THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 268, No. 29, Issue of October 15, pp. 21800-21810, 1993
Printed in U.S.A.

Skelemin, A Cytoskeletal M-disc Periphery Protein, Contains Motifs of Adhesion/Recognition and Intermediate Filament Proteins*

(Received for publication, June 9, 1993, and in revised form, June 24, 1993)

Maureen G. Price‡ and Richard H. Gomer§¶

From the Department of Biochemistry and Cell Biology and the §Howard Hughes Medical Institute, Rice University, Houston, Texas 77251-1892

In striated muscle, myofibrils are anchored to an interconnecting cytoskeleton of desmin intermediate filaments. Skelemin (195 kDa) may be a link between myofibrils and the intermediate filament cytoskeleton. Skelemin partitions with desmin to the insoluble cytoskeleton, and increases the thickness of reconstituted intermediate filaments. Concentrated at the M-disc periphery, skelemin may also contact myosin filaments. We used immunoscreening to isolate a mouse muscle cDNA which encodes a protein with a calculated molecular mass of 185 kDa. Anti-skelemin antibodies bound to the protein products of each of three nonoverlapping regions of the open reading frame. Antibodies directed against the protein products of each one-third of the cDNA react with a 195-kDa muscle protein and stain the M-disc indistinguishably from the original anti-skelemin antibodies, suggesting that the cDNA encodes skelemin. A single skelemin mRNA is detected in muscle but not non-muscle tissues, consistent with immunostaining results. Skelemin is a member of a family of myosin-associated proteins containing fibronectin type III and immunoglobulin superfamily C2 motifs. Skelemin is unique in this family in having intermediate filament core-like motifs, one near each terminus. We hypothesize that skelemin could interact with myosin or myosin-associated proteins through its fibronectin and/or immunoglobulin motifs, and with intermediate filaments through intermediate filament-like motifs.

Little is known about the proteins that direct the molecular morphogenesis of subcellular structures. One of the best systems for such studies is the semicrystalline sarcomere of striated muscle and the cytoskeleton attached to it. Sarcomeres are surrounded by a three-dimensional cytoskeleton of desmin intermediate filaments (for reviews, see Price and Sanger (1983), Price (1991), and Small *et al.* (1992)). These intermediate filaments form a ring around each Z-disc, link the Z-discs of adjacent sarcomeres, and additionally form longitudinal connections among the Z-discs in a myofibril. In some striated muscles, transverse filaments of unknown composition are attached to the M-discs (for review, see Price (1991)). The anchored lattice of desmin filaments may contribute to the limitation of cell distortion in both axes, to the maintenance of a regular architecture, and to the coordinated contraction/relaxation of mature myofibrils, all of which are important for muscle function (reviewed in Wang (1985)).

The exosarcomeric lattice is formed from the initially randomly arranged intermediate filaments of myogenic cells by a process of molecular remodeling and selective attachment (reviewed in Price (1991)). To examine the molecular basis of the interaction between intermediate filaments and myofibrils, we began an analysis of intermediate filament-associated proteins in mammalian striated muscle. One of the most prominent proteins which is co-purified with desmin in the intermediate filament cytoskeleton of bovine heart is skelemin. Skelemin is a protein that migrates on SDS-polyacrylamide gels as a band with an apparent molecular mass of 195 kDa and an isoelectric point near 5.0 (Price, 1987). Immunoblotting and immunofluorescence staining with polyclonal rabbit antibodies showed that skelemin is expressed in mammalian skeletal, cardiac, and smooth muscle cells, as well as in myoepithelial cells. It is absent in non-muscle cells other than certain glial cells in the brain.[1] Immunoblotting of extracts and residues shows that skelemin is co-fractionated with desmin throughout a series of extractions with seven different buffers, and like desmin, is relatively insoluble. Purified skelemin causes a 30% increase of the diameter of intermediate filaments reconstituted in their presence, as determined by electron microscopy (Price, 1984, 1987). These results led to the suggestion that skelemin associates with the desmin intermediate filament cytoskeleton.

Previously reported cytoskeletal proteins associated with the myofibril, including spectrins, ankyrin, synemin, and paranemin, are all found at the periphery of the Z-discs (for reviews, Small *et al.* (1992), Lazarides and Capetenaki (1986), Yang *et al.* (1990), and Price (1991)). In this location, these proteins could interact with the transverse desmin filaments wrapped around the Z-disc periphery as well as with the longitudinal desmin filaments. In contrast, skelemin is concentrated at the periphery of each M-disc as shown by immunofluorescence of muscle longitudinal and cross-sections, and by skelemin's partial extraction in a solution that solubilizes M-disc components (Price, 1987). From purification and from immunoblotting results, we estimated that there are roughly 200 molecules of skelemin per sarcomere.

The M-disc is the region in the middle of the myosin bundle, and thus the sarcomere, in which the bare zones of the bipolar myosin filaments are linked together by MM-creatine kinase and possibly also by other proteins (for discussions, see Thor-

* This work was supported in part by a Basic Research Grant from the Muscular Dystrophy Association and a grant from the John F. Smiekel Foundation. 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. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) Z22866.

‡ To whom correspondence should be addressed. Tel.: 713-527-4872; Fax: 713-285-5154.

¶ Assistant investigator of the Howard Hughes Medical Institute.

[1] M. G. Price, S. K. Malhotra, and S. Einheber, unpublished observations.

nell *et al.* (1990a) and Noguchi *et al.* (1992)). An intriguing question concerning the anchorage of the intermediate filament lattice to the myofibrils is whether the longitudinal desmin filaments, which must be over 2-μm long in order to link two consecutive Z-discs, are anchored at any point along the sarcomere. The partitioning of skelemin to the insoluble intermediate filament cytoskeleton and its ability to bind to desmin-containing filaments suggested that skelemin might anchor the desmin filaments to myosin or another M-disc component. Skelemin may also be associated with, or even comprise, the fragile, narrow filaments that link the M-discs of neighboring myofibrils in some muscles (Pierobon-Bormioli, 1981; Street, 1983; Wang and Ramirez-Mitchell, 1983), as suggested by anti-skelemin staining of thin strands linking isolated pairs of myofibrils (Price, 1987).

An insight into a possible binding site of skelemin came from an immunohistochemical study of mammalian heart development (Thornell *et al.*, 1990b). Skelemin was specifically localized at the M-discs of early (4 week bovine fetal) myofibrils, at the time when desmin-containing intermediate filaments are still in random arrays in the cytoplasm. At no stage of myogenesis were detectable amounts of skelemin found to be not associated with myofibrils or in M-disc striations appreciably broader than the mature ones. Several other M-disc components, myomesin (Grove *et al.*, 1984) and M-protein (Masaki and Takaiti, 1974), are first detected only after skelemin is already localized. These findings suggest that skelemin has a specific binding site at the periphery of the M-disc that is present in early myofibrils containing actin and myosin but lacking myomesin, M-protein, and attached desmin filaments.

Myosin is the major myofibrillar protein with which skelemin could be associated. There are about a dozen other proteins associated with the myosin-containing thick filament (for review, see Epstein and Fischman (1991)). These include, but are not limited to, the titin proteins of different animal species: *Caenorhabditis elegans* twitchin (Benian *et al.*, 1989), vertebrate titin (Labeit *et al.*, 1990, 1992), *Lethocerus indicus* p800 (Lakey *et al.*, 1990), and *Drosophila* projectin (Ayme-Southgate *et al.*, 1991; Fyrberg *et al.*, 1992), as well as C-protein (Einheber and Fischman, 1990), and 86-kDa protein (Fischman *et al.*, 1991), M-protein (Noguchi *et al.*, 1992), myosin light chain kinases of nonstriated muscle (Olson *et al.*, 1990; Shoemaker *et al.*, 1990; Gallagher *et al.*, 1991). Sequence analysis shows that these putative myosin-associated proteins contain two different types of ~90-residue motifs previously found only in secreted or cell-surface adhesion/recognition molecules. These motifs are the fibronectin type III motif (for reviews, see Hynes (1990), Ruoslahti (1988), and Odermatt and Engel (1989)), called motif I in the myosin-associated family, and the immunoglobulin superfamily C2 type motif (for reviews, see Hunkapiller and Hood (1989), Williams (1987), and Williams and Barclay (1988)), called motif II in the myosin-associated family (Benian *et al.*, 1989).

The cytoskeletal protein that skelemin appears to associate with is desmin, the muscle intermediate filament protein (Price, 1984, 1987). The members of the intermediate filament protein family also have a common motif, a helical core of ~310 residues, consisting of four closely linked α-helical regions (Fraser *et al.*, 1990; Steven, 1990). A striking feature of the α-helical regions is a heptad repeat, with the first and fourth residues being hydrophobic 75% of the time. The initial steps of polymerization occur through interactions of the α-helical core regions (Geisler *et al.*, 1992).

To gain insight into skelemin function, the deduced amino acid sequence was determined from mouse muscle cDNAs that were isolated by immunoscreening with polyclonal anti-skelemin antibodies (Price, 1987). Skelemin was found to be a member of the recently identified family described above, which includes cytoplasmic, myosin-associated proteins containing both fibronectin type III and immunoglobulin superfamily C2 type motifs. Unique in this family and consistent with its biochemical properties and localization, skelemin also contains desmin intermediate filament helical motifs. The motif content of skelemin suggests that skelemin might bind to myosin-containing thick filaments via fibronectin type III motifs and/or immunoglobulin superfamily C2 type motifs and to desmin filaments via desmin-like motifs.

MATERIALS AND METHODS

Isolation of Skelemin cDNAs and DNA Sequencing—Skelemin cDNA clones were identified by conventional immunoscreening of a λZAP SK$^+$ (Stratagene, La Jolla, CA) cDNA library of poly(A)$^+$ RNA from adult C57Bl/10J mouse hind limb skeletal muscle (a gift from Dr. Joel A. Pearlman, Baylor College of Medicine, Institute for Molecular Genetics). A total of approximately 1×10^6 clones were plated at 5×10^4 plaques/150-mm dish. Following Young and Davis (1983), nitrocellulose replica filters were incubated overnight at room temperature with polyclonal rabbit anti-skelemin serum (Price, 1987) at a 3000 × dilution in 150 mM NaCl, 10 mM Tris-HCl, pH 8.0, 3% non-fat milk, 0.5% bovine serum albumin, and 0.05% Tween 20. After washing in the above buffer, the filters were incubated with ^{125}I-protein A at 3.5×10^5 dpm/ml (ICN, Irvine, CA) in that buffer for 3 h, washed, and exposed to X-Omat AR5 film (Kodak, Rochester, NY) with a Cronex Lightning Plus intensifying screen (Du Pont), at −70 °C for 18–40 h. Plaques were picked and re-screened as above until clonal, typically three more times.

