Mammalian skeletal muscle C-protein: purification from bovine muscle, binding to titin and the characterization of a full-length human cDNA

DIETER O. FÜRST*, UWE VINKEMEIER and KLAUS WEBER
Max-Planck-Institute for Biophysical Chemistry, Department of Biochemistry, PO Box 2841, D-3400 Goettingen, FRG
*Author for correspondence

Summary

We report a fast method for the isolation of homogeneous C-protein from bovine skeletal muscle. In electron micrographs C-protein appears as short rods with a relatively uniform length of about 50 nm. Protein sequencing shows a single N-terminal sequence. Radio-labelled C-protein strongly decorates titin II and myosin rods but not myosin heads. Binding to titin II is retained in preparations lacking titin-associated proteins. Antibodies to bovine C-protein were used to screen a Agt11 cDNA library constructed from fetal human skeletal muscle. Clone HC38 is 3833 bp long and encodes a protein of 1138 amino acid residues. The start of the predicted sequence fits the N-terminal sequence of the bovine protein. All partial sequences obtained from the bovine protein (348 residues) and the sequence deduced from a partial chicken cDNA (Einheber and Fischman, 1990) can be aligned along the human sequence. The sequences of human and chicken C-proteins share 50% identity and 70% similarity. Along the repeat patterns of the human protein the fibronectin (Fn)-like domains are better conserved than the immunoglobulin (Ig)-like domains. Regions of strong divergence between chicken fast C-protein and human slow C-protein may represent differences in C-protein isoforms.

Key words: C-protein, fibronectin repeats, muscle, myosin, sarcomere, titin.

Introduction

Titin is an important component of the myofibril (for reviews see Wang, 1985; Maruyama, 1986; Trinick, 1991). Immunoelectron microscopy with a bank of 14 monoclonal antibodies shows that titin molecules have half-sarcomere length and span the distance from the Z band to the M band (Furst et al., 1988, 1989). Isolated titin II molecules have a narrow length distribution of about 900 nm but lack the Z-band anchoring domain due to the proteolysis that is necessary to extract the native molecules. Purified titin II molecules are a string carrying a single globular head. This seems to reflect two M band proteins, which have been described as titin-associated proteins (Nave et al., 1989). On the basis of these structural results one can discern functionally distinct parts of the titin molecule. The Z-band binding region is currently not directly available to molecular analysis. Over the I band portion titin molecules show high elasticity (Maruyama et al., 1985; Furst et al., 1988; Itoh et al., 1988; Whiting et al., 1989) and are then arranged parallel to the A band (Trinick et al., 1984). Finally the molecules end within the M band where they seem firmly embedded via specific M-band proteins (Nave et al., 1989). cDNA cloning of a part of titin situated in the A band has shown a regular pattern of 100-residue repeats, which reflect similar domains in immunoglobulins (class II domains) and fibronectin (class I domains) (Labeit et al., 1990). These 100-residue repeats probably explain certain electron micrographs indicating that titin is built from a linear array of 4.3 nm globular domains (Trinick et al., 1984; Whiting et al., 1989). Titin molecules seem to have an additional repeat pattern of 42 nm, which may be important in understanding the titin-A band disposition. Several monoclonal antibodies identify in immunoelectron microscopy a 42 to 43 nm repeat pattern in the A band (Furst et al., 1989). These repetitive epitopes seem to coincide with some of the striations of the A band known to harbor two myosin-associated proteins: C-protein and 86K protein (Sjöström and Squire, 1977; Craig and Offer, 1976; Dennis et al., 1984; Bähler et al., 1985a,b; Furst et al., 1989).

