Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 2763–2767, May 1982
Biochemistry

Isolation and characterization of rat skeletal muscle and cytoplasmic actin genes

(α- and β-actin/Southern blot/R loop/S1 nuclease mapping/DNA sequence determination)

U. NUDEL*, D. KATCOFF*, R. ZAKUT*, M. SHANI*, Y. CARMON*, M. FINER*, H. CZOSNEK*, I. GINSBURG†, AND D. YAFFE*

Departments of *Cell Biology and †Neurobiology, The Weizmann Institute of Science, Rehovot 76100, Israel

Communicated by Norman Davidson, January 1, 1982

ABSTRACT **Southern blots of rat genomic DNA indicate the existence of at least 12 *Eco*RI DNA fragments containing actin gene sequences. By using specific probes and stringent conditions of hybridization, it was found that only one of these fragments contains sequences of the skeletal muscle α-actin gene. Recombinant bacteriophages originating from eight different actin genes were isolated from rat genomic DNA libraries. One of them, Act 15, contains the skeletal muscle actin gene. Another clone, Act 1, contains a gene coding for a cytoplasmic actin, identified tentatively as the β-actin gene. Both genes have a large intron very close to the 5′ end of their transcribed region, followed by several small introns. DNA sequence analysis and comparison with the available data on actin genes in other organisms indicated an interesting relationship between the positions of introns and the evolutionary relatedness. Several intron sites are conserved from at least the echinoderms to the vertebrates; others appear to be present in some actin genes and not in others.**

Actins are among the most prevalent proteins in eukaryotic cells. In addition to their major role in muscle contraction, actins are involved in all forms of cell and organelle motility. At least six different vertebrate actins have been identified by amino acid sequence analysis: skeletal muscle actin (α-actin) and cardiac muscle, stomach smooth muscle, aorta smooth muscle, and the "cytoplasmic" nonmuscle actins (β- and γ-actins) (1–8). The amino acid sequences of the different actins have been highly conserved during evolution. No differences have been found between the amino acid sequence of skeletal muscle actins of several mammals and the chicken (5, 6). Even actins from remote evolutionary sources such as rabbit skeletal muscle and yeast actin differ by only 10–12% in their amino acid sequences (7, 9).

Thus, the actin gene family can provide an excellent model to study the evolution of genes. In addition, because different actin genes are expressed in different cells, a comparison of the structure of actin genes and their organization may be helpful in understanding how they are controlled. The skeletal muscle actin gene is of particular interest because studies with muscle cell cultures have shown that this gene becomes expressed during terminal differentiation (8). Thus, structural analysis of this gene and its flanking regions may provide essential information for understanding the mechanism of its expression during development. In the present communication, we report the isolation and characterization by electron microscopy, S1 mapping, and partial sequence determination of the rat skeletal muscle actin gene and a nonmuscle actin gene. This study shows that while several intron sites were conserved faithfully from echinoderms to mammals, the skeletal muscle and the cytoplasmic actin genes differ in at least one intron site.

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

Preparation of Spleen DNA and Southern Blots. High molecular weight DNA was prepared from spleens of Wistar rats, as described by Gross-Bellard *et al.* (10). The DNA was digested with restriction enzymes and blot hybridized as described in Fig. 1 (11).

Screening of the Rat Genomic DNA Libraries. Two rat genomic DNA libraries were used. One is a partial *Eco*RI digest library and the other is an *Hae* III partial digest library—both gifts from T. Sargent, L. Jagodzinski, and J. Bonner. They were screened for the actin genes as described (12, 13) by using nick-translated probes of plasmid p749 or its purified 570-base pair (bp) insert.

R Loop Analysis. R loop analysis was carried out as described by Kaback *et al.* (14).

