Arch. Pharm. Res. Vol.21, No.6, pp. 712-717, 1998

Nucleotide and Deduced Amino Acid Sequences of Rat Myosin Binding Protein H (MyBP-H)

Jaehoon Jung, Jihyun Oh and Kyunglim Lee*

College of Pharmacy, Ewha Womans University, Seoul, 120-750, Korea

(Received August 18, 1998)

The complete nucleotide sequence of the cDNA clone encoding rat skeletal muscle myosin-binding protein H (MyBP-H) was determined and amino acid sequence was deduced from the nucleotide sequence (GenBank accession number AF077338). The full-length cDNA of 1782 base pairs(bp) contains a single open reading frame of 1454 bp encoding a rat MyBP-H protein of the predicted molecular mass 52.7kDa and includes the common consensus 'CA_ _ TG' protein binding motif. The cDNA sequence of rat MyBP-H show 92%, 84% and 41% homology with those of mouse, human and chicken, respectively. The protein contains tandem internal motifs array (-FN III-Ig C2-FN III- Ig C2-) in the C-terminal region which resembles to the immunoglobulin superfamily C2 and fibronectin type III motifs. The amino acid sequence of the C-terminal Ig C2 was highly conserved among MyBPs family and other thick filament binding proteins, suggesting that the C-terminal Ig C2 might play an important role in its function. All proteins belonging to MyBP-H member contains 'RKPS' sequence which is assumed to be cAMP- and cGMP-dependent protein kinase A phosphorylation site. Computer analysis of the primary sequence of rat MyBP-H predicted 11 protein kinase C (PKC) phosphorylation site, 7 casein kinase II (CK2) phosphorylation site and 4 N-myristoylation site.

Key word : Myosin binding protein H (MyBP-H), Immunoglobulin superfamily C2, Fibronectin type III

INTRODUCTION

Vertebrate striated muscle sarcomere is consisted of the thick filaments which interact with the thin filaments to produce the cross-bridge of the A-band resulting in muscle contraction. Although several abundant thick filament-associated proteins have been known, myosin is the reigning component of the thick filament. MyBP-C, MyBP-H (previously termed 86 kDa protein) and MyBP-X (slow type isoform of MyBP-C) have been identified as a family of myosin binding proteins which regulate the function of myosin (Starr and Offer, 1993; Vaughan *et al.*, 1993; Weber *et al.*, 1993; Yamamoto, 1984). In addition, titin, twitchin, telokin and smooth muscle myosin light chain kinase (smMLCK) have been identified as myosin-binding protein (Benian *et al.*, 1989; Kobayashi *et al.*, Labeit *et al.*, 1990, 1992; Olson *et al.*, 1990; Shirinsky *et al.*, 1983).

MyBP-H gene is known to be expressed exclusively in skeletal muscle, but it is also expressed during Purkinje fibers formation (Schiaffno, 1997). MyBP-H has an extensive homology with MyBP-C which contains the conserved immunoglobulin superfamily C2 and fibronectin type III motifs, as do other members of MyBPs, such as titin, twitchin, telokin, and skelemin (Smith and Xue, 1997; Vaughan *et al.*, 1993). The H-proteins generally have a tandem array between two Ig C2 and between two FN III repeat, while the C-proteins have a pattern of seven Ig C2 and three FN III motifs (Vaughan *et al.*, 1993). The function of these repeat structure has not yet been determined, although the Ig C2 set of the C-terminal 14 kDa peptide has been known to play a central role in binding myosin with high affinity (Okagaki *et al.*, 1993).

In vitro studies have shown that the addition of purified rabbit skeletal muscle MyBP-C to synthetic thick filaments decreases the critical concentration of myosin required for filament formation and produces filaments of greater length and more uniform diameter than those formed in the absence of MyBP-C (Davis, 1988). Additional studies have shown that both of the MyBP-C and MyBP-H are able to bind to the complex of myosin and actin filaments to modulate the myosin ATPase activity (Moos and Feng, 1980; Yamamoto, 1984). Tatiana *et al.* (1997) extracted a portion of MyBP-C from chicken skeletal muscle to cause an increased number of MyBP-H bound to the myosin.

