MOLECULAR AND CELLULAR BIOLOGY, May 1983, p. 787–795
0270-7306/83/050787-09$02.00/0

Vol. 3, No. 5

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

PETER GUNNING,[1] PHYLLIS PONTE,[1] HIROTO OKAYAMA,[2] JOANNE ENGEL,[1] HELEN BLAU,[3] AND LARRY KEDES[1]*

Departments of Medicine,[1] Biochemistry,[2] and Pharmacology,[3] Stanford University School of Medicine and Veterans Administration Medical Center, Palo Alto, California 94304

Received 29 November 1982/Accepted 24 January 1983

cDNA clones encoding three classes of human actins have been isolated and characterized. The first two classes (γ and β, cytoplasmic actins) were obtained from a cDNA library constructed from simian virus 40-transformed human fibroblast mRNA, and the third class (α, muscle actin) was obtained from a cDNA library constructed from adult human muscle mRNA. A new approach was developed to enrich for full-length cDNAs. The human fibroblast cDNA plasmid library was linearized with restriction enzymes that did not cut the inserts of interest; it was then size-fractionated on gels, and the chimeric molecules of optimal length were selected for retransformation of bacteria. When the resulting clones were screened for actin-coding sequences it was found that some full-length cDNAs were enriched as much as 50- to 100-fold relative to the original frequency of full-length clones in the total library. Two types of clones were distinguished. One of these clones encodes γ actin and contains 100 base pairs of 5′ untranslated region, the entire protein coding region, and the 3′ untranslated region. The second class encodes β actin, and the longest such clone contains 45 base pairs of 5′ untranslated region plus the remainder of the mRNA extending to the polyadenylic acid tail. A third class, obtained from the human muscle cDNA library, encodes α actin and contains 100 base pairs of 5′ untranslated region, the entire coding region, and the 3′ untranslated region. Analysis of the DNA sequences of the 5′ end of the clones demonstrated that although β- and γ-actin genes start with a methionine codon (MET-Asp-Asp-Asp and MET-Glu-Glu-Glu, respectively), the α-actin gene starts with a methionine codon followed by a cysteine codon (MET-*CYS*-Asp-Glu-Asp-Glu). Since no known actin proteins start with a cysteine, it is likely that post-translational removal of cysteine in addition to methionine accompanies α-actin synthesis but not β- and γ-actin synthesis. This observation has interesting implications both for actin function and actin gene regulation and evolution.

Actin is a highly conserved protein which participates in a wide variety of cellular functions in eucaryotes including muscle contraction, ameboid movement, cytokinesis, and mitotic division (9, 20). Amino acid sequencing data have demonstrated that the range of actin isotypes now observed (27) evolved from two major classes of actin, "cytoplasmic" and "muscle." All organisms thus far examined express a cytoplasmic actin form which is usually utilized to construct the cellular microfilaments (27, 25). In the case of simple eucaryotes like *Dictyostelium*, *Physarum*, and *Saccharomyces cerevisiae*, only one type of cytoplasmic actin is expressed (5, 8, 16, 28). In mammals, there are two cytoplasmic actins (β and γ) (26), and in amphibians there are at least three (25). The second class of actin proteins, the muscle (or "α-like") actins, are found in birds and mammals. It follows that the original α-like actin gene evolved from the cytoplasmic type at some time before the divergence of birds and mammals. However, the precise time at which distinct α actin genes and proteins arose is still unclear. In mammals, four different tissue-specific muscle-actin isotypes have been found: skeletal actin (or α), cardiac actin, and two smooth muscle types, aortic actin and stomach actin (27).

All of the actin proteins which have been sequenced have an acidic amino acid at the

amino terminus (25). Gene and cDNA sequence data have shown that an initiation methionine codon directly precedes the acidic amino acid codon in the case of the lower eucaryotes *Saccharomyces cerevisiae* (8, 16) and *Dictyostelium* (5). However, in all published reports to date regarding protostome and deuterostome actin genes this methionine codon is always followed by a cysteine codon. This is true of all six actin genes in *Drosophila melanogaster* (7). It is also of interest that although all these *Drosophila* genes encode cytoplasmic-like actin proteins, some are expressed in muscle. This observation raises two important points. First, it demonstrates that the role of actin in muscle function is not restricted to the α-like isoform; cytoplasmic-like actins can perform the same role. Second, it suggests that the α-like isotype arose after the divergence of deuterostomes and protostomes.

Recently, sequence data of the chicken and rat α-actin genes and the human cardiac actin gene have also shown that the amino-terminal acidic residue is preceded in the primary translation product by a cysteine residue (6, 11, 29). Furthermore, two genes encoding cytoplasmic-like actins in the sea urchin have recently been shown similarly to have a cysteine codon at their amino termini (2) (W. Crain, personal communication). Thus, since all protostome and deuterostome actin cDNAs or genes sequenced to date encode an amino-terminal cysteine, yet none of the sequenced actin proteins has amino-terminal cysteines, one must conclude that cysteine is removed at some point after translation. These discoveries suggest that the transient presence of an amino-terminal cysteine is a general property of all protostome and deuterostome actins.

