MOLECULAR AND CELLULAR BIOLOGY, Oct. 1985, p. 2720–2732
0270-7306/85/102720-13$02.00/0

Vol. 5, No. 10

Evolution of the Functional Human β-Actin Gene and Its Multi-Pseudogene Family: Conservation of Noncoding Regions and Chromosomal Dispersion of Pseudogenes

SUN-YU NG,[1,2] PETER GUNNING,[1] ROGER EDDY,[3] PHYLLIS PONTE,[1†] JOHN LEAVITT,[2] THOMAS SHOWS,[3] AND LARRY KEDES[1*]

The MEDIGEN Project, Department of Medicine, Stanford University School of Medicine and Veterans Administration Medical Center, Palo Alto, California 94305[1]; Armand Hammer Cancer Research Center, Linus Pauling Institute of Science and Medicine, Palo Alto, California 94306[2]; and Department of Human Genetics, Roswell Park Memorial Institute, New York State Department of Health, Buffalo, New York 14263[3]

Received 10 April 1985/Accepted 13 June 1985

We have assigned six members of the human β-actin multigene family to specific human chromosomes. The functional gene, *ACTB*, is located on human chromosome 7, and the other assigned β-actin-related sequences are dispersed over at least four different chromosomes including one locus assigned to the X chromosome. Using intervening sequence probes, we showed that the functional gene is single copy and that all of the other β-actin related sequences are recently generated in evolution and are probably processed pseudogenes. The entire nucleotide sequence of the functional gene has been determined and is identical to cDNA clones in the coding and 5′ untranslated regions. We have previously reported that the 3′ untranslated region is well conserved between humans and rats (Ponte et al., Nucleic Acids Res. 12:1687–1696, 1984). Now we report that four additional noncoding regions are evolutionarily conserved, including segments of the 5′ flanking region, 5′ untranslated region, and, surprisingly, intervening sequences I and III. These conserved sequences, especially those found in the introns, suggest a role for internal sequences in the regulation of β-actin gene expression.

Cytoskeletal β-actin is one of the most abundant cellular proteins found in mammalian and avian nonmuscle cells. It is the major component, together with cytoskeletal γ-actin, of the microfilamentous structures found in these cells (4). Functionally, cytoskeletal actin has been implicated in intracellular movement of organelles, cytokinesis, and cell motility (55). In addition, sequential mutations in the β-actin protein have been associated with a parallel increase in the tumorigenicity of human cells (25). However, the precise mechanism of β-actin function in these cellular processes remains enigmatic.

One approach to a better understanding of the functions of β-actin involves the reintroduction of specifically modified β-actin genes into nonmuscle cells to observe the effects of the programmed alterations. This approach first required isolation and characterization of the gene. In addition, since the human β-actin gene is a member of a large multigene family, a prerequisite to interpreting the results of such studies requires a full accounting of the number of functional genes. There are at least 20 different β-actin gene sequences in the human genome (41), and similarly sized familes are found in the rat and mouse (41; Gunning, unpublished data). Sequence analysis has demonstrated that several of the human β-actin genes are processed pseudogenes (34; J. N. Engel, Ph.D. thesis, Stanford University, Palo Alto, Calif., 1982). This, in addition to other data, has led us to propose that the majority of the human (and rodent) β-actin genes are in fact pseudogenes (41). However, it is difficult to exclude the possibility that more than one functional β-actin gene exists. Nucleotide sequencing of all the human β-actin genes would solve this problem, but is impractical. We have therefore sought an alternative strategy to address these issues.

As a first step, we recently isolated an intron-containing human β-actin gene that is expressed in human fibroblasts (26). In this report we describe our studies of the chromosomal distribution of this gene and the other related β-actin sequences and present a complete structural analysis for the human β-actin gene. We demonstrate that there is but a single functional human β-actin gene, that it is located on chromosome 7, and that the remaining copies are intronless, have been generated recently, and have been randomly integrated into the genome.

We have determined the entire nucleotide sequence of the human β-actin gene and compared it with the corresponding sequences of the rat and chicken genes. Several noncoding regions of these genes, including the 5′ flanking regions and two of the five introns, are under strong evolutionary pressure to retain specific sequences. Conserved segments of an intron in the 5′ untranslated region (UTR) resemble potential transcription enhancer elements. This strong sequence conservation may reflect evolutionary pressure to retain a particularly strong transcription promoter function, since β-actin is one of the most abundant cellular transcripts in mammals and birds.

MATERIALS AND METHODS

Chromosome mapping. Parental cell lines, construction of human-mouse somatic cell hybrids, and identification of human chromosomes within the hybrids have been described previously (15, 47). DNA was isolated from hybrid cells and their parents, digested with *Eco*RI, size fractionated on 0.8% agarose gels, and transferred to nitrocellulose as described previously (35). Two DNA probes were used for chromosome

* Corresponding author.
† Address: California Biotechnology Inc., Mountain View, CA 94043.

mapping. A 1.5-kilobase (kb) *Sac*I DNA fragment was isolated from the functional human β-actin gene (26) and used as a probe to specifically detect this gene. The second probe was derived from the 3′ UTR of the human β-actin cDNA as a 118-base-pair *Hae*III DNA fragment which was used to specifically detect all of the human β-actin-related sequences.

Construction of 3′ UTR subclone. Three *Hae*III DNA fragments were isolated from the previously subcloned 3′ UTR of a human β-actin cDNA (pHFβA-3′UT) (41). Fragment 1 extends from codon 366 to base 211 in the 3′ UTR; fragment 2 covers base 212 to base 329 in the 3′ UTR, and fragment 3 extends from base 330 through the end of the 3′ UTR (base 594) to the vector *Bam*HI site. The middle 118-base-pair *Hae*III DNA fragment was directly cloned into the *Sma*I site of pHP34 (43). This subclone is denoted pHFβA-3′UT-HH. For convenience, the probe derived from this subclone is referred to as human β-actin specific (HβAS). The DNA fragment was isolated from this subclone by digestion with *Eco*RI, and the purified fragment was self-ligated by incubation with T4 ligase (15) before nick translation.

Construction of intervening sequence (IVS) region subclones. Restriction fragments of the human β-actin gene clone pM1(β1)-2 were subcloned by blunt-end ligation of the appropriate DNA fragments into the *Sma*I site of pHP34 (43) as described previously (41). Restriction fragments with protruding 5′ ends were converted to blunt ends by using the Klenow fragment of *Escherichia coli* DNA polymerase I. Restriction fragments with protruding 3′ ends were converted to blunt ends by using the 3′ exonuclease activity of bacteriophage T4 DNA polymerase. The subcloned DNA fragments (see Fig. 3 for map locations) were recovered from the plasmids after digestion with the endonuclease *Eco*RI. The fragments were purified by electrophoresis and self-ligated before nick translation.

DNA isolation and Southern blot analysis. DNA was isolated from HeLa cells and digested with various restriction enzymes as described previously (7). Digested DNA was size fractionated on 0.7% agarose gels (7) and transferred to nitrocellulose (50). Nick-translated DNA fragments (45), 10^8 dpm/μg, were hybridized to nitrocellulose filters and washed exactly as described previously (15).

Nuclease S1 mapping. The 400-base-pair *Bst*NI-*Bst*NI DNA fragment spanning the putative mRNA cap site was end labeled with [γ-^{32}P]ATP (ICN Pharmaceuticals Inc.) in the presence of polynucleotide kinase (New England Nuclear Corp.) according to standard procedures (30). The end-labeled DNA fragment was then digested with endonuclease *Xho*I. A 127-base-pair *Xho*I-*Bst*NI* DNA fragment, purified from an agarose gel with Schleicher & Schuell NA-45 DEAE membrane, was divided into two samples. The first sample was used for DNA sequencing (30). The second sample, after denaturation at 80°C, was hybridized to total RNA from human cells at 60°C for 3 h in hybridization buffer containing 80% formamide, 40 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (pH 6.4), 0.4 M sodium chloride and 1 mM EDTA (pH 8.0). After hybridization, ice-cold nuclease S1 buffer and 1,000 U of nuclease S1 (Sigma Chemical Co.) were added to each sample immediately. After incubation at 37°C for 15 min, the reaction was stopped by the addition of ammonium acetate and EDTA. Protected DNA was precipitated with an equal volume of isopropanol, washed with 70% ethanol, vacuum dried, and dissolved in 80% formamide, 0.05% bromphenol blue-xylene cyanol, and 1 mM EDTA. Samples were elctrophoresed at 40 V cm^{-1} on 8% acrylamide–8 mM urea thin sequencing gels in parallel with chemical sequencing cleavage fragments.

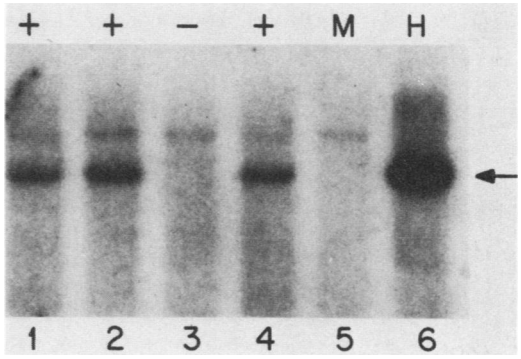

FIG. 1. Detection of the functional human β-actin gene (*ACTB*) in human-mouse somatic cell hybrids. DNA was isolated from somatic cell hybrids (lanes 1 through 4) and the parental mouse (lane 5) and human (lane 6) cell lines. After digestion with *Eco*RI, the DNA was electrophoresed on a 0.8% agarose gel, transferred to nitrocellulose, and hybridized with the human β-actin gene-specific probe (*Sac*I fragment, see Materials and Methods). The human β-actin gene is located on a 14-kb DNA fragment (indicated by the arrow), and the mouse DNA fragment migrating more slowly on the gel is approximately 20 kb in size.

DNA sequencing. Restriction fragments were sequenced by the method of Maxam and Gilbert (30). Sequence data were then managed with the GEL program (IntelliGenetics, Inc.).

DNA sequence alignments. Comparison of DNA sequence data was managed with the IFIND program (IntelliGenetics, Inc.) based on the Wilbur and Lipman algorithm (56). All DNA sequence alignments were carried out with a gap penalty setting of 4.

