Immunological identification of the genes encoding the four myosin heavy chain isoforms of Caenorhabditis elegans
(fusion peptides/epitope mapping/thick filament assembly/unc-54 gene)

DAVID M. MILLER*, FRANK E. STOCKDALE†, AND JONATHAN KARN

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom

Communicated by Sydney Brenner, November 21, 1985

ABSTRACT The nematode Caenorhabditis elegans produces four distinct myosin heavy chain (MHC) isoforms, A, B, C, and D. The MHC A and MHC B proteins are coordinately expressed in the body wall muscle and are incorporated into different regions of a single kind of thick filament. MHC C and MHC D are exclusively produced in the pharyngeal muscle. Previous studies of mutations that affect MHC B have shown that this isoform is encoded by the unc-54 gene. Three other MHC genes, myo-1, myo-2, and myo-3, were isolated from a C. elegans genomic library by hybridization with fragments of the unc-54 gene. We have now established the MHC isoform encoded by each gene. Restriction fragments from each of these genes were cloned in the plasmid expression vector pUR288, producing fusion proteins between Escherichia coli β-galactosidase and portions of the MHC rod domains of each gene. The hybrid proteins were screened with a panel of 18 isofomspecific monoclonal antibodies. The results demonstrate that myo-1 encodes MHC D, myo-2 encodes MHC C, and myo-3 encodes MHC A.

Multiple isoforms of myosin heavy chain (MHC) are expressed in the muscle tissues of most organisms (1–5). In many systems, the activation of the MHC genes encoding these isoforms is developmentally controlled (2, 3) and subject to neural (4) or hormonal regulation (3). Little is known, however, about the mechanisms of differential MHC gene expression or about the specific contributions of individual MHC isoforms to muscle physiology and structure.

In the small soil nematode Caenorhabditis elegans four distinct MHC isoforms, A, B, C, and D, are expressed in a tissue-specific and developmentally regulated manner. MHC A and MHC B are produced in the body muscle cells, whereas MHC C and MHC D are synthesized exclusively in the pharyngeal musculature (5–8). Studies of larval morphogenesis have indicated that MHC A and MHC B accumulate in a constant ratio, suggesting that their expression is coordinately regulated (9). There is no evidence that embryonic stages of C. elegans express unique MHCs, unlike vertebrate embryos (3).

The C. elegans MHC isoforms can be distinguished by monoclonal antibodies raised against native nematode myosin (7, 10). Immunohistochemical analysis of muscle tissue and isolated thick filaments has shown that MHC A and MHC B are incorporated into different domains of a single kind of body wall muscle thick filament. MHC A occupies the central region of each thick filament and therefore would appear to participate in the initiation of filament assembly. MHC B is present in the lateral portions of the filament and is exclusively involved in the reactions that elongate the structure (10).

Studies of numerous mutations that affect the MHC B and produce disorganized body wall muscle have shown that the unc-54 locus encodes MHC B (5, 8, 11, 12). Three other myosin heavy chain genes, myo-1, myo-2, and myo-3, have been isolated from bacteriophage λ genomic libraries probed with labeled restriction fragments from the unc-54 gene. Preliminary sequence analysis of these clones suggested that each encodes a distinct heavy chain (11, 13). However, mutations linking these MHC isoforms and the myo-1, myo-2, or myo-3 genes have not been found. Therefore other approaches were needed to determine if these genes are transcribed, and if so, which of the additional isoforms (MHC A, C, or D) they produce.

We have answered these questions by using a plasmid vector to express portions of each of the cloned MHC genes in Escherichia coli as fusion proteins with β-galactosidase (14). The hybrid proteins were screened with a panel of specific monoclonal antibodies (7, 10) to identify epitopes corresponding to each MHC isoform. The results demonstrate that myo-1 encodes MHC D, myo-2 encodes MHC C, and myo-3 encodes MHC A. The identification of the nematode MHC genes should facilitate the analysis of the mechanisms regulating the expression of the corresponding MHC isoforms and their assembly into muscle thick filaments.