We have isolated essentially full-length human cDNA clones encoding α, β, and γ actin. Surprisingly, sequence analysis of these clones demonstrates that whereas human sarcomeric actins have an amino-terminal cysteine residue, the cytoplasmic β and γ actins do not. This finding has interesting implications for both actin gene evolution and for post-translational processing of actin.

MATERIALS AND METHODS

RNA isolation. RNA was prepared from adult human skeletal muscle as described previously (3). The RNA was further purified by digestion with proteinase K (Beckman Instruments, Inc., Fullerton, Calif.) followed by phenol-chloroform (1:1) extraction (19). Polyadenylated mRNA was isolated from this RNA preparation by the method of Aviv and Leder (1).

Construction of cDNA libraries. The construction of the simian virus 40-transformed human fibroblast cDNA library has been described previously (18) and follows the method of Okayama and Berg (17). The human muscle cDNA library was constructed exactly as described by Okayama and Berg (17), with one exception. In our hands, *Escherichia coli* DNA ligase has proved to be unreliable, and we therefore substituted T4 DNA ligase (New England Biolabs, Inc., Beverly, Mass.), 1 mM dithiothreitol, and 1 mM ATP for *E. coli* DNA ligase, 0.1 mM β NAD, 10 mM $(NH_4)_2SO_4$, and 50 μg of bovine serum albumin per ml in the ligation reaction (17). The recombinants constructed from 2 μg of muscle polyadenylated mRNA generated 5,400 transformants. Duplicate master plates of the muscle library were harvested, and plasmid DNA was prepared (3).

Screening cDNA libraries. About 3,000 fibroblast library colonies and the entire muscle library were screened by the Grunstein and Hogness method (10). Actin-coding clones were identified by hybridization to a nick-translated chick β-actin cDNA insert as described previously (4).

General methods. Plasmid DNA preparation, restriction enzyme digests, agarose gel electrophoresis, isolation of DNA fragments from gels, Southern blotting, and hybridizations were exactly as described previously (4). Mini-lysates were prepared by the method of Holmes and Quigley (12). Nick-translations followed that of Rigby et al. (21).

Isolation of full-length β- and γ-actin cDNAs. The fibroblast library in the form of plasmid DNA was digested either with *Hin*dIII or with *Sal*I. Both digests were run on a 0.6% agarose gel with size standards. DNA was visualized with ethidium bromide, and regions of the gel corresponding to full-length actin cDNAs were excised. DNA was isolated from the gel slices and religated by the ligation procedure described above, with the DNA concentration below 1 μg/ml to favor intra- over intermolecular ligation. Bacterial transformants obtained by transforming *E. coli* HB101 with the ligation mixture were screened for actin-coding sequences as described above.

DNA sequence analysis. The 5′ ends of the cDNA clones were sequenced as described by Maxam and Gilbert (14).

RESULTS

Isolation of fibroblast actin clones. Approximately 3,000 clones in the human fibroblast library were screened for actin-coding sequences, and 8 clones, designated pHF1 through pHF8, were isolated. To identify which clones were derived from identical mRNAs, we took advantage of an intrinsic property of all clones obtained by the Okayama-Berg method (18). Since the clones are constructed starting at the polyadenylic acid tail and extending in the 5′ direction, all clones derived from the same mRNA should be identical at their 3′ ends (except for small differences in their polyadenylic acid length) and vary only in the length of primer extension by reverse transcription. By using the partial restriction digest method (23), we were able quickly to obtain a detailed restriction fragment comparison of these eight clones.

The restriction fragment digests of five of the eight clones were identical, although varying in their overall lengths, and the restriction map of the longest of these, pHF5, is shown in Fig. 1.

