Human Myosin-Binding Protein H (MyBP-H): Complete Primary Sequence, Genomic Organization, and Chromospinal Localization

ARTICLE *in* GENOMICS · MAY 1993

Impact Factor: 2.28 · DOI: 10.1006/geno.1993.1136 · Source: PubMed

6 AUTHORS, INCLUDING:

Franz E. Weber
University of Zurich
119 PUBLICATIONS **3,995** CITATIONS

GENOMICS 16, 34–40 (1993)

Human Myosin-Binding Protein H (MyBP-H): Complete Primary Sequence, Genomic Organization, and Chromosomal Localization

KEVIN T. VAUGHAN,*,[1] FRANZ E. WEBER,* THOMAS RIED,†,[2] DAVID C. WARD,†
FERNANDO C. REINACH,*,‡ AND DONALD A. FISCHMAN*,[3]

*Department of Cell Biology and Anatomy, Cornell University Medical College, New York, New York 10021; †Department of Genetics, School of Medicine, Yale University, New Haven, Connecticut 06510; and ‡Department of Biochemistry, Instituto Quimica, University of Sao Paulo, Sao Paulo, Brazil

Received September 22, 1992; revised November 27, 1992

Vertebrate striated muscle contains a set of myosin-associated proteins with discrete distributions in the A-band. Some of these proteins, including MyBP-H and MyBP-C, are characterized by a series of internal repeats (motifs) with homology to either the C2-set of the immunoglobulin superfamily or the fibronectin type III repeat. These repeats are predicted to be involved in protein–protein interactions within the myofibril. The cDNA sequence, the genomic organization, and the chromosomal localization of the human homologue of MyBP-H are presented. The 1.8-kb cDNA encodes a 52-kDa polypeptide containing two Ig-C2 and two type III repeats. The mRNA is expressed in a skeletal muscle-specific pattern. A 28-kb region of genomic sequence has been isolated that encompasses the 5′ and 3′ ends of the cDNA. This region includes 4.2 kb of upstream sequence with a potential promoter and 14 kb of downstream sequence containing the polyadenylation site. The chromosomal assignment was made by high resolution chromosomal *in situ* hybridization. This method maps the gene to chromosome 1q32.1. The repeat structure described previously in chicken MyBP-H and MyBP-C was also detected in human MyBP-H. The primary sequence of the C-terminal Ig-C2 motif and its predicted secondary structure have been extensively conserved in MyBP-H homologues and other members of the MyBP family. This Ig-C2 motif has been implicated in myosin binding. © 1993 Academic Press, Inc.

INTRODUCTION

Of the three filament systems in striated muscle, thick filaments appear to be the most complex, based on subunit composition. In addition to the myosin heavy chain (MyHC) and myosin light chains (MLC1, MLC2, MLC3), a series of abundant accessory proteins are found at discrete positions along the filament length. In the crossbridge-bearing zone, the C-region of the A-band was shown by Offer and colleagues to contain an abundant protein called C-protein (Offer, 1972; Offer *et al.*, 1973; Craig and Offer, 1976; Bennett *et al.*, 1986; Squire, 1981). This protein has been renamed MyBP-C to reflect its myosin-binding properties and emphasize the relationship between members of a family of myosin-binding proteins.

The MyBP family includes MyBP-C, MyBP-X, and MyBP-H (Dennis *et al.*, 1984; Bennett *et al.*, 1986). Originally described as contaminants of myosin preparations (Starr and Offer, 1971), these proteins are found in a series of transverse stripes spaced at 43-nm intervals (Craig and Offer, 1976; Dennis *et al.*, 1984; Bahler *et al.*, 1985a; Bennett *et al.*, 1986). MyBP-C is the most abundant of the family and is found in a 1:8 molar ratio with myosin heavy chain (Offer *et al.*, 1973). In the chicken, MyBP-H (formerly known as the 86-kDa protein) is found in an approximately equimolar ratio with MyBP-C (Bahler *et al.*, 1985b). In the rabbit, MyBP-H (formerly known as H-protein) is far less abundant (Starr and Offer, 1982). MyBP-X (previously termed X-protein), the least abundant member of this group (Starr and Offer, 1982), is thought to be the slow fiber-type isoform of MyBP-C (Reinach *et al.*, 1982).

The functions of this protein family in muscle contraction and development are not known. Speculations about its function range from a role in the regulation of myosin activity (Hofmann *et al.*, 1991a,b) to a role in myofibrillar assembly (Einheber *et al.*, 1990; Schultheiss *et al.*, 1990; Epstein and Fischman, 1991). To identify functionally significant primary sequences in this protein family, the cDNA encoding chicken MyBP-C was isolated and characterized (Einheber and Fischman, 1990; Fischman *et al.*, 1991; Okagaki *et al.*, 1993). In addition, the chicken MyBP-H has been cloned and analyzed, assuming that primary sequences responsible for myosin binding in the C-region of the A-band might be conserved (Vaughan *et al.*, 1993).

[1] Present address: Cell Biology Group, Worcester Foundation for Experimental Biology, Shrewsbury, MA.

[2] Present address: University of Heidelberg, Institut fur Humangenetik, Heidelberg, Germany.

[3] To whom correspondence should be addressed.

0888-7543/93 $5.00

Comparison of the deduced amino acid sequences of chicken MyBP-C and MyBP-H revealed a highly conserved C-terminal region of 400 amino acids (Vaughan *et al.*, 1993). Nested within this region, a series of conserved repeats was observed, resembling the C2-set of the immunoglobulin superfamily (Williams, 1987; Williams and Barclay, 1988) or the fibronectin type III repeat (Kornblihtt *et al.*, 1985; Petersen *et al.*, 1983). The C-terminal C2 motif is notable since its predicted secondary structure (Chou and Fasman, 1978) resembles the crystal structure of the immunoglobulin fold (Amzel and Poljak, 1979) and has been shown to possess myosin-binding activity in MyBP-C (Okagaki *et al.*, 1993).