Correspondence to: Kyunglim Lee, College of Pharmacy, Ewha Womans University, Seoul, 120-750, Korea

They have noticed that although MyBP-H is far less abundant protein than MyBP-C in both skeletal and cardiac muscle, the binding of MyBP-H can inhibit the binding of MyBP-C to the reconstituted filaments.

In this study, we report the full-length nucleotide sequences of a rat skeletal muscle MyBP-H. We have also analyzed the secondary structure of the deduced amino acid sequence compared to the MyBP family proteins and other muscle binding proteins.

MATERIALS AND METHODS

cDNA library construction and screening of rat skeletal MyBP-H cDNA

Total cytoplasmic RNA from rat skeletal muscle was prepared by the acid guanidium thiocyanate-phenol-chloroform extraction method (Chomczynski and Sacchi, 1987; Oh *et al.*, 1995: Park *et al.*, 1995). Poly (A)+ mRNA was purified by poly A Tract® mRNA isolation system from Promega (USA). The cDNA was made using Zap-cDNA synthesis kit (Stratagene, La Jolla, CA, USA). The size-fractionated cDNA was ligated into the EcoRI- and XhoI-cut pJG4-5 library vector that contains a Gal1 promoter, SV40 T antigen nuclear localization sequence, the B42 activation domain and the HA1 epitope at the NH_2-terminus of restriction site. The ligated DNA was transformed into Sure cells (Stratagene, La Jolla, CA, USA) by electroporation, using a Bio-Rad electroporator according to the manufacturer's instructions. The transformed colonies of 9.7×10^6 were obtained from the plates. The cDNA was screened by the Yeast Two-hybrid system to find out proteins interacting with the 4^{th} cytoplasmic domains of (Na, K) ATPase. Therefore rat MyBP-H clone was initially isolated as the (Na, K) ATPase-interacting protein.

DNA sequencing

cDNA sequencing was performed by customer service of Bioneer (Korea), using automatic sequencer. The sequencing primers are as follows: next to the *Eco*RI site primer 5'- TACCCTTAtGATGTGCCA-3', (-) strand sequencing primer of pJG 4-5 5'-AAGCTTCTCGAG-TTTTTT-3', primer R1 5'-ATCCTCAGGTCTGGAGGAC-3' (from 1548 to 1567), primer F3 5'-TAGACCAGCCTGT-CCACAT-3' (from 512 to 531), primer 8R-2 5'-AAGGA-GGGCGGTTCTGAG-3' (from 1170 to 1188), primer 10F-4 5'-CTATTGGATGTCTGGGGT-3' (from 859 to 877), primer 10F-3 5'-GCTGACCTTGGAGGA- TGAG-3' (from 264 to 283), primer A 5'-AGCGAGACTTCTC-AGAACCG-3' (from 1160 to 1180).

Analysis of the nucleotide and deduced amino acid sequences

The nucleotide sequence data of rat MyBP-H appear in the GenBank database under accession No. AF 077338. Nucleotide sequence similarity searches were performed by Entrez Advanced BLAST search program of current GenBank sequences. Multalign software (ver 5.3.3 copyright I.N.R.A France 1989, 1991, 1994, 1996) was used to compare multiple alignment and SIM program from ExPASy home page Prosite was used for binary sequence alignment in World Wide Web. Protomics tools, ScanProsite-Protein against PROSITE and TRANSFAC browser were provided to scan the nucleotide sequence, peptide pattern, rule and matrix entries in PROSITE from the University of Minesota Biological science net working service. The 3D-structure image was downloaded in PDB (Protein Data Bank; http://www.pdb.bnl.gov/).

RESULTS AND DISCUSSION

Nucleotide and deduced amino acid sequence

The nucleotide sequences of the cDNA encoding rat MyBP-H (1782 bp) contain the entire 1454 bp coding region corresponding to the protein of 484 amino acids. To date, there have been no reports indicating nucleotide binding property of MyBP-H, but it does contain the predicted common consensus sequence of 'CA_ _TG' which recognizes basic helix-loop-helix (bHLH) proteins as found in MyoD and E 2A protein complexes (Blackwer and weintraub, 1990). The deduced amino acid sequence revealed APEP, ASVST, ASEST, KAAI calculated periodic, tandem, separated amino acid alphabet repeats, and two fibronectin type III and two immunoglobulin superfamily C2 motifs, while MyBP-C protein revealed three FN III and seven Ig C2 motifs (Fig. 1).

