

# Mammalian cytoplasmic actins are the products of at least two genes and differ in primary structure in at least 25 identified positions from skeletal muscle actins

(smooth muscle/brain/thymus/protein sequence/isoelectric focusing)

JOËL VANDEKERCKHOVE AND KLAUS WEBER

Max-Planck-Institut für Biophysikalische Chemie, 3400 Göttingen, Federal Republic of Germany

Communicated by F. Lynen, November 28, 1977

**ABSTRACT** Muscle and cytoplasmic actins from several species have been compared by extensive fingerprint analysis and by partial amino acid sequence determination with the known amino acid sequence of rabbit muscle actin. Although complete sequences have not been established, the following characteristics are apparent. (a) Cytoplasmic actins are the products of two different genes. The difference seen in isoelectric focusing studies is probably determined only by the nature of the three amino-terminal acidic residues. (b) Mammalian cytoplasmic actins are exceedingly similar and perhaps identical. (c) Cytoplasmic actins may differ by at least 25 amino acid replacements from rabbit muscle actin. These replacements have been identified for calf thymus actin; however, other cytoplasmic actins show the same replacements. (d) The replacements always involve—except for the first five residues—neutral amino acid residues. (e) The replacements are not randomly distributed. Residues 18–75 are constant whereas residues 2–18 and 259–298 show many substitutions. (f) The main component of smooth muscle actin from chicken gizzard shows the charge characteristics found at the amino terminus of the less acidic cytoplasmic actin species. In the rest of the polypeptide chain, gizzard actin resembles skeletal muscle actin, although two substitutions of the cytoplasmic type have been identified. (g) Heart muscle actin is very similar to skeletal muscle actin. Only two amino acid replacements have been found; they are of the cytoplasmic type. (h) Skeletal muscle actins from chicken and beef have not shown a replacement.

Biochemical and electron microscopic studies showed originally that actin is a typical protein of muscle tissue and that "actin-like" proteins occur also in nonmuscle cells (for a review see ref. 1). In 1972, Bray proposed, on grounds of restricted fingerprint analyses, that "actin-like" proteins from nonmuscle cells are true actins and are extremely similar, if not identical, to skeletal muscle actin (2). Although these studies suggested that actin is a protein typical of eukaryotic cells and extremely well conserved during evolution, the question remained as to how similar different cytoplasmic actins are, both to each other and to muscle actin at the primary sequence level. This is a particularly intriguing question because *in vivo* the polymerization of muscle and nonmuscle actins can be quite different. In muscle, actin is highly polymerized and forms the thin filament which is arranged in nearly crystalline ordered sarcomeric arrays. By contrast, nonmuscle cells express actin in the form of microfilaments which are usually much less ordered and sometimes can aggregate laterally to form bundles of various thickness (for a review see ref. 1). Furthermore, different cell lines and different tissues can differ dramatically in the extent to which they express strong microfilament bundles although

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

the total cellular actin content seems rather similar (see, for instance, refs. 1 and 3).

Comparative fingerprint analysis of brain actin and of muscle actin from chicken indicated several changes in the tryptic peptide pattern (4). In addition, the determination of the complete amino acid sequence of rabbit muscle actin (5) and studies on some cyanogen bromide fragments of actins from human heart, human platelets, and beef brain have allowed Elzinga and his coworkers to assert the existence of different actin genes within the same organism (5, 6). More recently, isoelectric focusing studies of different actins in urea solution have shown that actin from skeletal muscle gives a single species ( $\alpha$  actin) while a variety of nonmuscle actins give rise to two different species ( $\beta$  and  $\gamma$  actin) (7, 8).

Because of the importance of actin for both muscle and cytoplasmic activity, it is necessary to establish how many kinds of muscle and cytoplasmic actins there are and how different these actins are both from each other and from muscle actin. Here we summarize our results on the partial amino acid sequences of three muscle actins, four cytoplasmic actins, and one smooth muscle actin. The data obtained establish clearly that different classes of actins can be distinguished by their primary sequences. In addition, our results on cytoplasmic actins, together with previous data on brain actin (5), show that the cytoplasmic actins are astonishingly similar to each other, but differ in at least 25 amino acid residues from the muscle sequence. Finally, cytoplasmic actins are the products of at least two different genes, the major amino acid differences being localized at the amino-terminal end. Identification of these replacements allows us to propose an explanation for the different actin forms observed in isoelectric focusing studies.

