Organization of the human skeletal myosin heavy chain gene cluster

(muscle genes/yeast artificial chromosomes/physical map)

Sung-Joo Yoon*, Stephanie H. Seiler†, Raju Kucherlapati*, and Leslie Leinwand††

Departments of *Molecular Genetics and †Microbiology and Immunology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461

Communicated by Frank Lilly, October 5, 1992

ABSTRACT Myosin is an important structural and enzymatic component of skeletal muscle. Multiple myosin isoforms are encoded by a multigene family and are expressed in different developmental stages and fiber types. In humans and mice, skeletal myosin heavy chain (MYH) genes are clustered on a single chromosome (17p and 11, respectively). Since the structural organization of the gene cluster may affect its expression as well as shed light on MYH genetic alterations, a physical map of the human MYH gene cluster was constructed. Nine yeast artificial chromosomes containing MYH genes were isolated and used to construct a contiguous set (contig) of overlapping yeast artificial chromosomes. This contig encompasses a genetic marker mapped to 17p13.1. Six MYH genes were located within a 500-kilobase segment of human DNA. The order of the genes within this cluster does not correspond to the developmental pattern of expression of individual members.

Myosin is the molecular motor of muscle and is composed of a pair of myosin heavy chains (MYHs) and two pairs of nonidentical light chains. At least seven sarcomeric MYH isoforms are expressed in mammalian skeletal muscle, including embryonic, perinatal, fast IIA, fast IIB, fast IId, extraocular, and slow (β cardiac) forms (1–3). MYH genes are expressed at different times during development and in different fiber types (4–8). The MYH isoform content of any given muscle can also vary in response to physiological changes such as innervation or hormonal perturbation (for a review, see ref. 9), and quantitative differences in myosin ATPase activity have been correlated with the contractile velocity of the muscle (10).

The cardiac and skeletal MYH genes are clustered on two different chromosomes in mice and humans. The α and β cardiac MYH genes are separated by only 4 or 5 kilobases (kb) of DNA on human and mouse chromosomes 14 (11–13). With the exception of the slow/β cardiac MYH gene, the skeletal MYH genes are clustered on human chromosome 17 and mouse chromosome 11 (12, 14, 40). Linkage analysis with a polymorphic probe proximal to an adult skeletal MYH gene (MYH2) localized this gene to human 17p13 (15). In addition, two other human skeletal MYH genes (embryonic, MYH3; perinatal, MYH8) have been mapped to 17p (16, 17). Several lines of evidence suggest that all or most of the vertebrate skeletal MYH genes are in close genetic and physical proximity to each other. Interspecific backcrosses and pulsed-field gel electrophoresis (PFGE) analysis in mice demonstrated that three MYH genes (embryonic, perinatal, and fast IIB) are clustered on mouse chromosome 11 within 370 kb of each other (12, 18). Two additional mouse MYH genes, fast IIA and a gene expressed abundantly in several adult skeletal muscles (termed MdMs), are separated by about 5 kb of DNA (19). The order and organization of the human skeletal MYH genes are not known.

The clustering of MYH genes is in distinction to the dispersed location of members of other contractile protein multigene families such as those encoding actin and myosin light chains. These differences in gene organization of the different families have led to the proposal that clustering of MYH genes has a regulatory significance (39). Although characterization of the regulatory elements of cardiac MYH genes has begun (20, 21), little is known about the regulatory sequences within the skeletal MYH genes. If the globin gene cluster serves as a model, it seems likely that the organization of the skeletal MYH genes will be important for their regulation. In addition to identifying domains or elements regulating expression of the skeletal MYH genes, a physical map will be essential in identifying gross structural mutations in skeletal MYH genes responsible for muscle disease. Mutations in the human β cardiac MYH gene have recently been demonstrated in several kindreds with the autosomal dominant disease familial hypertrophic cardiomyopathy (22), lending support to the hypothesis that mutations in the structural components of muscle may be responsible for some forms of muscle genetic disease.

