Volume 11 Number 11 1983 Nucleic Acids Research

Isolation and characterization of cDNA clones for human skeletal muscle α actin

A.Hanauer[1], M.Levin[2], R.Heilig[1], D.Daegelen[2], A.Kahn[2] and J.L.Mandel[1]*

[1]Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Institute de Chimie Biologique, 11 Rue Humann, 67085 Strasbourg Cedex, and [2]Institut de Pathologie Moléculaire, Unité 129 de l'INSERM, Centre Universitaire de Cochin Port Royal, 75674 Paris Cedex 14, France

Received 14 April 1983; Accepted 6 May 1983

SUMMARY

Two cDNA libraries corresponding to polyA$^+$ RNA from human adult skeletal muscle have been constructed by cloning in the PstI site of pBR322. Skeletal α actin cDNA clones have been isolated and characterized. Three of these plasmids have overlapping inserts which together contain the complete 5' non-coding and protein-coding region and part of the 3' untranslated region. Determination of the sequence of the cloned cDNA confirms the complete conservation in human of the amino-acid sequence of skeletal α actin compared to the rabbit or rat proteins. The 5' untranslated region, but not the 3' untranslated region, shows good homology with the corresponding one in the rat gene. Analysis of changes at silent sites within the protein-coding region suggests that the divergence of skeletal and cardiac α actin took place much earlier than the mammalian radiation. The plasmids described here have been used as probes to detect the homologous gene among the about thirty actin sequences present in the human genome.

INTRODUCTION

The actin genes constitute a complex gene family in mammals. At least 6 closely related actin proteins are synthesized in various cell types (1). One α actin is specific for skeletal muscle, another for the cardiac muscle, and two types are found in smooth muscle. The β and γ actins are constituents of the cytoskeleton of non muscle cells. There is evidence that, at least in mouse, different actin isoforms are expressed in fetal and adult skeletal muscle (2). The sequences of the various actins have been extremely well conserved during evolution : for instance there are no changes in amino acid sequence between chicken and mammalian skeletal α actin (1, 3), and DNA probes corresponding to Drosophila, Dictyostelium or Acanthamoeba actins can be used to detect actin sequences in mammals (4, 5, 6). The gene family however, appears even more complex when examined at the genomic DNA level. It has been estimated, by Southern-blot analysis or by cloning of genomic DNA that as much as 20 to 30 actin related sequences exist in the human genome (4, 7), which might be scattered on various chromosomes (8). This complicates the identification and cloning of those actin sequences in

genomic DNA which are expressed in a given cell type. Thus out of several cloned human genomic actin sequences, only the cardiac α actin gene has been identified (5) together with β and γ like actin sequences (6, 9).

We report here the construction of a cDNA library corresponding to polyA$^+$ RNA from a human adult skeletal muscle, and the identification and sequencing of clones corresponding to skeletal α actin. We have established the sequence of the entire 5' non-coding and protein-coding regions and of part of the 3' non-coding region. The cloned sequences were used as probes to identify the genomic fragments which are likely to correspond to the expressed genes. These probes might prove useful for the cloning of this gene and for its localisation on the human chromosome map.

MATERIAL and METHODS

Construction of cDNA libraries.

Two libraries were made by inserting "C-tailed" double stranded cDNA (ds-cDNA) in the plasmid pBR322 "G-tailed" at its PstI site.

a) "Classical" method (10) : PolyA$^+$ RNA, prepared as previously described (11) from quadriceps muscle of a patient undergoing amputation for an osteosarcoma, was copied with AMV reverse transcriptase in the presence of actinomycin D, and the Klenow fragment of DNA polymerase I was used for the synthesis of the second strand. After S1 nuclease treatment, the ds-cDNA was fractionated on a 5-20 % sucrose gradient and fractions containing molecules longer than about 700 bp were pooled and used in the tailing reaction with deoxynucleotidyl transferase in the presence of dCTP. Reaction conditions used throughout were derived from standard procedures (10, 12). 20 ng of tailed ds-cDNA were annealed to pBR322 dG tailed at the PstI site (a generous gift of Dr. J. Bloch). After transformation of E.coli C600 5K, about 15000 colonies were obtained, 90 % of which were ampicillin-sensitive.

b) The technique of Land et al. (13) was followed with very few modifications, using the same polyA$^+$ RNA preparation as above. Single stranded cDNA was tailed with dCMP residues and the second strand synthesis was primed with oligodeoxyguanylate. ds-cDNA was fractionated by sedimentation on a 5-20 % sucrose gradient as above and further treated as in the "classical" method. After annealing with the plasmid vector and transformation, 7000 colonies were obtained for 50 ng of dscDNA.

Characterization of α actin cDNA clones.

Screening was performed on colonies in ordered array transfered to Whatman 541 paper as described (14). Mini preparations (15) of positive

clones were analysed for insert length. Sequencing was performed by the Maxam and Gilbert method (16) using 5' end-labelled fragments subjected to strand separation (17).

Southern blotting.