To further elucidate functionally conserved primary sequences in this family of myosin-binding proteins, we have cloned and characterized the human homologue of MyBP-H. We have also cloned the gene encoding human MyBP-H and mapped its position in the human genome using high-resolution fluorescence *in situ* hybridization.

MATERIALS AND METHODS

Library screening. A human fetal muscle λgt10 cDNA library (Koenig *et al.*, 1987) was screened by standard methods (Sambrook *et al.*, 1989) using the chicken MyBP-H cDNA as probe (Vaughan *et al.*, 1993). Each clone was taken through three rounds of screening until pure, and λ DNA was purified using the Qiagen Lambda Mini Kit system (Qiagen Inc., Chatsworth, CA). The cDNAs were compared by restriction analysis and shown to overlap in the 3′ ends by DNA sequencing.

DNA sequencing. The longest clone was sequenced to completion after subcloning into M13, taking advantage of convenient restriction sites, and using custom oligonucleotide primers (Research Genetics, Huntsville, AL). The ends of the shorter clone were sequenced to confirm overlap. A combination of Sequenase (USB, Cleveland, OH) and Taquence (USB, Cleveland, OH) was used in areas of long G–C stretches. Gel reading, alignment of gel readings, sequence analysis, and sequence comparisons were performed using PCGENE (Mountain View, CA) programs. Comparisons with current releases of GenBank and Dayhoff databases were made using FASTA (Pearson and Lipman, 1988). Sequence similarity was assessed using DIAGON with a window of nine amino acids (Staden, 1982).

Genomic cloning. Genomic clones were isolated from a *Sau*3A human genomic library in λEMBL-3A (Clontech Laboratories, Palo Alto, CA). The DNA of λ-11 was restriction digested with *Xho*I, and the 5′ 3.2- and 3′ 2.5-kb fragments were subcloned into pUC. The relative positions of these fragments were confirmed through Southern blotting with the cDNA. From the 5′ *Xho*I fragment, the 1260 *Hin*dIII–*Pst*I fragment was sequenced to completion using custom oligonucleotides as primers.

In addition, a 1040-kb *Kpn*I–*Pst*I fragment was isolated from the 3′ 2.5-kb *Xho*I fragment of clone L-11 and sequenced to completion.

Northern blotting. A human MTN (Multiple Tissue Northern) blot was obtained from Clontech Laboratories (Palo Alto, CA). Equivalent RNA loading was confirmed by ethidium bromide staining of the gel before transfer and the blot after transfer.

Fluorescence in situ hybridization. Metaphase spreads were prepared from 5-bromodeoxyuridine (BrdU)-synchronized cultures (Speit, 1984; Fan *et al.*, 1990) of a healthy donor, with the following modifications: cells were incubated with methotrexate (10^{-7} *M*, final concentration) for 17 h at 37°C. Cells were collected by centrifugation, resuspended in FCS-free medium, and then washed twice with FCS-free medium. The block was released by adding BrdU (30 μg/ml, final concentration) for 5 h at 37°C.

DNA of both overlapping phage clones was labeled with biotin–11-dUTP (Sigma, St. Louis, MO) in separate nick-translation reactions (Sambrook *et al.*, 1989). Eighty nanograms of labeled DNA from each reaction was coprecipitated with 10 μg of human placental DNA, 5 μg salmon sperm DNA, and yeast RNA, respectively. DNA was resuspended in 10 μl of hybridization solution (50% formamide, 2× SSC, 10% dextran sulfate) and denatured for 5 min at 76°C. The repetitive sequences were allowed to preanneal for 30 min at 37°C. The probe was then added to previously denatured metaphase chromosome preparations (denatured in 70% formamide, 2× SSC at 80°C for 2 min, then dehydrated through an ethanol series), and a coverslip (18 mm^2) was added and sealed with rubber cement. Hybridization was performed at 37°C for 16 h. The coverslip was removed and the slides were washed in 50% formamide, 2× SSC at 42°C (3× 5 min), followed by three washes in 0.1× SSC at 60°C (3× 5 min). To reduce nonspecific background fluorescence, the slides were incubated in 4× SSC containing 3% BSA for 30 min at 37°C. The BrdU-induced R-banding pattern was generated through a fluorescence photolysis step. Slides were stained with Hoechst dye 33258 at room temperature for 15 min, briefly washed in water, and exposed to the light of a UV mercury lamp. Finally, the slides were incubated in 2× SSC at 60°C for 1 h. The biotinylated probe signals were detected with avidin–DCS–isothiocyanate (Vector Laboratories, Burlingame, CA), followed by three washes in 4× SSC, 0.1% Tween 20 at 37°C (5 min each). The chromosome specimens were counterstained with propidium iodide and mounted in antifade. Images were obtained with a Zeiss Axioskop epifluorescence microscope coupled to a cooled CCD camera (Photometrics, Tucson, AZ). The sequentially recorded gray scale images were pseudocolored and merged using custom computer software.

RESULTS

cDNA cloning and protein sequence. Using the cDNA encoding the chicken skeletal muscle MyBP-H as probe (Vaughan *et al.*, 1993), a human fetal skeletal muscle cDNA library (generous gift of Dr. L. Kunkel, Harvard Medical School) was successfully screened, yielding two overlapping clones. These clones were shown by restriction analysis and DNA sequencing to possess common 3′ ends and different degrees of 5′ extension. The longest clone (1786 bp) was shown to contain a single open reading frame of 1429 bp, beginning at base 28 and terminating at base 1458 (Fig. 1). The second clone, shorter by 600 bp, showed identity with the longer to the extent of their overlap. The deduced protein is 476 amino acids in length and has a predicted mass of 51,986 Da. These cDNA clones were concluded to encode the human muscle MyBP-H by virtue of their high DNA and protein sequence identity with the chicken homologue (see Discussion).

