Volume 11 Number 6 1983 Nucleic Acids Research

The nucleotide sequence of the rat cytoplasmic β-actin gene

U.Nudel, R.Zakut, M.Shani, S.Neuman, Z.Levy and D.Yaffe

Department of Cell Biology, The Weizman Institute of Science, Rehovot, 76100 Israel

Received 5 January 1983; Accepted 1 March 1983

ABSTRACT

The nucleotide sequence of the rat β-actin gene was determined. The gene codes for a protein identical to the bovine β-actin. It has a large intron in the 5' untranslated region 6 nucleotides upstream from the initiator ATG, and 4 introns in the coding region at codons specifying amino acids 41/42, 121/122, 267, and 327/328. Unlike the skeletal muscle actin gene and many other actin genes, the β-actin gene lacks the codon for Cys between the initiator ATG and the codon for the N-terminal amino acid of the mature protein.

The usage of synonymous codons in the β-actin gene is nonrandom, and is similar to that in the rat skeletal muscle and other vertebrate actin genes, but differs from the codon usage in yeast and soybean actin genes.

INTRODUCTION

The actins constitute a family of highly conserved proteins found in all eukaryotes. On the basis of the amino acid sequence at least 6 different mammalian actins were identified: skeletal muscle, heart muscle, 2 smooth muscle actins and the cytoplasmic β- and γ-actins. There are 4-6 amino acid replacements between the different muscle type actins; 4 amino acid replacements between the 2 cytoplasmic actins and 25 amino acid replacements between the cytoplasmic and the skeletal muscle actins. There are no amino acid replacements between bovine, rabbit, rat and chicken skeletal muscle actins (1-6). While the muscle actins play a major role in muscle contraction, the cytoplasmic actins are involved in many forms of cell motility and in cytoskeletal functions. The vertebrate non-muscle β- and γ-actins are considered functionally and evolutionarily to be more closely related to the actins found in the lower, unicellular, eukaryotes. In previous publications we described the nucleotide sequence of the rat skeletal muscle actin gene, and the isolation and preliminary characterization of the β-actin gene (6, 7). On the basis of these data and the available data on the structure of actin genes isolated from other organisms we suggested that the separation between the skeletal muscle and cytoplasmic actin genes occurred within the deuterostome

phylogenetic branch (6-8). Here we present the complete nucleotide sequence of the β-actin gene, confirming our earlier suggestion on the structure of this gene.

MATERIALS AND METHODS

Isolation of DNA

DNA from recombinant bacteriophages was prepared as described by Yamamoto et al. (9). Plasmids were prepared by the alkaline extraction method (10) and purified by centrifugation to equilibrium in cesium chloride-ethidium bromide gradients.

Subcloning of DNA fragments

The 8 kb EcoR1 digested DNA fragment was isolated from an agarose gel, purified on a DEAE-cellulose column and ligated into the EcoR1 site of pBR322. The two subclones containing the EcoR1-HindIII DNA fragments were constructed in the same way, except that the insert was ligated to pBR322 from which the small EcoR1-HindIII fragment had been removed.

DNA sequencing

DNA fragments were end-labelled either by filling in with reverse transcriptase, or by the exchange reaction using Klenow enzyme, or by labelling the 5' end with polynucleotide kinase. The fragments were sequenced by the technique described by Maxam and Gilbert (11).

Endonuclease S1 mapping

End-labelled, single- or denatured double-stranded DNA fragments were hybridized to polyA+ RNA from dividing myoblasts, containing β- and γ-actin mRNA but no detectable amounts of skeletal muscle actin mRNA (12), and were treated with endonuclease S1, as described by Berk and Sharp (13). The sizes of the protected fragments were determined by electrophoresis on polyacrylamide-urea sequencing gels (11).

Computer analysis of sequence homology

A two-dimensional dot matrix homology analysis was done as described by Konkel et al. (14) and modified by Unger and Zussman (in preparation).

RESULTS AND DISCUSSION

The Coding Region of β-Actin Gene

We have previously described the isolation, from a rat gene library, of a recombinant bacteriophage (Act 1) containing a gene coding for the rat cytoplasmic β-actin (6, 7). A 8 kb EcoR1 digested DNA fragment containing the

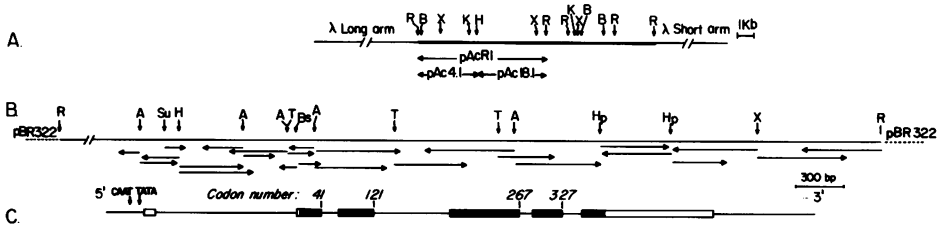


Fig. 1. The structure of the rat β-actin gene

(A) A restriction endonuclease map of the genomic DNA clone Act 1. pAcR1, pAc4.1 and pAc18.1 are the DNA fragments that were subcloned in pBR322. (B) The strategy of sequencing the structural gene and its flanking region. Most of the sequence was determined using plasmids pAc18.1 and pAc4.1, except for the HindIII site region which was determined using pAcR1. Restriction enzymes used in A and B were: AvaI (A); BamH1 (B); BstN1 (Bs); HgiA1 (Hg); HindIII (H); HpaII (Hp); KpnI (K); EcoR1 (R); Sau3 (Su); TagI (T); XbaI (X). (C) Schematic representation of the structure of the rat β-actin gene. Black bars: coding region; open bars: untranslated region; thin lines: introns and flanking region.