Primary sequence analysis

The predicted molecular weight is 52.7 kDa and the charge clusters are distributed mainly in the region of 30 to 60 residues. Compositional analysis showed a large number of alanine (9.1%) and proline (9.5%) in the N-terminal region. In spite of the fact that the anomalous electrophoretic mobility in SDS gels which is a property of the A-P-A-P motif in the N-terminus of the 86 kDa protein, we have not found the A-P-A-P repeat sequences in rat MyBP-H.

Multiple alignment analysis of the MyBP proteins from various species has revealed a high degree of amino acid identity, particularly in the C-terminal 40 kDa fragments. The rat H-protein showed 92% DNA sequence homologous to the mouse H-protein. H- and C-proteins of human and those of chicken share the identity with the H-protein of rat as follows; 84% peptide identity plus 17% conserved motif sequence, 51% peptide identity plus 15% conserved motif sequence, 41% peptide identity plus 14% conserved

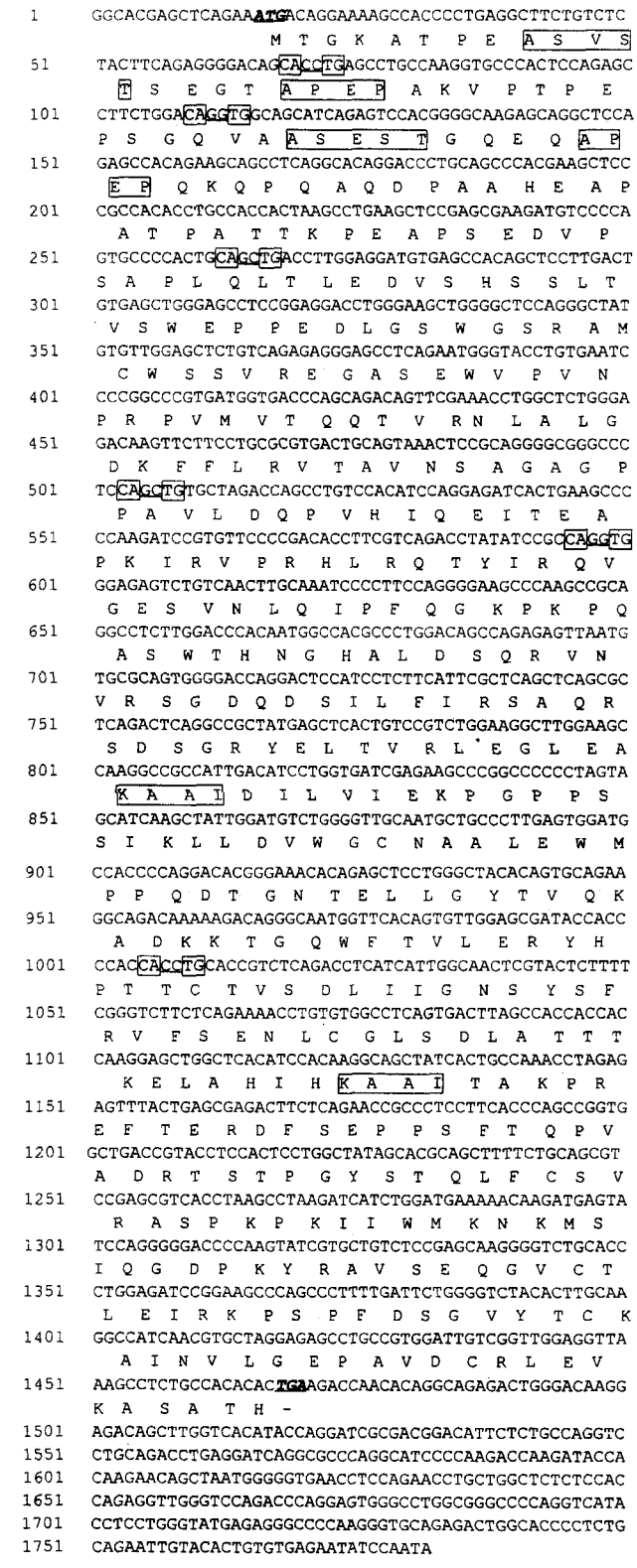

Fig. 1. Nucleotide and deduced amino sequences of rat myosin binding protein H (MyBP-H). The nucleotides that are common consensus sequence, CA--TG, are boxed and underlined. APEP (18-21, 44-47), ASVST (9-13), ASEST (35-39), KAAI (263-266, 370-373) amino acid alphabet repeats between FN III (78-164, 274-359) and Ig C2 (196-255, 403-463) motifs are boxed.

Table I. The homology percentage of the amino acid sequence and the conserved motif homology of human and chicken MyBP-H, and human and chicken MyBP-C compared with rat MyBP-H

	MyBP-H	MyBP-C	Conserved motif			
			FN III		Ig C2	
			MyBP-H	MyBP-C	MyBP-H	MyBP-C
HUMAN	84%	41%	12%	18%	8%	15%
CHICKEN	51%	42%	16%	18%	12%	16%

motif sequence, and 42% peptide identity plus 14% conserved motif sequence, respectively (Table I).