Human genomic DNA was prepared from white blood cells by a modification of the Gross-Bellard technique (18) (extractions were performed with a phenol/ chloroform/isoamyl alcohol mixture rather than with phenol alone and ethanol precipitations replaced the dialysis steps). DNA was digested with 5 to 10 fold excess of restriction enzyme and completion of digests was monitored with phage λ DNA added to an aliquot of the reaction mixture. Gel electrophoresis, transfer to Diazobenzyloxymethyl (DBM) paper and hybridization to nick translated probes were as described (19).

RESULTS

Construction of muscle cDNA libraries and screening for α actin sequences.

PolyA$^+$ RNA from human adult skeletal muscle was used to establish corresponding cDNA libraries by cloning in the PstI site of pBR322 after G-C tailing. Two different methodologies were followed in parallel. The "classic" method (10), with self priming for second strand synthesis and S1 nuclease treatment, yielded about 15 000 clones. A smaller library (7000 clones) was constructed using the method described by Land et al. (13), where single stranded cDNA is dC tailed and second strand synthesis is primed with oligodeoxyguanylate, thus avoiding the S1 nuclease treatment, which results in loss of 5' sequences and often in sequence rearrangements (2, 20, 21).

The PstI insert from a mouse skeletal α actin cDNA clone (p91, ref. 22) was used as a probe to screen ordered arrays of bacterial clones. In the first library, about 2 % of the colonies gave a positive signal and 12 clones were further analyzed for insert lengths. The two longest clones (αact 1 and αact 2, fig. 1) had inserts of 1200 and 1100 bp respectively. While αact 1 could be excised with PstI, αact 2 had only one PstI site remaining. Both inserts had single restriction sites for BglII, PvuII and XbaI (Fig. 1 line a). The mouse α actin cloned sequence contains also a PvuII site separating the insert into 5' and 3' fragments of similar length (520 and 630 bp), and we used the separated fragments as probes to analyse the orientation of the two human actin plasmids. From the results (not shown), we could conclude that αact 1 is more extended in the 5' direction (with respect to the mRNA sequence) and αact 2 is more extended in the 3' direc-

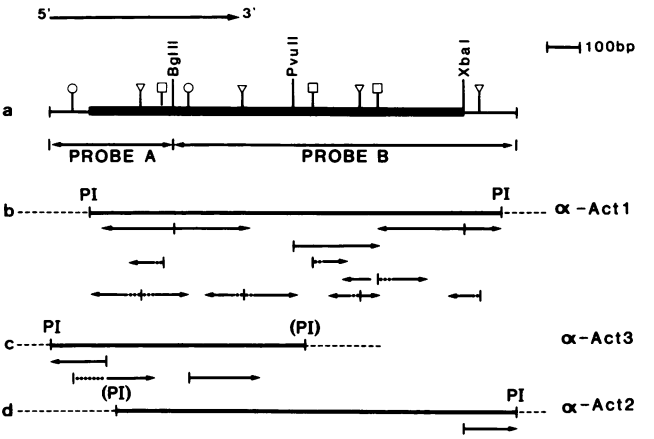


Fig. 1 : Map of skeletal α actin cDNA. a) Combined map of α actin cDNA as derived from sequence analysis of clones α Act 1, 2 and 3. The heavy line corresponds to the protein coding sequence. Sites for restriction enzymes used in mapping or sequencing are indicated (▽ HinfI, □ TaqI, ○ AvaI), as well as probes A and B used in the genomic blotting experiments. b, c and d) location of inserts in the plasmids α act 1, 3 and 2 with respect to the cDNA map. pBR322 sequences are represented by the broken lines and PI corresponds to PstI sites. PstI sites which have not been reconstituted are shown in parenthesis. Sequencing strategy is summarized below the map of each insert. The restriction sites used for 5'-end labeling are indicated by vertical bars and the direction and extent of sequencing are shown by the arrows. Sequences which have not been determined in a particular experiment are indicated by the dotted lines.

tion (Fig. 1, lines b and d). We then screened the second cDNA library using the 5' half of the mouse α actin insert as a probe and tested 10 clones for total insert length and for the length of the 5' PstI-BglII fragment. Four independent clones (αact 3 to 6) had PstI-BglII 5' fragments of identical length (370 bp, while corresponding fragments in αAct 1 and 2 are 260 and 180 bp long respectively), suggesting that they contain the complete 5'mRNA sequence. The inserts in αact 3 to 6 are however smaller than in αact 1 and 2, an indication of extensive sequence loss on the 3' side. Location of TaqI, RsaI, PvuII and BglII sites were identical in comparable regions of all 6 clones (αact 1 to 6, results not shown) suggesting that they indeed correspond to the same (or to very closely related) RNA species. The plasmids αact 1, 2 and 3 were further used for DNA sequencing.

Sequence of skeletal α actin cDNA.