structural gene was subcloned in pBR322 (Fig. 1a). The two HindIII-EcoR1 DNA fragments of the resulting plasmid, pAcR1, were also subcloned in pBR322. The 3 recombinant plasmids pAcR1, pAc 18.1, and pAc 4.1 were used to sequence the structural gene and its flanking DNA as described in Fig. 1b. Most of the sequence was determined by sequencing both DNA strands or by sequencing the same strand using two or more partially overlapping fragments.

The nucleotide sequence (Fig. 2a) codes for a protein which is identical to the bovine β-actin (5). (The amino acid sequence of rat β-actin has not been determined.) This again demonstrates the great conservation of actins during evolution.

We have previously shown that the coding sequence of the skeletal muscle actin gene begins with codons for Met-Cys which are absent from the mature protein and are probably cleaved off after translation (6). The same 2 codons are present in the chick skeletal muscle actin gene (15), in human cardiac muscle actin gene (16) and in all six drosophila actin genes (17), as well as in the nematode (18) and sea urchin actin genes (19). It is of interest that this Cys codon is absent from the β-actin gene. In addition, the β-actin is one amino acid shorter than the skeletal muscle actin (5). Alignment of the nucleotide sequence of the 2 genes suggests a deletion in the β-actin gene of two codons following the initiator ATG:

