MOLECULAR AND CELLULAR BIOLOGY, Jan. 1986, p. 15–25
0270-7306/86/010015-11$02.00/0

Vol. 6, No. 1

# The Complete Sequence of the Mouse Skeletal α-Actin Gene Reveals Several Conserved and Inverted Repeat Sequences Outside of the Protein-Coding Region

MICKEY CHIEN-TSUNG HU, SANDRA B. SHARP, AND NORMAN DAVIDSON*
*Department of Chemistry, California Institute of Technology, Pasadena, California 91125*

Received 17 June 1985/Accepted 24 September 1985

**The complete nucleotide sequence of a genomic clone encoding the mouse skeletal α-actin gene has been determined. This single-copy gene codes for a protein identical in primary sequence to the rabbit skeletal α-actin. It has a large intron in the 5′-untranslated region 12 nucleotides upstream from the initiator ATG and five small introns in the coding region at codons specifying amino acids 41/42, 150, 204, 267, and 327/328. These intron positions are identical to those for the corresponding genes of chickens and rats. Similar to other skeletal α-actin genes, the nucleotide sequence codes for two amino acids, Met-Cys, preceding the known N-terminal Asp of the mature protein. Comparison of the nucleotide sequences of rat, mouse, chicken, and human skeletal muscle α-actin genes reveals conserved sequences (some not previously noted) outside of the protein-coding region. Furthermore, several inverted repeat sequences, partially within these conserved regions, have been identified. These sequences are not present in the vertebrate cytoskeletal β-actin genes. The strong conservation of the inverted repeat sequences suggests that they may have a role in the tissue-specific expression of skeletal α-actin genes.**

The actins represent a multigene family of highly conserved proteins found in all eucaryotes. Differences in amino acid sequence among the various actins have shown that at least six different isoforms are expressed in vertebrates (52, 53). Two striated muscle isoforms, skeletal α and cardiac α (52), and two smooth muscle isoforms (53) are found in the contractile apparatus of muscle fibers, whereas two cytoskeletal isoforms, β and γ, are present in the cytoskeleton of all cells (51). These actin proteins are extremely conserved in amino acid sequence.

Actin gene expression is tissue specific and developmentally regulated (27, 29, 32, 35). By studying the structural organization of the actin gene family, one can begin to look for the controlling elements which modulate the expression of these genes during development. Here we present the complete nucleotide sequence of the single genomic copy of the mouse skeletal α-actin gene. The coding region of this gene is interrupted by five introns which are located in the same positions as introns previously identified in other vertebrate skeletal α-actin genes (13, 54). A comparison of the nucleotide sequences of several vertebrate skeletal α-actin genes reveals several blocks of highly conserved sequences in the 5′-flanking region and in both the 5′- and 3′-untranslated regions. Interestingly, the conserved sequences in the 5′-flanking region and within the first untranslated exon can potentially form several hairpin loops by base pairing between adjacent inverted complementary sequences. These regions do not correspond to the potential hairpin structure in the corresponding portion of the rat cytoskeletal β-actin gene (33). Furthermore, it is possible to form long hairpin loops within the first intron and one stem loop in the 3′-untranslated region upstream from the putative polyadenylation signal ATTAAA. These interesting secondary structures are apparently not present in the vertebrate cytoskeletal β-actin genes. To our knowledge, this is the first description of potential secondary structures in the first intron and among the highly conserved sequences in the 5′-flanking region and both the 5′- and 3′-untranslated regions of vertebrate skeletal α-actin genes. Since the species compared (avian and mammalian) have been separated for more than 250 million years (12), these results indicate a strong selective constraint to conserve these sequences and suggest that they may have an important role in the tissue-specific expression of the skeletal α-actin genes.

## MATERIALS AND METHODS

**Materials.** Restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, *Escherichia coli* exonuclease VII, and *E. coli* DNA polymerase I large fragment (Klenow) were purchased from Bethesda Research Laboratories, Boehringer Mannheim Biochemicals, or New England BioLabs. Avian myeloblastosis virus reverse transcriptase was obtained from Life Sciences, Inc. Sp6 RNA polymerase and placental RNasin were obtained from Promega Biotec. Radioactively labeled nucleotides were purchased from Amersham Corp. or New England Nuclear Corp. Unlabeled nucleotides were obtained from P-L Biochemicals, Inc. Synthetic oligonucleotides were synthesized by S. Horvath, Caltech, on an automated DNA synthesizer (21) and purified by electrophoresis through a 20% polyacrylamide–8 M urea preparative gel in Tris borate-EDTA buffer. BALB/c genomic DNA was provided by T. Hunkapiller. The *Drosophila* actin genomic clone DmA2 (Dm5C in reference 14) and a 3′-untranslated region of rat skeletal α-actin cDNA (15) were provided by B. Bond and L. Garfinkel, respectively.

**Isolation of genomic actin clones and restriction mapping.** A cosmid genomic library of BALB/c mouse sperm DNA, constructed and kindly provided by M. Steinmetz at Caltech, was screened by colony hybridization (49) by using an actin-coding region probe isolated from the *Drosophila* actin genomic clone DmA2 (Dm5C). From 16 positive clones, 1 was tentatively identified to contain the skeletal α-actin gene by hybridization with the conserved (29), isotype-specific (37), 3′-untranslated region of a rat skeletal α-actin cDNA.

* Corresponding author.

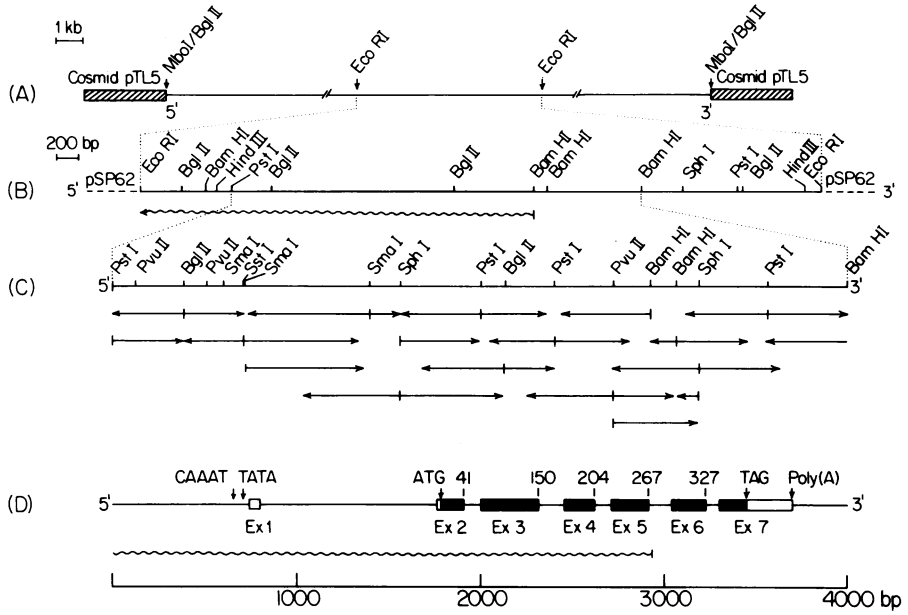


FIG. 1. Structure of the mouse skeletal α-actin gene. (A) Cosmid clone containing the genomic DNA encoding the mouse skeletal α-actin gene. (B) Restriction endonuclease map of the 6.8-kb *Eco*RI DNA fragment. One squiggly arrow represents an SP6 antisense transcript of a *Bam*HI-*Eco*RI fragment that was subcloned from the 6.8-kb *Eco*RI fragment. (C) Detailed restriction map of the mouse skeletal α-actin gene and flanking DNA. Fragments were subcloned into M13 vectors and sequenced by dideoxy chain termination as indicated by arrows. (D) Schematic representation of the structure of the mouse skeletal α-actin gene. Solid boxes represent coding exons, open boxes indicate transcribed untranslated regions, and solid lines coincide with introns and flanking DNA. Numbers above the exons correspond to codon positions (Fig. 2). The squiggly line represents the SP6 antisense transcript as mentioned above.

Southern blot (48) analysis with the *Drosophila* actin probe localized the mouse α-actin-coding region to a single 6.8-kilobase (kb) *Eco*RI fragment in the cosmid clone. This fragment was subcloned into the *Eco*RI site of plasmid pSP62-PL (28), provided by D. Melton, Harvard University.