S1 Nuclease Mapping. Uniformly ^{32}P-labeled plasmid DNA was prepared by growing the bacteria in low-phosphate medium containing 2 mCi of ^{32}P per 50 ml (1 Ci = 3.7×10^{10} becquerels). The specific activity of the DNA obtained was $0.5–1 \times 10^6$ cpm/μg. The endonuclease S1 procedure of Berk and Sharp (15) was followed.

DNA Sequence Determination. DNA fragments were end-labeled by filling in the 5′ protruding restriction sites with [α-^{32}P]deoxynucleotides, using reverse transcriptase, and the sequence was determined by the Maxam and Gilbert procedure (16).

Identification of Plasmid p72 by mRNA Selection and Translation. Plasmid p72 was linearized by digestion with *Eco*RI, and 5 μg was bound to a nitrocellulose filter. After baking for 2 hr at 80°C, the filter was used to select specific RNA from total mRNA of proliferating myoblasts (containing small amounts of fibers) as described by Ricciardi *et al.* (17). The selected RNA was translated in a rabbit reticulocyte lysate cell-free system, and the products were analyzed by isoelectric focusing and polyacrylamide/NaDodSO$_4$ gel electrophoresis (18). The predominant translation product of the RNA selected by p72 comigrated with the β-actin marker. No radioactive spots were detected in the α- and γ-actin positions. Under the same conditions, p749 bound translatable mRNAs for α-, β-, and γ-actin.

RESULTS AND DISCUSSION

Hybridization of Southern Blots of Rat DNA with Probes for Actin mRNA Sequences. Plasmid p749 contains an insert

Abbreviations: bp, base pair(s); kb, kilobase; NaCl/Cit, 0.15 M NaCl/0.015 M sodium citrate, pH 7.

that specifies codons for amino acid residues 171–360 of actin mRNA; plasmid p106 contains a 241-bp insert homologous to the entire 3′ untranslated region of the α-actin mRNA (19, 20). Plasmid p749 hybridizes to muscle and to nonmuscle actin mRNAs from various animals; plasmid p106 hybridizes specifically to striated muscle (skeletal and cardiac) actin mRNAs of mammals (20). Plasmid p72 contains a double-stranded cDNA insert of ≈250 bp, made on an RNA template extracted from rat brain (21). The insert of this plasmid was identified as a nonmuscle β-actin cDNA because it hybridized specifically, under stringent conditions, to β-actin mRNA.

The three plasmids were used to probe Southern blots of rat spleen DNA digested by the restriction enzyme *Eco*RI. Plasmid p749 hybridized to at least 12 DNA fragments (Fig. 1). By increasing the stringency of washing of the blots, it was possible to identify the DNA fragment sharing the best homology with the probe: two of the hybrids [5.7- and 6.9-kilobase (kb) fragments] were stable at 72°C in 0.015 M NaCl/0.0015 M sodium citrate, pH 7 [0.1 × NaCl/Cit (0.15 M NaCl/0.015 M sodium citrate, pH 7)]. At still higher stringency (76°C in 0.1 × NaCl/Cit), only the 5.7-kb fragment hybrid could be detected on the blot. Plasmid p106 hybridized to both the 5.7- and the 6.9-kb DNA fragments.

In a similar set of experiments, we found that plasmid p72 hybridized to several DNA fragments. Two of the hybrids (8.0 and 6.9 kb) were stable at 72°C in 0.1 × NaCl/Cit (Fig. 1). It seems that both p106 and p72 hybridize to a 6.9-kb fragment. We do not know whether both probes hybridize to the same fragment or to two fragments of similar size.