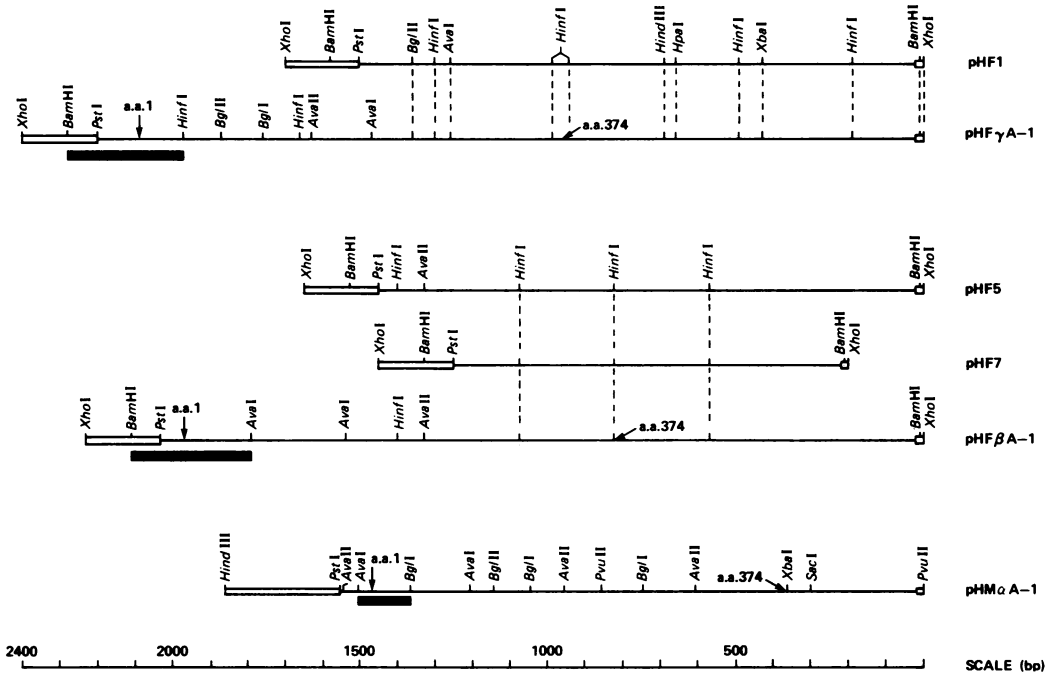

FIG. 1. Restriction maps of actin-encoding human cDNA clone inserts. The open box on the left of each map indicates the linker DNA fragment used in the cloning. The open boxes on the right represent the fragment of primer DNA abutting the polyadenylic acid tail of the mRNA sequence. The bars indicate the segments used for sequencing the 5′ ends of pHFβA-1, pHFγA-1, and pHMαA-1.

Two other clones displayed common restriction sites distinguishable from the pHF5 type, and the restriction map of the longest of these, pHF1, is also shown in Fig. 1. The remaining clone, pHF7, was identical to the pHF5 class with the exception of a 200-base pair deletion at its 3′ end (Fig. 1). Sequence analysis has subsequently shown pHF5 and pHF7 otherwise to be identical (data not shown). Since pHF5 has a polyadenylation signal sequence at its 3′ end, whereas pHF7 does not, we suspect that pHF7 is a deletion mutant created during cloning rather than deriving from a shorter mRNA.

Preliminary sequence analysis of these clones indicated that they code for actin, but that the longest of them extended only to about amino acid 150 (data not shown). Since our initial clones proved to be less than full length, we devised a strategy to determine whether the cDNA library contained any full-length copies of the two mRNA classes (expected to be ca. 2.1 kilobases [kb]) and then to isolate them. The library, in the form of plasmid DNA, was linearized with either of two restriction enzymes, *Hin*dIII or *Sal*I (18). Agarose gel-separated digests revealed that most of the plasmids were 2.9 kb (equal to the vector size alone) to 5.0 kb and thus contained inserts ranging in size from very small up to 2.1 kb (Fig. 2). The strong staining band at 2.9 kb probably reflects a considerable proportion of vector that contained no insert or cDNA inserts which were almost completely removed by the enzyme digestions. In addition, a number of stained bands can be seen larger than 2.9 kb. These probably correspond to cDNA-vector chimeras which are either full-length cDNAs or result from strong reverse transcription termination sites within abundant mRNAs (see below).

After restriction enzyme digestion and gel separation, we blotted the library onto nitrocellulose paper and hybridized with a nick-translated actin-coding probe. This experiment revealed that the actin cDNA clones in the library could be separated into two classes. In one class the cDNA insert is cut with *Hin*dIII and in the other it is not. Furthermore, since the probe hybridized to actin clones as great as ca. 5.3 kb in length, we concluded that full-length actin inserts existed in the library.

When digested with *Hin*dIII, the class of cDNA clones that contains a *Hin*dIII site within the cDNA (a pHF1 type; see Fig. 1) gave rise to a limited variety of different-length actin-coding fragments ranging from ca. 0.8 to ca. 1.9 kb. Since the vector *Hin*dIII site is located just

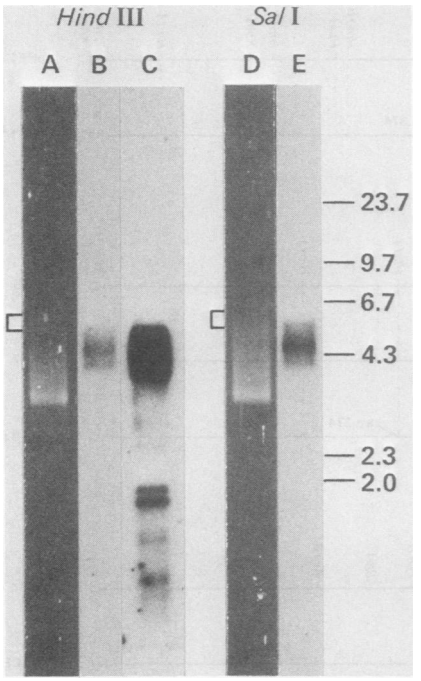

FIG. 2. Visualization of actin cDNA length heterogeneity in the human fibroblast library. Two 1-μg samples of the fibroblast cDNA library were digested with *Hin*dIII (lanes A, B, and C) or with *Sal*I (lanes D and E). The linearized plasmids were electrophoresed on a 0.5% (wt/vol) agarose gel and stained with ethidium bromide (lanes A and D). The DNA was transferred to nitrocellulose paper and allowed to hybridize to nick-translated chick β-actin cDNA insert as described in Materials and Methods. The resulting autoradiograms were obtained after exposures of 12 h (lanes B and E) or 4 days (lane C). The brackets indicate the regions of a parallel gel which was used to recover full-length actin cDNAs (see the text). The relative migration of coelectrophoresed DNA-size standards is shown at the right in kilobase pairs.