C-protein has been characterized extensively as a myosin-associated protein recognizing the rod portion of the myosin molecule (Moos et al., 1975; Starr and Offer, 1978) and its disposition along the A band has been well characterized (Craig and Offer, 1976; Dennis et al., 1984; Bennett et al., 1986). Several reports indicate that C-protein could play a role in modulating muscle contraction (Offer et al., 1973; Moos et al., 1978; Moos and Feng, 1980; Hartzell and Titus, 1982) or...
in thick filament assembly (Offer et al., 1973). Our results on some repetitive titin epitopes opened the possibility that C-protein could connect as a missing link the titin strings at multiple sites to the A band (Fürst et al., 1989). During our studies on C-protein and its possible interaction with titin, Einheber and Fischman (1990) reported a partial cDNA clone encoding about 80% of the fast isoform of chicken C-protein. Both C-protein and titin belong to a superfamily of proteins built from domains that share sequence similarity with immunoglobulin (class II domains) and fibronectin (type I domains).

Here we describe a fast and convenient procedure for the purification of C-protein from bovine slow muscle, which has facilitated the physical-chemical characterization of the molecule. We report a strong and specific interaction of the molecule. We report a strong and specific interaction of the molecule.

Materials and methods

Purification of bovine skeletal muscle C-protein

Bovine muscle (Musculus iliacus) was removed immediately after slaughter. It was chopped into small pieces, quickly frozen in liquid nitrogen and stored at −70°C. After quick thawing of ~100 g of this material in ice-cold LSB (low salt buffer: 100 mM KCl, 2 mM MgCl2, 5 mM EGTA, 1 mM 2-mercaptoethanol, 1 mM NaN3, 10 mM Tris-maleate, pH 6.8) containing 2 mM Na2P2O7% the tissue was homogenized for 2 × 30 s with a Polytron homogenizer. The following protease inhibitors were present in LSB and all subsequent buffers: 1 mM PMSF, 10 mg/ml trypsin inhibitor II (Trypsin, Sigma Chemical Co., St. Louis, MO, USA), and 5 mM E-64 (E3132, Sigma). Myofibrils were harvested (15 min at 3,000 g), washed 3 times with LSB, and resuspended in extraction solution (0.6 M KCl, 2 mM MgCl2, 2 mM EGTA, 1 mM 2-mercaptoethanol, 1 mM NaN3, 10 mM imidazole-HCl, pH 7.0) for 35 min. The supernatant obtained after centrifugation (20,000 g for 50 min) was extensively dialyzed against buffer A (2 mM EGTA, 1 mM 2-mercaptoethanol, 1 mM NaN3, 50 mM Tris-maleate, pH 7.9) containing 250 mM KG. The enzyme:substrate ratio was 1:600 (w/w). After 60 min at room temperature C-protein was completely converted to a 110,000 Mr fragment. The reaction was stopped by the addition of 3 mM PMSF (Serva) and 0.1 mM TLCK (Sigma, T7254). This material was gel filtered on an FPLC Superose S12 column equilibrated in the same buffer. The fractions containing the 110,000 Mr fragment were pooled and concentrated by ultrafiltration (Centricon 30, Amicon) to a final concentration of approximately 0.3-0.4 mg/ml.

Iodination of C-protein

About 250 μl of the central portion of the Superose S12 peak, containing 80 μg C-protein (in buffer A without 2-mercapto-ethanol) were labelled with 125I (Na 125I, #IMS3.0, Amer sham) using one bead of iodination reagent (Iodo Bead, Pierce, Rockford, IL, USA) as specified by the manufacturer. After incubation for 15 min at room temperature the mixture was passed through a PD10 gel filtration column (Pharmacia) equilibrated in PBS (137 mM NaCl, 3 mM KCl, 1.5 mM KH2PO4, 8 mM Na2HPO4, 1 mM Na3citrate, pH 7.3), containing 3% bovine serum albumin (BSA) to remove free iodine. Labelling specificity was monitored by gel electrophoresis and autoradiography.