MATERIALS AND METHODS

Plasmid Construction. Enzymes for recombinant DNA experiments were obtained from New England Biolabs or Boehringer Mannheim or were prepared in this laboratory. HindIII fragments from bacteriophage λ clones of nematode MHC genes were cloned in the lacZ′ expression vector pUR288 (14). The vector was cleaved with HindIII and treated with alkaline phosphatase prior to fragment insertion. Recombinants were recovered after transformation of E. coli (supF′, F′lacZ′M15 lacF′) and selection for ampicillin resistance. Colonies were screened immunologically as described below. The λ clones were SG24/1, unc-54; SG25/32, myo-1; SG25/39, myo-2; and SG25/23, myo-3 (13). The structures of the immunologically reactive clones were established by gel analysis of HindIII-digested plasmids.

Immunological Methods. Rabbit antiserum to C. elegans myosin was provided by A. R. MacLeod (8). Contaminating reactivity to bacterial antigens was removed from the serum by the method of Young and Davis (15). Monoclonal antibodies to nematode MHCs were used as ascites fluids or as purified IgG (10). Horseradish peroxidase coupled to rabbit antibodies to mouse IgG was from DakoPatts.

Abbreviations: MHC, myosin heavy chain; bp, base pair(s); S1 and S2, subfragments 1 and 2 of MHC; LMM, light meromyosin.

*Present address: Department of Zoology, North Carolina State University, P.O. Box 7617, Raleigh, NC 27695-7617.
†Present address: Stanford University School of Medicine, Room M211, Stanford, CA 94305.
Immunoreactive clones were detected by a modification of the method of Helfman et al. (16). Colonies were replica plated onto nitrocellulose filters soaked in isopropyl \(\beta\)-D-thiogalactoside (Sigma) at 200 \(\mu\)g/ml and grown on H-media plates overnight. The colonies were lysed with chloroform vapor and soaked in 0.01% NaDodSO\(_4\)/150 mM NaCl/0.1 mM phenylmethylsulfonyl fluoride/25 mM Tris-HCl, pH 7.4, for 1 hr to remove bacterial debris. The replicas were blotted with Whatman 3 MM paper moistened with 5 mM MgCl\(_2\)/150 mM NaCl/50 mM Tris-HCl, pH 7.4, to remove the NaDodSO\(_4\). The filters were then immersed in blocking solution (3% bovine serum albumin/5% goat serum/75 mM NaCl/30 mM Tris-HCl, pH 7.4) and allowed to react with anti- 

**RESULTS**

Nematode MHC Gene Family. The unc-54 gene was the first member of the C. elegans MHC gene family to be cloned and sequenced (8, 11). Mutations that deleted portions of the unc-54 gene were used to distinguish the unc-54 clone from other members of the MHC family and to demonstrate that unc-54 encodes MHC B (8). Comparisons between the deduced amino acid sequence of the unc-54 heavy chain and other known MHC sequences, including that of rabbit skeletal myosin, showed that MHCs are well conserved, especially in the head region (11). To identify the remaining MHC genes, a genomic library from C. elegans was screened with probes derived from the unc-54 sequences encoding both the head and rod regions of MHC. Three new genes, myo-1, myo-2, and myo-3, were isolated.

The structures of the MHC genes, determined by DNA sequencing, are shown in Fig. 1. Only the sequence of the unc-54 gene is complete, but extensive regions of the other three genes have been sequenced (J.K., I. Maruyama, and N. J. Dibb, unpublished data). The coding sequences of the genes are highly homologous, with the S1 regions showing 82% matching amino acids and the S2 and LMM regions showing 53% matching amino acids. However, each gene includes uniquely placed introns and distinct sequence features (13).

**Design of the Experiments.** The structural studies of the nematode MHC genes strongly suggested that each gene specifies a separate myosin isoform. To demonstrate that these genes are transcribed and to determine which isoform they produce (MHC A, MHC C, or MHC D) we expressed fragments of each MHC protein in bacteria (14) for reaction with isoform-specific monoclonal antibodies (7, 10).