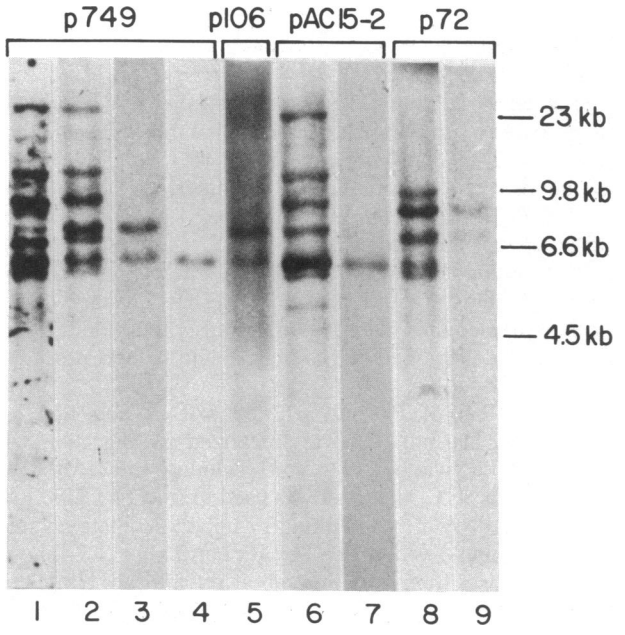

FIG. 1. Southern blots of rat spleen DNA hybridized to various actin DNA probes. Fifteen micrograms of rat spleen DNA samples were digested with *Eco*RI, fractionated on 0.7% horizontal agarose gels, transferred to nitrocellulose filters, and hybridized with nick-translated DNA inserts as indicated. Hybridization was done in 50% formamide/5 × NaCl/Cit/0.1% polyvinyl pyrrolidone/0.1% Ficoll/1% bovine serum albumin/200 μg of *E. coli* DNA per ml at 42° for 40 hr. After the hybridization, the blots were washed as follows: tracks 1, 5, and 6, 90 min in 2 × NaCl/Cit at 50°C; tracks 2 and 8, 60 min in 0.1 × NaCl/Cit at 50°C; tracks 3, 7, and 9, 30 min in 0.1 × NaCl/Cit at 72°C; track 4, 15 min in 0.1 × NaCl/Cit at 76°C. Two DNA preparations (each from several spleens) were used for the Southern blots. One was used in track 1 and the other in tracks 2–9. The variation in some bands may be due to genetic polymorphism.

Structure of the Rat Skeletal Muscle Actin Gene. The rat genomic DNA libraries were screened with labeled p749. Twenty recombinant phages were isolated and analyzed. Their restriction patterns and their hybridization to p749 showed that they originated from eight different actin genes. One of the recombinant phages, Act 15, which was isolated from the *Hae* III rat DNA library, was identified by several criteria as containing the gene coding for skeletal muscle actin: (*i*) Act 15 hybridized strongly to plasmid p106, the striated muscle actin-specific probe. (*ii*) Act 15 hybridized strongly to the *Eco*RI 5.7-kb spleen DNA fragment to which plasmid p749 hybridized under stringent conditions (Fig. 1). This fragment is insensitive to DNase I treatment in nuclei of nonmuscle tissues and is DNase I sensitive in nuclei of differentiated muscle cells (unpublished data), thus indicating that the gene is active in skeletal muscle (22). (*iii*) Partial sequence determination of the coding region of Act 15 showed codon sequences specific for the amino acid sequence of skeletal muscle actin (Table 1).

Restriction mapping revealed that Act 15 contains a 14.7-kb insert. The structural gene sequences are within a terminal 3.7-kb *Eco*RI fragment. This fragment is smaller than the 5.7-kb fragment observed in the genomic Southern blot because the phage was isolated from an *Hae* III library in which new terminal *Eco*RI sites were constructed. Eleven kilobase pairs flanking the 5′ end of the gene are included in the inserted DNA, whereas the terminal 165 bp of the 3′ untranslated end are missing (deduced from the sequence of p106). The 3.7-kb fragment was subcloned in the *Eco*RI site of pBR322, and the recombinant plasmid pAC15.2 was used for structural analysis.

