Isolation and characterization of a cDNA clone encoding avian skeletal muscle C-protein: An intracellular member of the immunoglobulin superfamily

(thick filament/myosin/myofibril/fibronectin repeats/neural cell adhesion molecule)

STEVEN EINHEBER AND DONALD A. FISCHMAN

Department of Cell Biology and Anatomy, Cornell University Medical College, New York, NY 10021

Communicated by Alton Meister, December 11, 1989

ABSTRACT C-protein is a thick filament-associated protein located in the crossbridge region of vertebrate striated muscle A bands. Its function is unknown. To improve our understanding of its primary structure, we undertook the molecular cloning of C-protein mRNA. We describe the isolation and characterization of a cDNA clone, λ C-86, that encodes $\approx 80\%$ of the fast isoform of C-protein in the chicken. Sequence analysis of the insert revealed that C-protein, although an intracellular, nonmembrane-associated protein, is a member of the immunoglobulin superfamily. Like several cell surface adhesion molecules that belong to this superfamily, C-protein contains sequence motifs that resemble immunoglobulin domains and fibronectin type III repeats. Computer searches using the C-protein sequence also lead to the identification of related domains in chicken smooth muscle myosin light chain kinase that have not been reported previously.

C-protein ($M_r \approx 140,000$) is an abundant thick filament protein of unknown function found in all vertebrate striated muscles (1). It is distributed in seven to nine stripes spaced 43 nm apart (2–5) in the crossbridge-bearing, middle third of each half A band and binds to the subfragment 2 (6) and light meromyosin (7) portions of myosin heavy chain (MHC). In addition to MHC, C-protein binds F-actin (8) and native thin filaments *in vitro* (9, 10). Its interaction with regulated thin filaments is Ca^{2+} -sensitive (9, 10).

Based on its distribution in the A band and myosin- and actin-binding properties, it has been suggested that C-protein may modulate muscle contraction (1, 8, 11). This hypothesis is supported by several studies that have shown that C-protein modifies the activity of the actin-activated myosin ATPase (1, 12–14). However, C-protein may play a more structural role in stabilizing and aligning thick filaments or in regulating their assembly (1).

Fast, slow, and cardiac muscle isoforms of C-protein have been identified in adult vertebrates (13, 15–17) and additional forms are expressed during embryogenesis (18). Classification of these proteins as isoforms is based on their similar molecular weights, myofilament-binding properties, effect on actomyosin ATPase activity, and localization in the A band (13). However, they differ in amino acid composition (13, 15), antigenicity (13, 15), ability to undergo phosphorylation (11), peptide maps (13), and patterns of distribution along the thick filament (4, 5). It is still a matter of debate whether the differences exhibited by these molecules justify their designation as unique proteins or members of an isoform family.

Understanding of C-protein function and the structural relationships between C-protein isoforms has been hindered by a lack of primary sequence information. To date, there have been no reports of the amino acid sequence of C-

protein. In this paper we describe the isolation and characterization of a λ gt11 partial cDNA clone encoding the fast isoform of C-protein in the chicken.* Sequence analysis of the clone reveals that C-protein, although an intracellular protein, is a member of the immunoglobulin superfamily (19) and exhibits regions of structural similarity to a number of extracellular and cell surface adhesion molecules. A preliminary report of this study has been presented (20).

MATERIALS AND METHODS

A λ gt11 cDNA library (21) was constructed from poly(A)⁺ RNA (22, 23) isolated from the pectoralis muscle (PM) of 1-week posthatch chickens. The library was screened (24) with a rabbit polyclonal antiserum directed against adult chicken PM C-protein that had been purified by sequential DEAE column chromatography (1, 16) and SDS/PAGE (25). Specificity of the antiserum was proven by immunoprecipitation and immunoblot analyses (20). Primary antibody was detected with horseradish peroxidase-conjugated goat anti-rabbit IgG (E-Y Labs, San Mateo, CA) followed by diaminobenzidine color reaction.

Hybrid selection was performed by a modification of the method described by Miller *et al.* (26). *In vitro* translations were performed as described by Bouché *et al.* (27), and the translation products were immunoprecipitated with monoclonal antibodies (mAbs) based on the methods of Kessler (28).

For RNA blotting, total RNA (27) was fractionated on a 1% agarose/formaldehyde gel, transferred to GeneScreen filters (New England Nuclear, Boston, MA), hybridized with ³²P-labeled probe, and washed according to the instructions of the manufacturer. cDNA inserts were radiolabeled with [α -³²P]dCTP using a random oligo-labeling kit (IBI, New Haven, CT).