beyond the 5′ end linker region and the cDNA *Hin*dIII site is towards the 3′ end of pHF1 (see Fig. 1), the actin-coding fragments released from the vector are all cleaved at this 3′ *Hin*dIII site, and their length heterogeneity must reflect the various points at which reverse transcription stopped. The second actin cDNA clone type (a pHF5 type; see Fig. 1) lacks a *Hin*dIII site in the cDNA. Thus, the actin clones in this second class are linear chimeric molecules with no fragments released. They yield a smear on gels due to a range of molecular sizes, the greatest being ca. 5.0 kb. The prominent band of 4.4 kb is presumably due to the presence of a strong reverse transcription stop. *Sal*I does not cut within either class of actin cDNA sequences. Therefore, the distribution of *Sal*I-linearized actin sequences reveals a strong band of hybridization around 4.4 kb due to a strong stop signal and a continuous range of hybridization up to 5.3 kb.

From the data of Fig. 2, it is clear that only a very small fraction of reverse transcripts of the pHF5 class reach the expected full-length actin cDNA size of 5.0 kb, whereas the opposite appears to be true of the pHF1 class. Strong hybridization bands are seen representing less than full-length cDNA sizes for both actin types. This indicates that preferred premature termination sites for reverse transcriptase exist within both the pHF1 and pHF5 type mRNAs. Thus, it is not surprising that our initial screening efforts failed to yield full-length actin cDNA clones.

Isolation of full-length fibroblast actin cDNAs. The results of the *Hin*dIII and *Sal*I digests suggested a strategy for enriching for full-length clones of both types. Since *Sal*I did not cut within the inserts of either actin cDNA subclass, linearized plasmids of 5.0 kb should be greatly enriched for full-length actin cDNAs and should not contain nonactin cDNAs that might be cut into smaller fragments by the enzyme or actin cDNAs which are less than full length.

After *Sal*I digestion of total library DNA, we extracted linear DNA from the gel regions containing the longest actin cDNAs (see Fig. 2). We used this *Sal*I-digested, size-selected DNA to transform bacteria as described in Materials and Methods. Of the 3,300 clones screened, 22 were actin clones. All 22 clones were identical with respect to their restriction maps with *Hin*fI, *Taq*I, and *Hin*dIII but differed in length at their 5′ ends. Of the 22, 7 were found to be of equal size and maximal length. The restriction map of one of these, pHFγA-1, a member of the pHF1 class, is shown in Fig. 1. Size selection resulted in a 10-fold enrichment for the pHF1 type actin clone (22/3,300 = 0.7%) compared with the frequency of this type in the total plasmid population (2/3,000 = 0.07%). Apparently, a significant proportion of the reverse transcripts of the pHF1 type approach full length, as we predicted from direct visualization of actin clone length in the library (Fig. 2).

In contrast, when the library was digested with *Hin*dIII and size-selected before transformation, only 4 of the 1,600 clones screened were actin clones. The four clones were identical with respect to their restriction maps, but differed in length at their 5′ end. The longest, pHFβA-1, corresponds to the pHF5 class (Fig. 1). Thus, the pHF5 type was not greatly enriched in the size-selected population (4/1,600) relative to its proportion in the total library (6/3,000). This result indicates that only a small fraction of the pHF5 type of reverse transcripts approach full length. This was anticipated because we estimated from the autoradiograph of hybridization of

actin probe to the *Hin*dIII-linearized library (Fig. 2) that only 1 to 2% of the pHF5-type cDNAs approach full length. Given the frequency of all pHF5-type sequences in the plasmid population (0.2%), we calculate that we would have had to screen between 25,000 and 100,000 non-size-selected clones to obtain a full-length clone such as pHFβA-1. Size selection thus resulted in a 50- to 100-fold enrichment of full-length pHF5 clones relative to their concentration in the library.

The failure to obtain any pHF5-type clones by using the *Sal*I-digested DNA is probably due to a difference in the length distributions of these two cDNAs. The longest pHF5 cDNA insert (pHFβA-1) is 2.1 kb, whereas there proved to be a significant population of pHF1 inserts of 2.3 kb or greater (Fig. 2). Our gel excisions in the case of the *Sal*I digestion probably excluded the pHF5 cDNAs.

First 16 amino acids encoded by fibroblast actin cDNAs. A comparison of the first 16 amino acids encoded by pHFγA-1 to the published protein sequence (27) conclusively identifies this clone as encoding human γ actin. Similarly, pHFβA-1 is identified as encoding human β actin (27). Within both clones, the initial methionine codon is immediately followed by the identical amino acid sequence previously defined for the corresponding mammalian protein. These two human fibroblast actin cDNAs are the first higher eucaryotic actins found to lack a cysteine residue as their first amino acid. A full analysis of the DNA sequence of these cDNA clones will be presented elsewhere (P. Ponte, P. Gunning, and L. Kedes, manuscript in preparation).