Gene structure. These cDNAs were used to screen a human genomic library, resulting in the isolation of two overlapping clones covering a total of 28 kb (Fig. 2). These genomic clones were restriction mapped and Southern blotted to identify the positions of corresponding sequences in the cDNAs. Clone L-11 contains an 11-kb *Hin*dIII fragment that encompasses the entire cDNA sequence (Fig. 2, bottom section). The positions of relevant *Xba*I, *Pst*I, *Sph*I, and *Kpn*I sites are indicated; *Pst*I sites reflecting a subset of the numerous sites in the cDNA and genomic clones.

Sequencing of clone L-11 in the region corresponding to the 5′ end of the cDNA suggests that the cDNA for human MyBP-H contains the entire coding sequence. Upstream of the first base of the cDNA, a putative TATA box was detected, indicating a potential region of

```
CACCCTGGCCCCCCAGCCAGTCCAGCAATGATGGAAAAAAACACCTCCGAGGGCCCTGCctgcagTCCAGAG   72
                              M  M  E  K  N  T  S  E  G  P  A  C  S  P  E    15

GAGACCGCATCTGAATCTGCCAAGGTGCCCACAGCAGAGCCTCCCGGAGAAGTGGCAGTATCAGAGTCCACC  144
 E  T  A  S  E  S  A  K  V  P  T  A  E  P  P  G  E  V  A  V  S  E  S  T    39

AGAGAAGAGCAGGTGCCCAAGCCGCACGGCCCTGCCCCACAGGCCCCTACAGCCTCCACAGCCACTAAGCCT  216
 R  E  E  Q  V  P  K  P  H  G  P  A  P  Q  A  P  T  A  S  T  A  T  K  P    63

GCACCCCCAAGTGAAGATGTCCCCAGTGCCCCACTGCTGCTGACCCTGGATGATGTGAGCAGCAGCTCTGTG  288
 A  P  P  S  E  D  V  P  S  A  P  L  L  L  T  L  D  D  V  S  S  S  S  V    87

ACTGTGAGCTGGGAGCCCCCAGAGAGGCTGGGGAGGCTGGGCCTCCAGGGCTATGTGCTGGAGCTctgcagA  360
 T  V  S  W  E  P  P  E  R  L  G  R  L  G  L  Q  G  Y  V  L  E  L  C  R   111

GAGGGAGCCTCGGAGTGGGTGCCTGTGAGTGCCCGGCCCATGATGGTGACCCAGCAGACTGTGCGGAACCTG  432
 E  G  A  S  E  W  V  P  V  S  A  R  P  M  M  V  T  Q  Q  T  V  R  N  L   135

GCTCTGGGAGACAAGTTCCTCCTGCGCGTGTCTGCAGTGAGTTCTGCAGGGGCTGGCCCGCCGGCCATGCTG  504
 A  L  G  D  K  F  L  L  R  V  S  A  V  S  S  A  G  A  G  P  P  A  M  L   159

GACCAGCCCATCCACATCCGAGAGAACATTGAGGCCCCCAAGATCCGTGTCCCCCGCCACCTCCGTCAGACC  576
 D  Q  P  I  H  I  R  E  N  I  E  A  P  K  I  R  V  P  R  H  L  R  Q  T   183

TACATCCGCCAGGTGGGAGAGACGGTCAACCTGCAAATCCCCTTCCAGGGGAAGCCTAAGCCTCAGGCCACA  648
 Y  I  R  Q  V  G  E  T  V  N  L  Q  I  P  F  Q  G  K  P  K  P  Q  A  T   207

TGGACCCACAACGGCCATGCCCTGGACAGCCAGCGGGTGAgcatgcGCACCGGGGACCAGGACTCCATCCTC  720
 W  T  H  N  G  H  A  L  D  S  Q  R  V  S  M  R  T  G  D  Q  D  S  I  L   231

TTCATTCGCTCGGCCCAGCGCTCCGACTCTGGCCGCTACGAGCTCACTGTGCGCGTGGAAGACCTGGAGGCC  792
 F  I  R  S  A  Q  R  S  D  S  G  R  Y  E  L  T  V  R  V  E  D  L  E  A   255

AAGGCAGTCATTGACATCCTGGTGATTGAGAAACCTGGACCCCCCAGCAGCATCAGGCTCCTGGACGTCTGG  864
 K  A  V  I  D  I  L  V  I  E  K  P  G  P  P  S  S  I  R  L  L  D  V  W   279

GGCTGCAATGCTGCTCTTCAGTGGACGCCACCCCAGGACACAGGCAACACAGAGCTCCTGGGCTACATGGTG  936
 G  C  N  A  A  L  Q  W  T  P  P  Q  D  T  G  N  T  E  L  L  G  Y  M  V   303

CAGAAGGCAGACAAAAAGACAGGGCAATGGTTCACAGTGCTGGAGCGCTACCACCCAACCACCTGCACCATC 1008
 Q  K  A  D  K  K  T  G  Q  W  F  T  V  L  E  R  Y  H  P  T  T  C  T  I   327

TCTGACCTCATCATCGGCAACTCGTACTCCTTCCGGGTCTTCTCAGAAAACCTGTGTGGACTCAGCACCTCG 1080
 S  D  L  I  I  G  N  S  Y  S  F  R  V  F  S  E  N  L  C  G  L  S  T  S   351

GCCACCGTCACCAAGGAGCTCGCCCACATCCAGAAGGCAGATATTGCTGCCAAACCTAAAGGGTTTATTGAG 1152
 A  T  V  T  K  E  L  A  H  I  Q  K  A  D  I  A  A  K  P  K  G  F  I  E   375

CGAGACTTCTCAGAAGCCCCCTCATTCACCCAGCCCCTGGCTGACCACACCTCCACCCCTGGCTACAGCACC 1224
 R  D  F  S  E  A  P  S  F  T  Q  P  L  A  D  H  T  S  T  P  G  Y  S  T   399

CAGTTGTTCTGCAGTGTCGAGCTTCACCAAGCCAAGATCATCTGGATGAAAAACAAGATGGAGATCCAGGGC 1296
 Q  L  F  C  S  V  E  L  H  Q  A  K  I  I  W  M  K  N  K  M  E  I  Q  G   423

AACCCCAAATACCGCGCCCTCTCTGAGCAAGGCGTCTGCACCCTAGAGATCCGGAAACCCAGCCCCTTTGAT 1368
 N  P  K  Y  R  A  L  S  E  Q  G  V  C  T  L  E  I  R  K  P  S  P  F  D   447

TCTGGGGTCTACACCTGCAAGGCCATAAATGTGCTGGGGGAGGCATCTGTGGACTGCCGGCTGGAGGTCAAA 1440
 S  G  V  Y  T  C  K  A  I  N  V  L  G  E  A  S  V  D  C  R  L  E  V  K   471

GCCTCAGCCGCACACTGAAGAGTCACGGGAGGAGGCTGTGATGAGAGCCAGGTGCTGCAGCCTGAGGGGCC  1511
 A  S  A  A  H                                                            476

AGGTCCACCAGGCCTACAGTCAAACTCCAGAGATGCCCAAGACCCCCGCTTCCATAGCGACGGCTGATGGAA 1583
CCGCCCCTggtaccTGCTGGCCTCCCTTCCCCAGAGGCTGGTGTCCAGAACTCAGGAATGGGCCTGGTAGGC 1655
CCCAGGCCAGACTAACTGGGCTCAAGGGGTGCTGGAAGGCGCAGAGATTGGAGTGCCCTGCGGAGTTGCACT 1727
CTGGGTGGGAAGCACTCAAATAAAGATGCGTGGTGTTAACAGTGAAAAAAAAAAAAAAA              1786
```