Over the decade, cDNA cloning and primary sequence comparison have shown that many myosin-associated proteins in both vertebrates and invertebrates exhibit similar repeat structure including immunoglobulin superfamily C2 and fibronectin type III motifs which are primarily observed in adhesion molecules, growth factor receptors and proteins involved in viral binding uptake (Cunningham *et al.*, 1987) (Fig. 2A). Projectin, smooth muscle myosin light chain kinase (MLCK), telokin, twitchin, and titin which belong to a novel family of cytoskeleton-associated protein kinase contain Ig C2 and FN III motifs. In our study we have also identified both Ig C2 and FN type III repeats in rat MyBP-H protein. These motifs are found in the globular domain which is involved in isoform-specific interaction of the myosin binding proteins (MyBPs) with skeletal and cardiac myosin (Alyonycheva *et al.*, 1997; Holness and simmons, 1994).

Multiple alignment analysis of the C-terminal Ig C2 repeats of MyBP proteins, telokin, and twitchin has shown that they have highly conserved sequences except the 'RKPS' sequence (Fig. 2B) which is a predicted cAMP- and cGMP-dependent protein kinase A (PKA) phosphorylation site (Glass *et al.*, 1983, 1986). Although the C-terminal Ig C2 motif has been implicated to have the binding activity, its presence without FN III motif is not sufficient to target MyBP-H and -C correctly to the A-band. Thus FN III motif is required for this process, implying that MyBPs are associated with the thick filament through its C-terminus containing both Ig C2 and FN III motifs (Gilbert *et al.*, 1996). Interestingly, all MyBP-H family proteins appear to have the PKA phosphorylation site but other muscle binding proteins do not. Therefore, it seems likely that MyBP-H family proteins are regulated differently from other muscle binding proteins to play a physiological role.

We have found that the deduced amino acid sequence of rat MyBP-H contains a number of prophesied phosphorylation site. Those are 11 protein kinase C (PKC) phosphorylation site, 7 casein kinase II (CK2) phosphorylation site, and 4 N-myristoylation site (Kishimoto *et al.*, 1985; Saijo *et al.*, 1997; Woodget