The restriction endonuclease map of the 6.8-kb *Eco*RI fragment was determined by single and double enzyme digests. Subsequently, the 3.9-kb *Bam*HI-*Eco*RI fragment was subcloned into the *Bam*HI and *Eco*RI sites of the plasmid pSP62-PL (see Fig. 1B) and mapped in finer detail by digestion with more restriction endonucleases.

**Localization of the promoter region of the α-actin gene.** Several different restriction endonuclease digests of the 6.8-kb *Eco*RI fragment were probed with a 20-base oligonucleotide (5'-GCCCAACACCCAAATATGGC-3') containing the sequence of the CAAT promoter homology, highly conserved between the skeletal α-actin genes of chickens and rats (34). The oligonucleotide was 5' end labeled with polynucleotide kinase and [γ-$^{32}$P]ATP, and hybridization was performed directly in the dried agarose gel as described previously (39). A linearized chicken skeletal α-actin genomic clone (provided by C. Ordahl [13]) and *Eco*RI-linearized SP6 vector were used as positive and negative controls, respectively.

**M13 cloning and DNA sequencing.** Appropriate restriction fragments from the 6.8- and 3.9-kb inserts were subcloned, in opposite orientations, into the multiple cloning sites of M13 mp18 and M13 mp19 RF vectors, transforming first into *E. coli* JM101 for high efficiency and replating with *E. coli* JM109 (*recA*$^-$) to prevent sequence changes.

Single-stranded M13 templates were sequenced by the dideoxy chain termination procedure (41) with an [α-$^{35}$S]dATP (500 Ci/mmol) label as described by Biggin et al. (3), with the following modifications. (i) The synthetic pentadecanucleotide (5'-TCCCAGTCACGACGT-3') and the hexadecanucleotide (5'-GGGTAACGCCAGGGTT-3') were used as sequencing primers. (ii) The dideoxy sequence reactions were carried out in 50 mM NaCl–7 mM Tris hydrochloride (pH 7.4)–10 mM $MgCl_2$–3 mM dithiothreitol. (iii) The final concentrations of unlabeled nucleotides in each sequence reaction were as follows: A reaction, 25 μM dCTP, 25 μM dGTP, 25 μM dTTP, 20 μM ddATP; C reaction, 8 μM dCTP, 32 μM dGTP, 32 μM dTTP, 50 μM ddCTP; G reaction, 32 μM dCTP, 8 μM dGTP, 32 μM dTTP, 50 μM ddGTP; T reaction, 32 μM dCTP, 32 μM dGTP, 8 μM dTTP, 50 μM ddTTP. (iv) After electrophoresis, the gel was immediately dried, without fixing, for 1 h at 80°C and autoradiographed.

Computer analysis of sequence homology was done as described by Hunkapiller et al. (21).

**Primer extension analysis.** Polyadenylated [poly(A)$^+$] RNA from a differentiated culture of the mouse myogenic cell line BC3H-1 (42) was isolated by guanidine thiocyanate extraction (9) and two cycles of oligo(dT)-cellulose chromatography (1). A synthetic 42-base oligonucleotide (5' - AGAGCCGTTGTCACACACAAGAGCGGTGGTCTC GTCTTCGTC-3') complementary to a portion of the coding sequence, spanning positions 1038 through 1079 in exon 2, was 5' end labeled with polynucleotide kinase and [γ-$^{32}$P]ATP (3,000 Ci/mmol) and used as an extension primer. One picomole of the labeled oligonucleotide ($10^6$ cpm/μg) was denatured by heating at 80°C for 10 min in 40 μl of 98% formamide containing 1 μl of 0.5 M EDTA. Five micrograms of poly(A)$^+$ RNA in 10 μl of 200 mM sodium piperazine-*N*,*N*'-bis(2-ethanesulfonic acid) (pH 6.4)–2 M NaCl–5 mM EDTA was added, and the mixture was incubated at 37°C for 12 h. The nucleic acids were ethanol precipitated from ammonium acetate and reprecipitated from sodium acetate. The dried pellet was suspended in 25 μl of 100 mM Trishydrochloride (pH 8.3)–40 mM KCl–20 mM $MgCl_2$–10

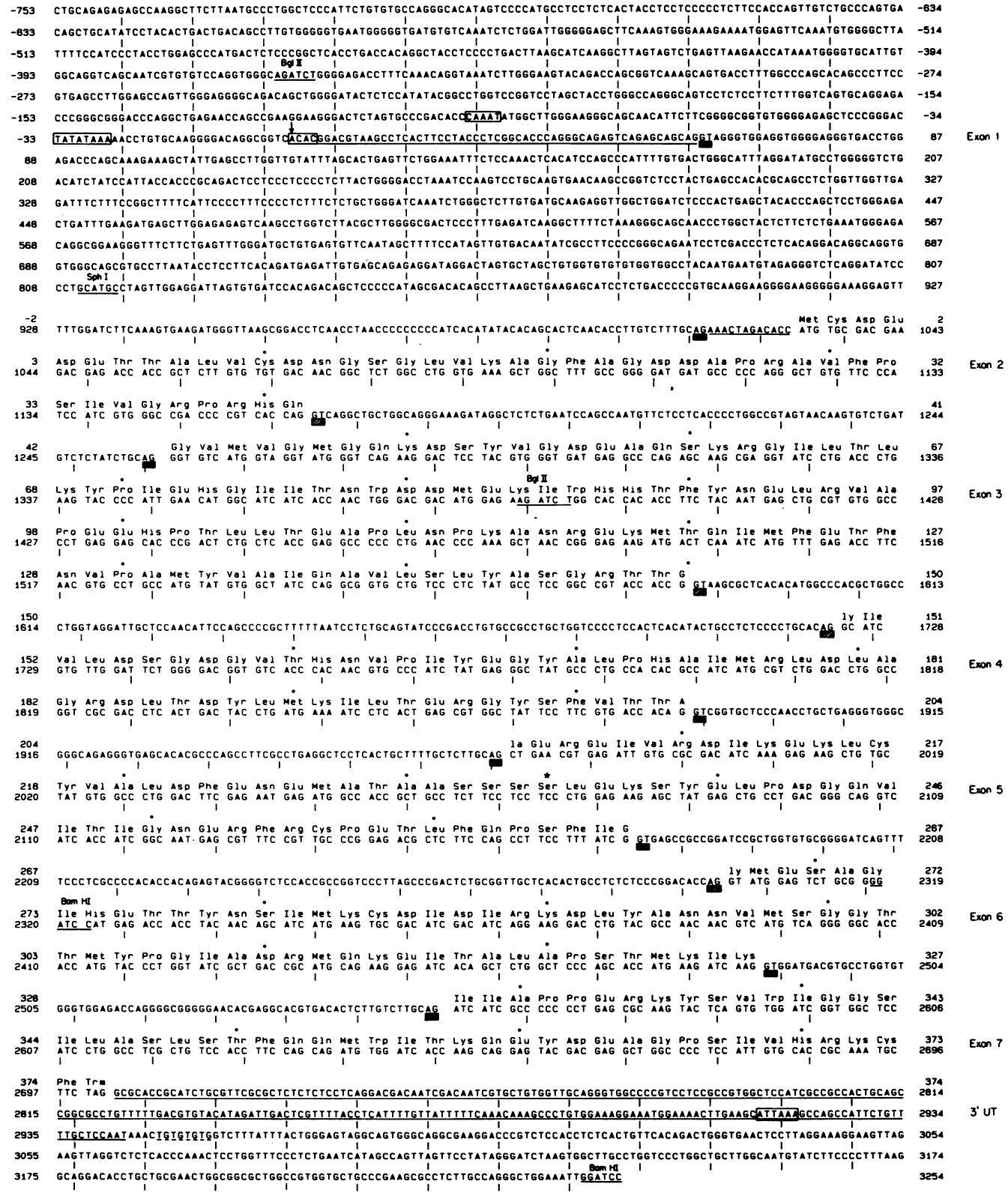


FIG. 2. Nucleotide sequence of the mouse skeletal α-actin gene. Numbers in the left and right margins, respectively, refer to the first and last nucleotides or amino acids in each line. Negative numbers indicate nucleotide positions upstream from the transcription start site. The deduced amino acid sequences encoded by the exons are indicated above the nucleotide sequence in the three-letter amino acid code and numbered as described previously (23, 52). An asterisk indicates an "extra" serine residue between codons 234 and 235, which has been designated position 234a (23, 52). A hatched box represents a splice junction border, and a vertical arrow denotes the cap site which is assigned as nucleotide number 1. The CAAAT, TATA, and putative polyadenylation signal ATTAAA are indicated by the boxes. The 5'- and 3'-untranslated gene regions are underlined, and a G+T-rich stretch downstream from the putative polyadenylation site (16, 26) is underscored with a broken line. The restriction sites used for the exonuclease VII mapping described in the legend to Fig. 3 are marked and underlined. Abbreviations: Trm, termination codon; 3'UT, 3'-untranslated region.