An electron micrograph of the R loops formed between the linearized plasmid pAC15.2 and rat skeletal muscle mRNAs is shown in Fig. 2 *A* and *B*. Thirty-one molecules were measured. At least five intervening sequences were detected (Fig. 2*C*). At the 5′ end of the gene, a large double-stranded loop was seen in all of these molecules and was interpreted as a large intervening sequence measuring 750 (±65) bp. An R loop representing an exon could never be detected in the 5′ direction of this intron, though a small exon must exist there because the RNA homologous to this exon is responsible for making the large loop. Five more exons (5′ to 3′ direction) were measured to be 150 (± 35) bp, 300 (±50) bp, 340 (±45) bp, 175 (±35) bp, and 200 (±45) bp, respectively (Fig. 2*C*).

For further analysis of the number and length of the exons, S1 nuclease mapping was performed. The 3.7-kb fragment con-

Table 1. Comparison of amino acid sequences in various actins

Source	Residue 259	266	271	278	286	296	298	357	364
Actins									
Skeletal muscle (rabbit)	Thr	Ile	Ala	Tyr	Ile	Asn	Met	Thr	Ala
Cardiac muscle (bovine)	—	—	—	—	—	—	Leu	Ser	—
Smooth muscle (chick gizzard)	—	—	—	—	—	—	Leu	Ser	—
β and γ (bovine)	Ala	Leu	Cys	Phe	Val	Thr	Leu	Ser	Ser
Genes									
Act-15	—	—	ND	—	—	—	—	—	—
Act-1	Ala	Leu	Cys	Phe	Val	Thr	ND	ND	ND

Amino acid replacements (between residue 259 and the COOH terminus) in actins isolated from the indicated sources (from ref. 6) and codons for corresponding amino acids in Act 1 and Act 15 genes are shown. —, sequences identical to the skeletal muscle actin; ND, not determined. Amino acid sequences unique to skeletal muscle actin are underlined.

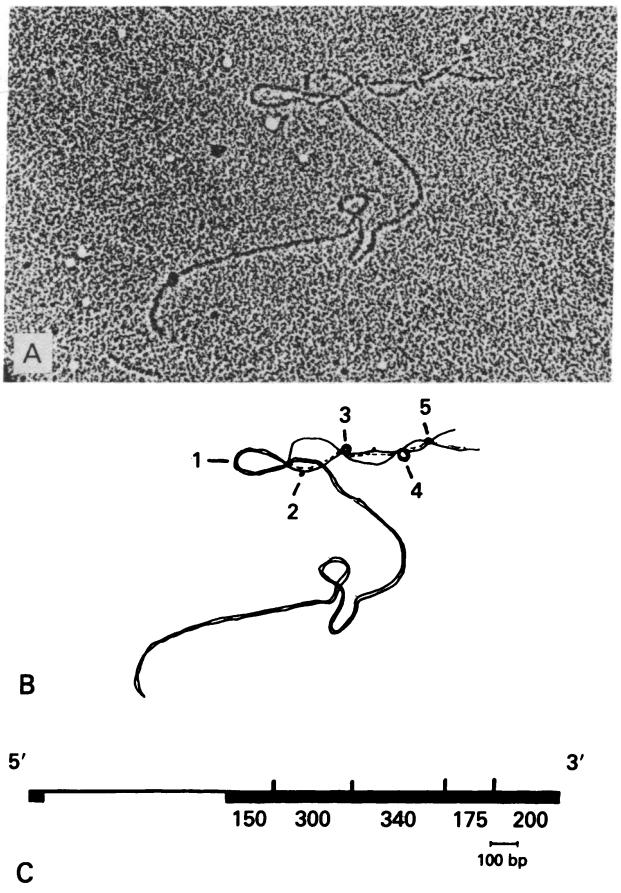

FIG. 2. (*A*) Electronmicrograph of the R loops formed between plasmid pAC15.2 and rat skeletal muscle mRNA. (*B*) The interpretive drawing of the molecule shown in *A*. Single and double solid lines, single- and double-stranded DNA, respectively; dashed line, RNA. In several molecules, some small introns did not pair with the complementary DNA but formed a small blip that appeared consistently in the same place. The visible introns are numbered (5′ → 3′). (*C*) Schematic representation of the structure of the α-actin structural gene, obtained by measuring 31 molecules. Black bars, exons; thin lines, introns whose size was measurable; ▮, intron too small to be measured.

taining the actin gene was uniformly labeled with [^{32}P]phosphate and hybridized to rat muscle poly(A)-containing RNA. Seven exons were detected by this technique (Fig. 3). Their sizes were 330, 220, 191, 180, 160, 142, and 55 bp.