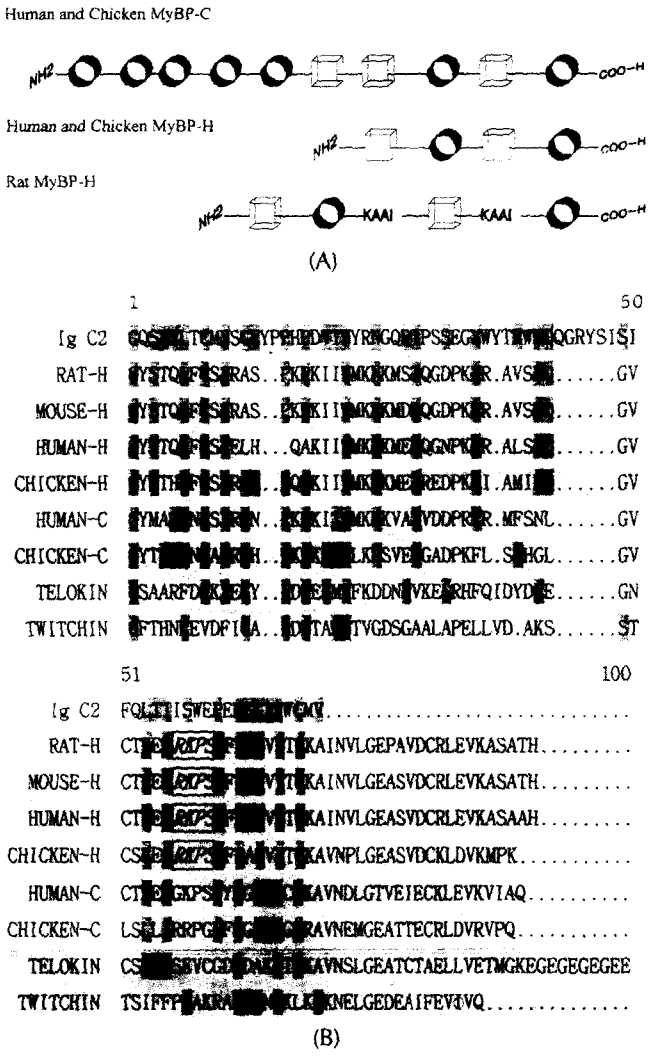

Fig. 2. (A) Schematic representation of MyBP proteins. Regions of homology to the Ig C2 (○) or FN III (■) are shown for full length MyBP-C and MyBP-H. (B) Multiple alignment of the C-terminal Ig C2 repeats of MyBP proteins, telokin and twitchin, and Ig C2 family. The 'RKPS' sequence, predicted cAMP- and cGMP-dependent protein kinase A (PKA) phosphorylation site (amino acid sequence number 449~452), is boxed. The shad- ed area represents the conserved sequences.