Analysis of the AUG initiator region of these two mRNA sequences reveals that in both the AUG is preceded at position −3 by a purine and followed at position +4 by a G (Fig. 3). This is characteristic of functional initiation sites for protein synthesis (13). The two initiation regions are otherwise nonhomologous, and neither mRNA contains the consensus sequence CAA Pu AUG found at the initiation site of five of the six *Drosophila* actin genes (7).

Isolation of muscle actin cDNA clones. To evaluate the human muscle cDNA library, we first analyzed the library in the form of plasmid DNA. Linearization with either *Eco*RI or *Hin*dIII revealed that most of the plasmids varied in size from 2.5 kb (corresponding to vector alone) to 4.5 kb (Fig. 4). A large number of bands were visualized above the background smear. Hybridization of these size-fractionated cDNAs with a nick-translated probe to actin-coding region revealed three size classes, the longest of which, 4.3 kb (Fig. 4), contains a cDNA insert of 1.6 kb. This is similar to the expected size of mammalian α-actin mRNA (15, 22). The hybridized actin band corresponds to a prominent band on the ethidium bromide-stained gel of the linearized muscle library (but not in the fibroblast library, allowing for the difference in vector size [see Fig. 3]). This supports our supposition that visible bands in the linearized cDNA libraries represent either products prematurely terminated by strong reverse transcription stops or full-length transcripts of abundant mRNAs.

The entire library of 5,400 clones was screened with a probe to the actin-coding region, and 30 positive clones were identified. Restriction endonuclease analysis of the 30 clones demonstrated that they fell into three size classes corresponding to the three hybridizing actin bands in Fig. 4. The three classes represent transcripts of increasing length with a common 3′ end. Eight clones fell into the longest class of maximal 4.3-kb size. The restriction map of one of these (pHMαA-1) is shown in Fig. 1. It has essentially no restriction sites in common with either human β- or γ-actin cDNAs.

Sequence analysis of pHMαA-1 revealed that, like the chick and rat α-actin mRNAs (6, 29), the human muscle actin cDNA encodes a cysteine

```
pHFβA-1      ... CTC ACC ATG GAT GAT GAT ATC GCC GCG CTC GTC GTC GAC AAC GGC TCC GGC ATG TGC ...
                 MET Asp Asp Asp Ile Ala Ala Leu Val Val Asp Asn Gly Ser Gly Met Cys ...

pHFγA-1      ... GCC GCA ATG GAA GAA GAG ATC GCC GCG CTG GTC ATT GAC AAT GGC TCC GGC ATG TGC ...
                 MET Glu Glu Glu Ile Ala Ala Leu Val Ile Asp Asn Gly Ser Gly Met Cys ...

pHMαA-1  ... GAC ACA ATG TGC GAC GAA GAC GAG ACC ACC GCC CTC GTG TGC GAC AAT GGC TCC GGC TTG GTG ...
                 MET Cys Asp Glu Asp Glu Thr Thr Ala Leu Val Cys Asp Asn Gly Ser Gly Leu Val ...
```

FIG. 3. Amino-terminal amino acid sequences derived from human cDNA clones. The 5′ ends of the two fibroblast cDNA clones, pHFβA-1 and pHFγA-1, and the muscle cDNA clone pHMαA-1 were sequenced by using the segments shown in Fig. 1. The derived amino acid sequences are identical with the corresponding mammalian β-, γ-, and α-actin proteins, and they unambiguously identify the coding capacity of the clones. Note the presence of a cysteine codon immediately after the initiator methionine codon in α-actin (boxed) but not in β- or γ-actin mRNA.

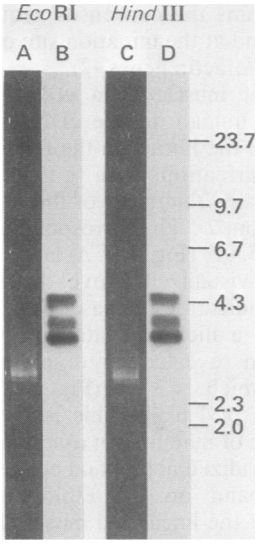

FIG. 4. Visualization of actin cDNA length heterogeneity in the human muscle library. Two 1-μg samples of the muscle cDNA library were digested with either *Eco*RI (lanes A and B) or with *Hin*dIII (lanes C and D). The linearized plasmids were analyzed exactly as described in the legend to Fig. 2. Ethidium bromide staining of the gel is seen in lanes A and C, and the corresponding autoradiograms of actin hybridization are seen in lanes B and D. The position of coelectrophoresed size standards are at the right in kilobase pairs.

residue after the initiator MET residue. However, neither of these amino acids is detected in any mammalian actin proteins (Fig. 5) (27). The next 17 amino acids encoded by this clone are identical to those of mammalian α-actin protein (27). These results suggest that the α-actin genes of birds and mammals encode a penultimate amino-terminal cysteine not found in the final protein product.