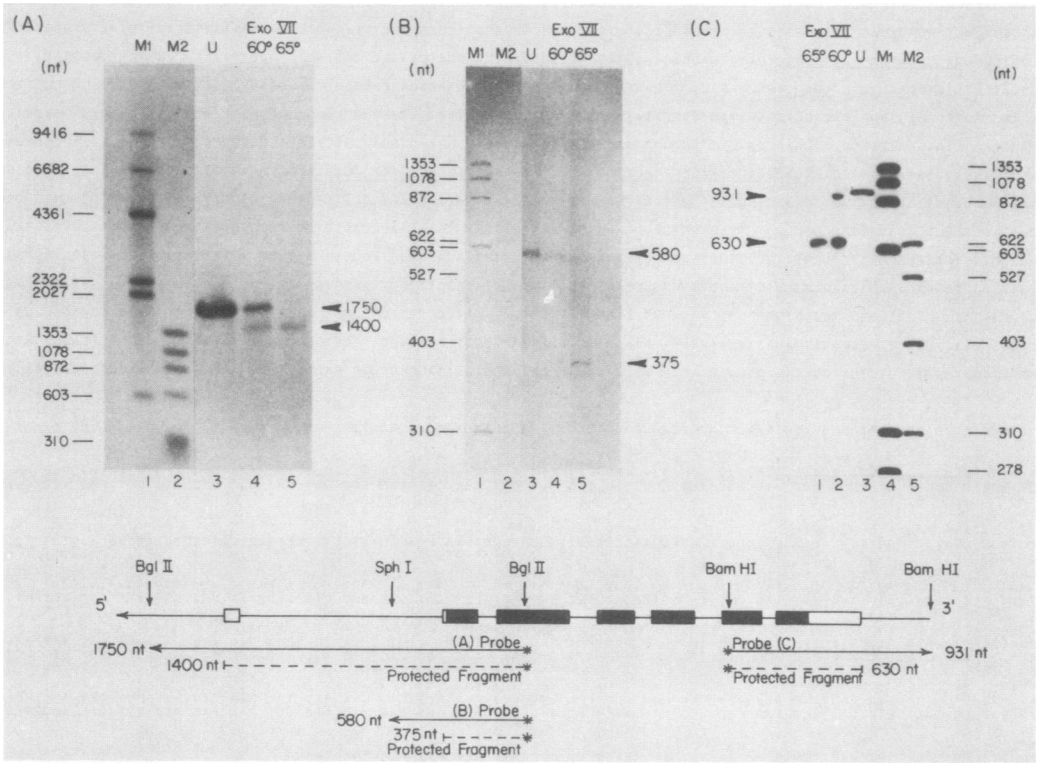


FIG. 3. Exonuclease VII mapping of the mouse skeletal α-actin gene. Total RNA (50 μg) from a differentiated culture of BC3H-1 cells was hybridized at 60 or 65°C with approximately 5 ng of each probe as shown in the schematic diagram (lower panel). The hybridized samples were treated with 3 U of exonuclease VII (1 h, 45°C), electrophoresed on an alkaline agarose gel (A) or a 5% polyacrylamide–8 M urea sequencing gel (B, C), and autoradiographed. (A) Lanes: 1, size markers, *Hin*dIII fragments of γ DNA; 2, size markers, *Hae*III fragments of φX174 DNA; 3, undigested probe of 1,750-nucleotide *Bgl*II-*Bgl*II fragment; 4 and 5, protected products from samples hybridized at 60 and 65°C, respectively. (B) Lanes: 1, size markers, *Hae*III fragments of φX174 DNA; 2, size markers, *Hpa*II fragments of PBR 322; 3, undigested probe of 580-nucleotide *Sph*I-*Bgl*II fragment; 4 and 5, protected products from samples hybridized at 60 and 65°C, respectively. (C) Lanes: 1 and 2, protected products from samples hybridized at 65 and 60°C, respectively; 3, undigested probe of 391-nucleotide *Bam*HI-*Bam*HI fragment; 4, size markers, *Hae*III fragments of φX174 DNA; 5, size markers, *Hpa*II fragments of PBR 322. nt, Nucleotide.

mM dithiothreitol and heated to 65°C for 10 min. This heating step substantially improved the resolution of extension products over background, as observed by Fornwald et al. (13). An equal volume of a solution containing deoxynucleoside triphosphates (1 mM each), 750 U of RNasin per ml, and 40 μg of actinomycin D per ml was added, and the primer extended with avian myeloblastosis virus reverse transcriptase (500 U/ml) at 42°C for 1 h. The reaction was terminated by the addition of EDTA to 10 mM, and the RNA was degraded by treatment with DNase-free RNase (50 μg/ml) at 40°C for 1 h. After ethanol precipitation, the reactions were suspended in formamide-dye buffer, denatured by boiling for 3 min, and electrophoresed on a 6% polyacrylamide–8 M urea sequencing gel in Tris borate-EDTA buffer. DNA sequencing reactions were used as size markers. Extended products were detected by autoradiography of the dried gel.

**Exonuclease VII mapping.** For exonuclease VII mapping, actin-gene-containing plasmid DNA fragments were isolated and labeled with either polynucleotide kinase and [γ-$^{32}$P]ATP (3,000 Ci/mmol) for 5'-end mapping, or *E. coli* DNA polymerase I large fragment (Klenow) and [α$^{32}$P]dCTP (400 Ci/mmol) for 3'-end mapping.

DNA-RNA hybridization was performed as described above in the primer extension analysis by using 50 μg of total RNA from a differentiated culture of BC3H-1 cell line and about 5 ng of labeled (ca. $10^6$ cpm/μg) DNA fragment. Hybridizations were carried out at 60 or 65°C for 3 h. Each hybridization mixture was diluted into 10 volumes of 30 mM KCl–10 mM Tris hydrochloride (pH 7.4)–10 mM EDTA, chilled on ice, and incubated at 45°C for 1 h with 10 U of *E. coli* exonuclease VII per ml (2). After ethanol precipitation, the exonuclease VII-resistant material was electrophoresed on a 5% polyacrylamide–8 M urea sequencing gel or alkaline agarose gel (24) and autoradiographed.

**In vitro transcription with SP6 RNA polymerase and RNase mapping.** Synthesis of the complementary-strand SP6 probe and RNase mapping were carried out as described by Melton et al. (28).

## RESULTS

**Isolation and mapping of the mouse skeletal α-actin gene.** A cosmid clone containing the skeletal α-actin gene was isolated from a BALB/c genomic cosmid library as described in Materials and Methods. The location and orientation of the gene within a single 6.8-kb *Eco*RI fragment were established by restriction endonuclease mapping with 5' and 3' fragments of the *Drosophila* actin-coding sequence and the rat α-actin 3'-untranslated sequence as probes.

The 6.8-kb *Eco*RI fragment containing the entire skeletal α-actin-coding and flanking sequences was isolated from the 41-kb insert of the cosmid clone (Fig. 1A) and subcloned into the *Eco*RI site of the plasmid vector pSP62 (Fig. 1B). Subsequently, the putative promoter region of the gene was

localized to about 100 base pairs within the *Sma*I-*Sst*I region (Fig. 1C and D) by hybridization with the 20-base oligonucleotide probe containing the sequence of the CAAT promoter homology which is highly conserved between the skeletal α-actin genes of chickens and rats (34) (data not shown). This result confirmed that the subcloned *Eco*RI fragment contained the 5′-flanking sequence of the gene. Also, it suggested that the position of the transcription initiation site should be close to the *Sst*I site. The detailed restriction endonuclease map of the actin-coding and flanking regions in Fig. 1C was used to choose the DNA fragments to be subcloned into M13 and mp18 and mp19 vectors for sequencing. From these subclones we determined the linear sequence of 4,007 nucleotides, on both strands independently (Fig. 2).