Purification of myosin, myosin-rod and myosin S1

Myosin was purified from bovine skeletal muscle essentially as described (Margo ssian and Lowey, 1982). Briefly, 100 g of fresh bovine muscle were homogenized and extracted in 300 ml high salt buffer (0.3 M KC1, 150 mM potassium phosphate, 20 mM EDTA, 1 mM MgCl2, 1 mM ATP). Extraction was stopped by the addition of three volumes of water and centrifugation (3,000 g, 10 min). To precipitate the myosin, 1.8 l water were added to the supernatant. After settling for 3.5 hours the supernatant was decanted and the protein was collected by centrifugation (9,000 g, 15 min). The pellet was dissolved in 90 ml 1 M KC1, 25 mM EDTA, 60 mM potassium phosphate, pH 6.5 and dialyzed overnight against 0.6 M KC1, 25 mM potassium phosphate, pH 6.5, 10 mM EDTA, 1 mM dithiothreitol (DTT). Actomyosin was removed by the addition of 110 ml of water followed by centrifugation (72,000 g, 1 hour). 1.3 l of water were added to the supernatant and after several hours of settling time the precipitate was collected by centrifugation (12,000 g, 10 min). The resulting crude myosin was dissolved in 3 M KC1, 10 mM potassium phosphate, pH 6.5, and dialyzed against 0.6 M KC1, 50 mM potassium phosphate, pH 6.5, 1 mM DTT (storage buffer). In order to remove C-protein, the crude myosin was purified by ion exchange chromatography on a DEAE-Sephadex A-50 column (20 cm × 1.5 cm). Proteins were eluted with a linear gradient from 0 to 500 mM KC1 in 150 mM potassium phosphate, pH 7.5, 10 mM EDTA. Myosin-containing fractions were pooled and after overnight dialysis against 40 mM sodium pyrophosphate, pH 7.5, 1 mM DTT and further processed on a DEAE-cellulose (DE-52) column (30 cm × 1.5 cm). Protein was eluted with a linear NaCl gradient (0 to 0.5 M in 20 mM sodium pyrophosphate, pH 7.5, 1 mM DTT). Myosin was precipitated by dialysis against a low salt buffer (40 mM KC1, 10 mM potassium phosphate, pH 6.5, 1 mM DTT) and stored at −20°C as a 5 mg/ml solution in storage buffer containing 50% glycerol.

Chymotryptic digestion was used to prepare the myosin subfragments S1 and rod (Margo ssian and Lowey, 1982).
Myosin filaments were generated by dialysis against 120 mM NaCl, 20 mM sodium phosphate, 1 mM EDTA, pH 7.0. A sample (8 ml) of this myosin suspension at a concentration of 2.8 mg/ml was treated at room temperature with a final α-chymotrypsin (Sigma, C3142) concentration of 0.05 mg/ml. The reaction was stopped after 10 min by the addition of PMSF (Serva) to 0.1 mM. The soluble S1 fraction was separated from the insoluble rods by centrifugation (60,000 g, 1 hour). The rod-containing pellet was dissolved in 6 ml of 0.6 M KCl, 50 mM potassium phosphate, pH 7.0; 18 ml of ethanol were added to denature residual myosin. After stirring for 4 hours the precipitate was collected by centrifugation (8,000 g, 30 min) and resuspended in the same buffer as above. Residual ethanol was removed by dialysis against the same buffer. Finally the solution was clarified by high speed centrifugation (90,000 g, 30 min) and dialyzed against a low salt buffer (30 mM KCl, 10 mM potassium phosphate, pH 7.0) to precipitate the myosin rods. This material was redissolved in a minimal volume of high salt buffer (0.6 M KCl, 50 mM potassium phosphate, pH 7.0) and used for the overlay assays. The soluble S1-containing supernatant from above was precipitated by the addition of ammonium sulfate. The pellet of the 48-60% fraction was dissolved in 0.1 M KCl, 10 mM imidazole-HCl, pH 7.0, 1 mM DTT, 0.3 mM EGTA, and chromatographed on a Superose S12 FPLC column equilibrated in the same buffer.

**Purification of titin**

Titin II from bovine muscle was purified as described by Nave et al. (1989). Limited proteolysis of titin II with trypsin and the subsequent removal of titin-associated proteins by gel filtration was also performed as described (Nave et al., 1989).