A

```
 -1   2                                  10                                                  20
Met Asp Asp Asp Ile Ala Ala Leu Val Val Asp Asn Gly Ser Ala Met Cys Lys Ala Gly Phe Ala Gly
ATG GAT GAC GAT ATC GCT GCG CTC GTC GTC GAC AAC GGG TCC GCC ATG TGC AAG GCC GGC TTC GCG GGC
                Ava 1     30                                                  40
Asp Asp Ala Pro Arg Ala Val Phe Pro Ser Ile Val Gly Arg Pro Arg His Gln
GAC GAT GCT CCC CGG GCC GTC TTC CCC TCC ATC GTG GGC CGC CCT AGG CAC CAG GTAAGTGACCCTTTACTTT

                                                                                  Gly Val Met Val Gly
GGGAGTGGCAGCCCTAGGGTTTTCTTGGGGGTCGATGCCAGTGCTGAGAACGTTGTTCTCCTCCGCAG GGT GTG ATG GTG GGT
                50                                                  60
Met Gly Gln Lys Asp Ser Tyr Val Gly Asp Glu Ala Gln Ser Lys Arg Gly Ile Leu Thr Leu Lys Tyr
ATG GGT CAG AAG GAC TCC TAC GTG GGC GAC GAG GCC CAG AGC AAG AGA GGC ATC CTG ACC CTG AAG TAC
 70                                 80                                                  90
Pro Ile Glu His Gly Ile Val Thr Asn Trp Asp Asp Met Glu Lys Ile Trp His His Thr Phe Tyr Asn
CCC ATT GAA CAC GGC ATT GTA ACC AAC TGG GAC GAT ATG GAG AAG ATT TGG CAC CAC ACT TTC TAC AAT
                          100                                                 110
Glu Leu Arg Val Ala Pro Glu Glu His Pro Val Leu Leu Thr Glu Ala Pro Leu Asn Pro Lys Ala Asn
GAG CTG CGT GTG GCC CCT GAG GAG CAC CCT GTG CTG CTC ACC GAG GCC CCT CTG AAC CCT AAG GCC AAC
                    120
Arg Glu Lys Met Thr Gln
CGT GAA AAG ATG ACC CAG GTCAGTATCCTGGGTGACCCTCCCCTTCTTATTGGGTCAACTTCTCAGCACGCCCTTCTCTAATTGT

CTTTCTTCTGCCATGTCCCATAGGACTCTCTTCTATGAGCTGAGTCTCCCTTGGAACTTTGCAGTTTCTGCTCTTTCCCAGATGAGGTCTT

TTTTTCTCTCGATCGCCTTTCTGACTAGGTGTTTAAACCCCTACAGTGCTGTGGGTGTAGGTACTAACAATGGCTCGTGTGACAAAGCT

AATGAGGCTGGTGATAAATGGCCTTGGAGTGTGTATTCAGTAGATGACAGTAGGTCTAAATGGAGCCCCTGTCCTGATACTCCCAGCACAC

TTAACTTAGCTGTGTTCTTGCCCTCTTTGCATGTCTCACTCAAATCTATCCTTACAGTCTCACCTGCCCTGAGTGTTTCTTGTGGCTTTAG
                                                                                  130
                                    Ile Met Phe Glu Thr Phe Asn Thr Pro Ala Met Tyr Val Ala
GAGCTTGACAATACTGTATTCCTTTCTCTACAG ATC ATG TTT GAG ACC TTC AAC ACC CCA GCC ATG TAC GTA GCC
                  140                                               150
Ile Gln Ala Val Leu Ser Leu Tyr Ala Ser Gly Arg Thr Thr Gly Ile Val Met Asp Ser Gly Asp Gly
ATC CAG GCT GTG TTG TCC CTG TAT GCC TCT GGT CGT ACC ACT GGC ATT GTG ATG GAC TCC GGA GAC GGG
     160                                 170                                             180
Val Thr His Thr Val Pro Ile Tyr Glu Gly Tyr Ala Leu Pro His Ala Ile Leu Arg Leu Asp Leu Ala
GTC ACC CAC ACT GTG CCC ATC TAT GAG GGT TAC GCG CTC CCT CAT GCC ATC CTG CGT CTG GAC CTG GCT
                                    190                                             200
Gly Arg Asp Leu Thr Asp Tyr Leu Met Lys Ile Leu Thr Glu Arg Gly Tyr Ser Phe Thr Thr Thr Ala
GGC CGG GAC CTG ACA GAC TAC CTC ATG AAG ATC CTG ACC GAG CGT GGC TAC AGC TTC ACC ACC ACA GCT
                        210                                             220
Glu Arg Glu Ile Val Arg Asp Ile Lys Glu Lys Leu Cys Tyr Val Ala Leu Asp Phe Glu Gln Glu Met
GAG AGG GAA ATC GTG CGT GAC ATT AAA GAG AAG CTG TGC TAT GTT GCC CTA GAC TTC GAG CAA GAG ATG
          230                 234a                240
Ala Thr Ala Ala Ser Ser Ser Ser Leu Glu Lys Ser Tyr Glu Leu Pro Asp Gly Gln Val Ile Thr Ile
GCC ACT GCC GCA TCC TCT TCC TCC CTG GAG AAG AGC TAT GAG CTG CCT GAC GGT CAG GTC ATC ACT ATC
250                                 260
Gly Asn Glu Arg Phe Arg Cys Pro Glu Ala Leu Phe Gln Pro Ser Phe Leu G
GGC AAT GAG CGG TTC CGA TGC CCC GAG GCT CTC TTC CAG CCT TCC TTC CTG GGTAAGTTGTAGTCTCGTCCCTT
                                                                                  270
                                                                    ly Met Glu Ser Cys Gly
CTCCATCTAAAGGTGACCAATGCTGGAGGCCACACTGTAACTCTGATCTCTTGCTTTTCCTTTCAG GT ATG GAA TCC TGT GGC
                                    280                                             290
Ile His Glu Thr Thr Phe Asn Ser Ile Met Lys Cys Asp Val Asp Ile Arg Lys Asp Leu Tyr Ala Asn
ATC CAT GAA ACT ACA TTC AAT TCC ATC ATG AAG TGT GAC GTT GAC ATC CGT AAA GAC CTC TAT GCC AAC
                  300                                               310
Thr Val Leu Ser Gly Gly Thr Thr Met Tyr Pro Gly Ile Ala Asp Arg Met Gln Lys Glu Ile Thr Ala
ACA GTG CTG TCT GGT GGC ACC ACC ATG TAC CCA GGC ATC GCT GAC AGG ATG CAG AAG GAG ATT ACT GCC
    320
Leu Ala Pro Ser Thr Met Lys Ile Lys
CTG GCT CCT AGC ACC ATG AAG ATC AAG GTAAGCAGCCTTAGCCTGGACCCATAGTGGGGTGTGGTCAGCCCTGTAGTTGTAG
                                                                                  330
                                                                      Ile Ile Ala Pro Pro
CCAACTCTCTTGGCTTAAGGAACAACCCAGCATCCAGAATGCTCACAATCACTGTCTTGCTTTCTTCAG ATC ATT GCT CCT CCT
                                    340                                             350
Glu Arg Lys Tyr Ser Val Trp Ile Gly Gly Ser Ile Leu Ala Ser Leu Ser Thr Phe Gln Gln Met Trp
GAG CGC AAG TAC TCT GTG TGG ATT GGT GGC TCT ATC CTG GCC TCA CTG TCC ACC TTC CAG CAG ATG TGG
                  360                                               370
Ile Ser Lys Gln Glu Tyr Asp Glu Ser Gly Pro Ser Ile Val His Arg Lys Cys Phe Ter
ATC AGC AAG CAG GAG TAC GAT GAG TCC GGC CCC TCC ATC GTG CAC CGC AAA TGC TTC TAG GCGGACTGTTA
```