**Amino-acid-coding region of the mouse skeletal α-actin gene.** The complete nucleotide sequence of the mouse skeletal α-actin gene with the 5′- and 3′-flanking regions is shown in Fig. 2. Exons within the protein-coding region and the introns separating them were initially assigned mainly by homology with the rat skeletal α-actin genomic sequence (54) and in part by comparison with the partial cDNA sequence for the carboxy-terminal portion of the mouse protein (29). This assignment was supported by the fact that the sequences at the presumed exon-intron junctions are in accordance with the consensus sequence for splice sites (7). In addition, the expected lengths of exons 2, 3, 4, and 5 were confirmed experimentally within an accuracy of ±3 nucleotides by hybridizing an SP6 anti-sense transcript of a *Bam*HI-*Eco*RI fragment (Fig. 1B) to poly(A)$^+$ RNA from differentiated BC3H-1 cells and RNase mapping (data to be presented in the Ph.D thesis by M. C.-T. Hu at the California Institute of Technology). The translated amino acid sequence for this interpretation of the structure of the mouse skeletal α-actin gene is identical to that of rats, rabbits, and chickens. (The amino acid sequence of mouse skeletal α-actin has not been directly determined.)

The coding sequence begins with codons for two amino acids, Met and Cys, which are absent from the mature protein. They are followed by the codon for Asp (GAC), the known N-terminal residue of striated muscle actin. The same two codons preceding the codon specifying the N-terminal amino acids are found in the human (20), rat (54), and chicken (13) skeletal α-actin genes and in the human (19) and chicken (8) cardiac α-actin genes. Interestingly, these two codons are also found in all six *Drosophila* actin genes (14) and sea urchin actin genes (10, 43), but the Cys codon is absent in vertebrate cytoskeletal β-actin genes (22, 33, 38). It has been suggested that the Met-Cys dipeptide is removed by posttranslational processing (54).

Although the derived sequence of the primary skeletal α-actin is 377 residues, we have numbered the amino acids in Fig. 2 in conformity with the numbering system suggested by Lu and Elzinga (23) and Vandekerckhove and Weber (52), which yields 374 positions. Of the three additional positions, two are the Met-Cys dipeptide at the N-terminus, while the third is an "extra" serine residue between positions 234 and 235 which has been designated 234a (23, 52).

The coding region of the mouse skeletal α-actin gene is split by five introns (Fig. 1D) at codons specifying amino acids 41/42 (IVS 2), 150 (IVS 3), 204 (IVS 4), 267 (IVS 5), and 327/328 (IVS 6). These intron positions are identical to those for the corresponding genes of chickens and rats (13, 54). The length and positions of the exons and introns may be deduced from Fig. 2. Previously, Zakut et al. (54) have reported that a potential splice site (CAG/GTA) is present 32

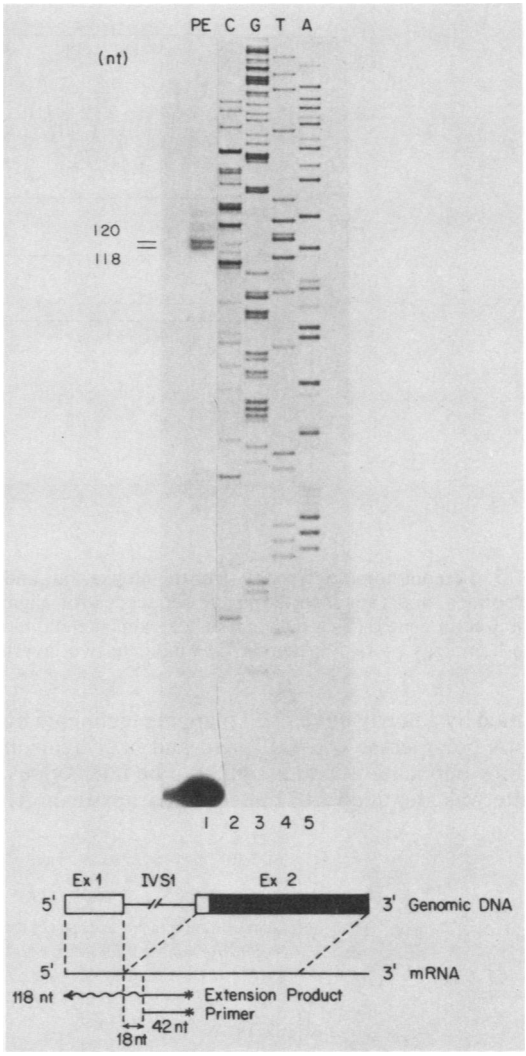


FIG. 4. Identification of the 5′ end of the mouse skeletal α-actin mRNA by primer extension. The 5′-end-labeled, 42-base oligonucleotide complementary to the coding sequence, from positions 1038 to 1079 (Fig. 2), was hybridized with 5 μg of poly(A)$^+$ RNA from a differentiated culture of BC3H-1 cells and extended by using reverse transcriptase as described in the text. The extension products were electrophoresed on a 6% polyacrylamide–8 M urea sequencing gel and autoradiographed. The diagram (lower panel) shows the product expected from full-length elongation of mRNA. Lanes: 1, primer extension products; 2 through 5, sequencing ladders used as size markers. nt, Nucleotide.

base pairs downstream from the CG/GT splice site at codon 150 in the rat skeletal α-actin gene. Our results do not reveal this potential splice site in the mouse gene. This is not surprising, because use of the extra splice site in the rat gene would produce an actin with an insert of 11 amino acids, and no such product has yet been detected.

**Sequence of the 5′-untranslated region of the mouse skeletal α-actin gene.** Although the actin amino acid sequence data and cloned cDNA partial sequence (29) could be used to identify the translated regions of the gene, independent means were required for delineating the 5′-untranslated region. The 5′ borders of the first, untranslated, exon and the second exon were approximately determined by exonuclease VII mapping (Fig. 3A and B). In addition, the precise assignment of the transcription initiation site, ACAC, was

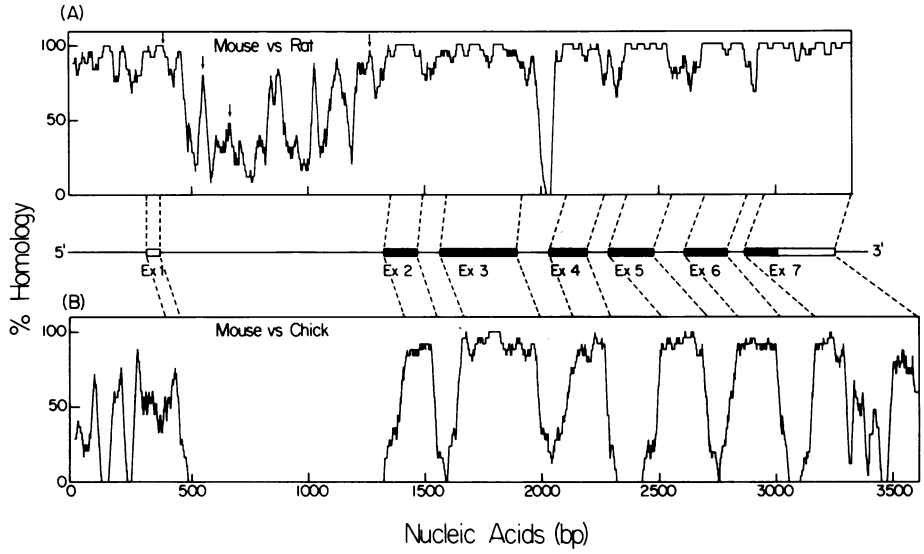


FIG. 5. Percent homology profiles for the mouse, rat, and chicken skeletal α-actin genes. In this analysis, the sequence search string was 25 nucleotides, and gaps inserted in the sequences for alignment purposes were scored as regions of 0% homology. (A) Mouse versus rat skeletal α-actin genes; (B) mouse versus chicken skeletal α-actin genes. The structure of the mouse skeletal α-actin gene is shown between the two homology plots. Vertical arrows indicate two inverted duplications as described in the text. bp, Base pairs.

confirmed by a newly developed mapping technique by using T4 DNA polymerase (M. C. T. Hu and N. Davidson, submitted for publication) and assigned to be nucleotides −1/1. This site was identified 1,031 nucleotides upstream from the initiator ATG codon (Fig. 2). We assigned the 5′ border of the first intron to nucleotides 58/59 by matching the length of the primer extended product (Fig. 4) with the positions determined above for the 5′ borders of the first and second