The partial sequencing of Act 15 showed the existence of introns at codons coding for amino acids 327, 267, 204, 150, and 41. The sizes of these introns were 76, 134, 104, 180, and 86 bp, respectively. Thus, the S1 nuclease mapping and the DNA sequence analysis revealed an intron at the codon coding for amino acid 204 that was not revealed by electron microscopy. The sequence analysis data (from amino acid 15 to the 3′ end of the cloned DNA) excluded the existence of more introns in this region.

Based on the data from R loop mapping, S1 nuclease mapping, and DNA sequence analysis, we propose the structure for rat skeletal muscle actin gene as presented in Fig. 4.

Isolation and Gross Structure of a Nonmuscle Actin Gene. One of the recombinant phages (Act 1), which was isolated from the *Eco*RI library, hybridized to plasmids p749 and p72 but not to p106 DNA. Restriction mapping and partial sequence analysis of Act 1 DNA identified an 8-kb *Eco*RI fragment as probably containing the entire structural region of the gene. The inserted DNA in the phage contained about 3.8-kb flanking the gene on the 5′ end and about 8 kb on the 3′ end. The 8-kb *Eco*RI fragment that contained the structural gene was subcloned into the *Eco*RI site of pBR322 for further structural analysis. The size of this fragment is the same as that of one of the two *Eco*RI fragments of rat spleen DNA that hybridized with p72 at high stringency (Fig. 1).

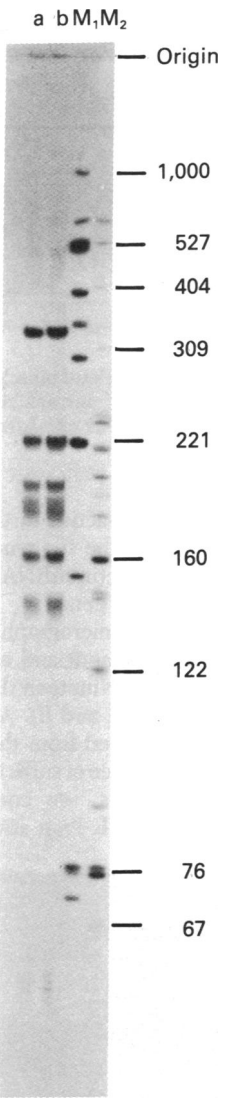

FIG. 3. Sizing of the exons of Act 15 by S1 nuclease mapping of α-actin mRNA using uniformly labeled plasmid pAC15. One microgram (track a) and 5 μg (track b) of rat skeletal muscle polyadenylated RNA were hybridized to the labeled 3.7-kb fragment containing the α-actin gene. After S1 nuclease treatment, the products were analyzed on a 5% polyacrylamide sequence analysis gel. ^{32}P-End-labeled pBR322, cut either with *Hinf*/*Eco*RI (track M_1) or with *HPA* II (track M_2) were used as size markers.

Hybridization of p72 or p749 with Southern blots of DNA from DNase I-treated nuclei showed that the 8-kb fragment is preferentially sensitive to treatment with DNase I in nuclei of cells synthesizing β-actin (i.e., brain, proliferating myoblasts, and differentiated muscle cultures). This indicates that this fragment contains an active actin gene (unpublished data).