Both strands of the λ C-86 insert were sequenced in M13mp18 and M13mp19 (29) using the dideoxy method (30) with M13 universal and custom-synthesized primers (Cornell Department of Microbiology and Research Genetics, Inc., Huntsville, AL). DNA sequences from individual gel readings were aligned and manipulated using the programs of Staden (31). Searches of the Dayhoff protein sequence database were performed with FASTA (32) and pairwise alignments of sequences were generated with LFASTA (32).

RESULTS

In the primary screen of 10⁵ recombinants from the λ gt11 library, four immunoreactive clones were isolated. The clone

Abbreviations: MHC, myosin heavy chain; PM, pectoralis muscle; N-CAM, neural cell adhesion molecule; smMLCK, smooth muscle myosin light chain kinase; Fn, fibronectin; mAb, monoclonal antibody.

*The sequence reported in this paper has been deposited in the GenBank data base (accession no. M31209).

designated λ C-86, which contained a 3-kilobase (kb) insert, exhibited the strongest immunoreactivity and was selected for further characterization. In addition to the antiserum used to screen the library, another polyclonal antiserum directed against gel-purified C-protein, an affinity-purified C-protein-specific polyclonal antibody prepared from antiserum against crude (DEAE column-purified) C-protein, and the fast C-protein-specific mAb, MF-1 (16), reacted with the protein encoded by λ C-86. The epitope recognized by MF-1 has been localized to a 95-kDa chymotryptic peptide of C-protein (33).

Further proof that λ C-86 encoded C-protein was obtained by hybridization of the insert to RNA fractions enriched in C-protein mRNA isolated from methylmercury hydroxide/agarose gels (unpublished data) and a message-selection assay (Fig. 1). A pUC18 plasmid (29) containing the λ C-86 insert, pC-86, specifically selected mRNA from total PM RNA that directed the translation of C-protein in a rabbit reticulocyte lysate (Fig. 1A). This translation product was identified as C-protein by its comigration on SDS/polyacrylamide gels with adult chicken PM C-protein and its immunoprecipitation with mAb MF-1 (Fig. 1A, lane 4). As controls in these message-selection experiments, other plasmids were tested for their ability to select mRNA. A plasmid containing a cDNA insert encoding MHC, pMHC1 (35), efficiently selected myosin mRNA (Fig. 1B), whereas a pUC18 clone containing a cDNA of an unidentified low molecular weight RNA and pUC8 (36) without an insert failed to select translatable mRNA (data not presented).

In Northern blot analysis, the λ C-86 insert specifically hybridized to RNA of about 4 kb from 19-day embryonic, 1-week posthatch, and adult chicken PM and to a message of similar size from adult rat quadriceps muscle (Fig. 2). These results are consistent with studies of C-protein expression using immunological techniques (17). Hybridization was not

FIG. 1. Analysis of translation products from RNA hybrid-selected by pC-86. One-week posthatch chicken PM poly(A)⁺ RNA and hybrid-selected RNA translation products were immunoprecipitated with either MF-1 (C-protein specific) or MF-20 (MHC specific) (34) mAbs. (A) Lane 1, poly(A)⁺ RNA translation products; lane 2, immunoprecipitation of poly(A)⁺ RNA translation products with MF-1; lane 3, translation products of RNA hybrid selected by pC-86; lanes 4 and 5, immunoprecipitation of translation products from RNA hybrid selected by pC-86 with MF-1 and MF-20, respectively. (B) Lane 1, poly(A)⁺ RNA translation products; lane 2, immunoprecipitation of poly(A)⁺ RNA translation products with MF-20; lane 3, translation products of RNA hybrid selected by the MHC cDNA clone pMHC1; lanes 4 and 5, immunoprecipitation of translation products from RNA hybrid selected by pMHC1 with MF-20 and MF-1, respectively. The positions of MHC (M) and C-protein (C) are indicated. The time of film exposure for A, lanes 4 and 5, is 1.5 times that of the other lanes. The low molecular mass bands in lanes 3 of A and B were endogenous components of the translation system. [Reproduced from ref. 20 with modification and permission (copyright Alan R. Liss).]