Most of the sequence presented in Fig. 2 is derived from the αact 1 plasmid (from residues 99 to 1348). However the 5' end of the sequence was

```
                                                                  1
                                                          (G)n-ACCGCAGCGGACAGCGCCAAGTG
                                                                                  100
AAGCCTCGCTTCCCTCCCGCGGCGACCAGGGCCCGAGCCGAGAGTAGCAGTTGTAGCTACCCGCCCAGAAACTAGACACA

ATG TGC GAC GAA GAC GAG ACC ACC GCC CTC GTG TGC GAC AAT GGC TCC GGC CTG GTG AAA
Met Cys Asp Glu Asp Glu Thr Thr Ala Leu Val Cys Asp Asn Gly Ser Gly Leu Val Lys
                                                  200
GCC GGC TTC GCC GGG GAT GAC GCC CCT AGG GCC GTG TTC CCG TCC ATC GTG GGC CGC CCC
Ala Gly Phe Ala Gly Asp Asp Ala Pro Arg Ala Val Phe Pro Ser Ile Val Gly Arg Pro

CGA CAC CAG GGC GTC ATG GTC GGT ATG GGT CAG AAA GAT TCC TAC GTG GGC GAC GAG GCT
Arg His Gln Gly Val Met Val Gly Met Gly Gln Lys Asp Ser Tyr Val Gly Asp Glu Ala
                  300
CAG AGC AAG AGA GGT ATC CTG ACC CTG AAG TAC CCT ATC GAG CAC GGC ATC ATC ACC AAC
Gln Ser Lys Arg Gly Ile Leu Thr Leu Lys Tyr Pro Ile Glu His Gly Ile Ile Thr Asn
                                                                      400
TGG GAT GAC ATG GAG AAG ATC TGG CAC CAC ACC TTC TAC AAC GAG CTT CGC GTG GCT CCC
Trp Asp Asp Met Glu Lys Ile Trp His His Thr Phe Tyr Asn Glu Leu Arg Val Ala Pro

GAG GAG CAC CCC ACC CTG CTC ACC GAA GCC CCC CTC AAT CCC AAG GCC AAC CGC GAG AAG
Glu Glu His Pro Thr Leu Leu Thr Glu Ala Pro Leu Asn Pro Lys Ala Asn Arg Glu Lys
                                          500
ATG ACC CAG ATC ATG TTT GAG ACC TTC AAC GTG CCC GCC ATG TAC GTG GCC ATC CAG GCC
Met Thr Gln Ile Met Phe Glu Thr Phe Asn Val Pro Ala Met Tyr Val Ala Ile Gln Ala

GTG CTG TCC CTC TAC GCC TCC GGC AGG ACC ACC GGC ATC GTG CTG GAC TCC GGC GAC GGC
Val Leu Ser Leu Tyr Ala Ser Gly Arg Thr Thr Gly Ile Val Leu Asp Ser Gly Asp Gly
                  600
GTC ACC CAC AAC GTG CCC ATT TAT GAG GGC TAC GCG CTG CCG CAC GCC ATC ATG CGC CTG
Val Thr His Asn Val Pro Ile Tyr Glu Gly Tyr Ala Leu Pro His Ala Ile Met Arg Leu
                                                                      700
GAC CTG GCG GGC CGC GAT CTT ACC GAC TAC CTG ATG AAG ATC CTC ACT GAG CGT GGC TAC
Asp Leu Ala Gly Arg Asp Leu Thr Asp Tyr Leu Met Lys Ile Leu Thr Glu Arg Gly Tyr

TCC TTC GTG ACC ACA GCT GAG CGC GAG ATC GTG CGC GAC ATC AAG GAG AAG CTG TGC TAC
Ser Phe Val Thr Thr Ala Glu Arg Glu Ile Val Arg Asp Ile Lys Glu Lys Leu Cys Tyr
                                          800
GTG GCC CTG GAC TTC GAG AAC GAG ATG GCG ACG GCC GCC TCC TCC TCC TCC CTG GAA AAG
Val Ala Leu Asp Phe Glu Asn Glu Met Ala Thr Ala Ala Ser Ser Ser Ser Leu Glu Lys

AGC TAC GAG CTG CCA GAC GGG CAG GTC ATC ACC ATC GGC AAC GAG CGC TTC CGC TGC CCG
Ser Tyr Glu Leu Pro Asp Gly Gln Val Ile Thr Ile Gly Asn Glu Arg Phe Arg Cys Pro
                  900
GAG ACG CTC TTC CAG CCC TCC TTC ATC GGT ATG GAG TCG GCG GGC ATT CAC GAG ACC ACC
Glu Thr Leu Phe Gln Pro Ser Phe Ile Gly Met Glu Ser Ala Gly Ile His Glu Thr Thr
                                                                     1000
TAC AAC AGC ATC ATG AAG TGT GAC ATC GAC ATC AGG AAG GAC CTG TAT GCC AAC AAC GTC
Tyr Asn Ser Ile Met Lys Cys Asp Ile Asp Ile Arg Lys Asp Leu Tyr Ala Asn Asn Val

ATG TCG GGG GGC ACC ACG ATG TAC CCT GGG ATC GCT GAC CGC ATG CAG AAA GAG ATC ACC
Met Ser Gly Gly Thr Thr Met Tyr Pro Gly Ile Ala Asp Arg Met Gln Lys Glu Ile Thr
                                          1100
GCG CTG GCA CCC AGC ACC ATG AAG ATC AAG ATC ATC GCC CCG CCG GAG CGC AAA TAC TCG
Ala Leu Ala Pro Ser Thr Met Lys Ile Lys Ile Ile Ala Pro Pro Glu Arg Lys Tyr Ser

GTG TGG ATC GGC GGC TCC ATC CTG GCC TCG CTG TCC ACC TTC CAG CAG ATG TGG ATC ACC
Val Trp Ile Gly Gly Ser Ile Leu Ala Ser Leu Ser Thr Phe Gln Gln Met Trp Ile Thr
                  1200
AAG CAG GAG TAC GAC GAG GCC GGC CCT TCC ATC GTC CAC CGC AAA TGC TTC TAG ACACACT
Lys Gln Glu Tyr Asp Glu Ala Gly Pro Ser Ile Val His Arg Lys Cys Phe ***
                                                                    1300
CCACCTCCAGCACGCGACTTCTCAGGACGACGAATCTTCTCAATGGGGGGGCGGCTGAGCTCCAGCCACCCCGCAGTCAC

TTTCTTTGTAACAACTTCCGTTGCTGCCATCGTAAACTGACACAGTGTTT-(C)n
```