FIG. 1. cDNA sequence of human MyBP-H and deduced amino acid sequence. The cDNA sequence and deduced amino acid sequence are presented using the one-letter code system. The position of the termination codon (TGA bp 1456) is underlined as is the putative polyadenylation signal (bp 1744). *Pst*I sites identified in Fig. 2 at nucleotides 60 and 353 are presented in lowercase letters, as is the *Kpn*I site at nucleotide 1592. This sequence can be obtained from GenBank under Accession number L05606.

transcription initiation (data not shown). In addition, the genomic sequences between the TATA box and ATG at base 28 contain termination codons in all three reading frames, suggesting that the ATG codon at position 28 is the first potential position for initiating translation. Genomic sequences in the 735 bp upstream of the putative TATA box possess eight consensus E boxes (data not shown), sequence motifs frequently found upstream of muscle-specific genes (Blackwell *et al.*, 1990). Restriction mapping indicates that clone L-11 contains

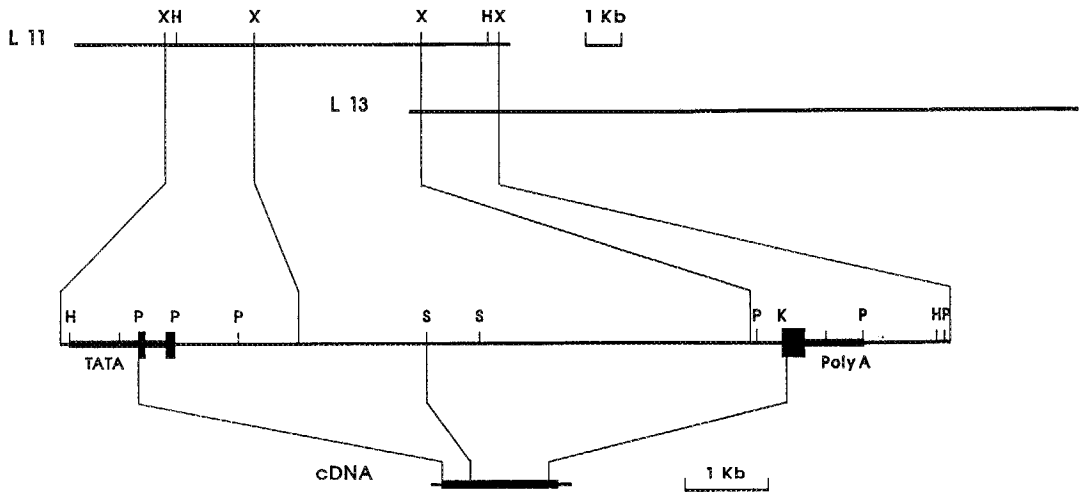

FIG. 2. Genomic cloning of human MyBP-H. Two overlapping genomic clones covering 28 kb were isolated from a human genomic library using the full-length cDNA as probe. Sequences corresponding to the ends of the cDNA were mapped and sequenced from clone L-11. These sequences (indicated by thick lines in the bottom panel) can be obtained from GenBank under Accession number L05607. The positions of relevant *Hin*dIII (H), *Pst*I (P), *Sph*I (S), and *Kpn*I (K) sites are indicated. Clone L-13 overlaps with clone L-11 in the region corresponding to the last exon of the cDNA.

an additional 3.2 kb of sequence upstream of these E boxes, suggesting that this clone may contain additional regulatory elements for this gene.