The cDNA inserts in pBluescript were autoexcised from λZAP with the R408 helper virus (Short *et al.*, 1988). Plasmid DNA isolation, subcloning, bacterial transformation by the $CaCl_2$ method, and restriction mapping were done by standard protocols (Sambrook *et al.*, 1989). Double-stranded DNA sequencing was performed on both strands by the dideoxy termination method with ^{35}S-dATP (Du Pont-New England Nuclear, Boston, MA) using modified T_7 DNA polymerase (U. S. Biochemical Corp.) following the manufacturer's directions. The 16–18-mer sequencing primers were synthesized by Rice University, Department of Biochemistry and Cell Biology, Oligos, Etc. (Guilford, CN), and Genosys (Houston, TX).

To obtain cDNA clones encoding the full-length of skelemin, approximately 1×10^6 clones of the poly(A)$^+$ cDNA library and approximately 1×10^6 clones of a random-primed mouse hind limb muscle cDNA library (also a gift from Dr. Joel A. Pearlman) were screened with oligonucleotides following Sambrook *et al.* (1989). The oligonucleotides (Oligos, Etc.) were 18-mer exact matches to sequences near the 5′ and the 3′ end of the longest cDNA clone that had been isolated by immunoscreening. The oligonucleotides were end-labeled with [γ-^{32}P]ATP using the Prime-It kit (Stratagene), and hybridized at an optimal temperature following McGraw *et al.* (1990). To determine the amount of additional skelemin cDNA in these clones relative to the longest original skelemin cDNA, PCR[2] using Taq polymerase (Perkin-Elmer Cetus, Norwalk, CT) was done with primers flanking the vector insertion site and with primers corresponding to sequences from the longest original clone.

Sequence Analysis—The predicted amino acid sequence of skelemin was compared with the PIR release 33 (Protein Information Resource) and SWISS-PROT release 23.0 by the FASTA (Pearson and Lipman, 1988) and SEARCH (Lawrence and Goldman, 1988) programs. Alignment of the immunoglobulin C2 and fibronectin type III motifs was approximated with the Treealign (Hein, 1990) and the Altschul and Erickson (1986) programs, and optimized by visual inspection. After the motifs had been defined, the comparison searches against the two protein data bases were repeated using defined portions of the skelemin amino acid sequence. Isoelectric points were analyzed by the CHARGEPRO program (Intelligenetics, Mountain View, CA) and the possible sites of post-translation modification in the predicted amino acid sequence of skelemin were suggested by the PROSITE program, release 9.2 (Bairoch, 1989). Structural predictions were made with several standard algorithms

[2] The abbreviations used are: PCR, polymerase chain reaction; kb, kilobase; N-CAM, neural cell adhesion molecule.

(Chou and Fasman, 1978; Garnier *et al.*, 1978). Hydropathy, hydrophilicity, and hydrophobic moment profiles were obtained by standard algorithms (Eisenberg *et al.*, 1984; Hopp and Woods, 1983; Kyte and Doolittle, 1982).

Northern Blots—Total RNA was isolated from the liver, limb skeletal muscle, heart, and whole brain of adult mice, from skeletal muscle of the waterbug *L. indicus*, and from COS cells, baby hamster kidney cells, and *Dictyostelium* vegetative cells by the guanidinium thiocyanate-phenol-chloroform single-step extraction method (Chomczynski and Sacci, 1987). RNA molecular weight standards (Life Technologies Inc.) and ~5 μg of each RNA were electrophoresed on a 1.25% agarose, 2.3 M formaldehyde gel. Gels were stained with ethidium bromide (Sambrook *et al.*, 1989) and photographed prior to transfer. Gels were then blotted to Duralon UV membrane (Stratagene) with 10 × SSC. Membranes were incubated with 1×10^5 cpm/ml of the 0.63-kb *Eco*RI fragment of skelemin cDNA labeled with [α-^{32}P]dCTP by the random-hexamer method (Boehringer Mannheim). Hybridization was done in a 10% polyethylene glycol-containing solution (Amasino, 1986) at 42 °C. The membranes were washed twice for 15 min in 2 × SSC, 0.1% SDS, then one to three times with 0.1 × SSC, 0.1% SDS at room temperature. To verify that equal amounts of each sample were loaded and that there was no apparent degradation of the ribosomal RNA, the membranes were stained with methylene blue (Monroy, 1988) after autoradiographs had been obtained. Autoradiography was done with preflashed X-Omat AR5 film (Kodak) with a Cronex Lightning Plus intensifying screen (Du Pont), at −70 °C for 5–40 h.

Bacterial Expression of Defined Regions of Skelemin—The original 3.84-kb skelemin cDNA fragment was expressed as a fusion protein with 121 residues of β-galactosidase. Log-phase XL1-Blue (Stratagene) bacteria transformed with pBluescript (Short *et al.*, 1988) containing the 3.84-kb in-frame skelemin cDNA were treated with 10 mM isopropyl-β-D-thiogalactopyranoside for 20 min. Controls were untransformed bacteria and those transformed with pBluescript alone or with a 2.7-kb mitoskelin cDNA insert that had been isolated using our anti-mitoskelin antibodies (Price and Gomer, 1989). The bacteria were harvested by centrifugation and boiled in Laemmli (1970) sample buffer.

Smaller regions of skelemin were also expressed as fusion proteins, with 22 residues of the bacterial ompT outer membrane protein at their amino termini (Grodberg *et al.*, 1988). Three coding regions representing roughly the amino, middle, and carboxyl one-third portions of skelemin were prepared from the 5.612-kb cDNA (see Fig. 1) by PCR using Taq polymerase (Perkin-Elmer Cetus) with pairs of oligonucleotide primers that incorporated a *Sal* site in the sense primer, and either a *Sal*I or a *Bam*HI site as well as a stop codon in the antisense primer. For the amino one-third of skelemin, the primers were 5′-GAGTCGACCATGTCTCTGCCCTTTTATCAG-3′ and 5′-GAGTCGACTTAGGGTGCTGAAGGATGACT-3′. This PCR product was inserted into pET 12 via the 5′ *Sal*I site and a *Bam*HI site that is 40 bases upstream of the antisense primer, yielding a fusion protein that began at residue 1 and ended at residue 584. For the middle one-third of skelemin, the primers were 5′-GAGTCGAC CGTCATTCCTGGCCCCCCC-3′ and 5′-GAGTCGACTTATGGA GGTGTTTCAGCCAC-3′, yielding a fusion protein that began at residue 615 and ended at residue 1120. For the carboxyl one-third, the primers were 5′-GAGTCGACCGGCACCAAAGAGGTTGTG-3′ and 5′-GAGGATCCTTACTTGGACTTCTGGTTGCC-3′, yielding a fusion protein that began at residue 1121 and ended at residue 1666. The pBluescript vector containing the 5.612-kb skelemin insert was denatured at 95 °C for 5 min, then Taq polymerase was added for 25 cycles of PCR with 1 min denaturation at 95 °C, 1 min of annealing at 55 °C, and 2.5 min of extension at 72 °C. PCR products were gel purified with Gene-Clean (BIO-101, Inc., La Jolla, CA). Due to the ineffective cleavage of *Sal*I sites near the ends of DNA, the PCR products were initially ligated into the pBluescript vector that had been restriction-digested with *Eco*RV and then ddT-tailed (Holton and Graham, 1991). PCR products recovered by restriction digestion were subsequently ligated into the pET 12a bacterial expression vector (Novagen, Madison, WI) that had been restriction digested with either *Sal*I or with both *Sal*I and *Bam*HI. Constructs were selected and amplified in DH5α bacteria. Expression of the skelemin regions in BL21(DE3)pLysS (Novagen, Madison, WI) bacteria was induced for 3 h with 0.4 mM isopropyl-β-D-thiogalactopyranoside (Studier *et al.*, 1990). Immunoblotting with the original anti-skelemin antibodies (Price, 1987) diluted 2,000 × was done to determine the best methods and colonies for expression of skelemin regions of the correct molecular weight. The induced skelemin proteins were purified by excision from preparative 10% SDS-polyacrylamide gels and electroelution (Hunkapiller *et al.*, 1983). The amount of recovered protein was estimated by comparison with a gradient of known amounts of bovine serum albumin on SDS-polyacrylamide gels (Laemmli, 1970).

Preparation of Antibodies Against Recombinant Skelemin Regions—Antibodies against the three nonoverlapping regions of skelemin (see Fig. 1) were prepared in female New Zealand White rabbits by Cocalico Biologicals, Inc. (Reamstown, PA) following previously described methods for producing the original anti-skelemin antibodies (Price, 1987). For the first immunization, 5 μg were injected into the popliteal lymph nodes and 15–20 μg were injected subcutaneously and intramuscularly (Goudie *et al.*, 1966). At 4 weeks, 100–125 μg of antigen was injected via only the latter two routes. At 5 weeks, sera were found to contain specific antibodies as assayed by immunoblotting and immunofluorescence of skeletal muscle. Serum was prepared 3 more times at 10-day intervals. Antibodies were partially purified by precipitation with 50% ammonium sulfate, dialysis against 10 mM $NaPO_4$ buffer, and clarification by centrifugation.

Immunoblotting and Immunofluorescence—Immunoblotting of proteins electrophoresed on 10% polyacrylamide-SDS gels was done as previously described (Price, 1987). The blots were stained with ammonium sulfate-purified anti-skelemin antibodies at 2,000 × dilution or with anti-skelemin region antibodies at 500–750 × dilutions, and then with ^{125}I-protein A (Du Pont-New England Nuclear) as previously described (Price, 1987).