Fig. 2 : Sequence of α actin cDNA. The sequence presented combines those obtained with plasmids α act1, 2 and 3. (G_n) and $(C)_n$ correspond to homopolymeric tails in α act 3 and α act 2 respectively.

derived from αact 3 and the 3' end of the sequence was derived from the αact 2 plasmid (see Fig. 1). There are no base changes between each pair of plasmids, in the overlapping sequences, which are 370 bp and 110 bp long for the 5' and 3' regions respectively. It is thus justified to combine the sequences derived from the three plasmids into a single one.

From the sequence presented in Fig. 2, an amino acid sequence corresponding exactly to the known rabbit or rat α actin sequences (1) can be derived. The initiation codon is located at position 104 (in the mature protein the first two residues, methionine and cysteine have been cleaved). Thus there is a rather long 5' non-coding sequence (103 nucleotides, devoid of any other ATG triplet). The 3' non-coding sequence, following the termination codon TAG at position 1235) is certainly not complete since it lacks the polyadenylation signal (usually AATAAA) which should precede the polyA tracks by about thirty nucleotides. Since the length of skeletal α actin mRNA in mammals is about 1650 nucleotides, including the polyA tail, and by comparison with the length of the 3' non coding regions in rat skeletal α actin and in human cardiac α actin mRNAs (23, 5) (240 to 250 Nucleotides), we estimate that about 120 bp are lacking in our cDNA sequence.

The observation that four clones (out of 10 examined from the second cDNA library) have apparently identical limits on their 5' side, suggested that they are indeed complete in this region. We have checked this point further in a primer extension experiment. A 45 nucleotide long single stranded HhaI-AvaI fragment derived from the 5' region of the αact 3 insert, 5'end-labelled at the AvaI site (Fig.1), was used as a primer in a reverse transcription reaction with muscle polyA$^+$ RNA as template. The product of the reaction was analysed on a denaturing urea/polyacrylamide gel, in parallel with a sequence ladder, allowing the exact determination of lengths (Fig. 3). The major doublet bands correspond to an elongation of 14-15 nucleotides upstream from the 3' HhaI end of the primer (position 15 in Fig. 2) while a minor doublet is longer by 3 to 4 nucleotides. This result confirms that the sequence in Fig. 2 is complete on the 5' side, and suggests that a minor cap site might exist a few nucleotides upstream the major one.

Analysis of genomic DNA.

Human DNA preparations from several individuals were cut with HindIII, EcoRI or PvuII and analysed by Southern blotting, using two contiguous actin fragments as hybridization probes. Probe A corresponds to the PstI-BglII fragment of clone αact 3 and contains residue 1 to 360 of the cDNA-sequence, thus including all the 5' non coding sequence; Probe B contains the rest of