FIG. 2. Northern blot analysis of chicken and rat RNA probed with λ C-86. The total RNA samples were isolated from 13-day embryonic chicken PM (lane 1), 19-day embryonic chicken PM (lane 2), 1-week posthatch chicken PM (lane 3), adult chicken PM (lane 4), 1-week posthatch chicken brain (lane 5), 13-day embryonic chicken brain (lane 6), and rat quadriceps muscle (lane 7). One microgram of chicken total RNA and 4 μ g of rat total RNA were loaded per lane. The positions of 28S rRNA and 18S rRNA are indicated. [Reproduced from ref. 20 with modification and permission (copyright Alan R. Liss).]

detected against total RNA from chicken brain. The length of the RNA detected by the probe is sufficient to encode a 140-kDa protein. Thus, the λ C-86 insert represents $\approx 75\%$ of the fast C-protein mRNA molecule.

The sequence of the λ C-86 insert is 3045 nucleotides long. It contains a single open reading frame that encodes 992 amino acids (Fig. 3) with a calculated molecular mass of 111,372 Da ($\approx 80\%$ of the C-protein molecule). The insert does not include about 30 kDa of the amino-terminal end of the protein. A putative poly(A) signal (AATAAA) (37) occurs near the end of the 68-nucleotide 3' untranslated region (Fig. 3) but the clone lacks a poly(A) tail.

Existence of a single base before the start of the open reading frame of the insert suggests that the encoded protein is not in the reading frame with β -galactosidase in the λ gt11 vector (21) and thus may not be expressed as a fusion protein. Expression of the encoded protein may reflect internal initiation of translation within the fusion transcript (38).

The derived amino acid sequence of λ C-86 includes a region near the amino-terminal end that is almost identical to a partial amino acid sequence of a 5-kDa chymotryptic fragment (distinct from the 95-kDa fragment containing the MF-1 epitope) of adult chicken fast C-protein (F. C. Reinach, D.A.F., and M. Elzinga, unpublished data; ref. 33). Fifteen of the 17 amino acids of the peptide match the derived amino acid sequence of λ C-86, thus providing conclusive proof that λ C-86 encodes C-protein (Fig. 3). The reason for the two mismatches is unknown but may reflect inaccuracies in the peptide sequence, differences between the sequences of adult and 1-week-old chicken PM C-protein, or allelic variation.

One of the most striking features of the derived amino acid sequence of λ C-86 is that it contains six regions that resemble the sequence motifs of immunoglobulin-like domains (19) (Figs. 4 and 5). The proportion of identical amino acids among the six immunoglobulin-like domains ranges from 13% (domain III vs. domain VII) to 37% (domain IV vs. domain VII). Tryptophan, isoleucine, tyrosine, and valine residues that correspond to positions 143, 169, 180, and 202 of domain I, respectively, are conserved in all of the domains (Fig. 4A).

In comparison with other members of the immunoglobulin superfamily, the immunoglobulin-like domains of C-protein are most similar to domains III and IV of chicken neural cell