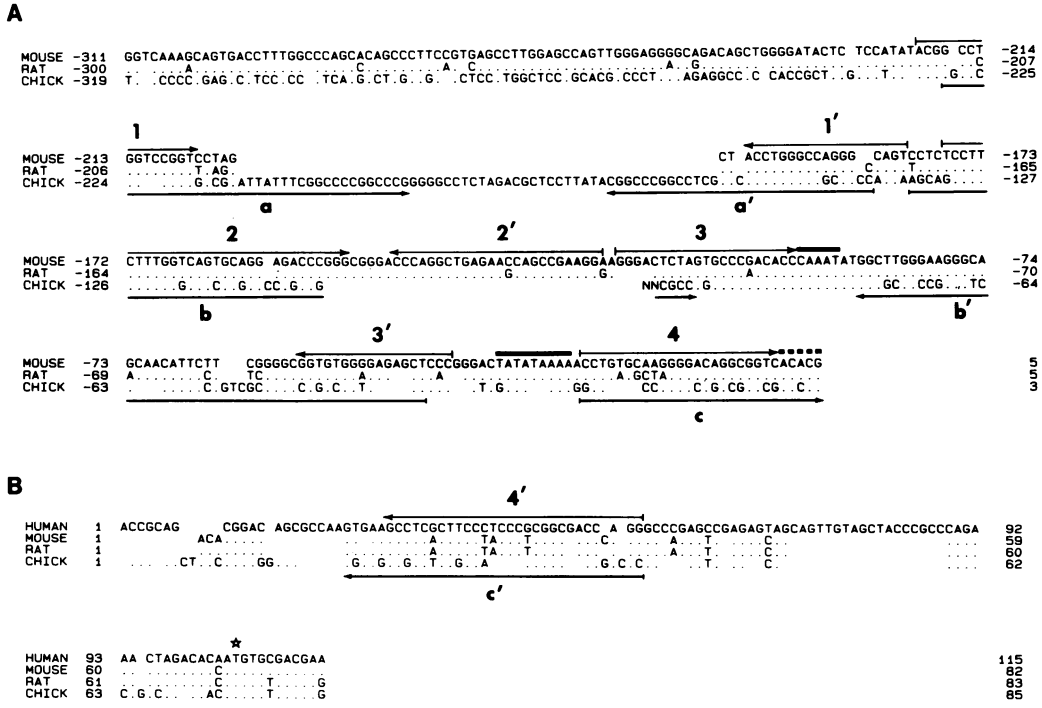


FIG. 6. Comparison of the nucleotide sequence of the 5′-flanking and the 5′-untranslated regions of vertebrate skeletal α-actin genes. (A) Alignment of the 5′-flanking region sequences of mouse, rat, and chicken skeletal α-actin genes. (B) Alignment of the 5′-untranslated region sequences of human, mouse, rat, and chicken skeletal α-actin genes. Dots indicate identity with the first sequence listed. Blanks indicate that gaps have been introduced during the alignment for maximal homology. The CAAAT box and TATA box are highlighted with solid bars. A broken bar indicates the transcription initiation site, and an asterisk indicates the initiation codon ATG. Horizontal arrows above the sequences represent the adjacent inverted complementary sequences of rodents which are indicated numerically, whereas horizontal arrows underneath the sequences represent the adjacent repeats of chicken sequences, which are indicated alphabetically.

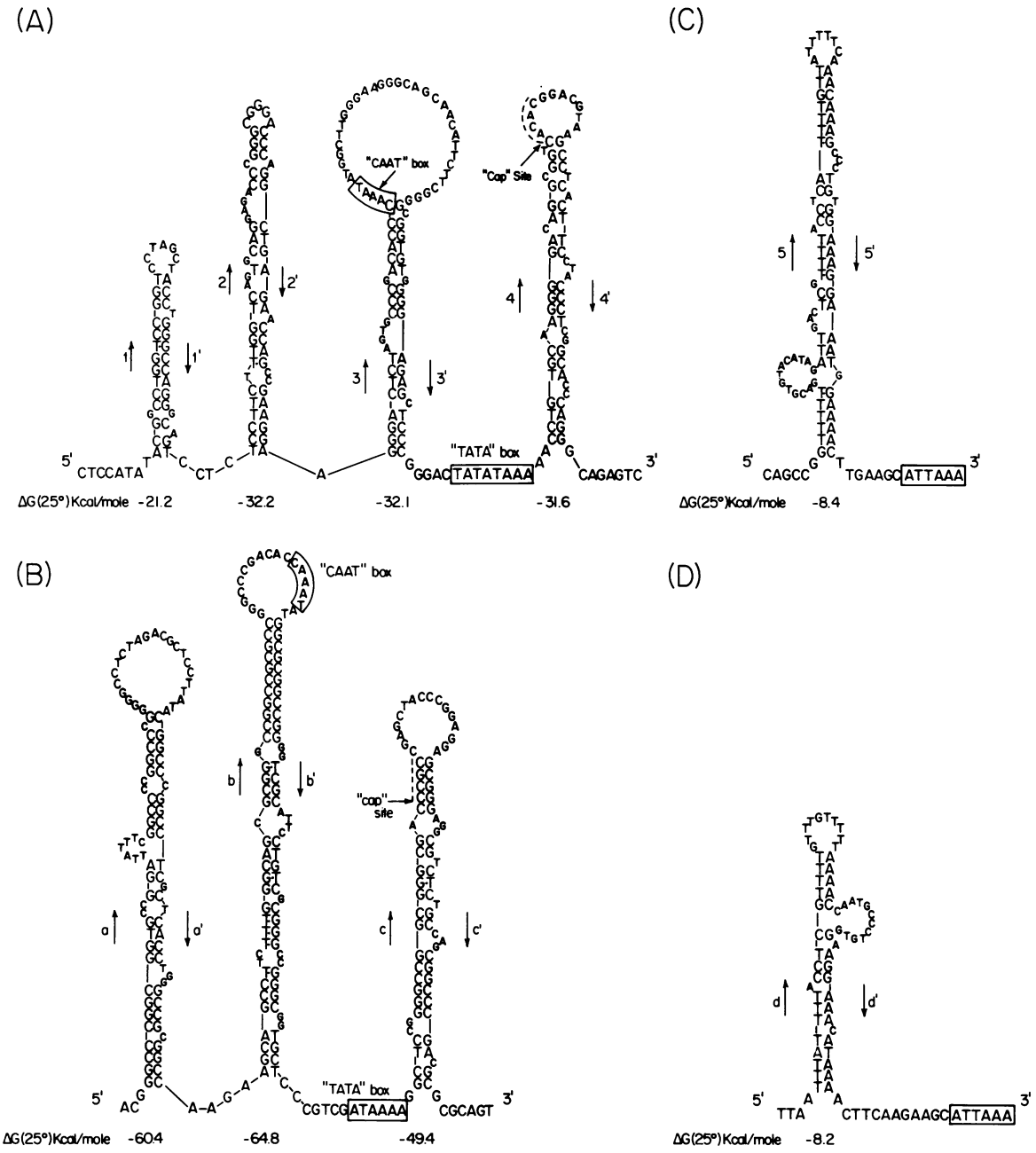


FIG. 7. Predicted inverted repeat structures in the 5'-flanking region and the 5'- and 3'-untranslated regions of the mouse, rat, and chicken skeletal α-actin genes. (A) Four potential configurations are shown in the 5'-flanking and the 5'-untranslated region of the mouse skeletal α-actin gene. Similar configurations can be found in the rat gene. (B) Three potential configurations are shown in the 5'-flanking and the 5'-untranslated region of the chicken skeletal α-actin gene. (C) A potential stem-loop is demonstrated in the 3'-untranslated region of the mouse skeletal α-actin gene. A similar structure can be found in the rat gene. (D) A potential stem-loop is shown in the 3'-untranslated region of the chicken skeletal α-actin gene. The indicated free energy values for the base-paired regions were calculated by the method of Tinoco et al. (50). Note that the free energies for the base-paired regions in the 5'-flanking region were estimated by the same method, assuming that the stacking energies of DNA base pairs are similar to those of RNA base pairs. CAAAT, TATA, and putative polyadenylation signal ATTAAA are indicated by the boxes.

exons. The sequences at the determined borders of the first intron are in agreement with the consensus splice site sequences (7). Thus, the first intron is 961 nucleotides long and interrupts the 5'-untranslated region 12 nucleotides upstream from the initiator ATG codon.

A canonical promoter sequence TATATAAA (5) was identified at nucleotides −33 to −26 (Fig. 2), and a CAAAT sequence (12) was located 91 nucleotides upstream from the transcription initiation site. The positions of these regulatory sequences in the promoter region correspond well with those of similar sequences found upstream from the 5' cap site of other eucaryotic genes (7).