In each gene the entire S2 region and most of the LMM region is encoded by a single long exon. Eighteen monoclonal antibodies are available, most of which recognize epitopes in the MHC rod (7, 10) (D.M.M., G. C. Berliner, and H. F. Epstein, unpublished data). It seemed likely, therefore, that

---

**Fig. 1.** Structure of C. elegans MHC genes. For each gene, coding regions are indicated as solid bars, introns as gaps. Only the unc-54 sequence is complete. Regions in the other genes that have not been sequenced are indicated by open boxes. In myo-3, there are gaps in the subfragment 1 (S1) and light meromyosin (LMM) regions of the sequence. Underneath the coding sequence maps are indicated the positions of the S1/S2 junction (after amino acid 862) and the S2/LMM junction (after amino acid 1183). Each of the gene maps has been aligned in the S2 region, which is represented as a striped box. A HindIII cleavage map for each gene is also shown. HindIII fragments are referred to by the indicated sizes [in base pairs (bp)] throughout the text and in Fig. 2.
fused proteins derived from this region would be detected in immunoassays with at least some of these antibodies.

The construction of suitable clones was greatly simplified by our observation that the rod regions of each MHC gene contained numerous short HindIII (A→AGCTT) fragments. All the HindIII sites showed the same translational phasing because each derived from the frequent appearance of the dipeptide sequence Lys-Leu (AAG-CTT). The HindIII fragments were cloned in a single expression vector, pUR288 (14). This plasmid allows insertion of DNA restriction fragments into a polylinker located near the 3' end of the lacZ gene (with an appropriate phasing).

Detection of Fusion Proteins. Colonies expressing fusion proteins were identified by immunoassays (16) of nitrocellulose filter replicas either with rabbit antisera to nematode myosin (8) or with monoclonal antibodies (7, 10). Clones expressing fusion peptides encoded by most of the HindIII fragments from the rod region of each MHC gene were isolated in this manner (Fig. 2 and Table 1).

The induced fusion proteins constituted a major fraction of the bacterial protein and migrated on NaDodSO₄/polyacrylamide gels in the expected high molecular weight range between β-galactosidase (Mt = 116,000) and nematode MHC (Mt = 226,000) standards (Fig. 2A). Most of the clones contained a single inserted HindIII fragment (as determined by restriction endonuclease mapping). For these clones, the relative sizes of the hybrid proteins are directly related to the length of the MHC HindIII insert. The coding sequence of the 2156-bp HindIII fragment from myo-2 is interrupted near its 3' end by a short intron (Fig. 1). Translation of the fusion peptide is presumed to stop near this junction. The molecular weight of the fusion peptide encoded by the 117-bp HindIII fragment of the unc-54 gene is anomalously large because this clone contains multiple 117-bp fragment inserts (data not shown). In a second case of multiple fragment insertion, a novel fusion peptide was produced by linking of the 111-bp and 1054-bp fragments from the unc-54 gene (Fig. 2A).

Immunological Identification of MHC Genes. Immunoblotting was used to define the reactivity of the fusion peptides with each of the monoclonal antibodies. Some examples are shown in Fig. 2; complete results are presented in Table 1. The patterns of reactivity of the monoclonal antibodies establish that each of the four MHC genes encodes one of the four MHC isoforms.

Antibody 5-6 is highly specific for an epitope in the S2 region of MHC A (10). This antibody reacts only with the hybrid protein encoded by the 460-bp HindIII fragment from the S2 domain of the myo-3 gene and does not recognize epitopes encoded by homologous fragments from the other three MHC genes (Fig. 2B). Similarly, antibody 9.2.1, which is specific to MHC C (7), preferentially recognizes the large fusion peptide encoded by the 2156-bp HindIII fragment from the myo-2 rod domain (Fig. 2C). The reaction of the MHC B-specific antibodies, 5-2, 5-3, and 12.1, with specific fragments from unc-54 is in agreement with the earlier assignment of MHC B to this gene (Table 1). The occurrence of direct amino acid sequence repeats within the unc-54 rod region (18) may account for the reaction of antibody 12.1 with an epitope encoded by both the 117-bp and the 294-bp HindIII fragments of the unc-54 gene.