	10	20	30	40	
AspSerCysSerPheSerIleAspValGluAlaProArgSerSerGluGlyAsnValLeuGlnAlaPheLysArgThrGlyGluGlyLysAspAspThrAlaGlyGluLeuAspPheSer					12
GGACAGCTGCTCCTTCAGCATCGACTGCAGGCCCCCGCTCCAGTGAAGGCATGTCTCGAGGCCTCAAGCGCACGGGTGAAGCAAGATGACATCGCAGGAGAGCTTGACATTGACC					24
* *	50	60	70	80	
GlyLeuLeuLysLysArgGluValGlnValGluGluLysLysLysLysAspGluAspAspGlnPheProProGluIleTrpGluLeuLeuLysGlyValThrLysLysSerGluTyr					24
GGGCTGCTCAAGAAGAGGAGCTTCAGCTGCGAGCAGAAGAAGAAGAAGAACAGGACGATCAGTTCCCCCCCGAGATCTGGGAGCTGCTGAAGGGCGTCAACAAGAAGAGOGAATAC					24
	90	100	110	120	
GluArgIleAlaPheGlnTyrGlyIleThrAspLeuArgGlyMetLeuLysArgLeuLysLysValHisValGluProLysLysSerGluAlaPheIleArgLysLeuAspProAlaTyr					36
GAACGGATGCGCTTCCAGTAACGGATCAACGGACCTAGGGGCGATGCTCAAACGGCTCAAAAGGTGCAAGTGGAGGCCAAAAAGAGCGCAAGCGTTTCATCGGAAGTTGGACCGGCTAC					36
	130	140	150	160	
GlnValAspLysGlyAsnLysIleLysLeuValValGluLeuSerAspProAspLeuProLeuLysTrpTyrLysAsnGlyGlnLeuLeuLysSerThrLysTyrValPheGluAsn					48
CAGTGGGACAAAAGGAACAAAATCAAACTGGTGGAGCTGAGCGACCCCGACCTGCCCTTAAAGTGTGTAAGAAGCGGGCAGCTGCTGAACCCAGCAGCAAGTACGCTTTTGAGAAC					48
	170	180	190	200	
ValGlyLeuLysArgIleLeuThrIleHisLysCysSerLeuAlaAspAlaAlaTyrGluCysArgValAsnAspGluLysCysPheThrGluValPheValLysGluProProVal					60
GTGGGGCTGAAGCGGATCTCCACCTCCATAATGCTCTCTGCGCCGACGACCGCGCTTAAAGTGTGTAAGAAGCGGGCAGCTGCTGAACCCAGCAGCAAGTACGCTTTTGAGAAC					60
	210	220	230	240	
ThrValValArgGlyLeuGluAspGlnGlnValValGlyAspArgValValLeuGluAlaGluValSerGluGluGlyAlaGlnValMetTrpLeuLysAspGlyValAspValThr					72
AACGCTGGTGGCTGGCTGGAGGACAGCAGCTGGTGGGGGACCGGCTGGTGTGGAGCGGAGGCTCCGAGGAGGGGGCGGAGTCACTGTGGCTGAAGGACGGGCTGGACGCTGAGC					72
	250	260	270	280	
ArgAspAspAlaPheLysTyrArgPheLysLysLysLysHisPheLeuIleIleAsnGluAlaGluLeuSerAspSerAlaHisTyrLysIleMetThrAsnGlyGlyGluSer					84
CGGGAAGCAGCGCTTAACTACCGCTTAAAGAAGGACGGCAAAAGCATTTCTCATCATCAAGAGAGGGAGCTCAGCGACAGCCCACTACAAAATCATGACCAACGGGGGAGAGC					84
	290	300	310	320	
GluAlaGluLeuSerValGluGlnLysGlnLeuValLeuGlnAspMetAlaGluLeuThrValLysAlaSerGluGlnAlaValPheLysCysGluValSerAspGluLysValThr					96
GAGCGGAGCTCAGCGTGAAGAGAAGCAGCTGGAGGTGCTGCAGGACATGGCGGATCTGACGGTGAAGGGTGGAGCGGCGGTGTTAAAGTGGAGGTGTCGGAAGGAGAGGTGACC					96
	330	340	350	360	
GlyArgTrpPheArgAsnGlyValGluValLysProSerLysArgIleHisIleSerHisLysGlyArgGlyPheHisLysValIleAspAspValArgProGluAspGluGlyAspTyr					108
GGCAGTGGTTCGCAACCGGGTGGAGTCAAAACCCAGAGGATCCACATCTCACACAAAGGAGGTTCACAAAGCTGGTGTGATCGAAGAGCGCCCAAAATCCACCTGGAGTCTCGGGGAAG					108
	370	380	390	400	
ThrPheIleAspGlyTyrAlaLeuSerSerAlaLysLeuAsnPheLeuGluIleLysValGluTyrValProLysGlnGluProProLysIleHisLeuAspCysThrGlyLys					120
ACCTTCATCCGGGACGGATACGGCTTCTCACTGTCCGAAAGCTCAACTTCTCGAGATCAAAAGTGAAGTACGTCGCCCAACAGAGCGCCCAAAATCCACCTGGAGTCTCGGGGAAG					120
	410	420	430	440	
AlaAlaGluAsnThrIleValValValAlaGlyAsnLysValArgLeuAspValProIleSerGlyGluAlaAlaProThrValThrTrpLysArgGlyGlyPheThrAlaThr					132
GGGGCGAGAACACCATGCTGGTGGTGGCGGGGAACAAAGTGGCGCTGAGCGTCCCATCTCGGGGAGCGCGCGCCACCGTCACTGGAAGAGGGGGATCAGCTGTTCAACGGCGAGC					132
	450	460	470	480	
GluGlyArgValHisIleAspSerGlnAlaAspLeuSerSerPheValIleGluSerAlaGluArgSerAspGluGlyArgTyrCysIleThrValThrAsnProValGlyGluAspSer					144
GAGGGCGGGTGCACATCGACTCTCAGGGGAGCTCAGCAGCTTTGTGATTGAGAGCGGGAGCGAGGATGAGGGGGCTACTGATCAGCGTCACCAACCGGCTGGGGGAGGACTCC					144
	490	500	510	520	
AlaThrLeuHisValArgValValAspValProAspProProGlnSerValArgValThrSerValGlyGluAspTrpAlaValLeuSerTrpGluAlaProProPheAspGlyGlyMet					156
GCCACCTTACAGTCCGGTGGTGGAGCTCCGGACCCCGCGAGTCCGTTCCGGTGACGTCGTTGGGGAGGATTGGGGGCTGCTCAGTGGGAGGGCGGCCATTGAGCGGGGATG					156
	530	540	550	560	
ProIleThrGlyTyrLeuMetGluArgLysLysGlySerMetArgTrpMetLysLeuAsnPheGluValPheProAspThrThrTyrGluSerThrLysMetIleGluGlyValPhe					168
CCCATCACCGGCTACCTGATGGAGCGCAAGAAGAGGCTCCATCGGCTGGATGAAGCTGAAGTTCGAGGTTTCCCGACACCACCTACGAGTCCACCAAAATGATCGAGGGCTCTTC					168
	570	580	590	600	
TyrGluMetArgValPheAlaValAsnAlaIleGlyValSerGlnProSerLeuAsnThrGlnProPheMetProIleAlaProThrSerGluProThrHisValValLeuGluAspVal					180
TACGAGATCGGGCTCTTCGCGCTCAACGCCATCGGGGTGCGCAGCCAGCTCAACAGCGAGCCCTTTATGCCATCGGCGGACGAGGAAACGACCCACGCTGGTGTGGAGGACGTC					180
	610	620	630	640	
ThrAspThrThrAlaThrIleLysTrpArgProProGluArgIleGlyAlaGlyGlyValAspGlyTyrLeuValGluTrpCysArgGluGlySerAsnGluTrpValAlaAlaAsnThr					192
ACCGACACCGCCGCCCATCAATGAGAGCCCGGAGCGCATGGGGGGGGCGGGCTGACGGATACCTGGTGGAGTGGTGGCGGAGGATCCAATGAATGGTGGCGGCCAACACG					192
	650	660	670	680	
GluLeuValGluArgCysGlyLeuThrAlaArgGlyLeuProThrGlyGluArgLeuLeuPheArgValIleSerValAsnMetAlaGlyLysSerProProAlaThrMetAlaGlnPro					204
GAGTGTGGTAGCGCTGTGGGCTGACGGCGCGGGCTCTCCCAAGGGGAGCGGCTGCTCTTCGGGTTCATCAGGCTCAACATGGGGGCAAAAGGCCCGCCCGCCCATGGCCCAACCC					204
	690	700	710	720	
ValThrIleArgGluIleValGluArgProLysIleArgLeuProArgHisLeuArgGlnThrTyrlleArgArgValGlyGluGlnValAsnLeuValIleProPheGlnGlyLysPro					