The DNA sequence determination showed that Act 1 is a gene coding for a cytoplasmic actin (Table 1). The sequence of a fragment of 105 bp of p72 was also determined and found to be identical, except for four nucleotides (perhaps due to a mistake in sequencing), with a segment coding for the 3′ untranslated region of the Act 1 gene. There are no amino acid re-

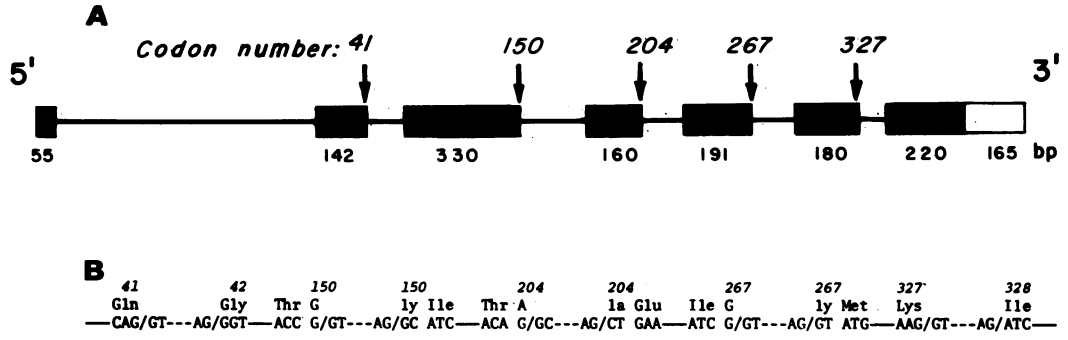

FIG. 4. (*A*) Structure of rat skeletal muscle α-actin gene deduced from R loop mapping, S1 mapping, and partial DNA sequence analysis. Solid bars, exons; open bar, part of the 3′ untranslated region that is missing from Act 15; thin lines, introns. (*B*) Nucleotide sequence at the exon–intron junctions. —— Exon; ---- intron. The numbers above *A* and *B* indicate the position of introns (codon number) in the sequenced portion of the gene. Numbering is according to ref. 6.

placements between β- and γ-actins in the sequenced region (6). However, on the basis of the sequence identity with p72, which binds specifically to β-actin mRNA, it seems most likely that Act 1 contains the β-actin gene.

Fig. 5*A* shows an electron micrograph of an R loop formed between the 8-kb *Eco*RI fragment and actin-enriched mRNA from proliferating myoblasts. Nineteen different hybrid molecules were measured (Fig. 5 *A* and *B*). A schematic representation of the structure suggested from those measurements is shown in Fig. 5*C*. (In the absence of sufficient confirmatory data from DNA sequence analysis, we consider it a tentative scheme.) The analysis of the R loop structures indicates the existence of at least five exons. Similar to the α-actin gene, the largest intron (820 ± 45 bp) is very close to the 5′ end of the transcribed region, and the first exon at the 5′ end could not be detected. The length of the other exons (5′ to 3′) were 100 (±10), 200 (±25), 490 (±70), and 710 (±80) bp long, respectively. By allowing for another exon about 50–100 bp long at the 5′ end, the total exon length measures about 1,600 bp. It should be noted that this is about 30% smaller than the length of the mRNA coding for β- or γ-actin, estimated from denaturing gels (19, 20). Such discrepancy in measurements of R loops has already been reported (e.g., ref. 23).