**C-protein binding assay**

Native protein preparations were spotted onto nitrocellulose membranes (BA-85; Schleicher and Schuell, Dassel, Germany) and air-dried for 30 min at 37°C. Nitrocellulose strips were blocked with 10% fetal calf serum (FCS; Boehringer Mannheim) in Buffer B (0.05% Tween-20, 3 mM 2-mercaptoethanol, 1 mM NaF, 50 mM Tris-HCl, pH 7.9, 150 mM NaCl) for 60 min at ambient temperature with constant shaking. Individual strips were incubated at room temperature for 2.5 hours with radiiodinated C-protein (see above) in 1 ml of buffer B supplemented with 3% FCS. After extensive washing with buffer B plus 3% FCS (at least 8 changes within 24 hours) the strips were air-dried and autoradiographed on Fuji RX-100 film.

**Protein-chemical studies on bovine C-protein**

Automated sequencing was performed on an Applied Biosystems sequenator (model A470) and a Knauer sequenator (model 810). Both instruments were operated with on-line PTH amino acid analyzers. In addition to standard protein-chemical procedures we made use of the methods described by Vandeserkhove and coworkers. These involved the transfer of C-protein from SDS-PAGE to a PVDF membrane, the treatment of the blots with endoproteinases and the subsequent separation of peptides by reverse phase HPLC (Bauw et al., 1987).

**Electron microscopy of bovine C-protein**

A drop of a 10 μg/ml solution of C-protein in buffer A containing 250 mM KCl and 50% glycerol was applied to freshly cleaved mica. Removal of excess liquid by drawing one edge of the mica across filter paper was followed by standard low-angle rotary shadowing using tantalum/tungsten at 5° and carbon at 90°. Replicas were floated off on a surface of distilled water and collected on copper grids (Nave et al., 1989).

**Cloning of human skeletal muscle C-protein**

The human skeletal muscle cDNA library in λgt11 (prepared from poly(A)+ RNA from skeletal muscle of a fetus in the 22nd week of gestation) was kindly provided by Dr. H. Arnold (University of Hamburg, FRG). The initial antibody screen of the library was performed as described (Stick, 1988). For the subsequent screens (to obtain a full-length cDNA clone) 5’ fragments were labelled using the T7 Quick Prime Kit (Pharmacia). Hybridization was carried out in 5 × SSPE, 10 × Denhardt’s solution, 0.5% SDS and 100 mg/ml sonicated salmon sperm DNA at 65°C. The final wash was at 65°C in 0.1% SDS, 1 × SSC. Inserts were subcloned according to standard procedures (Sambrook et al., 1989) into plasmids Bluescript (Stratagene) and M13mp18 and 19 (Yanish-Perron et al., 1985) for sequencing. Plasmids were purified on Qiangen columns (Qiagen). Sequencing was carried out by the dideoxy chain termination reaction according to Sanger et al. (1977) using the Sequenase version 2.0 DNA sequencing kit (United States Biochemical Corporation). Both stands were fully sequenced.

**Miscellaneous procedures**

A polyclonal antibody against bovine C-protein was raised in rabbits. Specificity of the antibody was checked by immunofluorescence and immunoblotting on whole muscle samples as described (Furst et al., 1988).

Analytical ultracentrifugation of purified C-protein was done in a model E centrifuge (Beckman Instruments Inc., Palo Alto, CA) with Flossmann optics and a scanner. C-protein at 0.2 mg/ml in buffer A containing 500 mM KCl was used. Centrifugation at 20°C was at 48,000 revs/min. The sedimentation coefficient was corrected to the s20,w value.