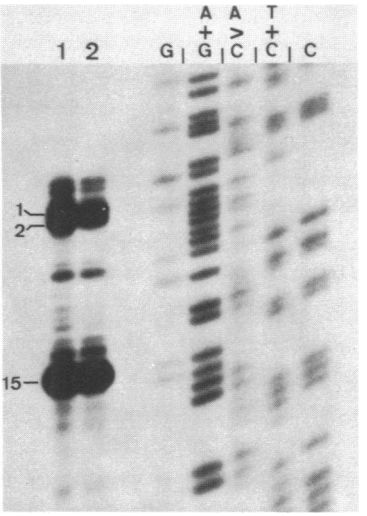
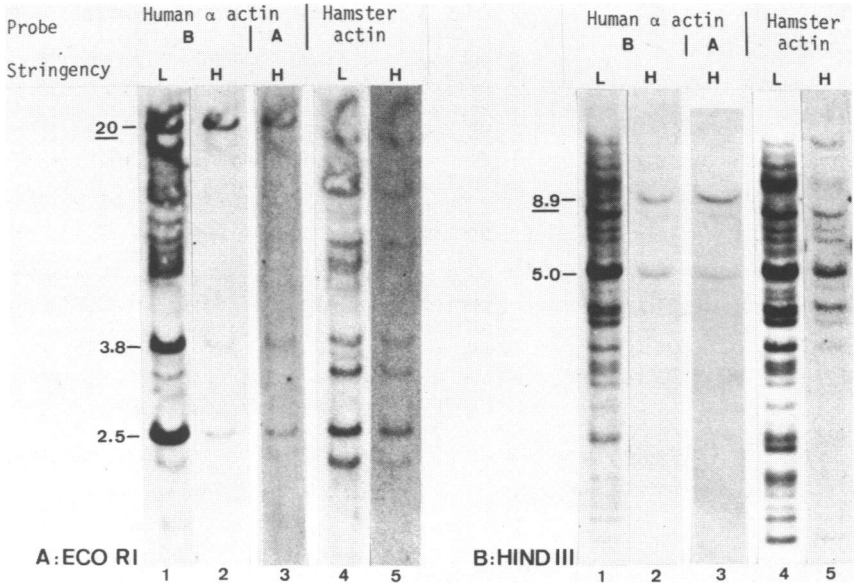


Fig. 3 : Analysis by primer elongation of the 5' end of α actin mRNA. 1 pmole of a 45 nucleotide long single stranded fragment 5' labelled at an AvaI site (position 59 in Fig.2) and extending to the HhaI site at position 15, was used as primer for reverse transcription of 2 µg of muscle polyA$^+$ RNA, with 12 and 60 units of reverse transcriptase (lanes 1 and 2 respectively), under conditions described previously (17). The products of the reactions were analysed on a 8 % polyacrylamide-urea gel, together with sequencing reactions performed on the AvaI-BglII fragment labeled at the AvaI site (position 56 with respect to the sequence in Fig. 2). The positions of the unelongated primer (15) and of the major product of elongation (2, 1) are indicated.

the cDNA sequence (fragment BglII-Pst of αact 2, see fig. 1). The DNA blots were hybridized to either probes in 40 % formamide, 0.9 M NaCl, at 42°C i.e. conditions of relatively low stringency. Filters were then washed either at low stringency (60°C 2 x SSC - 0,1 % SDS) or at high stringency (68°C, 0.2 x SSC - 0,1 % SDS) before autoradiography. A complex pattern consisting of more than 25 bands was obtained with all 3 restriction enzymes when using the low stringency washing conditions (Fig. 4 A1, B1, C1). After the high stringency wash, only two to three bands remained in each case. For HindIII and EcoRI, the same bands appeared with both probes A and B ie, a 8.9 kb and a 5.0 kb band for HindIII, a 20 kb, 3.8 and 2.5 kb bands for EcoRI (Fig. 4A, B). The two HindIII bands and the two smaller EcoRI bands had been detected previously as the most stably hybridizing bands with a heterologous mouse α actin probe while the 20 kb EcoRI fragment was apparently not seen after a high stringency wash (7). With PvuII, two bands are seen in common with probes A and B : a major 2,7 kb band and a faint 7.4 kb band. With probe B an additional strong band is seen at 7.9 kb (Fig. 4C). Since probe B contains an internal PvuII site, the 7.9 kb band seen only with this probe thus corresponds to the actin sequence located in 3' of this site, the 2.7 kb band common to probe A and B should correspond to the sequence present on the 5' side of the PvuII site.

The fact that the same two HindIII or three EcoRI bands are revealed, after high stringency washing, with the two contiguous probes poses a problem of identification of those fragments which correspond to the expres-