DNA sequence analysis of the region from the 5′ *Hin*dIII site to the second *Pst*I site (indicated in Fig. 2) reveals the position of the first and second exons of the gene as well as the first intron (data not shown). The intron sequence has several termination codons in the correct reading frame, suggesting that this is not an example of an optional exon (Nassim *et al.*, 1990). Sequencing of clone L-11 in the region corresponding to the 3′ end of the cDNA demonstrates that the *Kpn*I–*Pst*I fragment contains part of the last exon encoding the 3′ untranslated sequences up to the consensus polyadenylation signal. The overlap of clone L-13 in the 3′ end of clone L-11 indicates that an additional 14 kb of genomic sequence are present downstream from the last exon in clone L-13.

Tissue distribution. The mRNA encoding human MyBP-H is expressed in a skeletal muscle-specific pattern. A human MTN blot (Clontech) was probed with the full-length cDNA (Fig. 3). Equivalent RNA loading was confirmed by ethidium bromide staining of the gel before transfer and the blot after transfer. Only skeletal muscle RNA contained a hybridizing RNA species, approximately 1.8 kb in length and consistent with the length of the cDNA. As in the chicken, the skeletal muscle cDNA does not hybridize with RNA from cardiac muscle or smooth muscle (placenta) (Vaughan *et al.*, 1993).

Chromosomal localization. To identify the chromosome containing the human MyBP-H gene, human–hamster somatic cell hybrids were screened using the full cDNA sequence or the 3′ untranslated sequence (data not shown). After confirming a distinct human-specific pattern of hybridization following *Bgl*II digestion, the panel was probed with the 3′ untranslated sequence of the cDNA. The pattern of hybridization in the two overlapping panels of hybrids mapped the gene to human chromosome one.

More precise localization of the MyBP-H gene was determined by fluorescence *in situ* hybridization (FISH) onto high-resolution banded chromosomes (Arnold *et al.*, 1992). To generate a simultaneous fluorescent banding pattern for gene mapping studies, the methotrexate synchronization procedure was followed with subsequent BrdU incorporation. This procedure reveals an R-banding pattern and allows high-resolution band assignment (Arnold *et al.*, 1992).

Upon isolation and characterization of the genomic clones (described above), FISH was performed using a combination of the two genomic clones. This allowed labeling of genomic sequences covering 28 kb. Fifteen metaphases were investigated: 11 displayed two hybridization signals on both homologs of chromosome 1q32, 3 metaphases showed incomplete hybridization, giving either only one hybridization signal on both homologs or

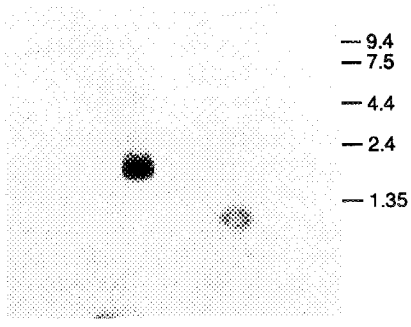

FIG. 3. Tissue distribution of human MyBP-H. A human MTN blot was probed with the cDNA encoding human MyBP-H. RNA samples were isolated from: lane (**1**), pancreas; lane (**2**), kidney; lane (**3**), skeletal muscle; lane (**4**), liver; lane (**5**), lung; lane (**6**), placenta; lane (**7**), brain; lane (**8**), heart; and lane (**9**), RNA markers. The relative mobilities of RNA molecular weight standards are indicated.