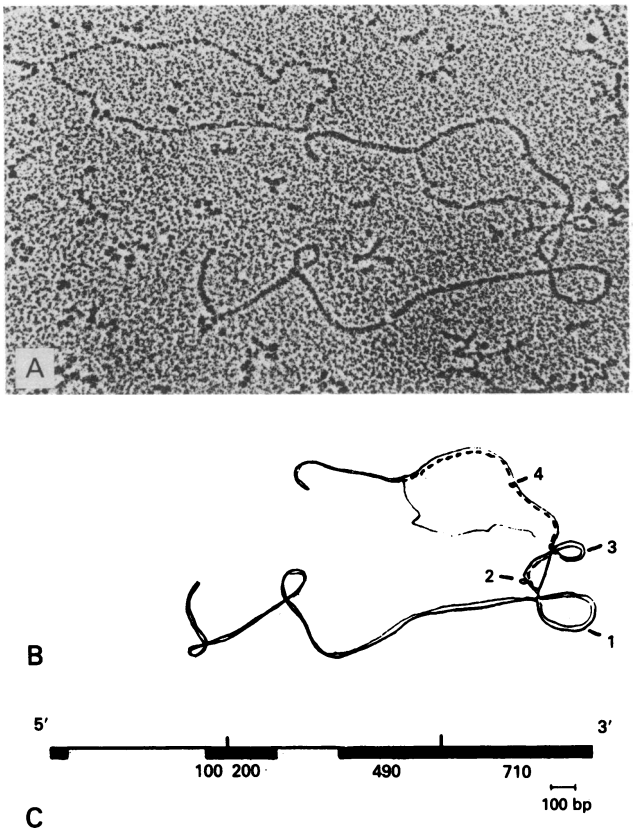

FIG. 5. (*A*) Electronmicrograph of an R loop formed between plasmid pACR1 and mRNA from dividing myoblast. (*B*) Interpretive drawing of the molecule shown in *A*. (*C*) Schematic representation of the gross structure of rat β-actin gene as suggested from measuring 19 molecules. See the legend to Fig. 2.

The sequence of Act 1 has been determined from codon 120 to codon 152 and from codon 225 to codon 300. At these regions there are introns at codon 121 (size of intron unknown) and at codon 267 (90 bp). Thus, in the regions in which the sequences of both α- and β-actin genes were determined, there is one intron in common (codon 267), one intron found in Act 15 but not in Act 1 (codon 150), and one in Act 1 but not in Act 15 (codon 121).

Several actin genes in other organisms have been investigated. Although these data are still fragmentary, it is already apparent that the actin gene family may provide interesting information relevant to the evolution, preservation, and possible role of introns. Seventeen actin-like genes have been found and sequenced in the slime mold *Dictyostelium discoidea*, but no introns were found (24, 25). Yeast has one actin gene, which has been completely sequenced. It contains only one intron, at codon 4 (7, 9). Partial sequencing is available also for the six *Drosophila* actin genes. Each contains several introns at different positions (26). None of these introns is at a position found in the present work. On the other hand, partial sequencing of an actin gene isolated from sea urchin shows introns at codons 41, 121, 204, and 267 (refs. 27 and 28; W. R. Crain and A. D. Cooper, personal communication). Thus, the available data suggest that

Table 2. Intron sites in actins (codons)

Rat		Sea
α	β	urchin*
41	?	41
—	121	121
150	—	—
204	?	204
267	267	267
327	?	—

—, No intron; ?, sequence unknown.
* W. R. Crain and A. D. Cooper, personal communication; refs. 27 and 28.

although there seems to be no homology between the positions of introns in *Drosophila* (which is a protostome) and those of the echinoderm and mammal actin genes, several introns have been preserved along a wide range of evolution in the deuterostomes (Table 2). Notably, an intron at position 121 is present in the sea urchin gene and rat β-actin gene, but not in rat α-actin gene, whereas an intron at position 150 is present in rat α-actin gene but not in the β-actin and echinoderm actin genes. Partial sequencing of chicken α-actin indicates identity of intron sites with pAC15 (C. P. Ordahl, personal communication).

Several investigations suggested that introns separate DNA sequences coding for different functional domains of the protein (29, 30). Introns have been found so far at a minimum of 10 sites in the actin gene family. Although the amino acid sequence of the various actins has been highly conserved during evolution, some of the introns are found in homologous positions but others are not. More information on structure–function relationships of the actin molecule and on the positions of introns in the various actin genes may provide important data for the examination of that hypothesis.