For immunofluorescence, thin strips of mouse hind limb skeletal muscle were taken from an adult BALB/C mouse that had been sacrificed by cervical dislocation. Muscle strips were stretched and plunged into liquid nitrogen, then placed on a layer of O.C.T. embedding compound (Miles, Elkhart, IN) in a mold set on dry ice. After filling the mold with O.C.T, the specimen was quick-frozen by immersion in liquid nitrogen. Longitudinal and cross-sections of 4–8 μm thickness were cut with a Accu-Edge blade (Miles, Elkhart, IN) in an IEC cryotome (Needham Heights, MA) and picked up on glass coverslips. Sections were fixed for 10 min in 3.7% formaldehyde in phosphate-buffered saline (137 mM NaCl, 3 mM KCl, 9 mM Na_2HPO_4, 1 mM KH_2PO_4, pH 7.45), soaked for 10 min in 0.01 M glycine to quench the fixative, soaked in several changes of phosphate-buffered saline, 0.05% Nonidet P-40, and stained for 1 h at room temperature in dilutions (between 2 × and 40 ×) of ammonium sulfate-purified antibodies directed against whole bovine skelemin or against defined regions of mouse skelemin. Following three 15-min washes in phosphate-buffered saline/Nonidet P-40, the sections were stained for 45 min at room temperature with a 200 × dilution of affinity-purified fluorescein-labeled goat anti-rabbit IgG $F(ab)_2$ (Cappel/Organon Technika, Durham, NC) containing 1 unit/100 μl of rhodamine-labeled phalloidin (Molecular Probes, Eugene, OR), washed again, and mounted in 1 mg/ml *p*-paraphenylene diamine in 90% glycerol, pH 9 (Johnson and Araujo, 1981). Phase-contrast/epifluorescence microscopy was done with a Nikon Microphot FX (Nikon, Melville, NY) with a 60 × Nikon (Garden City, NY) or a 63 × oil-immersion lens (Carl Zeiss, Thornwood, NY). Photographs were taken on T-MAX 3200 film (Kodak) and developed with D-76 (Kodak, Rochester, NY) following the manufacturer's instructions.

Skelemin Isolation and Peptide Sequencing—The skelemin-enriched urea extract of the cytoskeletal residue of bovine myocardium was prepared by sequential extraction in six different salt solutions as described (Price, 1984, 1987) except that the extractions were completed in 3 days. Proteins separated on 10% SDS-polyacrylamide gels were transferred to a polyvinylidene difluoride-derivatized Immobilon membrane (Millipore, Bedford, MA) (Yuen *et al.*, 1989) and stained with 0.1% amido black as described (Jain *et al.*, 1992). Skelemin bands, pooled to give 200–400 μg (1–2 nmol), other bands, and a blank (control) strip of membrane were excised, and aliquots of each were labeled with anti-skelemin antibodies (Price, 1987) for positive identification. The NH_2-terminal sequence of whole skelemin could not be obtained by standard methods, presumably due to blockage, so proteolytic peptides were prepared. Following described methods (Jain *et al.*, 1992) based on a combination of several protocols (Aebersold *et al.*, 1987; Bauw *et al.*, 1989), membrane strips were treated with 0.5% polyvinylpyrrolidone (40-kDA) to prevent protease binding, and then proteins were digested *in situ* with 10 μg/ml L-1-tosyl-amido-2-phenylethyl chloromethyl ketone-treated bovine trypsin (Boehringer Mannheim) in 0.1 M Tris-HCl, pH 8.5, at 37 °C overnight. Solubilized tryptic fragments were separated by reverse-phase high performance liquid chromatography (model 2350, Isco, Lincoln, NE) on a 4.6 × 250-mm 5-μm bead size C8-type column

(Dynamax 300A, Rainin, Woburn, MA), using a gradient from 0.1% trifluoroacetic acid/water to 0.08% trifluoroacetic acid/acetonitrile. Peptide sequencing was done at the Baylor College of Medicine Protein Sequencing Facility with an Applied Biosystems (Foster City, CA) Model 477A gas-phase protein sequenator equipped with an in-line model 120-A phenylthiohydantoin analyzer.

RESULTS

Isolation of Skelemin cDNA Clones—To examine the primary structure of the skelemin protein, we isolated and sequenced skelemin cDNA (Fig. 1). Five strongly immunoreactive clones were identified by anti-skelemin immunoscreening of 10^6 clones from a λZAP expression library of adult mouse skeletal muscle cDNA. The largest clone contains a 3.84-kb cDNA insert that is entirely open reading frame, as determined by DNA sequencing (Fig. 1, from 875 to 4716 bases). A protein migrating with the approximate molecular mass of 180 kDa expressed by bacteria transformed with this construct was detected by anti-skelemin antibodies but not by other polyclonal antibodies (Price and Gomer, 1989) or by preimmune serum (data not shown). Screening the two cDNA libraries with oligonucleotides that hybridize near the ends of the 3.84-kb cDNA yielded five different additional clones. A 5.612-kb cDNA was shown by PCR, restriction mapping, and sequencing to overlap with the 3.84-kb cDNA (Fig. 1).

The complete nucleotide sequence of this cDNA and the predicted amino acid sequence of skelemin are shown in Fig. 1. The 104 bases of 5′-untranslated region include 2 stop codons, in 2 different reading frames. The first ATG codon was identified as the start codon by the similarity of the region containing it to the Kozak consensus sequence for initiation of translation (Kozak, 1991). This methionine codon is within a sequence with 7 of 13 bases identical to the Kozak consensus sequence, including the critical −3 position being a purine. There is a single large open reading frame of 4998 bases ending with a TGA stop codon. The putative 3′-untranslated region of 510 bases includes an in-frame stop codon 66 bases downstream and out-of-frame stop codons at 20 and 90 bases downstream, in addition to numerous other stop codons. It also contains a single AATAA consensus recognition sequence for polyadenylation approximately 20 bases upstream from the poly(A) tail, the most common position relative to the poly(A) tail (Proudfoot, 1991).

Confirmation That the cDNA Encodes Skelemin—Indication that the open reading frame of the 5.612-kb cDNA encodes skelemin was obtained from sequencing skelemin peptides. Several peptides from purified bovine myocardial skelemin were found to match portions of the deduced amino acid sequence of mouse skelemin; two overlapping peptides form a 14/15 residue match (Fig. 1). The mismatches could be due to species and/or muscle-type differences.

To additionally confirm that the 5.612-kb cDNA encodes skelemin, polyclonal antibodies directed against three nonoverlapping regions of the open reading frame were produced (see Fig. 1). The approximate amino, middle, and carboxyl one-third regions of the open reading frame were expressed as proteins of 67, 55, and 60 kDa, respectively, each as about 30% of the total bacterial protein (Fig. 2). The original anti-skelemin antibodies, which are directed against bovine myocardial skelemin (Price, 1987), reacted with the protein regions encoded by each one-third of the cDNA (Fig. 2). Conversely, the antibodies directed against the nonoverlapping amino, middle, and carboxyl thirds of the protein product of the cDNA each reacted specifically with the 195-kDa skelemin of mouse skeletal muscle, as do the original anti-skelemin antibodies (Fig. 3). Each of the antibodies to the nonoverlapping protein regions encoded by the cDNA reacts only with whole skelemin of muscle or with the original antigen, not with the other regions of the protein encoded by the cDNA (data not shown).

The original anti-skelemin antibodies and all three antibodies directed against the nonoverlapping protein regions encoded by the cDNA likewise stain a narrow band overlying the M-disc in longitudinal sections of mouse skeletal muscle (Fig. 4). The antigen stained with each of the antibodies appears to be concentrated at the periphery of the myofibrils, as indicated by the pattern of stained striations going in and out of focus as the plane of focus is changed. Staining of muscle cross-sections with each of the antibodies directed against the nonoverlapping protein regions encoded by the cDNA shows rings of approximately 0.7–1.7 μm diameter, the diameter of myofibrils. This suggests that the protein regions recognized by each of these three antibody preparations is concentrated at the myofibril periphery (Fig. 5).

We have thus observed that antibodies directed against the amino, middle, and carboxyl nonoverlapping thirds of the protein encoded by the 5.612-kb cDNA all stain a protein with the molecular weight and the subcellular distribution in muscle of skelemin, and that the original anti-skelemin antibodies recognize all three thirds of the encoded protein. This strongly indicates that the 5.612 cDNA encodes skelemin. Furthermore, this cDNA appears to encode the entire skelemin molecule, since the open reading frame would encode 1,666 amino acids having a calculated molecular weight of 185.215. This is 95% the size of the 195-kDa skelemin band detected in immunoblots of mammalian muscle (see Fig. 3). Most muscle proteins migrate with apparent molecular weights greater than the molecular weights predicted by the amino acid sequence (see Bryan *et al.* (1989)).

Skelemin Transcript Is Expressed in Muscle Tissue—A single skelemin mRNA of ~5.8 kb that is present in heart muscle and skeletal muscle tissue but not in liver or whole brain of the adult mouse was detected by labeling blots of total RNA with a skelemin cDNA fragment (Fig. 6). The largest skelemin cDNA that we have isolated (Fig. 1) has an exceptionally short 5′-untranslated region (0.1 kb) and by comparison to the skelemin mRNA it appears to lack about 200 bases of 5′-untranslated region. An ~5.8-kb mRNA identical in size to that in the mouse muscle tissues was also detected in chicken skeletal muscle (Fig. 6). Similar amounts of RNA from the various mouse tissues and chicken muscles were loaded, as judged from the 18 S and 28 S rRNA detected with ethidium bromide and methylene blue (Monroy, 1988) staining, but only about 10% as much probe hybridized to the chicken skelemin transcript. The skelemin cDNA probe did not hybridize to any transcript in the muscle of an invertebrate, the water beetle *L. indicus*, nor to a transcript in COS cells, baby hamster kidney cells, or the motile vegetative cells of the simple eukaryote *Dictyostelium discoideum* (not shown). The results suggests that there is a single skelemin transcript, and that the muscle-specific expression pattern of skelemin mRNA is similar to that previously observed for the skelemin protein (Price, 1987).