To understand the organization of the human MYH gene cluster, yeast artificial chromosome (YAC) clones containing MYH genes were isolated and analyzed, revealing that six distinct genes are located within a 500-kb interval. Five correspond to previously identified MYH genes and/or cDNAs: embryonic; perinatal; two fast II MYH genes; and an adult skeletal MYH gene designated MYH2. The identity of the sixth gene has not yet been determined. The physical order of these genes does not correspond to their developmental program and may represent evolutionary expansion of the gene family to accommodate the physiological needs of mammalian muscle.

MATERIALS AND METHODS

Nomenclature. MYH genes have been given the numerical designations as follows: embryonic MYHemb, MYH3; perinatal MYHpes, MYH8; adult fast skeletal MYH (not subtype defined, but described as a genomic clone, AMHC10, in ref. 3), MYH2. YAC Screening. PCR-based screening (23) was used to isolate clones containing the human MYH locus from the Washington University human DNA YAC library (24) using MYHemb- and MYHpes-specific primers (MYHEM and MYHPM, respectively; see Table 1).

DNA Analysis. DNAs for PFGE were prepared by established procedures (25). Restriction fragments or intact yeast chromosomes in agarose plugs were separated in 1% agarose gel using a contour-clamped, homogeneous electric field apparatus (Bio-Rad) as described (26). DNA from yeast for conventional electrophoretic analysis was prepared by established procedures (27), and restriction enzyme-digested

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
yeast DNA was separated in a 0.8% agarose gel in 1× TAE (28). DNAs were transferred to GeneScreenPlus (DuPont/ NEN) and the filters were hybridized with appropriate probes in a solution containing 1 M NaCl, 1% SDS, and 10% dextran sulfate at 65°C. Fifty micrograms of sheared salmon sperm DNA per ml was added during prehybridization and with the denatured probe.

**Rescue of DNA Sequences from the Ends of YACs.** Three methods for end sequence rescue were used. In the first, referred to as vectorette or bubble PCR (29), yeast DNA was digested with Ava II, ligated with a synthetic vectorette (see below), and amplified with one primer corresponding to part of the vectorette and the other corresponding to the left (VL) or right (VR) end of the YAC vector. DNA ligation reactions were carried out under conditions described by Maniatis et al. (28). Amplification was conducted in a total volume of 50 μl containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 250 μM each of one of the four dNTPs, and 1 μM of each primer in a Perkin–Elmer thermal cycler under conditions of denaturation at 94°C for 4 min, 38 cycles of denaturation (92°C for 1 min), annealing (66°C for 30 sec), and extension (72°C for 2 min). The second method for end rescue is referred to as vector-Alu PCR, using the VL and VR primers and the Alu primer 517 or 559 (30). These reactions were carried out in 50-μl volumes containing 5 μg of yeast DNA in the reaction mixture described above. Amplification conditions were denaturation at 94°C for 4 min, followed by 35 cycles of denaturation (94°C for 1 min), annealing (65°C for 2 min), and extension (72°C for 5 min), followed by a final extension period of 10 min. The third method to generate end sequences is referred to as inverse PCR (31). Yeast DNA was digested with EcoRV (for left end) and Xma I (for right end), and the products were ligated under conditions that enhance intramolecular ligation and amplified using primers within the vector that are oriented in opposite directions.