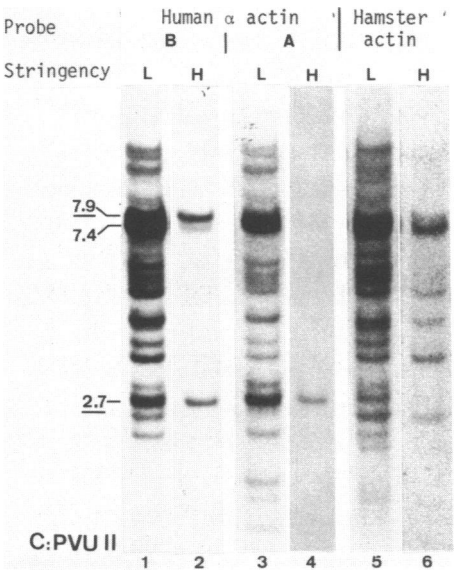


Fig. 4 : Identification of genomic fragments homologous to human α actin cDNA probes. Human DNA digested with EcoRI (panel A) HindIII (panel B) or PvuII (panel C) was electrophoresed on a 0.9 % agarose gel and blotted onto DBM paper. The blots were successively hybridized to nick translated α actin probes A and B (see Fig. 1) and to an hamster cytoskeletal actin probe (PstI fragment of pAct1, see (24)). Hybridization was performed for 18 h at 42°C as described (19) but with 40 % formamide and 5 % dextrane sulfate. Filters were then washed either under low (L) stringency conditions (2 SSC, 0.1 % SDS, 60°C) or under high (H) stringency conditions (0.2 SSC, 0.1 % SDS, 68°C) and exposed to an X ray film in the presence of an intensifying screen. The probes used and the washing conditions are indicated above each lane. The size of fragments discussed in the text are given (in kb) on the left of each panel, and are underlined for those which should correspond to the skeletal α actin gene.

sed skeletal α actin gene. The 20 kb EcoRI fragment, and the HindIII 8.9 kb fragments which appear as relatively minor band at low stringency, show very little decrease in signal intensity after the high stringency wash (their length is underlined in fig. 4 A and B). On the contrary, the 3.8 and 2.5 kb EcoRI bands and the 4.9 kb HindIII band, which appear under low stringency conditions as the major ones, elicit a drastic loss in intensity after washing in low salt at 68°C. Furthermore, the latter bands are also prominent with a heterologous hamster cDNA probe (24) corresponding to a cytoplasmic β or γ actin (Fig. 4 A and B lanes 4 and 5). Thus it is very likely that the fragments showing the more stable hybridization with the α actin probe, are the one which correspond to the skeletal α actin gene. The fact that bands which are not completely homologous to the α actin probe can appear as the major ones under low stringency conditions might be due to a combination of length differences (the 20 kb EcoRI fragment will not transfer as efficiently as the 2.5 and 3.7 kb ones), possible occurence of multiple fragments migrating at the same place, and perhaps also to differences in hybridization kinetics due to presence or absence of introns. However the fact that the same major bands appear under low stringency conditions with drosophila, hamster or human probes (ref. 4 and Fig. 4) suggests that they might correspond to an amplification of a single genomic fragment containing an actin sequence.

DISCUSSION

The isolation of cDNA clones corresponding to human skeletal α actin has allowed us to determine the sequence of the whole 5' non-coding and protein coding regions, and of part of the 3' untranslated region. The use of the cDNA cloning technique proposed by Land et al. (13) proved useful in yielding complete sequences on the 5' side. As expected from the known total conservation of the primary structure of skeletal muscle α actin, the amino-acid sequence which can be derived shows no changes with respect to other mammalians α actin sequenced at the nucleotide or amino acid level (1, 23). It is however possible to compare nucleotide sequences corresponding to mammalian α skeletal actin mRNAs in the regions or positions not involved in amino-acid coding. ie : 5' or 3' untranslated regions, and silent sites within the protein coding region.

The human α actin cDNA sequence determined here is characterized by a high G+C content (61.6 % in the coding region, 58.8 % for the combined 5' and 3' untranslated regions) compared to the 40 % G+C content of the human

TABLE I : Correlation between the G+C/A+T ratio at the third codon position and the bias against CpG, in cDNA sequences.

cDNA	$\frac{G+C}{A+T}$ ratio at the third codon position	CpG/GpC ratio in protein-coding regions
Human α actin	8.1	.98
Rat α actin	3.4	.78
Human cardiac α actin	2.0	.51
Human proopiomelanocortin	9.0	.79
Human α globin	7.9	.80
Human growth hormone	3.3	.45
Human β globin	1.9	.14

The values have been derived from sequences present in the EMBL Nucleic Acid Sequence database, or described in Ref. 23 and 5 for rat α actin and human cardiac α actin.

genome. In the coding region this is due to an extreme assymetry in the base choices for the third codon position : 89 % for G + C and only 11 % for A+T. A similar bias is found, to a lesser extent, in rat skeletal α actin mRNA (22.5 % A + U at the 3rd position) and in human cardiac α actin mRNA (33.6 % A + U). Another characteristic of this sequence is that, contrary to most sequences in vertebrate genomes, there is no bias against the CpG dinucleotide and the cloned sequences have a high frequency of restriction sites for "CpG" restriction enzymes, like HhaI, MspI, FnuDII or TaqI. In an analysis of α actin mRNA sequences and of a few other human mRNAs, we have found that a large excess of G + C at the third codon position is apparently correlated with an absence of bias against CpG dinucleotides (Table I). The pattern for human α skeletal actin resembles that found for α globin or proopiomelanocortin mRNAs, and is clearly different from that of β globin, growth hormone or cardiac α actin mRNAs. Since the gene specific differences in GC content and in the extent of the CpG bias are also often present in non-coding regions, including introns [as shown for instance in the human α and β globin genes (25, 26)] this suggests that the apparently stringent codon usage found in an mRNA such as the human α actin one, depends less on selective pressure related to protein synthesis mechanisms, but might rather reflect the presence of a gene in genomic regions which differ in their base composition over long distances [may be G or R bands (8, 27)]. It has been proposed recently that about 2% of the genome of vertebrates correspond to sequences which have a very high CpG content and are found unmethylated in all tissues (28). It will be interesting to know whether human α actin or α globin genes are found in this genome fraction.