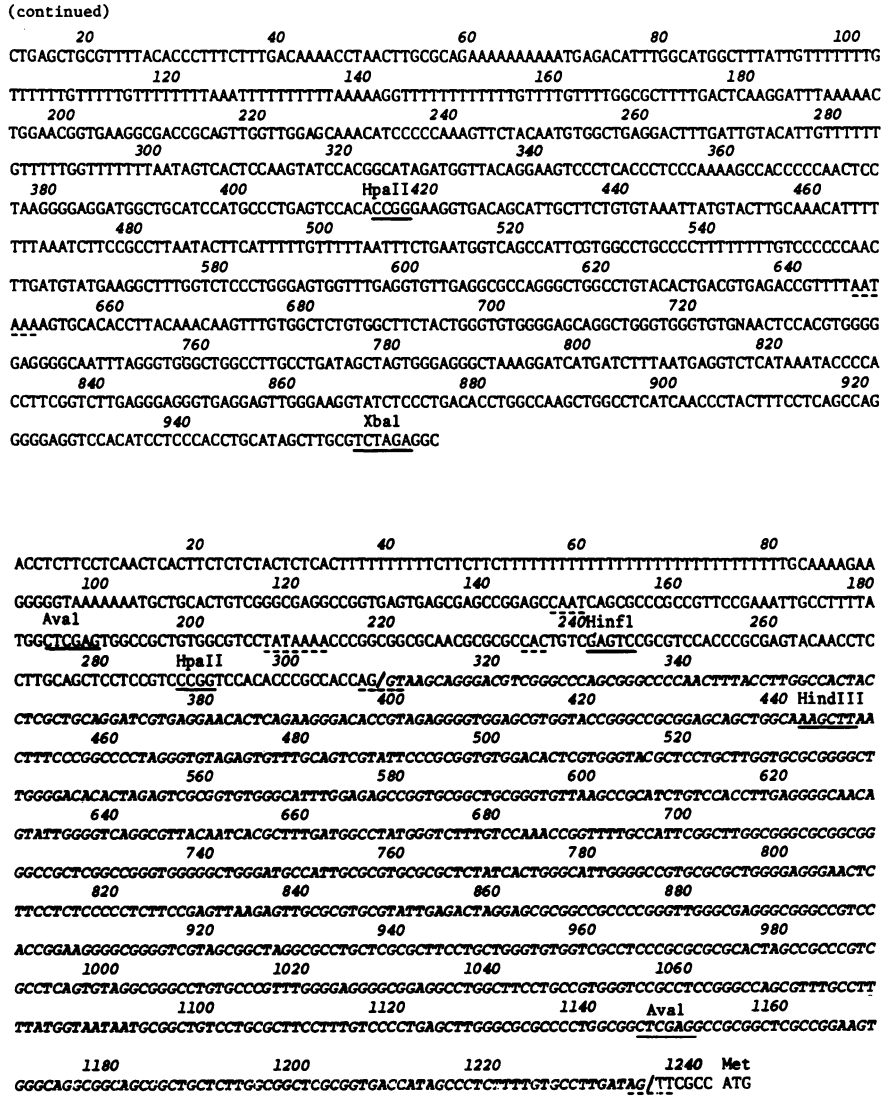


Fig. 2. Nucleotide sequence of the rat β-actin gene

(a) The coding and 3' untranslated regions; (b) the 5' untranslated region. The numbering of the amino acids (above) is according to Vanderkerckhove and Weber (5). Numbers in italics specify nucleotides in the 3' untranslated region (in a) and in the 5' untranslated region (in b). The restriction sites used for the endonuclease S1 mapping described in Fig. 3 are underlined. The CAAT sequence, TATA box, cap site, polyadenylation signal and exon/intron junctions in the untranslated regions are underlined with broken lines. The sequence of the introns is marked in italics.

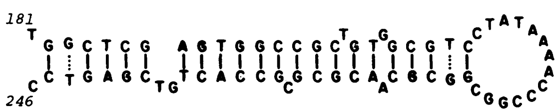
					1					5			
Skeletal			Met	Cys	Asp	Glu	Asp	Glu	Thr	Thr	Ala	Leu	Val
muscle	*CTCTTTGCAG*/AAACTAGACACC		ATG	TGT	GAC	GAG	GAC	GAG	ACC	ACC	GCT	CTT	GTG
			Met			Asp	Asp	Asp	Ile	Ala	Ala	Leu	Val
β	*TTTTGTGCCTTGATAG*/TTCGCC		ATG	---	---	GAT	GAC	GAT	ATC	GCT	GCG	CTC	GTC

The nucleotide sequence of the β-actin gene confirms our results obtained by partial DNA sequencing indicating that the coding sequence of the gene is interrupted by 4 introns at the codons specifying amino acids 41/42, 121/122, 267 and 327/328 (6, 7). The sequences of the splice sites are in accordance with the consensus sequence for splice sites (20, 21).

The 5' Untranslated Region

The electronmicrographs of R-loops formed between the β-actin gene and mRNA from dividing myoblasts indicated the existence of a large intron near the 5' end of the gene (7). From the complete nucleotide sequence it is evident that this intron is not located within the coding region. The signals associated with accurate initiation of transcription were identified at about 1000 nucleotides upstream from the initiator ATG (Fig. 2b). A sequence TATAAA (Goldberg-Hogness box, Refs. 21, 22) was identified at nucleotides 207-213 (Fig. 2b), and a CCAATC sequence (23) was found 60 nucleotides upstream from the TATA box sequence. A cap site sequence CAC was found 28 nucleotides downstream from the TATA box. The assignment of 5' and 3' borders of the large intron as well as the confirmation of the assignment of the cap site to nucleotide 235 were done by S1 endonuclease mapping, as described in Fig. 3. As can be seen in Fig. 2b, the large intron is 925 nucleotides long. It interrupts the 5' untranslated region 6 nucleotides upstream from the initiator ATG.