The VL primer used for Alu-vector PCR is a 29-mer of the composition 5′-CACCCTCTGGAGCAGCTGTCGGAC- CGC-3′ and is designated RK152. The VR primer used for vectorette PCR, designated LL384, is a 30-mer of the composition 5′-TTAAGCCGCAACGTCTTTATATTATCAC- TAC-3′. The VR primer used for both types of PCR reactions was designated RK151 and had the composition 5′-ATAT- AGGCGCCAGAAACCGCAACGTCTTTATATTATCAG- GAGAG-3′. RK151 had the partially complementary sequence 5′-CTCTCCTCCTGTAAGATCCCTTTGTCG- TACGAGAATCTGCTCTTCTCTT-3′. To generate the vectorette, oligonucleotides RK154 and RK155 were annealed and used for ligation with Ava II-digested yeast DNA. The vectorette primer designated RK153 is 5′-CGAATCCG- TAACCGGCTGGTACCGAATCTGCTCTTCTCTT-3′. These were annealed and used for ligation with Ava II-digested yeast DNA. The vectorette primer designated RK153 is 5′-CGAATCCG- TAACCGGCTGGTACCGAATCTGCTCTTCTCTT-3′. The following PCR primers were used in the inverse PCR reactions. For the left end, we used the primers RK156 (5′-GGTGGTTTAAG- GCCAAGA-3′) and RK204 (5′-GGCAGTCGTCGTG- GAATGGAC-3′). For the right end, we used RK205 (5′-GTGCGAACGCGCGATCTCAAG-3′) and LL385 (5′-GTAT- TGCTCCTTACTTCCCA-3′).

**PCR Conditions for Sequence-Tagged Sites (STSs).** PCR with gene-specific primers (Table 1) was carried out in 10 μl with 0.5 μM primers using 5 ng of yeast DNA or 100 ng of human DNA as controls in the conditions described above. Initial denaturation was at 94°C for 4 min, followed by 30 cycles of 94°C denaturation (1 min), 57°C (for MYHpn), or 62°C (for MYHEM, MYHPN) annealing (1 min), and 72°C extension. PCR reactions with other STS primers were carried out in 50 μl with 1 μM of each primer using 25 ng of yeast or 500 ng of human DNA. The PCR conditions for all STS pairs are initial denaturation at 94°C for 4 min, followed

<table>
<thead>
<tr>
<th>Table 1. STSs for the MYH gene cluster contig</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STS name</strong></td>
</tr>
<tr>
<td>A213C4</td>
</tr>
<tr>
<td>MYHE5</td>
</tr>
<tr>
<td>MYHEM</td>
</tr>
<tr>
<td>MYHE3</td>
</tr>
<tr>
<td>B120C11</td>
</tr>
<tr>
<td>MYHAS8</td>
</tr>
<tr>
<td>A288E1</td>
</tr>
<tr>
<td>MYHIla</td>
</tr>
<tr>
<td>MYHP5</td>
</tr>
<tr>
<td>MYHPM</td>
</tr>
<tr>
<td>MYHP3</td>
</tr>
<tr>
<td>A56B10R</td>
</tr>
<tr>
<td>B120C11R</td>
</tr>
</tbody>
</table>

*bp, Base pairs.

*Annealing temperature/time of extension (°C/min).

*This reaction did not have a designated extension period.
by 30 cycles of denaturation at 94°C for 1 min and annealing for 30 sec at temperatures shown in Table 1.

Probes. Total human DNA was used as a probe for PFGE analysis and for DNA fingerprint assays. Gene-specific probes, MYHem, MYHsm, MYHsa, and MYHb, were gel-purified fragments from the 3' untranslated regions of the corresponding genes and have been described (refs. 17, 32, and 33; I. Karsch-Mizrachi and L.L., unpublished data). The consensus 3' probe designated LL1000 was prepared by PCR with the perinatal cDNA as a template using primers from exons 37 and 39, which correspond to sequences identical in all published mammalian skeletal MYH genes. These primers were 5'-CAGGACACCAGCGCCCA-3' (sense) and 5'-TCCTCGGCTCTCCAGCTC-3' (antisense). End fragments rescued by PCR were digested with EcoRI to remove vector sequence, followed by gel purification. All probes were radiolabeled using the random primer method of Feinberg and Vogelstein (34).

Cell Lines. Two somatic cell hybrid cell lines, kindly provided by R. E. K. Fournier, were 7AE-13 and 7AC-2. Hybrids 7AE-13 and 7AC-2 contain intact human chromosomes 17 and 17q, respectively, in a rat background.

DNA Sequencing. Plasmid DNA was prepared and sequenced essentially as described (35). Sequence data were analyzed using the PRIMER program (Genome Center, Whitehead Institute, Cambridge, MA) to generate oligomers for use as primers in the PCR.