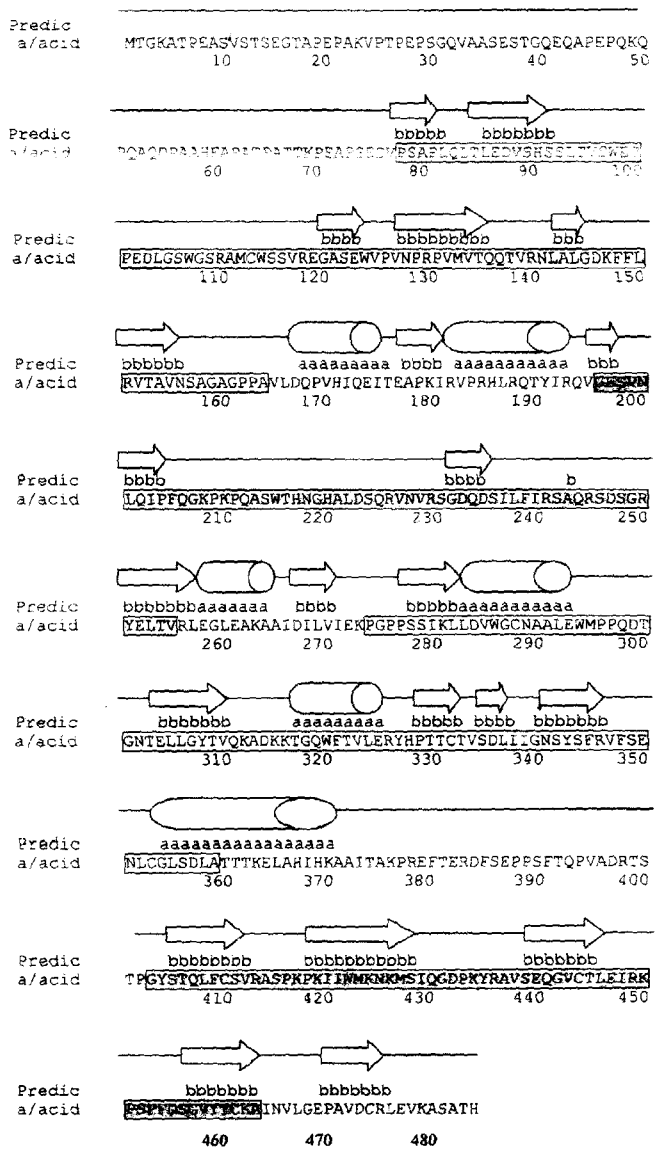

Fig. 3. (A) The predicted secondary structure of MyBP-H. The Ig C2 (□) and FN III (□) motifs are illustrated and the deduced amino acid sequence has 11 protein kinase C (PKC) phosphorylation site (2~4, 67~69, 116~118, 138~140, 224~226, 248~250, 254~256, 279~281, 344~346, 361~363, 411~413), 7 casein kinase II (CK2) phosphorylation site (12~15, 17~20, 39~42, 68~71, 85~88, 116~119, 322~325), 4 N-myristoylation site (31~36, 105~110, 287~292, 318~323).

et al., 1986) (Fig. 3A). During the contraction and the relaxation mechanism of the cardiac and muscle cells, Gs proteins are activated to have two distinct effects; it opens directly calcium channels in the plasma membrane and it activates membrane-bound adenylyl cyclase. The latter effect results in an increase in the intracellular cAMP, which activates cAMP-dependent protein kinase, which in turn phosphorylates and thereby activates at least three proteins; a plasma membrane protein, like Gs, a sarcoplasmic-reticulum protein, and myosin. Phosphorylation of myosin causes an increase in the rate at which the cross bridges cycle, resulting in an increase in the velocity of the contraction. From the sequence, we can imagine that PKA, PKC, and CK2 may regulate the MyBP-H proteins in the presence of the second messengers, cAMP and Ca^{2+} that stimulate the phosphorylation of myofibrillar proteins.

Secondary structure analysis

The immunoglobulin superfamily is the most abundant protein among the cell surface molecules, accounting for 50% of leukocyte surface glycoproteins (Holness and simmons, 1994). It is thought that the Ig domain maintains its stability that can be able to resist the

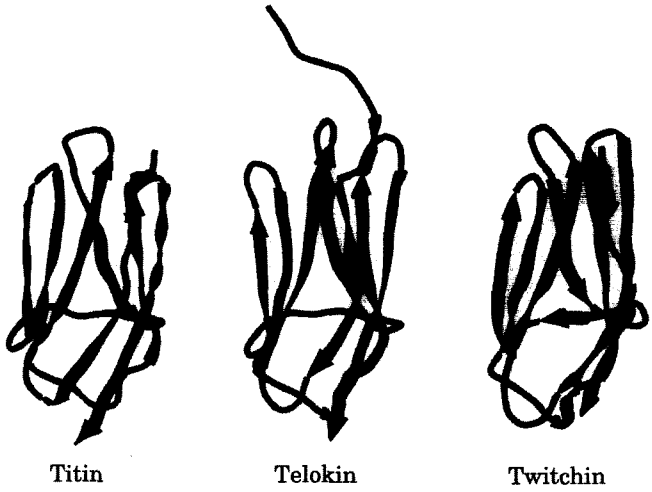

Fig. 3. (B) Immunoglobulin-like domain secondary structures of telokin (PDB code: 1tlk), twitchin (PDB code: 1wtl), and titin (PDB code: 1tit) are presented by solid black.

harsh proteolytic and oxidative environment of the extracellular circumstance. The Ig domain found in recognition/adhesion molecules has usually the beta barrel plan, but it seems that the one in adhesion molecules have longer beta strands and shorter connecting loop regions as compare to the one in antibody V domain. Adhesion molecules, peculiarly, do not have 'RKPS' sequence in the Ig C2 motif but still involved in protein binding through the C-terminal region. We have found that MyBP-H proteins do have 'RKPS' sequence between the two beta strands in the C-terminal Ig C2 motif (Fig. 3A). This implies that the binding property of MyBP-H may be different from that of adhesion molecules. From the predicted 2^{nd} structure of rat MyBP-H, we have found that the C-terminal Ig C2 motifs resemble to many other skeletal muscle myosin binding protein's immunoglobulin-like domain as well as adhesion molecules. For example, all of the immunoglobulin-like domain of telokin, titin and twitchin exhibit mainly beta sandwich architecture of the protein as shown in Fig. 3B.