Subsequent fine-structure analysis has revealed that 2 of the 30 muscle cDNA clones encode a distinct muscle actin. These clones have been demonstrated to encode human cardiac actin. The details and implications of finding cardiac actin cDNA in a library prepared from human adult skeletal muscle will be presented elsewhere (P. Gunning, P. Ponte, H. Blau, and L. Kedes, manuscript in preparation).

DISCUSSION

Evolution of human actin genes. We have demonstrated that the human fibroblast β- and γ-actin genes do not encode a cysteine immediately following the MET initiation codon, unlike human α actin (this work) or cardiac actin (11). This result strongly suggests that two discrete actin gene types gave rise to the deuterostome actins, but this result also raises an unsettling evolutionary paradox. The problem derives from the following considerations. First, the presence of a cysteine residue at amino acid 1 (referred to as "Cys+") is unique to protostome and deuterostome actin genes (2, 6, 7, 29), whereas all actin genes in more primitive organisms lack this residue (called "Cys-") (5, 8, 16). Second, the ubiquitous occurrence of both Cys+ and a mechanism to remove the amino-terminal cysteine in both protostomes and deuterostomes is too coincidental to be explained by convergent evolution. The Cys+ line of evolutionary argument therefore suggests that at some time before the divergence of protostomes and deuterostomes, two actin gene types already existed, both encoding a cytoplasmic-like actin, but only one of which had a cysteine at its amino terminus. No amino-terminal cysteine has ever been detected in any actin protein yet sequenced (see Discussion in reference 25). Therefore, it is likely that post-translational removal of this amino acid is necessary for normal actin function. Thus, one would predict that a mechanism for removal of this residue coevolved with the Cys+ actin genes. The evolutionary paradox arises from the fact that amino acid and nucleic acid sequence data show a closer relationship between the protostome Cys+ actins and the mammalian Cys- β and γ cytoplasmic actins.

A heterodox interpretation of the evolution of these two gene types based on the Cys+/Cys- data is shown in Fig. 5. According to this model, the *D. melanogaster* genes and at least two sea urchin cytoplasmic-like actin genes are evolutionarily more closely related to each other and to the mammalian α-actin gene than they are to human cytoplasmic β- and γ-actin genes. However, amino acid and nucleotide sequence analyses indicate that this is not the case. Unfortunately, intron location data shed no light on the evolutionary relationship of the actin genes (although such data have been used in an attempt to do so [29]) and thus does not help us out of the paradox. For example, *D. melanogaster* shares no intron locations with any deuterostome actin genes. Within the deuterostomes, the sea urchin Cys+ gene contains one intron found at the same site as in rat α but not β actin (amino acid 204), one intron location found in rat β but not α actin (amino acid 121), and two intron locations shared by all three (amino acids 41, 327), and it lacks one intron location shared by rat α and β actin (amino acid 267) and one intron found in only rat α actin (amino acid 150) (29). On the other hand, if it can be demonstrated that the Cys− lineage is represented in most other deuterostome actin families, then the use of the Cys+ and Cys− lineages may facilitate our

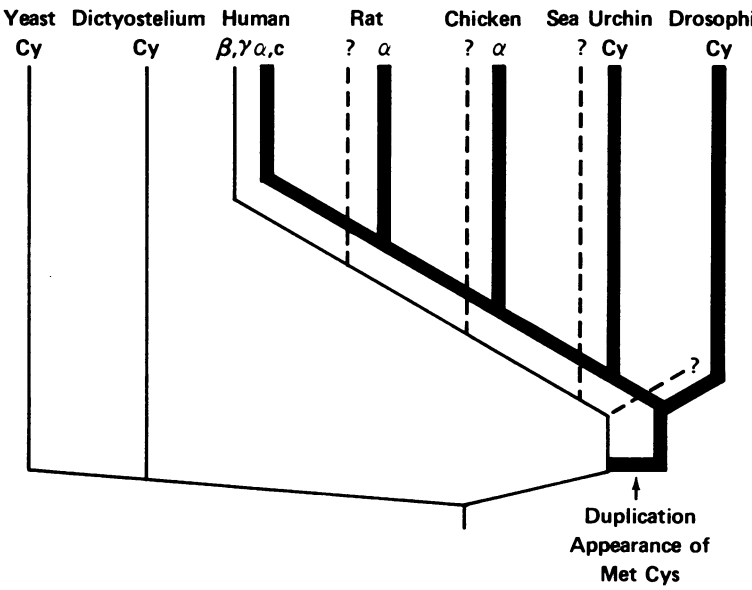

FIG. 5. Heterodox evolutionary relationship of Cys+ and Cys− actin genes. The archetype actin gene lineage, which presumably encoded a Cys− cytoplasmic-type actin (designated Cy), is represented by the thin line. The appearance of the Cys+ lineage, presumably deriving from an ancient duplication of a Cys− gene, is represented by the thick line. The vertical scale is not linear and merely represents relative evolutionary time. In mammals and birds, the muscle actins are called α (skeletal) and c (cardiac) and the cytoplasmic actins are termed β and γ. The broken lines represent uncertainties in the Cys− lineage due to absence of data. The relationship is heterodox because the amino acid and nucleic acid sequence data suggest that the genes designated Cy are more closely related to the β and γ actins than they are to the α actins.

understanding of the evolution of actin genes and the meaning of differences in intron location. Such data will also serve to address an alternative interpretation, namely, that the human β and γ actin genes derive from a Cys+ lineage but have lost this cysteine coding triplet at some stage of deuterostome evolution.