Skelemin Contains Repeats of Three Different Motifs including Adhesion Protein Motifs and Intermediate Filament Motifs, as well as Two Unique Regions—Isoelectric focusing of bovine skelemin indicated that it has a pI of 5.0 (Price, 1984). The predicted amino acid sequence of mouse skelemin contains slightly more acidic (13.8%) than basic (12.3%) residues, yielding a predicted isoelectric point of 6.08. There are approximately equal percentages of nonpolar (37.1%) and polar (31.9%) residues. Serine is the most abundant amino acid, comprising 9.3% of the total residues. Proline comprises 5.4%

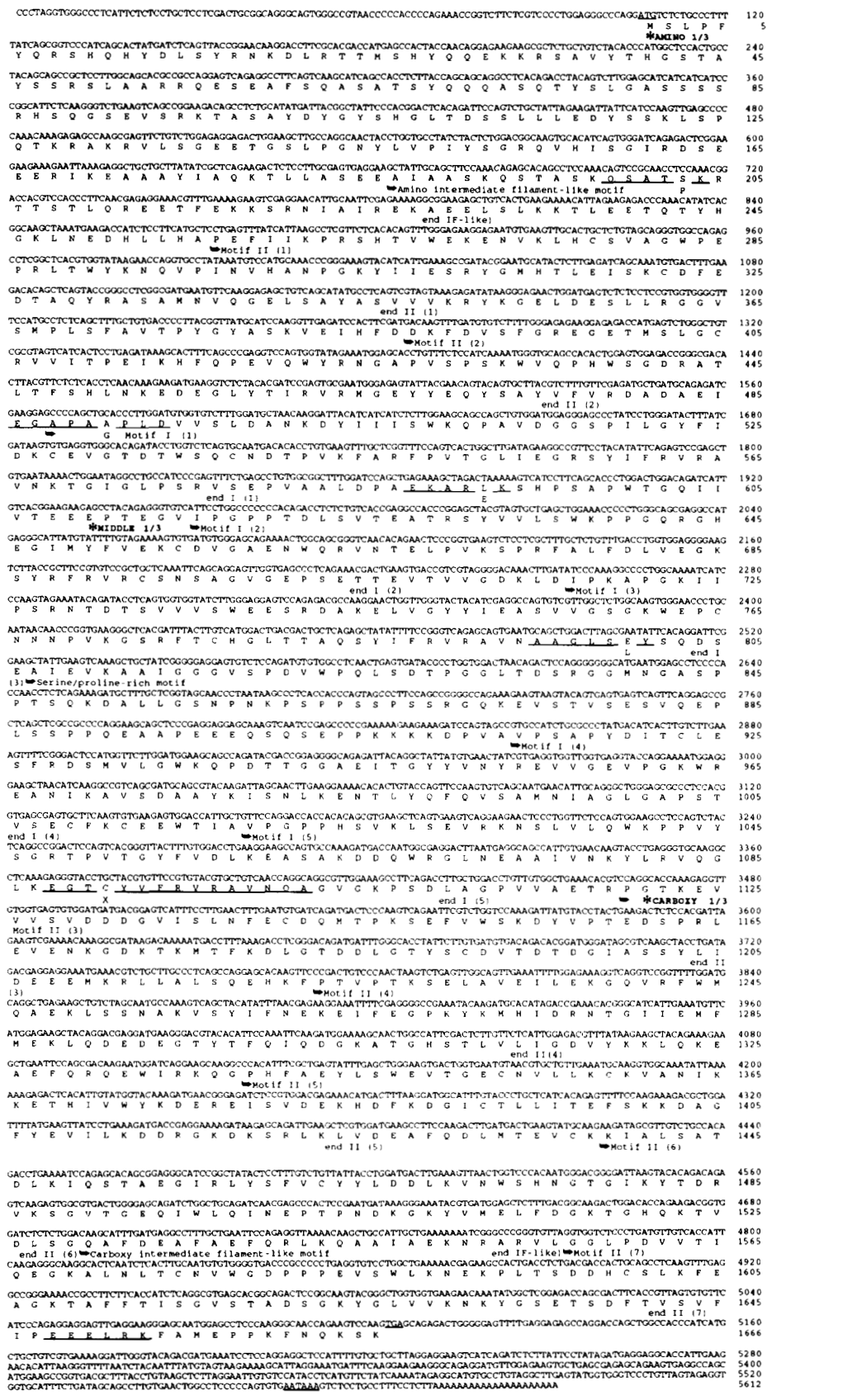

FIG. 1. **Complete nucleotide sequence and the predicted amino acid sequence of skelemin cDNA isolated from a mouse skeletal muscle cDNA library.** The putative ATG start codon, TGA termination codon, and polyadenylation site are *underlined*. Amino acid sequences confirmed by sequencing of bovine myocardial skelemin tryptic polypeptides are *underlined*, with the residues that differ from the predicted sequence shown below (*X* indicates that a cysteine cannot be detected by sequencing). The beginning of each nonoverlapping ~one-third region of skelemin used for production of polyclonal antibodies is indicated by an *asterisk*. The initial residue of each intermediate filament-like motif, motif I (immunoglobulin superfamily C2-like) and motif II (fibronectin type III-like), and of the serine/proline-rich region is indicated with an *arrow*; the last residue of each is also indicated (*end*). This sequence data is available from EMBL/GenBank/DDBJ under accession number Z22866.

of the total residues. The hydrophobic moment is negligible, and there are no strongly hydrophobic regions longer than about 15 residues. Secondary structure prediction algorithms suggest that skelemin is largely a β-sheet structure interrupted by short helices, turns, and random coils (Chou and Fasman, 1978; Garnier *et al.*, 1978). The PROSITE program (Bairoch, 1989) identifies 74 potential sites for phosphorylation in the predicted skelemin molecule, including those dependent on cAMP, protein kinase C, casein kinase II, and tyrosine kinase. The only other suggested post-translational modifications are

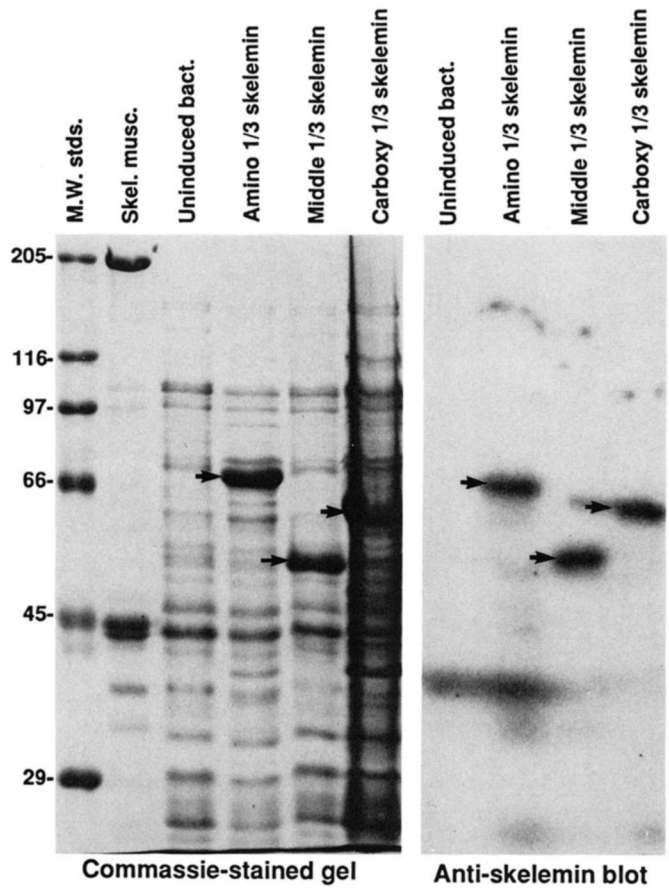

FIG. 2. **Three nonoverlapping protein regions encoded by the isolated cDNA are recognized by polyclonal antibodies directed against whole skelemin.** The Coomassie Blue-stained SDS-polyacrylamide gel shows total protein of bacteria that expressed the amino one-third (67-kDa), middle one-third (55-kDa), and carboxyl one-third (60-kDa) regions of the isolated cDNA (*arrows*). Molecular mass markers (in kDa) are indicated on the *left*. The corresponding blot stained with anti-skelemin antibodies shows specific labeling of each of the nonoverlapping regions (*arrows*).

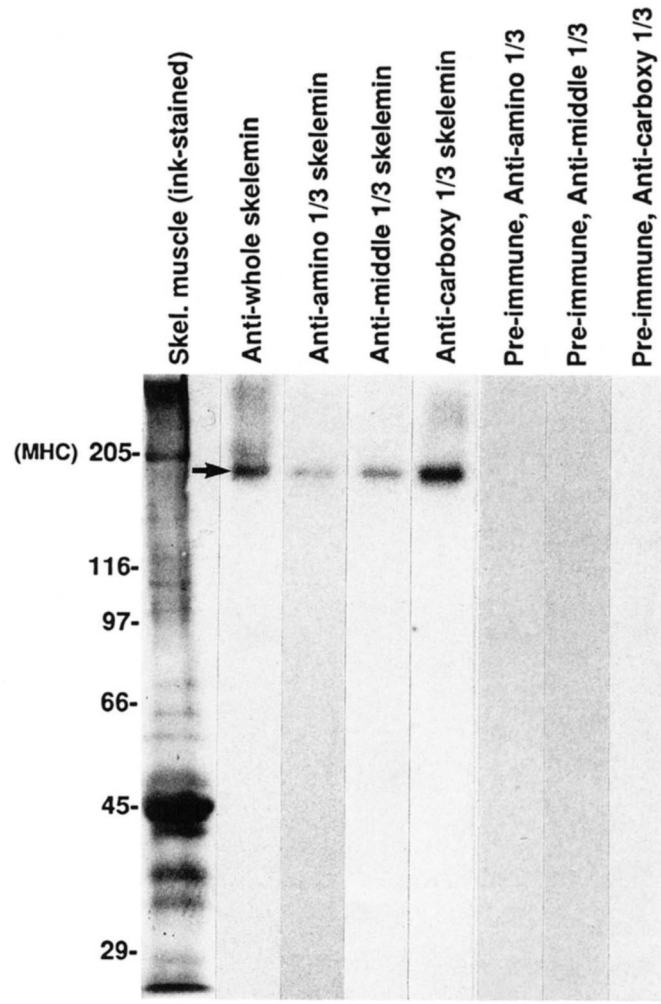

FIG. 3. **Polyclonal rabbit antibodies directed against each of the three nonoverlapping regions encoded by the isolated cDNA specifically label only the 195-kDa skelemin band (*arrow*) in immunoblots of mouse skeletal muscle.** This staining is comparable to that obtained with the polyclonal anti-skelemin antibodies. The ink-stained blot shows the position of the major skeletal muscle proteins. Myosin heavy chain (*MHC*, 205-kDa) and molecular mass markers (in kDa) are indicated on the *left*.

unlikely since skelemin is a cytoplasmic protein: glycosylation at asparagine in 4 potential sites and myristoylation at sites other than the amino-terminal residue.

Skelemin was found to consist largely of three different protein motifs and two unique regions as determined by comparison of the deduced amino acid sequence with those of proteins in the data banks (Fig. 7). There is a unique region at the amino terminus, short intermediate filament (desmin) helical core-like motifs near either end (Fig. 8*C*), five fibronectin type III-like motifs (Fig. 8*A*), seven immunoglobulin superfamily C2-like motifs (Fig. 8*B*), and a unique serine/proline-rich region in the center of skelemin. Between 7 and 36 residues are interspersed between the identified motifs. In addition, there is a highly (43%) charged 21-residue sequence at the carboxyl terminus.