**Sequence of the 3'-untranslated region of the mouse skeletal α-actin gene.** The location of the polyadenylation site was

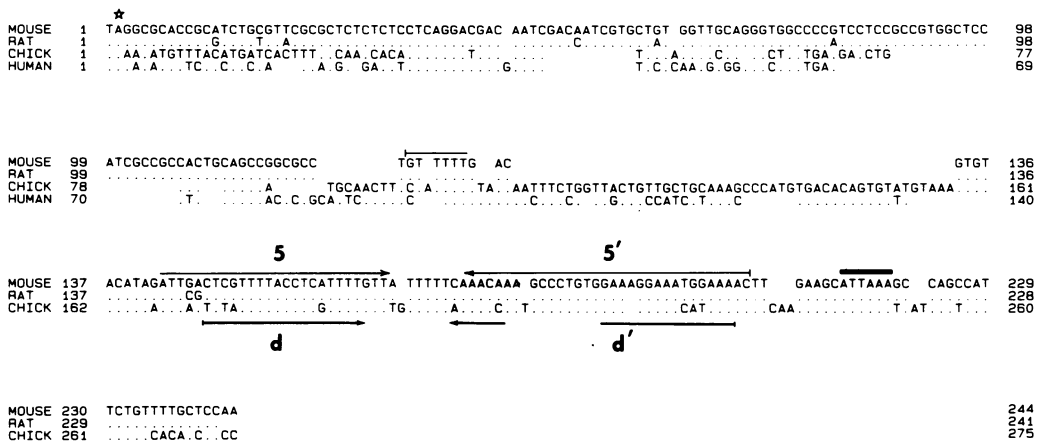


FIG. 8. Comparison of the nucleotide sequence of the 3'-untranslated regions of vertebrate skeletal α-actin genes. Alignment of the 3'-untranslated region sequences of mouse, rat, chicken, and human (partial) skeletal α-actin genes. Symbols are as described in the legend to Fig. 6. The putative polyadenylation signal ATTAAA is highlighted with a solid bar, and the termination codon TAG is indicated with an asterisk.

identified approximately at nucleotide 2944 (±5 nucleotides) of the sequenced gene (Fig. 2) by exonuclease VII mapping (Fig. 3C). The sequence ATTAAA was located 25 nucleotides upstream from the polyadenylation site. In addition, the sequence TGTGTGTGG was found 4 nucleotides downstream from the polyadenylation site, in agreement with the suggestion that a G+T-rich stretch downstream from the polyadenylation site is required for the correct 3'-end formation of mRNA (16, 26). Another potential polyadenylation signal AATAAA was also found at nucleotides 2942 through 2947, about 22 nucleotides downstream from the end of the putative polyadenylation signal ATTAAA identified above. We have no evidence that this second potential signal functions in the mouse.

Thus, the 3'-untranslated region of the mouse skeletal α-actin mRNA is about 245 nucleotides long [excluding the poly(A) tail]. It is about the same size as the 3'-untranslated regions of the rat (241 nucleotides [54]) and human (253 nucleotides [37]) skeletal α-actin mRNA but much shorter than that of rat cytoskeletal β-actin mRNA (670 nucleotides [33]). The length of skeletal α-actin mRNA in mammals is about 1,650 nucleotides (31, 37, 44), including the poly(A) tail. Comparison with the total length of transcribed sequences in the mouse and rat genes (about 1,450 nucleotides) would suggest that the poly(A) tail is about 200 nucleotides long.

**Copy number of the mouse skeletal α-actin gene.** Southern blot analysis of BALB/c genomic DNA digested with four different restriction endonucleases (*Pst*I, *Bgl*II, *Hin*dIII, and *Sst*I) demonstrated that genomic and cosmid fragments hybridizing with the skeletal α 3'-untranslated region-specific probe comigrate (data not shown). This result suggests that the gene is present in single copy in the mouse genome, in agreement with previous reports (29, 30).

## DISCUSSION

**Strong homology and interesting inverted repeat structures in the 5'-flanking region and both the 5'- and 3'-untranslated regions of vertebrate skeletal α-actin genes.** We have aligned the nucleotide sequence of the mouse skeletal α-actin gene with those of the rat and chicken by using percent homology profiles (Fig. 5). All alignments depend on introduction of gaps for maximal homology, and areas of high and low homology between two sequences are displayed as peaks and troughs, respectively. The coding sequences show a very high degree of homology (≥90%), as expected since the proteins themselves are identical. In comparison of rat with mouse, the intron sequences between coding exons are about 75% homologous, except for intron 3, for which the lengths differ by 49 nucleotides. The corresponding introns of the chicken are much more divergent in length and sequence. In comparing rat with mouse, the long intron following the 5' untranslated exon 1 shows sharp peaks of conserved and nonconserved regions.

Figures 5 and 6A show very high conservation (~85%) in the 5'-flanking region between the cap site and 300 nucleotides upstream of rat and mouse skeletal α-actin genes. The homology between chicken and mouse in the same region is also rather high (~60%). Nudel et al. (32) have found a similar degree of homology between the rat and chicken skeletal α-actin genes. We have also found considerable homology in the 5'-untranslated region. By introducing gaps for best alignment (Fig. 6B) there is (i) a high degree of homology between rat and mouse in the 5' untranslated region, (ii) a rather high degree of homology between human and rodents, except for three long inserts in the human gene, and (iii) a moderate degree of homology between chickens and rodents. Conserved sequences between chickens and rats around the CAAT box and about 46 to 59 nucleotides downstream from the cap site have been previously recognized by Ordahl and Cooper (34). A similar comparison of the 5'-untranslated regions of nonmuscle β-actins (human cDNA and rat genomic sequence) also shows a high degree of sequence conservation (38). We find, however, no cross-homology between α- and β-actin 5'-untranslated regions. A number of studies have suggested that the sequence and structure of the mRNA in the 5'-untranslated region have an important role in regulation of translation (11, 36, 40, 46). The fact that there are conserved sequences in the 5'-untranslated region in all of the vertebrate skeletal α-actin genes, but a different set of conserved sequences in the cytoskeletal β-actin genes, suggests that there may be developmentally specific translational regulatory mechanisms in muscle versus nonmuscle cells.

It is striking that a number of inverted repeat structures exist in the 5'-flanking and the 5'-untranslated regions of the rodent and chicken genes. These are indicated as inverted repeats in Fig. 6 and as remarkably stable hairpin structures

for a single strand in Fig. 7A and B. Some of these inverted repeat structures have been conserved between chickens and rodents. The species compared (avian and mammalian) have been separated for more than 250 million years (12), indicating a strong selective constraint to conserve these sequences and suggesting that the sequences may be biologically significant. If the primary transcripts actually initiate at the cap site, the structures shown in Fig. 7A and B would not occur in the RNA. These sequences could function as duplicated transcription factor binding sites, with the bound factors (presumably proteins) having opposite orientation at the two members of an inverted repeat as postulated by McKnight et al. (25) and by Giniger et al. (18). Alternatively, some single-strand DNA regions may be opened up during formation of a transcription bubble, and these hairpins could then form in the DNA as indicated in Fig. 7. Potential hairpin structures have also been found in the 5′-flanking region and the first untranslated exon in the rat cytoskeletal β-actin gene (33). They differ, however, from those found in the skeletal α-actin genes. For example, whereas the TATA box is presented within a loop of the rat β gene, it is found between stem-loop structures in the α genes. The existence of such differences between the cytoskeletal and muscle-specific actin genes raises the possibility that the conserved inverted repeats in the α genes are important for tissue-specific expression. To our knowledge, this is the first description of inverted repeats and possible secondary structure formation among the highly conserved sequences in the 5′-flanking and the 5′-untranslated regions of vertebrate skeletal α-actin genes.

A long sequence of about 110 nucleotides, including the putative polyadenylation signal ATTAAA, is highly conserved in the 3′-untranslated region of vertebrate skeletal α-actin genes (Fig. 8). It is noteworthy that two blocks of these highly conserved sequences in the 3′-untranslated region can form a stem-loop structure with estimated stabilities of −8.4 kcal/mol for mice and rats (Fig. 7C) and −8.2 kcal/mol for chickens (Fig. 7D). These structures are immediately upstream from the ATTAAA polyadenylation signal. It has been suggested that the inverted repeat at the 3′ end of sea urchin histone mRNA is important for the generation of the histone mRNA 3′ termini (5, 6). This putative structure does not act as a DNA cruciform, but exerts its function at the level of the RNA transcripts (4). It is conceivable that the potential hairpin structure upstream from the putative polyadenylation signal in skeletal α-actin plays a role in the correct 3′-end formation of skeletal α-actin mRNA.