The possibility that Cys+ β- and γ-actin genes are also present in humans must be considered, since we have demonstrated that there are at least 30 actin-coding fragments in the human genome (3, 4). However, using probes to the 3′ untranslated region of both our β- and γ-actin mRNAs, we can account for essentially all of these human actin genes. An analysis of the 3′ ends of the β- and γ-actin cDNAs shows that they are distinct and, therefore, define two nonoverlapping multigene subfamilies (Ponte et al., manuscript in preparation). Thus members of each of the two subfamilies must have evolved from a progenitor sequence and should be similar with respect to the coding capacity of their 5′ ends. This result indicates that if Cys+ cytoplasmic actin genes do occur in the human genome, they must comprise only a small fraction of the total number of actin genes.

On the role of cysteine at the amino terminus. Since the human fibroblast β- and γ-actin mRNAs do not encode a cysteine at amino acid 1, it can be confidently concluded that this residue is not essential for nonmuscle actin function. In contrast, the demonstration of this encoded residue in chick α (6), rat α (29), human cardiac (11), and now human α actin (this work), in addition to all six *Drosophila* actin genes (7), raises the possibility that this residue is intimately involved in muscle actin gene expression. If this cysteine is post-translationally removed in all cases, as it clearly is in chicken (6) and mammals (26, 29), then this residue may serve a regulatory rather than a structural role in actin expression. We predict that cysteine is removed before actin-myosin assembly since it would probably interfere with major actin-myosin interactions which involve the amino terminus of actin (24).

Advantages of the Okayama-Berg cloning procedure. Finally, we wish to draw attention to some aspects of the cloning procedure used. Both cDNA libraries were generated by the Okayama-Berg technique (17, 18). This method optimizes conditions for the reverse transcription of mRNAs and results in cDNA covalently linked to a plasmid vector. By avoiding standard nuclease trimming steps, the final cDNA clones obtained contain all of the original cDNA syn-

thesized (17). It is clear from our experiments that the efficiency with which full-length cDNA clones are obtained is quite variable, due to differences in the particular mRNA template used. Some mRNAs, such as human β actin, are very poorly reverse transcribed and have especially strong termination sites within the mRNA itself. In such cases, a full-length cDNA is rarely obtained. Others, such as human α- and γ-actin mRNAs, are very good templates. For example, in the case of α actin, 25% of all clones are close to full length (within 100 base pairs).

We have found that evaluation of the pooled plasmids derived from the cDNA library can greatly facilitate the ultimate isolation of full-length clones. These observations suggested two general methods for cloning full-length cDNAs of interest. Ideally, the library should be digested with one or more enzymes which cut the vector only once, do not cut within the cDNA of interest, but do cut other cDNAs. This will result in an enrichment of the frequency of cDNAs with the desired sequences in the region of the gel corresponding to full-length size. With a given restriction enzyme with a six-base recognition site, roughly half of a random population of 3.0-kb inserts might be expected to be cut within the inserts. Full-length clones can be excised from a gel, recircularized, and used for transfections. In the case of clones which rarely obtain full length due to strong internal reverse transcription stops or to very large mRNA size, this method increases the probability of obtaining such a rare clone from among the transfectants screened. Second, in this method the common starting point for reverse transcription of a given mRNA generates a limited number of sets of transcripts of homogeneous lengths. For abundant classes of mRNAs, these limited sets are visible as bands in ethidium bromide-stained gels after linearization of the library. By comparing libraries from different tissues or stages of differentiation, cDNAs to genes whose expression changes can be identified. DNA excised from such bands can be recircularized and used for transfection. Different clones selected at random should contain multiple examples of a common cDNA insert representing the regulated mRNA. Among the advantages of such an approach is the fact that no hybridization probe is required for the selection process. A manuscript providing a full demonstration of the power of this approach is in preparation (Gunning et al, in preparation).

ACKNOWLEDGMENTS

We thank Rob Maxson, Alex Mauron, Tim Mohun, and Karen O'Malley for many helpful discussions.

This work was supported in part by grants from the National Institutes of Health to L.K., to H.B. and to Paul Berg, by grants from the Muscular Dystrophy Association to L.K. and to H.B., by a grant from the Veterans Administration to L.K., by a grant from the March of Dimes Birth Defects Foundation to H.B., and by a grant from the American Cancer Society to Paul Berg. P.P. is a fellow of the American Cancer Society, and H.O. is a fellow of the Jane Coffin Childs Fund for Medical Research. J.E. was supported in part by the Medical Scientist Training Program.

