evolution. If splicing results from bringing together DNA segments during evolution to create new genes, as proposed by Gilbert (27), one can speculate that an efficient transcription or translation initiation sequence was fused to the actin coding region. Because the expressed actin gene(s) from *Dictyostelium* seems to lack an intervening sequence (4), the possibility exists that it was lost during evolution. In fact, as judged from the amino acid sequence, *Dictyostelium* actin (K. Weber, personal communication) must be considered as a higher evolved protein than yeast actin. Structural comparison of different actin genes will show whether our assumptions are correct.

The intron of the actin gene is 304 base pairs long and is extremely A+T-rich (70% A+T). No other obvious structural features have been detected except for a 3-fold repetition of the hexanucleotide 5′-G-T-G-A-T-A-3′ starting at position 201 of the intron (Fig. 2).

Structure of the Flanking Regions. The nucleotide sequences flanking the yeast actin and iso-1-cytochrome *c* genes (26) are compared in Fig. 5*A*. As mentioned above, there is a striking homology of the 12 nucleotides immediately upstream from the initiation codon. At positions −34 to −28 and −40 to −34 of the actin and iso-1-cytochrome *c* genes, respectively, potential binding regions for the 3′ end of 18S rRNA to the mRNAs exist (20, 21). A stable complex



with a free energy of formation ΔG of −9.1 kcal/mol (1 kcal = 4.184 kJ) (28) can be formed. It can also be seen in Fig. 2 that there is no other ATG codon upstream from the initiation codon at least up to position −107. This is characteristic for many eukaryotic mRNAs (29). Another interesting feature in the presumptive 5′-flanking region of the actin gene is the presence of the sequence 5′-T-A-T-A-G-A-A-T-A-3′ (position −88 to −80). It resembles the "Hogness-box" (5′-T-A-T-A-A-A-T-A-3′), which starts roughly 30 nucleotides in front of the capping site and is implicated to function as part of the eukaryotic RNA polymerase recognition site (21, 23–25). The iso-1-cytochrome *c* gene from *S. cerevisiae* has a similar structure (5′-T-A-T-A-A-A-A-3′) beginning 123 nucleotides upstream from the initiation codon (26). We have estimated the length of the actin mRNA to be about 1250 nucleotides (7). The messenger therefore contains about 120–130 untranslated nucleotides and it is possible that one of the three purines at position −58 to −56 (Fig. 2) is its capping locus. The cap structure in yeast mRNA is either $m^7G(5')pppA$- or $m^7G(5')pppG$- (30, 31).

In the 3′-flanking region we note another similarity between the two yeast genes (Fig. 5*B*). The termination codons are followed by pyrimidine-rich tracts and a common heptanucleotide, 5′-G-T-T-A-T-G-T-3′, beginning 22 and 37 nucleotides after the stop codon of the actin and iso-1-cytochrome *c* gene, respectively.

Amino Acid Composition of Yeast Actin. Provided the proposed excision points of the intron are correct, yeast actin has 374 amino acids and equals in length the mammalian cytoplasmic β and γ actins, which are one residue shorter than the muscle-type α actin (6). Protein sequence data have revealed that the NH_2-terminal region of different actins is highly variable (5, 6, 18). This region is acidic, and the NH_2-terminal residue in all actins whose amino acid sequences have been

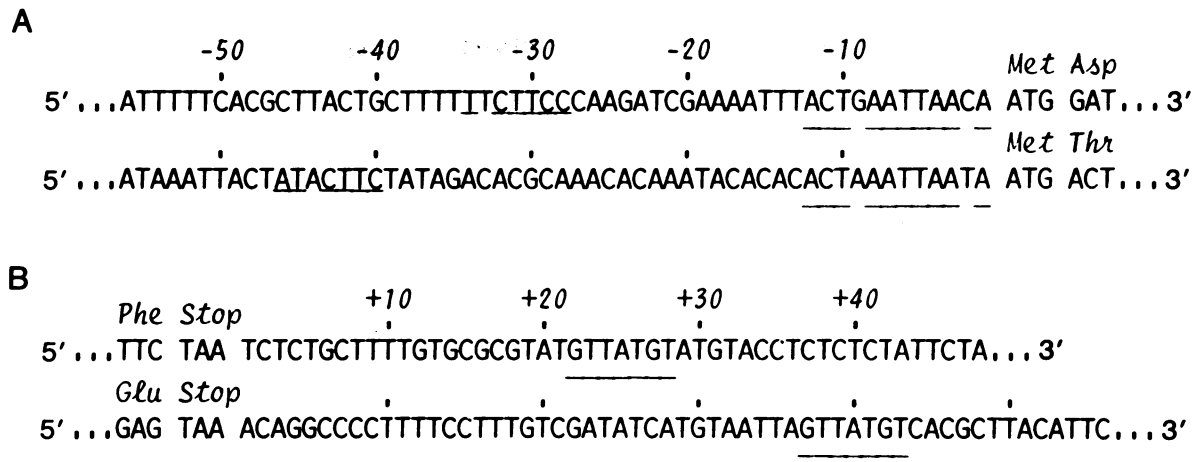


FIG. 5. Comparison of 5'- (*A*) and 3'- (*B*) flanking regions of yeast actin gene (upper sequences) and iso-1-cytochrome *c* gene (lower sequences) (26). Regions of homology immediately upstream from the initiation codons and following the termination codons as well as potential binding regions of the respective mRNAs to the 3' end of 18S rRNA are underlined.