RESULTS

Isolation of YACs. To understand the structural organization of the human MYH genes, clones were isolated from a human YAC library by B. Brownstein (Washington University) by PCR using primers (MYHEM; Table 1) that amplify a unique product from the embryonic skeletal MYH gene. This screen yielded nine YACs. A second screening of this library was conducted at the Genome Center of the Baylor College of Medicine using a pair of primers from the perinatal MYH gene (MYHPN). This screen yielded five YACs. The electrophoretic karyotype of each clone was obtained by PFGE and the sizes of the YACs containing human inserts were determined by hybridization to total human DNA and to MYH gene-specific probes. From both screenings, 3 clones had YACs containing human genomic DNA but failed to hybridize to the gene-specific MYHem or MYHsa probes. These clones (A258E1, A201F4, A120F5) were not subjected to further analysis. The sizes of YACs from the remaining 11 clones ranged from 135 to 410 kb. Results of hybridization of the YACs to human DNA are shown in Fig. 1.

Characterization of the YAC Clones. To begin to determine the degree of overlap between the clones, repetitive sequence fingerprints were determined by digesting total yeast DNA with EcoRI, blotting, and hybridization with total human DNA. In the case of YACs that were either indistinguishable or very close in size and in their fingerprint patterns, only one member of these pairs was used for further analysis. This eliminated A256D4, which was identical to A253D4, and eliminated A288F8, which was 20 kb smaller but entirely overlapping with A288E1. This process of elimination left five YACs that hybridized to the MYHem probe and four YACs that hybridized to the MYHsa probe.

Construction of the YAC Contig. Although the fingerprint patterns provided clues to the degree of overlap between different YACs, they were insufficient to construct a YAC contig. A simple method to construct YAC contigs is to generate a series of DNA markers internal to the YACs and from their ends followed by ascertainment of their presence in each of the YACs. Such an analysis permits ordering of the markers as well as of the YACs with respect to each other. The three types of markers we used to construct this contig include MYH gene-specific markers, internal markers generated from Alu-PCR products from individual YACs, and markers from the ends of individual YACs. YAC end sequences were obtained and used to probe each of the YACs by hybridization and/or PCR. If end sequences failed to hybridize to any YACs other than the one from which it was derived, it could represent one of two ends of the contig or be derived from a chimeric YAC. The latter feature was ascertained by determining whether the probe could hybridize to DNA from a somatic cell hybrid containing chromosome 17 as its only human component.

Three different methods were used to isolate end-specific sequences (see Materials and Methods). PCR products corresponding to the 18 ends of the nine YACs were generated. During an initial analysis, 14 of these products were hybridized to the nine YACs. Six of the products hybridized to all YACs, indicating that they contain repetitive sequences. Of the remaining eight end-specific probes tested, four (A288E1L, A288E1R, A56B10R, B120C11L) hybridized to subsets of the family of YACs, enabling us to infer that these probes are chromosome 17 specific and overlapping. These PCR products were cloned and partially sequenced to generate primer pairs that define STSs for this part of chromosome 17 (Table 1). The others (A213C4L, B120C11R, A23D7L, B27A12L) did not hybridize to any YACs other than the one from which they were derived. Primers that define B120C11R amplify the appropriate size product from the DNA of the somatic cell hybrid containing chromosome 17, suggesting that it constitutes one of the ends of the contig. Probe B27A12L was derived from a YAC that contained markers internal to the contig and did not hybridize to other members of the contig. We conclude that this probe is derived from a chromosome other than 17 and the YAC is chimeric. The other two end products, A213C4L and A23D7L, hybridized to the YACs from which they were derived but not to any others, suggesting that one was chimeric and one represents the true end of the contig. Since this determination had no effect on ascertaining the organization of the MYH gene cluster (see below), it was not investigated further. These data established the contig encompassing the MYH gene cluster. The level of chimerism observed (30%) was similar to that described in an earlier report about this library (36). We confirmed the nature of the contig by PFGE analysis of individual YACs (data not shown). The YAC contig established by all of these analyses is shown in Fig. 2. The estimated size of the contig is 650 kb.