RESULTS

Chromosomal location of the human β-actin gene. We have previously reported the isolation of the human β-actin gene as a πAN7 recombinant from a bacteriophage library of human fetal DNA and demonstrated that it is an expressed gene (26). This functional human β-actin gene is located on a 14-kb *Eco*RI DNA fragment in the human genome. A 6.5-kb fragment of the phage clone, containing the complete coding region and about 2 kb of 5′ flanking DNA, has been subcloned into the *Eco*RI site of pBR322. We derived a 1.5-kb *Sac*I DNA fragment from the cloned gene which covers the region from about 450 to 2,000 base pairs 5′ of the mRNA cap site. This DNA fragment hybridizes to a 14-kb DNA band in *Eco*RI-digested human DNA and to a 20-kb DNA band in *Eco*RI-digested mouse DNA (lanes 5 and 6, Fig. 1). Using this DNA probe, we were able to follow the segregation of the human β-actin gene with human chromosomes in human-mouse somatic cell hybrids.

Thirty-two human-mouse somatic cell hybrids were tested for the coordinate presence of the human β-actin gene (*ACTB*) and a specific human chromosome and chromosome-specific isozyme markers. After electrophoresis and Southern blotting of genomic cell hybrid DNA, the hybrids were scored for the presence or absence of the human gene (Fig. 1). Whereas all hybrid cell lines displayed a 20-kb mouse β-actin DNA fragment (lanes 1 through 4, Fig. 1), only a subset contained the 14-kb human *ACTB* band (lanes 1, 2, and 4, Fig. 1). Table 1 shows that in cell hybrids with different numbers and combinations of human chromosomes, all human chromosomes except chromosome 7 seg-

TABLE 1. Distribution of *ACTB* and human chromosomes in human-mouse cell hybrids

Cell hybrids[a]	*ACTB*[b]	Presence of human chromosome[c]:																							Trans-location[d]
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	
WIL-1	−	−	−	−	−	−	−	−	+	−	−	−	−	−	+	−	−	+	−	−	−	+	−	+	
WIL-5	−	−	−	−	+	−	−	−	−	−	+	−	−	−	−	−	−	+	+	−	−	+	−	+	
WIL-6	+	−	+	−	+	+	+	+	+	−	+	+	−	−	+	−	−	+	−	+	+	+	−	+	
WIL-7	−	−	+	+	−	+	+	−	+	−	+	+	−	+	+	−	−	+	+	−	−	+	−	+	
WIL-8	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	−	+	
WIL-8X	+	−	−	+	+	+	−	+	+	−	+	+	+	−	+	−	−	+	+	+	+	+	−	+	
WIL-14	+	+	−	+	−	+	−	+	+	−	+	−	+	−	+	+	−	+	−	−	−	−	−	+	
WIL-15	+	−	+	+	+	−	−	+	−	−	+	+	+	+	+	−	−	+	+	−	+	+	−	+	
REW-5	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	−	+	+	+	−	+	+	+	
REW-7	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	−	+	+	+	+	+	+	+	
REW-10	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	−	+	
REW-15	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	−	+	+	+	+	+	+	+	
ICL-15	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	+	−	−	+	+	−	−	
DUA-1CSAzF	+	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	
DUA-3BSAgA	+	−	+	−	−	−	−	+	+	−	−	−	−	+	+	−	−	+	−	−	−	−	−	−	
DUA-5BSAgA	−	−	−	+	−	+	−	−	−	−	−	+	−	−	+	−	−	+	+	−	−	+	−	−	
NSL-5	−	+	−	−	−	−	−	−	−	−	+	−	+	−	+	−	+	−	+	−	+	−	−	−	17/9 12q+
NSL-16	+	−	−	+	+	+	−	+	−	−	+	−	+	−	+	+	+	+	+	−	+	−	−	−	17/9
SIR-8	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	−	−	+	+	+	
JWR-26C	+	−	+	+	+	+	+	+	−	+	+	+	+	−	+	+	+	−	+	−	+	+	−	+	1/2
JVR-22	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	2/1
XTR-2	−	−	−	−	−	+	−	−	+	−	+	−	+	+	−	−	−	−	+	−	+	+	−	−	3/X
XTR-3BSAgB	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	+	−	−	−	3/X 10q−
XTR-11BSAgA	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	20p+
TSL-1	−	−	−	+	+	−	−	−	−	−	+	+	−	+	+	−	+	+	+	−	+	−	−	−	
TSL-2	−	−	+	−	−	+	+	−	−	−	+	−	+	−	−	−	−	−	+	−	+	+	−	+	17/3
XER-7	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	−	+	+	−	−	−	+	11/X
EXR-5CSAz	+	+	−	+	+	+	+	+	+	+	+	−	+	+	+	+	−	+	+	+	+	+	+	−	X/11
REX-11BSAgB	−	−	−	+	−	−	−	−	−	−	+	−	−	−	+	+	−	−	+	−	−	−	+	−	
ATR-13	+	+	+	+	+	+	+	+	+	−	+	−	+	+	+	+	+	+	+	+	−	−	−	−	5/X
DUM-13	+	+	+	+	−	+	+	+	−	−	+	+	+	−	+	−	+	+	+	+	+	+	+	−	X/15 15/X
JSR-17S	+	+	+	+	−	+	−	−	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	7/9
% Discordancy		22	22	25	22	25	31	0	25	44	28	25	25	31	22	19	44	28	38	25	41	44	41	25	

[a] Human-mouse somatic cell hybrids were isolated and characterized as described previously (47).
[b] The β-actin probe recognizing functional gene sequences was scored in cell hybrid genomic DNA after *Eco*RI digestion.
[c] Human chromosomes and enzyme markers already assigned to specific chromosomes were tested for all cell hybrids.
[d] Translocation chromosomes were well characterized rearrangements derived from reciprocal translocations observed in human parental cells.

regated discordantly with *ACTB*. The cell hybrid DUA-1 CSAzF, which retained only human chromosome 7 on a mouse background, was positive for *ACTB*, as was cell hybrid JSR-17 (48), which retained a 7/9 translocation (7pter→7q22::9p24→9pter) and the 7pter→7q22 region. This locates the *ACTB* gene in the pter→q22 region on human chromosome 7.

Chromosomal distribution of human β-actin-related sequences. The human genome contains at least 19 β-actin-related sequences in addition to the functional gene. It is of particular interest to determine the origin of these sequences. The two most likely explanations involve either duplications of the functional gene or generation of reverse-transcript processed pseudogenes. In the latter case processed genes might also be duplicated to further expand the gene family. Evaluation of the contribution of tandem duplication to the generation of the human β-actin gene family can be obtained from chromosome linkage analysis. If groups of these β-actin sequences are closely linked, then it is likely that they have resulted from tandem duplications. Conversely, if few or no β-actin gene sequences are linked, it is more likely that tandem duplications have had little or no role in the generation of this gene family. Earlier investigations (49) regarding the chromosomal dispersion of actin genes, relying on in situ hybridization, had concluded that the genes were dispersed. Since the probes used in those experiments detected all actin-coding sequences and did not distinguish among the actin isotypes, the issue of isotype clustering versus dispersal could not be addressed.

We examined the segregation of these additional β-actin sequences in human-mouse somatic cells containing different human chromosomes. To accomplish this we constructed a DNA probe that recognizes all of the human β-actin sequences but not those of the mouse. Since the 3′ UTR of the human β-actin gene is strongly conserved between humans and rodents (41, 42), DNA from the complete 3′ UTR did not discriminate between the human and mouse sequences. We therefore cut the human β-actin 3′ UTR into several DNA fragments by using the restriction endonuclease *Hae*III and examined the ability of the resulting fragments to hybridize with *Eco*RI-digested human and mouse DNA. Probes derived from the 5′ and 3′ ends of the human β-actin 3′ UTR hybridized strongly to all of the human and mouse β-actin genomic sequences (data not shown). However, a 118-base-pair *Hae*III DNA fragment derived from the middle of the β-actin 3′ UTR hybridized to all 20 human *Eco*RI β-actin gene sequences, but not to any derived from the mouse DNA (Fig. 2, lanes R and S). We designated this human β-actin-specific probe as HβAS.

The results of this experiment not only provided probes that would allow us to address the chromosome distribution of the β-actin-related sequences, but also provided an important insight into the evolution of these sequences. Since HβAS hybridizes to all the human β-actin sequences but not

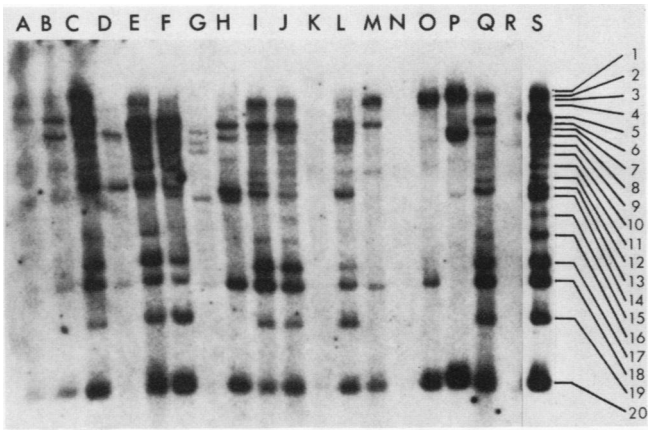

FIG. 2. Segregation of human β-actin gene sequences between human-mouse somatic cell hybrids. DNA was isolated from somatic cell hybrids (lanes A through Q), and the parental mouse (lane R) and human (lane S) cell lines. After digestion with *Eco*RI, the DNA was electrophoresed on a 0.8% agarose gel, blotted to nitrocellulose, and hybridized with the radiolabeled human-specific β-actin-specific DNA probe, HβAS. This probe recognizes 20 different-sized DNA fragments, and their numerical designation is shown on the right. The autoradiogram is the result of a 3-week exposure to Kodak XAR-5 film at −80°C, except for lane S, which was exposed for only 4 days.

those of the rodent, it is evident that all 20 genomic human β-actin sequences are more closely related to the functional human gene than they are to the mouse genes. This, in turn, strongly suggests that these copies of human β-actin sequences have been generated since the divergence of humans and rodents.

Over 30 different human-mouse somatic cell hybrid cell lines, each carrying a different subset of human chromosomes, were then assayed for the presence of the human β-actin-related sequences. DNA prepared from these cell hybrids plus the parental human and mouse cells was digested with *Eco*RI, size fractionated, and hybridized with the HβAS DNA fragment. Hybridization to a panel of human-mouse cell hybrid DNAs reveals complex banding patterns which differ between the different cell hybrids (Fig. 2, lanes A to Q). Pairwise comparisons of all these β-actin bands demonstrated that very few of them show any possible linkage (data not shown).