**Results**

**Purification and physicochemical characterization of bovine C-protein**

We have developed a new procedure for the purification of skeletal muscle C-protein. This method has the advantage of being considerably faster than previously published procedures (Offer et al., 1973; Starr and Offer, 1982). It involves only five principal steps, which can be performed within two days. These steps are: (1) a high salt extraction of washed myofibrils; (2) dialysis against a low salt buffer to remove most of the myosin and titin; (3) ammonium sulfate fractionation between 40 and 60% saturation; (4) desalting and batchwise DE-52 ion exchange chromatography; (5) gel filtration on FPLC Superose S12. As judged by Coomassie-blue stained SDS-PAGE (Fig. 1) the procedure provides C-protein totally free of contaminants. The purity of the preparation is also documented by the result of the analytical ultracentrifugation. C-protein sedimented as a single symmetrical peak with an s20,w of 4.66 S. This value is in excellent agreement with the sedimentation coefficient of 4.65 S reported by Offer et al. (1973). Protein sequencing showed a unique N-terminal sequence (see below).

In good agreement with the previous report on C-protein from rabbit skeletal muscle (Offer et al., 1973),
we found that circular dichroism spectra of the bovine C-protein show a typical $\beta$ spectrum (Fig. 2). Surprisingly, two earlier reports on the molecular shape of C-protein have revealed a rather heterogeneous picture. A study on C-protein from chicken heart indicated V-shaped particles with arm lengths of about 22 ± 4.5 nm (Hartzell and Sale, 1985). C-protein from chicken skeletal muscle appeared as rod-shaped, U-shaped or V-shaped within a size range of 15 to 40 nm (Swan and Fischman, 1986). In our hands rotary-shadowed molecules of bovine skeletal muscle C-protein show a relatively uniform morphology. The well-oriented molecules appear as short segmented rods with a length of about 50 nm and a diameter of 4 nm (Fig. 3). A possible explanation for these differences could be that we avoided spraying the molecules during sample preparation. We used instead the careful layering technique, which we found useful also in the case of titin (Nave et al., 1989).

Binding of radioiodinated C-protein to titin
Gel filtered C-protein was labelled with $^{125}$I to a specific activity of $4.5 \times 10^5$ to $6 \times 10^5$ cts/min per mg. Fig. 4 demonstrates the purity of the sample used for iodination and the specific incorporation of the label. Iodinated C-protein was used in a binding assay on native bovine titin II bound to nitrocellulose. Fig. 4 shows that radiolabelled C-protein strongly decorates purified titin II. To exclude the possibility that the binding observed was caused by interaction of C-protein with the two M band proteins present in the head of the titin II molecule (Nave et al., 1989) we treated titin II with trypsin and repurified the protein by gel filtration. Such titin II preparations have been shown to lack both titin-associated proteins and still retain an unchanged molecular length (see Nave et al., 1989). Fig. 4 shows that iodinated C-protein labelled “headless” titin with the same intensity as normal titin II. In control experiments in which unlabelled C-protein was added to the incubation mixture, the resulting autoradiographs showed a strong reduction in the titin decoration (results not shown).

The specificity of the C-protein binding assay was also documented with several control proteins. While serum albumin and myosin heads, prepared as chymotryptic S1, remained undecorated, the myosin rod and titin...
were strongly decorated (Fig. 4). This interaction does not require the intact C-protein as the thrombic derivative lacking the N-terminal 170 residues (see below) gives the same intense decoration of myosin rod and titin II (result not shown).

Protein sequence information on bovine C-protein
All partial protein sequences obtained on bovine C-protein could be aligned along the complete human C-protein sequence later predicted from a cDNA alone (Figs 5 and 6). Direct sequencing established the unique N-terminal sequence PEPTKKEENEVLAPAP (Fig. 6). CNBr treatment of the intact protein followed by reverse-phase HPLC provided only two pure fragments, which were characterized by extended sequencing. Additional information was collected from C-protein transferred from SDS-PAGE onto a PVDF membrane, which was subsequently treated either with trypsin or with endoproteinase Asp-N. Peptides released from the blot were subjected to HPLC and the resulting profiles were searched by sequence analysis for pure peptides.