Note Added in Proof. Further sequence determination of Act 1 has confirmed that it codes for β-actin. The gene contains an intron at codon 41 and codon 327 but not at codon 204.

We thank Ms. M. Engel, O. Saxel, S. Neuman, and Z. Levy for their excellent technical assistance and Drs. D. Kaback and Y. Shaul for introducing us to electron microscopy techniques. Supported by National Institutes of Health Grant R01 GM 22767 and by the Muscular Dystrophy Association, Inc., New York. U.N. is the incumbent of the A. and E. Blum Career Development Chair in Cancer Research.

1. Whalen, R. G., Butler-Browne, G. S. & Gros, F. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2018–2022.
2. Storti, R. V. & Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2346–2350.
3. Garrels, J. & Gibson, W. (1976) *Cell* **9**, 793–805.
4. Lu, R. & Elzinga, M. (1977) *Biochemistry* **16**, 5801–5806.
5. Vandekerckhove, J. & Weber, K. (1979) *FEBS Lett.* **102**, 219–222.
6. Vandekerckhove, J. & Weber, K. (1979) *Differentiation* **14**, 124–133.
7. Gallwitz, D. & Sures, I. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2546–2550.
8. Shani, M., Zevin-Sonkin, D., Saxel, O., Carmon, Y., Katcoff, D., Nudel, U. & Yaffe, D. (1981) *Dev. Biol.* **86**, 483–492.
9. Ng, R. & Abelson, J. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3912–3916.
10. Gross-Bellard, M., Oudet, P. & Chambon, P. (1973) *Eur. J. Biochem.* **36**, 32–38.
11. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
12. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
13. Nudel, U., Katcoff, D., Carmon, Y., Zevin-Sonkin, D., Levi, Z., Shaul, Y., Shani, M. & Yaffe, D. (1980) *Nucleic Acids Res.* **10**, 2113–2146.
14. Kaback, D. B., Rosebush, M. & Davidson, N. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2820–2824.
15. Berk, A. J. & Sharp, P. A. (1977) *Cell* **12**, 721–732.
16. Maxam, A. M. & Gilbert, W. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
17. Ricciardi, R. P., Miller, J. S. & Roberts, B. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4927–4931.
18. Hunter, T. & Garrels, J. I. (1977) *Cell* **12**, 767–781.
19. Katcoff, D., Nudel, U., Zevin-Sonkin, D., Carmon, Y., Shani, M., Lehrach, H., Frischauf, A. M. & Yaffe, D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 960–964.
20. Shani, M., Nudel, U., Zevin-Sonkin, D., Zakut, R., Givol, D., Katcoff, D., Carmon, Y., Reiter, J., Frischauf, A. M. & Yaffe, D. (1981) *Nucleic Acids Res.* **9**, 579–589.
21. Ginzburg, I., De Baetselier, A., Walker, M. D., Behar, L., Lehrach, H., Frischauf, A. M. & Littauer, U. Z. (1980) *Nucleic Acids Res.* **8**, 3553–3564.
22. Weintraub, H. & Groudine, M. (1976) *Science* **193**, 849–856.
23. Kinniburgh, A. J., Mertz, J. E. & Ross, J. (1978) *Cell* **14**, 681–693.
24. Kindle, K. L. & Firtel, R. A. (1978) *Cell* **15**, 763–778.
25. Firtel, R. A., Timm, R., Kimmel, A. R. & McKeown, M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 6206–6210.
26. Fyrberg, E. A., Bond, B. J., Hershey, N. D., Mixtes, K. S. & Davidson, N. (1981) *Cell* **24**, 107–116.
27. Dorica, D. S., Schloss, J. A. & Crain, W. R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5683–5687.
28. Schuler, A. M. & Keller, E. B. (1981) *Nucleic Acids Res.* **9**, 591–604.
29. Gilbert, W. (1978) *Nature (London)* **271**, 501.
30. Blake, C. C. F. (1981) *Nature (London)* **291**, 616.