Several interesting features, the biological significance of which is yet unknown, are found near the 5' region of the gene: 1) A track of 44 thymidilic acid residues (interrupted by three residues of cytidilic acid) is located 66 nucleotides upstream from the CAAT box. Long tracks of T were found also 5' to actin genes of Dictyostelium (24) but not 5' to the rat skeletal muscle actin gene (6). 2) A C+G-rich sequence with dyad symmetry, with only 20% mismatches is found at nucleotides 182-205 and 205-245. Formation of a stem and loop structure leaves the TATA box in the loop; the transcription initiation site is in the stem:



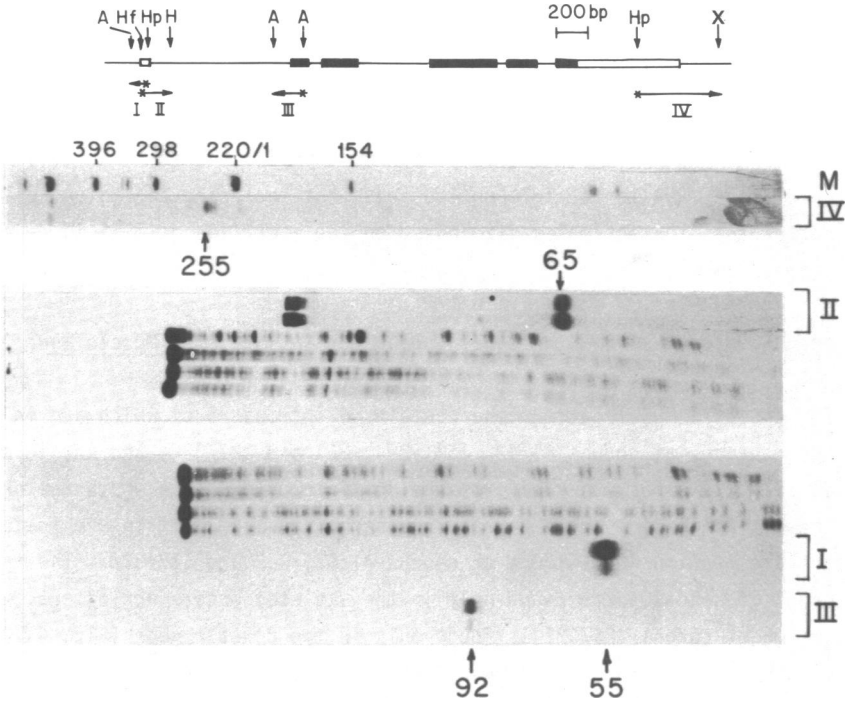


Fig. 3. Endonuclease S1 mapping of the cap site, the polyadenylation site and the borders of the large intron in the 5' untranslated region

Based on the data from the R-loop mapping and the DNA sequencing, the following DNA fragments were prepared for the S1 endonuclease mapping: (I) AvaI-HpaII (nucleotides 184-288; Fig. 2b); (II) HinfI-HindIII (241-442; Fig. 2a); (III) AvaI-AvaI (nucleotide 1145-codon 26; Figs. 2b and 2a, respectively); (IV) HpaII-XbI (414-958; Fig. 2b). Fragments I and III were labelled with polynucleotide kinase and fragments II and IV by filling in with reverse transcriptase. The noncoding strands or the denatured double-stranded fragments were hybridized with mRNA from proliferating myoblasts. Each fragment was hybridized with poly(A)+ RNA (0.2 and 1 µg/20 µl). The samples were then treated with S1 endonuclease (13) and electrophoresed on 6% polyacrylamide sequencing gels (11). The digestion products of fragments I, II and III were run on gels with 4 DNA sequencing reactions as size markers. The digestion product of fragment IV was run on a gel using pBR322 fragments as size markers. After electrophoresis, the gels were fluorographed.

3) The C+G content of the large intron (66.4%) is considerably higher than that of other introns (48 %). Simple sequences with alternating CG are abundant in this intron.

The 3' Untranslated Region

A potential polyadenylation signal AATAAA (25) was identified at nucleotides 646-651 (Fig. 2a). S1 endonuclease mapping (Fig. 3) indicated that the

polyadenylation site is ca. 20 nucleotides downstream from this site. Thus, the 3' untranslated region of β-actin mRNA is ca. 670 nucleotides long (excluding the polyA tail). It is much longer than the 3' untranslated region of the skeletal muscle actin mRNA (which is 241 nucleotides long). It contains several regions which are very rich in thymidilic acid. The full length of the β-actin mRNA is ca. 1870 nucleotides (not including the polyA tail). It is shorter than the estimated size based on the migration of mRNA in denaturing agarose gels (26).

Sequence Homology Between the Genes Coding for Rat Skeletal Muscle and β-Actin

The skeletal muscle actin gene contains 6 introns, 5 of which are in the coding sequence, at codons 41/42, 150, 204, 267 and 327/328, and one large intron 12 nucleotides upstream from the initiator ATG. Thus, both the skeletal muscle actin gene and the β-actin gene have a large intron in the 5' untranslated region and introns at codons 41/42, 267 and 327/328. The introns at codons 150 and 204 are found only in the skeletal muscle actin gene, while the intron at codons 121/122 is found only in the β-actin gene (Fig. 4). The splice sites of the three homologous introns in the coding region are exactly in the same position with respect to the reading frames. There are no amino acid replacements at these exon/intron junctions. It is of interest that a sea urchin actin gene was isolated that contained introns at codons 41/42, 121/122, 204 and 267 (19). Thus, this gene contained an intron which is found in the skeletal muscle actin gene but is absent from the β-actin gene (at codon 204), and an intron which is found in the β-actin gene and is absent from the skeletal muscle actin gene (at codons 121/122). This suggests that the skeletal muscle actin gene and the β-actin gene evolved from an actin gene which contained both introns.