Skelemin's amino-terminal region of 186 residues and the extremely serine/proline-rich central region of 107 residues (Figs. 1 and 7) are not significantly identical to any protein in the data bases. The amino-terminal has a predicted isoelectric point (10.4) far more basic than that of the entire skelemin molecule. This region is predicted to be more hydrophilic, have less β-sheet content, and to have a higher concentration of potential phosphorylation sites than other skelemin regions. It has an unusually high serine content, 18%, compared to the repeated motifs containing 4–11% serine (Fig. 1). The serine/proline-rich central region is 17% serine and 16% proline, and has near its end two repeats of the same charged amino acids (Fig. 1).

The presence of multiple ~90-residue long motifs that are significantly identical to either the fibronectin type III motif (Hynes, 1990; Odermatt and Engel, 1989; Ruoslahti, 1988) or to the immunoglobulin superfamily C2 motif (Hunkapiller and Hood, 1989; Williams, 1987; Williams and Barclay, 1988; Springer, 1990) places skelemin in a recently defined family of cytoplasmic proteins that resemble neural cell adhesion molecules (N-CAMs; Cunningham (1988), Edelman and Crossin (1991), and Hortsch and Goodman (1991) for reviews) and may all be associated in some way with myosin-containing thick filaments (Epstein and Fischman (1991) and Trinick (1992), for reviews). The fibronectin type III-like motifs in these cytoplasmic proteins have been called motif I and the immunoglobulin superfamily C2-like motifs have been called motif II (Benian *et al.*, 1989). The predicted skelemin amino acid sequence contains regions between 63 and 200 residues long with between 26 and 34% identity to the motifs I or motifs II of twitchin (Benian *et al.*, 1989), C-protein (Einheber and Fischman, 1990), titin (Fritz *et al.*, 1991), smooth muscle and non-muscle myosin light chain kinase (Olson *et al.*, 1990;

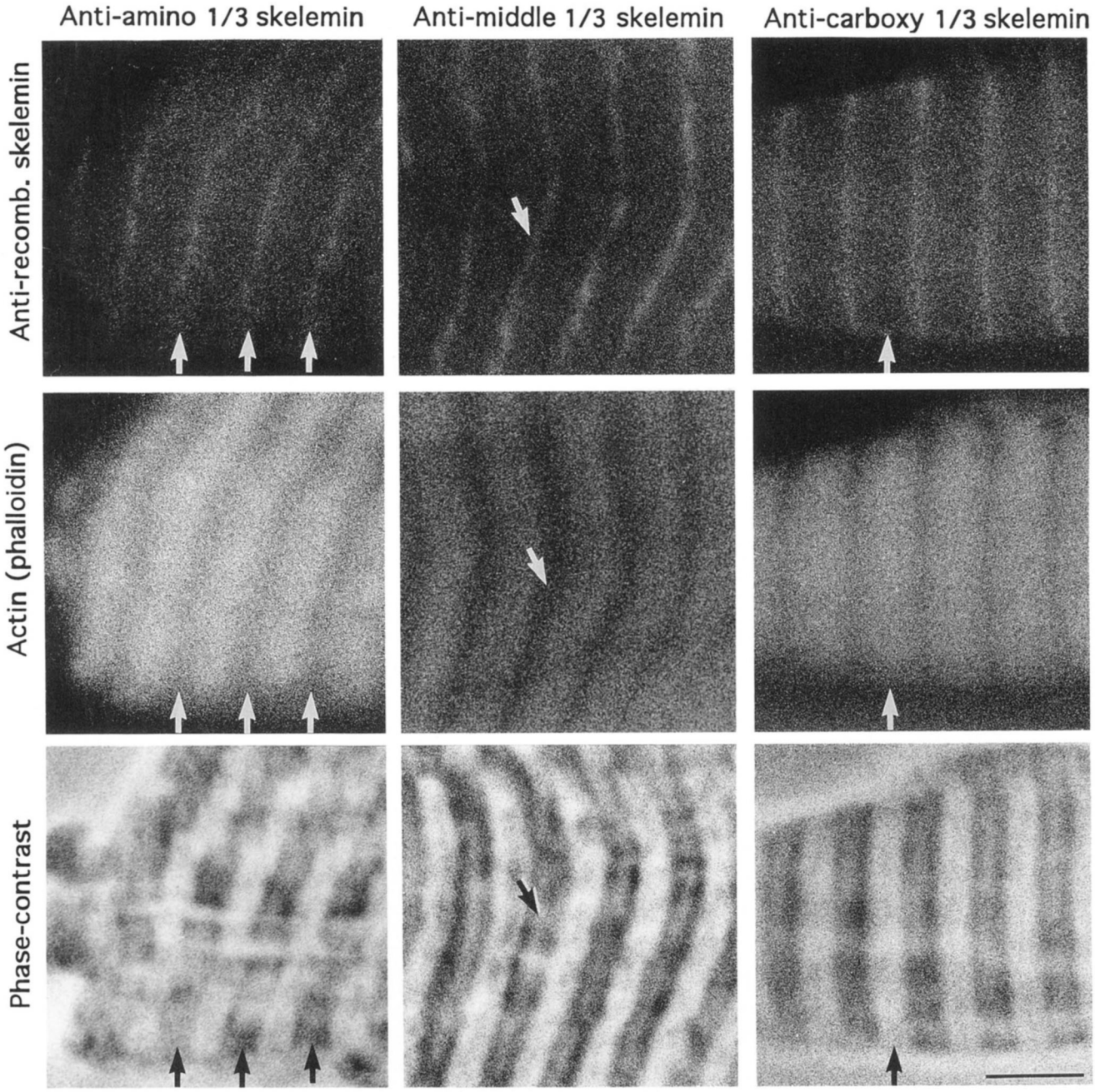

FIG. 4. **Specific immunofluorescence staining of the M-discs in longitudinal sections of mouse skeletal muscle was obtained with polyclonal antibodies directed against each of the three nonoverlapping protein regions encoded by the putative skelemin cDNA.** The *arrows* in each vertical column of 3 micrographs indicate the same structure: M-discs in the *left* and central column and the Z-line center of an actin-containing I-band in the *right* column. The *middle row* shows the broad bands of phalloidin-stained actin filaments that alternate with the skelemin-containing M-discs. The phase-contrast micrographs show the phase-dense myosin-containing A-bands, each of which is bisected by an M-disc (*arrows* in left column), and the phase-light actin-containing I-bands. *Bar* is 5 μm.

Shoemaker *et al.*, 1990; Gallagher *et al.*, 1991), and to the similar motifs of extracellular adhesion molecules such as N-CAM (Cunningham *et al.*, 1987; Hemperly *et al.*, 1986; Krieg *et al.*, 1989), N-CAM-L1 (Moos *et al.*, 1988), neuroglian (Bieber *et al.*, 1989), fasciclin II (Harrelson and Goodman, 1988), amalgam (Seeger *et al.*, 1988), and β4-integrin (Kajiji *et al.*, 1989; Suzuki and Naitoh, 1990), as well as to those of members of a recently defined family of receptor-linked protein tyrosine phosphatases with extracellular domains highly resembling N-CAMS: LAR and DLAR (Fisher *et al.*, 1991; Streuli *et al.*, 1989). The skelemin motifs are more like those of members of the cytoplasmic family of N-CAM-like proteins than those of the extracellular proteins of the immunoglobulin superfamily.

The five motif I and seven motif II repeats of skelemin were aligned with minimum gaps, and a consensus for each motif was devised using residues that are identical in more than half the motifs (Fig. 8, *A* and *B*). The skelemin consensus motif I and II were compared to a master consensus motif compiled from residues that are identical in more than half the consensus motifs I and II of six of the N-CAM-like putative myosin-associated proteins: twitchin (Benian *et al.*, 1989), rabbit skeletal muscle titin (Labeit *et al.*, 1990), projectin (Ayme-Southgate *et al.*, 1991), chicken C-protein (Einheber and Fischman, 1990), chicken smooth muscle myosin light

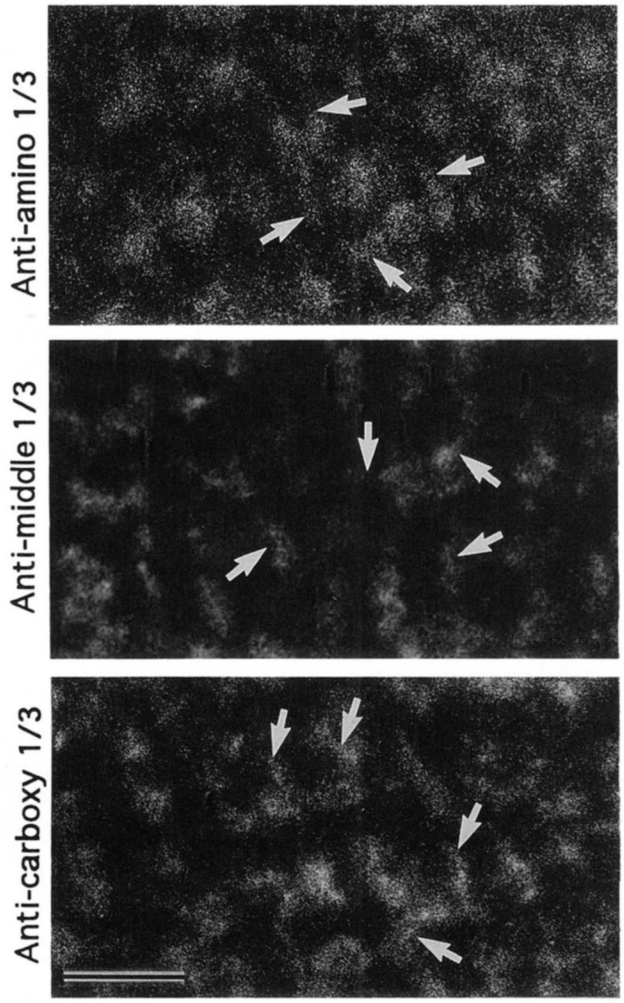

FIG. 5. **Specific immunofluorescence staining of the M-discs in cross-sections of mouse skeletal muscle was obtained with polyclonal antibodies directed against each of the three non-overlapping protein regions encoded by the putative skelemin cDNA.** *Arrows* indicate the staining at the periphery of a myofibril. *Bar* is 5 μm.