Each of the hybridizing human DNA fragments was examined for its cosegregation in cell hybrids with a specific human chromosome. We were able to determine for six of the human β-actin *Eco*RI fragments the percent discordancy for each chromosome among the various human chromosomes in the hybrid cell panels, and from these data six definitive chromosome assignments can be made (Table 2). The data confirm the assignment of band 5 (the functional gene, *ACTB*) to chromosome 7 and localize it to the 7pter→q22 region. The band 17 β-actin sequence also cosegregates with human chromosome 7. However, the band 17 sequence was not closely linked to *ACTB*, since its locus can be assigned to the 7q22→qter region.

At least one β-actin sequence, band 1, cosegregates with the X chromosome. Using a panel of X-autosomal translocations segregating in cell hybrids, we further delimited the location of band 1 to the Xq13→q22 region (data not shown). Bands 7 and 13 represent sequences that both segregate only with chromosome 5. Band 8 cosegregated only with chromosome 18.

Thus we have assigned 6 of the 20 human β-actin *Eco*RI fragments to specific chromosomes and, in several cases, to specific subchromosomal regions. We have not yet been able to assign the remaining β-actin sequences (bands) to specific chromosomes because of the complexities of their patterns and the possibility that some of the bands might represent comigrating fragments. However, there is little cosegregation of any of the β-actin gene sequences, and only one (band 17) cosegregates with the functional gene (band 5). This demonstrates that the majority of these β-actin-related sequences are not closely linked but, rather, are dispersed throughout the human genome. Since a number of these cell lines contain breaks in some of the human chromosomes, lack of cosegregation does not preclude location on the same chromosome, albeit at widely separated loci such as found with band 17 and the functional gene. Thus a set of 20 DNA sequences, apparently derived from the human β-actin gene some time after the divergence of mice and humans, is dispersed over the chromosomal landscape.

Sequence and organization of the human β-actin gene. To analyze the β-actin gene and examine its relationship to other actin genes, we have determined its nucleotide sequence. Figure 3 includes the sequencing strategy we used. Comparison of the genomic sequence with the cDNA sequence previously reported (42) enabled us to construct the structural organization of the human β-actin gene. The similarity of intron lengths between human and rat β-actin genes (37) has facilitated the mapping of these regions. The intron lengths for human, rat, and chicken β-actin genes are respectively 832,927, and 903 base pairs (IVS I); 134,87, and 320 base pairs (IVS II); 441,464, and 524 base pairs (IVS III); 95,88, and 306 base pairs (IVS IV); and 112, 124, and 355 base pairs (IVS V) (22, 37).

TABLE 2. Chromosome assignment of β-actin sequences[a]

β-Actin probe band no.	% Discordancy for human chromosome[b]:																						Region[c]
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	
1	27	30	30	24	43	32	24	27	43	35	30	35	32	30	35	49	24	41	32	43	24	38	0	Xq13 → q22
5	13	21	15	18	26	26	0	26	44	26	18	23	26	23	28	36	31	31	21	33	38	33	23	7pter → q22
7	26	26	30	30	0	26	22	13	30	30	26	30	17	30	22	35	35	35	26	35	35	26	39	
8	27	35	15	27	31	31	31	42	46	19	31	38	31	35	35	46	54	0	31	38	46	31	35	
13	27	20	27	27	0	20	25	25	36	25	25	30	27	34	32	39	41	20	23	34	30	34	32	
17	21	23	23	26	28	26	0	31	38	36	28	28	31	33	28	44	31	38	21	46	49	49	28	7q22 → qter

[a] Chromosome assignment panels consisted of at least 30 cell hybrids for each β-actin sequence. The panels did not always contain the same cell hybrids. For each band, the percent discordancy is listed for each chromosome. If a chromosome is present and a band is not (or the reverse), then the percent discordancy is indicated for the total number of hybrids in that panel. If a band and a chromosome cosegregate together in cell hybrids, then there is no discordancy ("0" discordancy) demonstrating that the β-actin sequence is encoded on the specific chromosome.

[b] Human chromosomes were identified by both karyotyping and previously mapped enzyme markers.

[c] With well-characterized translocation chromosomes, certain β-actin sequences segregated with specific chromosomal regions. Chromosome nomenclature followed that of the Paris Conference (34a).

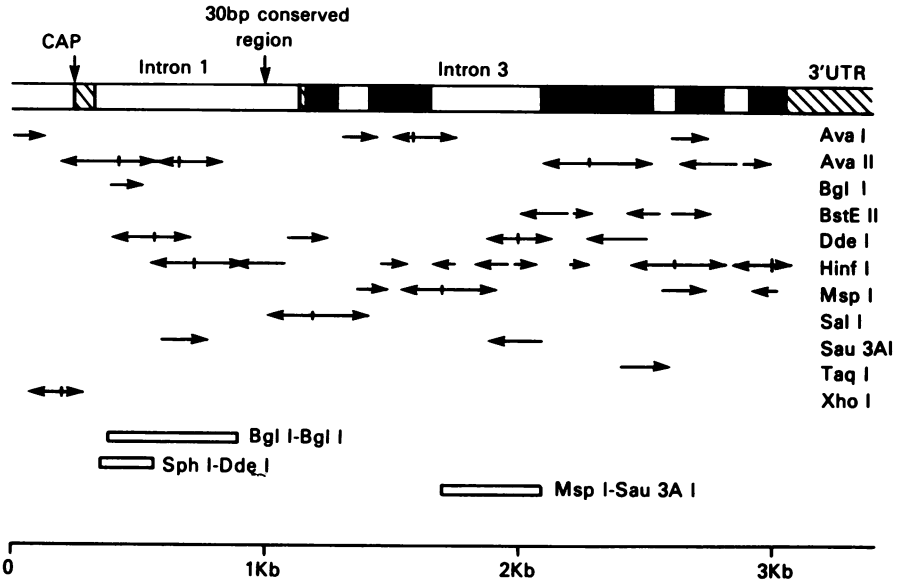

FIG. 3. Strategy employed for the determination of the nucleotide sequence of human β-actin gene in pM1β1-2. A restriction map of the 6.5-kb insert in pM1β1-2 has been published by Leavitt et al. (26). The clone has about 2 kb of DNA 5′ of the mRNA cap site and about 1.5 kb of 3′ UTR and πAN 7 sequences downstream of the translation termination codon. Protein-coding regions are depicted by solid boxes, UTRs are depicted by hatched boxes, and noncoding regions are depicted by empty boxes. The map shown only indicates the sites used for sequencing and does not represent a complete restriction map of every enzyme indicated. The map positions of IVS I and IVS III subclones from which intron probes are derived are indicated. These three IVS subclones are named pHβA-IVS I (*Sph-Dde*), pHβA-IVS I (*Bgl*I), and pHβA-IVS III, respectively.

The DNA sequence of the 2,826 base pairs of the human β-actin gene plus 240 base pairs of its 5′ flanking region is presented in Fig. 4. This transcription unit consists of 84 base pairs of 5′ UTR, 1,128 base pairs of protein-coding sequences interrupted by five introns (1,614 base pairs total), and about 591 base pairs of 3′ untranslated region. We have not included a complete DNA sequence of the region downstream of the termination codon since this gene was isolated as a πAN7 recombinant. The sequences in this region have been reported previously (42) and are inferred from the 3′ UTR of the β-actin cDNA clone.

The functional β-actin gene is single copy in the human genome. When a DNA fragment derived from the 3′ UTR of β-actin was used as a hybridization probe in genomic Southern blot experiments, we detect about 20 human genomic DNA fragments (41) (Fig. 2). Several of these β-actin sequences are known to be processed pseudogenes, as has been found by molecular cloning and DNA sequence analysis (33, 34; Engel, Ph.D. thesis). Accordingly we used DNA segments derived from IVS I and IVS III of the functional gene as hybridization probes to determine how many of the 20 β-actin coding sequences also contained intron sequences.

The probes (Fig. 3) were hybridized to genomic DNA cleaved with a variety of restriction endonucleases. The results with the IVS III probe are presented in Fig. 5. The number of hybridizing DNA fragments in each digest is consistent with the presence of but a single β-actin gene in the genome. *Pst*I and *Sac*I each generate two hybridizing fragments, as expected since they each cut once in the intron probe used in the experiment. The IVS I hybridization probe, derived from an *Sph*I-*Dde*I fragment near the 5′ end of IVS I (Fig. 3), also hybridizes to single genomic DNA fragments (data not shown).

The size of the *Eco*RI genomic fragment that hybridizes in Fig. 5 is about 14 kb. Our previous results, obtained with a hybridization probe containing 5′ region sequences (26), revealed two *Eco*RI hybridizing fragments of 14 and 6.4 kb. The 6.4-kb *Eco*RI fragment (band 11, Fig. 2) has been cloned and is known to be nonfunctional, based on preliminary DNA sequence analysis (Engel, Ph.D. thesis). This earlier result had left open the possibility that the 6.4-kb band might be an intron-bearing pseudogene. The results presented here exclude that possibility and demonstrate that the β-actin-coding sequence on the 6.4-kb *Eco*RI fragment lacks introns.

In sum, the results from genomic hybridization experiments with two different hybridization probes derived from two different regions of the β-actin gene demonstrate that there is a single chromosomal locus for the human β-actin gene and that the β-actin-related sequence of the 19 other *Eco*RI fragments are all recently generated, intronless, and dispersed on many chromosomes. From these results and those obtained from cloned copies of some of these *Eco*RI fragments, we conclude that these other *Eco*RI fragments are probably processed pseudogenes.

Mapping the human β-actin gene transcription unit. We localized the mRNA cap (or initiation) site of the β-actin gene first by S1 nuclease mapping (54) and then by DNA sequence comparisons. The DNA fragment we used for hybridization to human cellular RNA is an *Xho*I-*Bst*NI fragment, 127 bases long, extending from position −51 to base 76 in Fig. 4. The result of this experiment (Fig. 6) locates the region of the mRNA 5′ termini to within four nucleotides. Since S1-resistant DNA fragments migrate lower by a base on thin sequencing gels than do DNA fragments cleaved during chemical sequencing (12), the major S1-resistant DNA fragment terminates with the base A at position 1. The other protected fragments could represent inexact digestion products or protection by minor transcripts initiating at other positions.