## MATERIALS AND METHODS

**Actins.** The different actins were purified by standard procedures. Muscle actins were purified according to Spudich and Watt (9). Cytoplasmic actins were purified from calf thymus and bovine brain by affinity chromatography on immobilized pancreatic deoxyribonuclease (10) and from bovine brain and tissue culture cells (mouse 3T3 cells transformed by simian virus 40 and rat glioma tumor C<sub>6</sub> cells) by polymerization-depolymerization procedures (11). All actins were at least 90–95% pure as judged by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate. Isoelectric focusing studies yielded, as expected from previous reports (7, 8), exclusively  $\alpha$  actin for skeletal muscle actins and bovine heart actin and a mixture of  $\beta$  and  $\gamma$  actin for cytoplasmic actins with some variation in the relative intensity of the two species. Chicken gizzard actin gave predominantly one component migrating similarly to  $\gamma$  actin (12, 13).

Table 1. Localization of amino acid substitutions in eight different actins compared to rabbit skeletal muscle actin

Actin source	Residue no.																			
	16	17	76	89	106	129	153	162	176	201	225	259	266	271	278	286	296	298	357	364
	Skeletal muscle actin																			
Rabbit	<i>Leu</i>	<i>Val</i>	<i>Ile</i>	<i>Thr</i>	<i>Thr</i>	<i>Val</i>	<i>Leu</i>	<i>Asn</i>	<i>Met</i>	<i>Val</i>	<i>Asn</i>	<i>Thr</i>	<i>Ile</i>	<i>Ala</i>	<i>Tyr</i>	<i>Ile</i>	<i>Asn</i>	<i>Met</i>	<i>Thr</i>	<i>Ala</i>
Beef (80%)	<i>Leu</i>	<i>Val</i>	<u><i>Ile</i></u>	<i>Thr</i>	<i>Thr</i>	<u><i>Val</i></u>	<i>Leu</i>	<i>Asn</i>	<i>Met</i>	<i>Val</i>	<i>Asn</i>	—	—	—	—	—	<i>Asn</i>	<i>Met</i>	<i>Thr</i>	<i>Ala</i>
Chicken (76%)	<i>Leu</i>	<i>Val</i>	<u><i>Ile</i></u>	<i>Thr</i>	<i>Thr</i>	—	<i>Leu</i>	<i>Asn</i>	<i>Met</i>	<i>Val</i>	<i>Asn</i>	—	—	—	—	—	<i>Asn</i>	<i>Met</i>	—	<i>Ala</i>
	Heart muscle actin																			
Beef (71%)	<i>Leu</i>	<i>Val</i>	<i>Ile</i>	<i>Thr</i>	<i>Thr</i>	<i>Val</i>	<i>Leu</i>	<i>Asn</i>	<i>Met</i>	<i>Val</i>	—	—	—	—	—	—	<i>Asn</i>	<i>Leu</i>	<u><i>Ser</i></u>	<u><i>Ala</i></u>
	Smooth muscle actin																			
Chicken gizzard (87%)	<i>Leu</i>	<i>Cys</i>	<i>Ile</i>	<i>Ser</i>	<i>Thr</i>	<i>Val</i>	<i>Leu</i>	<i>Asn</i>	<i>Met</i>	<i>Val</i>	<i>Asn</i>	<i>Thr</i>	—	—	—	—	<i>Asn</i>	<i>Leu</i>	—	—
	Cytoplasmic actins																			
Bovine thymus (97%)	<i>Met</i>	<i>Cys</i>	<i>Val</i>	<i>Thr</i>	<i>Val</i>	<i>Thr</i>	<i>Met</i>	<i>Thr</i>	<i>Leu</i>	<i>Thr</i>	<i>Gln</i>	<i>Ala</i>	<i>Leu</i>	<i>Cys</i>	<i>Phe</i>	<i>Val</i>	<i>Thr</i>	<i>Leu</i>	<i>Ser</i>	<i>Ser</i>
Bovine brain (78%)	<u><i>Met</i></u>	<u><i>Cys</i></u>	<i>Val</i>	<i>Thr</i>	<i>Val</i>	<i>Thr</i>	—	—	<i>Leu</i>	<i>Thr</i>	<i>Gln</i>	<i>Ala</i>	<i>Leu</i>	<u><i>Cys</i></u>	<u><i>Phe</i></u>	<i>Val</i>	<i>Thr</i>	<u><i>Leu</i></u>	<i>Ser</i>	<i>Ser</i>
Mouse SV101 cells (62%)	<i>Met</i>	<i>Cys</i>	<i>Val</i>	<i>Thr</i>	<i>Val</i>	—	—	—	—	<i>Thr</i>	—	—	—	—	—	—	<i>Thr</i>	<i>Leu</i>	—	<i>Ser</i>
Rat C6 cells (62%)	<i>Met</i>	<i>Cys</i>	<i>Val</i>	<i>Thr</i>	<i>Val</i>	—	—	—	—	<i>Thr</i>	—	—	—	—	—	—	<i>Thr</i>	<i>Leu</i>	—	<i>Ser</i>