We have previously compared the location of introns in the rat skeletal muscle and β-actin genes with that of actin genes from other organisms and discussed the possible evolutionary implications (6, 8). We suggested that the skeletal muscle actin probably evolved from a nonmuscle actin within the phylogenetic branch of the deuterostomes, implying that the vertebrate and the insect muscle actins evolved independently from nonmuscle actin genes.

Figure 5 shows the comparison of the nucleotide sequence of the rat skeletal muscle and β-actin genes by two-dimensional dot matrix homology analysis (14), using conditions at which a dot is printed if at least 6 out of 7 nucleotides are homologous. It is evident that under these conditions only the

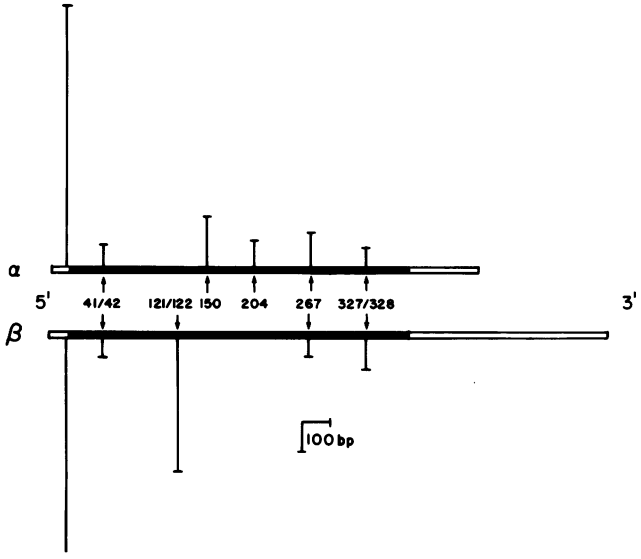


Fig. 4. Schematic comparison of the structure of the rat skeletal muscle and cytoplasmic β-actin genes

Black bars and open bars: the translated and untranslated regions of the exons, respectively. Vertical lines: introns. Numbers indicate the codons in the exon/intron splice junctions.

coding sequences show a high degree of homology. The regions between codons 1-40 and 122-150 are the most divergent, while the regions between codons 42-110, 204-267 and 335-374 show the highest degree of homology. It is also evident from Fig. 5 and from additional comparisons which we made using several different stringency requirements for homology (data not shown) that the 5' and 3' noncoding regions as well as the introns of the skeletal muscle and β-actin genes are very divergent, with only a few short sequences which may share homology. Part of this divergence (which is greater than that of the silent sites within the coding regions) is most probably due to deletions and/or insertions of sequences in the introns and in the untranslated regions. This is also reflected in the variation of sizes in these regions of the two genes.

We compared the nucleotide sequence of the 3' noncoding region of the rat β-actin gene to that of a cDNA clone of a human cytoplasmic actin gene containing 403 nucleotides of the 3' untranslated region (prepared by Hanukoglu et al. (27)). Most of the sequenced human 3' untranslated region shows $\geq$85% homology to the corresponding rat sequences.

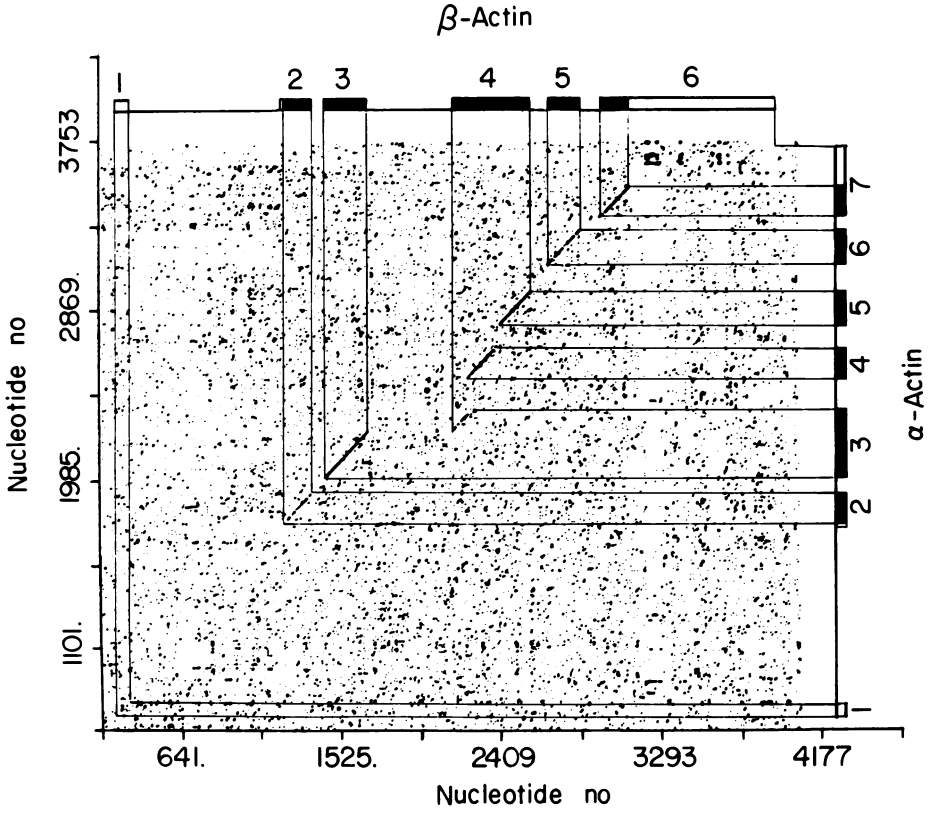


Fig. 5. Two-dimensional dot-matrix homology analysis of rat skeletal muscle and cytoplasmic β-actin genes