A number of antibodies tested showed multiple reactivity towards the different MHCs (7, 10). These antibodies showed

![Fig. 2](image-url)  
Fig. 2. Reaction of fusion peptides with specific monoclonal antibodies. Clones containing HindIII fragments from each of the MHC genes in pUR288 were grown in the presence of isopropl β-thiogalactoside and lysed with NaDodSO₄, and fusion proteins were separated on a NaDodSO₄/6% polyacrylamide gel. (A) Coomassie blue-stained gel. The numbers above each lane refer to the size of each HindIII insert in Table 1. Immunoblot of the gel shown in A reacted with antibody 5-6 (B), antibody 9.2.1 (C), and antibody 28.2 (D). Nematode myosin (MHC Mt = 226,000) and E. coli β-galactosidase (β-gal; Mt = 116,000) standards were run in the outside lanes of the gel, and the MHC is visible in each blot.
similar patterns of multiple reactivity in their reactions with the fusion peptides. For example, antibody 28.2 reacts strongly with MHC B on immunoblots but only weakly (e.g., visible only at high antibody concentration) with MHC A (10). Similarly, antibody 28.2 reacts most strongly with the fusion peptide from the 357-bp fragment of the unc-54 gene (Fig. 2D) but also exhibits a weak reaction with the fusion peptide from the 460-bp fragment of myo-3. Antibody 25.1 shows affinity for MHC B and MHC A and also reacts with the fusion peptides from the 357-bp fragment of unc-54 and the 460-bp fragment of myo-3. These patterns of antigenicity confirm the assignment of MHC A to the myo-3 gene. The reactivity of antibody 5-12 is consistent with the identification of myo-2 as the gene for MHC C; this antibody recognizes both MHC C and MHC B and reacts with the fusion peptide of the 2156-bp myo-2 fragment and the fusion peptide of the 1054-bp unc-54 fragment.

The evidence linking myo-1 to MHC D is less direct than the evidence linking myo-3 to MHC A and myo-2 to MHC C. MHC A and D migrate as a single band on NaDodSO4 Laemml gel systems that were not resolved on the original immunoblots used to classify the reactivity of the monoclonal antibodies (see * footnote of Table 1) (6). However, since antibody 5-25 reacts strongly with an immunoblots of the A + D MHC band from Laemml gels and with the epitope encoded by the 1577-bp fragment of myo-1 (Table 1), myo-1 must encode either MHC A or MHC D. Given that MHC A is clearly assigned to gene myo-3 by antibody 5-6 (see also Discussion), it follows that myo-1 must encode MHC D.

Six of the 18 antibodies we tested (5-4, 5-8, 5-14, 5-16, 5-17, and 5-22) did not react with any of the fusion peptides (data not shown).

Immunoblots of proteolytic digests of MHC B have mapped the epitopes recognized by antibodies 5-8, 5-22, and 5-13 to the carboxyl end of LMM (D.M.M., G. C. Berliner, and M. F. Epstein, unpublished data). This region is not represented in our fusion protein bank because it is interrupted by several introns. However, it is also possible that some of the nonreactive antibodies are sensitive to structural changes in the fusion proteins.

Taken together, our data suggest the following assignments: myo-1 encodes MHC D, myo-2 encodes MHC C, myo-3 encodes MHC A, and unc-54 encodes MHC B.

DISCUSSION

*C. elegans* Has Four Sarcomeric MHC Genes. We have utilized recombinant DNA and immunological approaches to identify the myosin isoforms encoded by each of the four cloned *C. elegans* MHC genes. Hybrid proteins representing most of the rod region (and all of the S2 region) of each MHC isoform were produced in *E. coli*. Specific monoclonal antibodies were allowed to react with these fusion proteins to identify DNA fragments encoding epitopes from each of the four MHC polypeptides.

Our results indicate that each nematode MHC isoform is encoded by a discrete gene. MHC isoforms in vertebrates also appear to be encoded by separate genes (1–3). In *Drosophila* at least two different MHC isoforms are produced from a single gene by developmentally regulated RNA splicing (19). It is likely that the four MHC genes unc-54, myo-1, myo-2, and myo-3 represent the entire nematode sarcomeric MHC gene family. No other MHC-like sequences were detected in exhaustive screens of genome banks from *C. elegans* or on Southern blots (J.K., unpublished observations) (11, 13). The reactivities of the monoclonal antibodies with the fusion peptides are also in agreement with this conclusion. We tested 12 different monoclonal antibodies (7, 10) that recognize epitopes encoded by fragments from all four genes. The patterns of reactivity that we observed are fully consistent with a one-to-one correspondence between the cloned MHC genes and the known myosin isoforms.