The substitutions found in the amino-terminal tryptic peptide of these actins (residues 1–18) are contained in Fig. 1. The residues are numbered according to the sequence of rabbit skeletal muscle actin (5). Numbers in parentheses give the percentage of the total amino acid sequences currently analyzed. The replacements that were reported by Elzinga and Lu (5, 21) are given in italics; those that were reported previously and also found by our approach are underlined. A dash indicates that the position has not yet been determined.

**Protein Chemical Procedures.** A full description of the methods used and the sequences obtained will be presented elsewhere. Briefly, actins were characterized by their tryptic peptides obtained by preparative fingerprint analysis using performic acid-oxidized protein. The peptides were separated on paper by a combination of electrophoresis and chromatography as described (14) and characterized by their amino acid composition and by their electrophoretic mobility according to the Offord plot (15). For long peptides, further enzymatic digestion was performed with, depending on the sequence of the corresponding muscle peptide, thermolysin, chymotrypsin, subtilisin, or *Staphylococcus aureus* protease. The resulting fragments were characterized as above. Peptides that showed an amino acid replacement when compared to the corresponding muscle peptide were further characterized by the dansyl-Edman degradation (16) and digestion with leucine aminopeptidase and carboxypeptidases A and B in order to locate the position of the amino acid exchange. The regions of the amino acid sequence corresponding to the insoluble tryptic core required further digestion with chymotrypsin, and the resulting peptides were then purified as described above. For those actins, that have been characterized most extensively (gizzard and thymus), a similar approach was used starting with chymotrypsin rather than trypsin and a characterization of the soluble CNBr cleavage products.

The amino-terminal tryptic peptides of the actins discussed in Fig. 1 were characterized by their amino acid composition and electrophoretic mobility. They were further fragmented with chymotrypsin, thermolysin, and *Staphylococcus aureus* protease. The sequences of those fragments with a free amino terminus were determined as described above and the sequence of the blocked peptides was obtained by a combination of carboxypeptidase A digestion, partial acid hydrolysis, and comparison with the electrophoretic mobilities of synthetic acetylated peptides.

**Limits of the Protein Chemical Approach Used.** Our ap-

proach relies heavily on the amino acid sequence for rabbit muscle actin reported by Elzinga and Lu and colleagues (5). We would like to note the degree of agreement with this sequence revealed in our studies. Our first screening of peptides relied on amino acid composition and the Offord plot. Only when amino acid exchange was suspected, was a detailed analysis performed (see above). Within these limitations we found very few disagreements with the earlier versions of the muscle actin sequence (17–19). These involved mainly amide assignments, and most of our results agree with the recently corrected sequence reported by Elzinga and Lu and coworkers (5). Specifically, we find an amide rather than an acidic residue in position 11 or 12 and in position 245. We also agree with an Ile rather than a Thr residue at position 308 and Glu instead of Gln at site 214. There is currently the possibility—however not yet proven—that the tryptophan residue is located not at position 74 but rather at position 78, 79, or 80. Thus, the substitution given in Table 1 at position 76 corresponds to position 77 in Elzinga's sequence. In aligning the peptides within the full sequence we again used the rabbit skeletal muscle sequence. Our results on different actins are expressed in an arrangement parallel to this sequence, although we show below that cytoplasmic actins and gizzard actin seem to be shorter by the amino-terminal residue, thus starting at residue 2.