Each dot represents a homology of 6 out of 7 successive nucleotides. The numbering of nucleotides is arbitrary, starting from the beginning of the sequenced region upstream from the structural genes. In the schematic structures of the genes the black bars and open bars represent the translated and untranslated parts of the exons, respectively. Thin lines represent introns. The numbers are of exons, starting from the 5' end of the genes.

Codon Usage

The great conservation of actin genes through a wide taxonomic range and the existence of several tissue-specific isoforms within the same organism make it possible to examine the question of whether codon usage is random, tissue-specific or genetically determined, by comparing the frequency of usage of synonymous codons in the various actin genes.

Table 1 shows the frequency of usage of the different codons specifying the same amino acids in various actin genes. It is clear that the usage of

Table 1. Codon usage in actin genes

		Human cardiac	Rat β	Rat α	Chick α	Sea urchin	Amoeba	Yeast	Soybean
Arg	CGA	0	1	2	0	1	0	0	3
	CGC	6	3	6	6	1	14	0	1
	CGG	3	3	2	1	0	0	0	0
	CGT	8	7	6	8	9	3	5	6
	AGA	1	1	0	1	2	0	13	6
	AGG	0	3	2	2	5	1	0	2
Leu	CTA	1	1	0	1	0	0	2	3
	CTC	6	6	5	5	11	18	0	9
	CTG	16	19	17	16	4	8	0	0
	CTT	0	0	2	1	10	3	2	11
	TTA	1	0	0	0	0	0	2	0
	TTG	3	1	2	3	2	0	19	4
Ser	TCA	0	1	2	0	3	0	3	5
	TCC	11	14	11	16	13	6	12	5
	TCG	0	0	1	0	0	17	0	0
	TCT	8	5	5	3	5	1	14	7
	AGC	5	5	4	4	4	0	0	7
	AGT	0	0	0	0	1	0	3	7
Thr	ACA	5	4	2	7	2	0	0	3
	ACC	13	15	20	16	16	22	8	6
	ACG	0	0	1	0	1	1	0	0
	ACT	8	7	4	4	3	3	13	9
Pro	CCA	1	2	2	4	10	0	13	7
	CCC	10	6	10	10	9	15	0	6
	CCG	3	0	2	0	0	3	0	0
	CCT	5	11	5	5	1	1	6	5
Ala	GCA	2	1	0	0	1	0	0	9
	GCC	13	18	15	18	19	23	11	8
	GCG	2	3	2	1	0	2	0	0
	GCT	12	9	12	10	10	4	15	11
Gly	GGA	2	1	0	1	13	2	0	8
	GGC	15	14	13	12	5	16	0	4
	GGG	3	2	5	5	0	0	0	3
	GGT	8	9	10	10	10	12	28	14

		Human cardiac	Rat β	Rat α	Chick α	Sea urchin	Amoeba	Yeast	Soybean
Val	GTA	2	2	0	1	1	0	0	3
	GTC	10	5	5	8	10	15	12	7
	GTG	6	13	16	12	3	6	0	5
	GTT	3	2	0	0	9	2	14	13
Ile	ATA	0	0	0	0	0	0	0	1
	ATC	18	20	25	22	27	24	16	11
	ATT	12	8	5	8	2	3	14	16
Lys	AAA	3	3	5	3	1	0	6	7
	AAG	16	17	14	16	18	19	12	13
Asn	AAC	5	6	9	9	9	9	9	4
	AAT	7	3	3	3	0	0	0	4
Gln	CAA	3	1	1	4	1	1	14	7
	CAG	8	11	10	7	12	10	0	2
His	CAC	7	7	8	7	7	10	7	3
	CAT	2	2	1	2	2	0	3	7
Glu	GAA	7	4	3	7	6	0	26	12
	GAG	21	21	25	21	20	28	1	17
Asp	GAC	14	18	18	16	15	18	9	6
	GAT	8	5	4	6	8	3	11	17
Tyr	TAC	10	10	10	8	15	14	14	2
	TAT	6	5	6	8	0	1	0	11
Cys	TGC	3	4	4	5	5	4	0	3
	TGT	3	2	2	1	1	0	4	1
Phe	TTC	9	12	9	8	13	11	12	6
	TTT	3	1	3	4	1	2	2	5

Data for human cardiac actin gene were taken from Ref. 16; for rat α actin gene from Ref. 6; for chick α actin gene, Ref. 15; for sea urchin actin gene I, Ref. 19; for amoeba actin gene I, Ref. 32; for yeast actin gene, Refs. 33, 34; and for soybean actin gene, Ref. 35.

synonymous codons is not random. There is a preferential usage of C and T in the third position, while A is much less frequent at this position. The general pattern of codon usage is quite similar in the rat β-actin gene and in the rat and chick skeletal muscle actin genes and in the human cardiac actin gene. It is similar to the pattern of codon usage in other genes in

mammals (28). It is, however, different from that of actin genes in yeast and soybean. The pattern of codon usage in the sea urchin actin gene is closer to that of vertebrates than to that of yeast and soybean. These data are compatible with the suggestion that the pattern of codon usage is not tissue-specific. Rather it suggests phylogenetic differences in pattern of codon usage (28), which is probably closely related to the abundance of the various isoacceptor tRNAs (29-31).