ACKNOWLEDGMENTS

We thank Dr. Il Yeong Park for helping with the usage of PDB. This study was supported by the Most through the Women's University Research Fund.

REFERENCES CITED

Alyonycheva, T., Cohen-Gould, L., Siewert, C., Fischman, D. A. and Mikawa, T., Skeletal muscle-specific myosin binding protein-H is expressed in Purkinje Fibers of the cardiac conduction system. *Circ. Res.*, 80, 665-672 (1997).

Alyonycheva, T., Mikawa, T., Reinach, F. C. and Fischman, D. A., Isoform-speicfic interaction of the myosin-binding proteins (MyBPs) with skeletal and cardiac myosin is a property of the C-terminal immunoglobulin domain. *J. Biol. Chem.*, 272, 20866-20872 (1997).

Benian, G. M., Kiff, J. E., Neckelmann, N., Moerman, D. G. and Waterston, R. H., Sequence of an unusually large protein implicated in regulation of myosin activity in C. elegans. *Nature*, 342, 45-50 (1989).

Blackwell, T. K. and Weintraub, H., Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science*, 250, 1104-1110 (1990).

Cunningham, B., Hemperly, J., Murray, B., Prediger, E., Brackenbury, R. and Edelman, G., Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell surface modulation, and alternative RNA splicing. *Science*, 236, 799-806 (1987).

Daley, J., Southgate, R. and Ayme-Southgate, A., Structure of the drosophila projectin protein: isoforms and implication for projectin filament assembly. *J. Mol. Biol.*, 279, 201-210 (1998).

Duronio, R. J., Jackson-Machelski, E., Heucheroth, R. O., Olins, P. O., Devine, C. S., Yonemoto, W., Slice, L. W., Taylor, S. S. and Gordon, J. I., Protein N-myristoylation in *Escherichia coli*: reconstitution of a eukaryotic protein modification in bacteria. *Proc. Natl. Acad. Sci., USA*, 87, 1506-1510 (1990).

Einheber, S. and Fischman, D. A., Isolation and characterization of a cDNA clone encoding avian skeletal muscle C-protein : An intracellular member of the immunoglobulin superfamily. *Proc. Natl. Acad. Sci. USA*, 87, 2157-2161 (1990).

Fremisco, J. R., Glass, D. B. and Krebs, E. G., Optimal spatial requirements for the location of basic residues in peptide substrates for the cyclic AMP-dependent protein kinase. *J. Biol. Chem.*, 255, 4240-4245 (1980).

Glass, D. B. and Smith, S. B., Phosphorylation by cyclic AMP-dependent protein kinase of a synthetic peptide corresponding to the autophosphorylation site in the enzyme. *J. Biol. Chem.*, 258, 14797-14803 (1983).

Heierhorst, J., Probst, W. C., Kohanski, R. A., Buku, A. and Weiss, K. R., Phosphorylation of myosin regulatory light chains by the molluscan twitchin kinase. *Eur. J. Biochem.*, 233, 426-431 (1995).

Higgins, D. G., Labeit, S., Gautel, M. and Gibson, T. J., The evolution of titin and related giant muscle proteins. *J. Mol. Evol.*, 38, 395-404 (1994).

Holness, C. and Simmons, D., Structural motifs for recognition and adhesion in members of the immunoglobulin superfamily. *J. Cell. Sci.*, 107, 2065-2070 (1994).

Kishimoto, A., Nishiyama, K., Nakanishi, H., Uratsuji, Y. and Nomura, H., Studies on the phosphorylation of myelin basic protein by protein kinase C and

adenosine 3',5'-monophosphate dependent protein kinase. *J. Biol. Chem.*, 260, 12492-12499 (1985).