Two consensus sequences associated with eucaryotic RNA polymerase II promoters, TATA and CAAT, are found

```
CCCAGCACCC CAAGGCGGCC AACGCCAAAA CTCTCCCTCC TCCTCTTCCT CAATNCTCGC TCTCGCTCTT TTTTTTTTTC GCAAAAGGAG GGGAGAGGGG GTAAAAAAAT GCTGCACTGT  -121

CGGCGAAGCC GGTGAGTGAG CGGCGCGGGG CCAATCGCGT GCGCCGTTCC GAAAGTTGCC TTTTATGGCT CGAGCGGCCG CGGCGGCGCC CTATAAAACC CAGCGGCGCG ACGCGCCACC   -1

1 ACCGCCGA GACCGCGTCC GCCCCGCGAG CACAGAGCCT CGCCTTTGCC GATCCGCCGC CCGTCCACAC CCGCCGCCAG GTAAGCCCG GCCAGCCGAC CGGGGCATGC GGCCGCGGCC  117

CCTTCGCCCG TGCAGAGCCG CCGTCTGGGC CGCAGCGGGG GGCGCATGGG GGGGGAACCG GACCGCCGTG GGGGGCGCGG GAGAAGCCCC TGGGCCTCCG GAGATGGGGG ACACCCCACG  237

CCAGTTCGGA GGCGCGAGGC CGCGCTCGGG AGGCGCGCTC CGGGGGTGCC GCTCTCGGGG CGGGGGCAAC CGGCGGGGTC TTTGTCTGAG CCGGGCTCTT GCCAATGGGG ATCGCAGGGT  357

GGGCGCGGCG TAGCCCCCGC CAGGCCCGGT GGGGGCTGGG GCGCCATGCG CGTGCGCGCT GGTCCTTTGG GCGCTAACTG CGTGCGCGCQ GGGAATTGGC GCTAATTGCG GCTGCGGCCQ  477

GGGACTCAAG GCGCTAATTG CGGCTGCGTT CTGGGGCCCG GGGTGCCGCG GCCQGGGCQG GGGCGAAGGC GGGCTCGGTC GGAAGGGGTG GGGTCGCCGC GGCTCCCGGG CGCTTGCGCA  597

CTTCCTGCCC GAGCCGCQGG CCGCCCGAGG GTGTGGCCGC TGCGTGCGCG CGCGCGACCC GGCGCTGTTT GAAQCGGGCG GAGGCGGGGC TGGCGCCCGG TTGGGAGGGG GTTGGGGCCT  717

GGCTTCCTGC CGCGCGCCGC GGGGACGCCT CCGACCAGTG TTTGCCTTTT ATGGTAATAA CGCGCCGGCC CGGCTTCCTT TATCCCCAAT CGTGCGCGCG CCGGCGCCCC CTAGCGGCCT  837

AAGGACTCGG CGCGCCGGAA GTGGCCAGGG CGGGGGCGAC TTCGGCTCAC AGCGCGCCCG GCTATTCTCG CAG CTCACC ATG GAT GAT GAT ATC GCC GCG CTC GTC GTC GAC  949
                                                                                    MET Asp Asp Asp Ile Ala Ala Leu Val Val Asp
                                                                                                                            10

AAC GGC TCC GGC ATG TGC AAG GCC GGC TTC GCG GGC GAC GAT GCC CCC CGG GCC GTC TTC CCC TCC ATC GTG GGG CGC CCC AGG CAC CAG GTAGGGGAGCT  1050
Asn Gly Ser Gly Met Cys Lys Ala Gly Phe Ala Gly Asp Asp Ala Pro Arg Ala Val Phe Pro Ser Ile Val Gly Arg Pro Arg His Gln
                                  20                                  30                                  40

GGCTGGGTGG GGCAGCCCCG GGAGCGGGCG GGAGGCAAGG GCGCTTTCTC TGCACAGGAG CCTCCCGGTT TCCGGGGTGG GGGCTGCGCC GCTGCTCAGG GCTTCTTGTC CTTTCCTTCC  1170

CAG GGC GTG ATG GTG GGC ATG GGT CAG AAG GAT TCC TAT GTG GGC GAC GAG GCC CAG AGC AAG AGA GGC ATC CTC ACC CTG AAG TAC CCC ATC GAG CAC  1269
    Gly Val Met Val Gly Met Gly Gln Lys Asp Ser Tyr Val Gly Asp Glu Ala Gln Ser Lys Arg Gly Ile Leu Thr Leu Lys Tyr Pro Ile Glu His
                                  50                                  60                                  70

GGC ATC GTC ACC AAC TGG GAC GAC ATG GAG AAA ATC TGG CAC CAC ACC TTC TAC AAT GAG CTG CGT GTG GCT CCC GAG GAG CAC CCC GTG CTG CTG ACC  1368
Gly Ile Val Thr Asn Trp Asp Asp Met Glu Lys Ile Trp His His Thr Phe Tyr Asn Glu Leu Arg Val Ala Pro Glu Glu His Pro Val Leu Leu Thr
                      80                                  90                                  100

GAG GCC CCC CTG AAC CCC AAG GCC AAC CGC GAG AAG ATG ACC CAG  GTGAGTGGCC  CGCTACCTCT  TCTGGTGGCC  GCCTCCCTCC  TTCCTGGCCT  CCCGGAGCTG  1473
Glu Ala Pro Leu Asn Pro Lys Ala Asn Arg Glu Lys Met Thr Gln
            110                                 120

CGCCCTTTCT CACTGGTTCT CTCTTCTGCC GTTTTCCGTA GGACTCTCTT CTCTGACCTG AGTCTCCTTT GGAACTCTGC AGGTTCTATT TGCTTTTTCC CAGATGAGCT CTTTTTCTGG  1593

TGTTTGTCTC TCTGACTAGG TGTCTAAGAC AGTGTTGTGG GTGTAGGTAC TAACACTGGC TCGTGTGACA AGGCCATGAG GCTGGTGTAA AGCGGCCTTG GAGTGTGTAT TAAGTAGGCG  1713

CACAGTAGGT CTGAACAGAC TCCCCATCCC AAGACCCCAG CACACTTAGC CGTGTTCTTT GCACTTTCTG CATGTCCCCC GTCTGGCCTG GCTGTCCCCA GTGGCTTCCC CAGTGTGACA  1833

TGGTGCATCT CTGCCTTACAG  ATC ATG TTT GAG ACC TTC AAC ACC CCA GCC ATG TAC GTT GCT ATC CAG GCT GTG CTA TCC CTG TAC GCC TCT GGC CGT ACC  1935
                       Ile Met Phe Glu Thr Phe Asn Thr Pro Ala Met Tyr Val Ala Ile Gln Ala Val Leu Ser Leu Tyr Ala Ser Gly Arg Thr
                                                  130                                 140

ACT GGC ATC GTG ATG GAC TCC GGT GAC GGG GTC ACC CAC ACT GTG CCC ATC TAC GAG GGG TAT GCC CTC CCC CAT GCC ATC CTG CGT CTG GAC CTG GCT  2034
Thr Gly Ile Val Met Asp Ser Gly Asp Gly Val Thr His Thr Val Pro Ile Tyr Glu Gly Tyr Ala Leu Pro His Ala Ile Leu Arg Leu Asp Leu Ala
    150                                 160                                 170                                 180

GGC CCG GAC CTG ACT GAC TAC CTC ATG AAG ATC CTC ACC GAG CGC GGC TAC AGC TTC ACC ACC ACG GCC GAG CGG GAA ATC GTG CGT GAC ATT AAG GAG  2133
Gly Arg Asp Leu Thr Asp Tyr Leu Met Lys Ile Leu Thr Glu Arg Gly Tyr Ser Phe Thr Thr Thr Ala Glu Arg Glu Ile Val Arg Asp Ile Lys Glu
                      190                                 200                                 210

AAG CTG TGC TAC GTC GCC CTG GAC TTC GAG CAA GAG ATG GCC ACG GCT GCT TCC AGC TCC TCC CTG GAG AAG AGC TAC GAG CTG CCT GAC GGC CAG GTC  2232
Lys Leu Cys Tyr Val Ala Leu Asp Phe Glu Gln Glu Met Ala Thr Ala Ala Ser Ser Ser Ser Leu Glu Lys Ser Tyr Glu Leu Pro Asp Gly Gln Val
                  220                                 230                                 240

ATC ACC ATT GGC AAT GAG CGG TTC CGC TGC CCT GAG GCA CTC TTC CAG CCT TCC TTC CTG G  GTGAGTGGAG ACTGTCTCCC GGCTCTGCCT GACATGAGGG TTAC  2337
Ile Thr Ile Gly Asn Glu Arg Phe Arg Cys Pro Glu Ala Leu Phe Gln Pro Ser Phe Leu
            250                                 260

CCCTCGGGGC TGTGCTGTGG AAGCTAAGTC CTGCCCTCAT TTCCCTCTCAG  GC ATG GAG TCC TGT GGC ATC CAC GAA ACT ACC TTC AAC TCC ATC ATG AAG TGT GAC  2444
                                                  Gly Met Glu Ser Cys Gly Ile His Glu Thr Thr Phe Asn Ser Ile Met Lys Cys Asp
                                                                      270                                 280

GTG GAC ATC CGC AAA GAC CTG TAC GCC AAC ACA GTG CTG TCT GGC GGC ACC ACC ATG TAC CCT GGC ATT GCC GAC AGG ATG CAG AAG GAG ATC ACT GCC  2543
Val Asp Ile Arg Lys Asp Leu Tyr Ala Asn Thr Val Leu Ser Gly Gly Thr Thr Met Tyr Pro Gly Ile Ala Asp Arg Met Gln Lys Glu Ile Thr Ala
            290                                 300                                 310

CTG GCA CCC AGC ACA ATG AAG ATC AAG GTGGGTGTCT TTCCTGCCTG AGCTGACCTG GGCAGGTCAG CTGTGGGGTC CTGTGGTGTG TGGGGAGCTG TCACATCCAG GGTCCTC  2657
Leu Ala Pro Ser Thr Met Lys Ile Lys
    320

ACTGCCTGTC CCCTTCCCTC CTCAG ATC ATT GCT CCT CCT GAG CGC AAG TAC TCC GTG TGG ATC GGC GGC TCC ATC CTG GCC TCG CTG TCC ACC TTC CAG CAG  2760
                           Ile Ile Ala Pro Pro Glu Arg Lys Tyr Ser Val Trp Ile Gly Gly Ser Ile Leu Ala Ser Leu Ser Thr Phe Gln Gln
                                      330                                 340                                 350

ATG TGG ATC AGC AAG CAG GAG TAT GAC GAG TCC GGC CCC TCC ATC GTC CAC CGC AAA TGC TTC TAG                                          2826
Met Trp Ile Ser Lys Gln Glu Tyr Asp Glu Ser Gly Pro Ser Ile Val His Arg Lys Cys Phe Ter
                      360                                 370
```

FIG. 4. Nucleotide sequence of human β-actin gene. The nucleotides are numbered relative to the A of the mRNA cap site. Nucleotides 5′ of the A are designated by negative numbers. The 84-base-pair 5′ UTR (bases 1 to 78, bases 911 to 916) is underlined. The 30-base-pair highly conserved region (bases 752 to 781) of IVS I is also underlined. The amino acid sequence is numbered by the method of Lu and Elzinga (29).

at positions −29 and −89, respectively. Further upstream of the CAAT sequence, there are long stretches of both polypyrimidines and polypurines. Such sequences may contain S1 nuclease-sensitive sites as found in similar regions in other genes (8).