The current limitations of our approach are: First, if a peptide from one of the actins would contain more than one amino acid replacement but still have the same amino acid composition and electrophoretic and chromatographic properties as the corresponding muscle peptide, we would currently consider it as unchanged. Second, we have not investigated the possibility of post-translational modifications of some actin residues, which could influence the isoelectric points of the different actins. We assume that a full residue of methylhistidine is present, as indicated by others (8, 20). Thus, the number of amino acid exchanges between different actins given below is a lower estimate.



In both brain and thymus actin the two cytoplasmic species ( $\beta$  and  $\gamma$ ; see above and Fig. 1) are missing the amino-terminal residue typical for skeletal muscle actin. They differ from each other not only in their three amino-terminal residues (position 2–4), but also in position 10. Beyond position 17 we do not know if all the 19 replacements listed in Table 1 for thymus actin are typical for both the  $\beta$  and the  $\gamma$  actins since the majority of the replacement studies were done before we became convinced by the finding of the two amino-terminal tryptic peptides that  $\beta$  and  $\gamma$  are two different gene products. The yield of most of the peptides, together with the chemical nature of some of the substitutions, are arguments that in the majority of the places both  $\beta$  and  $\gamma$  actin have the same substitution. This is in agreement with fingerprint analysis of radioactively labeled  $\alpha$ ,  $\beta$ , and  $\gamma$  actins, which showed a high degree of homology between  $\beta$  and  $\gamma$ , although two peptides were found that were common to  $\alpha$  and  $\beta$  but not present in  $\gamma$  (8). Similar arguments have to be raised in discussing the possibility of more than two cytoplasmic actins. Minor components electrophoretically identical with either  $\beta$  or  $\gamma$  but distinct by one or the other conservative replacement could have escaped our analysis.

Two other cytoplasmic actins were isolated from tissue culture cells (3T3 mouse fibroblasts transformed by simian virus 40 and rat glial tumor cells C<sub>6</sub>) growing *in vitro* and partial sequences were determined. Here we can account for 62% of the total amino acid sequence (see Table 1), leaving the regions 119–177, 214–237, 254–289, and 336–358 currently unchecked. However, all the eight amino acid replacements found in the studied regions are the same as those found in bovine brain and thymus actin. Furthermore, no other amino acid exchanges have been found in these parts of the sequence. The one amino acid replacement reported so far for human platelet actin (6) again involves a position (residue 129) with the same replacement as in brain and thymus actin.

The results shown above make it likely, although currently (short of the total sequences) by no means yet proven, that the four mammalian cytoplasmic actins studied are identical in amino acid sequence. There is no indication for species specificity, although the sample covers beef, mouse, and rat actins. There is also no indication for tissue specificity outside the muscle tissues (see below) because we have used actins from brain, thymus, virally transformed fibroblasts, and rat glial tumor cells.

An important result involves the nature of the amino acid replacements found so far. Of the 21 replacements between residues 6 and 374 found in thymus actin, with only 10 residues (residues 336–345) unchecked, none involves a basic or acidic amino acid. Thus, the overall net charge and the charge pattern of the muscle and cytoplasmic actins are extraordinarily conserved at physiological pH. Only the five residues at the amino terminus show changes in the pattern of acidic residues typical for this region. It may well be that this high degree of conservation of charge may be dictated by the multiple interactions of all the actins with other proteins in the cytoplasm as well as in the muscle tissue (i.e., actin itself, tropomyosins, troponin, myosins, filamin,  $\alpha$  actinin, Z-line structures, and others; for a review see ref. 1).