When native C-protein at an enzyme to substrate ratio of 1:600 was treated with thrombin for 60 min, subsequent SDS-PAGE revealed a stable fragment with an apparent $M_r$ of 110,000 (Fig. 1B). Gel filtration on a Superose S12 column separated this large fragment from smaller $M_r$ material. The large fragment provided an N-terminal sequence of 57 residues in which only one position remained unidentified (Fig. 6). The small $M_r$ fraction from the mild thrombic digest was processed by various treatments, which included cleavage at tryptophan by BNPS-skatole, cleavage by CNBr, separate digests with endoproteinases Asp-N, Lys-C and Glu-C. Resulting fragments were processed by HPLC and sequenced. Fig. 6 shows that the cleavage site for thrombin lies 171 residues past the N terminus. Thus the large fragment covers the C-terminal part of the molecule. In line with the specificity of thrombin cleavage occurs between an arginine and a small amino acid (serine). While this site is present in human and bovine C-protein it is changed in the chicken protein (Fig. 6).

Isolation and sequence of the cDNA encoding human C-protein
Antibodies raised against bovine C-protein were used to screen a Agt11 cDNA expression library constructed...
from poly(A)+ RNA of fetal human skeletal muscle. The primary screen of about 10^5 plaques yielded 5 strongly reacting clones. All 5 clones harbored cDNA inserts of ~1.3 to 1.6 kb and restriction enzyme maps indicated descent from a single mRNA template. Sequencing of the largest insert revealed an open reading frame of 424 residues. This partial protein sequence was highly related to the protein sequence information provided by bovine C-protein. Two subsequent screening rounds of the human λgt11 library using 32P-labelled 5' fragments of the longest available clones were required to isolate a full-length cDNA clone (HC38).

Clone HC38 has a length of 3833 bp and reveals a single large open reading frame that encodes a protein of 1138 amino acid residues (Fig. 5). The coding region is flanked by 5'-(nucleotides 1 to 96) and 3'-untranslated sequences (nucleotides 3511 to 3833), respectively. It starts with the ATG codon at positions 97-99, which are part of a nucleotide sequence fitting the consensus sequence required for efficient translational initiation (Kozak, 1989). One in-frame TAA stop codon at residues 79-81 precedes this start codon. The 3414 bp coding region is terminated by a TAA stop codon at positions 3511-3513. This marks the beginning of a 3' untranslated sequence of 323 bp. Residues 3589-3815 indicate a possible polyadenylation signal. However, since a remnant poly(A) tail was not found this assignment is not unequivocally established. Allowing for a poly(A) tail of over 100 residues, the total cDNA sequence of 3833 bp is most likely a near-complete copy of the ~4.0 kb mRNA found by Northern blotting of mRNA from bovine skeletal muscle (result not shown).

The predicted amino acid sequence of human C-protein following the initiator methionine is in excellent agreement with the N-terminal sequence established directly on the bovine protein (Fig. 6). Human C-protein has a calculated isoelectric point of 7.6 and a chemical Mr of 127,806. This value is somewhat lower than the 140,000 usually estimated by SDS-PAGE.

**Alignment of the sequences of human, bovine and chicken C-proteins**

Fig. 6 shows a comparison of the different C-protein sequences that were either obtained by partial protein sequencing (bovine C-protein) or deduced from cDNA sequences (human and chicken C-protein). While the human cDNA is complete, the chicken cDNA clone lacks approximately 150 amino acid residues at the 5' end (see Einheber and Fischman, 1990). Direct sequencing of the bovine protein provided a total of 348 amino acid residues, which correspond to one third of the molecule. The majority of this sequence covers the N-terminal part of the protein. Comparison of human and bovine sequences reveals 92% identical and 96% homologous residues. The strongest difference resides in the peptide spanning residues 35-53 with 9 amino acid exchanges (5 conservative replacements). Human C-protein shares 50% sequence identity with the chicken fast isoform (the degree of similarity is about 70%).