The region surrounding the cap site is conserved in evolution (Fig. 7; see below), and our assignment of the 5′ terminus is consistent with that inferred from the alignment of human and rat sequences. This assignment is further strengthened by DNA sequence comparisons with the 5′ untranslated regions of two human β-actin pseudogenes (34). These intronless, reverse transcript-type pseudogenes are flanked by terminal repeats. The junction of these terminal repeats and the 5′ untranslated regions is within 1 base pair of the S1 nuclease-protected terminus (data not shown). Therefore it appears that these processed pseudogenes are full length.

We conclude that the human β-actin mRNA has an 84-base 5′ UTR. Our longest β-actin cDNA clone contains only 41 base pairs of 5′ UTR sequences (42); these are

identical to those of the corresponding 5′ UTR of the β-actin gene from nucleotides 44 to 78 and 911 to 916 in Fig. 4. Thus the location of the IVS I can be unambiguously assigned to lie between bases 78 and 79 of the untranslated 5′ leader sequence. This IVS I is the longest intron, 832 base pairs in length.

IVS II, III, IV, and V are located in the protein-coding region of the gene. The second intron is 134 base pairs in length and occurs between codons 41 and 42. The third intron is 441 base pairs in length and occurs between codons 121 and 122. The fourth intron is 95 base pairs in length and interrupts codon 267. The fifth intron is 112 base pairs in length and occurs between codons 327 and 328. These intron positions are identical with those of rat (37) and chicken (22) β-actin genes. In addition, the relative lengths of the various introns appear to be conserved between humans and rats. IVS I and IVS III, for example, are the longest in all three species. Thus the structural organization of the β-actin gene is well preserved throughout vertebrate evolution.

The human β-actin gene thus has six exons. The second exon contains 6 base pairs of 5′ UTR and the first 41 codons. The other exons contain 80, 146, 61, and 48 codons respectively. All 376 codons are identical to those from our cDNA clone (42). Furthermore, this coding sequence is identical to those from another human β-actin cDNA clone derived from

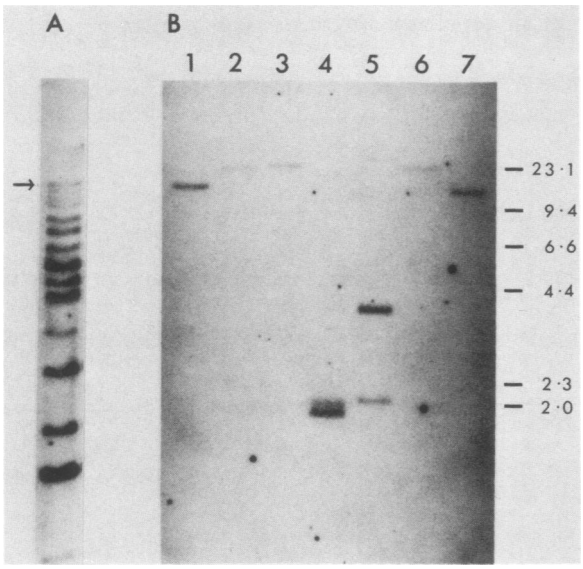

FIG. 5. Detection of the functional β-actin gene in the human genome. (A) Human DNA (5 μg) was digested with *Eco*RI, size fractionated on a 0.7% agarose gel, and blot transferred to a nitrocellulose filter. The filter was hybridized to the β-actin 3′ UTR probe (pHFβA-3′UT) and washed as described in Materials and Methods. The final wash was at 65°C in 0.5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate. The 14-kb genomic *Eco*RI fragment containing the functional β-actin gene is indicated by an arrow. (B) Human DNA (5 μg) was digested with the following: 1, *Eco*RI; 2, *Bgl*II; 3, *Hin*dIII; 4, *Pst*I; 5, *Sac*I; 6, *Xba*I; or 7, *Bam*HI. The DNA was then size fractionated on a 0.7% agarose gel and blot transferred to a nitrocellulose filter. The filter was hybridized to the IVS III subclone probe (*Msp*I-*Sau*3AI; Fig. 3) and washed as described in Materials and Methods. The final wash was at 65°C in 0.5 × SSC–0.1% sodium dodecyl sulfate. The IVS III probe has a higher G+C content (53%) than the 3′ UTR probe (42%) and as such would detect more divergent genomic sequences than the latter probe under identical hybridization conditions. Size markers (in kilobases) are indicted to the right of the autoradiogram.

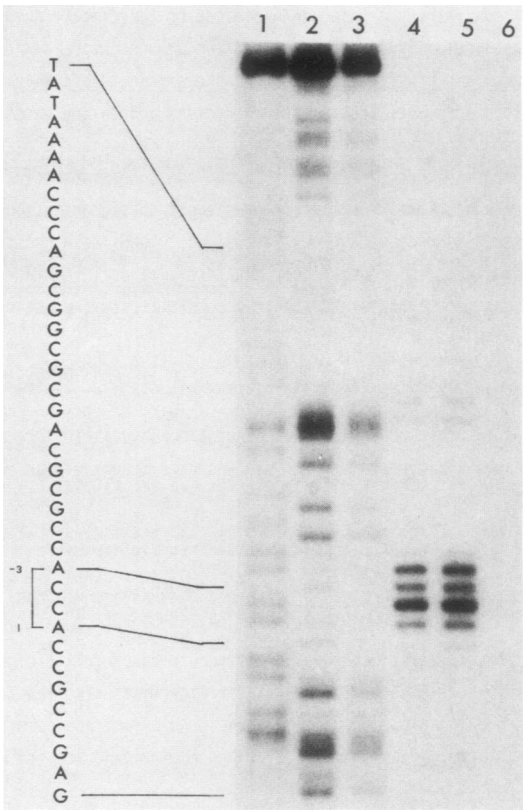

FIG. 6. S1 mapping of the human β-actin mRNA 5′ termini. The *Xho*I-*Bst*NI* fragment, spanning the mRNA cap site (see the text), was used for S1 analysis and chemical sequencing. Lanes 1 through 3 shows sequencing ladders resulting from base-specific modification reactions (1, A+G cleavage fragments; 2, C+T cleavage fragments; 3, C cleavage fragments). The nucleotide sequence of this region, based on chemical sequencing with this and other DNA fragments (Fig. 4), is presented to the left of the autoradiogram. Lanes 4 through 6 show S1-resistant fragments derived from the *Xho*I-*Bst*NI* probe after hybridization with (lanes 4 and 5) or without (lane 6) human total cellular RNA prepared as described previously (41).

mRNA isolated from a different tissue (19) and provides further evidence that this cloned gene is expressed in most, if not all, nonmuscle cells.

The amino acid sequence of β-actins from humans, rats, and chickens are identical (22, 37, 42). The codons are also highly conserved. The overall similarity of human and rat DNA sequences in the coding exons is 92%, and the overall similarity of human and chicken DNA sequences in the coding exons is 88%.

There are 22 to 24 amino acid differences between mammalian β-actin and various muscle actins (52). We found that the coding DNA sequences of the human β-actin gene were 86% similar to those of human α-skeletal actin (18) and 78% similar to human α-cardiac actin (17). In addition, the similarity with the first seven exons of the aortic smooth muscle actin sequence (51) is 80%. Since the human β-actin sequences are more closely related to the rat and chicken β-actin sequences than they are to any of the human muscle actin sequences, we conclude that the actin gene duplication that engendered the muscle actins and the β-actin genes apparently occurred before the divergence of mammals and birds.

Conservation of noncoding sequences: 5′ flanking region.

```
                                                        -132
CTCTCGCTCT TTTTTTTTTT CGCAAAAGGA GGGGAGAGGG GGTAAAAAAA  Human
* * *** *             *       ** ******
TTTTTTTTTT TTTTTTTTTT TGCAAAAG-- -----AAGGG GGTAAAAAAA  Rat
                                                   103

                                                        -84
TGCTGCACTG TC-GGCGAAG CCGGTGAGTG AGCGGCGCGG GGCCAATC-G  Human
           *     *                  ***    *       *
TGCTGCACTG TCGGGCGAGG CCGGTGAGTG AGCGAGCCGG AGCCAATCAG  Rat
                                                   153

                                                        -34
CGTGCGCCGT TCCGAAAGTT GCCTTTTATG GCTCGAGCGG CCGCGGCGGC  Human
  **          *                          *     * *
CGCCCGCCGT TCCGAAA-TT GCCTTTTATG GCTCGAGTGG CCGCTGTGGC  Rat
                                                   202

                                          1
GCCCTATAAA ACCCAGCGGC GCGACGCCGC ACCACCGCCG AGACCGCGTC  Human
 *           *       *  *  *    * *    *
GTCCTATAAA ACCCGGCGGC GCAACGC-GC GCCACTGTCG AGTCCGCGTC  Rat
                                                   251

                                                    65
CGCCCCGCGA GCACAGAGCC TCGCCTTTGC CGATCCGCCG CCC--GTCCA  Human
 **          *  **  *     ****     * *   *    *  **
C-ACCCGCGA GTAC--AACC TCC---TTGC AGCTCCTCCG TCCCGGTCCA  Rat
                                                   295

           78                                       916
CACCCGCCGC CAG  [     832 base pairs IVS     ]  CTCACC  Human
      *                                          *  *
CACCCGCCAC CAG  [     927 base pairs IVS     ]  TTCGCC  Rat
           308                                     1241
```

FIG. 7. Comparison of the 5′ flanking region and the 5′ UTR of the human β-actin gene with the rat β-actin gene. The sequences have been aligned for maximal homology. Mismatches and gaps are indicated by asterisks. The human sequence is numbered as in Fig. 4. The rat sequence is numbered as in Nudel et al. (37). The rat 5′ untranslated region (80 base pairs total) is between base 235 and 1241, interrupted by the 927-base-pair 5′ IVS (bases 309 to 1235).