A further interesting point is the distribution of the amino acid replacements along the linear sequences of skeletal muscle and thymus (cytoplasmic) actins. We have identified one region (residues 18–75) apparently without a single change, while another region (residues 259–298) and the region at the amino-terminus each shows a high degree of replacements. The three-dimensional structure of actin and detailed knowledge of its interaction with itself and other proteins should reveal the

importance of the individual amino acid replacements, their uneven distribution, their conservative nature, which avoids a change of charged residues, as well as the exchange pattern of acidic residues observed at the amino-terminus.

**The Main Component of Chicken Gizzard Actin Is Intermediate between Skeletal Muscle Actin and the  $\gamma$  Species of Cytoplasmic Actins.** Chicken gizzard actin was used in the hope that it might be a prototype of smooth muscle actins. Although currently only 87% of its amino acid sequence has been studied it is obvious that between residues 16 and 374 gizzard actin is very similar to skeletal muscle actin, showing only three replacements (positions 17, 89, and 298) with two (positions 17 and 289) being of the type also found in cytoplasmic actins (Table 1). The substitution at 89 is specific for gizzard, since here cytoplasmic actins and skeletal muscle actin are indistinguishable. It is important to note that in the studied sequences 11 positions are covered at which skeletal muscle actin and cytoplasmic actins differ but at which gizzard actin agrees with the skeletal muscle actin sequence. During our studies it was reported that adult gizzard actin contains two actin species upon isoelectric focusing which correspond to the  $\beta$  and  $\gamma$  cytoplasmic actins, with the  $\gamma$  species being much more abundant (12, 13). Our gizzard actin gives rise to two amino-terminal peptides, of which only the more abundant one has been studied (Fig. 1). In agreement with our previous interpretation of the isoelectric focusing data of cytoplasmic  $\beta$  and  $\gamma$  actin, the peptide carries the three amino-terminal glutamyl residues typical of cytoplasmic  $\gamma$  actin and is shorter, by one amino acid residue (the amino-terminal Asp), than the corresponding  $\alpha$  actin peptide of skeletal muscle (Fig. 1). In the rest of the amino-terminal sequence it resembles both muscle actin (positions 6, 10, and 16) and the  $\gamma$  component of cytoplasmic actin (positions 2, 3, 4, and 17). Preliminary results on the amino acid composition of the amino-terminal peptide found in minor amounts are indicative of the existence of a minor actin component similar to the  $\beta$  form of cytoplasmic actin, in agreement with the electrofocusing results (12, 13). Thus, gizzard actins could be coded for by two different genes as shown above for cytoplasmic actins, and the isoelectric focusing results can most likely be explained by the most amino-terminal residues. Currently we do not know if any function involves this cluster of acidic residues and their different arrangement in the different actins.

Isoelectric focusing studies on actins seem only to emphasize the type of amino-terminal sequence present, since exchanges of charged residues are limited to the amino terminus. Thus, for example, the procedure suggests a very strong relation between gizzard actins and cytoplasmic actins compared to muscle actin only because it cannot detect the numerous positions involving neutral amino acids in which the major component of gizzard is much more related to skeletal muscle actin than to the  $\gamma$  species of cytoplasmic actins.

**Bovine Heart Muscle Actin Differs at at Least Two Positions from Skeletal Muscle Actin.** For bovine heart muscle actin, 71% of its amino acid sequence (excluding residues 119–147, 214–237, 254–289, and 336–355) has been studied. These results (Table 1), together with the previous work of Elzinga and Lu (5), establish only two amino acid replacements so far, both of which are typical of cytoplasmic actins (positions 298 and 357). The close relationship between skeletal muscle and heart muscle actins is indicated by the fact that at 11 positions at which cytoplasmic actins differ from skeletal muscle actin, we do not detect these substitutions in heart muscle actin. In agreement with isoelectric focusing results (13) showing that heart muscle behaves like skeletal muscle actin, we found only

one amino-terminal tryptic peptide and this has the same composition as the corresponding muscle peptide.