FIG. 3. Complete nucleotide and derived amino acid sequences of the λ C-86 insert. The region of the derived amino acid sequence marked by a single underline matches that of a 5-kDa chymotryptic peptide of fast C-protein from adult chicken PM (see text). The asterisks indicate those positions that differ between the derived and peptide sequences. The peptide residues corresponding to positions 46 and 58 of the derived sequence are glycine and threonine, respectively. A putative poly(A) signal is indicated by a double underline.

adhesion molecule (N-CAM) (39) in which 11% (C-protein II vs. N-CAM IV) to 28% (C-protein IV vs. N-CAM III) of the amino acids match (Fig. 4A). Similar, although less extensive, levels of identity occur between the immunoglobulin-like domains of C-protein and two other adhesion molecules.

neural adhesion molecule L1 (L1) (40) and myelin-associated glycoprotein (MAG) (41, 42). Thirty percent identity was observed between C-protein IX (residues 913–980) and L1 III (residues 256–320); a similar level of identity was seen when comparing C-protein IV (residues 411–485) with MAG III

FIG. 4. Sequence alignments of the AC-86 domains. (A) Regions of the derived amino acid sequences of the six immunoglobulin-like domains encoded by AC-86 (C), an immunoglobulin-like domain detected in chicken smMLCK (M), and chicken N-CAM (N) domains III and IV. Positions are shaded where residues are identical in any two sequences except where they occur only in the N-CAM domains. Arrowheads indicate positions of highly conserved residues among the immunoglobulin domains of the C2-set (19). (B) Regions of the derived amino acid sequences of the three Fn-like domains encoded by AC-86 (C), a Fn-like domain detected in chicken smMLCK (M), and human Fn (F) type III repeats, III-3 and III-12. Positions are shaded where residues are identical in any two sequences except where they occur only in the human Fn repeats. Arrowheads indicate the positions of three highly conserved residues in Fn type III repeats (see text). The first and last amino acid positions of each sequence are indicated along the margins (refer to Fig. 5 for the actual boundaries of each domain).