The striking repeat pattern revealing immunoglobulin (Ig)-like domains (class II) as well as fibronectin (Fn)-like domains (class I) first reported for the chicken protein by Einheber and Fischman (1990) is equally well conserved in the human protein (Fig. 6). Residues 266-1125 of human and residues 125-980 of chicken C-protein show the pattern II-II-II-II-I-II-I-II-I-II-I-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-II-I...
Sequence of human muscle C-protein

Fig. 5.
Fig. 6. Sequence alignment of C-proteins emphasizing the arrangement of type I and type II repeats. (b) Residues obtained by direct protein sequencing of purified bovine C-protein. (h) The complete sequence of human C-protein derived from the cDNA (see Fig. 5). (c) The sequence of the chicken fast C-protein deduced from a partial cDNA clone (Einheber and Fischman, 1990). In the upper half of the figure all immunoglobulin-like domains (class II domains) are aligned. The more C-terminally situated regions of the repeats are separated by a space. The bottom half of the figure gives the alignment for all fibronectin-like domains (class I domains) and the subsequent C-terminal regions. The number at the beginning of each line indicates the number of the respective repeat in the linear sequence (see Fig. 5). The weak repeat —1 as well as repeats 1-9 are given for chicken and human proteins while the preceding weaker repeat —2 and the N terminus are currently only available for the human protein. The number at the end of the line gives the number of the last amino acid residue along the linear sequence of human C-protein (see Fig. 5). Highly conserved residues specific for class I or class II domains are marked by arrowheads below the 9th and 8th repeat, respectively. The arrow pointing down indicates the cleavage site for thrombin, which provides the large C-terminal fragment of MT 110,000 for bovine C-protein (see Results). The schematic diagram at the bottom outlines the arrangement of both repeat types in C-protein. I and II designate class I and II domains, respectively. The parentheses indicate that the two first repeats are significantly less well conserved than the other repeats.

Discussion

We have developed a fast and convenient procedure to isolate C-protein from bovine skeletal muscle. Approximately 10 mg of C-protein are isolated within 2 days from 100 g of muscle. The purified protein is homogeneous by various criteria. It provides a single band of apparent Mr 140,000 in SDS-PAGE and sediments as a homogeneous peak with an S20,w value of 4.66 S. More importantly automated sequencing shows a unique N-terminal sequence, which is preceded by the initiator methionine in the sequence of human C-protein
predicted from a cDNA clone. In addition all partial sequences obtained on the bovine protein are directly aligned along the human C-protein. In agreement with Offer et al. (1973), we find a circular dichroism spectrum dictated by \( \beta \)-structure. Metal-shadowed molecules of bovine C-protein show the appearance of short rods with indication for some segmentation. The length of the rods is around 50 nm and their diameter lies around 4 nm without correction for the metal layer.

Previous work of Offer et al. (1973) identified C-protein from rabbit skeletal muscle as a myosin binding protein. In agreement with the known location of C-protein to the A-band, Offer and colleagues showed that this binding occurs to the rod domain (Starr and Offer, 1978; Moos et al., 1975). Extending these studies we have now shown that radiolabelled C-protein binds strongly and specifically to myosin, myosin rod and titin II spotted on nitrocellulose, while myosin S1 and serum albumin are not decorated. Binding to titin II is independent of the two titin-associated proteins as shown by trypsinized and subsequently gel-filtered titin II, a preparation known to be devoid of associated proteins (Nave et al., 1989). Interestingly, binding to myosin rod and titin II is retained by the 110,000 fragment of C-protein obtained by treatment with thrombin. This fragment lacks the N-terminal 171 residues and retains all 9 strong repeat segments earlier identified by cDNA cloning of chicken C-protein (Einheber and Fischman, 1990). While these binding studies do not establish a stoichiometry they are the first in vitro support for a titin/C-protein/myosin rod interaction previously raised by immuneelectron microscopic results on some monoclonal antibodies detecting repetitive titin epitopes along the A-band. The position of these repetitive epitopes showed a 42-43 nm spacing coinciding in relative position with the known A band striations identified by C-protein and the 86K protein (Fürst et al., 1989). Although the major repeat distance along the titin string is around 4.3 nm, resembling a single 100-residue repeat (Trinick et al., 1984; Labeit et al., 1990), the 42-43 nm repeat may simply reflect a multiple of it provided by the specific linear arrangement of class I and II repeats. We expect that once longer titin sequences become known by cDNA cloning the linear repeat arrangement will offer a molecular explanation for the higher length repeat, which seems to relate to titin binding sites for C-protein and 86K protein, two major myosin binding proteins of the A-band.