determined so far is either aspartic acid or glutamic acid. Again, provided the initiation codon is assigned correctly, yeast actin has aspartic acid at its NH_2-terminus, followed by serine. Actins from *Physarum* (6) and *Dictyostelium* (K. Weber, personal communication) likewise start with an acidic residue that is followed by an amino acid with an uncharged polar group (glycine). In contrast, muscle and nonmuscle actins from higher eukaryotes have acidic amino acids in their first three or four positions (5, 6, 18). In Fig. 2 the amino acids of actin have been numbered as proposed by others (6, 16). Amino acids different from *Physarum* actin are underlined. There are 44 amino acid differences between the actins from these two lower eukaryotes. If one disregards the highly variable NH_2-terminal six amino acids, yeast actin differs from *Physarum* actin by 39 residues, from mammalian cytoplasmic γ actin (6) by 37 residues, and from rabbit skeletal muscle actin (18) by 44 residues. It is interesting to note that yeast actin differs from mammalian cytoplasmic actin to the same extent as from actins of the two other fungi, *Physarum* and *Dictyostelium*, whereas actins from the two slime molds exhibit a significantly higher degree of homology to mammalian cytoplasmic actins.

It is evident that the NH_2-terminal third of actin (except the first six residues) is highly conserved during evolution. In the region from amino acids 7 to 109 there are only three exchanges between yeast and *Physarum* actin and only two amino acid differences between yeast and cytoplasmic γ actin from vertebrates. The overall number of charged amino acids stays rather constant in all actins whose primary structures are known. In comparison to actins from *Physarum* and mammals, yeast actin, which might be considered as the archetype actin, displays eight changes in charged residues, in positions 68 (Arg → Lys), 167 (Ala → Glu), 262 (His → Gln), 274 (Asp → His), 275 (Gln → Glu), 291 and 310 (Glu → Asp), and 371 (His → Arg). There is a relative abundance of serine residues in yeast actin in comparison to all other actins whose sequences have been determined. In the structural development of actins from yeast to higher eukaryotes (including *Physarum* and *Dictyostelium*) there are four Ser → Ala changes (positions 114, 135, 144, and 170) and three Ser → Thr changes (positions 194, 201, and 323), in contrast to only one Gln → Ser change (position 232) and one Thr → Ser change (position 349).

It has recently been argued that the amino acid pattern at positions 17 (Val), 298 (Met), and 357 (Thr) is typical for skeletal muscle α actin and indicative for α actin's being the highest developed actin species (17). Interestingly, yeast actin has a Met in position 298 as well. The functional significance of the amino acid exchanges in different actins remains to be determined.

Note Added in Proof. The location and length of the intervening sequence has been confirmed by electron microscopy of hybrids between pYA208 DNA and yeast actin mRNA.

We thank Mrs. R. Seidel for excellent technical assistance. We are grateful to Dr. L. H. Kedes (Stanford) for his hospitality during the initial DNA sequence determination work. This work was supported by a grant from the Deutsche Forschungsgemeinschaft to D.G. (SFB 103, B7).

1. Goldman, R., Pollard, T. & Rosenbaum, J., eds. (1976) *Cell Motility*, Cold Spring Harbor Conferences on Cell Proliferation (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Vol. 3.
2. Korn, E. D. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 588–599.
3. Storti, R. V. & Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2346–2350.
4. Kindle, K. L. & Firtel, R. (1978) *Cell* **15**, 763–778.
5. Vandekerckhove, J. & Weber, K. (1978) *J. Mol. Biol.* **126**, 783–802.
6. Vandekerckhove, J. & Weber, K. (1978) *Nature (London)* **276**, 720–721.
7. Gallwitz, D. & Seidel, R. (1980) *Nucleic Acids Res.*, **8**, 1043–1059.
8. Hartwell, L. H. (1967) *J. Bacteriol.* **93**, 1662–1670.
9. Smith, H. O. & Birnstiel, M. (1976) *Nucleic Acids Res.* **3**, 2387–2398.
10. Maniatis, T., Jeffrey, A. & van de Sande, H. (1975) *Biochemistry* **14**, 3787–3794.
11. Richardson, C. (1965) *Proc. Natl. Acad. Sci. USA* **54**, 158–165.
12. Klenow, H., Overgaard-Hansen, K. & Patkar, S. A. (1971) *Eur. J. Biochem.* **22**, 371–381.
13. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
15. Sures, I., Levy, S. & Kedes, L. H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1265–1269.
16. Lu, R. C. & Elzinga, M. (1977) *Biochemistry* **16**, 5801–5806.
17. Koteliansky, V. E., Glukhova, M. A., Bejanian, M. V., Surguchov, A. P. & Smirnov, V. N. (1979) *FEBS Lett.* **102**, 55–58.
18. Vanderkerckhove, J. & Weber, K. (1979) *Differentiation* **14**, 123–133.
19. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4853–4857.
20. Hagenbüchle, O., Santer, M., Steitz, J. A. & Mans, R. J. (1978) *Cell* **13**, 551–563.
21. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. & Tizard, R. (1979) *Cell* **18**, 545–558.

22. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 784–788.
23. Goldberg, M. (1979) Dissertation (Stanford Univ., Stanford, CA).
24. Gannon, F., O'Hare, K., Perrin, F., LePennec, J. P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. & Chambon, P. (1979) *Nature (London)* **278,** 428–434.
25. Nishioka, Y. & Leder, P. (1979) *Cell* **18,** 875–882.
26. Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. L. & Hall, B. D. (1979) *Cell* **16,** 753–761.
27. Gilbert, W. (1978) *Nature (London)* **271,** 501.
28. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London) New Biol.* **246,** 40–41.
29. Kozak, M. (1978) *Cell* **15,** 1109–1123.
30. Scripati, C. E., Groner, Y. & Warner, J. R. (1976) *J. Biol. Chem.* **251,** 2898–2904.
31. DeKloet, S. R. & Andrean, A. G. (1976) *Biochim. Biophys. Acta* **425,** 401–408.