**Skeletal Muscle Actins of Higher Vertebrates Are Extremely Similar in Primary Structure.** More than 76% of the complete sequence of skeletal muscle actins from beef and chicken have been determined. (Regions 133–147, 254–289, and 336–358 of bovine muscle actin have not been studied; regions 119–147, 254–289, and 336–358 of chicken muscle actin have not been studied.) In agreement with a previous study on 55 residues in bovine muscle actin (5), we do not find an amino acid exchange when compared to rabbit skeletal actin, indicating an extremely high degree of similarity in primary structure. However, the possibility of a rare amino acid exchange remains, since a previous report has shown some differences between muscle actins from trout and rabbit (22).

The results of this study show an evolutionary development of cytoplasmic actins, smooth muscle actin, heart muscle actin, and skeletal muscle actin with two different genes expressed simultaneously for cytoplasmic and smooth muscle actins. Unless further actin species are found, the minimal number of actin genes expressed in a higher animal should be six.

**Note Added in Proof.** We have recently examined the amino acid sequence of the amino-terminal tryptic peptide of rabbit skeletal muscle actin. Instead of the sequence acetyl-Asp-Glu-Thr-Glu-Asp-Thr reported by others (5, 16–19), we find the sequence acetyl-Asp-Glu-Asp-Glu-Thr-Thr-. Thus the amino-terminal sequence of the cytoplasmic actins and gizzard actin (Fig. 1) should be compared to this revised sequence. The revised rabbit actin sequence has four contiguous negatively charged residues at the amino terminus, as compared to the three negatively charged residues at the amino terminus found here for both cytoplasmic actins and for gizzard actin.

We are indebted to Magda Puype for excellent technical assistance. We thank Dr. K. Zechel for the isoelectric focusing analysis of our actin preparations. J.V. is Bevoegdverklaard Navorsers at the Belgian National Fund for Scientific Research and was supported by a Short-Time EMBO Fellowship and also in part by the Alexander-von-Humboldt-Stiftung.

1. Pollard, T. D. & Weihing, R. R. (1974) *CRC Crit. Rev. Biochem.* **2**, 1–65.
2. Bray, D. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **37**, 567–571.
3. Pollack, R., Osborn, M. & Weber, K. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 994–998.
4. Gruenstein, E. & Rich, A. (1975) *Biochem. Biophys. Res. Commun.* **64**, 472–477.
5. Elzinga, M. & Lu, R. C. (1976) in *Contractile Systems in Non-Muscle Tissues*, eds. Perry, S. V., Margreth, A. & Adelstein, R. S. (Elsevier North-Holland Biomedical Press, Amsterdam), pp. 29–37.
6. Elzinga, M., Maron, B. J. & Adelstein, R. S. (1976) *Science* **191**, 94–95.
7. Whalen, R. G., Butler-Browne, G. S. & Gros, F. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2018–2022.
8. Garrels, J. I. & Gibson, W. (1976) *Cell* **9**, 793–805.
9. Spudich, J. A. & Watt, S. (1971) *J. Biol. Chem.* **246**, 4866–4871.
10. Lazarides, E. & Lindberg, U. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4742–4746.
11. Weber, K., Koch, R., Herzog, W. & Vandekerckhove, J. (1977) *Eur. J. Biochem.* **78**, 27–32.
12. Rubenstein, P. A. & Spudich, J. A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 120–123.
13. Izant, J. G. & Lazarides, E. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 1450–1454.
14. Vandekerckhove, J., Gielen, J. & Van Montagu, M. (1977) *J. Biol. Chem.*, **252**, 7761–7772.
15. Offord, R. E. (1966) *Nature* **211**, 591–593.
16. Bruton, C. J. & Hartley, B. S. (1970) *J. Mol. Biol.* **52**, 165–173.
17. Elzinga, M. & Collins, J. H. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **37**, 1–7.
18. Elzinga, M., Collins, J. H., Kuehl, W. M. & Adelstein, R. S. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 2687–2691.
19. Collins, J. H. & Elzinga, M. (1975) *J. Biol. Chem.* **250**, 5915–5920.
20. Bray, D. & Thomas, C. (1976) *J. Mol. Biol.* **105**, 527–544.
21. Lu, R. C. & Elzinga, M. (1977) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **36**, 898.
22. Bridgen, J. (1971) *Biochem. J.* **123**, 591–600.