The 5'untranslated region of the human α skeletal actin mRNA is slight-

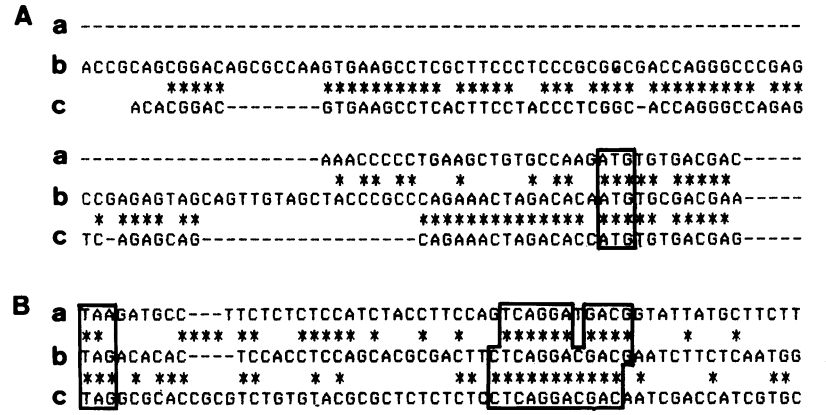


Fig. 5 : Sequence conservation between human and rat α actins in the 5' and 3' untranslated regions. Line a : human α cardiac actin (5), line b : human α skeletal actin, line c : rat α skeletal actin (23). The sequence of the 5' untranslated region for human α cardiac actin is unknown but for a few nucleotides proximal to the initiation codon. A) 5' untranslated region : The initiation codons are boxed. B) 3' untranslated region. The termination codons and a homology region are boxed.

ly longer than that of the rat homologous mRNA (103 versus 72 nucleotides) but regions of very significant homology can be found throughout (Fig. 5A). This suggests some functional importance, especially since in contrast the 3' untranslated (UT) sequences of the human and rat genes diverge right after the termination codon. Although very little homology is found in the 3' UT regions which can be compared, one common feature (shared also by the human α cardiac sequence) is constituted by a well conserved 12 nucleotides sequence located about 30 bp downstream from the termination codon (Fig.5B).

When comparing the amino acid coding sequences, one can distinguish between replacement sites (where a base change leads to an amino acid change) and the silent sites. Because of the conservation of the amino acid sequence, the replacement sites have to remain unchanged in skeletal α actin genes. Thus, in this case, only the analysis of changes at silent sites can give an idea of the divergence time for a pair of genes. When the rat and human actin protein coding sequences are compared, one finds 61 % silent changes (after correction for multiple events) (29). This is similar to the data obtained for other rat/human gene comparisons using the same calculation : insulin, 64 %, Growth Hormone, 71 %, Prolactin, 76 %) (29, 30). The 70 to 85 % silent site changes seen between the chicken α actin and the rat

TABLE II : Divergence at silent sites in α actin coding sequences.

α Actin Coding Sequences Compared	Silent Changes (in %)
Human skeletal / Rat skeletal	61
Human skeletal / Chicken skeletal	86
Human skeletal / Human cardiac	112
Rat skeletal / Chicken skeletal	70
Rat skeletal / Human cardiac	104

Silent changes (corrected for multiple events) were calculated as proposed by Perler et al. (29) using a computer program. Sequence data were from ref. 23 and 31 for the rat and chicken skeletal actin, and ref. 5 for the human cardiac actin. Divergence time is about 250 million years for the separation of birds and mammals, and about 80 million years for the mammalian radiation (see 29).

or human corresponding genes (Table II) support previous observations (29) that accumulation of these changes start to saturate after long times of divergence (>10^8 years). In contrast, the changes at silent site are highest between human cardiac α actin and either rat or human skeletal actin (>100%, see Table II). This analysis suggests that the divergence of cardiac and skeletal α actin genes occured much earlier than the mammalian radiation.

The cloned α actin cDNA sequence has allowed us to identify, in the human genome, restriction fragments most likely associated with the expressed gene. Since even under stringent conditions and using a small probe (360 bp) one always detects two fragments this suggests that there is one supplementary sequence in the human genome which either shares strong homology to the skeletal α actin mRNA, or more likely is present in multiple copies. Because of the complexity of the actin gene family, it might be difficult to identify among the many genomic clones which contain actin-like sequences, the ones which correspond to a specific expressed gene. Homologous cDNA probes might be useful for this task and should also allow to locate the gene on the human chromosome map.

ACKNOWLEDGEMENTS.