**Interesting features in the first intron of vertebrate skeletal α-actin genes.** Rat and mouse skeletal α-actin genes have a long first intron compared with chickens (i.e., 976, 961, and 111 nucleotides for rats, mice, and chickens, respectively). There are several highly conserved sequences in this intron between rats and mice, but the introns are quite divergent in other regions (Fig. 5A). The chicken first intron is quite G+C rich (82%) compared with rat introns (53.5%) and mouse introns (52%). We have been unable to find any sequence homology between the chicken and rodent introns. There are two inverted duplications within the rodent first intron. Both occur within the conserved sequences (arrows in Fig. 5A). Furthermore, an inverted repeat can also be found in the chicken first intron (shown as a hairpin in Fig. 9C). The remarkably stable hairpin structures in the vertebrate first intron (Fig. 9) may form in the primary transcript. If the splicing apparatus tracks along the intron in search of splice sites, it may be able to pass along the base of such hairpins. The hairpins would effectively shorten the intron, thereby

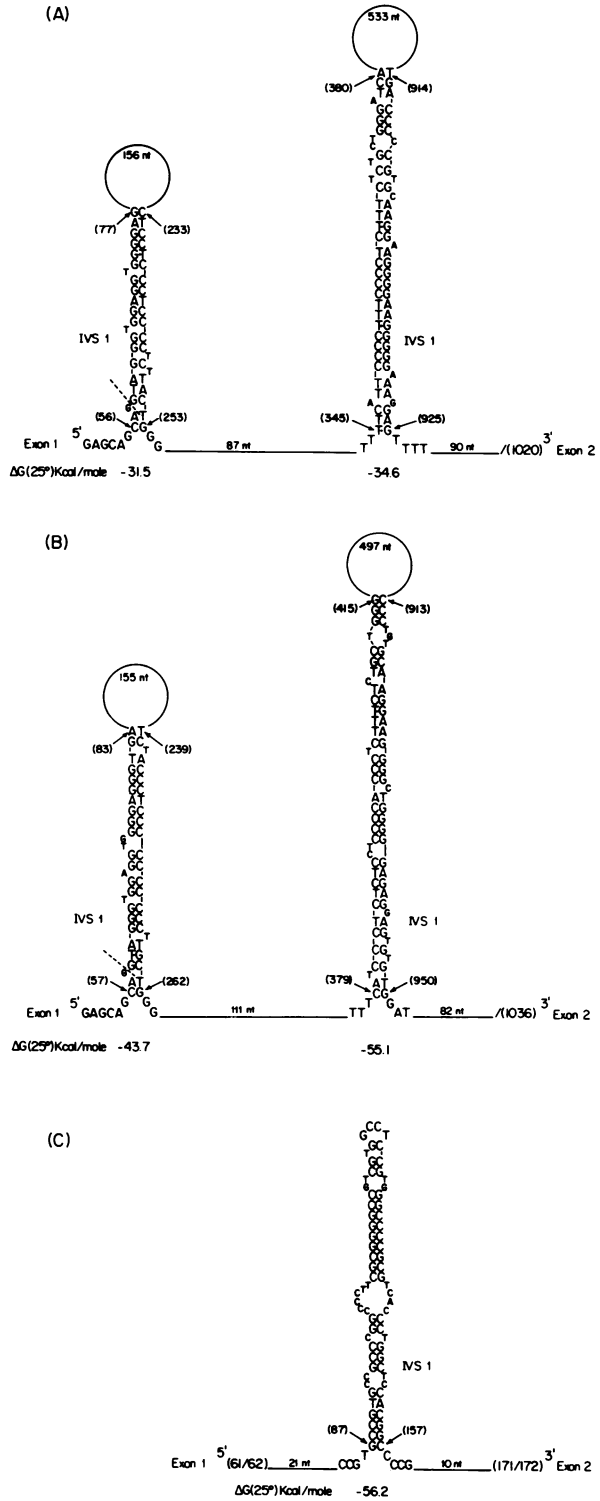


FIG. 9. Predicted inverted repeat structures in the first intron of rodents and chicken skeletal α-actin genes. (A) Two potential configurations in the first intron of mouse skeletal -actin gene. (B) Similar configurations in the first intron of rat skeletal α-actin gene. (C) One potential hairpin loop in the first intron of chicken skeletal α-actin gene. The indicated free energy values for the base-paired regions were calculated by the method of Tinococ et al. (50). Numbers within parentheses indicate the numbers of nucleotides downstream from the transcription initiation site.

expediting its excision by splicing. A similar mechanism has been proposed as one way to explain intermolecular splicing between two RNAs base paired through their intron sequences (47). Alternatively, we speculate that these inverted repeats may play a role of transcriptional enhancement in the regulation of tissue-specific gene expression because it has been proposed that a tissue-specific transcription enhancer element is located in the intron of a heavy-chain immunoglobulin gene (17). In fact, the putative "core" sequence in the heavy-chain gene intron is present as an inverted repeat (17).

In conclusion, we have found several conserved and inverted repeat sequences outside of the protein-coding region of the skeletal α-actin genes. It would be interesting to investigate whether these conserved inverted repeat sequences serve as regulatory elements in differentiated muscle cells.

## ACKNOWLEDGMENTS

We are grateful to Michael Steinmetz for generously providing a BALB/c genomic cosmid library, to Charles P. Ordahl for the chicken skeletal α-actin genomic clone, to Leonard I. Garfinkel for a rat α-actin cDNA, to Beverley J. Bond for the *Drosophila* actin DmA2 genomic clone, and to Tim Hunkapiller for BALB/c genomic DNA and help with the computer analyses. We also thank Joan Steitz for valuable discussions and David Solnick, Bruce J. Nicholson, Terry P. Snutch, Robert Little, and Nevis L. Fregien for their helpful comments on this manuscript.

This work was supported by a Public Health Service research grant from the National Institutes of Health.