Kost et al. (22) reported the presence of DNA sequence similarity in the 5′ flanking region of rat and chicken β-actin genes. We examined the extent of sequence similarity between the β-actin genes of these two species and the human sequences. The similarity between human and rat sequences is extremely high (Fig. 7); there is also significant homology between human and chicken sequences (data not shown). We calculated the degree of sequence similarity of this region of the β-actin genes by applying the method of Miyata et al. (32), which yields a similarity value $K_N(1)$ that is 0 for identical sequences and approaches a terminal value of 0.75 for random sequences. [$K_N(1)$ is the K_N value calculated based on method one; i.e., when a gap is found in any one of the aligned sequences, the corresponding site is excluded from the calculation (32).] In this region, from the mRNA cap site to about 170 base pairs upstream, the $K_N(1)$ value is 0.098 for the human-rat comparison and 0.302 for the human-chicken comparison. The human-rat $K_N(1)$ value is among the lowest yet found in comparisons of any 5′ flanking regions in mammals. This extraordinary conservation of DNA sequences in this 5′ region may be a reflection of essential functions associated with the constitutively and ubiquitously active β-actin gene promoter.

Conservation of noncoding sequences: 5′ and 3′ UTRs. We have reported the evolutionary conservation of the 3′ UTR of the human β-actin cDNA (42). Since the 5′ UTR of our β-actin cDNA clone was not full length, we had not been able to examine the sequence conservation of that region. We can now extend the sequence comparison between the human and rat 5′ UTR sequences (Fig. 7). The $K_N(1)$ value for this region, 0.179, is very low. Therefore this 5′ UTR is nearly as well conserved as the 3′ UTR of human and rat β-actin genes [$K_N(1)$ value of 0.135 (42)] and the 5′ UTR of human and rat skeletal actin genes [$K_N(1)$ value of 0.127 (18)].

Conservation of noncoding sequences: introns. Our comparisons of β-actin gene sequences have shown that they are exceptions to the general observation that UTR sequences are not well conserved. This finding, as well as the conservation of both the β-actin gene intron locations and their relative lengths, raised the possibility that the sequences of the introns might also be conserved.

We aligned the nucleotide sequences of the human and rat IVS segments by using the Wilbur and Lipman algorithm (56). There are strong sequence similarities in IVS I and IVS III (Fig. 8, 9, and 10). The IVS I alignments have a $K_N(1)$ value of 0.258 with a total of 392 base pairs matching and a minimal number of gaps inserted (Fig. 8). Although this $K_N(1)$ value is not as low as that for the 5′ flanking region or the 3′ UTR, it is highly significant when compared to values derived for other introns whose DNA sequences are known. For example, the large introns of the human and mouse β-globin genes which are 850 and 653 base pairs long, respectively, have a $K_N(1)$ value of 0.459 with 315 base pairs matching. Moreover, comparison of intron sequences of human insulin and metallothionein-I_A genes with their mouse homologs also did not reveal significant similarities [$K_N(1)$ values of at least 0.4]. In contrast, the introns of human and mouse protooncogene c-*fos* are more conserved than the 5′ flanking regions (53). However, since the conserved introns of β-actin and c-*fos* genes are dissimilar, it appears unlikely that they are directly involved in the regulation of β-actin and c-*fos* gene transcription by growth factors (5, 13).

Conservation of noncoding sequences: enhancer-like and potential Z-DNA sequences in IVS I. Within the conserved segment of IVS I is a 30-base-pair subsequence that is highly conserved from humans to chickens. This sequence, located 673 base pairs from the 5′ end and 129 base pairs from the 3′ end of the intron (underlined in Fig. 4), is shown in Fig. 9. There are only two mismatches, both of them pyrimidine changes, between the human and rat sequences. There are five additional mismatches between the human and chicken sequences, but the $K_N(1)$ value of this segment is still low (0.233). Moreover, in all three species the location of this sequence is conserved relative to the ends of the intron (legend to Fig. 9). Of interest is that this 30-base-pair segment contains sequences similar to those associated with viral enhancer elements. The simian virus 40 72-base repeat contains the sequence TGTGGAAA (24), and the more active murine sarcoma virus promoter-distal 73-base-pair repeat contains the sequence TGTGGTAA (23). The human, rat, and chicken β-actin introns all contain the sequence TATGGTAA within the 30-base-pair conserved region.

The IVS I sequences of human, rat, and chicken β-actin also contain several potential Z-DNA sequences (36). There are four short elements in both the human and rat genes (underlined in Fig. 8). The locations of three of these four elements are conserved between humans and rats. Short potential Z-DNA elements appear to be part of the simian virus 40 enhancers and of retroviral long terminal repeats (36). The potential Z-DNA elements in these β-actin introns might equally well function in concert with the 30-base-pair conserved region to mediate enhanced transcription. Demonstration that this 30-base-pair conserved region or the potential Z-DNA sequences have enhancer-like activities must await direct experimentation.

Neither this 30-base-pair sequence nor sequences related

FIG. 8. Comparison of IVS I sequences of the human β-actin gene with the rat β-actin gene. The sequences have been aligned for maximal homology. Mismatches and gaps are indicated by asterisks. There are 235 base pairs (in humans) and 362 base pairs (in rats) of IVS I sequences that do not align well with each other between the 5′ splice sites and the conserved regions shown in this figure. The human sequence is numbered as in Fig. 4. The rat sequence is numbered as in Nudel et al. (37). The potential Z-DNA-forming elements are underlined. The 30-base-pair highly conserved sequence (see Fig. 9) is enclosed in brackets. The $K_N(1)$ value for the 5′ subregion (293 base pairs; bases 313 to 605) is 0.284. The $K_N(1)$ value for the 3′ subregion (216 base pairs; bases 679 to 894) is 0.220. These two subregions are separated by a middle subregion with two long gaps of more than 10 nonmatching bases each. The latter region consists of 72 base pairs in humans (32 base pairs nonmatching with rats) and 49 base pairs in rats (9 base pairs nonmatching with humans). When comparing rat and chicken IVS I sequences (data not shown), we obtained $K_N(1)$ values of 0.466 and 0.347 for the 5′ and 3′ subregions, respectively.

```
Human   CCAGTGTTTGCCTTTIATGGTAATAACGCG    781
Rat     CCAGCGTTTGCCTTTIATGGTAATAATGCG   1095
Chick   GCAGCCATTGCCTTTIATGGTAATCGTGCG    -293
```

FIG. 9. A 30-base pair conserved region within β-actin gene IVS I sequence. Mismatches are indicated by asterisks. The human sequence is numbered as in Fig. 4. There are 673 and 129 base pairs of intervening sequences 5′ and 3′ to this segment in the human intron. The rat sequence is numbered as in Nudel et al. (37). There are 757 and 140 base pairs of intervening sequences 5′ and 3′ to this segment in the rat intron. The chicken sequence is numbered as in Kost et al. (22). There are 587 and 286 base pairs of intervening sequences 5′ and 3′ to this segment in the chicken intron. The eight-base sequences resembling viral enhancer elements are underlined.

to it are found in the other four introns. In fact, the longest sequences common to the five introns are only 4 nucleotides long. This observation and the fact that the sequences in the three smallest β-actin gene introns (IVS II, IV, and V) are fully diverged from their rat and chicken counterparts suggest that they contain no selected sequences.

Conservation of noncoding sequences: IVS III. IVS III, the second-longest intron of the β-actin gene, is also highly conserved in evolution, however. About 73% of its base pairs match those of IVS III of the rat (Fig. 10). The $K_N(1)$ value for this region, when compared with that for the rat, is 0.223. There is a 68-base-pair subsegment that is conserved between humans, rats and chickens (underlined in Fig. 10).

```
                                                                1474
GTGAGTGGCCCGCTACCTCTTCTGGTGGCCGCCTCCCTCCTTCCTGGCCTCCCGGAGCTGC    Human
  *   **  * ***********    *  ** *    *   **   ***** **** *
GTCAGTATCCTG-----------GGTGACCCTC-CCCTTCTTATTGGGTCAACTTCTCAGC    Rat
                                                                1740

                                                                1534
GC-CCTTTCTCACTGGTICTCTCTTCTGCCGTTTTCCGTAGGACTCTCTTCTCTGACCTGA    Human
* *  *     ******   *          * * *  *                 *   *
ACGCCCTTCTCTAATTGTCTTTCTTCTGCCATGTCCCATAGGACTCTCTTCTATGAGCTGA    Rat
                                                                1801

                                                                1595
GTCTCCTTTGGAACTCTGCAGGTTCTATTTGCTTTTTCCCAGATGAGCTCTTTTTCTGGTG    Human
      *       *    *   ****   *              *       ******
GTCTCCCTTGGAACTTTGCAGTTTCT----GCTCTTTCCCAGATGAGGTCTTTTT------    Rat
                                                          1852

                                                                1641
TTTGTCTC---------TCTGACTAGGTGTC-TAAG-----ACAGTGTTGTGGGTGTAGGT    Human
   *    *********              **   ******      *
TTTCTCTCGATCGCCTTTCTGACTAGGTGTTTTAAACCCCTACAGTGCTGTGGGTGTAGGT    Rat
                                                                1913

                                                                1701
ACTAACACTGGCTCGTGTGACAAGGCC-ATGAGGCTGGTGTAAAGCGGCCTTGGAGTGTGT    Human
      *                *  **            **  **
ACTAACAATGGCTCGTGTGACAAAGCTAATGAGGCTGGTGATAAATGGCCTTGGAGTGTGT    Rat
                                                                1974

                                                                1761
ATTAAGTAGGCGCACAGTAGGTCTGAACAGACTCCCCATCCCAAGAC-CCCAGCACACTTA    Human
   *     ** *           *  **  **   **   ** *  *
ATTCAGTAGATG-ACAGTAGGTCTAAATGGAGCCCCTGTCCTGATACTCCCAGCACACTTA    Rat
                                                                2034

                                        1795
-----GCCGTGTTCTTTGCACTTTCTGCATGTCCCCCGT----------------------    Human
*****  *        *  *  * *        * * **********************
ACTTAGCTGTGTTCTT-GCCCTCTTTGCATGTCTCACTCAAATCTATCCTTACAGTCTCAC    Rat
                                                                2094

                                                                1854
CTGGCCTGGCTGTCCCCAGTGGCTTCCCCAGTGTGAC-ATGGTGCATCTCTGCCT-TACAG    Human
   *    **   ** **       ****  **    *  **  *  **  **  *
CTGCCCTGAGTGTTTCTTGTGGCTTTAGGAGCTTGACAATACTGTATTCCTTTCTCTACAG    Rat
                                                                2155
```

FIG. 10. Comparison of IVS III sequences of the human β-actin gene with the rat β-actin gene. The sequences have been aligned for maximal homology. Mismatches and gaps are indicated by asterisks. The human sequence is numbered as in Fig. 4. The rat sequence is numbered as in Nudel et al. (37). The $K_N(1)$ value for the region between the two longest nonmatching gaps (from human base number 1437 to 1793) is 0.195. The underlined 68 base-pair region (from human base number 1491 to 1558) is also conserved in chickens. This sequence in chickens is TCCCTTTTTGCCATCTTTACAGGGTTTTCCTTTCCTGACCTGAGTCTCCTCTTTGCTGGACCTTGACAGGTT. The $K_N(1)$ values for this region are 0.147 for human-rat comparison and 0.358 for rat-chicken comparison. The conserved dyad symmetry element, CAAGGCC-N_{18}-GGCCTTG, is also underlined.