Genetic Implications. Now that the identity of each MHC gene is established it should be possible to apply genetic methods to further study the functions of the MHC isoforms.
In situ hybridization experiments show that the nematode MHC genes are not clustered but are distributed on three different chromosomes at discrete locations. The myo-3 gene hybridizes to a region of chromosome V that includes the sup-3 locus (20). Sup-3 mutations are suppressors of the uncoordinated phenotype conferred by unc-54 null alleles and by a missense mutation in the gene for paramyosin, unc-15 (21). Improved movement is correlated with the increased accumulation of MHC A, which apparently compensates either for the absence of MHC B or for defective paramyosin (6). Recent experiments show that the sup-3 mutant dpy-1 produces a new restoration deletion pattern in a region adjacent to the myo-3 gene and contains multiple copies of the myo-3 gene (22).

Epitope Mapping. The methods of proteolytic cleavage and immunoelectron microscopy have been employed to map the recognition sites of monoclonal antibodies to myosins from several organisms. Binding of antibodies to these epitopes has identified subregions of the myosin molecule that are involved in filament formation, ATPase activity, and the ability of myosin to catalyze movement in vitro (23, 24). We have mapped the epitopes of 12 monoclonal antibodies to fusion peptides encoded by MHC gene fragments ranging in size from 2150 to 117 bp. It should be possible to define the epitopes more accurately by expressing still shorter fragments. This information may be useful for correlating antibody decoration of native thick filaments with the proposed structure of the myosin rod (18) and with alternative models of myosin packing in the thick filament (25).

Thick Filament Assembly. MHC A and MHC B form homodimers, myosin A and myosin B (26), which are differentially incorporated into nematode body wall thick filaments (10). Myosin A occupies the medial portion of the structure, whereas myosin B is restricted to the polar regions. These locations are correlated with distinct domains and suggest unique functions for the myosins. The central base zone is composed of myosin A. In this region, myosin molecules are arranged in antiparallel-packed arrays (25). Filament elongation occurs with the bipolar addition of parallel oriented myosin molecules to this nucleating complex. Both myosin A and myosin B are incorporated into these lateral regions. Thus, myosin A must necessarily participate in the initial phase of filament assembly and can pack in both parallel and antiparallel orientations. On the other hand, myosin B is exclusively involved in parallel packing during the elongation of the filament and the reactions that terminate polymerization (10).

The complete nucleotide sequence of the MHC B gene has been previously determined and a structural model of the myosin B rod has been deduced. MacLachlan and Karn (18) have shown how the periodic array of hydrophobic residues specifies the coiled-coil structure of the rod and how the alternating bands of positive and negative charge on the surface of the rod would allow packing in the thick filament based on stagger of 98 residues (i.e., axial repeat of 145.5 A). These basic features are preserved in the known MHC A rod sequence, but limited regions of significant sequence divergence from the MHC B gene are present. We expect that further analysis of these sequences will begin to reveal the molecular basis for the different interactions of the myosin A and myosin B isoforms during the assembly of the nematode thick filament.

A consideration of the myosin sequences alone, however, is unlikely to provide a complete explanation of these events. Differential assembly of the myosins may also depend upon specific interactions with other thick filament components. In addition to myosin A and myosin B, the C. elegans body wall thick filament contains paramyosin (27). This rodlike protein occupies the core region of invertebrate muscle thick fila-

ments and interacts with myosin molecules on the surface (28). The unc-15 gene, which encodes paramyosin, has recently been cloned, making studies of its sequence possible (H. Kagawa, J.K., and S. Brenner, unpublished results). Biochemical and ultrastuctural studies have revealed that the core of the C. elegans thick filament contains another major component in addition to paramyosin (29). Finally, specific mutations that disrupt thick filament structure have identified genes that either encode as yet uncharacterized thick filament components or express functions required for thick filament assembly (30, 31). We expect that continued molecular and genetic studies of the muscle-defective mutants of C. elegans will help elucidate muscle structure and the mechanisms of muscle assembly.

We thank R. H. Waterston, S. Brenner, P. Matsudaira, A. Weeds, and N. J. Dibb for their advice and D. G. Albertson for communicating unpublished results. The authors gratefully acknowledge the contribution to this work of H. Epstein, in whose laboratory the monoclonal antibodies were prepared. D. M. M. was supported by grants from the Burroughs–Wellcome Foundation and the European Molecular Biology Organization and a National Institutes of Health Biomedical Research Support Grant to North Carolina State University. F. E. S. was a Fellow of the John Simon Guggenheim Memorial Foundation.