(42) (residues 254–322). Further resemblance to extracellular proteins known to mediate cell adhesion is reflected by the presence of an Arg-Gly-Asp (RGD) sequence (residues 432–434) in C-protein IV (43). The function of this tripeptide sequence in C-protein, if any, is unknown.

In addition to immunoglobulin-like motifs, the portion of C-protein encoded by AC-86 contains three regions that resemble fibronectin (Fn) type III repeats (44) (Figs. 4 and 5). All three include the conserved tryptophan characteristic of type III repeats, but only domains V and VIII exhibit the conserved tyrosine. A conserved leucine residue found in Fn type III repeats is also present in two of the C-protein repeats (Fig. 4B). Unlike Fn type III repeats, which contain only a single tryptophan, each of the C-protein Fn-like domains contains two additional tryptophans. The sequences of C-protein domains V, VI, and VIII exhibit between 26% and 32% identity to each other and 19% and 26% identity to the most similar type III repeats in human Fn (repeats III-3 and III-12) (44).

Another protein, chicken smooth muscle myosin light chain kinase (smMLCK) (38), was found in a computer search to exhibit significant similarity to the derived amino acid sequence of AC-86. Upon comparing the sequences, it was seen that smMLCK contains regions similar to the immunoglobulin-like (15–30% matching residues) and Fn-like (22–27% matching residues) domains of C-protein (Fig. 4). The existence of such domains in smMLCK has not been reported. Only one of each type of homology was observed in the published smMLCK sequence, which encodes the carboxyl-terminal 60% of the protein [these domains and an additional partial immunoglobulin-like domain have been identified independently by Benian and co-workers (45)]. No sequences related to the catalytic or calmodulin-binding domains of smMLCK were detected in C-protein.

DISCUSSION

The major finding of this study is that C-protein is an intracellular member of the immunoglobulin superfamily (19). Like several of the CAMs in this superfamily, such as

FIG. 5. Structural map of the immunoglobulin-like (clear boxes) and Fn-like (hatched boxes) domains encoded by AC-86. The lengths of domains I–IX are 90, 92, 91, 103, 96, 96, 97, 95, and 101 residues, respectively. Arrows indicate the amino acid corresponding to the first position of each domain.

N-CAM (39), L1 (40), and fasciclin II (46), C-protein contains immunoglobulin-like and Fn type III homologies. The immunoglobulin-like domains of C-protein and the CAMs are of the C2 type (19). C2 domains share structural features of variable and constant immunoglobulin domains and exhibit highly conserved residues at 14 positions, including a pair of cysteines that, in the variable and constant domains, form intrachain disulfide bonds between the β -sheets of the immunoglobulin fold (19, 47). Only C-protein IX has all of the conserved residues characteristic of C2 domains. The other immunoglobulin-like domains of C-protein contain most of the conserved residues and three have at least one cysteine in a conserved position (Fig. 4). Hydrophobic residues occur at the positions that lack the conserved cysteines, as found in similar domains of several other members of the immunoglobulin superfamily (19).

At present, it is unknown whether the secondary and tertiary structures of the C-protein immunoglobulin and Fn-related domains are similar to their counterparts in other proteins. Immunoglobulin domains (19, 47) and Fn type III repeats (48) are rich in β -sheet structure. Consistent with its content of immunoglobulin-like and Fn-like homologies, C-protein is 50% β -structure as determined by circular dichroism (1). However, the secondary structures of individual domains have not been established and there is no evidence that C-protein contains intrachain disulfide bonds (1). Conclusive proof that the tertiary structure of the immunoglobulin-like domains of C-protein resembles the immunoglobulin fold comprised of two β -sheets that typify variable and constant domains of the immunoglobulins will require x-ray crystallography.

In addition to C-protein, it was shown that chicken smMLCK also contains immunoglobulin and Fn-related domains. The inclusion of C-protein and smMLCK in the immunoglobulin superfamily takes on added significance in light of recent findings that three other thick filament-associated proteins have similar domains. These proteins are twitchin, the 600-kDa protein encoded by the *unc-22* gene in *Caenorhabditis elegans* (45, 49); titin, the megadalton-sized elastic protein of vertebrate striated muscle (50) (J. Trinick, personal communication); and an 86-kDa protein that is colocalized with C-protein in most of its stripes along thick filaments of chicken PM (51) (K. T. Vaughan, S.E., and D.A.F., unpublished data).