ACKNOWLEDGMENTS

Supported by NIH Grant RO1 GM 22767, the Muscular Dystrophy Association, Inc., U.S.A., the U.S.-Israel Binational Science Foundation, Jerusalem. Uri Nudel is the incumbent of the A. and E. Blum Career Development Chair in Cancer Research.

We wish to thank Dr. Y. Zussman and Mr. R. Unger for programming the 2-dimensional dot matrix homology computer analysis and Drs. L.L. Jagodzinsky, T. Sargent and J. Bonner for the rat gene libraries.

REFERENCES

1 Storti, R.V. and Rich, A. (1976) Proc. Natl. Acad. Sci. USA 73, 2346-2350
2 Whalen, R.G., Butler-Browne, G.S. and Gros, F. (1976) Proc. Natl. Acad. Sci. USA 73, 2018-2022
3 Garrels, J. and Gibson, W. (1976) Cell 9, 793-805
4 Elzinga, M. and Lu, R. (1976) in Contractile Systems in Nonmuscle Tissues (eds. Perry, S.V., Margreth, A. and Adelstein, R.), North-Holland, Amsterdam, p. 29
5 Vanderkerckhove, J. and Weber, K. (1979) Differentiation 14, 124-133
6 Zakut, R., Shani, M., Givol, D., Neuman, S., Yaffe, D. and Nudel, U. (1982) Nature 298, 857-859
7 Nudel, U., Katcoff, D., Zakut, R., Shani, M., Carmon, Y., Finer, M., Czosnek, H., Ginsburg, I. and Yaffe, D. (1982) Proc. Natl. Acad. Sci. USA 79, 2753-2767
8 Nudel, U., Zakut, R., Katcoff, D., Carmon, Y., Czosnek, H., Shani, M. and Yaffe, D. (1982) in Muscle Development: Molecular and Cellular Control (eds. Pearson, M.L. and Epstein, H.F.), Cold Spring Harbor Lab., New York, in press
9 Yamamoto, K.R., Alberts, B.M., Benzinger, R., Lawhorne, L. and Treiber, G. (1970) Virology 40, 734-744
10 Birnboim, H.C. and Doly, J. (1979) Nucl. Acids Res. 7, 1513-1523
11 Maxam, A.M. and Gilbert, W.R. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564
12 Shani, M., Zevin-Sonkin, D., Saxel, O., Carmon, Y., Katcoff, D., Nudel, U. and Yaffe, D. (1981) Dev. Biol. 86, 483-492
13 Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732
14 Konkel, D.A., Maizel, J.V. and Leder, P. (1979) Cell 18, 865-873
15 Fronwald, J.A., Kuncio, G., Peng, I. and Ordahl, C.P. (1982) Nucl. Acids Res. 10, 3861-3876
16 Hamada, H., Petrino, M.G. and Kakunaga, T. (1982) Proc. Natl. Acad. Sci. USA 79, 5901-5905

17 Fryberg, E.A., Bond, B.J., Hershel, N.D., Mixter, K.S. and Davidson, N. (1981) Cell 24, 107-116
18 Files, J.G., Carr, S. and Hirsch, D., submitted for publication
19 Cooper, A.D. and Crain, W.R., Jr. (1982) Nucl. Acids Res. 10, 4081-4092
20 Breathnach, R., Benoist, C., O'Hare, K., Gannon, G. and Chambon, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4853-4857
21 Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383
22 Goldberg, M.L. (1979) Ph.D. Thesis, Stanford University
23 Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 653-668
24 McKeown, M. and Firtel, R.A. (1981) Cell 24, 799-807
25 Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214
26 Katcoff, D., Nudel, U., Zevin-Sonkin, D., Carmon, Y., Shani, M., Lehrach, H., Frischauf, A.M. and Yaffe, D. (1980) Proc. Natl. Acad. Sci. USA 77, 960-964
27 Hanukoglu, I., Tanese, N.T. and Fuchs, E., J. Mol. Biol., in press
28 Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Pave, A. (1980) Nucl. Acids Res. 8, 749-762
29 Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3026-3031.
30 Ikemura, T. (1982) J. Mol. Biol. 158, 573-597.
31 Gouy, M. and Gautier, C. (1982) Nucl. Acids Res. 10, 7055-7074.
32 Nellen, W. and Gallwitz, D. (1982) J. Mol. Biol. 159, 1-18
33 Gallwitz, D. and Sures, I. (1980) Proc. Natl. Acad. Sci. USA 77, 2546-2550
34 Ng, R. and Abelson, J. (1980) Proc. Natl. Acad. Sci. USA 77, 3912-3916
35 Shah, O.M., Hightower, R.C. and Meagher, R. (1982) Proc. Natl. Acad. Sci. USA 79, 1022-1026