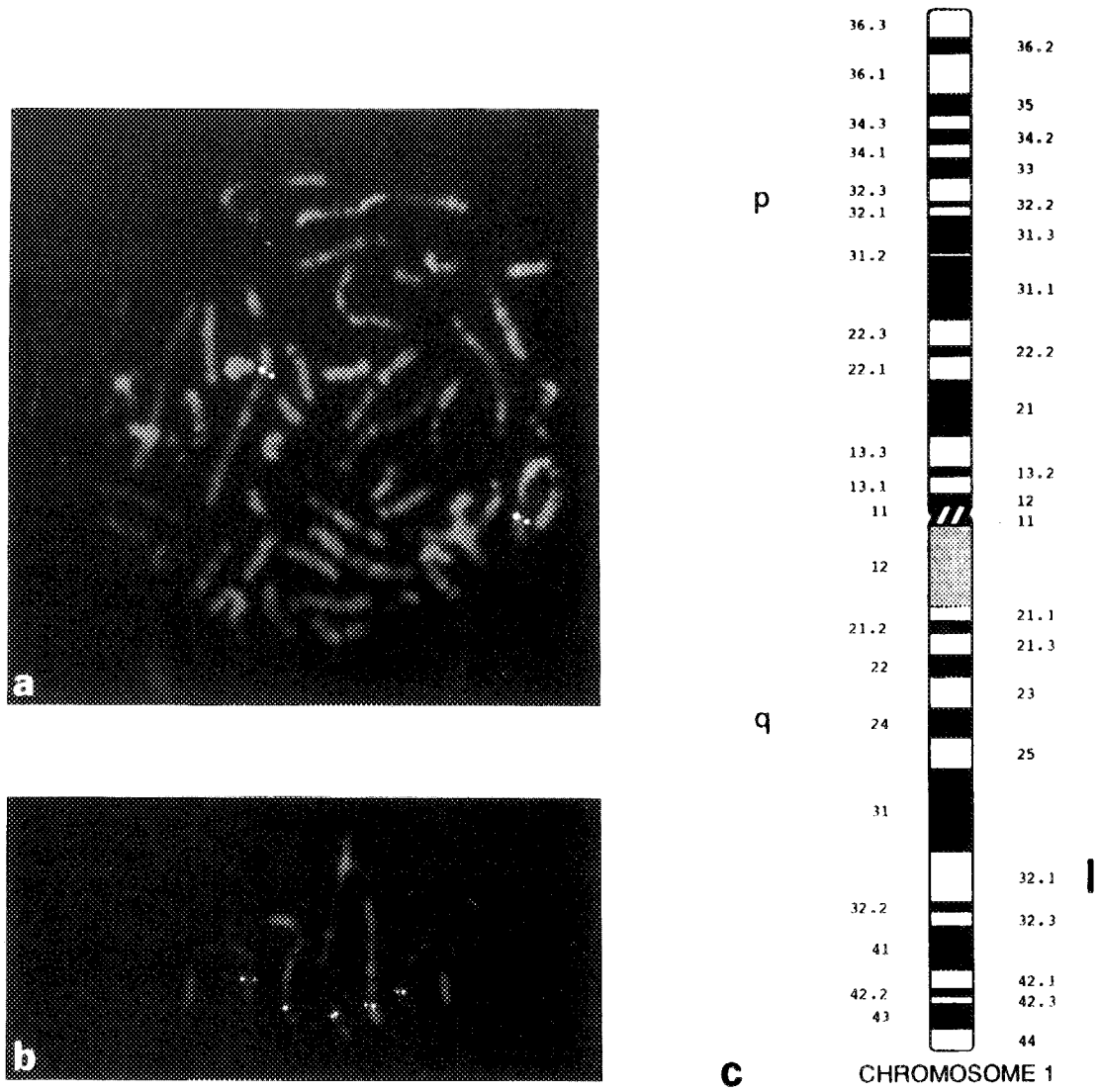

FIG. 4. Chromosome map position determined by fluorescence *in situ* hybridization. (**a**) Hybridization of the biotinylated phage clones for the human MyBP-H to a human metaphase spread. The DNA was detected with avidin–FITC. The chromosomes are counterstained with propidium iodide. The BrdU-induced R-banding pattern allows for the chromosome and band identification. The probe hybridized to chromosomal map position 1q32.1. (**b**) Examples of five different chromosome 1's from three different metaphase spreads. The signal consistently localizes to band 1q32.1. The observed R-banding pattern was induced using BrdU-induced replication and photolysis (see Materials and Methods). (**c**) Idiogram of chromosome 1, displaying the G-banding pattern with the 550-band resolution. The bar beside the idiogram denotes the chromosomal mapping position of the phage clones specific for human MyBP-H.

two signals on only one homolog. One metaphase did not reveal any hybridization signal. The chromosome map position of the phage clones encoding MyBP-H could be determined precisely when highly extended chromosomes were evaluated. For fine mapping, only chromosome 1 homologs that were at least 16 μm in length were analyzed. The signal consistently mapped to chromosome band 1q32.1 (Fig. 4). One entire metaphase is shown in Fig. 4a, and a composite of six different chromosome 1 fields is shown in Fig. 4b. The chromosome map position is displayed schematically in Fig. 4c.

DISCUSSION

We have isolated full-length cDNA clones that encode human MyBP-H and genomic clones that encompass this cDNA sequence. Evidence for complete cloning comes from a comparison of genomic and cDNA sequences in the region of likely transcription initiation. The first base of the cDNA is within 60 bp of a putative TATA box. In addition, the genomic sequence contains termination codons in all three reading frames upstream of the ATG codon at base 28, indicating that this ATG is likely to be the point of translation initiation. Upstream of the TATA box, eight consensus E box elements were detected that are thought to direct muscle-specific expression of a gene. Genomic clone L-11 contains an additional 3.2 kb of upstream sequence that is likely to contain additional regulatory elements. Analysis of sequences corresponding to the 3′ end of the cDNA indicates that genomic clones L-11 and L-13 overlap in the regions covering the polyadenylation site of the gene. The C-terminal *Kpn*I–*Pst*I fragment contains 3′ un-