Organization of the MYH Genes. To understand the organization of the individual members of the MYH gene family within the contig, DNA from each YAC was hybridized with
The relative distances of individual gene-specific primers (see (16, 32), probe corresponds to MYH2, p10-5, and A56B10R are within a 15-kb region. Open circles, location of individual MYH genes; closed circles, STSs derived from YACs; open boxes, DNA marker; closed boxes, ends that have not been analyzed; dashed lines, uncertainty about the length of that region; wavy lines, sequences from a chromosome other than 17.

MYH gene-specific DNA probes or PCR was carried out with gene-specific primers (see Table 1). These probes correspond to five distinct MYH genes and include the 3′ untranslated regions of the embryonic and perinatal skeletal MYH DNAs (16, 32), an adult fast skeletal MYH cDNA (33), and a cDNA from an MYH gene expressed primarily in adult skeletal muscle (L. Karsch-Mizrachi and L.L., unpublished data). The fifth probe corresponds to an adult skeletal MYH adult (MYH2) that was used earlier in linkage analysis and permitted its assignment to 17p13.1 (15). Because of the known genetic location, this probe allowed us to integrate the physical map of the MYH cluster into the genetic map of chromosome 17. These analyses permitted us to conclude that these five MYH genes are located within the contig. Two of the YACs, B72A12 and A51D4, contained the 5′ end of the MYHp but did not contain the 3′ end of this gene. A56B10 and B120C11 contained the 5′ and 3′ ends of the gene. Based on these data, the transcriptional orientation of MYHp gene, with respect to the contig, can be unambiguously established (left to right in Fig. 2).

To determine whether there were additional MYH genes in the YAC contig and to determine the order of the genes within the cluster, DNA from each of the YACs was digested with EcoRI and blot-hybridized with a probe derived from a highly conserved region corresponding to the myosin rod (LL1000). Results of this analysis are shown in Fig. 3. YACs A213C4, A24D2, A258D4, and A23D7, containing only the MYHem gene, yielded an 11-kb EcoRI fragment (Fig. 3, lanes b–e), consistent with the known organization of the MYHem gene (3). YAC A288E1, which is known to contain the MYHem and the MYHast genes (Fig. 3, lane f), shows the 11-kb band and another slightly larger (≈12 kb) band that must correspond to the MYHast gene. YAC B72A12 contains the MYHast and MYH as well as the 5′ end of the MYHp gene. This YAC (Fig. 3, lane g) has the MYHast band and, in addition, has two other bands 1.7 kb and 0.9 kb in length. YACs A56B10 and B120C11 contain bands corresponding to the MYHast, MYH, and MYHp genes. Thus, the 1.7-kb band must correspond to MYH and the 0.9-kb band must correspond to MYHp. YACs A56B10 and B120C11 contain two other bands (2.1 and 10 kb) that represent two additional MYH genes. The 2.1-kb EcoRI band corresponds to the adult skeletal MYH gene (MYH2) that had previously been used in linkage analysis (15). The relationship between the YAC contig and MYH2 was determined by hybridizing a single-copy sequence from MYH2 (p10-5 in ref. 3) to DNA from the nine YACs. Clones A56B10 and B120C11 showed positive hybridization and yielded an EcoRI band of appropriate size (data not shown). In addition, the STS A56B10R was detected to be present in a λ genomic clone containing the MYH2, thus defining the end of the YAC to occur within the MYH2 gene. The 10-kb EcoRI band corresponds to a sixth MYH gene whose identity is not yet known. Based on these data, we conclude that the order of the six MYH genes in this cluster is MYHast-MYHem-MYH- MYHp (MYH8)-MYH- MYH2. To establish the order of the MYH genes in the YACs, we establish the pattern in the human genome, human genomic DNA was digested with EcoRI and hybridized to LL1000. Several bands in the 10- to 12-kb size