The functions of the immunoglobulin and Fn-related domains in thick filament-associated proteins are unknown. Although one function may be to bind myosin, another

intriguing possibility is that the immunoglobulin-like domains confer homophilic and/or heterophilic binding properties analogous to their proposed roles in the CAMs (39, 40) and other members of the immunoglobulin superfamily (19). The concept of homophilic interactions between C-protein immunoglobulin-like domains would be compatible with models in which C-protein molecules encircle the thick filament and interact in an end-to-end or side-to-side fashion (2, 3) or extend outward to form links with C-protein molecules in adjacent thick filaments (1, 2, 52). Similarly, the previously hypothesized interactions between C-protein and titin (50) and 86-kDa protein (51) could be mediated by heterophilic interactions between the immunoglobulin-like domains of these proteins.

In considering such binding mechanisms, it is interesting to note that the arrangements of the immunoglobulin and Fn-like domains of N-CAM (39, 53), L1 (40), and fasciclin II (46) are similar to those of C-protein domains I–VI. All of these proteins exhibit Fn-like domains adjacent to a tandem array of immunoglobulin-like domains in their amino-terminal halves. A flexible hinge region, which may stabilize homophilic interactions, is located between these sets of domains in N-CAM (53). Given that C-protein contains a similar flexible region located in a pericentric position along the molecule (52, 54), it is conceivable that its hinge also occurs at this site and serves a comparable function.

Beyond the functional implications, the finding that C-protein, titin, twitchin, and smMLCK are members of the immunoglobulin superfamily may provide new insights into the evolutionary origins of the CAMs and immunoglobulins. It has been proposed that members of the immunoglobulin superfamily evolved by a combination of gene duplication and divergence from a primordial immunoglobulin domain (19). Williams (19) and Edelman (55) have further postulated that the immunoglobulins arose from an extracellular recognition system akin to that comprised by modern CAMs. Since at least two vertebrates (chicken and rabbit) and an evolutionarily primitive invertebrate (*C. elegans*) are now known to have intracellular cytoskeletal proteins that contain immunoglobulin-like domains, the possibility must be considered that these proteins evolved prior to or in parallel with the CAMs.

We express appreciation to F. Reinach and M. Elzinga for partial amino acid sequence analysis of PM C-protein. S.E. thanks F. Reinach for many helpful discussions, K. Wilkerson for guidance during preparation of the cDNA library, S. Goldfine for participating in preparation of the antisera and providing reticulocyte lysates, B. Cunningham for helpful discussions of the sequence comparisons and pointing out the RGD sequence, and G. Benian and J. Trinick for sharing their unpublished data. We also thank A. Popowicz, L. Van Houten, R. Mahajan, M. Rosenblum, K. T. Vaughan, and T. Milner for their help with the manuscript. This work was supported by grants from the National Institutes of Health (AR32147) and Muscular Dystrophy Association and was presented in partial fulfillment of Ph.D. requirements by S.E.