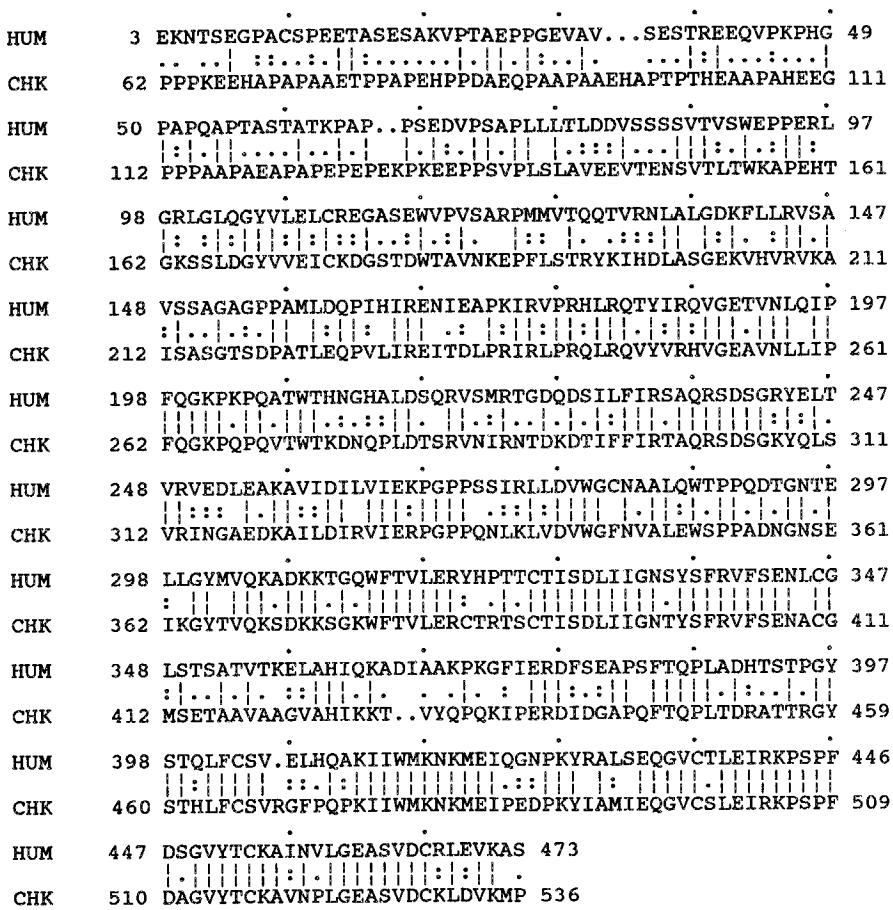

FIG. 5. Comparison of the deduced amino acid sequences of human (HUM) and chicken (CHK) MyBP-H. The C-terminal 470 amino acids of each protein are presented as aligned using the BESTFIT program of the GCG package. Identical amino acids are identified by "|"; strong conservative substitutions are identified by ":"; weak conservative substitutions are identified by ".". The Ig-C2 motifs (human—amino acids 185–264 and 391–476; chicken—amino acids 248–328 and 452–537) and fibronectin type III motifs (human—70–170 and 265–365; chicken—132–234 and 329–429) are highly conserved. The C-terminal Ig-C2 motif is predicted to contain the myosin-binding activity of the molecule.

translated sequences up to and including the polyadenylation signal.

The deduced protein sequence is clearly related to chicken MyBP-H (Vaughan *et al.*, 1993) (Fig. 5). These two proteins share 54% identity plus 15% conservative substitutions over their entire lengths. The C-terminal 400 amino acids of each protein share regions of almost perfect identity. This corresponds to the region of high identity shared between chicken MyBP-H and chicken MyBP-C. The N-termini of these proteins appear divergent. Most notably, human MyBP-H contains fewer of the alternating alanine–proline motifs that are thought to impart the anomalous SDS–PAGE mobility of chicken MyBP-H (Vaughan *et al.*, 1993).

Human MyBP-H is likely to be closely related to the rabbit MyBP-H. Although rabbit MyBP-H has an SDS–PAGE mobility of 74 kDa, sedimentation analyses suggest that the true mass is 52–54 kDa, close to the predicted mass of human MyBP-H (Starr *et al.*, 1982, 1983). Comparison of their amino acids compositions supports this hypothesis (data not shown).

The repetitive motifs demonstrated in chicken MyBP-H are also present in the human protein (Vaughan *et al.*, 1993). This block of 400 amino acids at the C-terminus of the protein contains two copies of the immunoglobulin C2 repeat and two copies of the fibronectin III repeat; NH2–Fn–C2–Fn–C2–COOH. This C-terminal Ig-C2 motif is likely to be involved in interactions with myosin, since the same C-terminal motif has been shown to contain the myosin-binding domain of chicken MyBP-C (Okagaki *et al.*, 1993). Comparison of this motif in human MyBP-H and chicken MyBP-H reveals remarkable sequence identity. It is likely that this repeat represents the myosin-binding domain in all of the members of this protein family.

To correlate this protein with a known human muscle disease, we mapped the position of the gene in the human genome. This mapping places the gene on chromosome band 1q32.1. Examination of the human gene atlas does not reveal a human muscle disease currently mapped to this chromosome band position.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institutes of Health (AR32147 to D.A.F. and HG00272 to D.C.W.), the Muscular Dystrophy Association, and The Rockefeller Foundation. T.R. received a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft, and F.E.W. a fellowship from the Revson-Winston Foundation. F. C. Reinach was a Biotechnology International Fellow of the

Rockefeller Foundation. The authors thank Madhvi Bhatt for providing the BrdU-banded chromosomes.

REFERENCES

Amzel, L. M., and Poljak, R. J. (1979). Three-dimensional structure of immunoglobulins. *Ann. Rev. Biochem.* **48:** 961–997.

Arnold, N., Bhatt, M., Ried, T., Ward, D. C., and Weinberg, J. (1992). Fluorescence *in situ* hybridization on banded chromosomes. *In* "Techniques and Methods in Molecular Biology: Non-radioactive Labeling and Detection of Biomolecules" (C. Kessler, Ed.), Springer-Verlag, Berlin, Heidelberg.

Bahler, M., Eppenberger, H. M., and Wallimann, T. (1985a). Novel thick filament protein of chicken pectoralis muscle: The 86 kd protein. II. Distribution and localization. *J. Mol. Biol.* **186:** 393–401.

Bahler, M., Eppenberger, H. M., and Wallimann, T. (1985b). Novel thick filament protein of chicken pectoralis muscle: The 86 kd protein. I. Purification and characterization. *J. Mol. Biol.* **186:** 381–391.

Bennett, P., Craig, R., Starr, R., and Offer, G. (1986). The ultrastructural location of C-protein, X-protein and H-protein in rabbit muscle. *J. Muscle Res. Cell Motil.* **7:** 550–567.

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

Chou, P. Y., and Fasman, G. D. (1978). Empirical predictions of protein conformation. *Ann. Rev. Biochem.* **47:** 251–276.