Finally, we report a complete C-protein sequence deduced from a human cDNA clone and have compared this sequence with the partial chicken C-protein sequence reported by Einheber and Fischman (1990). Just as in chicken C-protein, we find 9 strong repeat segments displaying homology with either immunoglobulin (Ig) C2 or fibronectin (Fn) type III domains. These two domain types are known for a large number of proteins, the so-called immunoglobulin superfamily. Interestingly, several muscle proteins are now identified as members of this superfamily. These are twitchin (Benian et al., 1989), which can be considered to be an invertebrate mini-titin (Nave and Weber, 1990; Nave et al., 1991), titin (Labeit et al., 1990) and 86K protein, a myosin-associated A-band protein (Einheber and Fischman, 1990). The overall degree of sequence conservation between the various members of the superfamily is rather low and this also often holds when corresponding domains of a given protein are compared. In human C-protein, for instance, only 15 residues within the group of Ig domains (type II) and 15 residues in the Fn domains (type I) are conserved (Fig. 6). The alignment of the deduced chicken and human sequences shows about 50% sequence identity and 70% similarity. Interestingly, however, the two classes of domains do not display the same degree of conservation. The Fn-type domains are significantly better conserved (on average 58% identity, 83% similarity) than the Ig-type domains (on average 49% identity, 70% similarity). The C-terminal regions of the repeats share less identical residues (44%) but have a comparable level of similarity (75%). There are only two regions where the described "normal" repeat pattern is perturbed. Repeat 1-1 has an insertion of 8 amino acids at positions 49-56 only in the chicken sequence. Human C-protein, on the other hand, has no insert of 18 residues (position 762-779) in repeat 6. C-protein was shown to exist in at least three isoforms in different muscle fiber types. These variants are characterized by their slightly different \( M_r \) values, amino acid composition (Callaway and Bechtel, 1981; Yamamoto and Moos, 1981; Locker and Wild, 1986) and reaction with specific monoclonal antibodies (Reinach et al., 1982). Until now, fast, slow and cardiac isoforms have been identified and their distribution in certain muscles studied (Reinach et al., 1982; Obinata et al., 1984; Bähler et al., 1985a,b). We have purified C-protein from bovine slow muscle and raised antibodies against it. In immunoblotting this antibody recognizes only one band and in immunofluorescence it stains predominantly slow fibers. Since this antibody was used to screen a cDNA library we think that we picked a clone representing the human slow isotype. The partial chicken clone is the fast isoform (Einheber and Fischman, 1990). Thus it is possible that some of the distinct sequences represent regions in which C-protein isoforms are distinct. This question can be answered once cDNA clones representing human fast C-protein and/or chicken slow C-protein become known. Finally, we note that although secondary structure prediction rules indicate \( \alpha \)-helix and \( \beta \)-sheet for human C-protein, the circular dichroism spectra identify a \( \beta \)-protein (Fig. 2; see also Offer et al., 1973). These experimental results are in perfect agreement with the structural features of a member of the immunoglobulin superfamily.

We greatly appreciate the help of Uwe Plessmann in obtaining the partial sequence data on bovine C-protein. Dr. H. Arnold, University of Hamburg, kindly provided the human skeletal muscle cDNA library. The sequence reported above is available from the EMBL/GenBank under accession number X66276.
References


(Rceived 20 March 1992 - Accepted 1 May 1992)