- Offer, G., Moos, C. & Starr, R. (1973) *J. Mol. Biol.* **74**, 653–676.
- Offer, G. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **37**, 87–95.
- Craig, R. & Offer, G. (1976) *Proc. R. Soc. Lond. Ser. B* **192**, 451–461.
- Dennis, J. E., Shimizu, T., Reinach, F. C. & Fischman, D. A. (1984) *J. Cell Biol.* **98**, 1514–1522.
- Bennett, P., Craig, R., Starr, R. & Offer, G. (1986) *J. Muscle Res. Cell Motil.* **7**, 550–567.
- Starr, R. & Offer, G. (1978) *Biochem. J.* **171**, 813–816.
- Moos, C., Offer, G., Starr, R. & Bennett, P. (1975) *J. Mol. Biol.* **97**, 1–9.
- Moos, C., Mason, C. M., Besterman, J. M., Feng, I. M. & Dubin, J. H. (1978) *J. Mol. Biol.* **124**, 571–586.
- Moos, C. (1981) *J. Cell Biol.* **90**, 25–31.
- Yamamoto, K. (1986) *FEBS Lett.* **208**, 123–127.
- Hartzell, H. C. & Titus, L. (1982) *J. Biol. Chem.* **257**, 2111–2120.
- Moos, C. & Feng, I. M. (1980) *Biochim. Biophys. Acta* **632**, 141–149.
- Yamamoto, K. & Moos, C. (1983) *J. Biol. Chem.* **258**, 8395–8401.
- Hartzell, H. C. (1985) *J. Mol. Biol.* **186**, 185–195.
- Callaway, J. E. & Bechtel, P. J. (1981) *Biochem. J.* **195**, 463–469.
- Reinach, F. C., Masaki, T., Shafiq, S., Obinata, T. & Fischman, D. A. (1982) *J. Cell Biol.* **95**, 78–84.
- Obinata, T., Reinach, F. C., Bader, D. M., Masaki, T., Kitani, S. & Fischman, D. A. (1984) *Dev. Biol.* **101**, 116–124.
- Takano-Ohmuro, H., Goldfine, S. M., Kojima, T., Obinata, T. & Fischman, D. A. (1989) *J. Muscle Res. Cell Motil.* **10**, 369–378.
- Williams, A. F. (1987) *Immunol. Today* **8**, 298–303.
- Einheber, S. & Fischman, D. A. (1989) in *Cellular and Molecular Biology of Muscle Development*, eds. Kedes, L. H. & Stockdale, F. E. (Liss, New York), Vol. 93, pp. 523–534.
- Young, R. A. & Davis, D. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 197–198.
- Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
- Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning: A Practical Approach*, ed. Glover, D. M. (IRL, Oxford), Vol. 1, pp. 49–78.
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
- Miller, J. S., Paterson, B. M., Ricciardi, R. P., Cohen, L. & Roberts, B. E. (1983) *Methods Enzymol.* **101**, 650–674.
- Bouché, M., Goldfine, S. & Fischman, D. A. (1988) *J. Cell Biol.* **107**, 587–596.
- Kessler, S. W. (1981) *Methods Enzymol.* **73**, 442–459.
- Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26**, 101–106.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Staden, R. (1982) *Nucleic Acids Res.* **10**, 4731–4751.
- Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
- Reinach, F. C. (1984) Dissertation (Cornell Univ. Graduate School of Medical Sciences, New York, NY).
- Bader, D., Masaki, T. & Fischman, D. A. (1982) *J. Cell Biol.* **95**, 763–770.
- Einheber, S., Raman, G., Nickowitz, R. & Fischman, D. A. (1988) *J. Cell Biochem., Suppl.* **12**, 366 (abstr.).
- Vieira, J. & Messing, J. (1982) *Gene* **19**, 259–268.
- Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
- Guerriero, V., Russo, M. A., Olson, N. J., Putkey, J. A. & Means, A. R. (1986) *Biochemistry* **25**, 8372–8381.
- Cunningham, B. A., Hemperly, J. J., Murray, B. A., Prediger, E. A., Brackenbury, R. & Edelman, G. M. (1987) *Science* **236**, 799–806.
- Moos, M., Tacke, R., Scherer, H., Teplow, D., Fruh, K. & Schachner, M. (1988) *Nature (London)* **334**, 701–703.
- Arquint, M., Roder, S., Chia, L. S., Down, J., Wilkinson, D., Bayley, H., Braun, P. & Dunn, R. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 600–604.
- Salzer, J. L., Holmes, W. P. & Colman, D. R. (1987) *J. Cell Biol.* **104**, 957–965.
- Ruoslahti, E. & Pierschbacher, M. D. (1986) *Cell* **44**, 517–518.
- Petersen, T. E., Skorstengaard, K. & Vibe-Pedersen, K. (1989) in *Fibronectin*, ed. Mosher, D. F. (Academic, New York), pp. 1–24.
- Benian, G. M., Kiff, J. E., Neckelmann, N., Moerman, D. G. & Waterson, R. H. (1989) *Nature (London)* **342**, 45–50.
- Harrelson, A. L. & Goodman, C. S. (1988) *Science* **242**, 700–708.
- Amzel, L. M. & Poljak, R. J. (1979) *Annu. Rev. Biochem.* **48**, 961–997.
- Odermatt, E. & Engel, J. (1989) in *Fibronectin*, ed. Mosher, D. F. (Academic, New York), pp. 25–47.
- Moerman, D. G., Benian, G. M., Barstead, R. J., Schrieffer, L. A. & Waterson, R. H. (1988) *Genes Dev.* **2**, 93–105.
- Wang, K. (1980) in *Cell and Muscle Motility*, ed. Shay, J. W. (Plenum, New York), pp. 315–369.
- Bahler, M., Eppenberger, H. M. & Walliman, T. (1985) *J. Mol. Biol.* **186**, 393–401.
- Hartzell, H. C. & Sale, W. S. (1985) *J. Cell Biol.* **100**, 208–215.
- Becker, J. W., Erickson, H. P., Hoffman, S., Cunningham, B. A. & Edelman, G. M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1088–1092.
- Swan, R. C. & Fischman, D. A. (1986) *J. Muscle Res. Cell Motil.* **7**, 160–166.
- Edelman, G. M. (1987) *Immunol. Rev.* **100**, 11–45.