Cooper, J., and Trinick, J. (1984). Binding and localization of AMP deaminase in rabbit psoas muscle myofibrils. *J. Mol. Biol.* **177:** 137–152.

Craig, R., and Offer, G. (1976). The location of C-protein in rabbit skeletal muscle. *Proc. R. Soc. London B.* **192:** 451–461.

Dennis, J. E., Shimizu, T., Reinach, F. C., and Fischman, D. A. (1984). Localization of C-protein isoforms in chicken skeletal muscle: Ultrastructural detection using monoclonal antibodies. *J. Cell Biol.* **98:** 1514–1522.

Einheber, S., and Fischman, D. A. (1990). 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.

Epstein, H. F., and Fischman, D. A. (1991). Molecular analysis of protein assembly in muscle development. *Science* **251:** 1039–1044.

Fan, Y-S., Davis, L. M., and Shows, T. B. (1990). Mapping small DNA sequences by fluorescence *in situ* hybridization directly on banded metaphase chromosomes. *Proc. Natl. Acad. Sci. USA* **87:** 6223–6227.

Fischman, D. A., Vaughan, K., Weber, F., and Einheber, S. (1991). Myosin binding proteins: Intracellular members of the immunoglobulin superfamily. In "Frontiers in Muscle Research" (E. Ozawa, T. Masaki, and Y. Nabeshima, Eds.), pp. 211–222, Elsevier Science Publishers, Amsterdam.

Hofmann, P. A., Greaser, M. L., and Moss, R. L. (1991a). C-protein limits shortening velocity of rabbit skeletal muscle fibres at low levels of Ca^{2+} activation. *J. Physiol. (London)* **439:** 701–715.

Hofmann, P. A., Hartzell, H. C., and Moss, R. L. (1991b). Alterations in Ca^{2+} sensitive tension due to partial extraction of C-protein from rat skinned cardiac myocytes and rabbit skeletal muscle fibers. *J. Gen. Physiol.* **97:** 1141–1163.

Koenig, M., Hoffman, E., Bertelson, C., Monaco, A., Feener, C., and Kunkel, L. (1987). Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell* **50:** 509–517.

Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, K., and Baralle, F. E. (1985). Primary structure of human fibronectin: Differential splicing can generate at least 10 polypeptides from a single gene. *EMBO J.* **4:** 1755–1759.

Nassim, F., Spears, P., Hoffman, H., Kuo, H., and Grabowski, P. (1990). A sequential splicing mechanism promotes selection of an optional exon by repositioning a downstream 5′ splice site in preprotachykinin pre-mRNA. *Genes Dev.* **4:** 1172–1184.

Offer, G. (1972). C-Protein and the periodicity in the thick filaments of vertebrate skeletal muscle. *Cold Spring Harbor Symp. Quant. Biol.* **37:** 87–93.

Offer, G., Moos, C., and Starr, R. (1973). A new protein of the thick filaments of vertebrate skeletal myofibrils: Extraction, purification and characterization. *J. Mol. Biol.* **74:** 653–676.

Okagaki, T., Weber, F. E., Vaughan, K. T., Mikawa, T., Reinach, F. C., and Fischman, D. A. (1993). The myosin-binding domain of skeletal muscle My BP-C (C-protein) resides in the C-terminal, immunoglobulin C2 motif. *J. Cell Biol.*, submitted for publication.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85:** 2444–2448.

Petersen, T. E., Thogersen, H. C., Skortengard, K., Vibe-Pedersen, K., Sahl, P., Sottrup-Jensen, L., and Magnusson, S. (1983). Partial primary structure of bovine plasma fibronectin: Three types of internal homology. *Proc. Natl. Acad. Sci. USA* **80:** 137–141.

Reinach, F. C., Masaki, T., Shafiq, S., Obinata, T., and Fischman, D. A. (1982). Isoforms of C-protein in adult chicken skeletal muscle: Detection with monoclonal antibodies. *J. Cell Biol.* **95:** 78–84.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," second ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Schultheiss, T., Lin, Z., Lu, M-H., Murray, J., Fischman, D. A., Weber, K., Masaki, T., Imamura, M., and Holtzer, H. (1990). Differential distribution of subsets of myofibrillar proteins in cardiac nonstriated and striated myofibrils. *J. Cell Biol.* **110:** 1159–1172.

Speit, G. (1984). Considerations on the mechanism of different Giemsa staining of BrdU-substituted chromosomes. *Hum. Genet.* **67:** 264–269.

Squire, J. (1981). "The Structural Basis of Muscular Contraction," Plenum, New York.

Staden, R. (1982). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Res.* **10:** 2951–2961.

Starr, R., and Offer, G. (1971). Polypeptide chains of intermediate molecular weight in myosin preparations. *FEBS. Lett.* **15:** 40–44.

Starr, R., and Offer, G. (1982). Preparation of C-protein, H-protein, X-protein, and phosphofructokinase. *Methods Enzymol.* **85:** 130–138.

Starr, R., and Offer, G. (1983). H-protein and X-protein: Two new components of the thick filaments of vertebrate skeletal muscle. *J. Mol. Biol.* **170:** 675–698.

Vaughan, K. T., Weber, F. E., Einheber, S., and Fischman, D. A. (1993). Molecular cloning of chicken MyBP-H (86kD Protein) reveals extensive homology with MyBP-C (C-Protein) with immunoglobulin C-2 and fibronectin type III repeats. *J. Biol. Chem.*, in press.

Williams, A. F. (1987). A year in the life of the immunoglobulin superfamily. *Immunol. Today* **8:** 298–303.

Williams, A. F., and Barclay, A. N. (1988). The immunoglobulin superfamily—Domains for cell surface recognition. *Annu. Rev. Immunol.* **6:** 381–405.