LITERATURE CITED

1. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidilic acid-cellulose. Proc. Natl. Acad. Sci. U.S.A. **69**:1408–1412.
2. **Cooper, A. D., and W. R. Crain.** 1982. Complete nucleotide sequence of a sea-urchin actin gene. Nucleic Acids Res. **10**:4081–4092.
3. **Engel, J. N., P. W. Gunning, and L. H. Kedes.** 1981. Isolation and characterization of human actin genes. Proc. Natl. Acad. Sci. U.S.A. **78**:4674–4678.
4. **Engel, J., P. Gunning, and L. Kedes.** 1982. The human genome contains multiple cytoplasmic actin genes. Mol. Cell. Biol. **2**:674–684.
5. **Firtel, R. A., R. Timm, A. R. Kimmel, and M. McKeown.** 1979. Unusual nucleotide sequences at the 5′ end of actin genes in Dictyostelium discoideum. Proc. Natl. Acad. Sci. U.S.A. **76**:6206–6210.
6. **Fornwald, J. A., G. Kuncio, I. Peng, and C. P. Ordahl.** 1982. The complete nucleotide sequence of the chick alpha-actin and its evolutionary relationship to the actin gene family. Nucleic Acids Res. **10**:3861–3876.
7. **Fyrberg, E. A., B. J. Bond, N. D. Hershey, K. S. Mixter, and N. Davidson.** 1981. The actin genes of drosophila: protein coding sequences are highly conserved but intron positions are not. Cell **24**:107–116.
8. **Gallwitz, D., and I. Sures.** 1980. Structure of a split yeast gene: complete nucleotide sequence of the actin gene in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U.S.A. **77**:2546–2550.
9. **Goldman, R. T. Pollard, and J. Rosenbaum (ed.).** 1976. Cell motility. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **Grunstein, M., and D. S. Hogness.** 1975. Colony hybridization: a method for isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. U.S.A. **72**:3961–3965.
11. **Hamada, H., M. G. Petrino, and T. Kakunaga.** 1982. Molecular structure and evolutionary origin of human cardiac muscle actin gene. Proc. Natl. Acad. Sci. U.S.A. **79**:5901–5905.
12. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmid. Anal. Biochem. **114**:193–197.
13. **Kozak, M.** 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. **9**:5233–5252.
14. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific cleavage reactions. Methods Enzymol. **65**:499–560.
15. **Minty, A. J., S. Alonso, M. Caravatti, and M. E. Buckingham.** 1982. A fetal skeletal muscle actin mRNA in the mouse and its identity with cardiac mRNA. Cell **30**:185–192.
16. **Ng, R., and J. Abelson.** 1980. Isolation and sequence of the gene for actin in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U.S.A. **77**:3912-3916.
17. **Okayama, H. and P. Berg.** 1982. High-efficiency cloning of full-length cDNA. Mol. Cell. Biol. **2**:161–170.
18. **Okayama, H. and P. Berg.** 1983. A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. Mol. Cell. Biol. **3**:280–289.
19. **Palmiter, R. D.** 1974. Magnesium precipitation of ribonucleoprotein complexes: expedient techniques for the isolation of undegraded polysomes and messenger ribonucleic acid. Biochemistry **13**:3606-3615.

20. **Pollard, T. D., and R. R. Weihing.** 1974. Actin and myosin and cell movement. Crit. Rev. Biochem. **2**:1–65.
21. **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.
22. **Shani, M., D. Zevin-Sonkin, O. Saxol, Y. Carmon, D. Katcoff, U. Nudel, and D. Yaffe.** 1981. The correlation between the synthesis of skeletal muscle actin, myosin heavy chain, and myosin light chain and the accumulation of corresponding mRNA sequences during myogenesis. Dev. Biol. **86**:483–492.
23. **Smith, H. O., and M. L. Birnstiel.** 1976. A simple method for DNA restriction site mapping. Nucleic. Acid. Res. **3**:2387–2397.
24. **Sutoh, K.** 1982. Identification of myosin binding sites on the actin sequence. *Biochemistry* **21**:3654–3661.
25. **Vandekerchkhove, J., W. Franke, and K. Weber.** 1981. Diversity of expression of non-muscle actin in amphibia. J. Mol. Biol. **152**:413-426.
26. **Vandekerckhove, J., and K. Weber.** 1978. At least six different actins are expressed in a higher mammal: an analysis based on the amino acid sequence of the amino-terminal tryptic peptide. J. Mol. Biol. **126**:783–802.
27. **Vandekerckhove, J., and K. Weber.** 1979. The complete amino acid sequence of actins from bovine aorta, bovine heart, bovine fast skeletal muscle, and rabbit slow skeletal muscle. Differentiation **14**:123–133.
28. **Vandekerckhove, J. and K. Weber.** 1980. Vegetative Dictyostelium cells containing 17 actin genes express a single major actin. Nature (London) **284**:475–477.
29. **Zakut, R., M. Shani, D. Givol, S. Neuman, D. Yaffe, and U. Nudel.** 1982. The nucleotide sequence of the rat skeletal muscle actin gene. Nature (London) **298**:857–859.