Overall, the evolutionary conservation of this region is even higher than that of IVS I. However, since this conserved intron is about 1.5 kb from the 5′ promoter region(s), it is less likely to have a major role in the regulation of actin gene expression. Nevertheless, such strong sequence conservation implies selection for some unknown function. There are several unusual features in IVS III that might be clues to such functions. First, this region has a G+C content of only 55% whereas the G+C content of IVS I is 76%. Second, IVS III is composed of 61% pyrimidines with four runs of 10 or more polypyrimidines. Third, there are several sequences in this intron with dyad symmetry. One of these, CAAGGCC-N_{18}-GGCCTTG in humans (underlined in Fig. 10) has 12 nucleotides out of 14 conserved between humans and rats. A similar sequence with dyad symmetry (GGGCCTG-N_{19}-CAGGCTC) is found in the chicken β-actin IVS III.

DISCUSSION

β-Actin is the product of a single functional gene. Since there are 20 or more DNA segments in the human genome that contain β-actin sequences, it was important to establish that the β-actin gene described here is the major, if not sole, functioning human β-actin gene. Several lines of evidence strongly support this conclusion. First, the mRNA encoded by this gene is identical to that of cDNAs cloned from a human fibroblast library (42) and from a human epidermal cell library (19). Thus this gene encodes the major β-actin mRNA in two human cell types.

Second, an allele of this gene can be expressed after transfection into mammalian cells. We have cloned and partially sequenced the two alleles, both being expressed, of the β-actin gene from the human HuT-14T cell line (26; C.-S. Lin, S.-Y. Ng, P. Gunning, L. Kedes, and J. Leavitt, Proc. Natl. Acad. Sci. USA, in press). One of the alleles contains a codon-altering point mutation at codon 244, and the other allele is identical to the β-actin gene reported here. These two genes were cloned by screening genomic libraries of HuT cells with a 5′ IVS I probe from our human β-actin gene. Since the IVS I is single copy, we conclude that the wild-type β-actin gene and the mutant gene(s) must be allelic copies. The mutant β-actin allele has been transfected into both human and rat-2 cells [J. Leavitt, P. Gunning, L. Kedes, and R. Jariwalla, Nature (London), in press], where it robustly expresses mutant β-actin. In addition, in blotting experiments with the RNA from the transfected rat-2 cells, we detect only a single discrete 2.1-kb human β-actin transcript. We conclude that the mutant β-actin allele and thus the nonmutant β-actin gene described in this manuscript are functional in human cells and are expressed under the regulation of a strong promoter.

The data presented in this paper strongly argue that, other than the functional actin gene, all of the other β-actin gene sequences detected by the human β-actin 3′ UTR probe are pseudogenes. The demonstration that all of the human β-actin gene sequences are more closely related to the functional human gene than they are to any of the β-actin genes of the mouse indicates that these human sequences were generated recently in evolution, but certainly after the divergence of mice and humans. These other human β-actin DNA fragments fail to hybridize with probes containing either of two β-actin intervening sequences. This implies that these other genes are intronless and, accordingly, are likely to be processed pseudogenes of the reverse transcript type. Furthermore, the nonlinkage of these gene sequences indicates that they are dispersed randomly throughout the genome and that tandem gene duplication has had little to do with their generation. In addition, direct sequence analysis of several of these gene sequences has confirmed that they are indeed intronless pseudogenes of the reverse transcript type (33, 34). We conclude that the human β-actin 3′ UTR probe, HβAS, defines a family of related gene sequences of which one is a functional gene and the remainder are intronless pseudogenes.

On the origin and dispersion of processed pseudogenes. We have shown by chromosome mapping that there is but a single copy of an expressed β-actin gene in the human genome and that the 19 (or more) β-actin-related sequences are not closely linked to this β-actin gene or to each other. From the data discussed earlier, it is highly likely that the human β-actin multigene family probably consists of a single expressed gene and at least 19 processed pseudogenes. The fact that a single, functional, β-actin gene, found on chromosome 7, is not closely linked with any of its related sequences is one of the first demonstrations in a multigene family of nonlinkage between an expressed gene and its related processed pseudogenes. This finding implies that the integration of processed pseudogenes in the human genome is probably random. Furthermore, the lack of detectable linkage of any of these β-actin sequences to each other indicates that tandem duplication has not been involved in the expansion of this family.

Based on our previous observation that the cytoskeletal actin genes, but not the sarcomeric actin genes, are associated with pseudogene families, we proposed the hypothesis that linked the expression of a gene in germ line cells to the production of large processed pseudogene families (41). In keeping with this hypothesis, other multigene families that consist predominantly of dispersed processed pseudogenes also appear to be expressed in germ cells. Examples include the multigene families encoding arginosuccinate synthetase (9), dihydrofolate reductase (1), glyceraldehyde 3-phosphate dehydrogenase (40), metallothionein (21), α-tubulin (28), and β-tubulin (27). It will be interesting to determine whether the level of mRNAs of a particular gene in germ cells correlates with the abundance of processed pseudogenes in the genome.

Use of multigene families for chromosome mapping. One unexpected outcome of this work was the relative ease and efficiency we observed in the mapping of a number of related DNA restriction fragments to specific chromosomes. We had initially expected that we would have to develop a gene-specific probe for each β-actin genomic DNA fragment (such as a flanking region probe) to distinguish restriction fragments in somatic cell hybrids. What we discovered was that selection of a species-specific probe allowed us to detect as many as 20 human-specific *Eco*RI fragments and to assign at least six of these fragments to specific chromosomes. For example, the mapping of *ACTB* to human chromosome 7 can be accomplished by the use of the HβAS probe alone. Thus small multigene families appear to provide a general approach for simultaneously mapping a number of restriction fragments from the same set of somatic cell hybrids. Furthermore, this approach is facilitated by providing multiple sets of data from the same genomic Southern blots. For example, we should be able to map members of a second multigene family (such as γ-actin) with these same blots by using a species-specific 3′ UTR probe. By comparing the two sets of data, we will also be able to determine whether various β- and γ-actin sequences cosegregate.

Knowledge of the chromosomal assignments of a set of dispersed restriction fragments, such as those generated from an analysis of the β- and γ-actin multigene families, will

have an important application for human chromosome mapping, or mapping of chromosomes for any species, when applied to somatic cell hybrids. We expect that at least half of the human chromosomes will contain either a β- or γ-actin gene sequence that can be readily visualized on a genomic Southern blot (Fig. 2). These loci could therefore serve as markers for specific human chromosomes in human-rodent somatic cell hybrids. Thus probes from β- and γ-actin, or a probe from any dispersed, moderately small multigene family, may provide a more convenient means of evaluating, at least initially, newly generated cell lines than conventional karyotyping or analysis of biochemical markers.

Evolution of intron segments. Eucaryotic actin gene isotypes share subsets of 15 known intron positions (46, 51), but no simple, parsimonious scheme can account for both the evolution of the various isotypes and the intron locations (16). Each gene has zero to seven introns. Therefore it has been proposed that different introns have been excised (or inserted) during evolution of various actin genes found in present-day eucaryotes (10, 51).

Mammalian sarcomeric (α-skeletal and α-cardiac) actin genes have two intron positions not found in the cytoskeletal β-actin genes (IVS III and IVS IV), and the cytoskeletal β-actin gene has a single unique intron position (IVS III) not found in α-skeletal and α-cardiac actin genes. We have shown here that the β-actin IVS III intron segment is evolutionarily well conserved. Similarly, the IVS III of rat and chicken α-skeletal actin genes is also conserved [the $K_N(1)$ value is 0.235 with 56 base pairs matching]. The preservation of these unique aspects of the sarcomeric versus the cytoskeletal actin gene organization throughout vertebrate evolution suggests that some intron segments have become functional domains of the respective transcription unit. Determination of the intron locations and structure of the other vertebrate cytoskeletal actin gene, γ-actin gene, may help us understand the evolution of intervening sequences.

Segments of genes can evolve at different rates. For example, we previously have reported evidence for separate units of selection within the 3′ UTRs of mammalian α-skeletal and α-cardiac mRNAs (14). The 3′ half of these 3′ UTRs shows stronger evolutionary conservation than the respective 5′ half. Similarly, IVS I, located in the 5′ UTR, has a nonconserved block, 235 base pairs in humans and 362 base pairs in rats, upstream of the conserved regions shown in Fig. 8. On the other hand, the entire 3′ UTR of the β-actin gene is uniformly conserved and appears to evolve as a single unit (42). Likewise, IVS III, the most highly conserved intron, appears to have been selected as a single unit, since the nucleotide differences between it and its rat homolog are randomly distributed.

Conserved intron segments. Our finding of highly conserved blocks of nucleotides in two of the five intervening sequences of β-actin genes raises the possibility that these segments have regulatory functions. Conserved internal regions have been reported previously, such as the internal transcriptional enhancer regions of immunoglobulin genes (2, 11, 39, 44). However, the locations of these enhancers were initially regarded as a peculiarity of the immunoglobulin gene loci. More recently, internal control regions have been detected (but yet undefined) for the adenovirus E1A gene (38), human globin genes (3), and chicken thymidine kinase gene (31). Any conclusion that the conserved β-actin intron sequences, especially those of IVS I, function as transcriptional enhancers must await direct experimentation. Nevertheless the evolutionary conservation of the immunoglobulin enhancer segments (6, 20) indicates that other transcriptional enhancers or *cis*-acting regulatory signals would be under selective pressure. It is interesting to note in this regard that the IVS I of both α- and β-globin genes are the most conserved introns of these genes. The IVS I of the human and mouse β-globin genes, for example, has 81 base pairs matching to give a $K_N(1)$ value of 0.302. Therefore these introns may well contain part of the proposed downstream regulatory elements (3).