## LITERATURE CITED

1. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69:**1408–1412.
2. **Berk, A. J., and P. A. Sharp.** 1978. Spliced early mRNAs of simian virus 40. Proc. Natl. Acad. Sci. USA **75:**1274–1278.
3. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and $^{35}S$ label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80:**3963–3965.
4. **Birchmeier, C., W. Folk, and M. L. Birnstiel.** 1983. The terminal RNA stem-loop structure and 80 bp of spacer DNA are required for the formation of 3′ termini of sea urchin H2A mRNA. Cell **35:**433–440.
5. **Birchmeier, C., R. Grosschedl, and M. L. Birnstiel.** 1982. Generation of authentic 3′ termini of an H2A mRNA *in vivo* is dependent on a short inverted DNA repeat and on spacer sequences. Cell **28:**739–745.
6. **Birchmeier, C., D. Schumperli, G. Sconzo, and M. L. Birnstiel.** 1984. 3′ editing of mRNAs: sequence requirements and involvement of a 60-nucleotide RNA in maturation of histone mRNA precursors. Proc. Natl. Acad. Sci. USA **81:**1057–1061.
7. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50:**349–383.
8. **Chang, K. S., K. N. Rothblum, and R. J. Schwartz.** 1985. The complete sequence of the chicken α-cardiac actin gene: a highly conserved vertebrate gene. Nucleic Acids Res. **13:**1223–1237.
9. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemstry **18:** 5294–5299.
10. **Cooper, A. D., and W. R. Crain.** 1982. Complete nucleotide sequence of a sea urchin actin gene. Nucleic Acids Res. **10:**4081–4092.
11. **Darveau, A., J. Pelletier, and N. Sonenberg.** 1985. Differential efficiencies of *in vitro* translation of mouse c-myc transcripts differing in the 5′ untranslated region. Proc. Natl. Acad. Sci. USA **82:**2315–2319.
12. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connel, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21:**653–668.
13. **Fornwald, J. A., G. Kuncio, I. Peng, and C. P. Ordahl.** 1982. The complete nucleotide sequence of the chick α-actin gene and its evolutionary relationship to the actin gene family. Nucleic Acids Res. **10:**3861–3876.
14. **Fyrberg, E. A., B. J. Bond, N. D. Hershey, K. S. Mixter, and N. Davidson.** 1981. The actin genes of *Drosophila*: protein coding regions are highly conserved but intron positions are not. Cell **24:**107–116.
15. **Garfinkel, L. I., M. Periasamy, and B. Nadal-Ginarcho.** 1982. Cloning and characterization of cDNA sequences corresponding to myosin light chains 1, 2, and 3, troponin-C, troponin-T, α-tropomyosin, and α-actin. J. Biol. Chem. **257:**11078–11086.
16. **Gil, A., and N. J. Proudfoot.** 1984. A sequence downstream of AAUAAA is required for rabbit β-globin mRNA 3′-end formation. Nature (London) **312:**473–474.
17. **Gillies, S. D., S. L. Morrison, V. T. Oi, and S. Tonegawa.** 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell **33:**717–728.
18. **Giniger, E., S. M. Varnum, and M. Ptashne.** 1985. Specific DNA binding of GAL4, a positive regulatory protein of yeast. Cell **40:**767–774.
19. **Hamada, H., M. G. Petrino, and T. Kakunaga.** 1982. Molecular structure and evolutionary origin of human cardiac muscle actin gene. Proc. Natl. Acad. Sci. USA **79:**5901–5905.
20. **Hanauer, A., M. Levin, R. Heilig, D. Daegelen, A. Kahn, and J. L. Mandel.** 1983. Isolation and characterization of cDNA clones for human skeletal muscle α-actin. Nucleic Acids Res. **11:**3503–3516.
21. **Hunkapiller, M., S. Kent, M. Caruthers, W. Dreyer, J. Firca, C. Giffin, S. Horvath, T. Hunkapiller, P. Tempst, and L. Hood.** 1984. A microchemical facility for the analysis and synthesis of genes and proteins. Nature (London) **310:**105–111.
22. **Kost, T. A., N. Theodorakis, and S. H. Hughes.** 1983. The nucleotide sequence of the chick cytoplasmic β-actin gene. Nucleic Acids Res. **11:**8287–8301.
23. **Lu, R. C., and M. Elzinga.** 1977. Partial amino acid sequence of brain acin and its homology with muscle actin. Biochemistry **16:**5801–5806.
24. **McDonell, M. W., M. N. Simon, and F. W. Studier.** 1977. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. **110:**119–146.
25. **McKnight, S. L., R. C. Kingsbury, A. Spence, and M. Smith.** 1984. The distal transcription signals of the herpesvirus tk gene share a common hexanucleotide control sequence. Cell **37:** 253–262.
26. **McLauchlan, J., D. Gaffney, J. L. Whitton, and J. B. Clements.** 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini. Nucleic Acids Res. **13:**1347–1368.
27. **Melloul, D., B. Aloni, J. Calvo, D. Yaffe, and U. Nudel.** 1984. Developmentally regulated expression of chimeric genes containing muscle actin DNA sequences in transfected myogenic cells. EMBO J. **3:**983–990.
28. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Manistis, K. Zinn, and M. R. Green.** 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12:**7035–7056.
29. **Minty, A. J., S. Alonso, M. Caravatti, and M. E. Buckingham.** 1982. A fetal skeletal muscle actin mRNA in the mouse and its identity with cardiac actin mRNA. Cell **30:**185–192.
30. **Minty, A. J., S. Alonso, J. L. Guenet, and M. E. Buckingham.** 1983. Number and organization of actin-related sequences in the mouse genome. J. Mol. Biol. **167:**77–101.
31. **Minty, A. J., M. Caravatti, B. Robert, A. Cohen, P. Daubas, A. Weydert, F. Gros, and M. E. Buckingham.** 1981. Mouse actin

messenger RNAs. J. Biol. Chem. **256**:1008–1014.
32. **Nudel, U., D. Greenberg, C. P. Ordahl, O. Saxel, S. Newman, and D. Yaffe.** 1985. Developmentally regulated expression of a chick muscle-specific gene in stably transfected rat myogenic cells. Proc. Natl. Acad. Sci. USA **82**:3106–3109.
33. **Nudel, U., R. Zakut, M. Shani, S. Neuman, Z. Levy, and D. Yaffe.** 1983. The nucleotide sequence of the rat cytoplasmic α-actin gene. Nucleic Acids Res. **11**:1759–1771.
34. **Ordahl, C. P., and T. A. Cooper.** 1983. Strong homology in promoter and 3′-untranslated regions of chick and rat α-actin genes. Nature (London) **303**:348–349.
35. **Ordahl, C. P., S. M. Tilghman, C. Ovitt, J. Fornwald, and M. T. Largen.** 1980. Structure and developmental expression of the chick α-actin gene. Nucleic Acids Res. **8**:4989–5005.
36. **Pelletier, J., and N. Sonenberg.** 1985. Insertion mutagenesis to increase secondary structure within the 5′ noncoding region of a eukaryotic mRNA reduces translational efficiency. Cell **40**: 515–526.
37. **Ponte, P., P. Gunning, H. Blau, and L. Kedes.** 1983. Human actin genes are single copy for α-skeletal and α-cardiac actin but multicopy for β- and γ-cytoskeletal genes: 3′ untranslated regions are isotype specific but are conserved in evolution. Mol. Cell. Biol. **3**:1783–1791.
38. **Ponte, P., S. Y. Ng, J. Engel, P. Gunning, and L. Kedes.** 1984. Evolutionary conservation in the untranslated regions of actin mRNAs: DNA sequence of a human beta-actin cDNA. Nucleic Acids Res. **12**:1687–1696.
39. **Purrello, M., and I. Balazs.** 1983. Direct hybridization of labeled DNA to DNA in agarose gels. Anal. Biochem. **128**:393–397.
40. **Saito, H., A. C. Hayday, K. Wiman, W. S. Hayward, and S. Tonegawa.** 1983. Activation of the c-myc gene by transformation: a model for translational control. Proc. Natl. Acad. Sci. USA **80**:7476–7480.
41. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
42. **Schubert, D., A. J. Harris, C. E. Devine, and S. Heinemann.** 1974. Characterization of a unique muscle cell line. J. Cell Biol. **61**:398–413.
43. **Schuler, M. A., P. McOsker, and E. B. Keller.** 1983. DNA sequence of two linked actin genes of sea urchin. Mol. Cell. Biol. **3**:448–456.
44. **Shani, M., U. Nudel, D. Zevin-Sonkin, R. Zakut, D. Givol, D. Katcoff, Y. Carmon, A. M. Frischauf, and D. Yaffe.** 1981. Skeletal muscle actin mRNA. Characterization of the 3′ untranslated region. Nucleic Acids Res. **9**:579–589.
45. **Shani, M., D. Zevin-Sonkin, Y. Carmon, H. Czosnek, U. Nudel, and D. Yaffe.** 1982. Changes in myosin and actin gene expression and DNase-I sensitivity associated with terminal differentiation of myogenic cultures, p. 189–200. *In* M. L. Pearson and H. F. Epstein (ed.), Muscle development. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
46. **Shih, C-K., M. Linial, M. M. Goodnow, and W. S. Hayward.** 1984. Nucleotide sequence 5′ of the chicken c-myc coding region: localization of a noncoding exon that is absent from myc transcripts in most avian leukosis virus-induced lymphomas. Proc. Natl. Acad. Sci. USA **81**:4697–4701.
47. **Solnick, D.** 1985. Trans splicing of mRNA precursors. Cell **42**:157–164.
48. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
49. **Steinmetz, M., A. Winoto, K. Minard, and L. Hood.** 1982. Clusters of genes encoding mouse transplantation antigens. Cell **28**:489–498.
50. **Tinoco, I. J., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **248**:40–41.
51. **Vandekerckhove, J., and K. Weber.** 1978. Mammalian cytoplasmic actins are the products of at least two genes and differ in primary structure in at least 25 identified positions from skeletal muscle actins. Proc. Natl. Acad. Sci. USA **75**:1106–1110.
52. **Vandekerckhove, J., and K. Weber.** 1979. The complete amino acid sequence of actins from bovine aorta, bovine heart, bovine fast skeletal muscle and rabbit slow skeletal muscle. Differentiation **14**:123–133.
53. **Vandekerckhove, J., and K. Weber.** 1984. Chordate muscle actins differ distinctly from invertebrate muscle actins. The evolution of the different vertebrate muscle actins. J. Mol. Biol. **179**:391–413.
54. **Zakut, R., M. Shani, D. Givol, S. Neuman, D. Yaffe, and U. Nudel.** 1982. Nucleotide sequence of the rat skeletal muscle actin gene. Nature (London) **298**:857–859.