Maniatis, T., Fritsch, E. F. and Sambrook, J., Molecular cloning : A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, Vol.2, pp.8.1-8.86, 1982.

Moller, W. and Amons, R., Phosphate-binding sequence in nucleotide-binding proteins. *FEBS Lett.*, 186, 1-7 (1985).

Moos, C. and Feng, I. N., Effect of C-protein on actomyosin ATPase. *Biochim. Biophys. Acta*, 632, 141-149 (1980).

Moos, M., Tacke, R., Scherer, H., Teplow, D., Fruh, K. and Schachner, M., Neural adhesion molecule L 1 as a member of the immunoglobulin superfamily with binding domains similar to fibronectin. *Nature*, 234, 701-703 (1988).

Okagaki, T., Weber, F. E., Fischman, D. A., Vaughan, K. T., Mikawa, T. and Reinach, F. C., The major myosin-binding of skeletal muscle MyBP-C (C protein) resides in the COOH-terminal, Immunoglobulin C2 motif. *J. Cell Biol.*, 123, 619-626 (1993).

Pinna, L. A., Casein kinase 2: an 'eminecne grise' in cellular regulation?. *Biochim. Biophys. Acta*, 1054, 267-284 (1990).

Sanger, F., Nicklen, S. and Conlson, A. R., DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA.*, 74, 5463-5467 (1977).

Seiler, S. H., Fischman, D. A. and Leinwand, L. A., Modulation of myosin filament organization by C-protein family members. *Mol. Biol. Cell*, 7, 113-127 (1996).

Silver, D. L., Vorotnikov, A.V., Watterson, D. M., Shirinsky, V. P. and Sellers, J. R., Sites of interaction between kinase-related protein and smooth muscle myosin. *J. Biol. Chem.*, 272, 25353-25359 (1997).

Smith, D. K. and Xue, H., Sequence profiles of immunoglobulin and immunoglobulin-like domains. *J. Mol. Biol.*, 274, 530-545 (1997).

Sobieszek, A., Andruchov, O. Y. and Nieznanski, K., Kinase-related protein (telokin) is phosphorylated by smooth-muscle myosin light-chain kinase and modulates the kinase activity. *Biochem. J.*, 328, 425-430 (1997).

Stull, J. T., Manning, D. R., High, C. W. and Blumenthal, D. K., Phosphorylation of contractile proteins in heart and skeletal muscle. *Fed. Proc.*, 39, 1552-1557 (1980).

Takagi, T. and Cox, J. A., Primary structure of the target of calcium vector protein of amphioxus. *J. Biol. Chem.*, 256, 19721-19727 (1990).

Vaughan, K. T., Weber, F. E. and Fischman, D. A., cDNA cloning and sequence comparisons of human and chicken muscle C-protein and 86 kD protein. symp. *Soc. Exp. Biol.*, 46, 167-177 (1992).

Vaughan, K. T., Weber, F. E., Ried, T., Ward, D. C., Reinach, F. C. and Fischman, D. A., Human myosin-binding protein H (MyBP-H): Complete primary sequence, genomic organization, and chromosomal localization. *Genomics*, 16, 34-40 (1993).

Vaughan, K. T., Weber, F. E., Einheber, S. and Fischman, D. A., Molecular cloning of chicken myosin-binding protein (MyBP) H (86kDa protein) reveals extensive homology with MyBP-C (C-protein) with conserved immunoglobulin C2 and fibronectin type III motifs. *J. Biol. Chem.*, 268, 3670-3676 (1993).

Weber, F. E., Vaughan, K. T., Reinach, F. C. and Fischman, D. A., Complete sequence of human fast- type and slow-type muscle myosin-binding-protein C (MyBP-C). *Eur. J. Biochem.*, 216, 661-669 (1993).

Weitkamp, B., Jurk, K. and Beinbrech, G., Projectin-thin filament interaction and modulation of the sensitivity of the actomyosin ATPase to calcium by projectin kinase. *J. Biol. Chem.*, 273, 19802-19808 (1998).

Yamamoto, K., Characterization of H-protein, a component of skeletal muscle myofibrils. *J. Biol. Chem.*, 259, 7163-7168 (1984).