![Fig. 2. Physical map of the skeletal MYH gene cluster. The top line shows the different probes that are used to establish the physical map. The relative distances between all markers are not known. Markers MYH2, p10-5, and A56B10R are within a 15-kb region. Open circles, location of individual MYH genes; closed circles, STSs derived from YACs; open boxes, DNA marker; closed boxes, ends that have not been analyzed; dashed lines, uncertainty about the length of that region; wavy lines, sequences from a chromosome other than 17.](image)

![Fig. 3. Location of MYH genes in different YACs. Total yeast DNA was digested with EcoRI and blot-hybridized with a MYH 3′ consensus sequence probe (LL1000). Lane: a, YPH252 (no YAC); b, A213C4; c, A24D2; d, A258D4; e, A23D7; f, A288E1; g, B72A12; h, A51D4; i, A56B10; and j, B120C11.](image)
range and all other bands observed in individual YACs were observed (data not shown). These results indicate that the organization of the MYH genes in the YAC contig represents the organization of these genes in the genome.

DISCUSSION

Nine YAC clones containing six distinct human skeletal MYH genes on chromosome 17 have been isolated and characterized. The physical map encompasses 650 kb and has been defined with 13 STSs. Two of the YACs, A288E1 and B120C11, which together comprise 500 kb of DNA, contain all of the MYH genes known to be located in this region.

MYH genes have been mapped to 17p by a number of different methods. Of the genes mapped, MYH2 is mapped with the greatest precision to 17p13.1 (15). Because this gene is contained in the YAC contig, it is possible to assign all six MYH genes to 17p13.1.

In the current study, the six MYH genes that have been found in the 650-kb YAC contig include embryonic, perinatal, three adult skeletal fast forms (either IIa, IIb, or IIId), and an unidentified gene. Four have been identified by comparison with cloned cDNAs. Two YAC fragments that hybridize to a consensus MYH probe are evident at the rightward end of the locus (YACs A56B10 and B120C11). One of these additional bands (2.1-kb EcoRI) is likely to correspond to an adult skeletal MYH locus from which a restriction fragment length polymorphism (p10-5) has been described and used in linkage analysis (3, 15). Its precise identity has not yet been determined and will require DNA sequence and RNA hybridization analysis, but preliminary data suggest it is an adult fast MYH form. The sixth MYH gene in the contig does not correspond to any of the available human genes or cDNA sequences and is a likely candidate for the extraocular MYH gene.

Several examples of related gene clusters include globin, proline-rich salivary proteins, keratin genes, homeobox, and the major histocompatibility complex. In the globin gene complex and the homeobox clusters, the organization of the genes reflects the developmental pattern of expression (37, 38). It has been suggested that structural organization within a cluster might reflect the functional aspects of individual members. The organization of the skeletal MYH genes does not follow their developmental expression order. For example, the embryonic and perinatal MYH genes are expressed sequentially during development but are not located adjacent to each other. The functional significance of the organization of MYH genes, if any, has yet to be elucidated.

Understanding the organization of the MYH genes may permit us to initiate studies of pathological aspects associated with defects of these genes. Examination of the organization of the MYH genes in patients who have dystrophic muscle may indicate that, in some of these patients, abnormalities in the individual members of the MYH gene family cause the pathology.

The manuscript was prepared by V. Gradus. We also acknowledge the contribution of somatic cell hybrid DNA from R. E. K. Fournier (Fred Hutchinson Cancer Center, Seattle) and the contribution of Ilene Karsch-Mizrachi in initiating the analysis of MYH sequences in the YAC clones. This work is supported by grants from the National Institutes of Health (GM33943 and HG00380 to R.K. and GM2990 to L.L.), a Wills Foundation grant to R.K., and a Cancer Center grant to the Albert Einstein College of Medicine (CA13330). L.L. is the recipient of an Irma T. Hirschl Award. The YAC libraries were kindly screened by Dr. B. Brownstein (Washington University) and by the Genome Center of the Baylor College of Medicine (Houston).