ACKNOWLEDGMENTS

This work was supported in part by grants from Veterans Administration (to L.K.), the American Cancer Society (NP-433 to J.L. and CD-62 to T.S.), and the National Institutes of Health (Public Health Service grants HD17031 to L.K., CA34763 to J.L., and GM20454 and HD05196 to T.S.).

We are grateful to Ching-Shwun Lin, Adrian Minty, and William Bains for thoughtful review and comments on the manuscript. Peter Evans provided invaluable technical assistance.

LITERATURE CITED

1. **Anagnou, N. P., S. J. O'Brien, T. Shimada, W. G. Nash, M. Chen, and A. W. Nienhuis.** 1984. Chromosomal organization of the human dihydrofolate reductase genes: dispersion, selective amplification, and a novel form of polymorphism. Proc. Natl. Acad. Sci. USA **81**:5170–5174.
2. **Banerji, J., L. Olson, and W. Schaffner.** 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell **33**:729–740.
3. **Charnay, P., R. Treisman, P. Mellon, M. Chao, R. Axel, and T. Maniatis.** 1984. Diffferences in human α- and β-globin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. Cell **38**:251–263.
4. **Clarke, M., and J. A. Spudich.** 1977. Nonmuscle contractile proteins: the role of actin and myosin in cell motility and shape determination. Annu. Rev. Biochem. **46**:797–822.
5. **Elder, P. K., L. J. Schmidt, T. Ono, and M. J. Getz.** 1984. Specific stimulation of actin gene transcription by epidermal growth factor and cycloheximide. Proc. Natl. Acad. Sci. USA **81**:7476–7480.
6. **Emorine, L., M. Kuehl, L. Weir, P. Leder, and E. E. Max.** 1983. A conserved sequence in the immunoglobulin J_κ-C_κ intron: possible enhancer element. Nature (London) **304**:447–449.
7. **Engel, J., P. Gunning, and L. Kedes.** 1982. Human cytoplasmic actin proteins are encoded by a multigene family. Mol. Cell. Biol. **2**:674–684.
8. **Evans, T., E. Schon, G. Gora-Maslak, J. Patterson, and A. Efstratiadis.** 1984. S1-hypersensitive sites in eukaryotic promoter regions. Nucleic Acids Res. **12**:8043–8058.
9. **Freytag, S. O., H. O. Bock, A. L. Beaudet, and W. E. O'Brien.** 1984. Molecular structures of human argininosuccinate synthetase pseudogenes. J. Biol. Chem. **259**:3160–3166.
10. **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.
11. **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.
12. **Green, M. R., and R. G. Roeder.** 1980. Definition of a novel promoter for the major adenovirus-associated virus mRNA. Cell **22**:231–242.
13. **Greenberg, M. E., and E. B. Ziff.** 1984. Stimulation of 3T3 cells induces transcription of the c-*fos* proto-oncogene. Nature (London) **311**:433–438.
14. **Gunning, P., T. Mohun, S. Ng, P. Ponte, and L. Kedes.** 1984. Evolution of the human sarcomeric-actin genes: evidence for units of selection within the 3′ untranslated regions of the mRNAs. J. Mol. Evol. **20**:202–214.
15. **Gunning, P., P. Ponte, L. Kedes, R. Eddy, and T. Shows.** 1984.

Chromosomal location of the co-expressed human skeletal and cardiac actin genes. Proc. Natl. Acad. Sci. USA **81**:1813–1817.
16. **Gunning, P., P. Ponte, H. Okayama, J. Engel, H. Blau, and L. Kedes.** 1983. Isolation and characterization of full-length cDNA clones for human α-, β-, and γ-actin mRNAs: skeletal but not cytoplasmic actins have an amino-terminal cysteine that is subsequently removed. Mol. Cell. Biol. **3**:787–795.
17. **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.
18. **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.
19. **Hanukoglu, I., N. Tanese, and E. Fuchs.** 1983. Complementary DNA sequence of a human cytoplasmic actin. Interspecies divergence of 3′ noncoding regions. J. Mol. Biol. **163**:673–678.
20. **Hayday, A. C., S. D. Gillies, H. Saito, C. Wood, K. Wiman, W. S. Hayward, and S. Tonegawa.** 1984. Activation of a translocated human c-*myc* gene by an enhancer in the immunoglobulin heavy-chain locus. Nature (London) **307**:334–340.
21. **Karin, M., and R. I. Richards.** 1982. Human metallothionein genes: primary structure of the metallothionein-II gene and a related processed gene. Nature (London) **299**:797–802.
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. **Laimins, L. A., P. Gruss, R. Pozzatti, and G. Khoury.** 1984. Characterization of enhancer elements in the long terminal repeat of Moloney murine sarcoma virus. J. Virol. **49**:183–189.
24. **Laimins, L. A., G. Khoury, C. Gorman, B. Howard, and P. Gruss.** 1982. Host-specific activation of transcription by tandem repeats from simian virus 40 and Moloney murine sarcoma virus. Proc. Natl. Acad. Sci. USA **79**:6453–6457.
25. **Leavitt, J., G. Bushar, T. Kakunaga, H. Hamada, T. Hirakawa, D. Goldman, and C. Merril.** 1982. Variations in expression of mutant β actin accompanying incremental increases in human fibroblast tumorigenicity. Cell **28**:259–268.
26. **Leavitt, J., P. Gunning, P. Porreca, S. Ng, C. Lin, and L. Kedes.** 1984. Molecular cloning and characterization of mutant and wild-type human β-actin genes. Mol. Cell. Biol. **4**:1961–1969.
27. **Lee, M. G., S. A. Lewis, C. D. Wilde, and N. J. Cowan.** 1983. Evolutionary history of a multigene family: an expressed human β-tubulin gene and three processed pseudogenes. Cell **33**:477–487.
28. **Lemischka, I., and P. A. Sharp.** 1982. The sequences of an expressed rat α-tubulin gene and a pseudogene with an inserted repetitive element. Nature (London) **300**:330–335.
29. **Lu, R. C., and M. Elzinga.** 1977. Partial amino acid sequence of brain actin and its homology with muscle actin. Biochemistry **16**:5801–5806.
30. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
31. **Merrill, G. F., S. D. Hauschka, and S. L. McKnight.** 1984. tk enzyme expression in differentiating muscle cells is regulated through an internal segment of the cellular *tk* gene. Mol. Cell. Biol. **4**:1777–1784.
32. **Miyata, T., T. Yasunaga, and T. Nishida.** 1980. Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc. Natl. Acad. Sci. USA **77**:7328–7332.
33. **Moos, M., and D. Gallwitz.** 1982. Structure of a human β-actin related pseudogene which lacks intervening sequences. Nucleic Acids Res. **10**:7843–7849.
34. **Moos, M., and D. Gallwitz.** 1983. Structure of two human β-actin-related processed genes one of which is located next to a simple repetitive sequence. EMBO J. **2**:757–761.
34a. **National Foundation.** 1971. Birth defects, original article series VIII, vol. 7. Standardization and human genetics. Alan R. Liss, Inc., New York.
35. **Naylor, S. L., A. Y. Sakaguchi, T. B. Shows, M. L. Law, D. V. Goeddel, and P. W. Gray.** 1983. Human immune interferon gene is located on chromosome 12. J. Exp. Med. **157**:1020–1027.
36. **Nordheim, A., and A. Rich.** 1983. Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. Nature (London) **303**:674–679.
37. **Nudel, U., R. Zakut, M. Shani, S. Neuman, Z. Levy, and D. Yaffe.** 1983. The nucleotide sequence of the rat cytoplasmic β-actin gene. Nucleic Acids Res. **11**:1759–1771.
38. **Osborne, T. F., D. N. Arvidson, E. S. Tyau, M. Dunsworth-Browne, and A. J. Berk.** 1984. Transcription control region within the protein-coding portion of adenovirus E1A genes. Mol. Cell. Biol. **4**:1293–1305.
39. **Picard, D., and W. Schaffner.** 1984. A lymphocyte-specific enhancer in the mouse immunoglobulin κ gene. Nature (London) **307**:80–82.
40. **Piechaczyk, M., J. M. Blanchard, S. Riaad-El Sabouty, C. Dani, L. Marty, and P. Jeanteur.** 1984. Unusual abundance of vertebrate 3-phosphate dehydrogenase pseudogenes. Nature (London) **312**:469–471.
41. **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.
42. **Ponte, P., S. 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.
43. **Prentki, P., and H. M. Krisch.** 1982. A modified pBR322 vector with improved properties for the cloning, recovery and sequencing of blunt-ended DNA fragments. Gene **17**:189–196.
44. **Queen, C., and D. Baltimore.** 1983. Immunoglobulin gene transcription is activated by downstream sequence elements. Cell **33**:741–748.
45. **Rigby, P. W., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling DNA to high specific activity *in vitro* by nick-translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
46. **Shah, D. M., R. C. Hightower, and R. B. Meagher.** 1983. Genes encoding actin in higher plants: intron positions are highly conserved but the coding sequences are not. J. Mol. Appl. Genet. **2**:111–126.
47. **Shows, T., R. Eddy, L. Haley, M. Byers, M. Henry, T. Fujita, H. Matsui, and T. Taniguchi.** 1984. Interleukin-2 (IL2) is assigned to human chromosome 4. Somat. Cell Mol. Genet. **10**:315–318.
48. **Shows, T. B., J. A. Brown, L. L. Haley, M. G. Byers, R. L. Eddy, E. S. Cooper, and A. P. Goggin.** 1978. Assignment of the β-glucuronidase structural gene to the pter→q22 region of chromosome 7 in man. Cytogenet. Cell Genet. **21**:99–104.
49. **Soriano, P., P. Szabo, and G. Bernardi.** 1982. The scattered distribution of actin genes in the mouse and human genomes. EMBO J. **1**:579–583.
50. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
51. **Ueyama, H., H. Hamada, N. Battula, and T. Kakunaga.** 1984. Structure of a human smooth muscle actin gene (aortic type) with a unique intron site. Mol. Cell. Biol. **4**:1073–1078.
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. A protein-chemical analysis of muscle actin differentiation. Differentiation **14**:123–133.
53. **van Straaten, F., R. Muller, T. Curran, C. Van Beveren, and I. M. Verma.** 1983. Complete nucleotide sequence of a human c-*onc* gene: deduced amino acid sequence of the human c-*fos* protein. Proc. Natl. Acad.Sci. USA **80**:3183–3187.
54. **Weaver, R. F., and C. Weissmann.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucleic Acids Res. **7**:1175–1193.
55. **Weeds, A.** 1982. Actin-binding proteins: regulators of cell architecture and motility. Nature (London) **296**:811–816.
56. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80**:726–730.