We wish to thank Drs. M. Buckingham, A. Minty and P. Soriano for gift of actin cDNA clones, Drs. M. Bellard, J. Bloch and G. Dretzen for useful advice, Dr. T. Reudelhuber for reading the manuscript and Dr. P. Chambon for constant support. We are grateful to Ms. C. Kloepfer for excellent assistance and to E. Badzinski and B. Boulay for preparing the manuscript. This work was supported by the INSERM (CRL 826032 and PRC 134029), the CNRS (ATP

006520/50) and the Fondation Simone et Cino del Duca.

*To whom all reprint request should be addressed.

REFERENCES

1. Vandekerckhove, J. and Weber, K. (1979) Differentiation 14, 123-133.
2. Minty, A.J., Alonso S., Caravatti, M. and Buckingham, M.E. (1982) Cell 30, 185-192.
3. Vanderkerckhove, J. and Weber, K. (1979) FEBS Lett. 102, 219-222.
4. Engel, J.N., Gunning, P.W. and Kedes, L. (1981) Proc. Natl. Acad. Sci. USA 78, 4674-4678.
5. Hamada, H., Petrino, M.G. and Kakunaga, T. (1982) Proc. Natl. Acad. Sci. USA 79, 5901-5905.
6. Moors, M. and Gallwitz, D. (1982) Nucleic Acids Res. 10, 7843-7849.
7. Humphries, S.E., Whittall, R., Minty, A., Buckingham, M. and Williamson, R. (1981) Nucleic Acids Res. 9, 4895-4908.
8. Soriano, P., Szabo, P. and Bernardi, G. (1982) EMBO Journal 1, 579-583.
9. Engel, J., Gunning, P. and Kedes, L. (1982) Mol. Cell. Biol. 2, 674-684.
10. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
11. Kahn, A., Cottreau, D., Daegelen, D. and Dreyfus, J.C. (1981) Eur. J. Biochem. 116, 7-12.
12. Cochet, M., Perrin, F., Gannon, F., Krust, A., Chambon, P., Mc Knight, G.S., Lee, D.C., Mayo, K.E. and Palmiter, R. (1979) Nucleic Acids Res. 6, 2435-2452.
13. Land, H., Grez, M., Hauser, H., Lindenmaier, W. and Schütz, G. (1981) Nucleic Acids Res. 9, 2251-2266.
14. Masiakowski, P., Breathnach, R., Bloch, J., Gannon, F., Krust, A. and Chambon, P. (1982) Nucleic Acids Res. 10, 7895-7903.
15. Ish-Horowicz, D. and Burke, J.F. (1981) Nucleic Acids Res. 9, 2989-2998.
16. Maxam, A., and Gilbert, W. (1980) in "Methods in Enzymology" (L. Grossman and K. Moldave, Eds.) Vol. 65, Academic Press, New York, pp. 499-560.
17. Heilig, R., Muraskovsky, R., Kloepfer, C. and Mandel, J.L. (1982) Nucleic Acids Res. 10, 4363-4382.
18. Gross-Bellard, M., Oudet, P. and Chambon, P. (1973) Eur. J. Biochem. 36, 32-38.
19. Bellard, M., Kuo, M.T., Dretzen, G. and Chambon, P. (1980) Nucleic Acids Res. 8, 2737-2751.
20. O'Hare, K., Breathnach, R., Benoist, C. and Chambon, P. (1979) Nucleic Acids Res. 7, 321-334.
21. Volckaert, G., Tavernier, J., Derynck, R., Devos, R. and Fiers, W. (1981) Gene 15, 215-233.
22. Minty, A.J., Caravatti, M., Robert, B., Cohen, A., Daubas, P., Weydert, A., Gros, F. and Buckingham, M.E. (1981) J. Biol. Chem. 256, 1008-1014.
23. Zakut, R., Shani, M., Givol, D., Neuman, S., Yaffe, D. and Nudel, U. (1982) Nature 298, 857-859.
24. Dodemont, H.J., Soriano, P., Quax, W.J., Ramaekers, F., Lenstra, J.A., Groenen, M.A.M., Bernardi, G. and Bloemendal, H. (1982) EMBO Journal 1, 167-171.
25. Liebhaber, S.A., Goossens, M.J. and Kan, Y.W. (1980) Proc. Natl. Acad. Sci. USA 77, 7054-7058.
26. Lawn, R.M., Efstratiadis, A., O'Connell, C. and Maniatis, T. (1980) Cell 21, 647-651.

27. Cuny, G., Soriano, P., Macaya, G. and Bernardi, G. (1981) Eur. J. Biochem. 115, 227-233.
28. Cooper, D.N., Taggart, M.H. and Bird, A.P. (1983) Nucleic Acids Res. 11 647-658.
29. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) Cell 20, 555-566.
30. Cooke, N.E., Coit, D., Shine, J., Baxter, J.D., Martial, J.A. (1981) J. Biol. Chem. 256, 4007-4016.
31. Fornwald, J.A., Kuncio, G., Peng, I. and Ordahl, C.P. (1982) Nucleic Acids Res. 10, 3861-3876.