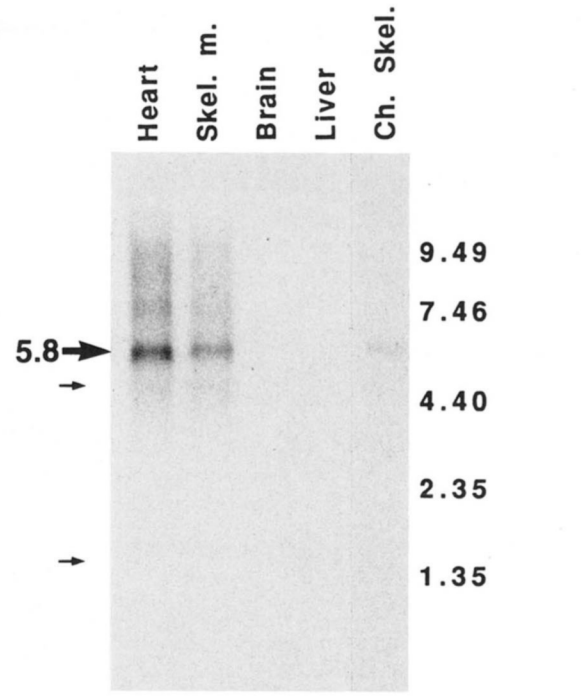

FIG. 6. **A skelemin transcript is expressed in muscle tissues.** This Northern blot containing equal amounts of total RNA isolated from the indicated mouse tissues (*heart, skel. muscle, brain,* and *liver*) and from chicken skeletal muscle (*Ch. skel.*) was probed with the 0.63-kb *Eco*RI skelemin fragment (bases 4084–4717). Mammalian and avian muscle tissues contain a ~5.8-kb skelemin mRNA (*large arrow*). The positions of the 28 S rRNA and 18 S rRNA (*small arrows, left*) and RNA molecular weight standards (*right*, in kb) are indicated.

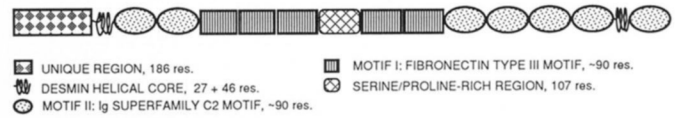

FIG. 7. **Schematic structural map of skelemin.** Shown are the linear position of the unique region, the two repeated motifs, the intermediate filament-like motifs, and the serine/proline-rich region.

chain kinase (Shoemaker *et al.*, 1990), and chicken 86-kDa protein (Vaughn *et al.*, 1993).

The five motif I repeats of skelemin, which are 90–95 residues long, are arranged in a tandem pair and triplet, as is found in other cytoplasmic N-CAM-like proteins (Benian *et al.*, 1989; Ayme-Southgate *et al.*, 1991; Fyrberg *et al.*, 1992; Labeit *et al.*, 1990; Trinick, 1992). They are 30–43% identical to one another (when only D and E are considered identical) and up to 46% similar to one another when conservative substitutions of V for I (Dayhoff *et al.*, 1978) are also considered. Thirteen residues are conserved and residues at 35 other positions are identical or are D/E conservative substitutions in three or four of the five motif I repeats in skelemin. The consensus motif I of the myosin-associated proteins can be aligned with the skelemin motif I consensus at 85% of the conserved residues and at 67% of the residues that are conserved in four of five skelemin motif I repeats. A 24-residue region (Fig. 8*A*) near the carboxyl-terminal of the motif I repeats of both skelemin and the myosin-associated proteins (Epstein and Fischman, 1991) is most similar to the type III repeat of fibronectin (Kornblihtt *et al.*, 1985). The two and three tandem repeats of motif I in skelemin provide the longest regions of high identity to a wide variety of proteins containing fibronectin type III-like motifs.

Seven 88–93-residue long motifs of skelemin have significant similarity to immunoglobulin superfamily C2 motifs; four of these motif II repeats form a tandem array near the carboxyl-terminal of skelemin (Fig. 7). The skelemin motif II repeats are less alike than are the motif I repeats, being up to 19% identical and up to 33% similar when conservative substitutions of hydrophobic residues are considered (Dayhoff *et al.*, 1978). Only 2 residues are conserved in the skelemin motif II repeats (Fig. 8*B*), as compared to there being no conserved residues in the 25 motif II repeats of twitchin (Benian *et al.*, 1989) and 3 residues conserved in the 6 motif II repeats of C-protein (Einheber and Fischman, 1990), one being the conserved tyrosine in the skelemin motif II repeats. Two short regions of the motif II repeats are the most identical, as has been noted in other members of the immunoglobulin superfamily (Fig. 8*C*). Four of the skelemin motif II repeats have 2 or 3 cysteines, but it is unlikely that disulfide bonding occurs in the cytoplasm.

Two intermediate filament-like motifs are found in the predicted skelemin molecule, one near either terminus. Immediately following the unique 186 amino-terminal region are 46 residues with 28% identity and 41% similarity to the carboxyl ~50% of helix 1b of chicken desmin (Geisler and Weber, 1982) (Fig. 8*C*). Near the carboxyl-terminal of skelemin there are 27 residues that are 22% identical and 33%

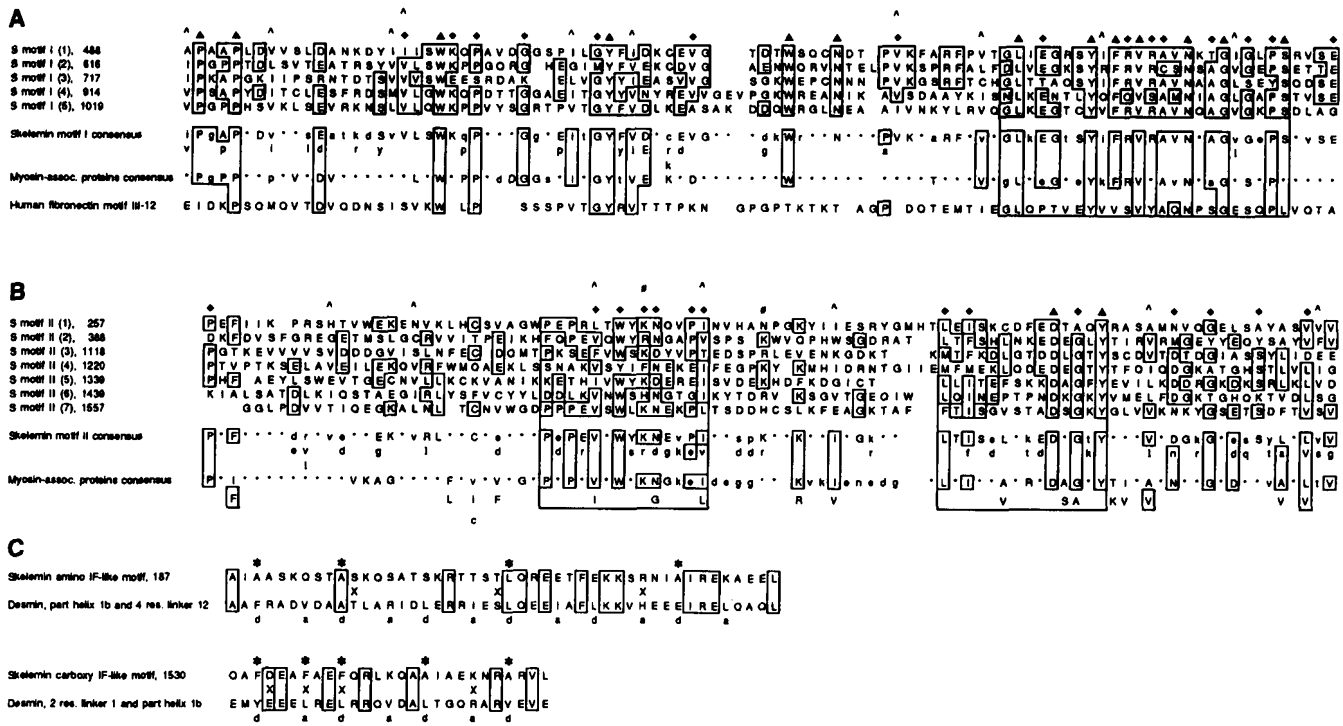

FIG. 8. **Alignments of the motifs I, II, and the intermediate filament-like motifs of skelemin (*S*).** The first amino acid of each motif is given. *Panel A*, the five fibronectin type III motifs, or motifs I, of skelemin are aligned with one another with *boxes* surrounding the residues that are identical or are D/E conservative substitutions in 3 or more motifs. The 13 conserved residues (*filled triangles*), the 15 residues identical in 4/5 motifs (*filled diamonds*), and the positions in which V or I occur 3 or more times (%) are indicated. The *larger box* indicates the most highly conserved region. The consensus motif I for skelemin is comprised of residues repeated 3 or more times (*upper case*) and residues repeated twice (*lower case*); *asterisks* mark other residues and *blanks* indicate gaps. The motif I consensus for the myosin-associated proteins was compiled from motifs I in twitchin, projectin, titin, myosin light chain kinase, C-protein, and 86-kDa protein. *Capital letters* indicate identical residues in 4 or more, and *lower case letters* indicate identical residues in 3 of the six protein's consensus motif I. The sequence of human fibronectin type III (repeat 12) (Kornblihtt *et al.*, 1985) is shown. *Boxes* surround the matched residues in the two consensus sequences and the fibronectin sequence. *Panel B*, the seven immunoglobulin superfamily C2-like motifs, or motifs II, of skelemin are aligned with one another with minimum gaps. *Boxes* surround residues that are identical, D/E, or K/R (the latter indicted with #) conservative substitutions in over half of the motifs. Conserved residues (*filled triangles*), residues present in 5 or 6 of the motifs (*filled diamonds*), and positions in which V/I substitutions (%) are present in over half the motifs are indicated. A consensus motif II for skelemin is derived from these *boxed* residues (*capital letters*) as well as from 2 or 3 residues that occur twice at a position (*lower case letters*). *Larger boxes* enclose two 14-residue regions with the greatest similarity. A master consensus motif II derived from six of the putative myosin-associated N-CAM-like proteins is compiled from residues identical in half or more of their consenus motifs II. *Boxes* surround matching residues of the two consensus motifs. *Panel C*, the two intermediate filament-like regions of skelemin are aligned with the corresponding regions of helix 1b from desmin of chicken smooth muscle (Geisler and Weber, 1982). *Boxes* enclose identical residues and D/E substitution; this and other conservative substitutions are indicated with an *X*. The *a* and *d* positions of the helix are indicated by *asterisks*.

similar to the amino ~30% of helix 1b of chicken desmin (Fig. 8*C*). The heptad repeat pattern of positions *a* and *d* generally being hydrophobic or uncharged residues (Fraser *et al.*, 1990) is well maintained in the carboxyl intermediate filament-like region of skelemin. Of the eight *a* or *d* positions in this intermediate filament-like region, five have hydrophobic residues in skelemin and one has a positively charged residue in both skelemin and desmin helix 1b.

DISCUSSION

Biochemical, immunolocalization, and developmental studies had suggested that skelemin might link desmin intermediate filaments to the myofibril (Price, 1984, 1987; Thornell *et al.*, 1990a). The ~5.6-kb mouse muscle cDNA described here appears to correspond to the skelemin transcript on the basis of five lines of evidence. First, protein expressed by an in-frame ~3.9-kb coding region strongly binds anti-skelemin antibodies. Similarly, anti-skelemin antibodies bind to three nonoverlapping recombinant proteins of 55–67 kDa which together account for the majority of the ~5.0-kb open reading frame (Fig. 2). Second, antibodies directed against each of three nonoverlapping regions of the protein product of the cDNA specifically label mouse skelemin on immunoblots (Fig. 3) and in longitudinal and cross-sections stain the M-discs of skeletal muscle indistinguishably from the antibodies directed against bovine skelemin (Figs. 4 and 5). Third, several bovine skelemin tryptic fragments closely match regions of the deduced amino acid sequence of skelemin (Fig. 1). The conservative substitutions and differences could be due to the existence of species and tissue isoforms. Fourth, the ~5-kb open reading frame is the appropriate size to encode skelemin (Fig. 3). The calculated molecular mass of 185.2 kDa is 95% of the apparent molecular mass of 195-kDa determined from SDS-polyacrylamide gel electrophoresis. Fifth, the skelemin transcript is detected in muscle tissues but not in whole brain or in liver (Fig. 6), consistent with previous immunochemical detection of skelemin protein in muscle tissues but not in neurons or in liver (Price, 1987).

An exciting result of the sequence analysis is that skelemin is a cytoplasmic member of the immunoglobulin superfamily. The rapidly expanding immunoglobulin superfamily includes many recognition molecules, including receptors, and adhesion molecules that appear to mediate cell-cell and cell-substrate association, differentiation, and proliferation (Cun-

ningham, 1988; Edelman and Crossin, 1991; Fischer *et al.*, 1991; Rutishauser, 1989; Springer, 1990; Williams and Barclay, 1988; Williams, 1987). The immunoglobulin superfamily C2 motif appears to have multiple binding properties (Williams and Barclay, 1988; Williams, 1987), which is not surprising considering the low level of identity among these motifs. It allows the homophilic binding of N-CAMs, and thus plays a critical role in neuronal outgrowth and pattern formation during embryogenesis (Frei *et al.* (1992), Hall *et al.* (1990), Rao *et al.* (1992), and references therein, Hortsch and Goodman (1991), Rutishauser (1989)). A few heterophilic associations of immunoglobulin superfamily C2 motifs have been demonstrated; the immunoglobulin superfamily C2 motifs of the N-CAM contactin (Ranscht, 1988) mediate binding to the extracellular matrix protein tenascin (Zisch *et al.*, 1992), while a similar motif of the intercellular adhesion molecule I-CAM binds an integrin (Diamond *et al.*, 1991). Many immunoglobulin superfamily members, such as the N-CAMs and members of a family of receptor-linked tyrosine phosphatases, also contain in their extracellular domains a second type of motif resembling the multifunctional fibronectin type III motifs (motif I in the cytoplasmic family). The multiple fibronectin type III motifs confer fibronectin's homophilic binding as well as its heterophilic binding to cells and extracellular components, and thus mediate fibronectin's roles in cell shape changes, migration, and adhesion (Barkalow and Schwarzbauer, 1991; Cole and Akeson, 1989; Haugen *et al.*, 1990; Hynes, 1990; Morla and Ruoslahti, 1992; Ruoslahti 1988). Little is known about the binding properties of fibronectin type III repeats in other proteins (Prieto *et al.*, 1992; Williams and Barclay, 1988; Williams, 1987).

The family of cytoplasmic proteins that resemble N-CAM due to the presence of both the fibronectin type III and immunoglobulin C2 motifs all appear to be myosin-associated, or at least thick filament-associated, in some way (Epstein and Fischman, 1991; Fischman *et al.*, 1991). Some of these cytoplasmic immunoglobulin superfamily proteins clearly have regulatory functions since they have kinase domains in their deduced sequences or have demonstrated kinase activity; these include the titin subfamily members (Benian *et al.*, 1989; Fyrberg *et al.*, 1992; Labeit *et al.*, 1992) and myosin light chain kinase of non-striated cells (Olson *et al.*, 1990; Shoemaker *et al.*, 1990; Gallagher *et al.*, 1991). C-protein may have a regulatory function (Einheber and Fischman (1990) for discussion), while the function of the closely related 86-kDa protein is unknown (Fischman *et al.*, 1991). All of these may also have mechanical functions, particularly the members of the titin subfamily (Wang *et al.*, 1991; Granzier and Wang, 1993) including nematode twitchin, mammalian titin (Labeit *et al.*, 1990), *Drosophila* projectin (Ayme-Southgate *et al.*, 1991; Fyrberg *et al.*, 1992), and *Lethocerus* p800 (Lakey *et al.*, 1990). In all family members, the fibronectin type III motifs are more similar to each other than are the immunoglobulin C2 motifs. Both motifs exhibit conserved residues in certain positions and have several separate regions of strong identity (Benian *et al.*, 1989; Epstein and Fischman, 1991; Williams and Barclay, 1988).

Structural determinations have shown that the immunoglobulin superfamily C2 motif and the fibronectin type III motif are similar, being the classic β-barrel with 3 and 4 β-strands parallel to one another, giving an immunoglobulin fold (Baron *et al.*, 1992; Holden *et al.*, 1992; Leahy *et al.*, 1992). The presence of both motifs in the family of intracellular N-CAM-like proteins suggests that the two motifs evolved from a common motif (see Bazan (1990) and Einheber and Fischman (1990) for discussion).

In addition to having a unique amino-terminal and serine/proline-rich region, skelemin is unique in the cytoplasmic N-CAM-like family in having desmin helical-like motifs. The 27- and 46-residue long motifs, which are similar to helix 1b, are found near each terminus. Desmin molecules self-associate via the helical regions, and helix 1b is crucial to intermediate filament self-association (Fraser *et al.*, 1990; Geisler *et al.*, 1992; McCormick *et al.*, 1991).

Skelemin is located at the midline of the bundle of myosin thick filaments from the time that desmin intermediate filaments are still in cytoplasmic arrays, through adulthood (Thornell *et al.*, 1990b). Skelemin is co-purified with intermediate filament proteins through seven different extractions and greatly increases the diameter of reconstituted intermediate filaments (Price, 1984, 1987). The existence of both putative myosin thick filament-association and intermediate filament-like regions in skelemin suggests a molecular mechanism whereby skelemin could link intermediate filaments to myofibrils. In support of this hypothesis, preliminary observations of skelemin expressed in cells that normally are skelemin-minus show skelemin binding to both the myofibril homologs (stress fibers) and intermediate filaments. Studies currently underway will allow us to delineate the respective binding domains.

Acknowledgments—We thank Christopher Brooks, Deborah Brock, Wei Chen, Carl Taphouse, John Platt, and Grace Wu for technical assistance, Diane Hatton for help preparing the manuscript, and Dr. Joel Pearlman for the two cDNA libraries and for advice. We are grateful to Drs. Guy Benian, Kathleen Beckingham, Henry Epstein, Donald Fischman, Michael Gustin, Michael Stern, and Kuan Wang for helpful discussions.

REFERENCES

Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. H., and Kent, S. B. H. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 6970–6974
Altschul, S. F., and Erickson, B. W. (1986) *Bull. Math. Biol.* **48,** 603–616
Amasino, R. M. (1986) *Anal. Biochem.* **152,** 304–307
Ayme-Southgate, A., Vigoreaux, J., Benian, G., and Pardue, M. L. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 7973–7977
Bairoch, A. (1989) *PROSITE: A Dictionary of Protein Sites and Patterns,* Fifth Ed., University of Geneva Press, Geneva
Barkalow, F. J. B., and Schwarzbauer, J. E. (1991) *J. Biol. Chem.* **266,** 7812–7818
Baron, M., Main, A. L., Driscoll, P. C., Mardon, H. J., Boyd, J., and Campbell, I. D. (1992) *Biochemistry* **31,** 2068–2073
Bauw, G., Van Damme, J., Puyve, M., Vanderkerkhove, J., Gesser, B., Ratz, G. P., Lauridsen, J. B., and Celis, J. E. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 7701–7705
Bazan, J. F. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 6934–6938
Benian G. M., Kiff, J. E., Neckelmann, N., Moerman, D. G., and Waterson, R. H. (1989) *Nature* **342,** 45–50
Bieber, A. J., Snow, P. M., Hortsch, M., Patel, N. H., Jacobs, J. R., Traquina, Z. R., Schilling, J., and Goodman, C. S. (1989) *Cell* **59,** 447–460
Bryan, J., Imai, M., Lee, R., Moore, P., Cook, R. G., and Lin, W. (1989) *J. Biol. Chem.* **23,** 13873–13879
Chomczynski, P., and Sacci, N. (1987) *Anal. Biochem.* **162,** 156–159
Chou, P. Y., and Fasman, G. D. (1978) *Adv. Enzymol.* **47,** 45–148
Cole, G. J., and Akeson, R. (1989) *Neuron* **2,** 1157–1165
Cunningham, B. A. (1988) *Adv. Cell Biol.* **2,** 25–54
Cunningham, B. A., Hemperly, J. J., Murray, B. A., Prediger, E. A., Brackenbury, R., and Edelman, G. M. (1987) *Science* **236,** 799–806
Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed) Vol. 5, pp. 345–352, National Biomedical Research Foundation, Washington, D. C.
Diamond, M. S., Staunton, D. E., Marlin, S. D., and Springer, T. A. (1991) *Cell* **65,** 961–971
Edelman, G. M., and Crossin, K. L. (1991) *Annu. Rev. Biochem.* **60,** 155–190
Einheber, S., and Fischman, D. A. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 2157–2161
Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 140–144
Epstein, H. F., and Fischman, D. A. (1991) *Science* **251,** 1039–1044
Fischman, D. A., Vaughan, K. T., Weber, F. E., and Einheber, S. (1991) in *Frontiers of Muscle Research, Myogenesis, Muscle Contraction, and Muscle Dystrophy* (Osawa, E., Masaki, T., and Nabashima, Y., eds) pp. 211–222, Elsevier/North Holland, Amsterdam
Fisher, E. H., Charbonneau, H., and Tonks, N. K. (1991) *Science* **253,** 401–406
Fraser, R. D. S., MacRae, T. P., and Parry, D. A. D. (1990) in *Cellular and Molecular Biology of Intermediate Filaments* (Goldman, R. D., and Steinert, P. M., eds) pp. 205–232, Plenum Publishing Corp., New York
Frei, T., von Bohlen und Halback, F., Wille, W. and Schachner, M. (1992) *J. Cell Biol.* **118,** 177–194
Fritz, J. D., Greaser, M. L., and Wolff, J. A. (1991) *Nucleic Acids Res.* **19,** 3747

Fyrberg, C. C., Labeit, S., Bullard, B., Leonard, K., and Fyrberg, E. (1992) *Proc. R. Soc. Lond. B* **249,** 33–40

Gallagher, P. J., Herring, B. P., Griffin, S. A., and Stull, J. T. (1991) *J. Biol. Chem.* **266,** 23936–23944

Garnier, J. D., Osguthorpe, J., and Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120

Geisler, N., and Weber, K. (1982) *EMBO J.* **1,** 1649–1656

Geisler, N., Schünemann, J., and Weber, K. (1992) *Eur J. Biochem.* **206,** 841–852

Goudie, R. B., Horne, C. H. W., and Wilkinson, P. C. (1966) *Lancet* **ii,** 1224–1226

Granzier, H. L., and Wang, K. (1993) *J. Gen. Physiol.* **101,** 235–270

Grodberg, J., Lundrigan, M. D., Toledo, D. L., Mangel, W. F., and Dunn, J. J. (1988) *Nucleic Acids Res.* **16,** 1209

Grove, B. K., Kurer, V., Lehner, C., Doetschman, T. C., Perriard, J.-C., and Eppenberger, H. M. (1984) *J. Cell Biol.* **98,** 518–524

Hall, A. K., Nelson, R., and Rutishauser, U. (1990) *J. Cell Biol.* **110,** 817–824

Harrelson, A. L., and Goodman, C. S. (1988) *Science* **242,** 700–708

Haugen, P. K., McCarthy, J. B., Skuitz, A. P. N., Furcht, L. T., and Letourneau, P. C. (1990) *J. Cell Biol.* **111,** 2733–2745

Hein, J. (1990) *Methods Enzymol.* **183,** 626–645

Hemperly, J. J., Murray, B. A., Edelman, G. M., and Cunningham, B. A. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 3037–3041

Holden, H. M., Ito, M., Hartshorne, D. J., and Rayment, I. (1992) *J. Mol. Biol.* **227,** 840–851

Holton, T. A., and Graham, M. W. (1991) *Nucleic Acids Res.* **19,** 1156

Hopp, T. P., and Woods, K. R. (1983) *Mol. Immunol.* **20,** 483–489

Hortsch, M., and Goodman, C. S. (1991) *Annu. Rev. Cell Biol.* **7,** 505–557

Hunkapiller, T., and Hood, L. (1989) *Adv. Immunol.* **44,** 1–63

Hunkapiller, M. W., Lujon, E., Ostrander, F., and Hood, L. E. (1983) *Methods Enzymol.* **91,** 227–236

Hynes, R. O. (1990) *Fibronectins,* Springer-Verlag, New York

Jain, R., Yuen, I. S., Taphouse, C. R., and Gomer, R. H. (1992) *Genes & Dev.* **6,** 390–400

Johnson, G. D., and Araujo, G. M. C. N. (1981) *J. Immunol. Methods* **43,** 349–350

Kajiji, S., Tamura, R. N., and Quaranta, V. (1989) *EMBO J.* **8,** 763–680

Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, I., and Baralle, R. E. (1985) *EMBO J.* **4,** 1755–1759

Kozak, M. (1991) *J. Cell Biol.* **115,** 887–903

Krieg, P. A., Sakaguchi, D. S., and Kintner, C. R. (1989) *Nucleic Acids Res.* **17,** 10321–10335

Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132

Labeit, S., Barlow, D. P., Gautel, M., Gibson, T., Holt, J., Hsieh, C.-L., Franke, U., Leonard, K., Wardale, J., Whiting, A., and Trinick, J. (1990) *Nature* **345,** 273–276

Labeit, S., Gautel, M., Lakey, A., and Trinick, J. (1992) *EMBO J.* **11,** 1711–1716

Laemmli, U. K. (1970) *Nature* **227,** 680–685

Lakey, A., Ferguson, C., Labeit, S., Reedy, M., Larkins, A., Butcher, G., Leonard, K., and Bullard, B. (1990) *EMBO J.* **9,** 3459–3467

Lawrence, C. B., and Goldman, D. A. (1988) *Comp. Appl. Biosci.* **4,** 25–31

Lazarides, E., and Capetenaki, Y. G. (1986) in *Molecular Biology of Muscle Development* (Emerson, C., Nadal-Ginard, B., Fischman, D., and Siddiqui, M. A. Q., eds) pp. 749–772, Alan R. Liss, New York

Leahy, D. J., Hendrickson, W. A., Aukhil, I., and Erichson, H. P. (1992) *Science* **258,** 987–991

Masaki, T., and Takaiti, O. (1974) *J. Biochem. (Tokyo)* **75,** 367–380

McCormick, M. B., Coulombe, P. A., and Fuchs, E. (1991) *J. Cell Biol.* **113,** 1111–1124

McGraw, R. A., Steffe, E. K., and Baxter, S. M. (1990) *Biotechniques* **8,** 674–678

Monroy, A. F. (1988) *Focus* **10,** 14

Moos, M., Tacke, R., Scherer, H., Teplow, D., Früh, K., and Schachner, M. (1988) *Nature* **334,** 701–703

Morla, A., and Ruoslahti, E. (1992) *J. Cell Biol.* **118,** 421–429

Noguchi, J., Yanagisawa, M., Imamura, M., Kasuya, Y., Sakuriai, T., Tanaka, T., and Masaki, T. (1992) *J. Biol. Chem.* **267,** 20302–20310

Odermatt, E., and Engel, J. (1989) in *Fibronectin* (Mosher, D. F., ed) pp. 25–47, Academic Press, New York

Olson, N. J., Pearson, R. B., Needleman, D. S., Hurwitz, M. Y., Kemp, B. E., and Means, A. R. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 2284–2288

Pearson, W. R., and Lipman, J. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 2444–2448

Pierobon-Bormioli, S. (1981) *J. Muscle Res. Cell Motil.* **2,** 401–413

Price, M. G. (1984) *Am. J. Physiol.* **246,** H566–572

Price, M. G. (1987) *J. Cell Biol.* **104,** 1325–1336

Price, M. G. (1991) *Adv. Struct. Biol.* **1,** 175–207

Price, M. G., and Gomer, R. H. (1989) *Cell Motil. Cytoskel.* **13,** 274–287

Price, M. G., and Sanger, J. W. (1983) *Cell Muscle Motil.* **3,** 1–40

Prieto, A. L., Andersson-Fisone, E., and Crossin, K. L. (1992) *J. Cell Biol.* **119,** 663–678

Proudfoot, N. (1991) *Cell* **64,** 671–674

Ranscht, B. (1988) *J. Cell Biol.* **107,** 1561–1573

Rao, U., Wu, X.-F., Gariepy, J., Rutishauser, U., and Siu, C.-H. (1992) *J. Cell Biol.* **118,** 937–949

Ruoslahti, E. (1988)*Annu. Rev. Biochem.* **57,** 375–413

Rutishauser, U. (1989) *Curr. Opin. Cell Biol.* **1,** 898–904

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual,* 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467

Seeger, M. A., Haffley, L., and Kaufman, T. C. (1988) *Cell* **55,** 589–600

Shoemaker, M. O., Lau, W., Shattuck, R. L., Kwiatkowski, A. P., Matrisian, P. E., Guerra-Santos, L., Wilson, E., Lukas, T. J., Van Eldick, L. J., and Watterson, D. M. (1990) *J. Cell Biol.* **111,** 1107–1125

Short, J. M., Fernandez, J. M., Sorge, J. A., and Huse, W. D. (1988) *Nucleic Acids Res.* **16,** 7583–7600

Small, J. V., Fürst, D. O., and Thornell, L.-E. (1992) *Eur. J. Biochem.* **208,** 559–572

Springer, T. A. (1990) *Nature* **346,** 425–434

Steven, A. C. (1990) in *Cellular and Molecular Biology of Intermediate Filaments* (Goldman, R. D., and Steiner, P. M., eds) pp. 233–266, Plenum Publishing Corp., New York

Street, S. F. (1983) *J. Cell Physiol.* **114,** 346–364

Streuli, M., Krueger, N. X., Tsai, A. Y. M., and Saito, H. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 8698–8702

Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol.* **185,** 60–89

Suzuki, S., and Naitoh, Y. (1990) *EMBO J.* **9,** 757–763

Thornell, L.-E., Carlsson, E., and Pedrosa, F. (1990a) in *The Dynamic State of Muscle Fibers* (Dirk, P., ed) pp. 369–383, Walter de Gruyter and Co., New York

Thornell, L.-E., Price, M. G., and Virtanen, I. (1990b) in *Developmental Cardiology: Morphogenesis and Function* (Clark, E. B., and Takao, A., eds) pp. 115–126, Futura Publishing Co., Mt. Kisco, NY

Trinick, J. (1992) *FEBS Lett.* **307,** 44–48

Vaughn, K. T., Weber, F. E., Einheber, S., and Fischman, D. A. (1993) *J. Biol. Chem.* **268,** 3670–3676

Wang, K. (1985) *Cell Muscle Motil.* **6,** 315–369

Wang K., and Ramirez-Mitchell, R. (1983) *J. Cell Biol.* **96,** 563–570

Wang, K., McCarter, R., Wright, J., Beverly, J., and Ramirez-Mitchell, R. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 7101–7105

Williams, A. F. (1987) *Immunol. Today* **8,** 298–203

Williams, A. F., and Barclay, A. N. (1988) *Annu. Rev. Immunol.* **6,** 381–405

Young, R. A., and Davis, R. W. (1983) *Science* **222,** 778–782

Yuen, S. W., Chui, A. H., Wilson, K. J., and Yuan, P. M. (1989) *Biotechniques* **7,** 74–82

Zisch, A. H., D'Alessandri, L., Ranscht, B., Falchetto, R., Winterhalter, K. H., and Vaughan, L. (1992) *J. Cell Biol.* **119,** 203–213